

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Computer aided design of selective cytochrome P450 inhibitors

Permalink

<https://escholarship.org/uc/item/9jt2k3jc>

Author

Verras, Andreas

Publication Date

2005

Peer reviewed|Thesis/dissertation

Computer Aided Design of Selective Cytochrome P450 Inhibitors

by

Andreas Verras

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

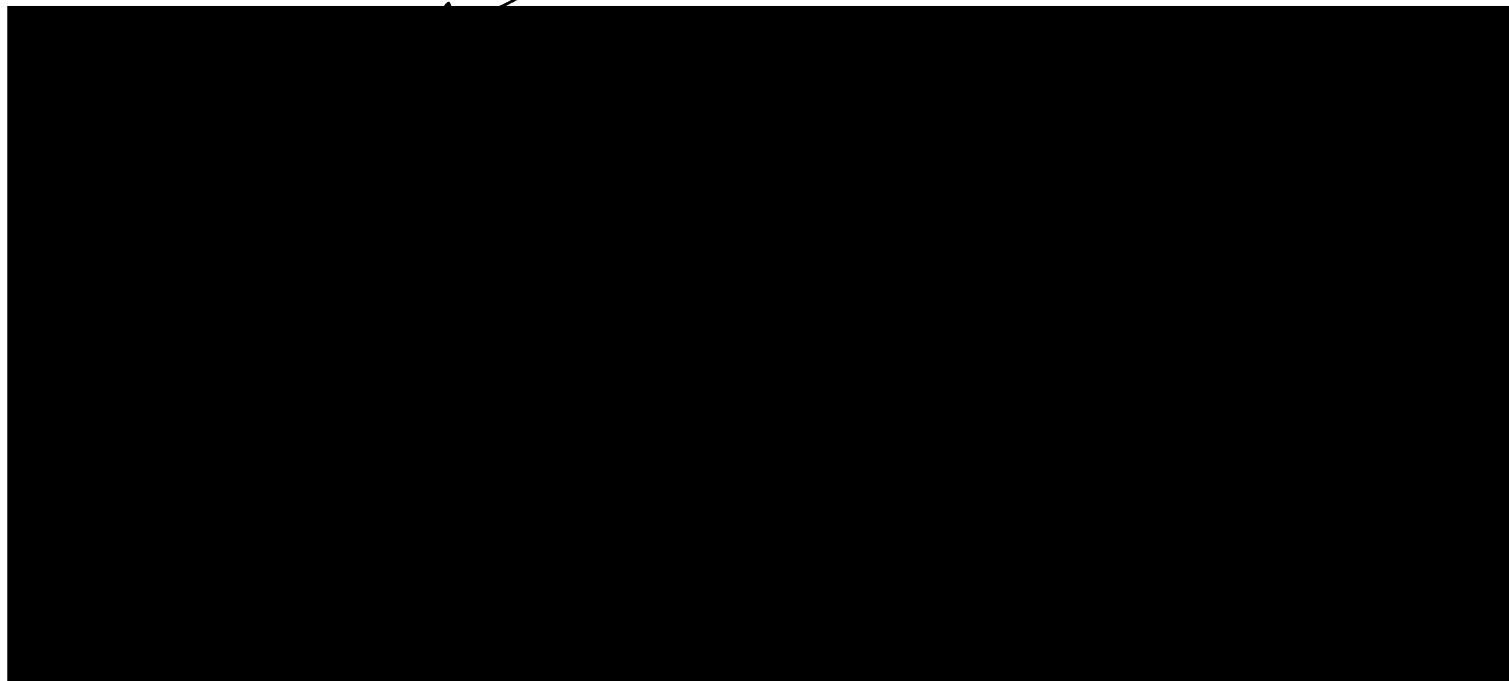
Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO



Preface

There are legions of people I would like to thank for making my graduate school studies fruitful and my experiences in San Francisco overwhelmingly positive. I would like to thank all the professors who helped guide my research, tutored me scientifically, and patiently helped at many stages during my stay at UCSF. Kip Guy, Volker Dotsch, Fred Cohen, and Almira Correia deserve thanks for sitting on my orals committee. Thanks to Peter Kollman for many fruitful discussions on computational chemistry and Ken Dill for my first rotation in a computational lab.

I would like to thank all the Kuntz and POM group members, past and present for numerous scientific and non-scientific discussions that made my time in lab a pleasure. Special thanks to Clinton Nishida, Wesley Straub, Geoffrey Skillman, and to every single person in the lab. They have all contributed to my research and personal life overwhelmingly. Also thanks to friends who were in different labs or graduate programs: Peter Chien, Adam Douglas, Zach Serber, and Nilesh Shah.

Christine Olsen, our departmental administrator has made my graduate degree much smoother than otherwise would have been possible. Thanks to her for reminding me I needed one more class to graduate well in advance of the deadline.

I would like to thank my friends outside of graduate school for much needed distraction and making me feel like San Francisco was my home: Joe Boyle, Bart Bernhart, Roberta Carrero, Nick Costarides, Caley Horan, Alex Ilnicki, Patrick Page, Julie Piacentine, TadpoleBear, and Ryan and Yuki Shaw.

I have immense gratitude to my family for their love and patience. My mother and father, Yianna and Athanasios, and my brother, Stathis, have always been close by with unconditional support.

I would like to thank Almira Corriea for agreeing to be on my thesis committee, always coming to see me speak, and for many delicious wine tastings.

Finally, I would like to thank Paul Ortiz de Montellano and Tack Kuntz, my graduate advisors, who have been instrumental in making me a scientist. Paul has always made himself instantly available to students and followed my work closely. He has patiently corrected my mistakes and been an exemplary advisor. Thanks to Tack for always hearing to my concerns, scientific and otherwise. I feel like I can approach him with any problem. I have been very lucky to have two excellent advisors who are role models for science and life.

Dedication

To my brother, Evstathios “Max” Verras, who is younger than me by two years,
but is someone I look up to.

To my mother and father, Yianna and Athanasios Verras, who have trusted and
encouraged me throughout graduate school.

Abstract

I have generated a virtual library of imidazole based compounds and used an altered form of the DOCK algorithm, CovDOCK, to screen these compounds against heme-containing proteins. I was able to generate high affinity ligands for P450cam. An L244A mutant of P450cam was generated and used to test our ability to generate selective ligands. Using our methods I was able to generate ligands which only inhibit the mutant enzyme.

The library was expanded with more substituted imidazoles and screened against a viable drug target, *Mycobacterium tuberculosis* CYP51. After one round of compound development I was able to find micromolar inhibitors. I also demonstrated that the methods and library are easily applicable to new protein systems.

To explore differences of the imidazole compounds in binding and inhibition, I carried out stopped flow kinetics experiments and further UV spectroscopic studies. I find that mono- and disubstituted imidazoles bind distinctly to wild-type P450cam. The natural substrate also affects both classes of ligands uniquely.

To explore the mitigated affinity for imidazole of the L244A mutant, I have done crystallography studies. I report the structures of the substrate-free mutant, and imidazole co-crystallized wild-type and mutant proteins. I present a possible explanation for the mitigated imidazole affinity based on structural factors observed when juxtaposing the wild-type and mutant structures.

Table of Contents

Preface.....	ii
Dedication	iv
Abstract.....	v
Table of Contents	vi
List of Figures.....	viii
List of Tables	x

Chapter 1 :

Introduction.....	1
--------------------------	----------

Chapter 2 :

Computer Assisted Design of Selective Imidazole Inhibitors for Cytochrome P450 Enzymes.....	49
--	-----------

Chapter 3 :

Expansion of the In Silico Imidazole Library and Application of Covalent DOCK to <i>M. tuberculosis</i>.....	87
---	-----------

Chapter 4 :

Differentiating Mono- and Di-Substituted Imidazole Ligands. Stopped Flow Analysis and Kinetics.....	116
--	------------

Chapter 5 :

Mitigated Imidazole Affinity: Crystal Structures of P450cam and P450cam L244A	146
--	------------

Chapter 6 :

Conclusions and Future Directions	185
--	------------

List of Figures

Chapter 1 :

Figure 1: Cytochrome P450 reaction cycle	5
Figure 2: P450 enzyme alignments.....	15
Figure 3: P450cam	16

Chapter 2 :

Figure 1: Covalent DOCK parameters	55
Figure 2: Covalent DOCK imidazole orientations	59
Figure 3: CovDOCK score components	60
Figure 4: Monosubstituted IC50 Selectivity.....	64
Figure 5: Difference Spectrum	69
Figure 6: K_s Values of the monosubstituted compounds	71
Figure 7: K_s Values of the disubstituted compounds	72

Chapter 3 :

Figure 1: CYP51 active site	93
Figure 2: Sample hits from first round of CovDOCK	95
Figure 3: CovDOCK score breakdown	96
Figure 4: Active site of CYP51 with bound 4-PIM	98
Figure 5: Second round CovDOCK hits	102

Chapter 4 :

Figure 1: Inhibitor affinity progression	122
Figure 2: Inhibition assays	124
Figure 3: Single point stopped flow	126
Figure 4: Initial dip at 416 nm	127
Figure 5: Substituted imidazole stopped flow	128
Figure 6: Comparison of two binding constants	129
Figure 7: Comparison of rates vs. ligand concentration	131
Figure 8: Stopped flow analysis of camphor	133
Figure 9: Two-state model of binding	141

Chapter 5 :

Figure 1: L244A crystals	154
Figure 2: Wild-type and L244A crystals	155
Figure 3: Temperature titrations of L244A crystal trays	158
Figure 4: L244A concentration titrations	159
Figure 5: Effects of imidazole on wild-type crystal morphology	161
Figure 6: Asymmetric unit of of P450cam L244A substrate-free protein	169
Figure 7: P450cam L244A substrate-free active site	170
Figure 8: P450cam L244A substrate-free active site waters	172
Figure 9: Wild-type P450cam imidazole-bound active site density	174
Figure 10: P450cam L244A imidazole-bound active site density map	176
Figure 11: Effects on imidazole binding by the L244A mutation	178

List of Tables

Chapter 1 :

Table 1: Currently Available P450 crystal structures	9
--	---

Chapter 2 :

Table 1: Monosubstituted IC50 values and CovDOCK scores	62
---	----

Table 2: Second round CovDOCK results	66
---	----

Table 3: Disubstituted IC50 values	68
--	----

Chapter 4 :

Table 1: Spectroscopic K_s values	120
---	-----

Table 2: K_s (μM) values vs. camphor concentration	135
---	-----

Chapter 5 :

Table 1: Data collection statistics for the L244A substrate-free protein	163
--	-----

Table 2: Data collection statistics for the wild-type P450cam crystallized in the presence of imidazole	165
--	-----

Table 3: Data collection statistics for P450cam L244A in the presence of imidazole ...	167
--	-----

1. Introduction

Overview

No enzyme family plays a more critical role in xenobiotic metabolism than that of the cytochrome P450 enzymes. Since their discovery in 1963 by Estabrook and coworkers [1], P450 enzymes have inspired the envy of chemists for their ability to perform synthetically difficult hydroxylations at unactivated carbons and have become a focal point for studies of xenobiotic metabolism.

Cytochrome P450 enzymes are best known for their role in phase I metabolism. They generally hydroxylate lipophilic compounds resulting in increased drug clearance, but P450 enzymes also have been implicated in toxic activation of compounds [2] and drug activation [3]. Furthermore, they play a critical role in several biosynthetic pathways such as steroid biosynthesis [4].

Because of their central role in metabolism, P450 enzymes are a frequent target of Absorption, Distribution, Metabolism, Excretion (ADME) studies [5]. The cost of drug development typically reaches hundreds of millions of dollars for a single drug [6] and eliminating candidates with poor pharmacokinetics or adverse drug-drug interactions has become critical in pre-clinical trial drug development. Understanding P450 specificity and function also has implications in bioremediation and the members of this enzyme superfamily are attractive candidates for bioengineering solutions to environmental detoxification [7].

Computational models have been applied to many aspects of these enzymes including inhibitor prediction [8], substrate prediction [9], profiling metabolism [10], determining mechanistic intermediates [11], and generating homology models [12].

Many of these studies are reviewed to provide the background of this thesis as well as suggest a context within which to frame the results and accomplishments of the research herein.

This thesis explores the application of structure-based drug design, specifically in inhibitor design against P450cam, but also contributes to our knowledge of P450 metabolite prediction, enzyme flexibility and promiscuity, and the engineering of selectivity in ligands. This is the cusp of a promising time in the P450 community. Computers continue to become more powerful and inexpensive, but, most importantly, we have witnessed an unprecedented growth of physiologically relevant P450 structural information in the last few years.

Atomic resolution structures of cytochrome P450 enzymes began with cytochrome P450cam. The structure was solved by Tom Poulos in 1985 [13] and has since been crystallized with dozens of ligands and with many mutations. Cytochrome P450cam has been highly characterized, expresses well, and has become an archetypal model for understanding P450 enzymes. Recently, a surface point mutation was reported that allows for more facile crystallization of this enzyme and promises to aid crystallographers in further characterizing this structure at even higher atomic resolution [14]. We exploit the properties of this surface mutation to aid in crystallizing an L244A mutant as will be discussed in Chapter 4.

Since the initial structure information of P450cam was deposited in the Protein Data Bank, many other P450 enzymes have joined the ranks of known structures. At first, these structures (P450 BM-3, CYP119, P450 nor, P450 EryF) [15-18] were all from

bacterial systems. The mammalian P450 enzymes, all membrane associated proteins compared to their soluble bacterial counterparts, remained elusive.

At the beginning of 2000 the first mammalian P450 crystal structure, that of cytochrome P450 2C5, was solved, [19]. This was a massive step forward in establishing a foothold on molecular resolution of human P450 enzymes. Not only was the structure critical for understanding a new class of P450 enzymes, but the methods used to allow its crystallization were now readily applicable to other mammalian isoforms. Deletion of the N-terminus in conjunction with several surface point mutations yielded soluble protein and these methods were then used to generate soluble P450 homologues for several human enzymes.

In the last year, two human structures have been reported. The structures of P450 2C9 and P450 3A4 were resolved to 2 Å [20, 21]. While cytochrome P450 enzymes metabolize thousands of xenobiotic compounds, P450 3A4 alone is responsible for approximately 60% of all phase I human drug metabolism [22]. With the wealth of structural information available, new directions and tools must be explored in an effort to increase our predictive powers in metabolism. Substrate and metabolism prediction is not a new science; a significant amount of research has already laid the foundation for future work and while much of it predates mammalian structural information, its inclusion is critical in realizing an accurate, fast method of metabolism profiling.

There are several inherent difficulties when modeling metabolite or inhibitor prediction of P450 enzymes. The reactive cycle of P450 enzymes highlights an immediate complexity. The mechanism of P450 enzymes, shown in figure one, illustrates the activated intermediates involved in catalysis. The first step is substrate binding to the

ferric complex, displacing water and changing the redox potential of the protein by approximately 100 mV [23], though the reduction potential of some P450 enzymes is independent of substrate binding [24, 25]. This is followed by a single electron transfer yielding the pentacoordinated ferrous complex. Molecular oxygen then binds and a second electron transfer occurs, resulting in oxygen bond cleavage and formation of the reactive ferryl oxygen species. Computational chemists are presented with a diversity of potential templates for modeling, for some of which there is little structural information. In predicting substrate specificity and the regioselectivity of metabolism, do we model the resting ferric state, which initially binds substrate, or the ferryl-oxygen species responsible for hydrogen abstraction? For our imidazole inhibitor studies (Chapter 2), the choice of intermediates is clear, that is the resting ferric state, but alternate inhibitors can act as suicide molecules and require activation by the ferryl oxygen species. Because of the complicated electrochemistry the species along the reaction cycle are distinct. There are significant electrostatic and steric differences in their attributes and conformational changes have been shown to occur in traversing the reaction cycle for at least some P450 enzymes [26]. Model choice can significantly affect prediction of substrates and metabolism.

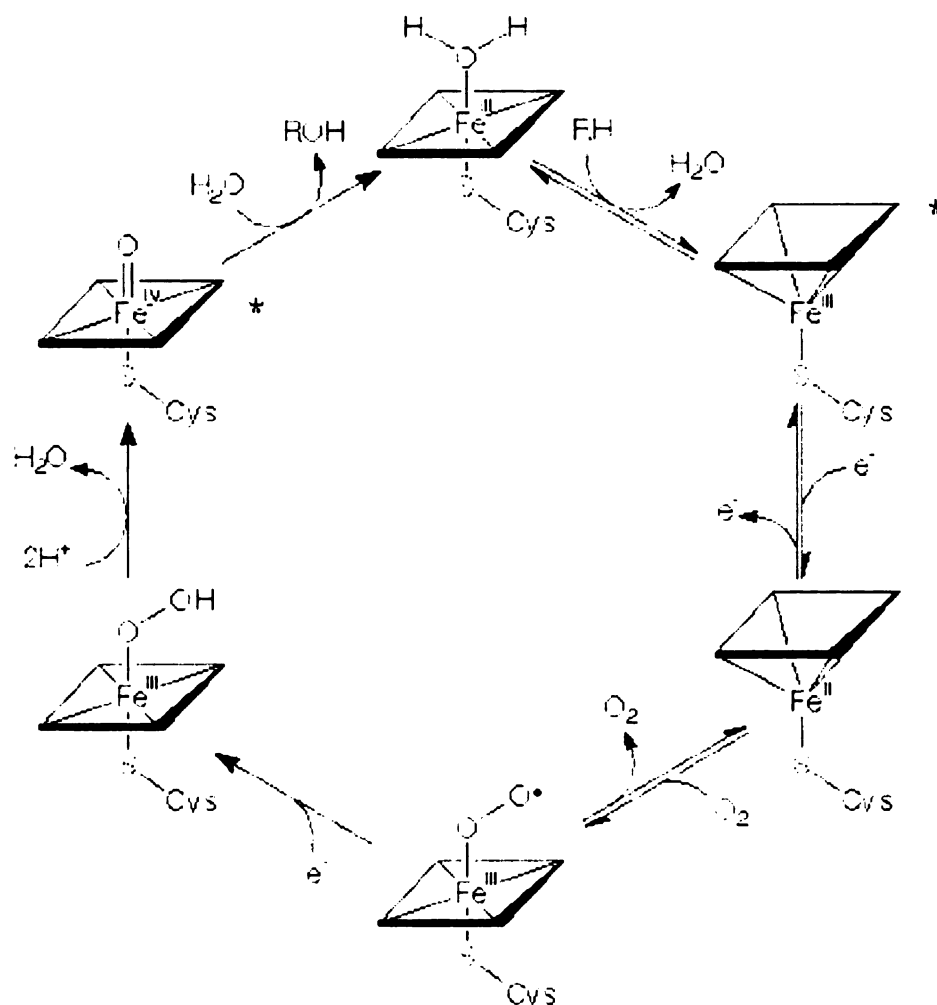


Figure 1. Cytochrome P450 reaction cycle. Species that are marked with an (*) are most frequently modeled in computational substrate or inhibitor prediction. The heme of cytochrome P450 is represented by the square with a central iron atom.

Another difficulty is the inherent flexibility of P450 enzymes. As was noted, a handful of P450 enzymes are responsible for metabolizing thousands of compounds and have thus evolved to be extremely promiscuous. Much of this promiscuity comes from the malleability of the active site side chains and breathing motions of the backbone.

This phenomenon will be discussed further in the introduction when we describe P450 structures and has been explored by our own research in Chapters 4 and 5.

Flexibility and substrate orientation also bring us to another consideration which must be minded in modeling P450-substrate interactions. At several steps in the catalytic cycle, uncoupling can occur [27] preventing the oxidation of substrate and instead resulting in the production of water or hydrogen peroxide. Several factors have been implicated in peroxide formation such as substrate-active site complementarity [28], poorly oriented substrates, or water occupancy of the active site [29]. For computational chemists this further confuses the practice of metabolism prediction. Not only must the substrate be able to bind, but the orientation of the substrate must be appropriate to allow for hydroxylation by the catalytically demanding P450 enzymes. The total volume occupied in the active site and the presence of waters further complicates efforts in substrate prediction.

Finally, another complication addressed by this thesis is the treatment of molecules that coordinate to the iron. Imidazole and other electron deficient aromatics are able to coordinate directly to the iron through sigma interactions and *p*-orbital backbonding. This will be discussed in more detail in Chapter two, but bears mention here as one of the difficulties in predicting ligand binding modes. Most scoring functions do not account for these complex interactions and can lead to mischaracterization of inhibitors as substrates.

The introduction is divided into three sections. We will begin by reviewing P450 structural information gleaned over the last twenty years from x-ray crystallography data and homology modeling. We will highlight some of the characteristics of P450 enzymes

which lead to their promiscuity and which present computational chemists with an assortment of technical challenges. Ligand based approaches predating structural information or including it implicitly will be considered. We will also focus on ligand based approaches which incorporate structural information explicitly, homology or otherwise. The third section will summarize the reviewed material and present the context and applicability of the research described within the thesis to inhibitor and substrate prediction. We also hope to highlight some of the ways in which ligand based models, molecular modeling, and structure-based approaches, working in tandem, can surmount the difficulties of predicting P450 protein-ligand interactions.

P450 Structures

Bacterial Enzymes

The first high resolution P450 crystal structure was that of P450cam [13, 30]. There was a pre-existing wealth of biochemical data on this enzyme that has since been expanded considerably. For these reasons P450cam has shaped the approaches we take to understanding the structure, function, and mechanism of these enzymes. Since determination of the structure of the wild-type protein, structures have been solved with alternate substrates [31, 32], inhibitors [33], mutations[34], and in the substrate free form [35]. With significantly more structures than any other P450 enzyme, P450cam remains the archetypal P450 model in the field.

After the solution of the P450cam structure, other prokaryotic P450 enzyme structures were solved by similar crystallographic methods, including those of P450 BM-3 [36], P450terp [37], and P450EryF[16]. A complete list of all structurally resolved

P450 enzymes, bacterial and mammalian, is shown in Table 1. While the overall identity of P450 enzymes between the various superfamilies is generally poor (about 20%), the quaternary structures of these proteins align well (Figure 2), revealing the presence of several conserved motifs.

Table 1. Currently available P450 crystal structures

Protein	PDB ID	Ligand	Resolution	References
<u>P450cam</u>				
ferric protein	1AKD	1-S camphor	1.8	[38]
ferrous protein	1CP4	phenyl radical	1.9	[39]
oxygen complex	1DZ4	camphor	1.6	[40]
ferryl oxygen (putative)	1DZ6	camphor	1.9	[40]
ferric protein	1DZ8	O ₂ , camphor	1.9	[40]
ferrous protein	1DZ9	oxygen, camphor	1.9	[40]
covalently modified protein	1GEM	n-butyl-isocyanide	1.5	[41]
	1GEM	n-butyl-isocyanide	2.0	[41]
	1GJM	n-2-(ferrocenylethyl)maleimide)	2.2	[42]
	1K2O	ruthenium-linker	1.65	[43]
	1QMQ	ruthenium-linker	1.55	[44]
	1LWL	fluorescent probe	2.2	[45]
	1O76	cyanide	1.8	[46]
	1PHA	UK-67254-13(+)	1.63	[33]
	1PHB	UK-67254-13(-)	1.6	[33]
	1PHC	none	1.6	[35]
	1PHD	2-phenylimidazole	1.6	[47]
	1PHE	2-phenylimidazole	1.6	[47]
	1PHF	4-phenylimidazole	1.6	[47]
	1PHG	metyrapone	1.6	[47]
L358P, C334A	1T85	camphor, CO	1.8	Poulos, to be published

WORK IN PROGRESS

L358P, C334A	1T86	camphor	1.9	Poulos, to be published
C334A	1T87	camphor, CO	1.8	Poulos, to be published
C334A	1T88	camphor	1.9	Poulos, to be published
T252A	2CP4	camphor	2.1	[48]
T252A	3CP4	adamantane	2.3	[48]
	4CP4	camphor	2.1	[48]
	5CP4	camphor	1.7	[49]
	2CPP	camphor	1.63	[30]
	3CPP	camphor, CO	1.9	[50]
	4CPP	adamantane	2.11	[32]
	5CPP	benzofuran adenine	2.08	[51]
D251N	6CP4	camphor	1.9	[49]
	6CPP	camphene	1.9	[32]
	7CPP	norcamphor	2.0	[51]
	8CPP	thiocamphor	2.1	[32]

P450 BM-3

Heme domain	1BU7	none	1.65	[36]
Heme-FMN domains	1BVY	none	2.03	[36]
Heme domain	1FAG	palmitoleic acid	2.7	[26]
Heme domain, T268A	1FAH	none	2.3	[52]
Heme domain, F393H	1JME	none	2.0	[53]
Heme domain	1JPZ	n-palmitoylglycine	1.65	[54]
Heme domain, F393A	1P0V	none	2.05	[55]
Heme domain, F393W	1P0W	none	2.0	[55]
Heme domain, F393Y	1P0X	none	2.0	[55]
Heme domain, A264E	1SMI	none	2.0	[56]
Heme domain, A264E	1SMJ	palmitoleic acid	2.75	[56]

108 10830

P450terp

1CPT none 2.3 [37]

P450 2C5

1DT6 none 3.0 [19]
1N6B sulfaphenazole 1.8 [57]
1NR6 diclofenac 2.1 [58]

P450 2C8

1PQ2 none (palmitic acid on surface) 2.7 [59]

P450 2C9

1OG2 none 2.6 [60]
1OG5 warfarin 2.55 [60]
1R9O flurbiprofen 2.0 [20]

P450 3A4

1TQN none 2.05 [21]
1W0E none 2.8 [61]
1W0F progesterone 2.65 [61]
1W0G metyrapone 2.74 [61]

007 11000

P450 2B4

1SU0 4-(4-chlorophenyl)imidazole 1.9 [62]

P450 2B5

1PO5 none 1.6 [63]

CYP 51

ferric protein
C37L, C151T, C442A

1E9X 4-phenylimidazole 2.1 [64]
1EA1 fluconazole 2.21 [64]
1H5Z none 2.05 [65]
1U13 none 2.01 Waterman, to be published
1X8V estriol 1.55 [65]

P450EryF

1EGY 9-aminophenanthrene (2 molecules) 2.35 [66]
1EUP androstenedione (2 molecules) 2.1 [66]
1JIN ketoconazole 2.3 [67]
1JIO 6-deoxyerythronolide b 2.1 [67]

P450Epok

1Q5D epothilone b 1.93 [68]
1Q5E none 2.65 [68]

P450Nor

WOLF BLOOM

P450 Oxyb

1LFK	none	1.7	[75]
1LG9	none	2.0	[75]
1LGF	none	2.2	[75]

P450 OxyC

1UED	none	1.9	[76]
------	------	-----	------

CYP 121

1N4G	iodopyrazole	1.8	[77]
------	--------------	-----	------

Cyp154A1

1ODO	4-phenylimidazole	1.85	[78]
------	-------------------	------	------

100% 100% 100%

With the recent solution of mammalian, membrane bound P450's this structural conservation is still observed. Not surprisingly, much of the variation in structure exists in the active site and contributes to the highly variable substrate specificities of these enzymes.

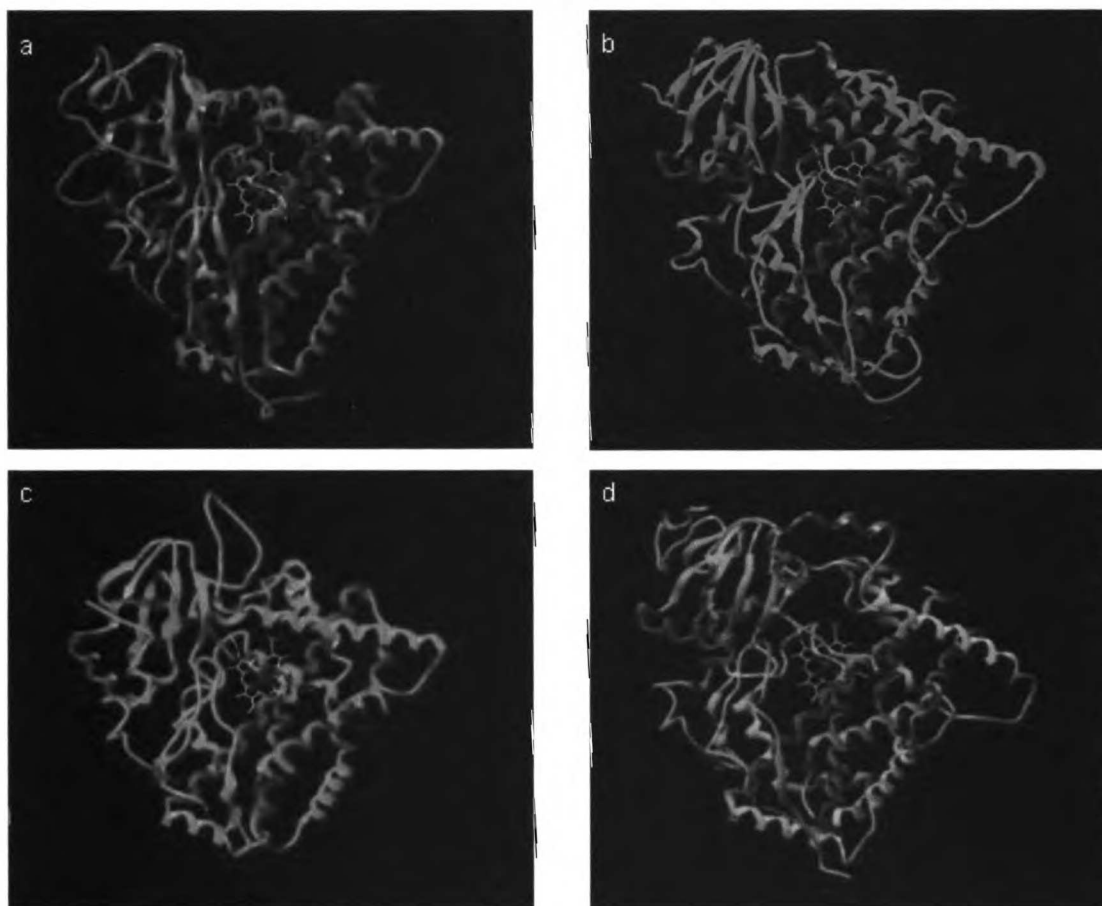


Figure 2. P450 enzyme alignments. a) P450cam, b) P450 BM-3, c) P450 EryF, and d) P450 3A4 aligned by superimposition of the heme groups. The overall structure is similar among the four enzymes despite less than 20% sequence identity.

The structure of P450 enzymes is primarily helical with a beta-pleated sheet region (Fig. 3). Unlike many proteins of its size it is not comprised of several distinct

domains, but instead of a single domain. A significant amount of insight regarding mechanism, electron relay pathways, and reductase partner interfaces has been gleaned from crystal structures; however, for the purpose of the introduction, we will focus on information contributing to understanding substrate specificity and regioselectivity. We will initially focus on P450cam.

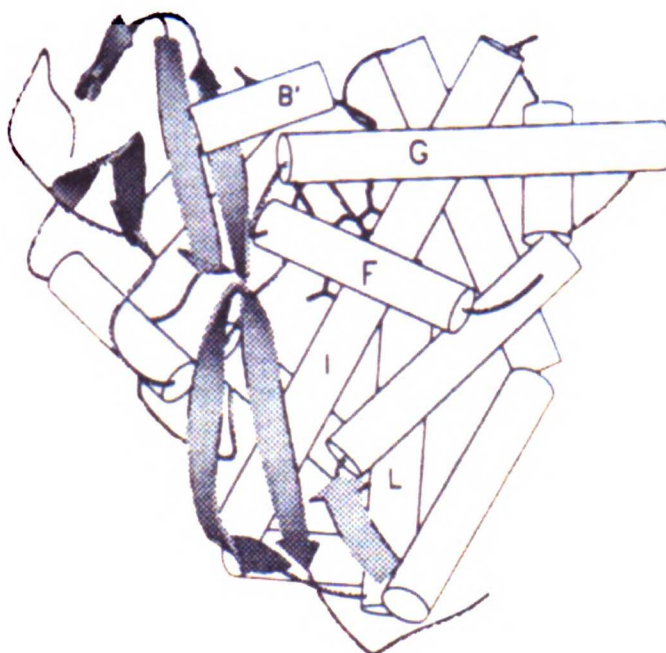


Figure 3. P450cam. Cartoon representation of P450cam with helices labeled according to P450 tertiary structure nomenclature.

While most of the helices in P450 enzymes are conserved between various families, the position of these helices frequently varies. For example, contrasting P450cam with P450EryF [79], we find that the position of the B' helix is significantly different between the two structures. In P450EryF the position of the B' helix is more distant from the heme, enlarging the active site cavity and completely removing its amino

UCSF LIBRARY
XEROX 15307

acid side chains from contributing to substrate binding. A long I helix runs down the length of the protein and contributes several residues to the active site. This feature is conserved amongst all P450 enzymes structurally characterized to date, eukaryotic and prokaryotic alike.

The Active Site Channel

The active site for the protein is buried in the center of the enzyme. A space filling model of P450cam shows no clear access channel to the heme. Several studies have been conducted in an effort to elucidate the path substrates must navigate to enter the active site and the protein movement that must occur concurrently to allow such access. The question has been approached computationally [80, 81] and crystallographically [43]. Using molecular dynamics and random forces on camphor beginning in the active site, Ludemann *et al.* computed a most likely path of entrance/exit for the ligand based upon relative energy barriers of expulsion. Structural approaches to channel elucidation have included co-crystallization of the protein with a ruthenium linker substrate that is bound at the heme and extends to the surface of the protein by a 9 carbon alkyl chain. By crystallizing the protein in the presence of this large ligand they were able to stabilize an open form of P450cam that resembles P450 enzymes with open access channels.

Both studies emphasized the importance of long salt bridging residues in forming an access channel of 22 Å length. Residues Arg186, Asp251, Lys178, and Asp182 were found to move while maintaining hydrogen bond and salt contacts, stabilizing the channel. At the protein surface, the channel was found to open between the F/G loop and

the B' helix. Those studies show the permutable nature of P450cam, flexible enough to bind 7, 9, and 13-mer alkyl chains. While the access channel must play a role in filtering potential substrates, the mechanism by which it does so remains elusive. Hydrophobicity seems to be a primary criterion, but to understand substrate specificity we must look at the active site.

Active Site Flexibility and Promiscuity

With the wealth of P450cam structures bound to a variety of substrates we are able to see structural permutations in a single enzyme and begin to understand the method by which this family adapts to a passel of substrates. Poulos et al. co-crystallized P450cam with UK-39671, a large compound found to be an inhibitor [33]. Large deformations of the active site relative to the native, substrate-bound conformation were necessary to bind this ligand. Tyr96, which in the substrate-bound structure forms a hydrogen bond to the camphor ketone group, is inverted in this structure, pointing up into the putative access channel.

In addition to side chain rearrangements, larger backbone motions have also been observed in co-crystals of P450cam with 1-, 2-, and 4-phenylimidazole [47]. With the exception of 2-phenylimidazole, the compounds coordinate to the heme iron through the nitrogen on the imidazole and the phenyl group makes Van der Waals contacts in the active site cavity usually occupied by camphor. Compared to the camphor structure the long, axial I helix is shifted in the presence of 1- and 4- phenylimidazole away from the heme. In the presence of 2-phenylimidazole, the I-helix is again shifted but in the opposite direction. As measured from the main chain atoms of the helix, the translations

are up to 2 Å in magnitude. The motions of the I helix are mirrored by large increases in the temperature factors in the region. Substrate induced changes have been observed in other P450 enzymes for which substrate free and ligand complexed structures have been resolved, including P450BM-3[15, 82] and CYP51 [64].

Of particular interest to this thesis are ligand-induced changes in the active site, for which there are abundant examples [65, 83]. One of the most striking examples of this phenomenon is illustrated by CYP119, a thermophilic P450 enzyme. The structure of this enzyme was solved in the presence of imidazole and 4-phenylimidazole [71] and, as the authors note, reveals an unprecedented rearrangement of the active site to adapt to each ligand. Accommodating both ligands is done not only by side chain conformer rearrangements, but also by unraveling of the F-G helix and backbone movements of up to 6 Å. For computational chemists pursuing structure-based design methods, these two structures perfectly illustrate the difficulties in substrate or inhibitor prediction. The CYP119 structure in the presence of imidazole, if assumed to be rigid, would suggest an inability to dock the larger 4-phenyl imidazole complex or lauric acid, which has been found to be oxidized by CYP119 [17].

We also observe ligand-induced changes in P450cam. This will be explored further in Chapter 2 and again in Chapter 5 through the use of crystallographic methods, which have yielded structures of P450cam and P450cam L244A in the presence and absence of imidazole.

Mammalian P450 Structures

The first mammalian structure solved was that of P450 2C5. This rabbit enzyme was crystallized by removing the N-terminal membrane anchor [19]. The soluble isozyme was resolved to 3.0 Å and represented a monumental breakthrough in P450 structure elucidation. It was followed by co-crystallization of the same construct with a derivatized dimethyl sulfaphenazole ligand resulting in an improved resolution of 2.3 Å. At this resolution the previously unresolvable F-G loop became visible as did the smaller F' and G' helices. The B-C loop and side chains were also better resolved. Overall, the tertiary structure of the first mammalian P450 resembled the soluble bacterial enzymes, even though the sequence identity was less than 20%.

In 2003, the first human P450 structure, that of P4502C9, was published [60]. Similar to P4502C5, the protein was a truncated construct that improved solubility and facilitated crystallization. P4502C9 is one of the prime actors in human metabolism and represented an immense resource for structural biologists predicting metabolism and drug-drug interactions. The structure was solved with bound warfarin, an anti-coagulant drug the enzyme is known to metabolize; however, the site of metabolism on the drug was positioned 10 Å away from the heme iron suggesting an inactive binding mode. The authors postulated a two-step binding mode, in which a subsequent rearrangement is necessary to bring the substrate adjacent to the heme or a second molecule must bind in a cooperative manner to enable oxidation.

The coordinates of another mammalian P450, P4502B5, were released in 2003 [63]. The enzyme was isolated as a homodimer and possessed a large open cleft running all the way to the heme. The channel was formed primarily from the B', C, F, and G

helices and appeared to represent an open form of the active site channel stabilized by helices contributed by the complementing homodimer. While in the same P450 superfamily as P4502C5, the structures are significantly different, illustrating the dynamic range of these enzymes necessary to promote metabolic flexibility.

Another member of the P450 2 superfamily, P4502C8, was solved in 2004 [59]. This enzyme has been implicated in the metabolism of taxol and cerivastatin, as well as retinoic and arachidonic acids. As is suggested by the enzyme's large substrates, the active site volume is twice that of P4502C5, occupying 1438 Å³.

Finally in June of 2004, structures of P4503A4 were deposited into the Protein Data Bank [61]. In humans, P4503A4 metabolizes more drugs than all other P450 enzymes combined. The protein was crystallized in the substrate-free form, with a bound substrate, and a bound inhibitor. Previous kinetic and binding studies suggested [84] that the enzyme could bind multiple substrates, and the crystal structure revealed a large active site in the enzyme consistent with this hypothesis.

In Chapter 3 we will show the ease with which our DOCKing methods and ligand library are applied to P450 systems other than P450cam. It is our hope that the methods described by the thesis in Chapter 2 for DOCKing imidazole based ligands into P450 enzymes will be applicable to mammalian systems.

Because of the vast numbers of substrates and the inability to crystallize all xenobiotics with all their respective metabolizing P450 enzymes, molecular modeling, docking, and molecular dynamics will always be necessary to elucidate protein-ligand interactions. Also, as we saw with P450cam, the structure of one protein-ligand complex does not immediately suggest a binding mode for alternate ligands; in fact, the static

structures from crystallography may seem to preclude the binding of known substrates. As with P450C9 and warfarin, the structure may yield suggestions for predicting metabolism, but not clearly indicate the mechanism of oxidation. For these reasons reconstructing molecular models remains an important pursuit.

Homology Modeling

Because of the inherent difficulties in crystallizing membrane bound proteins such as the mammalian P450 enzymes, homology modeling is often employed to determine structural information in the absence of crystallographic or NMR data. The process of homology modeling begins with a sequence alignment generated by methods such as ALIGN [85] or PSI-BLAST [86]. From this alignment one or more template proteins, whose structures are known, are chosen against which to align the target. Homology modeling is an iterative process that relies heavily on experimentally derived information to guide the process. In addition to requiring a template of at least 20% identity, mutational analysis and substrate and inhibitor information are used to guide the process of refinement. Several independent assessments of quality exist such as PROCHECK [87], PROSA [88], and WHAT IF [89]. These programs work by checking model characteristics including Ramachandran angles, atom-atom interactions, energy scores of interactions, and protein-solvent interactions. While these provide a method to assess the quality of a predicted structure, the true validity of any model must be judged by its ability to correctly suggest biochemical function, substrate specificity, or novel inhibitors.

An example of integrating experimental data with homology modeling can be found in Gilda Loew's molecular modeling of CYP4A11. The templates used were from

P450cam, P450terp, P450BM-3, and P450EryF, with a significant portion of backbone and sidechain information derived primarily from P450BM-3. Two models were generated with identical backbone coordinates, but different side chain conformers. Previous biochemical data identified lauric acid as a substrate for CYP4A11 and showed that the enzyme promoted the thermodynamically less favorable ω , rather than $\omega-1$ hydroxylation. Lauric acid was docked into the active sites of both models and they were submitted to 600 ps of molecular dynamics at 300° K. By PROSA, PROCHECK, and WHAT IF criteria, the models were indistinguishable in quality; however, as restraints in the molecular dynamics simulation were relaxed, one of the models was unable to anchor the lauric acid molecule and the substrate was expelled from the active site within the first 100 ps of the molecular dynamics trajectory. The other model, which was dynamically stable with respect to the ligand, was then used to implicate residues in the active site contributing to the unique regioselectivity of the 4A enzymes.

Following the solution of P450 2C9, a study by Afzelius *et al.* evaluated homology modeling techniques employed in the construction of several of these models, including single vs. multiple template alignments [90]. The authors used the crystal structures of P450's 2C9 and 2C5 as the metric by which to evaluate the homology models. A total of 15 homology models for each protein were generated using both single templates, 2C9 for 2C5 and vice versa, and multiple template alignments using P450cam, P450terp, P450BM-3, P450EryF, P4502C5, and P4502C9. The models were compared to their crystal structures by rmsd measurements and consensus principal component analysis derived from molecular interaction field descriptors of the protein active sites. This analysis gives both the overall fit of the molecules to their target

structures as well as a detailed side chain conformer description of the active site. Both quantifiers of fit illustrated that in all cases the models were more similar to their templates than to their target proteins. Using several proteins in the alignment resulted in higher similarity to the target, reinforcing the trend in the field toward multiple templates.

Structure-Based Metabolism Prediction, Ligand Based Approaches

Electronic Regioselectivity

One of the earliest computational models of metabolism was pioneered by Ken Korzekwa. With a dearth of protein structural data at the time, this model was able to predict metabolism profiles based upon the characteristics of the ligand, particularly electronic characteristics. Experimentally observed metabolism profiles, isotopic data, and atomic calculations displayed a tendency for oxidation following the potential for radical formation. That is N-dealkylation > O-dealkylation > secondary carbon oxidation > primary carbon oxidation. Using the AM1* semiempirical basis set, hydrogen abstraction transition sets were modeled for a set of compounds using *p*-nitrosophenoxy radical to model abstraction [91]. The heat of reaction and activation energy of radical formation correlated well with the experimental bond dissociation energies suggesting a method which worked quickly and was far more facile to model than active site transition states for each hydrogen in a compound.

This model was first applied to the metabolism prediction of nitrile bearing compounds [92]. Hydroxylation at the carbon alpha to the nitrile group resulted in a rapid loss of a cyanohydrin group as hydrogen cyanide. Experimental LD50 rates had been obtained for a set of nitrile compounds in mice [93]. The data correlated well with

logP, but for dinitriles or nitriles containing alkyl groups of four or more carbons, there was a poor correlation with hydrophobicity. A subset of 26 compounds was explored by the electronic model of hydrogen abstraction. The regioselectivity of hydroxylation was evaluated by comparing the potential for oxidation at all potential positions and followed by ranking the energy associated with radical formation at the position alpha to the nitrile, hydroxylation, and subsequent release of cyanide. The data correlated poorly to logP alone ($r = 0.59$), but with the inclusion of electronic parameters of radical formation the correlation improved to an r value of 0.85.

The model not only lacks any structural information of the enzyme, but there is no explicit knowledge of which P450 enzymes in particular contribute to hydroxylation. While ideally we hope to be able to predict the absolute regioselectivity for each compound and the corresponding P450 enzymes involved, the generality of this model suggested that electronic characteristics play a significant role in metabolism. For example, application of the same model with concurrent isotope effect analysis for benzylic hydroxylation indicated the relative steric contribution of P450 enzymes [94]. It was found that isotopically labeled 4-methoxy-3-methyl-phenol underwent benzylic hydroxylation for CYPs 1A2, 2B1, 2C8, 2C9, and 2E1, while 2B6 and 2B4 favored O-dealkylation. There are many examples of sterically and electrostatically driven P450 reactions, including the those of the CYP4A family [95] whose careful steric control results in fatty acid hydroxylation at the thermodynamically less favored omega position. To accurately predict substrate oxidation, protein structural information must be incorporated. Furthermore, to differentiate between substrates and inhibitors, one must be able to predict iron chelation of ligands such as imidazole.

Pharmacophores and 3D QSAR

In instances where there is little or no enzymatic structural information, pharmacophore and three dimensional quantitative structure-activity relationships (3D QSAR) provide techniques for predicting substrate specificity and regioselectivity. Pharmacophores are a set of features common to a series of active molecules used to deduce binding modes, specificity, and affinity [96]. Common pharmacophore groups include positive and negatively charged groups, hydrophobic groups, hydrogen bond acceptor and donor groups, and aromatic moieties. 3D QSAR techniques incorporate multiple drug descriptors including molecular weight, logP, and functionality. An often used 3D QSAR technique with cytochrome P450 enzymes is comparative molecular field analysis (CoMFA [97]), which overlays conformations of known active compounds and builds a molecular field grid about the molecules. The grid is then used in scoring new molecules. This approach enables us to estimate the enzyme active site topology in the absence of a crystal structure.

The promiscuous nature of P450 enzymes makes application of these techniques difficult. Both models begin with the assumption that compounds will be oriented similarly in the active site, an assumption that is not valid for many P450 enzymes. In fact, multiple hydroxylation sites on a single molecule by one P450 enzyme have frequently been observed [98], a finding that mandates that even one molecule may bind in multiple orientations. However, both techniques have shown reasonable success, more often with P450 enzymes with a relatively strict substrate specificity.

One of the strengths of pharmacophore and 3D-QSAR modeling is its ability to assimilate homology model information. Pharmacophore distances and CoMFA fields can be incorporated into homology models and data from homology models can be used to optimize overlaid conformation of training set molecules. For this reason, pharmacophore criteria are often explored in tandem with homology model structures. Often pharmacophore information is considered individually to remove model bias and then merged with protein structural data to produce a final model of ligand binding.

The combined pharmacophore-homology modeling technique was applied to CYP2D6 by de Groot et al. [99]. Initially they modeled 57 metabolic pathways based upon 40 substrates. The structures were built and energy minimized by the AM1-method. Debrisoquine and dextromethorphan, compounds with relatively low flexibility compared to the test set, were used as a template over which the other compounds were overlaid. A homology model of CYP2D6 was constructed using the structures of CYPs 101, 102, and 108 as templates. Molecular orbital (MO) calculations were carried out for all radical intermediates and hydroxylated products using the AM1 method.

The substrates were then docked into the CYP2D6 active site and a high complementarity was found between the pharmacophore model, which included a site of metabolism and three basic nitrogen atoms, and the protein active site, though both were obtained independently. The substrates were energy minimized with some restraints in the presence of protein. This highlights another important aspect of pharmacophore modeling and the importance of incorporating protein information. Because the final results are ultimately dependent on the conformations and alignment of the test set and because protein binding can permute the conformation of compounds relative to their free

state, energy minimization with a protein model is an important step toward a correct final conformation.

While MO calculations are important in deriving pharmacophores and in orienting the pharmacophore model in the active site, electronic calculations alone could not reproduce metabolic profiles. MO calculations generally favored O-dealkylation, as was seen with several of the compounds, but structural information was critical in elucidating the hydroxylation at alternate sites. The authors conclude that neither pharmacophore, homology model, nor electronic criteria alone are sufficient to predict metabolism and therefore reinforce the idea that these techniques must be merged to produce an accurate depiction of metabolism.

Structure-Based Design, Protein-Ligand Docking

In reviewing pharmacophore modeling I hoped to show that the various techniques described need not be distinct and in some cases are difficult to distinguish. Under the umbrella of structure-based design, I will include traditional ligand docking methods which include virtual screening methods such as DOCK [100], FlexX [101], GOLD [102], and AutoDock [103]. But I will also include molecular dynamics and associated quantum mechanical calculations.

Virtual screening is an attractive approach to substrate prediction or inhibitor design for several reasons. While structural information is required, no ligand test set is necessary though they can be employed for parameterization of scoring functions. Furthermore, depending on the scoring and sampling functions, thousands of compounds can be screened quickly. Because of the scale upon which this method can operate, it can

supplement high throughput assays and when the techniques are used in tandem, either in library design or assay organization, can significantly enrich successful compound discovery [104].

Virtual screening has not been applied extensively to P450 enzymes because of several inherent difficulties. For most of the above mentioned programs, single crystal structures are used to generate grids based upon rigid models of the active site. Once again, the large amount of promiscuity in P450 enzymes makes the assumption of a rigid active site potentially flawed. Also because the active site is generally modeled in its entirety, selection of its representation is critical. Docking to the ferryl-oxygen vs. ferrous forms can give dramatically different results. The electrostatics in the active site play a large role in virtual screening and have been calculated by several methods including semi-empirical INDO-ROHF calculations by Gilda Loew [105].

Early docking studies were performed with the first crystallized and best characterized P450 enzyme, P450cam. In 1997, De Voss et al. used the DOCK program to predict substrates for P450cam and a P450cam L244A mutant [106]. A total of 16 compounds were evaluated. Seven of the compounds predicted to bind to the wild-type were found to be substrates and compounds predicted not to fit were not metabolized. The L244A mutant was able to oxidize some of the larger compounds, incorrectly predicted to be substrates for the wild-type enzyme. Retroactive reparameterization of the DOCK scoring function resulted in accurate predictions for 15 of 16 compounds against the wild-type P450cam. It was found that changing the minimum contact distance for polar and nonpolar interactions from 2.3 to 2.4 Å and 2.8 to 2.9 Å,

respectively, resulted in better fits. With the new parameters the predictions for the L244A mutant also improved.

The interatomic distance De Voss et al. describe as a critical parameter illustrates approaches to modeling the malleability of the P450 enzymes. P450cam, found in *Pseudomonas putida*, oxidizes camphor which is the organism's sole source of carbon. As such, it is considered to be fairly specific, but studies have shown that even this enzyme can metabolize a variety of substrates [107].

As is often done in the computational community, surveys have been conducted of scoring functions and their applicability to distinct enzyme families [108-110]. Such a study was recently carried out for heme-containing proteins exclusively [111]. Using a test set of 45 protein-ligand complexes including several P450 enzymes, nitric oxide synthases, peroxidases, and hemoglobin the authors assessed the GOLD and Chemscore [112] scoring functions ability to return the appropriate ligands within 2 Å of their crystallized structures. The authors found generally a poorer success rate when docking to heme-proteins compared to another set of 139 diverse proteins[113]. For the heme-containing test set they found 57% of the complexes returned successfully by GOLD and 64% by Chemscore compared to 79% by both methods for the diverse protein test set. The authors attributed the discrepancy in predictive power to specific metal chelation factors, such as imidazole binding, not accounted for by the scoring functions, and the high lipophilicity of the P450 active sites. By reparameterizing the scoring function with empirically derived parameters for iron-acceptor contact atoms and modifying the lipophilic term of Chemscore, they were able to successfully return 73% of the corrected complexes.

Another example of scoring function validation has been done specifically with P450cam, exploiting the large amount of ligand-protein complex information for the enzyme [114]. The author used FlexX to dock a library of known P450cam substrates against the enzyme and then compared the GOLD, DOCK, FlexX, and PMF scoring functions. While PMF scores most often return correct binding modes, the GOLD derived ligand scores were found to correlate more closely with experimentally determined affinities. Using the combined methodologies of FlexX to position the ligands, PMF to nominate the top orientation, and GOLD to score it, the methods were applied to P450 3A4 to predict ligands. Preliminary results showed good success, though the scoring function could not discern between ligands and inhibitors.

One of the strengths of structure-based virtual screening in comparison with electronic characterization of metabolism and other methods which rely solely upon the ligand to predict regioselective oxidation [115, 116], is its ability to exclude compounds from consideration based upon steric constraints. Ideally one hopes that ligand docking would be able to correlate regioselectivity with specific P450 enzymes. Knowledge of metabolism at this level would enable design considerations that avoid drug-drug interactions or toxic activation of compounds by anomalously upregulated P450 enzymes (acetaminophen). For enzymes such as P450 2D6, which is absent in 5-9% of the Caucasian population [117], knowledge of compounds specifically metabolized by this isoform could have important implications in population subgroup therapy and pharmacogenomics.

While structure-based virtual screening has great potential, traditional methods do not capture the dynamic nature of protein-ligand binding. Induced fit effects such as side

chain reorganization or breathing motions in the secondary structure are difficult to model when high throughput screening is necessary. A study by Knegtel *et al.* quantitatively illustrates the effects of docking to a single structure [118]. Five HIV protease structures with co-crystallized ligands were used as a test set. DOCK was then employed to score each ligand against all the structures. While all the ligands scored well against their co-crystal structures, they frequently scored poorly against structures crystallized with alternate inhibitors. In some cases the inhibitors had a large positive score predicting that they could not bind at all to HIV protease. The authors were able to reproduce correct binding modes and computed energies based on energy-weighted or geometry-weighted averages of the composite structure and illustrated the importance of the target structure in virtual screening.

Molecular Dynamics

Several techniques exist by which to regenerate flexibility and motion from single static structures. Molecular dynamics is a computational method often employed for generating ensemble structures from one starting crystal structure or homology model. Because of the computational cost in time and computer power, large libraries cannot be explored by this method. Simulations usually involve single ligands, but can contribute significant information as to binding affinities, kinetics, and regioselectivity.

Kuhn *et al.* applied this method in evaluating the metabolism of sirolimus and a derivative by P450 3A4 using a combination of docking, molecular dynamics, and quantum chemical calculations [119]. Sirolimus is an immunosuppressant compound used in conjunction with cyclosporine to treat transplant patients [120] and was found to

have poor oral bioavailability [121]. A close derivative, everolimus, was found to have better bioavailability [122]. These are high molecular weight, flexible ligands generally larger than those evaluated previously by computational methods. Sirolimus and everolimus were docked into a P450 3A4 homology model [12] using experimentally known sites of metabolism as “anchors” for orientational sampling. The “anchors” were placed in the proximity of the heme and the rest of the molecule was then “grown” in using the “anchor-and-grow” functionality of DOCK. After simulated annealing, with restraints on the protein and protein-ligand distance, an unrestrained 60 ps equilibration was done followed by an unrestrained 160 ps simulation.

Radical intermediate energies were calculated at each hydrogen position in the ligands. The differences in energy were small, generally less than 1 kcal/mol, indicating the necessity of incorporating the protein contribution to regioselectivity. Protein-ligand conformations contributing to turnover were evaluated by distance and angle criteria. Reactive conformation criteria were established as a substrate hydrogen-ferryl oxygen distance less than 2.7 Å and an angle of 180° between the substrate carbon-abstracted hydrogen and ferryl oxygen. The criteria were applied to the simulation trajectory to give a probability of abstraction. They found that sirolimus had a higher probability of being found in a reactive conformation at the critical C52 position (25%, 8% respectively for the two criteria) compared to everolimus (3%, 4%). The authors were thus able to make qualitative predictions regarding the relative stability of everolimus compared with sirolimus, as well as lend insight into protein-ligand interactions for P450 3A4.

While the application of molecular dynamics is able to generate ensemble structures and estimate non-rigid active sites, it is computationally expensive and it is

difficult to envision its application to thousands of ligands. Instead it is better suited for exploring metabolic activity for prime candidates in a drug development pipeline and tailoring the molecules appropriately [123]. Alternate methods have been used to estimate flexibility while maintaining computational alacrity, one of which will be reviewed here.

Summary

A survey of the literature indicates that the majority of the studies that have been conducted on *in silico* screening of compounds against P450 enzymes have employed pharmacophore and 3D-QSAR methods. Docking based structural approaches, such as those presented in this thesis, are less well represented in the literature. Several factors are implicated in this disproportionate representation. First, the paucity of crystallographic information on physiologically relevant P450 enzymes has only recently been alleviated. Second, because of the physiological relevance of P450 enzymes and the relative ease with which ligand oxidation can be screened either with rat microsomes or single P450 enzymes *in vitro*, there is a wealth of ligand based data. These data naturally lend themselves to ligand-based techniques such as CoMFA or pharmacophore approaches. Last, as we have pointed out in our discussion of substrate-induced changes, the conformational liberty enjoyed by the P450 enzymes mandates the need for non-rigid or ensemble approaches to ligand docking. The majority of docking algorithms in use today target single enzyme active sites instead of an ensemble of structures. These enzymes are also considered as rigid to allow for quick screening of thousands of compounds.

One further consideration worth mentioning is the potential of P450 3A4, the prime actor in human metabolism, to bind multiple substrates [124, 125]. No docking, or pharmacophore method for that matter, has explicit methods for predicting multiple substrates simultaneously in a single active site, but to accurately predict P450 3A4 metabolism of small molecules, such an approach must be incorporated.

After pharmacophore modeling, molecular dynamics methods are most often employed in predicting P450 metabolism. While this method has the advantages of using an ensemble of structures, the ability to reproduce protein flexibility, and quantify reactive geometries necessary for oxidation compared to geometries resulting in uncoupled or inactive orientations, it is extremely time consuming.

In this thesis, I will begin by describing construction of a virtual library of imidazole inhibitors. To overcome DOCK's lack of a chelation term in the scoring function, I introduce a harmonic restraint which returns an accurate binding mode for our ligands. While I do not explicitly use pharmacophore modeling, our library was generated and aligned based on this imidazole group.

Because of the recent expansion in structural information for P450 enzymes, it is our hope that the library of compounds and methods described to dock them, will be easily applied to other P450 enzymes. To this end I expanded the library by including additional compounds based upon an alternate synthetic scheme, and docked these compounds into the P450 sterol 14 α -demethylase from *Mycobacterium tuberculosis*, a viable drug target.

Our studies illustrate the dynamic nature of the P450cam active site. Inhibitors predicted not to fit into a rigid model are able to bind to the wild-type P450cam active

site. We further explore this active site malleability by crystallographic studies and attempt to qualitatively differentiate between classes of inhibitors using kinetic assays.

References

1. Estabrook, R.W., D.Y. Cooper, and O. Rosenthal, *The Light Reversible Carbon Monoxide Inhibition of the Steroid C21-Hydroxylase System of the Adrenal Cortex*. *Biochem Z*, 1963. **338**: p. 741-55.
2. Jaeschke, H., et al., *Mechanisms of hepatotoxicity*. *Toxicol Sci*, 2002. **65**(2): p. 166-76.
3. Chen, C.S., et al., *Activation of the anticancer prodrugs cyclophosphamide and ifosfamide: identification of cytochrome P450 2B enzymes and site-specific mutants with improved enzyme kinetics*. *Mol Pharmacol*, 2004. **65**(5): p. 1278-85.
4. Waxman, D.J., *Interactions of hepatic cytochromes P-450 with steroid hormones. Regioselectivity and stereospecificity of steroid metabolism and hormonal regulation of rat P-450 enzyme expression*. *Biochem Pharmacol*, 1988. **37**(1): p. 71-84.
5. Wynalda, M.A. and L.C. Wienkers, *Assessment of potential interactions between dopamine receptor agonists and various human cytochrome P450 enzymes using a simple in vitro inhibition screen*. *Drug Metab Dispos*, 1997. **25**(10): p. 1211-4.
6. Kaitin, K., *Post-approval R&D raises total drug development costs to \$897 million*. 2003, Tufts Center for the Study of Drug Development.
7. Morant, M., et al., *Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation*. *Curr Opin Biotechnol*, 2003. **14**(2): p. 151-62.
8. Rao, S., et al., *A refined 3-dimensional QSAR of cytochrome P450 2C9: computational predictions of drug interactions*. *J Med Chem*, 2000. **43**(15): p. 2789-96.
9. Ekins, S., et al., *Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates*. *J Pharmacol Exp Ther*, 1999. **291**(1): p. 424-33.

10. Jones, J.P. and K.R. Korzekwa, *Predicting the rates and regioselectivity of reactions mediated by the P450 superfamily*. Methods Enzymol, 1996. **272**: p. 326-35.
11. Shaik, S., et al., *Two-state reactivity mechanisms of hydroxylation and epoxidation by cytochrome P-450 revealed by theory*. Curr Opin Chem Biol, 2002. **6**(5): p. 556-67.
12. Szklarz, G.D. and J.R. Halpert, *Molecular modeling of cytochrome P450 3A4*. J Comput Aided Mol Des, 1997. **11**(3): p. 265-72.
13. Poulos, T.L., et al., *The 2.6-Å crystal structure of Pseudomonas putida cytochrome P-450*. J Biol Chem, 1985. **260**(30): p. 16122-30.
14. Nickerson, D., L.L. Wong, and Z. Rao, *An improved procedure for the preparation of X-ray diffraction-quality crystals of cytochrome p450cam*. Acta Crystallogr D Biol Crystallogr, 1998. **54** (Pt 3): p. 470-2.
15. Ravichandran, K.G., et al., *Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's*. Science, 1993. **261**(5122): p. 731-6.
16. Cupp-Vickery, J.R. and T.L. Poulos, *Structure of cytochrome P450eryF involved in erythromycin biosynthesis*. Nat Struct Biol, 1995. **2**(2): p. 144-53.
7. Koo, L.S., et al., *Enhanced electron transfer and lauric acid hydroxylation by site-directed mutagenesis of CYP119*. J Am Chem Soc, 2002. **124**(20): p. 5684-91.

Shimizu, H., et al., *Crystal structures of cytochrome P450nor and its mutants (Ser286-->Val, Thr) in the ferric resting state at cryogenic temperature: a comparative analysis with monooxygenase cytochrome P450s*. J Inorg Biochem, 2000. **81**(3): p. 191-205.

Williams, P.A., et al., *Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity*. Mol Cell, 2000. **5**(1): p. 121-31.

Wester, M.R., et al., *The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-Å resolution*. J Biol Chem, 2004. **279**(34): p. 35630-7.

21. Yano, J.K., et al., *The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution*. J Biol Chem, 2004. **279**(37): p. 38091-4.
22. Wrighton, S.A., et al., *The human CYP3A subfamily: practical considerations*. Drug Metab Rev, 2000. **32**(3-4): p. 339-61.
23. Ruckpaul, K., H. Rein, and J. Blanck, *[Regulation mechanisms of the endoplasmic cytochrome P-450 systems of the liver]*. Biomed Biochim Acta, 1985. **44**(3): p. 351-79.
24. Kim, D. and F.P. Guengerich, *Selection of human cytochrome P450 1A2 mutants with enhanced catalytic activity for heterocyclic amine N-hydroxylation*. Biochemistry, 2004. **43**(4): p. 981-8.
25. Aguey-Zinsou, K.F., et al., *Electrochemistry of P450cin: new insights into P450 electron transfer*. Chem Commun (Camb), 2003(3): p. 418-9.
26. Li, H. and T.L. Poulos, *The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid*. Nat Struct Biol, 1997. **4**(2): p. 140-6.
27. Atkins, W.M. and S.G. Sligar, *Metabolic switching in cytochrome P-450cam: deuterium isotope effects on regiospecificity and the monooxygenase/oxidase ratio*. Journal of the American Chemical Society, 1987. **107**(12): p. 3754-3760.
28. Gelb, M.H., P. Malkonen, and S.G. Sligar, *Cytochrome P450cam catalyzed epoxidation of dehydrocamphor*. Biochem Biophys Res Commun, 1982. **104**(3): p. 853-8.
29. Loida, P.J. and S.G. Sligar, *Molecular recognition in cytochrome P-450: mechanism for the control of uncoupling reactions*. Biochemistry, 1993. **32**(43): p. 11530-8.
30. Poulos, T.L., B.C. Finzel, and A.J. Howard, *High-resolution crystal structure of cytochrome P450cam*. J Mol Biol, 1987. **195**(3): p. 687-700.
31. Strickler, M., et al., *Crystallographic studies on the complex behavior of nicotine binding to P450cam (CYP101)*. Biochemistry, 2003. **42**(41): p. 11943-50.

32. Raag, R. and T.L. Poulos, *Crystal structures of cytochrome P-450CAM complexed with camphane, thiocamphor, and adamantane: factors controlling P-450 substrate hydroxylation*. *Biochemistry*, 1991. **30**(10): p. 2674-84.
33. Raag, R., et al., *Inhibitor-induced conformational change in cytochrome P-450CAM*. *Biochemistry*, 1993. **32**(17): p. 4571-8.
34. Chen, X., et al., *Crystal structure of the F87W/Y96F/V247L mutant of cytochrome P-450cam with 1,3,5-trichlorobenzene bound and further protein engineering for the oxidation of pentachlorobenzene and hexachlorobenzene*. *J Biol Chem*, 2002. **277**(40): p. 37519-26.
35. Poulos, T.L., B.C. Finzel, and A.J. Howard, *Crystal structure of substrate-free Pseudomonas putida cytochrome P-450*. *Biochemistry*, 1986. **25**(18): p. 5314-22.
36. Sevrioukova, I.F., et al., *Structure of a cytochrome P450-redox partner electron-transfer complex*. *Proc Natl Acad Sci U S A*, 1999. **96**(5): p. 1863-8.
37. Hasemann, C.A., et al., *Crystal structure and refinement of cytochrome P450terp at 2.3 Å resolution*. *J Mol Biol*, 1994. **236**(4): p. 1169-85.
38. Schlichting, I., C. Jung, and H. Schulze, *Crystal structure of cytochrome P-450cam complexed with the (1S)-camphor enantiomer*. *FEBS Lett*, 1997. **415**(3): p. 253-7.
39. Raag, R., et al., *Formation, crystal structure, and rearrangement of a cytochrome P-450cam iron-phenyl complex*. *Biochemistry*, 1990. **29**(35): p. 8119-26.
40. Schlichting, I., et al., *The catalytic pathway of cytochrome p450cam at atomic resolution*. *Science*, 2000. **287**(5458): p. 1615-22.
41. Lee, D.S., et al., *Structural characterization of n-butyl-isocyanide complexes of cytochromes P450nor and P450cam*. *Biochemistry*, 2001. **40**(9): p. 2669-77.
42. Di Gleria, K., et al., *Covalent Attachment of an Electroactive Sulfhydryl Reagent in the Active Site of Cytochrome P450cam as Revealed by the Crystal Structure of the Modified Protein*. *J Am Chem Soc*, 1998. **120**(1): p. 46-52.

43. Dunn, A.R., et al., *Probing the open state of cytochrome P450cam with ruthenium-linker substrates*. Proc Natl Acad Sci U S A, 2001. **98**(22): p. 12420-5.
44. Dmochowski, I.J., et al., *Optical detection of cytochrome P450 by sensitizer-linked substrates*. Proc Natl Acad Sci U S A, 1999. **96**(23): p. 12987-90.
45. Dunn, A.R., et al., *Fluorescent probes for cytochrome p450 structural characterization and inhibitor screening*. J Am Chem Soc, 2002. **124**(35): p. 10254-5.
46. Fedorov, R., D.K. Ghosh, and I. Schlichting, *Crystal structures of cyanide complexes of P450cam and the oxygenase domain of inducible nitric oxide synthase-structural models of the short-lived oxygen complexes*. Arch Biochem Biophys, 2003. **409**(1): p. 25-31.
47. Poulos, T.L. and A.J. Howard, *Crystal structures of metyrapone- and phenylimidazole-inhibited complexes of cytochrome P-450cam*. Biochemistry, 1987. **26**(25): p. 8165-74.
48. Raag, R., et al., *Crystal structure of the cytochrome P-450CAM active site mutant Thr252Ala*. Biochemistry, 1991. **30**(48): p. 11420-9.
49. Vidakovic, M., et al., *Understanding the role of the essential Asp251 in cytochrome p450cam using site-directed mutagenesis, crystallography, and kinetic solvent isotope effect*. Biochemistry, 1998. **37**(26): p. 9211-9.
50. Raag, R. and T.L. Poulos, *Crystal structure of the carbon monoxide-substrate-cytochrome P-450CAM ternary complex*. Biochemistry, 1989. **28**(19): p. 7586-92.
51. Raag, R. and T.L. Poulos, *The structural basis for substrate-induced changes in redox potential and spin equilibrium in cytochrome P-450CAM*. Biochemistry, 1989. **28**(2): p. 917-22.
52. Yeom, H., et al., *The role of Thr268 in oxygen activation of cytochrome P450BM-3*. Biochemistry, 1995. **34**(45): p. 14733-40.
53. Ost, T.W., et al., *Structural and spectroscopic analysis of the F393H mutant of flavocytochrome P450 BM3*. Biochemistry, 2001. **40**(45): p. 13430-8.

54. Haines, D.C., et al., *Pivotal role of water in the mechanism of P450BM-3*. *Biochemistry*, 2001. **40**(45): p. 13456-65.
55. Ost, T.W., et al., *Oxygen activation and electron transfer in flavocytochrome P450 BM3*. *J Am Chem Soc*, 2003. **125**(49): p. 15010-20.
56. Joyce, M.G., et al., *A single mutation in cytochrome P450 BM3 induces the conformational rearrangement seen upon substrate binding in the wild-type enzyme*. *J Biol Chem*, 2004. **279**(22): p. 23287-93.
57. Wester, M.R., et al., *Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 Å resolution: evidence for multiple substrate binding modes*. *Biochemistry*, 2003. **42**(21): p. 6370-9.
58. Wester, M.R., et al., *Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: evidence for an induced fit model of substrate binding*. *Biochemistry*, 2003. **42**(31): p. 9335-45.
59. Schoch, G.A., et al., *Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site*. *J Biol Chem*, 2004. **279**(10): p. 9497-503.
60. Williams, P.A., et al., *Crystal structure of human cytochrome P450 2C9 with bound warfarin*. *Nature*, 2003. **424**(6947): p. 464-8.
61. Williams, P.A., et al., *Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone*. *Science*, 2004. **305**(5684): p. 683-6.
62. Scott, E.E., et al., *Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-Å resolution: insight into the range of P450 conformations and the coordination of redox partner binding*. *J Biol Chem*, 2004. **279**(26): p. 27294-301.
63. Scott, E.E., et al., *An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution*. *Proc Natl Acad Sci U S A*, 2003. **100**(23): p. 13196-201.
64. Podust, L.M., T.L. Poulos, and M.R. Waterman, *Crystal structure of cytochrome P450 14 α -sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in*

- complex with azole inhibitors*. Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3068-73.
65. Podust, L.M., et al., *Estriol bound and ligand-free structures of sterol 14alpha-demethylase*. Structure (Camb), 2004. **12**(11): p. 1937-45.
66. Cupp-Vickery, J., R. Anderson, and Z. Hatziris, *Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity*. Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3050-5.
67. Cupp-Vickery, J.R., et al., *Ketoconazole-induced conformational changes in the active site of cytochrome P450eryF*. J Mol Biol, 2001. **311**(1): p. 101-10.
68. Nagano, S., et al., *Crystal structures of epothilone D-bound, epothilone B-bound, and substrate-free forms of cytochrome P450epoK*. J Biol Chem, 2003. **278**(45): p. 44886-93.
69. Kudo, T., et al., *A positively charged cluster formed in the heme-distal pocket of cytochrome P450nor is essential for interaction with NADH*. J Biol Chem, 2001. **276**(7): p. 5020-6.
70. Oshima, R., et al., *Structural evidence for direct hydride transfer from NADH to cytochrome P450nor*. J Mol Biol, 2004. **342**(1): p. 207-17.
71. Yano, J.K., et al., *Crystal structure of a thermophilic cytochrome P450 from the archaeon Sulfolobus solfataricus*. J Biol Chem, 2000. **275**(40): p. 31086-92.
72. Park, S.Y., et al., *Crystallization and preliminary X-ray diffraction analysis of a cytochrome P450 (CYP119) from Sulfolobus solfataricus*. Acta Crystallogr D Biol Crystallogr, 2000. **56** (Pt 9): p. 1173-5.
73. Yano, J.K., et al., *Preliminary characterization and crystal structure of a thermostable cytochrome P450 from Thermus thermophilus*. J Biol Chem, 2003. **278**(1): p. 608-16.
74. Lee, D.S., et al., *Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from Bacillus subtilis. Crystallographic, spectroscopic, and mutational studies*. J Biol Chem, 2003. **278**(11): p. 9761-7.

75. Zerbe, K., et al., *Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis*. J Biol Chem, 2002. **277**(49): p. 47476-85.
76. Pylypenko, O., et al., *Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis*. J Biol Chem, 2003. **278**(47): p. 46727-33.
77. Leys, D., et al., *Atomic structure of Mycobacterium tuberculosis CYP121 to 1.06 Å reveals novel features of cytochrome P450*. J Biol Chem, 2003. **278**(7): p. 5141-7.
78. Podust, L.M., et al., *Comparison of the 1.85 Å structure of CYP154A1 from Streptomyces coelicolor A3(2) with the closely related CYP154C1 and CYPs from antibiotic biosynthetic pathways*. Protein Sci, 2004. **13**(1): p. 255-68.
79. Cupp-Vickery, J.R. and T.L. Poulos, *Structure of cytochrome P450eryF: substrate, inhibitors, and model compounds bound in the active site*. Steroids, 1997. **62**(1): p. 112-6.
80. Ludemann, S.K., V. Lounnas, and R.C. Wade, *How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms*. J Mol Biol, 2000. **303**(5): p. 797-811.
81. Ludemann, S.K., V. Lounnas, and R.C. Wade, *How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways*. J Mol Biol, 2000. **303**(5): p. 813-30.
82. Li, H. and T.L. Poulos, *Fatty acid metabolism, conformational change, and electron transfer in cytochrome P-450(BM-3)*. Biochim Biophys Acta, 1999. **1441**(2-3): p. 141-9.
83. Stout, C.D., *Cytochrome p450 conformational diversity*. Structure (Camb), 2004. **12**(11): p. 1921-2.
84. Shou, M., et al., *A kinetic model for the metabolic interaction of two substrates at the active site of cytochrome P450 3A4*. J Biol Chem, 2001. **276**(3): p. 2256-62.

85. Orcutt BC, D.M., Barker WC., *ALIGN*. 1982, National Biomedical Foundation, Georgetown University Medical Center: Washington, DC.
86. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. *Nucleic Acids Res*, 1997. **25**(17): p. 3389-402.
87. Laskowski, R.A.M., M.W.; Moss, D. S. Thornton, J. M., *PROCHECK: A program to check the stereochemical quality of protein structures*. *Journal of Applied Crystallography*, 1993. **26**: p. 283-291.
88. Sippl, M.J., *Recognition of errors in three-dimensional structures of proteins*. *Proteins*, 1993. **17**(4): p. 355-62.
89. Vriend, G., *WHAT IF: a molecular modeling and drug design program*. *J Mol Graph*, 1990. **8**(1): p. 52-6, 29.
90. Afzelius, L., et al., *Structural analysis of CYP2C9 and CYP2C5 and an evaluation of commonly used molecular modeling techniques*. *Drug Metab Dispos*, 2004. **32**(11): p. 1218-29.
91. Korzekwa KR, J.J.a.G.J., *Theoretical studies on cytochrome P-450-mediated hydroxylation: a predictive model for hydrogen atom abstractions*. *J Am Chem Soc*, 1990. **112**: p. 7042-7046.
92. Grogan, J., et al., *Modeling cyanide release from nitriles: prediction of cytochrome P450 mediated acute nitrile toxicity*. *Chem Res Toxicol*, 1992. **5**(4): p. 548-52.
93. Tanii, H. and K. Hashimoto, *Studies on the mechanism of acute toxicity of nitriles in mice*. *Arch Toxicol*, 1984. **55**(1): p. 47-54.
94. Higgins, L., et al., *An assessment of the reaction energetics for cytochrome P450-mediated reactions*. *Arch Biochem Biophys*, 2001. **385**(1): p. 220-30.
95. Okita, R.T. and J.R. Okita, *Cytochrome P450 4A fatty acid omega hydroxylases*. *Curr Drug Metab*, 2001. **2**(3): p. 265-81.

96. Leach, A.R., *Molecular Modelling: Principles and Applications*. 2nd ed. 1996, Essex, England: Addison Wesley Longman.
97. Cramer, R.D., 3rd, D.E. Patterson, and J.D. Bunce, *Recent advances in comparative molecular field analysis (CoMFA)*. Prog Clin Biol Res, 1989. **291**: p. 161-5.
98. Khan, K.K., et al., *Midazolam oxidation by cytochrome P450 3A4 and active-site mutants: an evaluation of multiple binding sites and of the metabolic pathway that leads to enzyme inactivation*. Mol Pharmacol, 2002. **61**(3): p. 495-506.
99. de Groot, M.J., et al., *Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6*. J Med Chem, 1999. **42**(9): p. 1515-24.
100. Kuntz, I.D., et al., *A geometric approach to macromolecule-ligand interactions*. J Mol Biol, 1982. **161**(2): p. 269-88.
101. Rarey, M., et al., *A fast flexible docking method using an incremental construction algorithm*. J Mol Biol, 1996. **261**(3): p. 470-89.
102. Jones, G., et al., *Development and validation of a genetic algorithm for flexible docking*. J Mol Biol, 1997. **267**(3): p. 727-48.
103. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function*. Journal of Computational Chemistry, 1998. **19**: p. 1639-1662.
104. Kick, E.K., et al., *Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D*. Chem Biol, 1997. **4**(4): p. 297-307.
105. Harris, D.L. and G.H. Loew, *Investigation of the proton-assisted pathway to formation of the catalytically active, ferryl species of P450s by molecular dynamics studies of P450eryF*. J Am Chem Soc, 1996. **118**(27): p. 6377-87.
106. De Voss, J.J. and P.R. Ortiz de Montellano, *Substrate docking algorithms and the prediction of substrate specificity*. Methods Enzymol, 1996. **272**: p. 336-47.

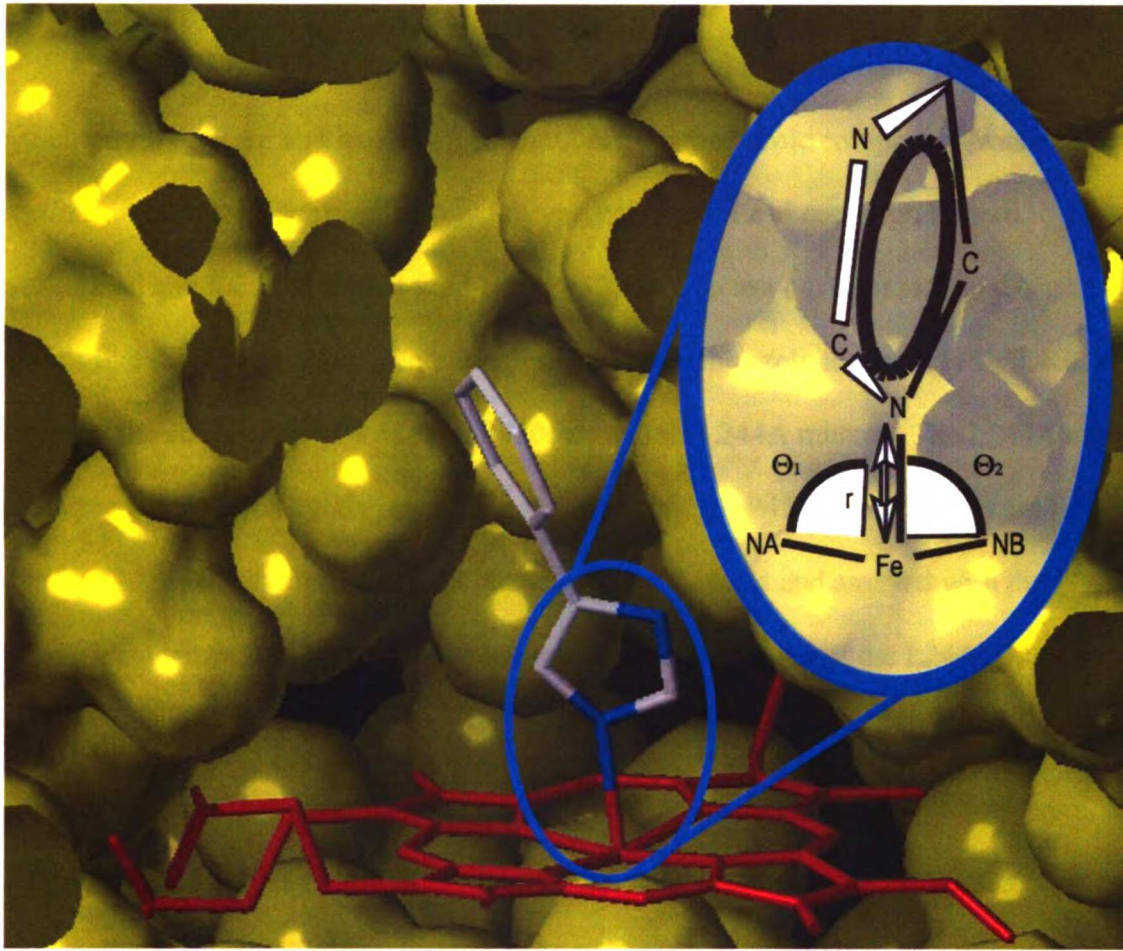
107. Zhang, Z., et al., *The substrate specificity of cytochrome P450cam*. *Bioorg Med Chem*, 1998. **6**(9): p. 1501-8.
108. Ha, S., et al., *Evaluation of docking/scoring approaches: a comparative study based on MMP3 inhibitors*. *J Comput Aided Mol Des*, 2000. **14**(5): p. 435-48.
109. Xing, L., et al., *Evaluation and application of multiple scoring functions for a virtual screening experiment*. *J Comput Aided Mol Des*, 2004. **18**(5): p. 333-44.
110. Kellenberger, E., et al., *Comparative evaluation of eight docking tools for docking and virtual screening accuracy*. *Proteins*, 2004. **57**(2): p. 225-42.
111. Kirton, S.B., et al., *Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins*. *Proteins*, 2005. **58**(4): p. 836-844.
112. Eldridge, M.D., et al., *Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes*. *J Comput Aided Mol Des*, 1997. **11**(5): p. 425-45.
113. Nissink, J.W., et al., *A new test set for validating predictions of protein-ligand interaction*. *Proteins*, 2002. **49**(4): p. 457-71.
114. Keseru, G.M., *A virtual high throughput screen for high affinity cytochrome P450cam substrates. Implications for in silico prediction of drug metabolism*. *J Comput Aided Mol Des*, 2001. **15**(7): p. 649-57.
115. Balakin, K.V., et al., *Kohonen maps for prediction of binding to human cytochrome P450 3A4*. *Drug Metab Dispos*, 2004. **32**(10): p. 1183-9.
116. Molnar, L. and G.M. Keseru, *A neural network based virtual screening of cytochrome P450 3A4 inhibitors*. *Bioorg Med Chem Lett*, 2002. **12**(3): p. 419-21.
117. Daly, A.K., et al., *Nomenclature for human CYP2D6 alleles*. *Pharmacogenetics*, 1996. **6**(3): p. 193-201.
118. Knegtel, R.M., I.D. Kuntz, and C.M. Oshiro, *Molecular docking to ensembles of protein structures*. *J Mol Biol*, 1997. **266**(2): p. 424-40.

119. Kuhn, B., et al., *Metabolism of sirolimus and its derivative everolimus by cytochrome P450 3A4: insights from docking, molecular dynamics, and quantum chemical calculations*. J Med Chem, 2001. **44**(12): p. 2027-34.
120. Kahan, B.D., *Efficacy of sirolimus compared with azathioprine for reduction of acute renal allograft rejection: a randomised multicentre study. The Rapamune US Study Group*. Lancet, 2000. **356**(9225): p. 194-202.
121. Yatscoff, R.W., et al., *Rapamycin: distribution, pharmacokinetics, and therapeutic range investigations*. Ther Drug Monit, 1995. **17**(6): p. 666-71.
122. Crowe, A., et al., *Absorption and intestinal metabolism of SDZ-RAD and rapamycin in rats*. Drug Metab Dispos, 1999. **27**(5): p. 627-32.
123. Venhorst, J., et al., *Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities*. J Med Chem, 2003. **46**(1): p. 74-86.
124. Shou, M., et al., *Activation of CYP3A4: evidence for the simultaneous binding of two substrates in a cytochrome P450 active site*. Biochemistry, 1994. **33**(21): p. 6450-5.
125. Korzekwa, K.R., et al., *Evaluation of atypical cytochrome P450 kinetics with two-substrate models: evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites*. Biochemistry, 1998. **37**(12): p. 4137-47.

Chapter 2

Computer Assisted Design of Selective Imidazole Inhibitors for Cytochrome P450 enzymes.

Submitted to J. Med. Chem.



LIBRARY
UNIVERSITY OF CALIFORNIA

Abstract

A modified version of the DOCK program has been used to predict inhibitors of cytochrome P450cam and its L244A mutant. A library of azole compounds was designed *in silico* and screened for binding to wild-type P450cam. Lead compounds were synthesized and found to inhibit wild-type P450cam. To test our approach to design ligands that discriminate between closely related sites, the azole library was DOCKed into both the active sites of wild-type P450cam and its L244A mutant. The L244A active site is predicted to be slightly larger than that of wild-type P450cam. Ligands predicted to have a high affinity for the mutant alone were synthesized and assayed with the recombinant enzymes. All of the compounds showed inhibition of the L244A enzyme (IC_{50} 6-40 μ M) and the compounds that were predicted to be too large to bind to the wild-type showed poor inhibition ($IC_{50} \geq 1$ mM). The binding mode was shown to be similar to that predicted by our modified version of DOCK by spectroscopic analysis. A discrepancy between the IC_{50} values and spectroscopic K_s values indicates that the spectroscopic binding constants do not accurately estimate inhibitory activity. This study, the first report of computer assisted ligand (drug) design for P450 enzymes in which the coordination bond between imidazole and the heme is explicitly considered in structural modeling, opens a promising design avenue because azole compounds are widely used as P450 enzyme inhibitors and drugs.

Introduction

Cytochrome P450 enzymes, which catalyze the monooxygenation of a broad diversity of substrates, play essential roles in both the metabolism of xenobiotics and the biosynthesis and catabolism of endogenous lipophilic factors [1]. Inhibition of P450 enzymes, and therefore of the synthesis of endogenous factors, is a promising avenue for the development of therapeutic agents. Two examples are the inhibition of estradiol synthesis by aromatase in estrogen-dependent breast cancer, and the possible control of retinoic acid levels through inhibition of P450-dependent retinoic acid metabolism [2, 3]. The inhibition of P450-dependent xenobiotic metabolism is also potentially desirable in situations in which xenobiotics are converted to toxic or carcinogenic metabolites [4]. In a different context, the rapid selection of drug candidates that minimize the potential for P450-based drug-drug interactions is becoming increasingly important in the pharmaceutical industry.

Conventional drug design approaches have led to the development of a range of P450 inhibitors that are currently used in the clinic, primarily as antifungal agents. These agents include miconazole, ketoconazole, and fluconazole. It is likely that structure-based drug design methods will soon be applicable to the design of P450-targeted agents, as X-ray crystallography is beginning to yield structural information on relevant enzymes. In recent years, the structures of *Mycobacterium tuberculosis* sterol 14-demethylase, membrane bound mammalian P450 enzymes, and even human P450 enzymes have been reported [5-7]. With this influx of structural data, computational methods that exploit

this information and methods that are easily translatable across various P450 enzymes must be developed.

I report the application of computational design methods for the generation of high affinity inhibitors for cytochrome P450cam. In order to determine the ability of the approach to discriminate between closely related active sites, I have expressed the L244A mutant, which only differs from the parent enzyme by three carbon atoms, and then generated inhibitors that preferentially target the mutant. While P450cam is a bacterial enzyme of little interest as a drug target, it has been well characterized and represents an ideal system in which to test and refine the necessary approaches. The prediction of inhibitors for P450 enzymes presents a particular challenge because of the relatively low specificity of P450 active sites. In previous efforts, I have made some progress in predicting the substrate specificity of P450cam [8]. However, this is the first instance in which such methods have been applied to the design of potential inhibitors of a structurally characterized P450 enzyme.

Results and Discussion

Virtual Library Design and a Modified DOCKing Algorithm. An initial virtual library of ligands was generated to screen as potential inhibitors for cytochrome P450cam. The choice of an imidazole scaffold was based upon the knowledge that imidazoles and other aromatic nitrogen heterocycles that coordinate to the iron offer an effective route to inhibition of P450 enzymes [9]. Past research into heme binding imidazoles indicates that substitution is limited to two positions [10]. Thus, variation of

the substituents at the 1 and 5 positions allows for the introduction of active site contacts and isoform specificity, whereas substitutions at the 2 or 4 positions introduce steric clashes with the heme moiety that preclude coordination of the imidazole to the heme iron atom. The nitrogen at position 3 of the imidazole must remain unsubstituted as it is the atom that actually coordinates to the iron (Fig. 1). Substituents for the virtual library were taken from the ACD (Available Chemical Database) as described in the experimental section. A monosubstituted imidazole library was then constructed with the substituent exclusively on the 1N position, as previous work suggested that a penalty exists for desolvating the unsubstituted 1N position [11]. However, the greatest library diversity was provided by the 1,5-disubstituted structures, which comprised 2550 of the total of 3508 structures.

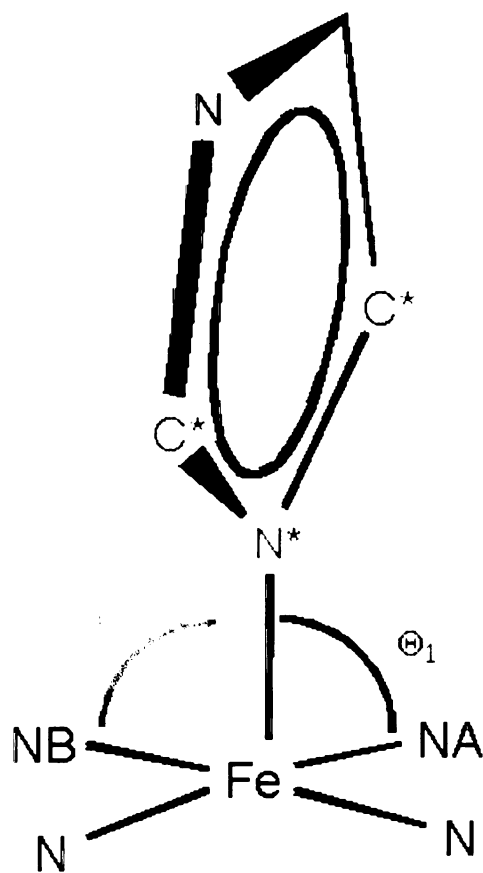


Figure 1. Covalent DOCK parameters. R is the distance from the heme iron to the imidazole N3. Θ_1 and Θ_2 are the angles as measured from nitrogens on the heme, to the heme iron, to imidazole N3. Atoms marked with an (*) have their van der Waal's contributions negated.

1875

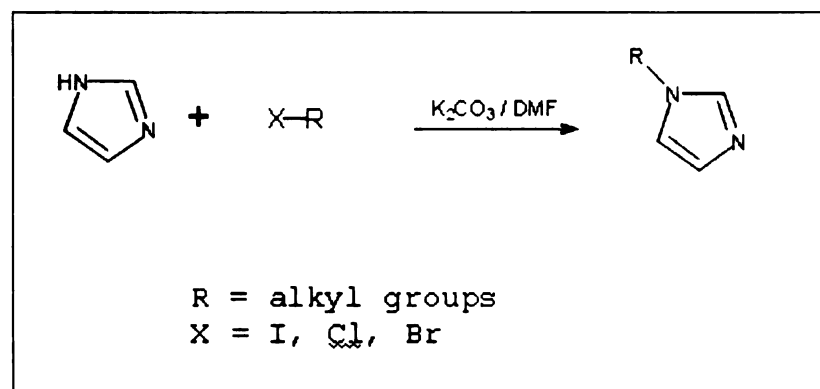
1875

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

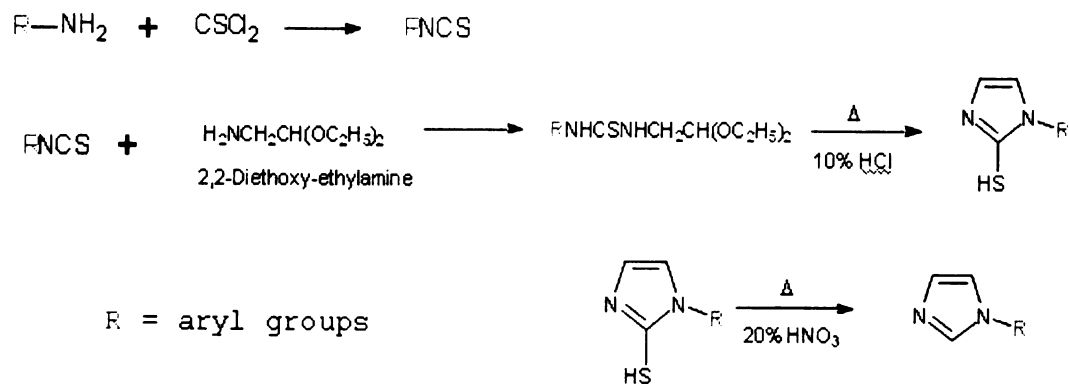
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

While substitution positions were based upon our knowledge of imidazole-heme chelation, the substituents were chosen based upon their synthetic feasibility. Three synthetic paths were chosen to yield 1-alkyl (Scheme 1) [12], 1-aryl (Scheme 2) [13], and 1,5-dialkyl (Scheme 3) [14] substituted imidazoles. By fabricating a synthetic library based on only three reaction pathways, we hoped to facilitate the synthesis of any of the compounds and insure that they could all be synthesized by a small number of reaction steps. The selection, charging, and joining of the fragments is detailed in the experimental section.

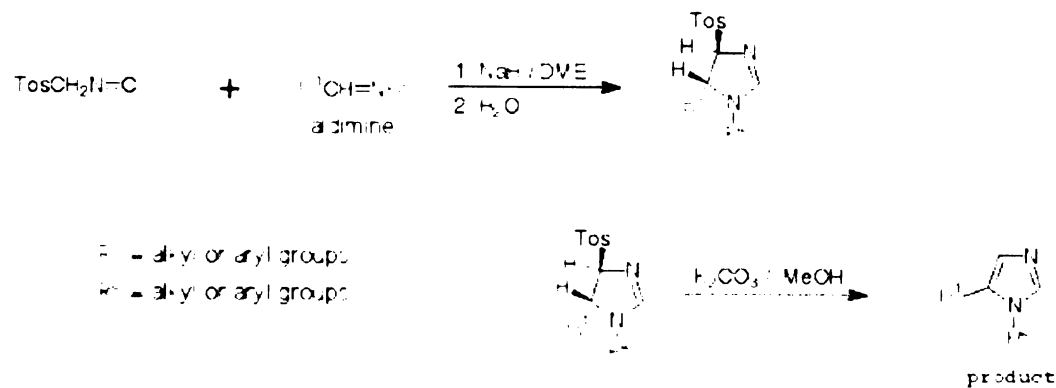
Scheme 1. Synthesis of 1-alkyl imidazoles.



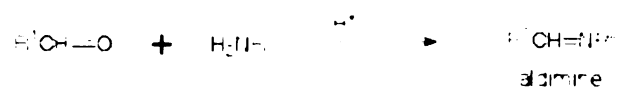
Scheme 2. Synthesis of 1-aryl imidazoles.



Scheme 3. Synthesis of 1,5-disubstituted imazoles.



Aldimine Synthesis



Because DOCK does not recognize the coordinate imidazole-iron bond, initial attempts to DOCK the library produced unreasonable inhibitor orientations. Even when the ligand was preoriented correctly, the correct orientation was lost during minimization. To overcome this problem, Covalent DOCK (CovDOCK) was used. By introducing a penalty for long imidazole N3-heme iron distances or improper angles, a binding mode was achieved that more correctly mimicked the binding observed in crystal structures of imidazole-P450 complexes (Fig. 2). The binding of imidazole-P450 complexes diverges to some extent from any fixed distance and angle [11, 15, 16], so the harmonic restraints were deliberately set to low values to allow for this variation.



Figure 2. CovDOCK imidazole orientations. The number one (blue), fifty (green), and one-fifty (red) ranked compounds when CovDOCKed against wild-type P450cam. The side chains have been stripped to better illustrate the placement of the imidazole relative to the heme.

The best ligand among the initial hits had a CovDOCK score of -15.9, as compared to the value of -30 expected for a strong ligand hit. The relatively low scores, however, were not due to the harmonic penalties used in CovDOCK but were a consequence of the small active site and the lack of electrostatic interactions. It is also to be noted that the van der Waals' interactions of three imidazole atoms (C2, N3, and C4) are ignored in order to allow for a quasi-covalent imidazole-iron distance. This further decreases the effective size of the ligands and their overall CovDOCK score. A break down of the

score contributions taken as absolute values to show their effective contributions to the overall CovDOCK score is illustrated in Figure 3. CovDOCK is able to return expected binding modes but does not dominate the score while doing so.

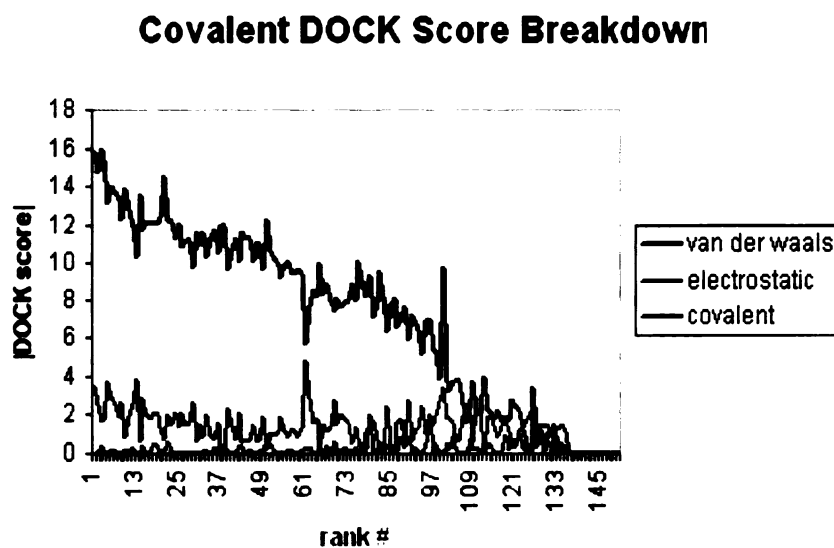
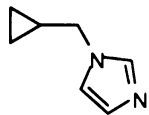
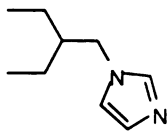
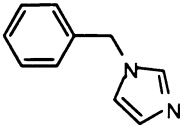
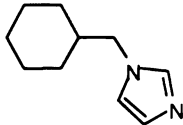


Figure 3. CovDOCK score components. All the components of the score have been shown in absolute value to better compare their magnitude.

Inhibition of P450cam. As might be expected, the hits pulled from the library were primarily small ligands. The 200 top scoring hits were clustered by Daylight fingerprints

(Daylight Comp., Santa Fe, NM), which effectively arranges them by functionality and not by 3D shape or conformation. A preference for monosubstituted imidazoles is immediately apparent in the preferred structure list despite the fact that monosubstituted ligands represent less than one third of the total virtual library. Representatives of several attractive clusters from among the top 200 ranked ligands were selected for experimental evaluation and were synthesized. Ligands were chosen for synthesis by manually inspecting the top 200 ranked compounds and selecting ligands that had good binding modes and maintained an imidazole-heme coordinate character similar to that found in the crystallographic data. The ability of the compounds to inhibit the oxidation of camphor to 5-*exo*-hydroxy-camphor was measured by gas chromatographic analysis of the product to starting material ratio. All of the mono-substituted compounds showed good inhibition of wild-type P450cam; most had IC₅₀ values in the low micromolar range (Table 1).

Table 1. Monosubstituted IC₅₀ Values and CovDOCK Scores^a

	IC ₅₀ μ M	CovDOCK scores
	4 \pm 2.0	-6.23
	0.8 \pm .3	-2.48
	16 \pm 11	-6.97
	1.5 \pm .5	-9.51

^a Monosubstituted ligands synthesized and their IC₅₀ values and CovDOCK scores against wild-type P450cam.

The mono-substituted compounds were then assayed against the L244A mutant and the results, as expected, indicated a preference for the wild-type. The results are shown on a log scale to better illustrate the differences in inhibitory activity between the wild type and mutant (Fig. 4). There is more space created by the L244A mutation that is not exploited by the monosubstituted compounds. While being able to generate high affinity

inhibitors in a first round synthesis is promising in itself, we were particularly interested in generating compounds selective for the L244A mutant by rational design.

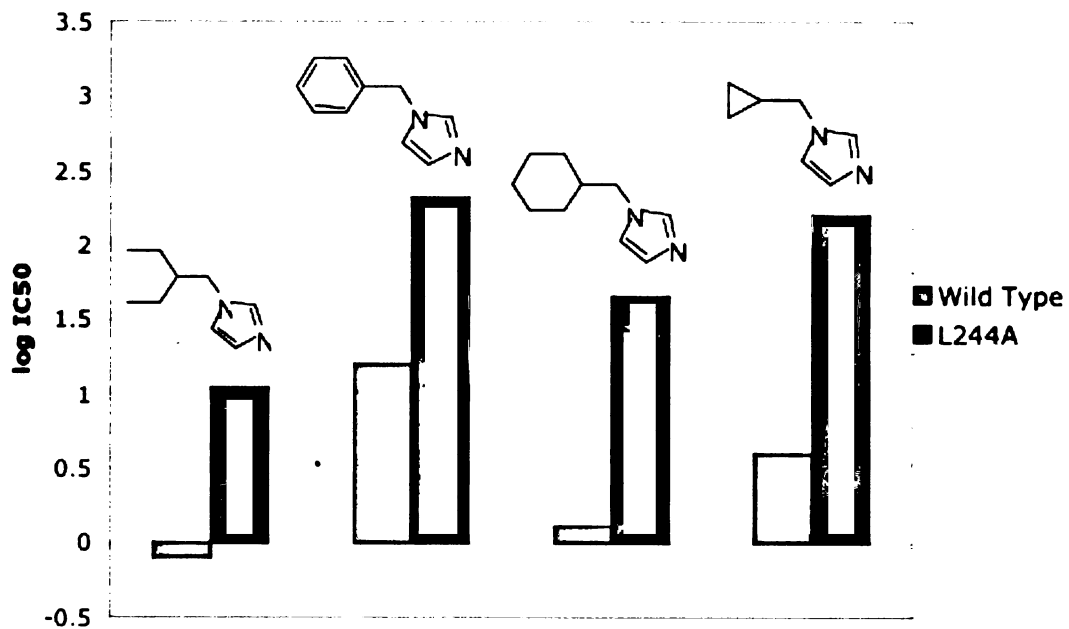
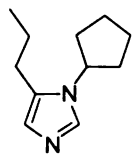
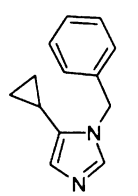
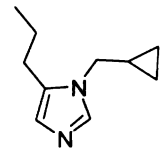
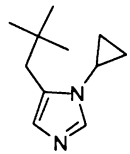
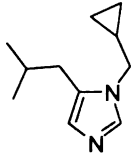


Figure 4. Monosubstituted IC₅₀ selectivity. Monosubstituted compound IC₅₀ values assayed against wild-type and the L244A mutant.

We used CovDOCK to search for inhibitors predicted to exhibit a high affinity for the L244A mutant and, simultaneously, a mitigated affinity for wild-type P450cam. The

L244A mutation was made in the 1PHD pdb structure using Sybyl and a grid was generated for CovDOCKing without minimization. The original virtual library was then covalently DOCKed into this L244A active site. In an effort to generate selective inhibitors, the best hits against the mutant were taken and CovDOCKed into wild-type P450cam. Large differences in the CovDOCK score (>50) were assumed to represent active site clashes and thus a potential poor fit. By comparing the CovDOCK results between the two active sites, we hoped to find the characteristics of ligands that would inhibit the L244A isozyme selectively. Gschwend and Kuntz [17] used a similar approach previously, predicting selectivity by computational screening between two strains of dihydrofolate reductase. Table 2 illustrates the scores for several of our compounds against the L244A mutant and the difference between the scores for the two forms of the enzyme. A major difference between compounds predicted to be inhibitors for the L244A vs. the wild-type is quickly apparent in the second substituent on the imidazole at the 5 position. While these disubstituted compounds comprise two thirds of the library they were not predicted to fit into the wild-type and when they were redocked into the wild-type they were found to have high scores due to steric clashes. Notably, the second substituent on the ligands predicted to bind to the L244A is generally small, in agreement with the small active site change due to mutating a leucine to an alanine. Based upon these results, the disubstituted compounds were predicted to be selective inhibitors and

Table 2. Second Round CovDOCK Results ^a

	RANK	Score Diff.		RANK	Score Diff.
	216	-195.1		195	-10.22
	57	-47.18		25	-73.15
	29	-49.99			

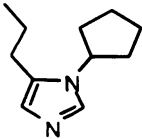
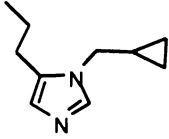
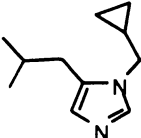
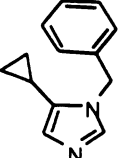
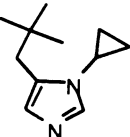
^a Disubstituted compounds rank against the L244A mutant as scored by CovDOCK and the difference in score when the same compound is Covalently DOCKed back into the wild-type enzyme.

five ligands were chosen for synthesis. The criteria for selection included good imidazole-heme coordination, a good CovDOCK score against the L244A mutant, and a poor score against the wild-type.

The disubstituted compounds, as predicted, have good affinity for the L244A protein (Table 3). When assayed against the wild-type they show almost no inhibition and even at a 1 mM concentration none of the disubstituted compounds completely inhibited catalytic activity. These results are very promising, as they suggest that we were able to predictively identify ligands complementary to the L244A active site but slightly too large to be able to inhibit the wild-type.

Spectroscopic Analysis. Binding of the compounds was evaluated spectrophotometrically. To insure a similar starting state, a difference spectrum was taken in which inhibitors were titrated into solution in the absence of camphor. All of the inhibitors exhibited a Type II spectral shift in which the water-coordinated Soret maximum at 417 nm decreased while the 424 nm imidazole-bound peak intensified. This shift is consistent with binding of a primary or aromatic nitrogen as the sixth ligand to the iron [18]. Figure 5, which illustrates the difference spectra obtained with 1-cyclopentyl-5-propylimidazole, is typical of all of spectra used to determine the K_s

Table 3. Disubstituted IC₅₀ Values^a

	L244A IC ₅₀ (μM)	WTYPE %Inhibition at 1mM
	6.5 ± 0.4	60
	36 ± 7	60
	10.1 ± 0.5	20
	41 ± 6	30
	129 ± 45	80

^a Disubstituted compounds synthesized and their inhibition of L244A and wild-type P450cam enzymes

values. A clear isosbestic point is observed and the transition to 424 nm is saturable.

The spectroscopic data agrees with the binding mode predicted by CovDOCK.

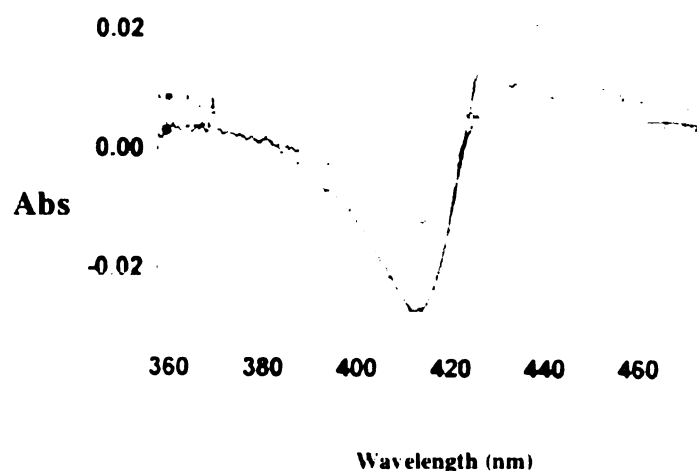


Figure 5. Difference spectrum. The difference spectrum of 1-cyclopentyl-5-propylimidazole. This Type II shift is seen with all the inhibitors.

The L244A P450cam is very similar to wild-type P450cam spectroscopically: both have a 450 nm peak in the reduced CO-bound state, and both enzymes absorb in the ferric state at the same wavelength, both in the high spin (390 nm; no sixth ligand) and low spin (416 nm; water coordinated) states. Furthermore the affinity for camphor, the natural substrate, remains unchanged (wild-type $K_s = 1.3 \pm 0.2 \mu\text{M}$, L244A $K_s = 1.5 \pm 0.2$

μM). The L244A mutant, as previously reported, is catalytically viable and, though it turns over camphor more slowly, exhibits a degree of uncoupling nearly identical to that of the wild-type enzyme [19]. This means that the NADH and O_2 consumed by the enzyme are used to the same extent in hydroxylation of camphor rather than in the production of H_2O_2 or H_2O . Nevertheless, an important difference was found in the affinity of the two forms of the enzyme for imidazole. The K_s of wild-type P450cam for unsubstituted imidazole is $7.5 \pm 1.4 \mu\text{M}$, whereas that of the L244A mutant is $304 \pm 49 \mu\text{M}$. This 43-fold difference was unexpected since the spectroscopic, substrate binding, catalytic, and uncoupling data suggest that the wild-type and mutant active sites are very similar. The loss of affinity for imidazole presumably reflects active site changes such as an increase in the active site water content, possible changes in the positioning of water molecules, and small perturbations of active site residues that do not alter the protein spectrum or its natural affinity for camphor. Reports of azole resistance amongst fungal strains are increasing [20, 21] and the L244A mutant may provide a useful model in this regard. Furthermore, the data indicate that our disubstituted inhibitors are able to compensate for the imidazole scaffold's mitigated affinity for the L244A active site.

The K_s values for the monosubstituted compounds against both the L244A and wild-type P450 enzymes are shown in Figure 6. A general trend toward selectivity is evident and expected given the difference in base affinities of the scaffolds. Surprisingly, the K_s values for the disubstituted compounds do not reflect the high selectivity observed in the IC_{50} assays (Fig. 7). While the IC_{50} assays are done in the presence of camphor ($250 \mu\text{M}$) it is still surprising that such a large discrepancy exists. In theory, if the disubstituted compounds are able to coordinate to the iron, they should prevent electron

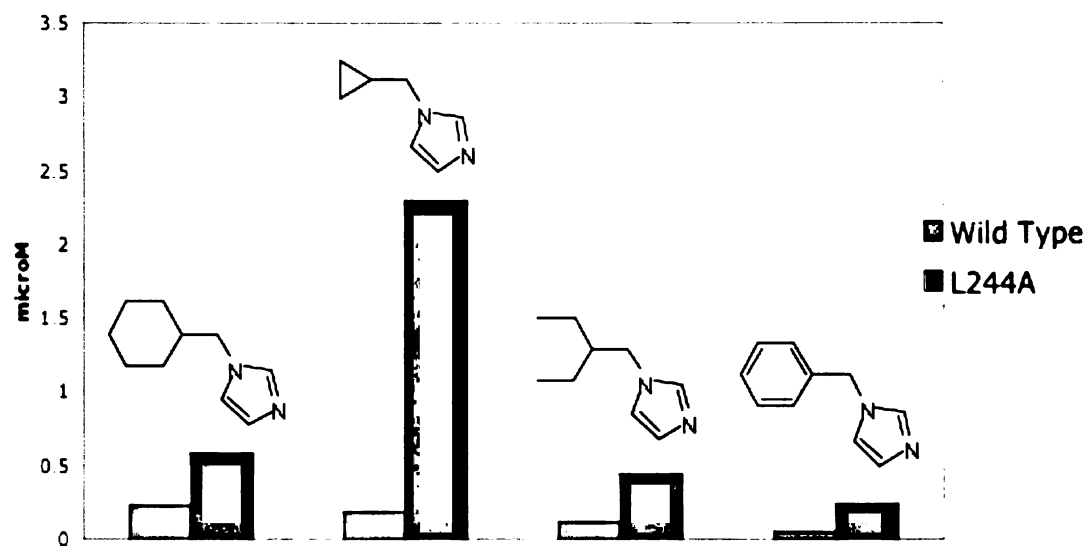


Figure 6. K_s values of the monosubstituted compounds. K_s values against wild-type P450cam and its L244A mutant.

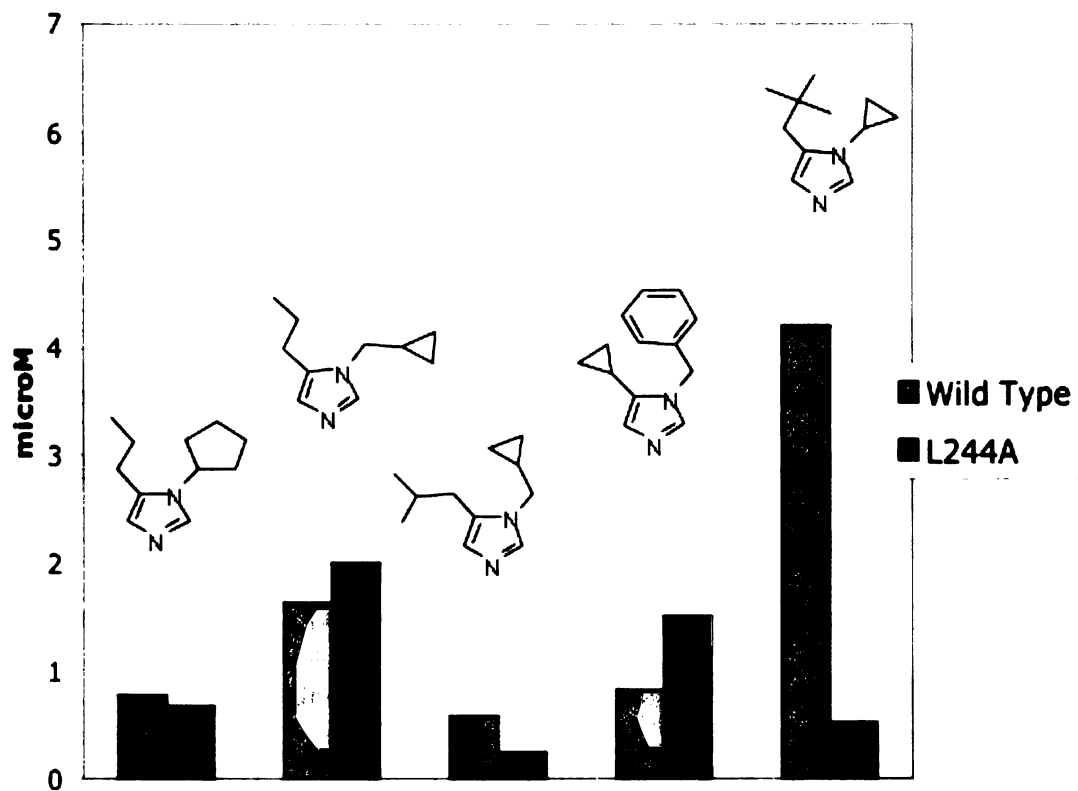


Figure 7. K_s values of the disubstituted compounds. K_s values against wild-type P450cam and its L244A mutant

transfer and catalysis; however, even at inhibitor concentrations up to 1 mM, activity is still seen in the presence of the disubstituted ligands.

Exploring the intricacies of these anomalies is continued further in Chapter 4 with kinetic analysis of the di- and mono-substituted inhibitors and again in Chapter 5 by crystallographic methods, as it is clear that this observation is related to a requirement of conformational changes in the P450cam active site. Support for this argument is provided by measurements of K_s values of ligands predicted to be too large to fit into either active site. Thus, 1,5-dicyclohexyl imidazole binds to both the wild-type ($K_s 90 \pm 55 \mu\text{M}$) and the L244A mutant ($K_s 20 \pm 4 \mu\text{M}$) even though, based on a rigid model of the protein, this large, disubstituted compound should be unable to bind to either the wild-type or L244A active site. Even more surprisingly ketoconazole was found to bind with high affinity to both the wild-type ($K_s 0.5 \pm 0.1 \mu\text{M}$) and L244A mutant ($K_s 0.6 \pm 0.1 \mu\text{M}$). No crystal structure of P450cam exists that suggests an active site anywhere large enough to accommodate this ligand. Thus a conformational change must occur.

The exact nature of the conformational changes, and the reason for the low inhibitory activity of some compounds that have low spectroscopic binding constants (K_s values), remain to be defined. One possibility is that the large inhibitors require a slow conformational change and are thus slow-binding, reversible inhibitors that adhere to the following kinetic model [22]:



However, we have not observed the time dependence in the K_s value determinations implied by this model. Thus, although conformational changes are clearly involved in the discrepancy between the K_s and IC_{50} values, definition of the precise relationship between these values requires further exploration.

Experimental Section

Protein Expression and Binding Assays

Cytochrome P450cam and the P450cam L244A mutant were expressed in *Escherichia coli* as previously reported [23]. Both proteins were stored with 100 μ M camphor at -70 °C. Camphor was removed before the proteins were used by twice eluting with 100 mM sodium phosphate through a PD-10 size exclusion column. The P450cam thus obtained has a water coordinated to the heme iron and is in the low spin state. Spectroscopic binding (K_s) assays were done in 100 mM potassium phosphate buffer at 37 °C by difference UV spectroscopy. The two cuvettes contained 0.5 μ M protein in 300 μ L of 100 mM potassium phosphate buffer. The potential inhibitors were dissolved in DMSO, with the exception of imidazole which was dissolved in water. DMSO never constituted more than 2% of the final sample volume. An equal amount of DMSO was added to the reference cuvette and the shift in absorbance from 416 to 424 nm as increasing concentrations of camphor were titrated into the cuvette was observed. The difference between 416 and 424 nm values was then recorded and fitted to a binding curve to obtain K_s values. K_s values were calculated by non-linear regression analysis using the equation $\Delta A = (\Delta A_{\max}[I])/(K_s + [I])$ where ΔA is the 416 trough to 424 peak

absorbance difference, ΔA_{max} is the difference in absorption at saturation, and $[I]$ is the inhibitor or ligand concentration [24].

The compounds reported in this study were synthesized except for imidazole, 1,5-dicyclohexyl imidazole, and ketoconazole. Imidazole and 1,5-dicyclohexyl imidazole were purchased from Aldrich (St. Louis, MO), and ketoconazole was obtained from Jansen Pharmaceuticals (Beerse, Belgium).

Gas Chromatographic Inhibition Assay and IC_{50} calculation. Catalytic activity was quantitated by gas chromatographic product analysis. The protein was eluted twice through a PD-10 column with 100 mM sodium phosphate to remove camphor. Assays were done under the following conditions: P450cam 5 nM, putidaredoxin 2 μM , putidaredoxin reductase 2 μM , camphor 250 μM , and NADH 1000 μM . The L244A protein was assayed at a concentration of 20 nM. The total reaction volume was 150 μL . Varying amounts of inhibitor were added in DMSO, but the DMSO volume never exceeded 3% of the total reaction volume. All assays were done in 100 mM potassium phosphate in a 20° C bath for 15 min. The reactions were then quenched with 20 μL of CH_2Cl_2 containing 2 mM β -thujone as an internal standard.

The CH_2Cl_2 was then withdrawn directly from the reaction mixture and injected onto an HP 5980 Series II Gas Chromatograph equipped with an Agilent Technologies (Palo Alto, Ca.) DB-10 capillary column. A temperature gradient from 40° to 280° C over 20 min was used. The β -thujone standard eluted at approximately 6 min, starting material eluted at approximately 6.2 min and product appeared at 9.9 min. The

difference between the product to starting material ratio was then used to calculate turnover. IC_{50} values were calculated by the Langmuir Isotherm equation [25]:

$$1 + (1 + (IC_{50}/[Inh]))).$$

Organic Synthesis. Two rounds of synthesis were done. Initially all the imidazoles synthesized were N-monoalkylated and were prepared as previously reported [12]. The syntheses of N-benzylimidazole and N-cyclopropylmethylimidazole are described here as typical procedures.

N-Benzylimidazole. Imidazole (100 mg, 1.4 mmol) was suspended in 5 mL of acetone under argon in a round bottom flask. Dry, powdered KOH (392.7 mg, 7 mmol) was added while stirring. The solution became beige. After 3 minutes a 190 μ L aliquot of benzyl bromide (1.57 mmol) was added dropwise and a white precipitate appeared. The acetone was removed on a rotary evaporator and the precipitate was resuspended in CH_2Cl_2 . Water was added, the solution pH was lowered to \sim 4 with concentrated HCl, and the product was extracted in the aqueous layer. The pH of the aqueous layer was then raised to \sim 10 with concentrated NaOH and the product was extracted with CH_2Cl_2 and crystallized directly from CH_2Cl_2 . The purity and identity were judged by TLC (ethyl acetate +1% triethylamine, R_f 0.1), mass spectrometry, and NMR: 17% yield; m.p. 73-74 $^{\circ}C$ (reported 72-74 $^{\circ}C$); 1H NMR ($CDCl_3$) 7.52 (1H, s, imidazole), 7.08 (1H, s, imidazole), 6.88 (1H, s, imidazole), 5.18 (2H, s, NCH_2), 7.14 benzyl (2H, d, aryl, $J=8.0$), 7.33 ppm (2H, m, aryl), and 7.32 (1H, m, aryl); ^{13}C NMR (DMSO) 137.4, 136.18, 129.83, 128.99, 128.27, 127.28, and 50.8; *EI-m/z* (low res.), 159.25 (M + H).

1-Methylcyclopropylimidazole. Imidazole (100 mg, 1.4 mmol) was suspended and deprotonated as described above. A 230 μ L aliquot of bromomethylcyclopropane (1.57 mmol) was added dropwise and the solution became a darker shade of beige. The reaction was worked up as described above and, after an acid-base wash, the solution was concentrated by rotary evaporation to give a yellow oil that was deemed sufficiently pure by TLC (ethyl acetate +1% triethylamine, R_f 0.1), mass spectrometry, and NMR for direct study. 50% yield, ^1H NMR (CDCl_3) 7.55 (1H, s, imidazole), 7.06 (1H, s, imidazole), 6.99 (1H, s, imidazole), 3.79 (2H, d, NCH_2 , $J=6.8$), 1.2 (1H, m, cyclopropyl), 0.68 ppm (2H, q, cyclopropyl, $J=8.0$), and 0.36 (2H, q, cyclopropyl, $J=4.2$); ^{13}C NMR (DMSO) 137.37, 128.62, 119.75, 50.85, 12.58, 4.08; *EI-m/z*, exp. 122.0843 (M^+) (calc. 122.0844, $\text{C}_7\text{H}_{10}\text{N}_2$).

The second round of ligands prepared as potential inhibitors of the P450cam L244A mutant were all 1,5-disubstituted compounds. These were prepared by a two-step synthesis utilizing the synthon tosylmethyl isocyanide [14]. The yields were generally low, less than 10%, but the reactions afforded sufficient product for the spectroscopic and kinetic assays. The synthesis of 1-benzyl-5-cyclopropylimidazole is given as a typical procedure.

1-Benzyl-5-cyclopropyl imidazole. To \sim 2 mL of dry molecular sieves was added 10 mL of anhydrous CH_2Cl_2 followed by benzylamine (1.46 mL, 13.1 mmol) and subsequently by dropwise addition of cyclopropylcarboxaldehyde (1 mL, 13.1 mmol). The disappearance of the aldehyde peak and appearance of the imine peak were monitored by infrared spectroscopy. After 2 h the reaction was filtered and rotary evaporated to give a clear yellow product that was used as obtained in the following step.

Dry potassium carbonate (9.55 g, 50 mmol) was added to tosylmethyl isocyanide (1.99 g, 10 mmol) suspended in 15 mL of anhydrous MeOH followed by dropwise addition of the imine product from the previous reaction. The resulting solution was allowed to stir for 24 h and was then filtered, the filtrate was subjected to rotary evaporation, and the residue was extracted thrice with ether (3 x 20 mL). The product was purified by flash chromatography on silica with a solvent system consisting initially of 300 mL of 50/50 (v/v) of hexane/ethyl acetate followed by 80/20 (v/v) of ethyl acetate/methanol. The desired product is one of the final substances to be eluted due the protonated imidazole nitrogen. Finally, the product was partitioned between acid and base as described above. Rotary evaporation of the solution then yielded a yellow oil that was pure by TLC (ethyl acetate/methanol 80/20, R_f 0.2) and NMR.

The following compounds were made by the appropriate procedure described above:

1-Methylcyclohexylimidazole: 22% yield; ^1H NMR (CDCl_3) 7.34 (1H, s, imidazole), 6.97 (1H, s, imidazole), 6.79 (1H, s, imidazole), 3.65 (2H, d, NCH_2 , $J=8.2$), 1.5-1.7 ppm (6H, m, cyclohexyl), and 1.05-1.19 ppm (5H, m, cyclohexyl); ^{13}C NMR (DMSO) 137.83, 128.11, 119.80, 52.01, 36.10, 29.82, 25.86, 25.09; *EI-m/z* exp. 164.1313 (M^+) (calc. 164.1313, $\text{C}_{10}\text{H}_{16}\text{N}_2$).

1-(2-Ethylbutyl)imidazole: 56% yield; ^1H NMR (CDCl_3) 7.42 (1H, s, imidazole), 7.03 (1H, s, imidazole), 6.85 (1H, s, imidazole), 3.81 (2H, d, NCH_2 , $J=8$), 1.63 (1H, septuplet, 2-ethylbutyl), 1.27 (4H, pentuplet, 2-ethylbutyl, $J=9.4$), and 0.87 ppm (6H, t, t-

CH₃, J=8.2); ¹³C NMR (DMSO) 137.63, 128.13, 119.71, 49.04, 36.10, 22.57, 10.34; *EI-m/z* exp. 152.1338 (M⁺) (calc. 152.1313, C₉H₁₆N₂).

1-Cyclopentyl-5-propylimidazole: 3% yield; ¹H NMR (CDCl₃) 7.49 (1H, s, imidazole), 6.79 (1H, s, imidazole), 4.33 (1H, pentuplet, cyclopentyl, J=8.0), 2.52 (2H, t, propyl, J=8.0), 1.6-1.95 (10H, m) and 1.05 (3H, t, propyl, J=7.2); ¹³C NMR (CDCl₃) 133.58, 131.72, 125.5, 55.72, 33.47, 26.27, 23.76, 21.45, 13.8; *EI-m/z* exp. 178.1469 (M⁺) (calc. 178.1470, C₁₁H₁₈N₂).

1-Methylcyclopropyl-5-propylimidazole: 11% yield; ¹H NMR (CDCl₃) 7.65 (1H, s, imidazole), 6.77 (1H, s, imidazole), 3.62 (2H, d, NCH₂, J=6.8), 2.53 (2H, t, propyl, J=8.0), 1.76, (2H, sextet, propyl, J=7.2), 1.21 (1H, m, cyclopropyl), 0.98 (3H, t, propyl, J=7.2) 0.34 (2H, q, cyclopropyl, J=5.2) and 0.24 (2H, q, cyclopropyl, J=4.4); ¹³C NMR (CDCl₃) 136.07, 132.62, 125.70, 49.46, 23.46, 19.27, 14.16, 10.91, 4.14; *EI-m/z* exp. 164.1313 (M⁺) (calc. 164.1313, C₁₀H₁₆N₂).

1-Methylcyclopropyl-5-butylimidazole: 13% yield; ¹H NMR (CDCl₃) 7.53 (1H, s, imidazole), 6.77 (1H, s, imidazole), 3.66 (2H, d, NCH₂, J=6.8), 2.41 (2H, d, butyl, J=7.2), 1.88 (1H, nonatet, butyl, J=6.8), 1.16 (1H, m, cyclopropyl), 0.66 (2H, q, cyclopropyl, J=8.4), and 0.33 (2H, q, cyclopropyl, J=4.8); ¹³C NMR (CDCl₃) 136.01, 130.32, 126.66, 49.04, 32.97, 27.81, 22.32, 11.12, 4.08; *EI-m/z* exp. 178.1470 (M⁺) (calc. 178.1470, C₁₁H₁₈N₂).

1-Benzyl-5-cyclopropylimidazole: 6% yield; ¹H NMR (CDCl₃) 7.42 (1H, s, imidazole), 6.84 (1H, s, imidazole), 5.15 (2H, s, methyl), 7.23-7.34 (5H, m, benzyl), 0.91 (1H, m, cyclopropyl), 0.85 (2H, m, cyclopropyl), and 0.68 ppm (2H, q, cyclopropyl,

J=4.4); ^{13}C NMR (CDCl_3) 137.161, 136.44, 134.24, 128.70, 128.40, 127.74, 125.72, 48.31, 5.51, 4.34; *EI-m/z* exp. 198.1153 (M^+) (calc. 198.1157, $\text{C}_{13}\text{H}_{14}\text{N}_2$).

1-Cyclopropyl-5-*tert*-butylimidazole: 9% yield; ^1H NMR (CDCl_3) 7.41 (1H, s, imidazole), 6.77 (1H, s, imidazole), 3.09 (1H, m, cyclopropyl), 2.13 (2H, s, *tert*-butyl), 1.0 (9H, s, *tert*-butyl CH_3), and 0.85-1.05 (4H, m, cyclopropyl); ^{13}C NMR (CDCl_3) 136.54, 129.62, 127.87, 37.20, 30.78, 29.32, 26.04, 6.25; *EI-m/z* exp. 178.1467 (M^+) (calc. 178.1470, $\text{C}_{11}\text{H}_{18}\text{N}_2$).

Library Design

Because azole compounds have long been known for their utility as P450 inhibitors [9], we chose an imidazole scaffold for this project. To create the virtual library, the imidazole scaffold was modeled in Sybyl (Tripos Inc., St. Louis, Mo.) and substituents were selected from the Available Chemical Database (ACD) (Molecular Design Systems, San Leandro, Ca.) based upon potential synthetic routes for creating 1-alkyl imidazoles [12], 1-aryl imidazoles [13], and 1,5-disubstituted imidazoles [14]. The corresponding primary and secondary halide, aniline, aldehyde, and primary amine reactants were extracted from the ACD using Merlin (Daylight Comp., Santa Fe, N.M.) and UCSelect (G, Skillman, PhD thesis, UCSF). Our primary criterion for reactant selection was synthetic feasibility. A virtual library of 582 primary and secondary alkyl halides, 378 anilines, 51 aldehydes, and 50 primary amines was compiled. The fragments were joined to their appropriate position on the imidazole scaffold using Sybyl. The final library contained 3508 compounds: 479 N-alkylated imidazoles, 376 N-aryl imidazoles, and

2550 1,5-disubstituted imidazoles. The compounds were then charged using Gasteiger charges from Sybyl.

Covalent DOCK. DOCK 4.0 [26] was employed to screen the library against cytochrome P450cam. Grid coordinates were taken from the 1PHD imidazole complexed structure determined by Poulos [11]. The charges for the heme were taken from the semi-empirical ZINDO calculations of Gilda Loew [27].

Covalent DOCK (CovDOCK) (G, Skillman, PhD thesis, UCSF), originally written to apply to serine proteases, was modified to model the imidazole-iron coordinate bond. CovDOCK adds harmonic restraints to the DOCK scoring function (eq.1):

$$\text{CovDOCK score} = k_1(\text{dist})^2 + k_2(\text{angleA})^2 + k_3(\text{angle B})^2; k_1 = 50, k_2 \text{ and } k_3 = .01 \text{ (eq.1)}$$

The three restraints used in this work were: a distance restraint from the iron to the imidazole N3 and two orthogonal angle restraints from nitrogens NA and NB on the heme, to the iron, to the imidazole N3 (Fig. 1). The van der Waals contributions of the nitrogen coordinating the imidazole to the heme as well as the carbons adjacent to the heme were set to zero. Removing the van der Waals contributions of these atoms was necessary to allow for the close contacts observed in crystal structures of P450 imidazole inhibitor complexes.

Conclusions

We have shown that it is possible, using structure-based design and currently available docking algorithms, to generate ligands of high affinity for both cytochrome P450cam and, more importantly, its L244A mutant. The CovDOCK method and azole libraries offer promise as a facile, modular system easily applicable to crystallographically characterized P450 enzymes. The library construction was based upon synthetic feasibility and is not limited to the 3500 compounds actually included in this study. Indeed, the library has recently been expanded to 4500 compounds (as detailed in chapter 3) and is completely amenable to the inclusion of alternative P450-coordinating scaffolds such as triazoles, pyridines, and other heterocycles, all of which would significantly increase the chemical diversity of the library.

Based upon the IC_{50} data, I was able to exploit small differences in the active sites of these two isozymes and synthesize two distinct groups of ligands with selectivity for either enzyme. The results are promising and, with the accelerating pace of determination of P450 crystal structures and the swift screening methods made possible by computational advances, one can envision that dozens, even scores of enzymes could be DOCKed to in an effort to generate isoform selectivity.

Although P450cam can hydroxylate a variety of non-natural substrates [18], the enzyme has a single natural substrate (camphor) and has clearly evolved to optimize its oxidation. P450cam has generally been thought to have a relatively rigid active site

because crystal structures of the enzyme with various substrates and inhibitors show only minor adjustments of the active site cavity [11, 28, 29]. Of course, it is evident that the protein must undergo substantial conformational changes to allow the entry of substrates into the active site cavity, as no entry channel is seen in the crystal structure. The spectroscopic results demonstrate that conformational mobility, even in an enzyme such as P450cam, is an important parameter in identifying ligands that differentiate between two similar P450 active sites. Also, while spectroscopic assays are frequently the initial test of binding to cytochrome P450 enzymes, our results suggest they can be misleading in inhibitor prediction. Studies measuring spectroscopic binding to microsomal P450 enzymes and *in vivo* activity have shown discrepancies between K_s and I_{50} values [30]. Our results suggest that discrepancies may also exist in single enzyme studies and that caution must be employed when screening ligands for inhibitory potential by spectroscopic means.

References

1. Ortiz de Montellano, P.R., *Cytochrome P450: Structure, Mechanism, and Biochemistry*. 2nd ed, ed. P.R. Ortiz de Montellano. 1995, New York: Plenum Press.
2. Fujii, H., et al., *Metabolic inactivation of retinoic acid by a novel P450 differentially expressed in developing mouse embryos*. *Embo J*, 1997. **16**(14): p. 4163-73.
3. Guengerich, P.F., et al., *Cytochrome P450 1B1: a target for inhibition in anticarcinogenesis strategies*. *Mutat Res*, 2003. **523-524**: p. 173-82.
4. McLemore, T.L., et al., *Metabolic activation of 4-ipomeanol in human lung, primary pulmonary carcinomas, and established human pulmonary carcinoma cell lines*. *J Natl Cancer Inst*, 1990. **82**(17): p. 1420-6.
5. Podust, L.M., et al., *Substrate recognition sites in 14alpha-sterol demethylase from comparative analysis of amino acid sequences and X-ray structure of Mycobacterium tuberculosis CYP51*. *J Inorg Biochem*, 2001. **87**(4): p. 227-35.
6. Williams, P.A., et al., *Crystal structure of human cytochrome P450 2C9 with bound warfarin*. *Nature*, 2003. **424**(6947): p. 464-8.
7. Williams, P.A., et al., *Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity*. *Mol Cell*, 2000. **5**(1): p. 121-31.
8. Zhang, Z., et al., *The substrate specificity of cytochrome P450cam*. *Bioorg Med Chem*, 1998. **6**(9): p. 1501-8.
9. Testa, B. and P. Jenner, *Inhibitors of Cytochrome P-450s and their mechanism of action*. *Drug Metab Rev*, 1981. **12**(1): p. 1-117.
10. Rogerson, T.D., C.F. Wilkinson, and K. Hetarski, *Steric factors in the inhibitory interaction of imidazoles with microsomal enzymes*. *Biochem Pharmacol*, 1977. **26**(11): p. 1039-42.

11. Poulos, T.L. and A.J. Howard, *Crystal structures of metyrapone- and phenylimidazole-inhibited complexes of cytochrome P-450cam*. *Biochemistry*, 1987. **26**(25): p. 8165-74.
12. Grimmet, M.R., *Comprehensive Heterocyclic Chemistry*. 1984, Oxford: Pergamon Press.
13. Johnson, A.L., et al., *The synthesis of 1-arylimidazoles, a new class of steroid hydroxylation inhibitors*. *J Med Chem*, 1969. **12**(6): p. 1024-8.
14. Leusen, J.A.W., J.; Oldenziel, O., *Base-Induced cycloaddition of sulfonylmethyl isocyanides to C,N double bonds. Synthesis of 1,5-disubstituted and 1,4,5-trisubstituted imidazoles from aldimines and imidoyl chlorides*. *Journal of Organic Chemistry*, 1977. **42**: p. 1153-1159.
15. Podust, L.M., T.L. Poulos, and M.R. Waterman, *Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors*. *Proc Natl Acad Sci U S A*, 2001. **98**(6): p. 3068-73.
16. Yano, J.K., et al., *Crystal structure of a thermophilic cytochrome P450 from the archaeon Sulfolobus solfataricus*. *J Biol Chem*, 2000. **275**(40): p. 31086-92.
17. Gschwend, D.A., et al., *Specificity in structure-based drug design: identification of a novel, selective inhibitor of Pneumocystis carinii dihydrofolate reductase*. *Proteins*, 1997. **29**(1): p. 59-67.
18. Schenkman, J.B., H. Remmer, and R.W. Estabrook, *Spectral studies of drug interaction with hepatic microsomal cytochrome*. *Mol Pharmacol*, 1967. **3**(2): p. 113-23.
19. De Voss, J.J. and P.R. Ortiz de Montellano, *Substrate docking algorithms and the prediction of substrate specificity*. *Methods Enzymol*, 1996. **272**: p. 336-47.
20. Kelly, S.L., et al., *The G464S amino acid substitution in Candida albicans sterol 14alpha-demethylase causes fluconazole resistance in the clinic through reduced affinity*. *Biochem Biophys Res Commun*, 1999. **262**(1): p. 174-9.

21. Kelly, S.L., et al., *An old activity in the cytochrome P450 superfamily (CYP51) and a new story of drugs and resistance*. *Biochem Soc Trans*, 2001. **29**(Pt 2): p. 122-8.
22. Morrison, J.F. and C.T. Walsh, *The behavior and significance of slow-binding enzyme inhibitors*. *Adv Enzymol Relat Areas Mol Biol*, 1988. **61**: p. 201-301.
23. Wagner, G.C., et al., *Apoprotein formation and heme reconstitution of cytochrome P-450cam*. *J Biol Chem*, 1981. **256**(12): p. 6262-5.
24. Rao, S.I., et al., *The lipoxygenase activity of myoglobin. Oxidation of linoleic acid by the ferryl oxygen rather than protein radical*. *J Biol Chem*, 1994. **269**(10): p. 7210-6.
25. Copeland, R.A., *Enzymes*. 2nd ed. 2000, New York: Wiley:VHC.
26. Ewing, T.J., et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. *J Comput Aided Mol Des*, 2001. **15**(5): p. 411-28.
27. Collins, J.R.C., D.L.; Loew, G.H., *Valproic acid metabolism by cytochrome P450: a theoretical study of stereoelectronic modulators of product distribution*. *Journal of the American Chemical Society*, 1991. **113**: p. 2736-2743.
28. Poulos, T.L., B.C. Finzel, and A.J. Howard, *High-resolution crystal structure of cytochrome P450cam*. *J Mol Biol*, 1987. **195**(3): p. 687-700.
29. Raag, R. and T.L. Poulos, *Crystal structures of cytochrome P-450CAM complexed with camphane, thiocamphor, and adamantane: factors controlling P-450 substrate hydroxylation*. *Biochemistry*, 1991. **30**(10): p. 2674-84.
30. Lavrijsen, K., et al., *Induction potential of antifungals containing an imidazole or triazole moiety. Miconazole and ketoconazole, but not itraconazole are able to induce hepatic drug metabolizing enzymes of male rats at high doses*. *Biochem Pharmacol*, 1986. **35**(11): p. 1867-78.

Chapter 3

Expansion of the In Silico Imidazole Library and Application of Covalent DOCK to *M. tuberculosis* CYP51

Prologue to Chapter 3

Establishing that CovDOCK could produce good inhibitors for P450cam was promising, however, the imidazole library and CovDOCK were ultimately intended to be used in designing viable drug candidates. Collaborators at Vanderbilt University solved the crystal structure of a P450 enzyme, CYP51 from *Mycobacterium tuberculosis*, that could putatively be targeted to generate drug candidates. They provided the coordinates and we agreed to collaborate to generate inhibitors for this system of potential utility in the treatment of tuberculosis or that might contribute to our understanding of the treatment of this disease.

I was excited to apply my methods to this system, but also excited to be working on this project as it was my first real collaboration. Xiaoti Zhou, a post doctoral researcher in our lab, was to do the synthetic work and Galina Lepesheva, in Dr. Waterman's lab at Vanderbilt, was to carry out the assays. My job was to use CovDOCK to screen our virtual library of imidazole compounds and try to predict compounds to synthesize.

Dr. Zhou left UCSF shortly after the project began, but we did get through one round of compound development and succeed in finding some micromolar ligands for Cyp51. Most of the synthetic work presented here is his and the spectroscopic binding assays were done by our collaborators at Vanderbilt.

Abstract

The versatility of CovDOCK and the applicability of the imidazole library are shown by their application to CYP51. We move from a proof-of-principle system (P450cam) to a viable drug target, a P450 involved in *Mycobacterium tuberculosis* cell wall synthesis necessary for the bacteria's survival. The imidazole library is expanded to take advantage of specific interactions with the new target and retroactively expanded again after initial organic synthesis revealed problems with the solubility of our newly added imidazole compounds. While the active sites, apart from the heme, vary significantly between the test protein (P450cam) and the new target (Cyp51), the methods are easily amenable to change and several hits are predicted. Spectroscopic UV measurements show that the ligands possess micromolar affinities.

Introduction

Solution of the structure of P450 sterol 14 α -demethylase, or CYP51, in 2001 afforded the first structurally characterized P450 enzyme that was also a viable drug target [1]. CYP51 catalyzes the oxidative elimination of the 14 α methyl group of lanosterol, dihydrolanosterol, and obtusifoliol [2-4]. CYP51 is critical in the biosynthetic pathways essential for growth and survival of *Mycobacterium tuberculosis*. In *Mycobacterium bovis* and *Mycobacterium smegmatis*, two mycobacteria which are closely related to *M. tuberculosis*, azole impaired growth has been shown to correlate with azole spectral affinity for CYP51 [5, 6].

Tuberculosis infection has been cited as the major cause of death in developing countries [7]. After decades of a declining incidence of tuberculosis infection in the United States, HIV/AIDS infection and multi-drug resistant strains resulted in a strong resurgence of the infection in the late 1980's and early 1990's [8].

The treatment regime for tuberculosis most often involves a three-drug cocktail that includes isoniazid and is administered for over six months [9]. Drug resistant strains are frequently observed, including strains resistant to isoniazid [10-12]. Developing novel therapeutics thus remains essential to combating tuberculosis in the United States and beyond.

For us, this enzyme system was an excellent target against which to apply our CovDOCK methods and screen our virtual library. While the CYP51 active site is distinct from that of P450cam, the library should be applicable to new enzymes and easily modified to exploit specific characteristics of CYP51. There are two structures

available for CYP51, both of which are complexed with imidazole or triazole-based inhibitors [1]. We show that, using our methods, we are able to predict inhibitors for CYP51 that bind with low micromolar affinity. The best predicted ligand has an affinity of 18.8 μM .

Results and Discussion

The overall structure of *Mycobacterium tuberculosis* CYP51 is similar to P450cam and all P450 enzymes characterized to date. It contains a long I-helix running down the central axis of the protein, the topology of the helices and beta-pleated regions are similar to that of P450cam, and it has a buried active site. As is often the case with P450 enzymes, clear distinctions between P450 families are only apparent in the active site.

CYP51 has been crystallized in the presence of 4-phenylimidazole (4-PIM) and fluconazole (Fig. 1). We used DOCK 4.0 [16] to generate spheres so as to better visualize the active sites. While the structures of both the compounds complexed with the protein are very similar, we find that the active site topology varies significantly depending on which inhibitor is present. 4-PIM opens a cavity to the right of the inhibitor (Fig. 1a) even though it does not occupy that volume. When observed from the same angle, we see that fluconazole inserts its triazole substituent into the channel on the left, lengthening that channel.

Because DOCK grids rigidly represent the active site that is selected, the choice of which crystal structure to use will heavily affect our results. We chose to work with

the 4-PIM co-crystal structure for several reasons. The 4-PIM co-crystal structure is about 400 Å³ larger than that of the fluconazole bound structure. The larger active site may represent a more open conformation that would allow us to explore more ligands and larger ligands. The 4-PIM structure is slightly better resolved to 2.1 Å vs. 2.21 Å for the fluconazole structure. Finally, the 4-PIM structure was solved by co-crystallization, while the fluconazole bound complex was solved by soaking substrate free crystals with the ligand.

First Round CovDOCK Results

The original unmodified library was screened using CovDOCK. The library was reoriented by superimposing the imidazole moieties of each ligand over the imidazole of the co-crystallized 4-PIM using Sybyl and CovDOCKed without further preparation. Initial CovDOCK scores ranged from -16.8 to -7.5 over the top 150 ranked ligands. The ligands selected were all 1-N monosubstituted ligands in agreement with the known ligands for CYP51 (fluconazole, 4-PIM, ketoconazole) that are also monosubstituted (Fig. 2).

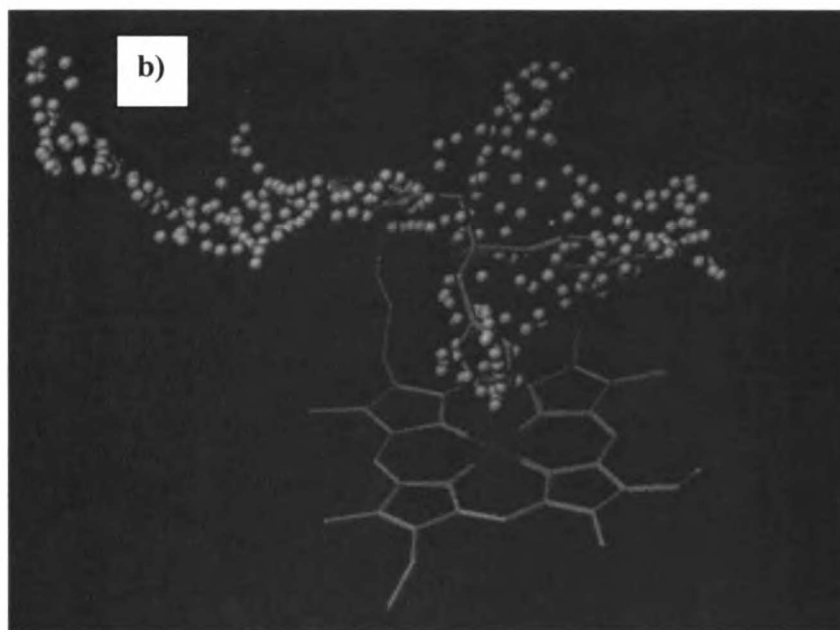
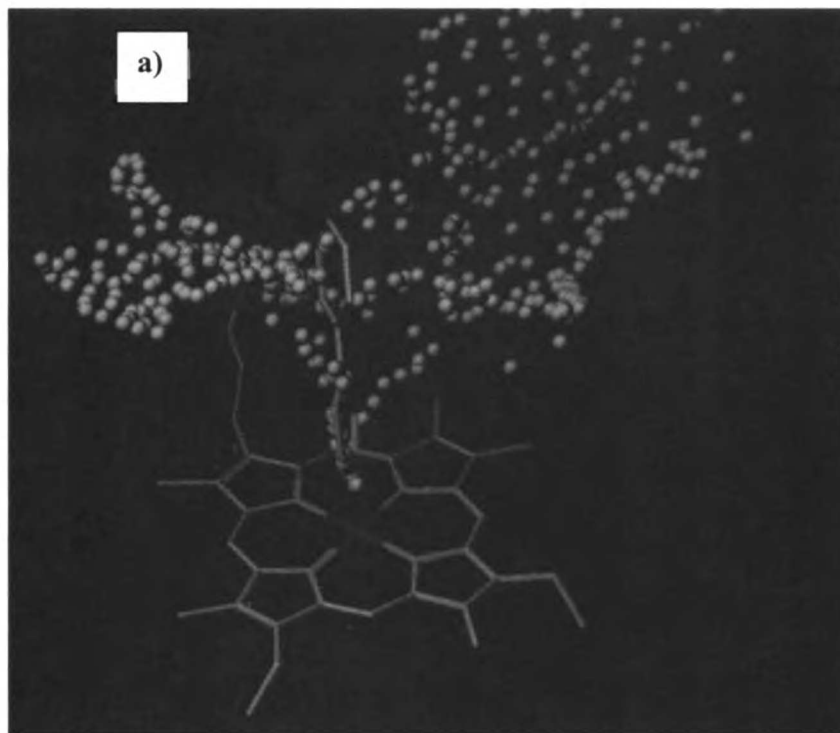


Figure 1. CYP51 active site. The CYP51 active sites represented by spheres (Spheregen). The structures correspond with a) 4-phenyl imidazole and b) fluconazole co-crystallized CYP51.

When juxtaposing the top ranked compounds for CYP51 with the top ranked compounds for P450cam, we find that the ligands for the former are generally larger. This is in agreement with the larger active site of CYP51 relative to P450cam. It is somewhat surprising to see that the top predicted ligands score slightly worse despite their larger size. The top 150 ligands for P450cam range in scores from -18 to -6 and are smaller in size. Because the scoring scheme in DOCK is based on an additive force field and because we make no correction for the number of atoms in a ligand, we expect to find that larger ligands generally score better. By observing the break down of the CovDOCK score we are able to shed light on this anomaly (Fig. 3). The overall scores in figure three have been converted to absolute values to better illustrate their individual contributions to the overall score, but for the top compounds the Van der Waals and electrostatic contributions should be a negative value. The covalent penalty, on the other hand, is simply a harmonic restraint (see Chapter 2) that drives the correct binding orientation and is always a positive or penalizing score. For P450cam, the overall contribution of the covalent score was generally small, about 1; however, when CovDOCK is applied to CYP51 we find that the covalent score is a significant portion of the overall score, hovering around 7 for most of the top ranked ligands.

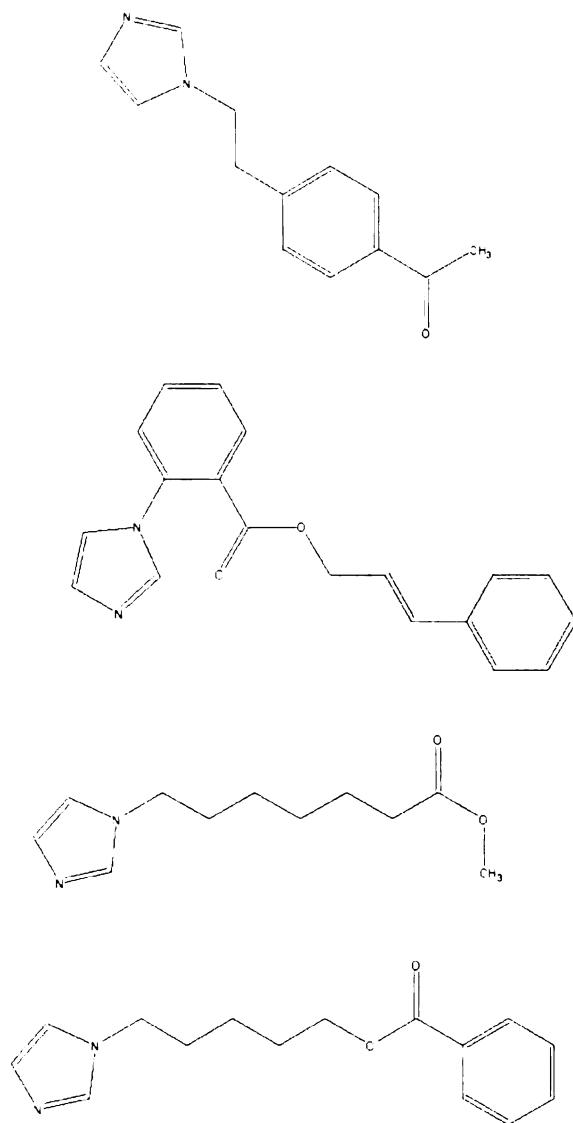


Figure 2. Sample hits from the first round of CovDOCK. From top to bottom the 1st, 10th, 20th, and 30th ranked ligands. Note all the compounds selected are 1-N monosubstituted imidazoles.

CovDOCK score break down first round CYP51

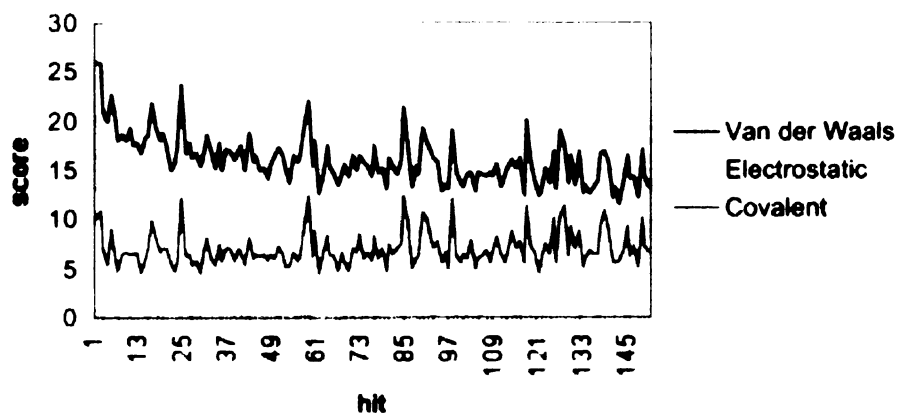


Figure 3. CovDOCK score break down. The score contributions are represented as absolute values to better compare their magnitude. As is typical of many P450 active sites, the predominant forces are Van der Waals interactions; however, the covalent contribution to the score is not negligible.

To further deconvolute these unusually high covalent dock scores, 4-PIM was docked back into the structure without minimization or orientational sampling. DOCK returned a large score, 146.9, indicating a steric clash. 4-PIM was then broken into fragments, an imidazole and phenyl ring, and the fragments were DOCKed individually returning scores of 156.33 and -10.3 respectively. CovDOCK negates the Van der Waals radii of the three atoms closest the heme iron on imidazole (C2, N3, and C4) to allow for the close distances observed between imidazole and heme during coordination. Closer inspection revealed a clash between C5 and threonine 206 in CYP51. Such close contacts in crystal structure coordinates are not uncommon and are often due to choice of force field during the crystallographic refinement. To circumvent the problem, the Van der Waals radii of the offending atoms in Thr206 were halved and the DOCK grids were recomputed.

First Library Expansion

One characteristic particular to the CYP51 binding site is a histidine residue in close proximity to the heme iron which is able to hydrogen bond to the ligands. In P450cam, 1-PIM is a superior inhibitor because the substituent on the 1-N nitrogen partially offsets the cost of desolvating that free nitrogen. For CYP51, however, we find that 4-PIM is a better inhibitor. CYP51 exploits the unsubstituted 1N nitrogen by the formation of a hydrogen bond to histidine 257 (Fig. 4).

To take advantage of this hydrogen bond, the library of 3508 imidazole compounds was supplemented with 4-substituted imidazoles. As with the original library design, synthetic feasibility was a priority and the substituents were chosen from the

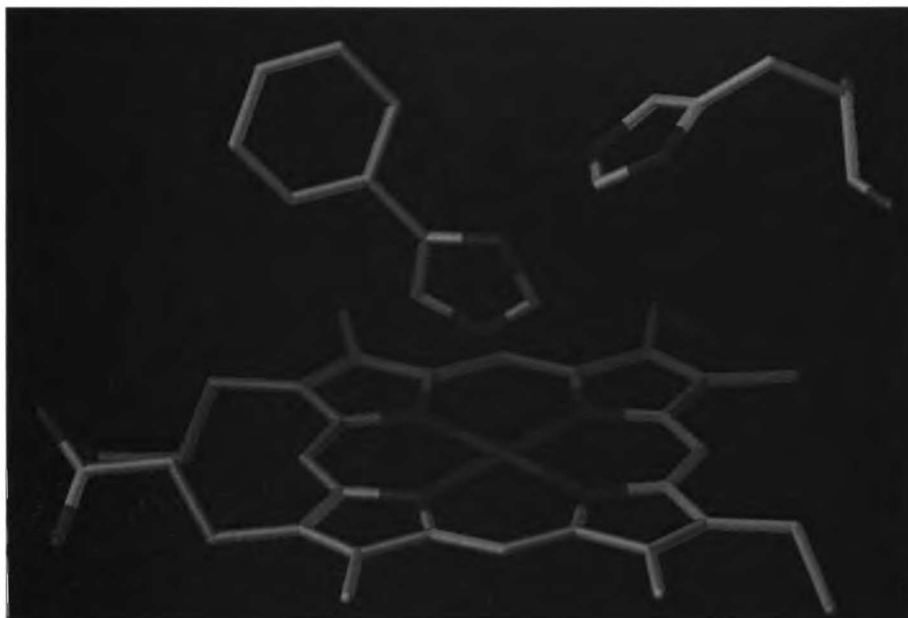


Figure 4. Active site of CYP51 with bound 4-PIM. Histidine 257 is able to hydrogen bond to the unsubstituted imidazole nitrogen.

ACD (Available Chemical Database) based on Scheme 1. Using UC Select [17] 323 primary amines were filtered from the ACD. The fragments were joined in Sybyl and charged with Gasteiger charges.

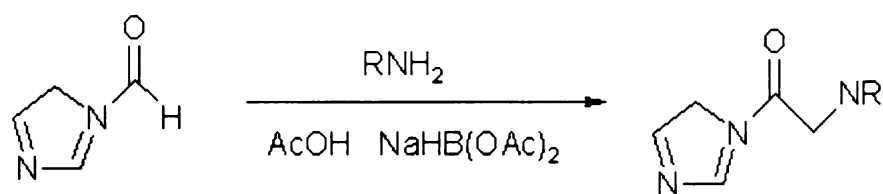
Second Round CovDOCK Results

With the inclusion of more singly substituted imidazole compounds and the corrected Van der Waals radii for the Thr206, a second round of CovDOCK was done. Immediate improvements were evident in both the overall scores (-27.85 to -18.95 for the top 150 hits) and better imidazole-heme geometries. The covalent contribution to the overall score also dropped to about 2. It is also notable that while the newly added 4-substituted imidazoles comprise less than 10% of the library they represent more than 80% of the top 150 ranked hits (Fig. 5).

Organic Synthesis of Second Round Hits

From the top 150 hits, 4 compounds were chosen by eye for synthesis (Fig. 6). The four compounds were chosen to show some diversity of shape and size, but were also nominated because of the availability of their reactants. 1-Aminobenzylimidazole was also synthesized, though it did not appear in the hit list. While monitoring the synthesis of 1-aminobenzylimidazole by TLC (thin layer chromatography) it became clear that the

Scheme 1. Synthesis of amine linked 4-substituted imidazoles



R = alkyl

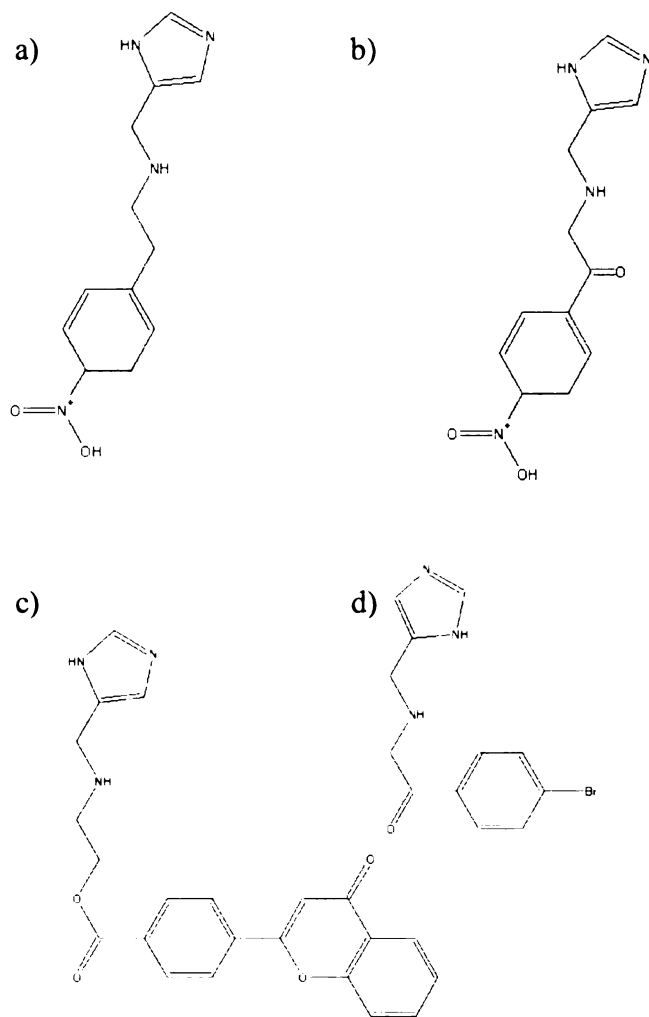


Figure 5. Second round CovDOCK hits. Sample hits from the second round of covDOCK with the supplemented library. a) 1st b) 10th c) 20th and d) 30th ranked hits.

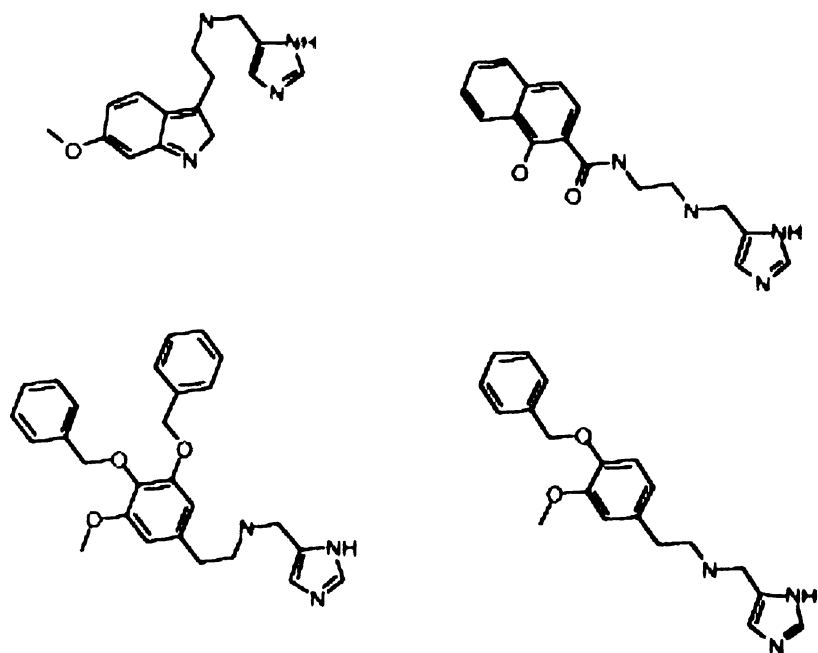


Figure 6. Ligands chosen for synthesis

resulting secondary amine product was positively charged (data not shown). They did not migrate significantly on the TLC plate and, as a result, were difficult to purify. In retrospect this was not a surprising find. The pKa values of secondary amines would lead to a protonated species at neutral pH. The active site of CYP51, like most P450 enzymes, is primarily hydrophobic. We thus discarded the library of imidazoles substituted through an amine linker and began the design of a second library.

Second Library Expansion

To circumvent the problems of a charged moiety due to a secondary amine in our compounds, the library was once again modified to include 4-substituted imidazoles, while avoiding the secondary amine that resulted in formally charged compounds. A post doctoral researcher in our lab, Dr. Xiaoti Zhou, had begun synthesizing compounds for CYP51 based upon a similar synthetic scheme to instead generate imidazoles substituted in the 4-position via an amide linker (Scheme 2).

Scheme 2 is similar to Scheme 1. The scaffold and substituent reactants are identical with the exception of the scaffold's conjugation to a solid phase. Initially, the reformatted library of secondary amine linked compounds, described earlier, was processed to yield the amide linked compounds. However, regenerating amides from secondary amines *in silico* proved to be much more difficult than originally imagined. Bond distances, order, and angle were more easily generated from scratch than

attempting to modify the 4-substituted library. As a result, the 4-carboxyl imidazole and the primary amines were joined as previously described using Sybyl and charged with Gasteiger charges. The final total of 4-substituted amide linked imidazoles was 322.

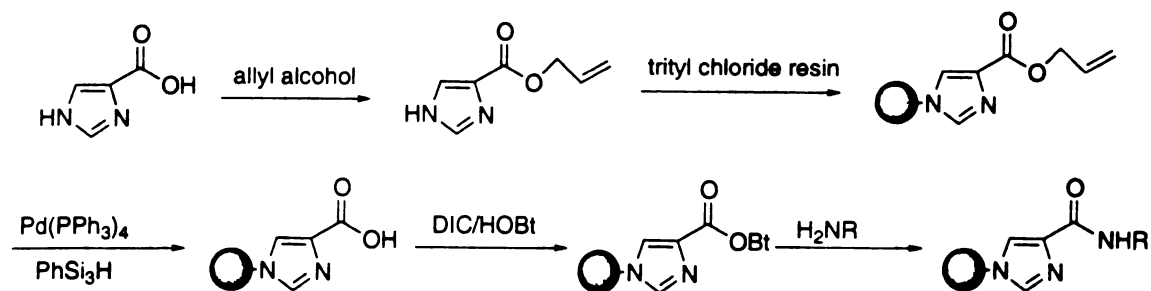
CovDOCK Results for Second Library Expansion

The library was screened using CovDOCK. The results were very similar to the screening of the secondary amine linked expanded library. The overall scores ranged from -27.8 to -20 and the ligands showed appropriate imidazole-heme orientations. The covalent contribution to the overall score was low (~2) and the majority (~80%) of the top hits were the newly added, amide-linked compounds. Figure 7 shows some of the ligands chosen for synthesis.

Organic Synthesis and Assays of Amide-Linked Compounds

The compounds from figure six were chosen by eye for synthesis from the top scoring ligands. Compounds were sent to our collaborators at Vanderbilt University (Waterman lab) who conducted spectroscopic binding assays. With the exception of compound **5**, all compounds give type II shifts indicating imidazole N3-heme iron coordination.

Scheme 2. Solid phase synthesis of amide linked 4-substituted imidazoles.



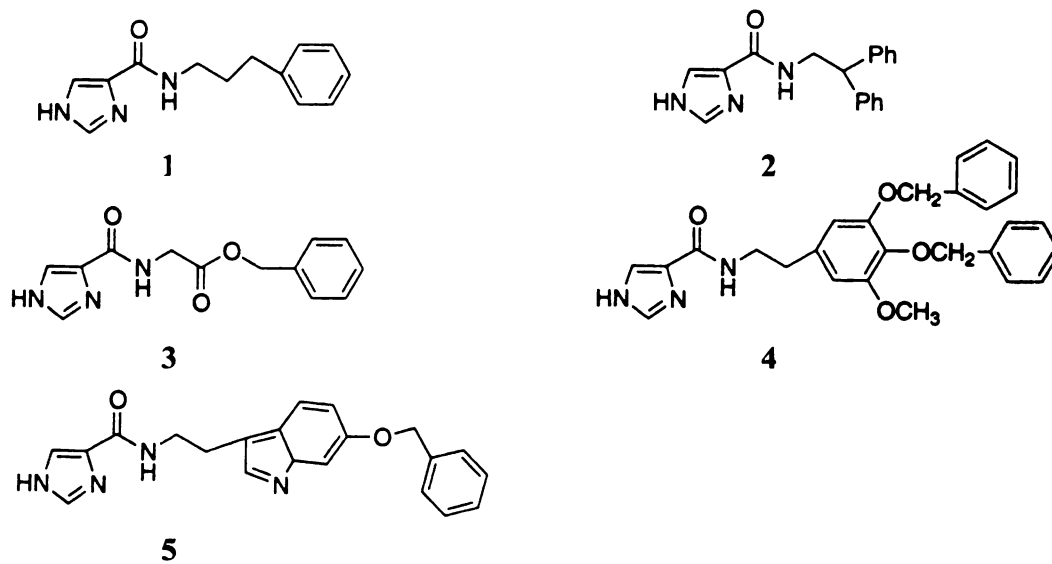


Figure 7. Second round CovDOCK hits. The above compounds were chosen for synthesis based upon their active site complementarity and reactant availability.

Compound **5** induces a type I spectral shift at low concentration, but then causes an increase at 424 nm with increased concentration. This suggests that the ligand binds as a substrate initially, displacing water and resulting in a pentacoordinated heme iron, but as ligand concentration is increased, a 424 nm peak is observed suggesting nitrogen-heme coordination. The apparent K_s for this compound, as calculated by the type II shift is weak ($40 \mu\text{M}$) and the maximum absorbance (0.0153 ± 0.0005) only corresponds to about 20% of the full, expected spectral response.

Overall, the K_s values for the compounds are modest. Ketoconazole has a K_s value of $6.3 \pm 1.1 \mu\text{M}$, while compounds 1-5 do not bind as tightly (21.4, 18.8, 74.4, 262.9, and $64.8 \mu\text{M}$, respectively). Because the binding affinities were not better than those of a known inhibitor, IC50 assays were not conducted. While K_s and IC50 values do not always agree (Chapter 2), spectroscopic analysis remains one of the primary screens for inhibitor affinity.

Experimental Section

Virtual Library Expansion

The virtual library was expanded to include 4-substituted imidazoles. The 4-carboxyimidazole scaffold was modeled in Sybyl (Tripos Inc., St. Louis, Mo.) and substituents were selected from the Available Chemical Database (ACD) (Molecular Design Systems, San Leandro, Ca.) based upon potential synthetic routes for creating 4-substituted imidazoles via an amine and amide linker. The corresponding primary amine reactants were extracted from the ACD using Merlin (Daylight Comp., Santa Fe, N.M.) and UCSelect (G. Skillman, PhD thesis, UCSF). Our primary criterion for reactant selection was synthetic feasibility. A total of 322 compounds generated by this manner were added to the library of 3508 previously generated imidazoles (Chapter 2). The fragments were joined to their appropriate position on the imidazole scaffold using Sybyl. The compounds were then charged using Gasteiger charges in Sybyl.

Organic Synthesis

4-aminobenzylimidazole. To a 50 mL flask, 4-formylimidazole (500 mg) was dissolved in 20 mL of dichloroethane. Benzylamine (397.6 μ L) was added drop-wise following the addition of 1.593 g of $\text{NaBH}(\text{OAc})_3$. Acetic acid (0.3 mL) was then added and the reaction was allow to procede for 3.5 h. Thin layer chromatography (80/20 EthOAc/hexanes) indicated the hydrophilicity of the products. There was no migration of the product spot from the baseline. Because of this, further purification and characterization was not attempted.

Amide Linked 4-Substituted Imidazoles

Allyl 4-imidazolecarboxylate. To a 500 mL flask, 3.0 g of 4-imidazolecarboxylic acid and 5.09 g of *p*-toluenesulfonic acid was added in 40 mL of allyl alcohol. The mixture was stirred at reflux under oxygen overnight. The allyl alcohol was removed under vacuum and the product was purified by column chromatography (CH₂Cl₂/MeOH 10/1) to provide pure product with 65% yield.

Attachment of allyl 4-imidazolecarboxylate to solid phase resin. At room temperature, 319 mg of allyl 4-imidazolecarboxylate with trityl chloride resin (853 mg), dichloroethane (3 mL), DMF (3 mL), and triethylamine (0.8 mL) was shaken for 24 h. The resulting resin was filtered and washed with DMF (3 x 10 mL), MeOH (3 x 10 mL), and CH₂Cl₂ (3 x 10 mL). The resin was dried under vacuum. Resin (990 mg), PhSiH₃ (1.95 g), Pd(PPh₃)₄ (340 mg), and 6 mL of dry CH₂Cl₂ were stirred at room temperature for 2 h, then filtered and washed with DMF (3 x 10 mL), MeOH (3 x 10 mL), and CH₂Cl₂ (3 x 10 mL). The beads were dried *in vacuo* to give the trityl chloride residue conjugated 4-imidazolecarboxylate.

General procedure for preparation of imidazole compounds. Beads conjugated with 4-imidazolecarboxylate (200 mg) were shaken with dicyclohexyl carbodiimide (94 mg) and 1-hydroxybenzotriazole (113 mg) in 4 mL of CH₂Cl₂ at room temperature overnight followed by filtration and rinsing with DMF (3 x 10 mL), MeOH (3 x 10 mL), and CH₂Cl₂ (3 x 10 mL). The beads were then dried *in vacuo*. Dried resin (100 mg) was

shaken with amine compounds in 4 mL of CH₂Cl₂ for 20 h followed by filtration and washing with the same conditions. The beads were then shaken with 5% TFA in CH₂Cl₂ for 30 min and washed with MeOH (3 x 10 mL) and CH₂Cl₂ (3 x 10 mL). The filtrate was concentrated to give the final product.

Data for **1**. NMR (¹H, CDCl₃): δ 8.48 (s, 1H), 7.92 (s, 1H), 7.29-7.14 (m, 5H), 4.09 (br, NH), 3.40 (t, *J*=7.2 Hz, 2H), 2.66 (t, *J*=7.6 Hz, 2H), 1.92 (t, *J*=7.2 Hz, 2H); NMR (¹³C, CDCl₃): δ 158.13, 141.40, 134.63, 128.60, 128.55, 125.45, 126.12, 120.85, 39.59, 33.26, 30.67; HRMS *m/e* 229.1213, calcd 229.1215.

Data for **2**. NMR (¹H, CDCl₃): δ 8.52 (s, 1H), 7.82 (s, 1H), 7.29-7.17 (m, 10H), 5.16 (br, NH), 4.35 (t, *J*=8.0 Hz, 1H), 4.03 (d, *J*=8.0 Hz, 2H); NMR (¹³C, CDCl₃): δ 157.43, 141.85, 134.62, 128.81, 128.56, 128.12, 127.00, 120.93, 50.34, 44.40; HRMS *m/e* 291.1377, calcd 291.1371.

Data for **3**: NMR (¹H, CDCl₃): δ 8.36 (s, 1H), 7.92 (s, 1H), 7.39-7.36 (m, 5H), 5.21 (s, 2H); NMR (¹³C, CDCl₃): δ 138.27, 128.58, 128.94, 67.25. HRMS *m/e* 259.0954, calcd 259.0956.

Data for **4**: NMR (¹H, CDCl₃): δ 7.47-7.25 (m, 12H), 6.51 (s, 1H), 6.46 (s, 1H), 5.03 (s, 2H), 4.99 (s, 2H), 3.78 (s, 3H), 3.63 (t, *J*=6.8 Hz, 2H), 2.82 (t, *J*=6.8 Hz, 2H); NMR (¹³C, CDCl₃): δ 153.87, 152.84, 137.94, 137.22, 136.26, 135.27, 135.06, 128.80, 128.67,

128.39, 128.09, 127.73, 108.01, 106.48, 75.45, 75.41, 56.30, 40.40, 36.49; HRMS *m/e* 457.2013, calcd 457.2001.

Data for **5**: NMR (^1H , CDCl_3) δ 9.08 (s, 1H), 7.62 (s, 1H), 7.52-7.26 (m, 6H), 7.17 (d, $J=2.0$ Hz), 7.09 (s, 1H), 6.91 (dd, $J=8.8, 2.4$ Hz 1H), 5.07 (s, 2H) 3.73 (t, $J=7.2$ Hz, 2H), 3.03 (t, $J=7.2$ Hz, 2H); NMR (^{13}C , CDCl_3): δ 152.97, 137.71, 128.53, 127.84, 127.76, 112.80, 112.08, 102.27, 71.19, 39.45, 25.44. HRMS *m/e* 362.1249, calcd 362.1266.

K_s Assays

The K_s titration experiments were carried out at Vanderbilt University using a 5 μM concentration of CYP51 with 0.2, 2, 5, and 10 μM stock solutions of the inhibitors in DMSO.

Conclusions

CYP51 from *Mycobacterium tuberculosis* represented an excellent target to apply our CovDOCK methods and screen our imidazole library. The coordinates became available shortly after our methods had been refined against P450cam and its importance as a drug target was verified by experiment. Furthermore, the specific interactions in the active site allowed us to finely tune our library to exploit contacts to the imidazole ligands not present in P450cam and tune our library to CYP51 specifically.

Initially, the 4-substituted imidazoles were conjugated to imidazole through an amine linker. Synthesis of 4-aminobenzylimidazole revealed that compounds linked

thusly were extremely polar and were removed from the virtual library. Modifying our virtual library was easily done and 4-substituted imidazoles (linked via an amide moiety) were substituted in the place of polar, amine linked compounds. The substituent reactants for both library additions were the same (primary amines).

The final library was used to predict compounds to synthesize. The solid phase synthesis was done by Dr. Xiaoti Zhou and the compounds were sent to our collaborators at Vanderbilt University where they were spectroscopically assayed against CYP51. Drug design is ideally an iterative process, including several rounds of design, synthesis, and assays. In our first round of inhibitor design, we were able to generate inhibitors for CYP51, one as potent as 18.8 μM . While this affinity is not as good as that of ketoconazole (6 μM), it is acceptable for a single round of design and synthesis. These findings suggest that our methods could easily and quickly be applied to P450 enzymes which are viable drug targets.

References

1. Podust, L.M., T.L. Poulos, and M.R. Waterman, *Crystal structure of cytochrome P450 14alpha-sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors*. Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3068-73.
2. Fischer, R.T., et al., *Mechanistic studies of lanosterol 14 alpha-methyl demethylase: substrate requirements for the component reactions catalyzed by a single cytochrome P-450 isozyme*. J Lipid Res, 1989. **30**(10): p. 1621-32.
3. Trzaskos, J.M., R.T. Fischer, and M.F. Favata, *Mechanistic studies of lanosterol C-32 demethylation. Conditions which promote oxysterol intermediate accumulation during the demethylation process*. J Biol Chem, 1986. **261**(36): p. 16937-42.
4. Aoyama, Y., et al., *Deformylation of 32-oxo-24,25-dihydrolanosterol by the purified cytochrome P-45014DM (lanosterol 14 alpha-demethylase) from yeast evidence confirming the intermediate step of lanosterol 14 alpha-demethylation*. J Biol Chem, 1989. **264**(31): p. 18502-5.
5. McLean, K.J., et al., *Azole antifungals are potent inhibitors of cytochrome P450 mono-oxygenases and bacterial growth in mycobacteria and streptomycetes*. Microbiology, 2002. **148**(Pt 10): p. 2937-49.
6. Guardiola-Diaz, H.M., et al., *Azole-antifungal binding to a novel cytochrome P450 from Mycobacterium tuberculosis: implications for treatment of tuberculosis*. Biochem Pharmacol, 2001. **61**(12): p. 1463-70.
7. Zumla, A., et al., *The tuberculosis pandemic--which way now?* J Infect, 1999. **38**(2): p. 74-9.
8. Schneider, E., M. Moore, and K.G. Castro, *Epidemiology of tuberculosis in the United States*. Clin Chest Med, 2005. **26**(2): p. 183-95.

9. Pust, R.E., *Tuberculosis in the 1990's: resurgence, regimens, and resources*. South Med J, 1992. **85**(6): p. 584-93.
10. Basso, L.A., et al., *Mechanisms of isoniazid resistance in Mycobacterium tuberculosis: enzymatic characterization of enoyl reductase mutants identified in isoniazid-resistant clinical isolates*. J Infect Dis, 1998. **178**(3): p. 769-75.
11. Blazquez, J., et al., *Genetic characterization of multidrug-resistant Mycobacterium bovis strains from a hospital outbreak involving human immunodeficiency virus-positive patients*. J Clin Microbiol, 1997. **35**(6): p. 1390-3.
12. Agdamag, D.M., et al., *Characterization of clinical isolates of Mycobacterium tuberculosis resistant to drugs and detection of RpoB mutation in multidrug-resistant tuberculosis in the Philippines*. Int J Tuberc Lung Dis, 2003. **7**(11): p. 1104-8.
13. Schuller, D.J., et al., *Crystal structure of human heme oxygenase-I*. Nat Struct Biol, 1999. **6**(9): p. 860-7.
14. Sugishima, M., et al., *Crystal structure of rat heme oxygenase-1 in complex with heme*. FEBS Lett, 2000. **471**(1): p. 61-6.
15. Schuller, D.J., et al., *Crystal structure of heme oxygenase from the gram-negative pathogen Neisseria meningitidis and a comparison with mammalian heme oxygenase-I*. Biochemistry, 2001. **40**(38): p. 11552-8.
16. Ewing, T.J., et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. J Comput Aided Mol Des, 2001. **15**(5): p. 411-28.
17. Skillman, A.G., *Structure Based Design of Combinatorial Libraries*, in *Pharmaceutical Chemistry*. 1999, University of California San Francisco: San Francisco. p. 315.

Chapter 4

Differentiating Mono- and Di-Substituted Imidazole Ligands. Stopped Flow Analysis and Kinetics.

Abstract

In Chapter 2 I reported that the 1-N monosubstituted imidazoles and the 1,5-disubstituted imidazoles that I synthesized have similar spectroscopic properties when assayed against the wild-type P450cam enzyme. However, when turnover assays were conducted in the presence of camphor, I found that the monosubstituted compounds alone are effective inhibitors of wild-type P450cam.

Stopped flow experiments illustrate that the binding kinetics are distinct between these two classes of compounds. Stopped flow experiments with camphor and imidazole also show that the binding is heavily dependent on the ligand size. Ligands differing by as few as four atoms (imidazole and 1-methyl cyclopropyl imidazole) are also distinct in their binding profiles. Spectroscopic binding assays in the presence of camphor reveal that the natural substrate affects the binding of the mono- and disubstituted ligands distinctly. While I was unable to fit the stopped flow data to a kinetic model, I was able to qualitatively gain insight into binding. The substituent's specific trends are used to suggest a binding model.

Introduction

In the course of a graduate thesis it is almost inevitable that an experiment will not only challenge an initial hypothesis, but also seemingly fall outside of our canonical knowledge. Such was the case with our imidazole-based inhibitors. In Chapter 2, I described our development of CovDOCK, its application to P450cam and the L244A P450cam mutant, the synthesis of several compounds, and their characterization by both UV and IC50 assays [1].

We found that the disubstituted imidazoles, predicted to bind to the L244A mutant, were also able to bind to the wild-type. This was confirmed by UV assays that not only suggest a binding affinity, but because of the particular type II UV shift (416 to 424 nm) confirms the binding mode [2]. The absorbance increase at 424 nm corresponds with imidazole coordination to the heme iron and it is observed with all of our inhibitors, indicating a single binding mode for all of them. It was not surprising that our disubstituted imidazoles should be able to bind to the wild-type P450cam. P450 enzymes are notoriously promiscuous and even P450cam [3], responsible for carbon salvaging for *Pseudomonas putida*, is known to accept multiple substrates [4]. In the presence of a second substituent, I postulate that the protein must present a slightly different, larger active site.

The big surprise came when I did IC50 measurements to confirm that these ligands were inhibitors. As was expected, the monosubstituted compounds were inhibitors of both P450cam and its L244A mutant. The disubstituted compounds were good inhibitors for the L244A mutant, but did not inhibit wild-type P450cam even at concentrations of up to 1 mM.


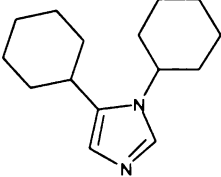
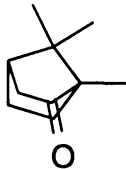
In an attempt to distinguish between the two classes of ligands, both of which are able to bind in the absence of substrate, we undertook a kinetic analysis that might suggest a model capable of explaining these counterintuitive findings. To this end several stopped flow binding assays were carried out with the mono- and disubstituted imidazoles as well as with camphor and imidazole.

Results and Discussion

Imidazole Affinity

Given our initial finding that the affinities of the disubstituted compounds, as measured by UV methods, were similar for P450cam and P450cam L244A (see Chapter one), we measured the affinities of imidazole and 1,5-dicyclohexyl imidazole versus the wild-type and mutant proteins (Table 1).

Table 1. Spectroscopic K_d values. The values for imidazole, 1,5-dicyclohexyl imidazole, and camphor are shown.

	P450cam WT (μM)	L244A (μM)
	7.53 ± 1.4	303 ± 49
	88.0 ± 56	19.6 ± 3.7
	1.5 ± 0.3	1.7 ± 0.3

The L244A affinity for the natural substrate camphor is almost identical to that of the wild type; however, imidazole, the scaffold for all of our ligands, shows a large difference in affinity between the wild-type and the L244A proteins. Across the board, the disubstituted compounds all bind similarly to both the wild-type and the L244A mutant. 1,5-dicyclohexyl imidazole, which is bigger than all our predicted ligands and larger than a rigid active site should allow in either protein, begins to show some

selectivity, as we would have predicted. The larger, disubstituted ligands should, if anything, bind preferentially to the slightly enlarged active site in the L244A mutant.

The difference in imidazole affinity as measured by our K_s assays is about 40-fold. It is unrealistic to employ this difference in affinity as a correction, dividing the affinity of all the disubstituted compounds against the L244A mutant by 40 and thus achieving selectivity. We can, however, follow the “progression” of a single ligand from scaffold to single substituent to a second substituent and observe the binding energy as measured by K_s (Fig. 1). Initially with imidazole there is a large difference in affinity that is not mitigated by a single substitution. The first substituent shows that the ligands gain affinity similarly for both enzymes. Upon addition of the second substituent we see that the trend changes. The L244A continues to gain affinity, while the wild-type enzyme actually loses some affinity for a doubly substituted imidazole. The ligands presented in

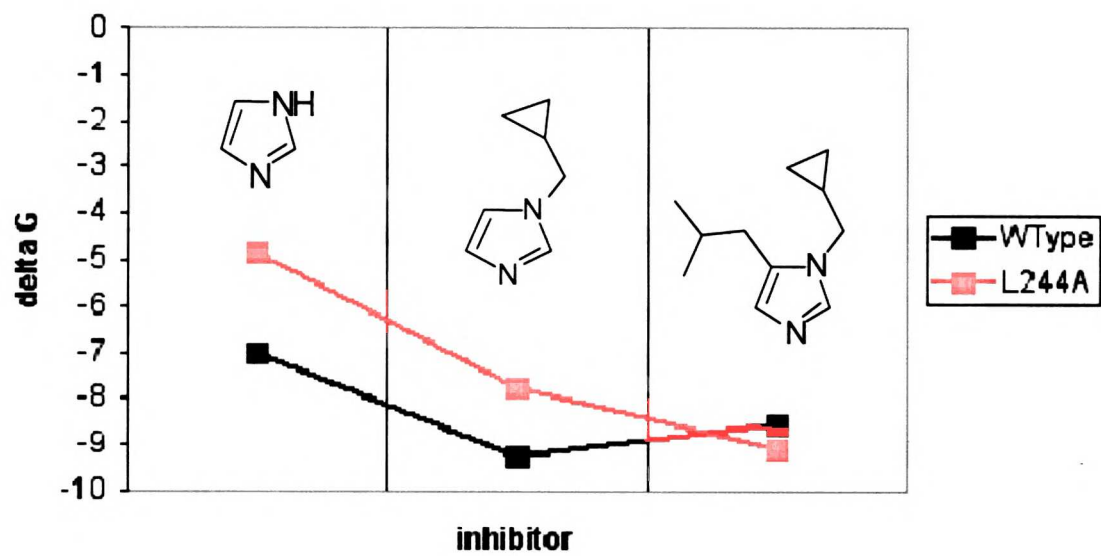


Figure 1. Inhibitor Affinity Progression. Monitoring the affinity from imidazole to 1-methyl-cyclopropyl imidazole to 1-methyl-cyclopropyl-5-(2-methylpropyl) imidazole.

Figure 1 are the most pronounced case in which the affinity actually becomes slightly poorer for the wild-type upon the second addition; however, for most of the ligands the overall trend is the same. The L244A protein continues to gain affinity after two substituents are added while the wild-type enzyme gains little affinity and in some cases loses affinity for the second substituent. These data, along with the K_s value of the 1,5-dicyclohexyl imidazole ligand, suggest an upper limit for the flexibility of wild-type P450cam, an upper limit that is modestly expanded by the L244A mutation.

Slow Tight Binding Inhibitors

More enigmatic than the altered imidazole affinity is the discrepancy between the UV and IC50 measurements. The UV measurements not only suggest binding, but a very specific binding mode. With the iron coordinated by the nitrogen, one expects that electron transfer should not be initiated and that there should be no turnover [5].

Inhibition is observed for the mono- and disubstituted imidazoles against the L244A. However, the monosubstituted compounds show good inhibition of wild-type P450cam, but the disubstituted compounds show no inhibition even at concentrations as high as 1mM (Fig. 2). These data, while in agreement with our initial CovDOCK predictions (Chapter two) are difficult to reconcile with the spectroscopic binding data.

A potential model to explain these findings was suggested by Dr. Susan Miller. Because the UV assays were based upon a titration, while the IC50 assays were based on kinetic assays, the inhibitors could be functioning as slow, tight binders. Slow, tight binders were originally proposed by Morrison et al. [6] to describe protease inhibitors

which irreversibly bound to protein before they could function as inhibitors. The kinetic model used to describe their behavior is:



Where **E** is the enzyme, **I** is the inhibitor, **EI** is the enzyme inhibitor complex, and **EI*** is the activated enzyme inhibitor complex necessary for inhibition. **EI** is not a viable inhibitory complex and can be out competed by substrate. The **EI*** formation is time dependent and dependent on the presence of substrate.

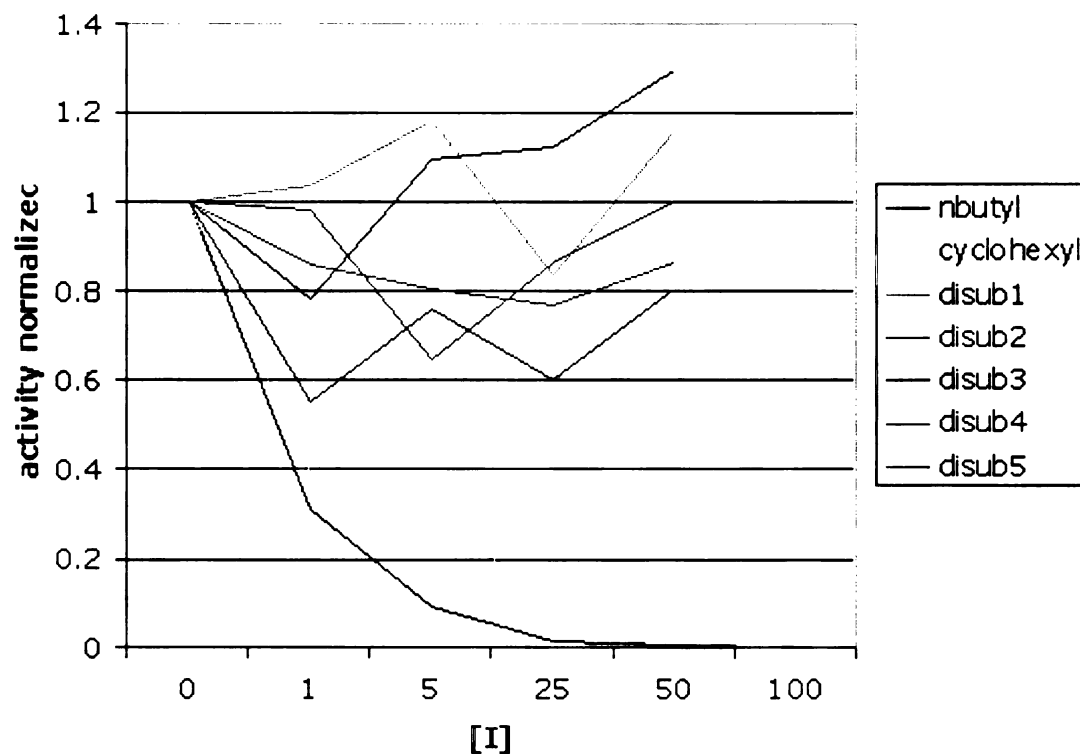


Figure 2. Inhibition Assays. Comparison of all di-substituted and two mono-substituted inhibitor profiles. The data has been normalized to a max activity of one to better compare the inhibition effects. The di-substituted compound's profile, while noisy, extends to 1 mM with an activity of 1 ± 0.4 .

While our inhibitors coordinate through the nitrogen, we are able to compete them off with camphor. The bond formed is not covalent. Regardless, we conducted the appropriate assays to determine whether our inhibitors were functioning as slow, tight binders. The enzyme was preincubated with 1-N benzyl imidazole, 1-methyl propyl-5-(2-methylpropyl) imidazole, and DMSO for 5, 10, and 20 min before initiating the reaction by addition of camphor and NADH. Results (data not shown) showed little difference between the DMSO control and the mono- or disubstituted imidazoles.

Stopped Flow Experiments

To differentiate between the particulars of binding which result in one class of inhibitor being displaced by camphor, while the other class acts as an inhibitor even at relatively low concentrations; we carried out stopped flow experiments to elucidate the mechanism of action of our inhibitors.

We began with single point studies using P450cam that had been run thrice through a PD-10 column to remove any bound camphor. Single mixing studies were done at 20 °C with 0.4 μ M protein and 150 μ M inhibitors. The absorbance was monitored from 325 to 700 nm over 0.6 seconds. Saturation was observed for all ligands and generally occurred over the first 200 milliseconds (Fig. 3).

One characteristic of the spectra was a slight dip at 416 nm at the start of the reaction. This was more pronounced with the disubstituted compounds (Fig. 4), primarily because their on-rates were slower. Initially we attempted to quantitate the disappearance of water as the hexacoordinated ligand, but the process happens too quickly to measure for the monosubstituted compounds and is barely observable for the

disubstituted compounds. The spectra seemed to fit to a single exponential, but with further studies (presented later in the chapter) it was clear that they fit better to a two exponential or an A to B to C mechanism.

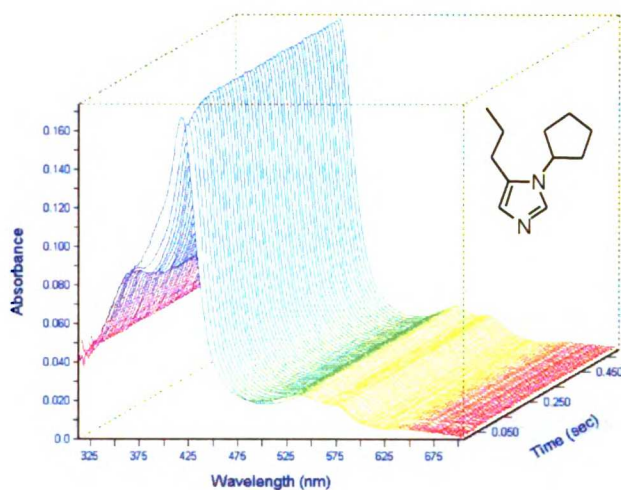


Figure 3. Single Point Stopped Flow. This pattern for 1-cyclopentyl-5-propyl imidazole is characteristic of most of the spectra. Protein ($0.2 \mu\text{M}$) was combined with $75 \mu\text{M}$ inhibitor. Saturation is observed over about 200 milliseconds. The overall spectra remain constant with a slight dip at 416 nm from water displacement.

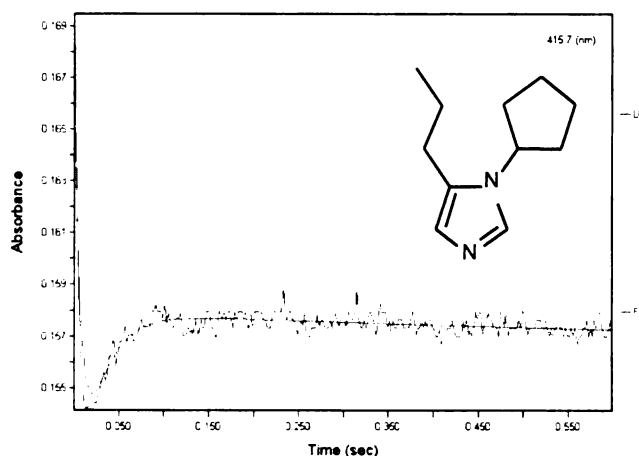


Figure 4. Initial Dip at 416 nm. Initial dip at 416 nm for 1-cyclopentyl-5-propyl.

It is difficult to draw any conclusions from the single-point experiments so we moved instead to multiple concentrations of ligand in an attempt to quantitate any kinetic differences between the mono- and disubstituted compounds. The experiments were done at 15 °C with a final protein concentration of 0.2 μ M after mixing. The ligands were mixed at final concentrations of 12.5, 50, 125, 250, and 500 μ M. At varying concentrations it became clear that neither the mono- nor the di-substituted compounds fit well to a single exponential. This alone was an interesting finding suggesting that the on rate of the ligands is more complicated than a simple A to B mechanism. Instead, the binding spectra can be fit to a two exponential function or an A to B to C mechanism with two rate constants (Fig. 5).

A single wavelength was monitored for all the compounds (424 nm) that corresponds to the heme iron-imidazole coordinated species [2]. Because we are fitting to a two exponential function we obtain two rate constants, k_1 and k_2 . In order to simplify the study we chose to analyze only one of the constants. The first rate, k_1 , is greater in magnitude and easier to reproduce. Furthermore, k_1 shows a larger ligand concentration dependence than k_2 and is more well behaved over the ligand range (Fig. 6).

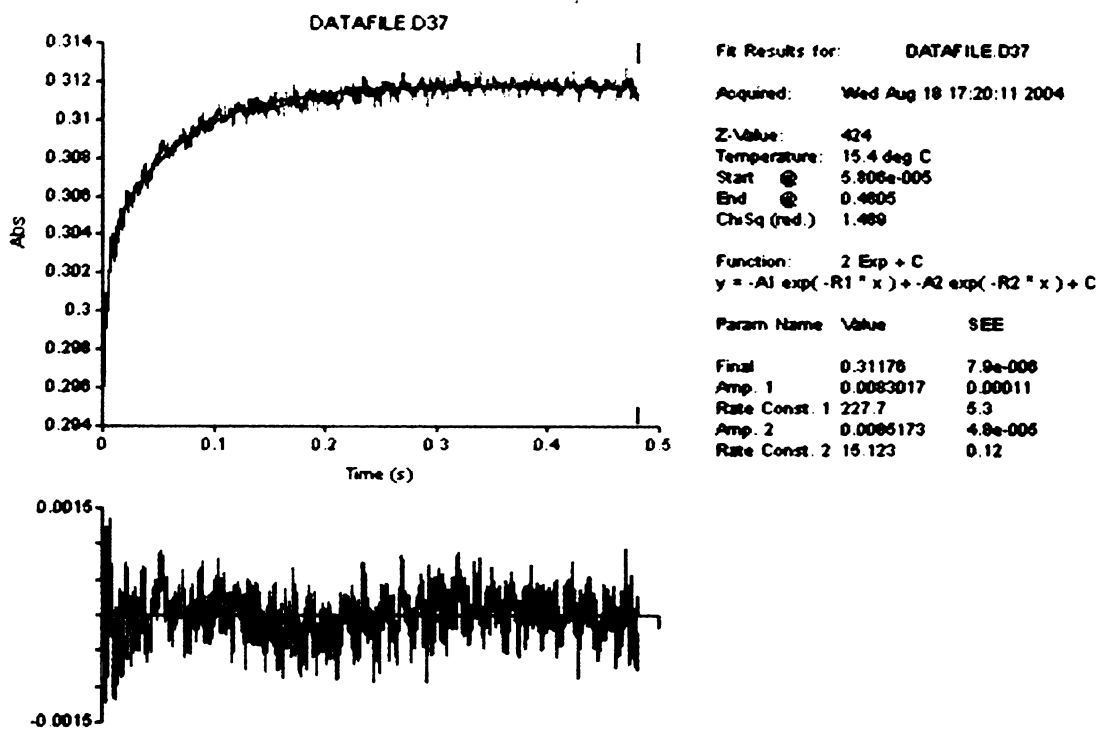


Figure 5. Substituted imidazole stopped flow. Binding spectra for 1-N-butyl imidazole monitored at 424 nm fit to a double exponential function.

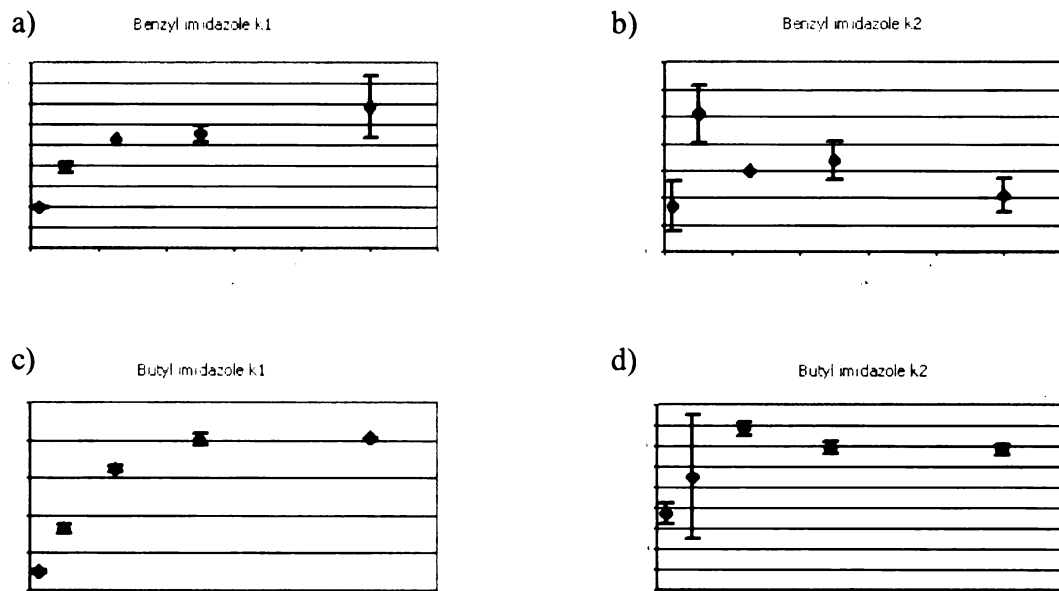


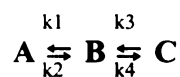
Figure 6. Comparison of two binding constants. Ligands 1-benzyl imidazole (a and b) and 1-nbutyl imidazole (c and d). The rate constant k_1 (a and c) compared with k_2 (b and d).

Plotting k_1 vs. ligand concentration reveals some general differences between the mono- and disubstituted compounds (Fig. 7). Most notable and pronounced, the k_1 for the monosubstituted compounds (Fig. 7a) is much greater than that for the disubstituted compounds (Fig. 7b). This suggests a slower initial state of complex formation that is consistent with the higher K_s values observed for the disubstituted compounds. The

monosubstituted compounds reach saturation faster than the disubstituted compounds. There is also a greater linearity to the inhibitor dependence for the disubstituted compounds, while the monosubstituted compounds show more of a hyperbolic progression with ligand concentration.

Overall the data reinforces the information gleaned from UV experiments. The monosubstituted compounds are tighter binders, but the disubstituted compounds are still able to bind. It also qualitatively illustrates some differences between these two classes of inhibitors.

Ultimately we would have liked to quantitatively describe the kinetics of these inhibitors by fitting them to a reasonable model. Several kinetic models have been applied to the plots of k_1 vs. ligand concentration without success, including a three phase A to B to C model which is solved as equation 1:



$$K_{obs} = \frac{k_3[I]}{[I] + k_{init}} + k_4 \quad \text{equation 1}$$

Unfortunately the poor fit of equation 1 (data not shown) indicates that this kinetic model does not appropriately describe our system.

Control experiments were conducted with camphor and imidazole. The binding of these ligands has been measured previously by UV spectroscopy and K_s values have

been calculated (1.3 μM and 7.5 μM , respectively). Imidazole produces a type II shift with an increase at 424 nm consistent with nitrogen coordination to the heme iron and is

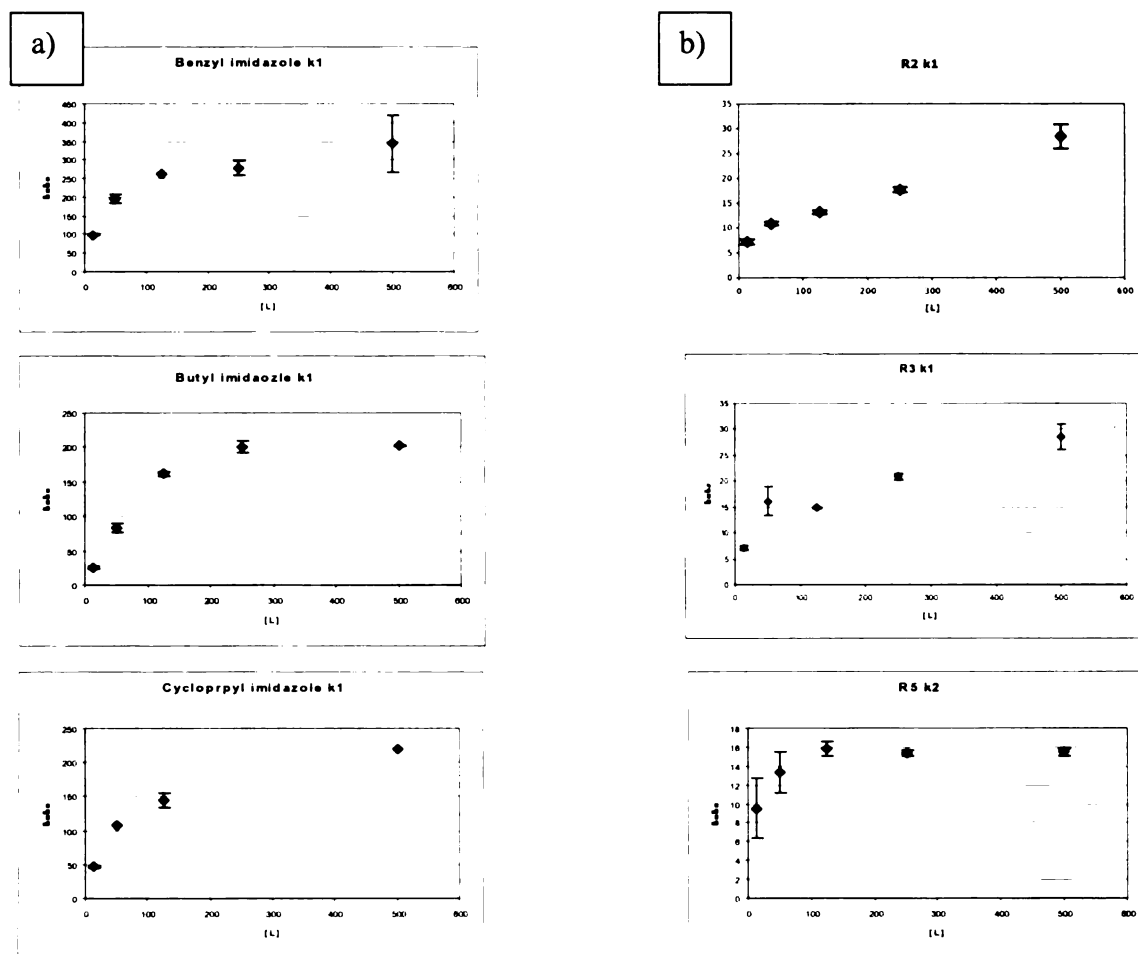


Figure 7. Comparison of rates vs. ligand concentration. Comparison of ligand rates (k_{obs}) vs. concentration (μM) for monosubstituted compounds (column a) and disubstituted compounds (column b) illustrating the greater linearity of the monosubstituted compounds.

identical to the shift produced by all the inhibitors discussed here. Camphor produces a type I shift [7], an increase at 390 nm caused by displacement of the coordinated water that results in a pentacoordinated heme iron. The kinetic profiles of both ligands were

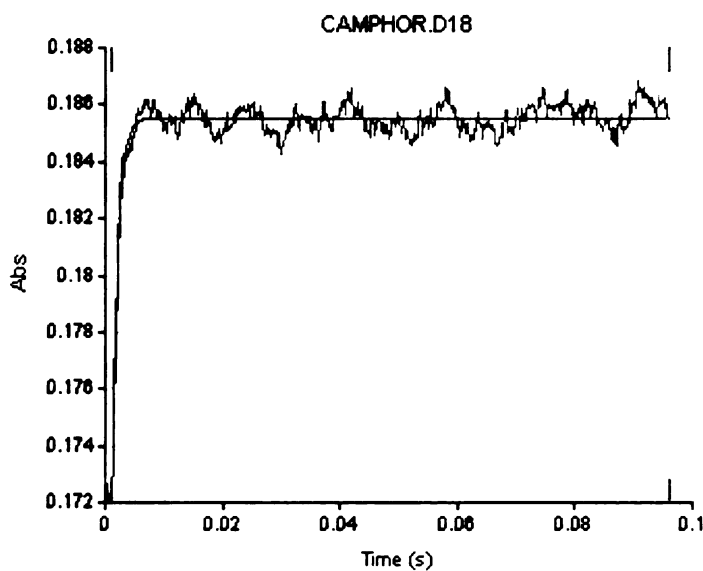
measured as with the substituted ligands. Single mixing experiments monitoring the absorbance at 424 nm for imidazole and 390 nm for camphor were conducted with 12.5, 50, 100, 250, and 500 μM concentration of each ligand. Camphor binds extremely quickly to the protein, faster than any of the substituted imidazole ligands, and is beyond the limit of detection of the instrument ($\sim 1000 \text{ s}^{-1}$) at concentrations greater than 125 μM ; the imidazole on-rate is relatively slow (5.8 s^{-1} at 125 μM ligand concentration).

These two control ligands display kinetic profiles different from both the mono- and disubstituted imidazole compounds, in that they can be fit to single exponential equation (Fig. 8). This is a somewhat surprising finding, particularly for imidazole which only varies by substitutions in one case by a group as small as a methyl cyclopropyl moiety. When plotted against ligand concentration the rate constant seems to show some linearity, as would be expected in an A to B mechanism. But when fit to a linear model (equation 2) [8] we find that the calculated K_d (equation 3) [9] values do not match the spectroscopically calculated K_s values.

$$k_{\text{obs}} = k_{\text{on}} [\text{L}] + k_{\text{off}} \quad \text{equation 2}$$

$$K_d = k_{\text{off}} / k_{\text{on}} \quad \text{equation 3}$$

The K_d values calculated for camphor and imidazole by this method are 29 μM and 183 μM respectively. The spectroscopically calculated K_s values are 1.3 μM for camphor and 7 μM for imidazole. While the ratios for these values are the same for both methods, potentially suggesting a systematic error, it is unfortunate that we cannot return the known results [10] with our methods. We can conclude, however, that the mechanisms



Fit Results for: CAMPHOR.D18

Acquired: Tue Oct 26 21:40:39 2004

Z-Value: 390
 Temperature: 15.3 deg C
 Start @ 0.001104
 End @ 0.095905
 ChiSq (red.) 1.905

Function: 1 Exp + C
 $y = -A \exp(-k \cdot x) + C$

Param Name	Value	SEE
Final	0.18554	1.5e-005
Amp.	0.045463	0.002
Rate Const.	1027.6	27



Figure 8. Stopped flow analysis of camphor. The absorbance change at 390 nm for camphor is able to be fit to a single exponential.

of binding for the natural substrate and imidazole are more similar to each other (single exponential) than to those of any of the substituted imidazoles (double exponential). Comparing the binding of imidazole in particular to its derivatized cousins, we can qualitatively see the impact even small substitutions have on the kinetics of binding. This further supports our hypothesis that imidazole substitution patterns, even small moieties, can dramatically affect binding

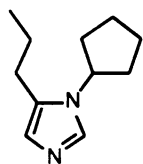
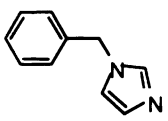
K_s Assays in the Presence of Substrate

All of the K_s assays were done with protein in which a water was bound to the heme iron as the starting point and were conducted by measuring the disappearance of a 416 nm, water coordinated peak and the appearance of a 424 nm, imidazole nitrogen-coordinated iron. The IC_{50} assays, by definition, were done in the presence of substrate (250 μ M camphor). Stop flow binding assays were also conducted with the heme iron in the water bound starting state. To understand the effect of substrate on the binding assays we carried out K_s assays for a disubstituted and monosubstituted imidazole at various concentrations of camphor.

UV assays were done in the presence of 5, 50, and 250 μ M camphor with 1-cyclopentyl-5-propyl imidazole and 1-benzyl imidazole (Table 2). Both compounds have K_s values lower than that of camphor, indicating that they are tighter binders. As we increase the concentration of camphor the observed K_s increases, as one would expect in a competition assay. 1-Benzyl imidazole, the tightest binder we developed, has an

affinity of 50 nM in the absence of camphor and at 250 μM substrate the apparent K_s is 7.4 μM . 1-cyclopentyl-5-propyl imidazole similarly increases its apparent K_s in the presence of camphor from 800 nm to 774 μM when assayed at 0 μM and 250 μM camphor, respectively. It is interesting that the affinities do not decrease similarly over equivalent concentrations of substrate.

Table 2. K_s values (μM) vs. camphor concentration.

	no camphor	$\sim 5 \mu\text{M}$ camphor	50 μM camphor	250 μM camphor
	0.8	3.1	198	774
	0.05	0.6	2.8	7.4

Instead we observe an increase in K_s of 100 fold for the mono-substituted imidazole and 1000 fold for the disubstituted imidazole. Not only does this suggest that the details of binding are different for the two classes of ligands, but also that the effect of substrate on binding is distinct for each class.

Experimental Section

Protein Expression and Binding Assays

Cytochrome P450cam was expressed in *Escherichia coli* as previously reported [11]. Both proteins were stored with 100 μ M camphor at -70 °C. Camphor was removed before the proteins were used by thrice eluting with 100 mM sodium phosphate through a PD-10 size exclusion column. The P450cam thus obtained has a water coordinated to the heme iron and is in the low spin state. Spectroscopic binding (K_s) assays were done in 100 mM potassium phosphate buffer at 37 °C by difference UV spectroscopy. The two cuvettes contained 0.5 μ M protein in 300 μ L of 100 mM potassium phosphate buffer. The potential inhibitors were dissolved in DMSO, with the exception of imidazole which was dissolved in water. DMSO never constituted more than 2% of the final sample volume. An equal amount of DMSO was added to the reference cuvette and the shift in absorbance from 416 to 424 nm as increasing concentrations of camphor were titrated into the cuvette was observed. For the spectroscopic assays in the presence of varying amounts of substrate, a stock solution of camphor in water was added prior to the imidazole compound titrations. The difference between 416 and 424 nm values was then recorded and fitted to a binding curve to obtain K_s values. K_s values were calculated by non-linear regression analysis using the equation 4:

$$\Delta A = (\Delta A_{\max}[I]) / (K_s + [I]) \quad \text{equation 4}$$

Where ΔA is the 416 trough to 424 peak absorbance difference, ΔA_{\max} is the difference in absorption at saturation, and $[I]$ is the inhibitor or ligand concentration [12].

Imidazole and 1,5-dicyclohexyl imidazole were purchased from Aldrich (St. Louis, MO).

Preincubation IC50 Assays

To determine whether our ligands were slow, tight binders we preincubated protein and cofactors minus camphor and NADH for 5, 10, and 20 min with 1-N benzyl imidazole and 1-methyl propyl-5-(2-methylpropyl) imidazole at ligand concentrations of 25, 50, 100, and 250 μM . The final reaction vessel contained P450cam 5 nM, putidaredoxin 2 μM , putidaredoxin reductase 2 μM , camphor 250 μM , and NADH 1000 μM . Camphor was added after the appropriate incubation time and after all aliquots had been preincubated, NADH was added to begin the reaction. The total reaction volume was 150 μL . Varying amounts of inhibitor were added in DMSO, but the DMSO volume never exceeded 3% of the total reaction volume. A control of DMSO alone was also included based upon the amount of DMSO necessary to achieve the appropriate ligand concentration per aliquot. All assays were done in 100 mM potassium phosphate in a 20°

C bath for 15 min. The reactions were then quenched with 20 μL of CH_2Cl_2 containing 2 mM β -thujone as an internal standard.

The CH_2Cl_2 was then withdrawn directly from the reaction mixture and injected onto an HP 5980 Series II Gas Chromatograph equipped with an Agilent Technologies (Palo Alto, Ca.) DB-10 capillary column. A temperature gradient from 40° to 280° C over 20 min was used. Turnover rates were then plotted for each inhibitor at various preincubation states.

Stopped Flow Assays

All stopped flow assays were done as single mixing experiments using a Shimadzu UV-2101PC double-beam absorbance spectrophotometer, and resulting data were analyzed using SPECFIT (Spectrum Software Associates, Chapel Hill, NC), a program that employs a singular value decomposition routine, global least-squares fitting, and factor analysis of the entire set of spectral data to yield values for kinetic or equilibrium binding constants for a specified model. When indicated, single wavelengths monitored were 390 nm for camphor binding assays, 424 nm for imidazole based compounds, and 416 nm for water-coordinated species. The binding assays were done at 15 or 20 °C as indicated in the text.

All assays were done in a buffer of 100 mM KPi with a final 2-fold diluted protein concentration of 0.2 μM at pH 7.4. Protein was concentrated in Amicon-2000 centrifugal concentrators. Camphor, imidazole, and mono- and disubstituted imidazole ligands were dissolved from a stock solution in DMSO into 100 mM KPi buffers at

concentrations of 25, 50, 100, 500, and 1000 μM to be diluted 2 fold during the single mixing experiments.

Conclusions

The finding that compounds could coordinate the heme-iron and not be inhibitors was as unexpected as it was difficult to resolve. The disubstituted compounds, whose K_s values were all at or below the natural substrate's affinity, showed good binding, but lacked any inhibitory properties in the presence of 250 μM camphor.

The initial hypothesis of slow, tight binding inhibitors was proposed primarily because of its precedence in the literature [13-15]. Our previous UV spectroscopy experiments had shown that the imidazole-based inhibitors could be displaced by camphor or removed by dilution. The coordination bond to the heme-iron was not an irreversible covalent species. The preincubation experiments with 1-N benzyl imidazole and 1-methyl propyl-5-(2-methylpropyl) imidazole compared with preincubation with DMSO alone did not suggest that there was a time dependence to the formation of a stable **EI*** inhibitory complex.

The stopped flow experiments reveal several differences between the mono- and disubstituted compounds. Primarily, that k_1 , the first half of the binding event as measured by conversion to the 424 nm species, is faster for the monosubstituted compared to the disubstituted compounds.

It is also important that the behavior of none of the substituted imidazoles can be fit to a single exponential binding curve, suggesting that the binding event is not a simple

A to B mechanism. This information is most notable against the backdrop of the controls, camphor and imidazole, which both display single exponential on-rates. It is surprising that even a single substituent on the imidazoles affects the method of binding.

While we were unable to fit the data to a specific kinetic model, and thus to quantitate the on- and off-rates for the ligands, we do find that the binding event is more convoluted than an A to B or even A to B to C mechanism. Furthermore, the UV spectroscopic assays done with 1-cyclopentyl-5-propyl imidazole and 1-benzyl imidazole in the presence of varying camphor concentrations highlight the fact that the effects of the substrate on the imidazole-based inhibitor binding is distinct for both the mono- and disubstituted models.

An early model that was considered and may be viable is shown in figure 9. In this model, the substrate-free form of the protein preferentially binds camphor and the

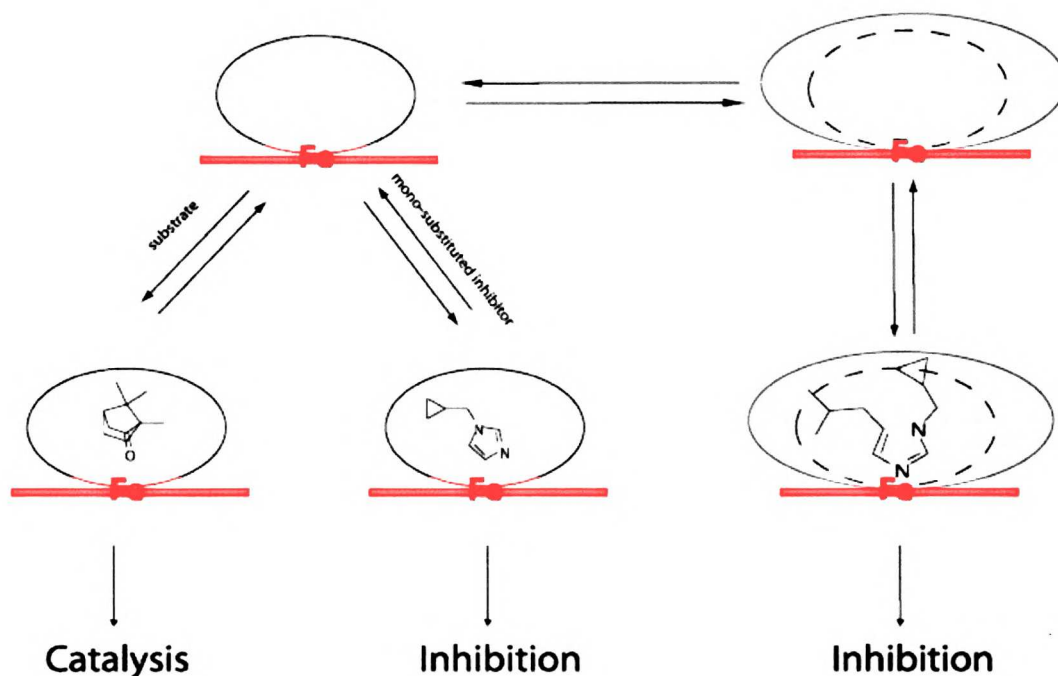


Figure 9. Two-state model of binding. The breathing motions that must be made to accommodate the disubstituted ligands are represented with the empty dotted and solid ellipses. The active site must expand (blue ellipse) to accommodate the larger ligands and in the presence of camphor the equilibrium is pulled toward the smaller active site.

monosubstituted ligands and requires little or no change in the active site structure [16, 17]. Larger ligands, however, require an active site that we know from substrate-bound and substrate-free crystal structures to be deformed and enlarged relative to the native state [18, 19]. It is clear, however, from the binding experiments that the protein is able to accommodate disubstituted imidazoles. The additional contacts, relative to the monosubstituted ligands, do not contribute to the overall binding (Fig. 2), most likely because of the energetic penalty the protein must endure to expand its active site. One

can then postulate a two-state active site, expanded and native. In the presence of camphor, we hypothesize, that the equilibrium would be drawn toward the left side of this scheme, away from an active site able to bind the disubstituted compounds.

References

1. Verras, A., I.D. Kuntz, and P.R. Ortiz de Montellano, *Computer-assisted design of selective imidazole inhibitors for cytochrome p450 enzymes*. J Med Chem, 2004. **47**(14): p. 3572-9.
2. Schenkman, J.B., H. Remmer, and R.W. Estabrook, *Spectral studies of drug interaction with hepatic microsomal cytochrome*. Mol Pharmacol, 1967. **3**(2): p. 113-23.
3. Poulos, T.L., *Cytochrome P450: molecular architecture, mechanism, and prospects for rational inhibitor design*. Pharm Res, 1988. **5**(2): p. 67-75.
4. Zhang, Z., et al., *The substrate specificity of cytochrome P450cam*. Bioorg Med Chem, 1998. **6**(9): p. 1501-8.
5. Ortiz de Montellano, P.R., *Cytochrome P450: Structure, Mechanism, and Biochemistry*. 2nd ed, ed. P.R. Ortiz de Montellano. 1995, New York: Plenum Press.
6. Morrison, J.F. and C.T. Walsh, *The behavior and significance of slow-binding enzyme inhibitors*. Adv Enzymol Relat Areas Mol Biol, 1988. **61**: p. 201-301.
7. Sharrock, M., et al., *Cytochrome P450cam and its complexes. Mossbauer parameters of the heme iron*. Biochim Biophys Acta, 1976. **420**(1): p. 8-26.
8. Segel, I.H., *Enzyme Kinetics : Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. 2nd ed. 1975, New York: John Wiley and Sons.
9. Copeland, R.A., *Enzymes*. 2nd ed. 2000, New York: Wiley:VHC.
10. Cole, P.E. and S.G. Sligar, *Temperature-jump measurement of the spin state relaxation rate of cytochrome P450cam*. FEBS Lett, 1981. **133**(2): p. 252-4.

11. Wagner, G.C., et al., *Apoprotein formation and heme reconstitution of cytochrome P-450cam*. J Biol Chem, 1981. **256**(12): p. 6262-5.
12. Rao, S.I., et al., *The lipoxygenase activity of myoglobin. Oxidation of linoleic acid by the ferryl oxygen rather than protein radical*. J Biol Chem, 1994. **269**(10): p. 7210-6.
13. Chang, C.F., et al., *Discovery of picomolar slow tight-binding inhibitors of alpha-fucosidase*. Chem Biol, 2004. **11**(9): p. 1301-6.
14. Bienvenue, D.L., B. Bennett, and R.C. Holz, *Inhibition of the aminopeptidase from Aeromonas proteolytica by L-leucinethiol: kinetic and spectroscopic characterization of a slow, tight-binding inhibitor-enzyme complex*. J Inorg Biochem, 2000. **78**(1): p. 43-54.
15. Kapoor, M., et al., *Slow-tight-binding inhibition of enoyl-acyl carrier protein reductase from Plasmodium falciparum by triclosan*. Biochem J, 2004. **381**(Pt 3): p. 719-24.
16. Poulos, T.L., B.C. Finzel, and A.J. Howard, *Crystal structure of substrate-free Pseudomonas putida cytochrome P-450*. Biochemistry, 1986. **25**(18): p. 5314-22.
17. Poulos, T.L., B.C. Finzel, and A.J. Howard, *High-resolution crystal structure of cytochrome P450cam*. J Mol Biol, 1987. **195**(3): p. 687-700.
18. Raag, R., et al., *Inhibitor-induced conformational change in cytochrome P-450CAM*. Biochemistry, 1993. **32**(17): p. 4571-8.
19. Dunn, A.R., et al., *Probing the open state of cytochrome P450cam with ruthenium-linker substrates*. Proc Natl Acad Sci U S A, 2001. **98**(22): p. 12420-5.

Chapter 5

Mitigated Imidazole Affinity: Crystal Structures of P450cam and P450cam L244A

Abstract

To further explore the mitigated L244A affinity for imidazole relative to wild-type enzyme, we crystallized the L244A mutant in the presence and absence of imidazole. The wild-type enzyme was also crystallized in the presence of imidazole. Because of the small size of imidazole and the lack of contacts it makes in the active sites of both proteins, I find that the electron density about the putative imidazole is not well resolved. Comparisons with the water coordinated heme irons for both proteins show that the distance from the heme iron to the first proximal atom shown in the electron density is not consistent with a water bound species. I use this and the density to argue that we have indeed resolved the structures with bound imidazole, though further experiments would be necessary to verify this conclusion.

The crystal structure of the substrate free P450cam L244A protein is also solved. This was the target against which our inhibitors were screened with CovDOCK in Chapter 2. The I helix of the L244A mutant is shifted toward the active site to compensate for the cavity made by the leucine to alanine mutation.

Introduction

In Chapter 2 we presented data which shows that introduction of an L244A mutation into P450cam results in abated affinity for our imidazole ligands. It is interesting that while the imidazole affinity is 40-fold worse for the mutant versus the wild-type the affinity for the natural substrate, camphor, is unchanged.

The L244A mutation is a single point mutant at the active site in the center of the enzyme that putatively enlarges the active site by removal of three atoms, facilitating the binding of our disubstituted ligands [1]. In this chapter, we will show crystallographic data of the substrate-free P450cam L244A mutant and show the structure we targeted for drug design in Chapter 2. Some interesting differences are found between the active site that we modeled for CovDOCK and the active site determined by crystallography.

The K_s for imidazole is $303 \pm 49 \mu\text{M}$ for the mutant compared to $7.5 \pm 1.4 \mu\text{M}$ for the wild-type P450cam. The forty-fold difference is significant and affects our ability to generate selective ligands when introducing a second substituent on imidazole. Because the primary antifungal agents currently used are azole-based compounds [2-4] and because the affinities of P450cam and P450cam L244A are unchanged for camphor (Chapter 2), this mutation has potential utility in studying P450 azole resistance. Azole resistance has been observed in several other P450 enzymes [5-7]. The discovery of mitigated affinity for imidazole in P450cam was somewhat serendipitous. The mutation was introduced to create a slightly altered active site against which to generate selective ligands, but in this chapter, we will explore the factors that may cause this mitigated imidazole affinity.

Crystallography was used to explore the structural factors contributing to the difference in imidazole affinity. A C334A/L244A double mutant was generated to facilitate crystallography of the mutant. The C344A modification is a surface mutation that prevents protein aggregation at the high concentrations necessary for crystallography [8] and allows more facile crystallization of the protein by enabling us to use the hanging-drop method rather than capillary diffusion to grow crystals [9].

Imidazole is a small ligand of only five atoms. The P450cam active site is itself small in comparison to those of other P450 enzymes (Chapters 1 and 3) and very few contacts are made between the active site residues and imidazole. To date only one P450 enzyme has been co-crystallized with imidazole (Chapter 1, Table 1). Cyp119 from *Sulfolobus solfataricus* was resolved to 2.69 Å with bound imidazole [10]. The Cyp119 active site collapses around the ligand and allows for distinct resolution of the imidazole. P450cam, despite being the best characterized P450enzyme [11], has not been crystallographically resolved with an unsubstituted imidazole. One of the obstacles in determining this co-crystal structure may be the small size of imidazole and the scarcity of contacts with the protein to stabilize a distinct, crystallographically resolvable complex.

I report the structures of P450cam and P450cam L244A in the presence of imidazole to 1.45 Å and 2.15 Å, respectively. Unfortunately, the density proximal to the heme in both cases is not fully resolvable as imidazole. By comparing the structure co-crystallized in the presence and absence of imidazole, several distinguishing characteristics are notable. The heme iron distance to the proximal electron density is distinct for imidazole co-crystallized structure versus the water coordinated structure and

the density suggests more than a single atom. Based on this putative imidazole binding mode, we are also able to suggest a cause of the mitigated imidazole affinity in P450cam L244A.

Results and Discussion

Construction of the L244A/C334A P450cam Double Mutant

In order to facilitate the crystallization of the L244A mutant, a second mutation was introduced (C334A). This cysteine to alanine mutation at the surface of the protein allows us to concentrate the protein enough to employ the crystallography hanging drop method without causing protein dimerization [9]. We also use the C334A mutation for the preparation of wild-type crystals. The P450cam C334A gene sequence was given to us by the Wong group at Oxford in a pCHC plasmid. The L244A primers were ordered with the following sequence:

5' GCCGACCAGTAAGGGCGCCACACATCC

3' GGATGTGTGGCGCCTTACTGGTCGGC

Underlined bases indicated the point of mutation and red bases are mismatched with the wild-type sequence. This primer also introduces a Nar I restriction site which we used to screen for successfully mutated plasmids. Quick change PCR with the P450cam C334A in pCHC and pcWori plasmids, followed by digestion with Nar I and Xbe I to screen for the presence of the L244A mutant did not show the appropriate scission profile.

Instead the L244A mutation, already present in a puc19 plasmid was used as the template and the following C334A primers were ordered:

5' GAGCGCGAAAACGCCGCCCCGATGCACGTCGAC

3' GTCGACGTGCATCGGGGGCGCGTTTTGCGCGCT

Underlined bases indicate the point of mutation, while bases colored red are not complementary to the wild-type sequence. The C334A oligo was then used as a primer with the P450cam L244A puc19 template. The quickchange reaction was successful as indicated by digestions with Nar I and Xbe I restriction enzymes (data not shown). The doubly mutated plasmid was sent for sequencing. The L244A and C334A (5' and 3') primers were used for sequencing. After the sequence was verified, it was cut out of puc19 using Nde I and Xbe I and ligated into pcWori for expression.

For the remainder of this chapter and including the introduction and abstract, the L244A mutant refers to the L244A/C334A double mutant. Wild-type refers to the C334A single mutant.

Protein Purification

The protein was expressed and purified as normal with one exception. Generally the protocol for P450cam purification involves two columns, a DEAE and a Q-sepharose column with an ammonium sulfate precipitation before the second column. Because, ultimately we will be crystallizing these proteins with an ammonium sulfate precipitant,

the precipitation step was omitted during protein purification and the protein was loaded directly on to the Q-sepharose column from the DEAE column. Generally, salt cuts done with similar precipitants as those used in crystal screening are not used during purification as they may upset the final concentrations of precipitants in the hanging drops.

After the second column, the R_z value for wild-type protein is generally between 1.3 and 1.4. For assay purposes this is sufficient, but to insure optimal purity for crystallography the protein was run through a third column. A gel filtration S-200 column was used with a running buffer of 50 mM KPi and 250 mM KCl at pH 7.4. This results in a final R_z greater than 1.5. Because the L244A protein does not fully convert into a 390 nm peak even in saturating concentrations of camphor, the L244A protein purity was determined by SDS-PAGE protein gels and its concentration was determined by a BCA (bicinchoninic acid) assay [12]. Also, contrary to the protocol for purifying P450cam, camphor was not included in any loading, eluting, or washing buffers.

After purification, the protein was concentrated to 50 mg/ mL in 50 mM KPi, 250 mM KCl, and 50 mM DTT. If at all possible, crystal trays were set up without prior freezing of the protein.

Crystal Screening

The protein was crystallized in 50 mM KPi, 250 mM KCl, and 50 mM DTT with ammonium sulfate ranging from 38-43% saturation at room temperature. Most crystals were observed about a median saturation of 40% ammonium sulfate. Initial crystal trays

were set up without ligand with P450cam and P450cam L244A. When calculating the concentration of the mutant protein by reduced CO spectrum, we found the concentration to be 4 times that calculated by the BCA assay (Fig. 1).

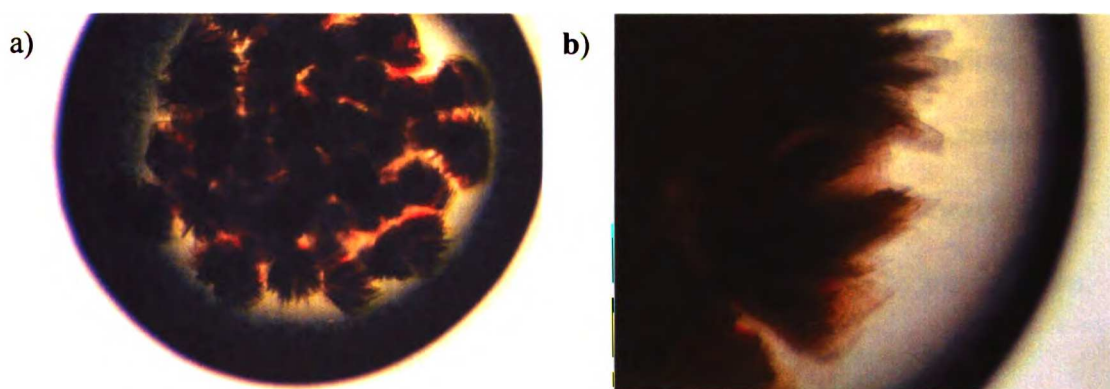


Figure 1. L244A crystals. Crystal droplets set up with 50 mg/mL P450cam L244A protein. The concentration is calculated by reduced CO spectrum. a) The overall droplet is saturated with protein. b) The crystals show distinct edges and are transparent.

The actual concentration of the protein in the droplet as calculated by the BCA assay is closer to 250 mg/mL than 50 mg/mL as suggested by the reduced CO spectrum. At this concentration we see the droplet is saturated with protein; distinguishing individual crystals is nearly impossible. We are able to observe distinct crystal edges at the edge of the droplet (Fig. 1b). All L244A concentrations were therefore determined by BCA analysis.

Despite the L244A protein having a single mutation buried deep in the active site, we find some surprising differences when comparing its crystal formation to that of the wild-type (Fig. 2). While the wild-type protein forms distinct, square and rectangular crystals, the L244A crystallizes as thin overlapping plates. This result is unexpected as surface contacts that might affect crystal packing should be unchanged by the mutation. It is possible that the L244A protein is less stable at room temperature, and that it precipitates under the crystallization conditions. The overlapping plates may then grow from an aggregation of protein rather than a nucleation event as in the wild-type.

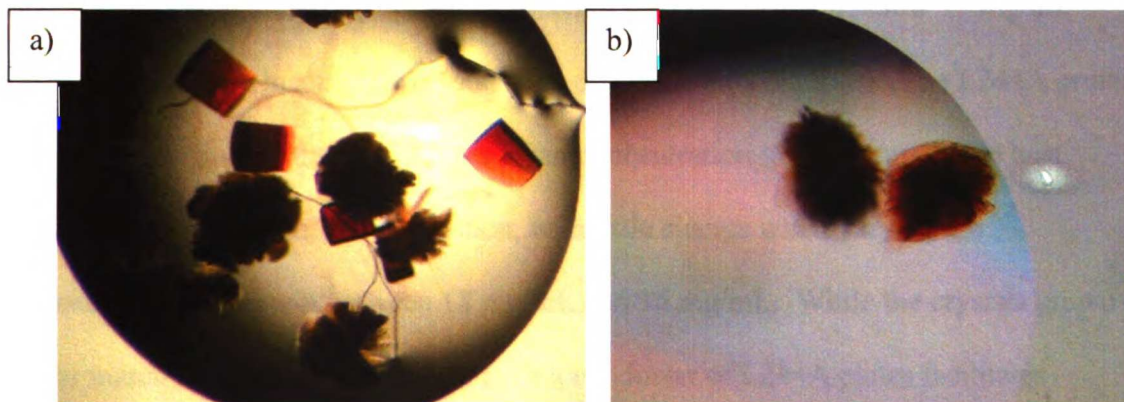


Figure 2. Wild-type and L244A crystals. a) Wild-type and b) L244A crystal morphologies under identical conditions.

Several different conditions were evaluated in an effort to grow better L244A crystals. Multiple trays were set up at identical concentrations and then incubated at various temperatures. Most crystals, of wild-type and mutant protein, grow over night; however, temperature has a great impact on the rate of crystal formation and crystal morphologies (Fig. 3). At room temperature, overlapping plates are observed. At 4 °C no crystals are observed even after two weeks. At 10 °C small crystal seeds are observed, but they do not grow larger even after two weeks. These trays, if transferred to room temperature grow crystals similar to trays incubated initially at room temperature. Trays set up at 15 °C show the most interesting behavior. The crystal morphology at this temperature changes from overlapping plates to star burst rods. The rods are able to diffract, but the thinness and small size of the crystals limits their resolution.

We also varied protein concentration in an effort to improve the crystals of P450cam L244A (Fig. 4). We already noted the large discrepancy in concentration when calculated by reduced CO spectrum versus BCA analysis. At concentrations closer to 50 mg/mL, there are subtle changes in the crystal morphologies. At 50 mg/mL several overlapping plate clusters are evident (Fig. 4a), while at 25 mg/mL we generally observe one or two crystal clusters (Fig. 4b). Lowering the concentration below 15 mg/mL generally results in incomplete crystals that do not diffract (Fig. 4c). The L244A protein was therefore finely titrated about a median concentration of 25 mg/mL. The best crystals are grown at this concentration, with little change when varying the concentration of protein between 18 mg/mL and 36 mg/mL. While the crystals grown never match the wild-type morphology, a lone cluster of L244A plates facilitates

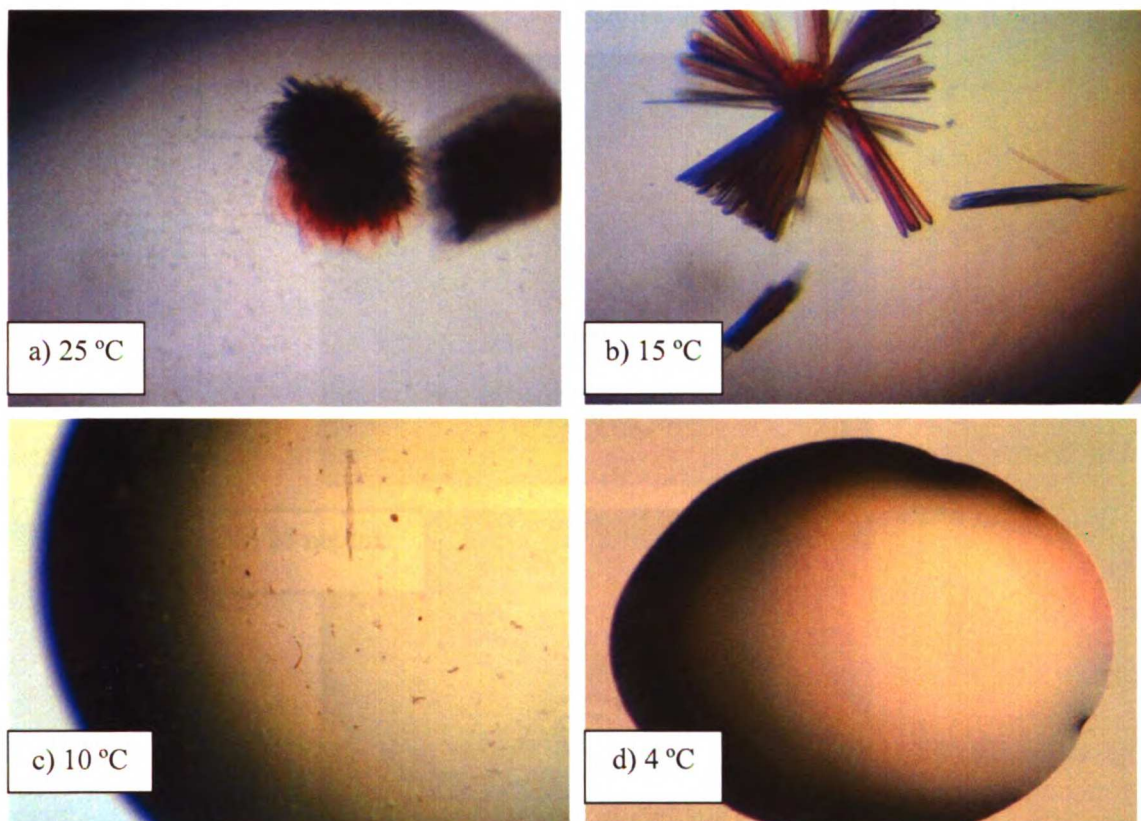


Figure 3. Temperature titrations of L244A crystal trays. Droplets incubated at a) 25 °C, b) 15 °C, c) 10 °C, and d) 4 °C.

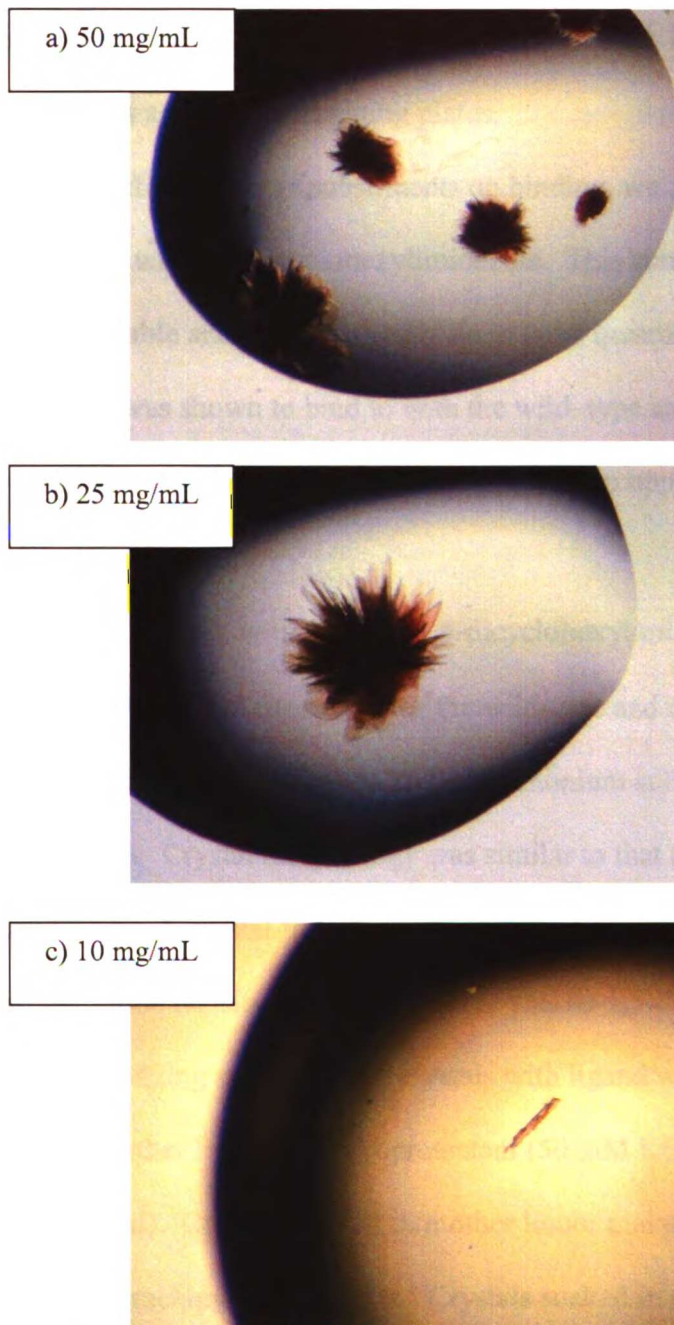


Figure 4. L244A concentration titrations. Crystal trays of P450cam L244A set up at a) 50 mg/mL, b) 25 mg/mL, and c) 10 mg/mL as determined by BCA analyses.

The effects of ligands on crystal formation have also been explored. Trays were set up in the aforementioned mother liquor with saturating amounts of camphor. The wild-type protein crystallizes as before, forming large square and rectangular crystals. The L244A mutant shows no improvement in crystal morphology in the presence of camphor and crystallizes again as overlapping plates.

To explore the effects of large substituents on binding, we attempted to co-crystallize both proteins with 1,5-dicyclohexylimidazole. This compound was chosen because (a) it is purchasable and therefore available in large quantities necessary for crystal screening, (b) it was shown to bind to both the wild-type and the L244A mutant with micromolar affinity, and (c) it is predicted to approach an upper limit of the wild-type active site flexibility.

Crystal trays were set up with 580 μM 1,5-dicyclohexylimidazole diluted from a stock solution in DMSO. Crystal formation for the wild-type and mutant only began at increased ammonium sulfate concentrations and so ammonium sulfate was screened at saturations from 46-52%. Crystal morphology was similar to that observed for the substrate free conditions. A crystal from the wild-type protein was solved to 1.6 \AA under these conditions, but initial density showed that there was no 1,5-dicyclohexylimidazole bound to the protein. Soaking substrate-free crystals with ligand was also attempted in both ligand saturated mother liquor and cryoprotectant (50 mM KCl, 25% ammonium sulfate, and 30% glycerol). Crystals soaked in mother liquor maintained their integrity, but showed significant cracking after minutes. Crystals soaked in cryoprotectant began to melt in under ten minutes.

Ultimately, the goal of the project was to understand what effects the mutation might have in mitigating imidazole affinity. To this end several screens were established with varying concentrations of imidazole in the droplet. Imidazole was added from a stock solution in water at final concentrations of 1, 2, 25, and 50 mM in the mother liquor (resulting in a 2-fold dilution in the droplet). At high concentrations of imidazole (25 or 50 mM) the crystal morphology for the wild-type began to resemble that of the mutant (Fig. 5). The mutant protein crystal morphology did not change at high concentrations of imidazole.

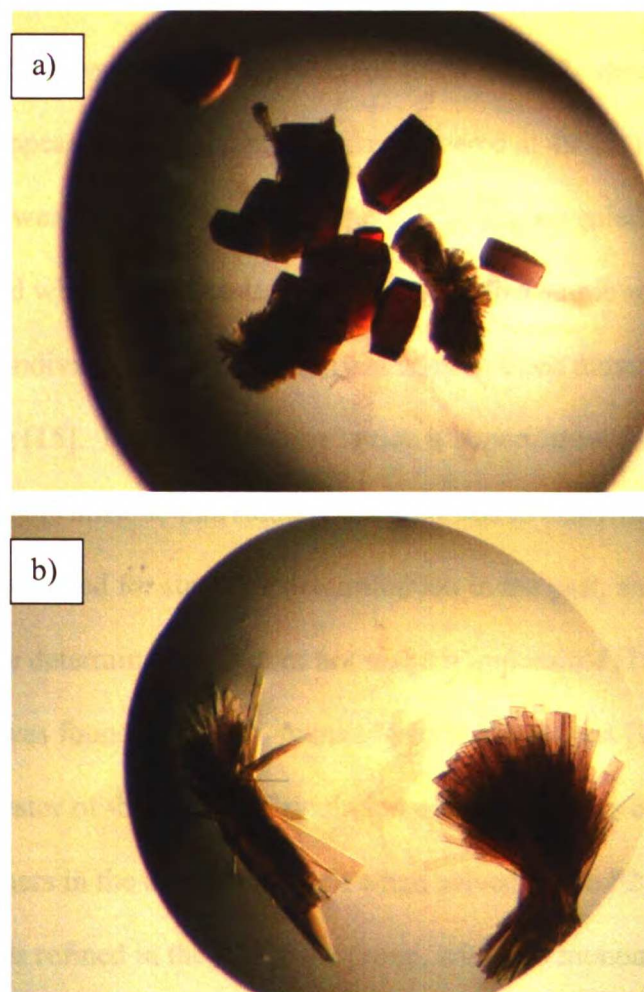


Figure 5. Effects of imidazole on wild-type crystal morphology. a) Droplet containing no imidazole in the mother liquor. b) Droplet containing 25 mM imidazole.

Crystals chosen for diffraction included: one substrate free form of the L244A mutant, one P450cam wild-type grown in the presence of 25 mM imidazole, and one L244A protein grown in the presence of 25 mM imidazole.

Data Collection and Processing

P450cam L244A, substrate-free protein

Crystal edges from the L244A overlapping plates were broken from the crystal clusters and looped. The average size of the plates was 0.30 x 0.30 x .02 mm. Data was collected over 180° with an oscillation angle of 1°. The final resolution was 2.15 Å (Table 1). Initially, the crystal data was processed in the $C2_12_12_1$ space group, but even though the density appeared appropriate, the R_{free} hovered at 49%. The data was reprocessed in the lower $P2_1$ space group and the crystal was revealed to be twinned. Twinning is observed when two crystals grow together into a single space group [13, 14]. If the lattices of the individual lattices perfectly overlap in three dimensions, the result is merohedral twinning [15]. In these cases one lattice is superimposed over the other and the intensity of the spots must be individually decomposed to describe each lattice. Twinned data has been used for structure determination in the past, and while it complicates structure determination, it does not make it impossible [16-18].

The protein was found to be 44% twinned by the Yeates test [19] for twinning with a twinning operator of $-h, -k, h+l$. Calculation of the Matthews coefficient [20] revealed two monomers in the asymmetric unit when solved in the $P2_1$ space group. When the protein was refined in the $P2_1$ space group, with two monomers (related by a pure translation), the R_{free} immediately dropped to 29%.

Table 1. Data collection statistics for the L244A substrate-free protein.

Data Collection	
Wavelength (Å)	1.115869
Resolution Range (Å)	50.0 - 2.15 (2.23 -2.15)
Unique Reflections	41265
Completeness (%)	91.9 (87.3)
Rsym	.073 (363)
I/σ	11.7(2.1)
Redundancy	2.0 (1.8)
Refinement	
	P2₁
Resolution Range (Å)	50.0 - 2.15 (2.23 -2.15)
Number of Reflections/Test Set	37919 (3556)/3792(356)
Rcryst/Rfree %	24.89(23.54)
A	57.704
B	59.853
C	114.124
Alpha	90
Beta	104.69
Gamma	90
Twin Fraction	0.38
Twin Operators	-h, -k, h+l

P450cam, imidazole-bound structure

Several crystals were collected for the wild-type protein in the presence of imidazole. The crystal used in data collection was square in shape with approximate dimensions of 0.3 x 0.3 x 0.2 mm. Data was collected over 180° with an oscillation angle of 1°. The wild-type crystals grown in the presence of imidazole did not diffract as well. There was greater streaking of the spots and they were more elliptically shaped. Also there was more noise and aberrant spots present in the wild-type crystals grown in the presence of imidazole compared to those grown without substrate (data not shown). Because of this, only 120° of the collected data was used in the final solution, but because of the high symmetry of the $P2_12_12_1$ space group, the redundancy is still high (Table 2). Wild-type crystals, despite differing from the mutant by only a single point mutation, crystallize in a different space group. We find that the wild-type is solved as $P2_12_12_1$ with unit cell dimensions $a = 36.667$, $b = 103.75$, and $c = 106.485$, in agreement with published literature [9].

Table 2. Data collection statistics for wild-type P450cam crystallized in the presence of imidazole.

Data Collection

Wavelength (Å)	1.60
Resolution Range (Å)	50.0 - 1.5 (1.55 -1.5)
Unique Reflections	65716
Completeness (%)	98.7 (96.8)
Rsym	0.053 (0.432)
I/σ	11.6 (2.74)
Redundancy	6.9 (6.5)

Refinement

P2₁2₁2₁

Resolution Range (Å)	50.0 - 1.5 (1.55 -1.5)
Number of Reflections/Test Set	64857 (6219) /6486 (622)
Rcryst/Rfree %	24.42 (21.34)
A	36.667
B	103.75
C	106.485
Alpha	90
Beta	90
Gamma	90

P450cam L244A, imidazole-bound structure

The final structure solved was that of the L244A mutant in the presence of imidazole. As before, the L244A protein crystallized in overlapping plates and an edge was broken off for diffraction. Average dimensions of the fragments were 0.30 x 0.30 x .02 mm. Data was collected over 180° with an oscillation angle of 1°. Unlike previous crystallizations of the L244A, these crystals were not twinned. The space group is $P2_1$ and the dimensions of the asymmetric unit are $a = 56.599$, $b = 58.784$, and $c = 57.441$ (Table 3). Compared to that of the L244A crystal structure solved in the absence of substrate and found to be twinned, the c axis of the asymmetric unit is half the length. In agreement with this smaller unit cell, we find that there was only one molecule in the asymmetric unit.

The final resolution for P450cam L244A crystallized in the presence of imidazole was 2.2 Å. The overall completeness and redundancy was a little bit low, particularly in the highest resolution bin (65.9 %), however, the data collections statistics were sufficient to solve a structure of this resolution.

Table 3. Data collection statistics for P450cam L244A in the presence of imidazole.

Data Collection

Wavelength (Å)	1.60
Resolution Range (Å)	50.0 - 2.2 (2.28 -2.2)
Unique Reflections	18968
Completeness (%)	91.4 (65.9)
Rsym	.087 (.294)
I/σ	11.6(2.74)
redundancy	3.2(2.5)

Refinement

P21

Resolution Range (Å)	50.0 - 2.2 (2.28 -2.2)
Number of Reflections/Test Set	17324 (1246) / 1732 (125)
Rcryst/Rfree %	28.35 (21.03)
A	56.599
B	58.784
C	57.441
Alpha	90
Beta	105.287
Gamma	90

Structures

P450cam L244A, substrate-free protein

The final L244A protein contained two monomers in the asymmetric unit and 163 water molecules. Residues E133 and K313 of monomer A were modeled as alanines because of poor density of their side chains, but both are surface residues. The protein monomers in the asymmetric unit were related by a pure translational transformation (Fig. 6), another indication of the twinning we previously described. The density about monomer A was slightly clearer than monomer B and this is the monomer we will refer to when describing details of the L244A, substrate-free mutant structure.

The overall structure is extremely similar to that of the wild-type protein with a global rmsd of less than 1.6 \AA^2 , somewhat surprising as the crystal morphologies, space groups, and unit cell dimensions are distinct between the two proteins. The primary differences between P450cam and P450cam L244A are in the active site. In Chapter 2, prior to any structural determination of the mutant, the active site was modeled simply as a three atom deletion, an alanine to leucine mutation. A difference map, calculated by superimposing the wild-type structure over the L244A density, shows that the mutation introduces a cavity smaller than that expected for the removal of three atoms (Fig. 7a). The largest difference between the proteins is on the helix near the mutation. The I helix for the L244A mutant is shifted in toward the active site (Fig 7b).



Figure 6. The asymmetric unit of P450cam L244a substrate-free protein. Two monomers are present in the asymmetric unit related by a pure translational transformation. The monomers are both colored by b-factor. Blue (1-10), Green (11-20), Yellow (21-30), Orange (>30).

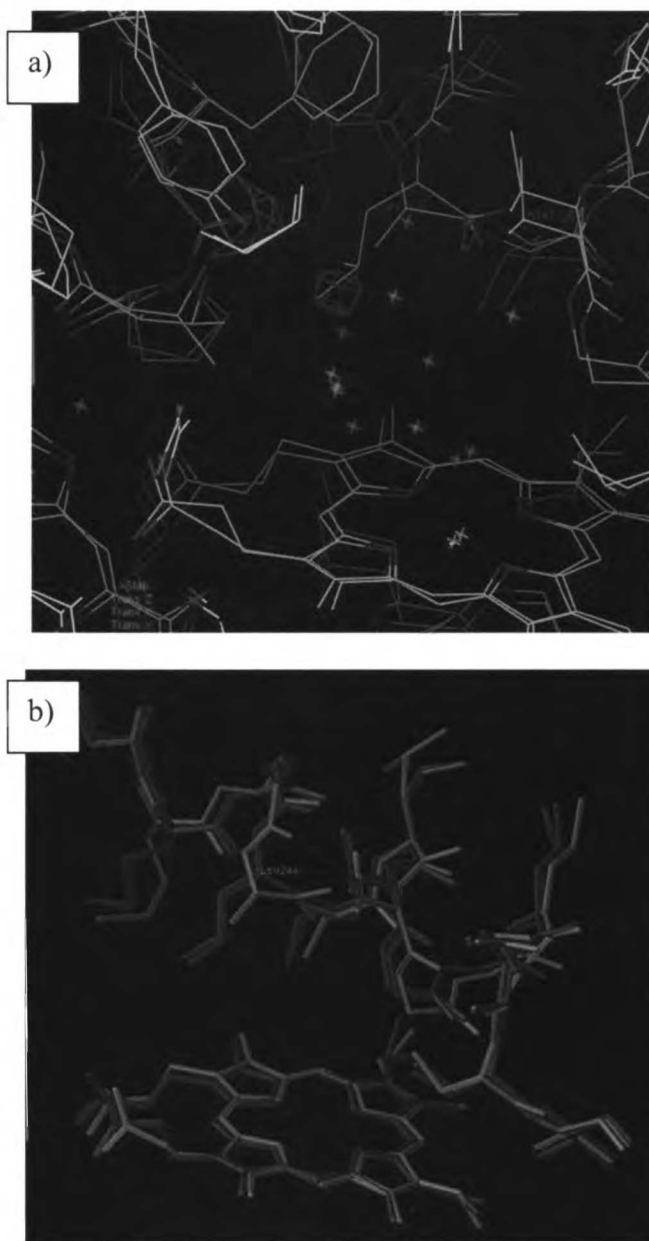


Figure 7. P450cam L244A substrate-free active site. a) Density difference map of the wild-type and mutant P450 enzymes showing. Red contour maps indicate points of deletion. Comparing the wild-type structure shows only a small patch of negative density at the leucine mutation site. b) Comparison of substrate-free P450cam (red), camphor bound P450cam (green), and substrate-free P450cam (L244A) (yellow). The helix shown is shifted about the L244A mutation to fill the gap created by the mutation.

Comparing the camphor bound [21] and camphor free [22] wild-type proteins shows little difference between the active site structures (Fig. 7b). It is interesting that despite all the induced fit effects described in Chapter 1, the natural substrate requires no changes in the active site to bind. The L244A mutant closely resembles the active-site state of the substrate-free wild-type. From our disubstituted ligand studies (Chapter 2), it is known that the L244A mutation is able to accommodate larger ligands, and must thus be more able to expand to accommodate those ligands than the wild-type.

Three possible effects may account for the altered imidazole affinity. One can hypothesize that the active site is sterically deformed, preventing optimal imidazole coordination, but with the substrate-free structure, no such changes are immediately evident. Alternatively, the heme might be puckered or skewed, preventing proper bonding to the imidazole, but we find that the heme overlaps identically between wild-type and mutant. Finally, one might imagine the water coordination in the active site to be distinct between the proteins and that the cost of solvating imidazole would be different between the mutant and wild-type. The mutant protein has seven water atoms clearly defined in the active site, while the wild-type has six (Fig 8a). The water atom coordinated to the heme iron is the same distance in both proteins (Fig. 8b) (this will be important in justifying our imidazole placement in the active site in the next section). The water placements alone do not suggest any difference in imidazole affinity between the two proteins and to understand the differences in imidazole affinity we must look at the imidazole-bound structures.

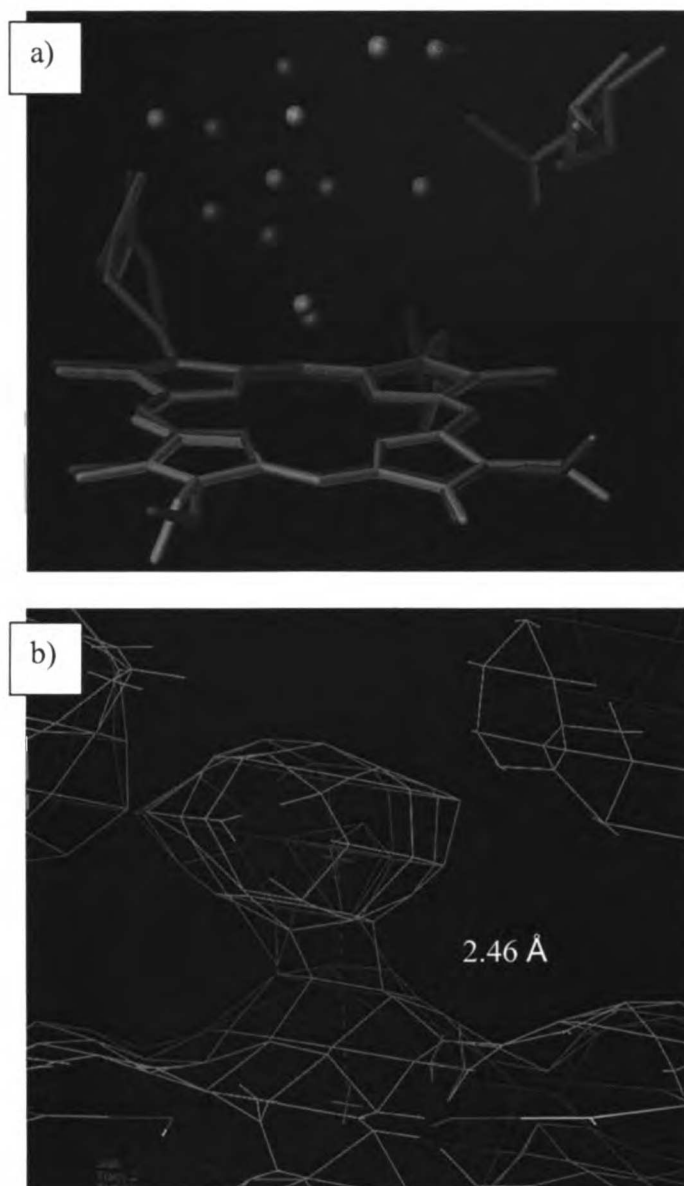


Figure 8. P450cam L244A substrate-free active site waters. a) Placement of the L244A (yellow) and wild-type (red) water atoms. b) Density map showing the distance to the iron coordinated water in the L244A mutant.

Imidazole Bound Structures for the Wild-Type and L244A Proteins

The wild-type protein was solved to 1.5 Å in the presence of imidazole. A total of 416 waters were added over 5 cycles of refinement. The overall structure is very similar to that of substrate-free wild-type, and even in the active site, no backbone or side chain rearrangements are evident. P450cam is the best characterized cytochrome P450 enzyme to date, but no structure exists of the imidazole-bound protein. The density around the imidazole ligand is incomplete (Fig. 9a). This is most likely due to the few contacts that are made with the small ligand. The clearest indication that the ligand is in indeed bound is the distance from the heme iron to the N3 of imidazole. If the ligand is refined rigidly with the imidazole N3-heme iron distance at 2.45 Å, and an fo minus fc map is generated, a spot of negative density becomes evident above the imidazole and a spot of positive density is seen closer to the heme (Fig. 9b). This data suggests that the ligand atom must be placed closer to the heme than a water molecule. When the Van der Waals parameters of the heme are adjusted to allow for closer distances between the imidazole and the heme iron, the fo minus fc maps looks more appropriate. The density also indicates that the proximal ligand is larger than a single water atom coordinated to the heme iron.

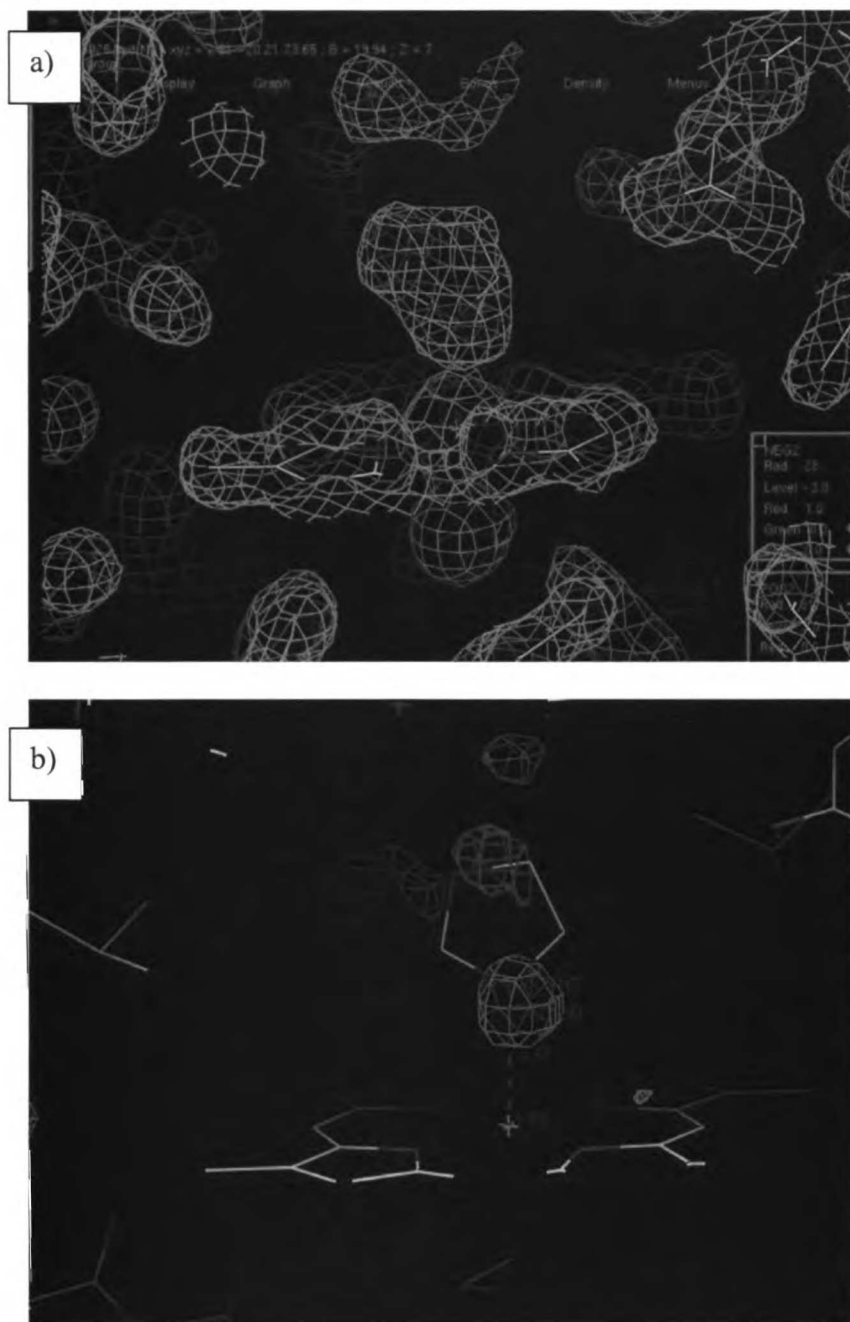


Figure 9. Wild-type P450cam imidazole-bound active site density. a) The electron density at the active site fits closely around the imidazole ligand. b) A difference map generated with the imidazole N3-heme iron distance at 2.45 Å. Negative density is only seen when the imidazole ligand is placed further from the heme than suggested by the density. Positive density suggests the imidazole coordinates to the heme closer than a water molecule.

Occupancies for the ligand and water atom were also refined. The occupancy of the N3 in imidazole is 1.0, while the other atom occupancies lie between 0.5 and 0.85. This is in agreement with our knowledge of imidazole coordination. Sigma bonding occurs between the N3 nitrogen and the heme. The heme iron dxy and dyz orbitals backbond with the electron deficient p-orbital of the imidazole N3. The other atoms in imidazole are free to move and co-crystal structures of other imidazole-P450 complexes have shown variation in the angle of the imidazole N3 nitrogen relative to the heme as well as deviation from perpendicularity of the heme-imidazole molecules [10, 23, 24]. These data, taken together, suggest that the imidazole is bound in the wild-type structure.

The L244A protein was also solved with 50 mM imidazole in the mother liquor. The final structure contained 116 water molecules added over 4 rounds of refinement. The density at the imidazole ligand is again ambiguous (Fig. 10) and we can only postulate that the imidazole is bound, again primarily on the same criteria by which we inferred its presence in the wild-type. We see here that the density, suspected to be imidazole, indicates that the imidazole-heme planes significantly deviate from perpendicularity. As with the wild-type, the distance from the N3 of the imidazole to the heme iron is less than would be expected if water was coordinated and the density suggests more than a single atom coordinated to the heme. Because dithiothreitol was used in the mother liquor and is known to coordinate to the heme iron, a round of refinement was performed with dithiothreitol present as the proximal ligand instead of imidazole. Difference maps suggest a poor fit for dithiothreitol (data not shown).



Figure 10. P450cam L244A imidazole-bound active site density map. The density of the imidazole suggests that the heme and imidazole planes are almost parallel.

The orientation of the imidazole as described by the density could certainly account for its mitigated affinity (Chapter 2). The L244A structure solved in the presence of imidazole closely resembles the L244A substrate-free structure. The movement of the I helix toward the heme to fill the space made by the L244A mutation is also present in the imidazole co-crystallized structure. While neither a leucine nor an alanine at position 244 interacts directly with the imidazole, other residues on the I helix do. The movement of the helix brings V247 in closer proximity to the heme and the imidazole (Fig. 11a). With the valine in such a close proximity to the heme, it is not possible for the imidazole to maintain perpendicularity to the heme plane (Fig. 11b). We hypothesize that this may be the cause of the mitigated affinity for imidazole observed with the L244A mutation.

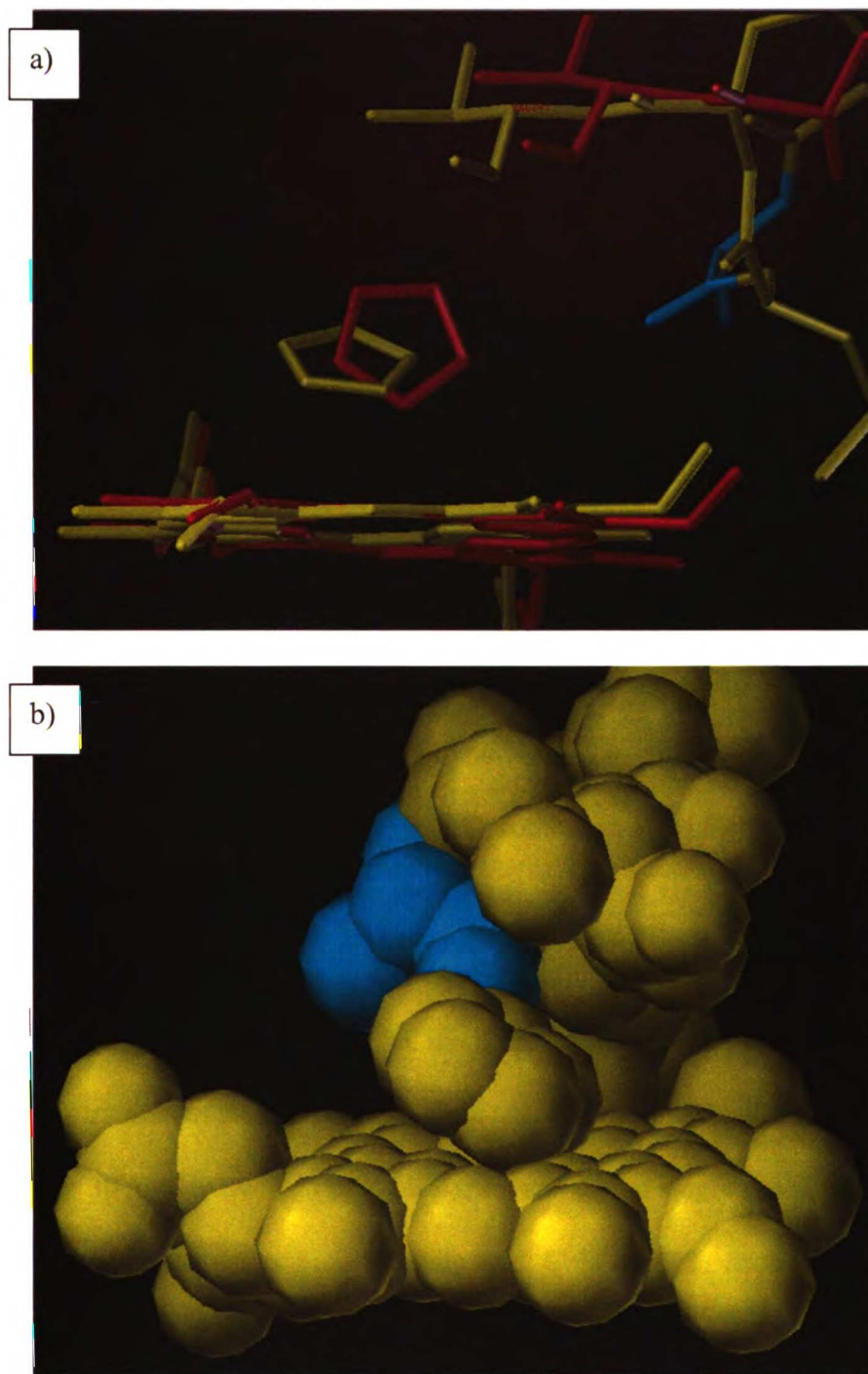


Figure 11. Effects on imidazole binding by the L244A mutation. a) Wild-type (red) and L244A (yellow) stick representations of the active site. The I helix is shifted toward the heme bringing valine 247 directly over the imidazole in the mutant protein. b) Space-filled model of the L244A showing the close proximity of V247 to the imidazole. Residue 244 is shown in light blue.

Experimental Section

Molecular Biology

Quick change PCR was performed in puc19. The primers included the restriction sites Nde I and Xbe I as well as a Nar I site introduced at the L244A mutation point that was used in screening for the presence of the mutant.

Protein Expression and Purification

Cytochrome P450cam and the P450cam L244A mutant were expressed in *Escherichia coli* as previously reported [25]. Proteins were purified as before with the exception that camphor was not included in any wash, elution, or loading buffers. No ammonium salt cut was performed in purifying the protein and an additional gel filtration column was run after the Q-sepharose column to yield protein sufficiently pure for crystallography. Protein was concentrated to 50 mg/mL in amicon-2000 centrifugal concentrators prior to storage. When possible, protein was used for crystal screening without freezing.

Protein concentration of the L244A mutant was determined by BCA analysis. Bicinchoninic acid was mixed 50:1 with CuSO₄ and 2 mL of the mixture was added to various concentrations of protein in small glass culture tubes. The tubes were placed in a 37 °C oven for 30 min, allowed to cool for 4 min, and then a single wavelength at 562 nm was recorded against a blank of bicinchoninic acid: CuSO₄ (50:1). Bovine serum albumin was used to calibrate the measurements. In order to insure that heme did not interfere in the colorimetric assay, cytochrome c was also used to calibrate the assay.

Crystal Screening, Data Collection, and Refinement

Crystal trays were set up at the temperatures indicated with mother liquor containing 50 mM KPi, 250mM KCl, 50 mM DTT, and ammonium sulfate ranging from 36-52%. Protein was combined 1:1 with mother liquor in the hanging drop. The cryoprotectant was 50 mM KCl, 25% ammonium sulfate, and 30% glycerol. Imidazole was added to mother liquor from a stock solution in water. All other ligands were added from a stock solution in DMSO. Crystals at room temperature generally formed overnight and completed crystal growth within two days. Crystals at 16 °C required 3-4 days to form and sometimes up to two weeks. Crystals at temperatures less than 16 °C formed over two weeks or not at all.

Data Collection and Processing

Data was collected at the Berkeley ALS (Advanced Light Source) on beamlines 8.3.1 and 8.3.2. Data reduction was done with DENZO [26] and MOSFLM [27]. Protein unit cell rotational and translational searches were done with CCP4 [28] and refinement was carried out with CNS [29]. Manual rebuilding of the proteins and visualization was accomplished with O [30].

Conclusions

The L244A mutant, surprisingly yielded crystals with different morphologies than the wild-type. The tertiary structures of the proteins are highly similar (rmsd < 2 Å);

however, they crystallize in different space groups ($P2_12_12_1$ for the wild-type, $P2_1$ for the L244A mutant).

The L244A crystals grown in the absence of ligand, were found to be twinned. The asymmetric unit contained two monomers related by a pure translational transformation. We were able to detwin the data using CNS and to resolve the substrate-free structure to 2.2 Å. The substrate free structure showed similar overall tertiary structure topology to that of the wild-type protein, except within the active site. The I helix, into which the L244A mutation was introduced, is shifted in the active site to fill the “hole” created by the leucine to alanine mutation. The final result of this rearrangement is an active site volume that differs from that of the wild-type by the space of one atom as opposed to three.

Both the wild-type and L244A proteins were crystallized in the presence of imidazole. While the electron density alone cannot place the ligand in the proteins with complete certainty, the electron density distance from the heme iron suggests that water is not coordinated in these structures. For the wild-type structure, which was refined to 1.5 Å, we are able to confidently refine occupancies of water and ligand atoms. The occupancies reinforce our knowledge of imidazole coordination; that is, the N3 nitrogen is relatively fixed, while the other atoms of the small ligand are without significant contacts in the active site and able to move more freely.

The putative binding mode of imidazole in the L244A mutant is nearly parallel to the plane of the heme. As observed for the substrate free form, the I helix is shifted toward the heme, bringing valine 247 in closer proximity to the heme than in the wild-type enzyme. With this valine jutting into the active site, the imidazole is unable to bind

perpendicular to the plane of the heme. This would certainly explain the altered imidazole affinity we observe between the wild-type and mutant enzymes.

References

1. Verras, A., I.D. Kuntz, and P.R. Ortiz de Montellano, *Computer-assisted design of selective imidazole inhibitors for cytochrome p450 enzymes*. J Med Chem, 2004. **47**(14): p. 3572-9.
2. Vanden Bossche, H., *Biochemical targets for antifungal azole derivatives: hypothesis on the mode of action*. Curr Top Med Mycol, 1985. **1**: p. 313-51.
3. Albengres, E., H. Le Louet, and J.P. Tillement, *Systemic antifungal agents. Drug interactions of clinical significance*. Drug Saf, 1998. **18**(2): p. 83-97.
4. Cha, R. and J.D. Sobel, *Fluconazole for the treatment of candidiasis: 15 years experience*. Expert Rev Anti Infect Ther, 2004. **2**(3): p. 357-66.
5. Lamb, D.C., et al., *The mutation T315A in Candida albicans sterol 14alpha-demethylase causes reduced enzyme activity and fluconazole resistance through reduced affinity*. J Biol Chem, 1997. **272**(9): p. 5682-8.
6. Rodero, L., et al., *G484S amino acid substitution in lanosterol 14-alpha demethylase (ERG11) is related to fluconazole resistance in a recurrent Cryptococcus neoformans clinical isolate*. Antimicrob Agents Chemother, 2003. **47**(11): p. 3653-6.
7. Kamai, Y., et al., *Characterization of mechanisms of fluconazole resistance in a Candida albicans isolate from a Japanese patient with chronic mucocutaneous candidiasis*. Microbiol Immunol, 2004. **48**(12): p. 937-43.
8. Nickerson, D.P. and L.L. Wong, *The dimerization of Pseudomonas putida cytochrome P450cam: practical consequences and engineering of a monomeric enzyme*. Protein Eng, 1997. **10**(12): p. 1357-61.
9. Nickerson, D., L.L. Wong, and Z. Rao, *An improved procedure for the preparation of X-ray diffraction-quality crystals of cytochrome p450cam*. Acta Crystallogr D Biol Crystallogr, 1998. **54** (Pt 3): p. 470-2.
10. Yano, J.K., et al., *Crystal structure of a thermophilic cytochrome P450 from the archaeon Sulfolobus solfataricus*. J Biol Chem, 2000. **275**(40): p. 31086-92.

11. Poulos, T.L. and R. Raag, *Cytochrome P450cam: crystallography, oxygen activation, and electron transfer*. *Faseb J*, 1992. **6**(2): p. 674-9.
12. Stoscheck, C.M., *Quantitation of protein*. *Methods Enzymol*, 1990. **182**: p. 50-68.
13. Catti, M. and G. Ferraris, *Twinning by Merohedry and X-ray Crystal Structure Determination*. *Acta Crystallogr A*, 1976. **32**: p. 163-165.
14. Donnay, G. and J.D.H. Donnay, *Classification of Triperiodic Twins*. *Can. Mineral*, 1974. **12**: p. 422-425.
15. Yeates, T.O. and B.C. Fam, *Protein crystals and their evil twins*. *Structure, Folding, and Design*, 1999. **7**(2): p. 25-29.
16. Breyer, W.A., et al., *On the molecular-replacement problem in the presence of merohedral twinning: structure of the N-terminal half-molecule of human lactoferrin*. *Acta Crystallogr D Biol Crystallogr*, 1999. **55** (Pt 1): p. 129-38.
17. Contreras-Martel, C., et al., *Crystallization and 2.2 Å resolution structure of R-phycoerythrin from *Gracilaria chilensis*: a case of perfect hemihedral twinning*. *Acta Crystallogr D Biol Crystallogr*, 2001. **57**(Pt 1): p. 52-60.
18. Redinbo, M.R. and T.O. Yeates, *Structure determination of plastocyanin from a specimen with a hemihedral twinning fraction of one-half*. *Acta Crystallogr D Biol Crystallogr*, 1993. **49**(Pt 4): p. 375-80.
19. Yeates, T.O., *Detecting and overcoming crystal twinning*. *Methods Enzymol*, 1997. **276**: p. 344-58.
20. Kantardjieff, K.A. and B. Rupp, *Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals*. *Protein Sci*, 2003. **12**(9): p. 1865-71.
21. Poulos, T.L., B.C. Finzel, and A.J. Howard, *High-resolution crystal structure of cytochrome P450cam*. *J Mol Biol*, 1987. **195**(3): p. 687-700.
22. Poulos, T.L., B.C. Finzel, and A.J. Howard, *Crystal structure of substrate-free *Pseudomonas putida* cytochrome P-450*. *Biochemistry*, 1986. **25**(18): p. 5314-22.

23. Raag, R., et al., *Inhibitor-induced conformational change in cytochrome P-450CAM*. *Biochemistry*, 1993. **32**(17): p. 4571-8.
24. Poulos, T.L. and A.J. Howard, *Crystal structures of metyrapone- and phenylimidazole-inhibited complexes of cytochrome P-450cam*. *Biochemistry*, 1987. **26**(25): p. 8165-74.
25. Wagner, G.C., et al., *Apoprotein formation and heme reconstitution of cytochrome P-450cam*. *J Biol Chem*, 1981. **256**(12): p. 6262-5.
26. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode*. *Methods in Enzymology*, 1997. **276**: p. 307-326.
27. Leslie, A.G.W., *Recent changes to the MOSFLM package for processing film and image plate data*. *Joint Newsletter on Protein Crystallography*, 1992. **26**.
28. 4., C.C.P., *The CCP4 suite: programs for protein crystallography*. *Acta Crystallogr D Biol Crystallogr*, 1994. **50**(760-763).
29. A.T.Brunger, P.D.A., G.M.Clore, W.L.Delano, P.Gros, R.W.Grosse-Kunstleve, J.-S.Jiang, J.Kuszewski, M.Nilges, N.S.Pannu, R.J.Read, L.M.Rice, T.Simonson, G.L.Warren, *Crystallography and NMR system: a new software suite for macromolecular structure determination*. *Acta Crystallogr D Biol Crystallogr*, 1998. **54**: p. 904-921.
30. Jones, T.A., et al., *Improved methods for binding protein models in electron density maps and the location of errors in these models*. *Acta Crystallogr A*, 1991. **47**: p. 110-119.

Chapter 6

Conclusions and Future Directions

Introduction

With any graduate thesis, the possibilities for future work are almost endless. Extensively describing the future directions I might imagine for the projects described in this thesis would be like writing a Christmas wish list. UCSF is a highly collaborative environment and my experience with graduate research is evidence of this. During my stay at UCSF I have done DOCK, molecular dynamics, organic synthesis, protein work, assay development, kinetics, and crystallography. Behind every new technique I learned were always professors, students, and post docs encouraging and tutoring me. Because of them, I could happily suggest a myriad of new techniques to apply to the research presented in this dissertation. Instead I will try to keep this last section brief and not deviate too far into fantasy.

CovDOCK and Library Design

In Chapter 2, we showed the development and application of CovDOCK. In Chapter 3, we furthered our work with this methodology by applying it to *Mycobacterium tuberculosis* CYP51[1] and expanding the virtual library with 4-substituted imidazoles.

CovDOCK was originally written by Geoffrey Skillman (Skillman, PhD Thesis, UCSF) into the scoring function of DOCK 4.0. Recently DOCK 5.0 was completed by Dimitri Moustakas (Moustakas, PhD Thesis, UCSF). With this version, DOCK was rewritten in the C++ object oriented language. This new design facilitates the inclusion of various scoring functions, minimizers, and other user designated modules. I would like to see a CovDOCK module written in C++ to increase its usage. Ideally CovDOCK

has applications not only for cytochrome P450 enzymes and heme containing enzymes, but also proteases and any enzyme system against which covalent inhibitors are targeted.

The library was expanded with 4-substituted imidazoles in Chapter 3, but the methods allow any type of library expansion. Triazoles and pyridines [2, 3] have shown to coordinate to heme iron similarly to imidazole and inclusion of these scaffolds could vastly augment the library size and diversity. No changes to the CovDOCK code would be required; the new libraries would simply have to be built with the appropriate atom numbering of the new scaffolds.

K_s and IC50 discrepancy

In Chapter 2, I reported a discrepancy between K_s and IC50 values for the disubstituted imidazole ligands against wild-type P450cam. While these ligands were shown to bind in the absence of camphor by spectroscopic measurements, inhibition assays showed that they did act as inhibitors. This anomaly has impact on P450 drug design. In Chapter 3, ligands were ruled out as inhibitors because their spectroscopic binding constants were no less than 18 μ M, but if K_s and IC50 values do not correlate, using spectroscopy as the criteria for proceeding with inhibitor assays is questionable.

In Chapter 4, kinetic analysis lends some insight into the different binding of these inhibitors and K_s assays in the presence of substrate differentiate the mono- and disubstituted imidazoles. Ultimately, to deconvolute these data more experiments are required. We have begun initial experiments with Chris Guilbert in Dr. Tom Jame's lab to computationally approach this phenomenon. Using an altered version of the PEDC (Path Explorations with Distance Constraints) algorithm which employs an rmsd

constraint [4], we are able to drive P450cam to an “open” form which resembles the open access channel found when the protein is co-crystallized with long ruthenium linked compounds. Next we hope to analyze the binding energetics of our mono- and disubstituted compounds by driving their entrance through the active site channel.

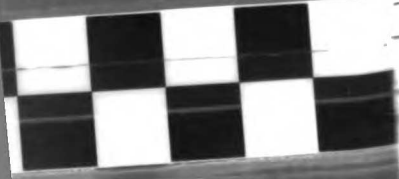
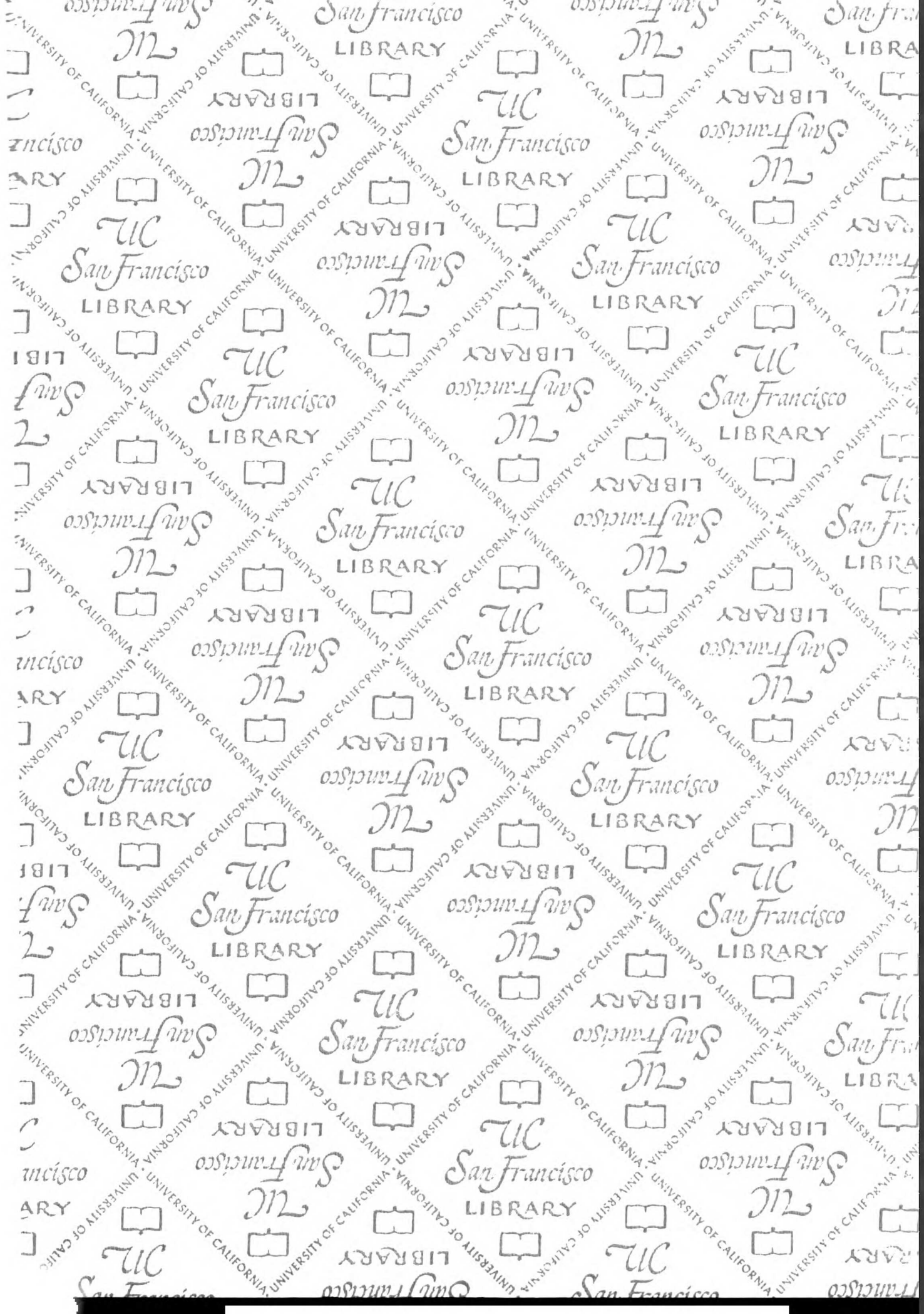
Mitigated L244A Affinity

Chapter 2 introduced the mitigated affinity of the L244A mutant against camphor and in Chapter 5 crystallography was used to explore the cause of this ablated affinity. While we are able to postulate at the structural factors which cause the discrepancy in imidazole affinity between wild-type and L244A proteins, the density of the active site ligands is not fully resolved for either enzyme.

Imidazole is a small, symmetric ligand with few contacts in the active site making its resolution difficult. There are larger ligands we could use to potentially reveal the source of the mitigated affinity. Ligands as small as 1-methylimidazole could be screened with wild-type and L244A crystals to afford a structure whose active site density is more defined.

1. Podust, L.M., et al., *Estriol bound and ligand-free structures of sterol 14alpha-demethylase*. *Structure (Camb)*, 2004. **12**(11): p. 1937-45.
2. Podust, L.M., T.L. Poulos, and M.R. Waterman, *Crystal structure of cytochrome P450 14alpha-sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors*. *Proc Natl Acad Sci U S A*, 2001. **98**(6): p. 3068-73.
3. Poulos, T.L. and A.J. Howard, *Crystal structures of metyrapone- and phenylimidazole-inhibited complexes of cytochrome P-450cam*. *Biochemistry*, 1987. **26**(25): p. 8165-74.
4. Guilbert, C. and D. Perahia, *A method to explore transition paths in macromolecules. Applications to phosphoglycerate kinase and hemoglobin*. *Computer Physics Communications*, 1995. **91**: p. 263-273.

UCSF LIBRARY



For reference

Not to be taken
from the room.

