Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7221–7225, October 1988
Biochemistry

Proteins from eight eukaryotic cytochrome P-450 families share a segmented region of sequence similarity

(cytochrome P450/consensus sequence/evolution/P-450cam/superfamily)

VERNON F. KALB AND J. C. LOPER

Department of Microbiology and Molecular Genetics, University of Cincinnati College of Medicine, 231 Bethesda Avenue, Cincinnati, OH 45267-0524

Communicated by M. J. Coon, July 5, 1988 (received for review March 30, 1988)

ABSTRACT **Proteins from eight eukaryotic families in the cytochrome P-450 superfamily share one region of sequence similarity. This region begins 275–310 amino acids from the amino terminus of each P-450, continues for ≈170 residues, and ends 35–50 amino acids before the carboxyl terminus. The region can be divided into four domains of sequence similarity, each possessing its own pattern of invariant, conserved, and variable amino acids. The four domains are 56, 20, 59, and 28 residues long and are connected by three shorter segments of limited sequence similarity. The number of residues in these short segments varies with the P-450 protein but ranges from 0 to 20 residues. Consensus sequences based on these similarities can be used to determine whether the sequence of an unidentified peptide resembles that expected for a P-450. Sequence similarities between proteins sometimes reflect constraints imposed by the requirements of a common function. The fourth domain of the P-450s, for example, contains an invariant cysteine that provides the axial thiolate ligand to the heme iron. Other relationships between the four domains and P-450 function can be examined by *in vitro* mutagenic procedures that alter the conserved amino acids or modify the distance between domains.**

The P-450 cytochromes are components in monooxygenase systems that catalyze a wide variety of oxidative reactions in prokaryotes and eukaryotes. These reactions include steps in the synthesis and degradation of such compounds as cholesterol, steroid hormones, and prostaglandins. P-450 proteins are also involved in the metabolism of drugs and in the activation and inactivation of carcinogens. A recent analysis of >60 P-450s from eukaryotes and one prokaryote led to the organization of the known P-450s into 10 different families with a total of 15 subfamilies. Amino acid sequences within a P-450 family are >36% similar, while sequences within a subfamily are ≈70% similar (1). Although sequence relatedness between P-450 families is low, these families are still grouped into one P-450 superfamily according to standard criteria (2).

The eukaryotic branch of the cytochrome P-450 superfamily arose >1000 million years ago (1, 3) and sequence similarities among these ancient families might identify conserved domains of structure or function. Since complete amino acid sequences had been published for proteins from eight eukaryotic P-450 families, we used a representative sequence from each family in a series of sequence comparisons. A multiple alignment of the eight sequences revealed one region of similarity shared by all. Sequence similarities in this region reside in four domains connected by short segments of variable length. The region will provide a focus for experiments that probe the relationship of amino acid sequence to structure and function in the P-450 superfamily.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

METHODS

Amino Acid Sequences. The source for each eukaryotic P-450 sequence used in these comparisons was as follows: LAω, rat liver (4); c, rat liver (5); 17α, bovine adrenal cortex (6); scc, bovine adrenal cortex (7); pcn1, rat liver (8); C21, bovine adrenal cortex (9); b, rat liver (10); 14DM, *Saccharomyces cerevisiae* (11). The bacterial P-450 sequence was P-450cam from *Pseudomonas putida* (12). Each amino acid sequence was derived from the corresponding nucleic acid sequence with the N-terminal methionine counted as residue 1. Some analyses used amino acid sequences of the Protein Identification Resource Protein Sequence Database.*

Construction of the Multiple Alignment. The multiple alignment was constructed after examination and manipulation of a number of different pairwise alignments for the eight P-450 proteins. We generated these alignments by using the Wilbur and Lipman algorithm (13) as implemented at Bionet in the IFIND program. The parameters WORD-LENGTH, GAP-PENALTY, WINDOW, DENSITY, and FAST were set at 1, 2, 30, LESS, and NO. The score for two aligned sequences equals the number of matching residues minus the number of gaps multiplied by the GAP-PENALTY. Similarities shared by all eight sequences were apparent only in the carboxyl-terminal portions of these proteins and only these portions are displayed in the multiple alignment.

The construction of the multiple alignment was not straightforward since different pairwise alignments sometimes yielded inconsistent results. One such case involved the three pairwise alignments of P-450C21, P-450c, and P-450pcn1. Although the alignment of P-450pcn1 with P-450C21 required one gap and the alignment of P-450c with P-450C21 required two gaps, the alignment of P-450pcn1 with P-450c did not require gaps. Such inconsistencies were eliminated by manually sliding or removing gaps in each pairwise alignment and then checking the multiple sequence alignment to find a configuration that yielded consistent gap placement and maximized the number of matching residues. This process removed closely positioned gaps, yielding a multiple alignment of eight sequences that contained only three gaps. The success of this approach suggests that pairwise alignment algorithms might be improved by incorporating an adjustable parameter that could be used to penalize closely positioned gaps.

Consensus Sequences. Sequence similarities are summarized by four tentative consensus sequences. To construct these consensus sequences, we assumed that each amino acid would appear an average of 1 time in 20 at any position in any of the eight aligned amino acid sequences. We then calculated (14) the expected frequency for every combination of eight amino acids that could occur at a single position in the multiple alignment. The 22 possible combinations range from the case where all eight amino acids are identical to the case

*Protein Identification Resource (1987) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 12.

where all eight are different. The 6 most common patterns and their expected frequencies are as follows: (*i*) no pairs of identical amino acids, 0.198; (*ii*) one pair of identical amino acids, 0.427; (*iii*) two pairs of identical amino acids, 0.229; (*iv*) three pairs of identical amino acids, 0.031; (*v*) three identical amino acids, 0.061; and (*vi*) three identical amino acids plus one pair of identical amino acids, 0.041. When 1 of these common patterns occurred, no consensus amino acid was selected. Each of the remaining 16 patterns is expected to occur at a frequency <1%. When 1 of these less common patterns occurred, we selected the identical amino acids in that pattern as the consensus amino acids at that position in the multiple alignment. We expected that a typical eukaryotic P-450 protein would match these consensus sequences at many but not necessarily all of the consensus positions.

Comparisons of Consensus Sequences to Proteins. Using a sliding window approach, each consensus sequence was compared to a target protein or to a randomly permuted version of the target protein. To accomplish this comparison, the first position in the consensus sequence was placed above the first amino acid in the sequence of the protein and the number of positions at which both sequences shared an identical amino acid was determined. The first position in the consensus sequence was then placed above the second amino acid in the protein and the number of matching residues was again determined. The stepwise movement of the consensus sequence was continued to the end of the protein. The highest number of matching residues found during this process became the score for that protein. When the target protein was a P-450, the consensus sequence was also compared to 100 randomly permuted variants of this P-450 protein. A mean and standard deviation were then calculated for the number of matching residues found during the steps in the comparisons to the randomized P-450 sequences. Using this mean and standard deviation, a z value could be calculated that quantitated the difference between the score for an unrandomized P-450 and the mean number of matching residues occurring in random variants of the P-450:

$$z = (\text{score} - \text{mean})/\text{standard deviation}.$$

An algorithm (15) was used to generate the random numbers that were required to permute (16) the amino acid sequences.

Consensus sequences were also compared to the proteins in the Protein Identification Resource database. The distribution of scores was plotted and the mean and standard deviation of these scores were determined.

RESULTS AND DISCUSSION

We performed pairwise alignments of eight eukaryotic cytochrome P-450 sequences to examine sequence similarities at the level of individual amino acid residues. Each of the eight chosen sequences was from a different P-450 family since we wished to find similarities common to all eukaryotic P-450s. These pairwise alignments were manually adjusted as described in *Methods* to construct a multiple sequence alignment.

The multiple alignment (Fig. 1) reveals the details of a region of sequence similarity found in all eight sequences. This shared region begins 275–310 amino acids from the amino terminus of each P-450 and continues to within 35–50 amino acids of the carboxyl terminus. Sequence similarities in this region reside in four domains that contain 56, 20, 59, and 28 amino acid residues. These domains are joined by short segments of dissimilar amino acid sequence that vary in size among the individual P-450 sequences. The four domains (A, B, C, and D in Fig. 1) define the region of sequence similarity found in all eight eukaryotic P-450 families. The boundaries of this region are marked by decreased sequence similarity. No consensus amino acid occurs in the 30 positions immediately before the start of the region. The yeast and mammalian sequences diverge at the end of the region.

Three of the four domains include elements described by previous workers. In the A domain, residues A-11 through A-27 correspond to a segment in the prokaryotic P-450cam that was demonstrated by x-ray crystallography to span the heme distal surface and to contain residues that contact the substrate camphor (17). In addition, antibodies directed against the dodecapeptide, which corresponds to residues A-8 through A-19 of P-450b, were shown to cross-react with P-450c (21). In the B domain, residues B-8 through B-20 correspond to a conserved tridecapeptide (18). The sequence D-7 through D-27 is the conserved peptide (19) that contains the thiolate ligand to the heme (22, 23). These previously identified peptides contain a number of invariant or nearly invariant residues; the multiple alignment has identified additional highly conserved residues, which include those at A-36, A-44, A-45, C-24, C-29, C-43, C-46, C-49, C-51, C-53, and C-54.

Four consensus sequences were derived from the aligned amino acid sequences and each contains its own pattern of conserved and invariant amino acids. These characteristic patterns do not arise merely because of some bias in the amino acid composition of P-450s as is shown by the data in Table 1. This table lists the scores obtained when the four consensus sequences were compared to eight eukaryotic P-450s and one bacterial P-450. For the eukaryotic P-450s, each score was always more than 8 standard deviations above the mean number of matching residues observed in randomized P-450 sequences. On the other hand, scores for the bacterial P-450cam were lower, especially the score of 5 obtained with the B consensus sequence. Results of the randomization experiments indicated that an average randomized variant of P-450cam contains about three segments with five matches to the B consensus sequence. Furthermore, the segment in P-450cam that best matched the B consensus sequence did not lie between the A and D domains. The highest scoring segment between these two domains had only four matches with the B consensus. This number of matches is expected to occur ≈13 times in an average "shuffled" P-450cam sequence. Thus, no segment in P-450cam matches the B consensus sequence to a significant extent.

The four consensus sequences were also tested for specificity by comparison against the proteins in the Protein Identification Resource database. A bimodal distribution of scores was observed (Fig. 2). Scores for all full-length eukaryotic P-450s fell into the high group while non-P-450 scores did not. These results indicate that the consensus sequences could be used to determine whether the sequence of an unidentified peptide resembled that expected for a eukaryotic P-450.

Scores for the bacterial P-450cam were not as easy to categorize as those for the eukaryotic P-450s. Using the D consensus sequence, the score for P-450cam was higher than that for any non-P-450 protein. When either the A or C consensus sequence was compared to P-450cam, the scores were higher than the scores for all but a few of the non-P-450 proteins, and the sum of the A and C scores for P-450cam exceeded the sum for any non-P-450 sequence. However, for the B consensus sequence, almost 60% of the non-P-450 proteins had scores that equaled or exceeded that of P-450cam. Consequently, as was also indicated by the results of the shuffling experiments described above, P-450cam does not contain a peptide that resembles the B consensus sequence to a significant degree.

All P-450 proteins share some properties, while other properties are specific to subsets of these proteins. For example, they all bind heme in a manner that yields a

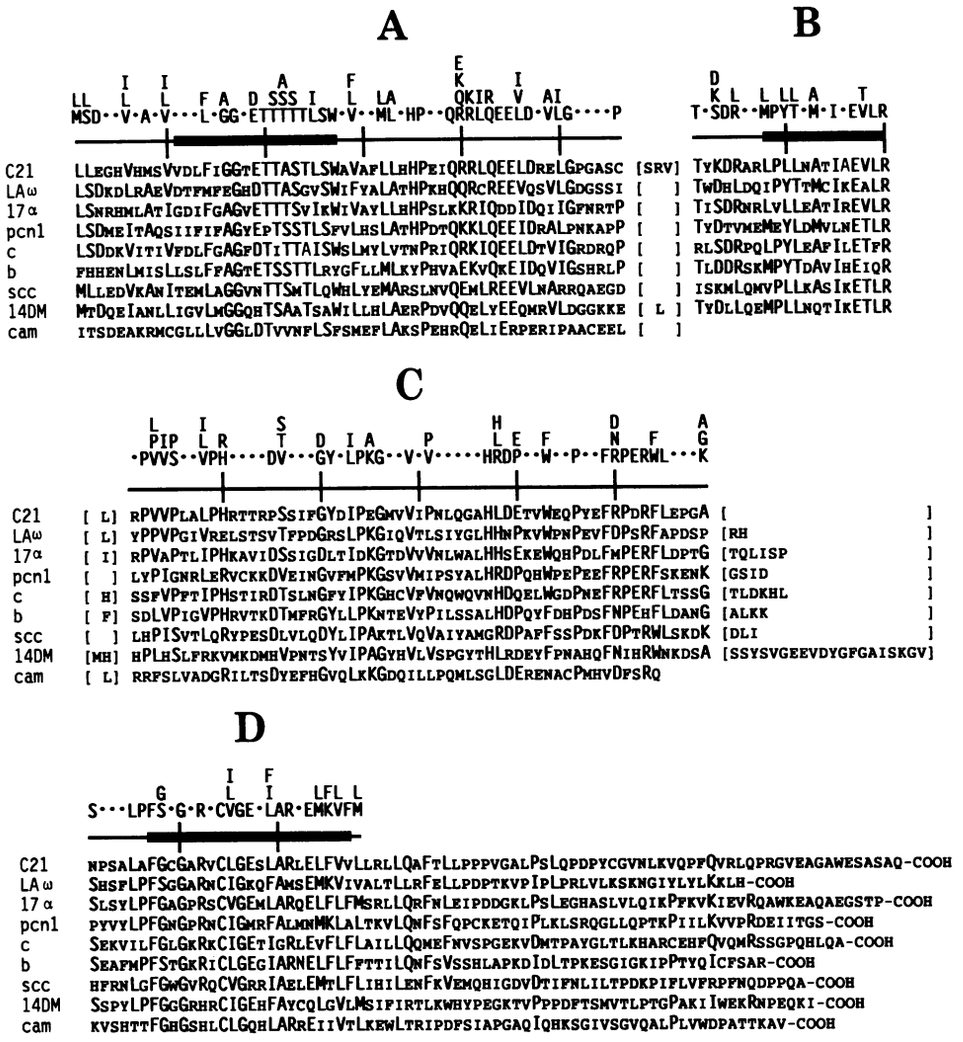

FIG. 1. Sequence alignment of the carboxyl-terminal portion of nine P-450s. These partial P-450 sequences begin 200–300 amino acids from the amino terminus of the parent P-450 but are otherwise complete from their beginning in the A domain on through the carboxyl terminus of the parent protein. Square brackets enclose residues that lie between domains. A horizontal line, positioned beneath the consensus sequence, spans each domain. A period in these consensus sequences identifies a position that lacks consensus amino acids. Boldface type indicates amino acids that match the consensus sequence. Vertical lines on the horizontal lines mark every 10th amino acid within each domain. The horizontal line that spans each domain thickens to mark previously reported regions of sequence similarity. These regions include a segment in P-450cam that traverses the heme distal surface (17) in the A domain; a conserved tridecapeptide (18) in the B domain; and a conserved cysteinyl peptide (19) in the D domain. The alignment for P-450cam was guided by the eukaryotic multiple alignment. The amino acid residue in each protein that begins the A domain is as follows: C21, Leu-275; LAω, Leu-305; 17α, Leu-287; pcn1, Leu-291; c, Leu-306; b, Phe-283; SCC, Met-310; 14DM, Met-299; and cam, Ile-234. The N-terminal methionine in the deduced amino acid sequence of each protein is residue 1 in this numbering system. The standard single letter notation is used (20). The I helix in P-450cam corresponds to residues A-2 through A-35, the J helix corresponds to residues A-36 through A-44, while the L helix starts at D-16 and continues for 20 residues (17). The highest percentage of invariant positions among the P-450s occurs in the first half of the conserved cysteinyl peptide in the D domain. This high density of invariant residues allows the peptide to be detected with relative ease in various P-450 sequences. Patterns for the A and C domains are less apparent since the density of invariant and highly conserved residues is relatively low in these domains.

characteristic difference spectrum in the presence of carbon monoxide and reducing agents (24). On the other hand, mitochondrial P-450s receive electrons from an iron–sulfur redoxin, while the microsomal P-450s receive electrons from a P-450 reductase. The microsomal P-450 reductase from one species can donate electrons to a P-450 from another species both *in vitro* (25) and *in vivo* (26). These common properties might be reflected in the sequence similarities identified in the multiple alignment displayed in Fig. 1.

These sequence similarities can be altered *in vitro* to examine the effects on structure and function. Highly conserved residues, such as the phenylalanines at C-63 and D-7, are attractive targets for site-directed mutagenesis. This approach allowed Liang *et al.* (27) to show that the phylogenetically conserved Phe-87 in the yeast iso-1-cytochrome *c* is involved in the transfer of electrons between hemoproteins.

Each of the four domains identified in the alignment might play a specific role in P-450 structure or function. A series of chimeric P-450 proteins could be constructed by swapping DNA sequences that encode these domains between different P-450 proteins. These constructs, when expressed *in vivo*, might allow the mapping of specific functions to specific domains and might also answer questions concerning the functional independence of the domains. The joints that connect the heterologous amino acid sequences in these chimeras would lie between the domains. Since segments between the domains vary in both size and sequence among the P-450s, any perturbation in tertiary structure caused by these joints is likely to be tolerated.

Furthermore, since the spacing between domains varies from one P-450 to another, it seems likely that changes to this spacing in a single P-450 would not severely disrupt at least some of the characteristic P-450 properties. It would be

Table 1. Scores from comparisons of the four consensus sequences to nine P-450s

Gene family	P-450	Consensus sequence A	B	C	D
XXI	C21	32 (14.1)	14 (10.6)	28 (14.6)	17 (11.1)
IV	LAω	29 (13.3)	11 (8.7)	22 (11.4)	16 (10.9)
XVII	17α	28 (12.3)	13 (10.1)	26 (13.9)	20 (13.7)
III	pcn1	31 (14.5)	10 (7.6)	24 (12.5)	15 (10.1)
I	c	31 (14.2)	12 (9.5)	26 (13.9)	16 (10.9)
II	b	26 (11.4)	12 (9.2)	27 (14.3)	18 (11.8)
XI	SCC	25 (11.3)	11 (8.6)	25 (13.1)	15 (10.1)
LI	14DM	23 (10.3)	12 (9.6)	19 (9.7)	15 (10.4)
CI	cam	13 (4.7)	5 (3.3)	12 (5.3)	12 (8.0)

The score and z value (in parentheses) were calculated as described. Gene family designations are those of Nebert and Gonzalez (3).

informative to alter the distance between the domains and determine the effects on heme binding, the difference spectrum, and the interaction with cytochrome P-450 reductase. Even though large variations in the distance between the four conserved domains might be tolerated for some P-450 properties, the three variable segments might have some functional significance with respect to another property such as substrate specificity and this could be tested.

The alignment may also aid in the design of synthetic polypeptides for antibody production. Polyclonal antibodies to a synthetic dodecapeptide in P-450b (corresponding to residues A-8 through A-19) cross-reacted with both P-450b and P-450c (21), although these two proteins share only five identical amino acids in this segment. A small panel of polyclonal antibodies that cross-reacts with many different P-450s might be used, for example, to isolate additional P-450 genes from expression libraries.

As additional P-450 sequences become available, it will be interesting to see how they further define the four linked domains of sequence similarity that we have found in all eight eukaryotic P-450 sequences. The prokaryotic P-450cam shares three of these domains. Additional prokaryotic sequences should help decide whether or not the differences seen in P-450cam are specific to that P-450 or result from the divergence between prokaryotic and eukaryotic P-450s. We expect that sequence data from new eukaryotic P-450 families may result in changes to the pattern. Such changes might include (*i*) minor modifications in consensus amino acid residues and (*ii*) resolution of the conserved domains into subdomains. Indeed, a partial deduced amino acid sequence for aromatase, the first known member of a newly discovered P-450 family (28), suggests that the C domain can be divided into two subdomains with the boundary occurring immediately after residue C-40.

Detailed information about the tertiary structure of the P-450 proteins will aid in understanding structure and function for this diverse protein superfamily. Poulos and his colleagues (17) have determined the crystal structure of P-450cam and have shown that the heme in this bacterial protein is sandwiched between two helices. Each helix contains a segment exhibiting sequence similarity to similarly positioned segments in mammalian P-450s. Poulos has suggested that the crystal structure of P-450cam might be useful in modeling the tertiary structure of eukaryotic P-450s. Such modeling requires the correct alignment of the two sequences (29) since incorrect alignments can lead to flawed three-dimensional structures (30). The multiple alignment presented in this work may be useful in this context even though P-450cam lacks a segment that resembles the B consensus sequence.

In their work on P-450 evolution and membrane topology, Nelson and Strobel (31, 32) recently presented an alignment of P-450 sequences that included the bacterial P-450cam and members from seven of the eight eukaryotic P-450 families that we have analyzed. They included a large number of sequences from the P-450I and P-450II families in their alignment, and the resultant intrafamily similarities tend to obscure similarities common to all P-450 families. Because their alignment was not characterized by using randomizations or consensus sequences, specific differences with our alignment are hard to evaluate. In any event, their alignment contains gaps and insertions in the carboxyl-terminal portion of the P-450 sequences that are not necessary by our analysis.

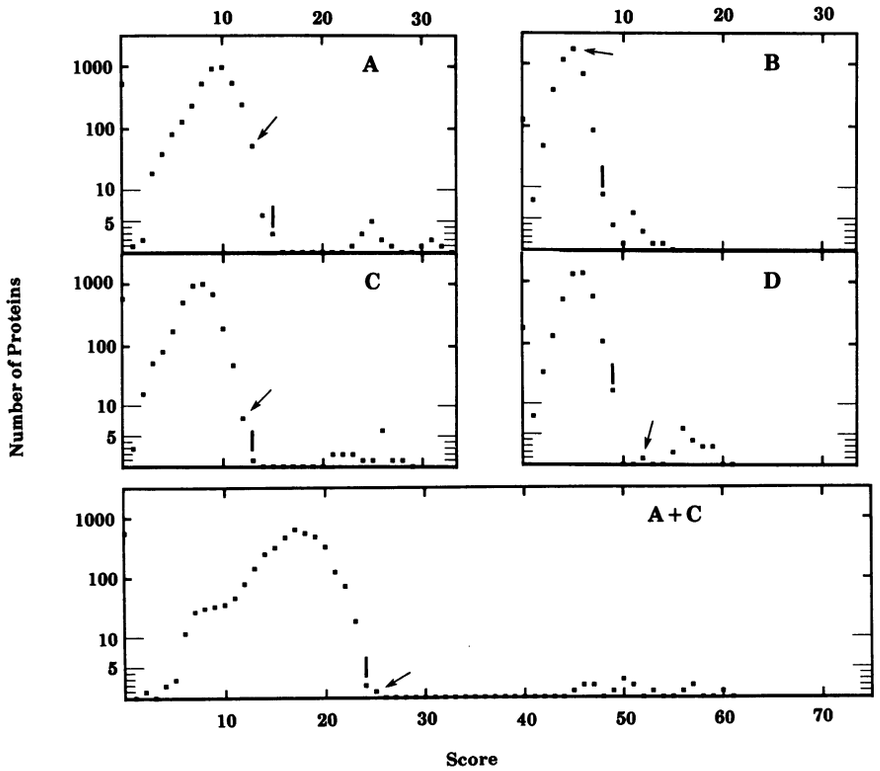

FIG. 2. Distribution of scores for proteins in the Protein Identification Resource database. Each consensus sequence was compared to each of the proteins in the database. Results of these comparisons using the A, B, C, and D consensus sequences are plotted in their respective panels. (*Bottom*) Number of proteins versus the sum of the A and C scores for these proteins. In each panel, a vertical bar marks the point corresponding to the highest score for a non-P-450 protein and an arrow marks the point corresponding to P-450cam. The vertical axis switches from arithmetic to logarithmic for values larger than 10. Mean scores (with standard deviations in parentheses) using the A, B, C, D, and A + C consensus sequences were 8.9 (2.9), 4.7 (1.2), 7.2 (2.5), 5.4 (1.6), and 16.1 (5.1), respectively.

Furthermore, we detect sequence similarity common to all eight eukaryotic P-450 families only in the carboxyl-terminal portion.

One proposed mechanism for gene evolution suggests that exons encode functional domains that assort to form new genes via recombination within flanking introns (33). These domains are short amino acid segments that carry out a limited function, such as the binding of heme by the peptide encoded in the central exon of the β-globin gene (34). Although the pattern of exon organization varies from one P-450 gene family to another (23), there are examples of introns interrupting each of the four domains. At least two of these introns interrupt eukaryotic P-450 genes in regions that correspond to known functional domains in the bacterial P-450cam. One intron splice position interrupts the codon for residue A-17 in members of the P-450I family (35). In P-450cam, this residue is thought to be one of the two residues in the oxygen binding site that contacts molecular oxygen (36). Another intron splice position interrupts the codon for residue D-10 in members of the P-450IIB subfamily (37). In P-450cam, this codon lies between residues D-7 and D-15, which provide hydrophobic contacts to the heme proximal surface (17).

A segmented pattern of sequence similarity also occurs in the globin superfamily. Bashford *et al.* (38) aligned and analyzed all known globin sequences. The 226 sequences included not only the hemoglobins and myoglobins of higher animals but also globins from insects, invertebrates, and plants. Sequence similarities in these globins can be found in six domains that are 16–21 residues long. Thus, in both the P-450 and globin superfamilies, evolution has conserved short domains that contain specific patterns of invariant, conserved, and variable amino acid residues. The absence of smaller domains in these superfamilies may imply that there is a lower size limit for heritable elements of protein structure or function.

We thank David J. Lipman and F. Peter Guengerich for helpful discussions while the manuscript was in preparation. Parts of this work were supported by grants from the U.S. Environmental Protection Agency and the University of Cincinnati, Groundwater Research Institute. Computer resources were provided by BIONET National Computer Resource for Molecular Biology.

1. 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.
2. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
3. Nebert, D. W. & Gonzalez, F. J. (1987) *Annu. Rev. Biochem.* **56**, 945–993.
4. Hardwick, J. P., Song, B.-J., Huberman, E. & Gonzalez, F. J. (1987) *J. Biol. Chem.* **262**, 801–810.
5. Sogawa, K., Gotoh, O., Kawajiri, K. & Fujii-Kuriyama, Y. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5066–5070.
6. Zuber, M. X., John, M. E., Okamura, T., Simpson, E. R. & Waterman, M. R. (1986) *J. Biol. Chem.* **261**, 2475–2482.
7. Morohashi, K., Fujii-Kuriyama, Y., Okada, Y., Sogawa, K., Hirose, T., Inayama, S. & Omura, T. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4647–4651.
8. Gonzalez, F. J., Nebert, D. W., Hardwick, J. P. & Kasper, C. B. (1985) *J. Biol. Chem.* **260**, 7435–7441.
9. Yoshioka, H., Morohashi, K., Sogawa, K., Yamane, M., Kominami, S., Takemori, S., Okada, Y., Omura, T. & Fujii-Kuriyama, Y. (1986) *J. Biol. Chem.* **261**, 4106–4109.
10. Fujii-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K. & Muramatsu, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2793–2797.
11. Kalb, V. F., Woods, C. W., Turi, T. G., Dey, C. R., Sutter, T. R. & Loper, J. C. (1987) *DNA* **6**, 529–537.
12. Unger, B. P., Gunsalus, I. C. & Sligar, S. G. (1986) *J. Biol. Chem.* **261**, 1158–1163.
13. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 726–730.
14. Snedecor, W. W. & Cochran, W. G. (1980) *Statistical Methods* (Iowa State Univ. Press, Ames), pp. 107–130.
15. Wichmann, B. & Hill, D. (1987) *Byte* **12**, 127–128.
16. McLachlan, A. D. & Boswell, D. R. (1985) *J. Mol. Biol.* **185**, 39–49.
17. Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C. & Kraut, J. (1985) *J. Biol. Chem.* **260**, 16122–16130.
18. Ozols, J., Heinemann, F. S. & Johnson, E. F. (1981) *J. Biol. Chem.* **256**, 11405–11408.
19. Gotoh, O., Tagashira, Y., Iizuka, T. & Fujii-Kuriyama, Y. (1983) *J. Biochem.* **93**, 807–817.
20. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *Eur. J. Biochem.* **5**, 151–153.
21. Frey, A. B., Waxman, D. J. & Kreibich, G. (1985) *J. Biol. Chem.* **260**, 15253–15265.
22. Black, S. D. & Coon, M. J. (1986) in *Cytochrome P-450*, ed. Oritz de Montellano, P. R. (Plenum, New York), pp. 161–216.
23. Morohashi, K., Sogawa, K., Omura, T. & Fujii-Kuriyama, Y. (1987) *J. Biochem.* **101**, 879–887.
24. Omura, T. & Sato, R. (1964) *J. Biol. Chem.* **239**, 2370–2378.
25. Aoyama, Y., Yoshida, Y., Kubota, S., Kumaoka, H. & Furumichi, A. (1978) *Arch. Biochem. Biophys.* **185**, 362–369.
26. Oeda, K., Sakaki, T. & Ohkawa, H. (1985) *DNA* **4**, 203–210.
27. Liang, N., Pielak, G. J., Mauk, A. G., Smith, M. & Hoffman, B. M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1249–1252.
28. Simpson, E. R., Evans, C. T., Corbin, C. J., Powell, F. E., Ledesma, D. B. & Mendelson, C. R. (1987) *Mol. Cell. Endocrinol.* **52**, 267–272.
29. Barlow, D. J., Blundell, T. L., Edwards, M. S., Sibanda, B. L., Sternberg, M. J. E., Taylor, W. R. & Thorton, J. M. (1986) in *Protein Engineering*, eds. Inouye, M. & Sarma, R. (Academic, New York), pp. 29–39.
30. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E. & Thornton, J. M. (1987) *Nature (London)* **326**, 347–352.
31. Nelson, D. R. & Strobel, H. W. (1987) *Mol. Biol. Evol.* **4**, 572–593.
32. Nelson, D. R. & Strobel, H. W. (1988) *J. Biol. Chem.* **263**, 6038–6050.
33. Gilbert, W. (1978) *Nature (London)* **271**, 501.
34. Craik, C. S., Buchman, S. R. & Beychok, S. (1981) *Nature (London)* **291**, 87–90.
35. Gonzalez, F. J., Kimura, S. & Nebert, D. W. (1985) *J. Biol. Chem.* **260**, 5040–5049.
36. Poulos, T. L., Finzel, B. C. & Howard, A. J. (1987) *J. Mol. Biol.* **195**, 687–700.
37. Suwa, Y., Mizukami, Y., Sogawa, K. & Fujii-Kuriyama, Y. (1985) *J. Biol. Chem.* **260**, 7980–7984.
38. Bashford, D., Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196**, 199–216.