Volume 279, number 1, 5–8 FEBS 09357 February 1991

ADONIS 001457939100075O

Evidence for gene duplication forming similar binding folds for NAD(P)H and FAD in pyridine nucleotide-dependent flavoenzymes

James H. McKie and Kenneth T. Douglas

Department of Pharmacy, University of Manchester, Manchester, M13 9PL, UK

Received 13 November 1990; revised version received 4 December 1990

For pyridine nucleotide-dependent flavoenzymes, binding both FAD and NAD(P)H on a single amino-acid chain, we have found a high degree of internal sequence similarity for certain regions of the FAD and NAD(P)H binding portions of the chain for any given protein. This was the case for a range of enzyme classes, including disulphide oxidereductases (such as glutathione reductase, trypanothione reductase, lipoamide dehydrogenase, mercuric reductase), mono- and dioxygenases, nitrite reductase, alkyl hydroperoxidase and NADH dehydrogenase from *E. coli*. This provides strong support for gene duplication as the origin of at least part of the FAD and NAD(P)H recognising domains of such enzymes.

Flavoenzyme; Gene duplication; Glutathione; Lipoamide; Reductase; Trypanothione

1. INTRODUCTION

From detailed X-ray diffraction analyses of glutathione reductase from human erythrocytes, Schulz et al. [1] pointed out the similarity of the FAD- and NADP-domains of this enzyme with respect to their chain-folding and ligand-binding modes. They suggested that this might indicate a distant evolutionary relationship, perhaps even a gene duplication [1,2,4]. In conflict with this, no sequence homology was detected between the FAD- and NADPH-binding regions of this enzyme [3] — a hint of sequence level homology was suggested, however, from comparisons of residues 23–31 (part of the FAD-domain) with residues 190–198 (of the NADPH-domain) [3].

In modelling the coenzyme-binding regions of a related enzyme [5], trypanothione reductase, on the framework of the atomic coordinates for human erythrocyte glutathione reductase [8] we searched sequence data banks for other NADPH- (or NADH-) dependent flavoenzymes having sequence similarities to the FAD regions of glutathione reductase (human [6], mouse [6] and *E. coli* [7] enxymes).

From this analysis it became clear that for NAD(P)H-dependent flavoenzymes, from a wide range of mechanistic classes, there is a high degree of sequence similarity in the FAD- and NAD(P)H-binding region of the amino-acid chain.

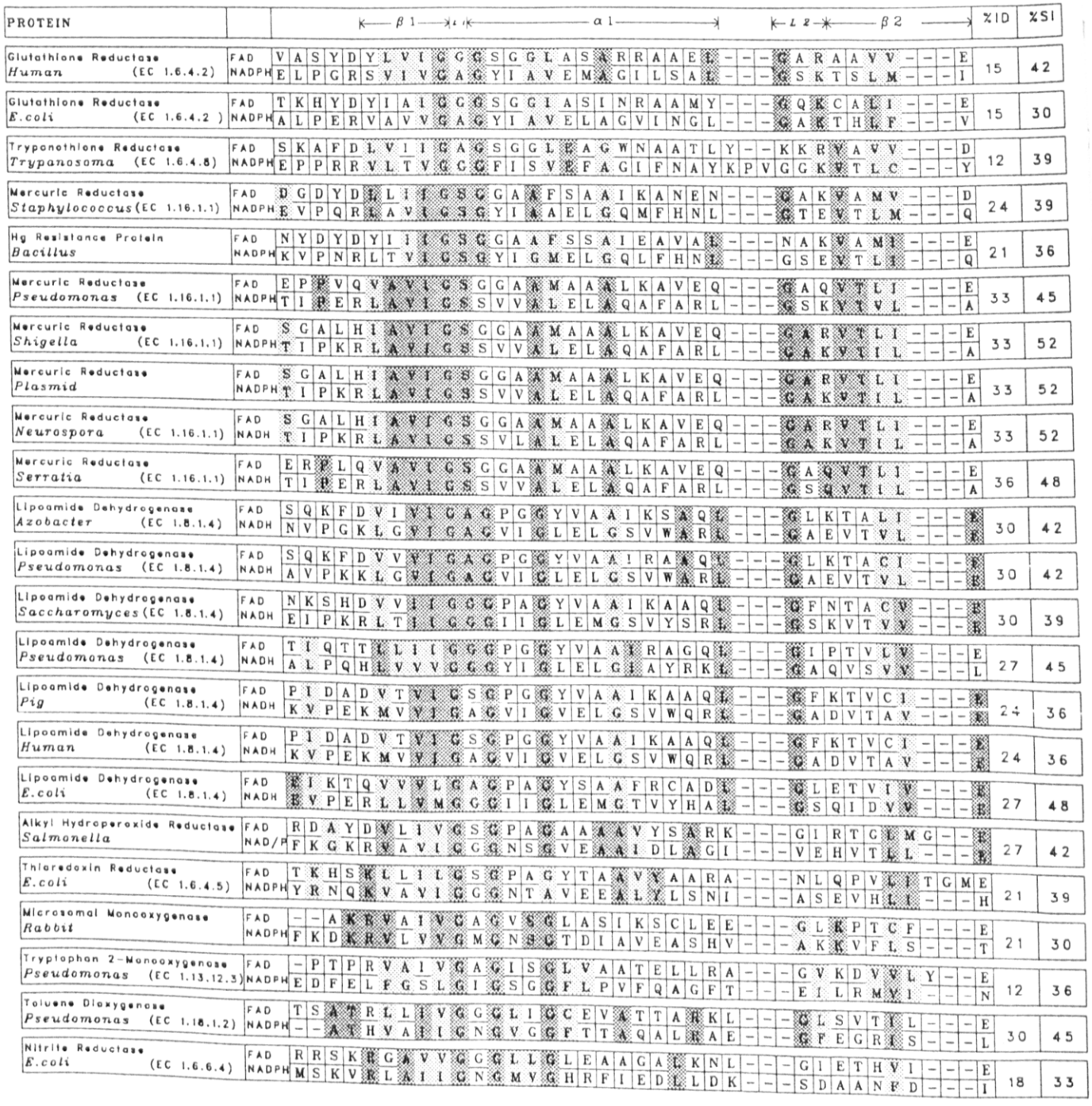
Correspondence address: K.T. Douglas, Department of Pharmacy, University of Manchester, Manchester, M13 9PL, UK

2. MATERIALS AND METHODS

Initial scans of the OWL composite sequence database [9] were performed using Sweep [10] and the PAM250 matrix with the protein sequences of human and *E. coli* glutathione reductase (EC 1.6.4.2), and *Trypanosoma congolense* trypanothione reductase. Sequences that showed high similarity with one or more of these proteins such as *Staphylococcus aureus* mercuric reductase (EC 1.16.1.1) and *Azotobacter vinelandii* lipoamide reductase (EC 1.8.1.4), were used to rescan the OWL database to locate additional sequences. This process produced sequences that showed marked similarities at both the FAD and the NAD(P)H 'GXG' sites. The predicted FAD regions were aligned with one another within each protein sequence. Similarly NAD(P)H regions were aligned. In this way FAD and NAD(P)H sites on each protein chain were assigned. Next, FAD and NAD(P)H regions for each protein were aligned against each other by eye and by means of the PAM matrix. Using the LUPES software [9], template motifs of the 'GXG' regions for FAD, the NAD(P)H, and the combined FAD + NAD(P)H sequences were made and used to make additional searches of the OWL database. The only protein sequences used were those which showed pairs of 'GXG' sites on the same chain that were predicted to correspond to FAD- and NAD(P)H-binding domains. Molecular graphics studies were performed using QUANTA software on a Silicon Graphics 240D model computer graphics system.

3. RESULTS AND DISCUSSION

Table I shows the sequences of 33-residue fragments of the FAD- and NADPH-binding domains of a number of NADPH- (or NADH-) dependent flavoenzymes. For each enzyme, the FAD-binding sequence is aligned optimally with the NADPH-binding sequence for that same enzyme. Complete identities between FAD- and NADPH-linked sequences range from 12–36%, and if similar amino acids (with respect to size, likely charge-state, hydrophobicity etc.) are also paired (similarity defined as indicated in the legend to Table I) the degrees



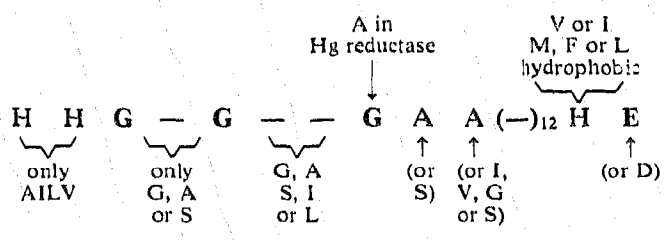
Identical residues are shown against a dark box; similar residues are shown against a grey stippled box. The following groups of amino acids were regarded as similar for this table (M, I, V, L, C, F); (E, D); (K, R); (A, G); (S, T). Percentage identity is calculated for the number of identities between the FAD and NAD(P)H sequences of 33 amino acids given. Percentage overall similarity is calculated using the number of pairwise identical or similar residues in the sequences given.

of sequence similarity over all of the enzymes range from 30% to 52%. The enzymes include examples from various mechanistic families, e.g. disulphide oxidoreductases (glutathione reductase, trypanothione reductase, thioredoxin reductase, lipoamide dehydrogenase), mercuric reductases, nitrite reductase, alkyl hydroperoxidase, monooxygenases as well as dioxygenases. The lipoamide dehydrogenases represent a wide evolutionary range and the similarities in sequence between FAD and NAD sites are clustered around particular parts of the chain, and this distribution equally is relatively invariant throughout the enzymes available in Table I. Taken

together, the high degree of sequence similarity within any one protein between FAD- and NAD(P)H-recognition sequences and the distribution of the conserved regions clearly support the idea of a gene duplication as originally suggested by Schulz [1].

The $\beta_1\alpha\beta_2$ region covered by the sequences in Table I corresponds to the region of the protein which interacts with the adenosine pyrophosphoryl moiety of the coenzyme (Fig. 1). The fragments for FAD and NADPH are separated on any given protein sequence by 140–190 residues, with the FAD domain N-terminal relative to the NAD(P)H domain. For glutathione reductase (human) the X-ray crystal structure unambiguously assigns these sites. The close sequence analogy between glutathione reductase (human erythrocute) and *T. congolense* trypanothione reductase (~60%) locates the sites for this enzyme. Assignment of the other FAD and NADPH sequences of Table I was done on the basis of homology to these sequences. Alignments were made of the putative FAD-binding sequences for all of the enzymes of Table I with identical and similar residues defined as in this table. Similarly, putative sequences for NADPH binding in these enzymes were aligned and compared.

FAD sequences have been compared previously [11, 12] and Wierenga et al. [12] have suggested that a FAD fold similar to that of glutathione reductase and *p*-hydroxybenzoate hydroxylase may occur in ferredoxin-$NADP^+$ oxidoreductase [13], lipoamide dehydrogenase and in mercuric reductase. They also noted some similarity in sequence for D-amino acid oxidase and NADH dehydrogenase in the region of the $\beta_1\alpha\beta_2$ units covered by the alignments in Table I. Our analysis of this more extensive data set confirms close sequence homology in this $\beta_1\alpha\beta_2$ region for FAD enzymes of a wide range of mechanistic classes. Wierenga and Hol [14] pointed out that the **G X G X X G** consensus was found in many nucleotide-binding proteins including flavoenzymes, pyridine-nucleotide dependent enzymes and protein kinases [14]. Based on analysis of the data in Table I, there are clearly alternatives to the C-terminal glycine of this consensus. A summary of the consensus sequence of the $\beta\alpha\beta$ FAD region in Table I, looking only at the FAD sequences, is:



(where H = a hydrophobic residue).

The NAD(P)H sequences for this set of enzymes, also show homology. Of the **G X G X X G** consensus only the first G is totally conserved, the second G can be replaced by S in some mercuric reductases; the final G is not conserved at all. We are analysing currently these preferences in terms of the known geometries of these binding sites on glutathione reductase and relative to non-FAD-dependent NAD(P)H enzymes.

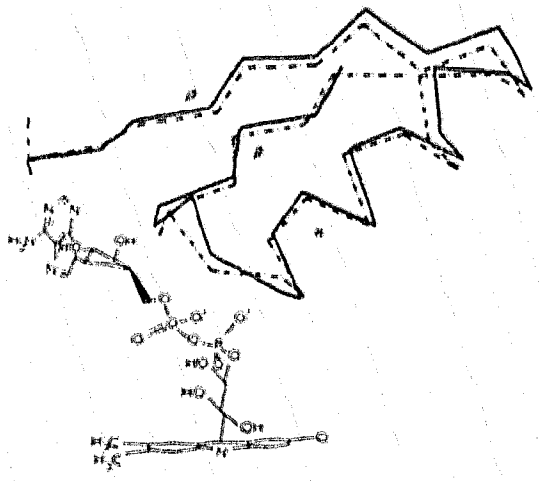


Fig. 1. Computer graphics generated view of the $\beta_1\alpha\beta_2$ region of the FAD- and NADPH-binding regions of human erythrocyte glutathione reductase described in Table I. The $\beta\alpha\beta$ motif for the NADPH pocket is superimposed on that for the FAD pocket (solid and dotted lines). The relationship to the bound FAD coenzyme is indicated. Coordinates used were from the Brookhaven Protein Data Bank for the X-ray diffraction data set described in reference [8].

In summary, the FAD and NADPH sequences of a wide range of NAD(P)H-dependent flavoenzymes show marked homology with each other for any given enzyme in support of the gene duplication suggested by Schulz [1]. In view of the well-defined common 3D fold used by FAD and NADPH in human glutathione reductase and by FAD in *p*-hydroxybenzoate hydroxylase, it is likely that, for the rather limited section ($\beta_1\alpha\beta_2$) of the total FAD (or NAD(P)H) fold concerned, a common 3D motif will be used for ligand recognition by the nucleotide-binding sequences of Table I.

Acknowledgements: We are grateful to the Wellcome Trust for support and to Prof. G.E. Schulz for provision of the coordinates for GSSG bound to human erythrocyte glutathione reductase.

REFERENCES

[1] Schulz, G.E., Schirmer, R.H. and Pai, E.F. (1980) in: Flavins and Flavoproteins, Proc. Int. Symp. 6th., pp. 557–567.
[2] Schulz, G.E., Schirmer, R.H., Sachsenheimer, R. and Pai, E.F. (1978) Nature 273, 120–124.
[3] Untucht-Grau, R., Schirmer, R.H., Schirmer, I. and Krauth-Siegel, R.L. (1981) Eur. J. Biochem. 120, 407–410.
[4] Schulz, G.E. (1980) J. Mol. Biol. 138, 335–347.
[5] Shames, S.L., Kimmel, B.E., Peoples, O.P., Agabian, N. and Walsh, C.T. (1988) Biochemistry 27, 5014–5019.
[6] Tutic, M., Lu, X., Schirmer, R.H. and Werner, D. (1990) Eur. J. Biochem. 188, 523–528.
[7] Greer, S. and Perham, R.N. (1986) Biochemistry 25, 2736–2742.

[8] Karplus, P.A. and Schulz, G.E. (1987) J. Mol. Biol. 195, 701–729.
[9] Bleasby, A.J. and Wooton, J.C. (1990) Prot. Eng. 3, 153–159.
[10] Ackrigg, D., Bleasby, A.J., Dix, N.I.M., Findlay, J.B.C., North, A.C.T., Parry-Smith, D., Wooton, J.C., Blundell, T.L., Gardner, S.P., Hayes, F., Islam, S., Sternberg, M.J.E., Thornton, J.M., Tickle, I.J. and Murray-Rust, P. (1988) Nature 335, 745–746.
[11] Williams, Jr, C.H., Arscott, L.D. and Schulz, G.E. (1982) Proc. Natl. Acad. Sci. USA 79, 2199–2201.
[12] Wierenga, R.K., Drenth, J. and Schulz, G.E. (1983) J. Mol. Biol. 167, 725–739.
[13] Sheriff, S. and Herriott, J.R. (1981) J. Mol. Biol. 145, 441–451.
[14] Wierenga, R.K. and Hol, W.G.J. (1983) Nature 302, 842–844.