

## MOLECULAR MODELING OF MAMMALIAN CYTOCHROME P450s

Fred K. Friedman<sup>1</sup>, Richard C. Robinson<sup>1</sup> and Renke Dai<sup>2</sup>

<sup>1</sup>Laboratory of Metabolism, National Cancer Institute, Bethesda, MD 20892, <sup>2</sup>Bristol-Myers Squibb, Pharmaceutical Candidate Optimization, Princeton, NJ 08543

### TABLE OF CONTENTS

1. Abstract
2. Introduction
3. P450 Structure
  - 3.1. Diversity of sequence
  - 3.2. Conservation of tertiary structure
4. Homology modeling
  - 4.1. Choice of template
  - 4.2. Sequence alignment
  - 4.3. Construction of backbone coordinates
  - 4.4. Loop insertion
  - 4.5. Adjustment of sidechain conformation
  - 4.6. Structure Refinement
5. Model validation
6. Conclusion and perspective
7. References

### 1. ABSTRACT

The cytochrome P450 enzymes collectively metabolize a wide range of xenobiotic and endogenous compounds. The broad substrate specificity of this superfamily derives from the multiplicity of P450s whose unique substrate specificity profiles reflect underlying differences in primary sequence. Experimental structures of P450s, where available, have provided great insight into the basis of substrate recognition. However, for those mammalian P450s whose structures have not been determined, homology modeling has become an increasingly important tool for understanding substrate specificity and mechanism. P450 modeling is often a challenging task, owing to the rather low sequence identity between target and template proteins. Although mammalian P450 models have previously been based on bacterial P450 structures, the recent advent of mammalian P450 structures holds great potential for generating more accurate homology models. Consequently, the substrate recognition properties of several mammalian P450s have been rationalized using the predicted substrate binding site of recently developed models. This review summarizes the major concepts and current approaches of molecular modeling of P450s.

### 2. INTRODUCTION

The cytochrome P450s comprise a superfamily of hemoprotein enzymes that oxidatively metabolize a wide variety of xenobiotic compounds such as drugs and carcinogens, and endogenous compounds such as steroids, prostaglandins, and fatty acids (1, 2). Their clinical relevance is evident from the fact that P450s metabolize >80% of marketed drugs. Inactivation, inhibition, and activation of these enzymes through P450-mediated drug biotransformation pathways may cause toxicity, drug

interactions, or decreased therapeutic efficiency. Furthermore, specific forms of P450 catalyze the conversion of certain compounds, like benzopyrene, into carcinogens which can form DNA-adducts that intercalate into double-stranded DNA and cause mutations, resulting in the encoding of mitogenic proteins with amino acid substitutions that result in their becoming oncogenic. Therefore, a major target of P450 research is to discover the forms of P450 that catalyze drug metabolism and carcinogen activation, and to design inhibitors of these P450s. Elucidation of the mode of substrate recognition by these P450s, using the molecular modeling process described in this paper, is a powerful tool in this research.

Although the substrate specificities of many mammalian P450s have been characterized, the structural basis of substrate recognition is not fully understood. However, the unique substrate specificities of individual P450s must reflect the composition and orientation of amino acid residues in the substrate binding site. The availability of crystal structures for substrate-free and -bound P450s has provided insights into the nature of the substrate-protein interaction. Such recently determined structures include rabbit P450 2B4 without substrate (3); rabbit P450 2C5 with progesterone (4), 4-methyl-N-methyl-N-(2-phenyl-2H-pyrazol-3-yl)benzenesulfonamide (5), and diclofenac (6); and human P450 2C9 with warfarin (7). However, such crystal structures are unavailable for the vast majority of mammalian P450s, leaving homology modeling as the only alternative for providing a structure that could be used to rationalize the observed substrate specificity. In addition to visualizing the conventional lock and key picture of a P450-substrate interaction, a homology model could also be used to explain structural transitions that correspond to the substrate induced conformation

changes that are often observed in P450 crystals (4-11).

Homology modeling is based on the presumed resemblance of the unknown P450 structure (the target) to a known P450 structure (the template). There is strong evidence for this premise, since similar major secondary structural components are observed in both bacterial and mammalian P450 crystals (3-13), and conserved sequences that govern this structural conservation are found to varying degrees within the P450 superfamily. Most simply stated, the goal of homology modeling is to apply the sequence similarity between target and template, to fashion a model structure. Prior to the availability of mammalian crystal P450s, early models of mammalian P450s were based on bacterial P450 structures (8-13). Since homology modeling entails mapping the one dimensional sequence of the target onto the three dimensional structure of the template, the most challenging aspect of modeling was the low sequence similarity (identity < 28%) between the mammalian P450 target and the bacterial P450 template. With the recent advent of mammalian P450 crystal structures (3-7), human P450 homology models have been constructed and used to elucidate their substrate recognition properties. Such models will be especially useful for predicting small molecule recognition by clinically relevant P450s. Consequently, the models can be used to identify potential drug-drug interactions early in the drug discovery and development process, and to predict the phenotype of new drug candidates by identifying P450s that potentially recognize these compounds. As noted above, molecular models can similarly be used to identify chemoprevention agents, and P450 inhibitors for P450-associated cancers and diseases.

This review presents current concepts of homology modeling of mammalian P450s, and complements several previous reviews (14-17). Recent progress utilizing the newly available mammalian P450 crystal structures is discussed, along with the steps and approaches to construction of a P450 model, including the all-important question of model validation. The inherent limitations of mammalian P450 modeling are discussed, along with recent findings on the role of protein dynamics in P450 structure and the implications of conformational heterogeneity.

### 3. P450 STRUCTURE

#### 3.1. Diversity of sequence

The number of P450 isoforms continues to grow, with > 1000 for named sequences for animals (see, for example, the web site at <http://drnelson.utm.edu/CytochromeP450.html>). Based on amino acid sequence identity, these are classified into about 40 families, whose members have > 40% sequence identity. A family is divided into subfamilies whose members are more closely related and have greater than 55% sequence identity. Although there is considerable overlap between P450 substrate recognition profiles, each individual P450 exhibits a unique profile that reflects the uniqueness of the residues within its substrate binding site. In addition to sequence identity, P450s are also classified according to their interaction with the redox partners during catalysis:

class I P450s (bacterial and mitochondrial) interact with an FAD containing reductase and iron-sulfur protein, while classes II P450s (microsomal) interact with a FAD and FMN-containing reductase. Furthermore, while the bacterial P450s are soluble, mammalian microsomal P450s are attached to the endoplasmic reticulum. P450 sequence diversity thus governs structural and functional diversity in three areas: 1) the substrate recognition site, 2) the binding site for the redox partner and 3) the membrane binding domain.

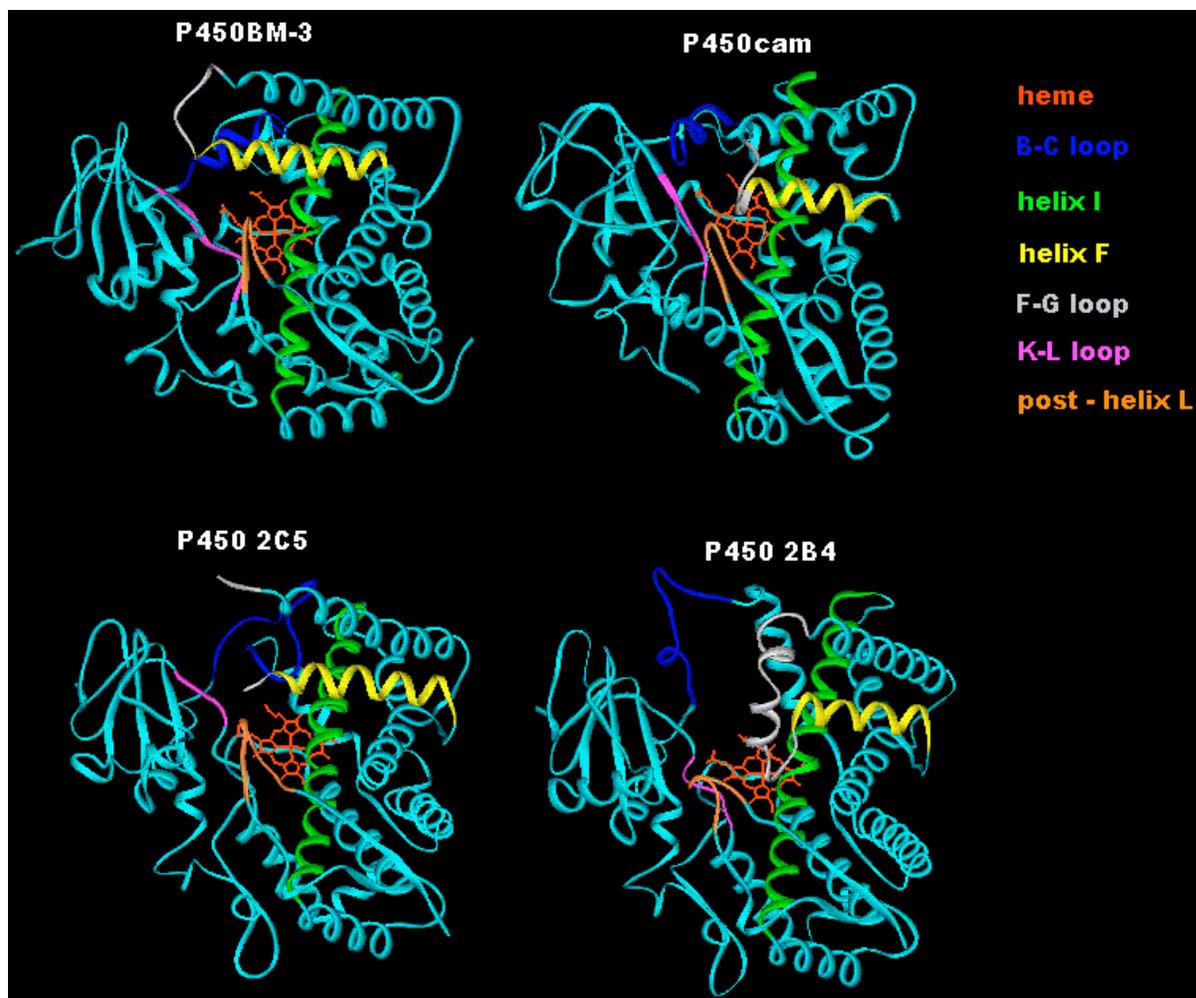
#### 3.2. Conservation of tertiary structure

Of the known P450s crystal structures, bacterial P450cam (12) and P450BM-3 (8), and rabbit P450 2C5 (4) have been the most often used templates for homology modeling. The overall resemblance of these P450s to each other and to the newly reported rabbit P450 2B4 (3), is evident from their structures, presented in Figure 1. Figure 2 shows a structure-based multiple sequence alignment of these P450s using VAST, a tool that compares protein structure neighbors (18, 19). A number of conserved structural features are readily evident. First, there is a topology dichotomy, with the heme wedged between a helix rich domain (right side) and a region containing  $\beta$ -sheets (left side). Although numerous structural similarities among these P450s are readily evident, a sampling of conserved features is highlighted in this figure: the long helix I crosses the heme on the side that encompasses the substrate binding site; helix F runs across the front face in this view; a finger-like beta sheet structure that is post-helix L and before the C-terminus, that includes substrate binding residues; a stretch between helices K and L also binds substrate; and the heme-distal peptide containing heme-binding Cys, directly under the heme.

Despite their variant substrate specificities and differential interaction with electron transfer partners and membrane, these four P450 crystal structures reveal an overall similarity in structural motifs. Nevertheless, structurally variable regions are also readily evident, and some of these are highlighted. The sequence connecting the F and G helices is one such region; note that this is a loop in P450cam and P450BM-3, and mainly a helix in P450 2B4. Another important difference around the substrate binding region is the stretch between helices B and C: this includes a small B' helix in P450cam and P450BM-3, but corresponds to a rather extended structure in P450 2B4. Neither of these two regions is observed in the P450 2C5 crystal (4); the implications of modeling regions lacking a template will be discussed later in the section on loop construction.

### 4. HOMOLOGY MODELING

The major objective in homology modeling is to provide a structural rationale for the observed substrate recognition properties of individual P450s. The basic premise of homology modeling is that structure of the target P450 resembles a known P450 crystal structure. A comprehensive review of the techniques used in modeling the structure of a protein from its homology to a protein of known structure is discussed in ref. 20. It should be noted that among the first successful uses of homology modeling,



**Figure 1.** Crystal Structures of P450s. Two bacterial P450s, P450cam and P450BM-3, and two rabbit P450s, P450 2C5 and P450 2B4 are shown. Their respective PDB accession codes are 2CPP, 2HPD, 1DT6 and 1PO5. Selected major helices and substrate-binding regions are identified by letter as denoted in the color key. Some P450s exhibit a small helix (B' helix) in the region between helices B and C (B-C loop). Likewise the F-G loop may be helical, as seen for P450 2B4.

30 years ago, was the prediction of the structure of alpha lactalbumin from lysozyme. The alpha-lactalbumin structure was predicted in 1974; the x-ray structure was determined almost 20 years later and was completely superimposable on the homology-modeled structure.

Overall, construction of a three dimensional protein structure based on the known structure entails: 1) choice of structural template(s), 2) amino acid sequence alignment, 3) assignment of backbone coordinates, 4) insertion of loops, 5) assignment of local sidechain conformation, and 6) structure refinement through energy minimization and molecular dynamics. Computational and/or experimental criteria are then applied to validate the model. An overview of the homology modeling process is shown in Figure 3.

#### 4.1. Choice of template

The first decision in homology modeling is choice of template. The earliest modeling work was performed using the sole available crystal structure of bacterial P450cam (12). Substrate recognition site regions

in mammalian P450s were suggested based on their alignment with P450cam (21), and these regions are still sometimes referred to when constructing models. With the publication of a P450BM-3 structure (8), this class II P450 soon became the template of choice, since it interacts with NADPH cytochrome P450 reductase and has greater sequence similarity with mammalian P450s. Modeling studies based on these bacterial P450s are detailed in our earlier review (15). With the advent of a rabbit P450 2C5 structure (4), this mammalian P450 was increasingly adopted as the template for human P450 models, which have primarily been P450s within the same family, such as P450 2A4 (22), P450 2B1 (23), P450 2C9 (24, 25) and P450 2D6 (26, 27).

Most simply, the chosen template can be a single P450. Alternatively, the template may be derived from several P450s, as was done for P450 2B1 (28), P450arom (29), and P450 2D6 (30). Conversely, the template sequence may be aligned with several sequences from the same P450 family as the target P450 (22-27, 31-34). Incorporation of

## Molecular Modeling of Cytochrome P450

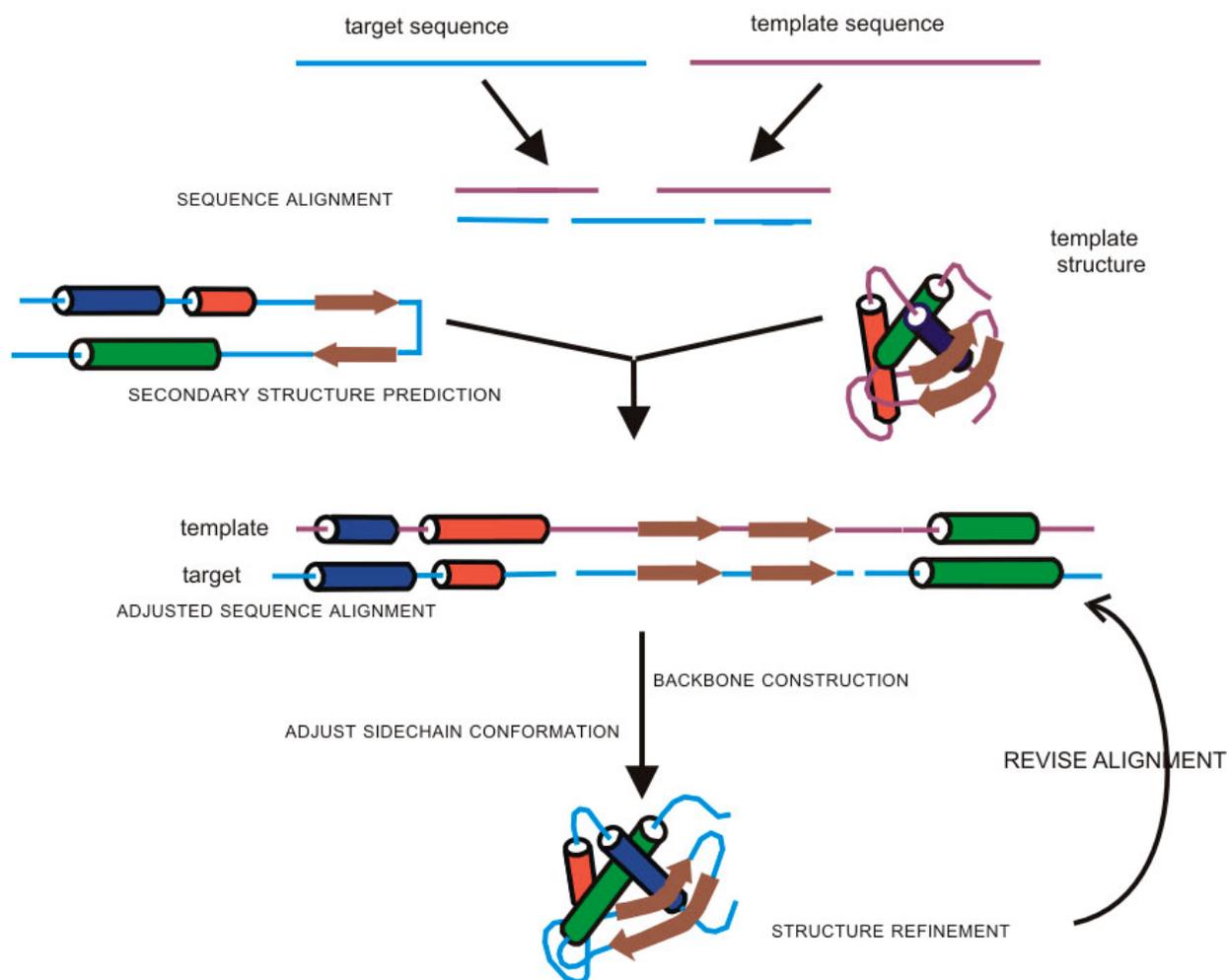
P450cam	P P H V P E H L V F D F D M Y N P S N L S A G <b>V Q E A W A</b> V L Q E S N V P D L V W T R C N G G H W I A T R - <b>G Q L I R E</b>	73
P450BM-3	I K E M P Q P K T F G E L K N L P L L N T D K <b>P V Q A L M K I A D E L</b> G E I F K F E A P G R V T R Y L S - <b>S Q R L I K E</b>	60
P450 2C5	K L P P G P T P F P I I G N I L Q I D A K D - <b>I S K S L T K F S E C Y</b> G P V F T V Y L G M K P T V V L H - <b>G Y E A V K E</b>	67
P450 2B4	G K L P P G P S P L P V L G N L L Q M D R K <b>G L L R S F L R L R E K Y</b> G D V F T V Y L G S R P V V W L C - <b>G T D A I R E</b>	67
	A	B
P450cam	<b>A Y E D</b> - Y R H F S S E C P F I P <b>R E A G E A Y</b> - D F I P T S M D P P E - - <b>Q R Q F R A L A N Q V V G</b> - - <b>M P V V D K L</b>	127
P450BM-3	<b>A C D</b> - - E S R F D K N L S <b>Q A L K F V R D F A</b> - G D G L F T S W T H E <b>K N W K K A H N I L</b> L P S F S Q Q A - - <b>M K G Y</b>	115
P450 2C5	<b>A L V D</b> - L G E E F A G R G S V P I L E K V S K - G L G I A F S N A - - K T <b>W K E M R R F S L M T</b> L R N F G M G K R S I	123
P450 2B4	<b>A L V D Q A E A F S G R G K I A V V D P I F Q Y G V I F A N</b> - - - - <b>G E R W R A L R R F S L A T M R D</b> F G M G K R S V	123
	B'	C
P450cam	<b>E N R I Q E L A C S L I E S L R P Q</b> - - - G Q C N F T E D Y A E P F P I R I F M L L A G L P - - - - - E E D I	174
P450BM-3	<b>H A M M V D I A V Q L V Q K W E R L N A D E H I E</b> - <b>V P E D M T R L T L D T I G L C G F</b> N Y R F N S F Y R D Q P H P F I	174
P450 2C5	<b>E D R I Q E E A R C L V E E L R K T</b> - N A S P C D P - T F I L G C A P C N V I C S V I F H - N R F D Y K D E E F L K M	180
P450 2B4	<b>E E R I Q E E A R C L V E E L R K S K G</b> - A L L D N - <b>T L L F H S I T S N I I C S I V F</b> G K R F D Y - K D <b>P V F L R L L</b>	180
	D	E
P450cam	<b>P H L K Y L T D Q M T R P D G S M T</b> - - - - - <b>F A E A K E A L Y D Y L I P I I E Q R R Q K</b> - - - -	214
P450BM-3	<b>T S M V R A L D E A M N K L Q R A N</b> - - - - P D D P A <b>Y D E N K R Q F Q E D I K V M N D L V D K I I A D R K A S G E</b> - -	228
P450 2C5	<b>E S L H E N V E L L G T P W L Q V Y N N F P A L L D Y F P G I H K T L L K N A D Y I K N F I M E K V K E H Q</b> K L L D V N	240
P450 2B4	<b>D L F F Q S F S L I S S F S S Q V F E L F S G F L K H F P G T H R Q I Y R N L Q E I N T F I G Q S V E K H R A T L D P S</b>	240
	F	G
P450cam	P G T D <b>A I S I V A N</b> - G Q V N - G - - - R P I T <b>S D E A K R M C G L L L V G G L D T V V N F L S F S M E F L A K S P E</b>	269
P450BM-3	Q S D D <b>L L T H M L N G K D P E T G</b> - - - E P L D <b>D E N I R Y Q I I T F L I A G H E T T S G L L S F A L Y F L V K N P H</b>	285
P450 2C5	N P R D <b>F I D C F L I K M E Q E N N</b> - - - L E F T <b>L E S L V I A V S D L F G A G T E T T S T T L R Y S L L L L L K H P E</b>	297
P450 2B4	N P R D <b>F I D V Y L L R M E K D K S D P S S E F H H Q N L I L T V L S L F F A G T E T T S T T L R Y G F L L M L K Y P H</b>	300
	H	I
P450cam	<b>H R Q E L I E R P</b> - - - - - <b>E R I P A A C E E L L R R F S</b> - L V A D G R - I L T S D Y E F - H	308
P450BM-3	<b>V L Q K A E E A A R V L V</b> - D P V P S <b>Y K Q V K Q L K Y V G M V L N E A L R L W P T A P A F S L</b> - Y A K E D T V L G G	343
P450 2C5	<b>V A A R V Q E E I E R V I G R H R S P C M Q D R S R M P Y T D A V I H E I Q R F I</b> D L L P T N L P H A V T R D V R F - R	356
P450 2B4	<b>V T E R V Q K E I E Q V I G S H R P P A L D D R A K M P Y T D A V I H E I Q R L G D L I P F G V P H T V T K D T Q F</b> - R	359
	J	K
P450cam	G V Q L K K G D Q I L L P Q <b>M L S G L</b> D E R E N A C P M H V D F S R Q K V - - - - - S H T T F G H G S H L C L G	359
P450BM-3	E Y P L E K G D E L M V L <b>I P Q L H R D K T I W G D D V E E F R P E R F E N P</b> - S A I P Q H A F K P F G N G Q R A C I G	402
P450 2C5	N Y F I P K G T D I I T S <b>L T S V L H D E K A F P N P K V F D P G H F L D E S G N F K K S D Y F M P F S A G K R M C V G</b>	416
P450 2B4	G Y V I P K N T E V F P V <b>L S S A L H D P R Y F E T P N T F N P G H F L D A N G A L K R N E G F M P F S L G K R I C L G</b>	419
	K'	*
P450cam	<b>Q H L A R R E I I V T L K E W L T R I P D F S I A P</b> - - - G A Q I Q H K S - G I V S G V Q A L P L V W D P A T T K A V	414
P450BM-3	<b>Q Q F A L H E A T L V L G M M L K H F D F E D H T</b> - - - - N Y E L D I K E - T L T L K P E G F V V K A K S K K I P L G	456
P450 2C5	<b>E G L A R M E L F L F L T S I L Q N F K L Q S L V E P</b> - K D L D I T A V V N G F V S V P P S Y Q L C F I P I - - - - -	469
P450 2B4	<b>E G I A R T E L F L F F T T I L Q N F S I A S P V P P E D I D L T P R E S</b> - G V G N V P P S Y Q I R F L A R - - - - -	472
	L	

**Figure 2.** Multiple sequence alignment for P450s. The sequences of the P450s shown in Figure 1 were aligned using VAST (18, 19). Conserved helical sequences are colored blue and denoted by letter. The heme iron-binding cysteine is denoted with an asterisk.

sequence-structure information from related P450s improves the target-template sequence alignment to minimize misalignments that occur when a single sequence is used. When evaluating potential templates, P450s from the same subfamily or family are preferred. Although more distantly related P450s from other families can be used, this will improve modeling of regions strongly conserved for

the entire P450 superfamily, modeling will be less reliable for the more variable regions. In general, sequence identities of at least 30% are required to generate homology models (35). However, the threshold is lowered by considering multiple sequences within P450 families.

Interestingly, a human P450 2D6 comparative



**Figure 3.** Schematic of homology modeling process, whereby the one dimensional sequence of the target is converted to a 3D structure. This is accomplished by mapping the target sequence onto the template structure via a sequence alignment that is guided by aligning the known helices on the template with the predicted helices of the target.

modeling exercise used bacterial P450s as well as the mammalian rabbit P450 2C5 as templates (27). Four models were constructed: one was based on the four bacterial templates; another used P450 2C5 alone; and the remaining two were based on the four bacterial P450s and P450 2C5, utilizing two different alignments. Based on energetic considerations such as solvent accessibility, hydrogen bond and dihedral angles, templates derived from both the mammalian and four bacterial P450s, yielded a higher quality model than those derived from using P450 2C5 alone, in terms of both main chain stereochemistry and amino acid residue environment. However, using these criteria, the model based on the four bacterial P450s was nearly as good.

#### 4.2. Sequence alignment

After P450 template selection, alignment of the target and template sequences is the most critical step in homology modeling. Alignments may be solely sequence-based, or can incorporate other information. For example, a predicted secondary structure is often used to guide the sequence-based alignment. One such algorithm is the PHD

method (36), which predicts a secondary structure derived from similar sequences in known protein structures, in conjunction with a neural network technique. Some other automated secondary structure prediction and sequence alignment tools are JPRED (37), SAM-T99 (38), and SSpro (39, 40). Some of these are readily accessible via a web-based interface. Alignment algorithms depend on user-adjustable parameters that determine the degree of sequence similarity and gap placement.

Whatever algorithm is used, adjustments in the automated alignment, whether automated or manual, are essential to ensure that the proposed alignment will ultimately yield a target P450 whose major structural elements (primarily helices) resemble those of the template P450. For example any gaps within predicted helices must be eliminated to ensure their integrity. This process thus applies additional criteria to align not only the sequence but also the major structural elements of template and target.

The importance of obtaining a good structure-

## Molecular Modeling of Cytochrome P450

based alignment cannot be underestimated. Ambiguities in the alignment can contribute to significant misalignments that have serious consequences; for example active site residues may be incorrectly identified. This problem may be exacerbated upon considering that the substrate recognition properties of a P450 might be reasonably rationalized with such an incorrect model.

### 4.3. Construction of backbone coordinates

When a single P450 template is used, the proposed alignment is applied to project the target sequence onto the template backbone. Such direct one-to-one mapping of target onto template  $\alpha$ -carbons is relatively straightforward, especially for highly similar regions without gaps. When using multiple P450s as templates, partial sequences from the target are projected onto aligned regions of one of the templates. Thus the template in this approach derives from the combined coordinates of several structures, as exemplified by a P450 2D6 model (30). Similarly, models of P450 2C9 (25) and P450 2D6 (27) constructed were based on the comparative modeling method using Modeller (Molecular Simulations Inc.).

### 4.4. Loop insertion

When the target has an insertion whose coordinates cannot be derived from the template, the usual homology procedures cannot be applied. Rather, the principles of protein folding must be considered to construct the missing coordinates. A simple approach is to constrain the ends and use molecular dynamics and energy minimization to search for a reasonable conformation. Another approach is to search a structural database for a peptide fragment whose sequence is similar to that of the insertion, and to apply its structure to the insertion. The inserted peptide is then subjected to energy minimization and molecular dynamics to search for a conformationally acceptable structure in the context of the target environment. This approach was taken with modeling of P450s 2B1 (31, 41) and 1A2 (32). Of course the ambiguity in the coordinate assignments of these residues depends on the size of the insertion. The accessible conformational space of smaller insertions is limited and nearby residues will be minimally perturbed. In contrast the wider conformational space of larger insertions can result in more extensive structural perturbations.

One of the more prominent insertion problems in P450 modeling is the region between the F and G helices, which is often of variable length. This region is thought to interact with the membrane and possibly serve as a conduit for the entrance of substrates from membrane into the P450 interior (15, 31, 42). Such a gatekeeper role, with its control of access into the P450, would make this region a determinant of P450 substrate selectivity. A small F-G loop is present in P450cam, whereas it is a bit larger in P450BM-3. In P450 2B4 this region is transformed into a prominent helix (3). This region is absent in the first P450 2C5 crystal structure (4) used for modeling many mammalian P450s. Although more recent P450 2C5 structures (5, 6) have these coordinates, their variability among P450s portends a greater degree of uncertainty in

homology modeling of this region.

### 4.5. Adjustment of sidechain conformation

Although construction of the  $\alpha$ -carbon backbone is readily accomplished by replacing template sidechains with those of the target, this invariably results in unfavorable interactions between neighboring sidechain. These are initially relaxed via energy minimization, as the structure changes to yield more favorable dihedrals for the backbone and sidechains. Inspection of this structure at this early stage may reveal some obvious problems that would be most easily resolved by readjusting the alignment.

The above process can be applied in a more automated manner. Thus spatial constraints were imposed while generating a model of P450 2D6 (27), using a comparative modeling method. Energetically favorable scoring functions compared the solvent accessibility, hydrogen bonding, and dihedral angles of the template and target P450s, and were used to develop a structure with constraints.

### 4.6. Structure refinement

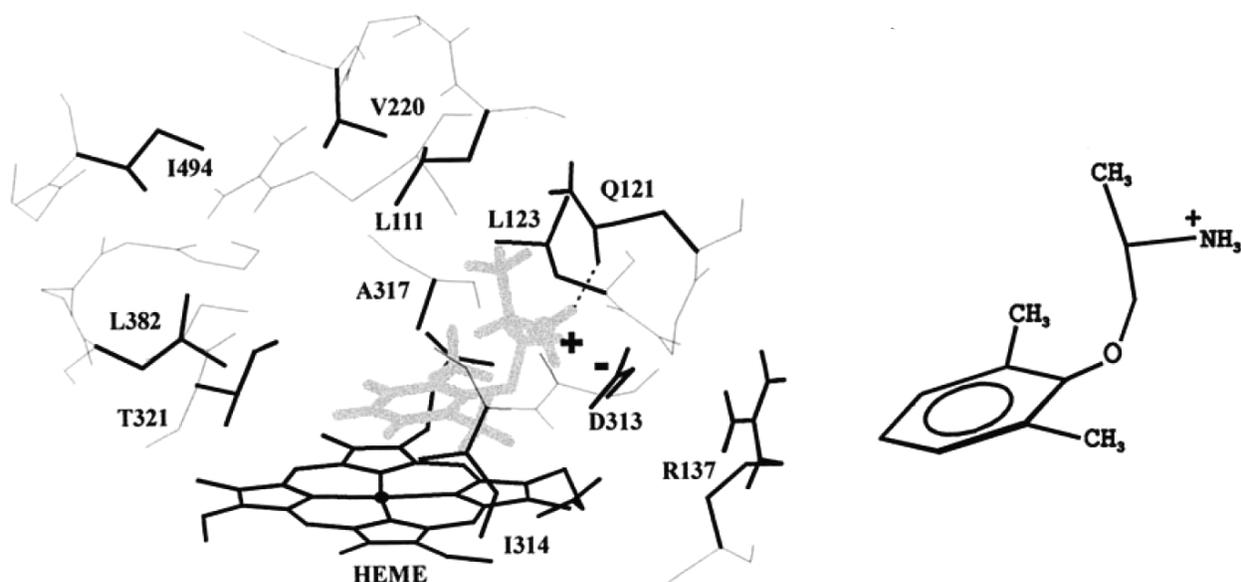
Initially the entire model structure is energy minimized. Since this yields a nearby, rather than global, energy minimum, molecular dynamics is invoked to sample conformational space and seek energetically favorable structures (22, 31, 32, 43). Different software packages employ varying energy expressions and algorithms, and consequently may yield software dependent structural output. During dynamics, the minimum energy structure must be perturbed by heating to a higher temperature and the system is then equilibrated. Cycles of molecular dynamics and energy minimization are often used to refine the initial structure to achieve a lower energy state. The total energy is monitored during these molecular simulations; a constant value indicates a stable structure that is reasonably equilibrated and suited for analysis of P450-ligand interactions.

Structural snapshots are taken during the dynamics simulations in order to sample the conformational space accessible to the P450. Either the final structure or an average of several structures obtained from the final phase of the dynamics run, is used for subsequent work. Thus, while the earlier homology modeling steps yield a structure similar to the template coordinates, molecular dynamics addresses a protein folding problem: the search for an energetically reasonable structure that is more suitable for analyses of P450-substrate interactions.

The power of this approach is illustrated by a study of an R97A mutation in P450 2C9 (44). The native arginine is directed toward the heme propionate groups, and dynamics simulation of the wild type P450 shows a stable structure. In contrast, the mutation compromises the experimental stability of the mutant, which was reflected in dynamics runs in which the structure became unstable.

## 5. MODEL VALIDATION

The "health" of the proposed structure is computationally



**Figure 4.** Predicted active site of the P450 1A2 complex with mexiletene. The antiarrhythmic drug (right) is shown in the active site (left). The protein backbone is shown along with substrate pocket sidechains that engage in hydrogen bonding, hydrophobic or charge interactions with the substrate.

assessed to ensure its reasonableness from the perspective of protein conformation. Typically the backbone and sidechain conformations are examined to ensure that they are consistent with those observed in other proteins, and are energetically permissible. As with the previous stages, a major problem may require readjusting the alignment and repeating the previous steps of the modeling process.

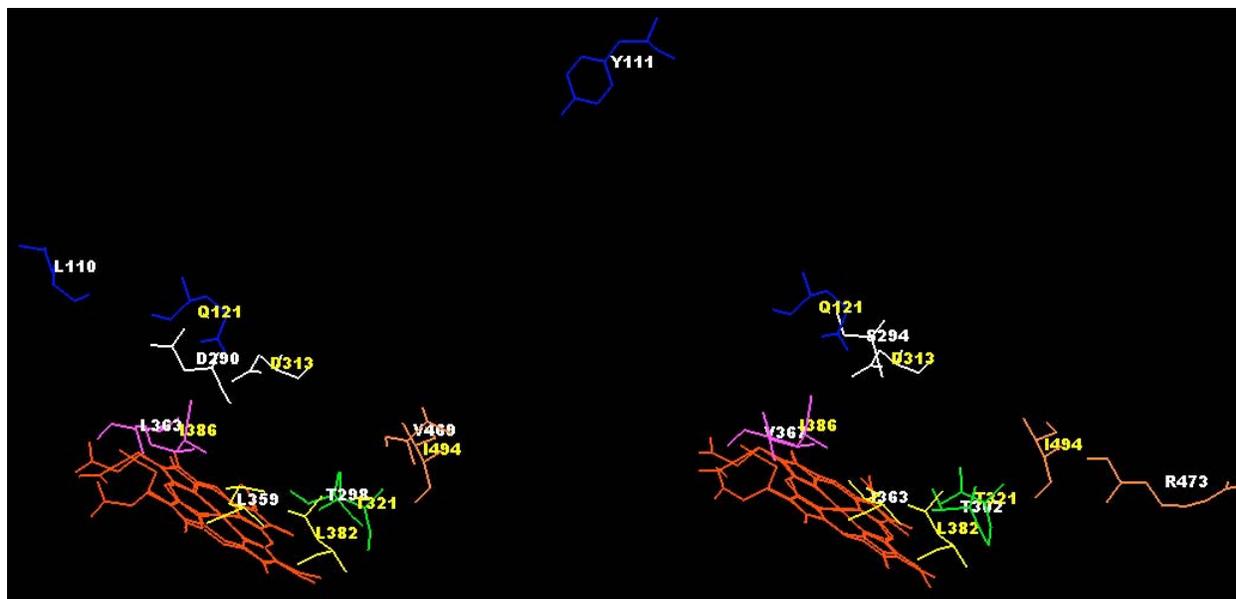
Subsequent model evaluation usually centers on the active site, and involves computational and/or experimental approaches. Computational validation is performed by docking of known substrates and/or inhibitors to the active site, and then rationalizing the substrate specificity and stereoselectivity of the P450. For example, P450 2B1 models successfully explained the site-specificity and stereoselectivity of steroid hydroxylation by this P450 (23, 28). The predictive value of a model can also be assessed by its ability to explain the relative activities of a set of substrates or inhibitors (31, 32, 45). Such validations typically gauge the energetics of the interactions between small molecules and a P450, on the basis of complementary hydrogen bond, electrostatic and hydrophobic interactions.

Substrate docking begins with selection of an initial position and orientation for the ligand. A reasonable and straightforward approach is to base these on an observed product (46). However, the process can be more challenging and ambiguous when multiple products are observed or for P450s (such as P450 3A4) that are presumed to have rather large substrate binding pockets. Initial substrate placement is further complicated by protein flexibility, which allows for a range of conformations in the active site. Indeed, distinct structural differences are seen in the individual molecules of the P450BM-3 dimer (8). Keeping in mind that a given substrate may be susceptible to several stereospecific reactions, and that a single P450

can recognize a wide array of structurally dissimilar substrates, a typical fixed lock-and-key model may not be appropriate for some P450 forms. Furthermore, significant substrate-induced changes in the active site are evidenced by comparison of the substrate-free and -bound P450BM-3 crystal structures (8, 9) and substrate mobility within its active site (47). This point is also illustrated by binding of two different substrates to P450eryF, 6-deoxyerythronolide B and ketoconazole, which resulted in different active site structures (10, 11). Additionally, substrate-dependent conformational differences are observed among three P450 2C5 structures (4-6). These findings thus emphasize the difficulty in P450 docking analyses that use a single model conformation.

The substrate positioning problems may be partially overcome by docking the substrate in several different orientations (31, 32, 45). For example, compounds with different rotations were docked into the P450 active site to generate complexes whose structures were refined by energy minimization and molecular dynamics. Structures with the highest degree of complementary substrate-P450 binding interactions, were then chosen for subsequent work.

An example of inhibitor docking for model validation is shown in Figure 4 for the predicted interaction of the antiarrhythmic mexiletene with the human P450 1A2 active site (45). This inhibitor interacts with active site residues via three hydrogen bonds, a charge interaction, and hydrophobic bonding. Two related antiarrhythmics were likewise assessed and less extensive interactions were observed. The strength of the predicted interactions of these three compounds with the P450 1A2 active site was found to correspond to their experimental inhibitory potencies. In addition, mutagenesis of R108 (48) significantly changed the catalytic activity, which is consistent with its predicted



**Figure 5.** Comparison of active sites of a homology model with crystal structures. Residues in the predicted active site of human P450 1A2 (32) are shown along with those of the rabbit P450 2C5 (left) and 2B4 (right) structures shown in Figure 1. Hemes are shown in red. P450 1A2 was aligned with each P450 and corresponding residues in the two P450s are denoted by the same color. P450 1A2 residue names are labeled yellow, and names of crystal P450 residues are white.

position in the substrate binding pocket.

Site-directed mutagenesis of predicted active site residues is often used for experimental validation. Mutation of critical residues has been shown to either abolish catalytic activity (46, 48, 49), or alter the substrate region- or stereoselectivity (46, 49, 50). Other approaches to confirm that a residue is in the active site are less common. One example is an NMR study of P450 2D6, which used the measured distances between substrate protons and the heme iron to constrain model development (51). Consistent with previous predictions, the active site of the proposed model contained an acidic residue. The goal of another study was to assess a P450 2B1 model not in the active site but in its recognition site for NADPH dependent P450 reductase (52). Peptide mimics of the predicted site were synthesized and shown to compete with P450 for binding to reductase, in accordance with the proposed structural model.

In addition to performing computational docking and experimental validations, it is instructive to compare a proposed model to subsequently published crystal structures. Figure 5 illustrates this for human P450 1A2 (32). The positions of predicted residues in the substrate binding pocket are compared to the positions of the corresponding residues in the subsequently reported crystal structures of rabbit P450s 2C5 (4) and 2B4 (3). The model reasonably predicted the placements of most residues in this region. The greatest deviation between model and an experimental structure occurs for Q121 of P450 1A2, whose position reasonably matched that of L110 of P450 2C5 but not Y111 of P450 2B4. The origin of this discrepancy is the relatively open P450 2B4 structure in which the B' helix, which includes Y111, is displaced

relative to other P450s. This displacement can be readily seen in Figure 1. Since the P450 1A2 model was primarily based on a P450 BM-3 template, it is thus no surprise that the model does not agree with the P450 2B4 structure in this region. This observation makes the important point that since a model resembles its template, the model will structurally deviate from a P450 that is locally dissimilar to the template. This bias toward the template is also underscored in differences between the B-C loop in P450 2C9 model based on a P450 2C5 template, and a subsequently published P450 2C5 structure (44).

## 6. CONCLUSION AND PERSPECTIVE

The objective of molecular modeling is to gain a structure-based understanding of P450 function. A model can be used to interpret existing experimental results, predict the potential for P450 interactions with untested compounds, and ultimately guide future research. From the enzymological viewpoint, modeling can be applied to explain P450 catalytic mechanisms such as aromatase catalyzed hydroxylation and aromatization steps (29). From the biochemistry and biophysics perspective, P450 modeling provides a molecular basis to help elucidate the substrate/inhibitor specificities of individual P450s. Pharmacologically, P450 models can help elucidate the basic mechanism of P450-mediated drug biotransformation, including generation of reactive metabolites, and explain clinically important P450-mediated drug-drug interactions and drug toxicities. Computational evaluation of the potential of drug candidates to interact with P450s will be an invaluable tool for predicting drug-drug interactions at an early stage in the drug development process. Models may serve to guide the design of new drugs with desired metabolic profiles. This is especially true when a drug is

primarily metabolized by a polymorphic P450 that results in a metabolic deficiency, where models may be applied to explain the structural basis of the metabolic defect.

While Figure 1 depicts P450 structures in a particular crystal state, one must always bear in mind two prominent factors that allow for appearance of alternative structures: 1) protein dynamics at physiological temperatures and 2) substrate binding. The former results from the energy landscape of proteins (53), which allows for access to multiple conformations. Such conformational multiplicity has been inferred from P450 ligand binding measurements (54, 55, 56) and has been incorporated in models that explain the non-Michaelis-Menten kinetics of some P450s (57). The substrate binding effect is observed when comparing crystal structures of the same P450, in the absence and presence of different substrate(s). For example, structural comparison of the bacterial P450eryF bound to either 6-deoxyerythronolide B or ketoconazole respectively (10, 11), reveals flexibility in the substrate orientation and/or active site conformation. Substrate-induced changes in the active site are also observed among P450 2C5 crystals with bound progesterone, 4-methyl-N-methyl-N-(2-phenyl-2H-pyrazol-3-yl)benzenesulfonamide, or diclofenac (4-6). The ability to bind and metabolize these three different chemotype substrates reflects a degree of flexibility in the P450 substrate binding pocket. Additionally, in a crystal of human P450 2C9 with bound warfarin, there is a large active site cavity, where the substrate is positioned 11 Å away from the heme-iron; in this conformation the warfarin interaction with active site residues is not amenable to the observed 7-hydroxylation (7), and an alternate conformation must exist wherein the warfarin is more proximate to the iron. Moreover, the cleft shape in the P450 2B4 (3) crystal is the result of dimerization; the open conformation of this P450 again illustrates that different conformations may be induced. Besides dynamics and substrate binding, other factors may perturb P450 structure, such as cycling of the heme-iron redox state (47).

The observation of multiple P450 conformations, shows that molecular dynamics, which searches through a range of conformational space, is a critical step for not only evaluating the robustness of a proposed structure, but also for increasing modeling accuracy since this process mimics P450 dynamic interactions. A static conformational snapshot of a crystal structure thus may or may not be the conformation which represents the rate limiting step for P450-mediated drug metabolism. One must thus bear in mind that a homology model is based on a crystal structure that represents one of many structures, and indeed might not be the structure most relevant to the study at hand. Although imperfect, generation of a homology model is nevertheless a necessary first step for subsequently mining the range of P450 dynamic interaction with substrates and inhibitors.

Progress in generating reasonable and validatable homology models will be accelerated with the advent of mammalian P450 templates with higher amino acid sequences similarity. Computational improvements in

modeling and improved validation strategies for newly proposed models will undoubtedly improve the accuracy of structure prediction and allow for more accurate prediction of the metabolic and inhibition profiles of major human P450s. Site-directed mutagenesis will surely continue to be a powerful tool to identify functionally significant residues involved in substrate binding.

## 7. REFERENCES

1. Ortiz de Montellano, P: Cytochrome P450 Structure, Mechanism, and Biochemistry, Ed. Second Edition, Plenum Press, New York (1995)
2. Nelson D.R: Cytochrome P450 and the individuality of species. *Arch Biochem Biophys* 369, 1-10 (1999)
3. Scott, E.E., Y.A. He, M.R. Wester, M.A. White, C.C. Chin, J.R. Halpert, E.F. Johnson and C.D. Stout: An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution. *Proc Natl Acad Sci USA* 100, 13196-201 (2003)
4. Williams, P.A., J. Cosme, V. Sridhar, E.F. Johnson and D.E. McRee: Mammalian microsomal cytochrome P450 monooxygenase: Structural adaptations for membrane binding and functional diversity, *Mol. Cell* 5, 121-131 (2000)
5. Wester, M.R., E.F. Johnson, C. Marques-Soares, P.M. Dansette, D. Mansuy and C.D. Stout: Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 Å resolution: Evidence for multiple substrate binding modes. *Biochemistry* 42, 6370-6379 (2003)
6. Wester, M.R., E.F. Johnson, C. Marques-Soares, S. Dsijols, P.M. Dansette, D. Mansuy and C.D. Stout: Structure of mammalian cytochrome P450 2C5 complex with diclofenac at 2.1 Å resolution: Evidence for an induced fit model of substrate binding. *Biochemistry* 42, 9335-9345 (2003)
7. Williams, P.A., J. Cosme, A. Ward, H.C. Angove, D.M. Vinkovic and H. Jhoti: Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* 424, 646-648 (2003)
8. Ravichandran K.G., S.S. Boddupalli, C.A. Hasermann, J.A. Peterson and J. Deisenhofer: Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science* 261, 731-736 (1993)
9. 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* 4, 140-146 (1997)
10. Cupp-Vickery J.R. and T.L. Poulos: Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nat Struct Biol* 2, 144-153 (1995)
11. Cupp-Vickery, J.R., C. Garcia, A. Hofacre and K. McGee-Estrada: Ketoconazole-induced conformational

## Molecular Modeling of Cytochrome P450

- changes in the active site of cytochrome P450eryF. *J Mol Biol* 311, 101-110 (2001)
12. Poulos T.L., B.C. Finzel and A.J. Howard: High-resolution crystal structure of cytochrome P450cam. *J Mol Biol* 195, 687-700 (1987)
13. Hasemann C.A., K.G. Ravichandran, J.A. Peterson and J. Deisenhofer: Crystal structure and refinement of cytochrome P450terp at 2.3 Å resolution. *J Mol Biol* 236, 1169-1185 (1994)
14. Peterson J.A. and S.E. Graham: A close family resemblance: the importance of structure in understanding cytochromes P450. *Structure* 6, 1079-1085 (1998)
15. Dai, R., M.R. Pincus and F.K. Friedman: Molecular modeling of mammalian cytochrome P450s. *Cell Mol Life Sci* 57, 487-499 (2000)
16. Kirton, S.B., C.A. Baxter and M.J. Sutcliffe: Comparative modeling of cytochromes P450. *Adv Drug Deliv Rev* 54, 385-406 (2002)
17. Lewis, D.F.: P450 structures and oxidative metabolism of xenobiotics. *Pharmacogenomics* 4, 387-395 (2003)
18. Madej, T., J.F. Gibrat and S.H. Bryant: Threading a database of protein cores. *Proteins* 23, 356-369 (1996)
19. Gibrat, J.F., T. Madej and S.H. Bryant: Surprising similarities in structure comparison. *Curr Opin Struct Biol* 6, 377-385 (1996)
20. Vasquez, M., G. Nemethy and H.A. Scheraga: Conformational Energy Calculations on Polypeptides and Proteins. *Chem Rev* 94, 2183-2239 (1994)
21. Gotoh O: Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* 267, 83-90 (1992)
22. Gorokhov, A., M. Negishi, E.F. Johnson, L.C. Pedersen and L. Perera: Explicit water near the catalytic I helix Thr in the predicted solution structure of CYP2A4. *Biophys J* 84, 57-68 (2003)
23. Kumar, S., E.E. Scott, H. Liu and J.R. Halpert: A rational approach to re-engineer cytochrome P450 2B1 regioselectivity based on the crystal structure of cytochrome P450 2C5, *J Biol Chem* 278 1718-17184 (2003)
24. Afzelius, L., I. Zamora, M. Ridderstrom, T.B. Andersson, A. Karlen and C.M. Masimirembwa: Competitive CYP2C9 inhibitors: Enzyme inhibition studies, protein homology modeling, and three-dimensional quantitative structure-activity relationship analysis. *Mol Pharmacol* 59, 909-919 (2001)
25. de Groot, M.J., A.A. Alex and B.C. Jones: Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *J Med Chem* 45, 1983-1993 (2002)
26. Venhorst, J., A.M. ter Laak, J.N. Commandeur, Y. Funae, T. Hiroi and N.P. Vermeulen: Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities. *J Med Chem* 46, 74-86 (2003)
27. Kirton, S.B., C.A. Kemp, N.P. Tomkinson, S. St-Gallay and M.J. Sutcliffe: Impact of incorporating the 2C5 crystal structure into comparative models of cytochrome P450 2D6. *Proteins* 49, 216-231 (2002)
28. Szklarz, G.D., Y.A. He and J.R. Halpert: Site-directed mutagenesis as a tool for molecular modeling of cytochrome P450 2B1. *Biochemistry* 34, 14312-14322 (1995)
29. Graham-Lorence S., B. Amarnah, R.E. White, J.A. Peterson and E.R. Simpson: A three-dimensional model of aromatase cytochrome P450. *Protein Sci* 4, 1065-1080 (1995)
30. de Groot M.J., N.P. Vermeulen, J.D. Kramer, F.A. van Acker and d.K.G. Donne-Op: A three-dimensional protein model for human cytochrome P450 2D6 based on the crystal structures of P450 101, P450 102, and P450 108. *Chem Res Toxicol* 9, 1079-1091 (1996)
31. Dai R., M.R. Pincus and F.K. Friedman: Molecular modeling of cytochrome P450 2B1: model of membrane insertion and substrate specificity. *J Protein Chem* 17, 121-129 (1998)
32. Dai R., S. Zhai, X. Wei, M.R. Pincus, R.E. Vestal, and F.K. Friedman: Inhibition of human cytochrome P450 1A2 by flavones: a molecular modeling study. *J Protein Chem* 17, 643-650 (1998)
33. Lewis D.F.V., M. Dickens, B.G. Lake, P.J. Eddershaw, M.H. Tarbit and P.S. Goldfarb: Molecular modelling of the human cytochrome P450 isoform CYP2A6 and investigations of CYP2A substrate selectivity. *Toxicology* 133, 1-33 (1999)
34. Lewis D.F., B.G. Lake, M. Dickens, P.J. Eddershaw, M.H. Tarbit and P.S. Goldfarb: Molecular modelling of CYP2B6, the human CYP2B isoform, by homology with the substrate-bound P450BM-3 crystal structure: evaluation of CYP2B6 substrate characteristics, the cytochrome b<sub>5</sub> binding site and comparisons with CYP2B1 and CYP2B4. *Xenobiotica* 29, 361-393 (1999)
35. Przytycka T., R. Aurora and G.D. Rose: A protein taxonomy based on secondary structure. *Nat Struct Biol* 6, 672-682 (1999)
36. Rost, B: PHD: predicting one-dimensional protein structure by profile based neural networks. *Meth in Enzym* 266, 525-539 (1996)
37. Cuff J. A. and G. J. Barton: Evaluation and improvement of multiple sequence methods for protein

## Molecular Modeling of Cytochrome P450

- secondary structure prediction. *Proteins* 34, 508-519 (1999)
38. Karplus, K., C. Barrett and R. Hughey: Hidden Markov models for detecting remote protein homologies. *Bioinformatics* 14, 846-856 (1998)
39. Baldi, P., S. Brunak, P. Frasconi, G. Pollastri and G. Soda: Exploiting the Past and the Future in Protein Secondary Structure Prediction. *Bioinformatics* 15, 937-946 (1999)
40. Pollastri, G., D. Przybylski, B. Rost and P. Baldi: Improving the Prediction of Protein Secondary Structure in Three and Eight Classes Using Recurrent Neural Networks and Profiles. *Proteins* 47, 228-235 (2002)
41. Szklarz G.D., Y.A. He and J.R. Halpert: Site-directed mutagenesis as a tool for molecular modeling of cytochrome P450 2B1. *Biochemistry* 34, 14312-14322 (1995)
42. Hasemann C.A., R.G. Kurumbail, S.S. Boddupalli, J.A. Peterson and J. Deisenhofer: Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* 3, 41-62 (1995)
43. Auchus, R.J. and W.L. Miller: Molecular modeling of human P450c17 (17 $\alpha$ -hydroxylase/17,20-lyase): insights into reaction mechanisms and effects of mutations. *Mol Endocrinol* 13, 1169-1182 (1999)
44. Davies, C., K. Witham, J.R. Scott, A. Pearson, J.J. DeVoss, S.E. Graham and E.M. Gillam: Assessment of arginine 97 and lysine 72 as determinants of substrate specificity in cytochrome P450 2C9 (CYP2C9). *Drug Metab Dispos* 32, 431-6 (2004)
45. Wei, X., R. Dai, S. Shai, K. Thummel, F.K. Friedman and R.E. Vestal: Inhibition human liver cytochrome P-450 1A2 by class IB antiarrhythmics mexiletine, lidocaine, and tocainide. *J Pharm Exp Ther* 289, 853-858 (1999)
46. Melet, A., N. Assrir, P. Jean, M. Pilar Lopez-Garcia, C. Marques-Soares, M. Jaouen, P.M. Dansette, M.A. Sari and D. Mansuy: Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid. *Arch Biochem Biophys* 409, 80-91 (2003)
47. Modi S., M.J. Sutcliffe, W.U. Primrose, L.Y. Lian and G.C. Roberts: The catalytic mechanism of cytochrome P450 BM3 involves a 6 Å movement of the bound substrate on reduction. *Nat Struct Biol* 3, 414-417 (1996)
48. Hadjokas, N.E., R. Dai, F.K. Friedman, M.J. Spence, R.E. Vestal, B.J. Cusack and Y. Ma: Arginine to lysine 108 substitution in recombinant CYP1A2 abolishes methoxyresorufin metabolism in lymphoblastoid cells. *Br J Pharmacol* 136, 347-352 (2002)
49. Scott, E.E., Y.Q. He and J.R. Halpert: Substrate Routes to the Buried Active Site May Vary among Cytochromes P450: Mutagenesis of the F-G Region in P450 2B1. *Chem Res Toxicol* 15, 1407-1413 (2002)
50. Flanagan JU, JD Marechal, R Ward, CA Kemp, LA McLaughlin, MJ Sutcliffe, GC Roberts, MJ Paine, CR Wolf: Phe120 contributes to the regiospecificity of CYP2D6: Mutation leads to the formation of a novel dextromethorphan metabolite. *Biochem J* (in press) (2004)
51. Modi S., M.J. Paine, M.J. Sutcliffe, L.Y. Lian, W.U. Primrose, C.R. Wolf and G.C. Roberts: A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. *Biochemistry* 35, 4540-4550 (1996)
52. Omata, Y., R. Dai, S.V. Smith, R.C. Robinson and F.K. Friedman: Synthetic peptide mimics of a predicted topographical interaction surface: the cytochrome P450 2B1 recognition domain for NADPH-cytochrome P450 reductase. *J Prot Chem* 19, 23-32 (2000)
53. Frauenfelder, H. and D.T. Leeson: The energy landscape in non-biological and biological molecules. *Nat Struct Biol* 5: 757-9 (1998)
54. Koley, A.P., J.T. Buters, R.C. Robinson, A. Markowitz, F.K. Friedman: CO binding kinetics of human cytochrome P450 3A4. Specific interaction of substrates with kinetically distinguishable conformers. *J Biol Chem* 270, 5014-5018 (1995)
55. Koley A.P., R.C. Robinson and F.K. Friedman: Cytochrome P450 conformation and substrate interactions as probed by CO binding kinetics. *Biochimie* 78, 706-713 (1996)
56. Koley, AP, JT Buters, RC Robinson, A Markowitz, FK Friedman: Differential mechanisms of cytochrome P450 inhibition and activation by alpha-naphthoflavone. *J Biol Chem* 272, 3149-3152 (1997)
57. Atkins, W.M., R.W. Wang and A.Y. Lu: Allosteric behavior in cytochrome P450-dependent *in vitro* drug-drug interactions: A prospective based on conformational dynamics. *Chem Res Toxicol* 14, 338-347 (2001)

**Key Words:** Computational modeling, Homology models, Protein structure, Cytochrome P450, Drug metabolism, Review

**Send correspondence to:** Renke Dai, Bristol-Myers Squibb, Pharmaceutical Candidate Optimization, Princeton, NJ 08543, Tel: 609-252-4471, Fax: 609-252-6802, E-mail: renke.dai@bms.com