Protein Engineering vol.10 no.2 pp.119–129, 1997

Construction of a 3D Model of cytochrome P450 2B4

Yan-Tyng Chang[1,4], Oscar B.Stiffelman[1], Ilya A.Vakser[2], Gilda H.Loew[1], Angela Bridges[3] and Lucy Waskell[3]

[1]Molecular Research Institute, 845 Page Mill Road, Palo Alto, CA 94304, [2]The Rockefeller University, Box 270, 1230 York Avenue, New York, NY 10021 and [3]Department of Anesthesia and the Liver Center, University of California, San Francisco and the Department of Anesthesia, Veterans Administration Medical Center, San Francisco, CA 94121, USA

[4]To whom correspondence should be addressed.

A three-dimensional structural model of rabbit phenobarbital-inducible cytochrome P450 2B4 (LM2) was constructed by homology modeling techniques previously developed for building and evaluating a 3D model of the cytochrome P450choP isozyme. Four templates with known crystal structures including cytochrome P450cam, terp, BM-3 and eryF were used in multiple sequence alignments and construction of the cytochrome P450 2B4 coordinates. The model was evaluated for its overall quality using available protein analysis programs and found to be satisfactory. The model structure was stable at room temperature during a 140 ps unconstrained full protein molecular dynamics simulation. A putative substrate access channel and binding site were identified. Two different substrates, benzphetamine and androstenedione, that are metabolized by cytochrome P450 2B4 with pronounced product specificity were docked into the putative binding site. Two orientations were found for each substrate that could lead to the observed preferred products. Using a geometric fit method three regions on the surface of the model cytochrome P450 structure were identified as possible sites for interaction with cytochrome b_5, a redox partner of P450 2B4. Residues that may interact with the substrates and with cytochrome b_5 have been identified and mutagenesis studies are currently in progress.

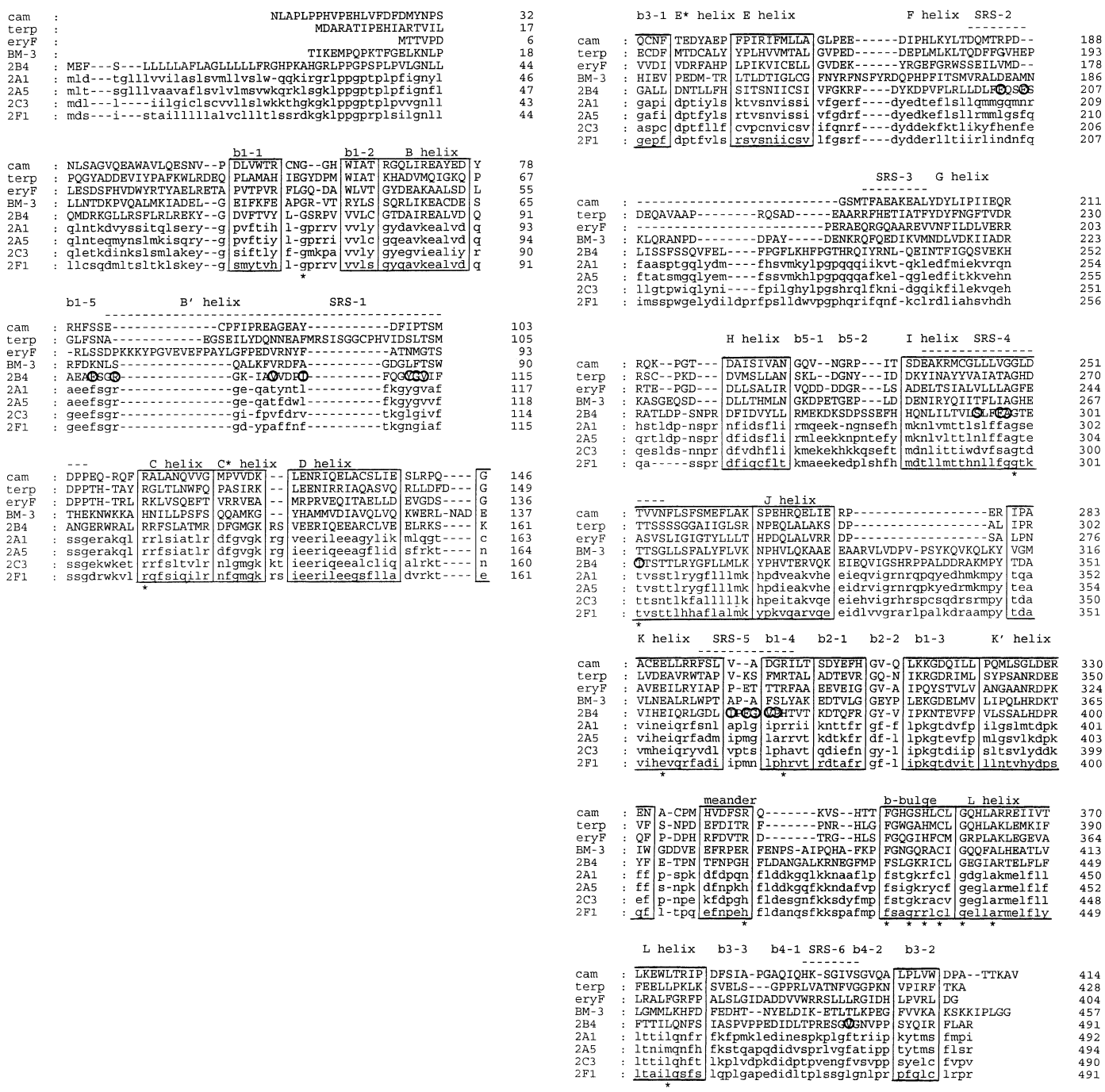
Keywords: androstenedione/benzphetamine/cytochrome b_5/homology modeling/P450 2B4

Introduction

The cytochromes P450 are a superfamily of enzymes that are responsible for the oxidative metabolism of many endogenous and exogenous compounds and have been isolated from plants, fungi, bacteria, insects and mammals. It has been commonly accepted that the enzymatic cycle of these enzymes includes substrate binding, electron transfer, oxygen binding and activation and substrate oxidation. However, there are many differences among the cytochromes P450. For example, the two electrons required in the enzymatic cycle in bacterial P450 systems, with the only exception of P450 BM-3, are derived from NADH; whereas in mitochondrial and microsomal P450s systems they are derived from NADPH. The redox partners involved in the electron transport chain for bacterial and mitochondrial cytochromes P450 are flavin adenine dinucleotide (FAD) containing reductase and iron–sulfur proteins (class I), whereas those involved in the electron transport for microsomal cytochromes P450 are reductases containing FAD and flavin mononucleotide (FMN) (class II). In addition, although all cytochromes P450 contain a heme prosthetic group which functions to generate a reactive oxygen during substrate oxidation, different cytochrome P450 isozymes exhibit overlapping but different substrate specificity and product regio- and/or stereospecificity. The elucidation of the molecular origin of the mechanistic and functional differences among these enzymes is difficult in large part because their three-dimensional structures are unknown. Experimental structural determination of the mammalian cytochromes P450 has proven difficult since they are membrane-bound proteins and in addition are difficult to purify to homogeneity. Model structures of a number of mammalian cytochrome P450 isozymes including 1A1 (Zvelebil *et al.*, 1991; Lewis and Moereels, 1992), 2B1 (Szklarz *et al.*, 1994), 2D6 (Koymans *et al.*, 1993), 3A4 (P450NF) (Ferenczy and Morris, 1989), 11α (Vijayakumar and Salerno, 1992), 14α (Morris and Richards, 1991; Boscott and Grant, 1994), 17α (Laughton *et al.*, 1990) and 19 (aromatase) (Laughton *et al.*, 1993) have been built based on the first available crystal structure of cytochrome P450, namely the soluble bacterial cytochrome P450cam (Poulos *et al.*, 1986). However, the low homology (<20% sequence identity) between cytochrome P450cam and mammalian cytochromes P450 makes the resulting sequence alignments and the 3D models of questionable intrinsic reliability (Poulos, 1991).

During the past few years, the crystal structures of three more bacterial cytochromes P450; including cytochrome P450 terp (Boddupalli *et al.*, 1992; Hasemann *et al.*, 1994), P450 BM-3 (Ravichandran *et al.* 1993; Li and Poulos, 1995) and cytochrome P450eryF (Cupp-Vickery and Poulos, 1994, 1995) have been solved. The use of these additional crystal structures as templates should improve the reliability of the resulting models. In particular, the inclusion of cytochrome P450 BM-3 as one of the templates is especially useful in homology modeling of mammalian cytochromes P450 since this isozyme is a class II enzyme and is known to resemble eukaryotic cytochromes P450 more closely in sequence and functional properties than it does other bacterial cytochromes P450 (Ruettinger *et al.*, 1989). For example, Ruan *et al.* (1994) reported two homology models of thromboxane synthase (CYP 5) based on two separate templates, P450cam and P450 BM-3, and concluded that that based on P450 BM-3 was a better model. Recently, Graham-Lorence *et al.* (1995) reported a model structure of cytochrome P450 aromatase (CYP19), based on three templates: cytochrome P450cam, terp and BM-3. We have also reported a model structure of the bacterial cytochrome P450choP enzyme using all four known structures including cytochrome P450cam, terp, BM-3 and eryF as templates in sequence alignment and model construction. The main purpose of our previous work on cytochrome P450choP was to develop and evaluate carefully strategies for constructing 3D models of cytochromes P450. In developing the cytochrome P450choP model structure, we have relied heavily on the

Table I. Multiple sequence alignments among P450cam, P450terp, P450BM-3, P450eryF, P450 2B4, P450 2A1, P450 2A5, P450 2C3 and P450 2F1



SCRs are shown with boxes. Substrate recognition sites (SRS) proposed for P450 2 family by Gotoh (1992) are indicated with dashed lines. Invariant or highly conserved residues among all cytochromes P450 are indicated with asterisks. Residues in contact with the cytochrome P450 2B4 substrate binding site are circled.

structurally conserved regions of the four templates when performing multiple sequence alignments. A simulated annealing/Monte Carlo method called Torso (Holm and Sander, 1991) was used to find the initial side chain conformation. Structural waters were added systematically using an in-house program, HYDRAS (Zhao *et al.*, 1996), in order to fulfil local hydrogen bonding networks and to stabilize the model structure, allowing unconstrained full protein energy minimization. Various protein analysis programs were used to evaluate the resulting model structure, including its backbone conformations (Quanta/Protein Health) (Molecular Simulations, 1994), the residue interactions (Prosa) (Sippl, 1993), local environment (Profiles 3D) (Luthy *et al.*, 1992) and residue contact (Whatif Quality Control) (Vriend and Sander, 1993). The results of these assessments, both for the model itself and when compared with assessments obtained for the four template structures,

Table II. The seven regions defined for local sequence alignment

Sequence	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7
	N terminus to B helix	β1–5 to D helix	β3–1 to E helix	F helix to J helix	K helix	β1–4 to K′ helix	Meander to C terminus
cam	1–77	78–141	142–175	169–277	278–295	296–333	334–414
terp	1–66	67–143	144–178	172–296	297–314	315–353	354–428
BM-3	1–64	65–129	130–171	159–293	294–328	329–368	369–457
eryF	1–54	55–131	132–148	159–270	271–288	289–327	328–404

indicated that the model cytochrome P450choP structure was nearly as robust as the four templates. Finally, this model P450choP structure was subjected to unconstrained full protein molecular dynamics simulation and found to be stable.

In this paper, we report the first application of this strategy, modified by an improved method of multiple sequence alignment, for the construction and assessment of a model structure of a mammalian cytochrome P450. The rabbit cytochrome P450 2B4 isozyme, also known as cytochrome P450 LM2, is expressed constitutively in lungs and is induced in the liver after treatment with phenobarbital (Grimm *et al.*, 1994). It was the first cytochrome P450 isozyme to be purified to apparent homogeneity from a mammalian source (Haugen and Coon, 1976) and to have the complete amino acid sequence reported (Heinemann and Ozols, 1983; Tarr *et al.*, 1983). As such, it has been subjected to extensive experimental studies. The report of this model structure of cytochrome P450 2B4 should aid in the design of experiments for further elucidation of the mechanistic relationship between the structure and function of this isozyme.

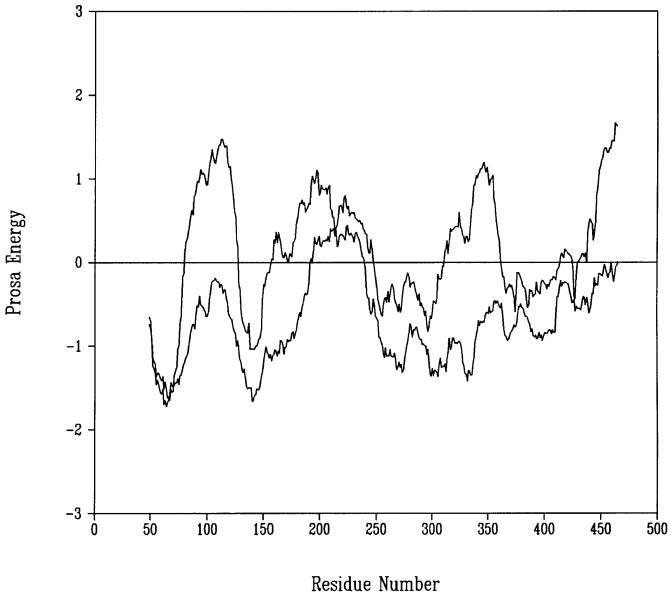


Fig. 1. Prosa energy plot. The top curve represents the residue interaction energy of the unrefined model. The bottom curve is for the refined model.

Construction of a model structure

In performing the sequence alignment of the target cytochrome P450 with the four templates including cytochrome P450cam, terp, BM-3 and eryF, we utilized the sequence alignment based on the structural similarity between the templates themselves obtained in our previous work (Chang and Loew, 1996). Structurally conserved regions (SCRs) among the four template proteins, indicated with boxes in Table I, were defined and identified as regions where the pairwise peptide backbone C_α atom r.m.s. deviations are small (typically <1 Å) when the four template structures are superimposed. The resulting alignment obtained among the templates was used to build both the model of cytochrome P450choP in the previous work and cytochrome P450 2B4 in this work.

The next step is to align the target sequence to the template sequences. In the previous work on cytochrome P450choP, this alignment was done utilizing the information of the predicted secondary structures of cytochrome P450choP and the automatic pairwise alignment between cytochrome P450choP and cytochrome P450eryF. However, because of the low sequence similarity between cytochrome P450 2B4 and the template sequences ($<$20%), the alignment between them cannot be performed reliably using such techniques alone. Instead, a new technique was used which found a more reliable alignment by utilizing the sequences of related cytochromes P450. In this technique, we have profited from the fact that the alignment of cytochrome P450 2B4 with its closely related proteins can be determined accurately by automated multiple sequence alignment techniques. Some of the related proteins exhibit a higher sequence similarity with the template proteins than does cytochrome P450 2B4. Therefore, a more confident alignment can be performed between those related cytochromes P450 and the template proteins than could be performed between cytochrome P450 2B4 alone and the template proteins. Once a related protein is aligned to the template proteins, cytochrome P450 2B4 can subsequently be aligned to the template proteins because its alignment to the related protein has already been determined. In addition, a choice was made to use local, rather than global, sequence alignments so that each region in the template proteins could be aligned to cytochrome P450 2B4 using only a related sequence that exhibited the highest similarity in that region. A multistep procedure was therefore developed to implement this idea. In the first step, a multiple sequence alignment of cytochrome P450 2B4 with 10 highly homologous sequences from the cytochrome P450 2 family, 2A1, 2A5, 2B2, 2C3, 2C4, 2C21, 2D1, 2E1, 2F1 and 2G1, was obtained. This alignment was performed by adopting the multiple sequence alignment among these 10 highly homologous sequences from the cytochrome P450 2 family already obtained by Gotoh (1992) and adding to it the complete sequence (491 amino acids) of rabbit cytochrome P450 2B4 isoform B0, retrieved from the Swiss-Prot database.

In the second step, seven local regions along the SCR-aligned template consensus sequence were defined as shown in Table II, each of which includes at least one SCR of the templates. Then, in the third step, in a series of pairwise alignments, each of the 11 cytochrome P450 2 sequences was aligned automatically with each of the seven regions of each

template protein using the Dayhoff mutation matrix (Dayhoff *et al.*, 1983). This third step resulted in 44 alignments for each of the seven regions. In the fourth step, the alignment scores and consistency among the five best alignments obtained for each region were examined and one of them was selected to guide the multiple sequence alignments between the group of 11 cytochrome P450 2 sequences and those of the four template proteins. The following seven pairwise alignments were selected for the seven regions: (i) 2A5–BM3, (ii) 2A5–cam, (iii) 2F1–terp, (iv) 2A1–BM3, (v) 2C3–BM3, (vi) 2C3–terp and (vii) 2C3–BM3, respectively. This resulting multiple sequence alignment was then used to build an initial model of the cytochrome P450 2B4 enzyme. Since the first 20 residues from the N-terminus of cytochrome P450 2B4 were suggested to be membrane bound (Uvarov *et al.*, 1994), the model structure begins at residue 25.

The remaining procedures for constructing the 3D coordinates were described in detail previously (Chang and Loew, 1996). Briefly, backbone atoms in the SCR regions were copied from one of the templates. In most regions, the structure from BM-3 was used. For the non-SCR regions, it can be seen in the alignment in Table I that there are no long insertions or deletions of cytochrome P450 2B4 relative to BM-3. Therefore, backbone atom coordinates in the non-SCRs for cytochrome P450 2B4 were copied directly from the cytochrome P450 BM-3 structure if the sequence lengths in the two proteins were the same. In the cases where the sequence lengths in these two proteins differ by a few residues in structurally unconserved regions, backbone atom coordinates were generated via loop search. Using Insight II (Biosym Technologies, 1994), 10 loop structures were generated for each non-SCR. One loop structure was then selected based on (1) the magnitude of the r.m.s. error of the two terminal residues in the searched segment and (2) visual inspection of steric interaction of a loop structure with its environment. Side chain conformations were generated using the rotamer library embeded in the program Torso (Holm and Sander, 1991). The best side chain conformation from this library of each residue was determined using a simulated annealing/Monte Carlo procedure. The ability of this procedure to predict correct side chain conformations was assessed in our previous study using the four template structures. It was found to predict more than 80% of χ_1 angles correctly to within 40° (Chang and Loew, 1996).

At this stage, an initial evaluation of the folding and steric contacts in the preliminary model structure was performed using the Protein Structure Analysis program (Prosa) (Sippl, 1993). Using this program, regions that may be misfolded due to alignment errors or have bad steric contacts are revealed by the presence of large repulsive (positive) residue interaction energy with the rest of the protein. As shown in the top curve of Figure 1, four regions including the B′ helix, the F–G helices, the K helix and the C-terminus region in this initial model were found to have large positive energy peaks and required modification. Therefore, the alignments and/or structures in these local regions were examined and varied gradually in order to lower the interaction energies. Specifically, a small modification of the sequence alignment of each of these local regions was performed to allow a smoother continuity with its neighboring regions and more matched residues between the target and the template sequences in the connecting regions. This modified sequence alignment then resulted in a shorter loop whose structure was determined by the following procedure. Candidate structures for each loop were found by searching the PDB databank according to the C_α distance matrix. Each candidate loop structure was then examined for the extent to which it reduced the bad steric contact of the residues in the loop with its surrounding when compared with the loop structure in the initial model. This procedure was repeated until no further improvement in the interaction energy could be obtained. Shown in the bottom curve of Figure 1 is the Prosa residue interaction energy of the improved model structure. The large positive energy peaks in the B′ helix, K

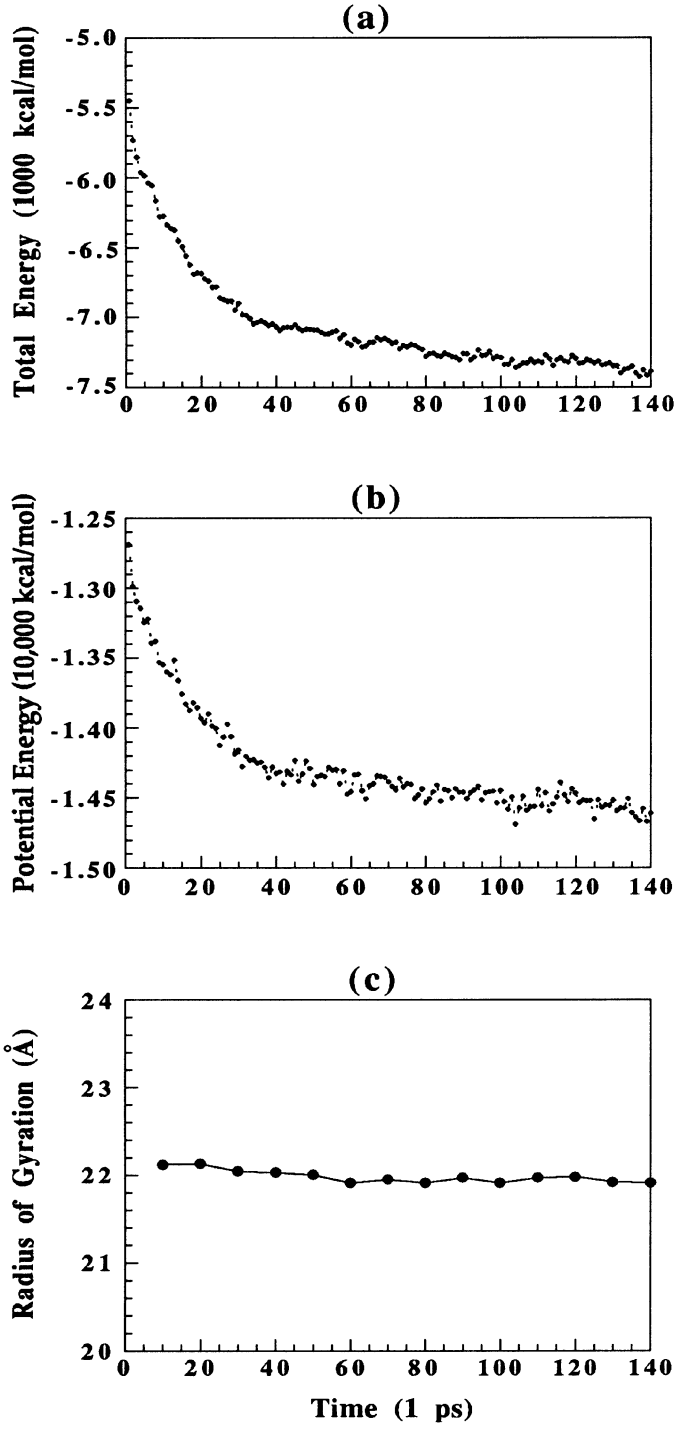


Fig. 2. (**a**) Total energy, (**b**) potential energy and (c) radius of gyration of the cytochrome P450 2B4 model during 140 ps of a molecular dynamics simulation.

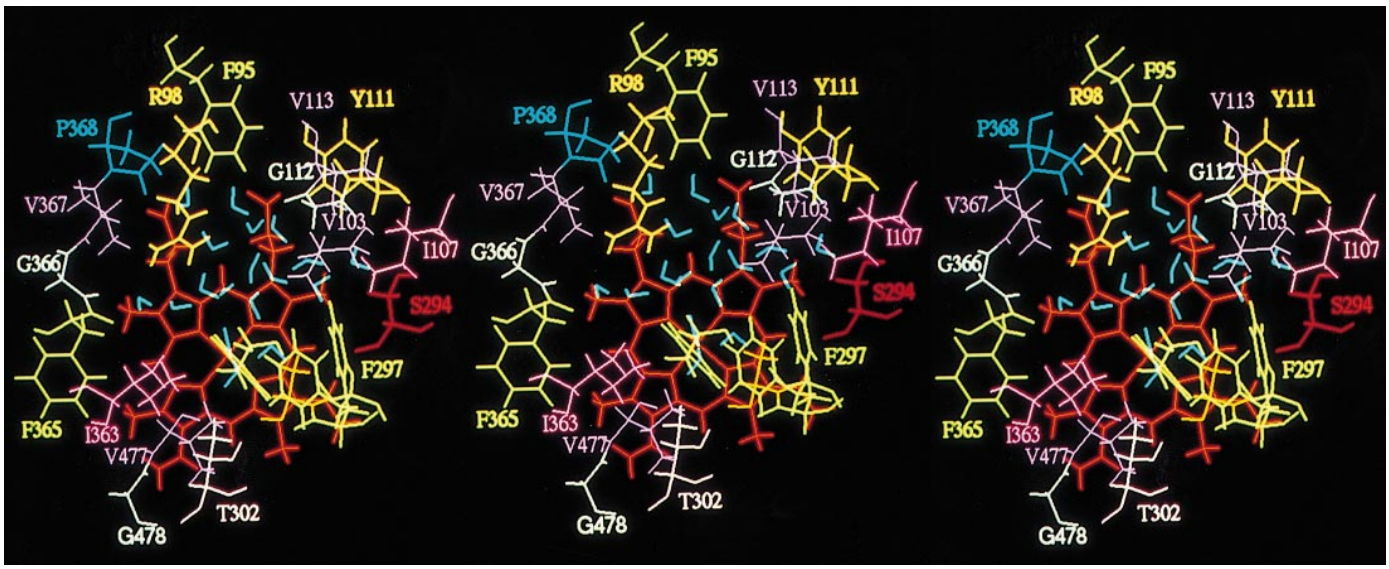


Fig. 3. Stereo view of the substrate binding site of the cytochrome P450 2B4 model looking from above the heme unit. The 18 water molecules occupying the putative substrate binding site are shown in blue. Their positions are predominately above the C and D rings of the heme and the two propionate groups.

helix and C terminus region in the initial model have disappeared and the interaction energies have become negative. The positive energy peak in the F–G helix region still exists but is lowered significantly. The final multiple sequence alignment resulting from this step among cytochrome P450 2B4, 2A1, 2A5, 2C3 and 2F1 and the four templates is shown in Table I. The nomenclature used by Hasemann *et al.* (1994,1995) for the helices and β-sheets was adopted in this table.

Constrained energy minimization of the side chains and structurally unconserved regions was then performed with this improved model using Amber 4.1 (Pearlman *et al.*, 1995). Structural waters, including buried and surface waters, were added to stabilize the model before full protein minimization. This step was performed using the HYDRAS program developed in our laboratory that examines the local residue environment for hydrogen-bonding and steric interactions between residues and waters. The ability of this method to predict the position of bound waters has been assessed by using it to add bound waters to the three known crystal structures of peroxidases (Zhao *et al.*, 1996) and comparing the results with those for crystallographic bound waters. The method was found to reproduce the crystallographic bound waters and to have additional waters mostly in the surface of the protein. Using this procedure, 782 structural waters were added to the cytochrome P450 2B4 model. Energy minimization of the water molecules and polar hydrogens in the amino acid residues was then performed to allow the best hydrogen-bonding network to form. Finally, an unconstrained full protein minimization using a non-bonding cut-off of 8 Å and a distance-dependent radial dielectric $\varepsilon = r$ was performed. All residues, the structural water molecules and the ferryl heme unit were allowed to move in the minimization for 5000 steps using combined steepest descent and conjugate gradient methods to obtain the final refined model.

Evaluation of the refined model

The refined model was then evaluated for its overall quality using available protein analysis programs. The principles and methods used in these programs were described previously

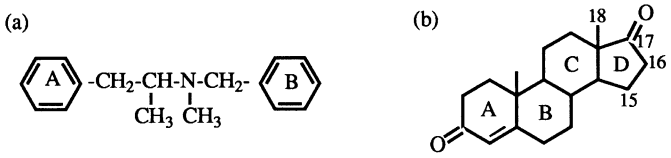


Fig. 4. Structures of (**a**) benzphetamine and (**b**) androstenedione

(Chang and Loew, 1996) and will not be repeated here. (1) Using Quanta/Protein Health, the backbone ϕ/ψ angles of ~93% of the residues were found to be in allowed region of a Ramachandran plot, compared with ~97% for the choP model and 98–99% for the four templates. (2) Using Prosa, a normalized *z* score over sequence length of 0.7 or higher was recommended for a good protein structure. The normalized *z* score was found to be 0.65 for this model, slightly below the recommended limit. In comparison, values of 0.97–1.12 were reported for the four templates and the cytochrome P450choP model (Chang and Loew, 1996). (3) Using Profiles-3D, the recommended criterion for a reasonable structure is that its normalized *s* score over sequence length is >0.45. The normalized *s* score was found to be 0.79 for this model, well above the recommended value. In comparison, the normalized *s* score was found to be 0.95 for the choP model and 1.00–1.12 for the four templates. (4) Using Whatif quality control, if the Whatif score is <–3, the model is definitely of low quality. If the value is >–1, it is assessed as a high quality structure. The Whatif score of this model was found to be –1.7. In comparison, a Whatif value of –0.5 to –0.75 was found for the four templates and –1.2 for the cytochrome P450choP model (Chang and Loew, 1996). The quality of a model with value between –1 and –2 is more difficult to assess. Since the Whatif score of the current model is in the ambiguous region (between –1 and –2), to aid in the assessment we subjected three crystal structures (pdb155c, pdb1abp and pdb1prh) in the PDB databank with resolution of 2.4, 2.5 and 3.5 Å to Whatif quality control with resulting scores of –2.25, –1.86 and –1.36, respectively. Therefore, the current model of P450 2B4, although not as good as high-resolution

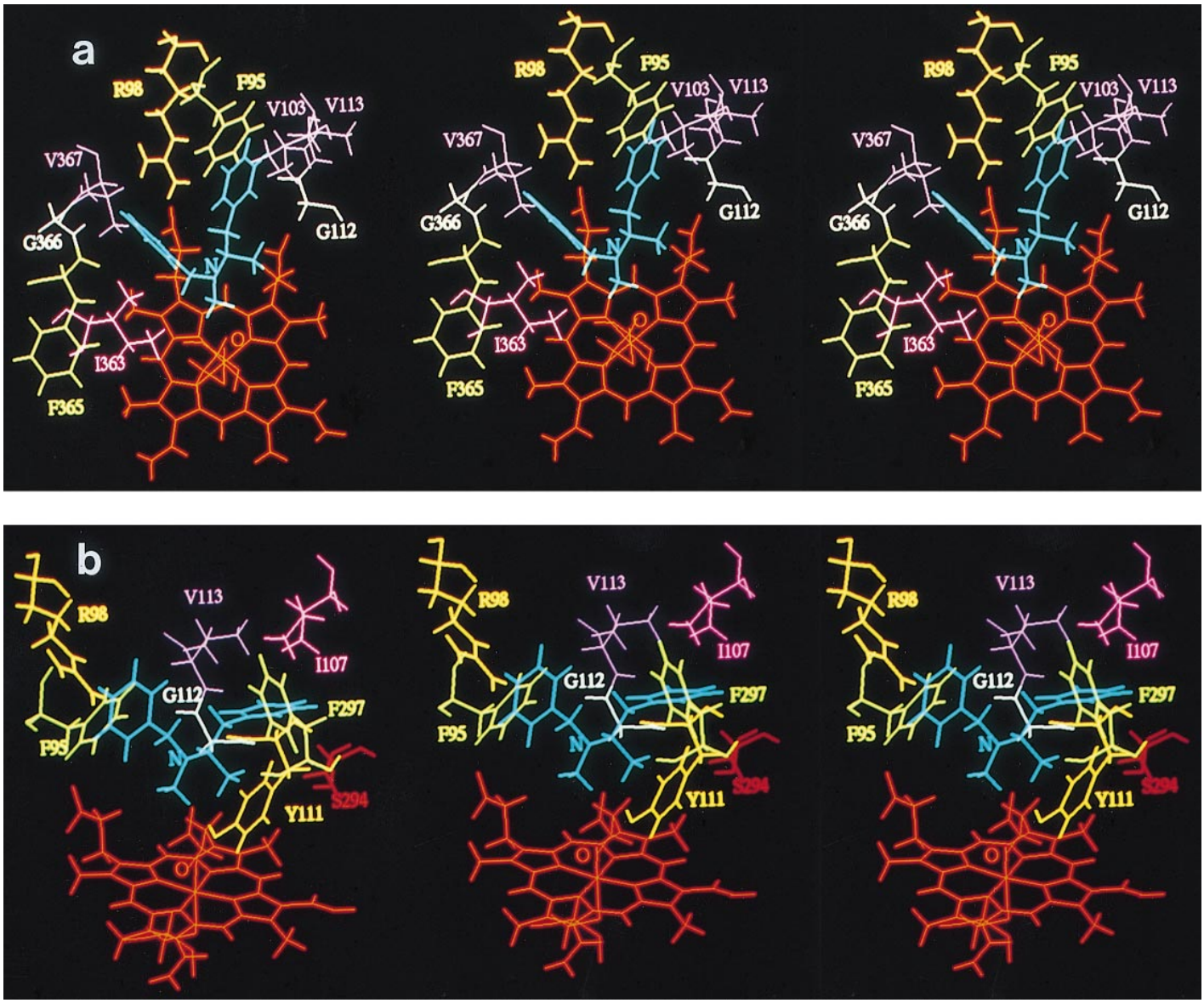


Fig. 5. Stereo views of two different docking orientations of benzphetamine (in blue) in the model cytochrome P450 2B4 binding site. The positions of the N atom of benzphetamine and the O ferryl oxygen atom of the heme (in red) are indicated. (**a**) Top, first orientation; (**b**) bottom, second orientation.

crystal structures and the P450choP model, is of comparable quality to these X-ray structures.

In our previous work, we reported for the first time a model cytochrome P450 structure that is stable when subjected to 120 ps of totally unconstrained molecular dynamics (MD) simulation (Chang and Loew, 1996). This refined cytochrome P450 2B4 model structure was therefore further evaluated for its stability by the same criterion. All residues, the ferryl heme unit and the structural water molecules were included in the molecular dynamics simulation performed using AMBER 4.1. A distance-dependent dielectric constant $\varepsilon = r$ and a long non-bonding cutoff distance of 15 Å were used. The model was kept at 300 K during the entire simulation. At the beginning of the simulation, all atoms were constrained in position with a finite harmonic force constant of $k = 10$ kcal/Å^2 for 1 ps. The constraint was gradually reduced to 5, 1, 0.5 and 0.1 kcal/Å^2 for a total of an additional 5 ps. The constraint was then removed completely and a 140 ps simulation was performed. Figure 2a and b show the total energy and the potential energy of this model during the course of the 140 ps simulation. It appears that the model has equilibrated and remains stable after about 80 ps of dynamic simulation. The stability of this model was further verified by its constant radius of gyration during the simulation, as seen in Figure 2c.

Identification and assessment of substrate binding site

The substrate binding site of the cytochromes P450 is located on the distal side of the heme unit where oxidation of the substrates by the ferryl oxygen occurs. Analysis of the distal pocket above the heme in the substrate-free 3D model of cytochrome P450 2B4 reveals the presence of 18 water molecules. This cluster of bound waters suggests the location of a possible binding site for the substrates. As shown in Figure 3, the residues that enclose the site of the 18 bound waters include F95, R98, V103, I107, Y111, G112 and V113 in the B′ helix region, F202 and F206 in the F helix, S294, F297, A298, T302 in the I helix, I363, F365, G366, V367 and

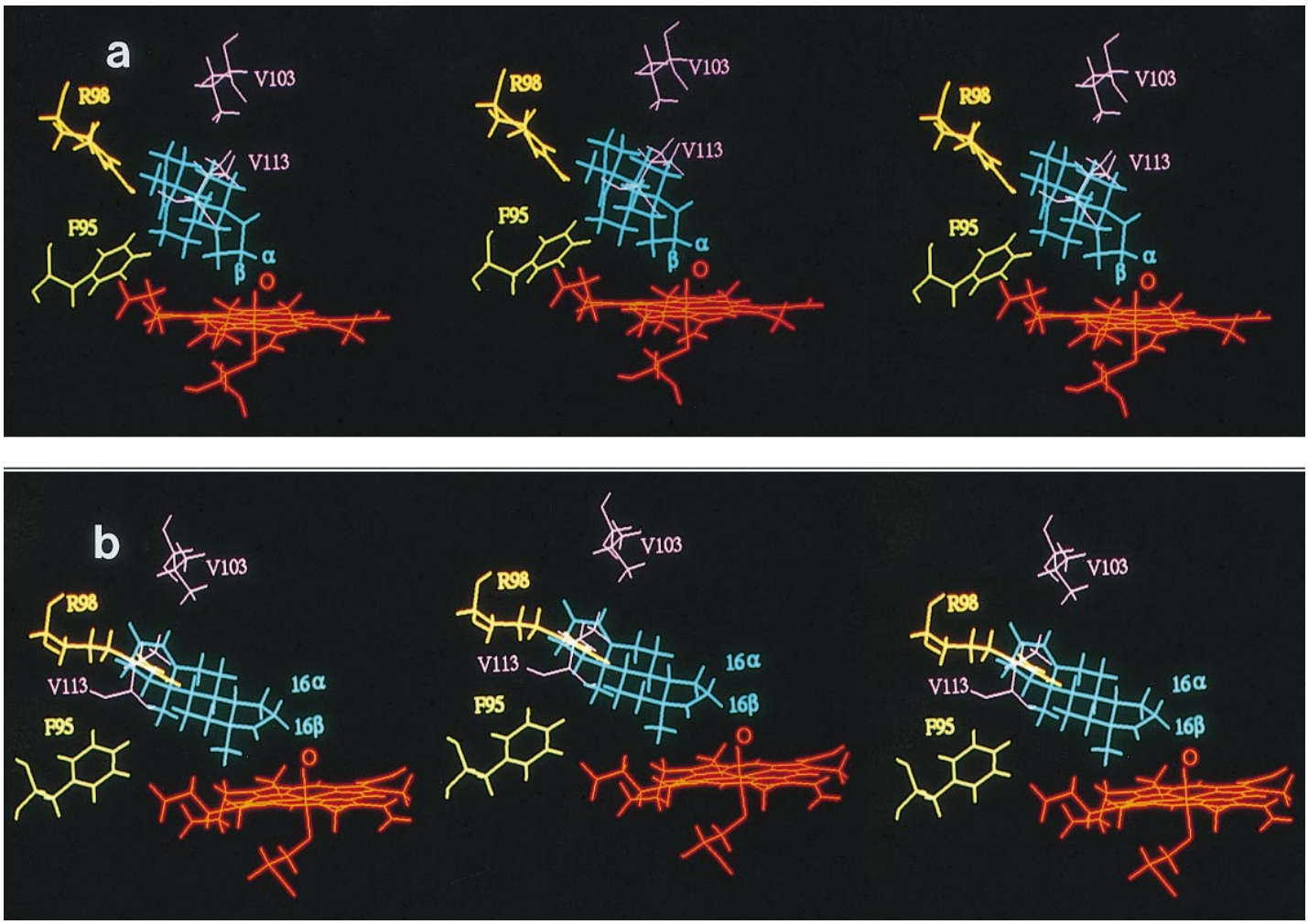


Fig. 6. Stereo views of two orientations of androstenedione (in blue) in the cytochrome P450 2B4 model binding site. The 16α and β hydrogen atoms and the ferryl oxygen atom of the heme (in red) are indicated. (**a**) Top, first orientation; (**b**) bottom, second orientation.

P368 between the K helix and β1–4 and V477 in the β4 hairpin region. These residues are indicated with circles in Table I. Examination of the binding site topology revealed that the space above rings C and D and the two propionate groups of the heme unit is more open than that above rings A and B.

Although we have assessed the internal consistency and stability of the model by a number of rigorous criteria, its usefulness depends heavily on correctly identifying the substrate binding site. One promising comparison is that except for F95, all of the residues comprising the putative binding site are also within the substrate recognition sites (SRSs) (indicated with dashed lines in Table I) proposed by Gotoh (1992) for the cytochrome P450 family 2.

To assess further the proposed binding site, we selected for study two known substrates of cytochrome P4502B4, benzphetamine and androstenedione, that are metabolized to known products. The extent to which these substrates could be docked in the substrate binding site in a configuration that was consistent with the formation of the preferred product was used as further validation of the model.

Cytochrome P450 2B4 metabolizes a wide variety of substrates including compounds such as methoxyflurane, nifedipine, *p*-nitroanisole, prostaglandin, lauric acid, *N*-methylcarbazole, chlorobenzene, *p*-nitrophenetole, 7-ethoxycoumarin, benzo[*a*]pyrene, lidocaine, testosterone, androstenedione and benzphetamine (Gruenke *et al.*, 1995). The *N*-demethylation of benzphetamine is the reaction most commonly used to measure the activity of cytochrome P450 2B4 in microsomes and reconstituted systems (Grimm *et al.*, 1994). *N*-demethylation possibly proceeds via hydroxylation of the methyl group, followed by formation of formaldehyde and the *N*-demethylated product. Since cytochrome P450-catalyzed carbon–hydrogen bond oxidation is believed to occur via hydrogen atom abstraction by the ferryl oxygen of cytochrome P450 followed by rapid OH radical rebound to the substrate, the preferred *N*-demethylation observed suggests that the *N*-CH_3 group of benzphetamine (structure shown in Figure 4a) has to be closest to the ferryl oxygen. Therefore, we determined whether benzphetamine could be docked in the substrate binding site in orientations that would lead to *N*-demethylation. To this end, we used the lowest energy optimized conformer of benzphetamine found in our laboratory using a conformational search and energy minimization (Harris, 1996). Two possible orientations of this conformer of benzphetamine in the cytochrome P450 2B4 binding site were found that could lead to *N*-demethylation. These two cytochrome P450 2B4–benzphetamine complexes were energy minimized with AMBER 4.1 using the same minimization protocols as described in the full protein minimzation of the protein itself.

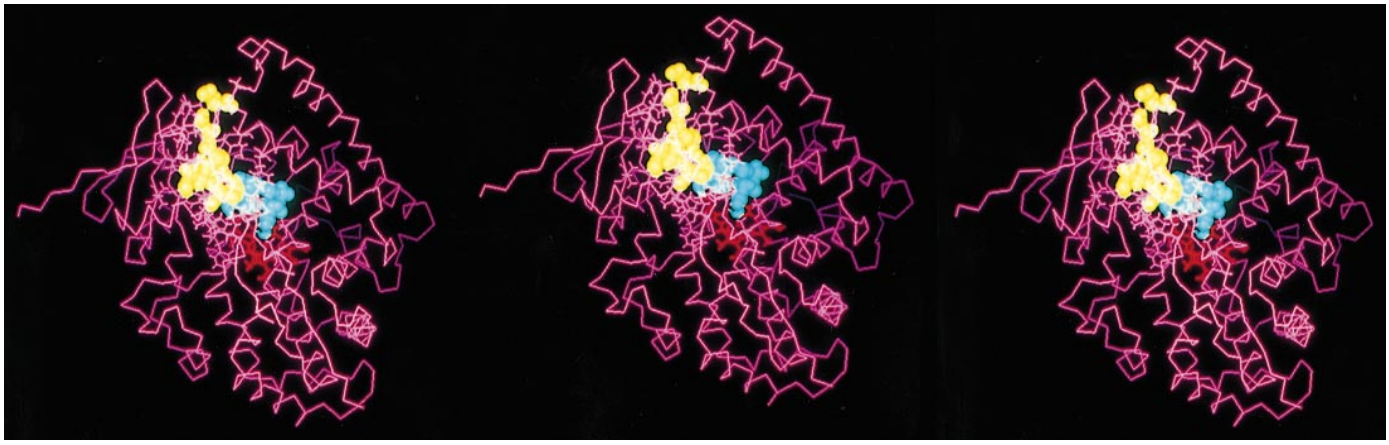


Fig. 7. Stereo view of the proposed substrate access channel indicated with the aggregation of water molecules from the substrate binding site leading to the surface. The 18 water molecules in the binding site are in blue. The other water molecules in the channel are in yellow.

Table III. Residues of cytochrome P450 2B4 that embrace the upper, middle and lower regions of the plausible substrate access channel with (in parentheses) the secondary structural location of each residue

Upper near the protein surface	L43 (A) S213 (F)	L70 (β1)	G71 (β1)	S72 (β1)	F212 (F)
Middle	M46 (A) V77 (β1) G476 (β4)	L51 (A) F206 (F)	L52 (A) I209 (F)	R73 (β1) R473 (β4)	V75 (β1) S475 (β4)
Lower near the heme	R98 (β1) P368 (β1)	P364 (β1) E387 (β1)	F365 (β1) F389 (β1)	G366 (β1)	V367 (β1)

Specifically, during the energy minimization, both the protein and the substrate were allowed to move. No constraint was imposed on any atom or bond in the complex. An examination of the structures of the complexes before and after the minimzation revealed little movement on the protein atoms (the r.m.s. deviation for all heavy atoms was <0.05 Å). The substrate atoms moved slightly, but the magnitudes are in general <0.5 Å. Figure 5 shows these two optimized orientations of benzphetamine in the cytochrome P450 2B4 binding site. In the first orientation (Figure 5a), the phenyl B ring of benzphetamine is in direct contact with R98, I363, F365, G366 and V367 and the phenyl A ring with F95, R98, V103 and G112. R98 bisects the two phenyl rings. In the second orientation (Figure 5b), the benzphetamine phenyl B ring now contacts F95, R98, G112 and V113, while the phenyl A ring makes contacts with I107, Y111, G112, V113, S294 and F297. In this second orientation, G112 and V113 bisect the two phenyl rings of benzphetamine. These two orientations overlap in a region surrounded by F95, R98, V103 and V113 and occupied by one of the phenyl rings of benzphetamine. In both orientations the *N*-methyl group is closer to the ferryl oxygen than any other part of benzphetamine, suggesting that *N*-demethylation can readily occur. The distance between the center of the ferryl oxygen and the center of the N atom of benzphetamine, *r*(O–N), in the first orientation is 4.4 Å and that between the ferryl oxygen and the *N*-methyl carbon atom, *r*(O–C), is about 2.9 Å. The *r*(O–N) and *r*(O–C) distances in the second orientation are 5.0 and 4.0 Å, respectively. The ability of these optimized benzphetamine–cytochrome P450 2B4 complexes to account for the observed product selectivity of benzphetamine provides additional support for the reliability of the cytochrome P450 2B4 model.

The other compound selected, androstenedione, has been used more recently as a specific probe of cytochrome P450 enzymes, because it is hydroxylated with pronounced regio- and stereoselectivity by many isozymes of cytochrome P450. Cytochrome P450 2B4 catalyzes 16β-hydroxylation, but very little 16α-hydroxylation of androstenedione (Figure 4b) (Grimm *et al.*, 1994). Thus, we have determined whether it too can be accommodated in the binding site in a manner consistent with its specific product formation. Since, as noted, the accepted mechanism of cytochrome P450-catalyzed carbon–hydrogen bond hydroxylation is hydrogen-atom abstraction by the ferryl oxygen followed by rapid OH radical rebound to the substrate, the question addressed was whether the optimized substrate–enzyme complexes could be found with the 16β H atom closest to the ferryl oxygen. To this end, the structure of androstenedione was constructed, energy minimized with Quanta/Charmm and used for docking into the cytochrome P450 2B4 substrate binding site. Two possible orientations of androstenedione were found that could lead to 16β-hydroxylation. The two cytochrome P450 2B4–androstenedione complexes were then energy optimized with AMBER 4.1. Figure 6a and b show the resulting optimized orientations of androstenedione. In both orientations, the $C_{16\beta}$ hydrogen is closer to the ferryl oxygen than any other atoms in androstenedione. In both orientations, the A ring of androstenedione is located in the common region of the two benzphetamine orientations. These two androstenedione orientations differ by an overall rotation such that the two methyl groups in androstenedione point in different directions in these two orientations. In the first orientation, the 16β hydrogen to ferryl oxygen distance [$r(H_\beta–O)$] is about 2.9 Å, shorter than the 4.0 Å distance between 16α and the ferryl oxygen [$r(H_\alpha–O)$].

In the second orientation, these distances are 2.1 Å [$r(H_\beta–O)$] and 3.2 Å [$r(H_\alpha–O)$]. The results imply that 16β-hydroxylation is favored over 16α-hydroxylation, in agreement with experimental data, and providing further support for the reliability of the model.

In addition to validation by substrate docking, determining the effect of mutation of key residues in the proposed binding site is another important way to assess the reliability of the model. Mutagenesis studies of cytochrome P450 2B4 to identify residues important for substrate binding have not been reported. However, a few studies have been performed for the highly homologous 2B1 (Aoyama *et al.*, 1989; He *et al.*, 1992; Halpert and He, 1993; Laethem *et al.*, 1994) and 2B11 isozymes (Hasler *et al.*, 1994), which share about 76% sequence identity with P450 2B4, that can be used to assess further the reliability of our model substrate binding site. These studies revealed the importance of I114, F206, I290, T302, V363, V367 and G478 for the binding of substrate to cytochrome P450 2B1 and V114, D290 and L363 for the binding of substrates to cytochrome P450 2B11. In our current model, the residues F206, T302, I363 and V367 in cytochrome P450 2B4 are in direct contact with the substrate binding region. I114, L290 and G478 are not in direct contact, but are very close to the proposed binding site. These comparisons support the reliability of the binding site and the 3D model of cytochrome P450 2B4. In addition, our model also identifies other residues in cytochrome P450 2B4 that may be important for substrate binding and could be candidates for mutagenesis studies. These include F95, R98, V103, I107, Y111, G112, V113, F202, S294, F297, A298, F365, G366, P368 and V477.

Identification of substrate access channel

The substrate binding site in the model structure is not completely buried. By tracing the positions of aggregated water molecules from the binding site region to the surface, a possible substrate access channel was identified. This channel is indicated by the positions of bound water molecules shown in Figure 7. As shown, the substrate access channel is narrower at the surface of the protein than in the middle and lower regions near the heme since the water population near the surface is smaller. The residues that form this channel are in the A helix, F helix, β1 and β4 regions. Table III lists the residues that have direct contact with this putative substrate access channel.

Identification of surface regions that interact with redox partners

The cytochromes P450 require the input of two electrons to complete the enzymatic cycle and to transform substrates into products. The ultimate sources of the two electrons are NADH for bacterial systems and NADPH for mitochondrial and microsomal systems. The transfer of these two electrons to the cytochromes P450 is through the interaction of cytochromes P450 with three classes of proteins including $Fe_2S_2Cys_4$ iron–sulfur proteins, flavin-linked reductases and cytochrome b_5, depending on the source and cellular location of the cytochrome

Fig. 8. The three plausible surface regions of 2B4 that may bind with cytochrome b_5. Top: region 1 (in green) is on the distal side of the heme (in red). Middle: region 2 (in magenta) is located on side of the heme propionate groups. Bottom: region 3 (in yellow) is on the proximal side of the heme.

Table IV. Surface residues of cytochrome P450 2B4 in each of the three regions suggested by a geometric fit algorithm to constitute the cytochrome P450 binding site for cytochrome b_5

Region 1	D47 (A)	R48 (A)	K49 (A)	G50 (A)	L51 (A)
	F171 (E*)	Y190 (F)	K191 (F)	L199 (F)	F203 (F)
	F206 (F)	E301 (I)	R308 (I)	L470 (b4)	T471 (b4)
	P472 (β4)	R473 (β4)	E474 (β4)	S475 (β4)	N479 (β4)
	P481 (β4)	P482 (β4)			
Region 2	K100 (B′)	I101 (B′)	V104 (B′)	D105 (B′)	F108 (B′)
	Q109 (B′)	V113 (B′)	I114 (B′)		
Region 3	M132 (C)	R133 (C)	D134 (C*)	K139 (C*)	R140 (C*)
	R271 (β5)	K274 (β5)	K276 (β5)	H354 (K)	K421 (m-β)
	R422 (m-β)	N423 (m-β)	F426 (m-β)	M427 (m-β)	P428 (m-β)
	F429 (β)	K433 (β)	R434 (β)	I435 (β)	L437 (β)
	E439 (L)	G440 (L)	R443 (L)	T444 (L)	F447 (L)

Shown in parentheses is the secondary structure location of each residue. (m-β) indicates the meander to β-bulge region and (β) indicates the β-bulge region.

P450. Understanding how cytochromes P450 interact with these redox partners is an important step in revealing the mechanisms and function of the mixed function oxidases. Among the three classes of the cytochrome P450 redox partners, the interaction between cytochromes P450 and cytochrome b_5 is especially interesting owing to the ability of cytochrome b_5 to either enhance or inhibit cytochrome P450-catalyzed activities (Omata *et al.*, 1994; Gruenke *et al.*, 1995). In addition, the only available crystal structure of a cytochrome P450 redox partner is of a water-soluble bovine liver cytochrome b_5 (Argos and Mathews, 1975; Durley and Mathews, 1994). Therefore, this cytochrome b_5 structure (pdb1cyo.ent from PDB databank) was used, together with our model cytochrome P450 2B4 structure, to identify possible binding site on cytochrome P450 for cytochrome b_5.

The method used to find surface regions for binding is a geometric fit algorithm for protein docking (Katchalski-Katzir *et al.*, 1992; Vakser, 1995). This method performs an exhaustive search for protein–protein structure complementarity. For high-resolution (X-ray) structures, the procedure takes advantage of the atom-size (1.0–1.7 Å) molecular surface details. In case of low-resolution (e.g. modeled) structures, the procedure averages small structural details (up to 6.8 Å) to predict the gross features of the protein complex (Vakser, 1995). Since one of the proteins involved is a model structure (i.e. cytochrome P450 2B4), the program was run at a low resolution (6.4–6.8 Å). The 100 lowest energy positions of the smaller cytochrome b_5 protein generated from this program were analyzed and three distinct clusters of cytochrome b_5 positions were found. These three clusters were then used to identify three surface regions (shown in Figure 8 with different colors) on the model cytochrome P450 2B4 structure that could possibly interact with cytochrome b_5. The first surface region is located on the distal face of the model cytochrome P450 2B4 and it includes residues in segments of the A helix, E* helix, F helix, I helix and in the β4 region. The second surface region is located on a face perpendicular to the heme plane of cytochrome P450 2B4, and it includes residues in the B′ helix region of cytochrome P450 2B4. The third surface region is located on the proximal face of cytochrome P450 2B4. The third face found in our study had previously been suggested to be the cytochrome b_5 binding site by Stayton *et al.* (1989) from manual docking experiments between the crystal structures of cytochrome P450cam and cytochrome b_5. Our study suggests that the surface residues of cytochrome P450 2B4 in this face that may contact cytochrome b_5 include residues in the C–C* helices, β5, K helix, meander, β-bulge and L helix. The surface residues in each of the three regions that may have direct contact with cytochrome b_5 are listed in Table IV.

These results were obtained using low-resolution geometric considerations only. The contributions of high-resolution structural details and physicochemical interactions to the binding of these two proteins were not included in this preliminary examination. Experimental studies using synthetic peptides corresponding to cytochrome P450 2B1 residues 116–134 to inhibit the interaction between cytochrome 2B1 and cytochrome b_5 suggested the important role of two basic residues, Lys122 and Arg125, of cytochrome P450 2B1 in cytochrome P450–cytochrome b_5 interaction (Omata *et al.*, 1994). The corresponding two basic residues, R122 and R125, of cytochrome P450 2B4 are located in the C helix region very close to the third surface region suggested from the geometric fit. As listed in Table IV, there are a number of other basic residues in the third surface region, including R133 in the C helix, K139 and R140 in the C* helix, R271, K274, K276 in the β5 region, K421 and R422 between the meander and β-bulge region, K433 and R434 in the β-bulge and R443 in the L helix region. These basic residues may also form charge pair interactions with acidic residues in cytochrome b_5. These residues, together with others listed in Table IV, are good candidates for mutagenesis studies. Studies of the mutation of some of these residues to alanine are in progress.

Conclusion

A three-dimensional homology model structure of rabbit cytochrome P450 2B4 (LM2) enzyme was constructed. The strategies used to construct this model were previously used for building and evaluating a bacterial cytochrome P450choP isozyme. In addition, a new method of sequence alignment between templates and target proteins was developed, which is based on local sequence alignment of a group of highly homologous sequences with the SCR-aligned template sequences. The resulting initial model structure was quickly evaluated using a protein analysis program (Prosa) to find possible misfolded regions due to alignment error or bad steric contact. These regions were then refined to obtain an improved model. The structurally non-conserved regions of this improved model were then energy minimized. Structural waters were added and optimized, allowing unconstrained full protein

minimization of the structure to be performed. This optimized structure was evaluated for its overall quality, using available protein analysis programs, and found to be satisfactory. The model structure was stable at room temperature during a 140 ps unconstrained full protein molecular dynamics simulation. The 3D model of cytochrome P450 2B4 was then used to identify and characterize three important regions related to its function; a substrate access channel, substrate binding site and candidate surface contact regions with a redox partner, cytochrome b_5. The substrate binding site identified contained many residues already shown by mutation studies to affect substrate binding in homologous members of the 2B family, 2B1 and 2B11. Moreover, when two different substrates, benzphetamine and androstenedione, that are metabolized by cytochrome P450 2B4 with pronounced product specificity, were docked into the putative binding site, two orientations were found for each substrate that could lead to the observed preferred products. These results provide evidence for the reliability of the model. Additional sites for mutation were also identified for further validation. Using a geometric fit method to dock the model cytochrome P450 2B4 structure with the known cytochrome b_5 crystal structure, three regions on the surface of the model cytochrome P450 2B4 structure were suggested to be possible sites for interaction with cytochrome b_5. Residues that may interact with cytochrome b_5 have been identified, and mutagenesis studies changing some of these residues to alanine are in progress.

Acknowledgments

Support for this work by the National Institutes of Health, Grants GM35533 and GM27943, and a VA Merit Review Grant is gratefully acknowledged. I.A.V. is grateful to Professor Andrej Sali for his support and helpful discussions. We also appreciate use of the Cray C-90 at the Pittsburgh Supercomputing Center under NSF Grant MCA93S007P, where the molecular dynamics simulations were performed.

References

Aoyama,T., *et al.* (1989) *J. Biol. Chem.*, **264**, 21327–21333.
Argos,P. and Mathews,F.S. (1975) *J. Biol. Chem.*, **250**, 747–756.
Biosym Technologies (1994) *Insight II, Version 2.3.5.* Biosym Technologies, San Diego.
Boddupalli,S.S., Hasemann,C.A., Ravichandran,K.G., Lu,J.-Y., Goldsmith, E.J., Deisenhofer,J. and Peterson,J.A. (1992*) Proc. Natl Acad. Sci. USA*, **89**, 5567–5571.
Boscott,P.E. and Grant,G.H. (1994) *J. Mol. Graphics*, **12**, 185–192.
Chang,Y.T. and Loew,G.H. (1996) *Protein Engng*, **9**, 755–766.
Cupp-Vickery,J.R. and Poulos,T.L. (1994) *Proteins*, **20**, 197–201.
Cupp-Vickery,J.R. and Poulos,T.L. (1995) *Struct. Biol.*, **2**, 144–153.
Dayhoff,M.O., Barker,W.C. and Hunt,L.J. (1983) *Methods Enzymol.*, **91**, 524–545.
Durley,R.C.E. and Mathews,F.S. (1994) *Refinement and structural analysis of bovine cytochrome b5 at 1.5 Å resolution*. Deposited in PDB databank, November 1994.
Ferenczy,G.G. and Morris,G.M. (1989) *J. Mol. Graphics*, **7**, 206–211.
Gotoh,O. (1992) *J. Biol. Chem.*, **267**, 83–90.
Graham-Lorence,S., Amarneh,B., White,R.E., Peterson,J.A. and Simpson,E.R. (1995) *Protein Sci.*, **4**, 1065–1080.
Grimm,S.W., Dyroff,M.C., Philpot,R.M. and Halpert,J.R. (1994) *Mol. Pharmacol.*, **46**, 1090–1099.
Gruenke,L.D., Konopka,K., Cadieu,M. and Waskell,L. (1995) *J. Biol. Chem.*, **270**, 24707–24718.
Halpert,J.R. and He,Y.-A. (1993) *J. Biol. Chem.*, **268**, 4453–4457.
Harris,D. (1996) personal communication.
Hasemann,C.A., Ravichandran,K.G., Peterson,J.A.and Deisenhofer,J. (1994) *J. Mol. Biol.*, **236**, 1169–1185.
Hasemann,C.A., Kurumbail,R.G., Boddupalli.S.S., Peterson,J.A. and Deisenhofer,J. (1995) *Structure*, **3**, 41–62.
Hasler,J.A., Harlow,G.R., Szklarz,G.D., John,G.H., Kedzie,K.M., Burnett,V.L., He,Y.-A., Kaminsky,L.S. and Halpert,J.R. (1994) *Mol. Pharmacol.*, **46**, 338–345.
Haugen,D.A. and Coon,M.J. (1976). *J. Biol. Chem.*, **251**, 7929–7939.
He,Y.-A., Balfour,C.A., Kedzie,K.M. and Halpert,J.R. (1992) *Biochemistry* , **31**, 9220–9226.
Heinemann,F.S. and Ozols,J. (1983) *J. Biol. Chem.*, **258**, 4195–4201.
Holm,L. and Sander,C. (1991) *J. Mol. Biol.*, **218**, 183–194.
Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesem,A.A., Aflalo,C. and Vakser,I.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2195–2199.
Koymans,L.M.H., Vermeulen,N.P.E., Baarslag,A. and den Kelder,G.M.D. (1993) *J. Comput.-Aided Mol. Des.*, **7**, 281–289.
Laethem, R.M., Halpert,J.R. and Koop,D.R. (1994) *Biochim. Biophys. Acta*, **1206**, 42–48.
Laughton,C.A., Neidle,S., Zvelebil,M.J.J.M. and Sternberg,M.J.E. (1990) *Biochem. Biophys. Res. Commun.*, **171**, 1160–1174.
Laughton,C.A., Zvelebil,M.J.J.M. and Neidle,S. (1993) *J. Steroid Biochem. Mol. Biol.*, **44**, 399–407.
Lewis,D.F.V. and Moereels,H. (1992) *J. Comput.-Aided Mol. Des.*, **6**, 235–252.
Li,H. and Poulos,T.L. (1995) *Acta Crystallogr.*, **D51**, 21–32.
Luthy,R., Bowie,J.U. and Eisenberg,D. (1992) *Nature*, **356**, 83–85.
Molecular Simulations (1994) *Quanta, Version 4.0.* Molecular Simulations, Burlington, MA.
Morris,G.M. and Richards,W.G. (1991) *Biochem. Soc. Trans.*, **19**, 793–795.
Omata,Y., Sakamoto,H., Robinson,R.C., Pincus,M.R. and Friedman,F.K. (1994) *Biochem. Biophys. Res. Commun.*, **201**, 1090–1095.
Pearlman,D.A. *et al.* (1995) *AMBER 4.1*. University of California, San Francisco.
Poulos,T.L. (1991) *Methods Enzymol.*, **206**, 11–30.
Poulos,T.L., Finzel,B.C. and Howard,A. J (1986) *J. Am. Chem. Soc.*, **25**, 5314–5322.
Ravichandran,K.G., Boddupalli,S.S., Hasemann,C.A., Peterson,J.A. and Deisenhofer,J. (1993) *Science*, **261**, 731–736.
Ruan,K.H., Milfeld,K., Kulmacz,R.J. and Wu,K.K. (1994) *Protein Engng*, **7**, 1345–1351.
Ruettinger,R.T., Wen,L.-P. and Fulco,A.J. (1989) *J. Biol. Chem.*, **264**, 10987–10995.
Sippl,M.J. (1993) *J. Comput.-Aided Mol. Des.*, **7**, 473–501.
Stayton,P.S., Poulos,T.L. and Sligar,S.G. (1989) *Biochemistry*, **28**, 8201–8205.
Szklarz,G.D., Ornstein,R.L. and Halpert,J.R. (1994) *J. Biomol. Struct. Dyn.*, **12**, 61–78.
Tarr,G.E., Black,S.D., Fujita,V.S. and Coon,M.J.(1983) *Proc. Natl Acad. Sci. USA*, **80**, 6552–6556.
Uvarov,V., *et al.* (1994) *Eur. J. Biochem.*, **222**, 483–489.
Vakser,I.A. (1995) *Protein Engng*, **8**, 371–377.
Vijayakumar,S. and Salerno,J.C. (1992) *Biochim. Biophys. Acta*, **1160**, 281–286.
Vriend,G. and Sander,C. (1993) *J. Appl. Crystallogr.*, **26**, 47–60.
Zhao,D., Gilfoyle,D.J., Smith,A.T. and Loew,G.H. (1996) *Proteins: Struct. Funct. Genet.*, **26**, 204–216.
Zvelebil,M.J.J.M., Wolf,C.R. and Sternberg,M.J.E. (1991) *Protein Engng*, **4**, 271–282.

Received June 12, 1996; revised August 27, 1996; accepted August 31, 1996