Biochem. J. (1990) 267, 839–841 (Printed in Great Britain)

BIOCHEMICAL JOURNAL LETTERS

Sequence similarity between *Pseudomonas* dihydrodiol dehydrogenase, part of the gene cluster that metabolizes polychlorinated biphenyls, and dehydrogenases involved in metabolism of ribitol and glucitol and synthesis of antibiotics and 17β-oestradiol, testosterone and corticosterone

There has been intense interest in characterizing the gene clusters found in *Pseudomonas pseudoalcaligenes* and *Pseudomonas putida* because they enable these organisms to metabolize completely various aromatic hydrocarbons such as biphenyl, polychlorinated biphenyl (PCB), toluene, and naphthalene [1–4] or convert them to less toxic forms that can be safely dispersed in the environment [5,6]. The genes for these enzymes have been cloned from *P. pseudoalcaligenes* [1] and *P. putida* [2,3], and they consist of several proteins that catalyse various steps in the oxidation of aromatic hydrocarbons to simpler molecules that can be used by the micro-organism as a source of carbon and energy.

We reported [7,8] that 17β-hydroxysteroid dehydrogenase [9], an enzyme important in the synthesis of 17β-oestradiol and testosterone, is homologous to actIII, an enzyme important in antibiotic synthesis in *Streptomyces coelicolor* [10,11]. Moreover, these two proteins share a common ancestor with rat corticosteroid 11β-hydroxysteroid dehydrogenase [8,12], *Escherichia coli* glucitol-6-phosphate dehydrogenase [13] and *Klebsiella aerogenes* ribitol dehydrogenase [14], enzymes that are used for metabolizing polyols as a source of carbon and energy, as well as to *Rhizobia* proteins that are part of the gene cluster that induces root nodulation in *Rhizobia's* host [15–17].

Recently, we searched the database for other members of this protein superfamily and uncovered a novel similarity between *P. pseudoalcaligenes* and *P. putida* dihydrodiol dehydrogenase and actIII, 17β-hydroxysteroid dehydrogenase, and other members of their protein superfamily. Later, as this manuscript was being completed, Zylstra & Gibson published the complete sequence of *P. putida* dihydrodiol dehydrogenase [3]. Their comparison of this sequence with *P. pseudoalcaligenes* dihydrodiol dehydrogenase indicated that the *N*-terminus of *P. pseudoalcaligenes* dihydrodiol dehydrogenase began at position 34 in the sequence determined by Furukawa *et al.* [1]. Thus, we adjusted the numbering for the residues in *P. pseudoalcaligenes* dihydrodiol dehydrogenase in the alignment with actIII and 17β-hydroxysteroid dehydrogenase and other members of this superfamily to conform to the numbering of Zylstra & Gibson [3]. In addition, we did a comparison of *P. putida* dihydrodiol dehydrogenase, which has about 65% sequence similarly to *P. pseudoalcaligenes* dihydrodiol dehydrogenase, with proteins in the actIII/17β-hydroxysteroid dehydrogenase superfamily.

The alignment of residues 6–259 of *S. coelicolor* actIII and residues 5–254 of *P. pseudoalcaligenes* dihydrodiol dehydrogenase is shown in Fig. 1. Out of 244 possible matches there are 70 (28.5%) identities and 46 (19%) conservative replacements. The ALIGN [18] score is 11.1 standard deviations higher than that obtained with 1000 comparisons of randomized sequences of these proteins. The probability (*P*) of getting such a score by chance is 6×10^{-29}. Table 1 summarizes the comparison scores for *P. pseudoalcaligenes and P. putida* dihydrodiol dehydrogenase with human 17β-hydroxysteroid dehydrogenase [9], *Escherichia coli* glucitol-6-phosphate dehydrogenase [13], *Klebsiella aerogenes* ribitol dehydrogenase [14], *Rhizobium meliloti nodG* protein [15,16], *Bradyrhizobium japonicum fixR* protein [16], *Eubacterium* bile acid 7-dehydroxylase [19], *Bacillus megaterium* glucose dehydrogenase [20] and mouse adipocyte p27 protein [21].

The region of similarity between *P. pseudoalcaligenes* dihydrodiol dehydrogenase and actIII comprises 250 residues, which is 92% of the 273 residues of the dihydrodiol dehydrogenase and 95% of the 261 residues of actIII. With the exception of human 17β-hydroxysteroid dehydrogenase, which contains 327 residues, the lengths of most of the other dehydrogenases listed in Table 1 vary from 244 residues (mouse adipocyte p27 protein) to 278 residues (*fixR*). In several proteins, the segments that are similar to dihydrodiol dehydrogenase are greater than 200 residues. These include glucitol-6-phosphate dehydrogenase 257-residue segment, bile acid 7-dehydroxylase 217-residue segment and *nodG* 228-residue segment, which constitute over 60% of these proteins. The other proteins have segments of at least 165 residues that are similar to dihydrodiol dehydrogenase. The proteins with the lowest ALIGN comparison scores are human 17β-hydroxysteroid dehydrogenase and rat 11β-hydroxysteroid dehydrogenase. The latter protein has ALIGN scores of 6.3 S.D. and 6.1 S.D. with the dihydrodiol dehydrogenases of *P. putida* and *P. pseudoalcaligenes*, respectively. Steroid affinity labeling of 17β-hydroxysteroid dehydrogenase has identified residues 204–223 as being part of the steroid-binding domain [22]. Thus, the similarity between 17β-hydroxysteroid dehydrogenase and dihydrodiol dehydrogenase decreases near the steroid-binding domain.

Most of the proteins listed in Table 1 are oxidoreductases. This similarity in enzyme function to the two *Pseudomonas* dihydrodiol dehydrogenases, combined with the strong ALIGN comparison scores, several of which are over 10 standard deviations, suggest that the dihydrodiol dehydrogenases share a common ancestor with the proteins in Table 1.

Metabolism of toluene and PCBs begins with conversion of these compounds to a dihydrodiol by a dioxygenase. For example, toluene is oxidized to *cis*-toluene dihydrodiol through the addition of molecular oxygen to the aromatic nucleus. This compound is then reduced by the dihydrodiol dehydrogenase to 3-methylcatechol, using NAD^+ as the cofactor. The similarities reported here suggest a likely location of the NAD^+ binding site. This information comes from Yamada & Saier's comparison [13] of the sequences of *E. coli* glucitol-6-phosphate dehydrogenase

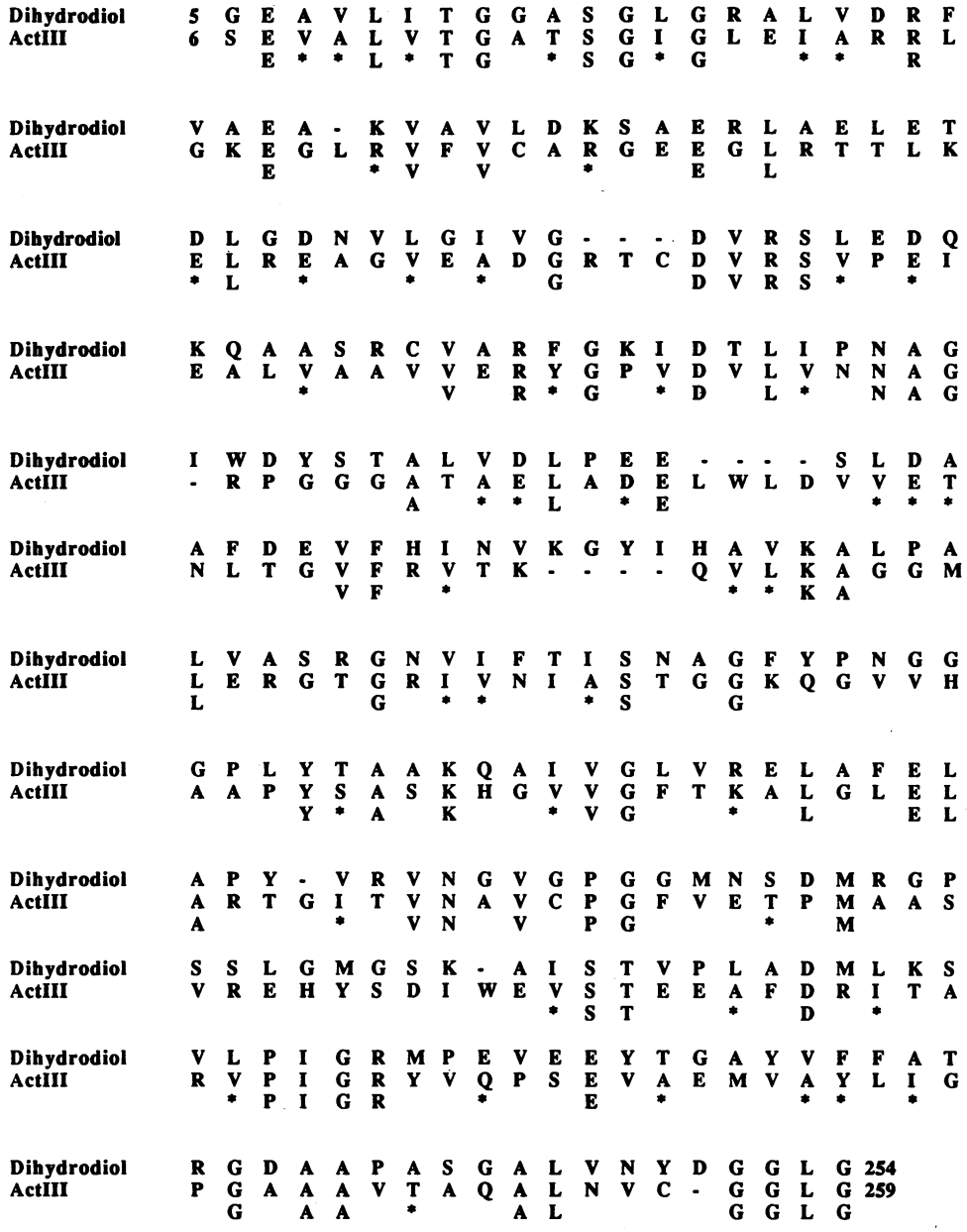

Fig. 1. Alignment of *Pseudomonas pseudoalcaligenes* dihydrodiol dehydrogenase with *Streptomyces coelicolor* actIII protein

Identities are noted in the space below the sequences; an asterisk (*) denotes conservative replacements. Out of 244 possible matches there are 70 (28.5%) identities and 46 (19%) conservative replacements. An ALIGN analysis of these sequences, with a gap penalty of 8, yields a score that is 11.1 standard deviations higher than that obtained with 1000 comparisons of randomized sequences of these segments. The probability of getting such a score by chance is 6×10^{-29}.

and *K. aerogenes* ribitol dehydrogenase with a consensus sequence of nucleotide binding domains from dehydrogenases, developed by Wierenga *et al.* [23], from which they concluded that the *N*-terminal 30 residues of these two dehydrogenases contain the nucleotide binding domain. Coleman *et al.* [19] and Jornvall *et al.* [24] reached a similar conclusion for the location of the nucleotide binding domain in bile acid 7-hydroxylase and for glucose dehydrogenase. It seems likely that the *N*-terminal part of dihydrodiol dehydrogenase contains the nucleotide binding domain.

The similarities reported here indicate that oxidoreductases for a diverse group of substrates that includes: metabolism of polyols (e.g. ribitol, glucitol), aromatic hydrocarbons (toluene, naphthalene, polychlorinated biphenyls), over a dozen *Streptomyces* antibiotics, and steroids (e.g. 17β-oestradiol, testosterone, corticosterone) are derived from a common ancestor. Moreover, the organisms containing these enzymes appear to use them for different biological functions. They can be used to supply a source of carbon and energy for cell growth (e.g. glucitol-6-phosphate dehydrogenase, ribitol dehydrogenase) or they can be used to synthesize molecules that act as intercellular signals (e.g. actIII, 17β-hydroxysteroid dehydrogenase, 11β-hydroxysteroid dehydrogenase) [8]. This is a very impressive example of how gene duplication and divergence can lead to a shift in the function of the enzyme from metabolism to that of synthesis. It is this shift that we proposed was important in the

Table 1. ALIGN comparisons of *Pseudomonas* dihydrodiol dehydrogenases with *Streptomyces coelicolor* actIII protein, human 17β-hydroxysteroid dehydrogenase, various dehydrogenases, *Rhizobia* proteins, and adipocyte p27 protein

The ALIGN analysis was used with the Dayhoff scoring matrix with a bias of 6 and a gap penalty of 8; 1000 comparisons of randomized sequences of these proteins were use for statistical analysis.

	Dihydrodiol dehydrogenase	
	P. pseudo-alcaligenes	*P. putida*
S. coelicolor actIII	11.1	12.0
K. aerogenes ribitol dehydrogenase	10.7	9.95
E. coli glucitol-6-phosphate dehydrogenase	9.3	8.1
Human 17β-hydroxysteroid dehydrogenase	7.7	6.35
Eubacterium bile acid 7-dehydroxylase	12.7	14.4
R. meliloti nodG protein	11.85	9.8
B. japonicum fixR protein	9.5	9.45
Mouse adipocyte p27 protein	10.3	11.55
B. megaterium glucose dehydrogenase	10.25	11.6

origins of stable molecules that were adapted for use in intercellular communications, as seen in *Rhizobia*–plant interactions and, of course, in steroid-mediated processes in mammals [8].

M. E. B. is a Visiting Scientist in the laboratory of Dr. Joseph J. Salvo in the Biological Sciences Laboratory, GE Research and Development, whose support is appreciated. I thank D. Abramowicz, J. Salvo, and J. Yates for comments on an earlier version of this manuscript.

Michael E. BAKER

Department of Medicine, M-023,
University of California, San Diego,
La Jolla, CA 92093, U.S.A.

Received 27 November 1989

1. Furukawa, K., Arimura, N. & Miyazaki, T. (1987) J. Bacteriol. **169**, 427–429
2. Irie, S., Doi, S., Yorifuji, T., Takagi, M. & Yana, K. (1987) J. Bacteriol. **169**, 5174–5179
3. Zylstra, G. J. & Gibson, D. T. (1989) J. Biol. Chem. **264**, 14940–14946
4. Taira, K., Hayase, N., Arimura, N., Yamashita, S., Miyazaki, T. & Furukawa, K. (1988) Biochemistry **27**, 3990–3996
5. Sayler, G. S., Reid, M. C., Perkins, B. K., Pagni, R. M., Smith, R. L., Rao, T. K., Epler, J. L., Morrison, W. D. & DuFrain, R. (1982) Arch. Environm. Contam. Toxicol. **11**, 577–581
6. Reineke, W. & Knackmuss, H.-J. (1988) Annu. Rev. Microbiol. **42**, 263–287
7. Baker, M. E. (1989) Mol. Endocrinol. **3**, 881–884
8. Baker, M. E. (1990) FASEB J. **4**, 222–226
9. Peltoketo, H., Isomaa, V., Maentausta, O. & Vihko, R. (1988) FEBS Lett. **239**, 73–77
10. Malpartida, F., Hallam, S. E., Kieser, H. M., Motamedi, H., Hutchinson, C. R., Butler, M. J., Sugden, D. A., Warren, M., McKillop, C., Bailey, C. R., Humphreys, G. O. & Hopwood, D. A. (1987) Nature (London) **325**, 818–821
11. Hallam, S. E., Malpartida, F. & Hopwood, D. A. (1988) Gene **74**, 305–320
12. Agarwal, A. K., Mondor, C., Eckstein, B. & White, P. C. (1989) J. Biol. Chem. **264**, 18939–18943
13. Yamada, M. & Saier, M. H., Jr. (1987) J. Biol. Chem. **262**, 5455–5463
14. Morris, H. R., Williams, D. H., Midwinter, G. G. & Hartley, B. S. (1974) Biochem. J. **141**, 701–713
15. Debelle, F. & Sharma, S. B. (1986) Nucleic Acids Res. **14**, 7453–7472
16. Fisher, R. F., Swanson, J. A., Mulligan, J. T. & Long, S. R. (1987) Genetics **117**, 191–201
17. Thony, B., Fisher, H.-M., Anthamatten, D., Bruderer, T. & Hennecke, H. (1987) Nucleic Acids Res. **15**, 8479–8499
18. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) Methods Enzymol. **91**, 524–545
19. Coleman, J. P., White, W. B., Lijewski, M. & Hyleman, P. B. (1988) J. Bacteriol. **170**, 2070–2077
20. Jany, K.-D., Ulmer, W., Froschle, M. & Pfleiderer, G. (1984) FEBS Lett. **165**, 6–10
21. Navre, M. & Ringold, G. M. (1988) J. Cell Biol. **107**, 279–286
22. Murdock, G. L., Chin, C.-C. & Warren, J. C. (1986) Biochemistry **25**, 641–646
23. Wierenga, R. K., Terpstra, P. & Hol, W. G. (1986) J. Mol. Biol. **187**, 101–107
24. Jornvall, H., von Bahr-Lindstrom, H., Jany, K. D., Ulmer, W. & Froschle, M. (1984) FEBS Lett. **165**, 190–196