CMLS, Cell. Mol. Life Sci. 57 (2000) 487–499
1420-682X/00/030487-13 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Review

Molecular modeling of mammalian cytochrome P450s

R. Dai[a], M. R. Pincus[b] and F. K. Friedman[a,*]

[a]Laboratory of Molecular Carcinogenesis, National Cancer Institute, Building 37 Room 3E-24, Bethesda (Maryland 20892, USA), Fax + 1 301 496 8419, e-mail: fkfried@helix.nih.gov
[b]Department of Pathology and Laboratory Medicine, Veterans Administration Medical Center, Brooklyn (New York 11209, USA) and Department of Pathology, SUNY Health Science Center, 450 Clarkson Avenue, Brooklyn (New York 11239, USA)

Received 28 September 1999; received after revision 25 November 1999; accepted 31 December 1999

Abstract. The cytochrome P450s are a superfamily of hemoprotein enzymes responsible for the metabolism of a wide variety of xenobiotic and endogenous compounds. The individual P450s exhibit unique substrate specificity and stereoselectivity profiles which reflect corresponding differences in primary sequence and tertiary structure. In the absence of an experimental structure, models for mammalian P450s have been generated by their homology with bacterial P450s of known structure. The rather low sequence identity between target and template proteins renders P450 modeling a challenging task. However, the substrate recognition properties of several P450s are consistent with recently developed working models. This review summarizes the major concepts and current approaches of molecular modeling of P450s.

Key words. Cytochrome P450; molecular modeling; protein structure; drug metabolism.

Introduction

The cytochrome P450s are a superfamily of hemoprotein enzymes which are responsible for the oxidative metabolism of a wide variety of xenobiotic compounds such as drugs and carcinogens, and endogenous compounds such as steroids, prostaglandins, and fatty acids [1, 2]. The types and amounts of P450s within a tissue are regulated by numerous factors including age, sex, nutrition, and exposure to certain P450 inducers. The individual P450 forms exhibit unique and overlapping substrate-metabolizing profiles that reflect the identity/similarity and orientation of amino acid residues in the substrate-binding site. In addition to a substrate-binding region that is P450 specific, mammalian P450s have other domains which correspond to functionalities that are common across P450s, e.g., recognition of redox partners such as NADPH cytochrome P450 reductase and cytochrome b_5, and a membrane-binding domain which anchors P450 to the endoplasmic reticulum.

Although the substrate specificities of many mammalian P450s have been characterized, the structural basis of substrate recognition is unclear, since an experimentally determined three-dimensional P450 structure has not been reported. In its absence, structural models for a number of P450s have been constructed based on the known structures of bacterial P450s [3–6]. The basic premise of P450 homology modeling is that the mammalian P450s share a basic three-dimensional structure with these bacterial P450s. Certain highly conserved regions are indeed found upon multiple sequence alignments of the P450 superfamily, especially within a family or subfamily, and secondary structural segments observed in bacterial P450s are also predicted to occur in mammalian P450s on the basis of secondary-structure prediction methods [7–14].

* Corresponding author.

The known structures of the bacterial P450s have thus been used as templates for constructing mammalian P450 models. P450cam was the first P450 whose structure was solved, and the early homology models were based on its structure. For example, a model of the human P450 3A4 active site was made by replacing selected P450cam residues with those of P450 3A4 [15]. As additional P450 crystal structures were reported, subsequent models were based on one or more of these templates. While some modeling studies focused solely on the substrate-binding site, most efforts were directed toward generating homology models of the entire P450. Since this entails mapping the one-dimensional sequence of the target onto the three-dimensional structure of the template, the most challenging aspect of P450 modeling arises from the low sequence similarity (identity $< 28\%$) between the mammalian P450 targets and the bacterial P450 templates.

The major objective in homology modeling is to provide a structural rationale for the observed substrate specificities of individual P450s. A map of the active site residues that govern substrate binding and orientation in the major drug-metabolizing human P450s could guide the design of new drugs that are specifically targeted by a given P450. The availability of a series of P450 models would also help in predicting the target P450s of new drugs, and would be an invaluable screening tool for identifying potential drug-drug interactions early in the drug development process. In addition to drug discovery, models can also be used to identify P450 inhibitors and chemoprevention agents for P450-associated cancers and diseases.

Several reviews on P450 modeling have recently appeared which present the computational methodologies and summarize previously published work [16–19]. Rather than present a comprehensive summary of the literature, this review instead presents the major concepts of homology modeling, as specifically applied to the P450s. Selected examples are presented to illustrate the various approaches taken in constructing a P450 model. The inherent limitations of using bacterial P450s as templates for mammalian P450s are discussed. Finally the all-important question of model validation is addressed.

Diversity of P450s

The number of distinct P450 isoforms identified thus far is > 750 [20]. These are classified into families, of which about 40 have been thus far identified, whose members have $> 40\%$ amino acid sequence identity. A family is in turn comprised of subfamilies whose members have $> 55\%$ sequence identity. The various P450s exhibit unique and overlapping substrate-metabolizing profiles that reflect the identity/similarity and orientation of amino acid residues in the substrate-binding site. In addition to differences in sequence, P450s are also classified according to their interaction with a redox partner during catalysis. Thus class I P450s (bacterial and mitochondrial) interact with an FAD-containing reductase and iron-sulfur protein, while class II P450s (microsomal) interact with an FAD- and FMN-containing reductase. P450s also differ in membrane attachment. While the bacterial P450s are soluble, mammalian microsomal P450s are attached to the endoplasmic reticulum but are not intrinsic membrane proteins.

Overview of bacterial P450 structure

The structures of four bacterial P450s have been solved. These are P450cam [3], P450BM-3 [4], P450terp [5], and P450eryF [6]. A ribbon diagram which represents the major structural features of these P450s is presented in figure 1. These P450s exhibit conserved structural features such as a four-bundle helical core consisting of helices D, E, I, and L [14, 19]. A secondary structure dichotomy is evident, with a helix domain-rich (right side) and a region containing β sheets (left side), with the heme wedged between these domains. Most apparent is a rather conserved core region surrounding the heme, which includes helices I and L and the residues containing the heme-binding Cys. Moving outward from the heme are the conserved D, E, J, and K helices, as well as other regions which are conserved to a lesser degree.

Homology modeling

The basic premise of homology modeling is that the target P450 sufficiently resembles a bacterial P450 to justify use of the latter as a template. Four steps are involved in constructing a three-dimensional target protein structure from a known structural template: (i) amino acid sequence alignment, (ii) assignment of backbone coordinates for the target protein, (iii) determination of local sidechain conformation, and (iv) structure refinement through energy minimization and molecular dynamics. Computational and/or experimental criteria are then applied to validate the model. A schematic for this process is shown in figure 2 and will be discussed stepwise.

Sequence alignment

Sequence alignment of the target and template P450 sequences is the most crucial step in construction of a homology model. The first decision is choice of tem-

plate. Early modeling work was performed using the sole available structure, that of P450cam. In this manner, substrate recognition site regions were suggested for ten mammalian P450s in family II based on their alignment with the P450cam sequence [10]. These regions are still used in guiding model development.

As more P450 crystal structures were solved, additional templates became available. Before extrapolating from these structures to construct a mammalian P450 model, it is instructive to consider how well the bacterial structures may be inferred from each other on the basis of sequence alignment. The sequence similarity among the crystallized P450s is low and the identity between any two is less than 20% [21]. As a result, automated sequence alignment for these P450s results in misalignment of some secondary structure elements [14, 19]. This is a sobering initial observation which underscores the limitations of depending solely on automated alignment algorithms, and the necessity of introducing additional information into the alignment process.

P450BM-3 is most often the P450 of choice when using a single P450 as template since (i) it exhibits $>20\%$ sequence similarity with mammalian microsomal P450s and (ii) as a class II P450 it is the only one that interacts with NADPH cytochrome P450 reductase during the catalytic cycle. Furthermore, the crystal structure of this P450 with its palmitic acid substrate [22] may be the preferred template for modeling substrate-bound mammalian P450s [23], provided that substrate-induced conformational changes are borne in mind. Thus P450BM-3 was used as a sole template to model P450s 1A [24, 25], 2A6 [26], 2B [27, 28], 2B6 [29], 2E1 [30, 31],

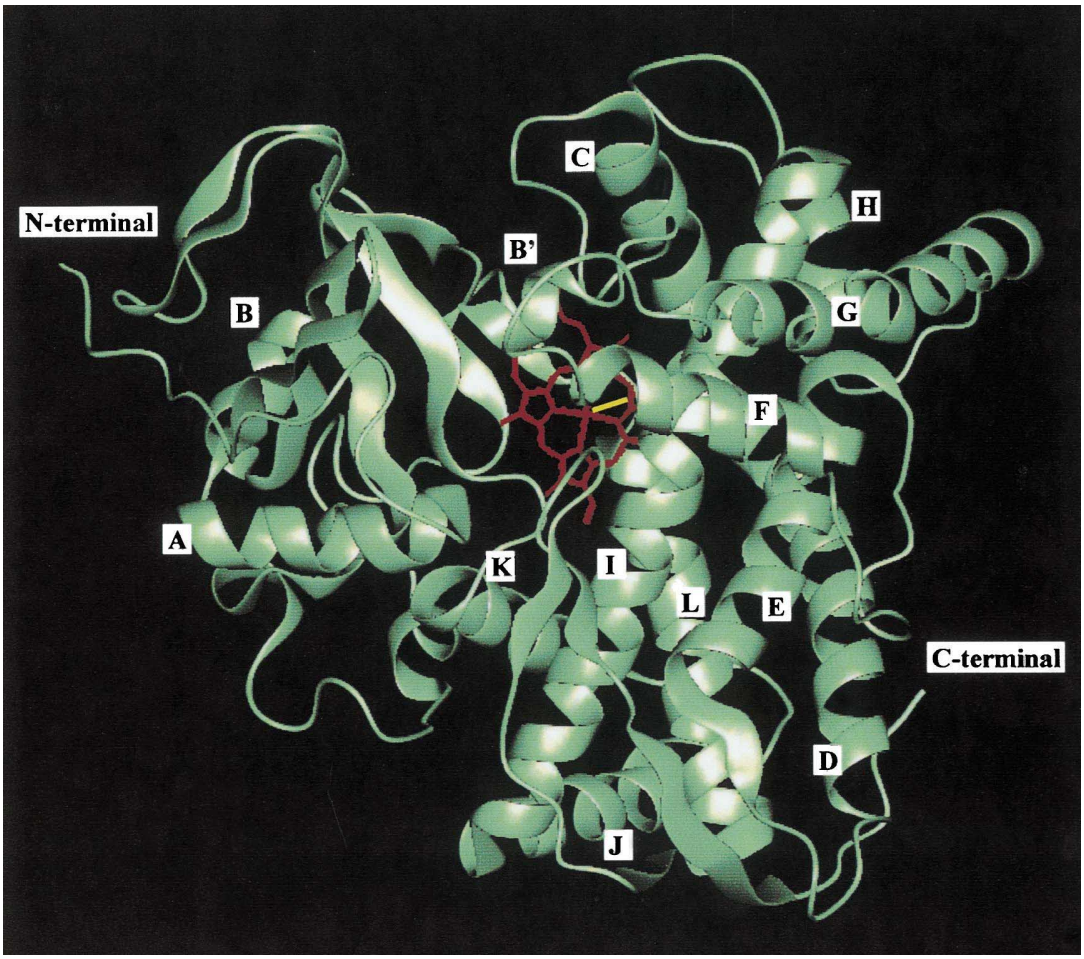

Figure 1. Structural representation of bacterial P450s. A composite structure was composed from the known structures of P450s cam, BM-3, terp, and eryF, using QUANTA software (Molecular Simulations, San Diego, Calif.). Major helices are identified by letter, the heme is red, and the heme iron-Cys bond is shown in yellow. The conserved core is primarily composed of helices D, E, I, and L and the region that contains the heme-binding Cys (behind heme, back side of P450). The substrate-binding region is on the heme face toward the front of the P450 and includes helices B′, F, and G.

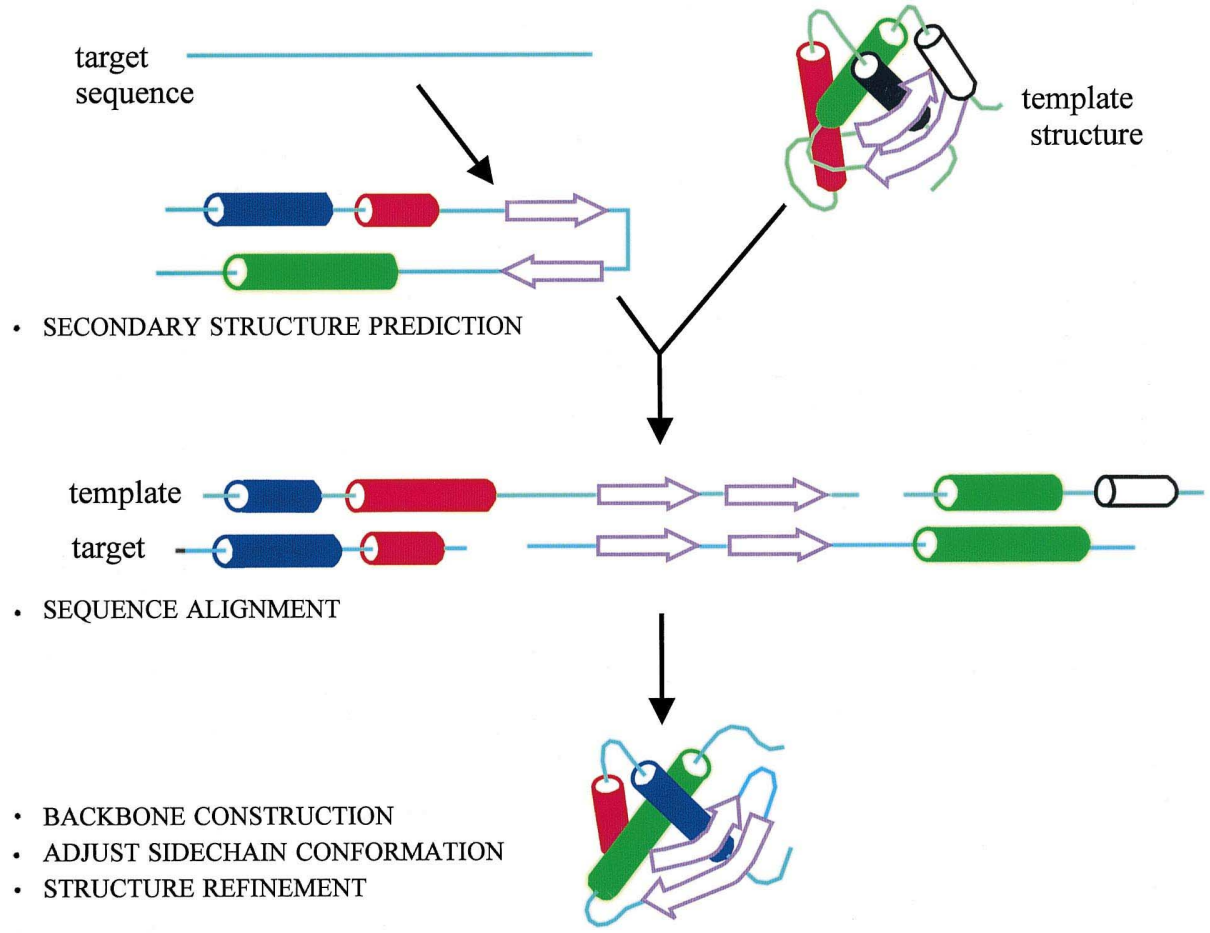

Figure 2. Schematic of homology modeling process, whereby one-dimensional sequence information from the target is converted to a three-dimensional structure. This is accomplished by mapping the target sequence onto the template structure via a sequence alignment that is guided by aligning the known helices on the template with the predicted helices of the target. Corresponding target and template helices are coded in the same color.

3A4 [32], 4A [33], P450c17 [34], prostacyclin synthase [35], and thromboxane A_2 synthase [36].

Sequence alignment algorithms are dependent on adjustable parameters whose input values determine the sequence similarity, placement of gaps, and the ultimate alignment. The automated alignment is often manually adjusted based on factors such as the predicted secondary structure profile of the target P450. This ensures that predicted helices in the target align with known helices in the template, and eliminates any gaps within these helices. From this overview of the alignment of single target and single template sequences, one can gather that the process has a degree of subjectivity, which ultimately translates into some ambiguity in the resulting model.

This ambiguity can be somewhat reduced by incorporating information from multiple sequence alignments. Thus models based on alignments with multiple bacterial P450 templates were constructed for aromatase [37], P450 2D6 [38], and P450 2B1 [39]. Conversely, multiple sequences from P450s related to the target P450 may be aligned with the template sequence(s). Thus sequences from the P450 superfamily were used to model several P450s [40] and aromatase [37], while sequences from the same P450 family were used to model P450s in family II [23] as well as P450s 2A6 [26], 2D6 [38], 2E1 [31], and 2B1 [28]. Alignment of sequences of the same P450 from different species was used to construct models of P450 1A2 [25] and P450c17 [34]. Such incorporation of sequence information from related P450s within a family or subfamily optimizes the sequence alignment between the template and target sequences and eliminates unreasonable gaps that occur when only the single template is used. Utilizing sequences within a superfam-

ily, whose members are more distantly related than those within a family, results in a more reliable alignment for regions that are conserved across all P450s, but a poorer alignment for the more variable regions that may be more family specific.

The reliability of the automated alignment is enhanced by manual adjustments which are based on additional information that relates sequence to structure. The importance of this step is evident from the observation that automated alignment results in misalignment of the known helices of the crystallized P450s [14]. It is therefore not surprising that alignment of target and template P450s should require some manual adjustment. This is usually accomplished using secondary-structure predictions of the target P450, as implemented with one of several algorithms. One such method is the PHD method [41, 42] which has proven more successful than approaches based solely on statistical methods. This approach generates a secondary structure by comparing the sequence of the target protein of unknown structure to a database of sequences with known structure, in conjunction with a neural network technique. Secondary-structure predictions of the target and related P450s are often used to identify consensus positions for the major structural elements, and the alignment adjusted to optimize the coincidence of these regions with the corresponding structures in the template(s). Experimental data are also sometimes incorporated into the alignment decision-making process. For example, when a site-directed mutant exhibits altered or reduced activity, the mutated residue is often presumed to be in the substrate-binding site [17, 39].

When it is possible to generate more than one reasonable alignment and it is difficult to choose among them, alternative structures may be constructed and compared. For example, two P450 2B1 models were compared in which an inserted loop was positioned between the D-E helices, or the F-G helices [43]. Site-directed mutagenesis studies of this P450 identified several critical residues which were presumed to be present in the active site. Since only the latter model included these critical residues in the active site, it was chosen as more appropriate. A representative alignment for human P450 2B1 with P450BM-3 is shown in figure 3 [28]. This was derived via an alignment using several family II P450s. On the basis of secondary-structure prediction by the PHD method, minor adjustments in the automated alignment, such as an insertion between the F and G helices, were required to ensure integrity of helices.

Construction of backbone coordinates for the target protein

With the proposed alignment in hand, the target sequence is projected onto the backbone of the template P450. The template can be a single P450 or a weighted average of the backbone coordinates of the crystallized P450s, as was done for P450s 2B1 [39] and 3A4 [44]. A more selective approach is to identify individually the higher-homology regions between the target sequence and each crystal template, and utilize a combination of coordinates from three or four crystal structures to reconstruct a new template. This approach was taken to model P450 2D6 [38] and 2B4 [45]. Construction of the target backbone is straightforward in regions where the alignment yields a direct one-to-one mapping of target onto template residue. The problems occur in modeling those regions for which there are no template atoms, such as gaps and the extended amino terminus. The latter region is thought to be a membrane anchor in mammalian P450s, but is not present in the soluble bacterial P450s. However, the structural uncertainty of this region is not critical since it is not involved in substrate recognition, which is the primary focus in P450 modeling studies.

Another major difficulty is generation of coordinates for insertions. Strictly speaking, since there is no template for these residues, structure prediction for the inserted residues is a problem in protein folding rather than homology modeling. The simplest approach is to merely apply an extended structure as an initial guess. Another approach is to search a structural database for a fragment whose sequence is similar to that of the insertion, and to base the initial insertion structure on the fragment structure. The inserted region is then subjected to energy minimization and molecular dynamics to search for a conformationally acceptable structure, as was performed with modeling P450s 2B1 [43] and 1A2 [25]. 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.

Local sidechain conformation

During construction of the target backbone, the replacement of sidechain residues often results in unfavorable interactions such as steric overlaps between atoms. Relaxing these bad sidechain contacts minimally requires repositioning sidechain atoms while fixing the backbone, to seek a local energy minimum. Subsequent release of the backbone constraints allows further changes in sidechain conformation as well as adjustments in the backbone. When these procedures cannot relax a local conformation from a high energy state, the original alignment may require adjustment. More refined techniques have also been applied to the prob-

lem of optimizing sidechain conformation along with the associated backbone structure. For example, in modeling P450s 2B4 [45] and 4A11 [46], the repulsive interaction energy of each residue with the rest of the protein was evaluated to identify local regions with questionable structures. Subsequent adjustment of the alignment and/or local structures resulted in a lower repulsion energy and improved models.

These conformational changes are performed by any of several software packages that employ molecular me-

```
P4502B1  meptillllallvgfllllvrghpksRGNFPPGPRPLPLLGNLLQLDRGG   50
P450BM3  ---------------------------TIKEMPQPKTFGELKNLPLLNTDK   24

P4502B1  LLNSFMQLREKYGDVFTVHLGPRPVVMLCGTDTIKEALvgqaedfsgrgt  100
P450BM3  PVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACdesrfdknlsqa   74
            -A-                              -B-

P4502B1  iaviepifkeygviFANGERWKALRRFSLATMRDFGMGKRSVEERIQEEA  150
P450BM3  lkfvrdfagdglftSWTHEKNWKKAHNILLPSFSQQAMKGYHAMMVDIAV  124
          -B'-                    -C-                   -D-

P4502B1  QCLVEELRksqgapldPTFLFQCITANIICSIVFGERFDY-----TDRQF  195
P450BM3  QLVQKWERlnadehieVPEDMTRLTLDTIGLCGFNYRFNSfyrdqPHPFI  174
                                  -E-

P4502B1  LRLLELFYRTFSLLSSfssqvfeffsgflkyfpgAHRQISKNLQEILDYI  245
P450BM3  TSMVRALDEAMNKLQRanpddp-----------AYDENKRQFQEDIKVM  212
           -F-                                    -G-

P4502B1  GHIVEKHRATLDPSAPRDFIDTYLLRMEKEKSnhhtefHHENLMISLLSL  295
P450BM3  NDLVDKIIADRKASGEQSDDLLTHMLNGKDpetgepl-DDENIRYQIITF  261
                               -H-

P4502B1  FFAGTETSSTTLRYGFLLMLKYPHVAEKVQKEIDQVIgshrLPTLDDRSK  345
P450BM3  LIAGHETTSGLLSFALYFLVKNPHVLQKAAEEAARVLvdp-VPSYKQVKQ  310
            -I-                        -J-            -J'-

P4502B1  MPYTDAVIHEIQRFSDLVPIGVPHRVTKDTMFRGYLLPKNTEVYPILSSA  395
P450BM3  LKYVGMVLNEALRLWPTAPAFSLYAKEDTVLGGEYPLEKGDELMVLIPQL  360
             -K-                                       -K'-

P4502B1  Lhdpqyf-DHPDSFNPEHFLDangalKKSEAFMPFSTGKRICLGEGIARN  444
P450BM3  Hrdktiwg-DDVEEFRPERFENpsa--IPQHAFKPFGNGQRACIGQQFALH  408
                                                 *     -L-

P4502B1  ELFLFFTTILQNFSVSSHlapkdidltpkesgigkipptyqicfsar      491
P450BM3  EATLVLGMMLKHFDFEDHtnyeldiketltlkpegfvvkakskkiplgg   457
```

Figure 3. Optimal sequence alignment between human P450 2B1 and P450BM-3. These alignments were derived using sequences of family II P450s. Regions that align with the highest significance are shown in uppercase letters. Helices are denoted by letters beneath the sequence. Helical residues presented in P450BM-3 or predicted in P450 2B1 are shown in bold. The heme iron-binding cysteine is denoted with an asterisk.

chanics to change atom positions and thus minimize energy. Since these programs differ in their energy expressions, force field parameters, and algorithms, the structural output may be somewhat software dependent. However, a comparison of P450 1A2 modeling using the two programs QUANTA (Molecular Simulations, San Diego, Calif.) and LOOK (Molecular Applications Group, Palo Alto, Calif.) yielded similar initial active site structures [25]. This comparison suggests that the differences among these packages is not critical, at least in generating an initial homology model.

Structure refinement

After generating initial backbone and sidechain conformations, the entire structure is energy minimized. Since global energy minimization only finds a nearby energy minimum for a given initial structure, molecular dynamic techniques are used to seek more energetically favorable structures by sampling a range of conformational space [25, 28, 37, 45, 46]. To accomplish this, the minimum energy structure, which pertains to a temperature of absolute zero, is first perturbed by introducing kinetic energy to heat the system to a higher temperature (such as 600 K) for a short time (up to 5 ps). The system is then equilibrated over a longer time period (40–200 ps) while maintaining a constant system energy near the desired temperature. Structural snapshots are taken at regular intervals during this equilibration, and either the final structure or an average of several structures observed near the end of the dynamics run is energy minimized. This structure may be subjected to another cycle of molecular dynamics to probe additional conformational space [25, 28]. The total energy is monitored during these molecular simulations, as a constant value indicates a stable structure in which the system is reasonably equilibrated and suited for analysis of P450-ligand interactions.

The molecular dynamics refines the initial structure to achieve a lower energy state. Thus, while the initial homology modeling steps yield a target structure based on the template coordinates, molecular dynamics is an extension of homology modeling that essentially addresses a protein-folding problem: how to use the initial structure of the target as a starting point to search for an energetically reasonable structure which is suitable for analyses of P450-substrate interactions and/or other structure-function questions. As previously mentioned, the low amino acid sequence similarity between the target mammalian P450s and the template bacterial P450s ($< 28\%$) decreases confidence in the initial model. Sequence identities of at least 30% are generally required to generate homology models although the threshold is lowered by considering multiple sequences within families [47]. We must also keep in mind that some mammalian P450s, such as P450 3A4, recognize a broad array of structurally dissimilar substrates and their active sites are thus probably not well represented by the classical lock-and-key model. Instead, conformational heterogeneity of P450 has been invoked as an explanation for broad substrate recognition, which derives from either the coexistence of multiple discrete conformational states [48] or, perhaps, a spectrum of conformations as is found in myoglobin [49]. In either event, 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.

Another point to consider is protein hydration, which has not been performed in most P450 homology modeling studies. In addition to bound waters at the protein surface, internal water molecules often stabilize local regions by hydrogen bonding to the protein backbone or sidechains. A basic hydration approach entails soaking the P450 in a larger water sphere, as was done for P450 2B1 [39]. However, diffusion of these external waters to the protein interior via molecular dynamics, would require a long and often inaccessible equilibration time. This problem was approached with an algorithm that introduces internal waters by optimizing local hydrogen-bonding and steric interactions [45, 46]. Addition of structural waters to the active site is especially appropriate when the substrate-free P450 is modeled, as this region would otherwise be perturbed by interactions among the surrounding residues. However, one must exercise caution when inserting internal waters, as their incorrect placement would restrict readjustments of their local environments during energy minimization and molecular dynamics.

A representative homology model for P450 2B1 [28] is shown in figure 4 along with the P450BM-3 template structure. Although, as expected, the target P450 resembles the template, differences are observed in some regions. For example, P450 2B1 exhibits a large insertion (18 residues) between the F and G helices relative to the P450BM-3 template, and a shortening of the G helix. Such differences are not surprising in view of the sequence differences between the two P450s. In addition, the bacterial P450 structure was obtained in the crystalline state, and crystal lattice forces at the protein surface might alter its structure. In contrast, the model structure was obtained in the absence of these forces, after the structural adjustments which followed molecular dynamics.

The large F-G insertion is often observed for mammalian P450s and this region has been proposed to interact with the membrane and possibly serves as a conduit for the entrance of hydrophobic substrates from the membrane to the adjacent substrate-binding site [19]. Figure 5 shows the P450 2B1 model from the

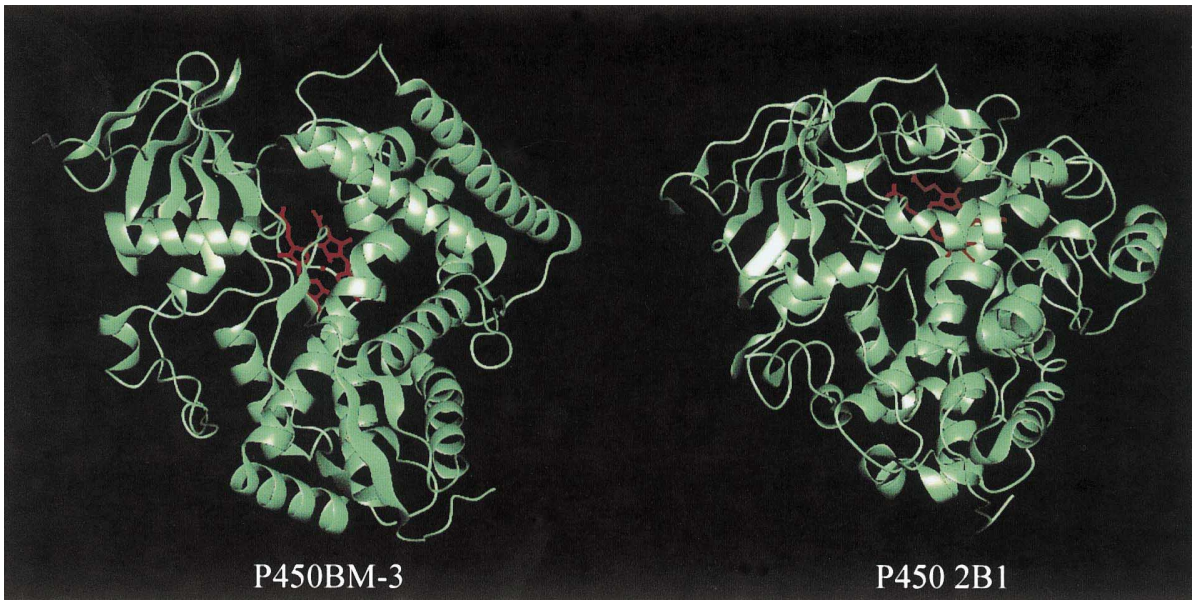

Figure 4. Structures of P450 2B1 target and P450BM-3 template.

perspective of membrane orientation. In this structure, a number of basic residues appear on the P450 surface adjacent to the membrane, and would be capable of electrostatic contacts with membrane phospho groups [28]. This view shows that three proximate hydrophobic regions are projected from the protein surface: the F-G loop, the N-terminal segment and a region before helix A. These regions may thus be inserted into the membrane to constitute a membrane-binding domain. Such a role for the pre-helix A region is supported by experiments with peptides corresponding to this segment in thromboxane A_2 synthase, which indicated that this region interacts with the membrane [50]. In addition, X-ray diffraction of liposome multilayers associated with P450 2B4 indicate that a large part of the molecule is embedded in the membrane [51], and studies of the interaction of this P450 with lipid monolayers also show that regions other than the N terminus bind the membrane [52]. Thus, although P450s are usually represented with the bulk as cytoplasmic and only the N terminus inserted into the membrane, these recent experimental results indicate that additional regions interact with the membrane, and suggest that the representation in figure 5 is more accurate than the common pictorials.

Model validation

The completed model is subjected to various computational tests to evaluate its 'health,' i.e., whether its structure is reasonable from the perspective of protein conformation. Such an assessment entails determining whether the backbone and sidechain conformations are energetically permissible and consistent with those observed in other proteins. Finally, the model is validated (i) computationally, by docking known substrates or inhibitors into the active site to evaluate its consistency with the known substrate/inhibitor specificity and stereoselectivity of the P450, and/or (ii) experimentally, by evaluating the effect of predicted active-site mutations on substrate binding or by other methods that assess whether particular residues are localized in their predicted region.

Model validation requires development of a rationale for the known substrate/inhibitor specificity of a P450. Indeed, most studies show docking of at least one substrate to complementary active-site residues. For example, the proposed mode of steroid docking into a P450 2B1 model explained the site specificity and stereoselectivity of hydroxylation [39]. Specificity prediction can also be assessed by docking several substrates or inhibitors, and evaluating whether the model explains their relative activities or inhibitory potencies [25, 28]. This computational validation approach offers the advantage that, unlike site-directed mutagenesis, the protein structure is not modified. Instead, the interactions of small molecules with active-site residues are evaluated, based on the number of hydrogen bonds, electrostatic interactions, and hydrophobic interactions. For a valid model, the degree of interaction for a set of substrates or inhibitors should agree with the relative

preferences of a P450 for these compounds. However, ligand docking is a challenging process that begins with selecting the best initial conformation and orientation for the docked ligand. A reactive substrate atom may simply be positioned near the catalytic heme-iron center, or in a position similar to that of the corresponding atom in a known crystal structure of a substrate-P450 complex. For example, since P450 2B1 primarily hydroxylates testosterone at the reactive C-16 position, the initial position of this atom relative to heme was similar to that of the C-5 of camphor in the crystal structure of P450cam [28].

A problem with the above procedures is that it assumes a rigid protein framework, whereas protein dynamics offers the potential for the active site to adopt a range of conformations, some of which may be more energetically favorable in the presence of a substrate. Indeed, the crystal structure of P450BM-3 shows a dimer with distinct structural differences between the two component molecules [4]. The substrate metabolism profiles of some P450s include several stereospecific reactions for the same substrate, as well as activity towards a wide array of structurally dissimilar substrates (e.g., 3A4), suggesting that a typical lock-and-key model may not always be appropriate. Significant changes in the active site upon substrate binding are evidenced by comparison of the substrate-free and substrate-bound P450BM-3 crystal structures [22], and substrate mobility in the active site [53, 54]. These observations obviously reflect some degree of flexibility in the substrate orientation and/or active-site conformation.

To better account for ligand and active-site flexibility, several ligand orientations can be evaluated [25, 28, 37]. For example, three orientations of substrate or inhibitor that were rotated 120° were docked into the active site to construct three substrate/inhibitor-P450 complexes

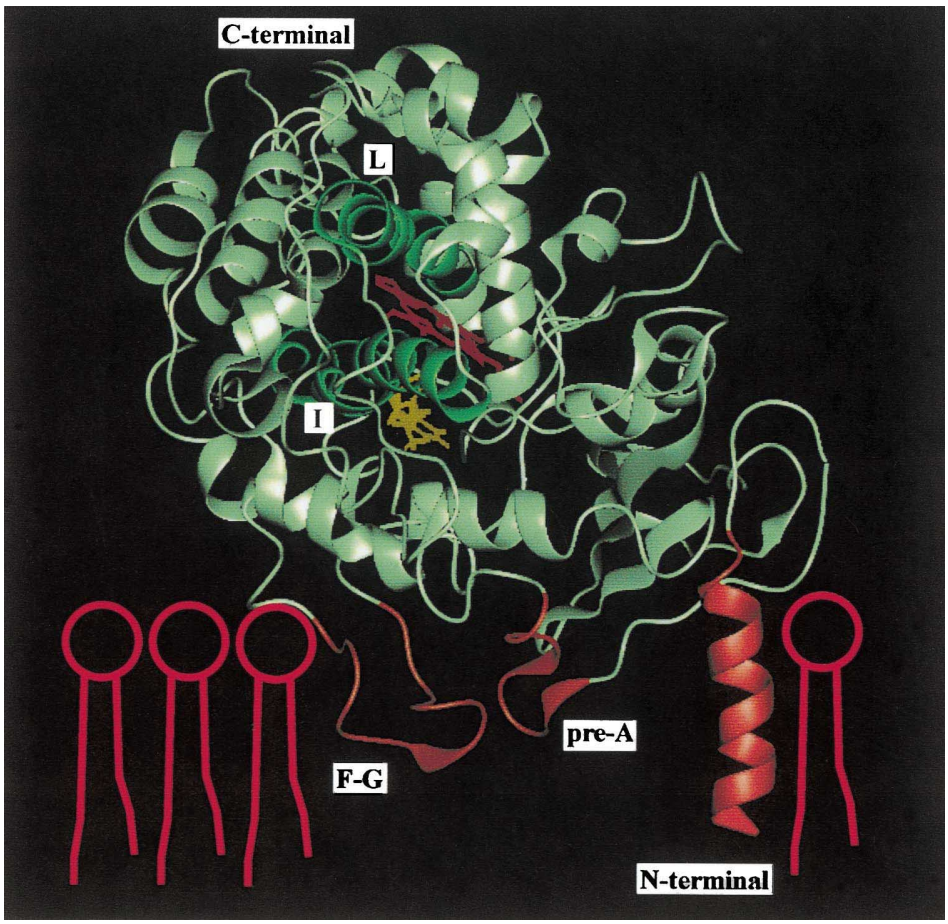

Figure 5. Putative interaction between P450 2B1 and membrane. An α-carbon trace of the P450 model is shown along with several membrane phospholipid molecules (red). Part of the F-G loop, pre-A helix region and N-terminal region comprise a putative membrane-binding domain; these are shown in brown. The substrate benzphetamine (yellow) is shown in the binding pocket. The heme is shown in red, and the I and L helices (bright green) are denoted to better orient the viewer to this structure.

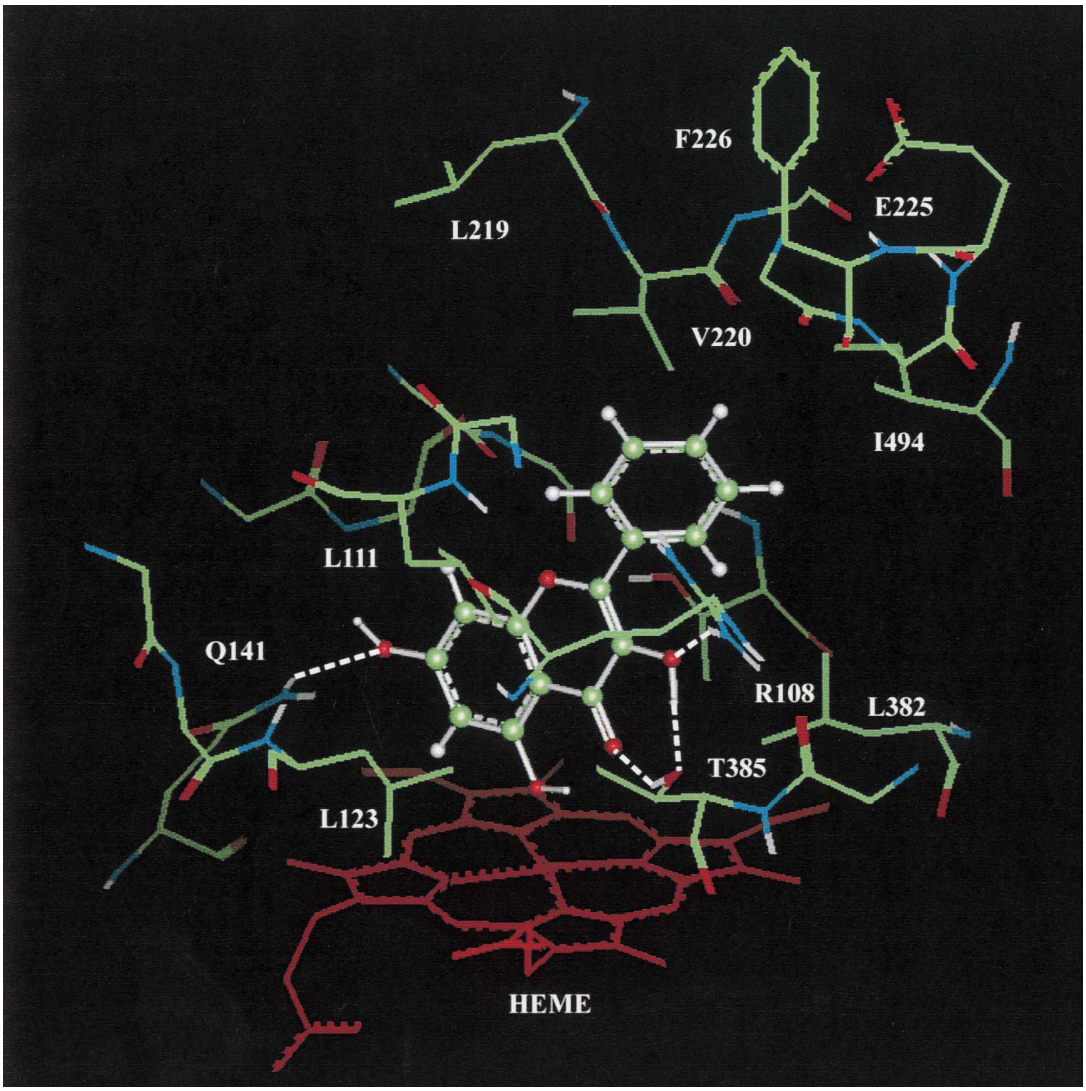

Figure 6. Predicted active site of the P450 1A2 complex with galangin. The flavonoid inhibitor galangin is shown in the active site. The protein backbone is shown along with hydrophobic or hydrogen-bonding sidechains within the substrate-binding pocket.

which were subsequently refined by energy minimization and molecular dynamics. [25, 28]. The structure of the complex with the strongest interaction between the ligand and the active site was used for subsequent substrate/inhibitor specificity analyses. Consideration of more substrate conformations will accordingly increase the chance of approaching the native interaction. However, since computation of additional docking configurations has the penalty of increased computational time, a practical balance must be sought between number of configurations and available computer resources. An example of inhibitor docking for model validation is shown in figure 6 for the predicted interaction of the flavonoid galangin with the human P450 1A2 active site [25]. This inhibitor interacts with active site residues via three hydrogen bonds and hydrophobic bonding. Two related flavonoids 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, a recent random mutagenesis study of this P450 [55] showed that E225 and F226 mutants significantly changed the catalytic activity, which is consistent with their positioning in the substrate-binding pocket (fig. 6). The model is further consistent with the results of a rat P450 1A2 photolabeling study [56] in which the substrate analog 4-aminobiphenyl was found to label a

peptide that included residues corresponding to L219 and V220, which appear in the predicted active site.

Site-directed mutagenesis is often used as a tool to validate and refine models that explain metabolism differences between wild-type and mutant P450s. This is exemplified by the V367A and V367L mutants of P450 2B1 [39], in which the former conferred androgen 6β-hydroxylase activity and the latter not only abolished this activity but also decreased androstenedione 16β-hydroxylase activity.

Direct methods to determine the positioning of particular residues are also possible, but such studies are not common. In a nuclear magnetic resonance (NMR) study of P450 2D6, the distances between substrate protons and the heme iron were measured, and used as a constraint in model development [57]. Consistent with previous predictions, the active site of the proposed model contained an acidic residue. Another study utilized a peptide corresponding to the P450 surface region that was predicted to interact with cytochrome b_5 [58]. When added to a mixture of these proteins, this peptide inhibited the protein-protein interaction, in accordance with the proposed structural model.

Concluding remarks and perspectives

The objectives of molecular modeling of mammalian P450s are to gain a structure-based understanding of P450 action which can be used to interpret existing experimental results and guide future research. From the biochemistry/biophysics perspective, P450 modeling provides a molecular structural basis to help elucidate the substrate/inhibitor specificities of individual P450s. From the enzymological viewpoint, modeling can be applied to explain P450 catalytic mechanisms such as aromatase catalyzed hydroxylation and aromatization steps [37]. P450 structural models can be applied to elucidate the basic mechanism of P450-mediated drug biotransformation, and explain clinically important drug toxicities and drug-drug interactions. Models could serve to guide the design of new drugs with desired metabolic and/or inhibition profiles. This is an important issue in the drug discovery arena where prior knowledge of biotransformation pathways of drugs under discovery/development would help predict the risk of drug-drug interactions. This is especially true if a drug is primarily metabolized by a polymorphic P450 that results in a metabolic deficiency, where models may also be applied to explain the structural basis of a polymorphism associated with a metabolic defect.

The pharmacological utility of P450 modeling is the discovery of specific inhibitors of individual P450s which may be used for chemoprevention of P450-associated cancers and diseases. Computational screening of structural databases for small molecules that bind P450cam has been used to identify new compounds that bind this P450 [59]. A validated mammalian P450 model should likewise successfully identify new substrate/inhibitor compounds.

Molecular modeling has been widely used to elucidate structure-function relationships, especially in ligand-receptor interaction for de novo inhibitor design. Although this review covers modeling of a receptor structure, a complementary quantitative structure-activity relationship approach can be used that is based on the structures of known substrates and/or inhibitors. For example, a pharmacophore-based approach was used to identify the functional chemical groups on small molecules which bind P450s 2D6 [60] and 3A4 [61]. Since these molecules bind the P450 active site, knowledge of the pharmacophore allows one to infer complementary structural information about the P450 active site. The major difference between protein- and ligand-based pharmacophore modeling is that the latter can generate new lead compounds without any knowledge of the P450 structure, and is useful if the only objective is discovery of these compounds. However, the coupling of a pharmacophore model with a protein homology model of P450 2D6 [62, 63] offers a most powerful approach that incorporates data from both P450 structure and metabolism, providing a combined model that satisfies the constraints of both active site structure and the substrate site reactivity, and holds great potential for predicting P450-mediated drug metabolism.

Computational improvements in modeling, the availability of new P450 crystal structures and improved validation strategies for newly proposed models will undoubtedly improve the accuracy of structure prediction and allow more accurate prediction of the metabolizing and inhibitory profiles of major human P450s. Site-directed mutagenesis has been employed as a powerful tool to identify functionally significant residues involved in substrate binding, and has aided development of some P450 models. Analyzing the interactions of different inhibitors with invariant residues in the P450 active site offers another approach to probe P450 active-site structure. Such model validation offers the advantage that a constant, native protein sequence is retained. Experimental data may be applied to constrain model development. For example, NMR-derived distances between the heme iron and codeine substrate protons were used in generating a P450 2D6 model [57]. The primary objective in many modeling studies is to provide a molecular rationale for the observed substrate specificity profiles of individual P450s. We would like to know why a particular P450 prefers one substrate over another as well as the basis for its stereoselectivity. Conversely, why is a given substrate metabolized by one P450 rather than another? Although current P450 mod-

els provide an explanation for some substrate specificities, a wider range of substrates per P450 needs to be evaluated. Modeling studies typically show how a known substrate docks into an active site, but the question as to whether a non-substrate is also capable of similarly binding this region is usually not addressed. A critique of the modeling arena would suggest the more rigorous validation criterion that the model be able to rationalize not only why a given molecule is a substrate, but also why other molecules are not substrates. This suggests that future modeling efforts address not only the 'why' but also the 'why not' aspect of substrate docking.

Acknowledgements. We are very grateful to Drs. Stanley V. Smith and Michael Bustin, and Mr. Richard C. Robinson for helpful discussions and suggestions.

1 Ortiz de Montellano P. (1995) Cytochrome P450 Structure, Mechanism, and Biochemistry, 2nd edn, Plenum, New York
2 Nelson D. R., Koymans L., Kamataki T., Stegeman J. J., Feyereisen R., Waxman D. J. et al. (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics **6:** 1–42
3 Poulos T. L., Finzel B. C. and Howard A. J. (1987) High-resolution crystal structure of cytochrome P450cam. J. Mol. Biol. **195:** 687–700
4 Ravichandran K. G., Boddupalli S. S., Hasermann C. A., Peterson J. A. and Deisenhofer J. (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science **261:** 731–736
5 Hasemann C. A., Ravichandran K. G., Peterson J. A. and Deisenhofer J. (1994) Crystal structure and refinement of cytochrome P450terp at 2.3 Å resolution. J. Mol. Biol. **236:** 1169–1185
6 Cupp-Vickery J. R. and Poulos T. L. (1995) Structure of cytochrome P450eryF involved in erythromycin biosynthesis. Nat. Struct. Biol. **2:** 144–153
7 Nelson D. R. and Strobel H. W. (1988) On the membrane topology of vertebrate cytochrome P-450 proteins. J. Biol. Chem. **263:** 6038–6050
8 Nelson D. R. and Strobel H. W. (1989) Secondary structure prediction of 52 membrane-bound cytochromes P450 shows a strong structural similarity to P450cam. Biochemistry **28:** 656–660
9 Edwards R. J., Murray B. P., Boobis A. R. and Davies D. S. (1989) Identification and location of alpha-helices in mammalian cytochromes P450. Biochemistry **28:** 3762–3770
10 Gotoh O. (1992) 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
11 Lewis D. F. and Moereels H. (1992) The sequence homologies of cytochromes P-450 and active-site geometries. J. Comput. Aided. Mol. Des. **6:** 235–252
12 Korzekwa K. R. and Jones J. P. (1993) Predicting the cytochrome P450 mediated metabolism of xenobiotics. Pharmacogenetics **3:** 1–18
13 Lewis D. F. (1995) Three-dimensional models of human and other mammalian microsomal P450s constructed from an alignment with P450102 (P450BM3). Xenobiotica **25:** 333–366
14 Hasemann C. A., Kurumbail R. G., Boddupalli S. S., Peterson J. A. and Deisenhofer J. (1995) Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure **3:** 41–62
15 Ferenczy G. G. and Morris G. M. (1989) The active-site of cytochrome P-450 nifedipine oxidase – a model-building study. J. Mol. Graph. **7:** 206–211
16 Lewis, D. F. V. (1996) Molecular modeling of mammalian cytochromes P450. In: Cytochromes P450: Metabolic and Toxicological Aspects, pp. 355–398, Ioannides C. (ed.), CRC, Boca Raton
17 Szklarz G. D. and Halpert J. R. (1997) Use of homology modeling in conjunction with site-directed mutagenesis for analysis of structure-function relationships of mammalian cytochromes P450. Life Sci. **61:** 2507–2520
18 Groot M. J. de and Vermeulen N. P. (1997) Modeling the active sites of cytochrome P450s and glutathione S-transferases, two of the most important biotransformation enzymes. Drug Metab. Rev. **29:** 747–799
19 Peterson J. A. and Graham S. E. (1998) A close family resemblance: the importance of structure in understanding cytochromes P450. Structure **6:** 1079–1085
20 Nelson D. R. (1998) Metazoan cytochrome P450 evolution. Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol. **121:** 15–22
21 Jean P., Pothier J., Dansette P. M., Mansuy D. and Viari A. (1997) Automated multiple analysis of protein structures: application to homology modeling of cytochromes P450. Proteins **28:** 388–404
22 Li H. and Poulos T. L. (1997) The structure of the cytochrome P450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat. Struct. Biol. **4:** 140–146
23 Lewis D. F. (1998) The CYP2 family: models, mutants and interactions. Xenobiotica **28:** 617–661
24 Lewis D. F. and Lake B. G. (1996) Molecular modelling of CYP1A subfamily members based on an alignment with CYP102: rationalization of CYP1A substrate specificity in terms of active site amino acid residues. Xenobiotica **26:** 723–753
25 Dai R., Zhai S., Wei X., Pincus M. R., Vestal R. E. and Friedman F. K. (1998) Inhibition of human cytochrome P450 1A2 by flavones: a molecular modeling study. J. Protein Chem. **17:** 643–650
26 Lewis D. F. V., Dickens M., Lake B. G., Eddershaw P. J., Tarbit M. H. and Goldfarb P. S. (1999) Molecular modelling of the human cytochrome P450 isoform CYP2A6 and investigations of CYP2A substrate selectivity. Toxicology **133:** 1–33
27 Lewis D. F. and Lake B. G. (1997) Molecular modelling of mammalian CYP2B isoforms and their interaction with substrates, inhibitors and redox partners. Xenobiotica **27:** 443–478
28 Dai R., Pincus M. R. and Friedman F. K. (1998) Molecular modeling of cytochrome P450 2B1: mode of membrane insertion and substrate specificity. J. Protein Chem. **17:** 121–129
29 Lewis D. F., Lake B. G., Dickins M., Eddershaw P. J., Tarbit M. H. and Goldfarb P. S. (1999) Molecular modelling of CYP2B6, the human CYP2B isoform, by homology with the substrate-bound CYP102 crystal structure: evaluation of CYP2B6 substrate characteristics, the cytochrome b_5 binding site and comparisons with CYP2B1 and CYP2B4. Xenobiotica **29:** 361–393
30 Lewis D. F., Bird M. G. and Parke D. V. (1997) Molecular modelling of CYP2E1 enzymes from rat, mouse and man: an explanation for species differences in butadiene metabolism and potential carcinogenicity, and rationalization of CYP2E substrate specificity. Toxicology **118:** 93–113
31 Tan Y., White S. P., Paranawithana S. R. and Yang C. S. (1997) A hypothetical model for the active site of human cytochrome P4502E1. Xenobiotica **27:** 287–299
32 Lewis D. F., Eddershaw P. J., Goldfarb P. S. and Tarbit M. H. (1996) Molecular modelling of CYP3A4 from an alignment with CYP102: identification of key interactions between putative active site residues and CYP3A-specific chemicals. Xenobiotica **26:** 1067–1086
33 Lewis D. F. V. and Lake B. G. (1999) Molecular modelling of CYP4A subfamily members based on sequence homology with CYP102. Xenobiotica **29:** 763–781

34 Burke D. F., Laughton C. A. and Neidle S. (1997) Homology modelling of the enzyme P450 17 alpha-hydroxylase/17,20-lyase – a target for prostate cancer chemotherapy – from the crystal structure of P450BM-3. Anticancer Drug Des. **12:** 113–123

35 Shyue S. K., Ruan K. H., Wang L. H. and Wu K. K. (1997) Prostacyclin synthase active sites: identification by molecular modeling-guided site-directed mutagenesis. J. Biol. Chem. **272:** 3657–3662

36 Ruan K. H., Milfeld K., Kulmacz R. J. and Wu K. K. (1994) Comparison of the construction of a 3-D model for human thromboxane synthase using P450cam and BM-3 as templates – implications for the substrate-binding pocket. Protein Eng. **7:** 1345–1351

37 Graham-Lorence S., Amarneh B., White R. E., Peterson J. A. and Simpson E. R. (1995) A three-dimensional model of aromatase cytochrome P450. Protein Sci. **4:** 1065–1080

38 Groot M. J. de, Vermeulen N. P., Kramer J. D., Acker F. A. van and Donne-Op d. K. G. (1996) 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

39 Szklarz G. D., He Y. A. and Halpert J. R. (1995) Site-directed mutagenesis as a tool for molecular modeling of cytochrome P450 2B1. Biochemistry **34:** 14312–14322

40 Lewis D. F. and Lake B. G. (1995) Molecular modelling of members of the P4502A subfamily: application to studies of enzyme specificity. Xenobiotica **25:** 585–598

41 Rost B. and Sander C. (1993) Prediction of protein secondary structure at better than 70- percent accuracy. J. Mol. Biol. **232:** 584–599

42 Rost B. and Sander C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Prot. Struct. Funct. Genet. **19:** 55–72

43 Szklarz G. D., Ornstein R. L. and Halpert J. R. (1994) Application of 3-dimensional homology modeling of cytochrome P450 2B1 for interpretation of site-directed mutagenesis results. J. Biomol. Struct. Dyn. **12:** 61–78

44 Szklarz G. D. and Halpert J. R. (1997) Molecular modeling of cytochrome P450 3A4. J. Comput. Aided Mol. Des. **11:** 265–272

45 Chang Y. T., Stiffelman O. B., Vakser I. A., Loew G. H., Bridges A. and Waskell L. (1997) Construction of a 3D model of cytochrome P450 2B4. Protein Eng. **10:** 119–129

46 Chang Y. T. and Loew G. H. (1999) Homology modeling and substrate binding study of human CYP4A11 enzyme. Proteins **34:** 403–415

47 Przytycka T., Aurora R. and Rose G. D. (1999) A protein taxonomy based on secondary structure. Nat. Struct. Biol. **6:** 672–682

48 Koley A. P., Robinson R. C. and Friedman F. K. (1996) Cytochrome P450 conformation and substrate interactions as probed by CO binding kinetics. Biochimie **78:** 706–713

49 Frauenfelder H. (1997) The complexity of proteins. In: Physics of Biological Systems: from Molecules to Species, pp. 29–60, Flyvbjerg H., Hertz J., Jensen M. H., Mouritsen O. G. and Snepper K. (eds), Springer, Berlin

50 Ruan K. H., Li D. W., Ji J., Lin Y. Z. and Gao X. L. (1998) Structural characterization and topology of the second potential membrane anchor region in the thromboxane A(2) synthase amino-terminal domain. Biochemistry **37:** 822–830

51 Miller J. P., Herbette L. G. and White R. E. (1996) X-ray diffraction analysis of cytochrome P450 2B4 reconstituted into liposomes. Biochemistry **35:** 1466–1474

52 Shank-Retzlaff M. L., Raner G. M., Coon M. J. and Sligar S. G. (1998) Membrane topology of cytochrome P450 2B4 in Langmuir-Blodgett monolayers. Arch. Biochem. Biophys. **359:** 82–88

53 Modi S., Sutcliffe M. J., Primrose W. U., Lian L. Y. and Roberts G. C. (1996) The catalytic mechanism of cytochrome P450 BM3 involves a 6 A movement of the bound substrate on reduction. Nat. Struct. Biol. **3:** 414–417

54 Lee H., Montellano P. R. O. de and McDermott A. E. (1999) Deuterium magic angle spinning studies of substrates bound to cytochrome P450. Biochemistry **38:** 10808–10813

55 Parikh A., Josephy P. D. and Guengerich F. P. (1999) Selection and characterization of human cytochrome P450 1A2 mutants with altered catalytic properties. Biochemistry **38:** 5283–5289

56 Yun C. H., Hammons G. J., Jones G., Martin M. V., Hopkins N. E., Alworth W. L. et al. (1992) Modification of cytochrome-P450 1A2 enzymes by the mechanism-based inactivator 2-ethynylnaphthalene and the photoaffinity label 4-azidobiphenyl. Biochemistry **31:** 10556–10563

57 Modi S., Paine M. J., Sutcliffe M. J., Lian L.Y., Primrose W. U., Wolf C. R. et al. (1996) A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. Biochemistry **35:** 4540–4550

58 Omata Y., Sakamoto H., Robinson R. C., Pincus M. R. and Friedman F. K. (1994) Interaction between cytochrome P450 2B1 and cytochrome b_5: inhibition by synthetic peptides indicates a role for P450 residues Lys-122 and Arg-125. Biochem. Biophys. Res. Commun. **201:** 1090–1095

59 DeVoss J. J., Sibbesen O., Zhang Z. P. and deMontellano P. R. O. (1997) Substrate docking algorithms and prediction of the substrate specificity of cytochrome P450(cam) and its L244A mutant. J. Am. Chem. Soc. **119:** 5489–5498

60 Ekins S., Bravi G., Binkley S., Gillespie J. S., Ring B. J., Wikel J. H. et al. (1999) Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors. Pharmacogenetics **9:** 477–489

61 Ekins S., Bravi G., Binkley S., Gillespie J. S., Ring B. J., Wikel J. H. et al. (1999) Three- and four-dimensional quantitative structure activity relationship analyses of cytochrome P-450 3A4 inhibitors. J. Pharmacol. Exp. Ther. **290:** 429–438

62 Groot M. J. de, Ackland M. J., Horne V. A., Alex A. A. and Jones B. C. (1999) Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. J. Med. Chem. **42:** 1515–1524

63 Groot M. J. de, Ackland M. J., Horne V. A., Alex A. A. and Jones B. C. (1999) A novel approach to predicting P450 mediated drug metabolism: CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. J. Med. Chem. **42:** 4062–4070