SEQUENCE HOMOLOGIES AMONG PYRIDINE NUCLEOTIDE-LINKED DEHYDROGENASES: POSSIBLE PARTIAL GENE DUPLICATIONS IN GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE

Paul C. ENGEL

Department of Biochemistry, University of Sheffield, Sheffield S10 2TN, England

Received 23 March 1973

1. Introduction

Homologous amino acid sequences neighbouring the essential cysteine residues of several dehydrogenases including glyceraldehyde-3-phosphate dehydrogenase (GPDH) suggest a possible common evolutionary ancestry [1–3], although, for such relatively short peptide homologies, the possibility of convergent evolution cannot be discounted. Although glutamate dehydrogenase (GDH) has no essential cysteine residues, the sequences surrounding the essential lysine 126 in GDH and lysine 212 in GPDH exhibit significant homology [4, 5]. Both these lysines react readily with pyridoxal-5′-phosphate.

GDH possesses within a single subunit both a catalytic site and one or more regulator sites specific for the catalytic coenzymes and their structural analogues ADP and GTP. Engel [6] suggested that the evolution of such regulator sites, perhaps better described as "homosteric" than "allosteric", might be most easily explained by partial gene duplication. The apparent duplication in GDH of an extensive region containing lysine 126 was adduced in support of this hypothesis.

The recent report [7] that GPDH also possesses a separate regulator site for ATP, ADP and NAD therefore prompted a re-examination of the amino acid sequence of this enzyme. The present paper reports evidence of a partial gene duplication in GPDH analogous to that found in GDH.

		111												
GPDH	a	GLY	GLY	ALA	LYS	Arg	VAL	Ile	Ile	Ser	Ala	PRO	Ser	Ala	ASP
	b	GLY				Lys		Val		Ser				Ala	ASP
	c	Ala				Lys		Val		Thr				Ser	Thr
		209													
GPDH	a	GLY	Ala	ALA	LYS	ALA	VAL	Gly	Lys	Val	Ile	PRO	Glu	Leu	ASP
	b										Ile				ASP
	c										Leu				Gln
		209													
LADH		Met	GLY	Cys	LYS	ALA	Ala								

Register shift = 98 residues.

Fig. 1. Internal comparisons of 2 sections of the amino acid sequence of GPDH. The sequences [8, 9] of pig (a), lobster (b) and yeast (c) GPDH are shown on separate lines where they differ. Capital letters are used to emphasise the internal homologies where they occur. A small segment of the sequence of LADH [10] is also shown for comparison.

2. Results

Fig. 1. shows an alignment of 2 sections of the GPDH sequences published by Harris and his colleagues [8, 9]. One of these, containing lys 212, has been shown previously [4–6] to be homologous with the duplicated sequence in GDH. In these two highly-conserved pieces of sequence 6 positions out of 14 are occupied by identical residues.

The same figure also shows part of the sequence of horse liver alcohol dehydrogenase [10]. Jörnvall detected significant in-register homology between GPDH and LADH from residues 1 to 60 but not beyond. In view of the homologies between GDH and GPDH it is perhaps significant that LADH also

```
                                                    Homology 2A
    239
a   VAL VAL ASP LEU Thr Cys  Arg Leu  Glu LYS
b                       VAL  Arg      Gly
c                       VAL  Lys      Asp

    321
a   VAL VAL ASP LEU Met VAL His Met  Ala Ser
b         Ile           Leu Lys      Met Gln LYS
c         VAL           Val Glu      Val/Ile Ala LYS

              Register shift = 82 residues.
```

```
                                                                    Homology 2B
    205
a   PRO ALA SER  Thr Gly Ala  Ala LYS ALA VAL Gly LYS Val Ile  Pro GLU
b         Ser                                                   Ile
c         Ser                                                   Leu

    249
a   PRO ALA Lys  Tyr Asp Asp Ile  LYS Lys  VAL Val LYS Gln Ala  Ser GLU
b   Glu  Cys  SER       Asp       ALA Ala  Met      Thr      Ser
c   Glu  Thr  Thr       Glu       Lys VAL  Val      Ala      Ala

                  Register shift = 44 residues.
```

```
                Homology 2C
    109
a   LEU Lys Gly Gly ALA LYS
b   Phe Lys Gly
c   Ile Asp Ala

    154
a   LEU Ala Pro Leu ALA LYS
b             Val
c             Leu

  Register shift = 45 residues.
```

```
                                                    Homology 2D
    5
a   VAL ASP Gly Phe GLY Arg Ile  Gly Arg LEU Val  THR Arg –
b   ILE                                    Leu
c   ILE                                    Met

    93
a   VAL Glu  Ser  Thr GLY Val Phe Thr Thr Met  Glu  Lys  Ala Gly
b   VAL Glu                      Thr Thr Ile  Glu  Lys       Ser
c   ILE ASP                      Lys Glu LEU  ASP  THR       Gln
```

```
    18
a   ALA Ala PHE Asn Ser GLY Lys Val Asp Ile  VAL Ala  Ile
b   ALA     LEU Ser Cys GLY –   Ala Gln VAL            Val
c   Ile     LEU Ser Arg Pro Asx Val Glx VAL            SER/THR

    109
a   ALA His LEU Lys Gly GLY Ala Lys Arg VAL Ile  Ile  SER
b   ALA     PHE Lys Gly             Lys      VAL       SER
c   Lys     Ile Asp Ala             Lys      VAL       THR
```

```
    31
a   Asn. Asp PRO Phe Ile  ASP Leu His  Tyr Met VAL
b                        Ala      Glu       Met VAL
c   Asx Asx PRO          Asx      Asp       Ala Ala

    120
a   –  Ala PRO Ser Ala ASP Ala Pro Met Phe VAL
b                  Ala ASP
c                  Ser Thr

          Register shift = 88–89 residues.
```

```
                                Homology 2E
271
Leu GLY TYR Thr Glu Asp Gln VAL Val Ser
                        Asp
                        Ala

315
Phe GLY TYR Ser Asn Arg Val VAL Asp Leu
Phe             Gln     Ile
Tyr             Thr     VAL

          Register shift = 44 residues.
```

Fig. 2. Further internal comparisons of the amino acid sequence of GPDH displayed as in fig. 1.

has lysine at position 212. LADH has no lysine at position 114, but there is lysine at position 113 (cf. fig. 1).

Whilst the partial gene duplication suggested by the evidence in fig. 1 may be related to the presence of regulator sites, attention must also be drawn to several other fragments of apparent sequence homology shown in fig. 2. In most of the positions not showing identity the alterations can be accounted for in terms of single base changes or by a simple chain of such mutations in those positions where the pig, lobster and yeast enzymes differ.

A further homology between beef liver GDH and pig liver GPDH has also been found and is shown in fig. 3. Five of the ten residues compared are identica The sequence of the 'B' peptide from *Neurospora* NADP-linked GDH [11] bears out this homology, adding an identity in a sixth position. With the exception of the change of threonine to valine, which requires two nucleotide base changes, all replacements can be accounted for by single base changes in the gen ome. It is interesting that the beef liver GDH fragment shows greater similarity to the fragment from GPDH than to the Neurospora GDH fragment.

3. Discussion

The homologies presented above contribute furthe evidence that the pyridine nucleotide-linked dehydro genase share common ancestry. Without further information from chemical modification studies and 3-dimensional structure determinations one cannot be confident that the amino acid duplications now documented in both GDH and GPDH represent the duplication of functional binding sites. On the evidence to date, however, it is attractive to speculate that duplication of a lysine-containing sequence in an ancestral protein is reflected in the present structures of GDH, GPDH and LADH, and that, while

```
                    164
Beef GDH            VAL PRO ALA PRO ASN Met SER Thr  GLY Glu

                    232
GPDH            a   VAL PRO Thr PRO ASN Val SER VAL Val  Asp
                b               PRO ASP
                c               Val Asx

Neurospora GDH      VAL PRO ALA Gly Asp Ile Gly VAL GLY Gly
'B' peptide
```

Fig. 3. Comparison of the amino acid sequences of GDH from beef liver [12] and *Neurospora* [11] with those of GPDH [8, 9 displayed as in fig. 1 and 2.

the duplicated sequence is perhaps non-functional in LADH, it has evolved a regulatory function in GDH and GPDH.

The different register shifts shown in figs. 1 and 2 may be an indication that the observed homologies have arisen by chance or through convergent evolution. On the other hand they may reflect repeated partial gene duplication in the evolution of these proteins. Homologous regions of DNA in the gene, once established, would facilitate the recurrence of such duplications. In this connection it may be noted that:

i) sequences 111... and 209... are homologous
ii) each of these sequences is homologous with a sequence nearer the carboxyl end, the register shift being in one case 44, and, in the other, 45, residues.
iii) homology 2D, which again involves residues 111–124, requires a register shift of 88, exactly twice 44.
iv) homology 2E, between 2 sequences near the carboxyl terminal, again requires a register shift of 44 residues.

These relationships may reflect random coincidence. Nevertheless it seems clear that the available dehydrogenase sequences should be subjected to a close and systematic scrutiny by the methods developed by Fitch [13]. Such a study should provide good evidence for or against the hypothesis of repeated duplication and is now in progress.

References

[1] J.I. Harris, Nature 203 (1964) 30.
[2] T.P. Fondy, J. Everse, G.A. Driscoll, F. Castillo, F.E. Stolzenbach and N.O. Kaplan, J. Biol. Chem. 240 (1965) 4219.
[3] J.J. Holbrook, G. Pfleiderer, K. Mella, M. Volz, W. Leskowac and R. Jeckel, European J. Biochem. 1 (1967) 476.
[4] E.L. Smith, M. Landon, D. Piszkiewicz, W.J. Brattin, T.J. Langley and M.D. Melamed, Proc. Nat. Acad. Sci. U.A. 67 (1970) 724.
[5] B.G. Forcina, G. Ferri, M.C. Zapponi and S. Ronchi, European J. Biochem. 20 (1971) 535.
[6] P.C. Engel, Nature 241 (1973) 118.
[7] J. Ovádi, M. Nuridsány and T. Keleti, Acta Biochim. Biophys. Acad. Sci. Hung. 7 (1972) 133.
[8] J.I. Harris and R.N. Perham, Nature 219 (1968) 1025.
[9] G.M.T. Jones and J.I. Harris, FEBS Letters 22 (1972) 185.
[10] H. Jörnvall, European J. Biochem. 14 (1970) 521.
[11] J.C. Wootton, G.K. Chambers, J.G. Taylor and J.R.S. Fincham, Nature 241 (1973) 42.
[12] K. Moon, D. Piszkiewicz and E.L. Smith, Proc. Nat. Acad. Sci. U.A. 69 (1972) 1380.
[13] W.M. Fitch, J. Mol. Biol. 16 (1966) 9.