

2008

Computer modeling of dapson-mediated heteroactivation of flurbiprofen metabolism by CYP2C9

Robyn Renee Ayscue
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Ayscue, Robyn Renee, "Computer modeling of dapson-mediated heteroactivation of flurbiprofen metabolism by CYP2C9" (2008). *Graduate Theses, Dissertations, and Problem Reports*. 2698.
<https://researchrepository.wvu.edu/etd/2698>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**COMPUTER MODELING OF DAPSONE-MEDIATED HETEROACTIVATION
OF FLURBIPROFEN METABOLISM BY CYP2C9.**

by

Robyn Renee Ayscue

A dissertation submitted to the Faculty of the School of Pharmacy
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Pharmaceutical Sciences

Approved by

Dr. Peter M. Gannett, Adviser
Dr. Timothy S. Tracy
Dr. Patrick Callery
Dr. Yongut Rojanasakul
Dr. Björn Söderberg

Department of Basic Pharmaceutical Sciences

Morgantown, West Virginia
2008

ABSTRACT

The occurrence of atypical kinetics in cytochrome P450 reactions can confound *in vitro* determinations of a drug's kinetic parameters. During drug development, inaccurate kinetic parameter estimates can lead to incorrect decisions about a lead compound's potential for success. It has become widely accepted that in certain CYP subfamilies more than one molecule can occupy the active site simultaneously, in some cases resulting in enhanced substrate turnover (heteroactivation). However, the specific mechanism(s) by which dual-compound binding results in heteroactivation remain unclear. It is known that orientation of the substrate in the active site, as dictated by interactions with active site residues, plays a large role in metabolic outcome. Effector compounds have been shown *in vitro* to alter substrate position in the active site. Here, data obtained via *in silico* methods including docking, molecular dynamics, semi-empirical and *ab initio* quantum mechanics indicate that direct interaction between effector and substrate can play a role in stabilizing the substrate in an alternative conformation conducive to oxidation. In this study a high-throughput screening computer model of heteroactivation of flurbiprofen metabolism by CYP2C9 has been developed for the purpose of elucidating key interactions between substrate, effector, and enzyme responsible for heteroactivation in this system, as well as to predict as yet unknown activators.

ACKNOWLEDGEMENTS

The author wishes to thank her advisor, Dr. Peter Gannett, for his patience, reassurance, encouragement and guidance in matters both scientific and non-scientific. The author would also like to thank Dr. Tim Tracy and his lab for their invaluable contribution of the *in vitro* component of this work. Also she wishes to extend thanks to those who have agreed to serve on her committee, including Dr. Tim Tracy, Dr. Patrick Callery, Dr. Yongut Rojanasakul, Dr. Björn Söderberg and Dr. Robert Haining. Lastly, the author wishes to offer sincere gratitude to family and friends, without whose love and support this work would not be possible.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES.....	vi
LIST OF TABLES	viii
CHAPTER I. Background	1
1.1. Overview of P450 Structure and Function.....	2
<u>1.1.1. Introduction</u>	<u>2</u>
<u>1.1.2. General structural characteristics of CYPs</u>	<u>7</u>
<u>1.1.3. Overview of behavior and structural characteristics of CYP2C9.....</u>	<u>7</u>
1.2. Kinetics of CYP2C9	12
1.3. Predictive Computational Models of Heteroactivation in CYP2C9.....	16
1.4. Research Objectives.....	18
<u>1.4.1. Overview</u>	<u>18</u>
<u>1.4.2. Investigate Effector-Substrate-Enzyme Interactions by Docking</u>	<u>19</u>
<u>1.4.3. Investigate Effector-Substrate-Enzyme Interactions by Molecular Dynamics.....</u>	<u>20</u>
<u>1.4.4. Statistical Model Development and External Database Screening..</u>	<u>21</u>
CHAPTER II. Literature Review	23
2.1. Introduction to <i>In Silico</i> Modeling of Atypical Kinetics in CYP2C9.....	24
2.2. Ligand-Based Models.....	25
2.3. Ligand-Receptor-Based Models.....	44
2.4. Receptor-Based Models.....	48
CHAPTER III. Molecular Dynamics Simulations of the CYP2C9-Flurbiprofen Complex With and Without Dapsone Present.....	55
3.1. Introduction.....	56
3.2. Materials and Methods.....	61
<u>3.2.1. Preliminary Structure Preparation.....</u>	<u>61</u>
<u>3.2.2. Molecular Mechanics/Molecular Dynamics Simulation.....</u>	<u>62</u>
<u>3.2.3. Analysis of MD Simulation.....</u>	<u>63</u>
3.3. Results and Discussion	63
CHAPTER IV. Determination of Effector-Active Site Binding in CYP2C9 with AutoDock3.0.....	73
4.1. Introduction.....	74
4.2. Materials and Methods.....	80
<u>4.2.1. Preliminary Structure Preparation and MD Simulation</u>	<u>80</u>
<u>4.2.2. AutoDock Screening.....</u>	<u>82</u>
<u>4.2.3. In-Silico-In-Vitro/CoMFA-predicted Data Correlation Evaluation.....</u>	<u>83</u>
<u>4.2.4. Quantum Mechanical Analysis of Docked-Effector-Substrate Interaction Energy.....</u>	<u>84</u>
<u>4.2.5. Visual Analysis</u>	<u>84</u>
4.3. Results and Discussion	85
CHAPTER IV. Determination of Effector-Active Site Binding in CYP2C9 with eHiTS	103

5.1. Introduction.....	104
5.2. Materials and Methods.....	111
5.2.1. Preliminary Structure Preparation	111
5.2.2. Receptor Training.....	111
5.2.3. Docking Protocol.....	112
5.2.4. Visualization and Analysis of Results.....	113
5.2.5. Data Fitting.....	113
5.3. Results and Discussion	113
CHAPTER VI. Development of a Statistical Model for Effectors of Flurbiprofen	
Metabolism in CYP2C9.....	143
6.1. Introduction.....	144
6.2. Materials and Methods.....	146
6.2.1. Cross-Validation of the Model.....	146
6.2.2. NCI Database Screening.....	147
6.3. Results and Discussion	148
CHAPTER VII. Summary and Conclusions.....	160
REFERENCES.....	164

LIST OF FIGURES

Figure 1.1. CYP Catalytic Reaction Cycle.....	3
Figure 1.2. Type I and Type II Binding Spectra.....	4
Figure 1.3. Heme Geometries of High-Spin and Low-Spin States.....	6
Figure 1.4. Tertiary Structure of CYP2C9 Highlighting Conserved Core Region.....	8
Figure 1.5. Enzyme Kinetic Profiles.....	14
Figure 2.1. Pharmacophore Model 1.....	27
Figure 2.2. Pharmacophore Model 2.....	28
Figure 2.3. Pharmacophore Model 3.....	29
Figure 2.4. Chemical Structure of Benzbromarone.....	35
Figure 2.5. Chemical Structures of Benzbromarone Analogs.....	37
Figure 2.6. Arg-108 Conformation in 1OG5 Crystal Structure of CYP2C9.....	40
Figure 2.7. Pi-Stacking Geometries.....	41
Figure 2.6. Arg-108 Conformation in 1R9O Crystal Structure of CYP2C9.....	42
Figure 3.1 Overlaid Distance Plots of Flurbiprofen 4'-H to the Heme Iron In the Presence of Dapsone , Dapsone to F476, and Dapsone to Flurbiprofen.....	64
Figure 3.2. H-bonds Formed During MD Between Dapsone and Active Site Residues of CYP2C9 in the Presence of Flurbiprofen	65
Figure 3.3. Pi-Stacking Interactions Between Dapsone, Phe-476 And Flurbiprofen in the Active Site of CYP2C9 During MD.....	67
Figure 3.4. Pi-Interactions Formed by Dapsone, Phe-476 and Flurbiprofen When the Substrate 4'-H-Fe Distance Is at Its Shortest During MD	69
Figure 3.5. Distance Plot of Flurbiprofen 4'-H to the Heme Iron in the Absence of Dapsone.	71
Figure 4.1. Initial Docking of Dapsone into CYP2C9 Empty Active Site by AutoDock.....	86
Figure 4.2. Initial Docking of Flurbiprofen into CYP2C9 Empty Active Site by AutoDock.....	87
Figure 4.3. Initial Docking of compound 19 into CYP2C9 Empty Active Site by AutoDock.....	88
Figure 4.4. Initial Docking of Compound 15 into CYP2C9 Empty Active Site by AutoDock.....	89
Figure 4.5. Initial Docking of Dapsone in CYP2C9 in the Presence of Flurbiprofen by AutoDock.....	91
Figure 4.6. Initial Docking of Compound 15 in CYP2C9 in the Presence of Flurbiprofen by AutoDock.....	92
Figure 4.7. Initial Docking of Compound 19 into CYP2C9 Active Site in the Presence of Flurbiprofen by AutoDock.....	93
Figure 4.8. Dapsone Docked into the Empty Active Site of CYP2C9 by AutoDock Using Recalculated van der Waals and Lennard-Jones Parameters.....	94
Figure 4.9. Flurbiprofen Docked into the Empty Active Site of CYP2C9 by AutoDock Using Recalculated van der Waals and Lennard-Jones Parameters.....	96
Figure 4.10. Flurbiprofen Docked into the Empty Active Site of CYP2C9 by AutoDock Using Re-assigned Heme Partial Charges.....	98

Figure 4.11. Dapsone Docked into the Empty Active Site of CYP2C9 by AutoDock Using Re-assigned Heme Partial Charges.....	99
Figure 4.12. Graphic Representation of <i>In Vitro</i> Flurbiprofen Metabolism by CYP2C9 as a Function of Substrate-Effector Complex Stability.....	102
Figure 5.1. Four Atomic Relationship Variables Illustrated.....	108
Figure 5.2. Hydrogen Bonding of Dapsone Docked Into The Untrained Receptor...	115
Figure 5.3. Pi- Interactions Between Flurbiprofen And Dapsone Docked In The Untrained Receptor.....	116
5.4. Pi- Interactions Between Aromatic Active Site Residues Of CYP2C9 And Dapsone In The Untrained Receptor.....	117
Figure 5.5. Overlay Of Dapsone Analogs Docked In The Alternative Pose In The Untrained Receptor.....	118
Figure 5.6. Overlay Of All Dapsone Analogs Docked In The Untrained Receptor..	121
Figure 5.7. Pi- Interactions Between Flurbiprofen And Dapsone Docked In The Manually Trained Receptor.....	123
Figure 5.8. Pi- Interactions Between Aromatic Active Site Residues Of CYP2C9 And Dapsone In The Manually Trained Receptor.....	124
Figure 5.9. Hydrogen Bonding of Dapsone Docked Into The Manually Trained Receptor.	125
Figure 5.10. Overlay Of Dapsone Analogs Docked In The Alternative Pose In The Manually Trained Receptor.....	127
Figure 5.11. Overlay Of All Dapsone Analogs Docked In The Manually Trained Receptor.	128
Figure 5.12. Pi- Interactions Between Aromatic Active Site Residues Of CYP2C9 And Dapsone In The Enrichment Trained Receptor.....	129
Figure 5.13. Pi- Interactions Between Flurbiprofen And Dapsone Docked In The Enrichment Trained Receptor.....	132
Figure 5.14. Optimized Orientation Of A Clustered Benzene Trimer.....	133
Figure 5.15. Overlay Of Dapsone Analogs Docked In The Alternative Pose In The Enrichment Trained Receptor.....	134
Figure 5.16. Overlay Of All Dapsone Analogs Docked In The Enrichment Trained Receptor.	135
Figure 5.17. eHiTS Activator Scores And Correlation Coefficients After Enrichment Training Of The Receptor.....	137
Figure 6.1. NCI Compounds in the Top 20 Identified by eHiTS As Activators of Flurbiprofen Metabolism by CYP2C9.....	150
Figure 6.2. NCI Compounds Identified by eHiTS as Activators Docked in the Active Site of CYP2C9 in the Presence of Flurbiprofen.	151
Figure 6.3. Pi-Interactions Formed by Compound 21 Docked in the Active Site of CYP2C9 in the Presence of Flurbiprofen.....	153
Figure 6.4. Docked NCI Compounds Identified by eHiTS as Activators of Flurbiprofen Metabolism by CYP2C9 Possessing Three Or More Rings.....	154
Figure 6.5. Chemical Structure of 3-methyl-1-phenyl-2-pyrazolin-5-one.	158

LIST OF TABLES

Table 4.1. Structures of Dapsone and Dapsone Analog Effectors.....	81
Table 5.1. The 32 Surface Point Types Recognized by eHiTS.....	107
Table 6.1. Cross-Validation Results.	149
Table 6.2. Activators of Flurbiprofen Metabolism by CYP2C9 Identified in the Seeded NCI Database by eHiTS.....	155

CHAPTER I

Background

1.1. Overview of P450 Structure and Function

1.1.1. Introduction

Cytochrome P450 enzymes are responsible for the metabolism of about 90% of xenobiotics¹, including pollutants, toxins, and therapeutic compounds, as well as certain endogenous compounds such as steroids². These enzymes are membrane-bound, and are principally found in the endoplasmic reticulum of hepatocytes, though they can also be found in considerable quantities in cells of other tissues, including the lungs, brain, and small intestine. Nearly all P450 enzymes are monooxygenases-- they transfer a single oxygen to the substrate compound using the reaction scheme described below (**Figure 1.1**), however the CYP superfamily can be divided into various families and subfamilies whose metabolic function is unique to specific physicochemical classes of substrates.

In the presence of substrate three types of binding mechanisms are known to occur, depending on the chemical composition of the substrate as well as the CYP isozyme involved in catalysis. These are Type I, Type II, and Reverse Type I. These mechanisms are defined by their respective UV spectra (**Figure 1.2**).

In a typical Type I interaction substrate enters the active site and in doing so displaces the bound water molecule, destabilizing the low-spin state and causing the low-spin-high-spin equilibrium to shift toward the high-spin state. With its 3d electrons now

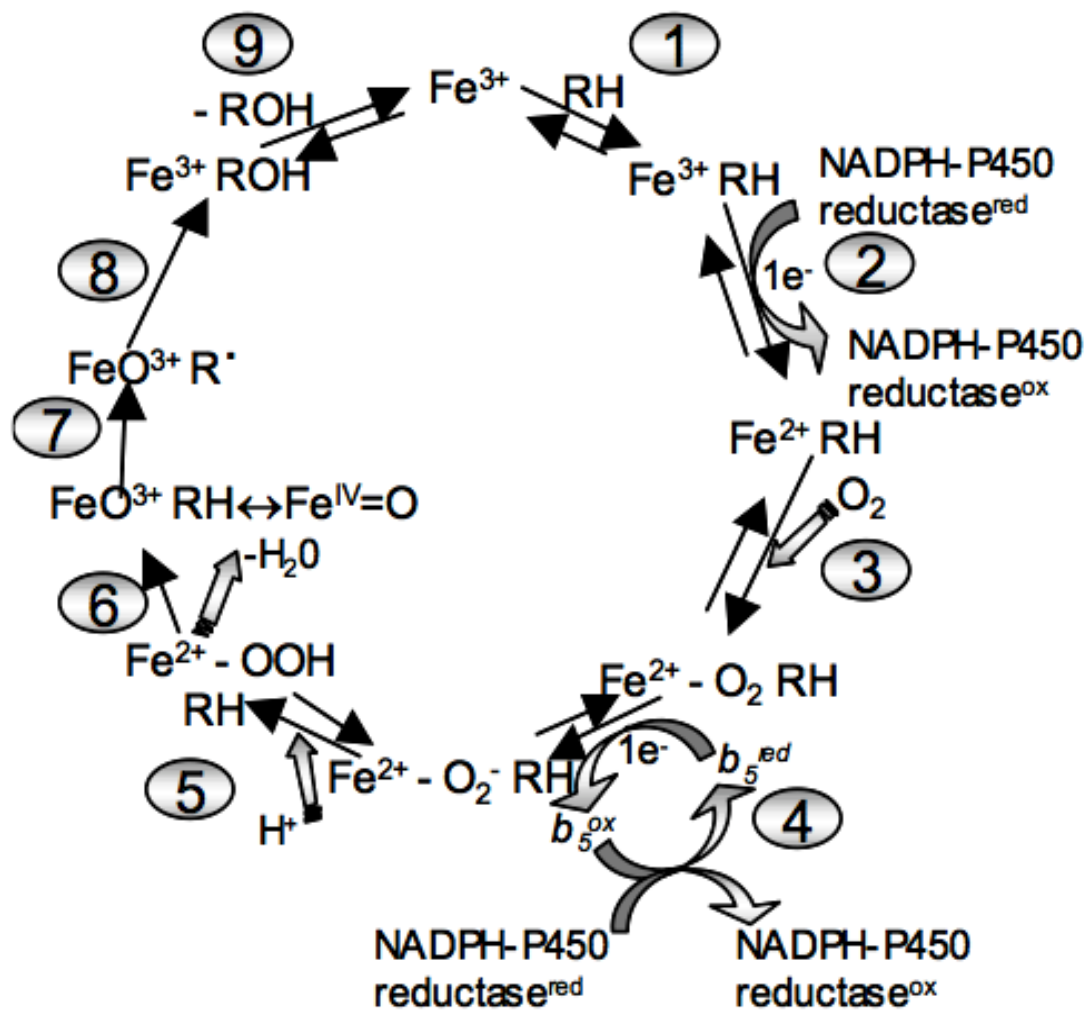


Figure 1.1. CYP Catalytic Reaction Cycle.

The CYP catalytic reaction cycle is illustrated. Adapted from Guengerich³.

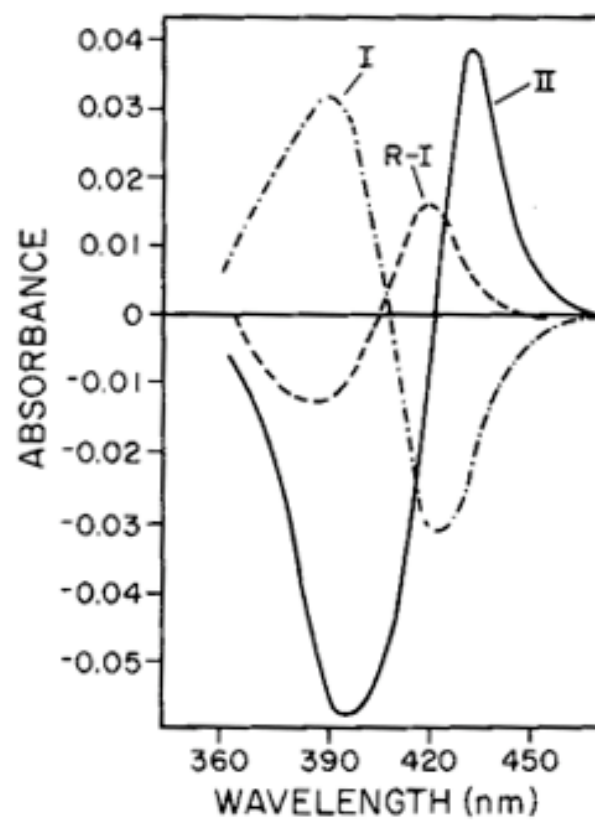
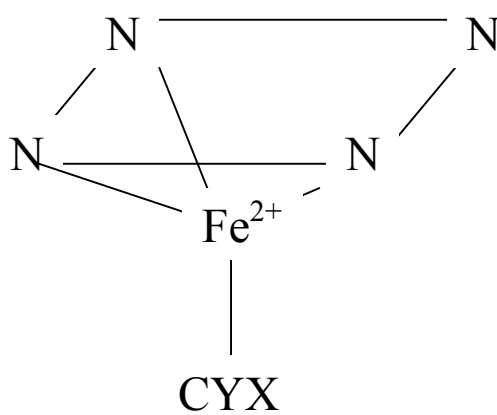


Figure 1.2. Type I and Type II Binding Spectra.

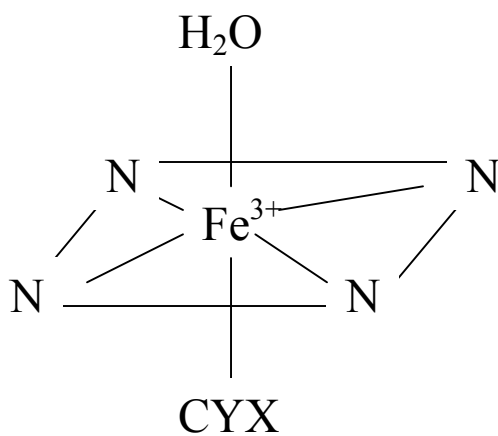
Type I and Type II binding spectra are shown. From Jefcoate et al⁴.

unpaired and redistributed, the ionic radius of the iron increases, rendering it too large to fit inside the porphyrin ring without distorting its geometry. Therefore the heme group "flutes" such that the iron atom pulls proximally out of the porphyrin plane (**Figure 1.3**). It is at this point that the substrate binds to the heme iron. It is generally believed that substrate binding elicits a conformational change in the enzyme that increases the binding affinity of electron-donating reductase. NADPH-cytochrome P450 reductase (CPR) is the obligatory redox transfer protein for providing the first electron of this two-electron process. However, the second electron may be provided by either CPR or cytochrome *b₅*. Electron transfer from the reductase to the heme iron changes its oxidation state from +3 to +2. Diatomic oxygen that diffuses into the active site can now bind to the ferrous iron. In the step that follows, CPR or *b₅* donates a second redox equivalent to the heme, and so creates a superoxide, which subsequently accepts a proton from either the side-chain hydroxyl group of a conserved threonine, or from a water molecule bound to that residue, heterolytically cleaving the dioxygen bond. The resulting structure collapses into excess water and what is speculated to be a highly reactive oxyferryl species in which iron exists in the +4 oxidation state. In an as-yet speculative reaction likely involving hydrogen extraction³, this oxygen is transferred to the substrate site of metabolism, simultaneously returning the heme iron to its resting 3+ state and releasing oxidized metabolite from the catalytic site.

In Type II binding the inhibitor displaces water as the sixth ligand and stabilizes iron in a ferric state, which



A



B

Figure 1.3. Heme Geometries of High-Spin and Low-Spin States.

Spin states Type I (A) and Type II (B) are illustrated. CYX is the cysteine residue connecting the heme iron to the apoprotein.

is not conducive to metabolism. Suicide inhibition does not produce a metabolite and renders the enzyme permanently inactivated. However, in certain cases such as diclofenac metabolism by CYP2C9, exhibit a Type II UV spectrum *in vitro*, yet are in fact metabolized and do not adversely affect the enzyme's metabolic ability⁵. The mechanism(s) responsible for this phenomenon is/are as yet unknown. Reverse Type I binding occurs with ligands possessing functional groups that coordinate weakly to the heme, and is thought to occur via a mechanism involving stabilizing the apoprotein conformation such that the ferric heme geometry is favored. It is thought to result at times from allosteric substrate binding⁴.

1.1.2. General structural characteristics of CYPs

The overall secondary structure is conserved among all human P450s, and consists of a "core region" of helices D, E, I, J, K, L and the heme, which is sandwiched between helices L and I (**Figure 1.4**). A cysteine residue links the heme moiety with the apoprotein by coordinating with the heme iron. Also highly conserved is a threonine (Thr-301) located in the I helix in the interior of the protein, thought to be involved in the transfer of protons in oxidative reactions with substrates. The iron-ligated cysteine is flanked by several residues in the sequence RXCXG, which is highly conserved also.

1.1.3. Overview of behavior and structural characteristics of CYP2C9

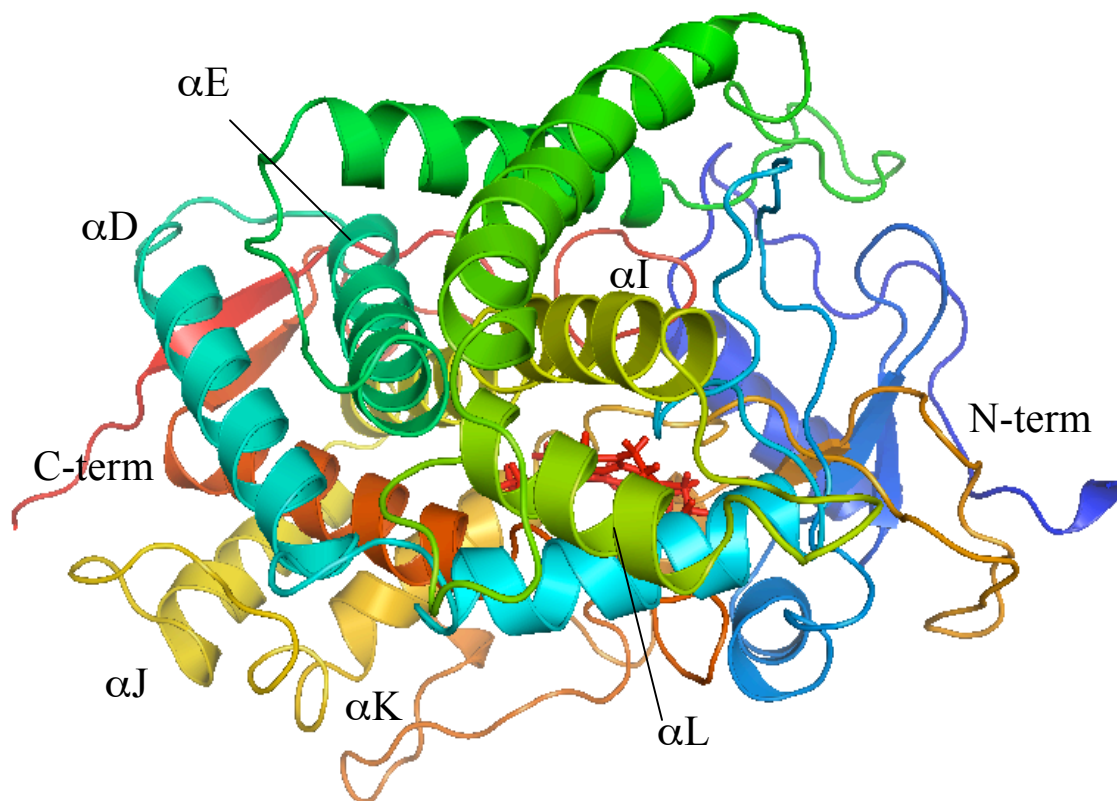


Figure 1.4. Tertiary Structure of CYP2C9 Highlighting Conserved Core Region.

The tertiary structure of CYP2C9 is illustrated and general secondary structures conserved among human CYPs are labeled. The heme is shown in stick-form colored crimson.

Of all CYPs involved in drug metabolism, CYP2C9 is one isozyme deserving closer study. CYP2C9-mediated oxidation occurs with nearly one-fifth of all currently marketed drugs ⁶, making mechanistic understanding of the enzyme's metabolic process of prime importance to drug development and design.

Substrates of CYP2C9 tend to be small, aromatic, weakly acidic compounds, such as phenytoin, ibuprofen, naproxen, tolbutamide, fluoxetine, and warfarin. The acidic group is typically about 7 Ångstroms from the site of oxidation ⁷. Occasionally, however, CYP2C9 metabolizes compounds that are neutral, such as benzo[a]pyrene, and some aliphatic compounds ⁸.

2C9 exhibits polymorphism in some individuals, with different allelic variants occurring in different ethnic populations. These differ from the wild-type by a single residue; the most common being R144C (*2), I359L (*3), and D360E (*5). All are slow metabolizers compared to the wild-type. The *2 variant occurs in about 20% of Caucasians, 9% in people of African descent, and less than 1% in Asians. ⁹. CYP2C9*3 is present in about 12% of Caucasians, < 2% in people of African descent, and 4% of Asians. The *5 phenotype is found only in people of African descent, and is very rare, occurring in less than 1% of the population ¹⁰. Due to the metabolic inefficiency of these polymorphic variants, drugs with narrow therapeutic indices will require an adjustment in dosage for patients with a non-wild-type CYP2C9 phenotype. Otherwise, toxicity or lack of therapeutic effect may result, depending on the drug being administered.

Structurally, CYP2C9 shares over 80% amino acid sequence identity to other 2C enzymes ¹¹. However, important residues form structural elements that create an active site topology unique to CYP2C9. Publication of the x-ray crystal structure (Protein

Databank Code 1R9O) by Wester in 2004 elucidated the roles of several key residues. it was observed that conserved residue Arg-108, which has been shown to be important in the oxidative functionality of the enzyme toward acidic substrates flurbiprofen, *S*-warfarin, and diclofenac ⁵, forms hydrogen bonds with I-helix residues Asp-293 and Asn-289. Arg-108 lies in a thirteen-residue segment in the BC-loop sandwiched between two glycine residues whose proton side chains render the region highly flexible. Residues 293 and 289 may aid in stabilizing Arg-108 such that its amino side chain faces the active site interior ¹². Asp-293 may also form hydrogen bonds with the backbone of a segment of the BC-loop composed of residues 111 through 114. In this segment Val-113 and Phe-114 are known to play an important role in *S*-warfarin binding to CYP2C9 based on mutagenesis studies ¹³. Supporting this theory is the fact that Asn-293 plays an additional role in maintaining overall tertiary structure ⁵. Residue 289 is an isoleucine in CYP2C19, which lacks the ability to metabolize the CYP2C9 substrate diclofenac, yet substitution of asparagine in place of Ile-289 confers that ability to CYP2C19 ¹⁴, as well as enables the enzyme to bind tightly to CYP2C9-selective inhibitor sulphaphenazole ¹⁵, which further cements Asn-289 as a residue of importance in CYP2C9.

Another highly-conserved arginine, Arg-97, lies within the same flexible helix region that Arg-108 does, and is in close proximity to the heme propionate groups. In one study it was found that mutation to alanine increases the K_m of diclofenac metabolism 13-fold but does not alter V_{max} ¹⁶. Another study involving ARG-97A mutagenesis concluded that Arg-97 plays a primarily structural role in retaining the heme moiety within the tertiary structure of the enzyme after observing a marked decrease in holoenzyme expression of ARG-97 mutants ¹⁷.

Despite the presence of several charged residues, the active site retains an overall hydrophobic character. Residues shown to interact with CYP2C9 substrates warfarin and flurbiprofen are Val-113, Phe-114, Leu-208, Ile-205, Leu-366, Phe-100, Phe-476^{12, 18, 19}. It is likely that given the tendency of CYP2C9 substrates to possess a high degree of aromaticity, which tends to confer hydrophobicity, at least some of these residues likely play a role in the substrate binding.

Water molecules in the active site are known to play a critical role in the P450 cycle³. Wester et al. identified three water molecules present in the active site of their x-ray crystal of CYP2C9 and bound substrate flurbiprofen; one of which is positioned near the C-terminus. Another crystallized water occupies a site near Thr-301 along the I-helix. The third sits between flurbiprofen and the heme iron at a distance of 2.4 Å away from the iron and 3.0 Å from the 4' hydrogen of flurbiprofen (the site of oxidation). This is consistent with experimental evidence in that flurbiprofen induces only a partial spin-state change in the heme iron upon binding, indicating that it does not entirely displace the water molecule originally bound there¹². Locuson and Tracy identified seven conserved protein-water binding sites within or near the active sites of several mammalian P450 isozymes including CYP2C9²⁰. Of the three active site waters identified by Wester only one, located proximal to the I-helix, was determined by Locuson and Tracy to occupy a high-density water binding site. The remaining non-heme-bound water in the crystal may be an artifact of the crystallization process, or it may play a role in sterically preventing the escape of the substrate once bound inside the active site, or play a role in the proton shuttle.

CYP2C9 is responsible for several known clinically relevant adverse drug-drug

interactions. Some examples are diclofenac 4'-hydroxylation inhibition by fluvastatin²¹; fluconazole inhibition of losartan metabolism²²; inhibition of phenytoin metabolism by isoniazid²³; and probenecid inhibition of naproxen demethylation²⁴, to name a few. These interactions can have especially severe repercussions for the patient when drugs with narrow therapeutic windows are involved. For example, warfarin, a potent blood thinner, is potentiated by fenofibrate²⁵, fluconazole²⁶; and amiodarone²⁷ due to their ability to inhibit warfarin hydroxylation by CYP2C9.

1.2. Kinetics of CYP2C9

Kinetic profiles are used to determine the intrinsic clearance of a drug, and hence the appropriate dosage *in vivo*. Due to the fact that the metabolism of a large number of xenobiotics falls mainly to only three subfamilies of CYPs (CYP3A4, 2D6, and 2C9), accurate estimation of intrinsic clearance is imperative. Miscalculation can put the patient at an increased risk of potentially adverse drug-drug interactions. Adding further complication is the fact that the effect one compound has on the metabolism of another is substrate-specific, and a compound that may be an activator for one substrate may in fact inhibit another. Clinically relevant problems arise when *in vitro-in vivo* correlative assays assume interactions occur between only one compound-- the substrate-- and the enzyme at any given time. A hyperbolic kinetic profile obtained for an enzyme-substrate system *in vitro* can become atypical in clinical trials when the participants simultaneously ingest other drugs to treat other maladies. This unforeseen change in drug kinetic parameters can result in adverse drug interactions that may persist even after cessation of treatment, particularly involving drugs with long half-lives.

Under most circumstances CYP2C9, along with other CYP isozymes, displays classical Michaelis-Menten kinetics (**equation 1.1**),

$$V_{\max} = \frac{K_m [S]}{K_m + [S]} \quad (1.1.)$$

where V_{\max} is the maximum rate of substrate turnover, K_m is the binding affinity constant, and $[S]$ is the substrate concentration. Enzyme-substrate systems displaying classical kinetics will yield a hyperbolic curve when velocity is plotted as a function of substrate concentration (**Figure 1.5**) Certain conditions, such as high substrate or salt concentrations, may result in CYP2C9 displaying atypical kinetics^{28,29}, however, a common cause of non-hyperbolic kinetics is allosterism, in which two compounds, either two molecules of substrate or of a substrate and an effector molecule, simultaneously bind to the enzyme. This allosteric phenomenon can lead to inhibition of substrate metabolism, biphasic kinetics, or activation of metabolism. Allosterism can not only cause drugs with a Michaelis-Menten profile to become kinetically atypical, but can also

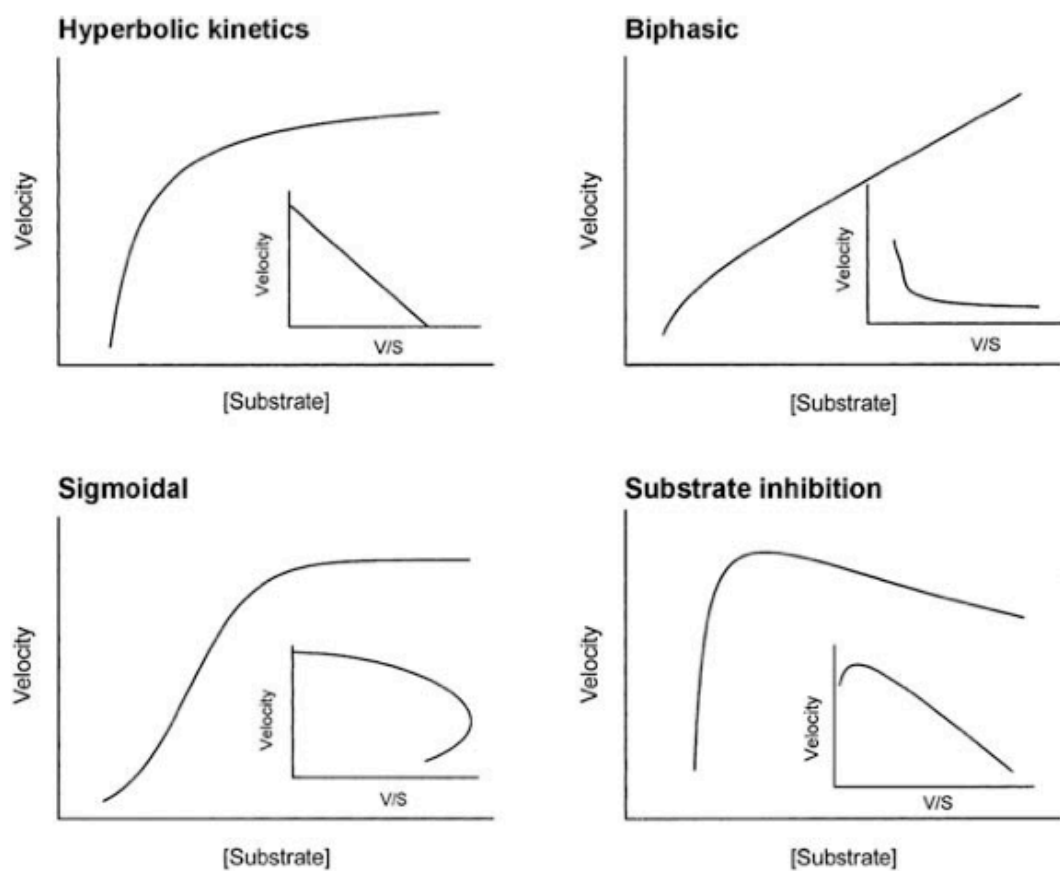


Figure 1.5. Enzyme Kinetic Profiles.

Enzyme kinetic profiles are illustrated by velocity curves and Eadie-Hoffstee plots (insets). From Atkins³⁰.

result in causing drugs known to possess an atypical kinetic profile to exhibit Michaelis-Menten kinetics.

One specific type of atypical kinetics is heterotropic positive cooperativity, also called heteroactivation, which is a form of allosterism in which a substrate and an effector compound bind simultaneously to the enzyme, resulting in increased substrate turnover. The effector compound itself is rarely metabolized, or forms very little metabolite. A model example of this phenomenon is dapsone activation of flurbiprofen metabolism by CYP2C9. Dapsone is used in the treatment of leprosy and in certain types of pneumonia peculiar to AIDS patients. Flurbiprofen is a non-steroidal anti-inflammatory drug (NSAID). Both are, individually, substrates of CYP2C9, forming *N*-hydroxydapsone and 4'-hydroxyflurbiprofen as metabolites, respectively. However, in the presence of dapsone, the rate of flurbiprofen 4'-hydroxylation increases by roughly 64% compared to flurbiprofen metabolism in the absence of dapsone, while the K_m decreases. Flurbiprofen has no effect on dapsone metabolism³¹⁻³³.

Polymorphism can determine the extent to which heteroactivation-related drug interactions will occur, or even if they will occur or not. Certain CYP2C9 variants also exhibit heteroactivation of flurbiprofen activation in the presence of dapsone. In the wild-type, dapsone activates flurbiprofen oxidation by about 64% compared to the rate of metabolism of flurbiprofen in the absence of dapsone. However, in CYP2C9*2 and *5 that figure increases to nearly 400%. CYP2C9*3 shows a nearly 1000% increase. Due to the poor oxidative ability of these three variants, the presence of dapsone may increase their ability to metabolize flurbiprofen at a rate more similar to that of the wild-type. [J.S. Aguilar, unpublished results]. In an other example, benzbromarone inhibits

flurbiprofen 4'-hydroxylation in wild-type CYP2C9, but activates flurbiprofen oxidation in the *3 variant ³⁴.

As with many forms of atypical kinetics, the mechanism(s) of heteroactivation is not well- understood, and accurate prediction of effector potential based on physicochemical attributes of compounds is not generally possible. Theories regarding the mechanism(s) of heteroactivation vary. Evidence of a simultaneous presence of two compounds in the active site of CYP3A4 and BM3 had been in existence since the mid 1990's ^{33, 35-37}, and it was proposed that this could be the case in CYP2C9 as well. Recent NMR studies have indeed shown dapsone and flurbiprofen bound simultaneously within the active site of CYP2C9. In addition, the distance between 4'-hydrogen of flurbiprofen and the heme iron decreased by 0.90 Å in the presence of dapsone compared to flurbiprofen alone ³⁸. By building on this and other previously published data our goal is to better understand the mechanism(s) behind the phenomenon of heteroactivation and develop a predictive computational model capable of identifying as yet unknown effector compounds.

1.3. Predictive Computational Models of Heteroactivation in CYP2C9

Atypical kinetics in CYP2C9 is difficult to anticipate due to the fact that no link between compound activity and physicochemical properties has yet been discovered ³⁹. Heteroactivation is one of the more difficult forms of atypical kinetics to predict. High-throughput *in vitro* assays designed to screen for compounds that cause non-classical kinetic profiles in CYPs are difficult to develop for heteroactivation, as it is practically impossible to screen every lead compound through an assay with every other therapeutic

already available. Often a lead compound has made its way through the drug design and development pipeline to the clinical phase before its effector potential is discovered, possibly resulting in its withdrawal from development and a wasted investment of both time and money of the pharmaceutical company.

Computational quantitative structure-activity relationship (QSAR) models are highly sophisticated and are useful for identifying key active site interactions responsible for metabolic activity that may elude visual analysis. Once these interactions are ascertained, a virtual high-throughput screening (vHTS or simply HTS) model can be developed to mine large databases of drug-like compounds for possible effectors. Thus, effector compounds can be identified at the beginning of lead compound design, saving time and money.

Prior to this work a comparative molecular field analysis (CoMFA) model had been developed by aligning dapsone and a series of dapsone analogs whose *in vitro* effector behavior was known. The alignment best correlating with *in vitro* activation of flurbiprofen oxidation was on the sulfone group. This model yielded a high correlation coefficient of determination ($R^2 = 0.80$), and descriptors important in the alignment were identified as sterics, x, y, and z components of the dipole and the total dipole of the system, and electrostatics [J.S. Aguilar, unpublished results]. While the correlation was satisfactory, the CoMFA technique does not lend itself to HTS.

Only one successful HTS model specific to CYP2C9 heteroactivation has been developed to date. Using the program Glide^{40,41}, a database of around 70,000 drug-like compounds was screened using a receptor structure consisting of CYP2C9 with substrate flurbiprofen docked in the active site. The top three conformers of the screened

compounds that scored below a certain energy threshold, and that had bound to a non-substrate binding site inside the enzyme were tested *in vitro* for activity on flurbiprofen metabolism by CYP2C9. The model was able to identify three weak activators of flurbiprofen metabolism by CYP2C9, as well as a then-unknown effector compound similar to dapsone in activation capacity, yet very structurally dissimilar to dapsone⁴². However, though the screening itself was fully automated, the process of identifying the effector compound using the binding modes of the flagged compounds was very user-intensive. While the model is good, the amount of human involvement in the process may render the methodology somewhat inefficient, particularly for large databases with hundreds of thousands of compounds.

The model can be improved to the point that the human element in screening can be lessened to a manageable amount by understanding the mechanism(s) of heteroactivation and weighting key parameters in the screening program accordingly. Computational screening models have been developed for 3A4 heteroactivation with relative success⁴³, suggesting that it is possible for CYP2C9.

1.4. Research Objectives

1.4.1. Overview

Computer modeling has been successfully applied to problems of lead compound generation and enzyme-drug interaction prediction⁴⁴. The same methods are applicable to the development of models for the use of predicting enzyme-substrate-effector combinations. Here, we have used data from kinetic, structural, mutational and active site

interactions, to develop a model that can predict the potential for a compound to interact cooperatively with an enzyme-substrate complex. In pursuit of this goal three aspects of modeling were explored and integrated: effector docking to identify likely effector-substrate-enzyme interactions, molecular dynamics to observe these interactions over a period of time, and high-throughput screening (HTS) of a large, external database following development of a statistical model using data obtained via the two techniques mentioned above, which involves the application of statistical methods to correctly weight certain interactions (i.e. “descriptors”) over others for the purpose of increasing predictive power and validate the resulting model. Upon completion of the work the expected outcome is a computer model capable of screening large drug databases and identifying as-yet unknown activators of flurbiprofen metabolism by CYP2C9.

1.4.2. Investigating Effector-Substrate-Enzyme Interactions by Docking

Docking is a fast and computationally cost-effective method of identifying binding modes for known substrates and inhibitors. By observing effector-substrate-enzyme interactions present in the docking results, docking a series of substrates into a CYP2C9-flurbiprofen receptor complex may reveal which descriptors are the most important for biological activity. Such information can later aid in directing the training of the receptor to weight these descriptors over others.

Docking has been used successfully in the development of several statistical models of CYP2C9 metabolism^{42, 43, 45-51}; its popularity due in large part to its low computational cost and its speed when screening large databases. However, a major drawback of docking is the fact that speed in computation requires some sacrifice of accuracy. For

example, in programs capable of high-volume screening the receptor must be held rigid in order to achieve cost-effectiveness, which can impair model development when the ligand requires a conformational change in the enzyme to bind correctly. This issue may be circumvented somewhat by using the most representative structure of the receptor (obtained from molecular dynamics simulations performed with the ligand bound in the active site), but this tactic is still no replacement for true flexibility, and the results of the docking run will inevitably reflect this shortcoming to at least some extent. Despite the drawbacks of this methodology, it was expected that binding modes determined by the docking program would be accurate to a sufficient degree that it would be possible to visually identify interactions present in the active site.

1.4.3. Investigating Effector-Substrate-Enzyme Interactions by Molecular Dynamics

Molecular dynamics (MD) simulations can be used to gain more detailed and accurate insight into receptor-ligand interactions. This method can allow full flexibility of the receptor, which may result in the discovery of active site interactions undetectable when the receptor is in a static state. Similarly, interactions identified by the docking program can be supported or refuted by MD. One of the most important advantages of using molecular dynamics is the ability to observe flexible receptor-ligand interactions. In addition, whereas docking programs output a static "snapshot" of the most likely binding mode of a ligand, in MD simulations the formation and dissolution of ligand-receptor interactions occur as a function of time. The addition of vector movement to all atoms in the system being investigated can be a major factor in determining the correct interactions when the receptor undergoes significant conformational change upon

binding.

A major disadvantage to MD, however, is that the cost in computation required to simulate the dynamic behavior of all atoms in both ligand and receptor (as well as individual water molecules in a solvated system) can be prohibitively large, and is therefore also very time-consuming. Because simulations of larger systems such as those involving whole enzymes can take up to three weeks or more, performing MD on a large number of compounds is not feasible. For these reasons dapsone alone was used as a representative of the training set and subjected to MD. NMR studies conducted prior to this work³⁸ established a relationship between the concomitant presence of dapsone in the active site of CYP2C9 and a decrease in distance between the site of oxidation of flurbiprofen to the heme iron. It was hypothesized that a MD simulation of CYP2C9 with flurbiprofen and dapsone both present in the active site would support the NMR findings, as well as illuminate specific effector-substrate-enzyme interactions responsible for them.

1.4.4. Statistical Model Development and External Database Screening

Unweighted scores for each docked effector calculated by both docking programs would undergo regression analysis to determine which docking program yields the best correlation with the *in vitro*/CoMFA-predicted biological activities. That program would then be used to generate a predictive statistical model. To this end, a correlation coefficient of determination (R^2) of > 0.50 would be deemed sufficient^{42, 43, 45-51}. For it to possess maximum predictive power, the model must be "trained" to weight important descriptors, which can be difficult and time-consuming, especially when the descriptors themselves may not actually be as important to the biological phenomenon as the binding

mode can lead the user to believe; in other words, the user must refine the model such that truly important descriptors are dominant over the "noise". Weight adjustment can be accomplished by either manually altering specific parameter files used by the docking program or by using an automated weighting function native to the program. "Signal-to-noise" ratio can be increased by one of many cross-validation techniques, such as leave-x-out algorithms or permuting training set biological activities, both of which are described in detail in Chapter VI. Weight training and cross-validation would be repeated until an R^2 of > 0.60 was achieved. Ultimately, once the statistical requirements have been met, the model can be validated by screening an external database, such as the National Cancer Institute (NCI) database of drug-like compounds. When screened, compounds identified by the docking program as probable activators of flurbiprofen metabolism by CYP2C9 would then be tested for their biological activity *in vitro*.

CHAPTER II

Literature Review

2.1 Introduction to *In Silico* Modeling of Atypical Kinetics in CYP2C9.

Unforeseen clinical complications involving ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) properties of a new drug are problematic for pharmaceutical companies, who spend an estimated \$800 million per year and an average of 12 years⁵² on the discovery and development of a lead compound. Of particular concern are atypical metabolic profiles, which can result in adverse drug-drug interactions and can be difficult to predict using *in vitro* HTS (high throughput screening) methods. The development of predictive statistical models of enzyme behavior *in silico* can allow early identification of potential problematic lead compounds at the beginning stages of drug discovery, reducing monetary, temporal, and labor-related costs. However, developing statistically valid HTS models with predictive power sufficient to maintain an acceptably low number of false positives / negatives has proven difficult⁵³. The majority of *in silico* kinetic models are concerned with inhibition^{46, 48, 54-56}; few have been developed for activation^{42, 49}.

In silico statistical model-building requires some knowledge of either the nature of the receptor binding site or the characteristics of ligands that bind to the receptor. As a result, *in silico* HTS models and the techniques used to generate them can be generally divided into three categories: ligand-based, ligand-receptor-based, and receptor-based. The ligand-based techniques are often employed before the other two, when little to nothing is known regarding a particular metabolic phenomenon besides the structures of the compounds responsible for it. Therefore, early models of CYP2C9 atypical kinetics were generated using ligand-based techniques. Ligand-receptor-based models appeared

when homology models of CYP2C9 were developed. However, because homology models are not experimentally determined and thus may be lacking in structural accuracy, statistical models of metabolism generated using them were often corroborated by developing the same statistical model via the time-tested ligand-based methods. When x-ray crystal structures of CYP2C9 were solved, these experimentally determined enzyme structures were reliable enough that they could, if desired, be used to the exclusion of ligand-based methods.

2.2 Ligand-Based Models

Due to the dearth of reliable 3D receptor structures, ligand-based models-- also known as pharmacophores-- are the most plentiful *in silico* models of P450 kinetic behavior. A pharmacophore is defined as a set of descriptors, usually three or four⁵⁷, arranged in a specific geometry that are responsible for receptor interactions leading to a particular enzyme behavior. Pharmacophores have been developed for several CYP isoforms including 2D6^{58, 59}, 3A4⁶⁰⁻⁶³, 2A6⁶⁴⁻⁶⁶, and 1A2⁶⁷⁻⁷⁰, but here the focus will be on 2C9.

One of the first predictive pharmacophore models of atypical kinetics in CYP2C9 was developed by Ekins et al. in 2000⁷¹. In addition to producing a 3D-QSAR, in which static compounds are aligned, this study also produced a 4D-QSAR, in which aligned compounds are allowed torsional freedom in order to obtain a superior alignment. The team constructed three initial pharmacophore models in Catalyst (Molecular Simulations, San Diego, CA), each built using a different training set. Multiple models composed of

multiple, diverse training sets increases the odds of discovering multiple and varied descriptors responsible for effector activity, which in this case was inhibition.

Model 1 used 9 inhibitors of tolbutamide 4-hydroxylation and diclofenac 4'-hydroxylation. These compounds were structurally diverse but had similar degrees of inhibition, differing by only 27-fold. This model produced four features presumed requisite for inhibition of diclofenac and tolbutamide metabolism: two hydrophobic groups, one hydrogen bond donor, and one hydrogen bond acceptor (**Figure 2.1**).

Model 2 was developed using 29 inhibitors of *S*-warfarin 7-hydroxylation, which had a greater degree of structural similarity than compounds used in Model 1, but whose inhibitor activities spanned two orders of magnitude. The resulting pharmacophore identified four features necessary for inhibition: one hydrophobic group, two hydrogen bond donors, and one hydrogen bond acceptor (**Figure 2.2**).

A third model was built using phenytoin- and bis-triazole-analogs of tolbutamide 4-hydroxylation, whose catalytic activities spanned only one order of magnitude. This model was therefore neither structurally diverse nor broad in range of activity. Model 3 produced three necessary features for inhibition: one hydrophobic group and two hydrogen bond acceptors. Interestingly, one hydrogen bond acceptor and one hydrophobic group were traits that were conserved in all three models (**Figure 2.3**).

All pharmacophores were validated in three ways: randomizing the inhibitors' activities with their structures (meaning that each given biological activity value was randomly assigned to a structure) and repeating pharmacophore generation; screening a

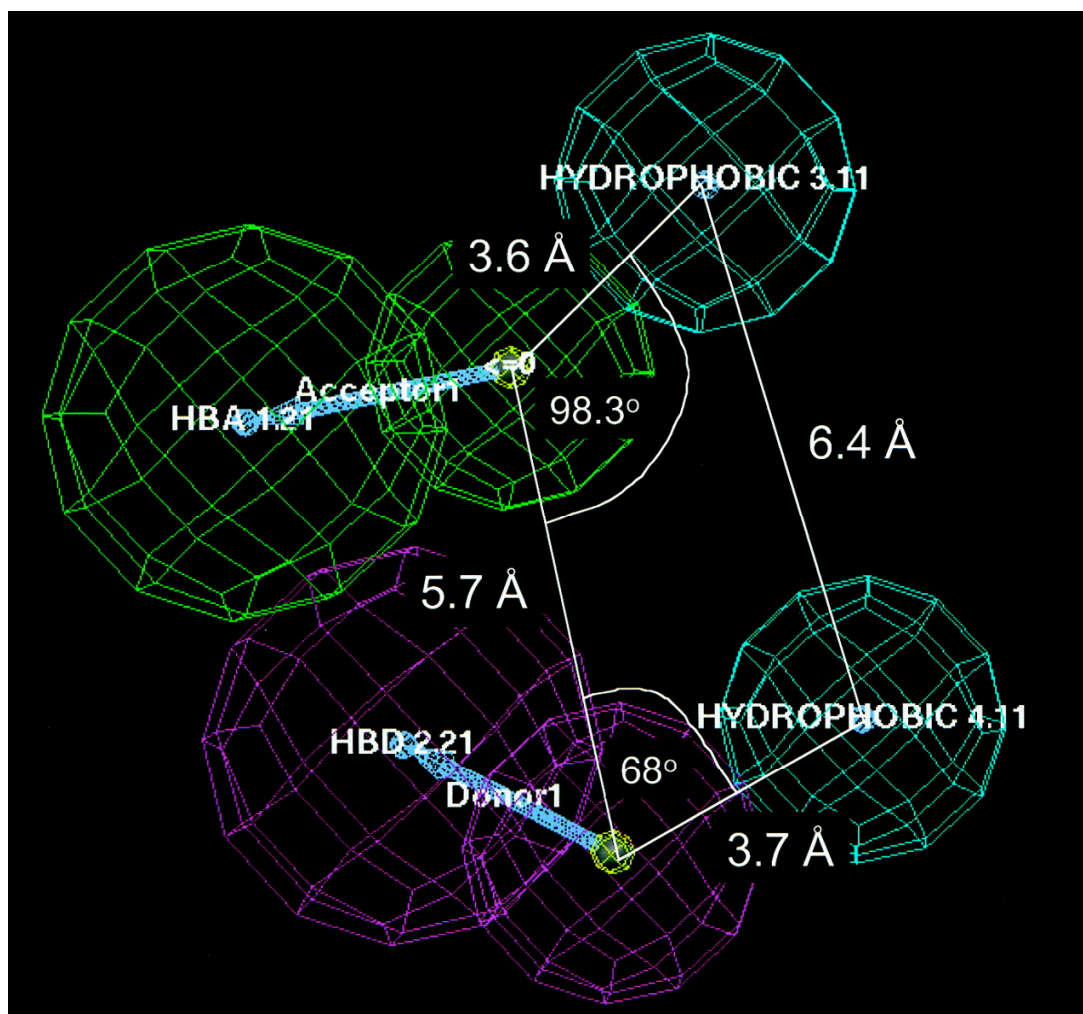


Figure 2.1. Pharmacophore Model 1.

Hydrophobic areas (cyan), hydrogen bond donor (HBD, purple), and a hydrogen bond acceptor feature (HBA, green) and a vector in the direction of the putative hydrogen bond donor are illustrated. From Ekins et al., 2000.

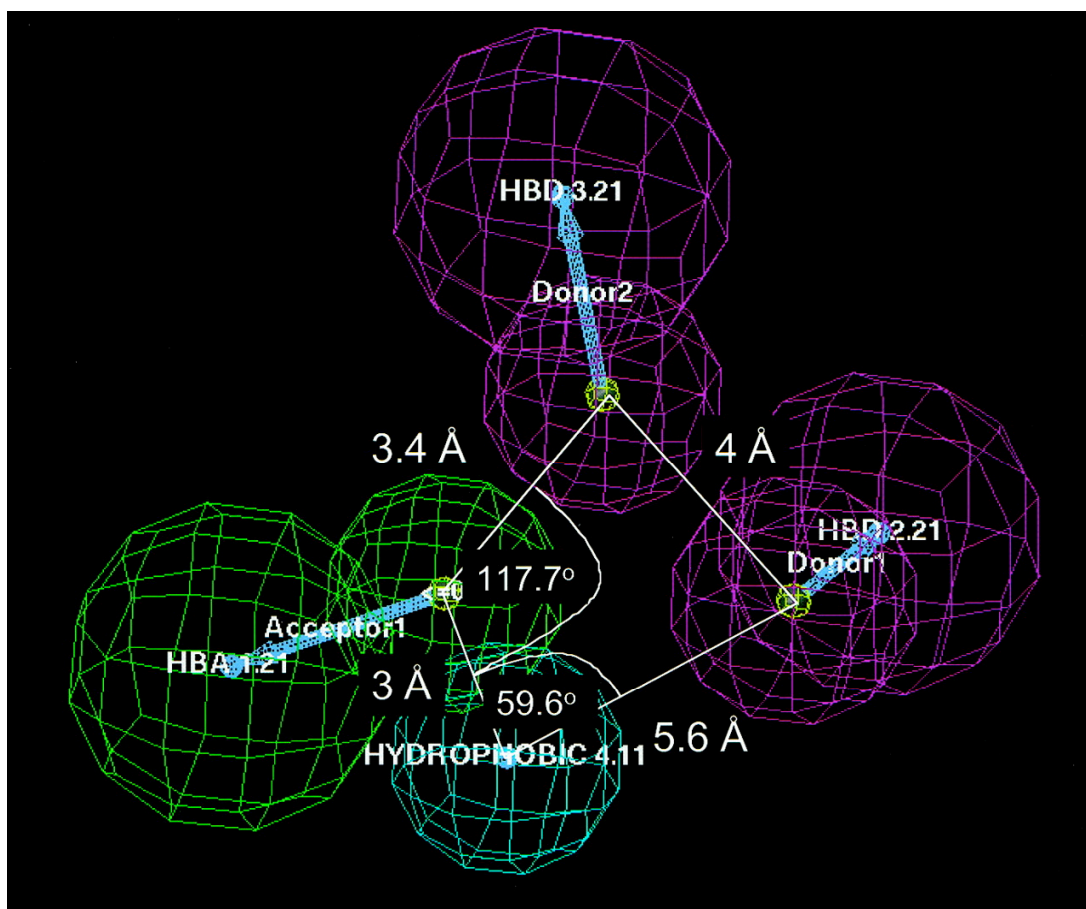


Figure 2.2. Pharmacophore Model 2.

Hydrophobic areas (cyan), hydrogen bond donor (HBD, purple), and a hydrogen bond acceptor feature (HBA, green) and a vector in the direction of the putative hydrogen bond donor are shown. From Ekins et al., 2000.

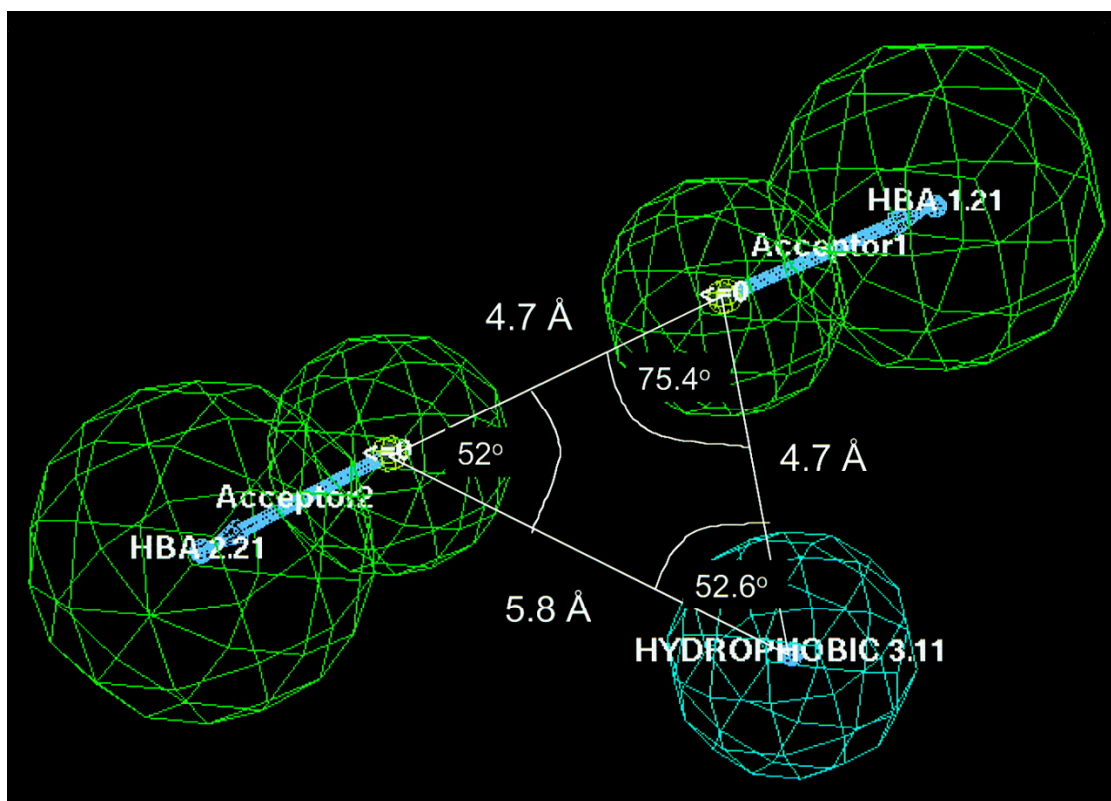


Figure 2.3. Pharmacophore Model 3.

Hydrophobic areas (cyan) and a hydrogen bond acceptor feature (HBA, green) and a vector in the direction of the putative hydrogen bond donor are illustrated. From Ekins et al., 2000.

test set of 14 inhibitors with known K_i values against the pharmacophore and conducting a partial least squares (PLS) calculation using a leave-one-out (LOO) algorithm, and a PLS calculation using a five-random-groups (5RG) algorithm repeated 100 times (5RG x 100). Activity permuting of the inhibitors in Model 1 resulted in a pharmacophore with a total energy cost similar to that of the null hypothesis (which states that there is no relationship between the pharmacophore generated and the activities of the compounds used to build it), indicating that any relationship that may be found between the data and the inhibitors' activities was likely only due to chance.

The Pearson's coefficient had decreased compared to that of the original Model 1, which would be expected in a non-predictive pharmacophore, and further suggests that the structure-activity-relationship in Model 1 is valid and not simply due to chance. However, the same descriptors were identified in this "null" pharmacophore as in the original Model 1; for a model to be considered valid its descriptors ought to differ from those of the null pharmacophore. In Model 2, the correlation coefficient decreased, the total energy costs of the null hypothesis and the permuted-activity model were the same, and their respective descriptors were different. Permuted-activity Model 3 had different descriptors and a lower Pearson's coefficient, but its total energy cost was not close to that of the null hypothesis. In Model 1, 12 of the 14 inhibitors were correctly predicted within 1 log unit residual using the fast fit function of Catalyst. In Model 2, 10 of 14 compounds were predicted within 1 log residual, also using the fast fit function. Allowing torsional rotation of the training set compounds (4D-QSAR) resulted in a LOO cross-validated correlation coefficient (q^2) of 0.54 for Model 1; the 3D-QSAR LOO q^2 was lower, at 0.40. In Model 2 the 4D LOO q^2 was 0.64. A q^2 greater than or equal to 0.30 is

considered significant at the 95% confidence limit⁷². This was a great improvement over the 3D LOO q^2 , which was statistically insignificant due to the fact that its q^2 value was less than 0.30. Model 3 did not generate a significant 3D nor 4D LOO q^2 value.

Validation with the 5RG x 100 algorithm is considered to be more credible than LOO, as the number of R^2 values used in calculating q^2 is not limited to the number of compounds in the test set as it is in LOO. 4D 5RG x 100 yielded a q^2 of 0.55 for Model 2, compared to an insignificant 3D 5RG x 100 q^2 value. The 4D approach did not improve the 5RG x 100 q^2 in Model 3, however, its 3D 5RG x 100 q^2 was significant, at 0.50. No significant 5RG x 100 q^2 could be obtained for Model 1 using either 3D nor 4D algorithms. The superior q^2 obtained from several 4D pharmacophores suggests that the rigid aligned conformer used in 3D-QSAR is not necessarily the most likely conformation of an effector compound. The study demonstrated that 4D QSAR may improve some models (as it did in Models 1 and 3), but actually reduced the predictive power of others. In addition, it appeared that different method of model validation can yield different results when determining models' significance. For example in Model 1, where the 5RG x 100 algorithm did not yield a significant correlation coefficient, but the LOO method did.

Another early predictive CYP2C9 pharmacophore was developed in 2001 by Rao et al.⁵⁴, and also modeled CYP2C9 inhibition of *S*-warfarin 7-hydroxylation. This CoMFA model consisted of 27 coumarin analogs, sulfonamides, carboxylate-containing drugs, and phenytoin. The training set compounds were manually aligned, allowing manual torsional rotation to obtain better superimposition. Manual alignment enabled the team to conform their model to complement site-directed mutagenesis studies as well as a homology model of CYP2C9 (at the time there was no x-ray crystal structure), which had

identified possible pi-stacking interactions between Phe-114 and aromatic rings of the inhibitor compounds. Sulfaphenazole, an outlier in an earlier CoMFA model by the same researchers, was presumed to interact with a positively charged residue; the LOO q^2 calculated when including sulfaphenazole in the model was 0.70. Interestingly, re-aligning sulfaphenazole in such a way that its aniline ring could pi-stack with Phe-114 prevented any electrostatic interaction with the protein, but the LOO q^2 calculated from this model was 0.72, suggesting that pi-stacking in this case may be at least as important as electrostatics in sulfaphenazole active site binding.

Descriptors deemed important in inhibitor binding identified by the re-aligned pharmacophore were a single strong negative charge interaction and one or two positive charge interactions. Hydrophobic interactions were not included in the CoMFA model, despite the data obtained from re-aligning sulfaphenazole. A test set of 14 compounds (10 of which were sulfonamides) that were not included in the training set was screened against the pharmacophore, which was able to predict the biological activities of the test set compounds in the words of the author to a “reasonable degree of accuracy”. Four sulfonamides were then arbitrarily selected from the test set and were included with the original training set to produce a new pharmacophore. This new model was used to screen the remaining 10 of the original 14 test set compounds, with the result of an increased ability of the model to predict sulfonamide activities. A final pharmacophore was developed by combining the original training set and all test set compounds, a total of 41 compounds. This model yielded a LOO q^2 of 0.65 when the original test set of 14 inhibitors was screened against it. All pharmacophores were cross-validated using PLS and a LOO algorithm, as well as a leave-20%-out algorithm, and by permuting biological

activities of the inhibitors in the test and training sets. For all pharmacophores the leave-20%-out method produced q^2 values similar to those calculated using the LOO method. Randomly assigning binding affinities of compounds used in the training sets resulted in pharmacophores with negative q^2 values, supporting the validity of the actual pharmacophore models developed by the researchers. The main value of this work lay in its ability to correctly identify potential active site interactions and binding modes of two structural classes of compounds (coumarin analogs and sulfonamides) in an enzyme structure that did not exist at the time, since no homology model nor crystal structure had yet been published. Also, this is the first *in silico* study involving CYP2C9 to explore the possibility of pi-stacking interactions as important to effector binding in the active site.

Locuson et al. conducted a study investigating inhibition of CYP2C9-mediated *S*-warfarin 7-hydroxylation by benzbromarone analogs⁷³. Here, by employing CoMFA and CoMSIA methodology, the team sought to accomplish three goals: 1) determine the degree to which anionic character affected benzbromarone analog binding affinity to CYP2C9, 2) develop a predictive pharmacophore for the purpose of screening unknown compounds that may precipitate drug-drug interactions, and 3) use the pharmacophore to identify active site residues contributing to specific interactions predicted by the pharmacophore. The latter was also useful for helping to verify residue conformations and the ligand binding mode in an x-ray crystal structure of CYP2C9 bound with *s*-warfarin, to distinguish what aspects of the crystal structure are likely to be artifacts of the crystallographic process. To build the initial pharmacophore, 17 minimized benzbromarone analogs with known binding affinity constants were aligned such that the primary site of oxidation, electronegative oxygens, and aromatic ring corresponding to

the brominated ring of benzbromarone were superimposed on *s*-warfarin. Torsions were manually adjusted for better overlap. These compounds were very similar in structure, but had activities ranging over three orders of magnitude. Five of the 17 analogs were anionic at physiologic pH, and both a neutral (fully protonated) training set and an anionic training set were used in pharmacophore generation to ascertain the importance of the role anionic character played in binding affinity.

Pharmacophores were made using both CoMFA and CoMSIA methods. CoMSIA differs from CoMFA in that it indexes the training set compounds by structural similarity and derives its descriptor fields from this rather than from calculating dipole moments. The models were cross-validated via a LOO algorithm and by assigning permuted biological activities to the training set. Negative q^2 values were calculated for the pharmacophore composed of permuted affinity constants, which indicates a pharmacophore that has no correlation with the data used in its construction, and bolsters the argument that the pharmacophores built using correct binding constants were not due to chance. LOO q^2 values were greater than 0.30 for all pharmacophores, further adding to the statistical validity of the models. Contour plots of the CoMFA models of both anionic and neutral training sets revealed that formal charge state was not an important factor for predictivity; the neutral test set generated a q^2 of 0.54 compared to the anionic set's 0.49, indicating a greater capability for accurate K_i prediction in the neutral set. The CoMFA models, which only calculated electrostatic and steric descriptors, predicted unfavorable steric interactions around the C-2 position of benzbromarone (**Figure 2.4**).

Analogs containing C-2 substituents had progressively lower affinities as the substituent size increased. According to the CoMFA data, benzbromarone analog binding

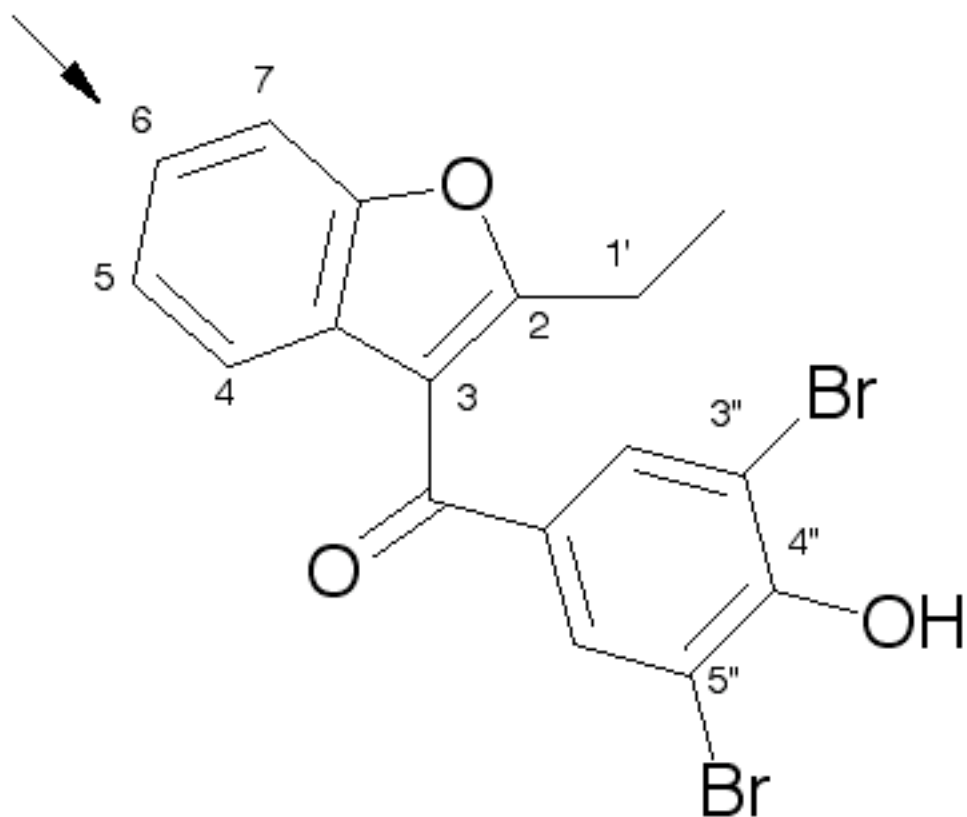


Figure 2.4. Chemical Structure of Benzbromarone.

Numbering scheme of benzbromarone used in the text is shown. The site of oxidation by CYP2C9 is designated by the arrow. From Locuson et al., 2004.

affinity could be attributed 40-46% to electrostatics, and 54-60% to sterics. There was no major difference in descriptors between the anionic and neutral CoMFA-generated pharmacophores, which further eroded the importance of formal state charge to binding affinity. CoMSIA-generated pharmacophores assessed not only sterics and electrostatics, but also hydrophobicity. In these models, a favorable hydrophobic interaction was predicted to exist above the bromated ring of benzbromarone. It was also noted that a CoMSIA model only including electrostatics and hydrophobicity produced better q^2 values (0.61 for the anionic training set, 0.60 for the neutral set) than a model including only electrostatics and sterics (0.48 for the anionic set, 0.61 for the neutral set). This implies that sterics are not as important a factor in binding affinity. Rather, that hydrophobicity and electrostatics are the driving forces in this case. However, sterics did play a greater role in the binding of a few benzbromarone analogs; compounds that lacked a formal negative charge but that were still somewhat acidic in nature had greater affinity if they contained bulky substituents on what corresponded to the brominated ring of benzbromarone than comparable analogs (**Figure 2.5**) devoid of the bulky groups. In addition, their binding affinities were only marginally less than formally anionic inhibitors. Warfarins, NSAIDs, and sulfaphenazoles were added to the set of benzbromarone analogs, which increased the structural diversity of the training set and brought the total number of compounds used to 58. These compounds were aligned in the same fashion as the 17 original molecules.

The resulting CoMSIA models agreed with the original, smaller models in that hydrophobicity was of greater importance to binding affinity than sterics. This was made apparent by the increase in q^2 from 0.64 in an electrostatics/sterics-only model to 0.75 in

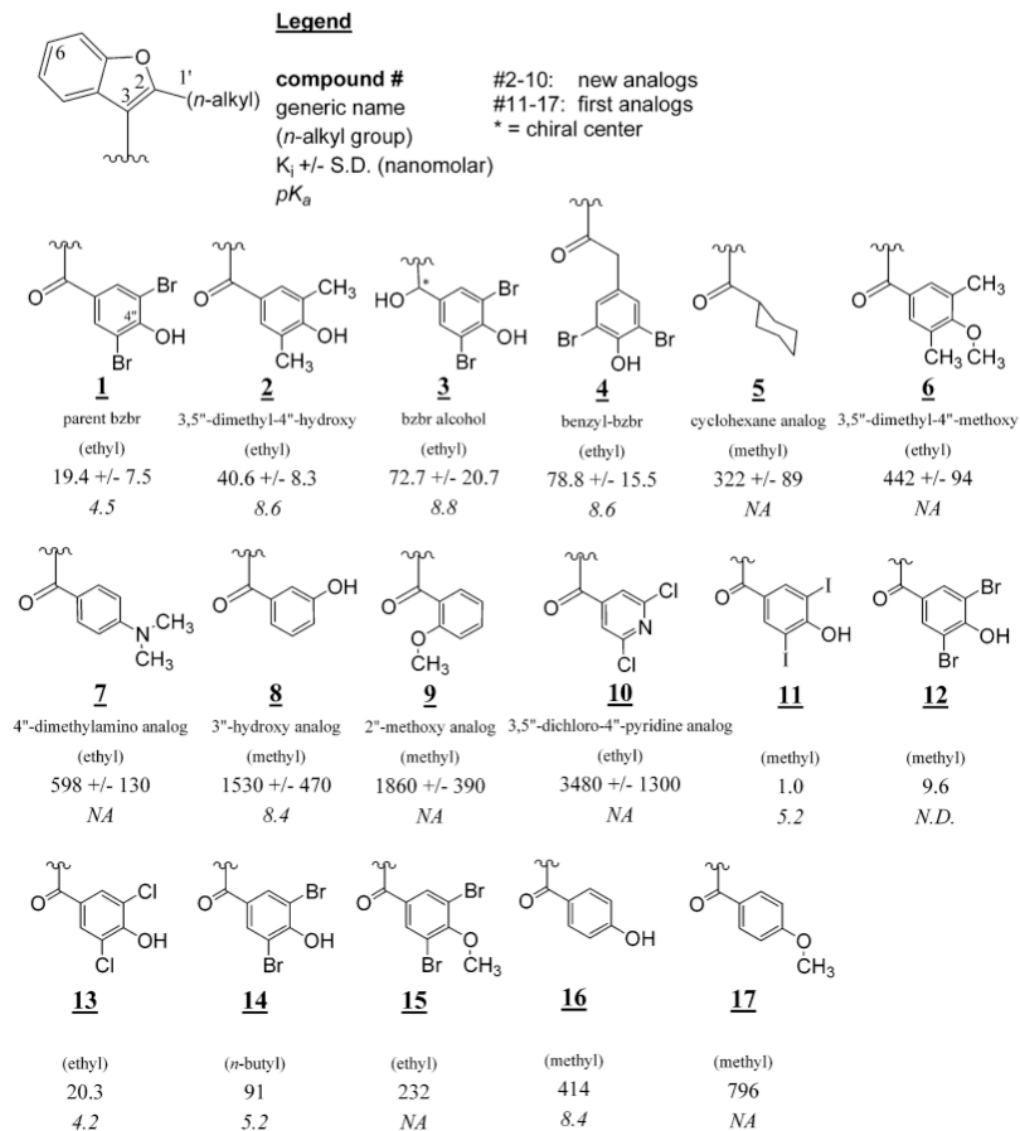


Figure 2.5. Chemical Structures of Benzbromarone Analogs.

Chemical structures used to develop the model are shown. From Locuson et al., 2004.

an electrostatics/sterics/hydrophobicity model. This pharmacophore predicted a favorable electrostatic interaction at the ketone of benzbromarone, which overlaid well upon oxygens belonging to sulfaphenazole and warfarin and suggested that a sufficiently electronegative functional group placed there would aid to increase the analog's binding affinity. Therefore, a general conclusion can be drawn that compounds better able to become polarized and redistribute electron density to such a strategically located oxygen will have lower K_i values, even if they are not formally anionic.

At the time of this publication an x-ray crystal structure of CYP2C9 co-crystallized with *S*-warfarin (1OG5) had become available¹⁹. Contour plots of the pharmacophores could aid in identifying specific residues in the active site by superimposing them onto the crystal structure. Being a competitive inhibitor, benzbromarone may bind to the same site as warfarin does prior to its metabolism. However, in the crystal structure, warfarin was bound in a hydrophobic pocket with its site of oxidation 7 Å from the heme iron-- too great a distance for metabolism to occur. While the pharmacophore could be superimposed onto the crystal structure in this binding site and still have all of its descriptor criteria fulfilled, a second binding site within the same pocket but closer to the heme iron for catalysis, also fulfilled the descriptors identified by the CoMSIA/CoMFA results. Though the crystal structure suggested otherwise, this second site was more likely to be the actual metabolite-producing binding site. In it, one favorable steric site was located near the site of warfarin metabolism, and when superimposed on the crystal structure lies nearly in contact with Phe-114, a residue known from mutagenesis studies to be important for metabolism in CYP2C9¹⁹. Assuming benzbromarone takes a similar pose to that of warfarin in the

crystal structure, favorable hydrophobic interactions between benzbromarone and either Phe-114 or Phe-476 are possible, as well as favorable hydrophobic interactions between the inhibitor's bromine substituents (or an equally bulky group on a benzbromarone analog) and hydrophobic residues such as Leu-208. Phe-114 and Phe-476 have both been identified as residues important to CYP2C9 metabolism¹³.

In 1OG5, Arg-108, which has been shown by mutagenesis studies to be critical to metabolism of several substrates by CYP2C9, including warfarin and benzbromarone⁷³ is oriented such that its cationic side-chain is directed away from the active site (**Figure 2.6**). This orientation is not conducive to productive substrate binding due to its large distance from the heme iron, however, given the geometry of the CoMSIA pharmacophore data, favorable hydrophobic interactions could occur between the benzoyl ring of benzbromarone and Phe-114 in a T-stacking geometry (**Figure 2.7**), which not only puts the site of metabolism about 3 Å directly above the heme iron, but also places its electronegative ketone oxygen 2.5 Å of an inward-facing side chain ammonium of Arg-108. It was postulated that perhaps the outward-facing, crystallized orientation of Arg-108 may be the result of high protein flexibility in that region, or even an artifact of the crystallization process. Later another crystal structure was released¹², in which flurbiprofen was bound in a productive pose with its anionic carboxylate group H-bonded to the side-chain of Arg-108 (**Figure 2.8**).

Egnell et al. developed the only pharmacophore model of CYP2C9 heteroactivation to date, using a test set of 36 inhibitors and activators of 7-methoxy-4-trifluoromethyl-coumarin (MFC)⁴⁹. Due to the lack of mechanistic knowledge of

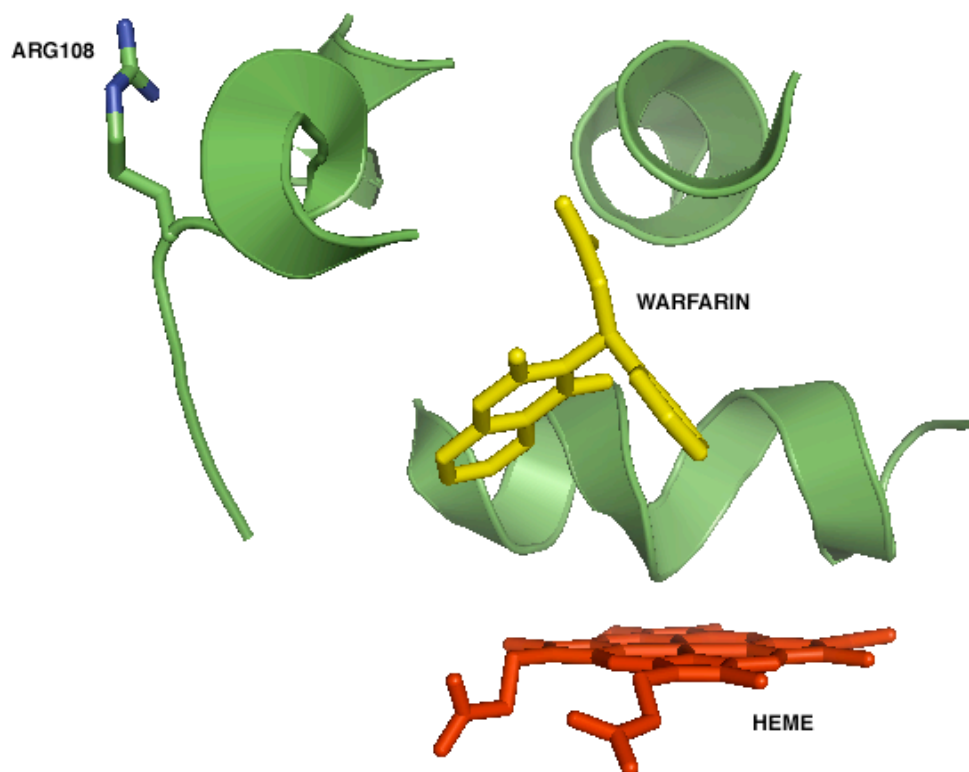


Figure 2.6. Arg-108 Conformation in 1OG5 Crystal Structure of CYP2C9.

The crystal structure of 2C9 co-crystallized with *s*-warfarin by Williams et al., ProteinDataBase code 1OG5 is shown. The substrate, warfarin, is in goldenrod and the heme is crimson. Note how the side chain of residue R108 is directed away from the active site cavity where it cannot readily interact with the substrate. From Williams et al., 2003.

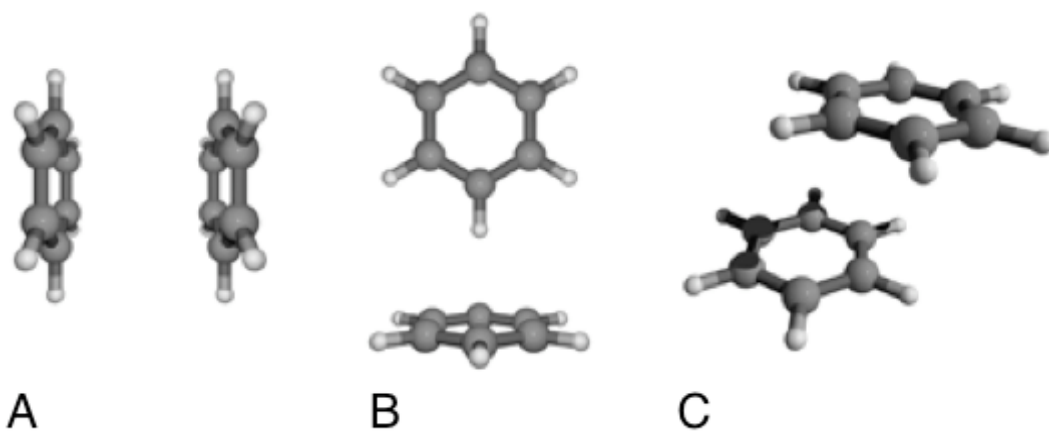


Figure 2.7. Pi-Stacking Geometries.

Stacking geometries of aromatic rings are illustrated: A) sandwich stacking, B) T-stacking, and C) displaced parallel stacking. From Sinnokrot et al., 2002⁷⁴.

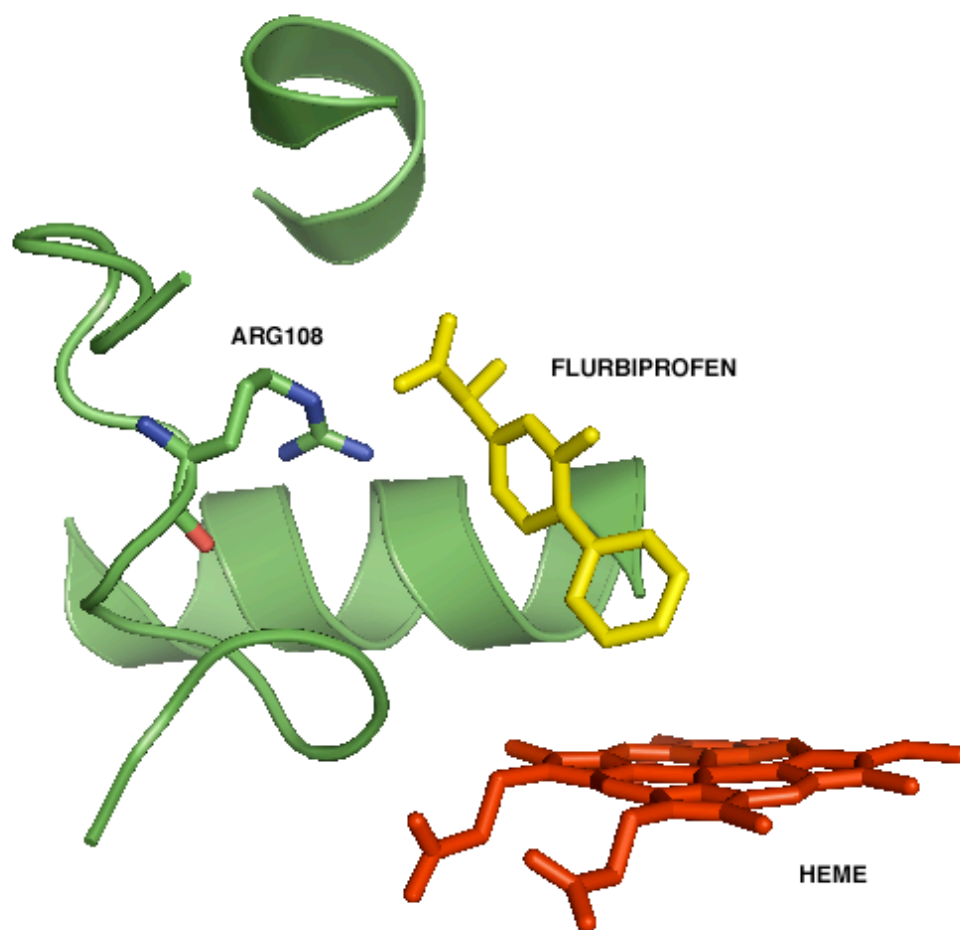


Figure 2.8. Arg-108 Conformation in 1R9O Crystal Structure of CYP2C9.

The crystal structure of 2C9 co-crystallized with flurbiprofen by Wester et al., ProteinDataBase code 1R9O is shown. The substrate, flurbiprofen is in goldenrod and the heme is crimson. Note how the side chain of residue R108 differs in orientation from that in the crystal structure by Williams (see Figure 2.6); here it comprises part of the active site topology and forms a hydrogen bond with the anionic carboxylate group of the substrate.

heteroactivation in 2C9, and hence an inability to calculate a pure heteroactivator affinity constant, $v_{150\%}$, the concentration of activator at which substrate turnover was 50% higher than v_{\max} , was substituted as an endpoint to measure the degree of metabolic activation. The pharmacophore was generated in Catalyst using a test set of 36 compounds that were either known activators or known inhibitors. Alignment of the test set compounds resulted in a pharmacophore that suggested hydrogen bond accepting groups, aromatic rings, and hydrophobic groups to be of prime importance for heteroactivation of MFC metabolism by CYP2C9. The correlation coefficient of determination, R^2 , was 0.50. Statistical validation was carried out using a LOO algorithm, as well as comparison of the total energy cost of the model to that of the null pharmacophore. A diverse database of 1504 compounds (including oral drugs and known P450 inhibitors) was screened against the pharmacophore. Thirty-seven of these were identified by the model as activators of CYP2C9-mediated MFC metabolism, with activation defined as an increase in metabolism greater than 20% at a concentration equal to or less than 125 μM . Among these were amiodarone, niclosamide, liothyronine, meclofenamate, zafirlukast, estropipate, dichlophenamide, hydroflumethiazide, mefenamic acid, and warfarin, all of which had $v_{150\%}$ values at concentrations below 10 μM . All identified compounds except for niclosamide were previously unknown activators of CYP2C9-mediated MFC metabolism. The top five most potent newly-identified activators could be easily superimposed onto the pharmacophore, matching both its chemical and geometric descriptors. Though the model correctly identified strong activators and inhibitors, it failed to correctly quantify the heteroactivity of compounds with more moderate potency. The binary nature of this pharmacophore model may be in

part due to the distribution of $v_{150\%}$ values of the training set, which contained strong activators and inhibitors, but no compounds of moderate biological activity.

2.3 Ligand-Receptor-Based Models

Both ligand-receptor- and receptor-based models require three-dimensional structures of both ligand(s) and receptor (enzyme), and are developed either by docking the ligand(s) into the active site or by applying a pharmacophore to the active site to identify specific residue interactions. For ligand-receptor-based methods receptor structures are usually obtained via homology modeling. Homology modeling is used when no crystal structure is available for a particular enzyme, but one does exist for an enzyme with a high degree of sequence similarity. Generally speaking, the higher the percentage of sequence similarity, the more reliable the homology model will be⁷⁵. The technique involves sequence alignment of the template enzyme crystal structure with the target enzyme, substitution of homolog residues for template residues, and then energy-minimizing the 3D geometry of the resulting structure. Prior to the development of CYP2C9 crystal structures, homology models of CYP2C9 were built using the crystal structure of CYP2C5⁷⁶, which shares a 77% sequence identity with CYP2C9, as a template.

The first study investigating atypical kinetics in CYP2C9 that involved homology modeling was conducted in 2000 by Afzelius et al⁴⁶. In this work both a 3D pharmacophore model and a homology model were constructed for 29 structurally diverse competitive inhibitors of diclofenac 4'-hydroxylation and 7-methoxy-4-

trifluoromethylcoumarin demethylation. The crystal structure of rabbit CYP2C5 served as a template, and the alignment between the two enzymes was adjusted manually using the Homology module of InsightII. Overlaying the backbones of the two structures produced an RMSd of 0.18 Å. The inhibitor compounds were energy minimized and docked into the model with the docking program GOLD [Gareth Jones, University of Sheffield, UK]. Following this, a pharmacophore model was developed using the resulting lowest-energy docked conformers of 21 of the same set of 29 inhibitors. Electrostatic and hydrophobic interaction energies were calculated using the program GRID [Molecular Discovery Ltd., University of Oxford, UK], and were subjected to statistical analysis with GOLPE. GRID, as its name implies, calculates a topology of the active site with each grid point defined by a specific ligand-receptor interaction energy. Interactions whose overall energies were less than 0.02 kcal/mol were excluded from analysis, and the remaining interaction variables were clustered together based on their respective docked conformers' geometric similarity to *s*-warfarin, phenytoin, progesterone, and sulfaphenazole. These four substrates have a known mechanism of binding and/or site of hydroxylation by CYP2C9, and therefore any competitive inhibitor such as those in the training set would likely take on a conformation similar to one of them when bound in the active site.

The eight compounds that were not used in the training set were used instead as a test set and were screened against the resulting pharmacophore. To test the validity of the pharmacophore model a LOO algorithm, a leave-20%-out algorithm, and a leave-50%-out algorithm were implemented. The LOO method calculated a Pearson's correlation coefficient of 0.947 and a q^2 of 0.730. The L-20%-O method yielded a q^2 of 0.672, and

the L-50%-O method produced a q^2 of 0.497. All of these q^2 values exceed 0.30 and are therefore statistically significant. In addition, binding affinities of the test set predicted by the pharmacophore were within 0.5 log units of their respective experimental values. The importance of Afzelius' research here lay in the development of the first CYP2C9 homology model, which enabled the future use of receptor-based methodologies for studying the enzyme's kinetics. The fact that a highly-predictive pharmacophore could be constructed using inhibitor conformers resulting from the homology model docking supported the accuracy of the homology model.

In a later study⁴⁷ Afzelius et al. conducted a very similar set of experiments, this time involving the use of a molecular descriptor-calculating tool called ALMOND [MIA, Perugia, Italy], which could process the docked ligand-receptor interactions predicted by GRID and compute which descriptors (i.e. hydrogen bond accepting, donating, or hydrophobic) were most involved in inhibitor binding. The methodology used in ALMOND did not require alignment of the inhibitor compounds, which circumvented the problems involved in trying to align a very structurally diverse set of compounds. Accurately predicting inhibition in CYP2C9 could be achieved, as could better understanding the mechanism by which inhibition occurred, using only one model.

Twenty-two inhibitors with known binding affinity constants were used as a training set and 12 additional inhibitors comprised the test set. Using the same methods employed in the first study, the compounds were docked into the active site of the CYP2C5-based homology model of CYP2C9. The GRID-calculated interactions of the resulting lowest-energy docked conformers were then input to ALMOND, which in turn calculated the most important descriptors based not only on the interaction energies

supplied by GRID, but also on the spatial distances between the atoms involved in those interactions. For each descriptor of every inhibitor the interaction with the lowest energy and whose participating atoms were within an arbitrary distance of one another was deemed the "maximum interaction", and a score was calculated by ALMOND using its energy as a function of distance. To validate the model, a PLS statistical analysis was conducted using both the LOO method and a random grouping algorithm, which ensured no one compound or group of compounds was driving the correlation. The analysis was performed for both the training set and the test set, correlating binding affinity constants with the ALMOND scores. For the 22-compound training set the correlation coefficient of determination was 0.81, the LOO q^2 was 0.62, and the random grouping q^2 was 0.59, indicating a model with significant predictive power. Binding affinities of nine out of 12 of the test set compounds were predicted within 0.3 log units, and two others were predicted within 0.5 log units.

Because the interaction grid points formed a topology of the CYP2C9 active site it was possible to observe which residues were responsible for which interactions. For various inhibitors used in the study, several residues were strongly implicated in binding, including Phe-114, Phe-476, and Asp-293-- all of which have been shown experimentally to be important in the catalytic activity of CYP2C9^{5,13}. The receptor-based methodology developed here made it possible to construct a predictive *in silico* kinetics model of CYP2C9 when prior knowledge of potential conformers of the compounds being screened is limited, and/or the breadth of structural diversity of the compounds is very wide.

2.4 Receptor-Based Models

Receptor-based models are generated using X-ray crystallographs (often referred to as "crystal structures"), which are preferred to homology models because that they are experimentally determined representations of the enzymes themselves. However, when modeling metabolism using crystal structures it is important to keep in mind that the structure may not be a perfect representation of enzymes *in vivo*. Crystallizing P450s is challenging due to the low solubility of the enzymes, and modifications such as mutating specific surface residues as well as truncating the hydrophobic N-terminal tail, which anchors the CYP to a cellular membrane. Crystal formation is also not undertaken at physiological conditions. The altered enzymes are tested *in vitro* to ensure that such modifications do not compromise catalytic activity, however, it is still possible that the overall conformation of the enzyme could be affected, which could in turn alter the mechanism of substrate metabolism. Two crystal structures have been solved for CYP2C9-- one co-crystallized with warfarin ¹⁹, and the other with flurbiprofen ¹².

When Hummel et al. obtained NMR data proving that dapsone-mediated activation of flurbiprofen metabolism by CYP2C9 required the simultaneous presence of both effector and substrate in the active site ³⁸, they conducted docking studies on another atypical kinetic system: benzbromarone inhibition of flurbiprofen metabolism in CYP2C9 ⁷⁷. Using MoViT v.8.0, in-house software developed by Pfizer, Inc. (LaJolla, CA), benzbromarone was docked into the crystal structure of CYP2C9 with flurbiprofen co-crystallized in the active site solved by Wester, and minimized to its lowest energy conformation. Flurbiprofen was excised from the crystal structure and re-docked into the active site to develop the final receptor structure. NMR data was used to orient both

flurbiprofen and benzbromarone in the active site prior to docking. Lowest-energy conformers of benzbromarone docked in two overlapping binding sites that were composed of the B-C loop, helices F,G, and I, and the C-terminal loop, which contains Phe476. Both of these possible effector binding sites are superimposable with that of dapsone when docked into the same receptor according to Wester. Another interesting note is that in the CYP2C9*3 variant benzbromarone changes from an inhibitor to an activator of flurbiprofen metabolism. MoViT requires a spatially fixed receptor, which eliminates the possibility of observing via docking whether or not the I359L mutation in the *3 variant would elicit a conformational change in the enzyme that would allow benzbromarone to dock in a conformation more conducive to encouraging flurbiprofen metabolism. (This could only be deduced via a very lengthy molecular dynamics simulation.) However, the docking results do show that sufficient space exists in the wild-type crystal structure as-is for both benzbromarone and flurbiprofen. Although Wester had already docked dapsone into this same receptor prior to this publication, Hummel's study reinforced Wester's findings and also suggested the role polymorphism may play in altering an effector's binding pose, causing it to change its effect on substrate metabolism. in the active site of CYP2C9, and aided in the identification of previously unknown effector binding sites within the greater active site of CYP2C9.

Kumar et al. built on the work done by Hummel, above, by including two amiodarone analogs in addition to benzbromarone. Amiodarone's effect on naproxen demethylation by CYP2C9 differs from that of three of its analogs, desmethyamiodarone, benzbromarone, and benz(meth)arone. *In vitro*, amiodarone activated naproxen demethylation at concentrations from 0.5 and 2.5 μM , but inhibited

metabolism at concentrations above 12.5 μ M. In contrast, benz(meth)arone activated naproxen metabolism in both alleles, and benzbromarone activated naproxen metabolism in the wild-type but inhibited it in the variant. Both desmethylamiodarone and benzbromarone inhibited naproxen metabolism at any concentration, but benz(meth)arone actually activated naproxen demethylation. Benzbromarone and benz(meth)arone not only affected the substrate turnover rate but also altered the kinetic profile of naproxen metabolism from biphasic to hyperbolic. It was hypothesized that different binding modes may be the cause of the discrepancies in the effects the amiodarone analogs had on naproxen metabolism. Docking studies were performed with the program Glide (Schrödinger, LLC, Portland, OR). The receptors were developed using the x-ray crystal structure solved by Wester ¹², and included CYP2C9-naproxen as well as CYP2C9-flurbiprofen and CYP2C9 alone as controls. Amiodarone, its three analogs, and naproxen were docked into the receptors and the top resulting scores and binding modes were compared. In the empty active site naproxen bound in a fashion similar to the co-crystallized flurbiprofen, with its methyl group-- the site of metabolism-- pointing toward the heme and its carboxylate moiety interacting with Arg-108. Amiodarone and its analogs, which are themselves substrates of CYP2C9, also bound in the empty active site of CYP2C9 in conformations in agreement with the formation of their respective metabolites. When docked in the presence of naproxen, benz(meth)arone oriented itself in such a way that it interacted with Phe-476 via one of its aromatic ring systems. Benzbromarone docked in roughly the same location as benz(meth)arone, near recognition site 6, but its ring systems were oriented oppositely compared to those of benz(meth)arone. It was thought that part of the mechanism behind benz(meth)arone's

activation of naproxen demethylation involved its ability to interact with Phe-476 via pi-stacking, which would hold the effector in place. Steric forces between benz(meth)arone and the substrate would crowd naproxen closer to the heme iron, thus increasing its rate of metabolism. Benzbromarone is a larger compound compared to benz(meth)arone due to its bromide groups; its size may restrict it to binding in such a way that it does not easily orient itself in a mode conducive to Phe-476 interaction and thus is not as readily stabilized in position inside the active site. In addition, benzbromarone is acidic at physiological pH, which may also play a role in determining its binding mode. This study was one of the few that computationally investigated atypical kinetics as a function of binding orientation in the active site of CYP2C9, and aided in the identification of specific effector-active site residue interactions that may be responsible for the effector's kinetics-altering properties.

Locuson et al. developed a screening protocol for activators of flurbiprofen metabolism in CYP2C9, also using Glide. Dapsone and nineteen of its analogs-- seven activators and twelve non-activators or inhibitors-- were chosen to develop the protocol based on their known *in vitro* effect on flurbiprofen metabolism and also due to their having a wide range of activities despite being structurally similar. The receptor was developed using Wester's x-ray crystal structure with flurbiprofen co-crystallized in the active site. Dapsone was docked into the active site along with flurbiprofen to act as a "placeholder" to prevent the active site from collapsing during the subsequent 2 ns molecular dynamics simulation using the Amber8 suite of modeling software. Molecular dynamics served to mitigate artifacts of the crystallization process that may affect enzyme conformation; the post-simulation structure would be more likely to accurately

reflect the enzyme conformation under physiological conditions. Dapsone was excised from the active site following the simulation, resulting in the finished receptor used for screening. The structures of the dapsone analog test set compounds were built and energy minimized using Sybyl 7.0 [Tripos, Inc., St. Louis, MO], and were docked into the active site using Glide and the top ten scored conformers for each effector were retained for analysis. No correlation was seen between the Glide scores and the degree of heteroactivation, however, the scores of the inhibitors tended to fall below that of the weakest activator. In addition, if flurbiprofen was removed from the receptor and the test set docked through the empty active site the compounds known to be activators preferred binding in a region occupied by dapsone during the molecular dynamics simulation (the "effector site") rather than the substrate binding pocket more proximal to the heme. The qualitative activities of five out of six heteroactivators were correctly predicted if one of the top three ligand conformations bound to the effector site. Using both the binding mode and the Glide score of the weakest activator as a cutoff, 11 out of 12 inhibitors were correctly predicted as such. To validate this method of effector activity prediction, the ZINC database⁷⁸ composed of drug-like compounds was chosen to be screened. The activities of the database compounds were unknown, and it was seeded with the 20 compounds comprising the test set as a control. As preparation for docking, up to four stereoisomers were generated for chiral compounds in the database and were included in the screen. This extended database was sieved through a molecular weight filter using Sybyl to limit the size of the compounds to a range of 120-350 amu. Of the filtered database, alternate ionization states were built for those molecules containing functional groups whose pKa values were between 5 and 9. Energy minimization was the final step

in the preparation of the database which could now be used as a validation set. Docking was carried out in two steps: first, torsions were held rigid so that the ligands could not deviate from their already optimized geometry. This screening, which was performed with flurbiprofen simultaneously present in the active site to cull ligands whose size would exceed that required to fit into the active site with the substrate, eliminated 76% of the ZINC compounds. In the second step the remaining molecules were docked into the empty CYP2C9 active site and torsional restraints were removed so that rotamers would also be scored. The top ten binding modes for each docked compound were analyzed to assess their binding preferences: whether or not they preferred the effector site or the substrate site. Only compounds that bound to the effector site in at least the top two scored conformations were considered to be heteroactivators. Ten of these compounds-- four whose Glide scores were superior to and six whose Glide scores were inferior to that of dapsone-- were selected for *in vitro* analysis of their biological activities. One compound was found to activate flurbiprofen metabolism by CYP2C9 by 46%, which was comparable to the activating activity of dapsone and was statistically different from the DMSO control ($p < 0.05$). Four other compounds appeared to be weak activators but were not statistically different from the DMSO control. Three others were found to be inhibitors, and two had no effect on flurbiprofen metabolism even at concentrations as high as 10 mM.

This was the first and at the time of this writing the only study to demonstrate that computational methods for the prediction of heteroactivation in CYP2C9 could be developed with success. It was also the first predictive computational method to identify a compound whose strong activating activity was unknown prior to this study. This

research also unearthed the presence of a specific binding site preferred by heteroactivators even when given the choice of docking in the substrate binding site, where favorable active site interactions are known to occur with compounds of similar chemical composition to the dapsons analogs. One drawback, however, is the amount of human involvement required to determine the binding mode for all compounds, which may be a limiting factor if the processed database to be screened is very large. Also, as with so many other models, the rigid receptor may cause many unknown activators to be missed, either because they exceed the space allotted in the active site during the first step of the screening process, or because residues that are required for activator binding may become part of the active site topology only with conformational change of the enzyme.

In conclusion, both pharmacophore and receptor-based modeling techniques have been used successfully to expand our understanding of the active site interactions responsible for atypical kinetics in CYP2C9. This work aims to build on that body of literature and add our own contribution to illuminating the mechanism(s) and predicting the likelihood of the occurrence of atypical kinetics, specifically heteroactivation.

CHAPTER III

Molecular Dynamics Simulations of the CYP2C9- Flurbiprofen Complex With and Without Dapsone Present

3.1. Introduction

Molecular dynamics simulation (MD) has been extensively used in a multitude of disciplines as a means of understanding the structural and dynamic behavioral properties of molecules at the atomic level. These include physics, engineering, chemistry, and biology. Molecular dynamics is generally considered to be a superior method compared to the Monte Carlo approach due to the greater amount of information the former provides; MD monitors fluctuations in the atoms' positions in three-dimensional space as well as their velocities per unit of time. (In contrast, the Monte Carlo method concerns itself primarily with a compound's spatial position without taking time dependence into account.)

These motions become increasingly manifold as the molecule grows in complexity, and macromolecules such as proteins display a broad range of characteristic motions. Essentially, they can be consolidated into three categories: local motion, medium-scale motion, and global motion. Local motion includes movement on the atomic level such as bond length, angle bending, and torsional rotation. The time-scale for such studies ranges from femtoseconds (10^{-15} s) to picoseconds (10^{-12} s). Examples of medium-scale motion are protein loop motion, rigid-body motion, and terminal-arm motion. Active site conformational changes and binding specificity are readily determined on this scale, which is typically in nanoseconds (10^{-9} s). The medium scale is the one most commonly employed by biological scientists, and certainly those studying enzymatic activity, as it offers a timescale sufficiently long for observing biological phenomena, but short enough so as to be computationally accessible. The third type of

motion is global motion, which encompasses protein folding/unfolding, subunit association, and helix-coil transition, operates on a timescale ranging from milliseconds (10^{-3} s) to hours (10^4 s). Though providing a great deal of information regarding overall protein functionality, this timescale is generally far too computationally expensive to have practical application, and is to-date never used for molecular dynamics. In this work we have chosen to use the medium scale, for the reasons indicated above.

In addition to an appropriate timescale, molecular dynamics simulations must possess a certain level of accuracy in order to be of practical use to scientists. The level of accuracy is determined by defining the equation of atomic motion in a Newtonian (classical) context, a stochastic context, or a context consisting of a mixture of quantum and classical mechanics (QM/MM). Unfortunately, accuracy is inversely proportional to computational expense.

The Newtonian approach is the least expensive method, as it simplifies a molecule by boiling its constituent atoms down to "balls", which are bound to other "balls" via "springs" (bonds). Electrons and their orbitals are not treated explicitly; rather, electronic effects are included in the "ball" concept via parameterization of specific, predetermined atom types in an empirical forcefield. These atom types are created according to the nature of chemical bonds between different elements as they are observed in vitro. Thus, the nature of the bond between elements incorporates the "electronic" aspect of the bonded atoms.

The equation describing the movement of these "balls on springs" is based on Newton's second law of motion:

$$F_i = m_i a_i = m_i \ddot{r}_i \quad (3.1)$$

Where F_i is the force acting on particle i , m_i is the mass of particle i , a_i is its acceleration, and \ddot{r}_i is the second derivative of the particle position with respect to time. Because F_i is also dependent on the gradient of the potential energy function, F_i can alternatively be expressed as:

$$F_i = -\nabla_i U(\mathbf{r}) \quad (3.2)$$

Where $U(\mathbf{r})$ is the potential energy function of particle i .

Although the accuracy of the simulation of electronic effects in molecular systems is limited by using a classical mechanical approach, it has been proven by countless molecular modeling studies spanning two decades to be effective in reproducing experimental phenomena and its shortcomings are known.

In purely Newtonian simulations the system exists in a vacuum, which makes for more simplistic--and hence faster--computing, but biomolecules in vitro/vivo generally exist in a solvated environment. Solvent effects can play a significant role in determining the motion of atoms in the biological system. These forces can be modeled using a Newtonian approach by solvating the system with explicit water molecules. However, explicit solvation drastically decreases computational affordability, and is not always practical, especially when much of the increase in computing is due to (and wasted on) water-water interactions. In an explicit model, the solvent molecules are independent

entities with their own force-variables, just as any atom in the biomolecule. By stripping these molecules of their individuality and reinterpreting them as an overall force, computational expense can be reduced to near- in vacuo conditions, yet solvent effects are still taken into account during the simulation. Thus, by representing solvation effects as a term within the mathematical definition of atomic motion of the biomolecule, one can include solvent effects on the system with reasonable accuracy, but without the expense of explicit solvent. This method, called stochastic dynamics, employs the Langevin equation (equation 1.3, below) to achieve that end:

$$m_i \ddot{r}_i = -\nabla_i U(\mathbf{r}) - m_i \beta_i v_i(t) + R_i(t) \quad (3.3)$$

Where $-\nabla_i U(\mathbf{r})$ is the gradient of the potential energy function, and incorporates classical mechanics (see equation 1.2), $m_i \beta_i v_i(t)$ represents the drag force due to solvent-induced friction on particle i , and $R_i(t)$ is a random force representing the probability of particle i colliding with solvent.

The third method of molecular dynamics is a composite of classical molecular mechanics coupled with a quantum mechanical approach (QM/MM). Unlike the stochastic method and Newtonian mechanics alone, the inclusion of quantum mechanics enables the algorithm to explicitly model electronic orbitals, greatly enhancing the accuracy of the simulation. This is achieved by employing the Schrödinger equation in the context of the overall energy of the system:

$$\hat{H}_{\text{eff}}\Psi(r, R_q, R_c) = E(R_q, R_c)\Psi(r, R_q, R_c) \quad (3.4)$$

Where \hat{H}_{eff} is the effective Hamiltonian for the system, Ψ is the wave function for the system, r is the coordinates of the electrons, R_q is the coordinates of the QM nuclei, and R_c is the coordinates of atoms having solely molecular mechanics parameters.

An explicit-orbital model has the capacity to simulate not only the movement of particles in space, but can also simulate the formation/cleavage of bonds between atoms-- a faculty that the classical and stochastic methods do not possess. Despite its superiority to other methods of modeling, using an exclusively quantum approach when simulating systems larger than 300 atoms becomes prohibitively expensive owing to the complexity of the equation shown in equation 1.4. To circumvent this problem, large compounds (including most biomolecules) are often divided into a quantum mechanical segment and a Newtonian mechanical segment. For example, when modeling an enzyme the quantum mechanical region typically consists of the active site residues thought to be involved in metabolism, the substrate, cofactor, and heme. The remainder of the enzyme is simulated using a classical forcefield instead, as it has little or no impact on the reaction mechanism itself.

To further decrease the overall computational cost, each atom belonging to the quantum region is electronically defined using a semi-empirical method that calculates an atomic electron density function. Algorithms for this purpose include AM1, MNDO, and PM3.

Mitigating the costliness of the QM/MM approach via semi-empirical methods has been successful to some extent, however, modeling large systems -- and

metalloenzyme systems in particular-- remains difficult. It is becoming increasingly feasible, however, with recent developments improving processor speed, and a rising availability as computers containing such processors become more financially affordable.

Regardless which methods specified above is employed, molecular dynamics is a powerful tool in the field of computer aided drug development. Unlike most docking methods, molecular dynamics allows the movement of the receptor as well as the ligand, which increases the likelihood of the ligand finding the correct binding mode(s). In the same vein, interactions responsible for the binding modes may be revealed even down to the electronic level if quantum mechanics are employed. The major drawback regarding the use of MD as a drug developing tool is the very trait that makes it useful over other techniques: time. Because of the sheer number of calculations required to be solved for each atom in the system, and because these calculations must be repeated for each time-step in the simulation, the real-time wait for MD results can be on the order of weeks to months.

3.2. Materials and Methods

3.2.1. Preliminary Structure Preparation

The effector, dapsone, was built with Sybyl (Tripos, St. Louis, MO) and energy minimized. Geometry optimization was performed at the AM1 level with Gaussian03 (Gaussian, Inc., Wallingford, CT). The resulting charges were fitted to Amber force field parameters using the resp algorithm in Antechamber from the Amber8 suite of programs. Dapsone was subsequently placed into the active site of the X-ray crystal structure of

CYP2C9¹² using AutoDock3.0^{79,80} and oriented according to the model determined by NMR³⁸. Flurbiprofen, the substrate, had been co-crystallized with the enzyme and its coordinates were left undisturbed. Using Amber's XLEaP program, the system was then solvated with a pre-equilibrated TIP3P periodic box whose water molecules extended roughly 8 Å from the solute. Chloride counterions were also added to neutralize the net charge, also with XLEaP.

3.2.2. Molecular Mechanics/Molecular Dynamics Simulation

On an SGI Fuel running Irix 6.5, the Sander module was used to minimize the energy of the solvated system, as well as to calculate all other molecular mechanics and molecular dynamics with the Cornell 94⁸¹ force field. The system was kept at a constant pressure for 1000 steps of minimization using a 2 ps timestep to relax the system and relieve steric clashes between non-bonded atoms. Following this, a 3 ps molecular dynamics simulation was run to allow the water molecules to fill areas of vacuum around the solute and come to an equilibrium density. Non-bonded atomic interactions were limited by a 9 Å cutoff and 25 kcal/mol · Å of positional restraints were applied to the solute to hold it fixed in space. Gradually the positional restraints were removed during the next 600 steps of equilibration until the solute was free to move randomly. The final dynamics run heated the system from 100 K to 300 K over 20 ps. Finally, a production run during which the system was held at a constant temperature of 300 K was performed for a 4 ns duration. The changes in the energetics of the system during the dynamics simulations were monitored using the Carnal module of Amber.

3.2.3. Analysis of MD Simulation

The trajectory of the production run was analyzed in MOIL-View⁸². Hydrogen bonding was determined by measuring the distance between the donor and acceptor atoms with a cutoff of 4.0 Å. Pi-stacking was determined by visual assessment of the geometry of the aromatic rings in relation to one another (i.e. sandwich stacked, T-stacked) and by measuring the distance between the closest carbons using a cutoff of 5 Å

74

3.3. Results and Discussion

Over the course of 4 ns, both H-bonding and pi-stacking appear to play important roles with regards to active site substrate positioning in the MD simulation of the dapsone-flurbiprofen-CYP2C9 system (**Figure 3.1**). For the first 1400 ps of the simulation the substrate is bound in the active site via H-bonding of its carboxylate group to the amine side chains of Arg-108 and Asn-204; the former is known via mutagenesis studies to be important in flurbiprofen metabolism by CYP2C9¹⁶. Dapsone is also bound in the active site via H-bonds, which are formed between one of its sulfone oxygens and the amine side chain of Asn-474, and more transient H-bonds are formed via its amine hydrogens to the backbone carbonyl oxygen of Thr-301 and via its sulfone oxygens to the hydroxyl side chain of Ser-209 (**Figure 3.2**). These electrostatic interactions possibly play an important role in the functionality of dapsone as an effector, since the substitution of a sulfide for the sulfone group in compound **8** results in loss of activity as an effector of flurbiprofen metabolism by

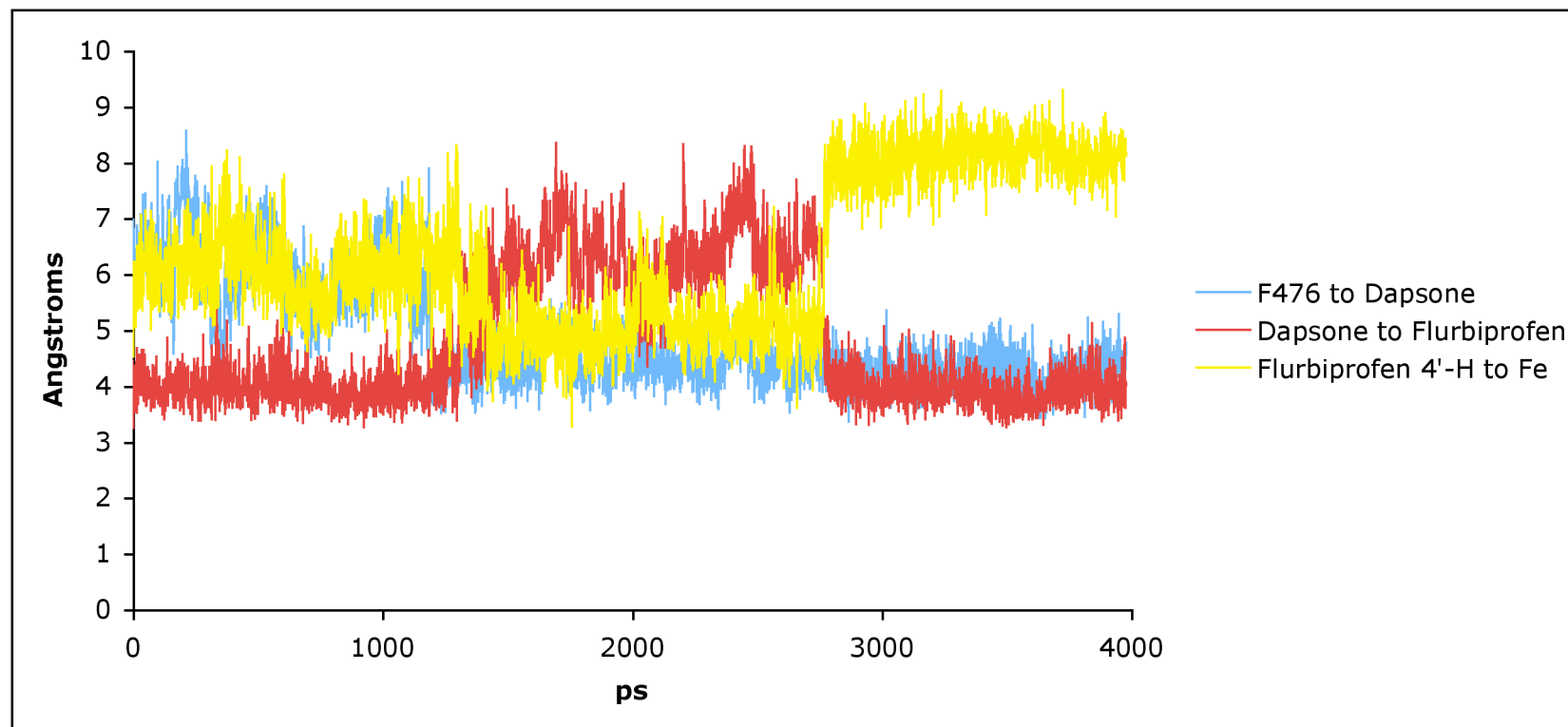


Figure 3.1. Overlaid Distance Plots of Flurbiprofen 4'-H to the Heme Iron In the Presence of Dapsone , Dapsone to F476, and Dapsone to Flurbiprofen.

Overlaid distance plots of flurbiprofen 4'-H to the heme, dapsone to F476, and dapsone to flurbiprofen are shown.

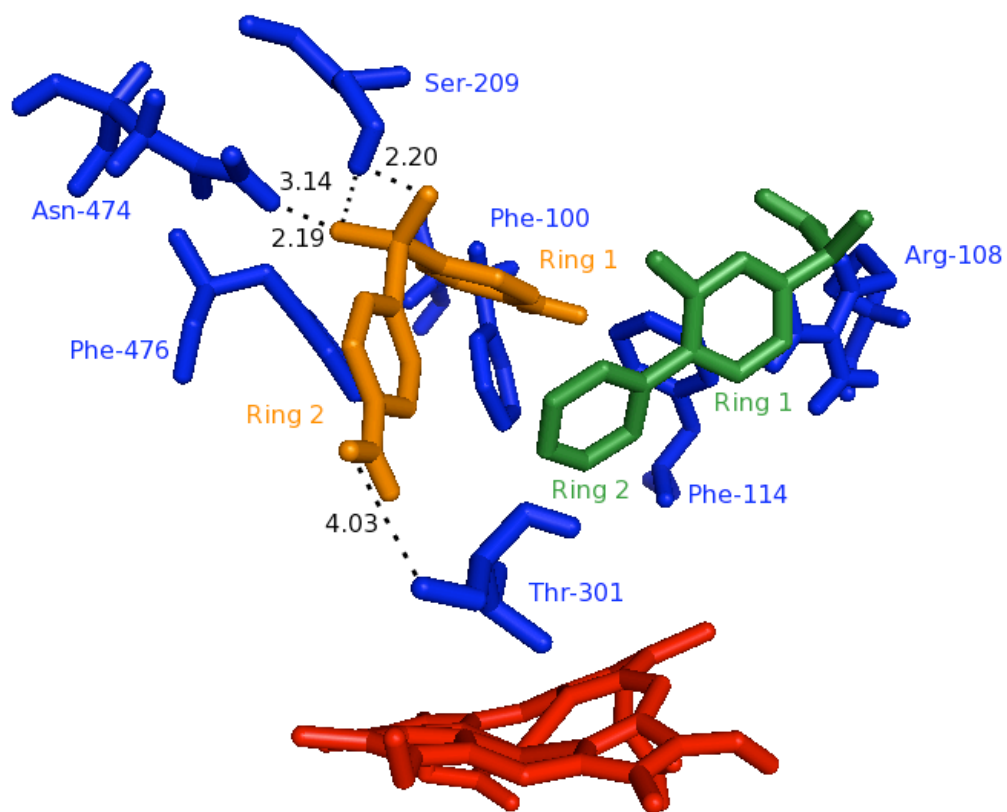


Figure 3.2. H-bonds Formed During MD Between Dapsone and Active Site Residues of CYP2C9 in the Presence of Flurbiprofen.

H-bonds formed between dapsone and active site residues during the MD simulation in the presence of flurbiprofen are illustrated. Dapsone is tangerine, flurbiprofen is forest, active site residues are royal blue, and the heme is crimson. H-bonds are denoted as black dotted lines and are measured in Ångstroms.

CYP2C9, and substitution of the sulfoxide with another H-bond accepting group like the sulfoxide in compounds **9** and **14** retains the activity of these compounds. Dapsone also formed the occasional H-bond with the carboxylate moiety of flurbiprofen, though these interactions were relatively weak, hovering around 4 Å and forming only in 370 frames out of the 4000 frame trajectory. In addition to forming electrostatic interactions, dapsone also appeared to pi-stack with Phe-476 as well as with flurbiprofen.

From 0 to 1300 ps the average distance between Ring 1 of dapsone and Ring 1 of flurbiprofen was 3.99 Å, which is well within the bounds of the cutoff for pi-stacking interactions. However, the average distance between the substrate 4'-H, where oxidation occurs, and the heme iron was 6.12 Å, which made metabolism highly unlikely to occur. Though both rings of flurbiprofen interacted with both rings of dapsone, Ring 2 of the effector did not associate with flurbiprofen as often as Ring 1, which stacked with one ring or another of the substrate in various dimer-like geometries and often with both substrate rings in a trimer-like geometry (**Figure 3.3**). Ring 1 of dapsone also appeared to trimerize with Phe-476 and Ring 1 of flurbiprofen, though at different intervals during the course of the trajectory dapsone favored association with one over the other. Up until the 1300 ps mark the average distance between Ring 1 of dapsone and the phenyl side chain of Phe-476 was 5.96 Å, but at 1300 ps they began to associate more closely, maintaining a distance between 3.5 Å and 5 Å until 2800 ps. This action resulted in the effector distancing itself from the substrate by an average of 6.15 Å over the course of the trajectory from 1300 ps to 2800 ps, and spans a range of distances from about 4 Å to 8.5 Å, which is also shown in the graph in **Figure 3.1**.

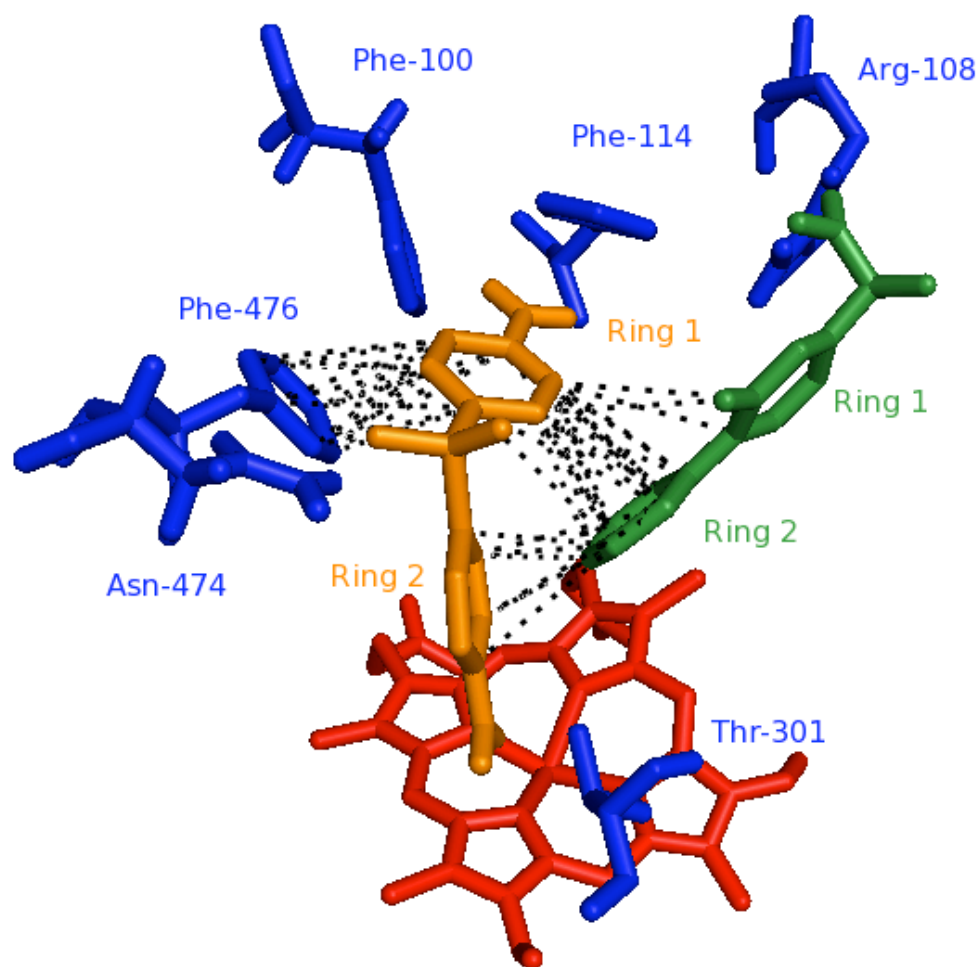


Figure 3.3. Pi-Stacking Interactions Between Dapsone, Phe-476 And Flurbiprofen in the Active Site of CYP2C9 During MD.

Pi-stacking interactions formed during the MD simulation between dapsone, Phe-476 and flurbiprofen in the active site of CYP2C9 are illustrated. Dapsone is colored tangerine, flurbiprofen is forest, active site residues are royal blue, and the heme is crimson. Pi-stacking interactions are denoted by black dotted lines.

Concurrently, the distance between the heme iron and the 4'-H of flurbiprofen shortened by 1 Å, to 5.12 Å, which is in agreement with experimental data obtained from NMR studies that showed the Fe-4'-H distance decreased by 0.9 Å in the presence of dapsone. At times the Fe-4'-H distance reached numbers as low as 3.27 Å (**Figure 3.4**).

Interestingly, at 1400 ps the H-bond between the carboxylate moiety of flurbiprofen and Arg-108 dissolved as the substrate moved closer toward the heme iron, though the substrate was able to maintain H-bonds with Asn-204 throughout the 1400-1800 ps time frame. H-bonding between Arg-108 and the substrate still occurred frequently, but were not constantly sustained. At 2800 ps Ring 1 of dapsone once again came into close proximity of Ring 1 of flurbiprofen and from this point to the end of the 4 ns trajectory the average distance between them was 3.92 Å, indicating a tighter interaction between the two compounds. During this same span of time, dapsone was also able to stack with Phe-476 and in doing so pulled flurbiprofen away from the heme to the point that it was no longer in position to be metabolized. It was at this point that the H-bond between the side chain of Arg-108 and the carboxylate group of the substrate re-solidified.

The fact that a shorter distance between flurbiprofen and dapsone corresponded to a larger distance between the 4'-H and the heme iron gives the appearance that dapsone should in fact be an inhibitor of flurbiprofen metabolism by CYP2C9. However, this is not supported experimentally. It may be that dapsone does actually pull flurbiprofen away from the heme iron, but only when the two compounds form very close interactions. Aromatic active site residues like Phe-476 may, by stacking with the effector, draw away from the substrate at times so that an optimum distance between

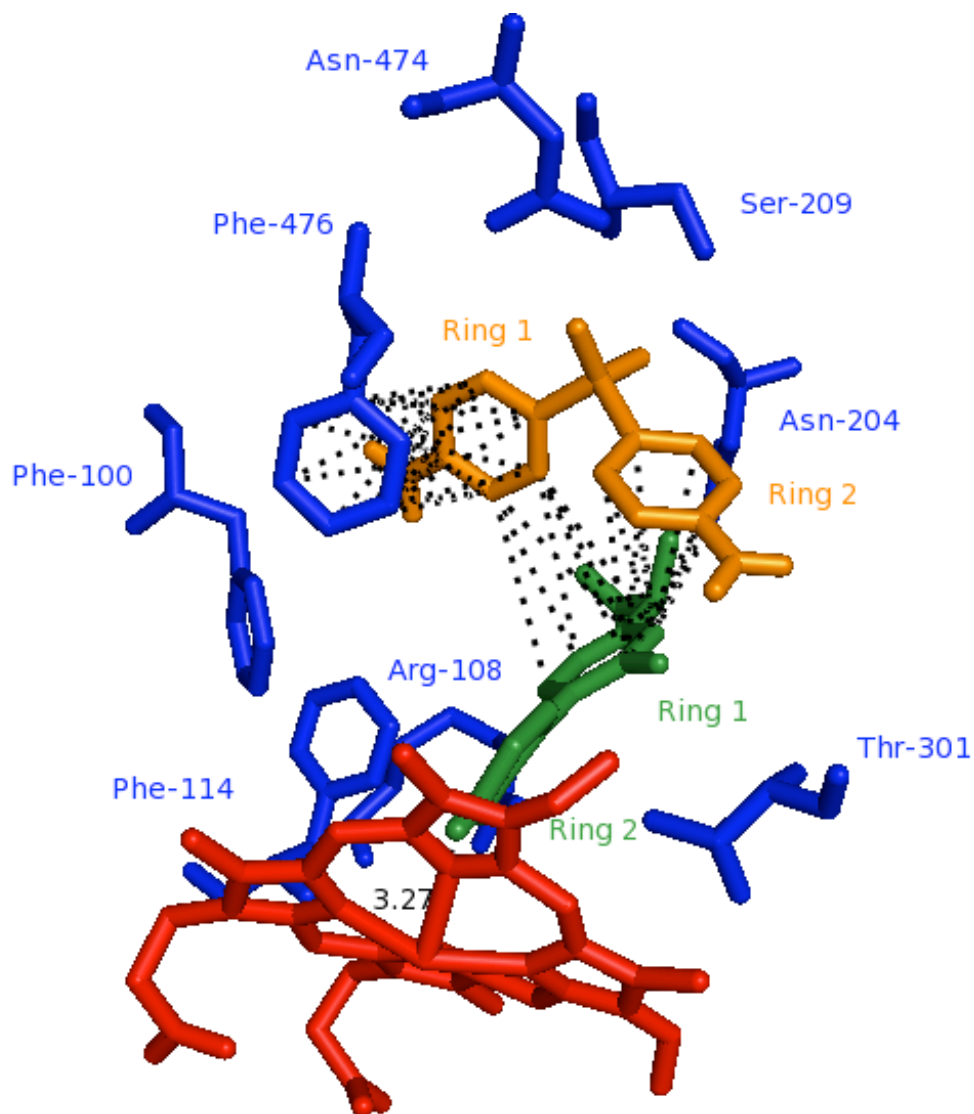


Figure 3.4. Pi-Interactions Formed by Dapsone, Phe-476 and Flurbiprofen When the Substrate 4'-H-Fe Distance Is at Its Shortest During MD.

Pi-interactions formed by dapson, flurbiprofen, and Phe-476 in the active site of CYP2C9 when the distance of the 4'-H of flurbiprofen to the heme iron is at its shortest are shown. Dapsone is colored tangerine, flurbiprofen is forest, active site residues are in blue, and the heme is green. Pi-stacking interactions and the 4'-H-Fe distance measurement are shown as dotted lines with measurements in Ångstroms.

dapsone and flurbiprofen is reached and metabolism can occur. At weaker interactions dapsone may stabilize the substrate in a position favorable to metabolite formation. It must be remembered that binding is a dynamic process and as such, the "closely stacked" and "moderately stacked" states of the flurbiprofen-dapsone complex exist in equilibrium, with the "moderately stacked" state predominating. This is consistent with QM/MM data discussed later in Chapter IV and with docking results obtained from eHiTS in Chapter V.

In contrast to the dapsone-flurbiprofen-enzyme complex, the MD simulation consisting of flurbiprofen alone in the active site of CYP2C9 did not show fluctuations in distance between the heme iron and the 4'-H of the substrate (**Figure 3.5**). In fact, the distance between these two atoms averaged 5.01 Å, which is a slightly shorter distance than that averaged by flurbiprofen in the presence of dapsone. In addition, the shortest distance between the 4'-H of the substrate and the heme iron (3.18 Å) was slightly shorter than that achieved by the 4'-H in the presence of dapsone. However, this does not mean that dapsone plays no role in the activation of flurbiprofen 4'-hydroxylation by CYP2C9. Rather, it could indicate that a longer MD simulation might be necessary to fully gain perspective on the behavior of flurbiprofen in the active site of CYP2C9 with and without dapsone present and the 4 ns time scale analyzed here may be insufficient to observe relevant interactions and substrate behaviors. Also, waters were excluded from the active site itself (though they were explicitly included externally as solvent). Active site waters are known to play a role in the catalytic P450 cycle, and the lack of them here may affect the behavior of flurbiprofen in these simulations. Another factor that could affect the outcome of the MD results is the use of a ferric heme

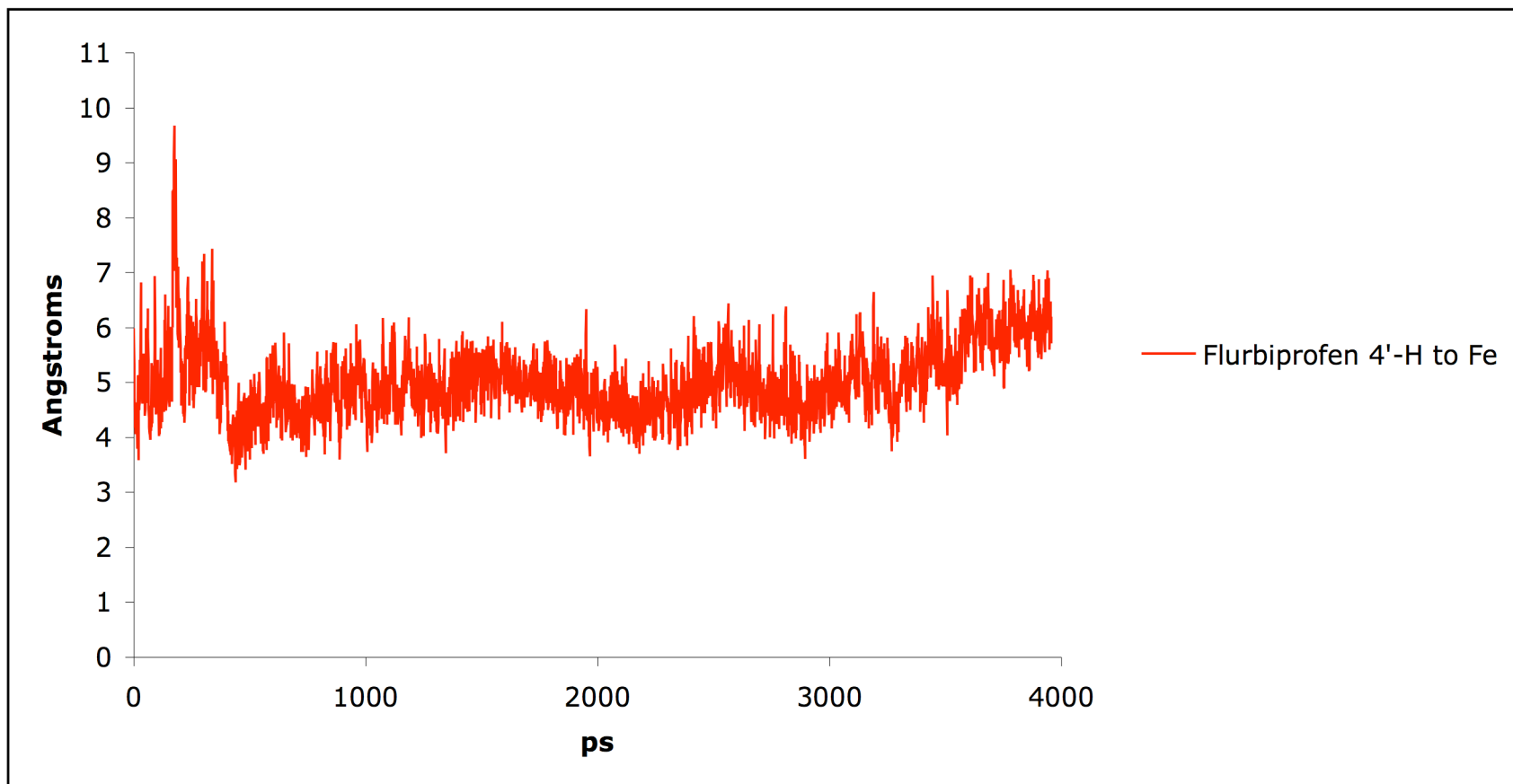


Figure 3.5. Distance Plot of Flurbiprofen 4'-H to the Heme Iron in the Absence of Dapsone.

The distance of the 4'-H of flurbiprofen to the heme iron over the course of the 4 ns MD simulation of CYP2C9-flurbiprofen with dapsone absent from the active site is illustrated.

iron. The most reactive species in the catalytic cycle is the Fe^{IV} oxoheme³, whose presence may cause the substrate to exhibit more accurate binding behavior. Overall, however, the MD simulation itself of the CYP2C9C-flurbiprofen-dapsone complex did illuminate aspects of possible mechanisms of activation and were in agreement with both external experimental data and computational data obtained using other methods in this work.

CHAPTER IV

Determination of Effector-Active Site Binding in

CYP2C9 with AutoDock3.0

4.1. Introduction

As of the writing of this work, many prefer AutoDock for ligand-receptor docking due to its reliability and its relatively low cost⁸³⁻⁸⁶. The AutoDock distribution is actually a suite of programs composed of three complementary units: AutoTors, AutoGrid, and AutoDock. An optional GUI interface called AutoDockTools is available separately. Each of these is described in further detail below.

AutoTors is a program responsible for defining the flexibility of the ligand. First, the ligand is divided based on bond torsion into rotatable "branches" that radiate from a fixed portion of the compound, the "root." Branches can be nested within one another, in which case they are numbered in order from the outermost bond to the root. It is generally assumed that all rotatable bonds will be allowed to move during the docking process, as this most closely mimics compounds in reality, but constraints can be applied to otherwise moveable bonds if the user so chooses. During the docking process, any conformer state that violates such a constraint will be immediately rejected. One instance in which constraints would be needed, however, is in the case of a large ligand that exceeds the maximum moveable torsions allowed by the docking algorithm. This limit can be raised by altering the appropriate header file for the docking program, but a greater number of allowed torsions brings with it a higher computational expenditure. In that regard it may be more economical to apply constraints, although thoughtfully so as to disrupt the natural flexibility of the ligand as little as possible. The output file from AutoTors is then used as the input file for the actual docking process. The program is not

applied to the receptor, as the receptor is held fixed during docking.

AutoGrid is another preparatory program, but applied exclusively to the receptor (also referred to as the "macromolecule"). As with most docking programs, the receptor is held fixed in space in the interest of conserving computational expense. AutoGrid calculates all potential energy interactions between the atom types present in a specified ligand and those in the macromolecule into atom-specific "map files" by referencing a comprehensive database included with the AutoDock suite. This database is composed of van der Waals radii for most atoms in the periodic table, several hundred possible atom-to-atom Leonard-Jones 12-6 potentials, and coefficients peculiar to AutoGrid's energy-calculating algorithm.

In the process of generating the grid maps, the program repeatedly surveys the active site topology and plots these potential interaction energies at finite grid points using a probe atom. The atom types present in the ligand dictate the chemical identity of the probe atom, whose chemical properties change depending on which atom type in the ligand it is representing. In addition to van der Waals interaction energies, electrostatic potential grid maps are also generated; the probe "atom" in this case, however, is actually a charge-- e , $+1.60219 \times 10^{-19}$ C. The spacing of the grid matrix is set by default to 0.375 Ångstroms, which, at roughly one-fourth the length of a carbon-carbon single bond, is fine enough to provide detailed interatomic information while limiting needless calculations. However, the grid point spacing can be made smaller by the user to achieve a higher resolution, or be made greater to reduce further computational cost. The grid boundaries are arbitrarily set by the user, requiring some knowledge of active site residues instrumental in ligand-receptor interaction, or at least the general location of the

binding site. While there is no limit concerning the dimensions of the grid box, a larger box volume will naturally hold a greater number of possible interaction energies, which results in a higher computing cost. By creating topological energy grid maps tailored for only the atom types present in a specific ligand, and also by delimiting docking site boundaries via its grid box, AutoGrid increases the overall efficiency of the docking process.

AutoDock, the namesake of the program suite, actually executes the docking algorithm. There are three main algorithms users can implement in a docking run: Monte Carlo simulated annealing, a genetic algorithm, and a Lamarckian local search algorithm used in conjunction with the genetic algorithm.

In simulated annealing, a ligand randomly orients and re-orientes itself as it samples the conformational space determined by the binding site of a fixed receptor. At each step a small random change is applied to each degree of freedom present in the ligand, including three-dimensional spatial translation and rotation of flexible bonds that alter dihedral angles. The energy of the resulting new ligand configuration is interpolated from the pre-calculated energy potential grid calculated by AutoGrid. If the new energy is more negative than that of the conformation in the previous step (and therefore more stable and likely to occur *in vitro*), then the new configuration is accepted and the process repeats itself to form a new random conformer. If the energy is more positive, however, it undergoes a temperature-dependent probability test to determine the likelihood of the conformer's *in vitro* existence at the user-defined temperature, T . The expression describing this probability is given in equation 4.1:

$$P(\Delta E) = e^{-[\Delta E/(k_B T)]} \quad (4.1)$$

where ΔE is the energy difference between the original ligand configuration and the new one, and k_B is the Boltzmann constant. The higher the temperature, the greater the probability of acceptance of higher-energy structures. Likewise, at lower temperatures the docking algorithm becomes more "stringent," rejecting a greater number of high-energy conformers.

The simulation is initiated at a high temperature, which provides the energy required to dislodge conformer atoms from their local minima. It is critical that the initial temperature be low enough that it is not too energetically permissive, but not so low that the ligand becomes trapped in a local minimum. The temperature is then lowered slowly, allowing new conformers a greater chance of settling into a new minimum with a lower internal energy than their immediate predecessors. The simulation proceeds in multi-step consecutive cycles, each conducted at a specific temperature and each lower in temperature than the last. A new cycle is initiated and the old one terminated after a set number of steps, and the cyclic schedule is determined by equation 4.2:

$$T_i = gT_{i-1} \quad (4.2)$$

where T_i is the temperature at cycle i , and g is a constant between zero and one. The process repeats until the user-defined total number of cycles has been reached. The result is the lowest-energy conformer of the ligand attainable by the algorithm in the allotted

number of cycles. If multiple conformers are desired for comparison, then the algorithm runs N number of "copies" of each temperature cycle to produce N number of resulting ligand conformers. Each copy runs independently of the others, resulting in configurations that are unrelated to and unbiased by one another.

While used extensively for ligand-receptor docking, the simulated annealing algorithm has its drawbacks. For one, it can be very slow, particularly in such cases when the receptor site is large, thus increasing the amount of conformational space to sample, or when docking a complex ligand containing many degrees of freedom. In addition, the reliability of the algorithm depends largely on its ability to find the global energy minimum for any given ligand, which requires fine-tuning of the temperature range in which the algorithm is run. Such refinement can be a lengthy and tedious task to implement. In such cases alternative algorithms are more efficient.

AutoDock is capable of conducting simulated annealing, however, it can alternatively employ a type of genetic algorithm peculiar to the program. This docking method employs a 2-point crossover and random mutation. An optional feature used in conjunction with the default genetic algorithm is a local searcher based on Solis' and Wets' optimization algorithm⁸⁷, which increases the precision of docking results, in that the binding mode corresponds to experimentally determined metabolite formation. In defiance of Mendelian genetics, the local searcher alters the phenotype, which results in a modification of the genotype. Because this approach to mutation entails adapting an individual to its environment and subsequently passing that new trait to its offspring, this algorithm is termed the "Lamarckian genetic algorithm." The Solis and Wets approach to conformational searching involves the determination of the probability of a given

conformer to change translationally or rotationally to a particular state. This variance is 1 for every initial conformer in the population, and is either doubled or halved depending on the number of consecutive free energy drops (successful mutations), or free energy jumps (failed mutations). AutoDock can use the classical Solis and Wets method described above, or implement a so-called "pseudo Solis and Wets" method, in which the variances are scaled to account for their relative magnitudes. In this study we employed the latter.

Regardless of the algorithm used to dock the ligand to the receptor binding site, a scoring function is required to rank the resulting conformers. AutoDock uses scoring function based on the principle of the hydrophobic effect. When ligand and receptor are separate from each other and both fully solvated, water molecules tend to organize around them by H-bonding with other water molecules and also with any H-bond donating or accepting functional groups present at the solute interfaces. When the ligand binds to the receptor, the waters once lining the binding site and also those surrounding the ligand are displaced into the bulk solvent, reducing the overall stabilizing effect of water-solute H-bond formation. AutoDock uses these solvent changes in entropy to estimate the binding energy of the ligand to the receptor, even though docking runs are, for the sake of computational economy, conducted in vacuo. This is possible due to Hess's law of heat summation, which states that the change in free energy between two states will be the same, regardless of the way in which those states were achieved. Therefore equation 4.3 can be used to calculate the binding of a given ligand, L , to a receptor, R , in solvation:

$$G_{binding,solution} = G_{binding,vacuo} + G_{solvation(LR)} - G_{solvation(L+R)} \quad (4.3)$$

Data obtained from the docking simulation itself is used to calculate $G_{binding,vacuo}$, and $G_{solvation(LI)}$ and $G_{solvation(E+I)}$ can be estimated by calculating the surface area of each atom isolated from solvent upon ligand-receptor complex formation and weighting that surface area with an atom-specific solvation parameter. Other energetic components are also taken into account and incorporated into the scoring function, such as loss of degrees of freedom of the ligand upon binding to the receptor, and classical molecular mechanics energy functions such as electrostatic, van der Waals, and H-bonding. Scores are in units of kcal/mol.

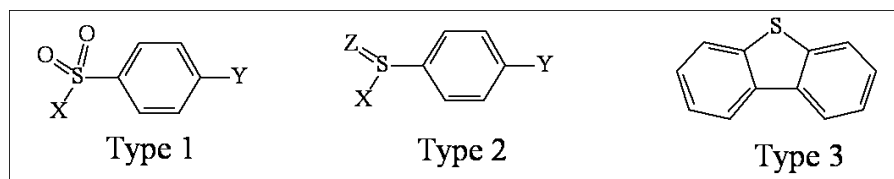
After a docking run is completed and the resulting conformers scored, conformers can be clustered according to RMSd (the cutoff for which is by default 0.5 Å but can be changed to suit the needs of the user). Conformers in the same cluster also tend to score within a small range of energy on the order of tenths of a kcal/mol. When clusters are plotted with the number of conformers per cluster as a function of energy (i.e. score), the resulting histogram provides a clear survey of the distribution of the results.

4.2. Materials and Methods

4.2.1. Preliminary structure preparation and MD simulation

Dapsone and 19 dapsone analogs consisting of the compounds listed in **Table 4.1**

Table 4.1. Structures of Dapsone and Dapsone Analog Effectors.



	Type	X	Y	Z	Name	% Change in Flurbiprofen 4'-Hydroxylation (observed)
1	1	4,6-CH ₃ -C ₄ H ₄ NH ₂ -2-NH-	-NH ₂	-	Sulfamethazine	-18.82
2	1	H ₂ N-	-NH ₂	-	Sulfanilamide	-17.84
3	1	C ₄ H ₄ NH ₂ -NH-	-NH ₂	-	Sulfadiazine	-17.05
4	1	2-CH ₃ -C ₃ H ₃ NO-4-NH-	-NH ₂	-	Sulfamethoxazole	-17.05
5	1	4-NO ₂ -C ₆ H ₄ -	-NH ₂	-	4-(4-nitrophenylsulfonyl)aniline	-10.58
6	3	-	-	-	Dibenzothiophene	-10.58
7	2	C ₆ H ₅	-H	NH	Diphenylsulfimine	-1.18
8	2	C ₆ H ₅	-H	-	Diphenylsulfide	1.18
9	2	C ₆ H ₅	-H	O	Diphenylsulfoxide	3.53
10	1	C ₆ H ₅ -NH-	-H	-	Diphenylsulfone	39.98
11	1	4-CH ₃ -C ₆ H ₄ -NH-	-CH ₃	-	Di-para-tolylsulfone	50.57
12	1	4-NH ₂ -C ₆ H ₄ -	-NH ₂	-	Dapsone	64.09
13	1	4-CH ₃ O-C ₆ H ₄ -	-OCH ₃	-	4,4'-Dimethoxyphenylsulfone	-25.0
14	2	4-CH ₃ CONH- C ₆ H ₄ -	-NHCOCH ₃	O	4,4'-Diacetamidophenylsulfoxide	5.03*
15	1	4-HOOC-C ₆ H ₄ -	-COOH	-	4,4'-Dicarboxyphenylsulfone	-5.0
16	3			-	Dibenzothiophenesulfone	17.82*
17	1	4-HO-C ₆ H ₄ -	OCH ₃	O	4-Hydroxyphenyl-4'-methoxyphenylsulfone	17.92*
18	1	4-CH ₃ CONH- C ₆ H ₄ -	-NHCOCH ₃	-	4,4'-Diacetamidophenylsulfone	8
19	1	4-HO-C ₆ H ₄ -	-OH	-	4,4'-dihydroxyphenylsulfone	95
20	1	4-CH ₃ O-C ₆ H ₄ -	-H	-	4-Methoxyphenylsulfone	44.84*

*non-experimental CoMFA-predicted values (model R² = 0.80) .

Structures of dapsone and the 19 dapsone analogs comprising the training set are shown.

Negative numbers indicate inhibition of flurbiprofen 4'-hydroxylation, whereas positive numbers indicate activation [J.S. Aguilar, unpublished results].

were built with Sybyl (Tripos, St. Louis, MO) and energy minimized. Geometry optimization was performed at the AM1 level with Gaussian 03 (Gaussian, Inc., Wallingford, CT). The effector, dapsone, was oriented in the active site of the X-ray crystal structure of CYP2C9¹² according to the model determined by NMR³⁸. Hexacoordinate heme parameters were obtained from the AMBER parameter database maintained by Richard Bryce at the University of Manchester (http://pharmacy.man.ac.uk/amber/cof/heme_all_inf.html)⁸⁸. Flurbiprofen, the substrate, had been co-crystallized with the enzyme. The flurbiprofen-dapsone-CYP2C9 complex was prepared for molecular dynamics using the Amber suite of programs and simulation were performed, both according to the protocol described in Chapter III Materials and Methods. Dapsone, or dapsone plus flurbiprofen were extracted from the resulting structure creating the active site used by the docking software. All dapsone analogs were docked into two active site structures: one with flurbiprofen present, and one without. Because *in vitro* NMR and metabolite data were known for dapsone and flurbiprofen³⁴, these two compounds were docked first, before docking the dapsone analogs, to ensure the quality of the model's parameterization; the model was considered poor and would require re-parameterization if the resulting docked poses of dapsone and flurbiprofen failed to conform to the poses established *in vitro*.

4.2.2. AutoDock Screening

The enzyme was imported into the GUI interface AutoDockTools, and an Amber-derived force field was applied. Non-polar hydrogens were merged, but polar hydrogens were allowed to remain explicit. Kollman united atom charges were then calculated and

added to the protein. Minimized dapsone analogs were also imported into AutoDockTools, merging non-polar hydrogens, and assigning Gasteiger charges for each compound. Compounds containing acidic functional groups were protonated or deprotonated depending on their respective pK_a values and a pH of 7.4. The active site boundaries for grid point calculations were defined by a 40 Å³ box centered on the fluorine of flurbiprofen (or on the corresponding coordinates for the empty enzyme). Dapsone and 19 dapsone analog effectors that had been used as a test set in a previous CoMFA analysis [J.S. Aguilar, unpublished results] (**Table 4.1**) were then docked into the active site of CYP2C9-flurbiprofen or CYP2C9 alone using a genetic algorithm. Flurbiprofen itself was also docked into the empty active site of CYP2C9. The resulting 50 docked conformations were scored using AutoDock's native scoring function and based on the lowest docking energy. Docking energy is defined as the sum of the final intermolecular energy and the final internal energy of the effector of each docked conformation of every effector.

4.2.3. In Silico- In Vitro/CoMFA-Predicted Data Correlation Evaluation

The correlation coefficient of determination relating the in vitro effector-mediated change in flurbiprofen metabolism and the docking results was calculated using Microsoft Excel. The r-squared value was calculated using three sets of data: 1) the lowest docking energy for each effector docked in the presence of flurbiprofen; 2) by the docking energy of the binding mode, where the greatest number of docked poses with an interconformer RMSd of 0.5 Å or less was judged to be either "productive" (i.e. causing activation) or "nonproductive". An effector was considered to be productive if it formed

favorable nonbonded interactions with flurbiprofen, which had been crystallized in an orientation that promoted 4'-hydroxyflurbiprofen formation. Favorable interactions were determined in part by the pharmacophore produced by an earlier CoMFA [J.S. Aguilar, unpublished results], and included H-bonding (determined by a distance cutoff of 4.0 Å between donor and acceptor atoms), and pi-interactions (determined by a distance cutoff of 5 Å between aromatic ring carbons⁷⁴); 3) a tally of the number of conformers comprising the binding mode was taken for each effector, and correlation coefficients (R^2) were calculated from both the number of conformers comprising the mode and the score corresponding to it. If the results for a particular effector were bimodal, the lowest energy mode was chosen for analysis.

4.2.4. Quantum Mechanical Analysis of Docked Effector-Substrate Interaction Energy

Flurbiprofen along with the binding mode of seven docked effectors were cut from the ligand-receptor complexes. Using HyperChem (Hypercube, Inc., Gainsville, FL) at the PM3 level of theory, energies of the substrate and effectors were calculated, as well as the substrate-effector complexes. This protocol served to determine whether the effector and/or substrate was in a more energetically favorable state alone or complexed with one another.

4.2.5. Visual Analysis

Interatomic measurement was conducted in PyMol (DeLano Scientific, Palo Alto, CA). Visualization of the results was performed in PyMol and AutoDockTools3.0.

4.3. Results and Discussion

In the initial docking runs, the heme iron was parameterized with the VDW term and Lennard-Jones potentials multiplied by a large scaling coefficient. Under these conditions, and in the absence of flurbiprofen, dapsone preferred to dock in a nonproductive conformation with sulfone oxygens pointing toward the heme iron (**Figure 4.1**). Flurbiprofen, when docked into the empty active site assumed a pose in which its carboxylate group associated closely to the heme iron; as in the case with dapsone, this conformation could not result in metabolism to the known metabolite (**Figure 4.2**). Knowing that the heme iron may be attracting highly electron-dense groups on the ligands via electrostatic effects, compounds **19** (formal charge -1) (**Figure 4.3**) and **15** (formal charge -2) (**Figure 4.4**) were docked in the empty active site to observe their conformations. In keeping with results seen with dapsone and flurbiprofen, the deprotonated oxygens of these two compounds oriented their negatively charged oxygens toward the heme iron. Interestingly, the docking energy of **15** was nearly an order of magnitude lower than that of **19**. The reason for this phenomenon is not known. Compound **15**, on the other hand, is a dianion and could form favorable H-bonding with Arg-108, thus driving the docking energy lower. This is not a likely scenario, however, due to the close proximity of the effector carboxylate group with that of flurbiprofen, which is energetically unfavorable. Because dapsone docked in a conformation that conflicted with previously published work¹², and because the docking energy of **15** was so extreme compared to similarly-structured **19**, it was decided to first dock these three

compounds in CYP2C9 in the presence of flurbiprofen for comparison before docking additional dapsone analogs in the empty active site.

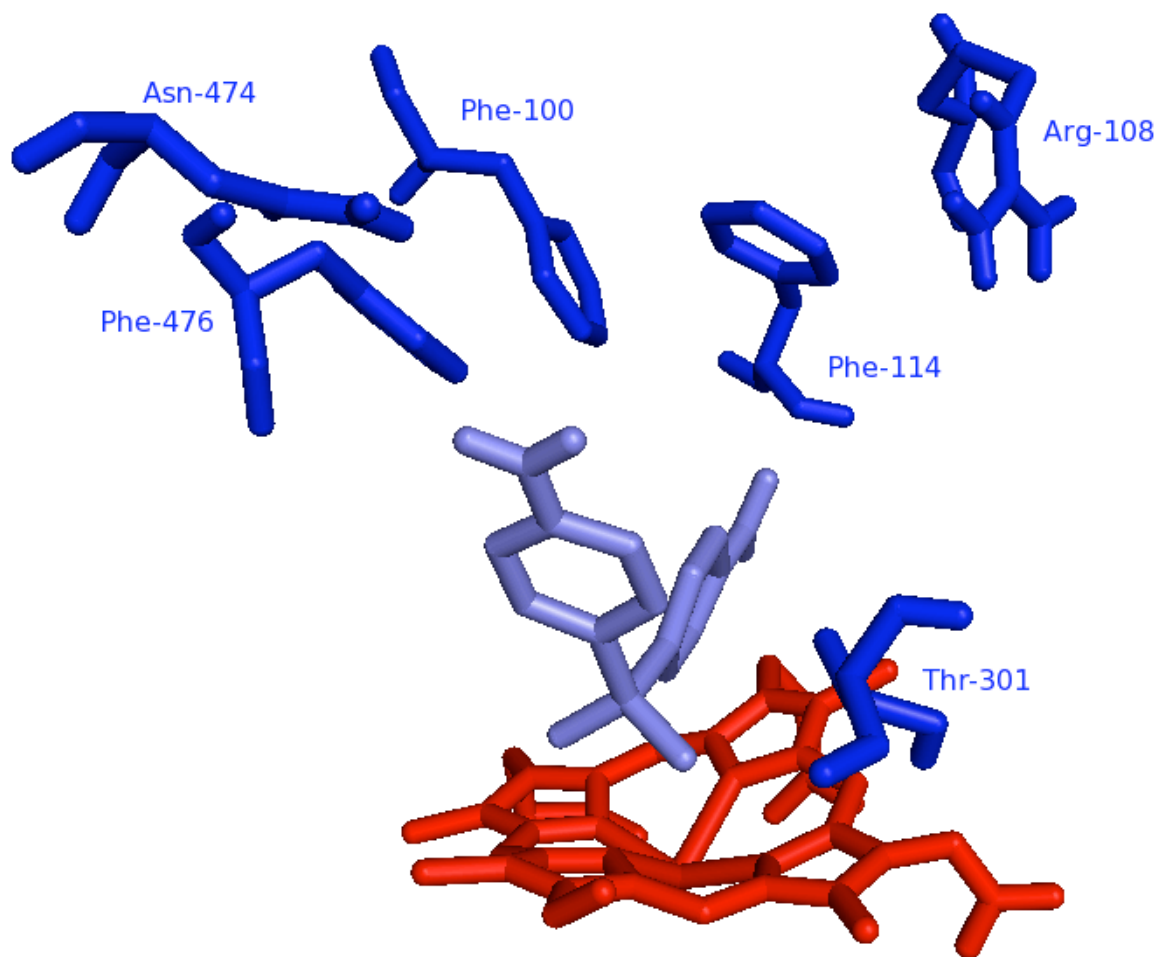


Figure 4.1. Initial Docking of Dapsone into CYP2C9 Empty Active Site by AutoDock.

Initial docking results of dapsone into the empty active site of CYP2C9 by AutoDock are shown. Active site residues are shown in royal blue, dapsone is periwinkle, and the heme is crimson.

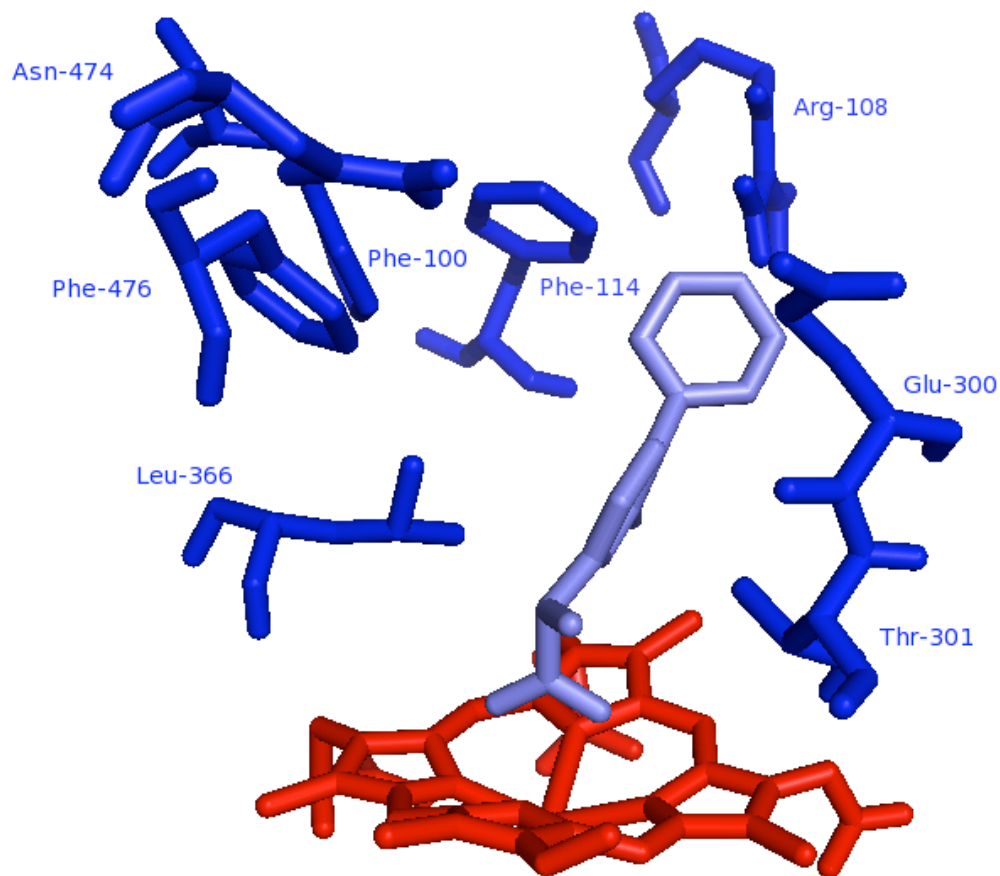


Figure 4.2. Initial Docking of Flurbiprofen into CYP2C9 Empty Active Site by AutoDock.

Initial docking results of flurbiprofen into the active site of CYCP2C9 by AutoDock are illustrated. Active site residues are royal blue, flurbiprofen is periwinkle and the heme is crimson.

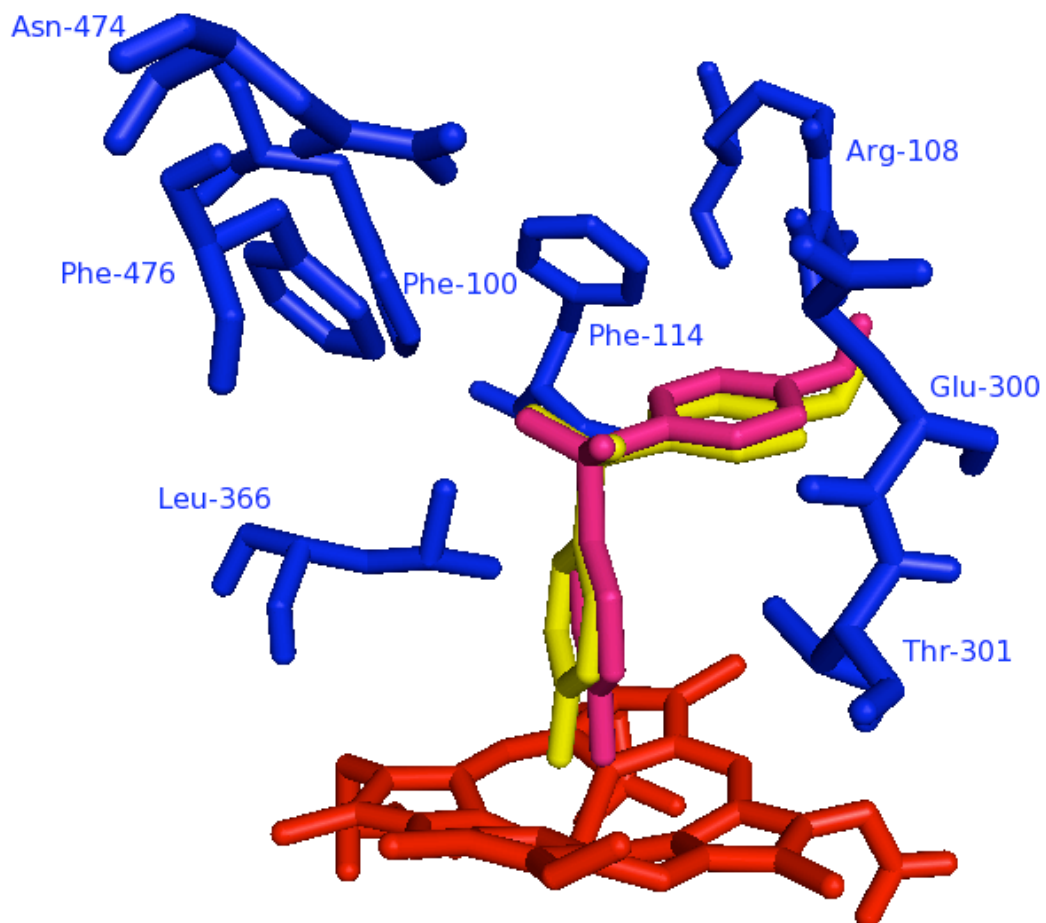


Figure 4.3. Initial Docking of compound 19 into CYP2C9 Empty Active Site by AutoDock.

Initial docking results of **19** into the empty active site of CYP2C9 as calculated by AutoDock are illustrated. The lowest energy conformer is displayed in yellow, the mode is magenta, active site residues are royal blue, and the heme is crimson.

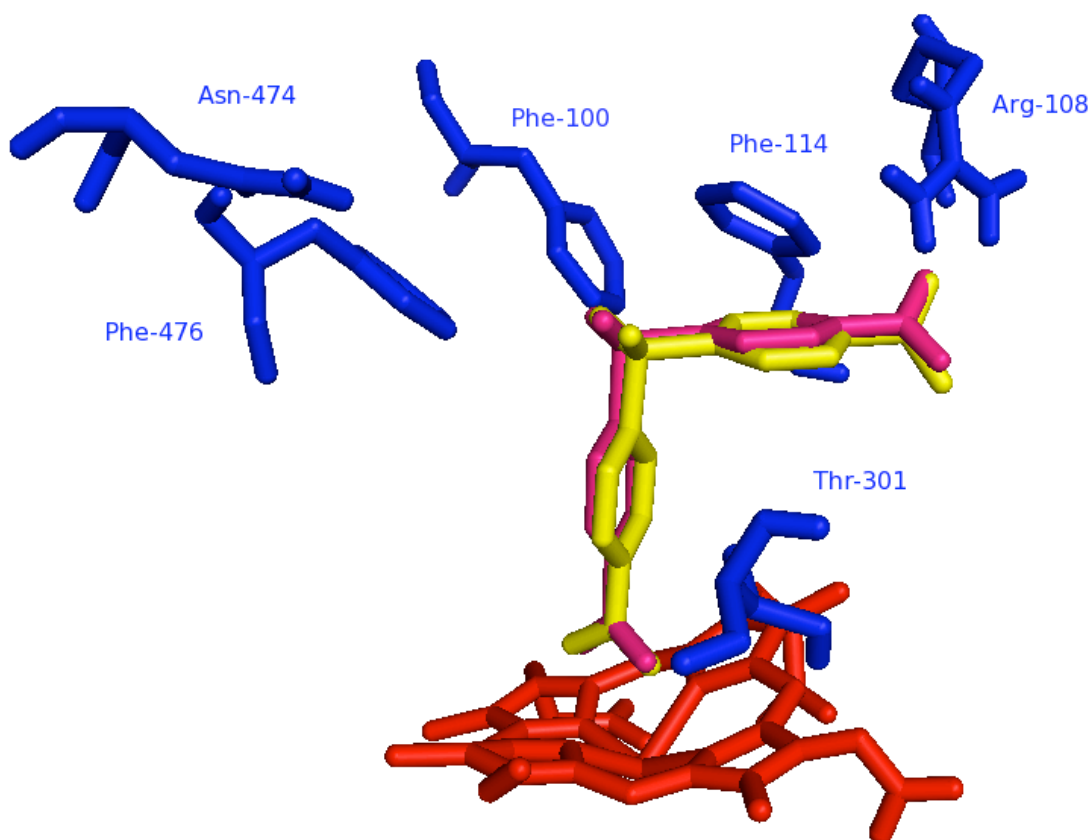


Figure 4.4. Initial Docking of Compound 15 into CYP2C9 Empty Active Site by AutoDock.

Initial docking results of **15** into the empty active site of CYP2C9 as calculated by AutoDock are illustrated. The lowest energy conformer is shown in yellow, the mode is magenta, active site residues are royal blue, and the heme is crimson.

When dapsonsone docked into the active site in the presence of flurbiprofen (**Figure 4.5**), the simultaneous presence of substrate in the active site largely mitigated this phenomenon. The same effect of flurbiprofen on the docking of dapsonsone applied to compounds **15** (**Figure 4.6**) and **19** (**Figure 4.7**). However, it is inconclusive as to whether this was due to some stabilizing interaction such as pi-stacking between effector and flurbiprofen (though in the case of **19** distances of less than 5 Å between their respective ring carbons do suggest this), or the substrate shielding the iron's charge, or more a matter of steric hindrance. Also possible are effector-active site interactions, such as stacking of both **19** and **15** with Phe-476 and H-bond formation between the sulfone oxygens of **15** to the amino side chain of Asn-474. The docking of dapsonsone produced two conformers: a lowest energy conformer and a mode conformer, unlike compounds **15** and **19**. The lowest energy conformer appeared to form stabilizing interactions with flurbiprofen in the way of pi-stacking, whereas the mode conformer formed interactions between active site residues: pi-stacking with Phe-476 and H-bonding with the side chain of Asn-474. The mode conformer does, however, form one interaction with the substrate by H-bonding to the carboxylate moiety via its amine group.

These questions were not pursued, since soon thereafter the error in the parameterization of the heme iron was identified. The VDW and Lennard-Jones values that had been miscalculated in the initial docking runs described above were re-calculated using the correct scaling coefficient, and in subsequent dockings into the empty enzyme dapsonsone bound in a conformation conducive to 4'-hydroxy metabolite formation (**Figure 4.8**). In this pose the lowest energy conformer differed very little from the that of the mode (RMSd = 1.27 Å) and the sulfone oxygens pointed away from the heme. Since this

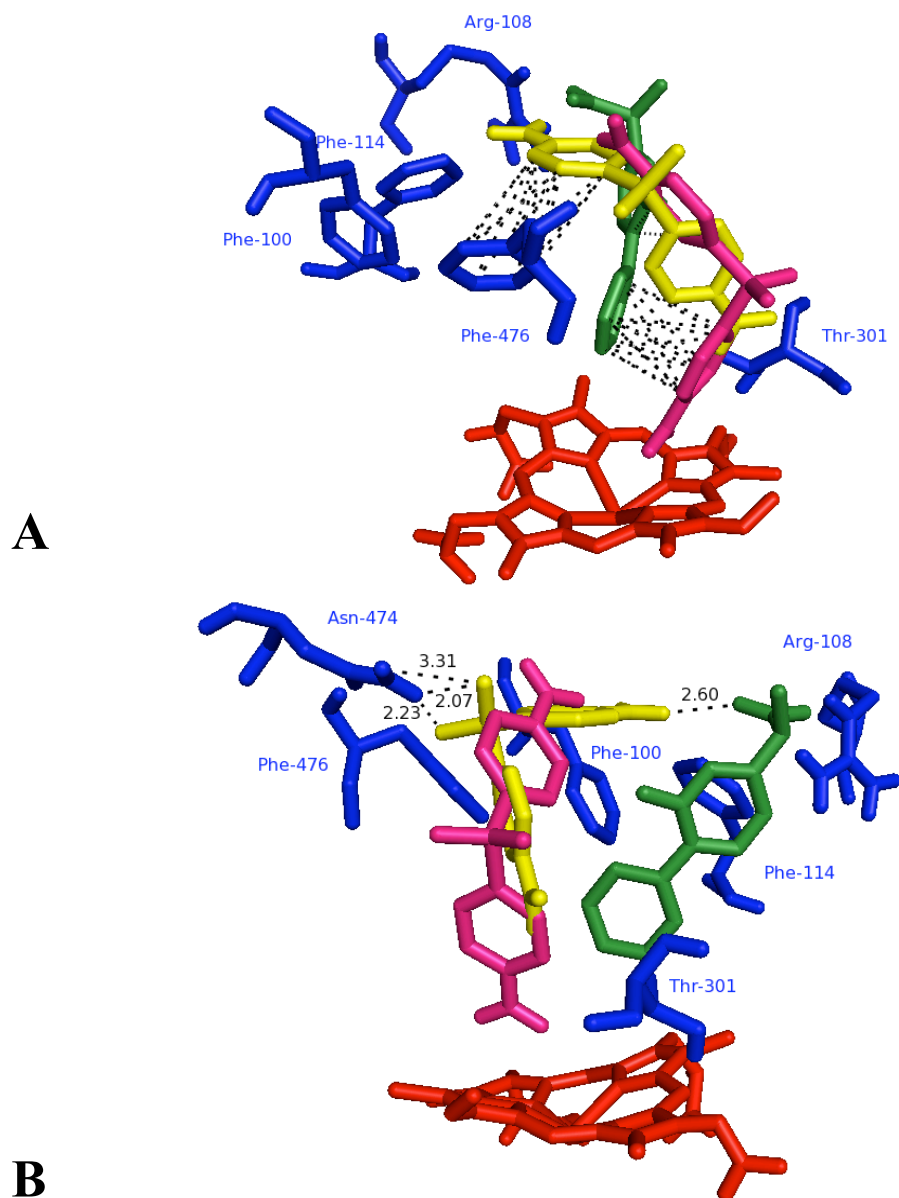


Figure 4.5. Initial Docking of Dapsone in CYP2C9 in the Presence of Flurbiprofen by AutoDock.

Initial docking of dapsone in the active site of CYP2C9 simultaneously occupied by flurbiprofen is shown. Pi-stacking (**a**) and H-bonding (**b**) interactions are denoted by dotted lines. Distances of H-bonds are in Ångströms. The lowest energy conformer is shown in yellow, the mode is magenta, flurbiprofen is forest, active site residues are royal blue, and the heme is crimson.

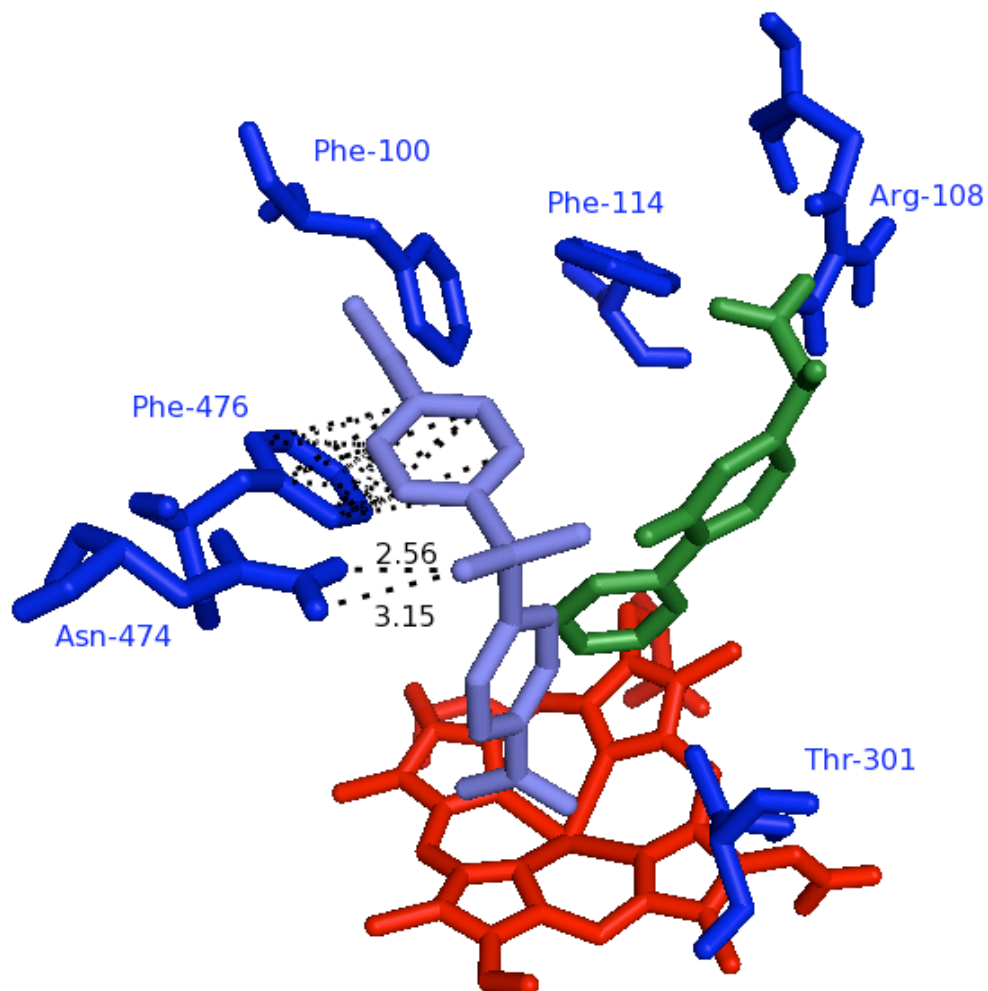


Figure 4.6. Initial Docking of Compound 15 in CYP2C9 in the Presence of Flurbiprofen by AutoDock.

Initial docking of **15** into the active site of CYP2C9 in the presence of flurbiprofen is illustrated. Compound **15** is colored periwinkle, active site residues are royal blue, flurbiprofen is forest, and the heme is crimson. Pi-stacking and H-bonding interactions are shown as dotted lines. Distances of H-bonds are in Ångstroms.

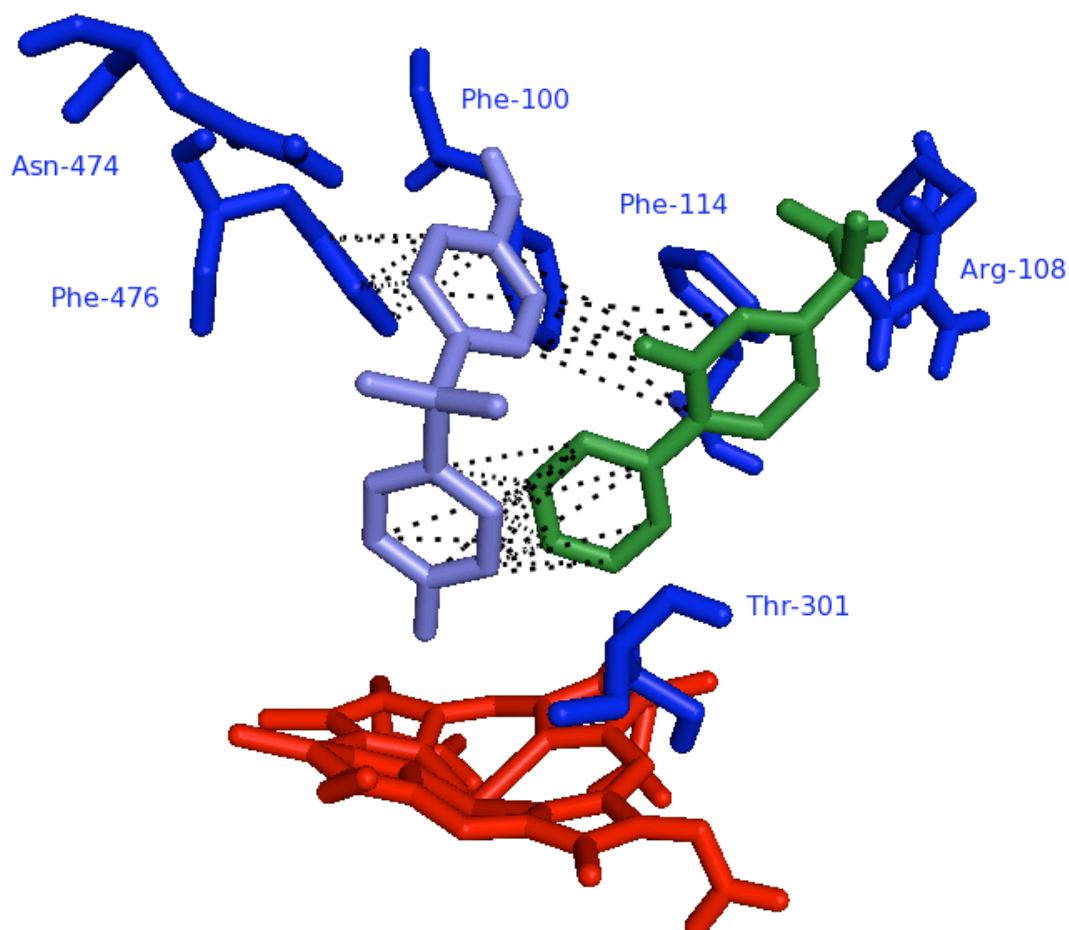


Figure 4.7. Initial Docking of Compound 19 into CYP2C9 Active Site in the Presence of Flurbiprofen by AutoDock.

Compound **19** initially docked into the active site of CYP2C9 in the presence of flurbiprofen is shown. Compound **19** is colored periwinkle, active site residues are royal blue, flurbiprofen is forest, and the heme is crimson. Pi-stacking interactions are denoted as dotted lines.

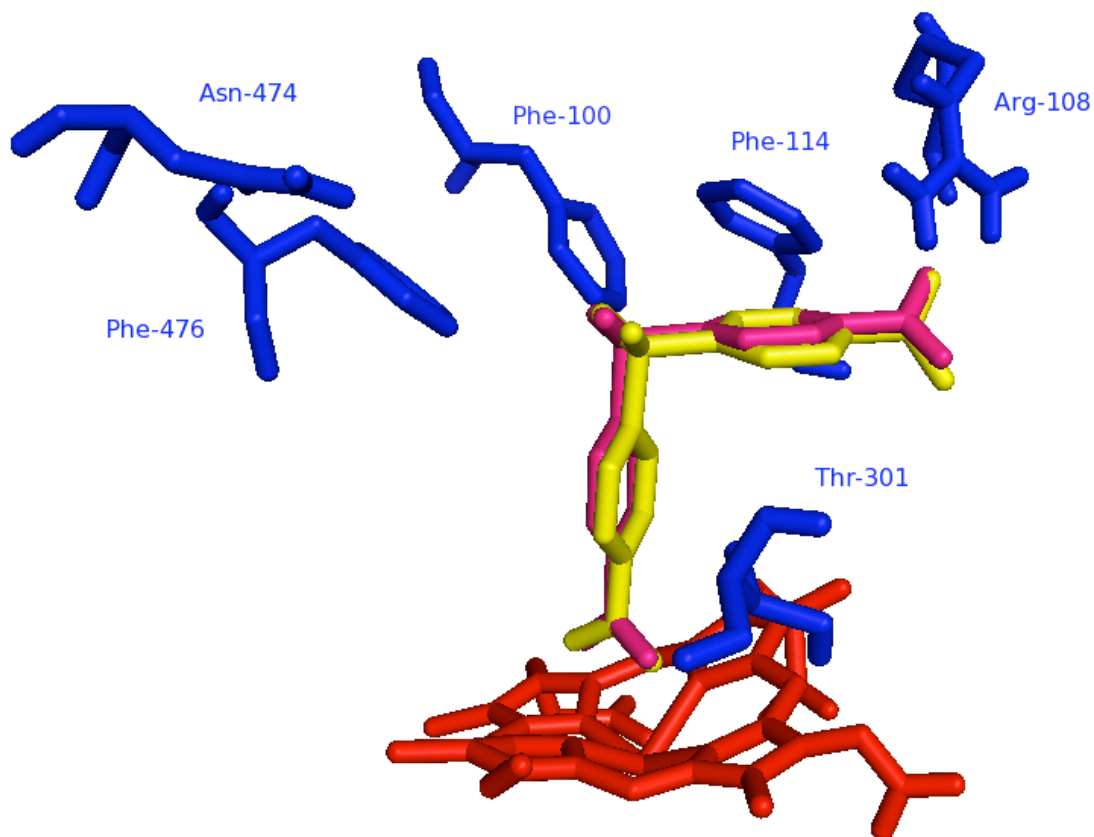


Figure 4.8. Dapsone Docked into the Empty Active Site of CYP2C9 by AutoDock Using Recalculated van der Waals and Lennard-Jones Parameters.

The docked conformation of dapsone in the empty active site of CYP2C9 using AutoDock with recalculated van der Waals and Lennard-Jones parameters is illustrated. Dapsone is shown in periwinkle, active site residues are royal blue, and the heme is crimson.

conformation orients the compound in such a way as to favor its catalysis by CYP2C9 to the primary metabolite, N-hydroxydapsone, the "sulfone up" pose is considered to be the "correct" orientation of dapsone within the CYP2C9 active site.

After correcting the scaling coefficient, flurbiprofen was also docked into the empty active site of CYP2C9 (**Figure 4.9**). Interestingly, its carboxylate moiety preferred to associate via hydrogen bonds to the side chain amino group of Asn-474 rather than Arg-108-- a residue to which binding is thought to be critical for flurbiprofen metabolism to occur. In this binding orientation, the distance between flurbiprofen's carboxylate group and the closest amine group of Arg-108 was 11.52 Å. Because the 4'-hydrogen was located 4.67 Å away from the heme iron, this was considered a nonproductive pose. These results suggest that the nonproductive binding mode is dominant over the productive one, however, it does not indicate that metabolism does not occur at all, or even that it occurs rarely. It possibly indicated that overall flurbiprofen metabolism may not be as efficient as it could be, but before that conclusion can be made one must take into account the fact that many factors contribute to metabolic efficiency that were not included in this model, such as the role of active site waters and reductase-enzyme electron transport.

Docking energies of dapsone and its analogs were fitted to a linear function to ascertain the degree of correlation between AutoDock's scores and the *in vitro*/CoMFA-predicted percent change in flurbiprofen metabolism. The result was that no correlation could be found between the docking energies of the dapsone analog effectors and the actual effector-mediated change in flurbiprofen catalysis as determined by the model containing the corrected heme parameters ($R^2 = 0.009$). However, based on the docked

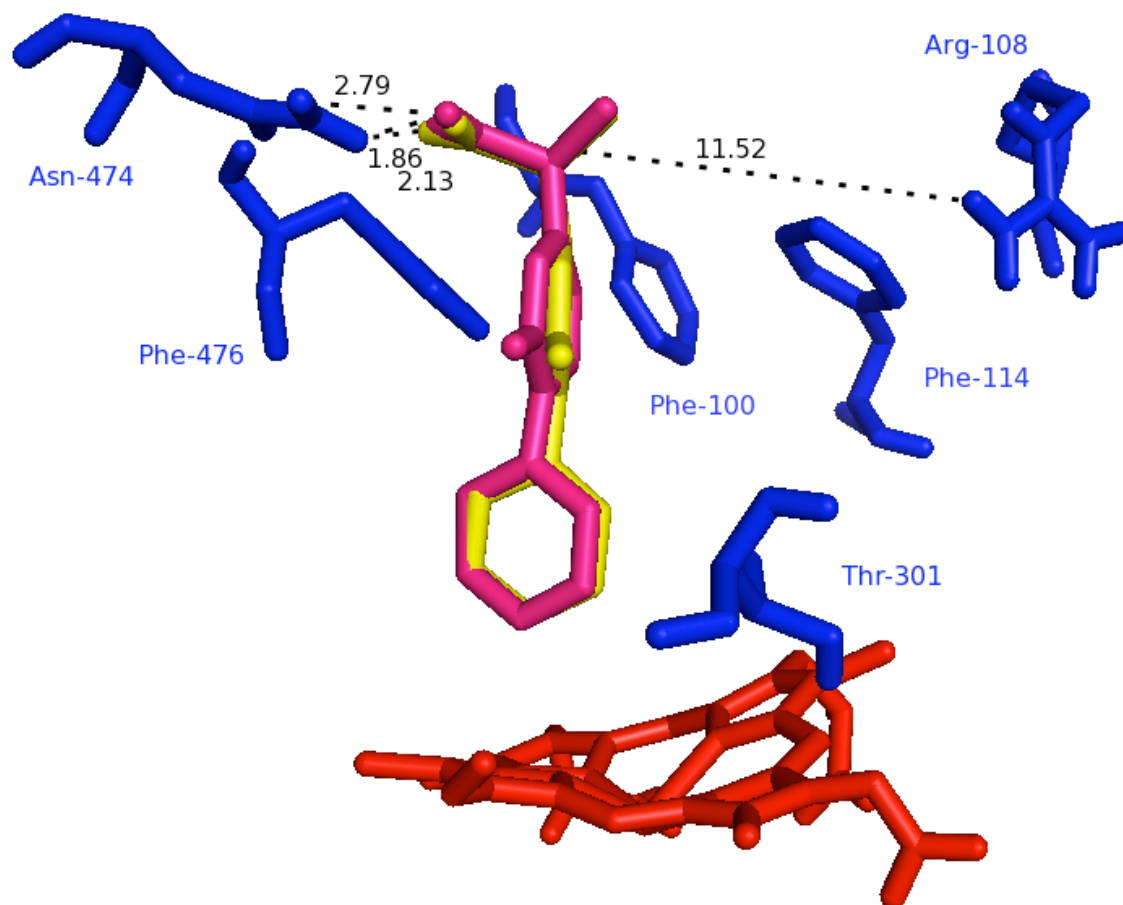


Figure 4.9. Flurbiprofen Docked into the Empty Active Site of CYP2C9 by AutoDock Using Recalculated van der Waals and Lennard-Jones Parameters.

Results of the docking of flurbiprofen into the empty active site of CYP2C9 by AutoDock using recalculated van der Waals and Lennard-Jones parameters are displayed. The lowest energy conformer is shown in yellow, the mode is magenta, active site residues are royal blue, and the heme is crimson. H-bonding interactions are shown for the mode only for clarity, and are denoted by dotted lines. Distances are in Ångstroms.

poses assumed by the dapsons effectors described above, it was postulated that the lack of in-vitro correlation with the docking energies may be due more in part to some parameterization error in the enzyme, rather than some inherent flaw in the program itself. Of particular concern was the heme, whose Fe+3 moiety is notoriously difficult to parameterize correctly using classical mechanical forcefields. Even after the VDW and Lennard-Jones terms were recalculated, it was discovered that AutoDock had not assigned solvation parameters nor partial atomic charges to any atoms in the heme moiety. Partial charges were included in the heme parameter files obtained from the AMBER parameter database, but were being overwritten by AutoDockTools in the course of pre-docking file preparation. A more recent set of heme forcefield parameters containing partial atomic charges differing from those used previously were applied⁸⁹ and solvation parameters were added to the receptor file manually by atom type. Afterward, the entire ligand set was re-docked using the same methods described above. The resulting poses found for flurbiprofen (**Figure 4.10**) and dapsons (**Figure 4.11**) were similar to those seen in the system without heme solvation parameters and partial charges. Despite the implementation of these new parameters, no correlation could be determined from the new docking results, using either the lowest scoring conformer of each effector ($R^2 = 0.03$) or the score corresponding to the mode ($R^2 = 0.01$), or using the mode conformer tally ($R^2 = 0.03$)

The difficulty in applying correct force field parameters to transition metals is ubiquitous in the field of molecular modeling. Our model may have been especially difficult to correctly assign hexacoordinate parameters to; given that docking different effectors may have displaced active site waters in different ways, those waters were

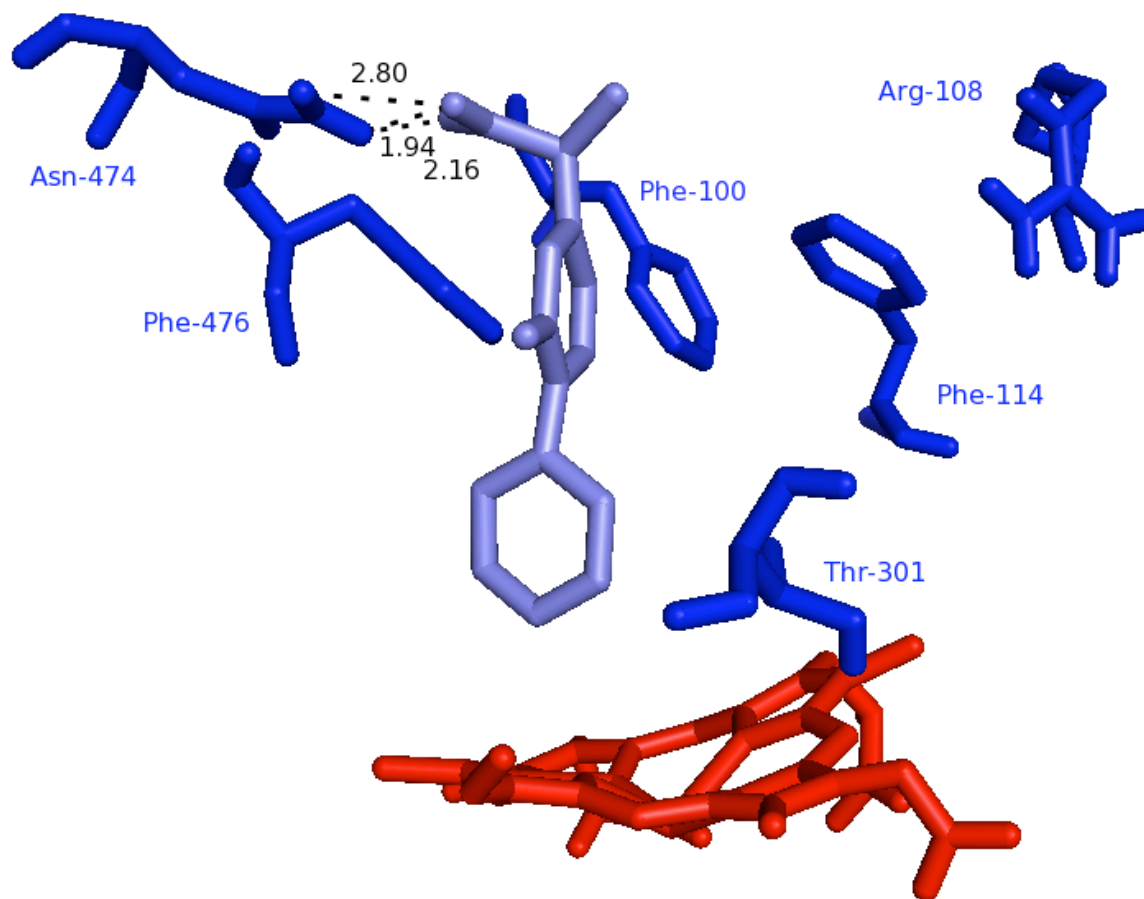


Figure 4.10. Flurbiprofen Docked into the Empty Active Site of CYP2C9 by AutoDock Using Re-assigned Heme Partial Charges.

The docked conformer of flurbiprofen as calculated by AutoDock after re-assigning partial charges to the heme is shown. Flurbiprofen is colored periwinkle, active site residues are royal blue, and the heme is crimson. H-bonding interactions are displayed as dotted lines, with distances in Ångstroms.

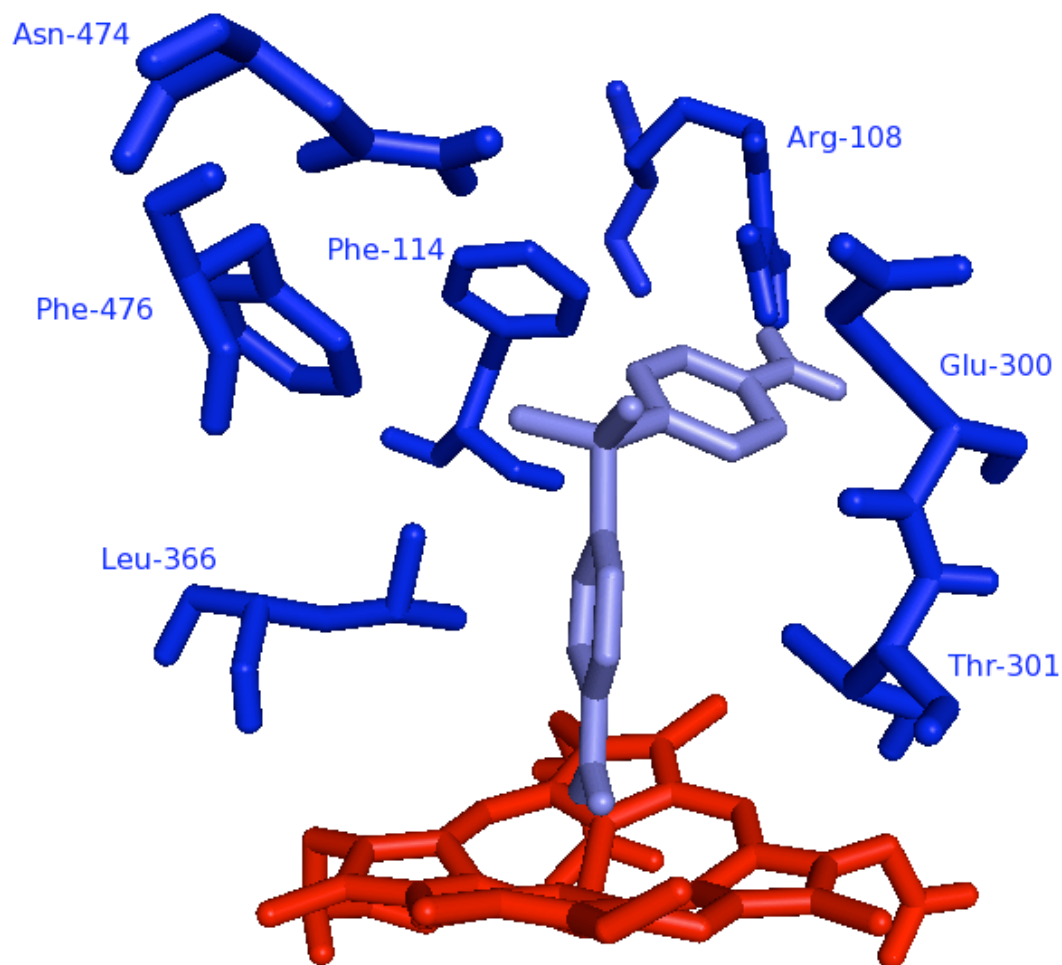


Figure 4.11. Dapsone Docked into the Empty Active Site of CYP2C9 by AutoDock Using Re-assigned Heme Partial Charges.

The docked conformer of dapsone as calculated by AutoDock after re-assigning partial charges to the heme is shown. Dapsone is colored periwinkle, active site residues are royal blue, and the heme is crimson.

removed from the active site to maintain consistency in docking. Among these was WAT600, which was the sixth ligand coordinated to the heme iron in the original 1R9O crystal structure.

Also adding to the difficulty in acquiring a solid *in silico-in vitro*/CoMFA-predicted correlation is the fact that the binding mode can be interpreted in two different ways. There is the option of perceiving the RMS cluster histogram as a distribution of equally viable poses, or as a kind of confidence test in which the mode is the conformation most likely to exist *in vitro*, and conformers in smaller clusters therefore deemed less likely to exist. If the results are taken as a survey of all possible conformers, then all RMSd-derived clusters with all of their respective conformers would need to be analyzed. This approach was not taken here, due to the precedence set in the literature^{12, 86, 90-92} to draw analysis from only the binding mode and/or the lowest scored conformer in the case of bimodality.

Scoring functions are difficult to implement successfully, and it has been shown^{91, 92} that binding modes are more indicative of the true binding orientation even though the binding mode often does not have the lowest docking energy in a given run. Therefore, in addition to data calculated by and obtained from AutoDock itself, docked effector binding mode-CYP2C9 complexes were subjected to semi-empirical QM analysis as a way to determine the accuracy of the docked results while bypassing AutoDock's scoring function. PM3 level of theory allows for a more explicit evaluation of electronic effects such as pi-stacking, which were of particular interest in the case of flurbiprofen-effector interactions given the presence and orientation of aromatic ring systems. Of the seven effectors chosen for analysis, five yielded a linear regression

correlation coefficient of 0.96 (**Figure 4.12**). Because the effectors were docked into a receptor in which flurbiprofen assumed a pose conducive to metabolism, any factor involved in orienting the effector in the active site is therefore also involved in heteroactivation. Between the PM3 data and visual assessment of ring orientation, the four activators lying along the trendline likely stabilize the substrate into a productive position via pi-stacking. One effector, **6**, had a destabilizing effect on flurbiprofen, which is consistent with its role as an inhibitor. The outliers, **19** and **13**, both contain hydrogen bond accepting groups, suggesting that these two compounds may exert their influence on flurbiprofen metabolism through interactions other than pi-stacking. Quantum mechanics data are generally accepted as more convincing than classical mechanics-based energy calculations such as those employed in docking programs⁹³, although, because the QM calculations were conducted in a vacuum, rather than in the active site of CYP2C9, active site-effector interactions that may in fact override effector-substrate interactions are not accounted for and therefore cannot be ruled out. Nevertheless, the data here suggest that AutoDock, in which electronic effects are implied in its forcefield, may not sufficiently or correctly account for them in its scoring function. The possibility also exists that the scoring function lacks a pi-stacking term entirely.

In conclusion, AutoDock provided credible effector binding poses, but a sufficient correlation between its results and the *in vitro*/CoMFA-predicted data could not be found. Therefore a decision was made to try a different docking program, eHiTS, and repeat the effector screening.

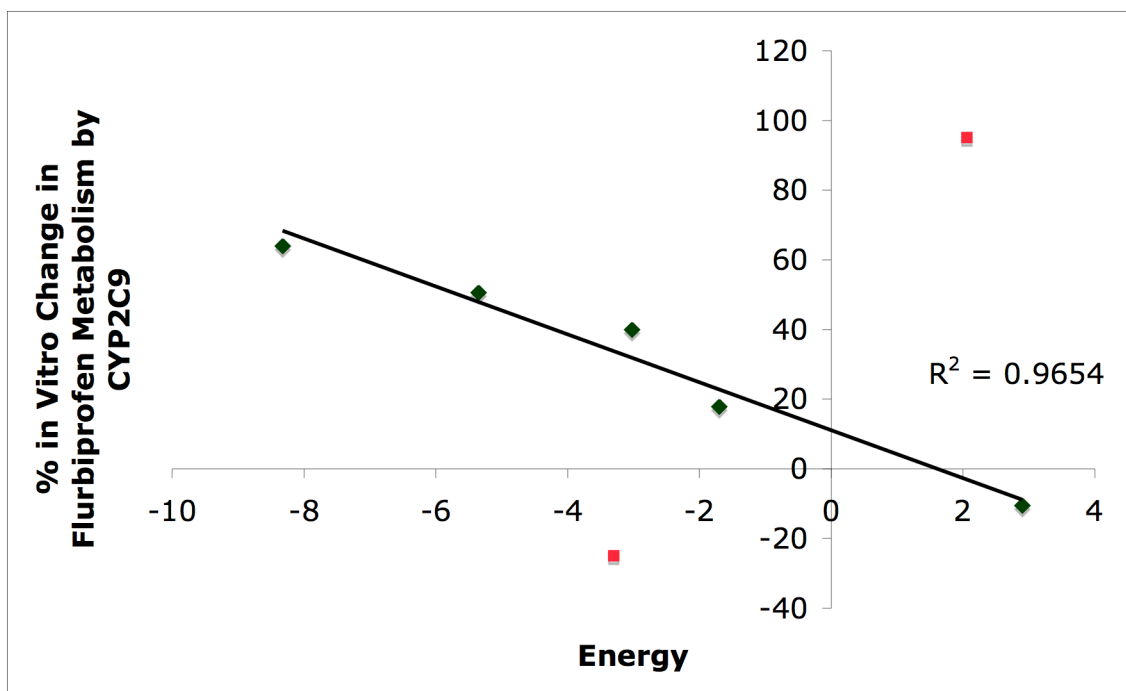


Figure 4.12. Graphic Representation of *In Vitro* Flurbiprofen Metabolism by CYP2C9 as a Function of Substrate-Effector Complex Stability.

The correlation between QM-calculated energetic stability of flurbiprofen-effector complexes and percent *in vitro* change in flurbiprofen metabolism by CYP2C9 is shown. Outliers are colored salmon and are not included in the trendline.

CHAPTER V

Determination of Effector-Active Site Binding in

CYP2C9 with eHiTS

5.1. Introduction

eHiTS is a docking program similar to AutoDock in that it involves a flexible ligand/rigid receptor methodology to conduct virtual HTS within an acceptable range of computational expenditure. Both use a grid of discrete interaction energies that form the topology of the active site. However, the algorithm employed by eHiTS for docking differs from that of AutoDock significantly. As in AutoDock, ligands are divided up into fragments separated by rotatable bonds, but eHiTS takes this exercise one step further and actually cleaves those bonds, separating the original molecule into rigid fragments (functional groups with fixed torsions) and flexible chains (functional groups with rotatable torsions). For greater computational economy both the active site topology and rigid fragments are more simplistically described by the program as polyhedral shapes, with each vertex corresponding to a specific chemical property. eHiTS subsequently docks the fragments in the active site individually rather than as a complete molecule. This approach enables a wide sampling of computational space inside the binding site that would require a much greater computing expenditure if using an intact ligand.

During docking all receptor-fragment combinations that do not result in steric violations are scored and retained for the next segment in the methodology. Even poorly-scored pairs are allowed to proceed to the next step due to the fact that their functionality in the whole compound may be simply to spatially separate the fragments that are contributing to the molecule's binding, rather than forming a binding interaction

themselves. All receptor-fragment pairs are archived in a table from which interactions may be retrieved if the program docks a compound containing the same rigid fragment as one that was previously docked. This tactic becomes especially cost-efficient when screening databases, which likely contain several similarly-structured compounds.

Next in the methodology is pose-matching, when receptor-fragment pairs are oriented spatially to one another such that the distances between them can accommodate the flexible chains that connect them. In this step each fragment is represented by a set of graph nodes, with each node corresponding to a possible fragment pose. The nodes of different fragments are considered to be connectable if there is no steric clash between the two poses and the distance between the poses lies within the interval set by the flexible connecting chain. Summing the respective scores of the fragments approximates the final score of the total molecule closely enough that those node pairs with lower summed scores can be discarded and only the most promising pairs retained for further processing. Following pose-matching, torsions of the flexible chains are rotated via energy-minimization such that they meet the fixed endpoints of the rigid fragments they connect at the lowest energy state possible, resulting in a reconstituted molecule. With the overall torsional energy factored in, a final score for the whole compound can be calculated.

The scoring function used by eHiTS is based on a combination of empirical and knowledge-based approaches. The empirical approach uses parameters (such as hydrogen bonding, van der Waals forces, etc.) weighted such that they can be correlated to

experimental data (for example, binding energies). Summing the weighted terms results in an approximation of the experimental data. Knowledge-based approaches employ statistics gathered from experimental data to develop criteria for preferred interactions. These criteria exist in the form of atomic pair potentials and are used to predict binding modes rather than binding energies.

Surface point energies composed of interaction data calculated via the above two approaches are used to determine the scores of ligands at each step of the eHiTS algorithm. eHiTS discerns 23 different types of surface points (**Table 5.1**), which exist at each vertex of the ligand and receptor polyhedral representations. The relationship between them is based on four variables (**Figure 5.1**): ligand-receptor interatomic distance, the respective angles formed by the atoms involved and a line representing interatomic distance, and the dihedral angle formed between the atoms involved. Interaction statistics are collected for each type of surface point and each interaction is composed of four variables, resulting in a 23 x 23 cell matrix with a four-dimensional probability field per cell. From this matrix interaction probabilities are calculated by fitting a third degree polynomial equation to the data in each cell using partial least squares regression. Converting these probabilities to docking energies is accomplished by assigning an energy scaling factor to each matrix cell based on the Boltzmann equation.

Table 5.1. The 32 Surface Point Types Recognized by eHiTS.

Surface Point Type	Definition
METAL	positively charged metal ion point
CHARGED_HPLUS PRIMARY_AMINE_HLP	positively charged hydrogen, e.g. Arginine primary amine hydrogen/lone-pair, e.g. $-\text{NH}_3^+$ or $-\text{NH}_2$
HDONOR	strong (primary) hydrogen bond donor H (polar-atom-H)
WEAK_HDONOR	weak (secondary) hydrogen bond donor H (polarized C-H)
CHARGED_LONEPAIR	lone pair of negatively charged group, e.g. PO_3^-
ACID_LONEPAIR	lone pair of an acid group, e.g. carboxylate
LONEPAIR	strong (primary) hydrogen bond acceptor lone pair
WEAK_LONEPAIR	weak (secondary) hydrogen bond acceptor lone pair
AMBIVALENT_HLP	donor H OR acceptor Lp depending on protonation state
ROTATABLE_H	rotatable-hydroxy donor H
ROTATABLE_LP	rotatable-hydroxy acceptor Lp
HYDROPHOB	H on aliphatic (chain) hydrophobic carbon
H_AROM_EDGE	H on hydrophobic carbon in aromatic ring (non-polarized)
WS_LIPO	H on weak secondary hydrophobic atom (e.g. carbon next to polar)
NEUTRAL	H/Lp on neutral atom (no recognized activity)
PI_AROMATIC	π electron of an aromatic ring
PI_RESON_POLAR	π electron on polar atom (N/O) in resonance chain, e.g. amide
PI_RESON_CARBON	π electron on carbon atom in resonance chain, e.g. amide
PI_SP2_POLAR	π electron on sp^2 polar atom (N/O) (non-resonating, non-arom)
PI_SP2_CARBON	π electron on sp^2 carbon atom (non-resonating, non-arom)
HALOGEN	lone electron pair of a halogen atom (F,Cl,I)
SULFUR	lone electron pair of a sulfur atom

Surface point definitions used by eHiTS to characterize atomic /electronic interactions are listed.

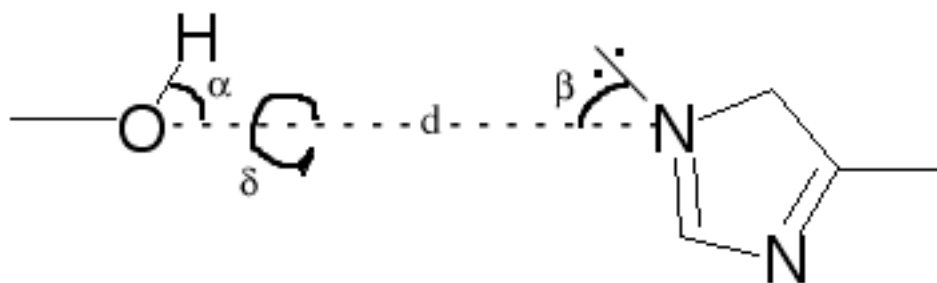


Figure 5.1. Four Atomic Relationship Variables Illustrated.

The relationship between ligand atoms and receptor atoms are determined by four variables: their interatomic distance, d ; angles α and β ; and torsion angle, δ . Lone pairs are also considered as “atoms” in this regard, as illustrated in this figure. From⁹⁴.

The final score includes the aforementioned derived scoring functions and calculated energy scaling factors as well as terms for sterics, binding pocket depth, receptor surface coverage, ligand conformational strain, ligand intramolecular interactions, and entropy. Scores are in units of the log binding affinity constant (pK_i).

The eHiTS scoring function takes into account the fact that different types of receptors require differently weighted parameters for accurate scoring by including specific weight sets for 97 different families of receptors based on the residue types that comprise the topology of the active site. If the receptor to be used does not belong to any of those families then a new set of parameter weights can be developed tailored specifically to the receptor in question via enrichment training. The result is a set of energy scaling factors (weights) for each parameter unique to the "trained" receptor that more accurately describe the energetic strengths of the parameters pertinent to the active site.

Training can be accomplished either manually by editing the default descriptor parameter file, or it can be automated by screening compounds of diverse but known biological activity (i.e. a training set). If automated, the process is as follows: First, the ligands are docked and the best poses are found. Then, using the best pose of each ligand, active site descriptor information is collected in the form of receptor and ligand surface point types and statistics on the frequency and geometry of their interactions. The residue names and numbers of the heavy atoms to which the surface points belong and the distances of the centers of the interacting atoms are also noted. This information is then

compared to that in the 97 pre-determined receptor types and then clustered according to the frequency with which the descriptor information matches that in a given receptor type on a per-residue basis. The relevant weights associated with the descriptor information in the pre-determined receptor type is then added to a "weights table" which will be applied to the trained receptor the next time a set of compounds is screened through it. Thus the resulting trained receptor is, in a way, a conglomerate of other receptors whose interaction energies have already been optimized. It must be stated that this clustering process is not based on biological activity per se, but rather on the residue types that comprise the topology of the active site and binding site similarities between the receptor to be trained and those that are already trained.

Screening databases through a receptor with eHiTS results in the best scored pose of a particular ligand from a user-arbitrated number of clusters. Each cluster is based on the ligands' root mean square deviation (RMSd) to each other, which can be changed to suit the requirements of the user. eHiTS outputs the top-scoring pose of each compound and concatenates the best poses of all compounds in the screened database into one file for analysis. An overall score is given for each compound as well as a decomposition of individual scoring parameters (H-bonding, hydrophobic interactions, etc.) used to compute the overall score, which allows the user to easily determine which interaction types are driving the binding pose for a particular ligand. The resulting poses can be observed along with pertinent receptor-ligand interactions using a companion visualizing program to eHiTS, CheVi. In addition to providing visual information regarding receptor-

ligand binding, CheVi also provides a decomposition of individual interaction types which are scored by the program according to their importance in binding.

5.2. Materials and Methods

5.2.1. Preliminary Structure Preparation

A series of dapsone analogs consisting of the first twelve compounds listed in **Figure 4.1** were built with Sybyl (Tripos, St. Louis, MO) and energy minimized. Geometry optimization was performed at the AM1 level with Gaussian 03 (Gaussian, Inc., Wallingford, CT). The effector, dapsone, was oriented in the active site of the X-ray crystal structure of CYP2C9¹² according to the model determined by NMR³⁸ by AutoDock3.0. Hexacoordinate heme parameters were obtained from the AMBER parameter database maintained by Richard Bryce at the University of Manchester (http://pharmacy.man.ac.uk/amber/cof/heme_all_inf.html)⁸⁸. Flurbiprofen, the substrate, had been co-crystallized with the enzyme. A 2 ns molecular dynamics simulation was conducted according to the methods in Chapter 4. At the end of the simulation dapsone was excised from the active site to create the receptor structure used for docking. Because eHiTS automatically takes all possible protonation states of both ligands and receptor into account, no preparatory action was taken to protonate/deprotonate ionizable ligands.

5.2.2. Receptor Training

Manual receptor training was undertaken by altering the parameters.cfg file of the eHiTS program to weight parameters deemed important for heteroactivation by a previously developed predictive CoMFA model [J. S. Aguilar, unpublished]. Parameter weight values changed were steric fit scoring weight (from 0.04 in the original parameter.cfg file to 1.0 in our altered file), surface contact van der Waals energy (from 1.0 to 2.0), strongest aromatic Pi stacking interaction energy (from 1.0 to 2.0), and strongest hydrogen bond energy (from 1.0 to 2.0). Automated receptor training was also conducted using an enrichment training algorithm in which known activators and inhibitors are submitted to the docking program. The dapsone analog test set listed in **Figure 4.1** was divided into activators and non-activators (which included inhibitors) and were submitted to eHiTS for the calculation of weighting terms.

5.2.3. Docking Protocol

Docking was performed using eHiTS 6.2 at the maximum accuracy setting. The docking area boundaries were set at 5 Å from the residues lining the active site of CYP2C9. All other settings were left at their default values. The program was run using the default parameter weight set, a manually trained receptor, and an enrichment-trained receptor. Initially, the 20-compound dapsone analog test set was screened through the model, followed by a validation set of the first 60 compounds from the NCI database⁹⁵, for whom biological activity data was not known. Following this, a second validation set consisting of the first 1000 compounds from the NCI database was screened. Both

validation sets were seeded with the dapsons analog test set compounds as a control. NCI compounds scored higher than dapsons were submitted to *in vitro* biological activity assays to determine the presence and/or degree of heteroactivation of flurbiprofen metabolism by CYP2C9.

5.2.4. Visualization and Analysis of Results

Receptor-ligand interactions were visualized and determined using CheVi (SimBioSys, Inc., Toronto, Ontario) and PyMol (DeLano Scientific, Palo Alto, CA). H-bonding was determined visually using a distance cutoff of 4.0 Å between donor and acceptor atoms, and pi-interactions were determined by a distance cutoff of 5 Å between aromatic ring carbons⁷⁴.

5.2.5. Data Fitting

Overall scores and decomposed interaction type scores were fitted to the experimental biological activity data using both linear and polynomial regression using Microsoft Excel 2004 for Macintosh.

5.3. Results and Discussion

Docking a test set of dapsons and 19 dapsons analogs into the active site of CYP2C9 using the default eHiTS scoring weights did not yield a significant R^2 (overall score $R^2 = 0.01$), indicating no correlation between either the overall eHiTS score or the individual decomposite interaction scores and the *in vitro*/CoMFA-predicted increase in

flurbiprofen metabolism. The binding mode of dapsone, however, was consistent with that found in the AutoDock results in Chapter 4 as well as with the pose found by Wester¹² in which the sulfone oxygens pointed away from the heme moiety toward the F-G loop of the enzyme. In this pose it is able to hydrogen bond with the ammonium side chain of Asp-474 by way of its sulfone oxygens (**Figure 5.2**). The amino group of Ring 1 was also able to form H-bonds by interacting with the substrate carboxyl moiety, further stabilizing the effector in this position. Another stabilizing interaction present in this pose is pi-stacking, which occurs between dapsone and the active site of CYP2C9 as well as with flurbiprofen. Ring 1 of dapsone stacked in an offset geometry via 16 pi-pi interactions with Ring 1 of flurbiprofen at an average distance of 4.22 Å (**Figure 5.3**). Dapsone also T-stacked via Ring 1 with Phe-476, and Ring 2 formed pi-interactions with this residue in an offset geometry. Distances between atoms involved in the rings stacking were well within an acceptable cutoff of 5 Å, averaging 4.05 Å from Ring 1 with a total number of 9 pi-pi interactions, and an average of 4.36 Å from Ring 2 with a total of 3 pi-pi interactions. However, the geometry of the ring overlap between dapsone and Phe-476 is not conducive to optimum pi-pi interactions due to the orientation of the participating rings; the rings were so severely offset that orbital overlap would be minimal (**Figure 5.4**). The stacking conformation of the dapsone-flurbiprofen pair, however, was more favorable due to a closer alignment of the contributing rings.

Though the majority of dapsone analogs docked with their sulfone oxygens pointed toward the F-G loop, eight assumed an "alternative" pose, with sulfones pointed toward the heme (**Figure 5.5**). Ring 2 of these effectors (**2, 4, 5, 11, 13, 15, 17 and 19**)

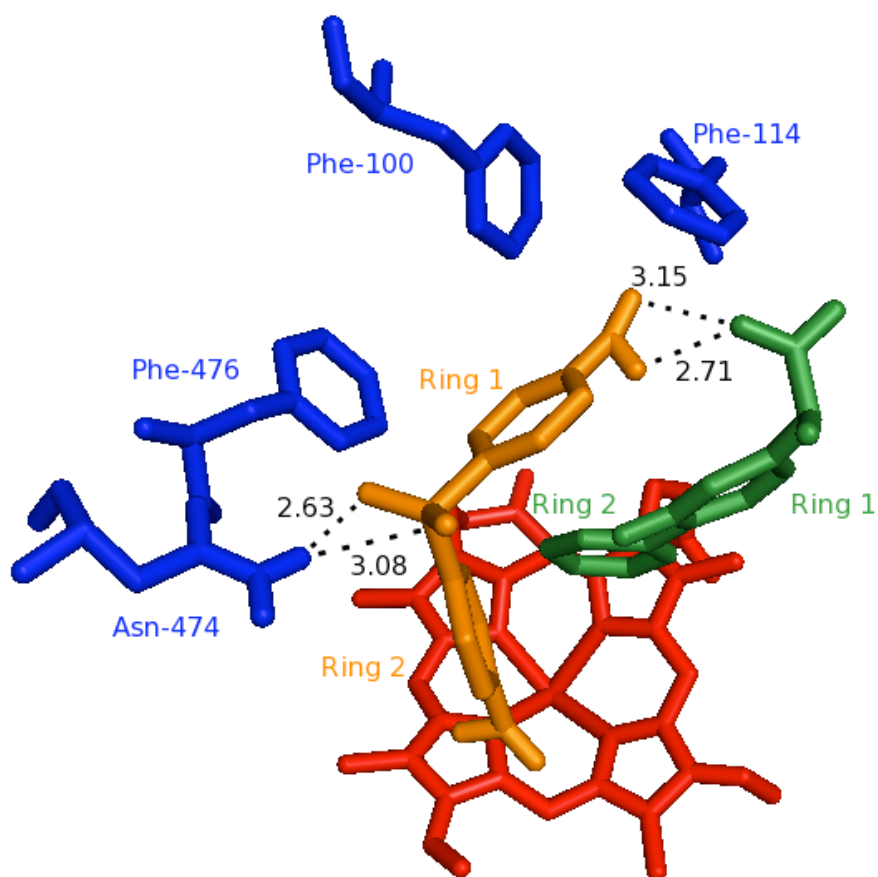


Figure 5.2. Hydrogen Bonding of Dapsone Docked Into The Untrained Receptor.

Hydrogen bonds formed between dapsone and the flurbiprofen-CYP2C9 complex using default scoring function weights are illustrated. Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue. Distances are in Ångstroms.

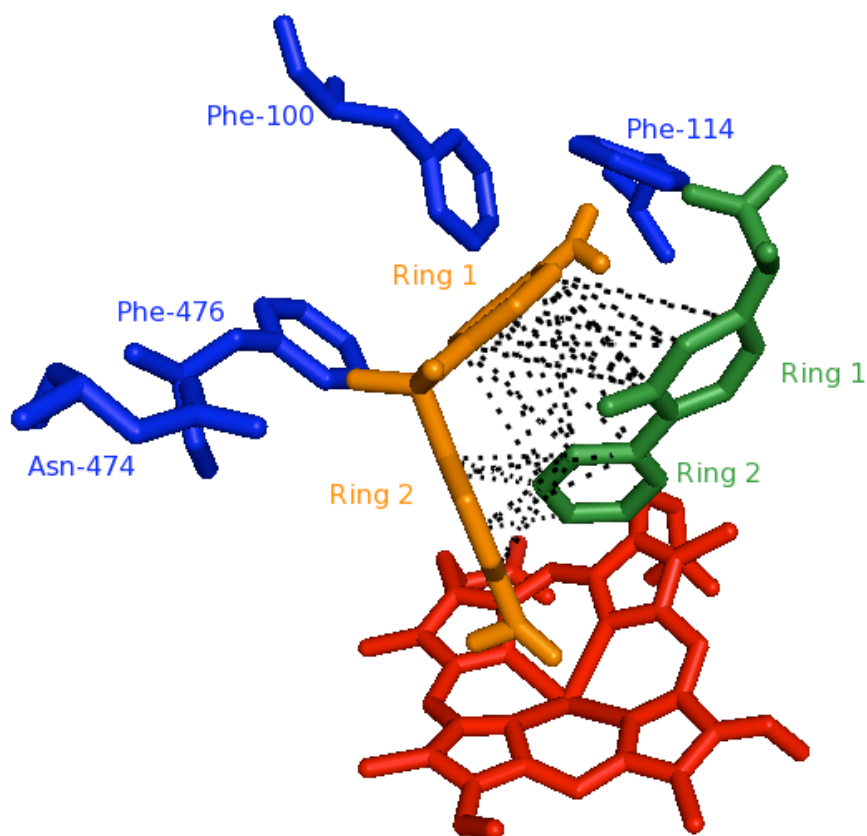
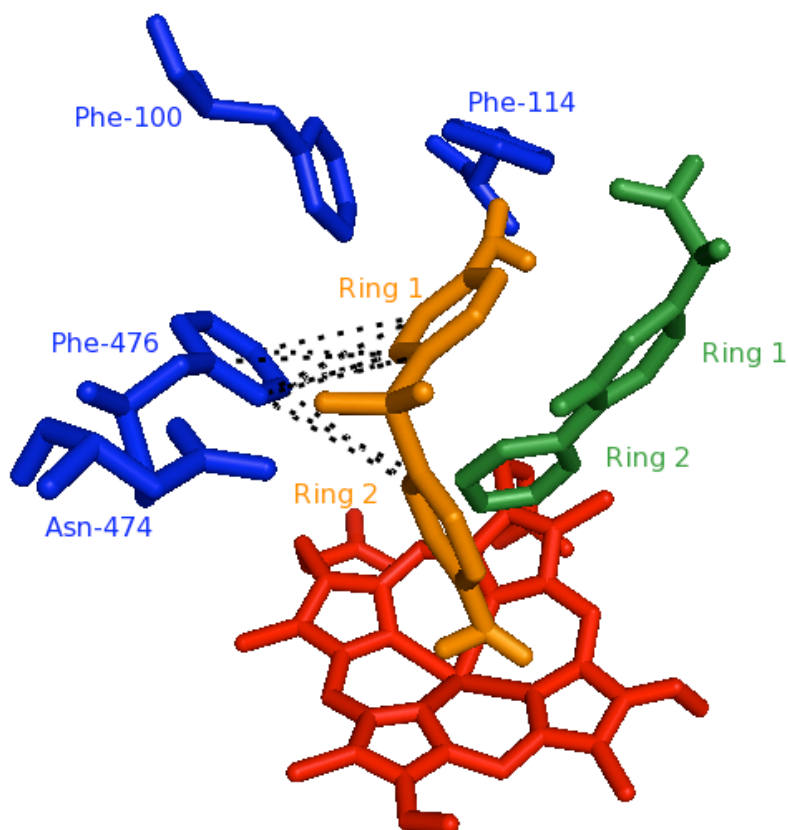


Figure 5.3. Pi- Interactions Between Flurbiprofen And Dapsone Docked In The Untrained Receptor.

Pi- interactions between dapsone and the flurbiprofen using default scoring function weights are illustrated. Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue. Black dotted lines indicate pi-interactions.



5.4. Pi- Interactions Between Aromatic Active Site Residues Of CYP2C9 And Dapsone In The Untrained Receptor.

Pi- interactions between dapsone and the aromatic active site residues of CYP2C9 using default scoring function weights are illustrated. Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue. Black dotted lines indicate pi-interactions.

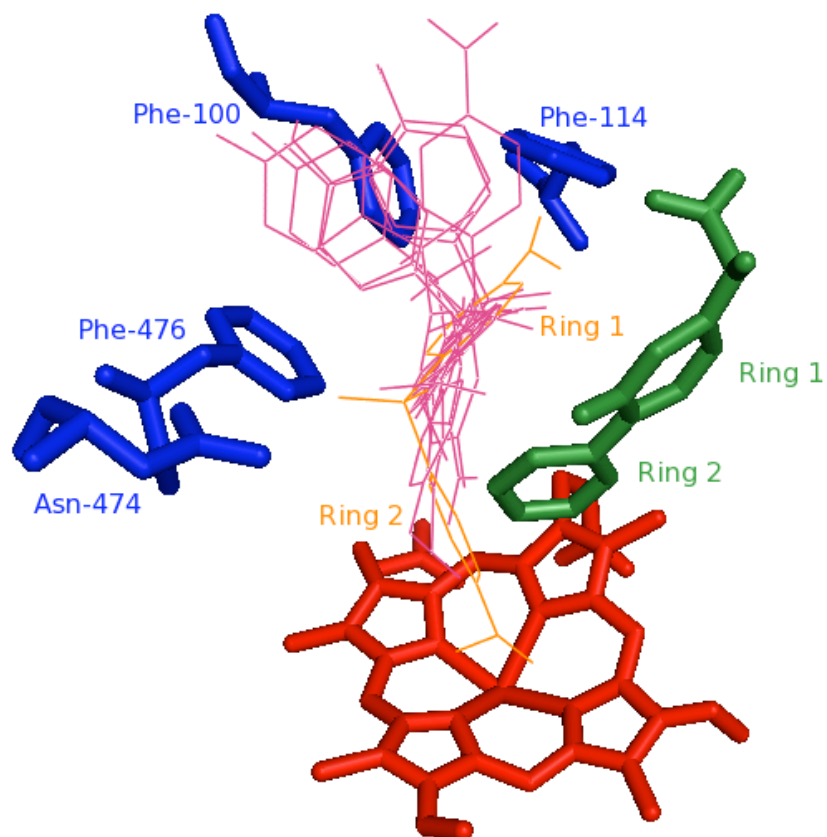


Figure 5.5. Overlay Of Dapsone Analogs Docked In The Alternative Pose In The Untrained Receptor.

Dapsone analog effectors that assumed an alternative pose in the untrained receptor are illustrated. Effectors are magenta and are overlaid as lines in the interest of clarity.

Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue.

stacked closely in an offset "T" geometry with Ring 1 of flurbiprofen. The same effector rings were T-stacked with Phe-476. Interestingly, all of the dapsone analogs taking the alternative pose were substituted with electron-donating groups in the *para* position of the ring involved in stacking with Phe-476/ flurbiprofen. It is likely that electron donating groups increased the electron density of the aromatic ring via resonance effects (-OH, -NH), or, to a lesser degree, inductive effects (-CH₃), resulting in stronger pi-pi interactions with other aromatic rings^{96,97}. Stronger stacking between the effector ring systems and aromatic active site residues like Phe-476 may make the favorability of the alternative pose energetically competitive with H-bonding of the sulfone oxygens to Asn-474.

Compounds that were *para* substituted with electron-donating groups that were also capable of accepting H-bonds formed electrostatic interactions with the side chain ammonium group of Asn-474, further stabilizing them in the alternative position. It should be noted that included among the activator compounds that docked in this position were the top three activators *in vitro*: (dapsone,**11** and **19**), and that the degree of *in vitro* heteroactivation for these three effectors corresponded to the electron-donating capacity of the *p*-substituents. The exception to this is **15**, which is a weak inhibitor of flurbiprofen metabolism (-5%), may have the potential to be an activator due the fact that carboxylates are strong electron donating groups, and thus would greatly increase the electron density of the effector aromatic ring. However, this compound is a dianion at physiological pH and the formal negative charge clashes with that of the substrate, rendering a close

proximity between the two compounds nearly impossible, and canceling out any activating qualities the effector may have. Assuming that the remainder of the inhibitors docking in the alternative pose are competitive inhibitors and therefore more likely to dock in the substrate binding pocket than they are in the "effector" pocket, it would appear that pi-stacking plays a very important role in activation of flurbiprofen metabolism by CYP2C9. As a corollary, electron-donating *p*-substituents on the aromatic rings involved in stacking with flurbiprofen will increase the strength of the pi-pi interaction between substrate and effector, thereby holding flurbiprofen more securely in a position conducive to 4'-oxidation. This data suggests that the probability that a given compound will induce flurbiprofen metabolism by CYP2C9 can be in part determined by the electron-donating capacity of its aromatic ring substituents.

Close stacking involving the same rings of dapsone analogs that formed pi-interactions with flurbiprofen was also observed with Phe-476. The alternative pose enabled the effectors to accomplish greater pi-orbital overlap with Phe-476 based on the more favorable geometry of the ring pair involved, which allowed a larger number of pi-interactions between rings with shorter distances between them compared to the ring geometry of the dapsone-Phe-476 pair. Effectors that did not assume the alternative pose preferred to dock with one of their aromatic rings in this region of dapsone Ring 1 (Figure 5.6).

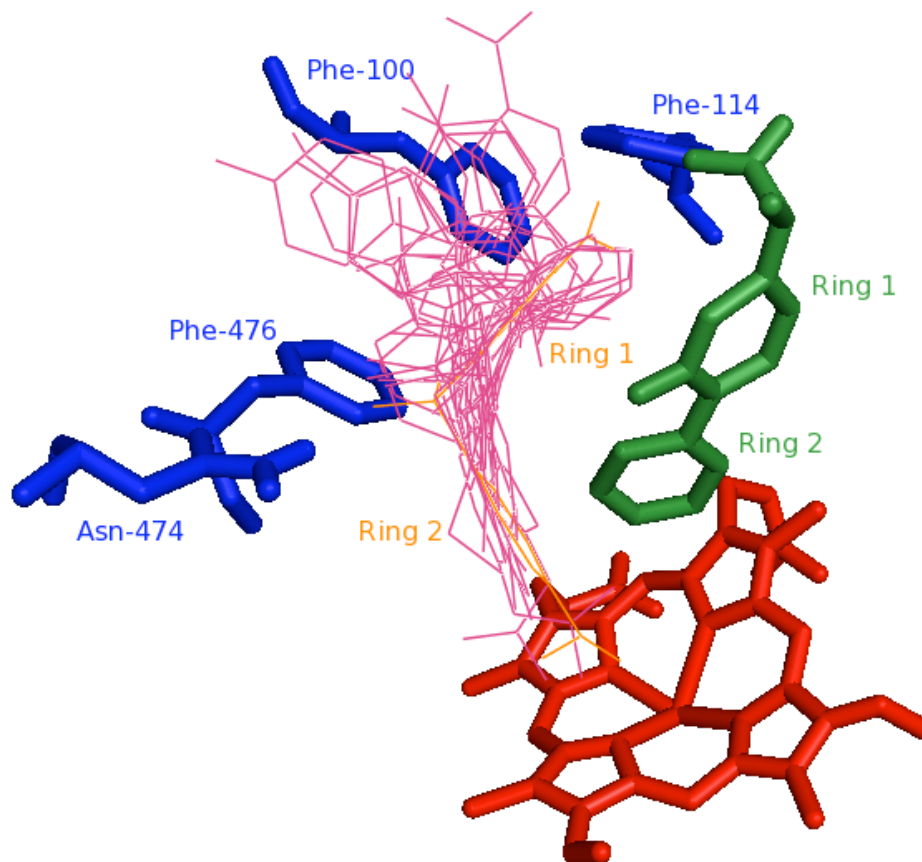


Figure 5.6. Overlay Of All Dapsone Analogs Docked In The Untrained Receptor.

Dapsone analog effectors docked in the active site of the untrained receptor are shown. Effectors are magenta and are overlaid as lines in the interest of clarity. Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue.

Following this it was decided to manually alter the default weights to reflect the importance of the specific descriptors identified by the predictive CoMFA model as most relevant to heteroactivation of flurbiprofen metabolism in CYP2C9. Although eHiTS does possess a built-in, automated weight-training function, the fact that some knowledge of important descriptors was already in-hand made the manual route a better choice to start with, as it enables the user to guide the scoring function more directly. As in the default weight eHiTS run, the docking results of the manually configured parameter file placed dapson in a conformation consistent with that seen in both the AutoDock results and the literature [Wester]. Several pi-interactions were observed in the results between dapson and flurbiprofen (**Figure 5.7**): Ring 1 of dapson stacked with both rings of flurbiprofen in an offset geometry that involved a total of 31 pi-interactions with an average distance of 4.37 Å, and Ring 2 of dapson T-stacked with Ring 2 of the substrate via 15 interactions averaging a distance of 4.26 Å. Pi- stacking between dapson and aromatic active site residues were limited to 9 interactions between Ring 1 and Phe-476, resulting in a T-stack geometry and having an average distance of 4.21 Å (**Figure 5.8**). H-bonds formed between the sulfone oxygen of dapson and the amino side chain of Asn-474, as well as between the carboxylate group of the substrate and the amino hydrogens of dapson Ring 1 (**Figure 5.9**).

Compared to the results using default weights, fewer effectors assumed the alternative pose, with sulfones pointing toward the heme, but the four that did (dapson,

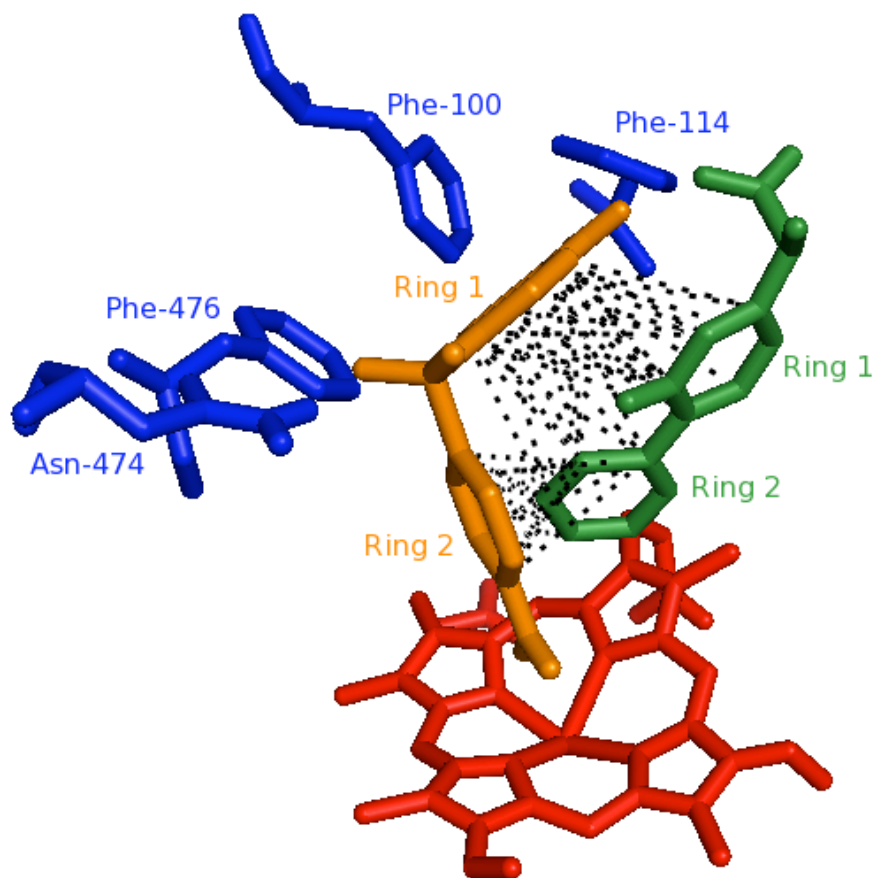


Figure 5.7. Pi- Interactions Between Flurbiprofen And Dapsone Docked In The Manually Trained Receptor.

Pi- interactions between dapsone and flurbiprofen using manually configured scoring function weights are illustrated. Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue. Black dotted lines indicate pi-interactions.

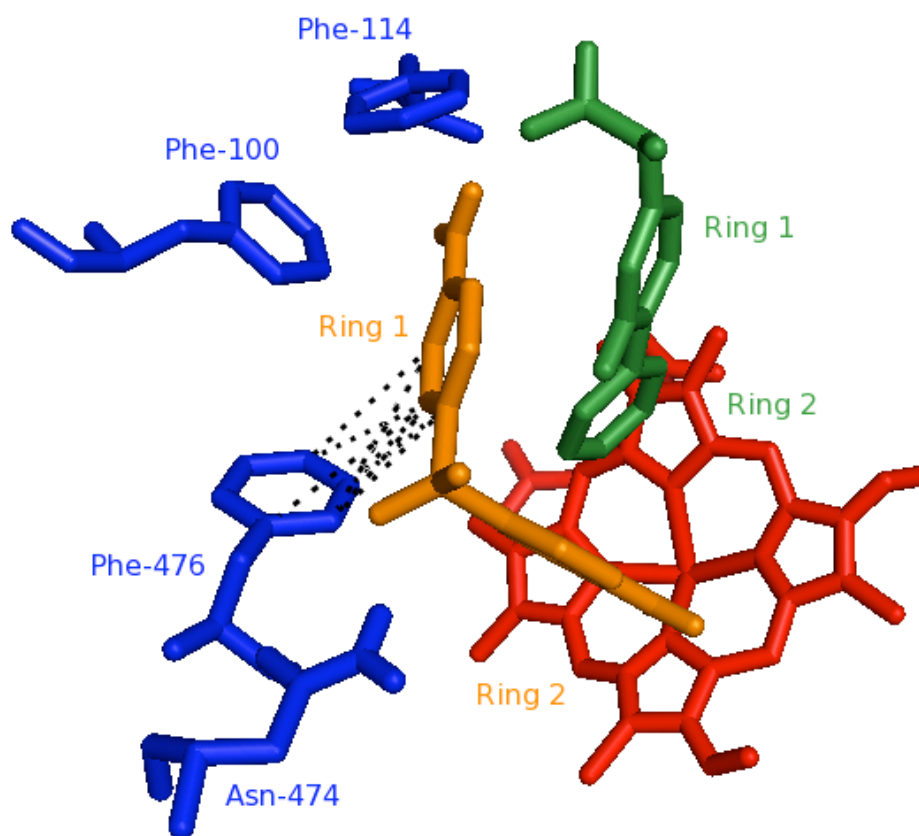


Figure 5.8. Pi- Interactions Between Aromatic Active Site Residues Of CYP2C9 And Dapsone In The Manually Trained Receptor.

Pi- interactions between dapsone and the aromatic active site residues of CYP2C9 using manually configured scoring function weights are illustrated. Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue. Black dotted lines indicate pi-interactions.

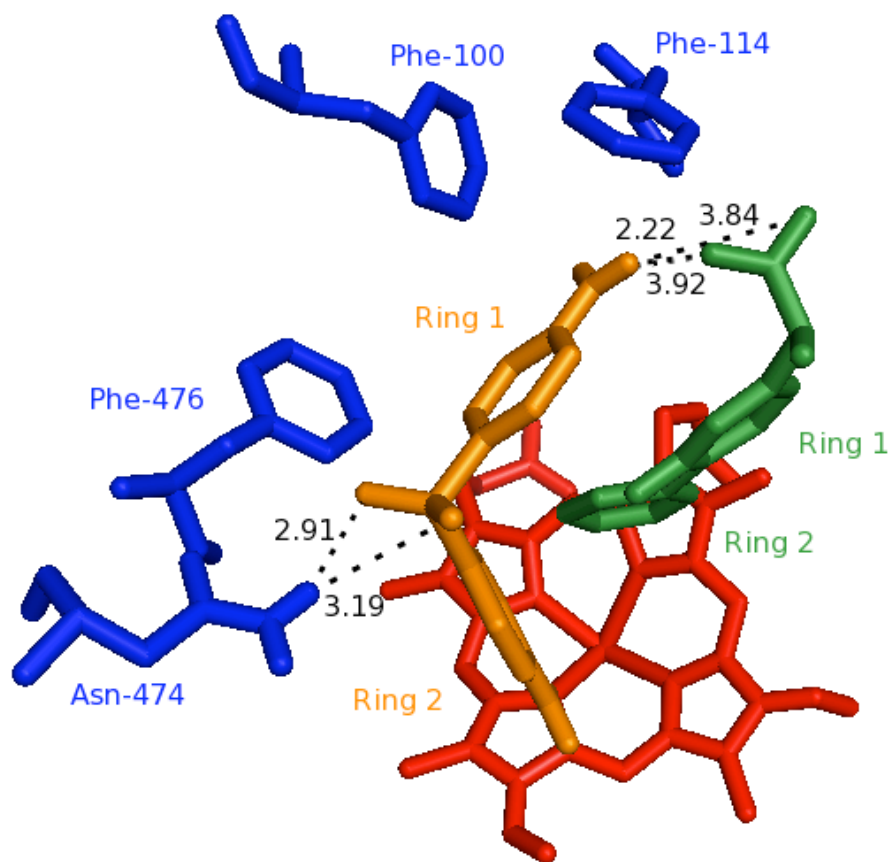


Figure 5.9. Hydrogen Bonding of Dapsone Docked Into The Manually Trained Receptor.

Hydrogen bonds formed between dapsone and the flurbiprofen-CYP2C9 complex using manually configured scoring function weights are illustrated. Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue.

Distances are in Ångstroms.

11, 15 and 19) were the same *para*- substituted compounds with electron-donating groups that docked in the alternative pose using the default weights. The rings of these effectors formed pi-interactions with Phe-476, Phe-114, and Phe-100 in addition to flurbiprofen (**Figure 5.10**). Superimposing the alternatively posed effectors with dapsone revealed a common site of ring overlap centered on Ring 1 of dapsone, where the greatest number of effector-substrate pi-interactions occurred. It is important to note that the center of effector overlap was not situated in a location that allowed for more interaction with active site residues, which suggests that interactions with flurbiprofen are a greater determinant of effector binding than those with active site residues. Effectors that docked with sulfones toward the F-G loop overlapped well with dapsone and were able to stack with both rings of flurbiprofen (**Figure 5.11**).

Scoring also improved somewhat, though all generated scores still lacked statistical significance; the decomposite interaction scores were slightly more promising compared to those obtained using the default weights, with the "Other" category having the best correlation of all categories ($R^2 = 0.37$, up from 0.27 using default parameters).

Automated enrichment training of the receptor was subsequently undertaken to assess the ability of the native weight-training function of eHiTS to produce docking scores more highly correlative to the *in vitro*/CoMFA-predicted data. In the results of this run dapsone docked in the alternative position (**Figure 5.12**), with sulfones directed toward the heme. In this pose Ring 1 of the effector readily stacked with Phe-100 in an

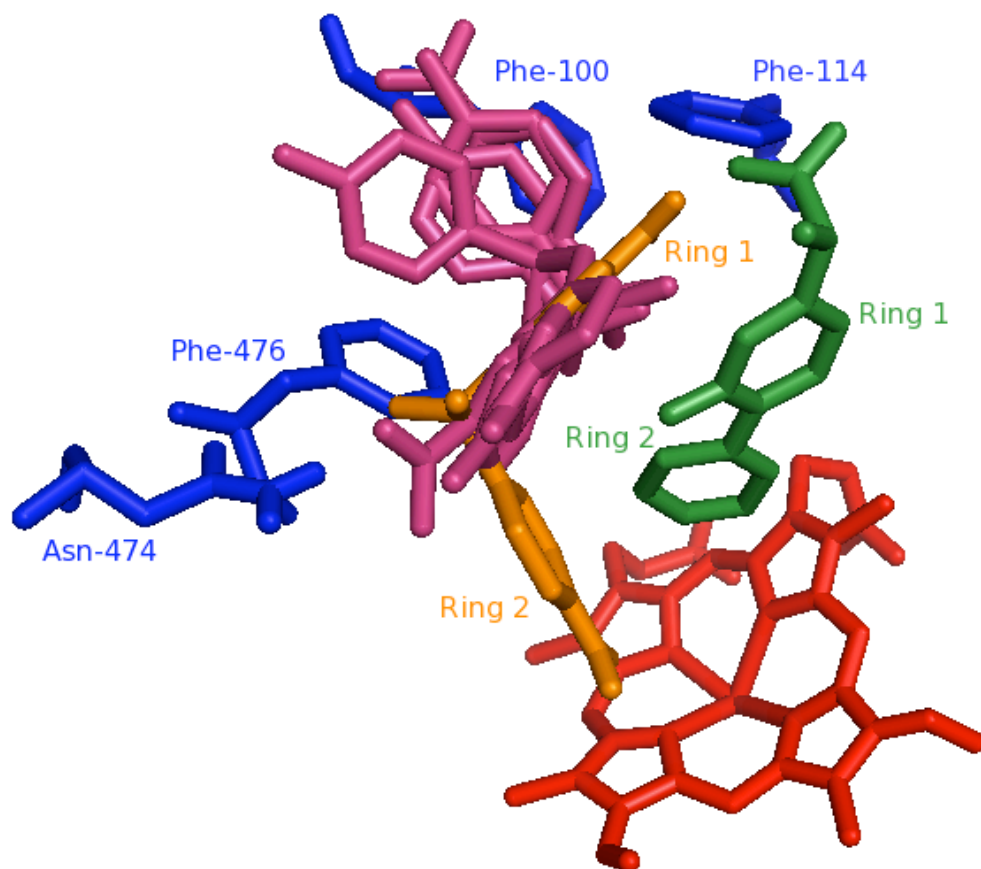


Figure 5.10. Overlay Of Dapsone Analogs Docked In The Alternative Pose In The Manually Trained Receptor.

Dapsone analog effectors that assumed an alternative pose in the manually trained receptor are illustrated. Effectors are magenta, flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue.

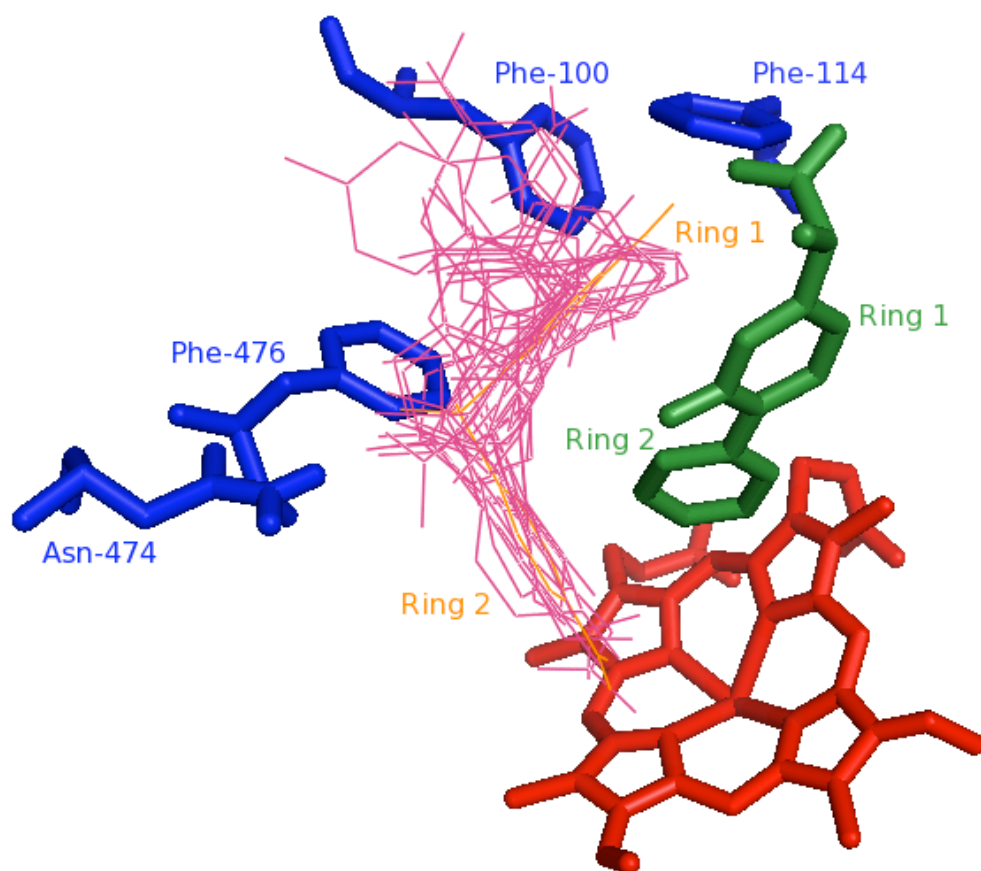


Figure 5.11. Overlay Of All Dapsone Analogs Docked In The Manually Trained Receptor.

Dapsone analog effectors docked in the active site of the manually trained receptor are shown. Effectors are magenta and are overlaid as lines in the interest of clarity.

Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue.

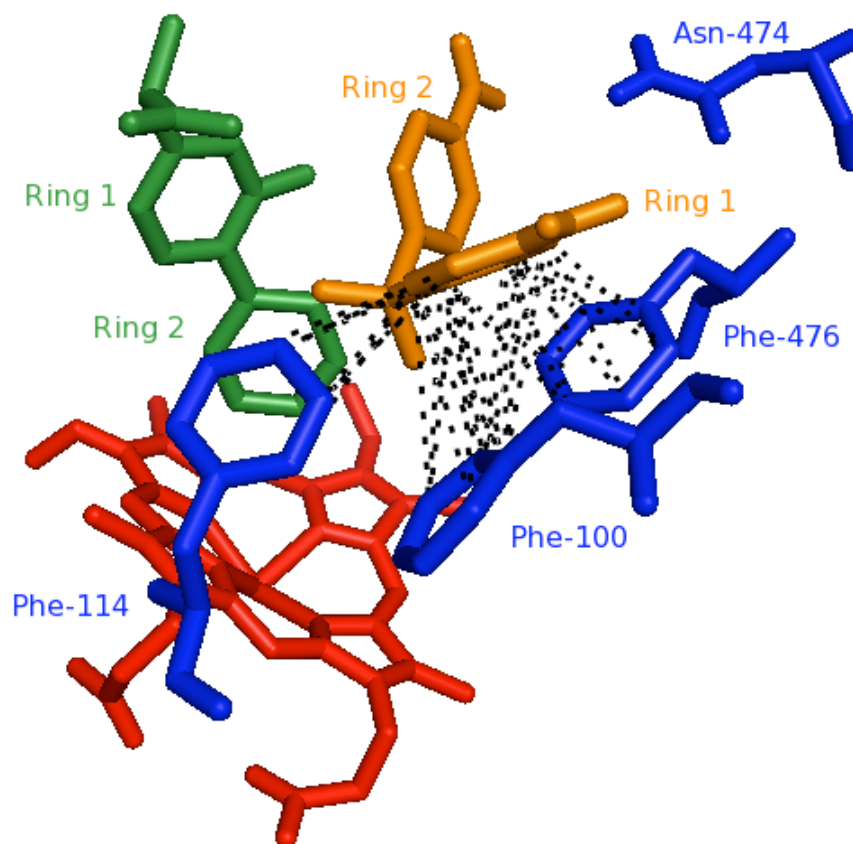


Figure 5.12. Pi- Interactions Between Aromatic Active Site Residues Of CYP2C9 And Dapsone In The Enrichment Trained Receptor.

Pi- interactions between dapson and the aromatic active site residues of CYP2C9 using enrichment-trained scoring function weights are illustrated. Flurbiprofen is forest, the heme is crimson, dapson is tangerine, and active site residues are in royal blue. Black dotted lines indicate pi-interactions.

offset geometry with an average distance of 4.41 Å via 14 interactions. Other pi-interactions with Phe-114 (average distance 4.01 Å with 8 interactions) were also observed via Ring 1 of dapson, as well as with Phe-476 (average distance 4.55 Å with 11 interactions), though poor pi orbital overlap with these residues meant that interactions with them were weak. Pi-interactions with flurbiprofen were still present between dapson Ring 2 and both rings of flurbiprofen (**Figure 5.13**) in a geometry resembling that seen in benzene trimers (**Figure 5.14**). Considering the large number of aromatic rings present in the active site in addition to aromatic active site residues when dapson or its analogs are bound simultaneously with flurbiprofen, higher order pi-stacking is a very conceivable possibility. No H-bonds were formed between dapson and flurbiprofen or Asn-474.

As in the previous two eHiTS runs, dapson analogs that assumed an alternative pose (**10** and **16**) were superimposable with dapson, with stacking occurring between the effectors and Phe-100 in an offset geometry as well as both rings of flurbiprofen (**Figure 5.15**). Neither effector contained electron-donating groups that would increase electron density in their aromatic ring systems. Effector compounds that docked with their sulfone groups directed at the F-G loop overlapped dapson Ring 2-- the same ring involved in the trimeric pi-stacking with flurbiprofen (**Figure 5.16**).

Only 5 of the 9 total activators in the training set were used in the enrichment training, due to the fact that the activity values of the other 4 were CoMFA-predicted. Because the correlation coefficient of the CoMFA model was high (0.80), a predictive

model can be built using manual methods such as parameter file editing without the need to differentiate the *in vitro* activity values from the CoMFA-generated values. This is especially true when the parameters being altered are descriptors identified by the CoMFA model as important to heteroactivation of flurbiprofen metabolism in CYP2C9. However, the enrichment training algorithm is complex enough that it may identify other descriptors that the CoMFA model missed. It was decided that, in the interest of preventing biasing the enrichment training with CoMFA data, the receptor should be trained using *in vitro* values only.

The overall scores that resulted from using the enrichment-trained receptor were more correlative to the *in vitro*/CoMFA-predicted data compared to both the manually configured run and the default parameter run. Again, the "Other" category possessing the best correlation coefficient ($R^2 = 0.44$). Despite the improvement in scores, there still existed no statistical correlation with the *in vitro*/CoMFA-predicted data. It was postulated that the low correlation could be due to the possibility that some compounds in the test set may elicit their effect on flurbiprofen metabolism by means of different mechanisms, for example inhibitors may be competitive inhibitors and prefer to dock in the substrate binding site, where eHiTS cannot dock them due to the rigidity of the receptor structure (in which flurbiprofen is present).

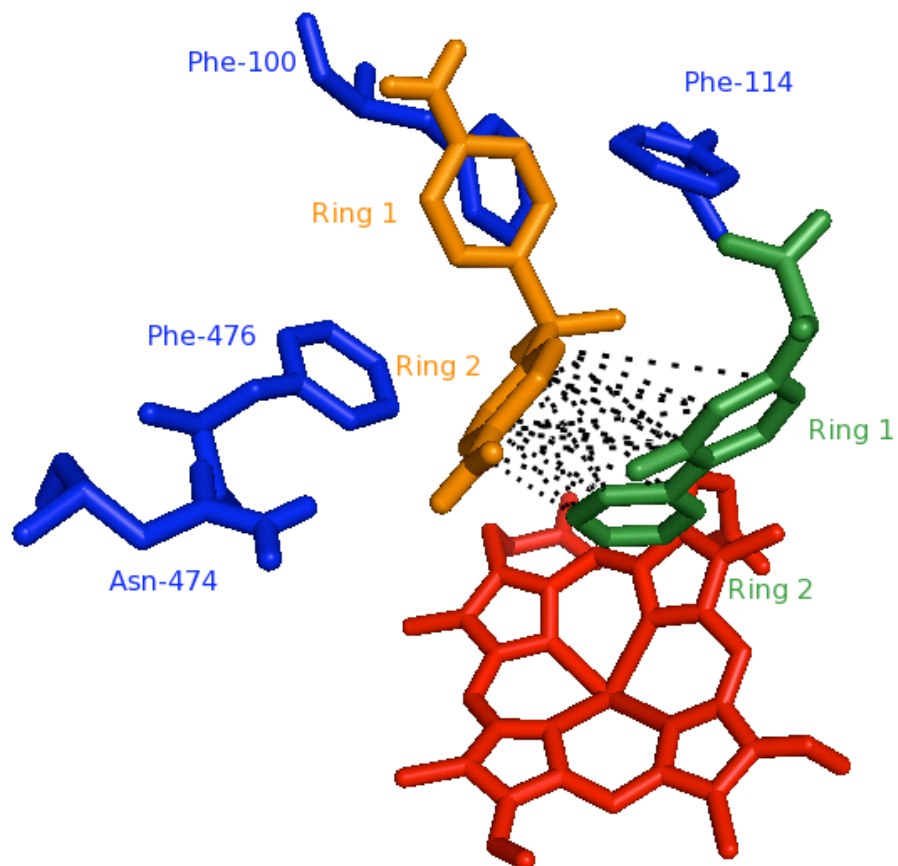


Figure 5.13. Pi- Interactions Between Flurbiprofen And Dapsone Docked In The Enrichment Trained Receptor.

Pi- interactions between dapsone and flurbiprofen using enrichment trained scoring function weights are illustrated. Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue. Black dotted lines indicate pi-interactions.

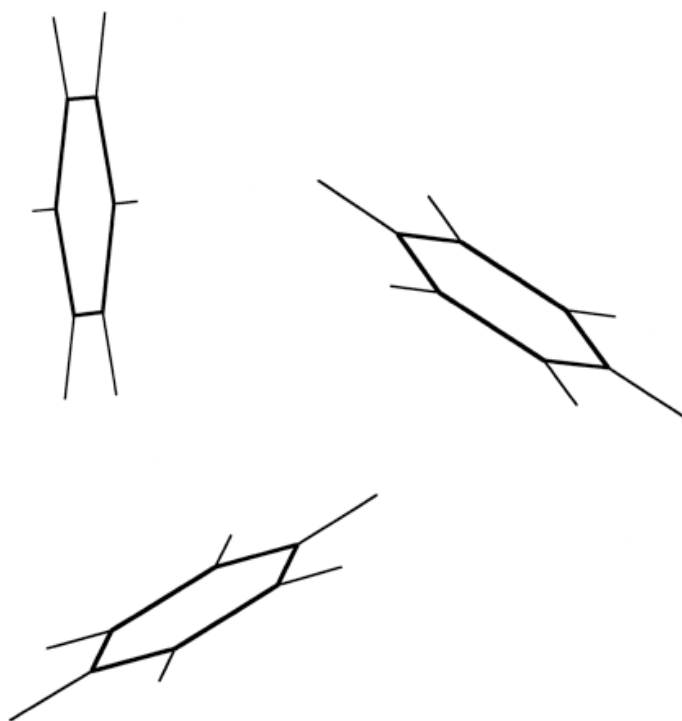


Figure 5.14. Optimized Orientation Of A Clustered Benzene Trimer.

Optimized orientation of a clustered benzene trimer is shown. From McGaughey et al ⁹⁸.

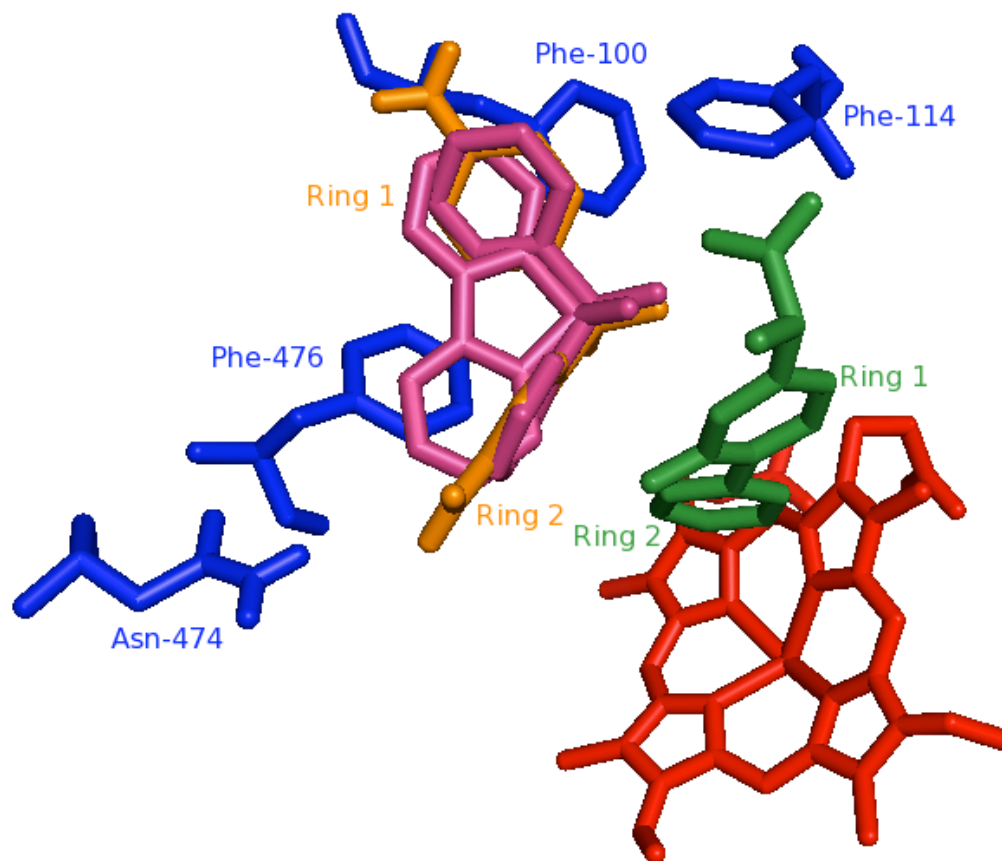


Figure 5.15. Overlay Of Dapsone Analogs Docked In The Alternative Pose In The Enrichment Trained Receptor.

Dapsone analog effectors that assumed an alternative pose in the enrichment trained receptor are illustrated. Effectors are magenta, flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue.

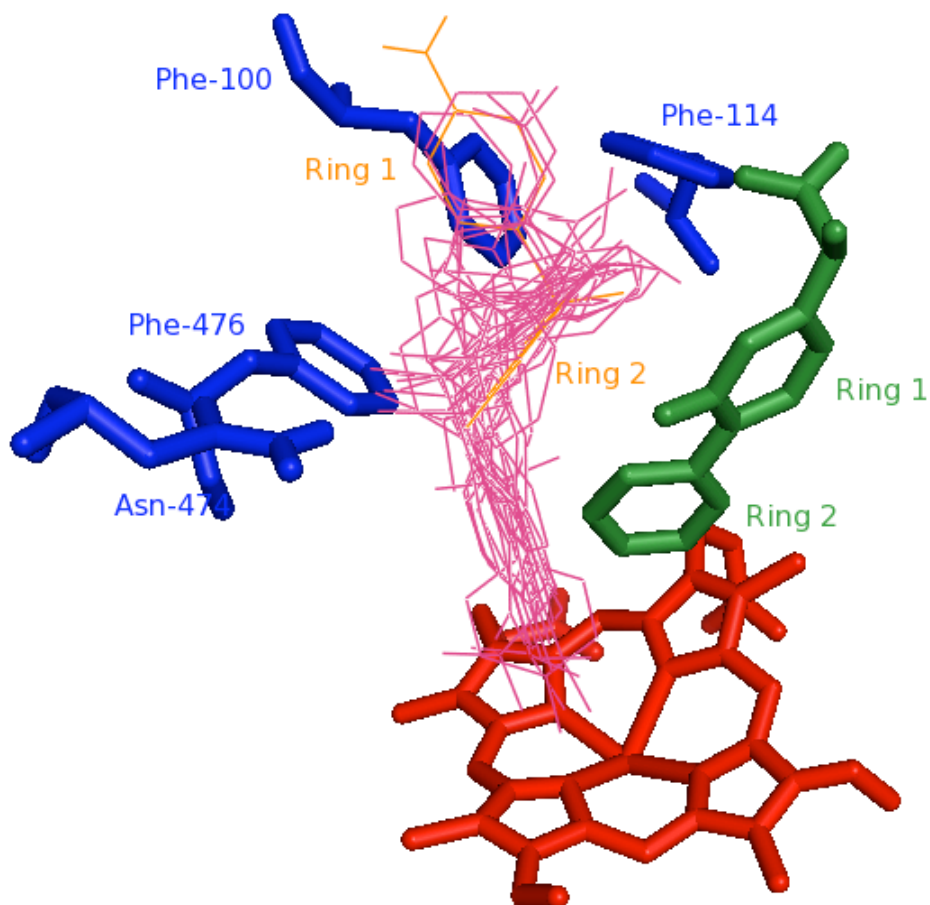


Figure 5.16. Overlay Of All Dapsone Analogs Docked In The Enrichment Trained Receptor.

Dapsone analog effectors docked in the active site of the enrichment trained receptor are shown. Effectors are magenta and are overlaid as lines in the interest of clarity.

Flurbiprofen is forest, the heme is crimson, dapsone is tangerine, and active site residues are in royal blue.

It stood to reason that calculating the R^2 of activators separately from non-activators (inhibitors and do-nothings) may improve the correlation since they are more likely to activate metabolism using similar mechanisms. Using this approach the R^2 of the "Other" category for activators in the test set dropped to 0.30. Thus far the regression analyses had been linear in nature. Applying a second-order polynomial equation to the data for activators lowered the R^2 of the "Other" category to 0.30. Interestingly, applying the same polynomial function to the overall score raised that R^2 value to 0.60, which is statistically significant (**Figure 5.17**).

The "Pi-stacking" category did not benefit from the polynomial function ($R^2 = 0.23$), which is at odds with the binding modes of the docked effectors. It is possible that the scoring function used to define the pi-stacking term may be insufficient, but it may be compensated for by other surface point descriptors used to calculate other terms that also contribute to the overall score. The likelihood is good that this is an example of an all-too-common phenomenon in which the native scoring function of a docking program possesses sophistication enough to determine the correct binding pose, but not enough to provide a correlative docking energy.

The most relevant correlation is that of the overall score due to the fact that eHiTS uses this score, and not the individual decomposite scores, to rank docked compounds

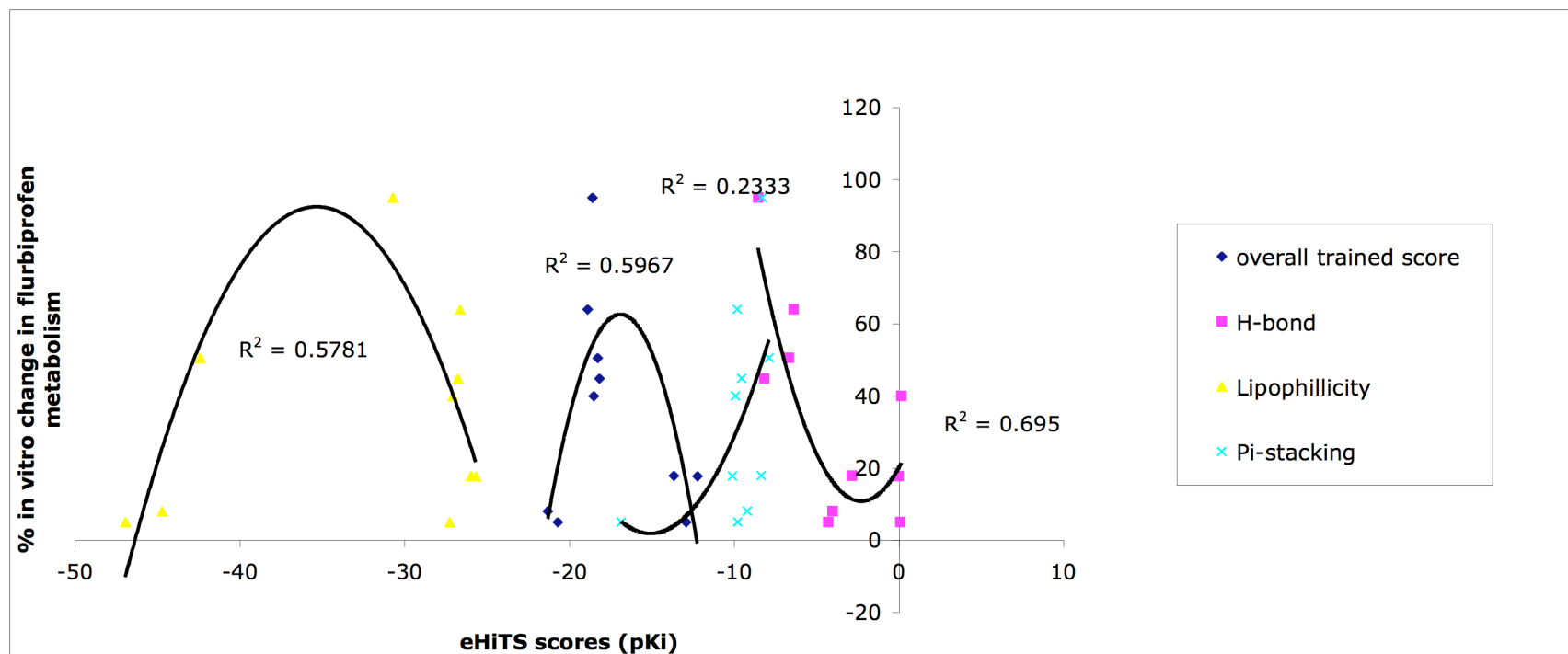


Figure 5.17. eHiTS Activator Scores And Correlation Coefficients After Enrichment Training Of The Receptor.

eHiTS activator scores and correlation coefficients obtained from the enrichment trained receptor are illustrated.

from a screened database; with a high correlation coefficient, eHiTS is more likely to rank as-yet unknown heteroactivators higher in the output results, making discovery of a new activating compound more likely in turn.

Any data can be fitted using the right function, however, and a good R^2 does not by default constitute a true relationship between the *in vitro*/CoMFA-predicted data and calculated scores. Rather, the equation must relate to a physical model. In this case the parabolic data suggests that as the substrate-effector interactions become stronger flurbiprofen turnover increases, until a point is reached at which the increasing strength of the interactions begins to reduce the rate of metabolism. At the point at which metabolism begins to increase, substrate-effector interactions would become dominant over effector-enzyme and substrate-enzyme interactions. This is reflected in all three sets of eHiTS docking results, in which effector compounds in general prefer to π -stack with the substrate rather than with active site residues. However, beyond a certain point those interactions become strong enough that they actually begin to slow the turnover rate. The reason for the reduction in metabolism could be due to π -interactions between flurbiprofen and the effector that may increase the electron density of the metabolized ring such that the 4'-hydrogen of the substrate becomes less acidic, and thus less inclined to react with the heme. Another possibility is supported by the MD simulation discussed in Chapter III, in which dapsone was shown to draw the substrate away from the heme when the distance between their rings was very close, but appeared to stabilize the

substrate in a position closer to the heme when the dapsone-flurbiprofen ring distances were somewhat greater.

It should be noted that when dapsone and flurbiprofen are each docked individually into the empty active site of CYP2C9, their resulting eHiTS scores (-3.94 pKi and -5.79 pKi, respectively) are both less negative than the eHiTS score for dapsone docked into the active site in the presence of flurbiprofen (-7.90 pKi). Also less negative is the score generated by docking flurbiprofen into a dapsone-CYP2C9 complex (-6.03 pKi). This supports the QM/MM data described in Chapter IV, and suggests several things. First, that dapsone energetically favors binding as an effector rather than binding as a substrate in the presence of flurbiprofen (this has been demonstrated in previous literature⁴²; second, that this preference is due to the presence of flurbiprofen and likely results from direct interactions with flurbiprofen; lastly, because the docking run was conducted inside the active site of CYP2C9 and not in a vacuum like the QM/MM calculations, the docking results indicate that the dapsone-flurbiprofen complex was not lower in energy than the compounds comprising it individually due to a lack of possibly important protein-effector interactions. Rather, active site residue-effector interactions do not outweigh substrate-effector interactions as factors that stabilize the dapsone-flurbiprofen complex in the active site of CYP2C9.

The fact that the score for dapsone docked in the presence of flurbiprofen was also lower than the score for flurbiprofen docked alone in the active site is a strong indicator that it is more energetically favorable for flurbiprofen to assume a metabolite-

producing binding pose in the presence of dapsone than it is in the absence of the effector. This points to dapsone-flurbiprofen interactions as being important factors in heteroactivation in this biological system, and suggests that dapsone activates flurbiprofen metabolism directly, as opposed to the hypothesis that dapsone may bring about heteroactivation indirectly by causing the substrate to interact with different active site residues.

5.4. Conclusion

Dapsone assumed both a "sulfone oxygens up" and an alternative, "sulfone oxygens down" pose. It is likely that dapsone has the ability to dock in either pose in the active site of CYP2C9. In addition to the added stabilizing effects of H-bonding with flurbiprofen carboxy oxygens, which was only observed in the non-alternative conformer, the "sulfones up" conformation may be more common due to a shorter average distance between stacking rings with superior stacking geometry as seen in the pi-interactions between dapsone and flurbiprofen. Interestingly, in all eHiTS runs pi-stacking between the dapsone and flurbiprofen was either as strong as or superior to that between dapsone and active site residues based on ring geometry and distance, leading to the conclusion that direct interaction with the substrate is at least in part responsible for dapsone activation of flurbiprofen metabolism by CYP2C9.

In all runs, effectors that assumed the alternative pose overlapped more closely with each other such that one of their aromatic rings was in position to better form pi-

interactions with Ring 1 of flurbiprofen. Effectors that docked with sulfones directed toward the F-G loop did not superimpose quite as well; they tended to deviate just far enough from the alternative conformer rings to be able to stack with Ring 2 of flurbiprofen in addition to or instead of Ring 1, depending on the effector. Higher-order stacking interactions such as trimers between the substrate and effectors could also be formed, as observed in the enrichment-trained run. All effectors engaged in stacking interactions with the substrate. Also, effector rings involved in those interactions regardless of the docked conformation were nearly always *p*-substituted with an electron-donating group, which, depending on the group, would add electron density to the attached ring and thus enable it to form stronger pi-interactions.

Substituent effects appeared to play an important role in determining the degree of heteroactivation elicited by a given effector compound, with increasing electron-donating ability of a substituent corresponding to greater *in vitro* heteroactivation. Alteration in electron densities of rings involved in activator-substrate pi-stacking resulted in a parabolic relationship between the *in vitro*/CoMFA-predicted data and overall scores calculated by eHiTS after enrichment training of the receptor; as heteroactivator-substrate pi-stacking interactions became stronger (due to substituent effects on the participating ring of the effector) flurbiprofen could be held more securely in a position conducive to 4'-oxidation. At a certain point, depending on the activator, the pi-interactions became strong enough that they altered the electron density of the substrate ring(s) such that the

4'- hydrogen becomes less inclined to react with the heme, resulting in a parabola when the data is fitted to a line.

CHAPTER VI

Development of a Statistical Model for Effectors of Flurbiprofen Metabolism in CYP2C9

6.1. Introduction

The correlation coefficient of determination calculated in Chapters IV and V, R^2 , is a measure of the internal validity of a statistical model. High R^2 values indicate that the training set data correlates well with the experimental values, and therefore indicate that the model is deemed accurate in predicting those compounds used in its development. However, R^2 does not provide a measure of the ability of the model to correctly predict activities of compounds not present in the test set. A cross-validated correlation coefficient, q^2 , provides a means to assess the overall predictive power of the model, i.e., how well the model can predict the biological activities of external compounds. There are several methods employed in cross-validating a model, the most common being the leave-x-out algorithm, and activity permutation.

The leave-x-out method ascertains whether or not the model is being driven by x number of compounds. The “x” in this case is not a variable in an equation per se, but rather an arbitrary number of compounds in the training set. The rationale behind this method is that a robust model should be able to be “rebuilt” sans x number of compounds from its original training set and still correctly predict the activities of those compounds that were left out of the rebuilt version. The validation is therefore carried out by building additional statistical models using the original training set of the model being validated, but in each test model one or more compounds in the training set is left out. Fitting each rebuilt model will result in a correlation equation. The activity value predicted by the original model for the compound left out of the rebuilt model is then substituted for x in the correlation equation obtained from fitting the rebuilt model, and is used to solve for

the dependent variable y . The result is \bar{y} , which ideally will be numerically close to y . This process is repeated until a model lacking at least one of every compound in the training set for the original model has been constructed and its data fitted. \bar{y} values calculated for the models can be correlated to the *in vitro*/CoMFA-predicted data by calculating a cross-validated correlation coefficient, q^2 , using equation 6.1:

$$q^2 = 1 - [\sum(y - \bar{y})^2 / \sum y^2] \quad (6.1.)$$

where y is the observed biological activity and \bar{y} is the cross-validated calculated activity. The q^2 value can range anywhere from -1 to 1, with values closer to 1 indicating a higher correlation between the calculated and observed activities and negative values indicating no correlation whatsoever-- literally, a correlation worse than "no model at all".

Generally speaking, the greater number of compounds left out of the cross-validation test models, the more reliable the test. However, the leave-one-out algorithm (LOO) has shown itself to suffice when the original training set consists of roughly 30 compounds or less, and has been successfully employed in the development and verification of several pharmacophore models^{47, 49, 54, 71, 73}. A large training set requires a more stringent algorithm (such as leave-20%-out) since more compounds in the set means more "noise" (independent variables that contribute very little to the biological activity, but in large numbers can appear deceptively strong). With enough "noise", the q^2 is more likely to approach 1.0 but may in fact be statistically irrelevant. Leaving out a greater number of compounds increases the "signal-to-noise" ratio, thus identifying the

factors most responsible for the biological activity and ensuring that it is they, not the "noise", that are actually driving the model.

Another method of cross-validation is by permuting the biological activities of the training set compounds. Permutation is defined as the repeated randomization of the activities of the training set compounds such that each value is no longer assigned to the original molecule. A model, deemed the "null hypothesis" due to the fact that the data is presumed to have no relationship to the observed activity, is then built using the permuted values. The correlation coefficient determined from this model should be statistically insignificant; if that is in fact not the case, then the actual, non-permuted, model may be due more in part to chance. The difference between the estimated biological activity generated by the model and the experimental activity, the total energy cost, is also calculated for both the permuted and the actual models. The greater the difference in total energy costs of the null hypothesis and the actual model, the smaller the likelihood that chance is responsible for correlation of the activities predicted by the actual model to those observed. A model that yields a poor q^2 can be refined and retested until a statistically significant q^2 can be obtained from it. Once the model is deemed sound, an external database of compounds with unknown biological activities can be screened, and those identified by the model as probable effectors can be subsequently evaluated *in vitro* to assess their actual biological activity.

6.2. Materials and Methods

6.2.1. Cross-Validation of the Model

The LOO algorithm was used to cross-validate the model. Eleven statistical models were generated by eHiTS using the enrichment-trained 2C9-flurbiprofen complex receptor, and test set activator compounds only; in each model one activator was not included in model development. The data was fitted using polynomial regression in Excel, and the resulting regression coefficients were applied to the overall score of the excluded compound as calculated by eHiTS. From this equation \bar{y} was solved. This process was repeated for every activator in the training set. All \bar{y} were summed and a q^2 was calculated using equation 6.1, above. It should be noted that, although the original model was generated using only the *in vitro* values, the cross-validation was performed using all 19 dapsone analogs.

6.2.2. NCI Database Screening

The National Cancer Institute (NCI) database of drug-like compounds was chosen as the external data set to be screened due to the fact that the compounds that comprised it were known to possess pharmacological properties, though their biological activity regarding flurbiprofen metabolism by CYP2C9 was unknown. In addition, many compounds could be obtained from commercial sources should they require *in vitro* metabolism studies. The first 858 compounds were arbitrarily excised from the database and were subjected to a molecular weight filter using the Concord module of Sybyl7.0 (Tripos, Inc. St. Louis, MO) that excluded compounds below 120 amu and those above 350 amu. Metallic or isotopic compounds were also excluded. Because dapsone analogs require aromaticity to activate flurbiprofen turnover, NCI compounds were required to possess at least one aromatic ring. Potential chirality of the compounds was included. The

resulting filtered database was converted to 3D coordinates and then screened by eHiTS using the enrichment-trained 2C9-flurbiprofen complex receptor with the same protocol described in Chapter V. Results of the screening were visualized using PyMol (DeLano Scientific, Palo Alto, CA) and CheVi.

6.3. Results and Discussion

The results of the LOO cross-validation can be viewed in **Table 6.1**. Substituting the value obtained by solving the fitted equation for each compound in the training set one-by-one yielded high correlation coefficients for every compound, which indicates that no single compound in the training set was driving the model and that the model was therefore robust. The q^2 value calculated from these coefficients was statistically significant at 0.6966, suggesting that the model possessed predictive power sufficient to reliably screen external databases for unknown heteroactivators. To this end, a small segment of the NCI database consisting of the first 60 compounds was screened through the model. As a control, the database was seeded with known dapsone analog activators from the test set used to develop the model. eHiTS ranked 6 out of 9 total heteroactivating dapsone analogs (67%) in the top 16 docked compounds and identified several NCI compounds as activators (**Figure 6.1**). The new compounds identified possessed structural similarities to dapsone (aromatic rings separated by one or more rotatable bonds), and took on a similar docked geometry to dapsone as well (**Figure 6.2**).

Table 6.1. Cross-Validation Results

Compound Left Out And Solved For	Regression Coefficients			\bar{y}	y	R^2
	A	B	C			
Diphenylsulfoxide (14)	-2.8317	-94.985	-732.80	21.568	5.03	0.5660
Diphenylsulfone (10)	-3.1442	-106.31	-832.03	58.411	39.98	0.6299
Diparatolylsulfone (11)	-3.0342	-102.53	-801.66	58.618	50.57	0.5888
Dapsone (12)	-2.7887	-94.089	-733.81	48.388	64	0.5711
Dacetamidophenylsulfoxide (14)	-2.5536	-87.450	-687.64	28.529	5.03	0.5918
Dibenzothiophenesulfone (16)	-3.5477	-121.46	-975.36	-20.336	17.82	0.6737
Hydroxyphenylmethoxyphenylsulfone (17)	-3.0033	-100.68	-778.18	36.725	17.92	0.6147
Diacetamidophenylsulfone (18)	-3.0808	-103.73	-809.50	2.111	8	0.5532
Dihydroxyphenylsulfone (19)	-2.3590	-79.502	-617.51	45.058	95	0.7047
Methoxyphenylsulfone (20)	-3.1611	-106.78	-835.21	61.186	44.84	0.6178

Leave-one-out cross-validation calculation results are shown. The regression coefficients were calculated for a second-order polynomial function, where $\bar{y} = Ax^2 + Bx + C$.

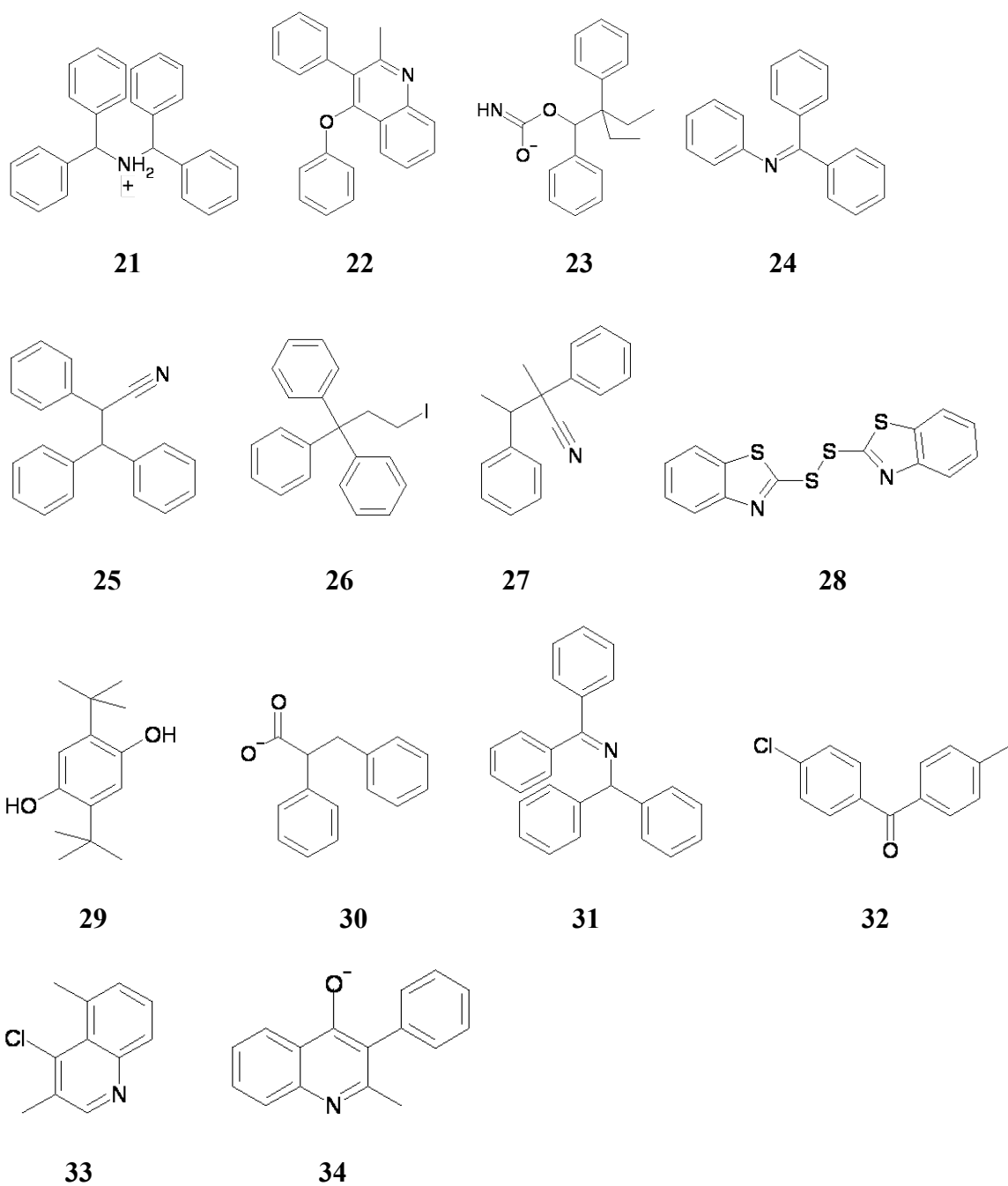


Figure 6.1. NCI Compounds in the Top 20 Identified by eHiTS As Activators of Flurbiprofen Metabolism by CYP2C9.

NCI compounds in the top 20 compounds identified by eHiTS as heteroactivators of flurbiprofen 4'-hydroxylation by CYP2C9 are illustrated.

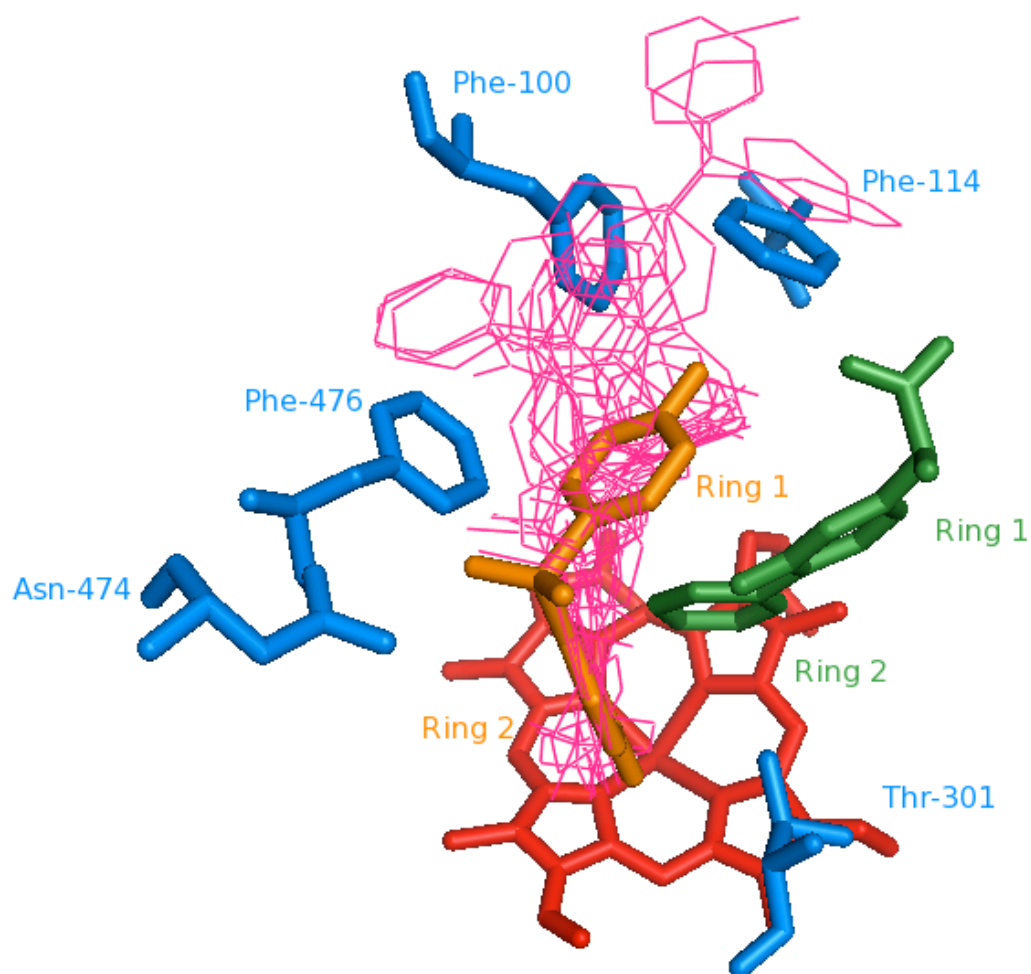


Figure 6.2. NCI Compounds Identified by eHiTS as Activators Docked in the Active Site of CYP2C9 in the Presence of Flurbiprofen.

NCI compounds identified as activators of flurbiprofen metabolism are shown overlaid with dapson for comparison. The NCI compounds are in magenta as lines, dapson is tangerine, flurbiprofen is forest, active site residues are marine, and the heme is crimson.

eHiTS ranked several NCI compounds with multiple rings (compounds **21**, **22**, **23**, **24**, **25** and **26**) higher than any of the dapsones analog activators. These compounds were able to interact more closely with active site residues than the dapsones analogs, and all but one (**26**) formed pi-stacking interactions with flurbiprofen simultaneously. This may in part be due to the greater number of rotatable bonds separating the aromatic rings involved in pi-stacking, which not only allows the effector to span the distance between the active site residues and the substrate, but also to rotate its rings to maximize orbital overlap. These multi-ring effectors tended to position themselves such that they were able to form interactions with both rings of flurbiprofen in an offset geometry, but if the structure of the effector prevented this type of pose, the effector would stack with Ring 1 of flurbiprofen rather than Ring 2, possibly because this arrangement enabled the effector to favorably interact with active site phenylalanine residues simultaneously. Interestingly, three of these six high-ranking multi-ring compounds (**21**, **23**, and **26**) formed a pinwheel-like trimer between one effector ring, Phe-100, and Phe-114 (**Figure 6.3**) and (**Figure 6.4**). The other three (**22**, **24**, and **25**) formed dimers with Phe-476. The ability of these effectors to form complex interactions with aromatic residues may stabilize them in position to better pi-stack with the substrate, which may account for their higher ranking as effectors by the docking program.

In vitro testing of the top 20 docked compounds from the NCI database seeded with dapsones analogs confirmed 4 compounds from the as activators of flurbiprofen metabolism by CYP2C9 (**Table 6.2**). Three of these compounds (**24**, **25**, and **28**) possessed activities ranging from a 4.26% increase in 4'-hydroxyflurbiprofen production

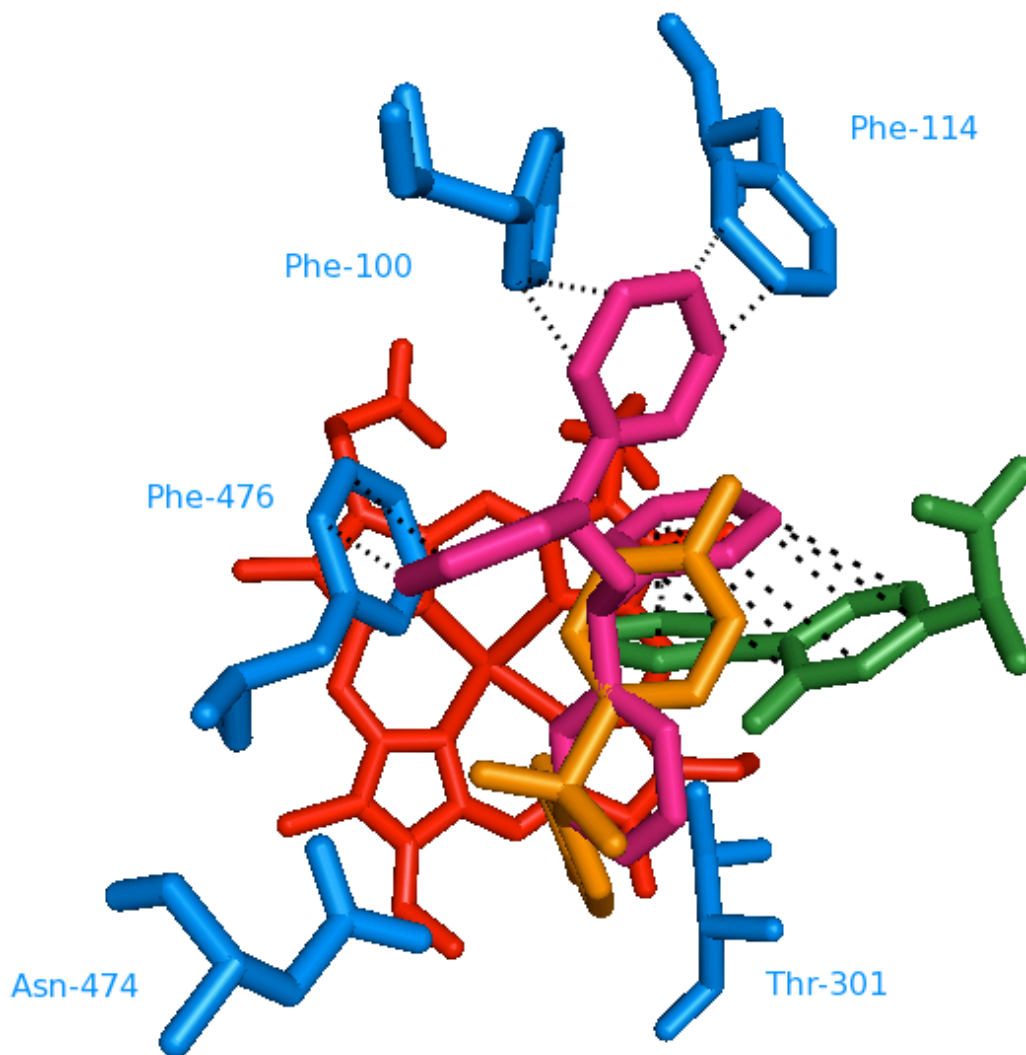


Figure 6.3. Pi-Interactions Formed by Compound 21 Docked in the Active Site of CYP2C9 in the Presence of Flurbiprofen.

Compound **21** is shown docked in the presence of flurbiprofen in the active site of CYP2C9. Dapsone is overlaid in comparison. **21** is magenta, dapsone is tangerine, flurbiprofen is forest, active site residues are marine, and the heme is crimson. Pi-interactions are denoted by dotted lines.

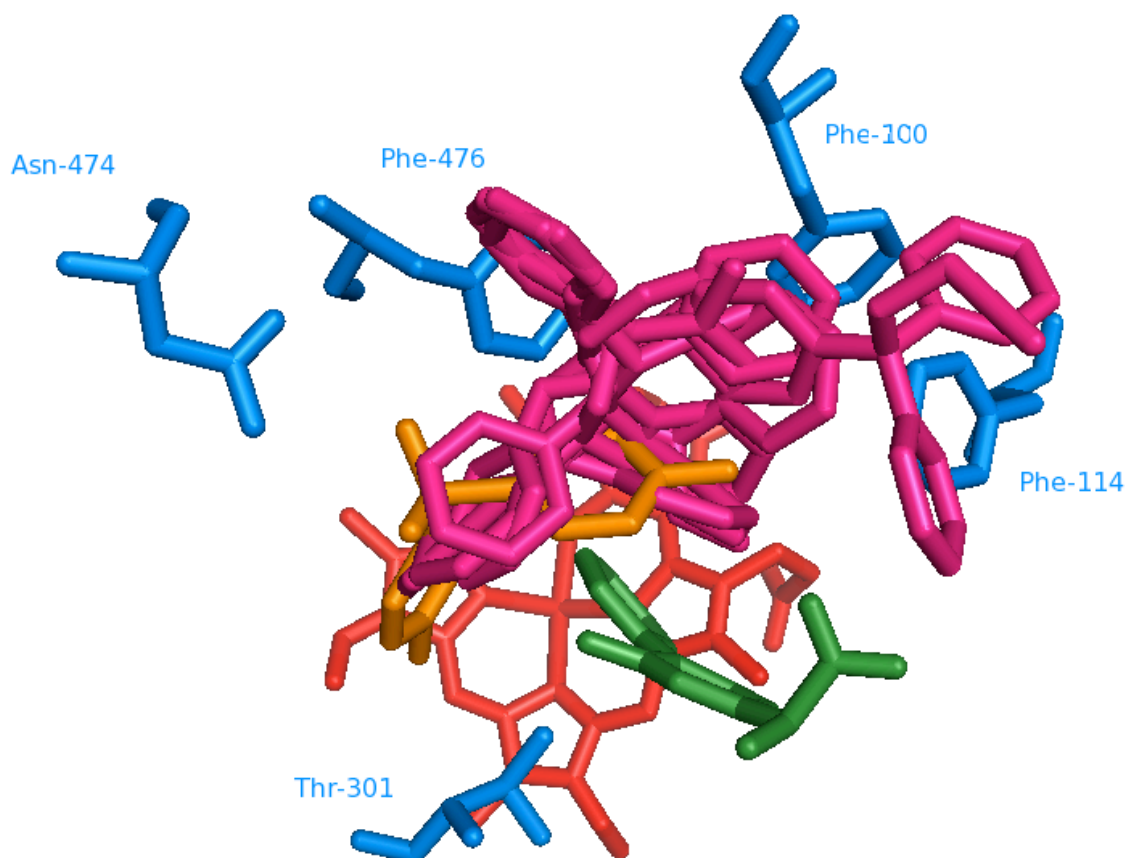


Figure 6.4. Docked NCI Compounds Identified by eHiTS as Activators of Flurbiprofen Metabolism by CYP2C9 Possessing Three Or More Rings.

Docked NCI compounds identified by eHiTS as heteroactivators of flurbiprofen metabolism by CYP2C9 possessing three or more rings are illustrated. Dapsone is overlaid for comparison. The NCI compounds are magenta, dapsone is tangerine, active site residues are marine, flurbiprofen is forest, and the heme is crimson.

Table 6.2. Activators of Flurbiprofen Metabolism by CYP2C9 Identified in the Seeded NCI Database by eHiTS.

Rank	Compound Number	Name	% change in flurbiprofen metabolism at 5 μ M effector concentration	% change in flurbiprofen metabolism at 10 μ M effector concentration
4	24	N-(diphenyl methylene) aniline	4.26	-7.95
5	25	α - β -diphenylhydrocinnamionitrile	5.23	-2.61
8	28	2,2'-dithiobis(benzothiazole)	7.82	-20.55
10	17	4-[(4-methoxyphenyl)sulfonyl] phenol	115.20	140.14
27	12	Dapsone	16.56	79.52
54	35	3-methyl-1-phenyl-2pyrazolin-5-one	-16.78	-30.50

Positive numbers indicate activation; negative numbers indicate inhibition.
Flurbiprofen concentration held constant at 10 μ M for all tests.

New activators of flurbiprofen 4'-hydroxylation by CYP2C9 identified by screening the seeded NCI database using eHiTS and their *in vitro* activities are listed.

to 7.82% at a low effector concentration of 5 μ M (flurbiprofen concentration was 10 μ M). For perspective, at this low concentration dapsone activated flurbiprofen metabolism by 16.56%, which marks the newly identified effectors as moderate activators. However, raising the effector concentration to 50 μ M turns these same 3 compounds into inhibitors of flurbiprofen metabolism. It is possible that these effectors have an affinities for the substrate binding site similar to that of flurbiprofen, but also have enough of an affinity for the effector binding site to serve as activators at concentrations lower than the concentration of substrate, when flurbiprofen is more likely to occupy the substrate binding site due to being present in solution in greater numbers compared to the effectors. At high effector concentrations the effector compounds outnumber the substrate in solution, increasing the chances that an effector will occupy the substrate binding site before flurbiprofen can, and thereby causing competitive inhibition.

The fourth activator compound identified by the model was one dapsone analog that was not part of the set of compounds used to develop the model due to its lack of *in vitro* activity values was scored by eHiTS as an activator stronger than dapsone. This compound, **17**, was ranked 10th whereas dapsone lagged behind at 27th place. The CoMFA model predicted **17** would activate flurbiprofen metabolism by CYP2C9 by about 18%, as seen in **Table 4.1**. However, *in vitro* testing revealed a much higher activity than the CoMFA model had predicted at 115.2%. This percent turnover increase was obtained using concentrations of 5 μ M activator and 10 μ M flurbiprofen. At effector concentrations of 50 μ M the rate of metabolism increases to 140.14%. Compared to the other three new activators, this compound may have a greater affinity for the effector

binding site and less affinity for the substrate binding site, or it may form stronger interactions with flurbiprofen in the active site that stabilize it in the effector binding pocket, or, most likely, a combination of these two mechanisms.

It is important to note that the model not only identified a compound not used in its development as an activator of flurbiprofen metabolism by CYP2C9, but also correctly ranked its activity above that of dapsone (which the CoMFA model that predicted its original activity value failed to accomplish, even with a correlation coefficient of 0.80). The model also correctly ranked a strong unknown inhibitor (**35**) of flurbiprofen metabolism toward the bottom of the list of 60 compounds at 54 (**Figure 6.5**) The compound inhibited substrate turnover by -16.78% at 5 μ M and -30.50% at 10 μ M.

The discovery of these four compounds indicated that the model is able to successfully identify potential effectors whose *in vitro* activity is unknown. With the establishment that the model possesses significant predictive power as determined by cross-validation and has practical application as demonstrated by the seeded NCI database screen, it is very possible that other, possibly stronger, effectors would be discovered by sampling larger portions of the database (only 60 of the total 740 filtered compounds were screened); a larger sampling pool increases the chances that the model would flag effectors in addition to the four already identified.

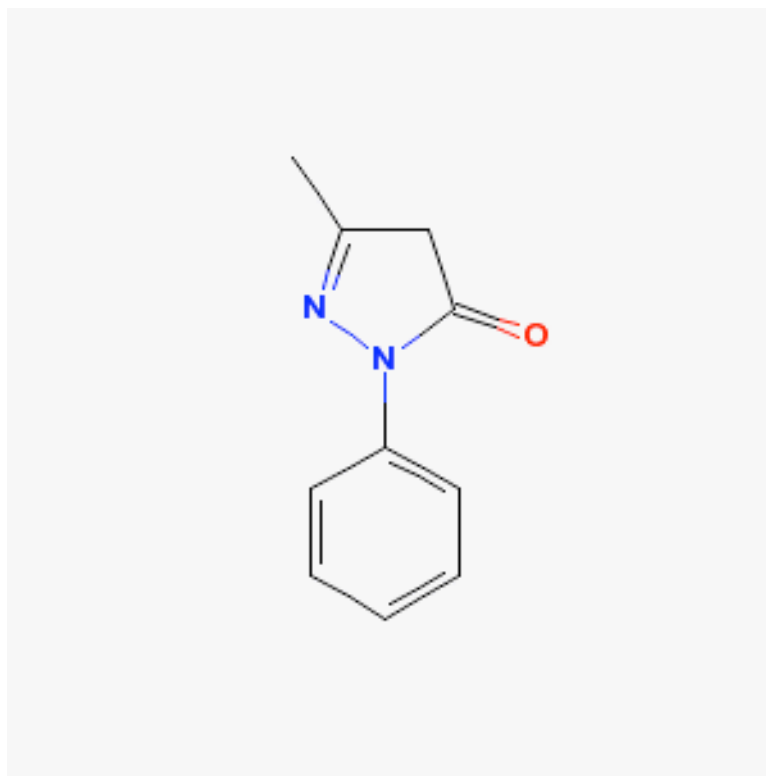


Figure 6.5. Chemical Structure of 3-methyl-1-phenyl-2pyrazolin-5-one.

The chemical structure of 3-methyl-1-phenyl-2pyrazolin-5-one (compound **35**) is shown.

CHAPTER VII

Summary and Conclusions

The modeling studies conducted and explained in this work have brought about greater understanding of the mechanisms responsible for dapsone-mediated activation of flurbiprofen metabolism by CYP2C9. It is especially worth noting that effector-substrate interactions, which had not been given much attention in previous literature, were shown to be of prime importance in effecting heteroactivation. Armed with this information, a statistical model capable of high-throughput screening of large databases has been developed, making the prediction of the likelihood of a given compound to possess activator qualities now possible for this biological system.

Docking studies using both AutoDock and eHiTS and molecular dynamics simulations revealed that both H-bonding and pi-stacking played important roles in the mechanism of heteroactivation, with Asn-474 and Ser-209 H-bonding to dapsone, and dapsone (as well as dapsone analogs) pi-stacking in various geometries with flurbiprofen and active site residues, particularly Phe-476, Phe-114, and Phe-100. Effectors typically formed higher-order stacking conformations involving the substrate and active site residues. MD simulations demonstrated that dapsone, when inside the active site of CYP2C9 in the presence of flurbiprofen, moved the 4'-H of the substrate closer to the heme iron by about 1 Å, which has been shown to occur experimentally via previous NMR studies¹⁰⁰. At times this distance reached values as low as 3.27 Å. Interestingly, the simulations also revealed that the 4'-H-Fe distance actually increased when pi-stacking rings of dapsone and flurbiprofen came within an average of about 4 Å of each other. As the substrate-effector distance widened to an average of 6.12 Å (it fluctuated between about 4 and 8.5 Å) flurbiprofen approached the heme more readily. During the time that dapsone was interacting more weakly with flurbiprofen, it was interacting more strongly with Phe-476.

This active site residue often formed trimer stacking orientations with the substrate as well as with dapson. There existed also a period of time during the simulation in which dapson closely associated with both Phe-476 and flurbiprofen, pulling the substrate away from the heme in the process to the point where metabolism was unlikely to occur. These states likely exist in equilibrium, with the weaker dapson-flurbiprofen interactive state predominating.

This data, which shows stronger effector-substrate interactions leading to a less-metabolite-producing scenario and weaker ones leading to a more productive scenario, is corroborated by the statistical model developed by eHiTS, which yielded a significant parabolic correlation between the interaction-based score it calculates for each effector and *in vitro* effector activity ($R^2 = 0.60$). The parabola suggests that the effector can stack with flurbiprofen with increasing strength up to a certain interaction energy with the effect that the substrate is increasingly better-stabilized in a position closer to the heme where it is readily oxidized. When the interaction energy between the two compounds exceeds a certain point, the effector is able to pull the substrate away from the heme, and it is less likely to be metabolized. The heteroactivator potency appeared to be related to the electron-donating capacity of functional groups attached *para* to the sulfone/sulfide/sulfoxide group of the effector rings. In general, the stronger the electron donating group, the greater electron-density of the effector ring, the stronger the interaction between that ring and the aromatic rings of flurbiprofen and active site phenylalanines.

The eHiTS scores also indicated that interactions between the substrate and the effector itself, not active site residues, are primarily responsible for heteroactivation of

flurbiprofen metabolism. The score generated from the docking run in which flurbiprofen was docked into a dapsone-CYP2C9 complex was superior to the scores of both the dapsone docked into the flp-CYP2C9 complex run and the run in which flurbiprofen was docked into CYP2C9 alone. The data suggests that dapsone plays a direct role in stabilizing flurbiprofen in a productive position in the active site. In addition, quantum mechanics studies of docked conformations of seven effector-flurbiprofen complexes revealed that in the case of all activators but two the complexes were more stable (i.e. had lower energies) than either flurbiprofen alone or the effector alone.

The statistical model developed using the automated training function in eHiTS not only yielded a significant cross-validated correlation coefficient of 0.70, but, upon screening 60 drug-like compounds from the NCI database, was also able to successfully predict four new activators of flurbiprofen 4'-hydroxylation by CYP2C9. One of these was shown *in vitro* to exceed the level of heteroactivity of dapsone, increasing flurbiprofen turnover by as much as 140%. The model also correctly ranked a strong unknown inhibitor (**35**) of flurbiprofen metabolism at 54, near the bottom of the list of a total of 60 screened compounds. The discovery of the activities of these five compounds indicated that the model is able to successfully identify potential effectors whose *in vitro* activity is unknown. With the establishment that the model possesses significant predictive power as determined by cross-validation, and has practical application as demonstrated by the seeded NCI database screen, it is very possible that other, possibly stronger, effectors would be discovered by sampling larger portions of the database (only 60 of the total 740 filtered compounds were screened); a larger sampling pool increases the chances that the model would flag effectors in addition to the four already identified.

Though exhibition of atypical kinetics in an enzyme is substrate-specific, the effector-substrate interactions discovered here may also apply to other biological systems, and the methodology used in developing the predictive statistical model could likely also be used for other systems with success. Exploration of such interactions in these systems would be worth pursuing in the future, particularly those involving drugs with narrow therapeutic indices such as warfarin. For the reasons given above, this work has increased the depth of knowledge with which drug-drug interactions stemming from atypical enzyme kinetics can be prevented.

REFERENCES

1. Shimada T, Yamazaki H, Mimura M, Inui Y, Guengerich FP. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: Studies with liver microsomes of 30 Japanese and 30 Caucasians. *J Pharmacol Exp Ther* 1994 Jul;270(1):414-23.
2. Rettie AE, Jones JP. Clinical and toxicological relevance of CYP2C9: Drug-drug interactions and pharmacogenetics. *Annu Rev Pharmacol Toxicol* 2005;45:477-94.
3. Guengerich FP. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* 2001 Jun;14(6):611-50.
4. Jefcoate CR. Measurement of substrate and inhibitor binding to microsomal cytochrome P-450 by optical-difference spectroscopy. *Methods Enzymol* 1978;52:258-79.
5. Dickmann LJ, Locuson CW, Jones JP, Rettie AE. Differential roles of Arg97, Asp293, and Arg108 in enzyme stability and substrate specificity of CYP2C9. *Mol Pharmacol* 2004 Apr;65(4):842-50.
6. Rendic S, Di Carlo FJ. Human cytochrome P450 enzymes: A status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab Rev* 1997 Feb-May;29(1-2):413-580.
7. Jones JP, Shou M, Korzekwa KR. Predicting the regioselectivity and stereoselectivity of cytochrome P450-mediated reactions: Structural models for bioactivation reactions. *Adv Exp Med Biol* 1996;387:355-60.
8. Miners JO, Birkett DJ. Cytochrome P4502C9: An enzyme of major importance in human drug metabolism. *Br J Clin Pharmacol* 1998 Jun;45(6):525-38.
9. Kirchheiner J, Brockmoller J. Clinical consequences of cytochrome P450 2C9 polymorphisms. *Clin Pharmacol Ther* 2005 Jan;77(1):1-16.
10. Dickmann LJ, Rettie AE, Kneller MB, Kim RB, Wood AJ, Stein CM, Wilkinson GR, Schwarz UI. Identification and functional characterization of a new CYP2C9 variant

(CYP2C9*5) expressed among african americans. *Mol Pharmacol* 2001 Aug;60(2):382-7.

11. Ridderstrom M, Zamora I, Fjellstrom O, 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* 2001 Nov 22;44(24):4072-81.
12. 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-A resolution. *J Biol Chem* 2004 Aug 20;279(34):35630-7.
13. Haining RL, Jones JP, Henne KR, Fisher MB, Koop DR, Trager WF, Rettie AE. Enzymatic determinants of the substrate specificity of CYP2C9: Role of B'-C loop residues in providing the pi-stacking anchor site for warfarin binding. *Biochemistry* 1999 Mar 16;38(11):3285-92.
14. Klose TS, Ibeanu GC, Ghanayem BI, Pedersen LG, Li L, Hall SD, Goldstein JA. Identification of residues 286 and 289 as critical for conferring substrate specificity of human CYP2C9 for diclofenac and ibuprofen. *Arch Biochem Biophys* 1998 Sep 15;357(2):240-8.
15. Jung F, Griffin KJ, Song W, Richardson TH, Yang M, Johnson EF. Identification of amino acid substitutions that confer a high affinity for sulfaphenazole binding and a high catalytic efficiency for warfarin metabolism to P450 2C19. *Biochemistry* 1998 Nov 17;37(46):16270-9.
16. Ridderstrom M, Masimirembwa C, Trump-Kallmeyer S, Ahlefeldt M, Otter C, Andersson TB. Arginines 97 and 108 in CYP2C9 are important determinants of the catalytic function. *Biochem Biophys Res Commun* 2000 Apr 21;270(3):983-7.
17. Davies C, Witham K, Scott JR, Pearson A, DeVoss JJ, Graham SE, Gillam EM. Assessment of arginine 97 and lysine 72 as determinants of substrate specificity in cytochrome P450 2C9 (CYP2C9). *Drug Metab Dispos* 2004 Apr;32(4):431-6.
18. Seifert A, Tatzel S, Schmid RD, Pleiss J. Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: The molecular basis of substrate binding and regioselectivity toward warfarin. *Proteins* 2006 Jul 1;64(1):147-55.

19. Williams PA, Cosme J, Ward A, Angove HC, Matak Vinkovic D, Jhoti H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* 2003 Jul 24;424(6947):464-8.
20. Locuson CW, Tracy TS. Identification of binding sites of non-I-helix water molecules in mammalian cytochromes p450. *Drug Metab Dispos* 2006 Dec;34(12):1954-7.
21. Transon C, Leemann T, Vogt N, Dayer P. In vivo inhibition profile of cytochrome P450TB (CYP2C9) by (+/-)-fluvastatin. *Clin Pharmacol Ther* 1995 Oct;58(4):412-7.
22. Kaukonen KM, Olkkola KT, Neuvonen PJ. Fluconazole but not itraconazole decreases the metabolism of losartan to E-3174. *Eur J Clin Pharmacol* 1998 Feb;53(6):445-9.
23. Miller RR, Porter J, Greenblatt DJ. Clinical importance of the interaction of phenytoin and isoniazid: A report from the boston collaborative drug surveillance program. *Chest* 1979 Mar;75(3):356-8.
24. Runkel R, Mroszczak E, Chaplin M, Sevelius H, Segre E. Naproxen-probenecid interaction. *Clin Pharmacol Ther* 1978 Dec;24(6):706-13.
25. Kim KY, Mancano MA. Fenofibrate potentiates warfarin effects. *Ann Pharmacother* 2003 Feb;37(2):212-5.
26. Seaton TL, Celum CL, Black DJ. Possible potentiation of warfarin by fluconazole. *DICP* 1990 Dec;24(12):1177-8.
27. Almog S, Shafran N, Halkin H, Weiss P, Farfel Z, Martinowitz U, Bank H. Mechanism of warfarin potentiation by amiodarone: Dose--and concentration--dependent inhibition of warfarin elimination. *Eur J Clin Pharmacol* 1985;28(3):257-61.
28. Wang RW, Newton DJ, Liu N, Atkins WM, Lu AY. Human cytochrome P-450 3A4: In vitro drug-drug interaction patterns are substrate-dependent. *Drug Metab Dispos* 2000 Mar;28(3):360-6.
29. Hutzler JM, Powers FJ, Wynalda MA, Wienkers LC. Effect of carbonate anion on cytochrome P450 2D6-mediated metabolism in vitro: The potential role of multiple

- oxygenating species. Arch Biochem Biophys 2003 Sep 15;417(2):165-75.
30. Atkins WM. Non-michaelis-menten kinetics in cytochrome P450-catalyzed reactions. Annu Rev Pharmacol Toxicol 2005;45:291-310.
 31. Hutzler JM, Hauer MJ, Tracy TS. Dapsone activation of CYP2C9-mediated metabolism: Evidence for activation of multiple substrates and a two-site model. Drug Metab Dispos 2001 Jul;29(7):1029-34.
 32. Hutzler JM, Kolwankar D, Hummel MA, Tracy TS. Activation of CYP2C9-mediated metabolism by a series of dapsone analogs: Kinetics and structural requirements. Drug Metab Dispos 2002 Nov;30(11):1194-200.
 33. Korzekwa KR, Krishnamachary N, Shou M, Ogai A, Parise RA, Rettie AE, Gonzalez FJ, Tracy TS. Evaluation of atypical cytochrome P450 kinetics with two-substrate models: Evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites. Biochemistry 1998 Mar 24;37(12):4137-47.
 34. Hummel MA, Dickmann LJ, Rettie AE, Haining RL, Tracy TS. Differential activation of CYP2C9 variants by dapsone. Biochem Pharmacol 2004 May 15;67(10):1831-41.
 35. Shou M, Grogan J, Mancewicz JA, Krausz KW, Gonzalez FJ, Gelboin HV, Korzekwa KR. Activation of CYP3A4: Evidence for the simultaneous binding of two substrates in a cytochrome P450 active site. Biochemistry 1994 May 31;33(21):6450-5.
 36. Harlow GR, Halpert JR. Analysis of human cytochrome P450 3A4 cooperativity: Construction and characterization of a site-directed mutant that displays hyperbolic steroid hydroxylation kinetics. Proc Natl Acad Sci U S A 1998 Jun 9;95(12):6636-41.
 37. Yoon MY, Campbell AP, Atkins WM. "Allosterism" in the elementary steps of the cytochrome P450 reaction cycle. Drug Metab Rev 2004 May;36(2):219-30.
 38. Hummel MA, Gannett PM, Aguilar JS, Tracy TS. Effector-mediated alteration of substrate orientation in cytochrome P450 2C9. Biochemistry 2004 Jun 8;43(22):7207-14.

39. McMasters DR, Torres RA, Crathern SJ, Dooney DL, Nachbar RB, Sheridan RP, Korzekwa KR. Inhibition of recombinant cytochrome P450 isoforms 2D6 and 2C9 by diverse drug-like molecules. *J Med Chem* 2007 Jul 12;50(14):3205-13.
40. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J Med Chem* 2004 Mar 25;47(7):1739-49.
41. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: A new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *J Med Chem* 2004 Mar 25;47(7):1750-9.
42. Locuson CW, Gannett PM, Ayscue R, Tracy TS. Use of simple docking methods to screen a virtual library for heteroactivators of cytochrome P450 2C9. *J Med Chem* 2007 Mar 22;50(6):1158-65.
43. Egnell AC, Houston JB, Boyer CS. Predictive models of CYP3A4 heteroactivation: In vitro-in vivo scaling and pharmacophore modeling. *J Pharmacol Exp Ther* 2005 Mar;312(3):926-37.
44. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* 2002 Jan 1;46(1):34-40.
45. Afzelius L, Zamora I, Masimirembwa CM, Karlen A, Andersson TB, Mecucci S, Baroni M, Cruciani G. Conformer- and alignment-independent model for predicting structurally diverse competitive CYP2C9 inhibitors. *J Med Chem* 2004 Feb 12;47(4):907-14.
46. Afzelius L, Zamora I, Ridderstrom M, Andersson TB, Karlen A, Masimirembwa CM. Competitive CYP2C9 inhibitors: Enzyme inhibition studies, protein homology modeling, and three-dimensional quantitative structure-activity relationship analysis. *Mol Pharmacol* 2001 Apr;59(4):909-19.
47. Afzelius L, Masimirembwa 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* 2002 Jul;16(7):443-58.

48. de Groot MJ, Alex AA, Jones BC. Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *J Med Chem* 2002 May 9;45(10):1983-93.
49. Egnell AC, Eriksson C, Albertson N, Houston B, Boyer S. Generation and evaluation of a CYP2C9 heteroactivation pharmacophore. *J Pharmacol Exp Ther* 2003 Dec;307(3):878-87.
50. 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* 2000 Aug;28(8):994-1002.
51. Schuster D, Laggner C, Steindl TM, Langer T. Development and validation of an in silico P450 profiler based on pharmacophore models. *Curr Drug Discov Technol* 2006 Mar;3(1):1-48.
52. Hodgson J. ADMET--turning chemicals into drugs. *Nat Biotechnol* 2001 Aug;19(8):722-6.
53. Stouch TR, Kenyon JR, Johnson SR, Chen XQ, Doweyko A, Li Y. In silico ADME/Tox: Why models fail. *J Comput Aided Mol Des* 2003 Feb-Apr;17(2-4):83-92.
54. 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* 2000 Jul 27;43(15):2789-96.
55. 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* 2005 Jul;33(7):873-8.
56. Lewis DF, Modi S, Dickins M. Structure-activity relationship for human cytochrome P450 substrates and inhibitors. *Drug Metab Rev* 2002 Feb-May;34(1-2):69-82.
57. van Drie JH. Pharmacophore discovery--lessons learned. *Curr Pharm Des* 2003;9(20):1649-64.

58. Strobl GR, von Krüedener S, Stockigt J, Guengerich FP, Wolff T. Development of a pharmacophore for inhibition of human liver cytochrome P-450 2D6: Molecular modeling and inhibition studies. *J Med Chem* 1993 Apr 30;36(9):1136-45.
59. 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 CYP2D6 inhibitors. *Pharmacogenetics* 1999 Aug;9(4):477-89.
60. 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* 1999 Jul;290(1):429-38.
61. 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* 2005 Mar;19(3):189-201.
62. Kriegl JM, Eriksson L, Arnhold T, Beck B, Johansson E, Fox T. Multivariate modeling of cytochrome P450 3A4 inhibition. *Eur J Pharm Sci* 2005 Apr;24(5):451-63.
63. Egnell AC, Houston B, Boyer S. In vivo CYP3A4 heteroactivation is a possible mechanism for the drug interaction between felbamate and carbamazepine. *J Pharmacol Exp Ther* 2003 Jun;305(3):1251-62.
64. Rahnasto M, Raunio H, Poso A, Wittekindt C, Juvonen RO. Quantitative structure-activity relationship analysis of inhibitors of the nicotine metabolizing CYP2A6 enzyme. *J Med Chem* 2005 Jan 27;48(2):440-9.
65. Poso A, Gynther J, Juvonen R. A comparative molecular field analysis of cytochrome P450 2A5 and 2A6 inhibitors. *J Comput Aided Mol Des* 2001 Mar;15(3):195-202.
66. Asikainen A, Tarhanen J, Poso A, Pasanen M, Alhava E, Juvonen RO. Predictive value of comparative molecular field analysis modelling of naphthalene inhibition of human CYP2A6 and mouse CYP2A5 enzymes. *Toxicol in Vitro* 2003 Aug;17(4):449-55.
67. Chohan KK, Paine SW, Mistry J, Barton P, Davis AM. A rapid computational filter for cytochrome P450 1A2 inhibition potential of compound libraries. *J Med Chem*

2005 Aug 11;48(16):5154-61.

68. Fuhr U, Strobl G, Manaut F, Anders EM, Sorgel F, Lopez-de-Brinas E, Chu DT, Pernet AG, Mahr G, Sanz F. Quinolone antibacterial agents: Relationship between structure and in vitro inhibition of the human cytochrome P450 isoform CYP1A2. *Mol Pharmacol* 1993 Feb;43(2):191-9.
69. Korhonen LE, Rahnasto M, Mahonen NJ, Wittekindt C, Poso A, Juvonen RO, Raunio H. Predictive three-dimensional quantitative structure-activity relationship of cytochrome P450 1A2 inhibitors. *J Med Chem* 2005 Jun 2;48(11):3808-15.
70. Lee H, Yeom H, Kim YG, Yoon CN, Jin C, Choi JS, Kim BR, Kim DH. Structure-related inhibition of human hepatic caffeine N3-demethylation by naturally occurring flavonoids. *Biochem Pharmacol* 1998 May 1;55(9):1369-75.
71. 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* 2000 Aug;28(8):994-1002.
72. Cho SJ, Tropsha A. Cross-validated R2-guided region selection for comparative molecular field analysis: A simple method to achieve consistent results. *J Med Chem* 1995 Mar 31;38(7):1060-6.
73. Locuson CW, 2nd, Rock DA, Jones JP. Quantitative binding models for CYP2C9 based on benzobromarone analogues. *Biochemistry* 2004 Jun 8;43(22):6948-58.
74. Sinnokrot MO, Valeev EF, Sherrill CD. Estimates of the ab initio limit for pi-pi interactions: The benzene dimer. *J Am Chem Soc* 2002 Sep 11;124(36):10887-93.
75. Lewis DF, Ito Y, Goldfarb PS. Investigating human P450s involved in drug metabolism via homology with high-resolution P450 crystal structures of the CYP2C subfamily. *Curr Drug Metab* 2006 Aug;7(6):589-98.
76. Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE. Mammalian microsomal cytochrome P450 monooxygenase: Structural adaptations for membrane binding and functional diversity. *Mol Cell* 2000 Jan;5(1):121-31.

77. Hummel MA, Locuson CW, Gannett PM, Rock DA, Mosher CM, Rettie AE, Tracy TS. CYP2C9 genotype-dependent effects on in vitro drug-drug interactions: Switching of benzbromarone effect from inhibition to activation in the CYP2C9.3 variant. *Mol Pharmacol* 2005 Sep;68(3):644-51.
78. Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005 Jan-Feb;45(1):177-82.
79. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *J Comput Aided Mol Des* 1996 Aug;10(4):293-304.
80. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. *Proteins* 1990;8(3):195-202.
81. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Jr., Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins and nucleic acids. *J Am Chem Soc* 1995;117:5179-97.
82. Simmerling C, Elber R, Zhang J. MOIL-view - A program for visualization of structure and dynamics of biomolecules and STO- A program for computing stochastic paths. *Modelling of Biomolecular Structure and Mechanisms* 1995:241-65.
83. de Graaf C, Pospisil P, Pos W, Folkers G, Vermeulen NP. Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. *J Med Chem* 2005 Apr 7;48(7):2308-18.
84. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL, 3rd. Assessing scoring functions for protein-ligand interactions. *J Med Chem* 2004 Jun 3;47(12):3032-47.
85. Hu X, Balaz S, Shelver WH. A practical approach to docking of zinc metalloproteinase inhibitors. *J Mol Graph Model* 2004 Mar;22(4):293-307.
86. Lozano JJ, Pastor M, Cruciani G, Gaedt K, Centeno NB, Gago F, Sanz F. 3D-QSAR methods on the basis of ligand-receptor complexes. application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *J Comput Aided Mol*

Des 2000 May;14(4):341-53.

87. Solis FJ, Wets RJ. Minimization by random search techniques. *Mathematical Operations Research* 1981;6:19-30.
88. Giammona DA. Ph. D. thesis. University of California, Davis; 1984.
89. Oda A, Yamaotsu N, Hirono S. New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J Comput Chem* 2005 Jun;26(8):818-26.
90. Shah NK, Rejto PA, Verkhivker GM. Structural consensus in ligand-protein docking identifies recognition peptide motifs that bind streptavidin. *Proteins* 1997 Jul;28(3):421-33.
91. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. Deciphering common failures in molecular docking of ligand-protein complexes. *J Comput Aided Mol Des* 2000 Nov;14(8):731-51.
92. Kontoyianni M, McClellan LM, Sokol GS. Evaluation of docking performance: Comparative data on docking algorithms. *J Med Chem* 2004 Jan 29;47(3):558-65.
93. Lyne PD, Walsh OA. Computer simulation of biochemical reactions with QM-MM methods. In: O. M. Becker, A. D. MacKerell Jr., B. Roux, M. Watanabe, editors. *Computational biochemistry and biophysics*. New York, Basel: Marcel Dekker, Inc.; 2001. .
94. Zsoldos Z, Reid D, Simon A, Sadjad BS, Johnson AP. eHiTS: An innovative approach to the docking and scoring function problems. *Curr Protein Pept Sci* 2006 Oct;7(5):421-35.
95. Voigt JH, Bienfait B, Wang S, Nicklaus MC. Comparison of the NCI open database with seven large chemical structural databases. *J Chem Inf Comput Sci* 2001 May-Jun;41(3):702-12.
96. Ringer AL, Sinnokrot MO, Lively RP, Sherrill CD. The effect of multiple substituents on sandwich and T-shaped pi-pi interactions. *Chemistry* 2006 May

3;12(14):3821-8.

97. Sinnokrot MO, Sherrill CD. Substituent effects in pi-pi interactions: Sandwich and T-shaped configurations. *J Am Chem Soc* 2004 Jun 23;126(24):7690-7.
98. McGaughey GB, Gagne M, Rappe AK. Pi-stacking interactions. alive and well in proteins. *J Biol Chem* 1998 Jun 19;273(25):15458-63.
100. Hummel MA, Gannett PM, Aguilar JS, Tracy TS. Effector-mediated alteration of substrate orientation in cytochrome P450 2C9. *Biochemistry* 2004 Jun 8;43(22):7207-14.