

The sequence homologies of cytochromes P-450 and active-site geometries

David F.V. Lewis^{a,*} and Henri Moereels^b

^aDepartment of Biochemistry, University of Surrey, Guildford, Surrey GU2 5XH, U.K.

^bDepartment of Theoretical Medicinal Chemistry, Janssen Research Foundation, B-2340 Beerse, Belgium

Received 26 April 1991

Accepted 2 January 1992

Key words: Cytochromes P-450; Sequence homology; Active site models

SUMMARY

The amino acid sequence alignment of 16 cytochrome P-450 proteins representative of the major families is reported. The sequence matching process has been carried out on the basis of maximum homology by residue type, retention of secondary structure and minimization of deletions/insertions except where additional loop regions exist. From the starting point of known reported sequence homology matching from the literature, a realignment on the basis of conserved residues involved in both structure and function gives rise to a self-consistent set of sequences which correlates with known mechanistic and structural data. Once fitted, these archetypal sequences form a straightforward template for the alignment of all P-450 subfamilies. Computer modelling of the active-site regions constructed from homology with the bacterial form of the enzyme (P-450_{CAM}) evinces the correct substrate specificity. Furthermore, the construction of the macromolecular assembly of components of the cytochrome P-450 system on the microsomal endoplasmic reticular membrane is presented from the evidence of site-directed mutagenesis, analysis by molecular probes, X-ray crystallography and molecular modelling.

INTRODUCTION

The cytochromes P-450 constitute a superfamily of over a hundred enzymes, present in all living systems, which are involved in the metabolism of fatty acids, prostaglandins, eicosanoids, steroids and a large diverse number of xenobiotic chemicals [1–5].

The cytochrome P-450 superfamily has evolved over a period of 1,500 million years from an ancestral P-450 gene [6–11] as shown diagrammatically in a simplified form in Fig. 1. The divergence of the various families and subfamilies of cytochrome P-450 proteins has been calculated [12] on the basis of the time scale required for the natural mutation of the P-450 gene to occur. The emergence of the major P-450 families from the primordial gene (presented in Fig. 1) shows that there

* To whom correspondence should be addressed.

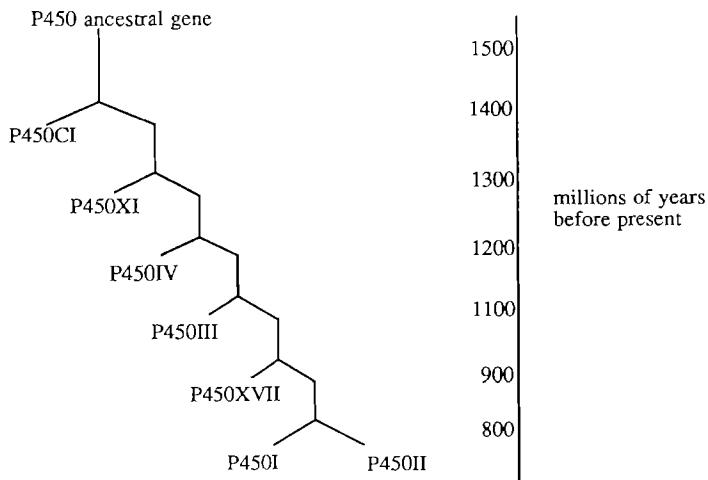


Fig. 1. Cytochrome P-450 superfamily phylogenetic tree (adapted from [7,9,12]).

is a rough correlation with the evolution of bacteria, plants and then animals, where the drug-metabolizing cytochromes P-450 are the most recently formed. Table 1 shows the classification and nomenclature of the cytochrome P-450 proteins, together with their known inducers and inhibitors [13,14]. This table also provides the new cytochromes P-450 nomenclature together with the previous terminology.

Currently, there is considerable interest in the function of the cytochromes P-450 due to their involvement in the activation of certain chemicals to carcinogenic and mutagenic species, in the formation of reactive oxygen intermediates and neo-antigens [15–18] and in the role of certain P-450 proteins pertaining to the development of enzyme inhibitors used in the treatment of cancer and other diseases [19,20]. Although a vast quantity of biochemical information has been generated on the cytochromes P-450 over the past twenty years, it is only recently that the three-dimensional structure of a bacterial form of the enzyme has been determined by X-ray crystallography [21–27]. However, many members of the cytochrome P-450 superfamily have been isolated, characterized and sequenced [28–37] such that several groups of workers have undertaken sequence homology matching using various techniques [38–42].

From the results of sequence alignments between a number of P-450 proteins and the bacterial form, cytochrome P-450_{CAM}, it has been possible to use computer modelling to assist in defining putative binding sites for cytochromes P-450 I and P-450 IIB [43], cytochrome P-450 IA1 [44], cytochrome P-450 IIE [45], cytochrome P-450 LI [46–48], cytochrome P-450 III [49], cytochrome P-450 XVII [50], cytochrome P-450 XIX [51] and cytochrome P-450 IV [42].

Due to the fact that there appear to be certain discrepancies and disagreements between sequence alignments reported by the 3 major groups of co-workers [38–41] a re-examination and modification of cytochromes P-450 protein sequence matching seems opportune, particularly when incorporation of results from site-directed mutagenesis experiments and molecular probe analyses is yielding important information regarding cytochrome P-450 structure and mechanism.

TABLE I
CYTOCHROME P-450 SUPERFAMILY

Family/subfamily and protein	Other literature names	Inducers	Inhibitors	Selectively catalysed reaction
P-450 I				
P-450 IA1	rat c, mouse P ₁ , rabbit form 6 human P ₁	PAHs e.g. benzo[a]pyrene	9-hydroxyellipticine	7-ethoxyresorufin O-deethylase
P-450 IA2	rat d, mouse P ₃ , rabbit form 4 human P ₃	isosafrole	α-naphthoflavone	Glu-P-1 N-hydroxylase
P-450 II				
P-450 IIA1	rat a	phenobarbital		progesterone 7α and testosterone 15α-hydroxylation
P-450 IIB1	rat b, rabbit form 2	phenobarbital	SKF-525A secobarbital, AIA	pentoxyresorufin O-depentylase
P-450 IID1	rat db ₁	noninducible	ajamalicine, quinidine	debrisquine 4-hydroxylase
P-450 IIE1	rat j, rabbit form 3a	ethanol isoniazid	diallyl sulphide	dimethylnitrosamine N-demethylase and p-nitrophenol oxidase
P-450 III				
P-450 IIIA1	rat pcn 1	pregnenolone 16α-carbonitrile	triacetyloando-mycin	testosterone 2β-hydroxylase
P-450 IIIA2	rat pcn 2	pregnenolone 16α-carbonitrile, dexamethasone	erythromycin	testosterone 6β-hydroxylase
P-450 IIIA4	human nf	pregnenolone 16α-carbonitrile	ethinylestradiol	nifedipine N-oxidase
P-450 IV				
P-450 IVA1	rat LAω	clofibrate	terminal acetylenic fatty acids	lauric acid 12-hydroxylase
P-450 IVA4	rabbit p2	progesterone		prostaglandin E ₂ ω-hydroxylase
P-450 XI				
P-450 XIA1	SCC	noninducible	trimethylsilylethyl pregn-5-enediol	cholesterol side-chain cleavage
P-450 XIB1	11β	noninducible		deoxycorticosterone 11β-hydroxylase
P-450 XVIIA1	17α	noninducible	cyclopropylamino androstenol	progesterone 17α-hydroxylase
P-450 XIXA1	aromatase	noninducible	4-hydroxy-androstenedione	androgen aromatase
P-450 XXIA1	C21	noninducible		17α-hydroxy progesterone C21 hydroxylase
P-450 LIA1	14DM	noninducible	ketoconazole	lanosterol C-14 demethylase
P-450 CIA1	CAM	noninducible	metyrapone	camphor 5-exohydroxylase

Modified from Ayerton [13] and Murray and Reidy [14].

METHODS

The amino acid sequences of 16 cytochrome P-450 proteins were collated from the original reports which can be found in previously reported studies [38,40] and matched with the aid of the CGEMA program [52] for sequence analysis using the cytochrome P-450_{CAM} sequence as a template, since certain structural motifs should be conserved. A list of helical, sheet and turn regions of P-450_{CAM} is shown in Table 2. The 16 P-450 sequences were selected as being representative of the 10 major families with certain relevant subfamilies of biochemical importance. These are listed in Table 3 using both types of nomenclature currently in use together with details of the original references of the protein sequences. Visual matching was employed to provide the best fit between the various sequences based on a number of criteria as outlined below:

Criteria for homology matching

- (1) Preservation of secondary and supersecondary structural motifs, especially helices.
- (2) Conservation of structurally and mechanistically important residues, for example, those involved in ion-pairing and electron transport.
- (3) The assumption of conservative changes based on molecular evolution relative to cytochrome P-450_{CAM}.
- (4) Preservation of residue block motifs to coincide with subdomains.
- (5) Elimination of gaps where possible, except where necessary for loop insertion and membrane-interactive segments.
- (6) Consensus between previous sequence homology matching studies with incorporation of the results of experimental work.
- (7) Preservation of heme-binding and substrate-binding regions.
- (8) Invariance of prolines due to their effect on tertiary structure.
- (9) Differences between substrate specificity of P-450s.
- (10) Differences between eukaryotic and prokaryotic P-450s, and between mitochondrial and endoplasmic reticular P-450s.

We believe that such considerations of protein structure and function should provide greater insight than straightforward homology matching using the standard computational methods [53–56].

Molecular modelling

The active-site geometries were generated from the 3D structure of cytochrome P-450_{CAM} (Brookhaven Protein Databank entry number 2CPP) after substitution of appropriate residues, according to the sequence alignment, using the SYBYL molecular modelling package (available from Tripos Associates). The active-site region of cytochrome P-450_{CAM}, comprising the heme group and the ligated Cys³⁵⁷ residue, the camphor substrate, a portion of the distal helix (I) containing residues 244–252, the helical turn region consisting of residues 96–101, and two small sections of β-sheet (residues 294–296 and 395, 396) were extracted from the entire structure for the purpose of modelling analogous regions of other P-450 proteins. A small degree of geometry minimization was employed to relieve unfavourable van der Waals contacts caused by the residue changes. SYBYL was also used to model the substrates' interactions with the relevant cytochromes P-450 by building appropriate substrates and fitting each with its respective active site. Interaction between components in the membrane-bound cytochrome P-450 assembly was carried out

TABLE 2
SUBSTRUCTURAL MOTIFS WITHIN THE CYTOCHROME P-450_{CAM} STRUCTURE

	Helix segments	Sheet segments (strand nos.)	Turn motifs
A	G37 - Q46	β_1 (1) D52 - C58	T ₁ P15 - V18
B	R67 - D77	β_1 (2) G60 - T66	T ₂ D27 - N30
BND	D77 - F81	β_5 (1) G146- F150	T ₃ N33 - A36
BPR	P89 - Y96	β_5 (2) I395 - V405	T ₄ E47 - V50
C	P106 - K126	β_5 (3) S382 - S397	T ₅ C58 - G61
D	L127 - Q145	β_2 (1) G226- V228	T ₆ F98 - S102
E	N149- L169	β_2 (2) G230- I233	T ₇ D104- Q108
F	D173- T185	β_3 (1) V295 - L301	T ₈ P170 - D173
G	T192 - K214	β_3 (2) G315- M323	T ₉ V228 - R231
H	D218 - N225	β_4 (1) Y305- H308	T ₁₀ R277 - I281
I	T234 - S267	β_4 (2) V310 - L312	T ₁₁ F307 - V310
J	S267 - Q276		T ₁₂ N328 - E331
K	R280 - F292		T ₁₃ C334 - H337
β_{10}	K324- N328		T ₁₄ T348 - G351
L	G359- I378		T ₁₅ G353 - L356
			T ₁₆ A384 - A387

using SYBYL with cytochrome b₅ and flavoredoxin structures taken from the Brookhaven Protein Databank (PDB entry numbers: 2 B5C and 4 FXN, respectively). All structures were produced on an Evans & Sutherland ESV20 graphics workstation.

RESULTS AND DISCUSSION

The matched amino acid sequences of 16 cytochrome P-450 proteins are presented in Table 3 using the bacterial form, cytochrome P-450_{CAM}, as a template. The secondary structural motifs of cytochrome P-450_{CAM} are indicated in Table 3 together with the residue numbering system of this form, whereas Table 2 gives the ranges of α -helical, β -sheet and turn regions of the bacterial enzyme. Figure 2 shows a linear mapping of the cytochrome P-450 sequence with secondary structure and binding domains. From the results of site-directed mutagenesis experiments, it has been demonstrated that the substrate, heme and membrane-binding regions are conversed throughout many P-450 proteins (Okhawa, personal communication, 1990) and the present alignment is in full agreement with those findings. Certain key residues, which should be conserved throughout the P-450 superfamily, assist in the alignment process and the procedure of homology matching has given rise to a number of other interesting facets of the structures which are in accordance with known features of cytochrome P-450 functionality. As the 16 P-450 proteins are from different sources – bacterial, fungal, mitochondrial and endoplasmic reticula of mammalia – it is useful to indicate the different components of electron transfer and other factors relevant to their function; these are presented in Table 4 for the 10 major families which make up the 16 individual protein sequences investigated. It is of interest to note that the similarities and differences of structure and function between these enzymes are brought out when their sequences are aligned according to the criteria used in this study.

TABLE 3
CYTOCHROMES P-450 SEQUENCE ALIGNMENTS^a

		A			
		20	40		
CRM	p.p.	TTETI . . . QSNANLAPL PPHFPEHLVFDFDMYNP SNLSAGVCEAMAVL . . . QESNVPDLW	CIA1	
b	rat	M . . . EP . . . SILLNLALL . . . VGFLLL . . . VRGHPKSRL . . . NF . . . PPGPRLPLLGNNQLL . . . DRGGLNSFM . . . QLRKEYGDVF	. . . DRGGLNSFM . . . QLRKEYGDVF	IIB1	
e	rat	M . . . EP . . . TILLNLALL . . . VGFLLL . . . VRGHPKSRL . . . NF . . . PPGPRLPLLGNNQLL . . . DRGGLNSFM . . . QLRKEYGDVF	. . . DRGGLNSFM . . . QLRKEYGDVF	IIB2	
3a	rab	AV . . . LGITVALLGIVMILLFISV . . . WKCQIHSSW . . . NL . . . PPGPFPLPITGNLLQ . . . DLKIDPKSGF . . . RLAEERFGPVF	. . . DLKIDPKSGF . . . RLAEERFGPVF	III E1	
c	rat	MPSVGGPFAITSATELLAVITF . . . LGRFWRVVT . . . RTHWPKG . . . LKS . . . PPGPWLPLFIGHVTL . . . GKNPHLSLT . . . KLSQQYGDVL	. . . GKNPHLSLT . . . KLSQQYGDVL	IA1	
d	rat	M . . . ASFQYIISLAPELLATAIF . . . LFRVFLRGT . . . RTQVPKG . . . LKS . . . PPGPWLPLFIGHVTL . . . GKNPHLSLT . . . KLSQQYGDVL	. . . GKNPHLSLT . . . KLSQQYGDVL	IA2	
p2	rab	. . . ALSPTRLPSLGSLLQVAELLGLLNLKAAGQYLHGRHLRLQQFP . . . PPGFMLLGHSHRFQNQD . . . ELERIQMVE . . . KF . . . PQA P	. . . PPGFMLLGHSHRFQNQD . . . ELERIQMVE . . . KF . . . PQA P	IWA4	
LA	rat	MSVALSSTRFTGSISGLFQVASVGLLLLLVKAVCFYLGQWLLKAFQGFS . . . PPGFIMFGHKGQFQGOK . . . ELQQIMTAVE . . . NF . . . PSAPP	. . . PPGFIMFGHKGQFQGOK . . . ELQQIMTAVE . . . NF . . . PSAPP	IWA1	
NF	hum	M . . . ALIP . . . DLA . . . MET . . . HILLASVLLVLYGTHSHGLFKLGI . . . PGPTPLFLGQNTLISY . . . HKGFNFDE . . . CHICKYGVW	. . . PGPTPLFLGQNTLISY . . . HKGFNFDE . . . CHICKYGVW	IIIA4	
PCN	rat	M . . . DLLS . . . ALT . . . LET . . . WLLAWVVLVLLYGFTRTHGLFKGQI . . . PGPKPLPFFGTVLYN . . . YNGLUKFDOVE . . . CHICKYGVH	. . . PGPKPLPFFGTVLYN . . . YNGLUKFDOVE . . . CHICKYGVH	IIIA1	
14DM s.c.		MSATKSIVGEALEWNVIGLHSFLALPAQRI . . . SLITI . . . FIVN . . . WMLQLYSLRKDRPPFLWVYI . . . PMWSAVV . . . YNGKMFVFEFE . . . LEQKQYGDVF	. . . YNGKMFVFEFE . . . LEQKQYGDVF	LIA1	
17a	bov	M . . . WLLL . . . AV . . . FLLTAYLFLPKTIGHSGA . . . KY . . . PRSLSPLVLGSLPFLP . . . RRQOCHNFF . . . KLEQKPYIY	. . . FLLTAYLFLPKTIGHSGA . . . KY . . . PRSLSPLVLGS LPFLP . . . RRQOCHNFF . . . KLEQKPYIY	XVII A1	
C21	bov	M . . . VLAG . . . LLLL . . . TLLA . . . GAHLWQ . . . R_WKLRLN . . . HL . . . PPLVPGFLHLLQ . . . PNLP1HLLS . . . LTOKLGPVY	. . . PPLVPGFLHLLQ . . . PNLP1HLLS . . . LTOKLGPVY	XXIA1	
SCC	hum	MLA_KQLPPIVSL . . . VKG_QT . . . FLSAPREGQLRVPTEGEGAI . . . S . . . TRS . . . PRPFNEIPSPGDNHLNLYHIRETG . . . KVHLHHQFQKPYIY	. . . PRPFNEIPSPGDNHLNLYHIRETG . . . KVHLHHQFQKPYIY	XIA1	
11b	bov	M . . . AL_WAKA . . . RVR . . . MAGPILS_LHEARLLGTRGAAPKAVLP . . . FEM . . . PRPGNQMMMLQIHKEQ . . . GSENMHDHN . . . QTGEQELP IF	. . . MAGPILS_LHEARLLGTRGAAPKAVLP . . . FEM . . . PRPGNQMMMLQIHKEQ . . . GSENMHDHN . . . QTGEQELP IF	XIB1	
AROM	hum	MVLEMNP1HYN1TSIVPEAMPAATMPVLLTGLFLLVNINYGESSI . . . PPGQY . . . M1GPLISH . . . GRFLUING1OSA . . . NYNNRVYGEFIR	. . . GRFLUING1OSA . . . NYNNRVYGEFIR	XIXA1	
		B	B'	C	D
CRM	p.p.	TRON . . . GHHWIA . . . RG . . . QLIREAY . . . EDYRHFSS . . . PFIPIREAGEAYDIFTSMOPPEQ . . . RQFRALANQW . . . GHPVWD . . . LENRQELASL I	. . . LENRQELASL I	CIA1	
b	rat	TWHL . . . GPRPVWMLCGTDT . . . KEALVGQE . . . DFSGRGT . . . TAVIE . . . PIFKEVGYFANGERWKA LRRFSLATMRDFGMGKRS . . . VEERI QEEAQ . . . LV	. . . VEERI QEEAQ . . . LV	IIB1	
e	rat	TWHL . . . GPRPVWMLCGTDT . . . KEALVGQE . . . DFSGRGT . . . TAVIE . . . PIFKEVGYFANGERWKA LRRFSLATMRDFGMGKRS . . . VEERI QEEAQ . . . LV	. . . VEERI QEEAQ . . . LV	IIB2	
3a	rab	TWYL . . . GSPRWVWLHGKVAKREMLHKKH . . . EFSGRGT . . . IPAF . . . REFKQG . . . IFRNGQTMDKTRAFSLTTLRQDGMKGS . . . NEDR1QKEAHF	. . . NEDR1QKEAHF	III E1	
c	rat	QIRI . . . GSPPWVWLQLNT . . . IQALVKQGD . . . DFKGRPDLYSFITL . . . ANQGSI . . . FNPDSGPMIARRRRLAQALKSFSI . . . ASDPTLASSYLEEHVSKAEAYLI	. . . ASDPTLASSYLEEHVSKAEAYLI	IA1	
d	rat	QIRI . . . GSPPWVWLQLNT . . . IQALVKQGD . . . DFKGRPDLYSFITL . . . TNQKSMTFNPDGSPVMAARRRLAQALKSFSI . . . ASDPTSVSSYLEEHVSKAEANHLI	. . . ASDPTSVSSYLEEHVSKAEANHLI	IA2	
p2	rab	WHL . . . GPKAYLIVDOPDYLKV1LGRS . . . DPKAPRN . . . YKLM . . . TPNI . . . BGVYGLLSDLGQTFWQHFRMLTPAFHYD . . . LKPVWGL . . . MVDSVQHML	. . . MVDSVQHML	IWA4	
LA	rat	RFWH . . . GSKAYLIVDOPDYLKV1LGRS . . . DPKAPRN . . . YKLM . . . TPNI . . . BGVYGLLSDLGQTFWQHFRMLTPAFHYD . . . LKPVWGL . . . KHMAD1RJML	. . . KHMAD1RJML	IWA1	
NF	hum	GFYD . . . GQCPVLA1TDPD1M1KTVLVE . . . Y . . . SVFTNRPFLP . . . VGFMSKA1S1A1EDEEMKRLRSLSSP1FTSKLKE . . . VP11AQYGDVLRNLR	. . . VP11AQYGDVLRNLR	IIIA4	
PCN	rat	GLFD . . . GQMLPFA1TDTM1KTVLVE . . . F . . . SVFTNRPFLP . . . VGIMKAVSAVDEEKMRYRALLSP1FTSGRLKEM . . . FP11EQYGD1LVKYLK	. . . FP11EQYGD1LVKYLK	IIIA1	
14DM s.c.		SFVLLGRMTV1YLGPKHGFEVFNKALRDLV . . . AFGVYAH1LTFPTVFGKGV1YDOPNSRLQME . . . QKVFVKGALTKEAFSKY . . . VPLIAEE . . . WYKFR	. . . VPLIAEE . . . WYKFR	LIA1	
17a	bov	SFLR . . . GSKT1M1V1GHHQLAEREVLLKKK . . . EFSGRPKVATLDL . . . SDNQK1G1AFADHGQHMLQHRLKLNALAFALKDQGNLK . . . LEKI1QNEANVL . . . D . . . F	. . . LEKI1QNEANVL . . . D . . . F	XVII A1	
C21	bov	RLRL . . . GLOEWLNLSKRT1EEAM1TRK . . . WVDAGRPQ1PSYKLV . . . SQD . . . C1 SLDGQSSLLKQHQLKRLTSLALLGTRSS . . . MEPNMDQ1QET . . . F	. . . MEPNMDQ1QET . . . F	XXIA1	
SCC	hum	REKL . . . QNVESVVIOPDEV . . . ALLRKSEGPHPERFL1PPIWVAYHQYQRP . . . IGVLLKSSA1WQD1RVALNQEVHAPEAT . . . KNFLPLDAVSDFVSVLH	. . . KNFLPLDAVSDFVSVLH	XIA1	
11b	bov	RYDV . . . GGRHVFMV1LPEDVER . . . QLOADSHIPQRM1LEPHILAYQARQHKG . . . GVFLLNGPQ1QRLDRRLRNLPDVLSPAL . . . QKYP1LVDGWARDPSQTLK	. . . QKYP1LVDGWARDPSQTLK	XIB1	
AROM	hum	VH1S . . . GEETL1 . . . ISKSSSMFH1MKH1N . . . HYSSRGSGKLGQ1G1MHEKG1 . . . IFNNPPEL1K1TRPFM1KALSGPGLV . . . RMV1TVAESLK1HLD . . . R	. . . RMV1TVAESLK1HLD . . . R	XIXA1	
		E	F	G	
CRM	p.p.	140 . . . QG . . . NFT . . . DYAREPFP1R1F1M1L . . . AGL . . . PEED1PHFLKYLTDQTM . . . R . . . PDGS1MTFREAKERAYDYL1P1 . . . CIA1	. . . CIA1		
b	rat	EELRKS . . . QGAPLDP1TFLQ . . . ITANI . . . CS1V6F6F6DFTYDTRQFLRLLFYLFR1TFSLLSSSSQVFEFFSOF1 . . . KVFPGAHRQ1SKNQEI1LDV1GHI . . . V	. . . V	IIB1	
e	rat	EELRKS . . . QGAPLDP1TFLQ . . . ITANI . . . CS1V6F6F6DFTYDTRQFLRLLFYLFR1TFSLLSSSSQVFEFFSOF1 . . . KVFPGAHRQ1SKNQEI1LDV1GHI . . . V	. . . V	IIB2	
3a	rab	EELRKT . . . QGQFDP1TFLQ . . . ITANI . . . CS1V6F6F6DFTYDTRQFLRLLFYLFR1TFSLLSSSSQVFEFFSOF1 . . . KVFPGAHRQ1SKNQEI1LDV1GHI . . . V	. . . KVFPGAHRQ1SKNQEI1LDV1GHI . . . V	III E1	
c	rat	SKFQLKMAEVGHDFPKYLV1WSV1N1C1 . . . FTRRYDHDDGELS1 . . . V1NSNFGEV1TGSY . . . PADF1P1L . . . RYLPNSSLDAFKD1KQFYSMCKL1 . . . CIA1	. . . CIA1		
d	rat	SKFQLKMAEVGHDFPKYLV1WSV1N1C1 . . . FTRRYDHDDGELS1 . . . V1NSNFGEV1TGSY . . . PADF1P1L . . . RYLPNSSLDAFKD1KQFYSMCKL1 . . . CIA2	. . . CIA2		
p2	rab	DHMEQ1QDSS1E1FQH1VSL1H1D1T1M1 . . . AFSYQ8SV1QD1RNSH1SY1C1 . . . ND1L1N1L1V1F1Y1R1N1V1F1Q1D1T1Y1 . . . NAVDFPVL . . . RYLPN1ALQ1P1F1ND1V1FL1S1L1Q1 . . . XVII A4	. . . XVII A4		
LA	rat	DKWQELAQ1QDSS1E1FQH1VSL1H1D1T1M1 . . . AFSYQ8SV1QD1RNSH1SY1C1 . . . ND1L1N1L1V1F1Y1R1N1V1F1Q1D1T1Y1 . . . NFSSN1H1N1P1Q1L1D1H1TD1V1K1 . . . CIA1	. . . CIA1		
NF	hum	REAEAT . . . QG1V1T1D1Q1D1V1G1F1G1V1M1D1V1 . . . T1S1F1G1N1D1S1N1Q1D1P1F1V1T1K1L1R1F1D1P1F1L1S1W1P1F1 . . . I1P1L1E1V1L1N1 . . . QFVPREV1TFLRK5	. . . QFVPREV1TFLRK5	IIIA4	
PCN	rat	QAEAT . . . QG1V1T1D1Q1D1V1G1F1G1V1M1D1V1 . . . T1S1F1G1N1D1S1N1Q1D1P1F1V1T1K1L1R1F1D1P1F1L1S1W1P1F1 . . . I1P1L1E1V1L1N1 . . . QFVPREV1TFLRK5	. . . QFVPREV1TFLRK5	IIIA1	
14DM s.c.		DSK1NFR1L1N1T1D1 . . . W1V1T1Q1P1E1T1F1A1S1L1G1K1P1M1R1L1D1 . . . K1Q1W1D1G1N1D1I1E1L1V1S1K1E1L1 . . . PL . . . EYRKR1D1Q1K1S1G1T1M1S1L1	. . . PL . . . EYRKR1D1Q1K1S1G1T1M1S1L1	LIA1	
17a	bov	LATCH . . . QEA1D1L1S1P1L1S1V1N1S1F1C1F1N1S1K1N1D1 . . . T1L1F1G1K1D1T1L1W1H1F1D1V1Q1D1M1K1H1D1M1S1 . . . K1Q1W1D1G1N1D1I1E1L1V1S1K1E1L1 . . . V	. . . V	XVII A1	
C21	bov	MWVA . . . QG1V1T1D1Q1D1V1G1F1G1V1M1D1V1 . . . T1S1F1G1N1D1S1N1Q1D1P1F1V1T1K1L1R1F1D1P1F1L1S1W1P1F1 . . . I1Q1L1D1W1P1F1 . . . R1F1P1G1S1L1R1Q1M1K1D1V1M1S1 . . . XXIA1	. . . XXIA1		
SCC	hum	R1I1K1Q1G1S1N1S1D1 . . . QD1R1F1P1F1A1S1L1G1K1P1M1R1L1D1 . . . T1W1F1G1Q1M1L1E1R1G1H1D1M1H1S1 . . . D1Q1W1D1G1N1D1I1E1L1V1S1K1E1L1 . . . V	. . . V	XIA1	
11b	bov	ARV1LQ1A1G1S1N1S1D1 . . . T1D1P1R1F1V1T1A1S1L1V1Y1E1R1G1L1 . . . T1Q1P1F1L1Q1H1K1M1K1S1T1V1G1H1 . . . F1V1P1R1K1T1K1D1W1A1D1V1L1S1K1D1Y1T	. . . F1V1P1R1K1T1K1D1W1A1D1V1L1S1K1D1Y1T	XIA1	
AROM	hum	LEE1V1N1S1 . . . Y1V1D1V1T1L1R1R1M1D1T1 . . . N1T1L1F1L1R1P1D1E1S1A1V1W1K1 . . . I1W1Q1G1F1D1A1C1L1E1 . . . PD1F1K1S1M1Y1K1Y1E1S1V1D1Q1D1K1D1A1E1V1L1E1 . . . XIXA1	. . . XIXA1		

^a The numbering system refers to the cytochrome P-450_{CAM} structure which is sequence 1. Helical and sheet regions in P-450_{CAM} are indicated above the sequence pattern and standard single-letter-code amino acid nomenclature has been used. The previous and most recent nomenclature of the P-450 proteins are shown at the start and end of the sequences, respectively. The listed sequences are taken from Nelson and Strobel [40] and Gotoh and Fujii-Kuriyama [38] though we found that some of the original sequences referenced in these studies differed from those reported. The protein sequences of

TABLE 3 (continued)
CYTOCHROMES P-450 SEQUENCE ALIGNMENTS^a

		H	β_2	240	I	260	
		220					
CAM	p.p.	EQRRQKPGT.....	DAISIVAN.....	GQWGRP..ITSDEAKMGLLVLGGDLTWNFLSFSNEFLAKSPHEHQELIER.....			CIA1
b	rat	EKHKRATLD.....	PSAPPDFIDTYLL.....	RMEKEKSNNHTEFHENLMISLLSLFPAFTETSTSTLRLYQFLMLKYPHVKEVQEIKDVGSHR.....			IIIB1
e	rat	EKHKRATLD.....	PSAPPDFIDTYLL.....	RMEKEKSNNHTEFHENLMISLLSLFPAFTETSTSTLRLYQFLMLKYPHVKEVQEIKDVGSHR.....			IIIB2
3a	rab	KEHHKSLD.....	PSAPPDFIDSLLI.....	EMEIKHSTEPLTYLENIAVTDAMHFFAGTTSTTLRLYQFLMLKYPHVKEVQEIKDVGSHR.....			III E1
c	rat	KEHYRTFE.....	KGHIRDITDSLLIEHC.....	QDRRLDENANVQLSDOKVITVFDLFGAFGDTITTAISMSLMLVLPVTPRQKQEEQELTVIGDR.....			IA1
d	rat	QEHYDFN.....	KNSICDITGALFCH.....	SENYKQDGLIPQVNLVIFDGFAGFETVITAIFRSILLLVTEPKWVKEIHEEDTVIGDR.....			IA2
p2	rab	RKAQLOQEGELEKVRIRKRLDFLVLLF.....		AKMENGSSLSDQDLRAEVDTFMEGHDTITASGVSHYFALATHPEHQHREEIQGLLGDGA.....			IVA4
LA	rat	RKDQDQNGAELEKVKQRRLDFLQLLLL.....		ARMENGQSLSDQDLRAEVDTFMEGHDTITASGVSHYFALATHPEHQHREEIQGLLGDGA.....			IVA1
NF	hum	VKRMKESRLDT.....	QKHRVDFQLMID.....	SQNSKETESHAKSLDLELVQSIIFIFQAGYETTSVSLVSHLSDPDKQKLOEEIDAVLPNKA.....			IIIA4
PCN	rat	VYRHKDFLMLNA.....		HNDSKDKEHTSALSDMTEIAAGSIIIFIFQAGYETTSVSLVSHLSDPDKQKLOEEIDAVLPNKA.....			IIIA1
14DM	s.c.	IKERRKNN.....	D1QDQLDIDSLSMK.....	NSTYK...DGVRHDTQEIARLILGVLMMQGQHTSAAATSANILHLAERPDQVQELEYEQMRVLDGGK.....			LIA1
17a	bov	EK QENFS.....	SDSITNLHLLIQLQKVNADNNAGPDDQSKLSSRHLATLGIDFGAFETVTSVSKVIIAYVLLHPSLQKRIODD100116FNR.....			XVII A1	
C21	bov	TRHKSIV.....	AGQMDHTDYLQG.VQPRVVEGPQG.....	LLGE..HVMSWLF1667ETTA1SLWVHPSLQKRIODD100116FNR.....			XVII A1
SCC	hum	QNFYMLRQK.....	GSMHDYDQILY.....	.RLLQDQSKHEDFIDKANVTEHLAGGWDITTSQHLYEHANLXQDMLRAEVLAHRQKQ.....			XIA1
11b	bov	QRIYQELA.....	LGHPWYHSIV.....	AELLRAADMTLDTIKANTIDLTAGSVDTTAFPLLNLFEELARNPVEQAVQAVQESLVAEARI.....			XIB1
AROM	hum	KRRIISTEEKLEEIM.....	DFATELIL.....	AEK..RGDLTREVNQI11EMI1APDTHMSVSFLFRFLIAKHPHWEAT1IKE1QT1VGE.R.....			XIXA1
		K	β_3	β_4	β_3	340	
		280	300	320			
CAM	p.p.	PERIPAAEELLRRFSLVADG.....	RILTSDYEFG.HGVL.....	KKGDQI1LPQMLSGLDERENA1PMHDFRSQ.....			CIA1
b	rat	L...PTLDRSKHMPYTDAVIHE1QRFSDLVLP1GVPVRVTKDTHF.RGYLL.....	PKNTIEWP1RSSALHDQPYFDHPDSFNPEHFLDANGA.....			IIIB1	
e	rat	P...PSLDDRTPYTDAVIHE1QRFIDLVPNSLPHEATDFTF.QGWVI.....	PKGTWV1PTLDSLLYDQEFDPPEKFKEHFNLNEEGK.....			IIIB2	
3a	rab	M...PSVDRVHMPYDAWHE1QRFIDLVPNSLPHEATDFTF.QGWVI.....	PKGTWV1PTLDSLLYDQEFDPPEKFKEHFNLNEEGK.....			III E1	
c	rat	Q...PLSDRDOLPYLEAF1LETFRHSSFPVFT1PHST1DTDSL.NGFWI.....	PKGH..VFVNOVQVHDEQELQDFPNEFRPERFLTSSGT.....			IA1	
d	rat	Q...PLSDRDOLPYLEAF1LETFRHSSFPVFT1PHST1DTDSL.NGFWI.....	PKEGCFIFINQHCVHDEQELQDFPNEFRPERFLTSSGT.....			IA2	
p2	rab	S...ITMEHLDQHPTT1PKERLRLYPPV.....	SVTQLSKPV1TFPDGSRSL.PKGVI1FLS1VYGLHNPWQMPNPEVFDPSRFAPDS.....			IVA4	
LA	rat	S...ITMEHLDQHPTT1PKERLRLYPPV.....	SVTQLSKPV1TFPDGSRSL.PKGVI1FLS1VYGLHNPWQMPNPEVFDPSRFAPDS.....			IVA1	
NF	hum	P...PTYDTVQHMEYLDWVNET1LRLFP1IAN.RLERVVKKDVIE1.NGMFI.....	PKGMVWMI1PSYALHHDQPKW1TEPEKFPLPERFSKNNK.....			IIIA4	
PCN	rat	P...PTYDTVQHMEYLDWVNET1LRLFP1IAN.RLERVVKKDVIE1.NGMFI.....	PKGMVWMI1PSYALHHDQPKW1TEPEKFPLPERFSKNNK.....			IIIA1	
14DM	s.c.	KE...LTYDQLEMPNLLQK1IET1RNHPLHS1LFRK1.....	KEHMVPTS1V1PAGYHVLVSGYTHLREYFPNHAOFN1HANKNSASSYVGEEVDYGF.....			LIA1	
17a	bov	T...PT1SDRNLVLLAE1TREVLR1RPW1LPIHKAV1DSS1.GDLT1.DKGTDVWNL1WALHSEKEQHDFLPMFPLDFTGT.....				XVII A1	
C21	bov	S...SRVITYKDRARLPLLNA1IAEVLR1RPW1PLALPHRT1PSSI1.FGYDI.....	PEGRWV1PHLQGAHDET1WQEPHEFRPDRFLEP.GA.....			XOIA1	
SCC	hum	G...DMATHLQLVPLLKAS1KET1RLHP1S1V1L.QRYLVLNDLVL.RDYH1..PAKT1LVQA1YALGRE1PTFFFDPENF1PTRLSDK.....				XIA1	
11b	bov	E...NPQR1AI1TELPL1RRAALKET1RLYPVG1TL.EREVSSD1LVL.QNYH1..PAGT1LV1KLVL1SLGRNPAVFARESYH1PQMLDQGS.....				XIB1	
AROM	hum	D...1KIDD1QLKVMENF1YESMRYOPWDLV.MRKALEDVI1.DGYPV.KKG1TN1LN1GRMRH..LEFPKPKNEFT1LENFAKV.....				XIXA1	
		L	β_5	400			
		360	380				
CAM	p.p.	KWSHTFGHSHL1LQGHALAR1.I1VT1K..EMLTR1.....	PDFS1APGAQ1QHKG1IVSGVGA..LPLVWDPATT1KAV.....				CIA1
b	rat	..LKOSEAFMFEST1GK11GEG1ARNE1FLFPTT1LONFSVS..SHLAPKD1DLTPKES..G1AKIP..PTYQ1QCSAR.....					IIIB1
e	rat	..LKOSEAFMFEST1GK11GEG1ARNE1FLFPTT1LONFSVS..SHLAPKD1DLTPKES..G1AKIP..PTYQ1QCSAR.....					IIIB2
3a	rab	..FKYSDYF1KPF5AGK1RV1V6EGLARNE1FLLSS1A1LQHFN1K..PLWQED1DLN1TV..GFGGRV..PRYK1QV1P1RS.....					III E1
c	rat	..LD1K1SEK1V1FLG1QK1R1GET1Q4RLEVFLF1LQGME1NPSGK1V1D..HTPRAY..GLT1KHAR1EHFQMG1R1SSQ1QH1Q1A.....					IA1
d	rat	..A1DK1LSEK1V1FLG1QK1R1GET1Q4RLEVFLF1LQGME1NPSGK1V1D..HTPRAY..GLT1KHAR1EHFQMG1R1SSQ1QH1Q1A.....					IA2
p2	rab	..AVH1KFL1PF5G1GK10KF1GKOF1AM1K1V1W1V1F1L1P1D..PDP..TR1PI1IARV..V1L1K1NG1V1L1R1R1K1H.....					IVA4
LA	rat	..PRH1KFL1PF5G1GK10KF1GKOF1AM1K1V1W1V1F1L1P1D..PDP..TR1PI1IARV..V1L1K1NG1V1L1R1R1K1H.....					IVA1
NF	hum	..N1D1P1Y11PF1G5G1P1G1M1F1L1V1L1Q1N1F1S1K..P1K1Q1P1L1S1Q1..G1L1Q1P1K1P1R1..W1L1K1N1P1D1S1A1.....					IIIA4
PCN	rat	..S1D1P1Y11PF1G5G1P1G1M1F1L1V1L1Q1N1F1S1K..P1K1Q1P1L1S1Q1..G1L1Q1P1K1P1R1..W1L1K1N1P1D1S1A1.....					IIIA1
14DM	s.c.	GAISKG1V1SS1P1L1P1G1P1R1..1G1H1F1A1Y1Q1G1V1L1S1F1I1T1L1K1W1H1P1E1G1..K1V1P1P1D1P1..F1T1M1V1L1..P1T1G1P1K1I1E1M1P1K1.....					LIA1
17a	bov	..Q1I1S1P1S1L1P1G1P1R1..1G1H1F1A1Y1Q1G1V1L1S1F1I1T1L1K1W1H1P1E1G1..K1V1P1P1D1P1..F1T1M1V1L1..P1T1G1P1K1I1E1M1P1K1.....					XVII A1
C21	bov	..N1P1S1A1P1S1L1P1G1P1R1..1G1H1F1A1Y1Q1G1V1L1S1F1I1T1L1K1W1H1P1E1G1..K1V1P1P1D1P1..F1T1M1V1L1..P1T1G1P1K1I1E1M1P1K1.....					XIA1
SCC	hum	..N1T1F1R1L1P1G1P1R1..1G1H1F1A1Y1Q1G1V1L1S1F1I1T1L1K1W1H1P1E1G1..K1V1P1P1D1P1..F1T1M1V1L1..P1T1G1P1K1I1E1M1P1K1.....					XIA1
11b	bov	..G1R1P1F1L1G1F1G1V1Q1R1A1E1M1T1I1N1H1N1F1R1..1G1H1F1A1Y1Q1G1V1L1S1F1I1T1L1K1W1H1P1E1G1..K1V1P1P1D1P1..F1T1M1V1L1..P1T1G1P1K1I1E1M1P1K1.....					XIB1
AROM	hum	..PY1R1F1Q1P1G1F1G1P1R1..1G1H1F1A1Y1Q1G1V1L1S1F1I1T1L1K1W1H1P1E1G1..K1V1P1P1D1P1..F1T1M1V1L1..P1T1G1P1K1I1E1M1P1K1.....					XIXA1

each cytochrome P-450 listed in this work are taken from the original reports which are as follows: P-450_{CAM}: [79], P-450_{SCC}: [80], P-450_G: [81], P-450_C: [82], P-450_{14DM}: [36], P-450_d: [83], P450_c: [84], P-450_{3a}: [33], P-450_{Lao}: [31], P-450_{p2}: [33], P-450_{PCN}: [85], P-450_{NF}: [30], P-450_{C21}: [86], P-450_{arom}: [29], P-450_{17a}: [32], P-450_{IIp}: [87]. The specifically catalysed reactions of individual cytochromes P-450 are shown in Table 1.

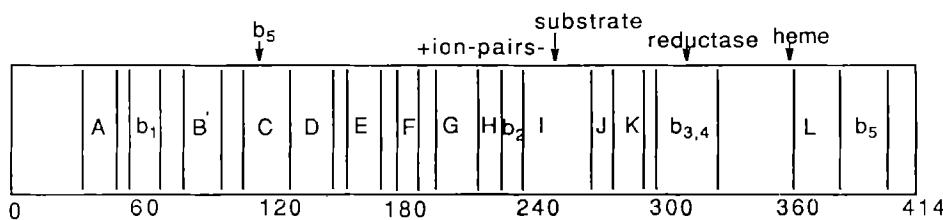


Fig. 2. Secondary structural segments and binding domains in P-450_{CAM}. The approximate position of residue numbers is indicated. A → L represent helical segments; b₁ – b₅ represent β-sheet segments (the remainder is composed of turns and random coils). The ion-pair between the F and I helices constrains the tertiary structure in the binding-site region which is composed of a hydrophobic stretch of the I helix and the heme binding pocket at the start of helix L.

It is found that a number of residues are either invariant or highly conserved between sequences and these are listed in Table 5 together with their possible or probable relevance.

Comparisons with previously reported sequence alignments

Sequence alignments between cytochromes P-450 have been reported by Nelson and Strobel [40], Edwards et al. [41] and Gotoh and Fujii-Kuriyama [38]. All of these alignments differ from each other and from the alignment presented in this work. The methods employed by Nelson and Strobel and Gotoh and Fujii-Kuriyama do not take into account the conservation of tertiary structure, whereas that reported by Edwards and co-workers involves alignment of the proposed helical regions of cytochrome P-450 proteins with that of the bacterial form of the enzyme.

More recently, Zvelebil et al. [44] have investigated multiple sequence alignments of 12 cytochromes P-450 and have used their results to construct a model of cytochrome P-450 IA1 based on the cytochrome P-450_{CAM} structure. Their approach utilizes a combination of alignment algorithms rather than the reliance on a single method and, as such, represents an advance on previous sequence homology matching. These workers have recognized, like ourselves, that the tertiary structure is essentially conserved in all cytochromes P-450, that loop regions are present in eukaryotic forms but not in the bacterial form, that charged residues tend to be conserved throughout cytochromes P-450 where they are required for ion-pairing and that about 40 residues at the N-terminus of eukaryotic cytochromes P-450 span the membrane. However, in this study, we have used experimental evidence from the results of site-directed mutagenesis and molecular probe

TABLE 4
TEN MAJOR FORMS OF CYTOCHROME P-450

Origin	Bacterial	Fungal	Mitochondrial	Endoplasmic reticular
P-450 family	CI	LI	XI,XVII,XIX,XXI	I,II,III,IV
Reducing agent	NADH	NADPH	NADPH	NADPH or NADH
Fe-S protein	putidaredoxin	none present	adrenodoxin	none present
Flavoprotein	reductase	P-450 reductase	reductase	P-450 reductase
Membrane	no	yes	yes	yes
Cytochrome b ₅	no	no	no	yes
Substrate(s)	camphor	lanosterol	steroids	xenobiotics, etc.

TABLE 5
RESIDUES INVARIANT OR HIGHLY CONSERVED IN P-450s

CAM no.	Type	Comments	CAM no.	Type	Comments
Pro ¹⁵	invariant	defines start of first turn into helix A	Phe ³⁵⁰	invariant	possibly π - π stacking interaction with His ³⁵⁵ for electron conduction
Gly ⁶⁰	invariant	structural prerequisite	Gly ³⁵³	invariant	defines heme-binding pocket
Arg ¹¹²	conserved	heme binding via propionate and electron transfer ion-pair with b ₅ or redoxin	His ³⁵⁵	conserved	charge relay from reductase to heme
Gly ²⁴⁹	conserved	hydrogen-bonded pair } defines substrate	Cys ³⁵⁷	invariant	heme ligation via sulphur (thiolate)
Thr ²⁵²	invariant	hydrogen-bonded pair } binding pocket	Gly ³⁵⁹	invariant	defines heme binding pocket
Glu ²⁸⁷	invariant	ion-pairs with Arg ²⁹⁰	Ala ³⁶³	conserved	defines heme binding pocket
Arg ²⁹⁰	invariant	ion-pairs with Glu ²⁸⁷	Leu ³⁷⁵	conserved	hydrophobic interaction with D helix
Arg ²⁹⁹	conserved	heme binding via propionate			

analysis to refine our alignment model such that key residues involved in redox and substrate binding interactions are optimally aligned across the cytochrome P-450 sequences.

Of the substrate binding-site regions, all reported sequence homology alignments agree on the I helical domain due to the fact that there are conserved residues close to the invariant Thr²⁵². However, differences occur between reported homology matchings for the B' helical turn because this region is more sensitive to substrate specificity differences between P-450 proteins and, therefore, it is more difficult to accurately predict alignment. Therefore, we feel that by consideration of the commonalities between cytochromes P-450, it is possible to arrive at more precise sequence homology matching which is self-consistent with the experimental evidence relating to cytochrome P-450 structure and function. It must be realized that the cytochromes P-450 have a low identity across the superfamily (15–20%) because of the wide variation in substrate specificity and this makes multiple sequence alignment difficult to achieve by a purely computational method.

The cytochrome P-450_{CAM} structure and its relevance to other P-450s

The 3D crystal structure of cytochrome P-450_{CAM}, as revealed by X-ray crystallography of the substrate-bound form (Fig. 3), shows a tertiary structure comprising a helical-rich region and a sheet-rich region from a span of 414 amino acid residues [23]. Of the 10 helical segments A-L, helices E, F, G and I form an antiparallel 'Greek key' helical bundle similar to that found in other globular proteins such as the globins, cytochrome c peroxidase, thermolysin, lysozyme and papain. Moreover, helices D and L show a parallel alignment where the latter comprises the heme binding region since it contains the invariant cysteine which ligates the iron atom. Close to this Cys³⁵⁷ there are a number of hydrophobic residues which define the heme pocket and are also either invariant or highly conserved in other P-450 proteins: these are Gly³⁵³ and Gly³⁵⁹, Leu³⁵⁸ and Leu³⁷⁵, Phe³⁵⁰ and Ala³⁶³. The Phe³⁵⁰ and a conserved basic residue, His³⁵⁵ are probably involved in electron transfer to the heme iron. In fact, it appears that the histidine lies at the centre of a charge relay system made up from Ser⁸³, His³⁵⁵ and the carboxylate moiety of the heme propionate. An analogous electron transport pathway (also known as a catalytic triad) is found in enzymes such as chymotrypsin and other serine proteases, where hydrogen bonding facilitates electron conduction through the apoprotein. Although the network of hydrogen bonds, which permeate the entire cytochrome, could act as potential electron conduits for the reduction process

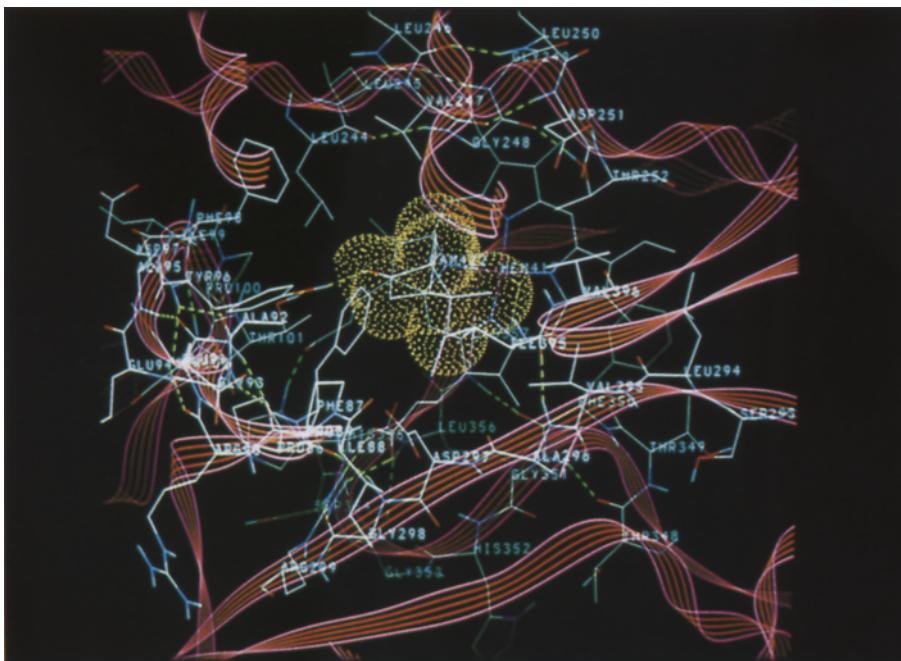


Fig. 3. The 3D structure of cytochrome P-450_{CAM} in the region of the active site. The camphor substrate is displayed as a dot surface and atoms are coloured according to type. Hydrogen bonds are shown as dashed lines.

necessary to P-450 function, an equally attractive possibility is provided by a number of aromatic residues which are arranged in a 'herringbone' pattern suggestive of a continuous channel of π -stacking interactions, at the optimum distance for electron tunneling, proceeding from the surface of the protein to the heme. In cytochrome P-450_{CAM} the aromatic rings are provided by Phe²⁵⁶, Phe²⁵⁹, Phe²⁶³, Phe²⁹², Phe³⁴⁰ and Phe³⁵⁰, and His²⁷⁰ and His³⁵⁵. In addition to Phe³⁵⁰ and His³⁵⁵ already mentioned above, Phe²⁵⁹ is also conserved in all P-450 proteins with the exception of P-450_{11B}. However, as the full complement of residues which comprise this aromatic channel is only conserved in some, but not all, cytochromes P-450, it is likely that they could only provide an alternative electron transport pathway in certain instances where it is mechanistically advantageous to have a secondary route to the 'normal' hydrogen-bonded conduit(s). Table 6 shows that there are different electron transfer components involved in the different P-450 systems and that reduction may occur via a flavoprotein (reductase) through a redoxin, or by cytochrome b₅, which utilizes reducing equivalents provided by NADPH or NADH. There must, therefore, be either one or two possible sites of interaction between the electron transfer component(s) and the cytochrome P-450. In the bacterial form, these have been identified as centering around Arg¹¹² and Lys³¹⁴. The first of these basic residues forms an ion-pair with a heme propionate and provides the closest point of contact between the heme at its proximal site and the protein exterior. Furthermore, it is highly conserved in all cytochromes P-450 with the exception of P-450_{14DM}, being present as either arginine or, in one instance, lysine. In microsomal P-450s, this residue is thought to form a site of interaction with cytochrome b₅ by acting in concert with other basic (positively charged) residues (Lys³⁴⁴, Arg⁷² and Arg³⁶⁴) in forming electrostatic attractions with corresponding carboxylate

TABLE 6
ELECTRON TRANSPORT PATHWAYS IN VARIOUS P-450 SYSTEMS

Bacterial	NADH	→	Putidaredoxin reductase	→	Putidaredoxin	→	P-450
Fungal	NADPH	→	cytochrome P-450 reductase			→	P-450
Mitochondrial	NADPH	→	adrenodoxin reductase	→	adrenodoxin	→	P-450
Microsomal	NADPH	→	cytochrome P-450 reductase			→	P-450
	or					→	P-450
	NADH	→	cytochrome b ₅			→	P-450

moieties on b₅ [57]. In other forms of P-450 proteins, this highly conserved arrangement of basic residues probably represents the point of interaction with redoxin [58–60]. As there is a tryptophan residue 4 residues upstream from the conserved arginine in all P-450s other than the bacterial or fungal forms, it is interesting to speculate as to the possibility of the involvement of this aromatic residue in electron transfer to the heme via a π-π stacking interaction.

Lys³¹⁴ has been shown to correspond with a point of electrostatic interaction between the flavoprotein reductase and liver microsomal cytochrome P-450 [61–65] from fluorescent probe experiments. It can be seen from Table 3 that this residue is well conserved in all P-450s with the exception of those involved in steroid biosynthesis. However, in these instances, the reductase is not directly coupled to cytochrome P-450 and therefore the invariance of this residue is not required because redoxin can interact at the Arg¹¹² site as mentioned previously. An inspection of the 3D structure of cytochrome P-450_{CAM} via molecular graphics shows that there is a hydrogen-bonded

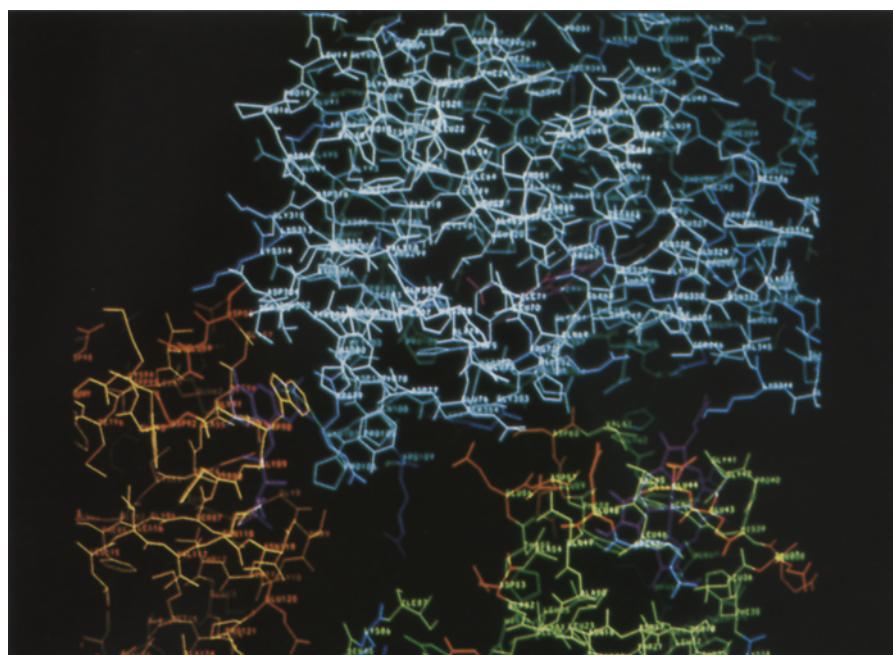


Fig. 4. A postulated mode of interaction between cytochrome P-450 (blue), redoxin (orange) and cytochrome b₅ (green) with hemes (on the cytochromes) and flavin on the redoxin shown in purple. Basic and acidic residues are displayed in blue and red, respectively, indicating the possible electrostatic interactions.

framework from Lys³¹⁴ through to the Ser⁸³-His³⁵⁵ charge relay at the heme edge, suggestive of an electron access channel from reductase to the heme iron. Figure 4 shows a possible interaction between cytochrome P-450, redoxin and cytochrome b₅.

Eukaryotic cytochromes P-450 possess a larger number of amino acids than the bacterial form and it is generally regarded that these additional residues form variable loop regions which serve to position the protein in the appropriate cell membrane. The sequence alignment presented in this work reveals the presence of 4 loop motifs in addition to an N-terminal helix which spans the membrane, in agreement with recently reported findings [66,67]. It is possible that the extra loop regions act as 'rafts' to stabilize the interaction between the cytochrome P-450 and the membrane phospholipid; the appearance of negatively-charged residues (aspartate and glutamate) in each of these additional segments could reinforce this association by electrostatic attraction to the phosphatidyl choline and phosphatidyl ethanolamine positively charged nitrogen atoms. Moreover, there could be a role for the phospholipid bilayer in the cytochrome P-450 oxidative reactions [68] by improving electron conduction, by providing better contact with other macromolecules in the redox chain, and by facilitating access of substrates to the cytochrome P-450 active site by transverse diffusion [69].

A number of factors contribute to the preservation of the overall tertiary structure of the cytochromes P-450. These are the invariant prolines which define interhelical turn regions, two internal ion-pairs and several glycine residues which enable folding of the apoprotein and, in particular, formation of the substrate and heme-binding domains. In the camphor structure, the glycine positions are 60, 249, 353 and 359, for protein folding (60), and for defining the hydrophobic pocket (249) and heme site (353 and 359), respectively, whereas Arg¹¹² and Arg²⁹⁹ hold the heme by ion-pairing with the propionate side chains and these basic residues are highly conserved in all P-450s. An ion-pair comprising Asp²⁵¹ and Arg¹⁸⁶, well-conserved throughout the cytochromes P-450, cooperatively assists in the protein folding which gives rise to the substrate binding site adjacent to the I helix, distal to the heme. Other ion-pairs which are wholly or partially conserved in the P-450 proteins correspond with Arg⁷²-Glu³³¹, Glu²⁸⁷-Arg²⁹⁰ and Asp²¹⁸-Arg²¹¹ in the P-450_{CAM} structure. Pro¹⁵ forms the start of the first turn in the bacterial cytochrome whereas Pro¹⁸⁷ defines an interhelical turn between the E and F helices which may represent part of the substrate access channel. Moreover, Pro²⁶⁸, which is invariant, links the turn between the helix I and helix J. Finally, prolines at P-450_{CAM} positions 335 and 408 are strongly conserved within other cytochromes P-450 and probably have a structural role in preserving the protein folding, though there is some evidence for their participation in the 'herringbone' stacking pattern of aromatic residues which could represent an electron conduit.

The substrate binding site(s)

Cytochrome P-450_{CAM} positions its endogenous substrate, camphor, within a hydrophobic cleft close to the heme such that stereospecific oxygenation occurs in the 5-exo position. In the crystal structure, camphor is held by a number of contacts with hydrophobic residues on the distal (I) helix and by a hydrogen bond formed with the hydroxyl group of Tyr⁹⁶. Other than those residues previously discussed, there is an invariant threonine* at position 252 which forms a hy-

*The appearance of proline instead of threonine at this position in P-450_{PCN} is thought to be a misreading of the cDNA information as the codon for proline (CCX) is similar to that of threonine (ACX) where X = any nucleotide base.

drogen bond with the conserved Gly²⁴⁹, though there is now growing evidence for the participation of the threonine hydroxy group in the oxygenation mechanism [70].

The distal stretch of helix I, corresponding to the primary region of binding to the camphor molecule in P-450_{CAM}, is highly variable within the matched sequences and thus may account for the substrate specificities of other cytochromes P-450. There is also variation at the secondary site corresponding to Tyr⁹⁶ and these differences entirely complement the differentiation between substrate specificities found in the various P-450 proteins under investigation. Active-site models have been proposed for many cytochromes P-450 which agree well with the range of substrate types found for the corresponding enzymes [15,16,43,45,47,48]. There has also been confirmatory evidence from the results of site-directed mutagenesis experiments [46,71-75] in certain cases which point to the appropriate regions of suspected substrate binding or access to the heme.

There is some controversy regarding the position of the substrate access channel in cytochrome P-450_{CAM}. However, the results of molecular dynamics simulations on the crystal structure indicate that flexibility in the F-G interhelical turn (Thr¹⁸⁵-Thr¹⁹²) which contains residues that form a hydrophobic cap over the bound substrate, gives rise to thermal fluctuations suggestive of a point of ingress. This region includes Arg¹⁸⁶ which ion-pairs with Asp²⁵¹ at the oxygen binding site. A 50-ps molecular dynamics simulation indicates that this latter residue in the distal helix is also capable of undergoing thermal motion at 300 K, which may point to its involvement in the P-450 reaction mechanism (Lewis, unpublished results). These findings are in agreement with those reported by Poulos and co-workers [22,23], who also note that Tyr⁹⁶, which hydrogen-bonds to camphor, shows a propensity for thermal fluctuation. An alternative substrate access channel, close to the heme, would appear to be a less favourable candidate due to a high proportion of polar residues, which would not be likely to transport a hydrophobic molecule such as camphor. However, if camphor is hydrated during transport, then polar residues at the access channel could participate in substrate entry to the active site by assisting in a favourable entropy change due to desolvation at the binding site.

Presumably, a conformational change occurs in the apoprotein when the substrate enters its access channel as suggested by molecular dynamics simulations and from the differences between substrate-bound and substrate-free cytochrome P-450_{CAM} crystal structures [22,23]. A conformational change also signals the reductase to transfer the first electron, once oxygen has entered the active site after the substrate has desolvated the heme environment. Exclusion of water molecules gives rise to a favourable entropy change which accounts for all of the free energy necessary to drive the catalytic process leading to hydroxylation of the substrate [76]. Calculations based on the changes in redox potential of cytochrome P-450_{CAM} which occur on substrate binding, suggest that the loss of between 10 and 16 water molecules is sufficient to explain an entropically-driven mechanism, and molecular modelling tends to support this finding (Lewis, unpublished results). It is interesting to speculate on the possibility of molecular recognition sites on the enzyme surface which facilitate movement of the correct substrate to its access channel entrance. Some circumstantial evidence for this has been reported by Gotoh et al. [77] and Mornon et al. [78] for steroid metabolizing cytochromes P-450. Apparently, there are possible steroid binding sites common to cytochrome P-450_{SCC} and prostatic steroid binding proteins which have been suggested from sequence alignments [79]. These putative binding-site regions are far removed from the heme pocket but are close to the likely substrate access channel entrance. Analogous regions also exist between

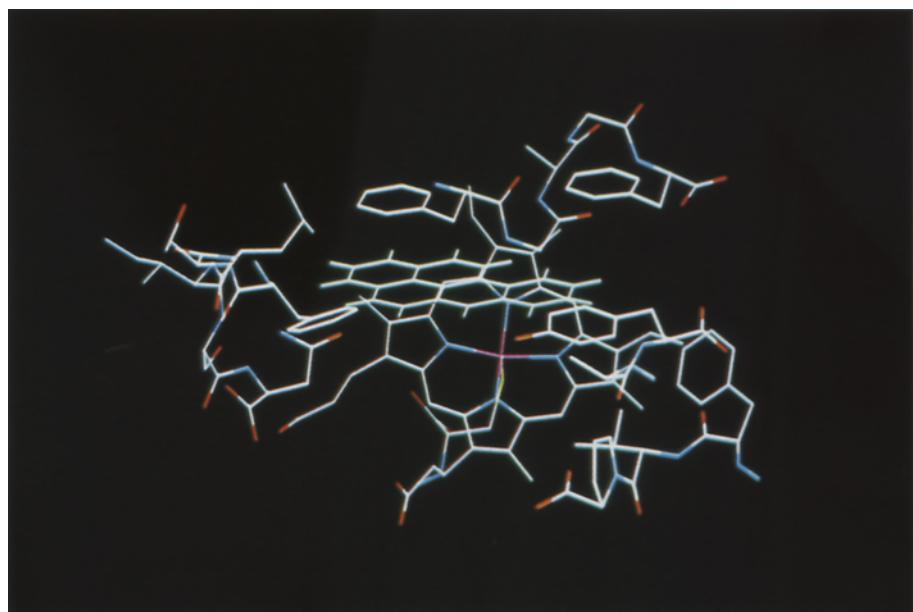


Fig. 5. A postulated active-site model for P-450 I based on the P-450_{CAM} active site. Atoms are coloured by type and the heme iron is shown in purple. The substrate benzo[a]pyrene is displayed in pale blue.

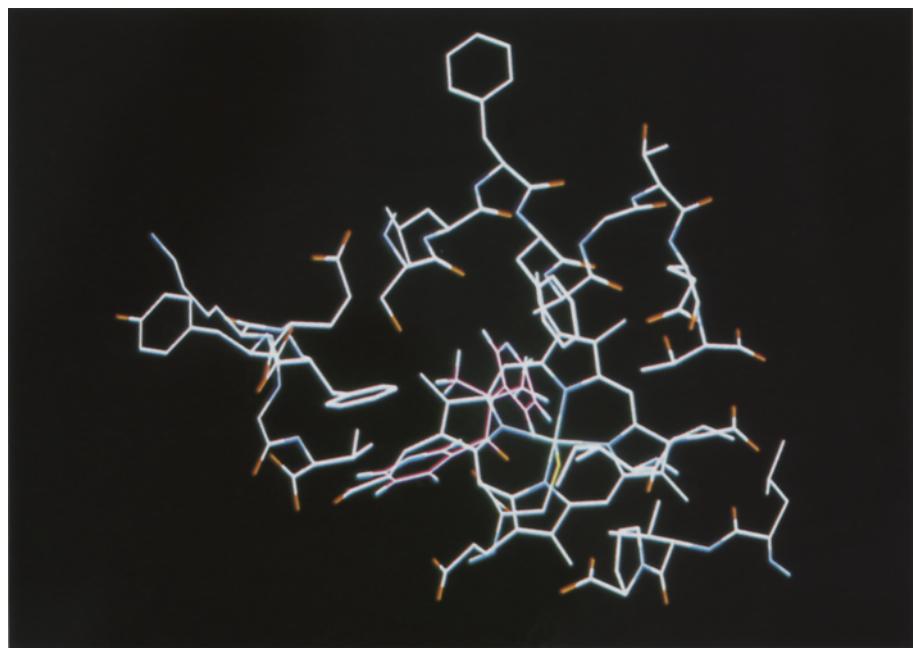


Fig. 6. A postulated active-site model for P-450 IIB based on the P-450_{CAM} active site. Atoms are coloured by type and the heme iron is shown in pale green. The substrate/inhibitor metyrapone is displayed with its carbon atoms shown in mauve.

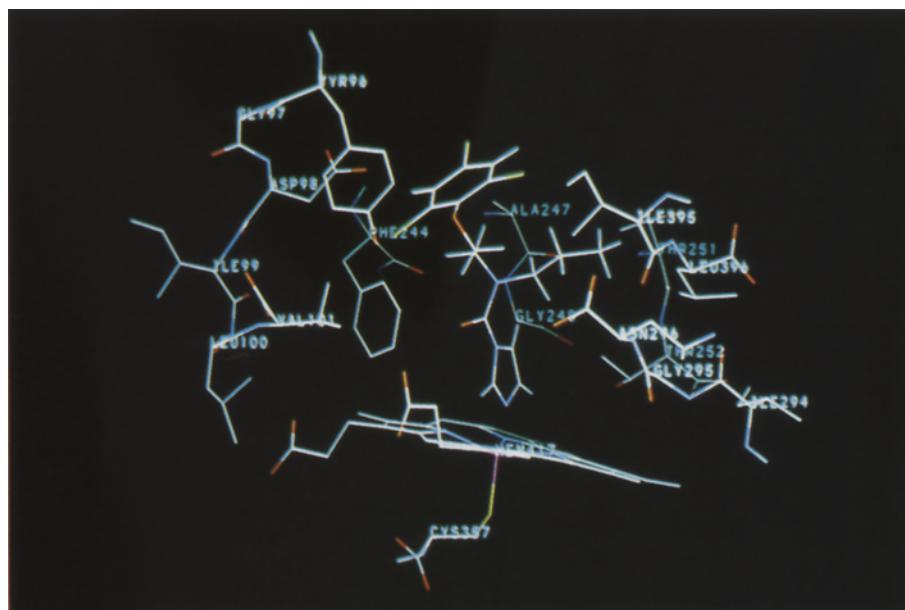


Fig. 7. A postulated active-site model for P-450 III based on the P-450_{CAM} active site. Atoms are coloured by type and residues are labelled by amino acid type using the P-450_{CAM} numbering system. The inhibitor prochloraz is shown with added hydrogens in pale blue.

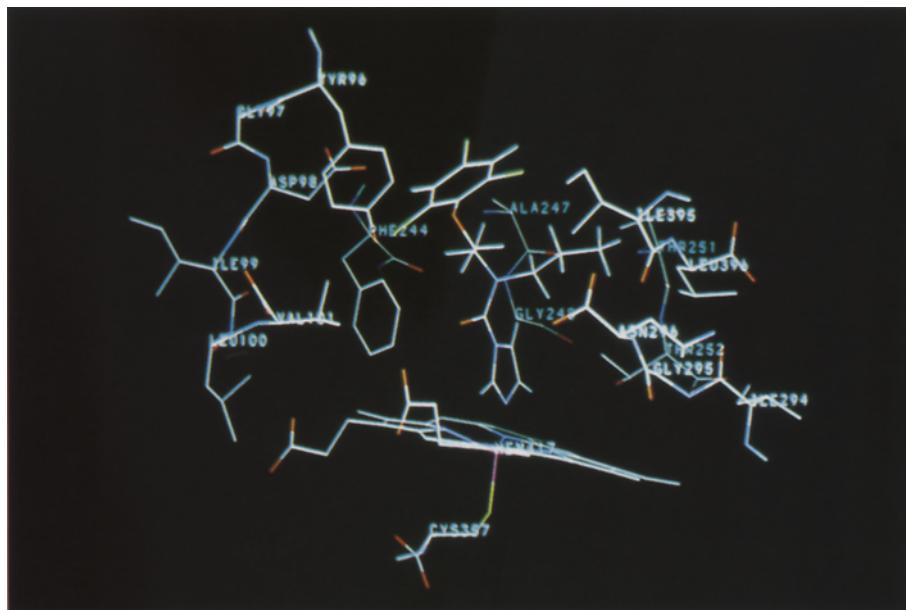


Fig. 8. A postulated active-site model for P-450 XIX based on the P-450_{CAM} active site. Atoms are coloured by type and residues labelled by amino acid type using the P-450_{CAM} numbering system. The inhibitor 4-hydroxyandrostenedione is shown docked in the active site with carbon atoms coloured in green.

steroid receptors and cytochrome P-450_{17a}, which are difficult to rationalize in terms of heme binding domains but could represent surface recognition sites.

The postulated substrates access channel in cytochrome P-450_{CAM} is lined with hydrophobic residues which can admit water molecules but will become desolvated when the camphor substrate enters. At the heme end of this channel there is a tyrosine residue (96 in the P-450_{CAM} sequence) which participates in substrate binding by the formation of a hydrogen bond between the phenolic hydroxy group and the carbonyl oxygen atom on camphor [22]. This key residue assists in the alignment of the substrate relative to the heme iron such that stereospecific oxygenation can occur. Although the majority of the amino acid residues which contact the camphor molecule are situated in the vicinity of a unique hydrogen-bonded turn in the I helix distal to the heme, Tyr⁹⁶ occurs at the start of a turn linking helices B' and C [22]. This region is also variable in the proposed sequence alignments and, as such, can account for the differences in substrate specificities between the cytochromes P-450 currently under investigation. The remainder of the hydrophobic cavity which comprises the active site in cytochrome P-450_{CAM} is made up by two small stretches of β -sheet containing hydrophobic residues Val²⁹⁵ and Ile³⁹⁵ [22]. The sequence alignments presented in Table 3 show that, of these two residues, the former is fairly well conserved throughout the 16 P-450 proteins. Taken together, the 4 areas corresponding to Tyr⁹⁶-Thr¹⁰¹, Gly²⁴³-Thr²⁵², Leu²⁹⁴-Ala²⁹⁶ and Ile³⁹⁵-Val³⁹⁶ can be used to identify substrate binding sites.

Substitution of the appropriate amino acid residues in P-450_{CAM} for those of eukaryotic cytochromes P-450 shows the topology of the putative substrate binding sites and Figs. 5-8 give these results for 4 cytochrome P-450 proteins matched in this work. In each case, a known substrate or inhibitor is docked within the active site and optimally aligned to indicate the degree of complementarity between substrate and binding-site residues. Although no geometry optimization has been used, it can be seen that the relevant substrate or inhibitor both fits the appropriate active site of the cytochrome P-450 for which it is specific and that certain key amino acid residues are likely to cooperatively assist in the binding process.

These initial findings will be employed to undertake more extensive modelling studies involving generation of the entire protein structure of various cytochromes P-450 which will be subsequently used to investigate the modes of binding of specific substrates in more detail.

REFERENCES

- 1 Guengerich, F.P. (Ed.), Mammalian Cytochromes P-450, Vol. 1 and 2, CRC Press, Boca Raton, FL, 1987.
- 2 Park, B.K. and Kitteringham, N.R., Progr. Drug Metab., 11 (1988) 1.
- 3 Guengerich, F.P., Cancer Res., 48 (1988) 2946.
- 4 Guengerich, F.P. and Macdonald, T.L., FASEB J., 4 (1990) 2453.
- 5 Ryan, D.E. and Levin, W., Pharmacol. Ther., 45 (1990) 153.
- 6 Nebert, D.W. and Gonzalez, F.J., Trends Pharmacut. Sci., 6 (1985) 160.
- 7 Nebert, D.W. and Gonzalez, F.J., Annu. Rev. Biochem., 56 (1987) 945.
- 8 Nebert, D.W., Nelson, D.R., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B., Levin, W., Phillips, I.R., Sato, R. and Waterman, M.R., DNA, 8 (1989) 1.
- 9 Gonzalez, F.J., Pharmacol. Rev., 40 (1989) 243.
- 10 Gonzalez, F.J. and Nebert, D.W., Trends Genet., 6 (1990) 182.
- 11 Gonzalez, F.J., Pharmacol. Ther., 45 (1990) 1.
- 12 Nelson, D.R. and Strobel, H.W., Mol. Biol. Evol., 4 (1987) 572.
- 13 Ayrton, A.D., Ph.D. Thesis, University of Surrey, U.K., 1989.

- 14 Murray, M. and Reidy, G.F., *Pharmacol. Rev.*, 42 (1990) 85.
- 15 Lewis, D.F.V., Ioannides, C. and Parke, D.V., *ATLA*, 18 (1990) 91.
- 16 Lewis, D.F.V., *Frontiers in Biotransformation*, 8 (1991) in press.
- 17 Ioannides, C. and Parke, D.V., *Drug Metab. Rev.*, 22 (1990) 1.
- 18 Parke, A.L., Ioannides, C., Lewis, D.F.V. and Parke, D.V., *Inflammopharmacology*, 1 (1991) 3.
- 19 VandenBossche, H., In Berg, D and Plemel, M. (Eds.), *Sterol Biosynthesis Inhibitors*, Ellis-Horwood, Chichester, 1988, p. 79.
- 20 Van Wauve, J.P. and Janssen, P.A.J., *J. Med. Chem.*, 32 (1989) 2231.
- 21 Poulos, T.L., Finzel, B.C., Gunsalus, I.C., Wagner, G.C. and Kraut, J., *J. Biol. Chem.*, 260 (1985) 16122.
- 22 Poulos, T.L., Finzel, B.C. and Howard, A.J., *Biochemistry*, 25 (1986) 5314.
- 23 Poulos, T.L., Finzel, B.C. and Howard, A.J., *J. Mol. Biol.*, 195 (1987) 687.
- 24 Poulos, T.L., In Miners, J.O. et al. (Eds.), *Microsomes and Drug Oxidations*, Taylor and Francis, London, 1988, pp. 159–167.
- 25 Poulos, T.L., *Pharmaceut. Res.*, 5 (1988) 67.
- 26 Poulos, T.L., *Nature*, 339 (1989) 580.
- 27 Raag, R. and Poulos, T.L., *Biochemistry*, 28 (1989) 917.
- 28 Guengerich, F.P., *Frontiers in Biotransformation*, 1 (1989) 101.
- 29 Means, G.D., Mahendroo, M.S., Corbin, C.J., Mathis, J.M., Powell, F.E., Meadelson, C.R. and Simpson, E.R., *J. Biol. Chem.*, 264 (1989) 19385.
- 30 Beaune, I.P.H., Umbenhauer, D.R., Bork, R.W., Lloyd, R.S. and Guengerich, F.P., *Proc. Natl. Acad. Sci. U.S.A.*, 83 (1986) 8064.
- 31 Hardwick, J.P., Song, B.-J., Huberman, E. and Gonzalez, F.J., *J. Biol. Chem.*, 262 (1987) 801.
- 32 Namiki, M., Kitamura, M., Buczko, E. and Dufan, M.L., *Biochem. Biophys. Res. Commun.*, 157 (1988) 705.
- 33 Khani, S.C., Zaphiropoulos, P.G., Fujita, V.S., Porter, T.D., Koop, D.R. and Coon, M.J., *Proc. Natl. Acad. Sci. U.S.A.*, 84 (1987) 638.
- 34 Song, B.-J., Gelboin, H.V., Park, S.-S., Yang, C.S. and Gonzalez, F.J., *J. Biol. Chem.*, 261 (1986) 16689.
- 35 Matsubara, S., Yamamoto, S., Sogawa, K., Yokotani, N., Fujii-Kuriyama, Y., Haniu, M., Shively, J.E., Gotoh, O., Kusunose, E. and Kusunose, M., *J. Biol. Chem.*, 262 (1987) 13366.
- 36 Kalb, V.F., Woods, C.W., Turi, T.G., Dey, C.R., Sutter, T.R. and Loper, J.C., *DNA*, 6 (1987) 529.
- 37 Corbin, C.J., Graham-Lorence, S., McPhaul, M., Mason, J.I., Mendelson, C.R. and Simpson, E.R., *Proc. Natl. Acad. Sci. U.S.A.*, 85 (1988) 8948.
- 38 Gotoh, O. and Fujii-Kuriyama, Y., *Frontiers in Biotransformation*, 1 (1989) 195.
- 39 Nelson, D.R. and Strobel, H.W., *J. Biol. Chem.*, 263 (1988) 6038.
- 40 Nelson, D.R. and Strobel, H.W., *Biochemistry*, 28 (1989) 656.
- 41 Edwards, R.J., Murray, B.P., Boobis, A.R. and Davies, D.S., *Biochemistry*, 28 (1989) 3762.
- 42 Black, J.D. and Coon, M.J., In Ortiz de Montellano, P.R. (Ed.), *Cytochrome P-450*, Plenum, New York, 1986, pp. 161–216.
- 43 Lewis, D.F.V., Ioannides, C. and Parke, D.V., *Chem.-Biol. Interact.*, 64 (1987) 39.
- 44 Zvelebil, M.J.J.M., Wolf, C.R. and Sternberg, M.J.E., *Prot. Eng.*, 4 (1991) 271.
- 45 Lewis, D.F.V., *Chem.-Biol. Interact.*, 62 (1987) 271.
- 46 Ishida, N., Aoyama, Y., Hatanaka, R., Oyama, Y., Imajo, S., Ishiguro, M., Oshima, T., Nakazato, H., Noguchi, T., Maitra, U.S., Mohan, V.P., Sprinson, D.B. and Yoshida, Y., *Biochem. Biophys. Res. Commun.*, 155 (1988) 317.
- 47 Morris, G.M., *J. Mol. Graphics*, 6 (1988) 135.
- 48 Morris, G.M., Richards, W.G. and Baldwin, B.C., In *Proceedings of the 8th International Symposium on Microsomes and Drug Oxidations*, Stockholm, 1990, Abstr. No. 281, p. 136.
- 49 Ferenczy, G.G. and Morris, G.M., *J. Mol. Graphics*, 7 (1989) 206.
- 50 Laughton, C.A., Neidle, S., Zvelebil, M.J.J.M. and Sternberg, M.J.E., *Biochem. Biophys. Res. Commun.*, 171 (1990) 1160.
- 51 Graham-Lorence, S., Mendelson, C.R. and Simpson, E.R., In *Proceedings of the 8th International Symposium on Microsomes and Drug Oxidations*, Stockholm, 1990, Abstr. No. 265, p. 128.
- 52 Moereels, H., De Bie, L. and Tollenaere, J.P., *J. Comput.-Aided Mol. Design*, 4 (1990) 131.
- 53 Kyte, J. and Doolittle, R.F., *J. Mol. Biol.*, 157 (1982) 105.
- 54 Garnier, J., Osguthorpe, D.J. and Robson, B., *J. Mol. Biol.*, 120 (1978) 97.

- 55 Chou, P.Y. and Fasman, G.D., *Annu. Rev. Biochem.*, 47 (1978) 251.
- 56 Sternberg, M.J.E., *Anti-Cancer Drug Design*, 1 (1986) 169.
- 57 Stayton, P.S., Poulos, T.L. and Sligar, S.G., *Biochemistry*, 28 (1989) 8201.
- 58 Dailey, H.A. and Strittmatter, P., *J. Biol. Chem.*, 254 (1979) 5388.
- 59 Simondsen, R.P., Weber, P.C., Salemme, F.R. and Tollin, G., *Biochemistry*, 21 (1982) 6366.
- 60 Matthew, J.B., Weber, P.C., Salemme, F.R. and Richards, F.M., *Nature*, 301 (1983) 169.
- 61 Schwarze, W., Bernhardt, R., Jänig, G.-R. and Ruckpaul, K., *Biochem. Biophys. Res. Commun.*, 113 (1983) 353.
- 62 Bernhardt, R., Makower, A., Jänig, G.-R. and Ruckpaul, K., *Biochim. Biophys. Acta*, 785 (1984) 186.
- 63 Bernhardt, R., Pommerening, K. and Ruckpaul, K., *Biochem. Int.*, 14 (1987) 823.
- 64 Bernhardt, R., Kraft, R., Otto, A. and Ruckpaul, K., *Biomed. Biochim. Acta*, 7 (1988) 581.
- 65 Bernhardt, R., Stiel, H. and Ruckpaul, K., *Biochem. Biophys. Res. Commun.*, 163 (1989) 1282.
- 66 Szczesna-Skorupa, E., Browne, N., Mead, D. and Kemper, B., *Proc. Natl. Acad. Sci. U.S.A.*, 85 (1988) 738.
- 67 Sato, T., Sakaguchi, M., Mihara, K. and Omura, T. In *Proceedings of the 8th International Symposium on Microsomes and Drug Oxidations*, Stockholm, 1990, Abstr. No. 287, p. 139.
- 68 Ruckpaul, K., Rein, H. and Blanck, J., *Frontiers in Biotransformation*, 1 (1989) 1.
- 69 Archakov, A.I. and Bachanova, G.I., *Cytochrome P-450 and Active Oxygen*, Taylor & Francis, London, 1990.
- 70 Shimada, H., Makino, R., Imai, M., Watanabe, Y., Matsushima-Hibiya, Y., Koga, H., Horiuchi, T. and Ishimura, Y., In *Proceedings of the 8th International Symposium on Microsomes and Drug Oxidations*, Stockholm, 1990, Abstract No. 288, p. 140.
- 71 Furuya, H., Shimizu, T., Hatano, M. and Fujii-Kuriyama, Y., *Biochem. Biophys. Res. Commun.*, 160 (1989) 669.
- 72 Furuya, H., Shimizu, T., Hirano, K., Hatano, M., Fujii-Kuriyama, Y., Raag, R. and Poulos, T.L., *Biochemistry*, 28 (1989) 6848.
- 73 Lindberg, R.L.P. and Negishi, M., *Nature*, 339 (1989) 632.
- 74 Atkins, W.M. and Sligar, S.G., *J. Am. Chem. Soc.*, 111 (1989) 2715.
- 75 Sligar, S.G., Egeberg, K.D., Sage, J.T., Morikis, D. and Champion, P.M., *J. Am. Chem. Soc.*, 109 (1987) 7896.
- 76 Griffin, B.W. and Peterson, J.A., *Biochemistry*, 11 (1972) 4740.
- 77 Gotoh, O., Tagashira, Y., Morohashi, K. and Fuji-Kuriyama, Y., *FEBS Lett.*, 188 (1985) 8.
- 78 Mornon, J.P., Bissery, V., Gaboriaud, C., Thomas, A., Ojasoo, T. and Raynaud, J.-P., *J. Steroid Biochem.*, 34 (1989) 355.
- 79 Unger, B.P., Gunsalus, I.C. and Sligar, S.G., *J. Biol. Chem.*, 261 (1986) 1158.
- 80 Morohashi, K., Sogawa, K., Omura, T. and Fujii-Kuriyama, Y., *J. Biochem.*, 101 (1987a) 879.
- 81 Fujii-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K. and Muramatsu, M., *Proc. Natl. Acad. Sci. U.S.A.*, 79 (1982) 2793.
- 82 Kawajiri, K., Watamabe, J., Gotoh, O., Tagashira, Y., Sogawa, K. and Fujii-Kuriyama, Y., *Eur. J. Biochem.*, 159 (1986) 219.
- 83 Kawajiri, K., Gotoh, O., Sogawa, K., Tagashira, Y., Muramatsu, M. and Fujii-Kuriyama, Y., *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 1649.
- 84 Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M. and Fujii-Kuriyama, Y., *Proc. Natl. Acad. Sci. U.S.A.*, 80 (1983) 3958.
- 85 Gonzalez, F.J., Nebert, D.W., Hardwick, J.P. and Kasper, C.B., *J. Biol. Chem.*, 260 (1985) 7435.
- 86 Chaplin, D.D., Galbraith, L.J., Seidman, J.G., White, P.C. and Parker, K.L., *Proc. Natl. Acad. Sci. U.S.A.*, 83 (1986) 9601.
- 87 Morohashi, K., Yoshioka, H., Gotoh, O., Okada, Y., Yamamoto, K., Miyata, T., Sogawa, K., Fujii-Kuriyama, Y. and Omura, T., *J. Biochem.*, 102 (1987b) 559.