REFERENCES
Ayad, S., Chambers, C. A., Berry, L., Shuttleworth, C. A., & Grant, M. E. (1985) *Biochem. J. 230*, 465-474.
Chu, M.-L., Mann, K., Deutzmann, R., Pribula-Conway, D., Hsu-Chen, C.-C., Bernard, M. P., & Timpl, R. (1987) *Eur. J. Biochem. 168*, 309-317.
Chu, M.-L., Conway, D., Pan, T.-C., Baldwin, C., Mann, K., Deutzmann, R., & Timpl, R. (1988) *J. Biol. Chem. 263*, 18601-18606.
Chung, E., Rhodes, R. K., & Miller, E. J. (1976) *Biochem. Biophys. Res. Commun. 71*, 1167-1174.
Engel, J., Furthmayr, H., Odermatt, E., Von Der Mark, H., Aumailley, M., Fleischmajer, R., & Timpl, R. (1985) *Ann. N.Y. Acad. Sci. 460*, 25-37.
Engvall, E., Hessle, H., & Klier, G. (1986) *J. Cell Biol. 102*, 703-710.
Friedman, M., Krull, L. H., & Cavins, F. (1970) *J. Biol. Chem. 245*, 3868-3871.
Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E., & Engel, J. (1983) *Biochem. J. 211*, 303-311.
Gibson, M. A., & Cleary, E. G. (1985) *J. Biol. Chem. 260*, 11149-11159.
Jander, R., Rauterberg, J., Voss, B., & von Bassewitz, D. B. (1981) *Eur. J. Biochem. 114*, 17-25.
Jander, R., Rauterberg, J., & Glanville, R. W. (1983) *Eur. J. Biochem. 133*, 39-46.
Jander, R., Troyer, D., & Rauterberg, J. (1984) *Biochemistry 23*, 3675-3681.
Laemmli, U. K. (1970) *Nature (London) 227*, 680-685.
Linsenmayer, T. F., Mentzer, A., Irwin, M. H., Waldrep, N. K., & Mayne, R. (1986) *Exp. Cell Res. 165*, 518-529.
Lunstrum, G. P., Kuo, H.-J., Rosenbaum, L. M., Keene, D. R., Glanville, R. W., Sakai, L. Y., & Burgeson, R. E. (1987) *J. Biol. Chem. 262*, 13706-13712.
Morris, N. P., Keene, D. R., Glanville, R. W., Bentz, H., & Burgeson, R. E. (1986) *J. Biol. Chem. 261*, 5638-5644.
Odermatt, E., Risteli, J., Van Delden, V., & Timpl, R. (1983) *Biochem. J. 211*, 295-302.
Rauterberg, J., Jander, R., Voss, B., Furthmayr, H., & Glanville, R. (1982) in *New Trends in Basement Membrane Research* (Kühn, K., Schöne, H. H., & Timpl, R., Eds.) pp 107-119, Raven Press, New York.
Timpl, R., & Engel, J. (1987) in *Structure and Function of Collagen Types* (Mayne, R., & Burgeson, B., Eds.) pp 105-143, Academic Press, New York.
Trueb, B., & Winterhalter, H. (1986) *EMBO J. 11*, 2815-2819.

Identification and Location of α-Helices in Mammalian Cytochromes P450

Robert J. Edwards,* Bernard P. Murray, Alan R. Boobis, and Donald S. Davies
Department of Clinical Pharmacology, Royal Postgraduate Medical School, London W12 0NN, U.K.
Received October 12, 1988; Revised Manuscript Received January 23, 1989

ABSTRACT: A model of the α-helical structure of mammalian cytochromes P450 is proposed. The location and sequence of α-helices in mammalian cytochromes P450 were predicted from their homology with those of cytochrome $P450_{cam}$, and these sequences were generally confirmed as helical in nature by using a secondary structure prediction method. These analyses were applied to 26 sequences in 6 gene families of cytochrome P450. Mammalian cytochromes P450 consist of approximately 100 amino acid residues more than cytochrome $P450_{cam}$. This difference was accounted for by three major areas of insertion: (1) at the N-terminus, (2) between helices C and D and between helices D and E, and (3) between helices J and K. Insertion 1 has been suggested by others as a membrane anchoring sequence, but the apparent insertions at 2 and 3 are novel observations; it is suggested that they may be involved in the binding of cytochrome P450 reductase. Only the mitochondrial cytochrome P450 family appeared to show a major variation from this pattern, as insertion 2 was absent, replaced by an insertion between helices G and H and between helices H and I. This may reflect the difference in electron donor proteins that bind to members of this cytochrome P450 family. Other than these differences the model of mammalian cytochromes P450 proposed maintains the general structure of cytochrome $P450_{cam}$ as determined by its α-helical composition.

Cytochromes P450 form a superfamily of hemoprotein isoenzymes that play a central role in the disposition of a wide variety of endogenous compounds and xenobiotics, many of which are toxic or carcinogenic (Black & Coon, 1987; Nebert & Gonzalez, 1987). In eukaryotes these enzymes are found bound to the membrane of the endoplasmic reticulum or mitochondrion.

In a recent review, the primary structures of 65 isoenzymes, from several species, were catalogued (Nebert et al., 1987). However, the three-dimensional structure of only one isoenzyme has been solved, that of cytochrome $P450_{cam}$, a soluble, bacterial protein (Poulos et al., 1985, 1987). Although the crystallization of bovine cytochrome $P450_{scc}$ has recently been reported (Iwamoto et al., 1988), the structure has yet to be solved. The primary structure of cytochrome $P450_{cam}$ shows only weak homology with the published sequences of the eukaryotic isoenzymes, except in several highly conserved regions, most notably in the area of the heme thiolate ligand (Poulos et al., 1987). Further structural information on the eukaryotic isoenzymes is difficult to obtain due to their membrane-bound nature.

In view of the wealth of information available on the sequences of these hemoproteins, it is pertinent to suggest ways in which this might be used to hypothesize upon the structure of the mammalian cytochromes P450 and to develop models upon which experimental work can proceed rationally. A number of such models have previously been proposed, using methods based upon either the local hydrophobicity of the

* Address correspondence to this author.

0006-2960/89/0428-3762$01.50/0 © 1989 American Chemical Society

enzymes (Hopp & Woods, 1981, 1983) or predictions of secondary structure (Chou & Fasman, 1978). However, the predictive value of these methods taken in isolation is extremely variable, being particularly inaccurate in the region of prosthetic groups. The resultant predictions led to the suggestion of a highly transmembranous arrangement for cytochrome P450 (Peterson et al., 1976; Nebert, 1979; Heinemann & Ozols, 1982; Tarr et al., 1983; Black & Coon, 1987). This structure is radically different from that of cytochrome $P450_{cam}$. Alternative models, in which a largely hydrophilic enzyme is attached to the membrane by a relatively small portion of its structure, have also been proposed (Ingelman-Sundburg, 1986; Nelson & Strobel, 1988). The hypothesis tested here is that since all of the cytochromes P450, including cytochrome $P450_{cam}$, have similar functions, they will have essentially the same three-dimensional structure. This has already been found to be the case for a number of proteins sharing similar functions (Richardson, 1981; Bajaj & Blundell, 1984; Blundell et al., 1987).

Cytochrome $P450_{cam}$ possesses a high α-helical content (211 of 414 residues), with a much lower proportion of other secondary structural features, and the enzyme may be roughly divided into two domains, one of which contains the majority of the helices. These helices thus define the basic topology of much of the enzyme. They are intimately involved in heme and substrate binding and should thus be a conserved feature in all members of the cytochrome P450 superfamily. The hypothesis that all cytochromes P450 share structural similarity was tested on the basis of homology between the primary sequence of these helices with that of corresponding areas in the mammalian isoenzymes. This was allied with the results of an independent prediction of secondary structure. The sequences of the isoenzymes from several of the gene families of mammalian cytochromes P450 were compared with the sequences of the helices of cytochrome $P450_{cam}$. Potential helices in the mammalian isoenzymes were identified on the basis of homology with the helices of cytochrome $P450_{cam}$ and the likelihood of the residues within these areas being present within an α-helix. The fact that an alignment of areas in the mammalian isoenzymes with cytochrome $P450_{cam}$ helices was possible by use of this algorithm demonstrates that there is significant structural similarity in the tertiary structure between mammalian cytochromes P450 and cytochrome $P450_{cam}$. Given the preponderance of α-helical content of cytochrome $P450_{cam}$, this should enable sufficient information to be obtained by using structural homology to form the basis for a model of the tertiary structure of the entire enzyme. This analysis also reveals areas of primary structure for which there are no orthologous counterparts in cytochrome $P450_{cam}$. Presumably, these regions are involved in functions of the mammalian cytochromes P450 quite different from those of cytochrome $P450_{cam}$, such as membrane anchoring, cytochrome P450 reductase or cytochrome b5 binding, and substrate binding.

Experimental Procedures

Location of α-Helices in Mammalian Cytochromes P450 by Sequence Homology with Cytochrome $P450_{cam}$ α-Helices. The derived amino acid sequences of members of cytochrome P450 gene families I, II, III, XI, XVII, and XXI (Nebert et al., 1987) were selected and aligned for maximum homology within each family. A computer algorithm for this was not used as the introduction of only a few spaces was necessary (Figure 1). Sequence homologies in families III and XI were determined by comparing the predicted amino acid residues at each aligned sequence position. The sum of homologous residues was calculated, and the result is expressed as a percentage of the aligned sequence length. In the other gene families, which contained more than two members, the same calculation was performed for all combinations of pairs of sequences, and the mean of these results is presented as a consensus homology for the aligned sequences in each gene family.

The rationale for the choice of the sequences was to use information on all known members of a gene family when available. Where the sequence of only one member of a family was available, the sequences of the orthologue in several species were combined (families XVII and XXI). Where the sequence of more than one member was available, sequences from just one species were combined (families II, III, and XI), unless full information was available for all members in more than one species and then all of this information was combined (family I). The algorithm described below required that a minimum of two sequences in a family be combined to produce predictions.

The primary structures of each of the α-helices of cytochrome $P450_{cam}$, as defined from the X-ray crystallographic structure (Poulos et al., 1987; Unger et al., 1986), were compared with each of the aligned families at every position throughout the length of the aligned sequences. The degree of sequence homology was assessed by summing the number of amino acid residues in the aligned sequences that matched, by identity and corresponding position, the amino acid residues in each cytochrome $P450_{cam}$ helix. The resulting data were ordered with respect to the areas yielding the greatest number of homologous amino acid residues and a mismatch penalty score was derived, zero for the area with the highest number of matches and increasing sequentially with decreasing matches. Sequence areas that scored a lower number of homologous amino acid residues than that of the 20th best were not considered in further calculations. Tied values were given equal penalty scores. These procedures were performed on an IBM-PC-compatible microcomputer using programs written in BASIC and analyzed by using a commerically available spreadsheet package.

The locations of regions orthologous to the cytochrome $P450_{cam}$ helices were assigned in each family. The following rules were applied: (a) helices were maintained in the order A to L; (b) overlapping of the helices by more than one amino acid was not permitted (the C-terminal residue of one helix can be the N-terminal residue of the next helix; this occurs in cytochrome $P450_{cam}$ between the I and J helices); (c) overlapping the heme thiolate ligand was not permitted. In practice, these instructions were followed: (i) The maximum number of helices in order (A to L) at positions giving zero penalty for each helix was assigned; if overlapping occurred, then the helix selected was the one with the lower mismatch score. (ii) Unassigned helices were located to give the lowest penalty scores but not such that they overlapped those already assigned. Those helices still unassigned incurred a penalty value of 20, and the total penalty of the prediction was then calculated. (iii) These unassigned helices were located, when possible, by displacing the previously assigned helices. These rearrangements were permitted only if they resulted in a reduction in the total penalty while maintaining the order of the helices.

De Novo Prediction of α-Helices. Prediction of secondary structure was performed by using the method of Garnier and co-workers (Garnier et al., 1978), as this method results in the unambiguous assignment of a single conformational state (helix, sheet, turn, or random coil) to each residue. The

Table I: Relative Homology of Amino Acid Sequences in Various Cytochrome P450 Families to the Helices of Cytochrome $P450_{cam}$[a]

helix	gene family						
	I	II	III	XI	XVII	XXI	sum
A	0	11	0	8	11	9	39
B	3	3	13	4	1	1	25
B′	20[b]	18	2	1	20[b]	2	63
C	2	0	1	0	2	2	7
D	0	0	1	9	0	0	10
E	0	0	0	8	4	5	17
F	9	7	20[b]	1	2	10	49
G	1	0	6	3	1	3	14
H	20[b]	2	0	0	0	20[b]	42
I	0	3	0	0	20[b]	0	23
J	0	3	0	1	5	4	13
K	2	2	0	0	0	0	4
L	0	0	2	1	0	0	3

[a] A value of zero represents the most homologous sequence and increasing values represent less homologous sequences. The positions of the helices are illustrated in Figure 1, and their choice is explained in the text. [b] Helix not assigned.

method was verified for the particular case of cytochromes P450 by the attempted prediction of the helical areas in cytochrome $P450_{cam}$ and comparing this with the correct helical/nonhelical configuration of each of the residues. When no assumptions were made concerning the overall proportion of α-helix in the secondary structure (all decision constants set to zero), 66% of all residues were correctly assigned. If, however, it was assumed that the proportion of α-helix was 20–50%, then the accuracy of prediction increased to 67%. Raising the assumed helicity to 50–100% (most appropriate for the 51% content of cytochrome $P450_{cam}$) resulted in a further slight improvement of prediction to 68%. The results of this final prediction are illustrated in Figure 2. Failures in the prediction consisted of false positives in the areas Nt–A, K–L, and L–Ct and an inability to predict the existence of the H helix. This latter failing is most probably due to suppression of helix prediction by the strong prediction of the β5 sheet and the 215–218 loop, as a discrete helix peak of 26 centinats is present in the appropriate location. Another characteristic feature is the bifurcation of the E, I, and L helices.

This algorithm thus shows a large improvement over that of Chou and Fasman (1978), which has an accuracy of 51% for the prediction of the structure of cytochrome $P450_{cam}$ (Haniu et al., 1982). The actual proportion of α-helical content in cytochrome $P450_{cam}$ is 51%, and the results of circular dichroism experiments on rat cytochromes P450c and P450d indicate helical contents of 50% and 60%, respectively (Haniu et al., 1986). Thus, the Garnier algorithm was used with decision constants for a helical content of 50–100% throughout. Individual predictions were then performed by using the primary sequences of the mammalian isoenzymes. The proportion of helix-assigned residues at each aligned sequence position within a family was then determined, and the results were compared with the predictions based upon homology.

Results

The positions of the helices in the families of mammalian cytochromes P450 predicted by homology with cytochrome $P450_{cam}$ are shown in Figure 1. Table I shows the total penalty values associated with each of the predicted helices. A penalty value of zero, representing the position in the aligned sequences yielding the greatest homology, was obtained for 37% of the predicted helices. Furthermore, 74% of the predicted helices were in the top 1% (penalty ≤ 4) and 86% were in the top 2% (penalty ≤ 9) of all possible sequence positions.

Table II: Frequency of Residues Predicted To Be in an α-Helical Conformation As Determined by the Algorithm of Garnier et al. (1978) in Areas Known To Be α-Helical in Cytochrome $P450_{cam}$ and for Areas Predicted To Be Helical, on the Basis of Homology, in the Mammalian Cytochromes P450[a]

helix	$P450_{cam}$	gene family						
		I	II	III	XI	XVII	XXI	av
A	90	1	0	40	50	81	30	33.7
B	36	47	91	0	46	100	85	61.5
B′	100	*b*	34	63	31	*b*	*b*	42.7
C	57	52	41	95	38	46	65	56.2
D	74	100	91	53	47	95	37	70.5
E	43	11	73	57	50	76	10	46.2
F	62	48	38	*b*	58	59	100	60.6
G	52	58	76	96	57	84	55	71.0
H	0	*b*	84	100	81	63	*b*	82.0
I	59	28	59	43	66	*b*	56	50.4
J	90	71	90	100	100	65	60	81.0
K	92	74	74	85	85	72	69	76.5
L	70	79	89	70	88	87	80	82.2

[a] Average values refer only to the mammalian enzymes. A comparison of each of these averages with the corresponding values obtained for cytochrome $P450_{cam}$ helices shows that the algorithm yields similar results in the majority of cases. All values are expressed as percentages of the total number of residues tested by the algorithm in each actual or predicted helix. [b] Helix not assigned.

This shows that there is, in general, good homology between the helical region in cytochrome $P450_{cam}$ and the mammalian cytochromes P450. However, in 8% of the cases no assignment of helices could be made.

An independent examination of the theoretical secondary structure of the areas predicted to be helices in the mammalian cytochromes P450 by use of the algorithm of Garnier et al. (1978) showed that, in general, they had high helix-forming ability (approximately 77%). It was also apparent that the pattern of predicted helicity throughout the regions orthologous to the E and I helices is similar to that predicted for the actual helices in cytochrome $P450_{cam}$ (Figure 3, compared with Figure 2), with a distinct bifurcation of both helices apparent. The difference in the frequency of predicted helical residues between cytochrome $P450_{cam}$ and the mammalian cytochromes P450 was less than 10% for helices C, D, E, F, I, and J and less than 20% for helices G, K, and L. Only the small helices A, B, B′, and H were markedly different (Table II).

The difference between the lengths of the primary structures of the mammalian cytochromes P450 and $P450_{cam}$ is approximately 100 residues. The positions of the predicted helices in the aligned mammalian cytochromes P450 were compared with the positions of the cytochrome $P450_{cam}$ helices to determine the distribution of this difference (Table III). Most of this could be accounted for by four major sites of sequence insertion in the mammalian cytochromes P450 (Nt–A, C–D, D–E, and J–K). The average length of insertion in these four regions ranged from 12.7 to 24.5 residues, compared with −3.7 to 7.0 in the other 10 areas.

The predicted secondary structure of the areas corresponding to these insertions was examined within individual families, but no consistent pattern was found, apart from the insert in the Nt–A region. This consisted of a strongly predicted α-helix of length 20–30 residues located at or near the amino terminus of the members of all six protein families, including the cytochrome P450 XI family from mitochondria. No analogous region was found in cytochrome $P450_{cam}$.

Discussion

Members of six families of mammalian cytochrome P450 have been analyzed for their α-helical content by comparison

Table III: Differences between the Predicted Interhelical Distances of Mammalian Cytochromes P450 and Those of Cytochrome $P450_{cam}$[a]

area	gene family						av
	I	II	III	XI	XVII	XXI	
Nt–A	34	5	57	1	11	28	22.7
A–B	−8	14	−10	28	0	−21	0.5
B–B′	1	16	9	4	−10	−9	1.8
B′–C	1	−9	−9	6	−10	18	−0.5
C–D	9	−1	22	4	24	18	12.7
D–E	42	21	1	1	21	21	17.8
E–F	−3	11	−5	−2	8	7	2.7
F–G	2	−6	−5	−2	−1	−5	−2.8
G–H	−2	7	6	13	2	−7	3.2
H–I	−2	2	−8	24	6	−7	2.5
I–J	0	3	3	0	6	8	3.3
J–K	19	15	67	18	15	13	24.5
K–L	14	11	−42	9	14	7	2.2
L–Ct	6	−2	4	2	9	10	4.8
sum	113	87	90	106	95	81	95.3

[a] These values were determined by counting the number of residues between each of the predicted helices (and adjacent N- and C-termini) and subtracting the number of residues in these regions in cytochrome $P450_{cam}$. When helices were not predicted in the mammalian cytochromes P450, the same calculation was performed but by using the adjacent predicted helices and the result divided evenly between the two interhelical areas. The values thus represent the positions of possible insertion or deletion in the mammalian isoenzymes compared to cytochrome $P450_{cam}$. The average values should represent the predicted structure of an idealized mammalian cytochrome P450.

of their primary sequences with that of cytochrome $P450_{cam}$ and by prediction of α-helical content by the algorithm of Garnier et al. (1978). The results enable a model for the helical structure of mammalian cytochromes P450 to be suggested, in which the helices equivalent to the cytochrome $P450_{cam}$ helices B, C, D, E, G, I, J, K, and L are conserved (Table I). These form the bulk of the helical-rich domain in cytochrome $P450_{cam}$ [Figure 1 in Poulos et al. (1987)]. The helices that appeared to be most poorly assigned, as measured by their high aggregate penalty scores, were A, B′, F, and H. Of these, only F and H are in the helical-rich domain. The H helix was well-defined in families other than I and XXI, and in these cases the helix could not be assigned because the adjacent helices, which appear to be well-defined, are too close together to accommodate this helix. The H helix may not, therefore, be conserved in all families of cytochrome P450. The precise assignment of the α-helices of mammalian cytochromes P450 is obviously not possible by use of these techniques, and this is the cause, at least in part, of the difference in the interhelical residue distances between the families of cytochromes P450 (Table III). However, any artifactual increase (or decrease) in the interhelical residue distance due to a degree of misalignment of a helix will be countered by an opposite change in the interhelical residue difference on the other side of the misaligned helix. Thus, over the span of two to three helices the interhelical distances should be similar among cytochrome P450 families. In general, this was found to be true (Table III). However, a number of discrepancies were found that could not be accounted for in this way.

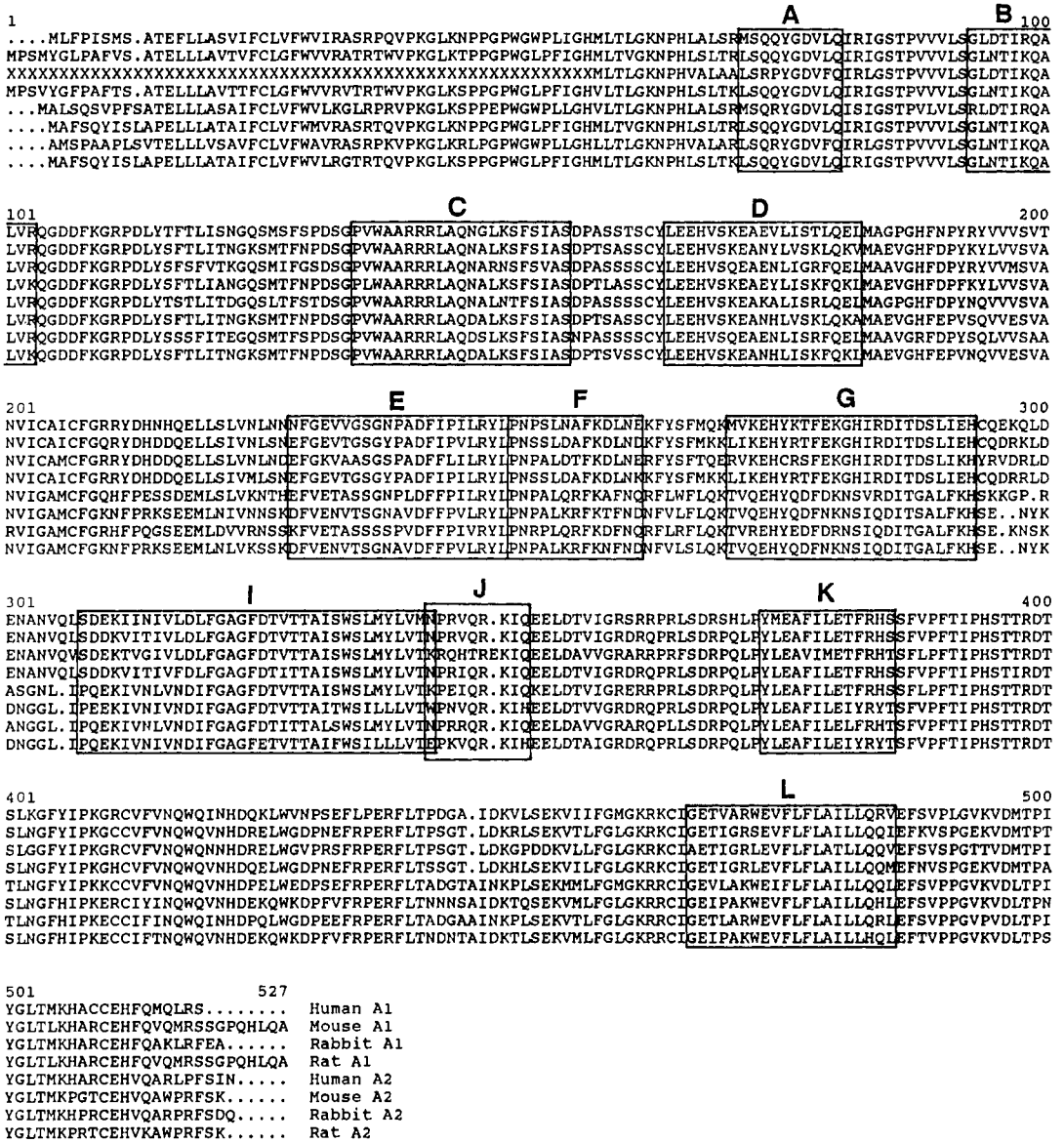
A general pattern of three areas of insertion was found that accounts for the difference in the length between cytochrome $P450_{cam}$ and the mammalian cytochromes P450 (Table III). These were (1) at the N-terminus (+22.7 residues), (2) between helices C and D and between helices D and E (+30.5 residues), and (3) between helices J and K (+24.5 residues). Observation 1 is in agreement with the work of Sakaguchi et al. (1987) and Szczesna-Skorupa et al. (1988), who have shown, respectively, that the N-termini of rabbit cytochrome P450 isoenzyme LM6 and the gene product of P450IIC2 (isoenzyme PBc2 or k) resemble signal peptides that are not cleaved during insertion of the cytochrome P450 into the membrane. Observations 2 and 3 appear to be novel and may represent binding sites for cytochrome P450 reductase or cytochrome *b*5.

Nelson and Strobel (1988) conclude that eukaryotic cytochromes P450 are anchored in the membrane by a pair of helices forming a transmembranous hairpin with the N-terminus at the cytoplasmic surface. This conclusion was based upon the fact that the N-terminal residue of rabbit cytochrome P450 LM2 can be labeled with fluorescein isothiocyanate (Bernhardt et al., 1983). In fact, this study was performed upon the solubilized protein, and thus no information can be obtained on the normal location of this residue, as a different pattern of labeling could well result with the membrane-bound enzyme. Additionally, examination of the amino acid composition of the proposed second transmembranous region [S2, Figure 1 in Nelson and Strobel (1988)] reveals a high proportion (22%) of proline residues that are classically considered as "helix breakers" (Levitt, 1978). This region is thus unlikely to be helical in character and, in fact, the series of relatively rigid bends in the peptide backbone caused by these conserved proline residues may be the means by which the protein is orientated so that the heme is held parallel to the plane of the membrane in accordance with electron paramagnetic resonance studies (Rich et al., 1979).

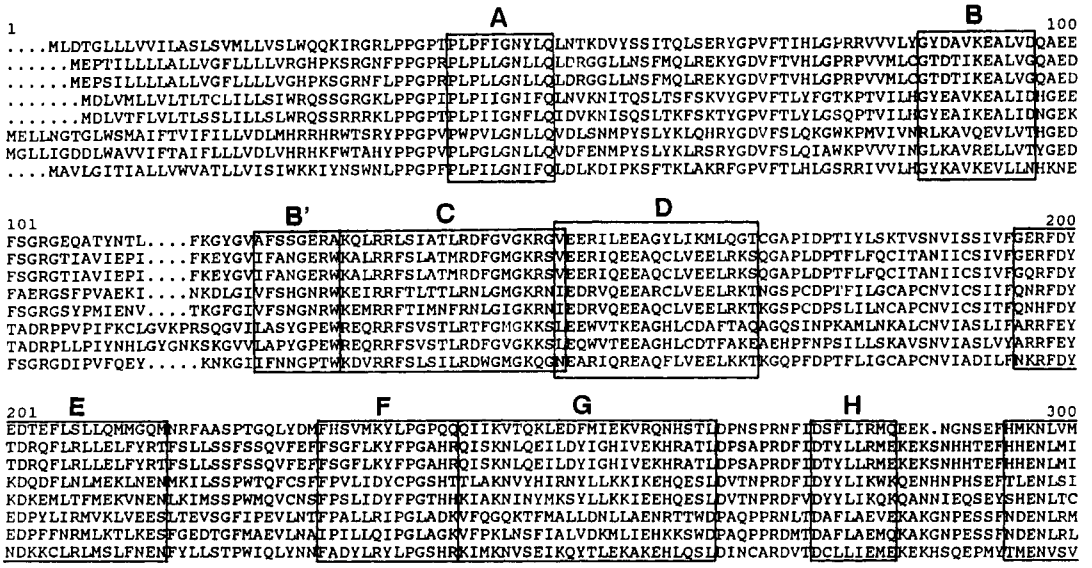
Some anomalous results concerning the pattern and length of insertions were apparent among individual families. Negative values (Table III), indicating a shortening of the mammalian cytochrome P450 compared with cytochrome $P450_{cam}$, were expected to be small. However, in family III there was a large apparent deletion, between the helices K and L, of 42 residues, most probably due to incorrect assignment of the K helix. The resultant predicted sequence is quite dissimilar to the predicted K helices for other mammalian cytochrome P450 families (Figure 1). In addition, there was an exceptionally large insertion of 67 residues between the J and K helices. A second possible position for the K helix at position 376, determined by homology with the predicted helices in the other mammalian families, would give insertion values of 35 and −10 residues for areas between the J and K and K and L helices, respectively, and this location is supported by the secondary structure prediction. This individual result appears to have distorted the average values calculated in Table III at the region of the K helix. The correct overall interpretation may be an extended area between helices J and K and between helices K and L.

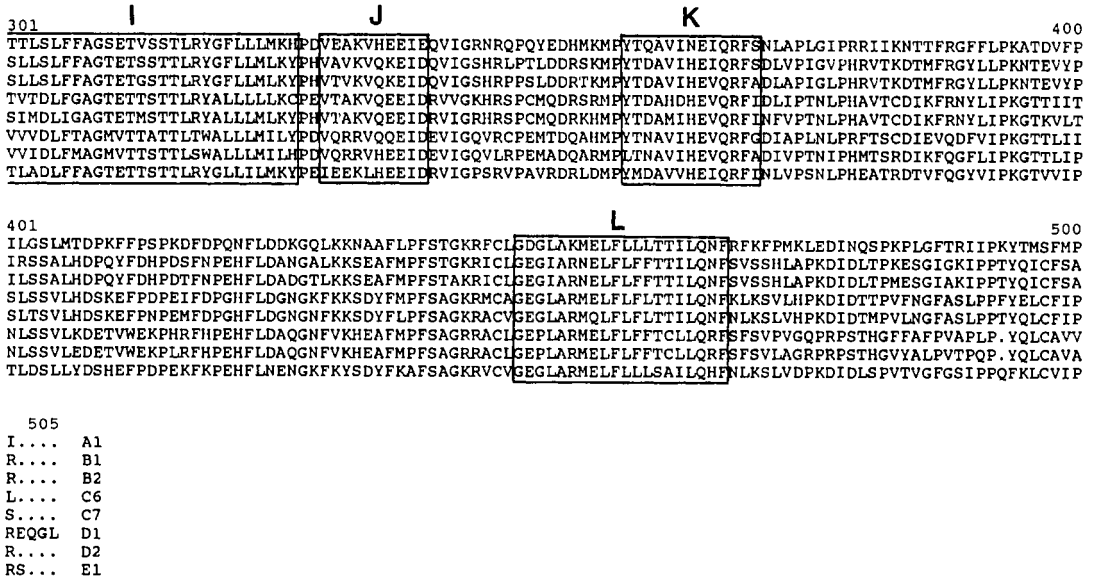
One discrepancy is unique to family XI, the only mitochondrial family, where the 2 insertion was absent, but appeared to be replaced by insertions between the G and H helices and the H and I helices (+37 residues). This may be related to differences between family XI cytochromes P450 and cytochromes P450 of the other families studied here with respect to binding of electron donor proteins, adrenodoxin in the case of the mitochondrial isoenzymes and cytochrome P450 reductase or cytochrome b5 in the case of the other isoenzymes. Therefore, the proposed G–I extended region in family XI may be associated with adrenodoxin binding, and the lack of a C–D extended region in this family and its presence in the reductase-utilizing isoenzymes may indicate that this region is involved in reductase binding.

FAMILY I



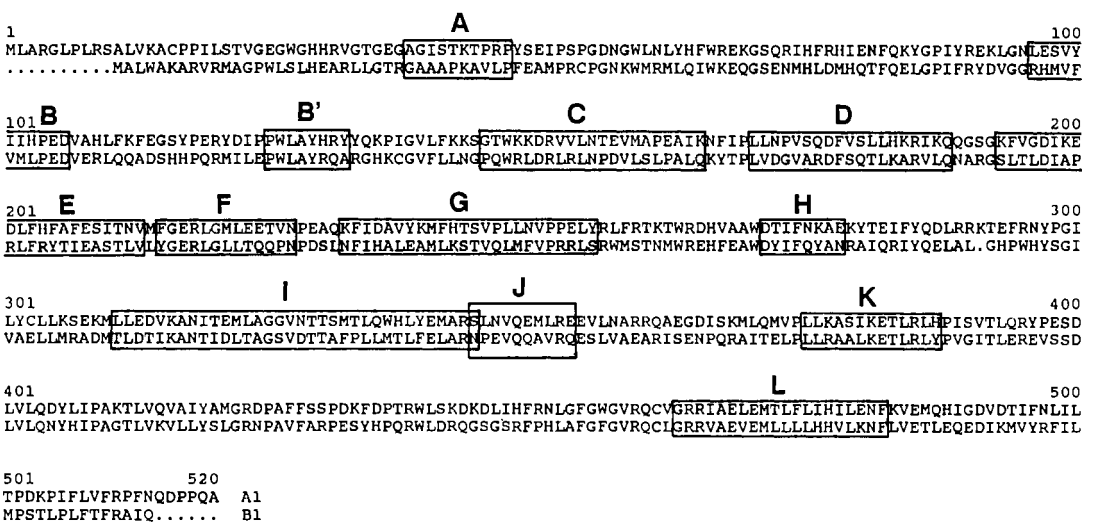
FAMILY II





FAMILY III

FAMILY XI



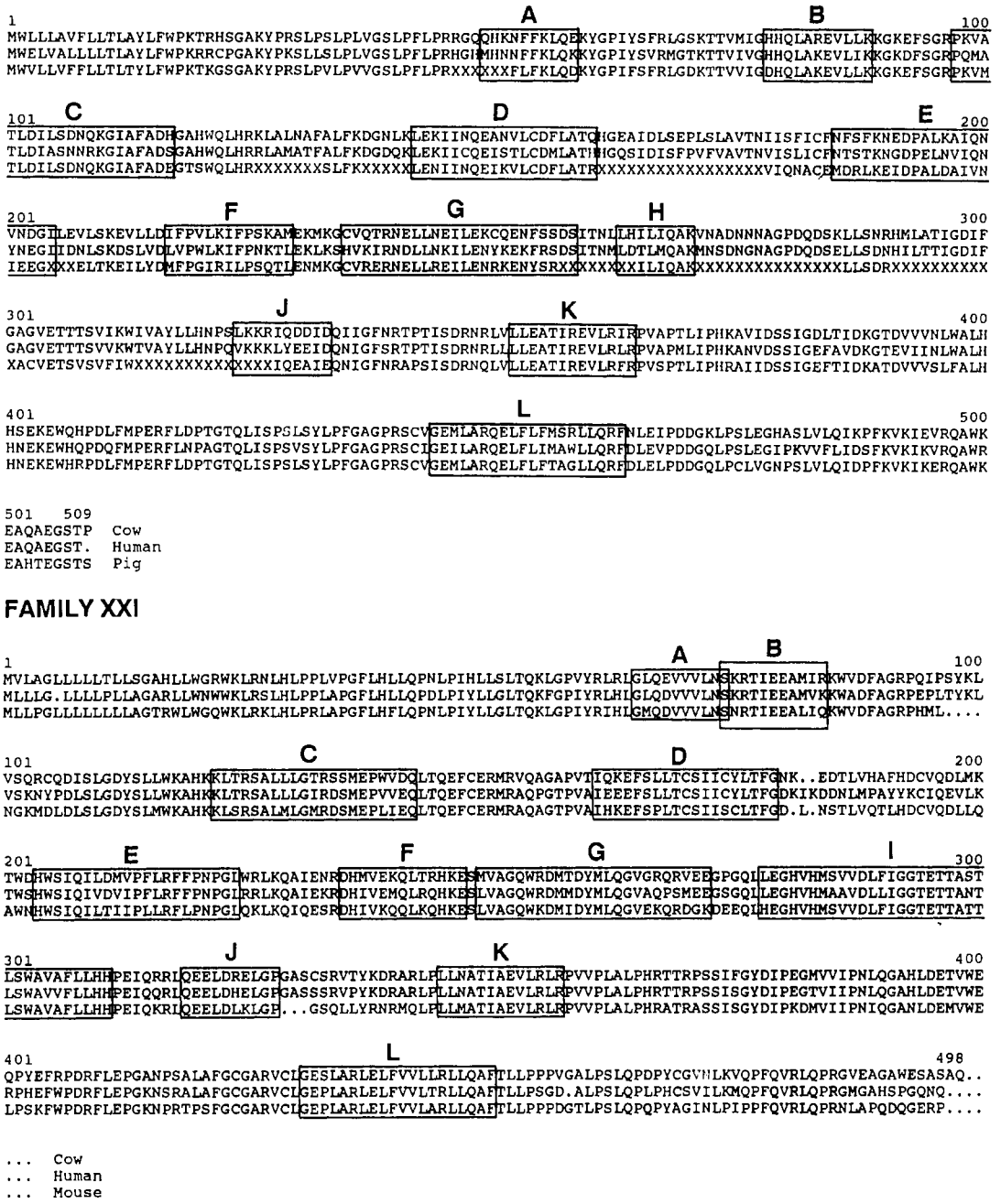


FIGURE 1: Aligned amino acid sequences of members of six cytochrome P450 gene families with the regions predicted to be helical, on the basis of limited homology, as described in the text. The derived sequences were obtained from the following genes: (Family I) Information from two genes, P450IA1 and P450IA2, from four species was used, i.e., the gene products from the P450IA1 gene of mouse (Kimura et al., 1984), human (Jaiswal et al., 1985), rabbit (Okino et al., 1985), and rat (Sogawa et al., 1984) plus the products of the P450IA2 gene from the same species (Kimura et al., 1984; Quattrochi et al., 1986; Ozols et al., 1986; Sogawa et al., 1985, respectively); consensus homology is 72%. (Family II) Derived sequences from the rat genes, P450IIA (Nagata et al., 1987), P450IIB1 (Suwa et al., 1985), P450IIB2 (Kumar et al., 1983), P450IIC6 (Gonzalez et al., 1986a), P450IIC7 (Gonzalez et al., 1986a), P450IID1 (Gonzalez et al., 1987), P450IID2 (Gonzalez et al., 1987), and P450IIE1 (Song et al., 1986), were used; consensus homology is 47%. (Family III) Both rat genes P450IIIA1 (Gonzalez et al., 1985) and P450IIIA2 (Gonzalez et al., 1986b), were used; sequence homology is 89%. (Family XI) Immature bovine forms of P450XIA1 (Morohashi et al., 1984) and P450XIB1 (Chua et al., 1987) were used; sequence homology is 35%. (Family XVII) The P450XVII gene from human (Chung et al., 1987), pig (Chung et al., 1987), and cow (Zuber et al., 1986) were used; consensus homology is 72%. (Family XXI) Genuine genes P450XXIA or P450XXIB, as appropriate, from mouse (Chaplin et al., 1986), human (White et al., 1986), and cow (Chung et al., 1986) were used; consensus homology is 73%. The sequences were aligned for maximum homology within each family by insertion of dots. The positions of areas predicted, by the homology method described in the text, to be equivalent to α-helices in cytochrome $P450_{cam}$ are shown as labeled boxes. The sequences of the cytochrome $P450_{cam}$ α-helices were obtained from Unger et al. (1986) and Poulos et al. (1987) and were GVQEAWAVLQ (A), RGQLIREAYED (B), PREAGEAY (B′), PEQRQFRALANQVVGMPVVDK (C), LENRIQELACSLIESLRPQ (D), NFTEDYAEPFPIRIFMLLAGL (E), DIPHLKYLTDQMT (F), MTFAEAKEALYDYLIPIIEQRRQK (G), DAISIVAN (H), TSDEAKRMCGLLLVGGLDTVVNFLSFSMEFLAKS (I), SPEHRQELIE (J), RIPAACEELLRRF (K), and GQHLARREIIVTLKEWLTRI (L).

Alignment of the complete amino acid sequences of cytochrome $P450_{cam}$ with the mammalian cytochromes P450 is complicated by the large difference in the lengths of the sequences and their relatively low degree of homology. However, this has been attempted by Black and Coon (1986) using four family II sequences from two species and, more ambitiously,

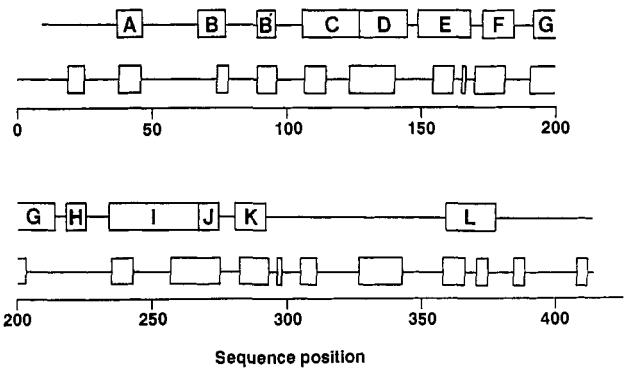


FIGURE 2: Comparison of the locations predicted to be α-helices in $P450_{cam}$ by the algorithm of Garnier et al. (1978) (lower lines) and the true locations as defined by the X-ray crystallographic structure (Poulos et al., 1987) (upper lines).

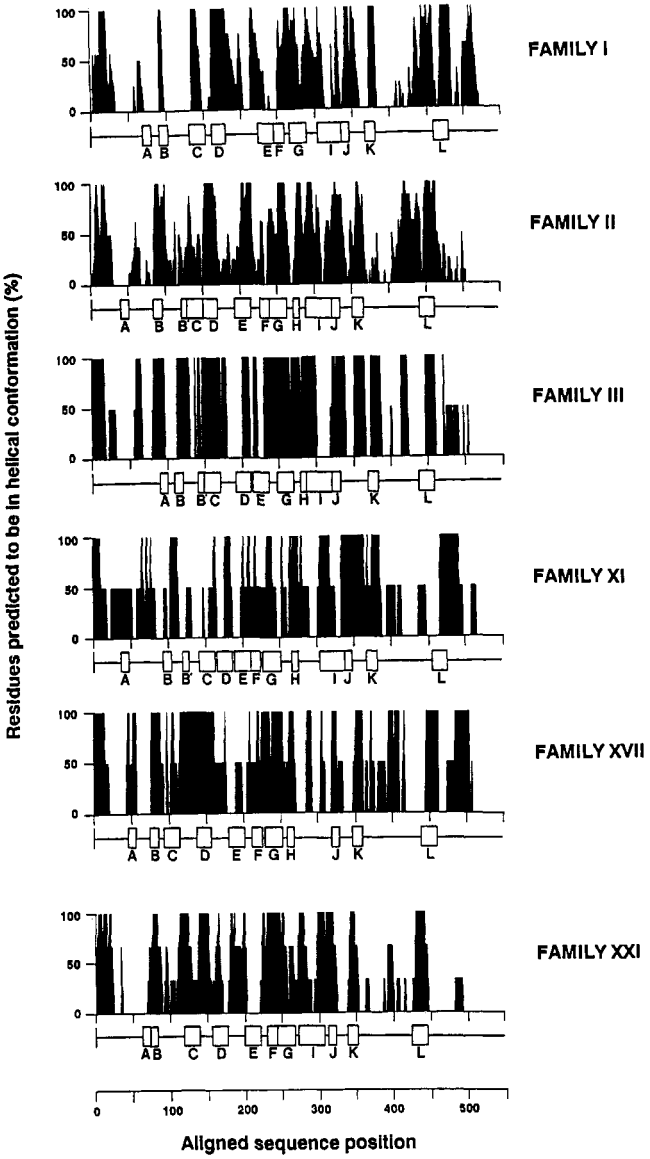


FIGURE 3: Comparison of the predicted helical content of six cytochrome P450 families determined by using the method of Garnier et al. (1978), depicted as filled histograms with the location of regions containing sequences homologous to those of the α-helices of cytochrome $P450_{cam}$ (lettered boxes).

by Nelson and Strobel (1988) using 34 sequences from a variety of families. The algorithm used by Nelson and Strobel allows for the insertion of long gaps in the sequences because each may be produced by a single mutational event (Gotoh, 1982). To maximize the degree of alignment between the sequences, large gaps were frequently inserted in the cytochrome $P450_{cam}$ sequences. As there was no consideration of secondary structure, the position of these gaps included the positions of known helices. This means that the maintenance of secondary structure, essential to retain the overall structure and function of a cytochrome P450, was sacrificed in favor of apparent homology. It is more likely that the secondary structure, and hence the overall structure, are maintained and that the areas accounting for the differences in sequence lengths are limited to regions possessing little or no secondary structure.

The approach used here was relatively naive in concept, in that only exact matches of amino acid residues were accepted in comparisons of homology. Evolution from a common ancestral cytochrome P450 sequence may have taken many forms including mutations to a chemically similar residue or to a dissimilar residue if that residue is not important for structure or function. Deletions and insertions within the helices and terminal extensions would occur only if structure and function are maintained and are thus less likely in helices. The results of the secondary structure prediction appear to support this view. Therefore, such changes were considered to be unlikely and have not been examined here. It is of interest that the N-terminus of the products of the mitochondrial cytochrome P450 XI family is very similar to that of the other mammalian cytochrome P450 families, despite the fact that this is cleaved during insertion into the mitochondrial membrane. This might suggest the evolution of divergent function from the same structural feature in the mammalian cytochromes P450.

The results of the present study support the model of cytochrome P450 as a largely hydrophilic protein associated with the membrane by an N-terminal hydrophobic tail that is likely to be helical in structure. The N-terminus could either be buried in or at the interior surface of the membrane. This model allows for lateral movement of the cytochrome P450 in the membrane and free association with electron-transport components.

Registry No. Cytochrome P450, 9035-51-2.

REFERENCES

Bajaj, M., & Blundell, T. (1984) *Annu. Rev. Biophys. Bioeng. 13*, 453–492.

Bernhardt, R., Dao, N. T. N., Stiel, H., Schwarze, W., Friedrich, J., Janig, G.-R., & Ruckpaul, K. (1983) *Biochim. Biophys. Acta 745*, 140–148.

Black, S. D., & Coon, M. J. (1986) in *Cytochrome P-450* (Ortiz de Montellano, P. R., Ed.) pp 161–216, Plenum Press, New York.

Black, S. D., & Coon, M. J. (1987) *Adv. Enzymol. Relat. Areas Mol. Biol. 60*, 35–87.

Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., & Thornton, J. M (1987) *Nature (London) 326*, 347–352.

Chaplin, D. D., Galbraith, L. J., Seidman, J. G., White, P. C., & Parker, K. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9601–9605.

Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.

Chua, S. C., Szabo, P., Vitek, A., Grzeschik, K.-H., John, M., & White, P. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7193–7197.

Chung, B.-C., Matteson, K. J., & Miller, W. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4243–4247.

Chung, B.-C., Picado-Leonard, J., Haniu, M., Bienkowski, M., Hall, P. F., Shively, J. E., & Miller, W. L. (1987) *Proc.*

Natl. Acad. Sci. U.S.A. 84, 407–411.
Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.
Gonzalez, F. J., Nebert, D. W., Hardwick, J. P., & Kasper, C. B. (1985) *J. Biol. Chem. 260*, 7435–7441.
Gonzalez, F. J., Kimura, S., Song, B.-J., Pastewka, J., Gelboin, H. V., & Hardwick, J. P. (1986a) *J. Biol. Chem. 261*, 10667–10672.
Gonzalez, F. J., Song, B.-J., & Hardwick, J. P. (1986b) *Mol. Cell. Biol. 6*, 2969–2976.
Gonzalez, F. J., Matsunaga, T., Nagata, K., Meyer, U. A., Nebert, D. W., Pastewka, J., Kozak, C. A., Gillette, J., Gelboin, H. V., & Hardwick, J. P. (1987) *DNA 6*, 149–161.
Gotoh, O. (1982) *J. Mol. Biol. 162*, 705–708.
Haniu, M., Armes, L. G., Yasunobu, K. T., Shastry, B. A., & Gunsalus, I. C. (1982) *J. Biol. Chem. 257*, 12664–12671.
Haniu, M., Ryan, D. E., Levin, W., & Shively, J. E. (1986) *Arch. Biochem. Biophys. 244*, 323–337.
Heinemann, F. S., & Ozols, J. (1982) *J. Biol. Chem. 257*, 14988–14999.
Hopp, T. P., & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3824–3828.
Hopp, T. P., & Woods, K. R. (1983) *Mol. Immunol. 20*, 483–489.
Ingelman-Sundburg, M. (1986) in *Cytochrome P-450* (Ortiz de Montellano, P. R., Ed.) pp 119–160, Plenum Press, New York.
Iwamoto, Y., Tsubaki, M., Hiwatashi, A., & Ichikawa, Y. (1988) *FEBS Lett. 233*, 31–36.
Jaiswal, A. K., Gonzalez, F. J., & Nebert, D. W. (1985) *Science 228*, 80–83.
Kimura, S., Gonzalez, F. J., & Nebert, D. W. (1984) *J. Biol. Chem. 259*, 10705–10713.
Kumar, A., Raphael, C., & Adesnik, M. (1983) *J. Biol. Chem. 258*, 11280–11284.
Levitt, M. (1978) *Biochemistry 17*, 4277–4285.
Morohashi, K., Fujii-Kuriyama, Y., Okada, Y., Sogawa, K., Hirose, T., Inayama, S., & Omura, T. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4647–4651.
Nagata, K., Matsunaga, T., Gillette, J., Gelboin, H. V., & Gonzalez, F. J (1987) *J. Biol. Chem. 262*, 2787–2793.
Nebert, D. W. (1979) *Mol. Cell. Biochem. 27*, 27–46.
Nebert, D. W., 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., & Waterman, M. R. (1987) *DNA 6*, 1–11.
Nelson, D. R., & Strobel, H. W. (1988) *J. Biol. Chem. 263*, 6038–6050.
Okino, S. T., Quattrochi, L. C., Barnes, H. J., Osanto, S., Griffin, K. J., Johnson, E. F., & Tukey, R. H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 5310–5314.
Ozols, J. (1986) *J. Biol. Chem. 261*, 3965–3979.
Peterson, J. A., Ebel, R. E., O'Keeffe, D. H., Matsubara, T., & Estabrook, R. W. (1976) *J. Biol. Chem. 251*, 4010–4016.
Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C., & Kraut, J. (1985) *J. Biol. Chem. 260*, 16122–16130.
Poulos, T. L., Finzel, B. C., & Howard, A. J. (1987) *J. Mol. Biol. 195*, 687–700.
Quattrochi, L. C., Pendurthi, U. R., Okino, S. T., Potenza, C., & Tukey, R. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6731–6735.
Rich, P. R., Tiede, D. M., & Bonner, W. D. (1979) *Biochim. Biophys. Acta 546*, 307–315.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Sakaguchi, M., Mihara, K., & Sato, R. (1987) *EMBO J. 6*, 2425–2431.
Sogawa, K., Gotoh, O., Kawajiri, K., & Fujii-Kuriyama, Y. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5066–5070.
Sogawa, K., Gotoh, O., Kawajiri, K., Harada, T., & Fujii-Kuriyama, Y. (1985) *J. Biol. Chem. 260*, 5026–5032.
Song, B. J., Gelboin, H. V., Park, S.-S., Yang, C. S., & Gonzalez, F. J. (1986) *J. Biol. Chem. 261*, 16689–16697.
Suwa, Y., Mizukami, Y., Sogawa, K., & Fujii-Kuriyama, Y. (1985) *J. Biol. Chem. 260*, 7980–7984.
Szczesna-Skorupa, E., Browne, N., Mead, D., & Kemper, B. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 738–742.
Tarr, G. E., Black, S. D., Fujita, V. S., & Coon, M. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6552–6556.
Unger, B. P., Gunsalus, I. C., & Sligar, S. G. (1986) *J. Biol. Chem. 261*, 1158–1163.
White, P. C., New, M. I., & Dupont, B. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5111–5115.
Zuber, M. X., John, M. E., Okamura, T., Simpson, E. R., & Waterman, M. R. (1986) *J. Biol. Chem. 261*, 2475–2482.