Possible steroid binding site common to adrenal cytochrome P-450scc and prostatic steroid binding protein

O. Gotoh, Y. Tagashira, K. Morohashi*† and Y. Fujii-Kuriyama*

*Department of Biochemistry, Saitama Cancer Center Research Institute, Ina-machi, Saitama 362, *Department of Biochemistry, Cancer Institute, Japanese Foundation for Cancer Research, Toshima-ku, Tokyo 117* and †*Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan*

Received 13 June 1985

By searching the entire PIR-protein-sequence data base, we have found that a dodecapeptide sequence in bovine adrenal cytochrome P-450scc is closely related to that in rat prostatic steroid binding protein. The two proteins belong to unrelated protein families, but both have steroids as substrates or ligands. Thus, the dodecapeptides may be important for substrate/ligand recognition in the individual proteins.

Cytochrome P-450 *Steroid binding protein* *Protein data base* *Sequence homology*

1. INTRODUCTION

Cytochrome P-450scc catalyzes side-chain cleavage of cholesterol, which is the primary reaction in the synthesis of various steroid hormones. The enzyme functions on mitochondrial membranes in the adrenal cortex. It shares various physicochemical properties with microsomal cytochromes P-450. However, mitochondrial cytochrome P-450scc shows a limited substrate specificity and receives electrons from NADPH-adrenodoxin reductase via an iron-sulfur protein, adrenodoxin, whereas its microsomal counterpart has a broad substrate specificity and receives electrons directly from a flavoprotein, NADPH-cytochrome P-450 reductase [1].

Prostatic steroid binding protein is the major secretory protein in rat prostatic fluid [2]. It consists of 2 heterologous subunits, each composed of 2 polypeptides, C1 and C3, and C2 and C3, respectively, linked by disulfide bridges [3]. The function of the prostatic steroid binding protein is unknown, but this protein has been suggested to act as a carrier in the process of steroid production [2].

Cytochrome P-450scc and prostatic steroid binding protein belong to different protein families, and there is no evidence for their relatedness, although they have the same specific substrates or ligands. This report shows that a short peptide sequence consisting of 12 amino acids is shared by these 2 proteins of different classes. This information may be useful for locating currently unknown substrate-binding sites in different families of proteins.

2. METHODS

Amino acid sequences were taken from the recent literature [4–9] or from the PIR data base [10]. This data base was searched for sequence similarity by the method of Smith and Waterman [11] modified as suggested in [12]. A mutation data matrix [13] at 100 PAM (100 accepted point mutations per 100 residues) was used as a measure of relatedness between aligned amino-acid elements. A penalty of $60k + 60$ was imposed on each gap of length k [12]. The searches were performed on a Hitachi S810-20 supercomputer at the University of Tokyo. The resulting similarity scores were converted into standard-deviation units by calibrating for length and optionally for compositions, and then sorted in descending order. Further analyses

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

were made in a TSS job on a Hitachi M280 computer.

3. RESULTS AND DISCUSSION

The used version (April, 1984) of the PIR data base [10] contained 3 cytochrome P-450 sequences (rat P-450b [14], rabbit P-450LM2 [15], and *Pseudomonas putida* P-450cam [16]). When the data base was searched against the bovine P-450scc sequence [4], the 2 mammalian P-450 scores were the highest, followed by that for rat prostatic steroid binding protein component 2 (C2) [17]. This protein is particularly interesting because like cytochrome P-450scc it recognizes steroids as substrates or ligands.

Fig.1 shows the region where the sequences of cytochrome P-450scc and component C2 are similar to each other. Of the 12 sites, 6 are identical, 4 are conservative changes, and 2 are neutral changes. On the basis of shuffling tests [18], we estimated that the probability of observing this degree of sequence similarity by chance would be less than 10^{-8}. Thus, it seems reasonable to consider that this region is most likely to be responsible, at least in part, for substrate recognition.

Of the 3 components of rat prostatic steroid binding protein, component C1 is most closely related to C2, showing 50% amino-acid sequence match [17]. However, considerable sequence divergence (33% matches) is observed in the region where C2 is similar to P-450scc (fig.1). This is consistent with the idea that the identified dodeca sequence is responsible for steroid binding, because only subunit C2-C3 retains the ability of steroid binding when the 2 subunits, C1-C3 and C2-C3,

```
Bovine Cytochrome P-450scc      210   I T N V M F G E R L G M   221
                                      * . . : * : * : * * : *
Rat steroid binding protein C2   69   I N K I M Y G D R L S M    80
                                      : : . :     * * * *   :
Rat steroid binding protein C1   69   V D Q M S N G D R L V V    80
```

Fig.1. Tentative steroid binding sequence common to bovine cytochrome P-450scc (upper) and rat prostatic steroid binding protein component C2 (middle). The corresponding sequence in rat binding protein component C1 is also shown (lower). The marks between the sequences indicate identity (*), conservative change (:), neutral change (.) and non-conservative change (blank).

Species		No.	Sequence		Ref.
Rat	P-450c	188	FDP-FKYLVVSVANVICAICFGRRYDHDDOELLSIVNL	224	[6]
Mouse	P1-450	188	FDP-YKYLVVSVANVICAICFGQRYDHDDQELLSIVNL	224	[9]
Rat	P-450d	185	FEP-VNQVVESVANVIGAMCFGKNFPRKSEEMLNLVKS	221	[5]
Mouse	P3-450	185	FEP-VSQVVESVANVIGAMCFGKNFPRKSEEMLNIVNN	221	[9]
Rat	P-450b	165	LDPTFLFQCIT-ANIICSIVFGERFDYTDRQFLRLLEL	191	[14]
Rat	P-450e	165	LDPTFLFQCIT-ANIICSIVFGERFDYTDRQFLRLLEL	191	[7]
Rabbit	P-450LM2	165	LDNTLLFHSIT-SNIICSIVFGKRFDYKDPVFLRLLDL	191	[15]
Rabbit	P-450c13		CDPTFLLFCVP-CNVICSVIFQNRFDYDDEKFKTLIKY		[8]
Rabbit	P-450c11		CNPTFILGAAP-CNVICSVIFQNRFDYTDQDFLSLMGK		[8]
Rabbit	P-450c12		CDPTFILGAAP-CNVICSVIFQNRFDYTDQDFLSLMGK		[8]
Bovine	P-450scc	199	ED-LFHFAFESITNV----MFGERLGMLEETVNPEAQK	232	[4]
P.putida	P-450cam	145	-------ACNF-TE-----DYAEPF--PIRIFMLLAGL	167	[16]

Fig.2. Part of the global sequence alignment of various cytochromes (P-450) including the tentative steroid binding site in P-450scc (underlined).

are dissociated [3]. We found that the other component, C3, is also related to C1 and C2, but the conserved sites are located in the NH_2-terminal regions (not shown).

The tentative steroid-binding dodecapeptide in P-450scc is located near the center of the molecule, close to the HR1 region [18]. The central regions of various molecular species of P-450 are highly variable. The assignment of correspondence shown in fig.2 is mainly based on positional information. Since the substrates of various P-450 species vary in size and form, positional correspondence may not directly indicate functional correspondence. Nevertheless, the sequences listed in fig.2 have the interesting feature that they are rich in cysteine. These cysteines might be involved in substrate binding as suggested in [16]. Moreover, the high variability in the sequence naturally explains the extraordinarily wide range of substrate specificities shown by the multiple forms of cytochrome P-450 [1].

We can say little about the evolutionary process by which the common sequence was acquired by otherwise unrelated proteins. Some movable elements or dynamic gene conversion mechanisms might be involved in the process. Alternatively, independent convergent evolution might lead to the current sequence similarity. Elucidation of genomic structures of P-450scc and prostatic steroid binding proteins may provide a clue to this problem.

ACKNOWLEDGEMENT

This work was partly supported by grants from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Sato, R. and Omura, T. eds (1978) Cytochrome P-450, Academic Press, New York.
[2] Heyns, W. and De Moor, P. (1977) Eur. J. Biochem. 78, 221–230.
[3] Heyns, W., Peeters, B., Mous, J., Rombauts, W. and De Moor, P. (1978) Eur. J. Biochem. 89, 181–186.
[4] Morohashi, K., Fujii-Kuriyama, Y., Okada, Y., Sogawa, K., Hirose, T., Inayama, S. and Omura, T. (1984) Proc. Natl. Acad. Sci. USA 81, 4647–4651.
[5] Kawajiri, K., Gotoh, O., Sogawa, K., Tagashira, Y., Muramatsu, M. and Fujii-Kuriyama, Y. (1984) Proc. Natl. Acad. Sci. USA 81, 1649–1653.
[6] Sogawa, K., Gotoh, O., Kawajiri, K. and Fujii-Kuriyama, Y. (1984) Proc. Natl. Acad. Sci. USA 81, 5066–5070.
[7] Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M. and Fujii-Kuriyama, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 3958–3962.
[8] Leighton, J.K., DeBrunner-Vossbrinck, B.A. and Kemper, B. (1984) Biochemistry 23, 204–210.
[9] Kimura, S., Gonzalez, F.J. and Nebert, D.W. (1984) J. Biol. Chem. 259, 10705–10713.
[10] Barker, W.C., Hunt, L.T., Orcutt, B.C., George, D.G., Yeh, L.S., Chen, H.R., Blomquist, M.C., Johnson, G.C., Seibel-Ross, E.I. and Ledley, R.S. (1984) Protein Sequence Database, Rel.1.
[11] Smith, T.F. and Waterman, M.S. (1981) J. Mol. Biol. 147, 195–197.
[12] Gotoh, O. (1982) J. Mol. Biol. 162, 705–708.
[13] Dayhoff, M.O. ed. (1978) Atlas of Protein Sequence and Structure vol.5, suppl.3, National Biomedical Research Foundation, Washington.
[14] Fujii-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K. and Muramatsu, M. (1982) Proc. Natl. Acad. Sci. USA 79, 2793–2797.
[15] Heinemann, F.S. and Ozols, J. (1982) J. Biol. Chem. 257, 14988–14999.
[16] Haniu, M., Armes, L.G., Tanaka, M., Yasunobu, K.T., Shastry, B.S., Wagner, G.C. and Gunsalus, I.C. (1982) Biochem. Biophys. Res. Commun. 105, 889–894.
[17] Parker, M., Needham, M. and White, R. (1982) Nature 298, 92–94.
[18] Gotoh, O., Tagashira, Y., Iizuka, T. and Fujii-Kuriyama, Y. (1983) J. Biochem. 93, 807–817.
[19] Haniu, M., Yasunobu, K.T. and Gunsalus, I.C. (1982) Biochem. Biophys. Res. Commun. 107, 1075–1081.