Volume 229, number 2, 383–387 FEB 05669 March 1988

Sequence of the gene encoding phosphoglycerate mutase from *Saccharomyces cerevisiae*

Malcolm F. White and Linda A. Fothergill-Gilmore

Department of Biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD, Scotland

Received 6 January 1988

The gene encoding yeast phosphoglycerate mutase was isolated, and its sequence was determined. The gene specifies a protein of 246 amino acids, and contains no introns. The sequence shows a strong codon bias. The upstream untranslated portion of the gene contains a CT-rich block such as is found in many highly expressed yeast genes, but does not have the associated CAAG sequence.

Phosphoglycerate mutase gene; Nucleotide sequence; *(Saccharomyces cerevisiae)*

1. INTRODUCTION

Phosphoglycerate mutases comprise a family of enzymes which catalyse reactions involving the transfer of phospho groups among the three carbon atoms of phosphoglycerates. There are at least five types of phosphoglycerate mutase that are kinetically and structurally distinct, but nevertheless have many features in common (e.g. see [1–3]). The enzyme from *Saccharomyces cerevisiae* has been extensively characterized, and the high-resolution crystallographic structure and amino acid sequence have been reported [4,5]. We present here the isolation and sequence determination of the gene encoding this enzyme, as a prerequisite for protein engineering studies to investigate its mode of action.

2. EXPERIMENTAL

The gene encoding *S. cerevisiae* phosphoglycerate mutase (GPM) has been obtained in the multicopy plasmid YEp13 by complementation [6], and was kindly given to us by D.G. Fraenkel. A restriction map of the plasmid was constructed, and the coding sequence was located with the use of oligonucleotide probes. A *Hin*dIII/*Sal*I restriction fragment containing the entire coding region together with several hundred base pairs of flanking sequences was subcloned into phage M13 for sequence determination by the dideoxy method. The use of the Sequenase DNA sequencing kit (US Biochemical Corp., Cleveland, OH) allowed sequences of about 500 bases to be determined from a single primer.

Correspondence address: L.A. Fothergill-Gilmore, Department of Biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD, Scotland

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number X06408

3. RESULTS AND DISCUSSION

3.1. *Restriction mapping of YEp13-GPM*

Restriction mapping of the recombinant YEp13 plasmid revealed an insert of 6 kbp of *S. cerevisiae* chromosomal DNA (fig.1). The phosphoglycerate mutase coding sequence was located within this insert by Southern blotting with two end-labelled oligonucleotides corresponding to the N- and C-terminal regions of the protein sequence. Both oligonucleotides hybridized specifically to a 1.3 kbp *Hin*dIII/*Sal*I restriction fragment which was therefore subcloned into phage M13mp19 for sequencing.

3.2. *Nucleotide sequence of GPM gene*

The strategy used to sequence the phosphoglycerate mutase gene is summarised in

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

fig.1. A contiguous stretch of 1154 bases was sequenced which includes an open reading frame of 741 bases encoding the enzyme phosphoglycerate mutase (fig.2). The DNA sequence largely agrees with the previously determined protein sequence, and the differences are discussed below.

Phosphoglycerate mutase is typical of glycolytic enzymes in yeast in being highly expressed. Like other highly expressed yeast genes [7], GPM exhibits an extreme codon bias. Only 31 codons are used, and of these 24 account for 93% of the amino acids (phosphoglycerate mutase has no cysteine residues). This compares with values of 94% for phosphoglycerate kinase, 95% for enolase I and 95% for alcohol dehydrogenase I. The upstream flanking sequence also has features similar to those of other highly expressed yeast genes. These include A residues at −1 and −3, and a 44 base A-rich region containing no G residues immediately adjacent to the initiating ATG codon [8]. Another feature found in many highly expressed yeast genes is a CT-rich block, here present at −83 to −109, but there is no associated CAAG sequence [9]. A TATA box at position −138 may be involved in the initiation of transcription, and the sequence CACACA at position −16 is found in several other yeast genes although its function is not known [9]. The downstream flanking region is AT-rich (73%), and contains two possible polyadenylation/transcription termination signals TAG...TAGT...TTT at positions 812 and 844 [10].

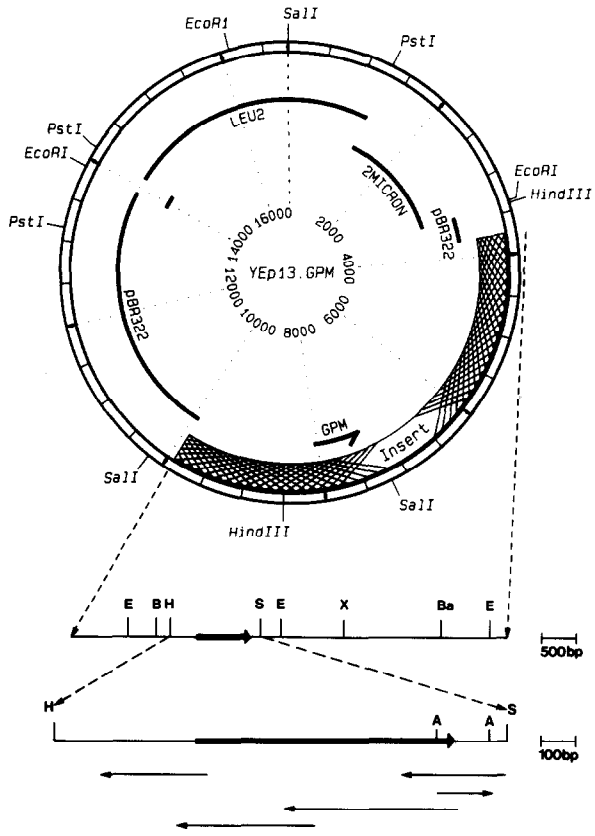

Fig.1. Restriction map of YEp13-GPM and sequencing strategy of GPM gene. The thick lines indicate the phosphoglycerate mutase coding sequence. Arrows show the direction and length of individual sequence determinations. Restriction enzymes: *Hin*dIII (H), *Sal*I (S), *Eco*RI (E), *Bgl*II (B), *Bam*HI (Ba), *Xho*I (X) and *Alu*I (A).

3.3. *Comparison of DNA and protein sequences*

A comparison of the amino acid sequence deduced from the DNA sequence with that determined by manual Edman degradation of proteolytic fragments of the protein [5] reveals several discrepancies (fig.2). Some of the differences probably relate to *S. cerevisiae* strain differences. The phosphoglycerate mutase used for the protein sequencing and X-ray crystallography was purified from dried baker's yeast, and was genetically heterogeneous as indicated by sequence microheterogeneities between residues 139 and 166 [5]. The sequence differences between residues 46 and 52 may indicate an additional heterogeneous region. The occurrence of sequence heterogeneities complicated the protein sequence determinations, and was the reason for wrongly assigning the position of the Asp-Leu sequence (residues 169–170). The other sequence discrepancies can be explained in terms of difficulties inherent in the manual Edman degradation procedure. For example, the amide assignments were largely based on the electrophoretic mobilities of peptides, and residue 99 was incorrectly assigned as glutamine because of the abnormally low p*K* of His-88. The interchange of the serine and valine residues at positions 172 and 176 resulted from having missed the labile serine residue after Edman degradation. Similarly, the sequence between residues 207 and 210 involves an incompletely cleaved proline and a labile threonine. The determination of the sequence of the C-terminal portion of the protein relied heavily

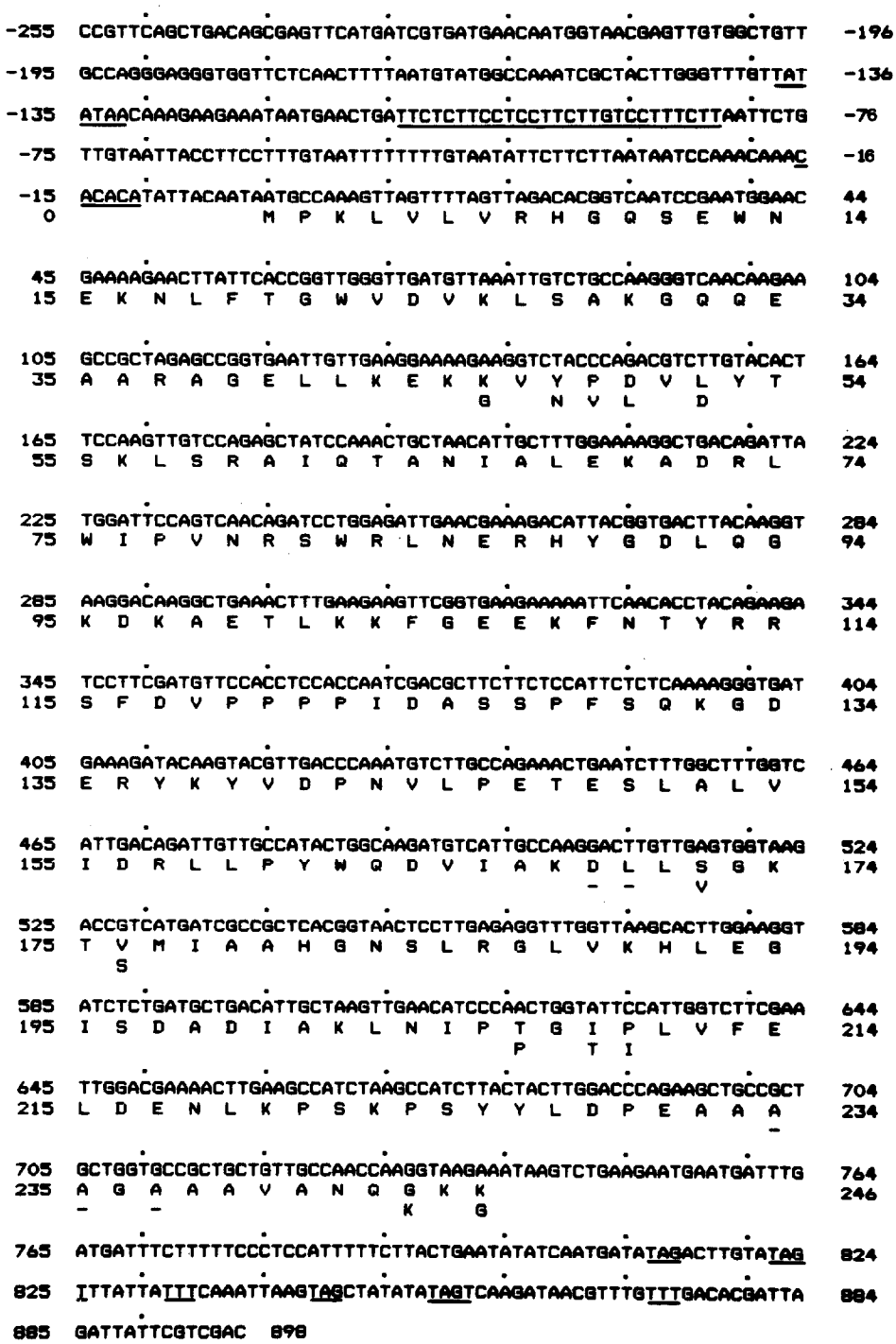

Fig.2. Nucleotide sequence and deduced amino acid sequence of GPM gene. The underlined sequences are discussed in the text. Residues different in the published protein sequence [5] are indicated below the deduced amino acid sequence, and are discussed in the text.

on overlapping peptic peptides, and the unusually large number of alanines between residues 232 and 239 (7 of 8 residues) was incorrectly identified. The presence of a glutamine at position 243 reduced the efficiency of Edman degradation, and the order of the C-terminal three residues was largely assigned on the basis of the complete resistance of this region to digestion by carboxypeptidases A, B and Y.

3.4. *Comparison with mammalian phosphoglycerate mutases*

Phosphoglycerate mutase exists as three different isoenzymes in mammalian tissues: type M

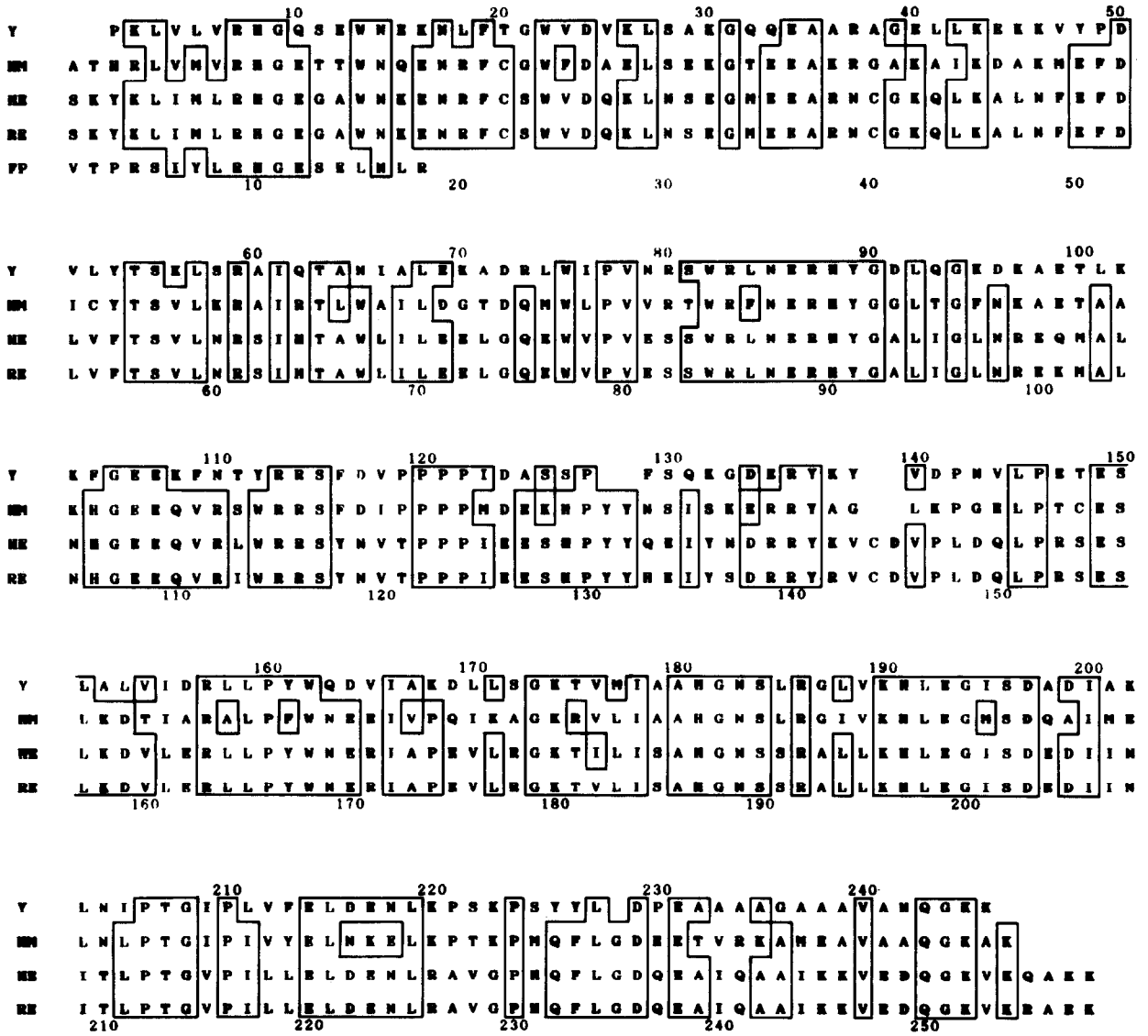

Fig.3. Comparison of amino acid sequences of yeast and mammalian phosphoglycerate mutases. The sequences are taken from the following references: human muscle phosphoglycerate mutase (HM) [14], human erythrocyte phosphoglycerate mutase (HE) [15], rabbit erythrocyte phosphoglycerate mutase (RE) [16] and fructose 2,6-bisphosphatase (FP) [13]. The numbering along the top of the sequences refers to the yeast sequence, and that below to the erythrocyte sequences. The boxed-in residues show the regions that are identical in three or more sequences. Gaps have been introduced in the Y and HM sequences to maximise homology.

predominates in muscle, type B in brain and type E in erythrocytes. The E type enzyme catalyses the synthesis and breakdown of 2,3-bisphosphoglycerate (2,3-BPG), in addition to catalysing the interconversion of 2- and 3-phosphoglycerates. This isoenzyme is frequently known as bisphosphoglycerate mutase, and plays a major role in regulating haemoglobin oxygen affinity as a consequence of controlling 2,3-BPG concentration. A comparison of the sequence of yeast phosphoglycerate mutase with the sequences of M and E type mammalian phosphoglycerate mutases is given in fig.3. These enzymes show the same slow rate of evolution as observed for other glycolytic enzymes [11], with about half of the residues identical between the yeast and mammalian enzymes. The two mammalian isoenzymes are as divergent from each other (49% different) as they are from the yeast enzyme, implying that the E isoenzyme experienced a period of rapid change after the gene duplication of an ancestral mutase gene before it settled down to the slow rate of change shown by a comparison of the rabbit and human E isoenzymes (3.5% different).

Most, but not all, of the active-site residues [4] are conserved. Thus, His-8 (yeast numbering) and His-181 which are involved in phospho transfer and Arg-59 which is required for binding the carboxyl group of the phosphoglycerates are identical in all sequences. However, Ser-11 and Thr-20 which in the yeast enzyme probably provide ligands for binding the transferred phospho group are not conserved. The replacement of Ser-11 by a glycine in the E isoenzyme is consistent with the greater instability of its phosphorylated form [12], and also with the relative ease with which the product 2,3-BPG can dissociate [1]. The replacement of Thr-20 by a cysteine explains why muscle but not yeast phosphoglycerate mutase is inhibited by thiol modifying reagents (review [11]).

The sequence of an active-site peptide from rat liver fructose 2,6-bisphosphatase has recently been reported, and its homology with the active site of phosphoglycerate mutases noted [13] (see fig.3). Like the mutases, fructose 2,6-bisphosphatase involves a phosphohistidine intermediate (His-8 of the yeast mutase), and the sequence similarities indicate that the enzymes have probably diverged from a common ancestor.

Acknowledgements: We are particularly grateful to Dr D.G. Fraenkel for the gift of the YEp13-GPM plasmid. We thank the Science and Engineering Research Council for financial support, and the Wellcome Trust for funding the OSWEL oligonucleotide synthesis facility. L.A.F.-G. is the recipient of a Wellcome Trust University Award.

REFERENCES

[1] Rose, Z.B. (1980) Adv. Enzymol. Relat. Areas Mol. Biol. 51, 211–253.
[2] Oh, Y.K. and Freeze, B. (1976) J. Bacteriol. 127, 739–746.
[3] Price, N.C., Duncan, D. and Ogg, D.J. (1985) Int. J. Biochem. 17, 843–846.
[4] Winn, S.I., Watson, H.C., Harkins, R.N. and Fothergill, L.A. (1981) Phil. Trans. R. Soc. Lond. B 293, 121–130.
[5] Fothergill, L.A. and Harkins, R.N. (1982) Proc. R. Soc. Lond. B 215, 19–44.
[6] Kawasaki, G. and Fraenkel, D.G. (1982) Biochem. Biophys. Res. Commun. 108, 1107–1112.
[7] Benetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026–3031.
[8] Hamilton, R., Watanabe, C. and De Boer, H.A. (1987) Nucleic Acids Res. 15, 3581–3593.
[9] Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J., Kingsman, S.M., Perkins, R.E., Conroy, S.C., Dunbar, B. and Fothergill, L.A. (1982) Nucleic Acids Res. 10, 2625–2637.
[10] Zaret, K.S. and Sherman, F. (1982) Cell 28, 563–573.
[11] Fothergill-Gilmore, L.A. (1986) Trends Biochem. Sci. 11, 47–51.
[12] Han, C.-H. and Rose, Z.B. (1979) J. Biol. Chem. 254, 8836–8840.
[13] Pilkis, S.J., Lively, M.O. and El-Maghrabi, M.R. (1987) J. Biol. Chem. 262, 12672–12675.
[14] Shanske, S., Sakoda, S., Hermodson, M.A., DiMauro, S. and Schon, E.A. (1987) J. Biol. Chem. 262, 14612–14617.
[15] Joulin, V., Peduxxi, J., Romeo, P.-H., Rosa, R., Valentin, C., Dubart, A., Lapeyre, B., Blouquit, Y., Garel, M.-C., Gooseens, M., Rosa, J. and Cohen-Solal, M. (1986) EMBO J. 5, 2275–2283.
[16] Yanagawa, S., Hitomi, K., Sasaki, R. and Chiba, H. (1986) Gene 4, 185–191.