Volume 301, number 1, 89–93 FEBS 10922 April 1992

Expansion of the mammalian 3β-hydroxysteroid dehydrogenase/plant dihydroflavonol reductase superfamily to include a bacterial cholesterol dehydrogenase, a bacterial UDP-galactose-4-epimerase, and open reading frames in vaccinia virus and fish lymphocystis disease virus

Michael E. Baker[a] and Rafael Blasco[b]

[a]*Department of Medicine, 0623B, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0623, USA* and [b]*Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA*

Received 28 January 1992; revised version received 27 February 1992

Mammalian 3β-hydroxysteroid dehydrogenase and plant dihydroflavonol reductases are descended from a common ancestor. Here we present evidence that *Nocardia* cholesterol dehydrogenase, *E. coli* UDP-galactose-4 epimerase, and open reading frames in vaccinia virus and fish lymphocystis disease virus are homologous to 3β-hydroxysteroid dehydrogenase and dihydroflavonol reductase. Analysis of a multiple alignment of these sequences indicates that viral ORFs are most closely related to the mammalian 3β-hydroxysteroid dehydrogenases. The ancestral protein of this superfamily is likely to be one that metabolized sugar nucleotides. The sequence similarity between 3β-hydroxysteroid dehydrogenase and the viral ORFs is sufficient to suggest that these ORFs have an activity that is similar to 3β-hydroxysteroid dehydrogenase or cholesterol dehydrogenase, although the putative substrates are not yet known.

Steroid dehydrogenase; Vaccinia virus; Evolution

1. INTRODUCTION

Computer analyses of protein sequences have uncovered surprising similarities between proteins that would not be expected to be homologs. Thus, 3β-hydroxysteroid dehydrogenase [1,2], which converts pregnenolone to progesterone, — a key step in the synthesis of steroid hormones — has a common ancestor with plant dihydroflavonol reductases [3], enzymes used to synthesize pigments [4–7]. Recently, the ancestry of 3β-hydroxysteroid dehydrogenase became even more interesting when the enzyme was found to be homologous to open reading frames (ORFs) in vaccinia virus [8,9] and fish lymphocystis disease virus [9,10]. The similarity to the vaccinia ORF was strong enough (35% identity over a 340 amino acid segment) to suggest that this ORF may have an activity similar to 3β-hydroxysteroid dehydrogenase, although a function for steroid metabolism in the life cycle of vaccinia virus is not immediately apparent.

In extension of these findings, we report here that two bacterial proteins with other enzyme activities are members of this superfamily. One is *Nocardia* cholesterol dehydrogenase [11]; the other is *E. coli* UDP-galactose-4-epimerase [12]. The homology of steroid dehydrogenases and dihydroflavonol reductases to a bacterial enzyme that recognizes UDP-galactose is surprising. In view of the strong sequence similarity between proteins that have separated at least 2 billion years ago from a common ancestor, it is likely that other members of this superfamily, possibly with other novel activities, will be found in organisms between bacteria and mammals.

Correspondence address: M.E. Baker, Department of Medicine, 0623B, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0623, USA. Fax: (1) (619) 534 1424.

2. RESULTS AND DISCUSSION

Table I summarizes the results of the ALIGN analysis in which *Nocardia* cholesterol dehydrogenase is compared with human 3β-hydroxysteroid dehydrogenase, *E. coli* UDP-galactose-4-epimerase, corn and petunia dihydroflavonol reductase, and the ORFs of vaccinia virus and fish lymphocystis disease virus. Most of the comparison scores are above 10 standard deviations ($P=10^{-23}$) which is why we propose that these proteins are homologs, that is descended from a common ancestor.

To better understand the evolution of these proteins, we constructed a phylogenetic tree, depicting the relative genetic distance for members of this diverse protein superfamily, using the progressive alignment program of Feng and Doolittle [14] and either distance matrix or parsimony methods [15]. Figs. 1 and 2 present a multiple sequence alignment and a tree derived from the

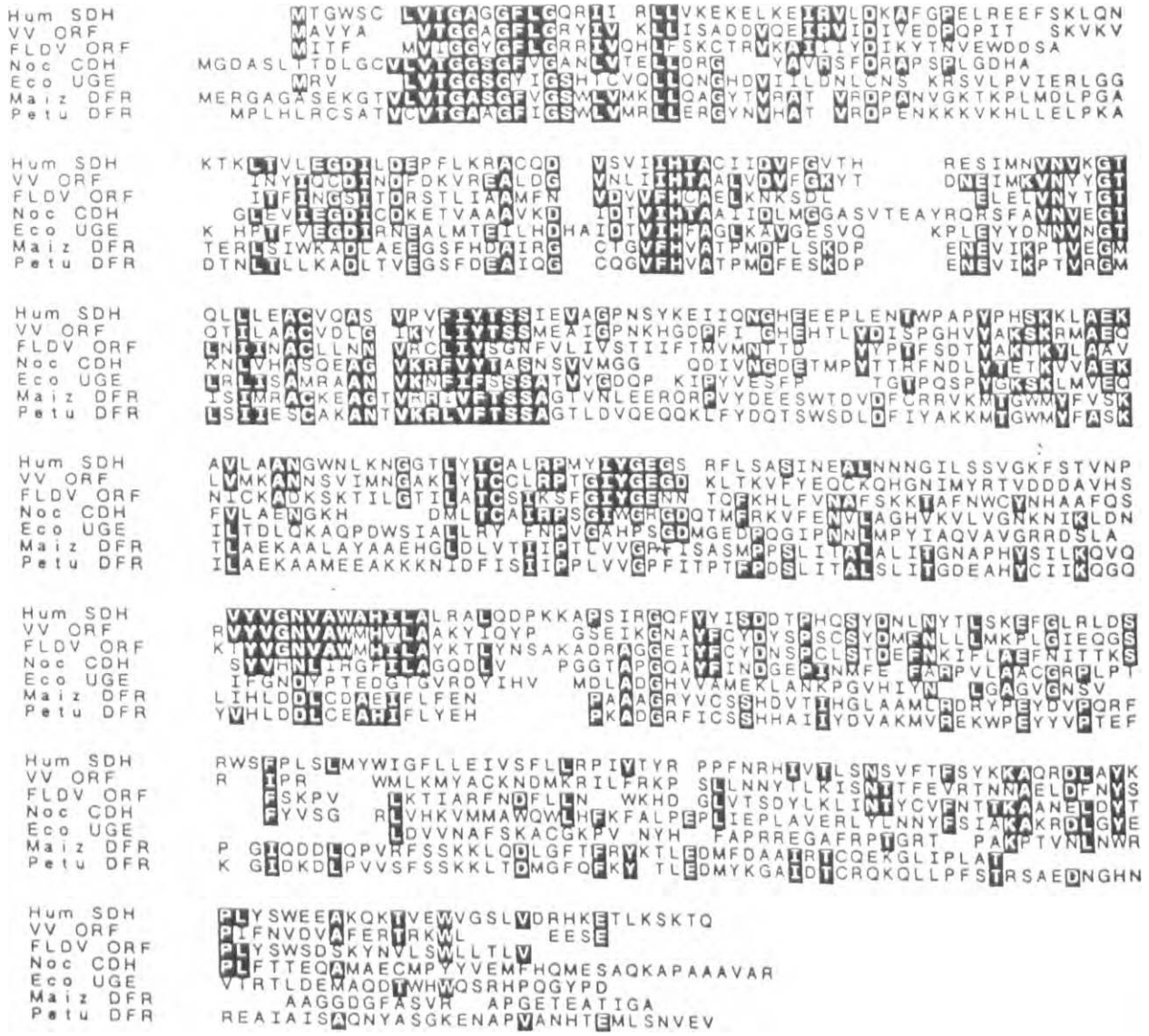

Fig. 1. A multiple alignment of mammalian 3β-hydroxysteroid dehydrogenases, open reading frames in vaccinia virus and fish lymphocystis disease virus, *Nocardia* cholesterol dehydrogenase, *E. coli* UDP-galactose-4-epimerase, and maize and petunia dihydroflavonol reductases. Sequences with three identities in one position are shown. In some instances the six sequences have two different amino acids conserved in three proteins at one position in the alignment. It is clear that the amino terminus is highly conserved among all six proteins and is likely to be essential for functioning of these proteins. Other areas where four to six amino acids are conserved also are likely to be important. Site-specific mutagenesis studies will help define the functions of these residues.

alignment. As expected, the tree shows the close grouping of the mammalian steroid dehydrogenases and the clustering of the plant dihydroflavonol reductases. Interestingly, the vaccinia virus ORF is the most closely related, among the superfamily members, to mammalian 3β-hydroxysteroid dehydrogenase.

2.1 *Similarity between mammalian 3β-hydroxysteroid dehydrogenase and* Nocardia *cholesterol dehydrogenase*

Although human 3β-hydroxysteroid dehydrogenase's principal activity is thought of as catalyzing the Δ^5-Δ^4 isomerization and C3 oxidation of pregnenolone to yield progesterone, this enzyme also catalyzes oxidation and reduction of C19 steroids at C3 [16,17]. Thus, the finding that a bacterial cholesterol dehydrogenase is homologous to human 3β-hydroxysteroid dehydrogenase is not too surprising in veiw of their similar oxidoreductase activity at C3 on a steroid substrate. The region of strong similarity between 3β-hydroxysteroid dehydrogenase and cholesterol dehydrogenase covers almost the entire sequence of each enzyme (Fig. 1).

2.2 *Similarity of cholesterol dehydrogenase to plant dihydroflavonol reductases*

The region of similarity between cholesterol dehydrogenase and the plant dihydroflavonol reductases covers the first 200 residues of each enzyme. The lack of se-

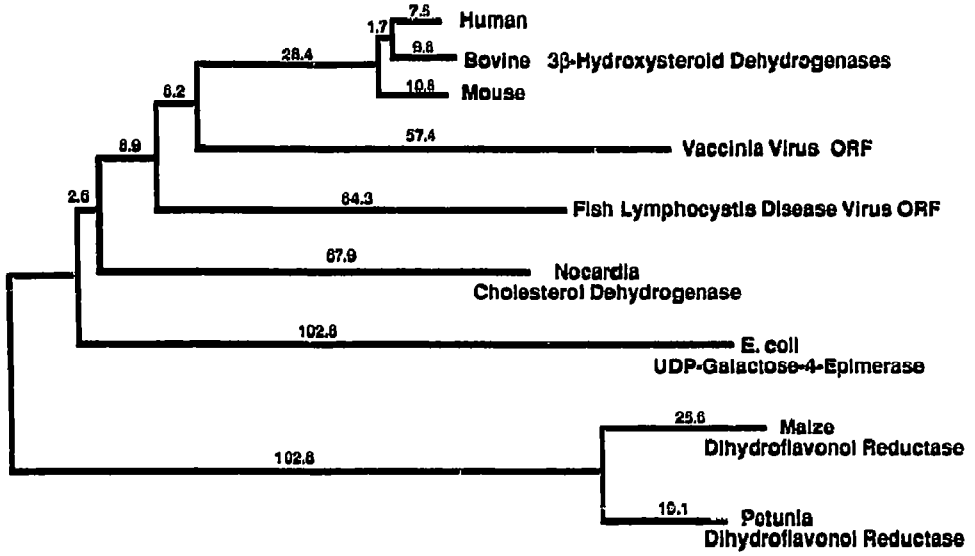

Fig. 2. An unrooted phylogenetic tree for the proteins aligned in Fig. 1. Branching points represent gene duplications. The lengths of the branches are proportional to the genetic distances calculated according to Feng and Doolittle [14].

quence similarity in the remaining ≈160 residues in the C-terminal part of these enzymes may arise from this part containing the substrate recognition domain, which, in view of the differences in the enzymes' substrates, would have the greatest differences in sequence and tertiary structure. However, the high ALIGN comparison scores indicate that cholesterol dehydrogenase and dihydroflavonol reductase are homologs, thereby supporting a conclusion reached by comparison of plant dihydroflavonol reductases with mammalian 3β-hydroxysteroid dehydrogenase [3].

Table I

ALIGN comparisons of *Nocardia* cholesterol dehydrogenase with human 3β-hydroxysteroid dehydrogenase, *E. coli* UDP-galactose-4-epimerase, plant dihydroflavonol reductase, open reading frames in vaccinia virus and fish lymphocystis disease virus

Protein	Comparison score (Standard deviation units)
Human 3β-hydroxysteroid dehydrogenase	14.15
Corn dihydroflavonol reductase	13
Petunia dihydroflavonol reductase	9.65
E. coli UDP-galactose-4-epimerase	10
Vaccinia virus ORF	12.55
Fish Lymphocystis disease virus ORF	12

Similarity between proteins was analyzed with the ALIGN program [13]. This program calculates the best alignment between any pair of sequences using the empirically derived Dayhoff scoring matrix and a penalty for breaking a sequence (gap penalty). The score for the two sequences is compared with that obtained from comparison of random permutations of the two sequences. The alignment score is the number of standard deviations by which the maximum score for the real sequences exceeds the average maximum score for the random. For the analyses reported here, from 500 to 2000 random permutations were used for the statistical analysis, and the Dayhoff matrix was used with a bias of 6 and a gap penalty of 8. The probability of getting a comparison score of 10 standard deviations by chance is 10^{-23} and, thus, it is very unlikely that the similarity between the two sequences is a result of convergent evolution. The more likely explanation is that the protein sequences are descended from a common ancestral gene and, therefore, are homologs.

The similarity of 3β-hydroxysteroid dehydrogenase and cholesterol dehydrogrnase, two steroid metabolizing enzymes, to dihydroflavonol reductases is of physiological interest because flavonoids bind to the mammalian estrogen receptor [18–21] and to type II binding sites in rat uterus [22], as well as enzymes that metabolize steroids [23] and prostaglandins [24]. Moreover, flavonoids have hormonal activity in mammals [18–23] and appear to protect against breast and colon cancer [25–27]. The similarity reported here suggests that there is an evolutionary linkage for some of these biological activities of flavonoids.

2.3 *Ancestry of dehydrogenases that have sugar nucleotides as substrates*

The most surprising finding is that mammalian 3β-hydroxysteroid dehydrogenase, bacterial cholesterol dehydrogenase, and plant dihydroflavonol reductase belong to the same superfamily as UDP-galactose-4-epimerase, an enzyme that recognizes a suger/nucleic acid substrate, modifying the configuration of the sugar's C4 hydroxyl group.

UDP-galactose-4-epimerase is an interesting dehydrogenase because it has a tightly bound nucleotide cofactor, which abstracts a proton from the substrate and then adds the proton in the opposite stereochemical configuration. The *E. coli* enzyme has homologs in other microorganisms including *Salmonella typhimurium* [28–30], *Streptomyces lividans* [31] and *Streptococcus thermophilus* [32]. Although these homologous

enzymes may have different nucleotide-diphosphate-sugar substrates, they appear to have common elements in their enzymatic mechanism of action [33]. Some homologs of UDP-galactose-4-epimerase act on CDP-hexoses. Thus, CDP-4-keto-3,6-dideoxy-D-galactose can be reduced by two different, but homologous enzymes, abequose synthase [29] and paratose synthase [30] to CDP-abequose and CDP-paratose, depending on the stereochemical addition of the proton. These reductases release the nucleotide cofactor after reduction of the substrate, in contrast to that of UDP- galactose-4-epimerase. Moreover, CDP-tyvelose epimerase, which epimerizes the C2 hydroxyl on paratose to yield CDP-tyvelose, is a member of this superfamily [30]. This family of sugar nucleotide converting enzymes is an excellent example of gene duplication and divergence to yield a family of enzymes with specificity for different sugar nucleotides and even different sites on the sugar substrate.

2.4 *Evolution of this protein superfamily*

What is surprising and not easily explainable, with the data available, are the events that led to a diverse family of proteins that include CDP-sugar dehydrogenases, UDP-C4 and CDP-C2 sugar epimerases, 3β-hydroxysteroid dehydrogenase, dihydroflavonol reductases, and virus ORFs. We consider it likely that the ancestral enzyme is one that recognizes sugar nucleotides, and that the flavonoid and steroid metabolizing enzymes arose later, for the following reasons. First, sugar nucleotides are likely to have formed in the prebiotic soup of compounds synthesized by chemical means before replicating life forms arose and would be part of the RNA world that is now thought to be the basis for the first life forms [34–37]. Sugar nucleotides are remarkably versatile compounds, and the enzymes that form, convert, and metabolize them have important roles in cellular biology. In addition to facilitating the metabolism of sugars for generating ATP, they are used in the synthesis of complex carbohydrates, sugar-lipids for membranes, and even glycosylation of aromatics such as steroids. This makes it likely that an ancestor of UDP-galactose-4-epimerase, using a sugar nucleotide substrate, would be present in the earliest microorganism, especially because these reactions can be accomplished under anaerobic conditions. All of this suggests that enzymes that catalyze reactions resembling those of UDP-galactose-4-epimerase are likely to be more ancient than enzymes that metabolize steroids or flavonoids.

2.5 *What is the function of the open reading frames in vaccinia virus and fish lymphocystis disease virus?*

The role(s) of the homologous ORFs in vaccinia virus and fish lymphocystis disease virus is not known. Vaccinia virus and fish lymphocystis disease virus belong to two different virus families, Poxviridae and Iridoviridae. Although both are large DNA viruses, their basic organization and replication strategies are different. Poxviruses are brick-shaped viruses, with a large genome (190 kb in vaccinia virus) that contains hairpin structures at the ends and inverted terminal repeats (for review see [38]). On the other hand, Iridoviruses are icosahedral viruses with circularly permuted and terminally redundant genomes [39–41]. We consider that those basic differences indicate that the two virus families represent independent evolutionary lineages. This idea is in good agreement with the phylogenetic tree, in which the vaccinia and fish lymphocystis disease viruses come in two independent branches. This branching pattern suggests that ancestral viruses for the two families picked up genes from their respective hosts, and accordingly, that the two viral genes are the result of two different hosts to virus transfer events. It is not known if other kinds of viruses contain homologs to these ORFs. Perhaps, identification of homologs in other viruses would help in resolving its genealogy.

In any event, as discussed in this report it is clear that the 3β-hydroxysteroid dehydrogenase superfamily is diverse, with a novel genealogy. Elucidation of the activities of these proteins and the identification of other members of the superfamily are likely to lead to information useful in treating endocrine and viral diseases.

REFERENCES

1 Luu-The, V., Lachance, Y., Labrie, C., Leblanc, G., Thomas, J.L., Strickler, R.C. and Labrie, F. (1989) Mol. Endocrinol. 3, 1310–1312.
2 Zhao, H.-F., Simard, J., Labrie, C., Breton, N., Rheaume, E., Luu-The, V. and Labrie, F. (1989) FEBS Lett. 259, 153–157.
3 Baker, M.E., Luu-The, V., Simard, J. and Labrie, F. (1990) Biochem. J. 269, 558–559.
4 Rohde, W., Barzen, E., Marocco, A., Schwarz-Sommer, Z., Saedler, H. and Salamini, F. (1987) Barley Genetics V, 533–541.
5 Schwarz-Sommer, Z., Shepherd, N., Tacke, E., Gierl, A., Rohde, W., Leclerq, L., Mattes, M., Berndtgen, R., Peterson, P.A. and Saedler, H. (1987) EMBO J. 6, 287–294.
6 Beld, M., Martin, C., Huits, H., Stuitje, A. and Gerats, G.M. (1989) Plant Mol. Biol. 13, 491–502.
7 Reddy, A.R., Britsch, L., Salamini, F., Saedler, H. and Rohde, W. (1987) Plant Sci. 52, 7–13.
8 Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P. and Paoletti, E. (1990) Virology 179, 247–266.
9 Blasco, R., Cole, N.B. and Moss, B. (1991) J. Virol. 65, 4598–4608.
10 Schnitzler, P. and Darai, G. (1989) Virology 172, 32–41.
11 Horinouchi, S., Ishizuka, H. and Beppu, T. (1991) Appl. Environmental Microbiol. 57, 1386–1393.
12 Lemaire, H.-G. and Muller-Hill, B. (1986) Nucleic Acids Res. 14, 7705–7711.
13 Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–545.
14 Feng, D. and Doolittle, R.F. (1987) J. Mol. Evol. 25, 351–360.
15 Felsenstein, J. (1988) Annu. Rev. Genet. 22, 521–565.
16 Lorence, M.C., Murry, B.A., Trant, J.M. and Mason, J.I. (1990) Endocrinology 126, 2493–2498.
17 Orlowski, J. and Clark, A.F. (1991) Endocrinology 128, 872–884.
18 Baker, M.E. (1991) Steroids 56, 354–360.
19 Shemesh, M., Lindner, H.R. and Ayalon, N. (1972) J. Reprod. Fert. 29, 1–9.

20 Shutt, D.A. and Cox, R.I. (1980) Endocrinology 52, 299–304.
21 Martin, P.M., Horwitz, K.B., Ryan, D.S. and McGuire, W.L. (1978) Endocrinology 103, 1860–1867.
22 Markaverich, B.M., Roberts, R.R., Alejandro, M.A., Johnson, G.A., Middleitch, B.S. and Clark, J.H. (1988) J. Steroid Biochem. 30, 71–78.
23 Ibrahim, A.-R. and Abul-Hajj, J.J. (1990) J. Steroid Biochem. Mol. Biol. 37, 257–260.
24 Degen, G.H. (1990) J. Steroid Biochem. 35, 473–479.
25 Adlercreutz, H., Hockerstedt, K., Bannwart, C., Bloigu, S., Hamalainen, E., Fotsis, T. and Ollus, A. (1987) J. Steroid Biochem. 27, 1135–1144.
26 Adlercreutz, H. (1984) Gasteroenterology 86, 761–766.
27 Setchell, K.D.R., Borriello, S.P., Hulme, P., Kirk, D.N. and Axelson, M. (1984) Am. J. Clin. Nutre. 40, 569–578.
28 Wyk, P. and Reeves, P. (1989) J. Bacteriol. 171, 5687–5693.
29 Verma, N. and Reeves, P. (1989) J. Bacteriol. 171, 5694–5701.
30 Houng, H.-S.F., Kopecko, D.J. and Baron, L.S. (1990) J. Bacteriol. 172, 4392–4398.
31 Poolman, B., Royer, T.J., Mainzer, S.E. and Schmidt, B.F. (1990) J. Bacteriol. 172, 4037–4047.
32 Adams, C.W., Fornwald, J.A., Schmidt, F.J., Rosenberg, M. and Brawner, M.E. (1988) J. Bacteriol. 170, 203–212.
33 Gabriel, O. (1978) Trends Biochem. Sci. 3, 193–195.
34 Pace, N.R. (1991) Cell 65, 531–533.
35 Joyce, G.F. (1989) Nature 338, 217–224.
36 Benner, S.A., Ellington, A.D. and Tauer, A. (1989) Proc. Natl. Acad. Sci. USA 86, 7054–7058.
37 Darnell, J.E. and Doolittle, W.F. (1986) Proc. Natl. Acad. Sci. USA 83, 1271–1275.
38 Moss, B. (1991) Annu. Rev. Biochem. 59, 661–688.
39 Goorha, R. and Murti, K.G. (1982) Proc. Natl. Acad. Sci. USA 79, 248–252.
40 Delius, H., Darai, G., Flugel, R.M. (1984) J. Virol. 49, 609–614.
41 Darai, G., Delius, H., Clarke, J., Apfel, H., Schnitzler, P. and Flugel, R.M. (1985) Virology 146, 292–301.