THE CONTRIBUTION OF HEME ELECTROSTATICS AND ACTIVE-SITE CONFORMATIONAL DYNAMICS TO CYP3A4-MEDIATED DEHYDROGENATION OF RALOXIFENE AND 4-HYDROXY-TAMOXIFEN

by

Kiumars Shahrokh

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Pharmacology and Toxicology

The University of Utah

May 2012

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Kiumars Shahrokh**

has been approved by the following supervisory committee members:

Garold S. Yost	, Chair	**12/19/2011** Date Approved
Thomas E. Cheatham III	, Member	**12/19/2011** Date Approved
Michael R. Franklin	, Member	**12/20/2011** Date Approved
Donald K. Blumenthal	, Member	**12/19/2011** Date Approved
Charles B. Grissom	, Member	**12/19/2011** Date Approved

and by **William R. Crowley** , Chair of the Department of **Pharmacology and Toxicology**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

The characterization of novel and reactive Phase I metabolites of xenobiotics, such as those frequently produced by P450 enzymes, is an area of interest that has led to increased research efforts during preclinical drug-testing and development. A key interest is improving our understanding of factors that contribute to competing Phase I reaction mechanisms, some of which produce stable products that can be further metabolized and excreted, and others that produce reactive metabolites capable of causing toxicities. Due to the high-energy nature of the P450 catalytic oxyferryl heme species, Compound I, P450 enzymes can also catalyze different oxidation reaction mechanisms, including dehydrogenation reactions. Dehydrogenation reactions are more difficult to predict than the more common P450 oxygenation and dealkylation reactions. Moreover, dehydrogenation mechanisms can compete with hydroxylation mechanisms to produce unstable desaturated electrophilic metabolites capable of forming potentially toxic biomolecular adducts. The work presented here focuses on improving existing computational tools for the prediction of P450 metabolism of two model substrates, raloxifene and 4-hydroxy-tamoxifen. These two compounds are FDA-approved selective estrogen receptor modulators currently used in the treatment of breast cancer. In Chapter 2 the development,

testing and refinement of molecular mechanics parameters for key species of the heme prosthetic group during the P450 catalytic cycle is presented. It is shown that the assignment of atomic partial charges for key heme species improves the identification of the sites of metabolism of raloxifene by CYP3A4. Building on this work, in Chapter 3 it is shown that despite using these new heme parameters, extensive quantum mechanics calculations to probe substrate reactivity, molecular dynamics of the enzyme structure to find representative active site conformations makes the greatest improvement in the identification of the sites of metabolism for 4-hydroxy-tamoxifen. In summary, this work identifies that heme electrostatics and enzyme conformational dynamics play important roles in enzyme function and that the ability to predict sites of metabolism for P450-substrates requires the integration of both for the improvement of future in silico tools.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

The application and refinement of calculations and mathematical models are central to the discipline of science. These calculations and models provide us with mental constructs that allow us to determine how well we understand the principles that govern the system we have chosen to study. In so doing, they also allow us to move from an observatory role to a more predictive role in our scientific endeavors. This distinction between roles of observation and classification, and analysis and prediction has separated the fields of classical physics, biology and chemistry throughout much of their history. However, with the complexity of the problems that are challenging modern research, and the vast array of sophisticated instrumentation and analytical methods available, present day researchers find themselves having to bridge these schisms and forge new interdisciplinary approaches.

In this spirit, the work presented in this dissertation shows the application, development and refinement of computational tools used currently in biophysical chemistry research, to a problem of biomedical relevance in the field of drug metabolism and toxicology: the contribution of electrostatic and thermodynamic factors to competing reaction mechanisms during the enzymatic metabolism of therapeutic drugs. The relevance of the problem investigated is that the

metabolites formed from competing reaction mechanisms can differentiate between the safe elimination of compounds, and the production of toxicities. Currently, these toxicities are identified experimentally by measuring the propensity of different novel compounds to form covalent with biological molecules[1, 2]. This type of screening occurs late during the preclinical phase of drug-development and can result in lead compound failure, projects being abandoned or costly attempts made to redesign the compounds[3-5]. Computational tools are used throughout the drug-design process to help reduce development costs and timelines[6]. It is therefore obvious that improvements in the accuracy of substrate and enzyme models, as well as the tools used to predict their interactions, can be of enormous benefit. Thus, the long-term goal of this work is to improve drug design and safety with the hope of a beneficial impact on human health.

The biological focus of this work is the human hepatic cytochrome P450 enzymes that are involved in the metabolism of the majority of drugs in use today, CYP3A4. This enzyme is capable of metabolizing a variety of potentially toxic compounds with diverse chemical structures via numerous high-energy reaction mechanisms to facilitate their elimination[7]. Even though some of these reaction mechanisms are well understood, the factors that favor certain competing reaction mechanisms are not[8-10]. The challenge presented by this type of problem is well suited for the application of computational models. The structural basis of enzyme-substrate interactions and catalysis can be experimentally difficult, if not impossible, to measure. Computational tools based

on quantum or classical mechanics can help in understanding these details at an electronic or atomic level. Although these tools have limitations[11], they can prove useful adjuncts to more traditional experiments. Thus, an integrated use of theory and experiment could therefore aid in the identification of shortcomings of the theoretical tools within a scaffold of empirical data derived from well-established experimental methodologies. Moreover, the refined use of the theoretical tools can provide models to help guide experiments. This work therefore exploits a unique opportunity for an interdisciplinary approach to apply, refine and develop new theoretical methodologies to increase the understanding of drug metabolism beyond what is measurable with experiments alone.

CHAPTER 1

INTRODUCTION

Overview of Classic Phases of Drug Metabolism

The metabolism and elimination of exogenous compounds is classically divided into three phases: Phase 1 activation reactions, Phase 2 conjugation reactions and Phase 3 transport and elimination[1]. Phase 1 reactions involve the "activation" of stable or unreactive atomic centers or bonds. These activation reactions can proceed via single or double electron oxidation reactions, and are commonly catalyzed by enzymes with highly energetic active sites, such as cytochrome P450 enzymes (CYP) and flavin monooxygenases. Cytochrome P450s require coenzyme reductase partners for the transfer of electrons, molecular oxygen and involve complex catalytic cycles[2]. P450 catalyzed reactions include hydroxylation and heteroatom dealkylation reactions, and provide a functional heteroatom for subsequent conjugation reactions. Phase 2 reactions involve the conjugation of endogenous polar nucleophilic molecules to Phase 1 activated atomic centers. These reactions increase the solubility of these lipophilic substrates and expedite their elimination. Phase 2 reactions are also catalyzed by enzymes, such as glutathione-S-transferases, and also require

cofactors such as glutathione. These conjugation reactions typically involve the incorporation of a lipophilic substrate and a hydrophilic cofactor into one product molecule with increased water solubility. Phase 3 reactions involve transport molecules such as ATP-cassette binding transporters that require an energy source to transport molecules against their concentration gradients across membranes into the kidney tubules or bile canaliculi[3].

From this rudimentary overview, it can be easily envisaged that the metabolism of drugs occurs through a complex interplay of different enzymes and cofactors. One of the most studied central players in this complex scheme is the family of hepatic P450 enzymes which are involved in the Phase 1 metabolism of the majority of drugs in use today. Since their discovery and characterization by Omura and Sato over 40 years ago[4-7] this family of enzymes and their heme-containing active sites have remained the focus of constant research. The identification of the different classes of hepatic P450s which exhibit significant differences in substrate selectivity and metabolism has contributed enormously to increased drug safety[8].

The Structure-Function Relationship of Hepatic P450 Enzymes and the Potential Contribution of Active-site Plasticity in Substrate Selectivity

P450 enzymes were first discovered as membrane-bound heme-containing proteins, identified as colored pigments from rabbit livers that demonstrated a unique light absorption characteristic at the unusually long wavelength of 450 nm in the presence of carbon monoxide[4-7,9]. Due to this

physical property, this class of enzymes became known as cytochrome P450s, cyto- meaning from the cell, -chrome, meaning a color. Since their first discovery, P450 enzymes have been identified as important biological catalysts that are found within almost all phyla of life[10] where they play key roles in many different biosynthetic and metabolic pathways, including the highly specific modifications of steroid and lipid molecules via numerous oxidation reaction mechanisms[11].

Hepatic cytochrome P450s demonstrate a degree of promiscuity in substrate selectivity typically not observed in P450s expressed in other cell types. This promiscuity in substrate recognition and the diversity of reaction mechanisms favors the role of these enzymes in protecting an organism with changing nutritional and environmental needs and conditions from the toxic build-up of chemically and structurally diverse endogenous and exogenous compounds[12]. However, this very promiscuity in substrate selectivity, and the diversity in competing chemical reaction mechanisms, poses one of the most significant hurdles in the prediction of site and mechanism of the metabolism of novel compounds[13,14].

The basis for the diversity of chemical reaction mechanisms catalyzed by P450 enzymes has been investigated using a vast array of experimental approaches[9]. These methods include site-directed mutagenesis of the enzyme in substrate-recognition sequences first identified by Gotoh[15], the use of UV-vis spectroscopy to identify changes in the oxidation state of the heme within the P450 active site[16,17], radical-clock and kinetic isotope studies[18] and mechanism-based inhibition studies[19,20]. However, the direct elucidation of the structure of

human hepatic P450s has been hampered by their membrane association, an obstacle which posed a significant hurdle to modern spectroscopic studies.

The elucidation of the three-dimensional structures of nonmembrane associated soluble bacterial P450[21] allowed structure-based methods to be pursued for the elucidation of the basis for substrate selectivity and drug metabolism. Homology models of many human hepatic P450 were constructed based on sequence similarity to these structures to investigate human drug metabolism[9]. However, these efforts produced mixed results and lacked significant predictive power for the metabolism of even the more substrate selective hepatic P450s, such as CYP2D6[22].

Since the turn of the century, advances in recombinant expression technologies and x-ray crystallography have permitted elucidation of an ever increasing number of P450 structures, including engineered mammalian P450s missing the transmembrane region[23-29]. These advances have facilitated the expression, crystallization and x-ray spectroscopic analysis of one P450 that is responsible for the majority of oxidative drug metabolism in humans: CYP3A4. In fact, x-ray crystallography of CYP3A4 in the presence of novel substrates is attempted routinely during drug development to provide insights into substrate binding and sites of metabolism (SOM)[30]. However, a number of factors that include crystallization artifacts, absence or ambiguity of x-ray electron density within the P450 active site and the complexity of the P450s catalytic cycle limit the utility of x-ray crystallographic structural data to consistently answer most questions on P450 metabolism.

The growing body of x-ray crystallographic-based structural information on engineered mammalian microsomal P450[31] has provided a great deal of insight into the potential structural dynamics of some of these enzymes[31]. For a few of the major human hepatic drug metabolizing enzymes, a number of different x-ray structures have been solved. For CYP3A4, there are seven published structures (PDB ID: 1TQN[32], 1WOE[33], 1WOF[33], 1WOG[33], 2VOM[29], 2JOD[29] and 3NXU[34]). Four (1TQN, 1W0E, 1W0F and 1W0G) are in a similar 'closed' state, the 1TQN structure has the highest resolution, the 1W0E, 1WOF and 1WOG are similar to one another and the 1TQN structure, but 1W0F and 1W0G have different molecules bound to them, a substrate progesterone and an inhibitor metyrapone, respectively. The more recent 2JOD, 2VOM and 3NXU structures demonstrate the potential plasticity of the active site of this enzyme when bound to either the substrate erythromycin, the inhibitor ritonavir or two molecules of the inhibitor ketoconazole, respectively[29].

The available x-ray structures have provided a great deal of information about differences in the structures of P450 enzymes, including the potential plasticity of these enzymes. However, the structural data from these x-ray studies provide us only with snap-shots of enzyme structures. Furthermore, the experimentally derived models may possess structural artifacts introduced by the crystallography conditions that may not be indicative of conformational plasticity that is of physiological relevance. The presence of these artifacts may be difficult to identify[35], especially when only one or a few structure(s) exist for the P450-substrate or -ligand binding mode of interest. Any unrecognized structural

artifacts could introduce significant errors in current structure-based methods that use these experimental models to understand the metabolism of other compounds[35].

The process of enzymatic catalysis is highly complex, involving substrate ingress and egress, complex electronic bond-breaking and bond–making processes and for CYP3A4, can involve intricate and distal interactions with multiple substrate molecules[29], cofactors and the lipid membrane[36]. However, one improvement that can be directly addressed with current computational tools and models is the evaluation of the contribution of electrostatic and thermodynamic factors to existing x-ray structures, and the potential impact of those structural changes on enzyme-substrate interactions.

The Role of the Heme Prosthetic Group in P450-Catalyzed Chemical Reactions

The accurate prediction of metabolism of hepatic P450 enzymes is not only complicated by their promiscuity in substrate selectivity and conformational plasticity but also by the diversity in oxidation reactions they catalyze. Oxidation reactions, such as oxygenation reactions that are catalyzed by P450, are defined as either the loss of electrons or an increase in the oxidation state of the molecule or atom. P450 “hydroxylation” reactions are the prototypical Phase 1 activation reaction catalyzed via P450 oxidation and involve the incorporation of one atom of oxygen from molecular oxygen into a hydrocarbon bond [9]. However, P450s are capable of catalyzing a number of diverse chemical reactions[37] and all of the observed reaction mechanisms are classifiable as oxidation reactions [9].

Since these enzymes can catalyze a number of high-energy reaction mechanisms other than incorporation of oxygen into a hydrocarbon bond, they are more accurately referred to as mixed-function oxidases[38]. The diversity in reaction mechanisms is made possible by the heme prosthetic group present in the P450 active site which facilitates controlled electron transfers and is responsible for their unique light absorbance characteristics that led to their discovery[39].

Heme prosthetic groups are specialized protein cofactors that incorporate porphyrin rings interacting with various metals, ligands and protein side chains. In the case of P450, this is an iron center which is coordinated to a sulfur atom in a cysteine residue. The catalytically competent high valence oxyferryl heme species, Compound I, produced within the active site during the catalytic cycle has recently been directly observed[40]. Even though the electronic changes to the heme within the P450 active site during ligand binding and substrate turnover have been extensively studied experimentally, the development of a consistent set of transferable parameters for computational modeling of key species of the heme that form during the P450 catalytic cycle has faced significant challenges. One of the major difficulties is the open-shell states of this iron-containing prosthetic group that requires complicated quantum mechanics-based methods to accurately describe. However, even though it is commonly accepted that changes in the electronic state of the heme play an important role during P450-catalysis, current models do not account for these changes in a consistent manner. Thus, an obvious improvement for computational approaches for the

prediction of P450-metabolism is the development and testing of a consistent and transferable set of parameters for the key heme species formed during the P450-catalytic cycle.

P450-Catalyzed Dehydrogenation Reactions

The heme prosthetic group provides the capacity for P450s to catalyze a number of high-energy reactions. Many of these reaction mechanisms lead to the safe elimination of numerous compounds. However, some reactions produce reactive metabolites. Identifying compounds that form reactive metabolites continues to be a major area of concern in the evaluation of the safety of new drugs[41]. In 2002, the FDA in collaboration with industry released official guidelines for the evaluation of the metabolites in safety testing during drug development[42]. Within these FDA MIST guidelines, it is stated: “Based on the nature of the chemical reactions involved, metabolites formed from Phase I reactions are more likely to be chemically reactive or pharmacologically active and, therefore, more likely to need safety evaluation”[42]. Thus, elucidating the factors that contribute to the formation of reactive metabolites via P450-catalyzed reactions, such as dehydrogenation continues to be of concern in drug safety and development.

Experimental studies of have shown that a variety of commonly used pharmaceutical compounds[43,44], natural products[45-47] and environmental contaminants [48-50]can produce reactive desaturated metabolites via P450-catalyzed dehydrogenation reactions. The data from these studies indicate that a

variety of substrates can undergo dehydrogenation reactions to produce a variety of unstable metabolites including quinone methides[51]. Unfortunately, no clear set of rules has yet been established to predict accurately the dehydrogenation of substrates during P450-mediated metabolism. Unlike some of the better understood reaction mechanisms with distinctly different SOM, dehydrogenation and oxygenations can compete at the same SOM of a substrate[52].

Using gas-phase quantum mechanics based calculations, a study by Shaik and coworkers of radical clock substrates investigated the factors that contribute to the hydroxylation versus desaturation of *trans*-2-phenyl-*iso*-propylcyclopropane [53]. This was a follow-up study to an earlier theoretical investigation by the same investigators of the rebound mechanism originally proposed by Groves and coworkers[54]. This earlier study used gas-phase quantum mechanics based calculations with models of experimental radical clock substrates, and a truncated iron-porphine model with a thiolate ligand representation of the heme to provide insights into the mechanism of P450-hydroxylation. These theoretical studies produced results in good agreement with experimental isotope-labeling and substrate-rearrangement data. The application of theory provided a scientific means to explain experimental results at a level that was not possible to do so with experiment alone, and validated the rebound mechanism proposed by Groves and coworkers, and the theoretical two-state reactivity and the gas-phase model system[54]. The follow-up study that focused on desaturation of the radical clock substrate used a similar model system and showed that the desaturation reaction mechanism can proceed spontaneously,

and is energetically more favorable under specific conditions than radical rebound[38]. These results required extrapolation from gas-phase calculations using truncated models to predict the factors that contribute to competing oxygenation versus dehydrogenation reactions during P450-mediated turnover of pharmaceutical compounds. The authors concluded that: "In sum: the oxidase-dehydrogenase mixed activity *occurs from the cationic intermediate species and requires electro-steric inhibition of the rebound process*[53]." Thus, to validate this hypothesis and to improve the currently available tools for predicting drug metabolism, an ideal experimental approach would use the biomedically relevant P450, e.g. CYP3A4 and therapeutic compounds that are metabolized via competing hydroxylation and dehydrogenation reactions.

Raloxifene and 4-Hydroxy-Tamoxifen are Useful Molecular Probes of P450 Enzyme-Substrate Interactions

Raloxifene

Raloxifene is an FDA-approved selective estrogen receptor modulator used in the treatment of breast cancer[55]. CYP3A4 has been shown to both hydroxylate and dehydrogenate this compound in a regioselective manner[56]. Specifically, CYP3A4-mediated hydroxylation of raloxifene produces 3'-hydroxy-raloxifene, and dehydrogenation putatively produces a di-quinone metabolite that is so unstable that it has only been observed as a nucleophile-conjugate[57], shown in Figure 1.1. This di-quinone is believed to be the species responsible for the mechanism–based inactivation of CYP3A4 by raloxifene[58,59]. Recent work by our lab has shown that molecular modeling of CYP3A4 and raloxifene

interactions was greatly improved by the assignment of a preliminary set of atomic partial charges, and allowed the identification of Phe215 as playing an important role in determining hydroxylation versus dehydrogenation[60]. Since it has thus been experimentally validated that models based on raloxifene and CYP3A4 were improved by the assignment of partial charges representative of one state of the heme, this model system may be appropriate for the development and refinement of a set of consistent and transferable heme electrostatic parameters for a key heme species formed during the P450 catalytic cycle.

4-Hydroxy-tamoxifen

4-Hydroxy-tamoxifen (4OHT), is a primary metabolite of tamoxifen (TAMX), which is also an FDA-approved selective estrogen receptor modulator used extensively in the treatment of breast cancer [61]. A number of P450 metabolites of TAMX, 4-hydroxy-tamoxifen (4OHT), endoxifen and Metabolite E, an oxidative metabolite found abundantly in the plasma of TAMX-treated patients, have been shown to bind the estrogen receptor (ER) with a much higher affinity than the parent compound[62-65]. TAMX is thus classified as prodrug, requiring specific P450-mediated modifications to produce pharmacologically active metabolites required to achieve the desired therapeutic effect. P450-mediated TAMX metabolism has been shown to be P450 specific[66]. A number of the metabolites, including 4OHT, also undergo P450 metabolism to form reactive secondary metabolites[67], via both hydroxylation and dehydrogenation reactions.

P450-catalyzed dehydrogenation of 4OHT to quinone methide products can result in the formation of covalent adducts with other biomolecules in a nonenzymatic manner. A scheme of TAMX metabolism from parent to 4OHT, 4OHT-quinone methide and a biomolecular adduct is shown in Figure 1.2. 4OHT is therefore a useful second molecular probe since it is a pharmacologically active drug metabolite that undergoes P450-mediated metabolism to form reactive products that pose a risk of toxicity, as is of concern in the MIST guidelines[68]. Furthermore, 4OHT poses a greater challenge to molecular modeling compared to raloxifene because it undergoes competing hydroxylation and dehydrogenation reactions at the same site of metabolism[67]. Thus, 4OHT can either validate the findings with raloxifene or provide insights into other mechanistic details of the determinants of P450-hydroxylation versus P450-dehydrogenation reactions for alternative types of substrates.

Development of Tools with Greater Predictive Power

Within the fields of physical and biophysical chemistry, a number of well-established computational approaches based on physical principles exist that can be used to provide insight into possible enzyme–substrate interactions. The computational tools that are currently available and applicable to enzyme-substrate interactions and catalysis can be broadly separated into quantum mechanics based methods for understanding complex electronic configurations and behavior in smaller model systems and molecular mechanics based

methods that allow modeling and sampling of much larger systems such as proteins.

Quantum mechanics calculations address the electronic structure of atoms and molecules, and can do so by representing the constituent electrons in an explicit manner. To accurately describe the quantum behavior of subatomic species, such as electrons, requires computationally expensive and sophisticated mathematics. Excellent and comprehensive reviews of this material may be found elsewhere[69]. Thus, even though quantum theory provides a useful approach for deriving properties based on the electronic structure of molecules, it is currently computationally too expensive to do so at the required level of theory, with a system that includes the entire P450.

Molecular mechanics based methods are an atom-based treatment of molecular models and include sophisticated potential energy functions and parameter sets. Molecular mechanics based methods include docking programs and molecular dynamics simulations (MD). Docking calculations allow the modeling of interactions between biological macromolecules and inhibitor or substrate molecules. MD simulations allow modeling of thermodynamic factors within molecular models at an atomic level. The potential energy functions of the modern MD simulations have provided many useful insights into structure-function relationships even though still suffering from some shortcomings[70].

Combinations of quantum mechanics and molecular mechanics methods have also been applied specifically to P450-catalysis based on experimentally derived structures. This work has been performed to date for only one P450, the

bacterial P450cam, for which multiple structures of key steps have been solved during the catalytic cycle with its substrate, camphor. This study has validated the application of mixed quantum mechanics and molecular mechanics (QM/MM)-based methodologies to evaluate both the electronic and structural factors involved in enzymatic catalysis and reaction mechanism for this bacterial P450. However, this type of approach also remains computationally too expensive for routine screening of chemical libraries.

The work presented in this dissertation is a combined empirical and theoretical investigation of *in silico* tools to predict the site of metabolism of two therapeutic drugs known to be metabolized to reactive electrophiles via both oxygenation and dehydrogenation reactions by CYP3A4. Standard *in vitro* incubation techniques coupled with mass spectroscopic analysis of products are utilized to determine differences in metabolism of model substrates, raloxifene and 4OHT. This experimental approach is then integrated with a computational analysis of the contributions of electronic changes in the heme prosthetic group during the P450 catalytic cycle, and contributions of thermodynamic entropy-driven processes to P450 structure and to enzyme-substrate. It is the ultimate goal of this research to improve the accuracy and reliability of computationally rapid and affordable tools for predictions of drug metabolism.

References

1. Meyer, U. Journal of Pharmacokinetics and Pharmacodynamics 1996, 24(5), 449-459.

2. Paine, M. J. I.; Scrutton, N. S.; Munro, A. W.; Gutierrez, A.; Roberts, G. C. K.; Wolf, C. R. 2005(115-148).

3. Klassen, C. D. McGraw-Hill Medical Publishing Division 2001. 471-490.

4. Omura, T.; Sato, R. The Journal of biological chemistry 1962, 237, 1375-1376.

5. Omura, T.; Sato, R. Biochimica et biophysica acta 1963, 71, 224-226.

6. Omura, T.; Sato, R. The Journal of biological chemistry 1964, 239, 2379-2385.

7. Omura, T.; Sato, R. The Journal of biological chemistry 1964, 239, 2370-2378.

8. Meunier, B.; de Visser, S. P.; Shaik, S. Chem Rev 2004, 104(9), 3947-3980.

9. Ortiz de Montellano, P. R.; De Voss, J. J. Kluwer Academic/Plenum Publishers, New York 2005, 185-245.

10. Stoilov, I.; Jansson, I.; Sarfarazi, M.; Schenkman, J. B. Drug Metabolism and Drug Interactions 2001, 18(1), 33-56.

11. Nebert, D. W.; Russell, D. W. The Lancet 2002, 360(9340), 1155-1162.

12. Guengerich, F. P. Aaps J 2006, 8(1), E101-111.

13. Afzelius, L.; Arnby, C. H.; Broo, A.; Carlsson, L.; Isaksson, C.; Jurva, U.; Kjellander, B.; Kolmodin, K.; Nilsson, K.; Raubacher, F.; Weidolf, L. Drug metabolism reviews 2007, 39(1), 61-86.

14. Sun, H.; Scott, D. O. Chem Biol Drug Des, 75(1), 3-17.

15. Gotoh, O. Journal of Biological Chemistry 1992, 267(1), 83-90.

16. Estabrook, R. W.; Franklin, M. R.; Hildebrandt, A. G. Annals of the New York Academy of Sciences 1970, 174(1), 218-232.

17. Estabrook, R. W.; Franklin, M. R.; Cohen, B.; Shigamatzu, A.; Hildebrandt,

A. G. Metabolism 1971, 20(2), 187-199.

18. Groves, J.; Ortiz de Montellano, P. R., Ed.; Springer US, 2005, p 1-43.

19. Zhou, S.; Yung Chan, S.; Cher Goh, B.; Chan, E.; Duan, W.; Huang, M.; McLeod, H. L. Clinical pharmacokinetics 2005, 44(3), 279-304.

20. Kent, U. M.; Jushchhyshyn, M. I.; Hollenberg, P. F. Current Drug Metabolism 2001, 2(3), 215-243.

21. Poulos, T. L.; Finzel, B. C.; Howard, A. J. Journal of Molecular Biology 1987, 195(3), 687-700.

22. Lewis, D. F. V.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Xenobiotica 1997, 27(4), 319-340.

23. Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Science (New York, NY 2004, 305(5684), 683-686.

24. Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. The Journal of biological chemistry 2006, 281(11), 7614-7622.

25. Scott, E. E.; Halpert, J. R. Trends in biochemical sciences 2005, 30(1), 5-7.

26. Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C.; Halpert, J. R.; Johnson, E. F.; Stout, C. D. Proceedings of the National Academy of Sciences of the United States of America 2003, 100(23), 13196-13201.

27. Scott, E. E.; White, M. A.; He, Y. A.; Johnson, E. F.; Stout, C. D.; Halpert, J. R. The Journal of biological chemistry 2004, 279(26), 27294-27301.

28. Sligar, S. G.; Makris, T. M.; Denisov, I. G. Biochemical and biophysical research communications 2005, 338(1), 346-354.

29. Ekroos, M.; Sjogren, T. Proceedings of the National Academy of Sciences of the United States of America 2006, 103(37), 13682-13687.

30. Sun, H.; Scott, D. O. Chemical biology & drug design 2010, 75(1), 3-17.

31. Johnson, E. F.; Stout, C. D. Biochemical and Biophysical Research Communications 2005, 338(1), 331-336.

32. Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Journal of Biological Chemistry 2004, 279(37), 38091-38094.

33. Williams, P. A.; Cosme, J.; Vinković, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Science 2004, 305(5684), 683-686.

34. Sevrioukova, I. F.; Poulos, T. L. Proceedings of the National Academy of Sciences 2010, 107(43), 18422-18427.

35. Søndergaard, C. R.; Garrett, A. E.; Carstensen, T.; Pollastri, G.; Nielsen, J. E. Journal of Medicinal Chemistry 2009, 52(18), 5673-5684.

36. Davydov, D. R.; Halpert, J. R. Expert Opinion on Drug Metabolism & Toxicology 2008, 4(12), 1523-1535.

37. Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. Chemical Reviews 2009, 110(2), 949-1017.

38. Kumar, D.; De Visser, S. P.; Shaik, S. Journal of the American Chemical Society 2004, 126(16), 5072-5073.

39. Omura, T.; Sato, R. Journal of Biological Chemistry 1964, 239(7), 2370-2378.

40. Rittle, J.; Green, M. T. Science 2010, 330(6006), 933-937.

41. Park, B. K.; Boobis, A.; Clarke, S.; Goldring, C. E. P.; Jones, D.; Kenna, J. G.; Lambert, C.; Laverty, H. G.; Naisbitt, D. J.; Nelson, S.; Nicoll-Griffith, D. A.; Obach, R. S.; Routledge, P.; Smith, D. A.; Tweedie, D. J.; Vermeulen, N.; Williams, D. P.; Wilson, I. D.; Baillie, T. A. Nat Rev Drug Discov 2011, 10(4), 292-306.

42. Baillie, T. A.; Cayen, M. N.; Fouda, H.; Gerson, R. J.; Green, J. D.; Grossman, S. J.; Klunk, L. J.; LeBlanc, B.; Perkins, D. G.; Shipley, L. A. Toxicology and applied pharmacology 2002, 182(3), 188-196.

43. Obach, R. S. Drug metabolism and disposition: the biological fate of chemicals 2001, 29(12), 1599-1607.

44. Sun, H.; Moore, C.; Dansette, P. M.; Kumar, S.; Halpert, J. R.; Yost, G. S. Drug metabolism and disposition: the biological fate of chemicals 2009, 37(3), 672-684.

45. Reilly, C. A.; Ehlhardt, W. J.; Jackson, D. A.; Kulanthaivel, P.; Mutlib, A.

E.; Espina, R. J.; Moody, D. E.; Crouch, D. J.; Yost, G. S. Chemical research in toxicology 2003, 16(3), 336-349.

46. Reilly, C. A.; Yost, G. S. Drug metabolism and disposition: the biological fate of chemicals 2005, 33(4), 530-536.

47. Reilly, C. A.; Yost, G. S. Drug metabolism reviews 2006, 38(4), 685-706.

48. Skiles, G. L.; Yost, G. S. Chemical research in toxicology 1996, 9(1), 291-297.

49. Skordos, K. W.; Laycock, J. D.; Yost, G. S. Chemical research in toxicology 1998, 11(11), 1326-1331.

50. Weems, J. M.; Cutler, N. S.; Moore, C.; Nichols, W. K.; Martin, D.; Makin, E.; Lamb, J. G.; Yost, G. S. Toxicol Sci 2009, 112(1), 59-67.

51. Fan, P. W.; Zhang, F.; Bolton, J. L. Chemical research in toxicology 1999, 13(1), 45-52.

52. Yost, G. S. Adv Exp Med Biol 2001, 500, 53-62.

53. Kumar, D.; de Visser, S. P.; Shaik, S. Journal of the American Chemical Society 2004, 126(16), 5072-5073.

54. Kumar, D.; de Visser, S. P.; Shaik, S. J Am Chem Soc 2003, 125(43), 13024-13025.

55. Clemett, D.; Spencer, C. M. Drugs 2000, 60(2), 379-411.

56. Yu, L.; Liu, H.; Li, W.; Zhang, F.; Luckie, C.; van Breemen, R. B.; Thatcher, G. R. J.; Bolton, J. L. Chemical research in toxicology 2004, 17(7), 879-888.

57. Moore, C. D.; Reilly, C. A.; Yost, G. S. Biochemistry 2010, 49(21), 4466-4475.

58. Chen, Q.; Ngui, J. S.; Doss, G. A.; Wang, R. W.; Cai, X.; DiNinno, F. P.; Blizzard, T. A.; Hammond, M. L.; Stearns, R. A.; Evans, D. C.; Baillie, T. A.; Tang, W. Chemical research in toxicology 2002, 15(7), 907-914.

59. Baer, B. R.; Wienkers, L. C.; Rock, D. A. Chemical research in toxicology 2007, 20(6), 954-964.

60. Moore, C. D.; Shahrokh, K.; Sontum, S. F.; Cheatham, T. E., 3rd; Yost, G. S. Biochemistry 2010, 49(41), 9011-9019.

61. Smigel, K. Journal of the National Cancer Institute 1998, 90(9), 647-648.

62. Coezy, E.; Borgna, J. L.; Rochefort, H. Cancer research 1982, 42(1), 317-323.

63. Jordan, V. C. Breast cancer research and treatment 1982, 2(2), 123-138.

64. Borgna, J. L.; Rochefort, H. The Journal of biological chemistry 1981, 256(2), 859-868.

65. Robertson, D. W.; Katzenellenbogen, J. A.; Long, D. J.; Rorke, E. A.; Katzenellenbogen, B. S. Journal of steroid biochemistry 1982, 16(1), 1-13.

66. Desta, Z.; Ward, B. A.; Soukhova, N. V.; Flockhart, D. A. The Journal of pharmacology and experimental therapeutics 2004, 310(3), 1062-1075.

67. Fan, P. W.; Zhang, F.; Bolton, J. L. Chemical research in toxicology 2000, 13(1), 45-52.

68. Baillie, T. A.; Cayen, M. N.; Fouda, H.; Gerson, R. J.; Green, J. D.; Grossman, S. J.; Klunk, L. J.; LeBlanc, B.; Perkins, D. G.; Shipley, L. A. Toxicology and Applied Pharmacology 2002, 182(3), 188-196.

69. Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. Chem Rev, 110(2), 949-1017.

70. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B. Angewandte Chemie (International ed 2006, 45(25), 4064-4092.

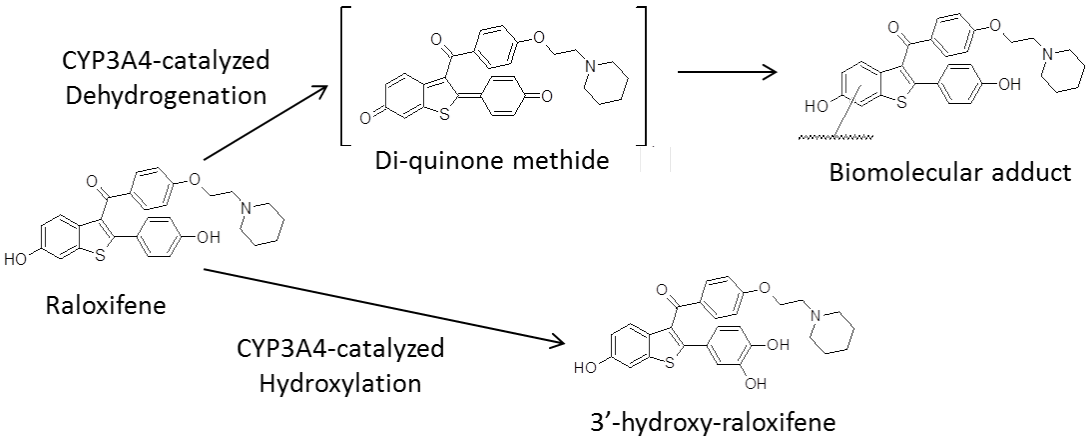


Figure 1.1: Scheme for CYP3A4 metabolism of raloxifene

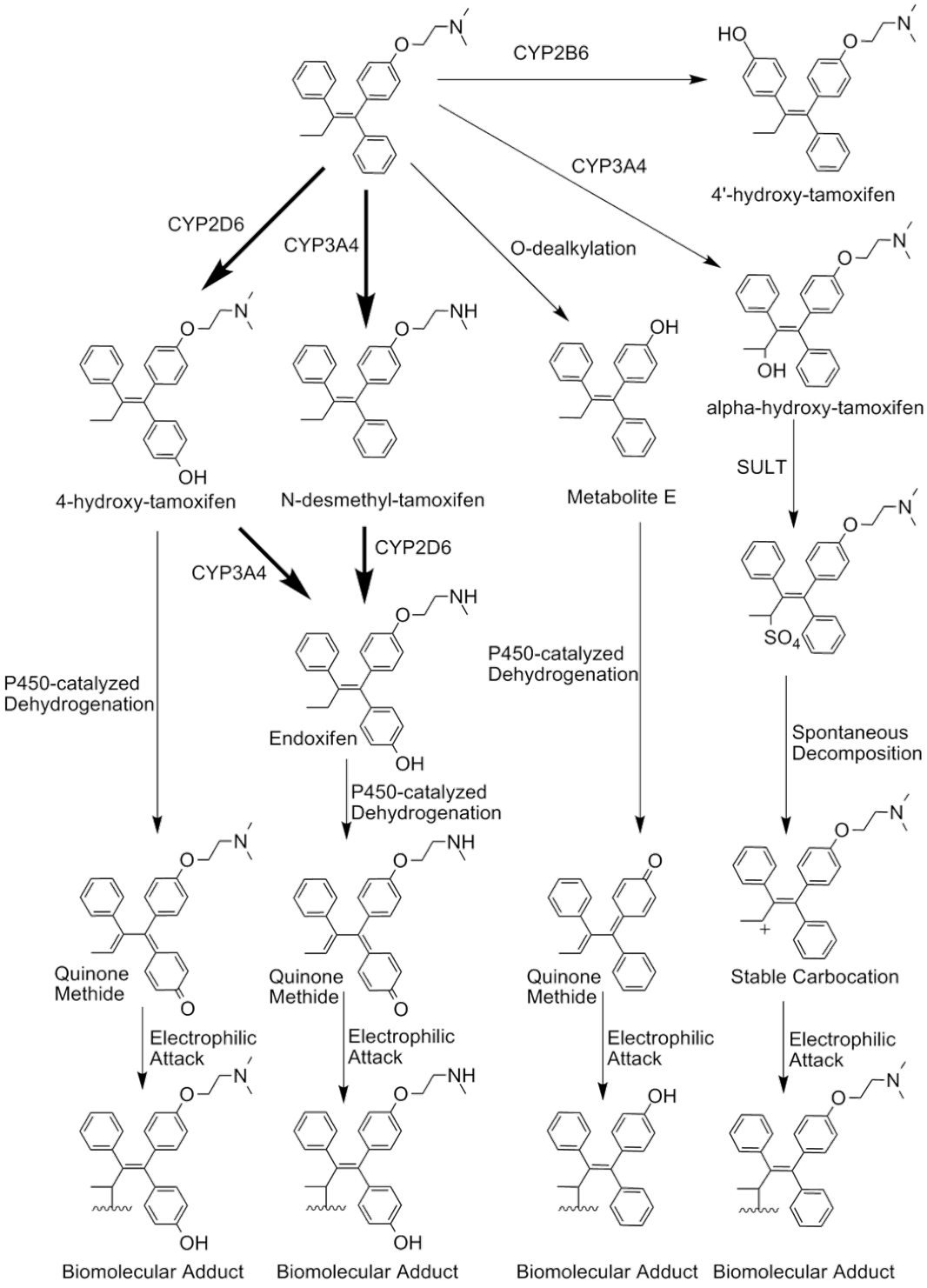


Figure 1.2: Scheme for oxidative metabolism of tamoxifen

CHAPTER 2

QUANTUM MECHANICALLY DERIVED AMBER-COMPATIBLE HEME PARAMETERS FOR VARIOUS STATES OF THE CYTOCHROME P450 CATALYTIC CYCLE

This chapter was reproduced with permission from the published paper from Shahrokh, K. Orendt, A. Yost, Garold S. and Cheatham III, T.E. *The Journal of Computational Chemistry.* **33.** 119-133.

DOI: 10.1002/jcc.21922

Quantum Mechanically Derived AMBER-Compatible Heme Parameters for Various States of the Cytochrome P450 Catalytic Cycle

Kiumars Shahrokh,[a] Anita Orendt,[b] Garold S. Yost,[a] and Thomas E. Cheatham, III*[c]

Molecular mechanics (MM) methods are computationally affordable tools for screening chemical libraries of novel compounds for sites of P450 metabolism. One challenge for MM methods has been the absence of a consistent and transferable set of parameters for the heme within the P450 active site. Experimental data indicate that mammalian P450 enzymes vary greatly in the size, architecture, and plasticity of their active sites. Thus, obtaining X-ray-based geometries for the development of accurate MM parameters for the major classes of hepatic P450 remains a daunting task. Our previous work with preliminary gas-phase quantum mechanics (QM)-derived atomic partial charges greatly improved the accuracy of docking studies of raloxifene to CYP3A4. We have therefore developed and tested a consistent set of transferable MM parameters based on gas-phase QM calculations of two model systems of the heme—a truncated (T-HM) and a full (F-HM) for four states of the P450 catalytic cycle. Our results indicate that the use of the atomic partial charges from the F-HM further improves the accuracy of docked predictions for raloxifene to CYP3A4. Different patterns for substrate docking are also observed depending on the choice of heme model and state. Newly parameterized heme models are tested in implicit and explicitly solvated MD simulations in the absence and presence of enzyme structures, for CYP3A4, and appear to be stable on the nanosecond simulation timescale. The new force field for the various heme states may aid the community for simulations of P450 enzymes and other heme-containing enzymes. © 2011 Wiley Periodicals, Inc. J Comput Chem 33: 119–133, 2012

Keywords: cytochrome P450 enzymes · heme force field parameters · molecular mechanics · RESP charges · AMBER · drug metabolism

Introduction

Biomolecular simulation is playing an increasingly important role as an adjunct to experiment in understanding the structural and dynamic factors involved in many biological processes at the atomic level.[1] With the ever increasing number of X-ray structures available for drug-metabolizing P450 enzymes, the development and refinement of structure-based computational tools for predicting metabolic pathways for lead compounds continues to be an area of intensive research.[2] The identification of sites of metabolism and further, the identification of the structural factors that can contribute to competing P450-catalyzed reaction mechanisms are important goals of rational drug design. Competing P450 reaction mechanisms can lead either to the formation of safe products for excretion or to the formation of unstable products that can lead to P450-related toxicities.[3] Thus, increasing the predictive power of computational tools to improve the accuracy of "structural alerts" for identifying compounds that can form reactive intermediates can have a very direct impact on overall drug safety and human health.[2,4]

The diversity in P450 substrate selectivity, sites of metabolism, and reaction mechanism is in part due to the large number of different subfamilies of hepatic P450s that are involved in drug metabolism. Each P450 subfamily exhibits unique functional characteristics in the metabolism of a variety of endogenous and exogenous compounds.[5] The ever growing number of experimentally derived structures of hepatic P450 clearly demonstrates that structure–function relationships can be established for these subfamilies. Even though hepatic P450 structures share a common fold,[5] the subclasses of hepatic P450 enzymes vary significantly in size and architecture. Furthermore, many of the experimentally derived structures of hepatic P450s both with and without substrates and ligands suggest significantly varying degrees of the plasticity of the active site between the different subclasses.[6]

Another major contributor to the diversity of reaction mechanisms catalyzed by P450 is the highly reactive nature of the catalytic heme oxyferryl species. Heme prosthetic groups, present in the P450 active site, are specialized protein cofactors that incorporate porphyrin rings interacting with various metals, ligands, and protein side chains. The unique chemical

[a] *K. Shahrokh, G. S. Yost*
Department of Pharmacology and Toxicology, University of Utah, Salt Lake City, Utah 84112
[b] *A. Orendt*
Center for High Performance Computing, University of Utah, Salt Lake City, Utah 84112
[c] *T. E. Cheatham, III*
Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112
tec3@utah.edu

Contract/grant sponsor: NIH; Contract/grant number: GM079383 (TEC3); Contract/grant sponsor: NIH; Contract/grant number: GM0742249 (G. Y.).

Additional Supporting Information may be found in the online version of this article.

structure of the heme facilitates a wide chemistry including—but not limited to—oxidation, dehydrogenation, and oxygen transport.[7] Pioneering work in the crystallization of bacterial cytochrome P450, specifically P450cam,[8] and the application of quantum mechanics (QM)-based methods to the heme has structurally identified and electronically characterized a number of the heme species populated during the P450 catalytic cycle.[9] Moreover, although this work has clearly shown that that gas-phase density functional theory (DFT) methods can provide reasonable results in agreement with experiment depending on the model system used,[10] and that only with the inclusion of the protein environment, from structures derived from experiment, combined with mixed quantum and molecular mechanics (QM/MM) simulations it is possible to accurately model the electronic state, and specifically the spin densities that are critical for understanding the contribution of structural factors during catalysis to these open-shell species to within experimental accuracy.[11] For an extensive review of the applications of different computational methods and utility of different model systems, see these excellent reviews.[10(a),11]

To better characterize and screen putative drug candidates, it is useful to understand how the drugs interact with P450s. Unfortunately, the acquisition of P450-bound X-ray structures for each new substrate is potentially very time consuming and not always possible. Thus, obtaining accurate crystal structures to facilitate the development of accurate QM/MM parameters for the major classes of hepatic P450 across the complete catalytic cycle, as previously performed for bacterial P450cam, remains a daunting task. Even given sufficient data, the use of QM/MM methods remains currently very computationally expensive to meet the demand for routine *in silico* structure-based screening of large libraries of compounds against multiple P450 structures. MM-based force field methods are computationally much less expensive than QM/MM methods and allow for adequate sampling of diverse chemical libraries of substrates and enzyme structures. With the use of molecular docking and scoring schemes, these MM-based methods can potentially identify the site of, and reaction mechanism of, drug metabolism. However, despite the availability of increasing numbers of P450 structures, the predictions of metabolic pathways for novel agents using MM-based computational tools and crystal structures have frequently been incorrect.[3,12] Although the modeling of enzyme–substrate interactions and substrate metabolism are complex processes involving additional factors such as substrate ingress, binding, catalysis, and product release and egress, one obstacle for these methods has been the absence of a consistent set of MM parameters for the heme species during the P450 catalytic cycle.

Even though the electronic changes to the heme within the P450 active site during ligand binding and substrate turnover have been extensively experimentally studied, the development of a consistent set of MM-based parameters for key species of the heme during the P450 catalytic cycle faces significant challenges due to open-shell states of this iron-containing prosthetic group. Earlier, simple and less computationally expensive approaches for the development of MM parameters for the resting heme state of the P450 catalytic cycle and also for the resting heme in cytochrome *c*[13] have been attempted using AM-1, Hartree-Fock (HF), and DFT levels of theory with a number of different basis sets.[14] Some of the limitations encountered with these models were the treatment of the heme side chains, the assignment of simple, but inaccurate, Mulliken charges, and the use of only a very small set of different heme states. Even with more accurate and consistent methods, it must be acknowledged that the development of a generally applicable heme MM model faces limitations because we cannot fully account for the entire set of potential influences on the heme structure and properties. Although we can control direct or covalent influences on the heme by developing separate models, that include additional bonded atoms, we cannot easily include indirect influences. These indirect influences include changes in the heme properties caused by the protonation states of nearby active site residues, changes in active site geometries or heme environment after substrate binding, and changes in active site hydration. Despite this inherent limitation of pairwise nonadditive MM models, such residue-based parameterizations have proven extremely useful to understand biomolecular structure, dynamics, and interactions.[15] Specifically in the case of metalloproteins, MM parameters were developed and successfully used to understand dynamic processes for iron-containing systems.[16] The bonded plus electrostatics model of a metal ion allows for the incorporation of the metal ion into MM force fields, such as AMBER and is based on the method first developed for zinc metalloproteins.[17]

Recent work from our lab has demonstrated that the assignment of more reasonable atomic partial charges for the resting ferric state of the heme greatly improves docking studies for raloxifene to the CYP3A4 P450 enzyme.[18] Further, the results facilitated the identification of key residues involved in shunting between two distinct P450 metabolic pathways.[18] Raloxifene is an anticancer therapeutic with a significant number of degrees of structural rotational freedom and heteroatoms capable of polar interactions. CYP3A4 is the most biomedically relevant subfamily of hepatic P450 studied, as CYP3A4 makes the largest contribution to overall P450-mediated hepatic drug metabolism.[9] CYP3A4 enzymes are able to perform this critical role because of promiscuity in substrate selectivity due to the size and flexibility of the active site.[19] These characteristics provide a very significant challenge to structure-based methods for the prediction of metabolism. The incorporation of a preliminary set of atomic partial charges for the pentacoordinate ferric state to the heme in the X-ray structure of CYP3A4 (PDB ID: 1W0E)[20] developed with a gas-phase model of an extended heme structure with a X-ray structure vastly improved the modeling of raloxifene within the CYP3A4 active site[18] over models lacking any charge assignment to the heme.

Here, we describe the development of a consistent, transferable, and robust set of MM parameters for different heme states of the P450 catalytic cycle using two distinct model systems of the heme, specifically a minimal "truncated" heme (T-HM) compared to the full-heme model (F-HM). Because of

computational cost, we have investigated only four commonly accepted heme states in the P450 catalytic cycle, as schematically shown in Figure 1, with the two separate model systems

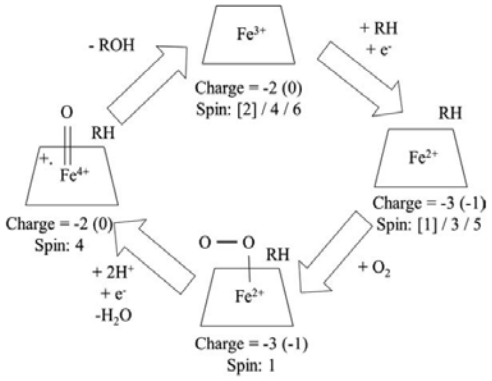


Figure 1. Schematic for describing the P450 catalytic cycle. Charge states for the extended F-HM are shown without parentheses. Charge values used for the T-HM are shown in parentheses. Spin values used for both T- and F-HM are shown without parentheses. Spin values only used for T-HM calculations are shown in brackets.

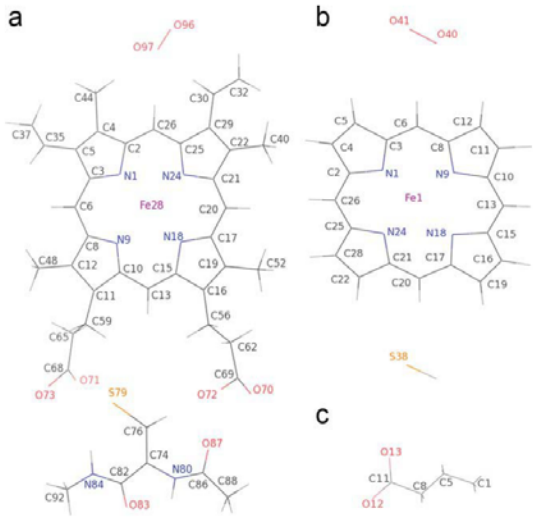


Figure 2. Atomic numbering scheme for a) the extended F-HM that includes a cysteine dipeptide (F-HM), b) the T-HM with a thiolate ligand, and c) propionic acid models. Hydrogen labels have been omitted for clarity.

illustrated in Figure 2. The goal was to identify the most appropriate, or most consistent, set of heme parameters for MM simulations of enzyme–substrate interactions. Our results indicate that the use of the RED-III-derived atomic partial charges[21] from the gas-phase F-HM, despite some charge-transfer artifacts due the absence of full shielding of the propionates in these gas-phase-based calculations, improves the accuracy of cytochrome P450 enzymes/substrate-docked predictions for raloxifene by CYP3A4 in docking experiments with AutoDock3 (Scripps Research Institute, La Jolla, CA)[22] using a distance-based scoring. Furthermore, docking with the new heme parameters and heme states suggests a strong influence of the choice of heme state on the docking results. Specifically, the nature and population of the binding mode changes depend on the applied heme state model. MD simulations on the heme molecules in solution suggest that the distinctive heme structure is maintained and motion about rotatable groups is not inhibited.

Methods

Calculations

Calculations on the heme models were based on two different model systems: the truncated porphine/thiolate (T-HM) and a full protoporphyrin macrocycle model with all its substituents including propionate side chains and a bound cysteine dipeptide ligand ligated to the iron (F-HM) as shown in Figure 2. The calculations were performed in a buildup manner from less (T-HM) to more complicated (F-HM). After optimization at the resting ferric high-spin state, interacting ligands (oxygen, etc.) and different iron spin states were added to represent the various stages of the catalytic cycle (Fig. 1). The final geometries were compared to experimental structures where known, and where unknown, careful consideration of homologous theoretical structures was given to eliminate potential errors and identify limitations.

A chief concern centered on the treatment and influence of the macrocycle substituents on the results, especially the propionate groups and the ligated proximal cysteine dipeptide. The model systems chosen were investigated in a manner consistent with QM-based MM parameter development approaches used with the current AMBER ff9X and ff10 force fields.[23] For charge derivation, this typically involves HF optimization of one or more geometries of the residue of interest using the overly polarized 6-31G* Pople-type basis set followed by restrained electrostatic potential (RESP) charge fitting.[24] To avoid influences due to the initial molecule orientation, the automated RED-III procedures were used.[21] The standard HF/6-31G* level of theory used for small organic molecules in AMBER force field development is not applicable to the iron–protoporphyrin heme model. Instead, DFT hybrid functionals such as B3LYP with LACVP were applied, as they have been previously shown to produce sufficiently accurate results for geometries at a reasonable computational cost.[10(a),11]

Geometry optimizations

Calculations were begun with a porphyrin ring at the HF/3-21//B3LYP/6-31G* level with the imposition of C4v symmetry constraints. The iron atom was then added along with the thiolate ligand and symmetry constraints were removed. Optimizations were performed at the B3LYP/LACVP (with 6-31G*[25] on the organic atoms and the LANL2DZ[26] basis set with frozen core electrons, as implemented in Gaussian 03) level for the high-spin ferric species. This species was then used to sequentially model the changes in the P450 catalytic cycle and the spin states of interest.

The representative species for the heme were selected based on the P450 catalytic cycle. This cycle begins with a

penta-coordinate ferric species (IC*), a species that is then reduced to a penta-coordinate ferrous species (OUS*), which then binds molecular oxygen. This dioxygen-bound species (DIOXY) is then reduced further through the addition of a proton and electron leading to expected cleavage of the oxygen–oxygen bond to form the catalytically active oxyferryl species, compound I (CPDI). Many of these species can exist in a low-, medium-, or high-spin state. Because of the reduced computational costs for the smaller truncated heme models, a number of these spin states were investigated for the penta-coordinate truncated heme models. The dioxygen-bound species is a closed shell singlet. For CPDI, MM parameters were only derived for the high-spin state, as it has been shown that the low-spin doublet state requires significantly more computationally demanding multireference configuration interaction[27] or CASSCF/CASPT2[28] methods to describe accurately. Additionally, some ambiguity remains as to whether this low-spin state is degenerate or lower in energy than the quartet state.[11]

The optimized T-HM penta-coordinate ferric high-spin sextet state (IC6) geometry was used to develop the F-HM model. The substituent groups were then added to the heme, including the propionate groups and a more extended model of the bound sulfur, through the inclusion of a cysteine dipeptide (i.e., a model *N*-methyl and acetyl blocked cysteine residue consistent with the AMBER force field[23]). The dihedral angles on the propionates were restricted to values consistent with the available set of published X-ray structures of mammalian P450. Restraints were applied because the positioning of the propionates in most of these experimental structures differs significantly from those observed in the P450cam structures. The values were calculated using the heme in the A chain of the following P450 structures for CYP3A4 (PDB ID: 1W0E, 1W0F, 1W0G, 1TQN, 2JOD, and 2VOM),[19,20,29] CYP2D6 (PDB ID: 2F9Q),[30] and CYP2B4 (PDB ID: 1PO5, 2BDM, and 1SUO).[6(b,e),31] These structures were chosen, as they show the plasticity of the P450 active site in the absence and presence of a number of structurally diverse ligands. The average dihedral angle defined by C11–C59–C65–C68 was thus frozen at 176.00° and the dihedral defined by C16–C56–C62–C69 was frozen at 174.50° (see Fig. 2 for atomic numbering). Optimizations were performed with the cysteine dipeptide in a number of different starting positions relative to the propionate side chains at the B3LYP/LACVP level for the ferric high-spin species. Because of the vagarities of Gaussian 03 implementation, only one structure converged. This structure was then used as the starting point for the other species in the P450 catalytic cycle.

The stability of the minima was confirmed with normal mode analysis and the stability of the wave functions was confirmed with addition of the "stable = opt" keyword.[32] Spin contamination was also verified to be in agreement with values in the literature[33] for the open-shell species (see Supporting Information Table 1).

For the T-HM, thiolate–porphyrine heme species, "tight" convergence criteria were used for another round of optimizations and the stability of the wave function and minima were confirmed again. Differences in energies were compared between structures optimized with these different criteria, differences of ~0.1 kcal/mol were observed for the smaller, more symmetric heme model. For the F-HM, the optimizations were carried out to the default G03 criteria.

Charge fitting

RESP charges were derived using RED-III[21] for the optimized geometries for all F-HM and T-HM species at the B3LYP/LACVP level. Intramolecular charge restrictions and constraints were applied to the blocked cysteine residue (consistent with the AMBER ff9X force field)[23] and molecular electrostatic potential computations were performed on multiple orientations of a single conformation of each system, based on atoms N1, Fe28, and S79 (see Fig. 2 for atom nomenclature).

Docking

AutoDock3.05 and AutoDockTools (Scripps Research Institute)[22] were used for all docking studies. Docking was performed consistent with our previous work.[34]

Using UCSF Chimera, the heme coordinates from the B3LYP/LACVP optimized structure for the F-HM and the PDB ID: 1WOE[20] coordinates were matched at the pyrrole nitrogen atoms of the heme to translate the F-HM and 1W0E[20] coordinates into a common frame of reference and saved. The coordinates for the heme from the X-ray study were removed and the coordinates for each charge and spin state of the F-HM were added to the prepped and matched 1W0E. In this manner, 1W0E files were generated to represent CYP3A4 during different phases of the P450 catalytic cycle for the ferric high and intermediate, O2-bound, and oxyferryl states. The B3LYP/LACVP RED-III-derived partial charges were then assigned to the atoms in each modified 1W0E template. The van der Waals nonbonded parameters for these values were taken from the AMBER ff99SB[35] and GAFF force fields,[36] so that the heme force field would be transferable between the AMBER ff9X force fields. The Fe values used in the previous docking study[34] and recommended in the AutoDock package for iron atoms were assigned: 1.3 Å radii (Rii) and well-depth 0.01 (epsii) and solvent parameters assigned with AutoDockTools (Scripps Research Institute).

AutoGrid3.05 (Scripps Research Institute)[22] was used to define the active site space of CYP3A4. A grid box was assigned to encompass the whole active site of CYP3A4 and to closely match the one used in the previous study[34] within the new frame of reference of the matched system. Within this frame of reference, the grid box was recentered at −4.225, 0.944, and 5.434 displacement from the default center assigned by AutoGrid3.05 with the dimensions of 68 × 52 × 62 and a resolution of 0.375 Å in each dimension.

In this study, we scored all docking results with less stringent criteria for identification of the site of metabolism for the penta-coordinate state from that commonly used with X-ray coordinates for the heme. This was done because in the QM-based optimizations the Fe atom is located below the plane of the heme in the high-spin state, whereas in X-ray structure assignments of mammalian P450, the Fe atom is determined to lie approximately in the plane of the heme. Thus, for the penta-coordinate states, we permitted a distance of <6.0 Å from the Fe to the known site of metabolism as a positive. For the hexa-coordinate states, we increased the stringency of this criteria for the catalytic oxygen present and scored a positive as a heavy atom at a known site of metabolism of the substrate lying within <4.5 Å from the proximal/activated oxygen atom.

Molecular dynamics

AMBER parameter development. GAFF atom types were used to build the heme frcmod files containing the necessary force field parameters and atom types were selected by similarity of the bond length and angle parameters in the force field that were closest to the QM structures. Specifically, cc and cd atom types were used to define Cα and Cβ; nc and nd atom types were used to define the pyrrole nitrogen atoms; cg was used to define C_{meso} based on bond length and was modified with existing cc–cc–cf, cd–cd–ha, and ce–ce–cf for missing angle parameters: cc/cd–cc/cd–cg, cc/cd–cg–ha, and cd–cg–cc, respectively. For the hexa-coordinate states, oa was used for the proximal and ob, for the distal axial oxygen atom. The propionates were built with c3, h1, c, and o atom types, and a CYP residue type was introduced for the axial proximal cysteinate ligand. This CYP residue was a standard CYS residue with the thiolate H atom removed, and RED-III atomic partial charges assigned. The CYP and HEM residues had the Fe—S and Fe—N bonds added manually in tleap.

QM-based coordinate scans Fe—S, Fe—N, and Fe—O bond and Fe—N—C, N—Fe—N, S—Fe—N, N—Fe—O, and Fe—O—O force constant parameters of the T-HM were developed by fitting B3LYP/LACVP scan results. Scans of bond coordinates were performed in 0.05 Å steps and scans of angle coordinates were performed in 2° increments. However, for certain coordinates for different heme states, these steps had to be reduced or not performed for the same number of steps from the equilibrium distance/angle. This we believe was caused by the asymmetry of the model system for certain heme states. The F-HM compound I model due to its size was only used for one scan along an increasing Fe—S—C angle coordinate, as the T-HM did not possess an equivalent coordinate.

$$V_{\text{MMbond}} = k(r - r_0) \quad (1)$$

$$V_{\text{MMangle}} = k(\theta - \theta_0) \quad (2)$$

Energies of QM scans of the T-HM along key bond and angle coordinates and the MM potential energy (V_{MM}) were fit using either eq. (1) or (2), and the force constant (k) was fit to minimize the sum of the square of the residual differences between the QM and MM energies. The barrier height for torsion terms involving the Fe that were absent from GAFF and ff99SB force fields was assigned as has been done previously for transition metals using the bonded model for metal ions.[37] An alternative approach that could be used is the MTK++[38] from AmberTools which creates a model structures of the metal ion's first coordination sphere from experimentally determined structures, and can then be used to determine charges, and bond and angle constants with *ab initio* calculations.

Implicit solvent simulations. To loosely demonstrate the stability of the various heme states, 10 ns-generalized Born implicit solvent[39] MD simulations of the various heme states with parameters derived from the DFT B3LYP optimized structures with AMBER-compatible RESP charges were performed. Langevin temperature coupling with a 1.0/ps coupling time at 300 K was applied with a 1-fs time step and 16-Å nonbonded interaction cutoff with AMBER.[40]

Explicit solvent simulations. All the MD simulations were carried out using AMBER 11[40] with the ff99SBildn force field.[41] Additional Na^+ and Cl^- ions[42] were added to net-neutralize each system, and then the model was solvated in an octahedral box of TIP3P[43] water with an extension of at least 12 Å from each side. Periodic boundary conditions were applied with the particle mesh Ewald (PME) method[44] to calculate the full electrostatic energy. The 10-Å nonbonded pair list was updated whenever any atom moved more than 0.5 Å since the last list update. SHAKE[45] constraints were applied on the bonds involving hydrogen (tolerance: 10^{-5} Å), and the time step was set to 2 fs. A cutoff of 9 Å was applied to the Lennard–Jones and direct space electrostatic interactions with a uniform density approximation included to correct for the long range van der Waals.

The system was first minimized without electrostatics for 300 steps, then with a restraint of 25 kcal/(mol Å^2) applied on the Cα (500 steepest descent cycles followed by 500 conjugate gradient cycles). This minimization was followed by 100-ps MD simulation with 25 kcal/(mol Å^2) positional restraints applied on the backbone Cα atoms, and the temperature was slowly increased from 0 to 300 K. The temperature was regulated using Langevin dynamics[46] with a collision frequency of 0.2/ps. Then, followed by five cycles of equivalent 500 steps of steepest descent cycles followed by 500 steps of conjugate gradient minimization, and 50-ps equilibrations with a restraint force constant of 5, 4, 3, 2, and 1 kcal/(mol Å^2), followed by final 500 ps equilibration with a force constant of 0.5 kcal/(mol Å^2) to equilibrate the density. The above equilibration steps were all carried out at constant volume, 300 K (Langevin dynamics, collision frequency: 1/ps) with 1-atm constant pressure, and this was followed by at least 25-ns production MD simulation for each model under equivalent conditions. All calculations were performed on the SGI Pople machine at the Pittsburgh Supercomputing Center and also at the University of Utah's Center for High Performance Computing various clusters.

Results

Geometry optimizations

As an initial test of the calculations, we compared the geometries obtained for the smaller T-HM with previously published values for a larger set of heme states due to the lower computational cost than for the more extended F-HM. Energy optimizations of the T-HM structure model were performed at the B3LYP[47]/LACVP level for high-, intermediate-, and low-spin states of two penta-coordinated T-HM structures, the ferric (Fe^{3+}) and ferrous (Fe^{2+}). Key geometries such as the Fe—S distances, average Fe—N distances, and Fe out of N-plane values are in good quantitative agreement with previously published work,[48] as shown in Figure 3. General trends that can be observed for these states are that the maximum Fe—S bond length is observed for the intermediate-spin state for both of these penta-coordinate states, whereas the average Fe—N distance and deflection of Fe from the N-plane both increase from low to high spin. The differences in energies of the different ferric penta-coordinate spin states were all within 1.5 kcal/mol, whereas for the ferrous species, the intermediate triplet spin state was found to be 9.2 kcal/mol higher, with the low-spin singlet state 14.0 kcal/mol higher than the high-spin

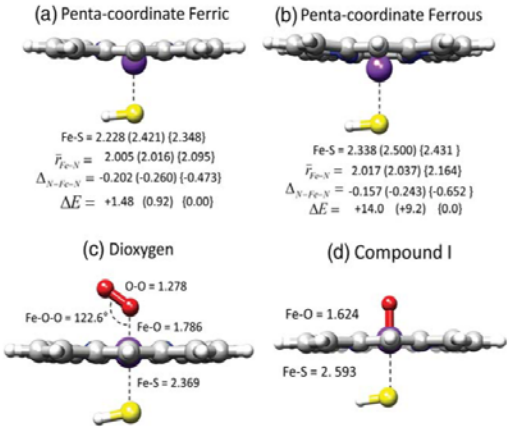


Figure 3. B3LYP/LACVP optimized geometry parameters of key heme species during P450-catalytic cycle (a–d) using the T-HM. Differences in parameters for different spin states are indicated as follows: low spin (medium/high spin). Calculated distances are in Angstroms. The relative difference in the sum of electronic and thermal energies is indicated for each species for which calculations were made using different spin states (kcal/mol). Adapted from Meunier et al. (2004).

quintuplet state again in good agreement with that in the literature,[48] as shown in Figures 3a and 3b.

For the T-HM hexa-coordinate states, the dioxygen-bound ferrous form of the T-HM, calculations were performed with the singlet state[48] and for compound I for the high-spin quartet state of compound I. Key distances and angles for the hexa-coordinate T-HM are shown in Figures 3c and 3d and are in good quantitative agreement with previously published work.[48]

Geometry optimizations at the same level of theory were then applied to the more extended model system. Comparison of the results for these calculations with previously published work is more difficult because the treatment of the substituents and the contribution to the free energy of the system by them can vary significantly due to the increased number of degrees of freedom of the extended F-HM system compared to the truncated T-HM.[11] Furthermore, the computational cost for the treatment of the larger system is much greater than that for the T-HM. Thus, the number of heme P450 catalytic cycle states that were modeled was limited to what could be performed at a reasonable computational expense for our largest model system, F-HM. For the F-HM penta-coordinate ferric and ferrous species, geometry optimizations were performed for the high- and intermediate-spin states. As we could not compare these differences directly to other published work, we compared observed trends in key geometries for the T-HM and also to that of published X-ray structure for P450cam. The heme in P450cam has the same substituents as observed in human hepatic P450, but X-ray data show a different orientation of one of the propionates. Key distances and angles from the gas-phase geometry optimizations for the F-HM penta-coordinate states are shown in Figure 4 alongside values from P450cam X-ray structures: PDB ID: 1DZ4 (ferric) and 1DZ6 (ferrous) for reference in blue. The trends in the geometries of the F-HM system are in agreement with those in the literature for similar more extended systems.

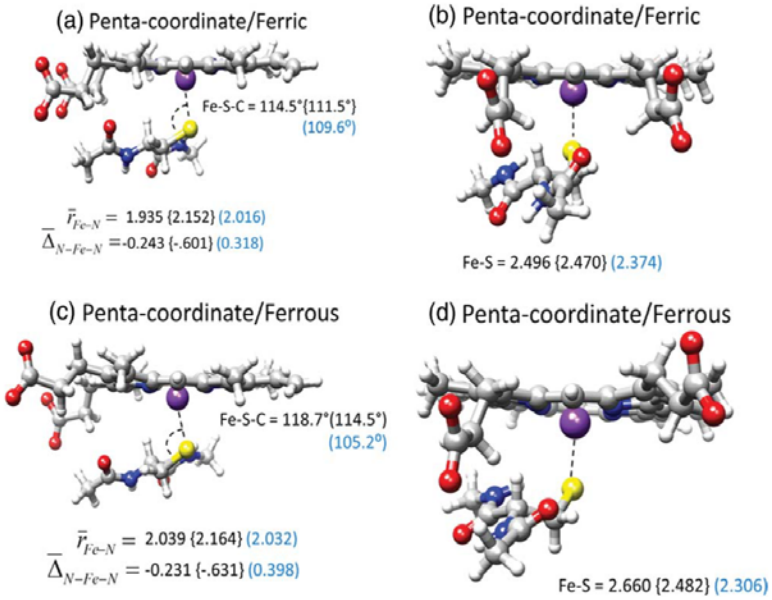


Figure 4. B3LYP/LACVP optimized geometry parameters of key heme species during P450-catalytic cycle (a–d) using the extended F-HM. Differences in parameters for different spin states are indicated as follows: intermediate spin (measurements from X-ray data shown in [blue a and b, PDB code: 1DZ4 and c and d, PDB code: 1DZ6]). Calculated distances are in Angstroms. Adapted from Meunier et al. (2004).

In brief, for the F-HM ferric intermediate- and high-spin states, the Fe—S distance is longer than that for the T-HM and the X-ray structure, the average Fe—N distances are shorter than the T-HM and in agreement with the X-ray structure value, deflection of the Fe from the N-plane is less for the intermediate spin and greater for the high-spin state than the T-HM while the X-ray value is within the limits of the values for both models. For the ferrous intermediate- and high-spin states, the Fe—S distance is longer but in closer agreement with the T-HM than the ferric state, and larger than that of the X-ray structure; the average Fe—N distances are shorter than the T-HM and in closer agreement with the X-ray structure value; the deflection of the Fe atom from the N-plane is less for the intermediate-

and high-spin states than the T-HM, while the X-ray value is in agreement and within the limits of the values from both models.

For the hexa-coordinate DIOXY state, we observe a longer Fe—S and Fe—O distances for the F-HM than the T-HM and the X-ray structure. In general, the values for the T-HM are closer to the X-ray structure as observed previously. The Fe—S—C angle more acute in the X-ray structures than in the F-HM-based calculations. Also, for the DIOXY-bound F-HM structure model, the Fe—O and O—O bond lengths are longer than what is observed in the X-ray structure PDB ID: 1DZ8.[8] The Fe—O—O bond angle is more acute in both the T-HM and F-HM than observed in the X-ray structure, whereas the Fe—S—C is more obtuse. The average Fe—N distance is very similar in both models and close to that of the X-ray structure.

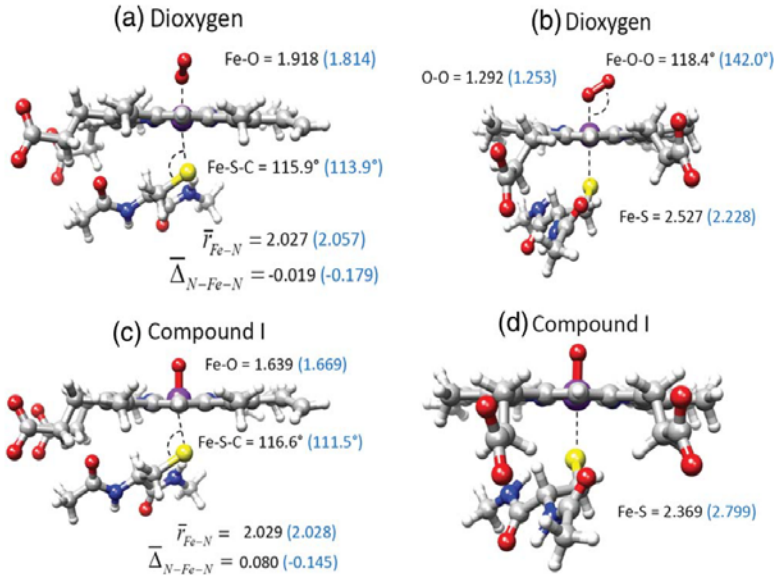


Figure 5. B3LYP/LACVP optimized geometry parameters of hexa-coordinate heme species during P450-catalytic cycle (a–d) with extended F-HM. Distances are in Angstroms. Measurements from X-ray data shown in blue (a and b, PDB code: 1DZ8 and c and d, PDB code: 1DZ9.). Adapted from Meunier et al. (2004). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

For the F-HM high-spin quartet CPDI structure, key distances and geometries are provided and are compared to the values from the P450cam structure PDB ID: 1DZ9[8] in Figures 5c and 5d. Despite the presence of some ambiguity of the electron density data,[49] this structure provides the only available X-ray structural data to date for CPDI in a P450 enzyme and the values from our calculations are in good agreement with key geometries inferred from this experimental data,[8,50] for F-HM. With the T-HM structure, Fe—O, R_{avg} Fe—N distances are also in good agreement with experiment. Both structure models underestimate the observed Fe—S bond length although the T-HM value is closer. This is not surprising because the models lack context of the protein which may alter bond lengths. Also, the Fe—S—C angles are slightly more acute in the F-HM than are observed in experiment. Both of our gas-phase model structures predict that in the absence of active site constraints the Fe will be slightly above the N-plane, which agrees with related calculations,[11] noting that this is contrary to what is found in the X-ray structure of P450cam.[8]

Charge and spin

As the results of our geometry optimizations were in reasonable agreement with the results of other published calculations and X-ray structures, we then used these geometries to derive atomic partial charges and compare the atomic partial charge values for these open-shell states for the T-HM and F-HM. Atomic partial charges were calculated with the RED-III[21] program that provides for multiple conformations and orientations in the fit, as described in greater detail in the 'Methods' section. A few general trends for the RED-III calculated atomic partial charges for the two-different model systems are given in Table 1. For all the heme states, a greater asymmetry is observed in the distribution of atomic partial charges for the F-HM structure models than is observed with the T-HM, due to the greater asymmetry of the F-HM.

For the penta-coordinate, high- and intermediate-spin states, similar positive charges are observed on the iron atom, along

Table 1. RED-III atomic partial charges at nonbackbone heteroatoms for (A) the extended F-HM and (B) the T-HM systems.

	CPDI	DIOXY	IC6	IC4	OUS5	OUS3
(A) F-HM						
N (sum)	0.0676	1.1168	−0.5112	−0.8098	−0.3531	0.0746
Fe1	0.262	0.0289	0.2492	0.048	0.1462	0.0297
S79	−0.4381	−0.1587	−0.4602	−0.3723	−0.5562	−0.6037
O71 and O73 (avg)	−0.65515	−0.71915	−0.5791	−0.6017	−0.7117	−0.7181
O70 and O72 (avg)	−0.65075	−0.7168	−0.64265	−0.6851	−0.72285	−0.7212
O (proximal)	−0.3729	−0.1174				
O (distal)		−0.1945				
(B) T-HM						
N (sum)	0.4311	0.5947	−0.5441	−0.0863	−1.4568	−0.3553
Fe1	0.033	0.131	0.2769	0.047	0.6999	0.0592
S38	−0.3835	−0.6807	−0.379	−0.424	−0.7309	−0.6977
O (proximal)	−0.3422	−0.2078				
O (distal)		−0.2295				

with similar net negative charges on the nitrogen atoms with both T-HM and F-HM structure models. For the sulfur atom, a greater negative charge is calculated for the F-HM structure model for the sextet, and a smaller negative charge is calculated for the quartet state compared to the T-HM. For the reduced penta-coordinate ferrous quintet and triplet states, greater charges are observed for the sulfur and iron atoms, and net negative charges are observed for the nitrogen atoms comparing the T-HM to the F-HM. The F-HM structure model predicts a net positive charge on the nitrogen atoms for the ferrous triplet state (0.0746) as opposed to the T-HM that predicts a net negative charge (−0.3553).

For the hexa-coordinate states, the dioxygen-bound structure is a singlet and therefore this F-HM structure model is in principle a good 'control' to look out for potential changes in the atomic partial charges due to shifts in unpaired electron-spin density due to the absence of shielding of the propionates by basic active site residues in these gas-phase calculations. However, with this state, we observe significant differences between the partial charges calculated with the different structural models. For the nonbackbone heteroatoms, we observe much greater net atomic partial charge on the sulfur, iron, and molecular-oxygen atoms with the T-HM than the F-HM. The nitrogen atoms exhibit greater asymmetry in their charge distribution while bearing positive charges in both models with the sum of these charges being almost double for the F-HM (1.1168) than the T-HM (0.5947).

For compound I, the F-HM predicts a greater positive charge on the iron atom, slightly greater negative charges on the sulfur and axial oxygen atoms, and a much lower net positive charge on the nitrogen atoms than the T-HM.

It has been previously shown that for gas-phase calculations of extended heme models, charge-transfer artifacts can be introduced due the absence of full shielding of the propionates by active site basic residues, and thus spin-density delocalization to the oxygen atoms of the propionate groups.[51] As we observed significant differences in atomic partial charges for T-HM and F-HM systems for each equivalent state of the catalytic cycle, we more closely investigated if the differences were due in major part to electron-spin delocalization to the propionates. As the DIOXY state is a closed shell system it should not suffer from any charge-transfer artifacts due to unpaired spin delocalization to the propionate oxygen atoms and should so provide a good reference for the values calculated for the other open-shell heme states. The ferrous open-shell penta-coordinate high- and intermediate-spin states produced similar negative atomic partial charge values for the propionate oxygen atoms as observed in the DIOXY state. However, the ferric ic4, ic6, and the CPDI states show differences in the atomic partial charge values compared with the DIOXY and ferrous states (Table 1).

Mulliken atomic spin densities

To better understand the delocalization of spin densities in these gas-phase calculations for the F-HM, we compared these values at key atoms to the T-HM. The T-HM cannot suffer from spin delocalization to propionates and has been observed to be a better model system for determining quantum effects such as spin densities in gas-phase calculations.[11] As given in Table 2, the sum of the Mulliken atomic spin densities is

Table 2. Mulliken atomic spin densities for the T-HM and the extended F-HM systems from the B3LYP/LACVP geometry optimizations.

	IC6	IC4	OUS5	OUS3	CPDI
F-HM					
Fe	3.662	2.453	3.661	1.969	1.187
S	0.116	0.393	0.108	0.108	−0.019
N (sum)	0.176	−0.025	0.171	−0.063	−0.007
Porph (—N)	−0.017	0.025	0.050	−0.018	−0.128
Cys (—S)	0.012	0.014	0.011	0.005	0.003
Prop O (sum)	1.052	0.140	0.001	−0.002	1.062
O					0.901
T-HM					
Fe	3.904	2.465	3.645	1.911	1.075
S	0.474	0.485	0.141	0.199	0.554
N (sum)	0.531	−0.013	0.166	−0.023	0.384
Porph (—N)	0.094	0.070	0.046	−0.080	0.059
O					0.943

N (sum) is the sum of the Mulliken spin densities on all the heme nitrogen atoms, Porph (—N) is the sum of the spin densities on the protoporphyrin (F-HM) or porphyrin (T-HM) minus the N (sum) value, Cys (—S) is the sum of the spin densities on the blocked cysteine ligand minus the spin density on the sulfur atom (F-HM), Prop O (sum) is the sum of the spin densities on the propionate oxygen atoms.

increased on propionate oxygen atoms for the open-shell species that show a difference in partial charges, and that spin density is delocalized on the propionate oxygen atoms more for the ic4, ic6, and CPDI than for ferrous states as previously observed.[51] Closer examination of the spin densities for the penta-coordinate states from the optimization calculations reveals very similar values for the iron atom predicted for the all the penta-coordinate states. For the ferric high-spin, we observe less spin density on the sulfur and greater spin density on the porphyrin for the F-HM compared with the T-HM. The ferric states show greater spin density on the nitrogen atoms and porphyrin for the T-HM, whereas for the ferrous high spin shows similar values between the two structure models and the ferrous intermediate values differ between the two structure models. For CPDI, we observe similar spin on the iron and oxygen with the two models but different spin on the sulfur, nitrogen atoms, and porphyrin between the two models.

As we calculated the atomic partial charges with RED-III, and this method uses multiple reorientations of the molecule to 'average' out the charges to make them more reproducible, we examined the Mulliken spin densities for each of these calculations. We found that there was an orientation dependence for the Mulliken atomic spin calculations (see Supporting Information Table 2 for the averages and standard deviation of Mulliken charges and spin for T-HM and F-HM CPDI). Therefore, we calculated the averages for Mulliken atomic spin densities for all reorientations made for each heme states during the RED-III charge fitting for both T-HM and F-HM and these are presented in Table 3. Comparison of the agreement between the average spin-density values from six reorientations for each

Table 3. Average Mulliken atomic spin densities from RED-III reorientation RESP calculations for (A) extended F-HM and (B) the T-HM.

	IC6 Avg	IC6 Std Dev	IC4 Avg	IC4 Std Dev	OUS5 Avg	OUS5 Std Dev	OUS3 Avg	OUS3 Std Dev	CPDI Avg	CPDI Std Dev
(A) F-HM										
Fe	3.814942	0.068763	2.632201	0.000106	3.660980	0.000067	1.968843	0.000066	1.077921	0.005778
S	0.224528	0.050995	0.181283	0.000103	0.107627	0.000052	0.107839	0.000026	0.316469	0.005664
Cys (w/NME&ACE)-	0.014580	0.000227	0.018785	0.000014	0.010580	0.000016	0.005450	0.000007	0.010583	0.000351
N (sum)	0.344437	0.078408	0.104677	0.000044	0.170377	0.000019	−0.062941	0.000081	0.481340	0.015486
Porph (—Fe)—[N (sum)]	0.000416	0.028386	0.159447	0.000023	0.048458	0.000017	−0.016633	0.000033	0.105598	0.082849
Prop (—O)	−0.042027	0.001366	−0.021686	0.000014	0.001281	0.000016	−0.002306	0.000026	0.066800	0.000222
O-prop (sum)	0.643122	0.152473	−0.074708	0.000023	0.000697	0.000001	−0.001761	0.000003	0.104380	0.099501
O (Opdl)									0.915887	0.004121
Spin annihilation										
Before	8.830883	0.033076	3.982900	0.000155	6.008900	0.000000	2.033867	0.000052	3.969400	0.070334
After	8.750933	0.000408	3.755733	0.000052	6.000000	0.000000	2.000600	0.000000	3.764667	0.007101
(B) T-HM										
Fe	3.903714	0.000022	2.465124	0.000166	3.644619	0.000013	1.910712	0.000094	1.075067	0.000080
S	0.474990	0.000104	0.484387	0.000031	0.141229	0.000030	0.199167	0.000086	0.554511	0.000058
(S)H	−0.003301	0.000005	−0.006393	0.000001	0.002452	0.000003	−0.006957	0.000003	−0.015107	0.000003
N (sum)	0.530832	0.000113	−0.013365	0.000133	0.166183	0.000039	−0.022582	0.000170	0.383719	0.000107
Porph (—Fe)—[N (sum)]	0.093765	0.000005	0.070246	0.000037	0.045518	0.000004	−0.080337	0.000145	0.059097	0.000010
									0.942711	0.000051
Spin annihilation										
Before	8.766000	0.000000	3.790833	0.000103	6.008900	0.000000	2.056233	0.000052	3.791667	0.000052
After	8.750100	0.000000	3.750500	0.000000	6.000000	0.000000	2.001333	0.000052	3.750800	0.000000

state to that of the original optimization indicates a closer agreement when comparing the average values, for the IC6 and CpdI states. However, the agreement between the F-HM and T-HM values for the ferric intermediate state appears to have worsened while differences between the ferrous penta coordinate remain relatively unchanged.

To identify the contribution of these differences in spin densities to the partial atomic charges was, we calculated the difference in charges and spin for key atoms and moieties of the heme molecule (see Table 4) and plotted the difference in charge versus the difference in spin between the T-HM and F-HM (T-HM value − F-HM value) for the iron, sulfur, and nitrogen atoms (Fig. 6) and fit these lines with a linear regression. No clear correlation is apparent between the differences of the RED-III charge and average Mulliken spin density, nor for Mulliken charges and Mulliken spin values from the optimizations, for any of these atoms.

For the propionate oxygen atoms, however, there does appear to be a decrease in negative charge from the average charge for the dioxy and ferrous states (average O71 and O73 value is −0.7163 and the average O70 and O72 value is −0.7203); this is may be due in part to the delocalization of unpaired electron density on the propionate oxygen atoms. However, the effect of this charge transfer to the rest of the system is unclear. Changes in charge for the iron, sulfur, nitrogen atoms, and other key moieties vary to such a degree between the T-HM and F-HM systems and heme states, even for closed shell systems, that the contribution of this charge-transfer amount to the rest of the system is impossible to identify.

Molecular docking

As we could not observe a direct correlation between the spin-density delocalization and the atomic partial charges calculated with RED-III, we tested the use of the charges empirically derived from the T-HM and the F-HM. Previous docking studies from our lab have shown that the incorporation of more accurate atomic partial charges for the heme with the x-ray structure of CYP3A4 (PDB ID: 1W0E)[20] improved the docking results and identified specific residues involved in substrate binding that were verified experimentally.[18] Specifically, we observed that with the better charge models, there was an increase in the number of conformations found through docking that are consistent with the observed metabolism of raloxifene. We further identified and experimentally confirmed a single residue (Phe215) within the active site that significantly contributed to competing metabolic pathways. Thus, to test the influence of the current set of B3LYP/LACVP optimized structure models and RED-III calculated RESP atomic partial charges for different heme states during the P450 catalytic cycle on the molecular docking results, we used the same CYP3A4 template modified with these new QM-based heme coordinates and charges for additional docking simulations. Here, dockings were performed with the F-HM penta-coordinate ferric high-spin [IC6 (F-HM)] and intermediate-spin [IC4 (F-HM)] geometries and RED-III-derived charges, F-HM penta-coordinate ferric high-spin [IC6 (Oda)] and intermediate-spin [IC4 (Oda)] geometries with previously published[14(a)] charges for the ferric penta-coordinate state, F-HM hexa-coordinate molecular-oxygen-bound [DIOXY (F-HM)] geometries with RED-III-derived partial charges, F-HM hexa-coordinate catalytic compound I [CPDI (F-HM)] geometries with RED-III-derived partial charges, and T-HM hexa-coordinate catalytic compound I [CPDI (T-HM)] RED-III-derived atomic partial charges assigned to the F-HM-derived geometry modified with gas-phase propionic acid-derived charges at the B3LYP/6-31G* level. In this work, we scored with slightly less stringent criteria

Table 4. Differences in charge and spin for key atoms and moieties of the T-HM and extended F-HM.

	CPDI	DIOXY	IC6	IC4	OUS5	OUS3
Diff charge						
Fe	−0.2290	0.1021	0.0277	−0.0010	0.5537	0.0295
S	0.0546	−0.5220	0.0812	−0.0517	−0.1747	−0.0940
N (sum)	0.3635	−0.5221	−0.0329	0.7235	−1.1037	−0.4299
Propionate (Diff chg)						
O71 and O73 (each)	−0.0612	0.0028	−0.1372	−0.1146	−0.0046	0.0018
O70 and O72 (each)	−0.0695	−0.0035	−0.0776	−0.0352	0.0026	0.0009
Sum of diff for all O	−0.1307	−0.0007	−0.2149	−0.1498	−0.0021	0.0027
Propionate (charge) non-O	1.2519	1.2191	1.2539	1.1751	1.2376	1.2495
Porph (—Fe) − [N (sum)] (CHG)						
F-HM	−0.2110	−1.7144	−0.1459	0.3702	−0.5986	−0.8131
T-HM	0.1028	−0.7450	0.4774	0.3013	0.0998	0.3404
Diff	0.3138	0.9694	0.6233	−0.0689	0.6984	1.1535
CYP (CHG)						
Sidechain − (S)	0.1666	−0.1933	0.1722	0.2766	0.1079	0.0558
Sidechain + (S)	−0.2715	−0.352	−0.288	−0.0957	−0.4483	−0.5479
CYP	−0.3859	−0.4664	−0.4024	−0.2101	−0.5627	−0.6623
Diff avg spin						
Fe	−0.0029		0.0888	−0.1671	−0.0164	−0.0581
S	0.2380		0.2505	0.3031	0.0336	0.0913
Cys ((w/NME & ACE)-(S))	−0.0257		−0.0179	−0.0252	−0.0081	−0.0124
N (sum)	−0.0976		0.1864	−0.1180	−0.0042	0.0404
Porph (—Fe) − [N (sum)]	−0.0465		0.0933	−0.0892	−0.0029	−0.0637
Propionate (spin)						
Prop (—O)	0.0668		−0.0420	−0.0217	0.0013	−0.0023
O-prop (sum)	0.1044		0.6431	−0.0747	0.0007	−0.0018
Porph (—Fe) − [N (sum)] (spin)						
F-HM	0.1056		0.0004	0.1594	0.0485	−0.0166
T-HM	0.0591		0.0938	0.0702	0.0455	−0.0803
Diff	−0.0465		0.0933	−0.0892	−0.0029	−0.0637
CYP(spin)						
S	0.3165		0.2245	0.1813	0.1076	0.1078
CYP ((w/NME & ACE)-(S))	0.0106		0.0146	0.0188	0.0106	0.0055
CYP ((w/NME & ACE)-(S))	0.3271		0.2391	0.2001	0.1182	0.1133

for identification of the site of metabolism for all model systems and atomic partial charges, allowing a distance of <6.0 Å from the Fe to the substrate heavy atom for the ferric states and <4.5 Å from the proximal/activated oxygen atom for predicting the site of metabolism. We relaxed the criteria for the 'resting' penta-coordinate state, as the QM calculations predict that the Fe atom is below the plane of the heme, whereas the x-ray structure used in our previous work assigns that Fe atom within the plane of the heme. We also increased the stringency of the criteria, correspondingly, for the hexa-coordinate heme parameters with the catalytic oxygen present.

For the F-HM ferric high-spin state, our results indicate that the use of the new RED-III-derived atomic partial charges produces results consistent with our previous work (see Supporting

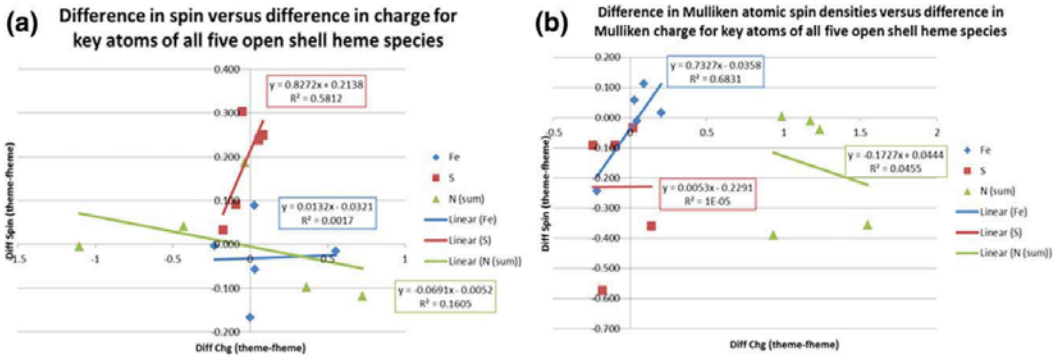


Figure 6. Differences in a) RED-III atomic partial charges versus average Mulliken atomic spin densities and b) Mulliken charges and atomic spin densities from geometry optimization calculations for the truncated heme (T-HM) and the full heme (F-HM) models. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Information Tables 3A–3C). Furthermore, the number of conformations that bind in a manner consistent with the observed metabolism is increased. Use of the ferric intermediate-spin parameters produced a greater number of large clusters containing only nonproductive binding states compared to the high-spin state, and this model also failed to produce any large clusters that predicted the observed hydroxylation of raloxifene (see Table 5 and Supporting Information Tables 3A–3C).

Docking results using the dioxygen-bound hexa-coordinate full heme geometries and RED-III-derived charges [DIOXY (F-HM)] also did not produce any significant population predicting the observed hydroxylation and dramatically reduced the number of nonproductive binding modes (see Table 5 and Supporting Information Tables 3A–3C). The use of the hexa-coordinate heme parameters for the catalytic oxyferryl CPDI produced as the highest populated and lowest energy clusters, binding modes supporting the observed dehydrogenation of raloxifene, in addition to one major cluster predicting nonproductive binding modes and a minor cluster predicting the observed hydroxylation of raloxifene (see Table 5 and Supporting Information Tables 3A–3C).

Table 5. Percentage of binding modes of AutoDock3.0 results for CYP3A4 (PDB: 1W0E) and raloxifene for known site of metabolism with different heme parameters.

Overall	Dehydrogenation	3′OH	Nonproductive
IC4 (F-HM)	0.474	0.000	0.526
IC6 (F-HM)	0.662	0.042	0.296
IC4 (Oda)	0.269	0.000	0.731
IC6 (Oda)	0.325	0.000	0.675
DIOXY (F-HM)	0.846	0.000	0.154
CPDI (F-HM)	0.521	0.068	0.411
CPDI (T-HM)	0.329	0.032	0.639

Comparison of docking studies performed with these new F-HM charges against previously calculated charges[14(a)] assigned to either of our QM-optimized penta-coordinate ferric heme geometries (ic4-Oda: our intermediate-spin ferric and ic6-Oda: our high-spin ferric with the previously published charges[14a]) shows that with this model system, we see improved results with our new charges (see Table 5 and Supporting Information Tables 3A–3C).

Comparison of docking studies with the RED-III-derived atomic partial charges for the hexa-coordinate full heme model of compound I [CPDI (F-HM)] versus the RED-III-derived atomic partial charges for the truncated heme compound I with charges for the gas-phase propionates added and the other macrocycle substituents left neutral [CPDI (T-HM)], indicate again that the CPDI (F-HM) charges produce improved docking results with this model system. The results suggest a profound dependence on the choice of heme model and catalytic state.

In summary of our docking studies, our results demonstrate that the incorporation of atomic partial charges for the heme in each of the different states calculated for the P450 catalytic cycle produce different results for each set of atomic partial charges for each heme state and model system, including different spin states for the same penta-coordinate ferric charge state, as well as for both hexa-coordinate, dioxygen-bound and the catalytic CPDI species. Surprisingly, for this particular substrate-P450 combination with this force field, despite charge-transfer artifacts identified earlier, the open-shell F-HM-derived charges for the resting ferric high-spin and the catalytic CPDI species produce results most closely in agreement with experimentally observed metabolism of raloxifene by CYP3A4. This may be in part due to the fact that the RED-III multiple reorientation-derived charges for F-HM-unshielded propionates exhibit a smaller charge transfer than observed elsewhere.[51] This would suggest that the extended model F-HM produces a better representation for the electrostatics of the heme than the T-HM despite the absence of shielding of basic active site residues in these gas-phase calculations. Interestingly, the docking study was also sensitive to the presence of molecular oxygen with the DIOXY (F-HM) parameters not supporting both of the observed mechanisms of P450 metabolism, whereas with the CPDI (F-HM) parameters for the catalytic compound I species produced binding modes suggestive of both hydroxylation and dehydrogenation of raloxifene.

Bond force constants and molecular dynamics

For performing molecular dynamics simulations of the heme alone or of the P450 enzymes, heme force constant parameters were required that were absent from the GAFF parameters. Specifically, these heme parameters were for the porphyrin moiety, and key chelations to the iron atom. Thus, scans for all the missing parameters were performed with the T-HM, except the Fe—S—C describing the angle for the iron bond angle to the proximal cysteine which was performed with the F-HM (Fig. 7).

The Fe—S and Fe—N force constants were obtained for the MM bond parameters by fitting the QM scan results and are relatively similar for the penta-coordinate and dioxygen-bound states (Fig. 7). By contrast, the estimated force constant values for these bonds for CPDI differed substantially. Furthermore, the Fe—O force constant for CPDI was also much greater than the force constant for the DIOXY state (Table 6).

For the angle terms, the force constants varied more between the different heme species. The hexa-coordinate states were similar in value for the Fe—N—C and N—Fe—O. The penta-coordinate and the dioxygen-bound states were closer in value for the N—Fe—N than CPDI. The C—S—Fe angle term was determined with a scan only due to computational cost using the F-HM of CPDI and fit in a similar manner (Table 6). An alternative strategy, when a complete set of bond and angle parameters are desired could be used by using the MTK++ code in AmberTools to mechanically derive the force matrix quantum.[38] This approach has the advantage over the method used here in that it avoids defining internal coordinates; however, the influence of small changes in bond and angle parameters introduced to the force constants that we have already developed and tested to be compatible with the AMBER ff99SB force field and GAFF in both MD and docking,

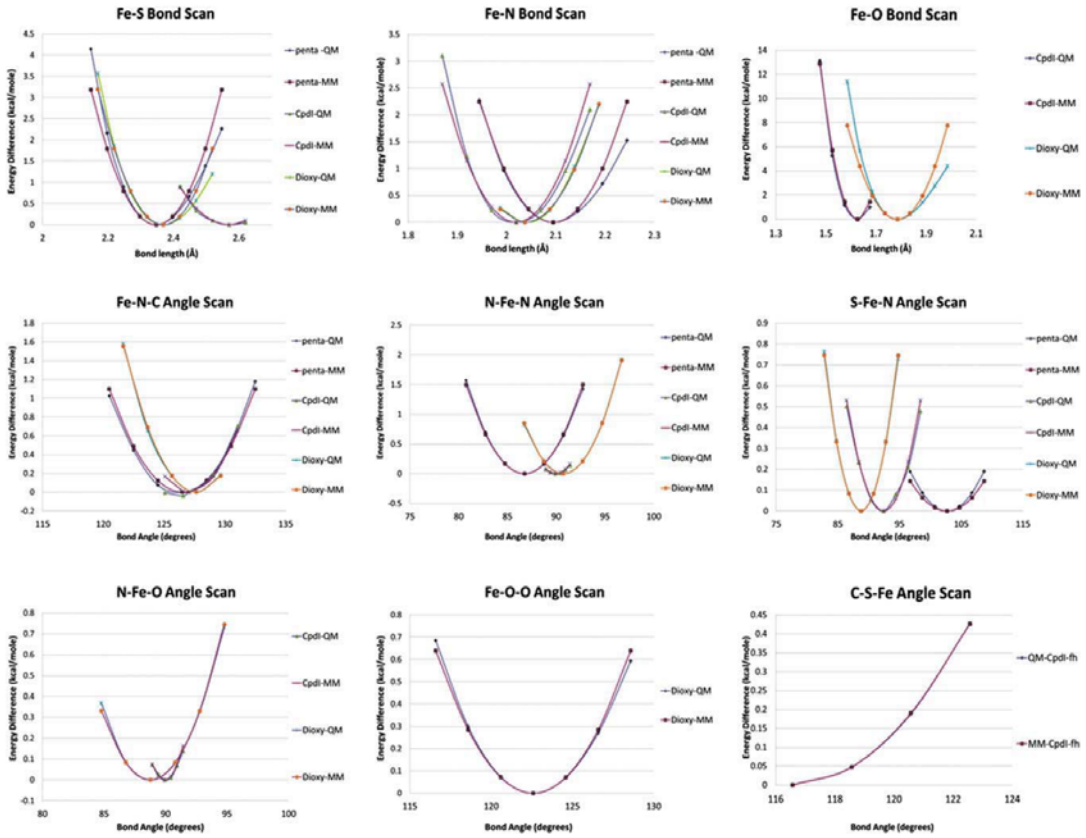


Figure 7. Fit of QM scans along key coordinates for the development of MM force constants. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

is likely minimal. We do recommend the use of this approach for the next generation of heme parameters.

To determine if the GAFF[36]-based heme models, with F-HM-based RESP charges and with the additional T-HM-based force constants were sufficient to maintain the heme structure, molecular dynamics simulations were performed on the various heme models with a generalized Born implicit solvent treatment (GB-MD). The initial GB-MD runs indicated that the penta-coordinate state required the addition of a number of dihedral angle energy terms to maintain the iron in an out-of-plane state as observed in the QM geometry optimizations. These dihedral angles were X–S–Fe–X, X–cg–cc–X, X–cg–cd–X, cc–nc–fe–nc, cd–nd–fe–nd, cc–nc–fe–SH, and cd–nd–fe–SH (see Supporting Information). For the hexa-coordinate states, the only extra dihedral angle energy term that was included was for X–cg–cd–X and X–cg–cc–X for CPDI and the DIOXY states.

Molecular dynamics of the F-HM in implicit solvent

Visual inspection of the average structures generated for the heme models over the entire 10-ns simulation appears reasonable in that the heme maintains its overall "planarity" and the expected movement is occurring with axial substituents and the cysteinate ligand (Figs. 8a–8c). A visual inspection of the MD trajectory determined that the larger fluctuations in the RMSD occur during movement of the cysteinate ligand. RMSD plots of 10-ns GB-MD simulations for each of the three different heme states (Fig. 8d) suggest that these parameters are sufficient to maintain the expected heme structure.

Table 6. Calculated MM force constant parameters for bond stretching and angle bending from fit of the energy profiles derived from T-HM all except *, which was derived with F-HM.

	ic6	DIOXY	CPDI
Bonds			
Fe—S	80	80	39
Fe—N	100	98	114
Fe—O		194	572
O—O		372	
Angles			
Fe—N—C	100	142	136
N—Fe—N	137	174	239
S—Fe—N	13	68	48
N—Fe—O		68	65
Fe—O—O		58	
Fe—S—C			39*

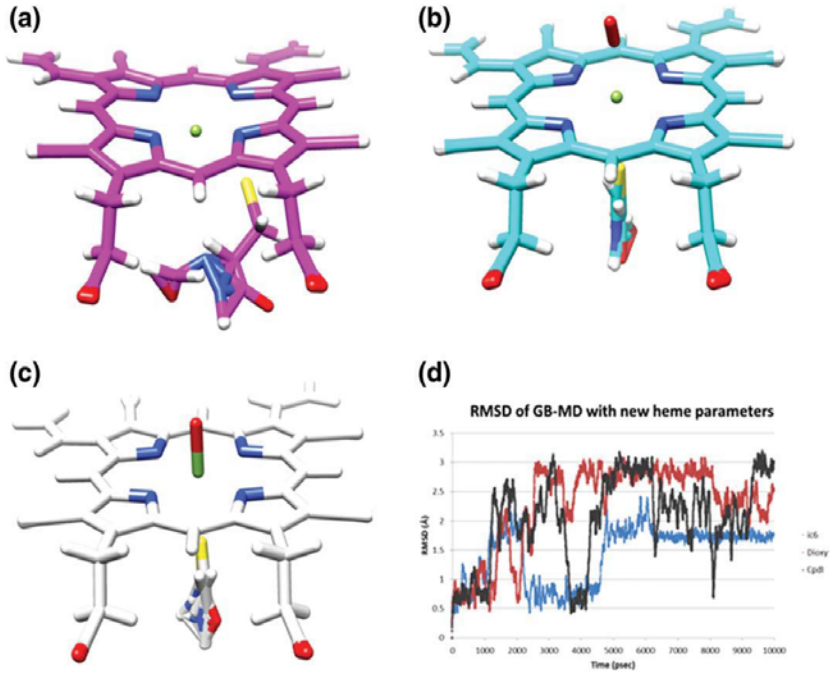


Figure 8. Average structure over 10-ns GB-MD simulation with RED-III derive atomic partial charges and new force constants: a) penta-coordinate ferric high-spin, b) hexa-coordinate dioxygen-bound, c) compound I heme states, and d) RMSD of three GB-MD simulations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Molecular dynamics of the CYP3A4 protein

MD simulation of the QM parameterized heme states in models of CYP3A4 (PDB ID: 1TQN and 1W0E) was also performed to assess the heme models. Key geometries of the heme for the F-HM structure model with CPDI over the course of full MD simulations of the protein in explicit solvent are shown in Figure 9. The heme Fe—O and Fe—S bond distances and the n-Fe—N, N—Fe—O and N—Fe—S angles appear to remain stable over the course of the ~25-ns MD simulation. The P450 structure appears to reach a steady state based on the RMSD plot 9(C) after 5 ns, and the 2D RMSD indicates that a steady state is reached after 10 ns that differs from that reached during the initial 10 ns. During these small conformational readjustments of the P450 model, the heme geometry maintains key equilibrium geometries with the new heme parameters. Similar steady-state results are observed for the 1W0E structure for CPDI and for both the dioxy and penta-coordinate ferric state with CYP3A4 models 1TQN and 1W0E (data not shown).

Discussion

Accurate description of the electronic ground state of different P450 heme species has been clearly shown to require full QM/MM treatments of the enzyme system based on experimental structures. Yet smaller and more tractable gas-phase models—despite displaying model system specific limitations—have been shown to provide useful insights into reaction mechanisms.[11] The smaller porphyrine–thiolate model system, such as the T-HM, has been especially useful in this manner and has proven to be surprisingly accurate producing electronic spin densities in good agreement with the results from QM/MM treatment of the full enzyme system.[51] As QM gas-phase calculations using truncated model systems have proven useful for the development of transferable parameters for classical MD force fields such as for AMBER[23]/GAFF,[36] simpler and less computationally expensive truncated models of the heme system has been used for earlier studies for MM parameter development for the resting heme state of the P450 catalytic cycle.

Here, we have tested the use of two model systems, T-HM and F-HM, for the development of MM parameters. The results of the geometry optimizations show that the T-HM more closely matches key axial geometries with available X-ray structures, but lacks some of the experimentally observed asymmetry. Both the F-HM and T-HM are closer in agreement on 'in plane' heme geometries with each other than available X-ray results. This is in agreement with other published work and raises the question of what effect these differences may produce in the charge and spin.

Our calculations show that the charges do vary between the two systems, and that charge transfer is observed, in agreement with previously published work,[51] but that these differences are smaller than that observed elsewhere when charges are derived using RESP with RED-III. Furthermore, the changes in atomic partial charges because of the charge transfer due

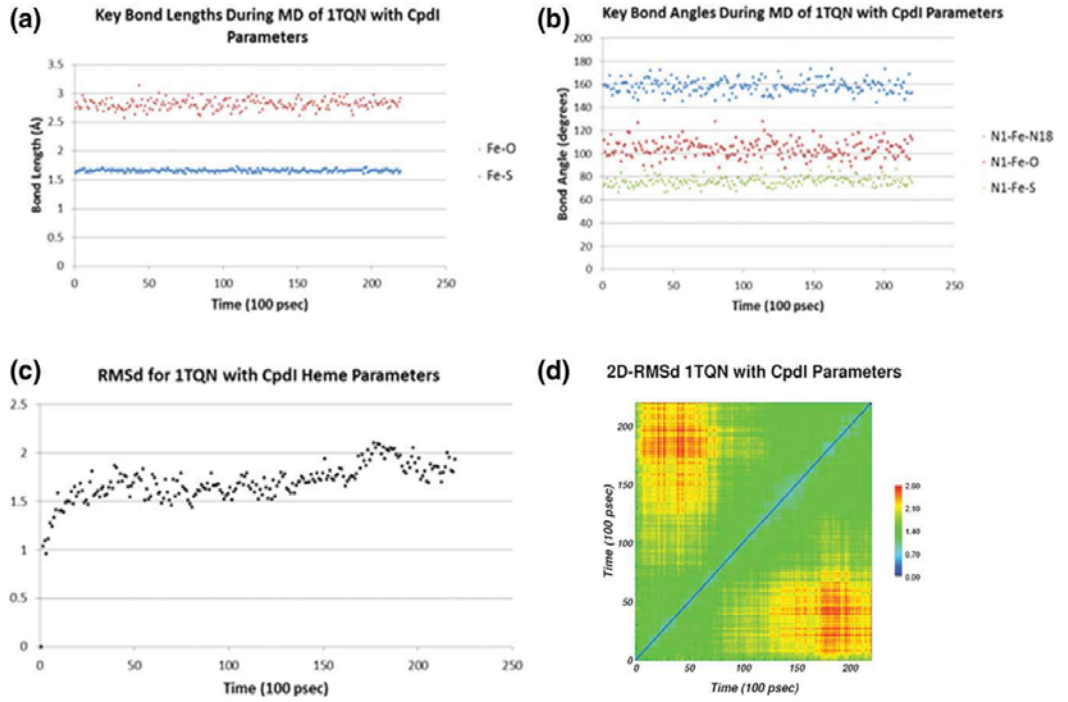


Figure 9. a) Key heme bond distances, b) key heme bond angles, c) 2D RMSD of Cα, and d) RMSD of MD simulation of 3A4 (PDB: 1TQN) with the parameters for compound I (CpdI). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

to "artificial" unpaired electron-spin density on the propionate oxygen atoms are difficult to correlate with changes in atomic partial charges of the entire system. We believe that this is inherit to differences between the model systems, as it is also observed to occur with the closed shell DIOXY state.

As no clear difference or superiority is observed in partial atomic charges derived from either model system, we empirically tested both the T-HM and F-HM-derived charges and previously published charges with a model system that had been experimentally confirmed in our lab. Our results indicate that with AMBER-based docking software that is sensitive to changes in geometries and charges derived from different spin states for the heme, despite the issues with charge transfer from the F-HM produce improved docking results.

Therefore, we developed and tested a set of MM parameters for/with the new AMBER ff99SBildn force field using the T-HM for the force constants and the F-HM charges. These models appear to be stable and produce reasonable results both in GB-MD and with MD of the full enzyme system in explicit solvent.

It remains to be seen how these parameters perform in simulations with other P450 enzymes, substrates, and ligands and which parameters will need additional refinement. However, on the basis of our current experience in MM applications, we recommend these B3LYP/LACVP-derived parameters and RESP-derived charges which are provided in the Supporting Information for the various P450-catalytic cycle heme states in improving P450 MM-based simulations.

Acknowledgments

Thomas E. Cheatham, III acknowledges financial support from NIH GM079383, computer time from the Center for High Performance Computing at Utah and from the NSF TeraGrid MCA01S027. Garold S. Yost acknowledges financial support from NIH GM0742249.

[1] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hunenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. van der Vegt, H. B. Yu, Angew Chem Int Ed 2006, 45, 4064.

[2] L. Afzelius, C. H. Arnby, A. Broo, L. Carlsson, C. Isaksson, U. Jurva, B. Kjellander, K. Kolmodin, K. Nilsson, F. Raubacher, L. Weidolf, Drug Metabol Rev 2007, 39, 61.

[3] H. Sun, D. O. Scott, Chem Biol Drug Design 2010, 75, 3.

[4] (a) Q. Wang, J. R. Halpert, Drug Metab Dispos 2002, 30, 86; (b) S. Ekins, Biochem Soc Trans 2003, 31(Pt 3), 611; (c) F. P. Guengerich, Chem Res Toxicol 2008, 21, 70.

[5] M. Otyepka, J. Skopalik, E. Anzenbacherova, P. Anzenbacher, Biochim Biophys Acta 2007, 1770, 376.

[6] (a) E. E. Scott, J. R. Halpert, Trends Biochem Sci 2005, 30, 5; (b) E. E. Scott, Y. A. He, M. R. Wester, M. A. White, C. C. Chin, J. R. Halpert, E. F. Johnson, C. D. Stout, Proc Natl Acad Sci USA 2003, 100, 13196; (c) E. E. Scott, Y. Q. He, J. R. Halpert, Chem Res Toxicol 2002, 15, 1407; (d) E. E. Scott, M. Spatzenegger, J. R. Halpert, Arch Biochem Biophys 2001, 395, 57; (e) E. E. Scott, M. A. White, Y. A. He, E. F. Johnson, C. D. Stout, J. R. Halpert, J Biol Chem 2004, 279, 27294.

[7] (a) D. E. Bikiel, L. Boechi, L. Capece, A. Crespo, P. M. De Biase, S. Di Lella, M. C. Gonzalez Lebrero, M. A. Marti, A. D. Nadra, L. L. Perissinotti, D. A. Scherlis, D. A. Estrin, Phys Chem Chem Phys 2006, 8, 5611; (b) A. Dey, Y. Jiang, P. Ortiz de Montellano, K. O. Hodgson, B. Hedman, E. I. Solomon, J Am Chem Soc 2009, 131, 7869; (c) P. R. Ortiz de

Montellano, Chem Rev 2010, 110, 932; (d) P. R. Ortiz de Montellano, J. J. De Voss, Cytochrome P450, 3rd ed.; Kluwer Academic/Plenum Publishers. New York, NY, USA, 2005, p. 183.
[8] I. Schlichting, J. Berendzen, K. Chu, A. M. Stock, S. A. Maves, D. E. Benson, R. M. Sweet, D. Ringe, G. A. Petsko, S. G. Sligar, Science 2000, 287, 1615.
[9] P. R. Ortiz de Montellano, R. Paul, Eds. Cytochrome P450, Structure, Mechanism and Biochemistry, 3rd ed.; Kluwer Academic/Plenum Publishers: New York, 2003.
[10] (a) S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, Chem Rev 2005, 105, 2279; (b) D. Harris, G. Loew, L. Waskell, J Inorg Biochem 2001, 83, 309; (c) C. M. Bathelt, L. Ridder, A. J. Mulholland, J. N. Harvey, J Am Chem Soc 2003, 125, 15004.
[11] S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar, W. Thiel, Chem Rev 2010, 110, 949.
[12] G. D. Szklarz, J. R. Halpert, J Comput-Aided Mol Des 1997, 11, 265.
[13] F. Autenrieth, E. Tajkhorshid, J. Baudry, Z. Luthey-Schulten, J Comput Chem 2004, 25, 1613.
[14] (a) A. Oda, N. Yamaotsu, S. Hirono, J Comput Chem 2005, 26, 818; (b) P. Rydberg, L. Olsen, P.-O. Norrby, U. Ryde, J Chem Theory Comput 2007, 3, 1765.
[15] (a) E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas, K. Schulten, Structure 2009, 17, 1295; (b) R. Lavery, K. Zakrzewska, D. Beveridge, T. C. Bishop, D. A. Case, T. Cheatham, III, S. Dixit, B. Jayaram, F. Lankas, C. Laughton, J. H. Maddocks, A. Michon, R. Osman, M. Orozco, A. Perez, T. Singh, N. Spackova, J. Sponer, Nucleic Acids Res 2010, 38, 299; (c) R. O. Dror, M. O. Jensen, D. W. Borhani, D. E. Shaw, J Gen Physiol 2010, 135, 555.
[16] (a) J. Kottalam, D. A. Case, J Am Chem Soc 1988, 110, 7690; (b) D. A. Case, J. A. McCammon, Annals N Y Acad Sci 1986, 482, 222.
[17] S. C. Hoops, K. W. Anderson, K. M. Merz, J Am Chem Soc 1991, 113, 8262.
[18] C. D. Moore, C. A. Reilly, G. S. Yost, Biochemistry 2010, 49, 4466.
[19] M. Ekroos, T. Sjogren, Proc Natl Acad Sci USA 2006, 103, 13682.
[20] P. A. Williams, J. Cosme, D. M. Vinkovic, A. Ward, H. C. Angove, P. J. Day, C. Vonrhein, I. J. Tickle, H. Jhoti, Science 2004, 305, 683.
[21] F. Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak, Phys Chem Chem Phys 2010, 12, 7821.
[22] G. M. Morris, D. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, J Comput Chem 1998, 19, 1639.
[23] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, J Am Chem Soc 1995, 117, 5179.
[24] C. I. Bayly, P. Cieplak, W. Cornell, P. Kollman, J Phys Chem 1993, 97, 10269.
[25] (a) P. C. Hariharan, J. A. Pople, Theor Chim Acta 1973, 28, 213; (b) M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees, J. A. Pople, J Chem Phys 1982, 77, 3654.
[26] (a) P. J. Hay, W. R. Wadt, J Chem Phys 1985, 82, 299; (b) T. H. Dunning, Jr., P. J. Hay, Methods of Electronic Structure Theory; Plenum: New York, 1977.
[27] A. Altun, D. Kumar, F. Neese, W. Thiel, J Phys Chem A 2008, 112, 12904.
[28] M. Radoń, E. Broclawik, J Chem Theory Comput 2007, 3, 728.
[29] J. K. Yano, M. R. Wester, G. A. Schoch, K. J. Griffin, C. D. Stout, E. F. Johnson, J Biol Chem 2004, 279, 38091.
[30] P. Rowland, F. E. Blaney, M. G. Smyth, J. J. Jones, V. R. Leydon, A. K. Oxbrow, C. J. Lewis, M. G. Tennant, S. Modi, D. S. Eggleston, R. J. Chenery, A. M. Bridges, J Biol Chem 2006, 281, 7614.
[31] Y. Zhao, M. A. White, B. K. Muralidhara, L. Sun, J. R. Halpert, C. D. Stout, J Biol Chem 2006, 281, 5973.
[32] (a) R. Seeger, J. A. Pople, J Chem Phys 1977, 66, 3045; (b) R. Bauernschmitt, R. Ahlrichs, J Chem Phys 1996, 104, 9047.
[33] M. Swart, A. R. Groenhof, A. W. Ehlers, K. Lammertsma, J Phys Chem A 2004, 108, 5479.
[34] C. D. Moore, K. Shahrokh, S. F. Sontum, T. E. Cheatham, G. S. Yost, Biochemistry 2010, 49, 9011.
[35] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Proteins 2006, 65, 712.
[36] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, J Comput Chem 2004, 25, 1157.
[37] (a) R. D. Hancock, Progress in Inorganic Chemistry, Vol. 37; Wiley, 1991; Hoboken, NJ, USA, p. 187; (b) K. M. Merz, J Am Chem Soc 1991, 113, 406.
[38] M. B. Peters, Y. Yang, B. Wang, L. S. Füsti-Molna'r, M. N. Weaver, K. M. Merz, J Chem Theory Comput 2010, 6, 2935.
[39] G. D. Hawkins, C. J. Cramer, D. G. Truhlar, J Phys Chem 1996, 100, 19824.
[40] D. A. Case, T. E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, J Comput Chem 2005, 26, 1668.
[41] (a) C. Simmerling, B. Strockbine, A. E. Roitberg, J Am Chem Soc 2002, 124, 11258; (b) K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, Proteins 2010, 78, 1950.
[42] I. S. Joung, T. E. Cheatham, III, J Phys Chem B 2008, 112, 9020.
[43] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, J Chem Phys 1983, 79, 926.
[44] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, J Chem Phys 1995, 103, 8577.
[45] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, J Comput Phys 1977, 23, 327.
[46] R. W. Pastor, B. R. Brooks, A. Szabo, Mol Phys 1988, 65, 1409.
[47] (a) A. D. Becke, J Chem Phys 1993, 98, 5648; (b) C. Lee, W. Yang, R. G. Parr, Phys Rev B 1988, 37, 785; (c) S. H. Vosko, L. Wilk, M. Nusair, Can J Phys 1980, 58, 1200; (d) P. J. Stephens, F. J. Delvin, C. F. Chabalowski, M. J. Frisch, J Phys Chem 1994, 98, 11623.
[48] B. Meunier, S. P. de Visser, S. Shaik, Chem Rev 2004, 104, 3947.
[49] R. Davydov, T. M. Makris, V. Kofman, D. E. Werst, S. G. Sligar, B. M. Hoffman, J Am Chem Soc 2001, 123, 1403.
[50] T. M. Makris, K. von Koenig, I. Schlichting, S. G. Sligar, J Inorg Biochem 2006, 100, 507.
[51] J. C. Schöneboom, S. Cohen, H. Lin, S. Shaik, W. Thiel, J Am Chem Soc 2004, 126, 4017.

Received: 25 May 2011
Revised: 28 July 2011
Accepted: 30 July 2011
Published online on 14 October 2011

CHAPTER 3

CONFORMATIONAL DYNAMICS OF CYP3A4 IMPLICATE THE IMPORTANT ROLE OF ARG212 COUPLED WITH THE OPENING OF INGRESS, EGRESS AND SOLVENT CHANNELS TO DEHYDROGENATION OF 4-HYDROXY-TAMOXIFEN

Abstract

Structural red-flags for P450 substrates are commonly used during drug development to identify sites of metabolism (SOM) that have the potential to produce high-energy toxic metabolites. P450-catalyzed dehydrogenation reactions involve the abstraction of two hydrogen atoms from two *different* SOM and can produce unstable reactive products. 4-hydroxy-tamoxifen (4OHT) is the active metabolite of tamoxifen, a commonly used anticancer drug, which can undergo competing hydroxylation and dehydrogenation reactions during its turnover by CYP3A4. *Ab initio* gas-phase electronic structural characterization of 4OHT identifies the sequence of hydrogen atom abstraction for dehydrogenation to an electrophilic quinone methide. To provide representative conformational sampling of CYP3A4, molecular dynamics (MD) simulations were applied along

multiple trajectories of the existing structure (PDB ID: 1TQN) to generate multiple templates of CYP3A4. Docking with both x-ray and MD-refined CYP3A4 structures incorporating QM-based penta- and hexa-coordinate heme parameters identifies binding modes that are supportive of the observed metabolism of 4OHT by CYP3A4, including dehydrogenation. Our models indicate that coupled structural contributions of the ingress, egress and solvent channels to the CYP3A4 active site geometries play key roles in the observed 4OHT binding modes. The models supporting dehydrogenation are experimentally validated and consistent with results from *in vitro* incubations of 4OHT with wild type CYP3A4 and the site-directed point mutant CYP3A4 R212A.

Introduction

Cytochrome P450 enzymes (P450) metabolize a diverse set of substrates via different oxidation reaction mechanisms[1,2]. As more and more experimental P450 structures with and without substrates become available[3-5], the emerging consensus is that the extent of promiscuity in P450 substrate selectivity has a structural basis which relates to both the varied enzyme active site architectures and their plasticity[6]. These characteristics vary greatly between the different subfamilies of hepatic P450[7-9]. While hepatic P450 promiscuity in substrate-selectivity favors the elimination of numerous endogenous and exogenous substrates whose accumulation can prove toxic to the organism[10], it presents a significant challenge to structure-based drug design. This promiscuity makes it

difficult to *a priori* predict if a putative lead compound will be adversely metabolized by P450s[11,12].

The characterization of novel and reactive Phase I drug metabolites, such as those produced by P450, is an area of increasing research efforts during preclinical drug-testing and development to help identify reactive metabolites[13-15]. Of key interest is improving our understanding of factors that contribute to competing Phase I reaction mechanisms, some of which produce stable products that can be further metabolized and/or excreted and others that produce reactive metabolites capable of inducing toxicities[16]. Due to the high-energy nature of the catalytic oxyferryl species, Compound I[17], P450s can catalyze a variety of different reaction mechanisms, such as dehydrogenation reactions, that can produce desaturated electrophilic metabolites capable of forming potentially toxic biomolecular adducts[18,19]. Although a number of tools exist to predict P450 metabolism, they demonstrate a mixed level of success depending on the class of substrate, P450 and reaction mechanism. Since P450 can catalyze a number of different oxidation reactions, including competing reaction mechanisms at the same site of metabolism (SOM), simply identifying the major metabolite is not adequate for drug safety screening.

Computational tools for predicting the site and mechanism of P450-metabolism are commonly used to provide structural alerts for a given set of lead compounds[11], the goal of which is to highlight and identify the potential for the generation of toxic metabolites early in the drug-design process. Ultimately, eliminating potentially reactive lead compounds earlier may reduce drug

development costs and timelines[13,14,20]. However, due to the diversity of the active site architecture and plasticity of the different subfamilies of hepatic P450, structure-based computational methods face a serious challenge in accurately predicting the SOM and the P450-catalyzed reaction mechanism (PRM), especially for CYP3A4 which is the hepatic P450 responsible for the majority of Phase I drug metabolism. Experimental data have demonstrated that CYP3A4 can accommodate a wide variety of structurally diverse ligands, substrates, and cofactors[8,21,22]. The currently available x-ray structures of CYP3A4 (PDB ID: 1W0E[23], 1W0F[23], 1W0G[23], 1TQN[24], 2J0D[8], 2V0M[8] and 3NXU[25]) support that this enzyme has a large plastic active site and suggests that the conformational dynamics of the CYP3A4 active site may have a direct impact on drug metabolism by influencing both the substrate selectivity and the specific reaction mechanism.

Our group has focused on dehydrogenation reactions due to their potential to metabolize substrates to reactive metabolites. Dehydrogenation reactions are more difficult to predict than standard P450 reactions and moreover, they can compete with hydroxylation mechanisms. Predicting them is made even more complicated since dehydrogenation may involve the abstraction of two hydrogen atoms from two *different* SOM, in contrast to the conventional ‘text-book’ rebound mechanism for hydroxylation which only involves one SOM[26]. Although dehydrogenation reactions are not as well studied as other P450-catalyzed reactions, mechanistic studies reveal the requirement for the formation of a stable cationic species and suggest that the enzyme active-site - substrate

interactions impose steric hindrance within the P450 active site so as to inhibit the rebound process to form a hydroxylated product[27]. To our knowledge, currently tools do not yet exist for predicting the susceptibility of substrates to be dehydrogenated and to form electrophilic products.

In previous work, we showed that molecular modeling combined with docking methods were able to assess enzyme-substrate interactions important in the P450-metabolism of raloxifene (RALX) by CYP3A4[28]. It was shown that the reliable prediction of enzyme-substrate interactions was greatly improved by the assignment of an initial set of chemically reasonable atomic partial charges for the resting high-spin ferric state of the heme. These theoretical studies correctly identified the SOM and were able to guide us to experimentally confirm the role of Phe215 in determining the selectivity of dehydrogenation versus oxygenation metabolic pathways. More recently, we have improved our heme models through the development of consistent and AMBER-compatible heme parameters for a number of heme states during the P450 catalytic cycle, including several penta-coordinate and hexa-coordinate states with dioxygen and the catalytically activated oxygen atom bound, Compound I[29]. We validated these molecular mechanics parameters for the modeling of P450-substrate interactions with raloxifene and CYP3A4[28]. With the new heme parameters, we were able to generate models that further improved our prediction of the SOM through a structure-based modeling approach.

Building on the previous work, here we investigate the P450 metabolism of 4-hydroxy-tamoxifen (4OHT) which, similar to raloxifene, undergoes both

hydroxylation and dehydrogenation reactions[30-32]. 4OHT is a primary metabolite of tamoxifen (TAMX), which is a FDA-approved selective estrogen receptor modulator (SERM) used extensively in the treatment of breast cancer [33] (Figure 3.1). 4OHT is a useful molecular probe since it is a pharmacologically active drug metabolite that can undergo P450-oxygenation and dehydrogenation to form reactive products, which is of concern in the MIST guidelines. Moreover, its metabolism is well understood and shown to be dependent on differences in the P450-subclass active sites[30]. Also, 4OHT provides an additional challenge for improving P450 models since unlike raloxifene, initiation of dehydrogenation reactions competes with hydroxylation reactions at the same SOM. To extend upon our previous approach, we use extensive quantum mechanics calculations to probe substrate reactivity for dehydrogenation, and molecular dynamics simulations to find representative enzyme active-site conformations (of the 1TQN PDB structure). With the incorporation of the recently published heme parameters[29], these CYP3A4 MD –refined structures provide models for the docking experiments that incorporate thermodynamic fluctuations of enzyme structure and changes in the electronic state of the heme.

We investigate if gas-phase quantum calculations at the HF/6-31G* and B3LYP/6-31G* level of the substrate structure can identify the dehydrogenation SOM and sequence of hydrogen abstraction. In addition to the gas-phase analysis, since experimental data indicate that P450 structure and dynamics may play a key role in metabolism, we have applied molecular dynamics (MD) methods to full P450 models in an attempt to better model enzyme-substrate

interactions. We used clustering analysis to produce representative structures from multiple MD trajectories, modified these MD-refined CYP3A4 structures with new heme parameters for a number of different steps of the P450 catalytic cycle and performed docking with the QM-optimized 4OHT structures. Our results indicate that including representative conformational sampling of enzyme and substrate configurations along with the use of penta- and hexa-coordinate QM-based heme parameters for docking leads to an improved signal-to-noise for identifying the binding modes that are supportive of the entire observed metabolism of 4OHT by CYP3A4. These models are experimentally validated in line with our previous work with *in vitro* incubations of 4OHT with wild type CYP3A4 and the site-directed point mutant CYP3A4 R212A.

Materials and Methods

Computational

All calculations were performed on the SGI Pople machine at the Pittsburgh Supercomputing Center and at the University of Utah's Center for High Performance Computing. The calculations involved *ab initio* QM, homology modeling, current state-of-the-art MD simulations with current force fields and explicit solvent, and molecular docking.

Substrate optimizations. All QM calculations were performed using Gaussian 03 software[34] in the gas-phase. All structures were built, calculations set up, and results analyzed using Gaussview[35]. Initial conformations of 4OHT were intuitively organized to arrange the central ring system in one of two

propeller conformations while the other functional groups were arranged to sample the conformational space as completely as possible. These starting structures were all optimized at the HF/6-31G* level of theory. Overlay of optimized structures using UCSF Chimera was used to identify redundant structures which were not used for further refinement. All optimized structures were also subjected to frequency analysis at the same level of theory as the optimization. All species with an amine moiety had optimizations for multiple minima performed with both 0 and +1 formal charges.

Dihedral scans. Scans of all ψ angles defined in Figure 3.2 were performed at the HF/6-31G* level of theory using one of the two lowest energy conformations of 4OHT, conformation A with a neutral charge. For ψ_{ETHYL}, additional scans were also performed at the B3LYP/6-31G* and MP2/6-31G* levels of theory.

Substrate bond stretching scans. Since relaxed scans produced non-physical re-association of the abstracted hydrogen atom with some other portion of the molecule, rigid scans of the three bonds involved in the dehydrogenation of 4OHT to a quinone methide (Figure 3.1) were performed for one conformation each intermediate. Only one conformation was used for these scans and the H9a and H9b naming is arbitrary (Figure 3.2). Rigid scans of just one conformation may have produced some conformation-dependent differences in the energies of H9a and H9b; hence, we used conformations based on the parent's A conformation that were identified, as described above, for each intermediate along the dehydrogenation pathway. HF/6-31G*//uB3LYP/6-31G* optimized

conformations of dehydrogenation reaction coordinate intermediates at stable minima, as determined via frequency analysis, were then subjected to a rigid scan of hydrogen CH bonds involved in the dehydrogenation of TAMX metabolites in 40 x 0.1 Å steps. Initially, relaxed scans were performed, but these were computationally much more intensive and produced spurious results in which the hydrogen atom recombined with another part of the molecule. Certain bonds that could not be successfully scanned as a stretch from the optimized coordinate were performed starting +4.0 Å to the optimized CH bond length of interest followed by 40 x (-)0.1 Å step scans to the return to the optimized structure geometry. The UB3LYP/6-31G*scans were performed using the SCF=QC method, NOSYMM constraints and the MAXCYCLE limit extended to 1000.

Substrate charge fitting. Restrained electrostatic potential (RESP) charges[36] were derived using RED-III[37] for the optimized geometries and molecular electrostatic potential (MEP) computations were performed on multiple orientations of four conformations for 4OHT in both charge states, based on atoms C9, O22 and N14 (see Figure 3.2 for atomic numbering).

Initial P450 loop refinement. For the x-ray structures, (PDB ID: 1TQN[24] and 1W0E[23]) used for docking and molecular dynamics simulations, missing residues in undefined loop regions were identified from the PDB file and added with UCSF MODELLER9.9[38]. For each missing loop, five models were built with the use of the 'automodel' class at the loop refinement level, and 'refine.fast' with the 'env.io.hetatm' set to 'True' to include the heme HETATM from the pdb file.

The structure with the lowest Modeller Objective Function value was selected for further refinement.

P450 molecular dynamics. MD was performed as described previously[29] except the current simulations were performed with our preliminary heme parameters[28]. Five independent MD simulations were performed for ~40 ns each with a different starting structure of the loop-refined (from the independent MODELLER loop refinements of) PDB ID: 1TQN[24] due to its higher resolution. The models were built with the ff99SB force field[39] with histidine (δ, ε or doubly protonated) states chosen by visualization (residue 27 HID, other histidines HIE). Heme parameters (shown in the supporting information) were applied and the cysteine-to-iron bond was manually added in tLEAP (residue 415 atom SG to residue 473 Fe). Charge was neutralized with chloride ions[40] and an additional 70 Na+ and Cl- ions were added to bring the salt concentration to ~200 mM. The models were solvated into a truncated octahedron unit cell with TIP3P water[41] extending at least 12 Å from any of the protein atoms and the positions of the ions randomized by randomly swapping with water molecules such that no ions were within 4 Å of each other and within 6 Å of the protein. All simulations employed AMBER 10.0 or 11.0 with a 9 Å cutoff with an additional 1 Å pairlist buffer, automated pairlist builds, the particle mesh Ewald method with default settings and a homogeneous long-range correction for the van der Waals interactions. MD simulations applied SHAKE[42] on all hydrogens and were integrated with a 2 fs time step. The models were equilibrated with 500 steps each of steepest descent and conjugate gradient minimization with 25.0 kcal/mol-

Å^2 restraints on the protein Cα atoms. Heating with constant volume MD followed for 50 ps starting at 100K ramped to 300K over 5000 steps with 25.0 kcal/mol-Å^2 restraints on the Cα atoms. Cycles of 1000 step minimization and 50 ps Berendsen-coupled[43] constant pressure and temperature MD (0.2 coupling times, 300K) followed as before with Cα positional restraints with gradually reduced force constants at each step of the cycle of 5.0, 4.0, 3.0, 2.0 and 1.0 kcal/mol-Å^2. Finally, a 500 ps constant pressure MD simulation was performed with 0.5 kcal/mol-Å^2 positional restraints on the protein Cα atoms, followed by production MD with the pressure/temperature coupling times increased to 5.0. Each trajectory was determined to have reached steady state with 2D RMSD plots. Visual inspection of the 2D RMSD plots was also used to determine the most reasonable number of clusters to search for. Analysis was performed with the *ptraj* module of AMBER11, each trajectory was searched for 4 clusters using the means clustering algorithm and an average representative structure was generated for each cluster in PDB format. From each trajectory, one representative structure was selected that represented the largest cluster and these were labeled m1-m5.

QM-heme parameters. All P450 models were prepared as described previously[29] with the following modifications; all structures were placed into a common frame of reference with the Matchmaker utility in UCSF Chimera tools[44] and saved. The existing heme coordinates were modified with QM-derived coordinates and force field parameters for the resting high-spin and Compound I[29] using LEAP in AMBER tools, and an energy minimization (imin = 1) was

performed for 1000 steps (maxcyc = 1000) with electrostatics turned off (type = 'ELEC', value1 = 0.0) and a cut-off of 9.0 Å (cut = 9.0) in implicit solvent was performed to relieve any steric clashes. Overlap check was performed with the check overlap utility in *ptraj*, and from this final structure a PDB format coordinate file was then generated and prepared for docking with the addition of Autodock3.0 parameters and RESP charges to the heme.

Docking. Four QM energy minimized structures for 4OHT with REDIII-derived RESP charges calculated using multiple conformation and reorientations for the parent molecule were used with each P450 model. The formal charge state was denoted as **foht-** as a neutral and **fohtq-** with a formal +1 charge due to protonation of the amine moiety. Then, 256 independent dockings were performed for each of the four conformations, and thus, 1024 dockings for each charge state of 4OHT with each P450 model were performed.

The refinement of the enzyme coordinates involved molecular dynamics simulations of the PDB ID: 1TQN[24] structure with five independent trajectories, and using clustering analysis of all of the structures generated to produce an average structure representative of the largest cluster of each trajectory (m1-m5). All of the MD-refined structures and the x-ray-based PDB ID: 1W0E[23] and 1TQN[24] were then modified with QM-based AMBER compatible heme parameters[29] and energy minimized to remove any steric clashes. These structures were then used with (***-wq**) and without (***-00**) assignment of atomic partial charges and scored using a distance-based scheme from the *iron* to heavy atom at a known SOM. The iron atom was used rather than the catalytic

oxygen for both penta- and hexa-coordinate states of the heme so as to differentiate the contributions of charges on the heme, sterics due to the presence of the catalytic oxygen for the hexa-coordinate Compound I versus the penta-coordinate ferric and the contributions of changes in the active site architecture to the docking results.

Docking was performed with AutoDock3[45] and Autodock tools[46] as previously published[28,29] with the following modifications. The grid box was recentered at -21.545, -20.757 and -11.334 displacement from the default center assigned by Autogrid3.05 with the dimensions of 68 x 72 x 82 and a resolution of 0.375 Å in each dimension. Consistent with our previous work, for each of the four minima of the substrate 4OHT, 256 independent genetic algorithm runs were performed producing a total number of 1024 docked conformations for each substrate and heme charge state. This was performed for the two x-ray-based (PDB ID: 1TQN and 1W0E) and the 5 MD-refined (m1-m5) P450 models, and for each P450 model docking studies were performed with charges assigned or not assigned to the heme for each catalytic cycle state. The docking results were analyzed using Autodock tools[46]. The results from each docking were clustered consistent with our previous work[28]. Autodock tools[46] was used to perform conformational clustering of Autodock3[45] results with the 2 Å RMSD cutoff for conformational similarity of each bin and used the following site of metabolism scoring criteria based on the following distances:

α-OH: Fe-C11 < 8 Å and Fe-R2 > 8 Å

Dehydrogenation: both Fe-C11 and Fe-R2 < 8 Å

N-demethylation: Fe-C27 or C28 < 8 Å.

For each repetition of each substrate in each P450 state, the total number of poses in every cluster above the 5% cut-off that represented a particular reaction mechanism was pooled. A number of ambiguous or nonproductive (**Non**) binding modes were also identified and these were all pooled together.

For each cluster that had more than 52 representative states (>5% of 1024 total), the lowest energy conformation was saved in PDB format and further analyzed with UCSF CHIMERA[44].

Statistics. Statistical analyses of each reaction mechanism for each substrate-P450 state involved a one-way ANOVA and a post-hoc test using the Tukey test. Analyses were performed using GraphPad Prism version 5.00 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com.

Channel mapping. CAVER[47] was applied to identify the channels present within the different P450 structures via the CAVER add-in to PYMOL[48]. The center selection of the P450 was based on the default setting and all structures were searched for 10 channels. In each of the runs, only 2-4 distinct channels were observed with the remaining channels being redundant to the initial set of 2-4 channels. Nomenclature was assigned consistent with Cojocaru et al. 2007[49].

Active-site mapping. Active sites were mapped with UCSF HOLLOW[50] using a cylinder type of search with a radius 6.5 Å. The axis for each channel was found for each structure with CAVER was defined from the activated oxygen of the hexa-coordinate heme (O1) to the Cα of the most appropriate residue

deemed by visual inspection. The offset to the Cα was also set at 0.0, -1.0 or -2.0 Å based again on visual inspection.

Molecular graphics. Images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081)[44] and the PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC as noted below.

Experimental

Materials: 4OHT, NADPH, silver oxide, glutathione (GSH), N-acetylcysteine (NAC) and raloxifene (internal standard) were obtained from Sigma-Aldrich (St. Louis, MO). All other chemicals were of analytical grade and obtained at the highest quality commercially available.

Synthesis of standards of dehydrogenated 4OHT product adducts. 4OHT was oxidized using 500 mg silver oxide in 2 ml anhydrous ACN with constant stirring at 30°C. Production of 4-hydroxy-tamoxifen quinone methide (4OHT-qm) was monitored by UV-vis absorbance at 280 nm. It was determined that 4OHT-qm had reached a steady-state maximum after ~8 min. Reactions were terminated by vacuum filtration of the mixture to remove excess silver oxide and purified by flash chromatography, during which the organic solution was added drop-wise to ½ volume aqueous phase (100 mM KPi, pH 7.4 with 25 mM NAC) to trap the dehydrogenated quinone methides as NAC adducts. These solutions were diluted to <5% organic phase and further purified using the Waters Oasis

WCX Cation Exchange SPE column. The eluted products were concentrated under $N_{2,}$ and reconstituted to a final volume 50 µL MeOH prior to analysis by LC/MS.

Incubations with 4-hydroxy-tamoxifen. The CYP3A4 wild type and R212A mutants were expressed, purified and kinetics characterized as described previously[51]. Incubations were performed as described previously[51], except 20 µM 4-hydroxytamoxifen was used as a substrate, and 5 mM NAC was used instead of 4 mM GSH. NADPH (5mM) was used to start the reaction. Reactions were allowed to proceed for 12 min at 37ºC and terminated with addition of 200 µL ice-cold acetonitrile with IS added (RALX 5 µM). The mixtures were then vortexed and centrifuged at 21,000xg for 5 min. Excess KPi buffer was added to dilute the organic phase to <5%, and products were purified using a Waters Oasis WCX SPE cartridge. The MeOH eluate was then concentrated to dryness under vacuum with a Savant SVC100 Speed Vac and reconstituted in 50µL MeOH for LC-MS analysis.

LC-MS analysis of 4OHT metabolites. LC-MS analysis of 4OHT metabolites was performed as described previously with the following[28,52] modifications: source temperature 385ºC and capillary voltage 10V. Chromatography was performed with Phenomenex Jupiter 5µ C4 (150mm x 2.00mm) reverse-phase column using a mobile phase consisting of Solvent A: acetonitrile and Solvent B: 20% MeOH, 0.1% formic acid (v/v/v). The analytical gradient was linear from 40% to 95% Solvent A over 15 min at a flow rate of 0.2 ml/min.

4-OHT and its hydroxylated and dehydrogenated metabolites were identified as described previously[53]. Briefly, 4-OHT was identified as the molecular ion at *m/z* 388.5, with an elution time of 5.6 min. MS^2 of this molecular ion with activation energy of 34.0% and an isolation width of 3 amu produced a major daughter ion at *m/z* 331. Hydroxylated 4OHT was identified as the molecular ion at *m/z* 404.5, with an elution time of 4.1 min. MS^2 of the *m/z* 404.5 with an activation energy of 42.0% and an isolation width of 3 amu produced a major daughter ion at *m/z* 386.0 representing the loss of a water. The dehydrogenation product 4OHT + NAC was identified as the molecular ion at *m/z* 549.0, with an elution time of 4.3 min. MS^2 of *m/z* 549.0 with activation energy of 20.0% and an isolation width of 5 amu, yielded a major daughter ion with *m/z* 386.0 representing the loss of NAC. MS^3 fragmentation of the 386.0 daughter ion with 42% activation energy and isolation width of 3 amu of this molecular ion yielded the major diagnostic ion with *m/z* 341.1, the diagnostic ion of 4OHT. Raloxifene was identified as *m/z* of 474.5 with an elution time at 4.4 min.

Results

Substrate geometry optimizations

The structure activity relationships for tamoxifen have been investigated previously in an effort to design the next generation of selective estrogen receptor modulators[54-59]. These studies illustrate the complex interplay of interactions with various estrogen receptors and other targets showing how small differences in the chemical structure can alter the pharmacological activity[60].

Early theoretical investigations at lower levels of theory had also provided insights into the complex conformational landscape produced by the central alkene covalently bound to three aromatic rings, which could occupy multiple minima[61,62]. However, recent investigations at HF and B3LYP levels of theory revealed one minima existed which was in agreement with experimental x-ray data[63]. Energy minimizations at the HF/6-31G* level – the conventional level of theory applied for molecular mechanical force field parameterization compatible with the nonpolarizable AMBER-based pairwise additive ff9X force fields that produce a slight and beneficial overpolarization - identified multiple minima within ~2 kcal/mol (Table 3.1 and Figure 3.3). The major difference differences in the optimized geometries at the HF/6-31G* level involved two sites: 1) the orientation of the aromatic rings and ethyl group attached via a central double bond from C7 to C8, which compose the central ring system; and 2) the oxy-methyl-amine (MTOXY) 'tail' composed of atoms O11, C12, C13, N14, C15 and C16 (refer to Figure 3.2 for nomenclature), which adopted one of two conformations (Table 3.1). The differences with the ring system (dihedral angle $\psi_{1\text{-}3}$) relate to different 'propellering' in one of two directions, and the ethyl group is oriented with C10 above or below the plane defined by the central carbons (C1, C7, C8, C9, C17 and C23).

Optimizations at the B3LYP/6-31G* level did not show the same difference in orientation of the MTOXY tail. Scans of ψ_{MTOXY} at HF/6-31G* and B3LYP/6-31G* level indicate that the minima observed at the HF level at ~87º is not observed at the B3LYP level. The B3LYP/6-31G* structures shared all the other

trends and showed similar energy differences. Since structures derived at the B3LYP-level of theory have been shown to be in good agreement with experimental structures[63], the DFT-based conformations were used for further modeling. The differences between these four geometries are shown qualitatively in Figure 3.3. Of the four conformations, denoted A-D, A and C share similar orientations of the ring systems as do B and D. The main differences between these two pairs relate to the orientation of the ethyl group at C9 and C10.

Mulliken population analysis of atomic spin densities for 4OHT open-shell dehydrogenation reaction intermediates

In an attempt to identify sites of metabolism and hydrogen abstraction, atomic spin densities of open shell species were calculated to assess the relative probabilities of unpaired electrons at specific sites on 4OHT. In addition, gas-phase *ab initio* quantum mechanics calculations probed the sequence of hydrogen abstraction involved in dehydrogenation of 4OHT. Previous gas-phase quantum mechanics based studies of the desaturation of a radical clock substrate and a truncated thiolate model of the heme/Compound I concluded that "*the oxidase-dehydrogenase mixed activity occurs from the cationic intermediate species and requires electro-steric inhibition of the rebound process*"[27]. Based on the hypothesis that the mixed activity occurs with the cationic intermediate species, we evaluated the applicability of this approach for identifying SOM of dehydrogenation reactions that involve two distinct SOM, specifically C9 and O29 for 4OHT, as opposed to other P450-catalyzed reactions that commonly involve

one hydroxylation or neighboring atomic centers such as epoxidation [19]. Also, since the major CYP3A4 metabolites of 4OHT have been identified experimentally as α-hydroxylation at C9, and N-demethylation at C15 and C16, and the major CYP2B6 4OHT hydroxylated metabolite at C26[30], we also evaluated atomic spin densities at these sites.

Gas-phase reaction coordinates for 4OHT dehydrogenation are shown in Figure 3.4 and for all four conformations of the first cationic species formed, we examined the Mulliken atomic spin densities to see if this approach would identify the site of the first hydrogen abstraction. The highest spin densities were observed at C8, which is not a commonly observed P450 SOM, and C26, the observed SOM for 4'-hydroxylation by both CYP3A4 (minor metabolite) and CYP2B6 (major metabolite)[30]. However, spin densities did not predict metabolism at C9 or at any of the carbon atoms bonded to N14, the SOM for the major metabolites produced by CYP3A4 from 4OHT. Mulliken atomic spin densities for all three intermediate neutral radical species were also examined for the single lowest energy conformation (A), to determine if these might identify the site of the second hydrogen atom abstraction in the dehydrogenation of 4OHT. The Mulliken atomic spin densities remained high at the site of hydrogen atom abstraction for each species, and at C7 for the radicals formed from aliphathic hydrogen abstraction. Thus, the Mulliken atomic spin densities from gas-phase QM calculations at the 6-31G* level of theory for the open-shell species formed along the putative dehydrogenation reaction coordinate do not identify the SOM for the CYP3A4 metabolism of 4OHT, except for the minor product 4'-hydroxy-

tamoxifen. *A priori* knowledge is therefore required to identify the two dehydrogenation SOM of 4OHT by CYP3A4.

Investigation of the sequence of hydrogen atom abstraction

Since the SOM for dehydrogenation with gas-phase calculations of the substrate alone were not useful, *a priori* knowledge of the SOM resulting in the formation of the desaturated quinone methide species identified experimentally[31] was used for further analysis. We selected conformation (A), one of the two lowest energy conformations, to perform analysis of the energy differences of all of the stable species formed along each putative reaction pathway for hydrogen abstraction (Figure 3.4). For this study, we utilized two distinct approaches to determine the order of hydrogen abstraction. Since it had been previously proposed that the formation of a stable cation was likely a key factor required to undergo dehydrogenation, approach 1 investigated the stability, relative energies and electron affinities of the putative intermediates determined to be involved in the formation of 4OHT quinone methide[31], while approach 2 calculated the bond dissociation energies (BDE) of these bonds.

Approach 1: Stability, relative energies and electron affinities of intermediates along putative l pathways of hydrogen atom abstraction for the formation of quinone methides from 4OHT. We performed geometry optimizations and frequency analyses for all putative gas-phase intermediates that could be formed from abstraction of the three hydrogen atoms putatively involved along the dehydrogenation reaction coordinate (Figure 3.4). All species

were determined to be at stable minima due to the absence of negative frequencies. The cation 1 intermediate is an open-shell doublet with a +1 charge formed from the parent. The subsequent radical intermediates can be formed from the abstraction of H29a, H9a or H9b, resulting in neutral open-shell doublets. The second cation 2 intermediate can be formed from any of the putative neutral radical intermediate species, has a +1 charge and is a singlet. The parent and product species are neutral closed shell systems. A summary of the difference in the sum of electronic and the thermal free energies relative to the lowest energy equivalent intermediate is given for each species with the corresponding electron affinity (Figure 3.4).

The largest observed energy difference was for the formation of the second cationic species from the abstraction of H29a which was ~30 kcal/mol higher in energy than either of the second cationic species formed from the initial abstraction of either allylic hydrogen atom (H9a or H9b). This species also has a greater electron affinity than either of the two cation 2 equivalents or of cation 1. These data suggest that dehydrogenation would be least favored to proceed via an initial hydrogen abstraction from the hydroxyl group based on the energies of the intermediates along the putative dehydrogenation pathways involving C9 and O29 explored here.

Approach 2: Bond dissociation energies (BDE). The BDE for the parent and the first cationic species formed show that, for conformation A, the H9a abstraction was the lowest energy (Figure 3.5). The BDE for the three neutral radical intermediate species and the second cationic intermediate (Cation 2)

indicate that the hydroxyl bond (H29a) is the lowest in energy for the second hydrogen atom abstraction. These results together suggest that the dehydrogenation of 4OHT likely proceeds with the abstraction of the allylic H9a followed by the abstraction of the hydroxyl H29a.

This gas-phase method provides a physical basis to more quantitatively score docking results to accurately identify the site and order of metabolism. These data suggest that 4OHT likely undergoes sequential abstraction of hydrogen atoms beginning with the aliphatic hydrogen. This underscores the importance of sampling and identification of all of the minima available at physiological temperatures. Furthermore, the QM-based reaction coordinate pathway energetic analysis and BDE provide a physical basis to support the postulated dehydrogenation order of 4OHT to be used to score our docking results.

Docking

Molecular docking was performed to investigate putative binding modes of the substrates in the active site of CYP3A4. To provide a more realistic sampling of active site geometries beyond that shown in the available crystal structures, molecular dynamics simulations were employed. The use of ensembles of representative configurations of enzyme-substrate structures sampled with modern MD simulation protocols when used with computationally affordable tools such as docking holds great promise for structure-based prediction metabolism of new chemical entities[64,65]. In spite of sampling

limitations and force field inaccuracies[66], MD simulations can fold small proteins[67] and accurately model protein-ligand interactions using modern simulation protocols and balanced force field treatments[68].

To confirm that the energy barriers between the minima were low enough for the minima to be thermally accessible at physiological temperature, scans of all the ψ* dihedral angles as defined in Figure 3.1 were performed. The results of these scans indicated that the energy barriers would indeed allow all of these minima to be occupied at physiological temperatures. Therefore, docking was performed with all four representative structures, and the results from all four dockings were combined.

Single docking runs only provide a “snap-shot” of enzyme- substrate interactions. However, this allows docking programs (e.g. Autodock3) that treat the coordinates for the biomolecule as static and use a united atom treatment to provide for fast sampling of the search space. However, identifying one pose out of many by visually inspecting the docking results can be time consuming, and subject to issues of signal-to-noise, which may not provide statistically significant results. A commonly applied metric for identifying potential SOM after docking is the use of distance-based scoring schemes, such as sites within 5 Å of the heme oxygen[12,20,51,69]. Autodock3[45] uses a united atom treatment and a static protein model. Moreover, since our heme models now include the catalytic oxygen atom in the hexa-coordinate heme state, the lowest energy docked structure within a family of similar poses may not necessarily correspond to the pose most amenable to catalysis. Therefore, we loosened the distance criteria for scoring

the SOM for the lowest energy conformation representative of each cluster of Autodock3 results. Specifically, a cutoff < 8 Å from the iron atom to the heavy atom representing the SOM was used for metabolism involving a single SOM, e.g. hydroxylation and N-demethylation, and for dehydrogenation, one of the two iron-to-SOM distances should be < 5.5 Å and the iron to the second SOM distance should be < 8 Å. Visual inspection of productive clusters showed that within each cluster/bin, structures could be found where the corresponding iron-SOM distance was < 5 Å.

We improved the sampling by performing multiple independent docking runs so as to allow for statistical analysis of the results. We also imposed an arbitrary signal-to-noise cutoff by only evaluating clusters which had >5% of the total population. In most cases, this reduced the number of clusters analyzed per 1024 docking runs to <10 clusters. All of the docking studies were repeated three times, independently, and analyzed for statistically significant differences for each reaction mechanism.

Preliminary docking studies using x-ray structures of CYP3A4 (PDB ID: 1W0E[23] and 1TQN[24]) with a single MM2 energy-minimized structure of 4OHT with Gasteiger charges assigned, no charge assignment to the heme atoms, consistent with our earlier work with RALX, did not identify binding modes that supported the observed metabolism of 4OHT. Conversely, docking studies with multiple conformations of 4OHT with the PDB ID:1W0E-based structure modified with QM-based heme parameters (1w0e-ic6-00/wq and 1w0e-cpdi-00/wq) identified a number of binding modes supportive of α-hydroxylation (α-OH) and

N-demethylation (NdM) with the penta-coordinate heme (Figure 3.6). Docking with this same P450 structure with the hexa-coordinate heme parameters for Compound I identified a binding mode supportive of both α-hydroxylation and dehydrogenation (dH1). The association of this binding mode with a specific reaction mechanism is ambiguous since no specific residues capable of producing the required steric hindrance to inhibit the hydroxyl rebound mechanism are observed. Interestingly, the number of configurations supportive of NdM were increased with the presence of the catalytic oxygen *without* atomic partial charges. This result suggests that, with this enzyme model the steric effect from the presence of the catalytic oxygen is decreased with the assignment of atomic partial charges that causes a change in the orientation of the flexible MTOXY moiety. Also, the number of non-productive modes in the major clusters significantly decreased with the use of the hexa-coordinate heme parameters, indicating that the presence of the catalytic oxygen atom has a greater steric contribution for 4OHT with this x-ray model.

A similar approach with the PDB ID: 1TQN-based structure modified with QM-based heme parameters (1tqn-ic6-00/wq and 1tqn-cpdi-00/wq), docking studies identified mostly non-productive modes and 1 productive mode supportive of α-hydroxylation (Figure 3.6).

To incorporate flexibility and dynamics in the active site structures, five different MD trajectories of the starting 1TQN structure were also evaluated. To choose representative structures for docking, clustering of the structures sampled at 1 ps-intervals by RMSD similarity for each trajectory was performed.

The average structure from the largest cluster was chosen for docking in each of the five trajectories. To remove artifacts from averaging, the structure was first minimized with the electrostatic interactions turned off to remove steric clashes.

Docking studies with the MD-based structure m2 modified with QM-based heme parameters (m2-ic6-00/wq and m2-cpdi-00/wq) also predicted α-OH, NdM as well as an alternate binding mode supportive of dehydrogenation (dH2). This binding mode suggested R212 as a single residue that could potentially produce electro-steric inhibition of the hydroxyl-rebound mechanism leading to dehydrogenation. The number of poses in the dH2 binding mode also significantly increased with the inclusion of hexa-coordinate Compound I coordinates *and* atomic partial charges. For this structure, the number of non-productive binding modes in the largest cluster also significantly decreased with the incorporation of hexa-coordinate Compound I coordinates.

Analysis of QM-based heme parameters and substrate charge state in docking

Inclusion of the heme charge *and* coordinate parameters, and alternate charge state for the substrate did not produce a significant “across the board” improvement for modeling 4OHT-CYP3A4 interactions with 1W0E, 1TQN or the MD refined models. Furthermore, significant model-specific differences were observed in binding modes supporting different reaction mechanisms, except α-OH. To ensure that the substrate charge state did not alter the docking results, these studies were repeated with charged 4OHT. Similar results were obtained

with neutral, and +1 4OHT with respect to identifying SOM and various reaction mechanisms. Significant differences for the same model, with and without charges, were also observed for a few dockings, but no significant improvement based on the charge state of 4OHT or heme parameters was observed across all CYP3A4 models. However, it should be noted that inclusion of the hexa-coordinate state never eliminated a binding mode identifying the SOM for any structure tested. *In fact, inclusion of these coordinates was required to identify the dH1 and dH2 binding modes*.

4OHT binding modes and CYP3A4 conformational dynamics and active site architecture

For CYP3A4 models that produced binding modes supportive of dehydrogenation, two modes were observed, labeled dH1 and dH2. The dH1 binding mode was similar in orientation to the binding mode that supported α-OH, except with 4OHT "flipped over", thus placing O29 in a position within hydrogen abstraction distance to the heme (Figure 3.7). However, with the dH1, no residues were observed to intervene between 4OHT and the catalytic center of the heme to provide the necessary steric inhibition of the rebound process. This makes interpretation of dehydrogenation versus hydroxylation ambiguous using this type of MM-based approach. In contrast, the dH2 binding mode clearly positioned Arg212 in an orientation that imposed steric and potentially electro-static inhibition of the rebound process. This was only observed with the m2 model. Further, the m2-based dH2 pose showed significant improvement in the

results predicting the involvement of the dehydrogenation with the inclusion of the QM-based heme charges.

Ingress, egress and solvent channels

Examination of the binding modes within the context of the active site of the MD-refined m2-cpdi and w0e-cpdi models show that for 4OHT, the opening and closing of the ingress and egress channel 4 provides a space for the substrate with respect to the catalytic site of the enzyme (Figure 3.7). In the m2 structure, channels 2b, 4 and S were determined to be open. The dH2 binding mode is only possible with the opening of channel 4, in such a manner so as to allow 4OHT to partially occupy it. The other major reaction mechanisms (i.e. α-OH, dH1 and NdM), are supported by binding modes that allow 4OHT to occupy a region of the active-site directly above the heme, that among the models tested, appears to be less malleable.

Table 3.2 shows the channels that were identified for each model and the reaction mechanisms determined to be supported by our scoring scheme. All of the channels present in the MD-refined structures are present in published CYP3A4 x-ray structures, but in different combinations, including channel 4, which is observed in another x-ray structure.

By comparing structure specific channels and binding modes (Table 3.2), it can be clearly seen that there is no one-to-one relation between channel opening and closing and substrate binding. For structures that do have channel 4 open, specifically m2, m3, m5 and PDB ID: 2V0M, the orientation and structure

of channel 4 was examined (Figure 3.8). For all four structures different volumes and paths defined for channel 4 were observed. The positioning of Arg212 toward the heme is similar for m2, m3 and m5, but for 2V0M[8], the Arg212 is oriented away from the heme (Figure 3.8). This would suggest that the use of this model would not have implicated the potential role of this residue in dehydrogenation. Therefore, both the channel 4 open state and Arg212 oriented toward the heme are required for a binding mode that supports dehydrogenation of 4OHT.

Analysis of the representative structures used here identifies that different regions of the CYP3A4 active site are perturbed by the open or closed state of ingress and egress channels to different degrees. These more malleable regions appear to have multiple states available to them. Thus, we recommend future docking studies of CYP3A4 with larger substrates, such as 4OHT, to include multiple structures with multiple combinations of open channels. In fact, using a number of CYP3A4 structures with different combinations of open channels may prove to be an important metric of achieving adequate sampling for identifying modes supportive of different SOM and P450-catalyzed reaction mechanisms.

Incubations and liquid-chromatography mass-spectroscopy analysis (LC-MS)

LC-MS analysis of incubations with CYP3A4 and the CYP3A4-R212A mutant showed a dramatically altered rate of dehydrogenation versus oxygenation, adding support to the results from the modeling studies. LC-MS semiquantitative analysis of incubations of 4OHT with CYP3A4 wild-type and

CYP3A4-R212A produced both hydroxylated and dehydrogenated metabolites. However, CYP3A4-R212A produced significantly less dehydrogenated product ($p=0.016$), while the amount of hydroxylated metabolite was essentially unalterered ($p=0.677$) (Figure 3.9).

Conclusions

The dH2 binding mode correctly implicated the potential role of Arg212 in the steric hindrance required to inhibit the rebound process of the dehydrogenation of 4OHT. Arg212 could possibly also be contributing some type of electrostatic repulsion of the cationic intermediate formed during the dehydrogenation reaction coordinate. Our docking results show that population of the dH2 binding pose is sensitive to both inclusion of atomic partial charges and Compound I coordinates. Thus, Arg212 may provide both steric and electrostatic hindrance of the hydroxyl rebound process. However, the possible contributions from the dH1 mode, with inhibition of the rebound process due to thermodynamic fluctuations of the active site architecture as a whole, cannot be ruled out. Nor can we rule out the possibility that other biologically relevant binding modes may exist that we have yet not identified.

Interestingly, the key residue involved in the dehydrogenation of 4OHT is different than that identified in previous work with RALX, suggesting that these two SERMs have diverse conformational landscapes in the CYP3A4 active-site; even though they are able to bind the ER in the same manner, their P450 metabolism requires very different binding modes.

QM-based energy minimizations identified multiple conformational minima for 4OHT that are likely thermally accessible at physiological temperatures. Therefore, performing MM-based studies requires sampling from all of these minima. The *ab initio* methods used here did not unambiguously identify the dehydrogenation SOM and *a priori* knowledge was required to do so. However, QM-based methods did consistently support the energetically favored sequence of hydrogen atom abstraction and provided a physical basis for the docking scoring scheme used here.

MD simulations of x-ray structure supportive of only one productive binding mode (PDB ID: 1TQN) produced models with ingress and egress channels observed to be present in other CYP3A4 experimental structures but in different combinations. These models, when used with computationally inexpensive docking tools, identified binding modes supportive of the entire metabolism of 4OHT by the most promiscuous and biomedically relevant hepatic P450, CYP3A4. One unique enzyme-substrate configuration was identified in a MD-based model that involved the potential interaction of a single residue, Arg212, and was supportive of the dehydrogenation reaction. Enzymatic incubations with a R212A mutant of CYP3A4 confirmed the role of this residue in the dehydrogenation of 4OHT. Use of these different models with QM-based heme parameters for the heme and charge states for 4OHT showed that for this hydrophobic substrate, the contribution of atomic partial charges was not as important as for RALX.

Comparison of these theoretical enzyme-substrate configurations have identified regions of the enzyme active site that differ in plasticity, and different binding modes supportive of the same reaction mechanisms. We postulate that the difficulties in identifying the SOM and PRM for CYP3A4 are related to the binding modes within the active-site. Depending on the size, rigidity and polarity of the substrate, multiple binding modes may be available supportive of the same PRM. This is further complicated by the fact that certain regions of the CYP3A4 active site appear to be more malleable than others. Thus, for some substrates, multiple binding modes may be available, and some of these may exist within these plastic regions which cannot be modeled using x-ray data alone. We speculate that these plastic regions, particularly those surrounding channel 4, which is formed with the long insertion in the F-G loop, unique to mammalian P450[5], may play an important part in the ability of a number of classes of P450 enzymes ability to accommodate chemically diverse substrates. Furthermore, the contribution of the cavities formed by the channels may play a more important role than that of individual residues for binding of more flexible and nonpolar substrates.

Based on our findings, we recommend that future computational CYP3A4 metabolism studies identify all the available minima for the substrate and verify that the interconversion between these minima is thermally possible so as to provide adequate sampling. We recommend that docking studies be performed with the penta- and hexa- coordinates, since inclusion of these parameters has been shown to help increase the observed number of productive binding-modes

for molecules with both rigid and flexible cores, albeit in different manners. Regarding models of the enzyme, it is recommended that in the refinement of existing experimental models, adequate sampling also be performed and considered when attempting to identify different states of the enzyme in regards to the opening and closing of channels. This is especially important for larger, more flexible substrate or ligand molecules that can occupy these volumes adjacent to the active-site. To identify binding modes supportive of the entire metabolism of CYP3A4 substrates, including minor reaction pathways such as dehydrogenation, that produce unstable and reactive metabolites, it is important to adequately sample the conformational space of the enzyme.

References

1. Furge, L. L.; Guengerich, F. P. Biochemistry and Molecular Biology Education 2006, 34(2), 66-74.

2. Ortiz de Montellano, P.; Voss, J.; Ortiz de Montellano, P. R., Ed.; Springer US, 2005, p 183-245.

3. Pochapsky, T. C.; Kazanis, S.; Dang, M. Antioxid Redox Signal 2010, 13(8), 1273-1296.

4. Halpert, J. R. Drug metabolism and disposition: the biological fate of chemicals 2011, 39(7), 1113-1121.

5. Otyepka, M.; Skopalík, J.; Anzenbacherová, E.; Anzenbacher, P. Biochimica et Biophysica Acta (BBA) - General Subjects 2007, 1770(3), 376-389.

6. Hendrychová, T.; Anzenbacherová, E.; Hudeček, J.; Skopalík, J.; Lange, R.; Hildebrandt, P.; Otyepka, M.; Anzenbacher, P. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 2011, 1814(1), 58-68.

7. DeVore, N. M.; Meneely, K. M.; Bart, A. G.; Stephens, E. S.; Battaile, K. P.; Scott, E. E. FEBS Journal 2011.

8. Ekroos, M.; Sjogren, T. Proceedings of the National Academy of Sciences of the United States of America 2006, 103(37), 13682-13687.

9. Skopalík, J.; Anzenbacher, P.; Otyepka, M. The Journal of Physical Chemistry B 2008, 112(27), 8165-8173.

10. Guengerich, F. P. Chemical research in toxicology 2007, 21(1), 70-83.

11. Afzelius, L.; Arnby, C. H.; Broo, A.; Carlsson, L.; Isaksson, C.; Jurva, U.; Kjellander, B.; Kolmodin, K.; Nilsson, K.; Raubacher, F.; Weidolf, L. Drug metabolism reviews 2007, 39(1), 61-86.

12. Sun, H.; Yost, G. S. Chemical research in toxicology 2008, 21(2), 374-385.

13. Smith, D. A.; Obach, R. S. Chemical research in toxicology 2009, 22(2), 267-279.

14. Baillie, T. A. Chemical research in toxicology 2007, 21(1), 129-137.

15. Baillie, T. A.; Cayen, M. N.; Fouda, H.; Gerson, R. J.; Green, J. D.; Grossman, S. J.; Klunk, L. J.; LeBlanc, B.; Perkins, D. G.; Shipley, L. A. Toxicology and Applied Pharmacology 2002, 182(3), 188-196.

16. Yost, G. S. Adv Exp Med Biol 2001, 500, 53-62.

17. Meunier, B.; de Visser, S. P.; Shaik, S. Chem Rev 2004, 104(9), 3947-3980.

18. Guengerich, F. P.; Macdonald, T. L. Accounts of Chemical Research 1984, 17(1), 9-16.

19. Ortiz de Montellano, P. R.; De Voss, J. J. Kluwer Academic/Plenum Publishers, New York 2005, 185-245.

20. Sun, H.; Scott, D. O. Chemical Biology & Drug Design 2010, 75(1), 3-17.

21. Davydov, D. R.; Davydova, N. Y.; Tsalkova, T. N.; Halpert, J. R. Archives of biochemistry and biophysics 2008, 471(2), 134-145.

22. Scott, E. E.; Halpert, J. R. Trends in biochemical sciences 2005, 30(1), 5-7.

23. Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Science (New York, NY 2004, 305(5684), 683-686.

24. Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The Journal of biological chemistry 2004, 279(37), 38091-38094.

25. Sevrioukova, I. F.; Poulos, T. L. Proceedings of the National Academy of Sciences of the United States of America 2010, 107(43), 18422-18427.

26. Ortiz de Montellano, P. R. Chemical reviews 2009, 110(2), 932-948.

27. Kumar, D.; De Visser, S. P.; Shaik, S. Journal of the American Chemical Society 2004, 126(16), 5072-5073.

28. Moore, C. D.; Shahrokh, K.; Sontum, S. F.; Cheatham, T. E., 3rd; Yost, G. S. Biochemistry 2010, 49(41), 9011-9019.

29. Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 3rd. Journal of computational chemistry 2011.

30. Desta, Z.; Ward, B. A.; Soukhova, N. V.; Flockhart, D. A. The Journal of pharmacology and experimental therapeutics 2004, 310(3), 1062-1075.

31. Fan, P. W.; Zhang, F.; Bolton, J. L. Chemical research in toxicology 2000, 13(1), 45-52.

32. Fan, P. W.; Bolton, J. L. Drug metabolism and disposition: the biological fate of chemicals 2001, 29(6), 891-896.

33. Smigel, K. Journal of the National Cancer Institute 1998, 90(9), 647-648.

34. Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.;

Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.;. Gaussian, Inc, Wallingford CT, 2004 2004.

35. Dennington, R.; Keith, T.; Millam, J. Semichem Inc, Shawnee Mission KS, 2009.

36. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. The Journal of Physical Chemistry 1993, 97(40), 10269-10280.

37. Dupradeau, F. Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. Phys Chem Chem Phys 2010, 12(28), 7821-7839.

38. Fiser, A.; Šali, A. In Methods in Enzymology; Charles W. Carter, Jr.; Robert, M. S., Eds.; Academic Press, 2003, p 461-491.

39. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Proteins: Structure, Function, and Bioinformatics 2006, 65(3), 712-725.

40. Smith, D. E.; Dang, L. X. The Journal of Chemical Physics 1994, 100(5), 3757-3766.

41. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. The Journal of Chemical Physics 1983, 79(2), 926-935.

42. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Journal of Computational Physics 1977, 23(3), 327-341.

43. Hermans, J.; Berendsen, H. J. C.; Van Gunsteren, W. F.; Postma, J. P. M. Biopolymers 1984, 23(8), 1513-1518.

44. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. Journal of computational chemistry 2004, 25(13), 1605-1612.

45. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Journal of computational chemistry 1998, 19(14), 1639-1662.

46. Sanner, M. F. J Mol Graphics Mod 1998, 17(February), 57-61.

47. Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J.

BMC Bioinformatics 2006, 7, 316.

48. DeLano, W. 2002.

49. Cojocaru, V.; Balali-Mood, K.; Sansom, M. S.; Wade, R. C. PLoS Comput Biol 2011, 7(8), e1002152.

50. Ho, B.; Gruswitz, F. BMC Structural Biology 2008, 8(1), 49.

51. Sun, H.; Moore, C.; Dansette, P. M.; Kumar, S.; Halpert, J. R.; Yost, G. S. Drug metabolism and disposition: the biological fate of chemicals 2009, 37(3), 672-684.

52. Yu, L.; Liu, H.; Li, W.; Zhang, F.; Luckie, C.; van Breemen, R. B.; Thatcher, G. R. J.; Bolton, J. L. Chemical research in toxicology 2004, 17(7), 879-888.

53. Fan, P. W.; Zhang, F.; Bolton, J. L. Chemical research in toxicology 1999, 13(1), 45-52.

54. Murphy, C. S.; Parker, C. J.; McCague, R.; Jordan, V. C. Molecular Pharmacology 1991, 39(3), 421-428.

55. Fang, H.; Tong, W.; Shi, L. M.; Blair, R.; Perkins, R.; Branham, W.; Hass, B. S.; Xie, Q.; Dial, S. L.; Moland, C. L.; Sheehan, D. M. Chemical research in toxicology 2001, 14(3), 280-294.

56. Philippe d, M.; Gilles, F.; Marc, P. Current Medicinal Chemistry - Anti-Cancer Agents 2004, 4(6), 491-508.

57. Hardcastle, I. R.; Rowlands, M. G.; Houghton, J.; Parr, I. B.; Potter, G. A.; Jarman, M.; Edwards, K. J.; Laughton, C. A.; Trent, J. O.; Neidle, S. Journal of Medicinal Chemistry 1995, 38(2), 241-248.

58. Gust, R.; Lubczyk, V. The Journal of Steroid Biochemistry and Molecular Biology 2003, 87(1), 75-83.

59. Menezes, I. R. A.; Leitão, A.; Montanari, C. A. Steroids 2006, 71(6), 417-428.

60. Maximov, P. Y.; Myers, C. B.; Curpan, R. F.; Lewis-Wambi, J. S.; Jordan, V. C. Journal of medicinal chemistry 2010, 53(8), 3273-3283.

61. Kuramochi, H. Journal of Medicinal Chemistry 1996, 39(15), 2877-2886.

62. Hossain, M. B.; van der Helm, D.; Schmitz, F. J.; Pordesimo, E. O.; Magarian, R. A.; Meyer, K. L.; Overacre, L. B.; Day, B. W. Journal of Medicinal Chemistry 1994, 37(11), 1670-1683.

63. Huang, M.-J. International Journal of Quantum Chemistry 2004, 96(4), 374-379.

64. Park, S.-J.; Kufareva, I.; Abagyan, R. Journal of computer-aided molecular design 2010, 24(5), 459-471.

65. Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Journal of Molecular Biology 1997, 266(2), 424-440.

66. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B. Angewandte Chemie (International ed 2006, 45(25), 4064-4092.

67. Simmerling, C.; Strockbine, B.; Roitberg, A. E. Journal of the American Chemical Society 2002, 124(38), 11258-11259.

68. Steinbrecher, T.; Labahn, A. Current Medicinal Chemistry 2010, 17(8), 767-785.

69. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Nature reviews 2004, 3(11), 935-949.

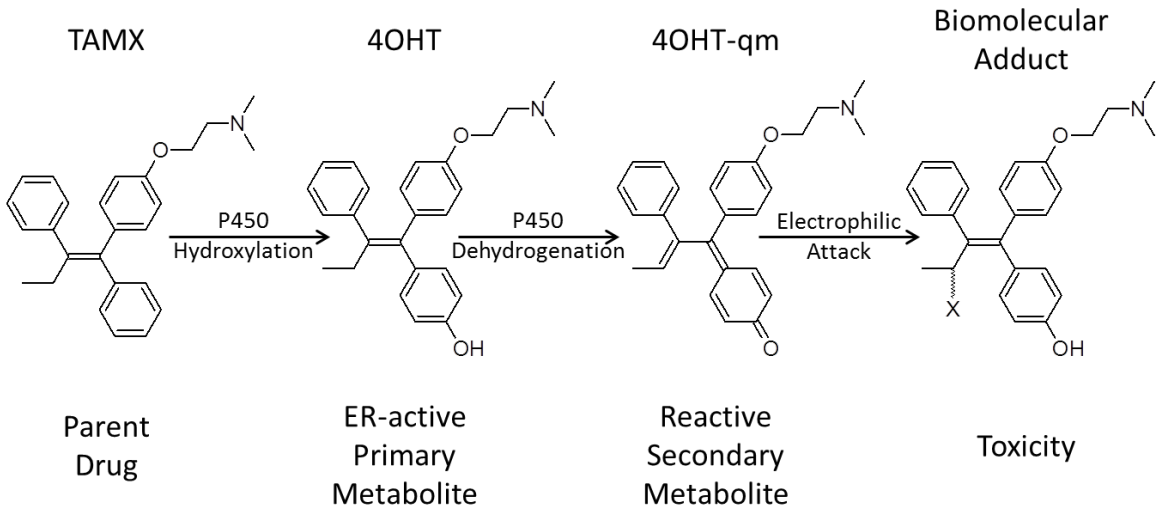


Figure 3.1: Scheme for one putative pathway of P450 mediated bioactivation of tamoxifen (TAMX) to estrogen-receptor (ER) active 4-hydroxy-tamoxifen via P450-catalyzed hydroxylation, followed by conversion to 4-hydroxy-tamoxifen quinone methide (4OHT-qm) which then forms an adduct with a biomolecule (X) via an electrophilic attack.

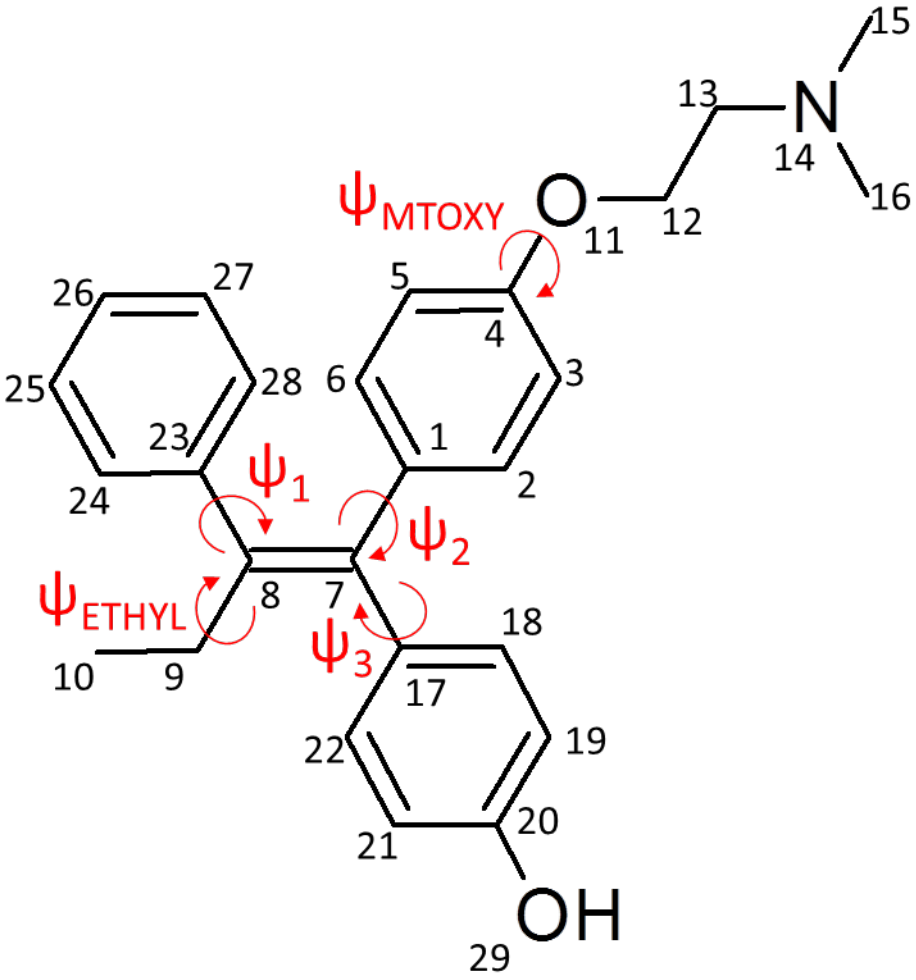


Figure 3.2: Atomic numbering and dihedral labeling scheme for 4-hydroxy-tamoxifen.

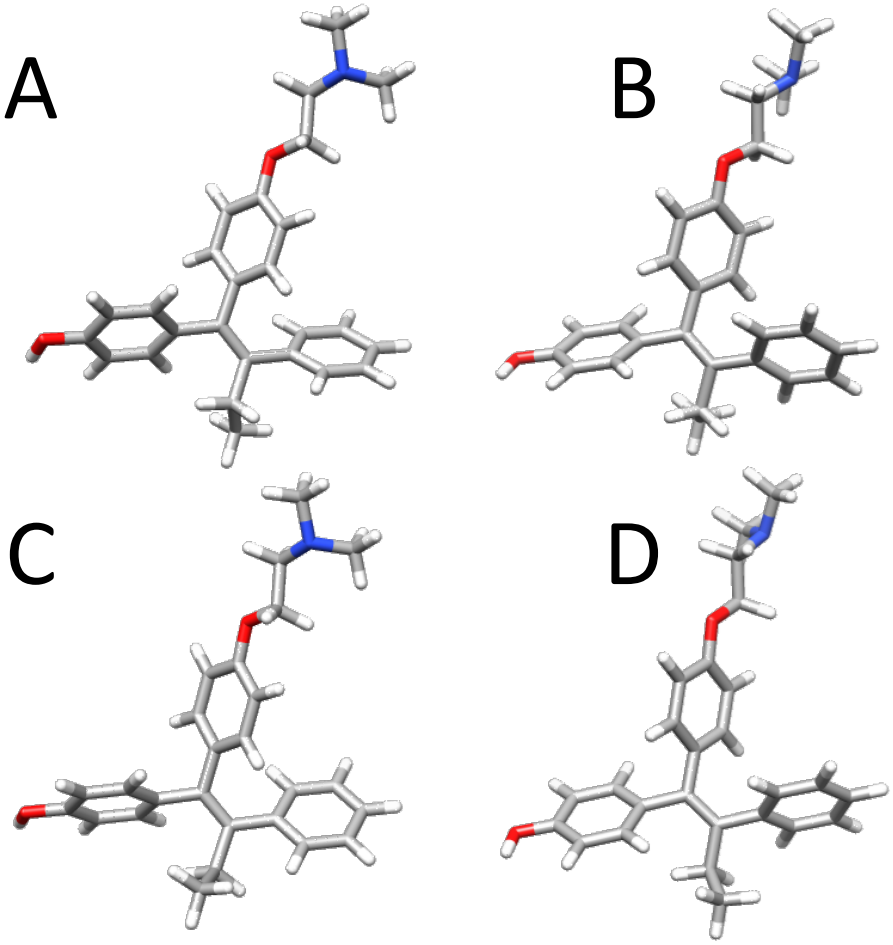


Figure 3.3: Four (A-D) different energy minimized geometries of 4-hydroxy tamoxifen identified at the B3LYP/6-31G* level of theory.

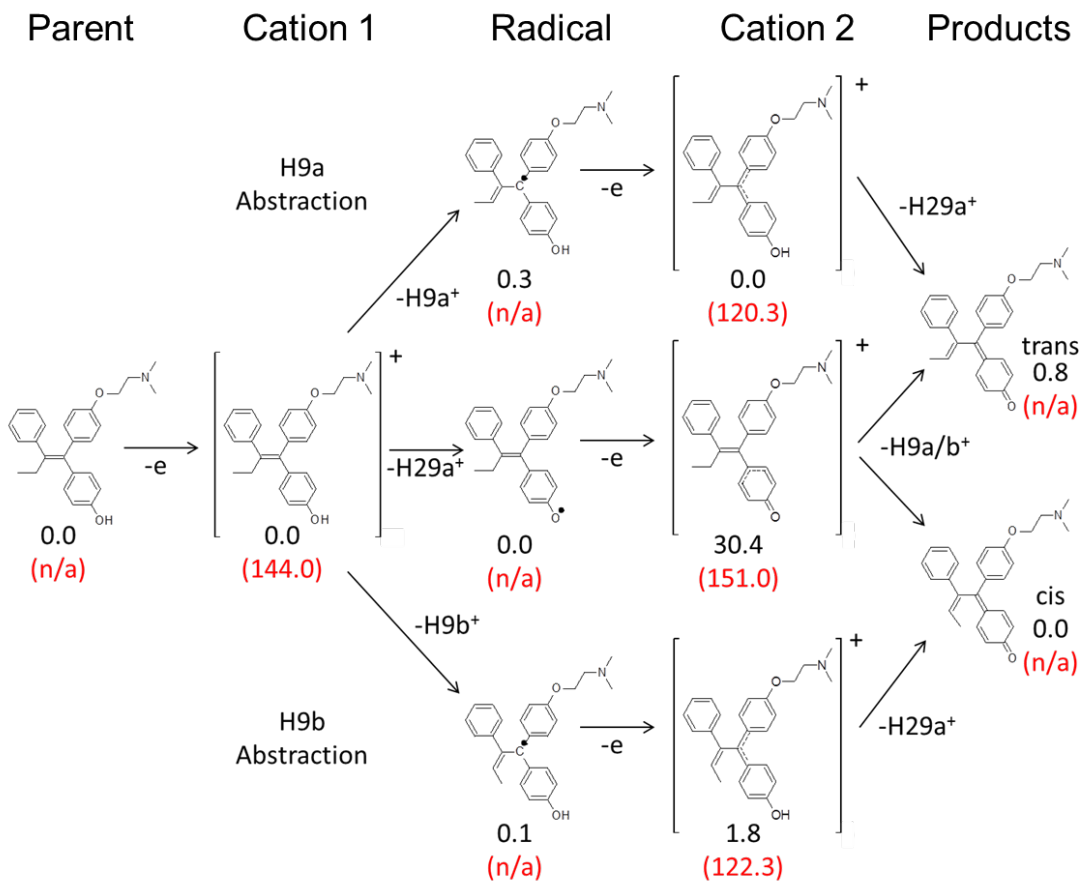


Figure 3.4: Scheme for alternate hydrogen abstraction pathways for 4OHT intermediates along the dehydrogenation reaction coordinate. Differences in the sum of electronic and thermal free energies for each species are shown relative to that of the lowest energy equivalent intermediate in black. Electron affinities are shown where applicable as relative energy to previous neutral intermediate in the same pathway in parentheses in red. All energies given in units of kcal/mol.

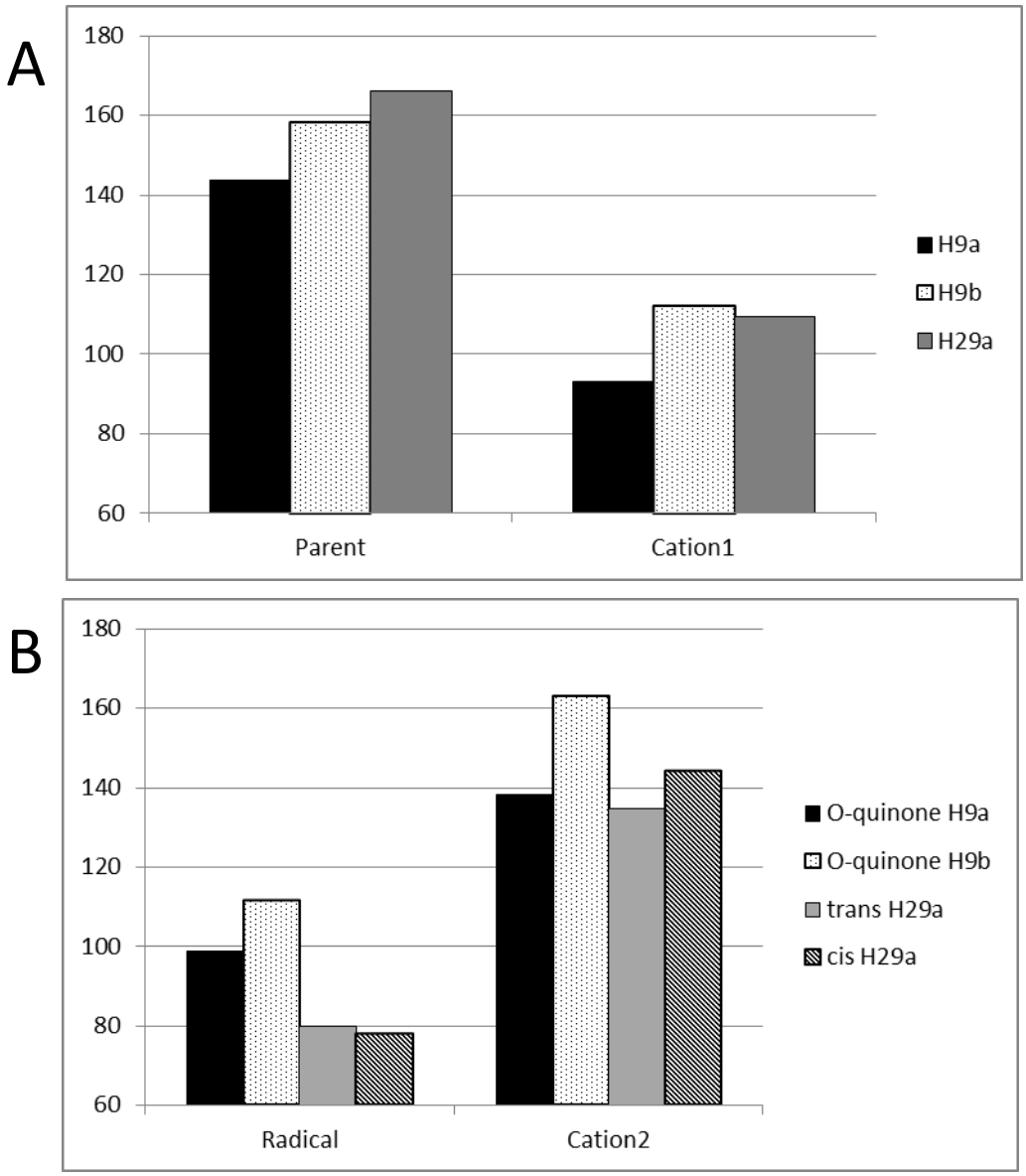
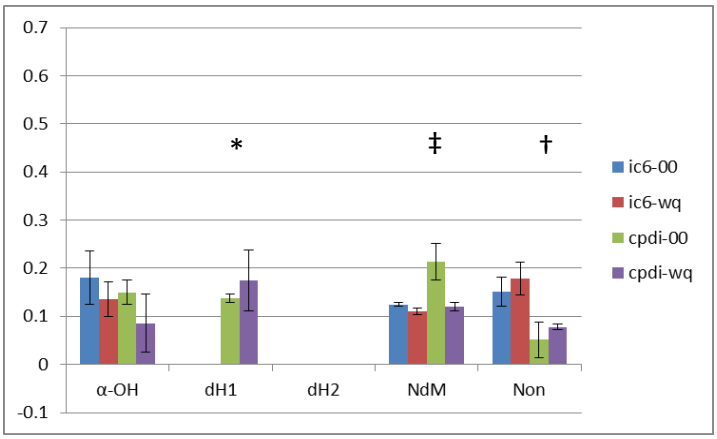


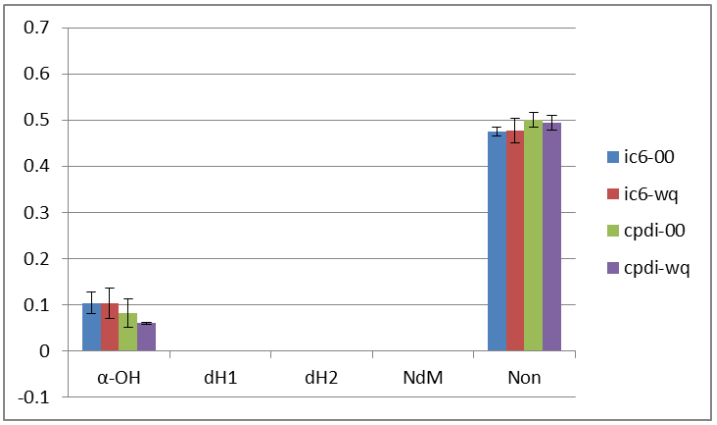
Figure 3.5: Final bond dissociation energies of carbon-hydrogen and oxygen-hydrogen bonds involved in the dehydrogenation of 4OHT at 4 Å extension for all intermediates on the dehydrogenation reaction coordinate: (A) first hydrogen abstraction, (B) second hydrogen abstraction. The legend indicates the hydrogen abstracted and for (B), the intermediate and the hydrogen abstracted.

Figure 3.6: Ratio of Autodock3 binding modes supportive of observed metabolism of 4-hydroxy-tamoxifen reaction mechanisms alpha-hydroxylation (α-OH) dehydrogenation modes 1 and 2 (dH1 & dH2), N-demethylation (NdM) and nonproductive and/or ambiguous modes (Non) with CYP3A4 experimentally derived x-ray (A) PDB ID: 1W0E, (B) PDB ID: 1TQN and (C) molecular dynamics refined m2 refined with quantum mechanics based heme parameters for resting high-spin (ic6) and Compound-I (cpdi) with (wq) and without (00) RESP charges assigned to the heme. Statistically significant differences are indicated with symbols: †, ‡ and *.

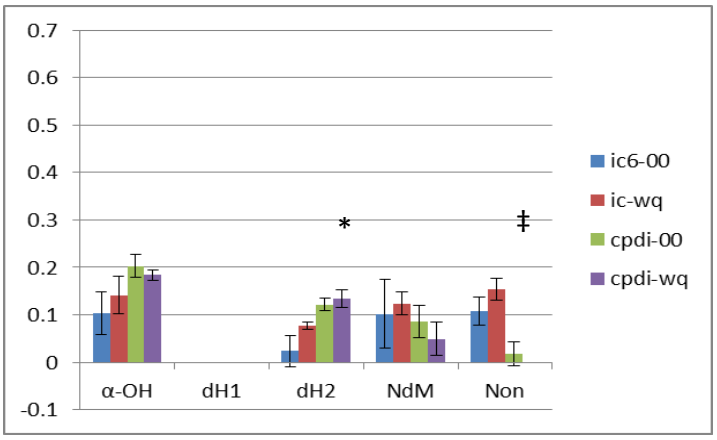
(A)


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
*
‡
†
α-OH
dH1
dH2
NdM
Non
ic6-00
ic6-wq
cpdi-00
cpdi-wq

(B)


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
α-OH
dH1
dH2
NdM
Non
ic6-00
ic6-wq
cpdi-00
cpdi-wq

(C)


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
*
‡
α-OH
dH1
dH2
NdM
Non
ic6-00
ic-wq
cpdi-00
cpdi-wq

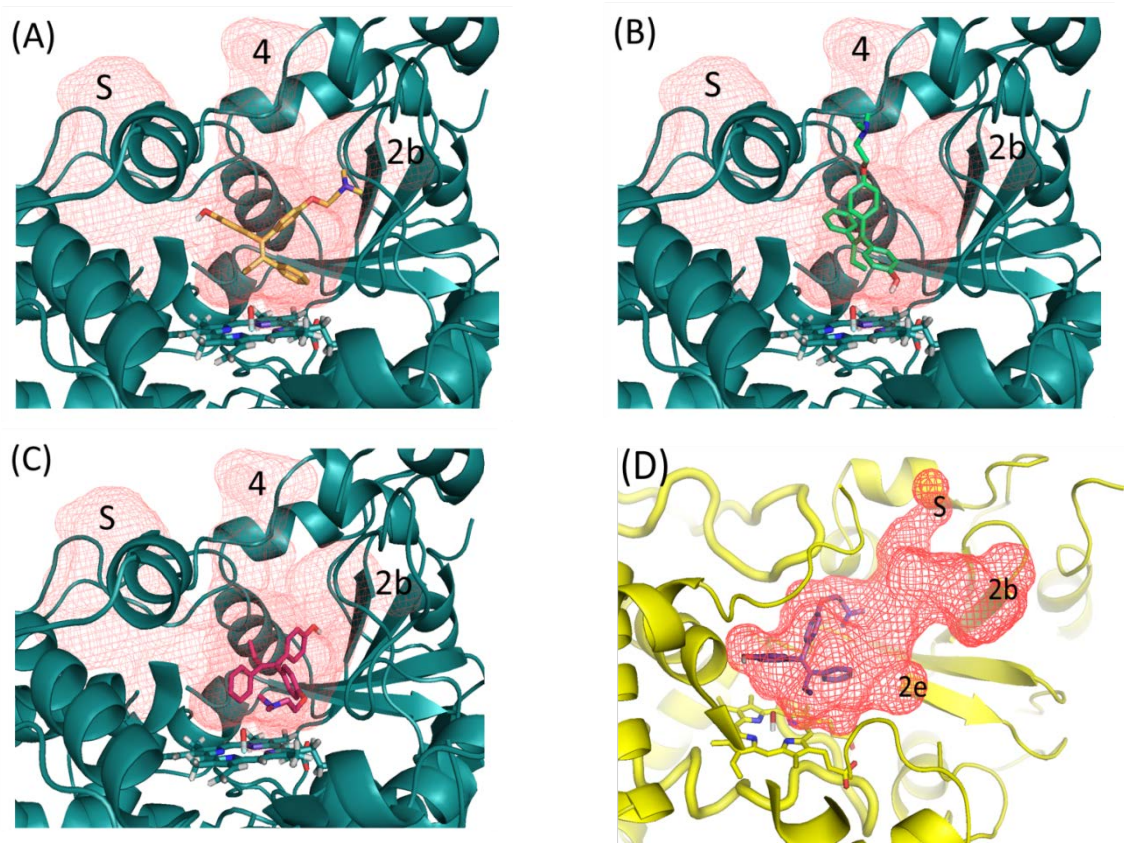


Figure 3.7: Active site and major channels (red grid) for structure m2-cpdi (cyan) and w0e-cpdi (yellow) shown with orientation of representative poses scored for different reaction mechanisms (A) α-hydroxylation [αOH], (B) dehydrogenation [dH2], (C) N-demethylation [NdM] and (D) dehydrogenation [dH1]. Volumes defining active-site and ingress/egress channels are shown as red mesh and labeled S (solvent), 2b, 2e and 4 channels. The 1w0e-cpdi and m2-cpdi model are shown at slightly different angles, and the peptide backbone in the foreground for all pictures has been hidden to allow the differences in the structure and orientation of channels and substrate to be seen more clearly. Active-site mapping performed with UCSF HOLLOW, figures generated with PYMOL.

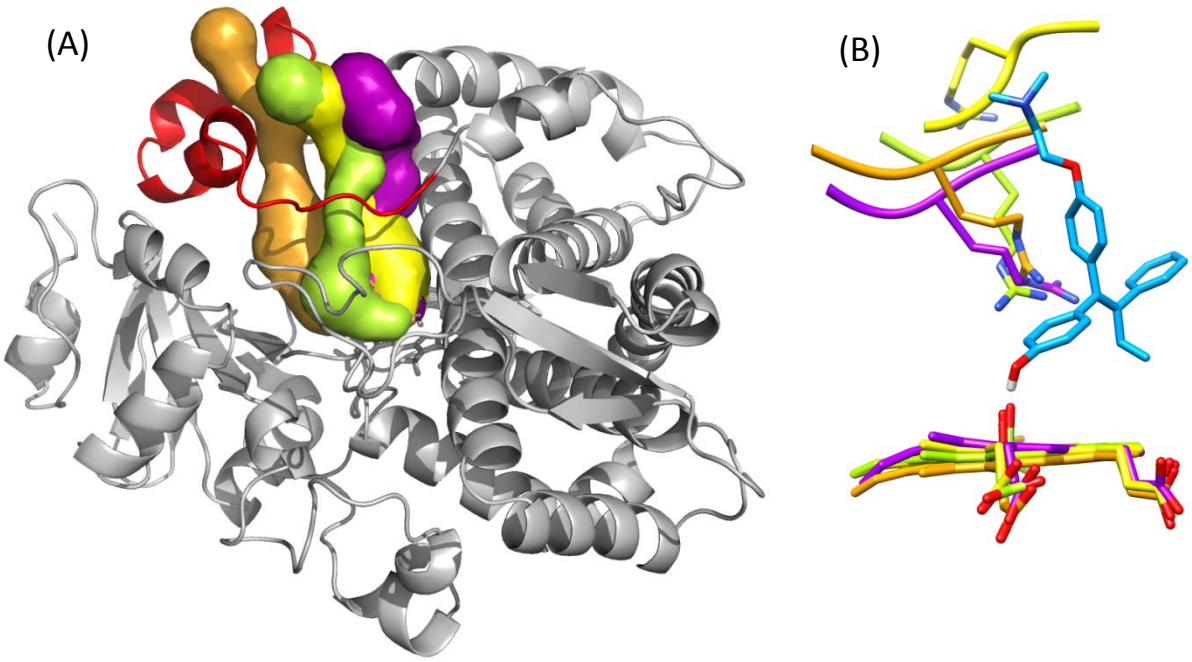


Figure 3.8: Differences in channel 4 and position of ARG212. (A) Volume and placement of channel 4 as defined by CAVER shown as a surface for each model m2 (lime green), m3 (purple), m5 (orange) and 2V0M (yellow). The CYP3A4 backbone is shown for the m2 model as a grey ribbon, the F-G loop (which includes the F' and G' helices) is shown in red. (B) Position of ARG212 and heme for m2 (lime green), m3 (purple), m5 (orange) and 2V0M (yellow) shown in reference to docked conformation of 4OHT (cyan) in dH2 pose. Hydrogen atoms have been hidden for clarity. Neighboring residues to ARG212 are shown only as ribbon for the back bone and atoms hidden also for clarity. Image generated (A) with PYMOL, and (B) with UCSF CHIMERA.

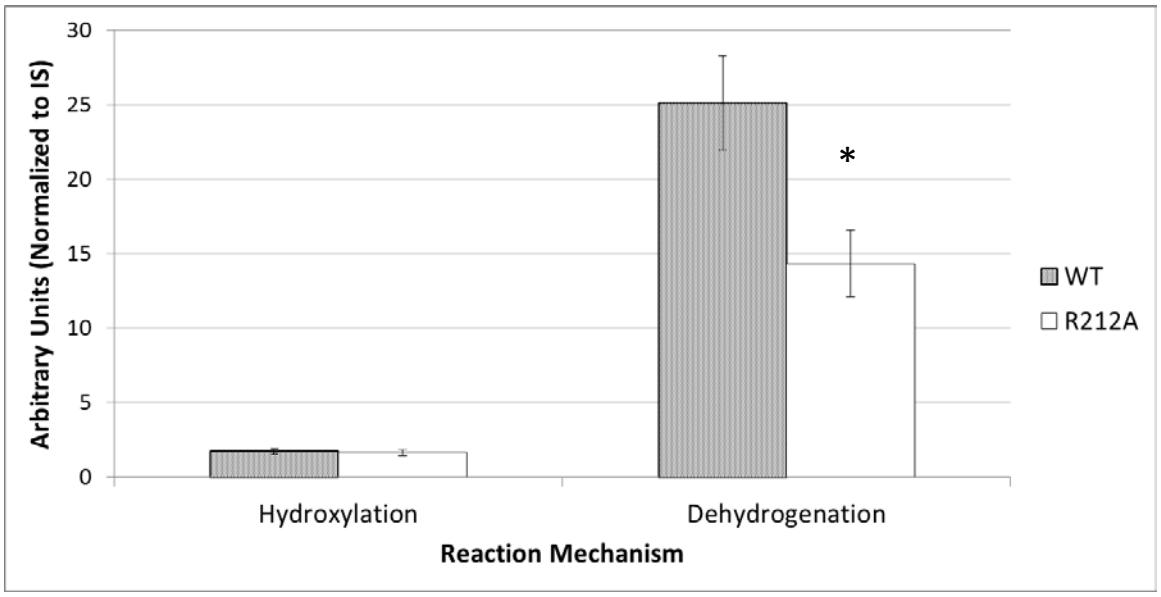


Figure 3.9: 4OHT relative dehydrogenation and hydroxylation 4OHT products of Wild-Type (WT) versus Mutant (R212A) CYP3A4**.** Statistical significance indicated with (*).

Table 3.1: Key dihedral angles of $\psi_{A\text{-}C,\ ETHYL\ \&\ MTOXY}$ and differences in zero-point energies (ΔZPE) and sum of electronic and free energies (ΔG) for four nonredundant optimized geometries of 4-hydroxy-tamoxifen. Conformations labeled (A-D) at results from calculations at (i) HF/6-31G*, (ii) B3LYP/6-31G* levels of theory. Differences in energy are shown in units of kcal/mol. Dihedral angles are shown in degrees.

(i)

HF/6-31G*	ψ_{ETHYL}	ψ_1	ψ_2	ψ_3	ψ_{MTOXY}	ΔZPE	ΔG
A	114.0	119.4	127.3	-64.1	1.8	0.0	0.0
B	81.2	-119.8	-127.5	69.1	87.3	3.2	1.9
C	-81.2	119.7	128.4	-68.8	1.8	2.0	1.5
D	-113.8	-119.5	-126.4	64.6	87.4	1.1	0.4

(ii)

B3LYP/6-31G*	ψ_{ETHYL}	ψ_1	ψ_2	ψ_3	ψ_{MTOXY}	ΔZPE	ΔG
A	115.7	127.9	134.7	-56.6	1.3	0.1	0.2
B	76.1	-136.5	-137.7	63.3	1.5	1.9	2.4
C	-76.1	136.0	137.1	-63.0	1.5	1.8	2.1
D	-115.4	-127.8	-134.5	56.5	1.0	0.0	0.0

Table 3.2: Identification and classification of (i) ingress/egress channels and (ii) binding modes supportive of different reaction mechanisms for refined molecular dynamics based (m1-5-cpdi) and x-ray-based (1w0e-cpdi & 1tqn-cpdi) models of CYP3A4.

(i)

model	# channel	2a	2b	2e	4	5	Solvent
m1-cpdi	2					5	S
m2-cpdi	3		2b		4		S
m3-cpdi	4		2b	2e	4		S
m4-cpdi	3		2b			5	S
m5-cpdi	4	2a	2b	2e	4		
1w0e-cpdi	4		2b	2e		5	S
1tqn-cpdi	3	2a		2e			S

(ii)

model	# productive modes	α-OH	dH1	dH2	NdM	Non
m1-cpdi	2				NdM	Non
m2-cpdi	3	α-OH		dH2	NdM	Non
m3-cpdi	1					Non
m4-cpdi	1				NdM	Non
m5-cpdi	2	α-OH	dH1		NdM	Non
1w0e-cpdi	3	α-OH	dH1		NdM	Non
1tqn-cpdi	1	α-OH				Non

CHAPTER 4

CONCLUSION

P450-mediated drug metabolism continues to be an area of intensive research even four decades after the discovery of this class of enzymes. The accumulating experimental evidence continues to provide insights into the functional diversity of these enzymes. Current experimental methods for structural studies of P450s, such as x-ray crystallography, only provide snap-shots of conformational states of the enzyme. For the P450s that have a number of x-ray structures available bound to chemically diverse ligands and substrates, the data provide insight into the potential plasticity of the active-sites of these enzymes. From these static snap-shots, the thermodynamic contributions to the structure of the enzyme cannot be determined. However, these structures do provide a framework of experimental evidence for the application for computational modeling techniques to predict the conformational dynamics of P450s. Improved models of P450 structure based on existing x-ray structures but accounting for thermodynamic contributions have been shown here to be important for accurately predicting drug metabolism. Thus, the application of computational tools for biophysical analysis in this manner can not only provide

insight into the basic biochemistry of enzyme-substrate interactions, but also have a beneficial impact on human health overall.

Recent work on P450 flexibility with molecular dynamics has identified the importance of the ingress and egress channels[1]. A number of MD-based studies in the absence[2,3] and presence[4] of the lipid bilayer using small substrates have shown that these channels can accommodate the movement of compounds to and from the active-site. Furthermore, data from spectroscopic analyses at normal and high pressures would indicate that family-specific differences exist in the malleability of the P450 active-site[5]. Unfortunately, even the most sophisticated current computational tools[6-9] for prediction of drug metabolism do not account for the fluctuations of the protein backbone nor the changes in the structure or the electronic state of the heme during the P450 catalytic cycle.

Based on the findings presented in this dissertation, both the QM-based parameters for changes in the heme during the P450 catalytic cycle, and the MM-based modeling of the thermodynamic fluctuations of the CYP3A4 active site, are required criteria to accurately model enzyme-substrate interactions. Each of these factors has been shown to play an important role in identifying binding modes that are supportive of the multiple pathways of the metabolism of 4OHT and raloxifene. In the case of raloxifene, the assignment of atomic partial charges to the heme increased the accuracy of our models. However, in the case of 4OHT, less dramatic effect due purely to the assignment of atomic partial charges was observed. This may have been due, in part, to the more polar and rigid 2-(4-hydroxyphenyl)-benzothiophen-6-ol core of raloxifene. This rigid core

structure may have induced a molecular reorientation which was more susceptible to local charge-charge interactions. It was shown that for 4OHT, a molecule with a more flexible core, the accurate modeling of the conformational dynamics of the enzyme structure was also important for identifying the binding modes suggestive of its entire observed metabolism. Accurately identifying binding modes supportive of the entire metabolism of novel compounds is important since minor pathways of metabolite production can be responsible for toxicities[10]. Unfortunately, most of the current tools that are used for predicting drug metabolism are only successful in identifying the major metabolite of the more substrate-selective hepatic drug metabolizing P450s, such as CYP2D6[6,9].

Our work with 4OHT and CYP3A4 shows that the active-site of CYP3A4 may differ in plasticity in specific key regions. Specifically, a binding mode was identified and experimentally confirmed for the dehydrogenation of 4OHT that involved a channel formed between the F'- and G'-helices in the F-G loop. This loop feature is unique to mammalian P450s[11], and this finding suggests that depending on the size, flexibility and polarity of the substrate, the binding mode supportive of catalysis may be dependent on reshaping of key regions of active site. Furthermore, it is important to stress that this feature of CYP3A4 was only observed through analysis of multiple MD trajectories of >40ns. Currently, most MD simulations use only one trajectory of shorter duration[2,9] to model the conformational dynamics of P450s. This multiple trajectory approach produced models of CYP3A4 that were then used with relatively unsophisticated docking

tools to identify the entire metabolism of 4OHT by CYP3A4, including binding modes supportive of its metabolism to a reactive quinone methide species.

These findings raise an intriguing question: Can the available computational tools now correctly identify the conformational dynamics of hepatic drug metabolizing enzymes? Based on the success of the approach developed and exemplified herein with raloxifene, 4OHT and CYP3A4, further investigation of this flexible characteristic of the major classes of hepatic P450s, CYP3A4, CYP2D6, CYP2E1, CYP2A6 and CYP1A2, with and without substrates, would be of great interest. These studies could identify regions of differing plasticity in the different P450 active sites. These data could then be used with model substrates to determine to what degree the plasticity of the active site played a role for different substrates and P450s. This new understanding would facilitate the development of classification scheme for P450s and their substrates. This classification scheme would determine how extensive a computational approach would be required to accurately determine metabolism of novel compounds as well as provide models that could be used with computationally affordable tools to rapidly screen chemical libraries.

The results from future research would not only be of benefit for drug-metabolism and safety but would also be of long-term benefit to a number of other areas of biomedical sciences. Experimental approaches that use site-directed mutagenesis, or directed evolution in altering P450 function, including application such as biosynthesis, rational drug-design targeting nonhepatic P450 such as for the treatment of prostate cancer, and P450 structural refinement

methods used in spectroscopy could all benefit from the identification of flexible motifs in the P450 active-site.

References

1. Cojocaru, V.; Winn, P. J.; Wade, R. C. Biochimica et Biophysica Acta (BBA) - General Subjects 2007, 1770(3), 390-401.

2. Li, W.; Liu, H.; Scott, E. E.; Gräter, F.; Halpert, J. R.; Luo, X.; Shen, J.; Jiang, H. Drug Metabolism and Disposition 2005, 33(7), 910-919.

3. Seifert, A.; Tatzel, S.; Schmid, R. D.; Pleiss, J. Proteins: Structure, Function, and Bioinformatics 2006, 64(1), 147-155.

4. Berka, K.; Hendrychová, T.; Anzenbacher, P.; Otyepka, M. The Journal of Physical Chemistry A 2011, 115(41), 11248-11255.

5. Hendrychová, T.; Anzenbacherová, E.; Hudeček, J.; Skopalík, J.; Lange, R.; Hildebrandt, P.; Otyepka, M.; Anzenbacher, P. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 2011, 1814(1), 58-68.

6. Li, J.; Schneebeli, S. T.; Bylund, J.; Farid, R.; Friesner, R. A. Journal of Chemical Theory and Computation 2011, 7(11), 3829-3845.

7. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. Journal of computational chemistry 2007, 28(6), 1145-1152.

8. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Journal of computational chemistry 1998, 19(14), 1639-1662.

9. Hritz, J.; de Ruiter, A.; Oostenbrink, C. Journal of medicinal chemistry 2008, 51(23), 7469-7477.

10. Baillie, T. A.; Cayen, M. N.; Fouda, H.; Gerson, R. J.; Green, J. D.; Grossman, S. J.; Klunk, L. J.; LeBlanc, B.; Perkins, D. G.; Shipley, L. A. Toxicology and Applied Pharmacology 2002, 182(3), 188-196.

11. Otyepka, M.; Skopalík, J.; Anzenbacherová, E.; Anzenbacher, P. Biochimica et Biophysica Acta (BBA) - General Subjects 2007, 1770(3), 376-389.