GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE: AMINO ACID SEQUENCE OF ENZYME FROM BAKER'S YEAST

G.M. Thelwall JONES and J. Ieuan HARRIS
Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Received 25 February 1972

1. Introduction

Glyceraldehyde 3-phosphate dehydrogenase* (GPDH) from the skeletal muscle of a wide range of organisms possesses very similar enzymic and molecular properties [1, 2]. Within mammals the primary structure is very highly conserved [3] and the amino acid sequence around the essential thiol group is identical in all the species that have been studied [3, 4]. More striking still is the considerable identity of the complete primary structure of enzyme from the muscle of such distantly related organisms as pig [5] and lobster [6]. Enzyme from baker's yeast on the other hand shows significant differences in properties, notably in the mode of binding of NAD [7–9], and in the reactivity of the essential thiol group towards alkylating agents [10, 11]. The latter observation is the more remarkable since the amino acid sequence of a 17-residue tryptic peptide containing the essential thiol group is identical in the yeast and mammalian enzymes [10].

As a further step towards the identification of the structural elements responsible for the differences in properties between the yeast and muscle enzymes, we report here a nearly complete amino acid sequence for the protein sub-unit of yeast GPDH.

2. Materials and methods

Glyceraldehyde 3-phosphate dehydrogenase (3X crystallised) was prepared from baker's yeast and shown to be homogeneous as previously described [12]. For sequence studies the enzyme was S-carboxymethylated with [2-^{14}C]iodoacetic acid in 8 M urea [13]. Peptide fragments that accounted for the entire protein chain of 331 residues were obtained from: a) total tryptic digests of *S*-[2-^{14}C] carboxymethylated protein; b) similar tryptic digests of non-radioactive *S*-carboxymethylated protein in which methionine residues had been reacted with [1-^{14}C] iodoacetamide in 8 M urea at pH 3.0; c) tryptic digests of lysine-blocked (*N*-trifluoroacetyl-[15] and *N*-maleyl-[16]) *S*-[2-^{14}C] carboxymethylated protein; d) cyanogen bromide cleavage (in 70% formic acid) of *S*-[2-^{14}C] carboxymethylated protein, by suitable combinations of gel-filtration on Sephadex columns (G-50, G-75 and G-100) and paper electrophoresis/chromatography (cf. [14]).

Peptide sequences were established by the dansyl-Edman method [17], and overlaps of many of the primary tryptic peptides were obtained from chymotrypsin and subtilisin digests of lysine-blocked fragments [18], and from pepsin digests of cyanogen bromide fragments. Overlaps that were not experimentally established have been assigned provisionally on the basis of apparent sequence homology with the pig and lobster muscle enzymes (cf. [5]). A detailed account of these procedures and results is available [19].

3. Results and discussion

A partial amino acid sequence for the sub-unit of yeast GPDH is given in fig. 1 in comparison with the

**Abbreviations*
GPDH: Glyceraldehyde 3-phosphate dehydrogenase.
ADH: Alcohol dehydrogenase.

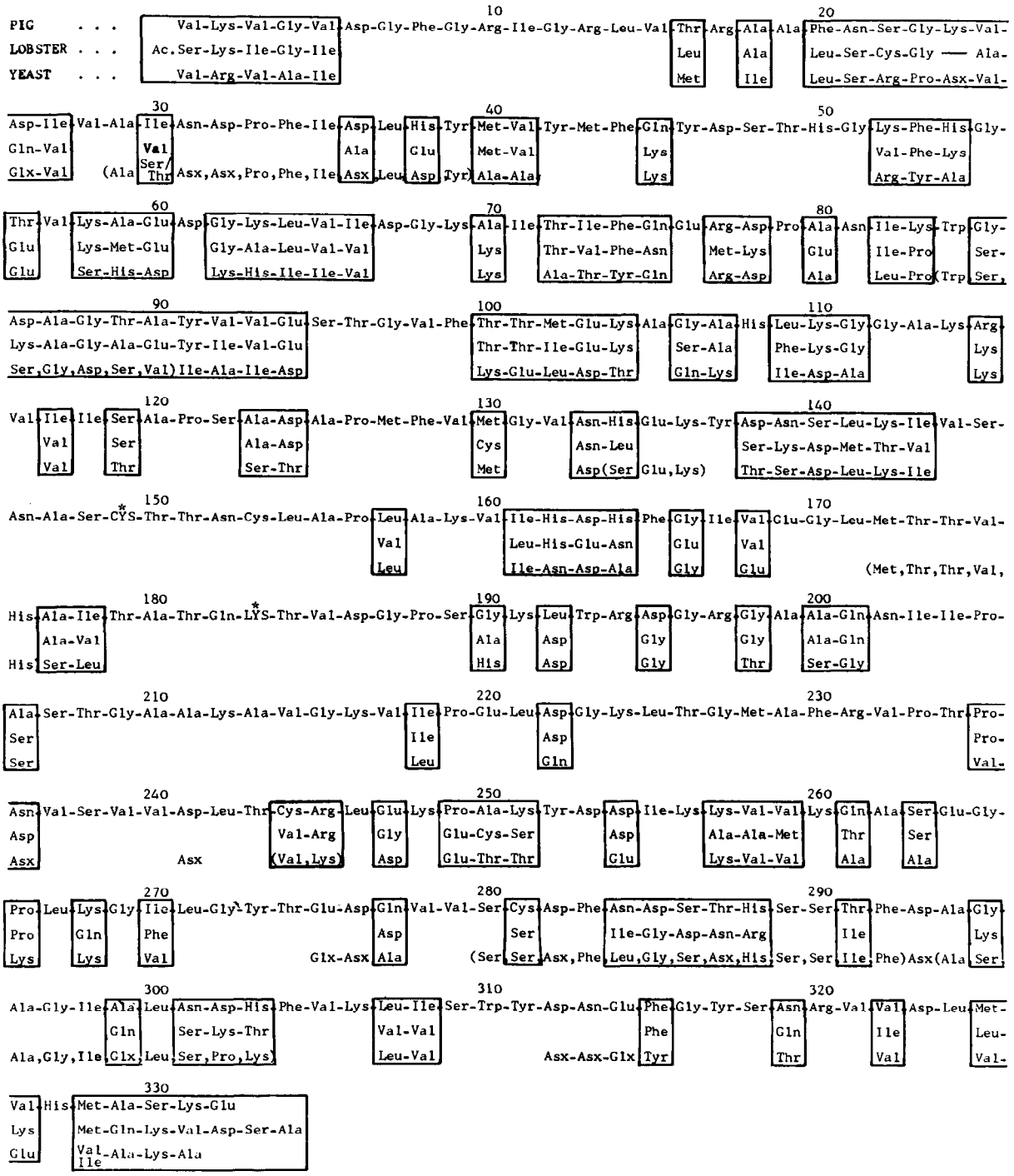

Fig. 1. Comparison of the amino acid sequence of GPDH from pig muscle, lobster muscle and baker's yeast. Established sequence differences are shown in the boxed regions. Sequences not experimentally determined for the yeast chain are given within parentheses and in a *provisional* order that maximises the sequence homology between the yeast and muscle enzymes. Asx and Glx denote residues in the yeast chain for which amide assignments were not unambiguously established. CŸS (149) and LŸS (183) are residues implicated in the active site.

complete sequences of the pig [5] and lobster [6] muscle enzymes. Incompleted sequences (shown within parentheses) are expressed provisionally in the order that maximizes sequence homology between the yeast and muscle enzymes. Peptides containing residue 30 gave fractional values for threonine and serine but it is not clear if this reflects the same kind of micro-heterogeneity as at residue 328 where valine and isoleucine were shown to occur in the ratio of 2:3 [19]. With these exceptions no evidence was found for similar but different sequences as would be expected from non-identical sub-units (cf. [20]). Moreover the facility with which the two amino acids at position 328 were detected emboldens us to believe that crystalline yeast GPDH, like its counterparts from muscle, is indeed a protein of four sub-units each containing 331 amino acids in essentially identical sequence.

A total of 277 residues have been placed in sequence by direct experiment (fig. 1) but, of these, 10 of the carboxylate side chains were not unambiguously assigned as free acids or amides. Comparison of the sequence of the pig and lobster muscle enzymes showed that 72% of the residues (241 out of 332) are identical [5]. Of the 267 residues that can be directly compared in all three species a total of 160 (60%) are identical and a similar value of 59% (195 out of 331) is obtained if the comparison is extended to include the 54 residues in the yeast chain that have been placed by homology. No other single protein from all three species has yet been sequenced but comparison of cytochrome *c* from pig muscle and yeast shows an identity of 62% where residues occur in both chains [21]. In the corresponding glyceraldehyde 3-phosphate dehydrogenases 68% of the residues are identical. Moreover it is of interest to note that a comparable degree of sequence identity (67%) exists between the yeast and lobster muscle enzymes showing a remarkably similar rate of molecular evolution of the pig and lobster lineages since their divergence.

These results show considerable overall homology in the primary structure of GPDH from yeast and muscle. This, in turn, implies a conservation of tertiary structure for which the notably higher than average conservation of sequence in certain parts of the protein chain may be particularly relevant. Thus, for example, 80% of the residues in a central section of the chain (residues 144 to 243) are identical in the three species. Significantly, therefore, this part of the chain contains the essential thiol group (residue 149) as well as lysine-183 shown to be the site of the S→N transacetylation reaction [22, 23] in the yeast [19] as well as the muscle [24, 25] enzymes. This reaction implies that these two residues can be juxtaposed in the three-dimensional structure and that the sequence of this region of the molecule may have been particularly strongly conserved because of its importance in maintaining the tertiary structure of the active site. It may also be significant that the highly conserved sequence around a lysine (residue 212) that reacts with pyridoxal phosphate in the rabbit muscle enzyme [26], is similar to the sequence around the lysine (residue 97) that reacts specifically with pyridoxal phosphate in glutamic dehydrogenase [27].

Aromatic residues in GPDH are notably conserved in all three species. Thus the 3 tryptophans, 8 of the tyrosines and 10 phenylalanines occur in corresponding positions in the three chains while phenylalanine residues at positions 53, 74 and 315 in the two muscle chains are changed to tyrosine in the yeast enzyme. It should also be noted that tyrosine-46 (followed by tyrosines 39 and 42) is the most reactive towards $K^{125}I_3$ in all three species [28] which again suggests a similar three-dimensional structure. In contrast, other residues such as cysteine and histidine are conserved to a much lower extent. Thus only 2 cysteines (residues 149 and 153) are common to all three species and of the eleven histidine residues that occur in the mammalian enzyme only four are common to the lobster and yeast enzymes. Significantly histidine-38, implicated in the catalytic activity of the rabbit enzyme [29], does not occur in either the lobster or yeast enzymes and if one assumes that enzyme from diverse sources possesses a similar basic reaction mechanism this particular histidine is unlikely to be an essential component of the active site.

Anaerobic conversion of carbohydrate to pyruvate is an essential metabolic step in virtually all living organisms and the extent to which the primary structure of GPDH has been conserved suggests that this enzyme (and in all probability other enzymes in the glycolytic pathway too) has existed essentially in its present form from a very early stage in the evolution of the species. A need for the internal packing of hydrophobic residues within sub-units, coupled with a requirement for the individual sub-units to interact in a highly specified manner so as to maintain the surface

structure of the active site in each of the sub-units, could well impose severe restrictions on the amino acid sequence of a multi-substrate enzyme such as glyceraldehyde 3-phosphate dehydrogenase. Indeed, recent studies of active site peptides from a variety of muscle (and liver) aldolases (e.g. [30, 31] are beginning to indicate a similarly high conservation of primary structure for this enzyme. Yeast aldolase [32] on the other hand appears not to be related to its counterparts in muscle and liver.

Comparison of the primary structure of GPDH from yeast and muscle has not enabled us to identify differences that could account for the changed reactivity of the essential thiol group, or the different NAD-binding properties of the yeast enzyme. Moreover there are no obvious features to explain the differences in sub-unit interaction between the muscle enzymes which exhibit negative co-operativity [7], and the yeast enzyme which exhibits both positive and negative co-operativity [8]; or, alternatively, obeys the allosteric model [9].

In comparing different NAD-linked enzymes there is (in addition to the segment of twelve residues with similar sequence in glyceraldehyde 3-phosphate and glutamic dehydrogenases, referred to earlier), an apparent similarity between the sequence of residues 52 to 63 in yeast GPDH and part of the tryptic peptide containing the reactive thiol group from yeast ADH [33]:

GPDH: -- Arg^{52}-Tyr-Ala-Gly-Glu-Val-Ser-His-Asp-Asp-Lys-His^{63}-

ADH: -Lys/Arg-Tyr-Ser-Gly-- Val-Cys-His-Thr-Asp-Leu-His-

This apart, the information at present available on the primary structure of different NAD-linked dehydrogenases (cf. [34]) has not revealed any decisive evidence to support the possibility that these enzymes could have originated from a common ancestral gene. It is, of course, entirely possible that sequence changes in glyceraldehyde 3-phosphate dehydrogenase may reflect complementary changes in other enzymes with which it forms coupled systems *in vivo* (lactic dehydrogenase in muscle and alcohol dehydrogenase in yeast) because of a possible need for protein–protein interactions. Clearly the resolution of these and of other phenomena concerned with the mechanism of action of NAD-linked dehydrogenases must await the solution of their detailed primary and three-dimensional structures.

References

[1] W.S. Allison and N.O. Kaplan, J. Biol. Chem. 239 (1964) 2140.
[2] J.I. Harris, in: Pyridine Nucleotide Dependent Dehydrogenases, ed. H. Sund (Springer, Berlin, 1970) pp. 57–70.
[3] R.N. Perham, Biochem. J. 111 (1969) 17.
[4] W.S. Allison and J.I. Harris, Abstr. 2nd Meet. European Biochem. Soc. Vienna (1965) 140.
[5] J.I. Harris and R.N. Perham, Nature 219 (1968) 1025.
[6] B.E. Davidson, M. Sajgò, H.F. Noller and J.I. Harris, Nature 216 (1967) 1181.
[7] A. Conway and D.E. Koshland, Jr., Biochemistry 7 (1968) 4011.
[8] R.A. Cook and D.E. Koshland, Jr., Biochemistry 9 (1970) 3337.
[9] K. Kirshner, E. Gallego, I. Schuster and D. Goodall, J. Mol. Biol. 58 (1971) 29.
[10] R.N. Perham and J.I. Harris, J. Mol. Biol. 7 (1963) 316.
[11] B. Eisele and K. Wallenfels, FEBS Letters 1 (1968) 25.
[12] P.J.G. Butler and G.M. Thelwall Jones, Biochem. J. 118 (1970) 375.
[13] J.I. Harris and R.N. Perham, J. Mol. Biol. 13 (1965) 876.
[14] J.I. Harris and H. Jörnvall, European J. Biochem. 13 (1970) 565.
[15] R.F. Goldberger and C.B. Anfinsen, Biochemistry 1 (1962) 401.
[16] P.J.G. Butler, J.I. Harris, B.S. Hartley and R. Leberman, Biochem. J. 112 (1969) 679.
[17] W.R. Gray, in: Methods in Enzymology, Vol. 11, ed. C.H.W. Hirs (Academic Press, New York, London, 1967) pp. 469–475.
[18] R.N. Perham and G.M.T. Jones, European J. Biochem. 2 (1967) 84.
[19] G.M. Thelwall Jones, PhD. Dissertation (University of Cambridge, 1969).
[20] H. Jörnvall, European J. Biochem. 16 (1970) 41.
[21] M.O. Dayhoff, Atlas of Protein Sequence and Structure, Vol. 4 (National Biomedical Research Foundation, Silver Spring, Maryland, 1969) D.213/214.
[22] L. Polgàr, Experientia 22 (1966) 232.
[23] J.H. Park, C.F. Agnello and E. Mathew, J. Biol. Chem. 242 (1966) 769.
[24] J.I. Harris and L. Polgàr, J. Mol. Biol. 14 (1965) 630.
[25] B.E. Davidson, European J. Biochem. 14 (1970) 545.
[26] B.G. Forcina, G. Ferri, M.C. Zapponi and S. Ronchi, European J. Biochem. 20 (1971) 535.
[27] E.L. Smith, M. Langdon, D. Piszkiewicz, W.J. Brattin, T.L. Langley and M.D. Melamed, Proc. Natl. Acad. Sci. U.S. 67 (1970) 724.
[28] J.O. Thomas and J.I. Harris, Biochem. J. 119 (1970) 307.
[29] J.S. Bond, S.H. Francis and J.H. Park, J. Biol. Chem. 245 (1970) 1041.
[30] I. Gibbons, P.J. Anderson and R.N. Perham, FEBS Letters 10 (1970) 49.

[31] S.-M. Ting, C.L. Sia, C.Y. Lai and B.L. Horecker, Arch. Biochem. Biophys. 144 (1971) 485.
[32] R.S. Jack and J.I. Harris, Biochem. J. 124 (1971) 68P.
[33] J.I. Harris, Nature 203 (1964) 30.
[34] H. Jörnvall, European J. Biochem. 14 (1970) 521.