Vol. 118, No. 2, 1984 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 30, 1984 Pages 655-662

SEGMENTAL HOMOLOGIES IN THE CODING AND 3' NON-CODING SEQUENCES OF RAT LIVER CYTOCHROME P-450e AND P-450b cDNAs AND CYTOCHROME P-450e-LIKE GENES

Markus Affolter and Alan Anderson

Centre de recherche sur le cancer
de l'Universite Laval, Hotel-Dieu de Quebec
and Departement de Biologie, Universite Laval,
Quebec, Quebec, Canada

Received December 19, 1983

SUMMARY. The nucleotide sequence of a cloned cDNA insert carried by pHDQ14 was determined and found to code for the 107 C-terminal amino acids of rat liver cytochrome P-450e. Comparison of the pHDQ14 cDNA sequence with those of cloned cDNAs for cytochrome P-450b and of 2 P-450e-like genes revealed segmental homologies that may have resulted from gene conversion. These results suggest that gene conversion may generate sequence variants of genes for rat liver cytochrome P-450s.

The extraordinary diversity of chemical structures to be found among the substrates of the mammalian mixed function oxidase system is explained in part by the existence of multiple molecular forms of the terminal oxidase, cytochrome P-450 (1-6). At least 8 rat liver cytochrome P-450 isozymes, including the closely related PB-inducible species P-450b and P-450e, have been purified to electrophoretic homogeneity (1-3). Moreover, subtle differences exist between PB-induced cytochrome P-450s (including P-450b and P-450e) of different rat strains and even within a single strain (4-6). Fujii-Kuriyama et al. cloned (7) and sequenced (8) 2 types of similar but distinct cDNAs coding for PB-inducible (Sprague-Dawley) rat liver cytochrome P-450s. These cDNAs were subsequently shown to code for cytochromes P-450b and P-450e by Yuan and coworkers, who determined the amino acid sequence of ~75% of each protein from Long-Evans rats (9). The known P-450b cDNA coding sequence covered all 491 codons except the first 5, whereas the cloned P-450e cDNA sequence extended from codon 52 to codon 358 (8). For the regions of overlap, cytochromes P-450b and P-450e are over 97% homologous at the protein

Abbreviation used: PB, phenobarbital.

0006-291X/84 $1.50

and cDNA levels (7,8). More recently, Mizukami et al. (10) have obtained evidence for the presence, in rat genomic DNA, of 6 or more different genes homologous to P-450b and P-450e cDNAs and have cloned and sequenced the exons of one of them (11). It would code for a 491-amino acid protein of which the deduced sequence differs from the cytochrome P-450e partial sequence (9) only by the substitution of methionine for lysine at position 473. Thus, associated with PB-inducible cytochrome P-450s, there is microheterogeneity both at the protein and DNA levels.

We report here the nucleotide sequence of a cloned cDNA fragment coding for the 107 C-terminal amino acids of the cytochrome P-450e protein (9). Comparison of our cDNA sequence with other known PB-inducible P-450 cDNA and genomic sequences revealed segmental homologies that could have arisen by gene conversion. Thus gene conversion may contribute to cytochrome P-450 gene polymorphism as it is thought to generate polymorphism in other systems (12-19).

MATERIALS AND METHODS

Male Sprague-Dawley rats (175-200 g) were obtained from Charles River Canada. Four days prior to sacrifice, a single rat was given one intraperitoneal injection (500 mg/kg) with Aroclor 1254 (a gift of Monsanto Chemicals) in corn oil. Total liver poly(A)$^+$ RNA was extracted (20,21). Double-stranded cDNA was then prepared by the method of Crabtree and Kant (22), slightly modified (23). The double-stranded cDNA was inserted into the Pst I site of pBR322 via oligo(dC)-oligo(dG) tails. After transformation of E. coli RR1, a differential hybridization procedure (23) was used to identify 2 clones containing putative PB-inducible cytochrome P-450 cDNA sequences. The cloned insert of one of them was used as a probe to identify 3 other clones containing homologous sequences. One such clone, carrying a recombinant plasmid designated pHDQ14, was retained for sequence analysis.

RESULTS AND DISCUSSION

The nucleotide sequence of the 726 base pair cDNA fragment carried by pHDQ14 (Figs 1,2) contains an open reading frame of 107 codons followed by a TGA stop codon, 385 bases of 3' non-coding sequence and a 17 base poly(A) tract. The open reading frame codes for the 107 C-terminal residues of the P-450e polypeptide sequence established by Yuan et al.(9) (Fig. 2). Mizukami and coworkers had suggested that strain differences might explain why their P-450e-like gene (henceforth designated as the P-450e^M gene), cloned from Sprague-Dawley rats, carries a methionine codon at position 473 (11), whereas

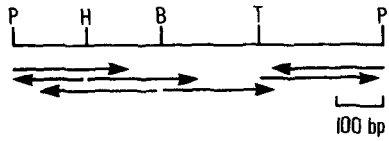

Figure 1. Restriction enzyme cleavage map showing the sites used in sequencing the insert of pHDQ14 by the chemical degradation method of Maxam and Gilbert (24). The lengths of the horizontal arrows indicate the extent of the sequence read from the restriction site. All restriction sites were 5' end labeled except BglII which was 3' end labeled. The sequence between the BglII and TaqI sites was also determined after 5' end labeling at the BglII site. Restriction enzyme sites used were: B, BglII; H, HindIII; P, PstI; T, TaqI.

the equivalent position in the cytochrome P-450e protein of Long-Evans rats is occupied by lysine (9). But, our cDNA (henceforth designated as P-450e^{L} cDNA) was cloned from a Sprague-Dawley rat, and it has a lysine codon at position 473 (Fig. 2). This, plus 6 other sequence differences noted (Fig. 2) clearly indicate that, at the gene level, P-450e^{L} and P-450e^{M} are different. They may represent non-allelic genes or they may be 2 allelic forms of the same gene. Non-allelic genes have recently been shown to be responsible for the synthesis of cytochromes P-450b and P-450e (6).

Although our cloned P-450e^{L} cDNA carries a poly(A) tract, it lacks the canonical (25) polyadenylation signal, AATAAA. Two related sequences, GGTAAA and AATCAC, end respectively 20 and 11 nucleotides upstream from the poly(A) tract (Fig. 2). The canonical signal had previously been found to be missing from the sequences of the P-450b cDNA and the P-450e^{M} gene (11), which are, in fact, identical to that of the P-450e^{L} cDNA over the last 118 bases preceeding the first A of the poly(A) tract (Fig. 2). In a rat liver cytochrome P-450d cDNA the canonical polyadenylation signal is also replaced by a variant (AATAGA) (ref. 23 and data not shown).

For the 3 partial sequences compared here, the cDNAs of P-450b and P-450e^{L} and the corresponding exons of P-450e^{M}, the nucleotide sequence divergences show a remarkable tendency to cluster into regions (Fig. 2). At 9 of 10 variant positions from nucleotide 9 to nucleotide 281 of our sequence P-450e^{L} and P-450e^{M} are identical and P-450b is different. The same is true at 7 of 8 variant nucleotide positions in the middle third of the 3'

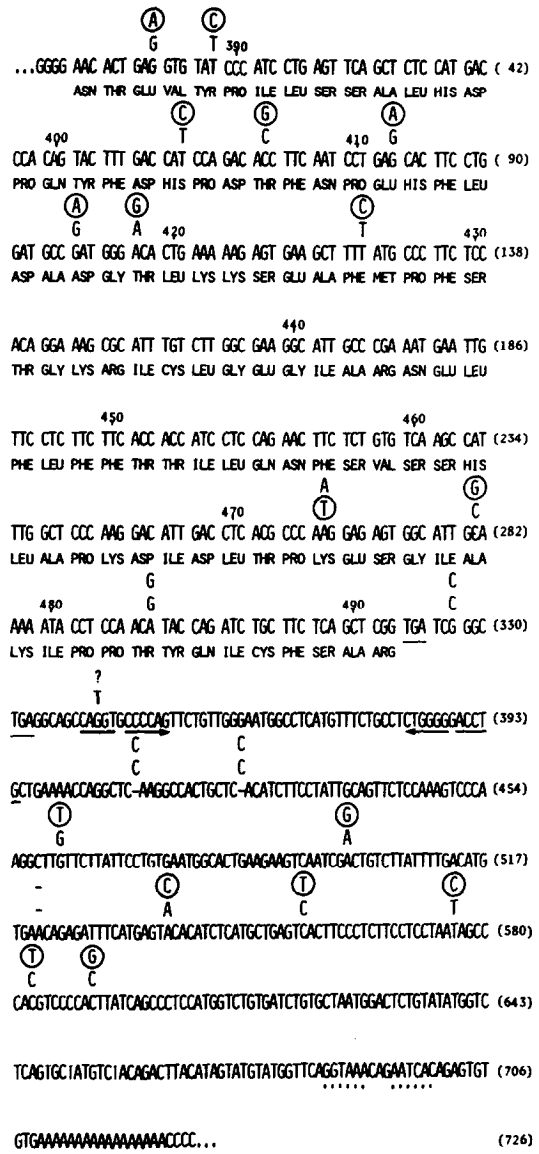

Figure 2. Nucleotide sequence of the cDNA insert of pHDQ14. The coding nucleotide sequence is given with the corresponding amino acid sequence shown below it. Codons are numbered according to the amino acid sequence of cytochrome P-450e (9). The numbers shown in parenthesis at the end of each line refer to the nucleotide sequence, numbered from the first nucleotide of the pHDQ14 cDNA insert. Differences with respect to the sequence of the P-450b cDNA (8,11) and/or the Mizukami P-450e^{M} gene (11) are shown above the nucleotide sequence at the variant positions: top, corresponding nucleotide in the P-450b cDNA (the P-450b cDNA sequence was obtained from a single cloned molecule up to position 428 of our sequence (8) and from another cloned molecule from there to the poly(A) tract (11)); middle, corresponding nucleotide in the Mizukami P-450e^{M} gene. Absence of a nucleotide at a given position is represented by a dash. Circled nucleotides at the variant positions are those that are unique to the P-450b cDNA or to the Mizukami P-450e^{M} gene; at positions where the P-450e^{L} cDNA is unique, none of the three nucleotides is circled. Nucleotides without indications above our sequence are the same in all three cytochrome P-450-related DNAs. The P-450b cDNA sequence is ambiguous at the 18th non-coding nucleotide (question mark at position 342);it was given as G in one report (8) and as T

non-coding region, from nucleotide 461 to nucleotide 591. Between these 2 nearly perfect blocks where the P-450b sequence tends to be unique, there is a region where, at 4 of 5 (or 5 of 5, one position is ambiguous, see Fig. 2) variant positions, P-450e^{M} and P-450b are the same and P-450e^{L} is unique. Thus, P-450e^{M} and P-450e^{L} are virtually identical at both ends and divergent in the middle portion of the sequence (Fig. 2). Formally, the present sequence of the P-450e^{M} gene for the region from near nucleotide position 290 to near nucleotide position 425 of our sequence can be derived from that of a non-allelic (6), but closely related gene, the gene for P-450b. Although other scenarios are possible, (involving, for example, the preferential accumulation of mutations between nucleotides 290 and 425 of what is now P-450e^{L}), this suggests that part of an ancestral P-450e^{M} sequence was recently replaced via gene conversion (12,26) by the corresponding part of P-450b (Fig. 3).

After our work was completed, Adesnik and coworkers (27,28) published additional rat cytochrome P-450 cDNA and genomic sequences. Kumar et al.(27) reported a partial (Long-Evans) cDNA sequence, carried by plasmid R17. The R17 cDNA encodes the 211 C-terminal amino acids of a P-450e-like protein and also carries the complete 3' non-coding sequence (27). For the region of overlap, the pHDQ14 and R17 cDNAs are identical but for 2 single base substitutions that would lead to 2 amino acid replacements; codons 438 and 476 are both GGC (glycine) in the P-450e^{L} sequence carried by pHDQ14, but are both GAC (aspartate) in the cDNA of R17 (27). Atchison and Adesnik (28) reported the nucleotide sequence of exons 2 to 9 of a P-450e^{M}-like gene (isolated from the same genomic library, that of Jagodzinsky and Bonner, as was employed by Mizukami et al. (11) to isolate their P-450e^{M} clone). This Adesnik P-450e^{M} gene carries a methionine (ATG) codon at position 473, but, its exonic sequence is otherwise identical to our pHDQ14 cDNA sequence, for

---

in another (11). The broken arrows under positions 340 to 351 and 383 to 394 indicate an almost perfect 12 base pair inverted repeat. Dotted lines under positions 685 to 690 and 694 to 699 indicate two sequences related to the canonical AATAAA polyadenylation signal. TGA codons are underlined.

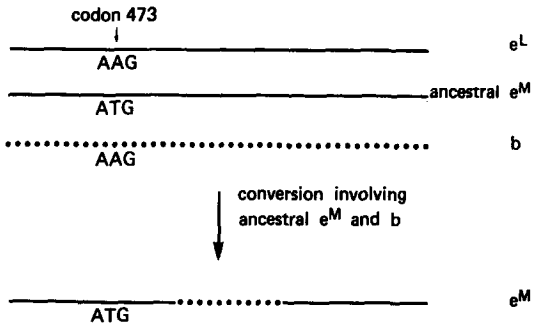

Figure 3. Possible origin by gene conversion of the (Mizukami) P-450e^M sequence from ancestral P-450e^M and P-450b sequences. The coding sequences are represented from the beginning of the pHDQ14 P-450e^L cDNA insert to the poly(A) tract, but are not drawn to scale. The beginning and the end of the conversion tract would both be within the ninth exon (11) of the Mizukami P-450e^M gene. A portion of the ancestral P-450e^M gene would have been replaced by the corresponding portion of the P-450b sequence, as shown. This model was derived from a comparison (Fig. 2) or our P-450e^L cDNA sequence with those of the P-450b cDNA inserts and the Mizukami P-450e^M gene; we have cloned and characterized by restriction mapping several P450b-type cDNA inserts each of which covers at least the entire region shown in figure 2 (data not shown). Atchison and Adesnik have since reported (28) a genomic sequence that we refer to as the Adesnik P-450e^M gene. For the region represented here, the Adesnik P-450e^M gene has the precise sequence expected for that of an ancestral pre-conversion P-450e^M gene, except for the presence of a single extra A at position 521, downstream from the conversion tract and also present (Fig. 2) in our P-450e^L cDNA but absent (11) from the P-450b and Mizukami P-450e^M sequences.

the region of overlap. In other words, the Adesnik e^M gene does not contain the block of P-450b-like sequence that is present in the Mizukami e^M gene (Figs 2,3). For the region in which we have proposed that a gene conversion event took place (between nucleotides 290 and 425 of our sequence; see Fig. 2), the Adesnik e^M sequence can be viewed as the pre-conversion ancestor of the Mizukami e^M sequence (Fig. 3). Perhaps the best evidence that can be obtained from DNA sequence analysis to confirm the occurence of a conversion event is to identify a pre-conversion ancestor, a post-conversion product and a donor gene. Based on the comparison of our P-450e^L sequence with the P-450b and Mizukami P-450e^M sequences (Fig. 2), we had identified the putative post-conversion product (the Mizukami e^M sequence) and a possible donor gene (the P-450b sequence). The Adesnik e^M sequence furnishes a possible pre-conversion ancestor, and thus provides substantial support for the gene conversion model. Note, however, that in a formal genetic sense, double unequal crossing over could also account for the results.

Gene conversion has been suggested (12-19) as a mechanism to explain similar examples of segmental homologies between related genes. Indeed, conversion-like events may have been involved in generating the particular combination of sequences compared here. But, it should be recalled, gene conversion is believed to act to maintain sequence homogeneity among (constant domains of) members of multigene families (12,28). Therefore, conversion events alone can not account (28) for the presence of a "variable region" (8,9) in P-450b and P-450e coding sequences. Rather, other factors, such as positive selection in favor of variation, with or without a localized increase in mutation rate in the variable region (29) would have to be invoked. Indeed, it is the existence of variable regions that permits the consequences of conversion-like events to be readily observed.

The extent to which gene conversion might generate cytochrome P-450 protein variants remains to be established. But one possible example is already available: the P-450e-like cDNA of Kumar et al. (27) contains, between codons 313 and 322 (not included in the cDNA insert of pHDQ14), a short block of P-450b-like sequence, involving 3 base substitutions and implying amino acid replacements at positions 321 and 322. Gene conversion could account for the generation of this hybrid coding sequence.

Acknowledgements. This research was supported by the National Cancer Institute of Canada and, in part, by the Natural Sciences and Engineering Research Council of Canada. We are grateful to Carmen Parent-Vaugeois, Martine Raymond, Pierre Baril, Michel Guertin, Bernard Turcotte, Mario Chevrette, Yves Labelle, Michel Lambert, Normand Paquette, Luc Belanger and Adolfo Ruiz-Carrillo for help, counsel or access to facilities, and to Elisabeth Lemay who typed the manuscript.

REFERENCES

1. Ryan, D.E., Thomas, P.E., Korzeniowski, D., and Levin, W. (1979) J. Biol. Chem. 254, 1365-1374.
2. Guengerich, F.P., Dannan, G.A., Wright, S.T., Martin, M.V., and Kaminsky, L.S. (1982) Biochemistry 21, 6019-6030.
3. Ryan, D.E., Thomas, P.E., and Levin, W. (1982) Arch. Biochem. Biophys. 216, 272-288.
4. Vlasuk, G.P., Ghrayeb, J., Ryan, D.E., Reik, L., Thomas, P.E., Levin, W., and Walz, G.F., Jr. (1982) Biochemistry 21, 789-798.
5. Walz, F.G., Jr., Vlasuk, G.P., Omiecinski, C.J., Bresnick, E., Thomas, P.E., Ryan, D.E., and Levin, W. (1982) J. Biol. Chem. 257, 4023-4026.
6. Rampersaud, A., and Walz, G.F., Jr. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 6542-6546.

7. Fujii-Kuriyama, Y., Taniguchi, T., Mizukami, Y., Sakai, M., Tashiro, Y., and Muramatsu, M. (1981) J. Biochem. (Tokyo) 89, 1869-1879.
8. Fujii-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K., and Muramatsu, M. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 2793-2797.
9. Yuan, P.-M., Ryan, D.E., Levin, W., and Shively, J.E. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1169-1173.
10. Mizukami, Y., Fujii-Kuriyama, Y., and Muramatsu, M. (1983) Biochemistry 22, 1223-1229.
11. Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M., and Fujii-Kuriyama, Y. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3958-3962.
12. Baltimore, D. (1981) Cell 24, 592-594.
13. Schreier, P.H., Bothwell, A.L.M., Mueller-Hill, B., and Baltimore, D. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 4495-4499.
14. Lalanne, J.L., Bregegere, F., Delarbe, C., Abastado, J.P., Gachelin, G., and Kourilisky, P. (1982) Nucl. Acid. Res. 10, 1039-1049.
15. Weiss, E.H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J., and Flavell, R.A. (1983) Nature (London) 301, 671-674.
16. Ollo, R., and Rougeon, F. (1983) Cell 32, 515-523.
17. Bentley, D.L., and Rabbits, T.H. (1983) Cell 32, 181-189.
18. Bregegere, F. (1983) Biochimie 65, 229-237.
19. Kourilsky, P. (1983) Biochimie 65, 85-93.
20. Iynedjian, P.B., and Hanson, R.W. (1977) J. Biol. Chem. 252, 655-662.
21. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1408-1412.
22. Crabtree, G.R., and Kant, J.A. (1981) J. Biol. Chem. 256, 9718-9723.
23. Affolter, M. (1983) These de maitrise, Universite Laval.
24. Maxam, A.M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
25. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature (London) 263, 211-214.
26. Fincham, J.R.S., Day, P.R., and Radford, A. (1979) Fungal Genetics, 4th Ed., pp.205-234, University of California Press, Berkeley.
27. Kumar, A., Raphael, C., and Adesnik, M., (1983)., J. Biol. Chem. 258, 11280-11284.
28. Atchison, M., and Adesnik, M. (1983) J. Biol. Chem. 258, 11285-11295.
29. Benoist, C., Mathis, D.J., Kanter, M.R., Williams, V.E. II, and McDevitt, H.O. (1983) Cell 34, 169-177.