Acta Cryst. (1995). D**51**, 21–32

Modeling Protein–Substrate Interactions in the Heme Domain of Cytochrome $P450_{BM-3}$

BY HUIYING LI AND THOMAS L. POULOS*

Departments of Molecular Biology and Biochemistry, and Physiology and Biophysics, University of California at Irvine, Irvine, CA 92717, *USA.*

(*Received* 17 *May* 1994; *accepted* 5 *August* 1994)

Abstract

The crystal structure of heme domain of the fatty acid monooxygenase, cytochrome $P450_{BM-3}$, consisting of residues 1–455 has been independently solved to $R = 0.18$ at 2.0 Å. The crystal form used, space group $P2_1$ with two molecules per asymmetric unit, is isomorphous with that form with residues 1–471 first described by Boddupalli *et al.* [Boddupalli, Hasemann, Ravinchandran, Lu, Goldsmith, Deisenhofer & Peterson (1992). *Proc. Natl Acad. Sci. USA*, **89**, 5567–5571] and used by Ravichandran, Boddupalli, Hasemann, Peterson & Deisenhofer [(1993). *Science*, **261**, 731–736] to determine the crystal structure. The substrate-access channel consists of a large, hydrophobic cleft that appears to be the most likely route taken by fatty acid substrates. Attempts to soak crystals in mother liquor containing a variety of fatty acid substrates yielded featureless difference Fouriers even though fatty acid substrates are known to bind with dissociation constants in the μM range. Modeling substrate–enzyme interactions reveals few contacts between the enzyme and substrate. More detailed modeling was carried out by subjecting both molecules in the asymmetric unit to extensive energy minimization. These studies reveal that the heme-domain active-site cleft can undergo a large conformational change that closes the access channel thereby providing enhanced protein–substrate interactions. These conformational changes are prevented from occuring by intermolecular contacts in the crystal lattice which lock the protein in the 'open' conformation.

Introduction

Although the fatty acid monooxygenase, cytochrome $P450_{BM-3}$, is a bacterial enzyme isolated from *Bacillus megaterium*, it more closely resembles in sequence and functional properties eukaryotic P450's than it does other bacterial P450's such as the camphor monooxygenase, $P450_{cam}$ (Ruettinger, Wen & Fulco, 1989). Like eukaryotic microsomal P450's, $P450_{BM-3}$ employs a diflavin protein (FMN + FAD) as a source of electrons while $P450_{cam}$ and other bacterial or mitochondrial P450's use a 2Fe–2S iron–sulfur protein as an electron donor. In $P450_{BM-3}$, however, the diflavin reductase and P450 are fused into a single 119 kDa polypeptide, making it a catalytically self-sufficient system requiring only O_2 and NADPH (Narhi & Fulco, 1986). The crystal structure of the $P450_{BM-3}$ heme domain has been determined (Ravichandran *et al.*, 1993) and, as will be described in this article, we have independently determined the same crystal structure with the exception that the heme domain employed in our studies consists of residues 1–455 while that of Ravichandran *et al.* (1993) comprises residues 1–471. A third P450 structure, $P450_{terp}$, which belongs to the $P450_{cam}$ family of P450's, has also been determined (Hasemann, Ravichandran, Peterson & Deisenhofer, 1994). The $P450_{terp}$ and $P450_{BM-3}$ structures, however, do not contain substrate while the best $P450_{cam}$ crystal structures contain either substrates, their analogues, or inhibitors (Poulos, Finzel & Howard, 1987; Poulos & Howard, 1987; Raag & Poulos, 1991). Since the usual procedure of soaking crystals in solutions containing substrate has not resulted in interpretable difference Fourier maps, we have turned to modeling and computational methods to gain some insights on how a fatty acid substrate interacts with $P450_{BM-3}$.

Material and methods

Protein preparation

The $P450_{BM-3}$ heme domain used in this study consists of residues 1–455. Narhi & Fulco (1987) showed that trypsin cleavage generates a heme domain consisting of residues 1–471. We chose to generate several C-terminal truncated versions of the heme domain in an attempt to improve crystal quality. Molecular modeling using $P450_{cam}$ as a guide indicated that 1–455 was about the minimal unit that could be generated without affecting the C-terminal

* To whom correspondence should be addressed.

ISSN 0907-4449

β structure. The *E. coli* expression system with the pT7 vector described previously for other $P450_{BM-3}$ products (Darwish, Li & Poulos, 1991) was employed to produce this 455-residue heme-domain protein. A new 3′-end primer was designed to place a TAG stop codon and an *Eco*RI site right after the coding position for Leu455, which was then used along with the original 5′-end primer containing a *Bam*HI site for a polymerase chain reaction (PCR) run to obtain a gene of the new heme domain. The following DNA manipulations, protein expression and purification were carried out according to the same protocol as described previously (Li, Darwish & Poulos, 1991).

The truncated heme-domain protein showed the same UV–visible absorption spectra as does the first 470-residue product (Li, Darwish & Poulos, 1991). A blue shift of the Soret peak from 417 to 390 nm was also observed in the presence of substrate indicating

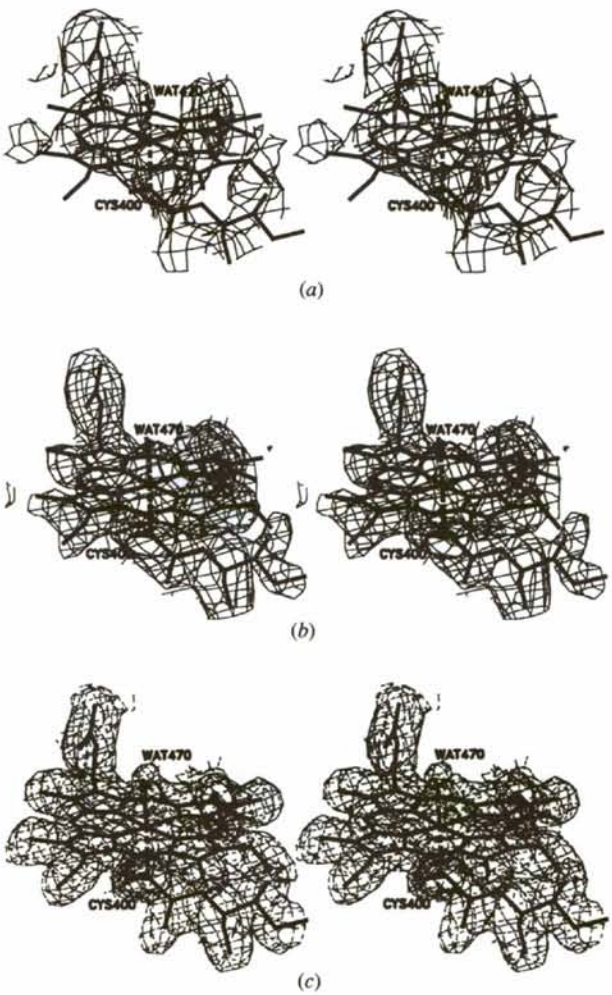

Fig. 1. Various electron-density maps (contoured at 1σ) of the heme group in the $P450_{BM-3}$ heme-domain structure. The final refined model is superimposed on (*a*) the original 3.0 Å MIR map, (*b*) the map with phases extended to 2.3 Å using *SQUASH*, and (*c*) the final 2.0 Å $2F_o - F_c$ map. This region is selected to illustrate how poorly it fits to the original MIR map and the dramatic improvement achieved with *SQUASH*. This and the remainder of the figures in this paper were generated with the programs *SETOR* (Evans, 1993) and *MOLSCRIPT* (Kraulis, 1991).

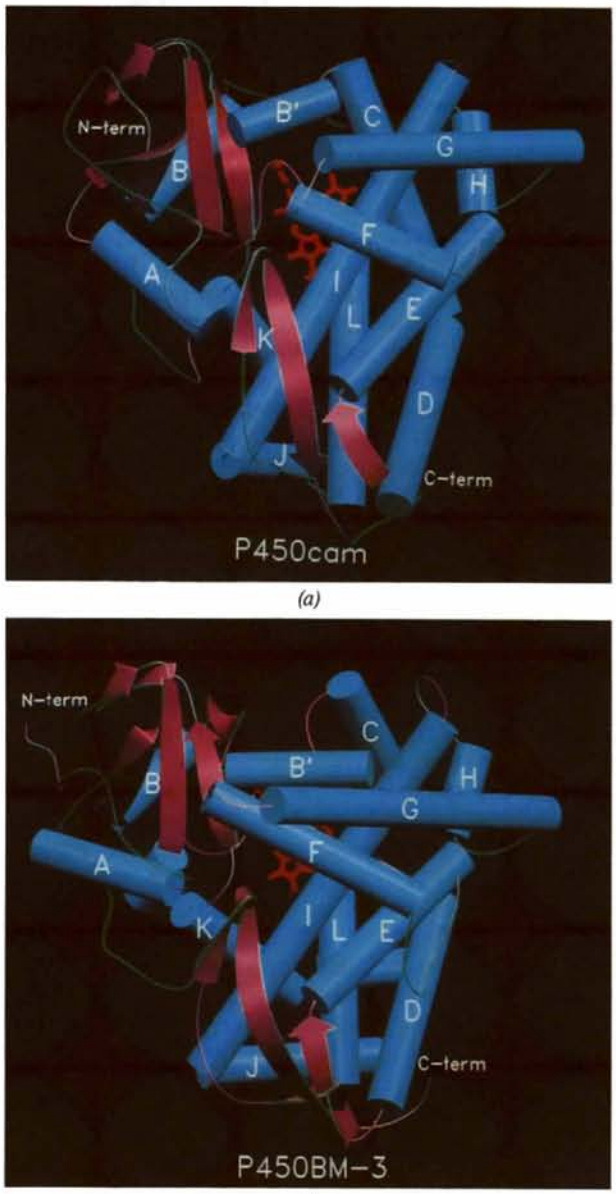

Fig. 2. Schematic diagrams of the overall topography for $P450_{cam}$ and $P450_{BM-3}$. The α-helices are represented by cylinders and the β-strands are ribbons. The C and N termini and the common helices are labeled. $P450_{BM-3}$ has more helical segments (*J′* and *K′*) which are drawn without labels. Although the two P450's share a similar basic fold, significant topological changes were found between the two structures.

Table 1. *Conditions for heavy-atom derivative preparation, statitics of data collection and phase refinements*

Derivative*	EMTS	MEHGCL1	MEHGCL2	HGCN	PTAMCL
Soaking conditions	1.3 m*M* 65 h	≈1.0 m*M*, 1 h soak in Hg-free solution ≈12 h	50 m*M* DTT ≈1.0 m*M* Hg 1 h back-soak ≈12 h	1.5 m*M* 70 h	2.0 m*M* 4 days
No. of observations	43171	84748	61534	37213	68240
No. of independent reflections	15792	25071	18012	15942	22502
R_{merge} † (%)	5.38	4.65	5.94	5.00	6.41
R_{iso} ‡ (%)	18.5	19.6	18.7	13.2	21.1
Highest resolution (Å)	3.00	2.73	3.00	3.22	2.85
Accumlated completeness (%)	59	76	65	82	68
R_{Cullis} §	0.632	0.634	0.589	0.650	0.618
Phasing power ¶	1.75	1.62	1.88	1.19	1.02
No. of phased reflections	11574	17366	11176	8871	8359
Resolution of phase refinement (Å)	3.0	2.9	3.0	4.0	4.0
No. of sites per molecule	1	1	1	1	2
Binding residue	Cys156	Cys156	Cys156	Cys156	Met118 Met185

The overall mean figure of merit for 14233 phased reflections to 3.0 Å is 0.64.

* EMTS, sodium ethylmercurithiosalicyclate; MEHGCL, methyl mercury chloride; HGCN, mercury(II) cyanide; PTAMCL, platinum(II) diamino dichloride.

† $R_{merge} = \sum_{hkl}\{\sum_i |I_i - \langle I\rangle|/\sum_i I_i\}$, where I_i is observed intensity of the ith observation. $\langle I\rangle$ is mean intensity of the *hkl* reflection.

‡ $R_{iso} = \sum_{hkl}||F_{PH}| - |F_P||/\sum_{hkl}|F_P|$, fractional isomorphous difference where F_P and F_{PH} are the structure factors of parent and derivative data set, respectively.

§ $R_{Cullis} = \sum||F_H| - (|F_{PH}| - |F_P|)|/\sum|F_H|$, where F_H is the calculated heavy-atom structure factor; F_P and F_{PH} are the structure factors of parent and derivative data, respectively.

¶ Phasing power = r.m.s. $|F_H|$/r.m.s. (lack of closure).

the binding of substrate to the 455-residue heme-domain protein.

Crystallization and data collection

Crystallization conditions were similar to those of Boddupalli *et al.* (1992) with some modification. The sitting-drop vapor-diffusion method was used with reservoir solutions containing 18% PEG 8000 plus 50 m*M* of $MgSO_4$ or 16% PEG 8000 plus 100 m*M* $MgSO_4$, both in 0.1 *M* PIPES buffer at pH 6.8. Each sitting drop was set up by mixing equal volumes of the reservoir solution and the protein solution of 40 to 50 mg ml^{-1}. The reducing agent dithiothreitol was not included. Crystal growth was initiated by streak seeding after a minimum of one day pre-equilibration at 278–280 K. Crystals exhibited two different habits with plates using the former reservoir solution and rods using latter. The crystal rods turned out to be the better form for data collection as they usually showed more uniform thickness. These crystals exhibiting slightly different habits belong to the same space group, $P2_1$, as reported by Boddupalli *et al.* (1992), with unit-cell dimensions $a = 59.5$, $b = 154.0$, $c = 62.4$ Å, $\beta = 95.0°$, and two molecules per asymmetric unit.

X-ray diffraction data was collected on a Siemens X-1000 area detector equipped with a rotating Cu-anode X-ray generator and focusing optics. Data collection was monitored by the program *FRAMBO* (SAXII Area Detector Software Release Notes for Version 1.21, Siemens Analytical X-ray Instruments, Inc., 1990) and data reduction was carried out with *XENGEN* (Howard *et al.*, 1987) to obtain the integrated and scaled structure factors.

Heavy-atom derivatives

About two dozen heavy-metal reagents were screened and four proved to be good candidates for isomorphous derivatives. The conditions for the preparation of heavy-atom derivatives as well as data-collection statistics are summarized in Table 1. Slow, stepwise addition of heavy-metal reagents to the crystal soaking solution was crucial to avoid crystal cracking or sudden decay in the X-ray beam. For the mercury derivatives, back-soaking of crystals in metal-free solutions was successful in prolonging the crystal lifetime in the X-ray exposure and reducing the background in the difference Patterson maps. Derivative data screening was monitored directly using the area detector. Approximately 100–200 frames of data were collected and processed and the isomorphous difference between parent and derivative data sets computed. The mean fractional isomorphous difference in the range of 0.10 to 0.25 and retention of isomorphous cell dimensions were considered acceptable minimum conditions for the data collection.

The metal-binding sites were interpreted from difference Patterson syntheses using the *PHASES* package (Furey, 1990). The Patterson prediction routine *PATPRED* and map-contour facility in the *XTALVIEW* package (McRee, 1992) proved very

handy in locating and confirming the self and cross Patterson vectors. Although the three mercury data sets share the common binding sites, including all three did show better statistics than using any one derivative on its own.

MIR phase refinement and initial model building

The common Hg-atom positions extracted from EMTS and MEHGCL derivative data sets were refined using the *PHASES* package with anomalous dispersion data for the MEHGCL2 derivative included. The platinum-binding sites were identified with difference Fourier syntheses using the mercury phase set. In Table 1 are also provided the heavy-atom refinement and phase-calculation statistics. The F_o electron-density map calculated with the solvent-leveled MIR phases at 3.0 Å was clear enough to reveal the molecular boundary and major secondary-structural features. Using $P450_{cam}$ as a guide, helices *D*, *E*, *I* and *L* were located. The model of $P450_{cam}$ with side chains removed was used to aid in tracing the remainder of the polypeptide chain using *FRODO* (Jones, 1978).

The electron density of the heme group in the initial MIR map was rather poor (Fig. 1). The heme group was positioned based on the $P450_{cam}$ model as well as the *I* and *L* helical segments, which were reasonably clear. The non-crystallogrtaphic symmetry-related molecule was located by building in a few helices and then aligning these with the $P450_{cam}$ polyalanine model using the Hg-binding sites as reference points. The non-crystallographic twofold lies approximately along the crystallographic *c* axis and the program package written by J. K. M. Rao (Jaskolski, Miller, Rao, Leis & Wlodawer, 1990) was used for electron-density map averaging. However, the averaged electron density did not show significant improvement probably due to a poorly defined protein envelope.

Crystallographic refinement and phase extension

Extensive rigid-body, positional, and/or simulated-annealing (slowcool protocol) refinements with *X-PLOR* (Brünger, 1992) were carried out interspersed with frequent model building with *FRODO* (Jones, 1978) or *TOM* (Cambillau & Horjales, 1987) using electron-density maps calculated with combined F_c and MIR phases. In addition, *SIGMAA* (Read, 1986) in the *CCP*4 program suite (Evans, 1991) was used for phase combination in order to reduce model bias.

The process of model building was fairly difficult using the 3.0 Å solvent-leveled MIR, or averaged and model phase-combined electron-density maps. An attempt at refining the heavy-atom derivative MIR phases against the partial model F_c phase using the *MLPHARE* routine in *CCP*4 resulted in an MIR map of better quality and aided in the location of a fourth Pt site. The map fitting was still not that straightforward until a phase extension to 2.3 Å was implemented using *SQUASH* (Zhang & Main, 1990*b*; Cowtan & Main, 1993). (A modified version of *SQUASH*, written by D. Schuller in our laboratory, was used.) A molecular envelope was carefully constructed from a partially refined model having 440 amino-acid residues. The non-crystallographic symmetry (NCS) was refined using the program *RAVE* (Jones, 1992) followed by 30 cycles of phase extension to 2.3 Å, imposing simultaneously the constraints of solvent leveling, histogram matching (Zhang & Main, 1990*a*), and NCS map averaging in each cycle. The new 2.3 Å electron-density map provided a remarkable improvement in the quality of the electron-density map (Fig. 1). Regions which were essentially invisible in any of the lower resolution maps (MIR, solvent-leveled, or averaged) were now clear. Approximately 90% of the side chains could be located easily. After refitting to the new map, the crystallographic *R* factor dropped from 0.30 to 0.25 for data between 10.0 and 2.5 Å.

The N-terminal region and the loop between helices *F* and *G* were the last two trouble spots in model building owing to the high mobility and different conformations adopted in the two molecules in the asymmetric unit, which made the averaged map almost featureless in these regions. The 3.0 Å MIR map was used to interpret these regions until the refinement was nearly complete. Individual temperature factors of all non-H atoms were refined with *X-PLOR*. Water molecules were included if the $F_o - F_c$ electron density was at least 3σ and if the water molecule was within hydrogen-bonding distance of another solvent or protein atom. Although an automatic procedure was used for locating solvent molecules, these were confirmed by visual inspection. The final refined model has an *R* factor of 18.4% and includes residues 1–455 in both molecules in the asymmetric unit and 390 water molecules. In Table 2 is provided the final refinement statistics.

Pseudo twofold symmetry

The orientation of the non-crystallographic twofold rotation being roughly parallel to the unit-cell *c* axis was confirmed by locating molecule *B* with a partially refined search model of molecule *A* using the molecular-replacement package *MERLOT* (Fitzgerald, 1988). The pseudo twofold is sandwiched by the two aromatic rings of Tyr166 from each of the two molecules. Close contacts between the two molecules make a symmetrical pattern of intermolecular interactions due to the pseudo symmetry. Between Val120 and Pro170 of both mol-

Table 2. *Statistics of data collection and refinement of cytochrome P450$_{BM-3}$ heme domain (residues 1–455)*

Cell parameters (Å, °)	$a = 59.5$, $b = 154.0$, $c = 62.4$, $\beta = 95.0$
Space group	$P2_1$
No. of observations	277100
No. of independent reflections	57597
R_{merge} (%)	7.9
Data completeness (%)	
48.0–3.9 Å	99
3.9–3.2 Å	94
3.2–2.8 Å	76
2.8–2.3 Å	55
2.3–2.0 Å	25
Number of reflections used in refinement with $F > 2\sigma(F)$	41253
Resolution range of refinement (Å)	10.0–2.0
R ($= \sum\|\|F_o\| - \|F_c\|\|/\sum\|F_o\|$)	18.4%
No. of protein atoms	7428
No. of water molecules	390
R.m.s. in bond lengths (Å)	0.008
R.m.s. in bond angles (°)	1.54
Estimated coordinate error (Å)	
SIGMAA plot	0.39
Luzzati plot	0.25

Table 3. *Hydrogen-bonding interactions in the pseudo twofold crystal-contact region*

Donor atom*		Molecule	Acceptor atom*		Molecule	Distance (Å)
NE	R132	*A*	OH	Y166	*B*	2.85
NH2	R132	*A*	OH	Y166	*B*	2.82
NH1	R134	*B*	OH	Y166	*A*	2.89
ND2	N134	*A*	O	R161	*B*	2.85
ND2	N134	*B*	O	R161	*A*	2.79
N	R161	*A*	OD1	N134	*B*	3.18
N	R161	*B*	OD1	N134	*A*	2.91
ND2	N163	*A*	O	R132	*B*	2.79
ND2	N163	*B*	O	R132	*A*	3.09
N	D168	*A*	OD1	D168	*B*	2.92
N	D168	*B*	OD1	D168	*A*	2.90
OD1	D168	*A*	OE1	Q169	*B*	2.65
OD2	D168	*A*	OE1	Q169	*B*	3.17
N	Q169	*A*	OD1	D168	*B*	3.10

Acceptor atom* Molecule *A*		Distance (Å)	Water	Distance (Å)	Acceptor atom* Molecule *B*	
O	Q125	2.75	O 485	2.67	OH	Y166
OH	Y166	2.65	O 658	2.75	O	Q125
NE2	Q128	2.70	O 490	3.05	NH1	R132
NH2	R132	2.67	O 649	2.81	OE1	Q128
ND2	N163	2.79	O 491	2.79	O	E137
OE2	E137	2.89	O 774	3.04	ND2	N163
O	R167	2.61	O 500	3.07	O	Y166
OD1	N134	2.70	O 564	3.11	O	N159
OD1	D136	2.92	O 656	2.94	OH	Y160
NH1	R132	2.78	O 830	2.78	OD1	D121
OD1	D121	3.18	O 831	2.88	NE	R132
O	D121	2.66	O 588	3.15	NH2	R132

* Atom labels used in the standard PDB coordinate file.

ecules is an extensive hydrogen-bonding network involving a few dozen water molecules. Seven residues form direct protein–protein hydrogen bonds while 12 water molecules mediate hydrogen bonds between the two molecules (Table 3).

Energy minimization of the crystallographic model

Energy-minimization calculations were carried out with *DISCOVER* (version 2.9) from Biosym Technologies. The force-field parameters were taken from a standard library provided by Biosym (consistent valence force-field) with the exception of the heme and Cys thiolate ligand parameters. Charges for the heme group and non-bonded parameters for the iron were taken from Paulsen & Ornstein (1991) while the parameters for the S—Fe bond, angles and force constants were taken from Collins, Camper & Loew (1991). Each molecule in the asymmetric unit, molecule *A* or molecule *B*, was independently minimized. Molecule *A* consisted of 7983 atoms which included 193 water molecules identified in the crystal structure. Molecule *B* consisted of 7962 atoms which included 187 water molecules. Non-bonded interactions were set to 0 at a cutoff distance of 17 Å with a switching function width of 2.0 Å. No cross terms were used and the dielectric was set to 1.0. Since only those water molecules found in the crystal structure were included, charged amino acids were made net neutral to compensate for the effect of bulk solvent. The following protocol was used for all energy-minimization runs. (1) All H atoms were allowed to move for 1000 steps. (2) All water molecules were allowed to move for 1000 steps. (3) All side-chain atoms plus waters were allowed to move for 2000 steps. In these three steps, the method of steepest descents was used. (4) In the last step, all atoms were allowed to move for 5000 steps using the method of conjugate gradients. If needed, the minimization was continued using the method of conjugate gradients until the average change in energy with respect to atomic position was less than 0.004 kcal Å^{-1}.

A model of myristic acid was docked to the minimized molecule *B* structure manually by manipulating only the substrate. It was assumed that C13 ($\omega - 1$ position), which is preferentially hydroxylated (Miura & Fulco, 1975; Matson, Hare & Fulco, 1977), will lie closest to the Fe atom. This is a rather 'soft' restraint, however, since the $\omega - 2$ and $\omega - 3$ positions are also hydroxylated (Miura & Fulco, 1975). P450$_{cam}$ was used as a guide by superimposing the C13 atom of myristic acid onto the C5 atom of camphor, the site of camphor hydroxylation (Gelb, Heimbrook, Malkonen & Sligar, 1982). Arg47 was found to be situated close to the carboxylate group of the substrate so the substrate was adjusted to maximize this interaction. After manual optimization, the model was energy minimized by first allowing the substrate alone to move followed by an all-atom minimization.

Results and discussion

Overall topography

Since Ravichandran *et al.* (1993) have described the P450$_{BM-3}$ heme-domain structure, we confine the

present discussion to those structural features relevant to substrate binding. As expected, all the secondary-structural features documented by Ravichandran *et al.* (1993) are observed in this structure. The heme-domain structure described therein consists of residues 1–471 and hence is 16 residues longer than the structure we have solved (residues 1–455). This is probably not important to the overall fold and properties of the heme domain since the last 14 residues in the 471-residue structure described by Ravichandran *et al.* (1993) are not visible in their electron-density maps.

When viewed perpendicular to the heme plane, the overall topography of the heme domain resembles the triangular prism of $P450_{cam}$ with an α-helix-rich domain occupying two of the three corners and a β-sheet-rich domain occupying the third (Fig. 2). One of the major differences is that the *F* and *G* helices and the loop in between protrude out of the triangular face, leaving a large cleft on one corner of the molecule. On the other side of the cleft is the $\beta 1$ pair (according to the nomenclature of secondary structure used in $P450_{cam}$) and a small stretch of 3_{10} helix (residues 16–20) flanked towards the N-terminus by a β-turn.

Substrate-access channel

This cleft defines a large opening that probably serves as an entry point for substrates. The hair-pin region of the $\beta 1$ pair in $P450_{BM-3}$ is oriented differently from that in $P450_{cam}$, stretching out and more parallel to the $\beta 3$ pair thereby allowing the formation of a five-stranded β-sheet including the loop between the *B* and *B'* helices. This is a more extensive β structure than is found in $P450_{cam}$ and the new orientation of the $\beta 1$ pair in $P450_{BM-3}$ contributes to the wider opening of the substrate-access channel. The space-filled models of this region in Fig. 3 provide a sharp comparison of $P450_{cam}$ and $P450_{BM-3}$. In $P450_{cam}$, the substrate-access channel is closed while in $P450_{BM-3}$ it is open. This region of the structure exhibits higher thermal motion in the substrate-free form of $P450_{cam}$ (Poulos, Finzel & Howard, 1986) suggesting that dynamical fluctuations are important for allowing the access channel to open and camphor to enter. A recent study by Raag, Li, Jones & Poulos (1993) on the structure of a large inhibitor bound to $P450_{cam}$ shows that this access channel can adopt alternate conformations, which is consistent with the idea that fluctuations are important to substrate binding.

The access channel in $P450_{BM-3}$ also exhibits high thermal motion (Fig. 4), especially the loop connecting the *F* and *G* helices which is partially disordered in the crystal structure. It is noteworthy that the loop between the *F* and *G* helices is also disordered in the crystal structure of cytochrome $P450_{terp}$ (Hasemann, Ravichandran, Peterson & Deisenhofer, 1994). This region of P450 consisting of the loop between *F* and *G* helices, the $\beta 1$ pair, and a stretch of peptide close the the N-terminus is likely to be an important structural feature for substrate access in many P450's.

Errors in atomic coordinates

Since the energy minimization and modeling studies to be described later argue for rather large conformational changes, it is important to first assess how well we know the position of those regions involved in the conformational switch. The overall error in atomic coordinates was estimated from a Luzzati plot (Luzzati, 1952) to be 0.25 Å while from a *SIGMAA* plot (Read, 1986), the error is 0.39 Å. As a further check, crystallographic refinement was carried out with those regions of interest excluded from all calculations. The omit electron-density maps shown in Fig. 5 illustrate that the electron density for all regions, with the exception of the *F/G* loop, is

Fig. 3. Space-filled diagrams of (*a*) $P450_{cam}$ and (*b*) molecule *B* of $P450_{BM-3}$, viewed along their respective substrate-access channels. The heme groups are shaded. Although this is a comparison between substrate-bound ($P450_{cam}$) and substrate-free ($P450_{BM-3}$) enzymes, there is no significant change in structure of the substrate-free form of $P450_{cam}$ (Poulos, Finzel & Howard, 1986). Panel (*c*) shows the same region of $P450_{BM-3}$ (molecule *B*) after the energy minimization. The heme is deeply buried due to the 'closing down' of the substrate-access channel.

well defined. Hence, major motions observed in the modeling and energy-minimization studies are not due to errors in the incorrect positioning of the models of these regions in the electron-density map. Finally, we have compared our structure with that solved by Ravichandran *et al.* (1993). The overall r.m.s. deviation for Cα atoms in molecule *A* is 0.76 Å and for molecule *B* it is 0.59 Å. However, if the loop connecting the *F* and *G* helices, residues 191–199, is excluded, the r.m.s. deviations for molecules *A* and *B* decreases to 0.22 and 0.24 Å, respectively. Residues 191–199 represent the most uncertain region of the structure which is reflected by high temperature factors (Fig. 4).

Conformational flexibility of the substrate-access channel

The conformational flexibility around the substrate-access channel in the heme domain of $P450_{BM-3}$ noted by Ravichandran *et al.* (1993) is observed in this structure. Two independent molecules in the crystal lattice adopt slightly different conformations in the substrate-access region. Nevertheless, in both molecules *A* and *B* the access channel is open. Modeling myristic acid, a 14-carbon fatty acid substrate, indicates that there is too much space in the pocket with few close protein–substrate interactions. This raises the possibility that the active-site cleft undergoes considerable motion upon substrate binding. To further model protein–substrate interactions, molecules *A* and *B* were energy minimized without substrate giving some unexpected results.

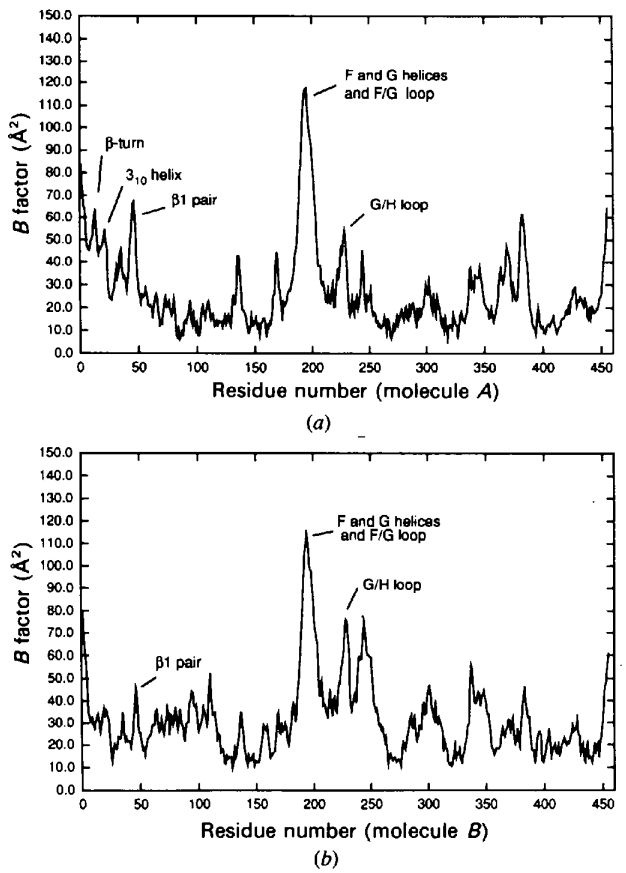

Fig. 4. A plot of the averaged crystallographic temperature factor (*B* factor) for main-chain atoms *versus* residue number for the two molecules of $P450_{BM-3}$. Those regions especially important for substrate entry with high *B* factors are labeled.

The r.m.s. differences of the main-chain atoms for each molecule before and after the energy minimization are shown in Fig. 6. Both molecules, and especially molecule *B*, undergo very large changes. As a control, we subjected the refined $P450_{cam}$ crystal structure to exactly the same minimization protocol. As shown in Fig. 7, only the loop connecting the *F* and *G* helices and another surface loop (around residue 343) that makes crystal contacts, undergo substantial motion in $P450_{cam}$. In $P450_{BM-3}$, the largest shifts (3.5–7.5 Å) are localized on the two sides of the active-site cleft. One side of the cleft that moves includes the *F* and *G* helices and the loop connecting these helices. The other side of the cleft that moves consists of an N-terminal turn (residues 9–14), a 3_{10} helix (residues 16–20), and the β1 pair, as illustrated in Fig. 8. The entire motion can be described as two different rotations. One is the rotation of *F* and *G* helices about the *H* helical axis. The second one is a rotation about the *B* helical axis of the polypeptide segment consisting of residues 9–23 and the β1 pair. The two rotations lead to an open–close motion of the substrate-access channel. The final conformation of molecule *B* resulting from the energy minimization represents the fully closed situation, which is depicted in the space-filled diagram in Fig. 3(*c*).

Molecule *A* exhibits a similar open–close motion after the energy minimization but to a lesser extent (Fig. 6). The crystal-packing constraints can explain the difference. The N-terminal stretch of the polypeptide in molecule *A* is free of any close crystal contacts which allows it to adopt a more closed conformation in the crystal structure. Molecule *B*, however, has extensive intermolecular contacts along the stretch consisting of residues 9–24, indicated by lower *B* factors in this region in molecule *B* compared to molecule *A* (Fig. 4). These crystal-packing contacts in molecule *B* prevent the N-terminal side of the cleft from closing in toward helices *F* and *G*. A comparison of the r.m.s. difference between molecules *A* and *B* before and after minimization is shown in Fig. 9. Although the two molecules are very similar in the crystal structure, the N-terminal region does adopt slightly different positions in the crystal structure. The minimization for each mol-

ecule, therefore, starts from a different position and clearly ends up with a more randomly distributed r.m.s. deviation between the two molecules than the starting models. The increased r.m.s. deviation may be interpreted in one of the two ways. First, there are a number of isoenergetic states available to the protein, especially for surface regions, and the final energy-minimized structure depends on the starting model. Second, the r.m.s. difference between molecules *A* and *B* after minimizing reflects errors in the technique. There are undoubtedly contributions from both interpretations.

Nevertheless, even if we consider the r.m.s. deviations between the two molecules after minimization as the 'noise' level, there are still significant motions in the substrate-access channel. Most importantly, the direction of motion is the same in both, namely, toward the closed conformation. The N-terminal side of the cleft in the two molecules adopts a very similar conformation after minimization, showing declined r.m.s. deviations (Fig. 9*b*). Moreover, in both minimized molecules, the 3_{10} helix (residues 16–20) and its flanking residues make direct van der Waals or hydrogen-bonding contacts with residues of the *F* helix. Most of these interactions are either weak or absent in the crystal structure. Listed in Table 4 are some representative interatomic distances showing the reduced conformational difference between two molecules resulting from the minimization. The large remaining r.m.s. deviation in the *F/G* loop region is mainly due to the different twists of the loop in the two minimized models. The loop itself does not make direct contact with the N-terminal side of the cleft. It should also be noted that the *F/G* loop is the most uncertain region in the crystal structure and has the most poorly defined electron density.

The role of crystal contacts

In order to test if the crystal contacts are indeed the reason for molecule *B* adopting a more open

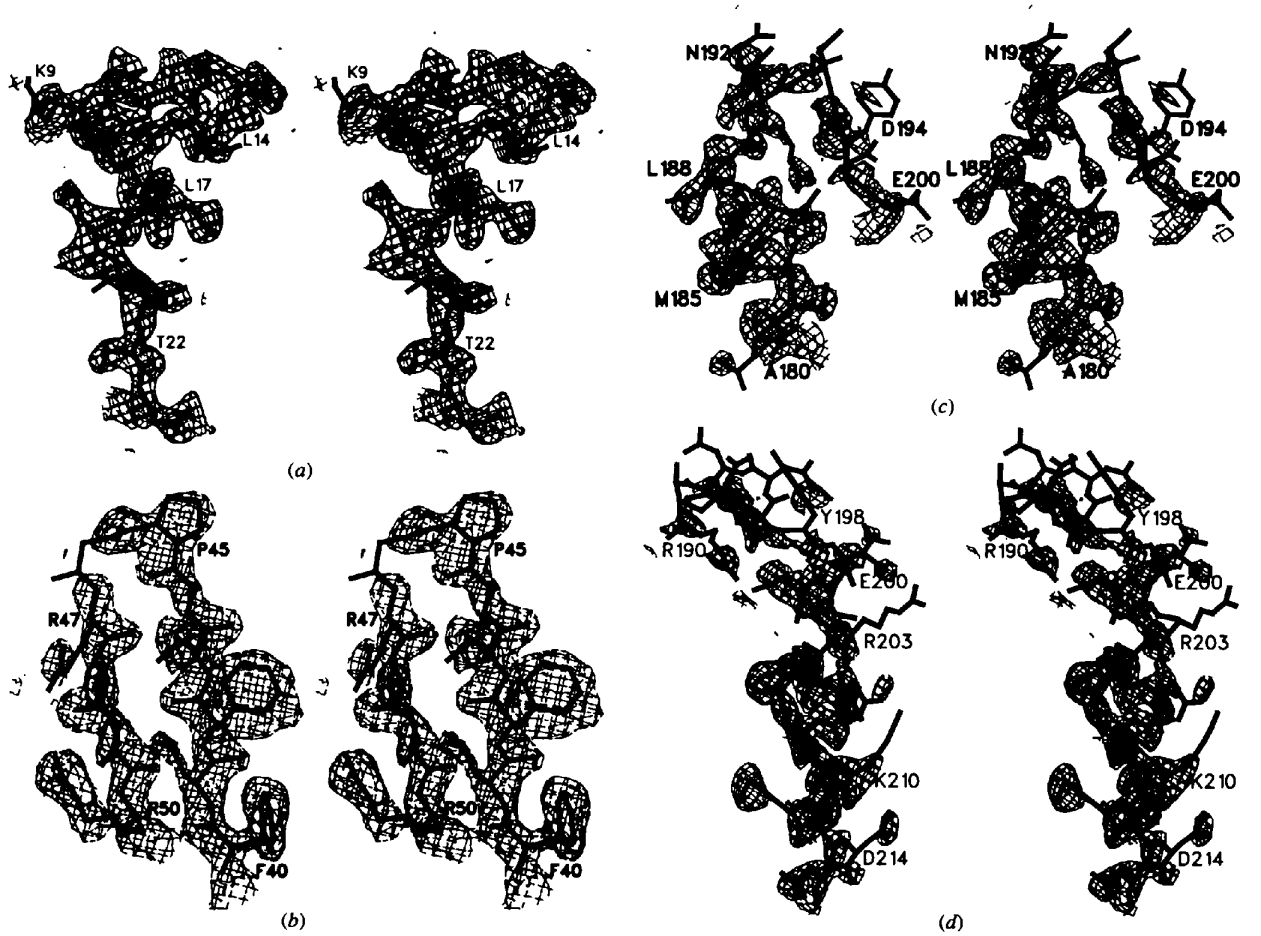

Fig. 5. The $2F_o - F_c$ omit maps (contoured at 1σ) calculated with *X-PLOR* by running 100 cycles of positional refinement with the appropriate residues excluded. (*a*) β-turn and 3_{10} helix region with residues 9–23 omitted. (*b*) β-1 pair with residues 40–50 omitted. (*c*) C-terminal portion of *F* helix and the *F/G* loop region with residues 180–200 omitted. (*d*) *F/G* loop and N-terminal portion of *G* helix with residues 190–214 omitted.

conformation, an energy minimization was carried out as with the other runs except those residues involved in crystal contacts were fixed. As shown in Fig. 10, the large motion in the N-terminal region observed in Fig. 6(*b*) no longer occurs. The *F* and *G* helices still move, but to a lesser extent. A single 2.9 Å intermolecular hydrogen bond, which was not modeled during the minimization, was found between the side chains of Q189 (molecule *B*) and K349 (molecule *A*) with no counterpart in the *F* helix of molecule *A*. Fixing residue 189 would surely prevent the *F* helix from moving but might over weight the constraint a single hydrogen bond can impose on the *F* helix in the crystal lattice. We chose, therefore, not to put any constraints on the *F* and *G* helices in the minimization except the small loop (224–226) after the *G* helix that makes some weak van der Waals interactions to the adjacent molecule. Fixing these residues did not keep *F* and *G* helices from moving. That means the crystal contacts at the C-terminal end of *G* helix, no matter how weak or strong, have only limited inpact on the motion of *F* and *G* helices. The crystal contacts observed in the *F*/*G*-loop region in our current model are quite weak. However, the possibility that the loop region does not make some strong intermolecular contacts in the lattice cannot be totally ruled out because of the high thermal motion of the loop in the crystal structure. Cryocrystallographic studies might be able to provide a 'cooler' structure which would aid in the unambiguous crystal-packing analysis. Nevertheless, results of the energy minimization and the crystal-packing simulation suggest that it is crystal packing that holds molecule *B*, and to a lesser extent molecule *A*, in the open conformation around the substrate-access channel. This can explain, in part, why we have been unsuccessful in getting fatty acid substrates to bind to heme-domain crystals. Soaking crystals or co-crystallization with μM levels of substrates provided no residual density in the difference Fourier maps, while substrates in the m*M* range would either cause crystal cracking or produce different forms which are too small to be of any use.

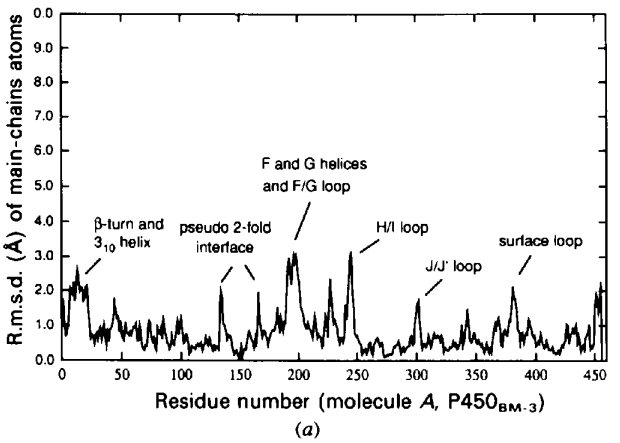

Fig. 6. The root-mean-square deviation (r.m.s.d.) of main-chain atoms for each molecule of $P450_{BM-3}$ before and after the energy minimization calculated with *X-PLOR*. Regions that experience large shifts are labeled.

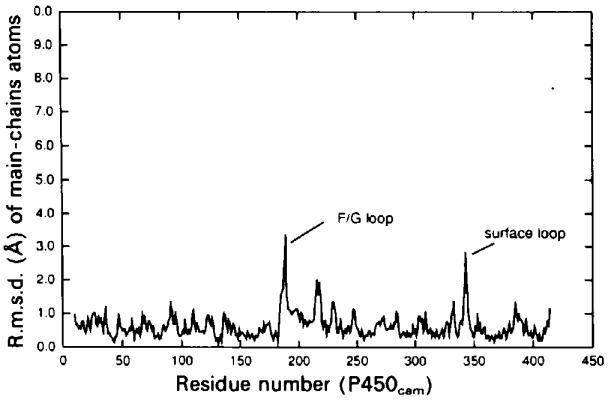

Fig. 7. The r.m.s. deviation of main-chain atoms of $P450_{cam}$ before and after the energy minimization. Only two regions that are labeled undergo significant changes.

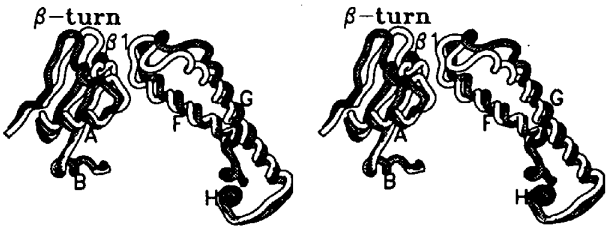

Fig. 8. A superimposition of the region in molecule *B* undergoing the conformational change before (gray) and after (clear) energy minimization. The helices, β-1, and a β-turn involved in the motion are labeled. Note that the open–close motion of the substrate-access cleft is accomplished by rotations about the *B* and *H* helical axes of the N-terminus and the *F* and *G* helix sides of the cleft, respectively.

Another observation relevant to the substrate binding and crystal contacts is the crystallization behaviour of a deletion mutant of the heme domain in which the first nine N-terminal residues were removed (Li & Poulos, unpublished results). Under the same crystallization conditions as that used for wild type, the substrate-free deletion product failed to produce crystals. However, in the presence of the fatty acid substrate, palmitic acid, a different, hexagonal crystal form was obtained that did not diffract well. The N-terminal nine-residue deletion product of the intact $P450_{BM-3}$ was observed as one of the trypsin cleavage products by Narhi & Fulco (1987), which showed poor substrate affinity. The deletion mutant of the heme domain behaves the same as its counterpart of the intact $P450_{BM-3}$. Lys9, which is missing in the deletion mutant, participates in the open–close conformational change of the proposed substrate-access cleft, although the side chain is not located inside the cleft (Fig. 5*a* and Fig. 8). Interestingly enough, the side chain of Lys9 in molecule *B* forms a 2.8 Å hydrogen bond with the carbonyl of Lys76 in the adjacent molecule *A*, becoming one of the residues in the N-terminal region that make crystal contacts which hold molecule *B* in the open conformation. Removal of this crystal contact in the deletion mutant is very likely a key factor in why the deletion mutant failed to crystallize in the monoclinic space group and instead, crystallized in a hexagonal space group.

Table 4. *Changes of some interatomic distances in substrate-access channel*

Interatomic pair*	Distance before minimization (Å)	Distance after minimization (Å)
Molecule *A*		
NE2 Q189···O L20	2.95	2.89
NE2 Q189···O T22	2.97	2.95
CB A191···CD2 L14	7.27	4.86
Molecule *B*		
NE2 Q189···O L20	5.13	2.96
NE2 Q189···O T22	4.90	2.97
CB 191···CD2 L14	12.86	4.12

* Atom labels used in the standard PDB coordinate file.

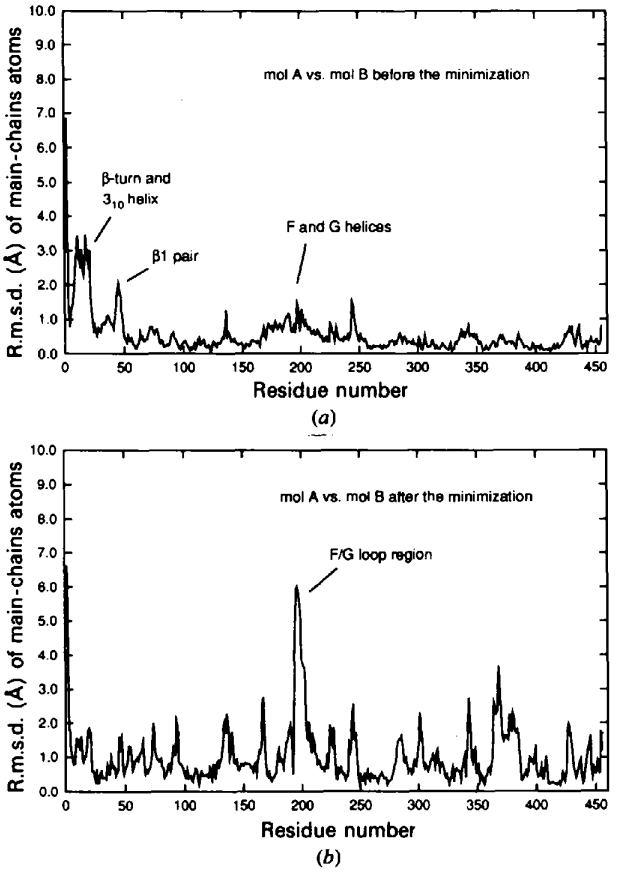

Fig. 9. The r.m.s.d. of main-chain atoms between molecules *A* and *B* of $P450_{BM-3}$ (*a*) before and (*b*) after the energy minimization. The non-crystallographic symmetry transformation was calculated with *X-PLOR* based on the coordinates of molecules *A* and *B* before and after the minimization independently.

Substrate-docking model

Even in the energy-minimized closed conformation of molecule *B*, a hydrophobic substrate-binding pocket is roomy enough to accommodate fatty acids of 12 to 18 C atoms without much readjustment of side chains. The docking model of myristic acid reveals some basic features that would be expected for the substrate–protein interaction in $P450_{BM-3}$. The substrates of $P450_{BM-3}$, long-chain fatty acids, alcohols and amides, saturated or partially unsaturated, share similar structural features. These all have a polar end of either carboxyl, hydroxyl, or amido groups, and a non-polar end (ω-end) of

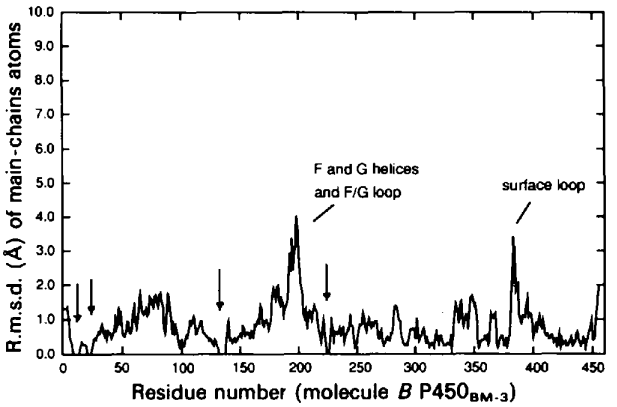

Fig. 10. The r.m.s. deviation of main-chain atoms of molecule *B* of $P450_{BM-3}$, resulted from the energy minimization with the following residues fixed: 12–14, 21–23, 162–167 and 224–226. These fixed regions are marked with arrows.

methyl or methylene groups. Most substrates are hydroxylated by $P450_{BM-3}$ at $\omega-1$, $\omega-2$ and $\omega-3$ positions with certain product distributions depending on the length of the carbon chain and the position of double bond(s) (Miura & Fulco, 1975; Matson, Hare & Fulco, 1977; Shirane, Sui, Peterson & Ortiz de Montellano, 1993).

Based on these considerations, myristic acid was docked into the binding cleft of the minimized structure with the methyl end close to the heme active center and the carboxyl group within hydrogen-bonding distance of the R47 side chain (Fig. 11). The substrate complex of $P450_{cam}$ also was used as a guide by assuming that the $\omega-1$ C atom of myristic acid, which is preferrentially hydroxylated (Miura & Fulco, 1975), will adopt a position similar to the main site of hydroxylation on camphor, C5. Resulting protein–substrate interactions were found to consist of three groups. The first group includes L20, P25, L29, R47 and Y51 that interact with the polar end of the substrate. Of particular interest are residues R47 and Y51, both having the potential of forming hydrogen bonds with substrates. The charge pairing between R47 and the carboxyl group of fatty acids is a further stabilizing factor. The second group is residues S72, A74, P329, A330 and M354. These residues make only weak van der Waals contacts with the hydrophobic middle portion of the fatty acid chain. The roomy and hydrophobic environment allows substrates to undergo quite flexible conformational changes in the middle section of the carbon chain in order to optimize contacts with the protein on both ends of the substrate molecule. The third group consists of residues F87, I263, T268, A328 and L437 which make possible contacts with the ω-end of substrates. One reason why hyroxylation reactions never occurs at the ω-position (Miura & Fulco, 1975; Shirane *et al.*, 1993) is that the phenyl ring of F87 partially blocks the substrate access to the heme iron center (Ravichandran *et al.*, 1993) so that the methyl end of the myristic acid model must 'curl up'. This places the C14 atom close to I263 and L437 while leaving C13 and C12 ($\omega-1$ and $\omega-2$) close the expected position of an iron-linked oxo atom that ultimately would insert into a substrate C—H bond. The overall picture that emerges from the docking model is that the extreme ends of the substrate, the polar 'head' and methyl 'tail', are rigidly fixed while the middle section of the fatty acid chain is buried in a hydrophobic cleft but remains relatively flexible.

Summary

The main conclusion from this study is that the heme domain of $P450_{BM-3}$ is very likely to be able to undergo a large conformational change leading to a closing of the substrate-access channel when substrates bind. Intermolecular contacts within the crystalline lattice hold the protein in the 'open' conformation which explains why we have been unable to bind fatty acid substrates in the crystals. When these contacts are removed, energy minimization of the structure results in a 'closing' of the access channel. Subsequent modeling of a fatty acid substrate binding to the active site results in a better set of protein–substrate interactions in the minimized structure than in the crystal structure.

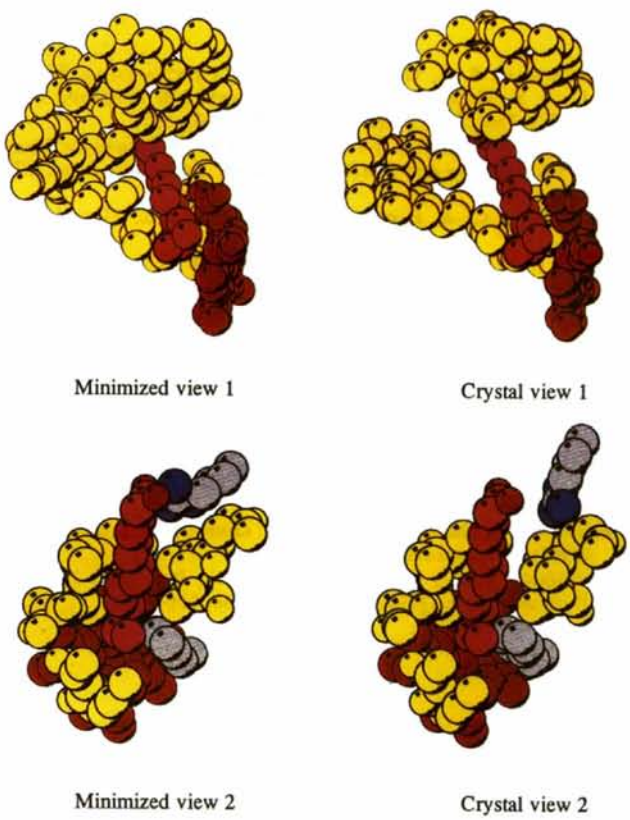

Fig. 11. Hypothetical models of the myristic acid [CH_3—$(CH_2)_{12}$—CO_2H] substrate complexed to molecule *B* of the $P450_{BM-3}$ heme domain before and after minimization. The color-coding scheme is as follows: substrate = purple; heme = red; protein = yellow; key groups that contact the substrate = gray; O atoms on substrate = red; N atoms = blue. The top two views illustrate the large motion of protein groups near the surface toward the substrate after energy minimization. The bottom views show the substrate region closer to the heme. Note that in this region, there is a much less pronounced movement of protein groups. Also note that Phe87 is close to the substrate and may be responsible, in part, for preventing hydroxylation of the terminal methyl group of the substrate, Arg47 is in a position to ion pair or hydrogen bond with the substrate carboxylate group.

The atomic coordinates and structure factors of this structure have been deposited with the Brookhaven Protein Data bank.*

This work was supported in part by National Institutes of Health Grant GM 33688. We would like to thank David Schuller for drawing our attention to the program *SQUASH* and for program modifications.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 2BMH,R2BMHSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: GR0387).

References

Boddupalli, S. S., Hasemann, C. A., Ravinchandran, K. G., Lu, J. Y., Goldsmith, E. J., Deisenhofer, J. & Peterson, J. A. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 5567–5571.

Brünger, A. T. (1992). *X-PLOR Version* 3.1. *A System for X-ray Crystallography and NMR*. New Haven, Yale Univ. Press, CT, USA.

Cambillau, C. & Horjales, E. (1987). *J. Mol. Graphics*, **5**, 174–177.

Collins, J. R., Camper, D. L. & Loew, G. H. (1991). *J. Am. Chem. Soc.* **113**, 2736–2743.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Darwish, K., Li, H. Y. & Poulos, T. L. (1991). *Protein Eng.* **4**, 701–708.

Evans, P. R. (1991). *Crystallographic Computing*, Vol. 5, edited by D. Moras, pp. 136–144. Oxford Univ. Press.

Evans, S. V. (1993). *J. Mol. Graphics*, **11**, 134–138.

Fitzgerald, P. M. D. (1988). *J. Appl. Cryst.* **21**, 275–278.

Furey, W. (1990). Abstract, American Crystallographic Association, Series 2, pp. 18–73.

Gelb, M. H., Heimbrook, D. C., Malkonen, P. & Sligar, S. G. (1982). *Biochemistry*, **21**, 370–377.

Hasemann, C. A., Ravichandran, K. G., Peterson, J. A. & Deisenhofer, J. (1994). *J. Mol. Biol.* **236**, 1169–1185.

Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. & Salemme, F. R. (1987). *J. Appl. Cryst.* **20**, 383–387.

Jaskolski, M., Miller, M., Rao, J. K. M., Leis, J. & Wlodawer, A. (1990). *Biochemistry*, **29**, 5889–5898.

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

Jones, T. A. (1992). *Molecular Replacement*, edited by E. J. Dodson, pp. 91–105. Warrington, England: SERC Daresbury Laboratory.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Li, H. Y., Darwish, K. & Poulos, T. L. (1991). *J. Biol. Chem.* **266**, 11909–11914.

Luzzati, P. V. (1952). *Acta Cryst.* **5**, 802–810.

McRee, D. E. (1992). *J. Mol. Graphics*, **10**, 44–47.

Matson, R. S., Hare, R. S. & Fulco, A. J. (1977). *Biochim. Biophys. Acta*, **487**, 487–494.

Miura, Y. & Fulco, A. J. (1975). *Biochim. Biophys. Acta*, **388**, 305–317.

Narhi, L. O. & Fulco, A. J. (1986). *J. Biol. Chem.* **261**, 7160–7169.

Narhi, L. O. & Fulco, A. J. (1987). *J. Biol. Chem.* **262**, 6683–6690.

Paulsen, M. D. & Ornstein, R. L. (1991). *Proteins*, **11**, 184–204.

Poulos, T. L., Finzel, B. C. & Howard, A. J. (1986). *Biochemistry*, **25**, 5314–5322.

Poulos, T. L., Finzel, B. C. & Howard, A. J. (1987). *J. Mol. Biol.* **195**, 697–700.

Poulos, T. L. & Howard, A. J. (1987). *Biochemistry*, **26**, 8165–8174.

Raag, R., Li, H., Jones, B. C. & Poulos, T. L. (1993). *Biochemistry*, **32**, 4571–4578.

Raag, R. & Poulos, T. L. (1991). *Biochemistry*, **30**, 2674–2684.

Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A. & Deisenhofer, J. (1993). *Science*, **261**, 731–736.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Ruettinger, R. T., Wen, L. P. & Fulco, A. J. (1989). *J. Biol. Chem.* **264**, 10987–10995.

Shirane, N., Sui, Z., Peterson, J. A. & Ortiz de Montellano, P. R. (1993). *Biochemistry*, **32**, 13732–13741.

Siemens Analytical X-ray Instruments Inc. (1990). SAXII Area Detector Software Release Notes. Version 1.21, Siemens Analytical Instruments Inc., Madison, Wisconsin, USA.

Zhang, K. Y. J. & Main, P. (1990*a*). *Acta Cryst.* A**46**, 41–46.

Zhang, K. Y. J. & Main, P. (1990*b*). *Acta Cryst.* A**46**, 377–381.