Volume 246, number 1,2, 1-5 FEB 06936 March 1989

Pyruvate decarboxylase is like acetolactate synthase (*ILV2*) and not like the pyruvate dehydrogenase E1 subunit

Jeremy B.A. Green

European Molecular Biology Laboratory, Meyerhofstrasse 1, D6100 Heidelberg, FRG

Received 23 January 1989

Protein sequences of pyruvate decarboxylase (PDC) derived from cloned yeast (*Saccharomyces cerevisiae*) and bacterial (*Zymomonas mobilis*) genes were compared with each other and with sequence databases. Extensive sequence similarities were found between them and with two others: cytochrome-linked pyruvate oxidase from *Escherichia coli* and acetolactate synthase (*ilvI* in *E. coli*; *ILV2* gene in *S. cerevisiae*). All catalyse decarboxylation of pyruvate using thiamine pyrophosphate (TPP) as cofactor. General overall similarity suggests common ancestry for these enzymes. None of the sequences was similar to the E1 component of pyruvate dehydrogenase from *E. coli* which also decarboxylates pyruvate with the help of TPP.

Thiamine pyrophosphate; Sequence homology; Protein sequence; Data base, FASTP; Data base, SWISS-PROT

1. INTRODUCTION

In this paper I present polypeptide sequence similarities that point to evolution of an enzyme family and provide important information for enzymologists pursuing structure/function analysis of any of these molecules. Although one pair of similarities that I have included were already known, I clearly demonstrate the existence of a larger family which may hereby be designated the pyruvate decarboxylase (PDC) family. I also describe an enzyme that is catalytically, essentially identical to PDC but is, by sensitive sequence criteria, structurally unrelated.

Thiamine pyrophosphate (vitamin B_1) is a cofactor whose biochemical functions and mechanistic role are well understood (e.g. [1,2]). Some of the enzymes that use it and whose genes have been cloned are listed (with references) in table 1. PDC links glycolysis to alcohol dehydrogenase and thence ethanol synthesis in the yeast *Saccharomyces cerevisiae* and the bacterium *Zymomonas mobilis*. Acetolatate synthase is part of the highly conserved isoleucine-valine biosynthesis pathway and is perhaps better known as the product of the *ilvI* (*Escherichia coli*) and *ILV2* (*S. cerevisiae*) genes. POX (cytochrome) is a membrane-bound enzyme of *E. coli* of unknown physiological role [10], while PDH E1 subunit is known to link glycolysis to respiration via acetyl-CoA. Other TPP-binding enzymes are known but genes for these four have been cloned. From the start-point of the yeast PDC sequence I show that PDC, ILV and POX sequences are highly similar while that of PDH E1 is quite different.

Correspondence address: J.B.A. Green, Embryogenesis Laboratory, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, England

Abbreviations: TPP, thiamine pyrophosphate; PDC, pyruvate decarboxylase; PDH, pyruvate dehydrogenase; POX, pyruvate oxidase; ILV, acetolactate synthase (isoleucine valine biosynthetic enzyme)

2. METHODS AND RESULTS

Using the program FASTP [11], I compared the *S. cerevisiae* pyruvate decarboxylase protein sequence derived from the nucleotide sequence of the gene *PDC1* [3] with the SWISS-PROT protein se-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

Table 1

Some TPP- and pyruvate-binding enzymes

Enzyme name (number)	Reaction catalysed	Cloned sequence	Refs
Pyruvate decarboxylase (EC 4.1.1.1)	Pyruvate = acetaldehyde + CO_2	SCPDC1 ZMPDC	3 4
Acetolactate synthase (EC 4.1.3.18)	2 pyruvate = acetolactate + CO_2	ECILVI ECILVG SCILV2	5 6 7
Pyruvate oxidase (cytochrome) (EC 1.2.2.2)*	Pyruvate + H_2O + ferricytochrome *b*1 = acetate + CO_2 + ferrocytochrome *b*1	ECPOXB	8
Pyruvate dehydrogenase, E1 subunit (EC 1.2.4.1)	Pyruvate + lipoamide = *S*-acetyldihydrolipoate + CO_2	ECACEE	9

* The IUB Enzyme Nomenclature Commission (1984) [19] recommends the name pyruvate dehydrogenase (cytochrome). However, 'pyruvate oxidase, (cytochrome)' is the name used in all the cited references as well as in the sequence databases. This enzyme need not be confused with EC 1.2.3.3 (also known as pyruvate oxidase) which produces acetyl-phosphate and hydrogen peroxide and which is also a TPP-requiring enzyme

quence database. (The SWISS-PROT protein sequence database at EMBL, Heidelberg, FRG, is based on the PIR protein sequence databank (National Biomedical Research Foundation, Washington, DC, USA) but includes translations of some EMBL nucleotide sequence database entries.) The similarity score of acetolactate synthase, encoded by the *ilvI* gene in *E. coli* [5] was sufficiently high to indicate a genuine structural likeness (see below). The *ilvI* protein sequence and that of an isoenzyme, *ilvG*, [6] also scored highly when I compared *Zymomonas* PDC DNA sequence [4] translation with the SWISS-PROT database. The next highest score, that for sheep vasopressin, was not sufficiently above the average score for the database to warrant further analysis.

The similarities detected between these peptide sequences by FASTP were not detected when comparing nucleotide sequences against the EMBL nucleotide database using FASTP.

The similarities between PDC and acetolactate synthase proteins suggested that other proteins with similar catalytic functions might likewise show structural resemblance. The sequence of *E. coli* POX (cytochrome) was already known to be similar to that of the ILV gene [8]. Fig.1 shows dot-matrix similarity plots generated by the DIAGON program [12] for the PDC, POX and ILV proteins. The program compares sequences on the basis of amino acid similarity using a statistically derived interchangeability score for all possible amino acid pairs [15,16]. The regions of extended similarity appear as diagonal lines on or near the main diagonal. They show varying degrees of interrelatedness and two particularly well-conserved domains detectable in almost all of the diagons.

The extensive similarities suggested that it would be possible to align all five sequences on a single axis. I did this using programs GAP and LINEUP [13] (fig.2). Note that there are some regions in which alternative alignments are equally plausible. To the computer-generated alignment I have added marks indicating positions of better-conserved residues (see fig.2 legend for details). Regions of greatest similarity can be seen at positions 100–200 and 545–600 of fig.2. There is also a more dispersed similarity between these two regions.

The E1 subunit of pyruvate dehydrogenase catalyses much the same reaction as PDC and might therefore be expected to have some structural resemblance. The DIAGON program [13] was used with a range of parameter values to try and detect regions of structural similarity between the PDH sequence [9] and any of the other TPP pyruvate enzymes. No regions of similarity were detectable (not shown). No significant similarity was detected either by the more sophisticated program of Argos [14], which uses physical as well as

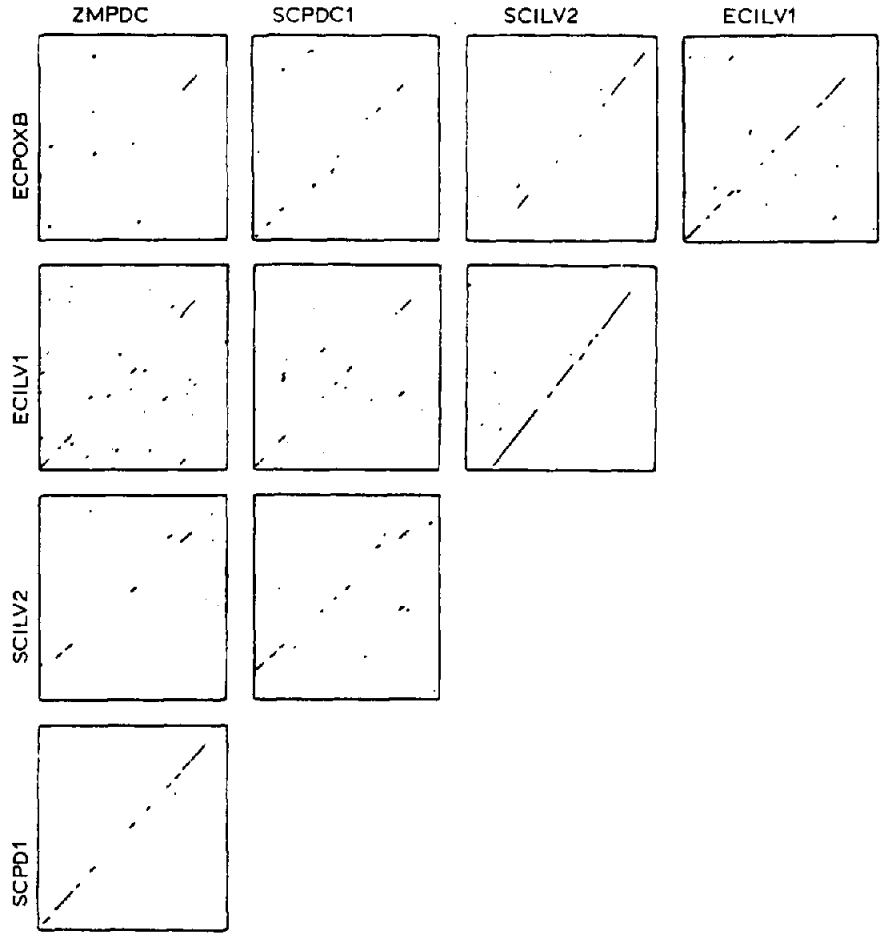

Fig.1. DIAGON [12] comparisons of TPP- and pyruvate-binding protein sequences. Axes represent the two sequences being compared; N to C direction is bottom-to-top and left-to-right. Complete sequences (lengths in the range 550–690 amino acids) were used for all comparisons and sequence names are according to table 1. Diagonal lines indicate regions of extended similarity. Span length: 25, 'match' threshold 275 (defaults 11 and 300; see [12] for scoring system).

evolutionary criteria and scans all 'window' sizes for comparison (see [14]). As a check against sequencing errors in the published PDH E1 sequence, comparisons were also made with translations in alternative reading frames but no sequence similarities were uncovered thereby.

3. DISCUSSION

Protein sequences of pyruvate decarboxylases from *S. cerevisiae* [3] and *Zymomonas mobilis* [4] were found to be similar to each other and to those of acetolactate synthase [5–7] and *E. coli* pyruvate oxidase (cytochrome) [8]. The similarities are obvious when the sequences are aligned and since they extend over most of the length of each protein it is reasonable to conclude that the similarity arises from common ancestry: these enzymes constitute a homologous family [17].

Ullrich [18] found that modification of PDC tryptophan blocks TPP binding. None of the yeast PDC tryptophan residues is conserved in more than one other sequence but two of the five tryptophans in yeast PDC lie in relatively well-conserved domains (positions 132 and 587 in fig.2). Few other relevant protein chemical experiments that would help characterise functional sites in these enzymes are to be found in the literature and more are needed.

The absence of similarity between *E. coli*

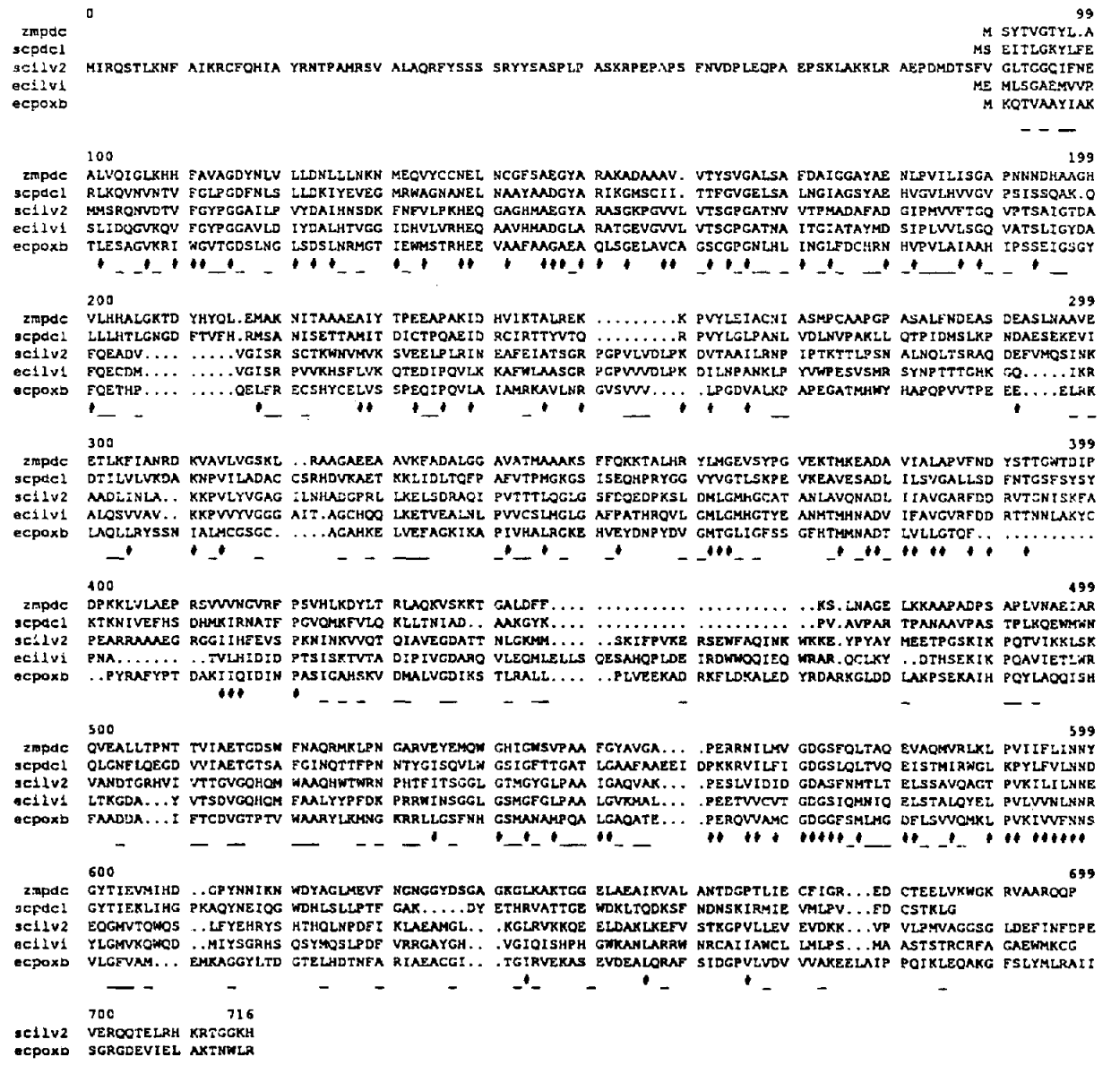

Fig.2. Alignment of five TPP pyruvate protein sequences. Made using GAP and LINEUP programs [13]. Numbering is arbitrary. Dots indicate spaces introduced to optimise alignment (corresponding to insertion or deletion during evolution). Symbols: (#) 4 or 5 matches; (–) 3 matches where a match means amino acids are identical or similar (i.e. scoring +2 or more on the similarity matrix of Dayhoff [15,16]).

pyruvate dehydrogenase E1 subunit (PDH E1) and any of the other enzymes is striking in view of the fact that it catalyses a virtually identical reaction. This may be an example of convergent evolution. An alternative explanation is that though PDH E1 may have had the same ancestor, its incorporation into a multienzyme complex has provided strong evolutionary pressures to make it fit better into the complex and enhance the catalytic advantage thereby gained. Primary sequence similarity to the other TPP pyruvate enzymes might thereby have been lost. As far as is known, none of the other enzymes

examined in this paper is part of such an enzyme complex.

Acknowledgements: Thanks to Patricia Kahn, Toby Gibson, Pat Argos, Brian Hartley, Fritz Zimmermann and Stefan Hohmann for practical help and comments. The author was supported via the Biotechnology Programme of the European Community.

REFERENCES

[1] Krampitz, L.O. (1969) Annu. Rev. Biochem. 38, 213–240.
[2] Mahler, H. and Cordes, E. (1966) Biological Chemistry, Harper and Row, New York.
[3] Kellermann, E., Seeboth, P.G. and Hollenberg, C.P. (1986) Nucleic Acids Res. 14, 8963–8977.
[4] Conway, T., Osman, Y., Konnan, J., Hoffmann, E. and Ingram, L.O. (1987) J. Bacteriol. 169, 949–954.
[5] Squires, C.H., Defelice, M., Devereux, J. and Calvo, J. (1983) Nucleic Acids Res. 11, 5299–5313.
[6] Lawther, R.P., Calhoun, D.H., Adams, C.W., Hauser, C.A., Gray, J. and Hatfield (1981) Proc. Natl. Acad. Sci. USA 78, 922–925.
[7] Falco, S.C., Dumas, K.S. and Livak, K. (1985) Nucleic Acids Res. 13, 4011–4027.
[8] Grabau, C. and Cronan, J.E. (1986) Nucleic Acids Res. 14, 5449–5460.
[9] Stephens, P.E., Darlison, M.G., Lewis, H.M. and Guest, J. (1983) Eur. J. Biochem. 133, 155–162.
[10] Chang, Y. and Cronan, J.E. (1986) J. Bacteriol. 167, 312–318.
[11] Lipman, D. and Pearson, W. (1985) Science 227, 1435–1441.
[12] Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961.
[13] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[14] Argos, P. (1987) J. Mol. Biol. 193, 385–396.
[15] Dayhoff, M., Barker, W. and Hunt, L. (1983) Methods Enzymol. 91, 524–545.
[16] Schwartz, R.M. and Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, vol. 5, suppl. 3, pp. 353–358, National Biomedical Research Foundation, Washington, DC.
[17] Reeck, G., De Haen, C., Teller, D., Doolittle, R., Fitch, W., Dickerson, R., Chambon, P., McLachlan, A., Margoliash, E., Jukes, T. and Zuckerhandl, E. (1987) Cell 50, 667.
[18] Ullrich, J. (1982) Ann. NY Acad. Sci. 378, 287–305.
[19] IUB Enzyme Nomenclature Commission (1984) Enzyme Nomenclature, Academic Press, Orlando, FL.