Volume 247, number 2, 312-316 FEB 07049 April 1989

The yeast pyruvate kinase gene does not contain a string of non-preferred codons: revised nucleotide sequence

Teresa McNally, Ian J. Purvis*+, Linda A. Fothergill-Gilmore and Alistair J.P. Brown*

Department of Biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD and **Institute of Genetics, University of Glasgow, Church Street, Glasgow G11 5JS, Scotland*

Received 15 February 1989

The sequence of the gene encoding pyruvate kinase from *Saccharomyces cerevisiae* was re-determined because of failures with oligonucleotide-directed mutagenesis experiments involving a region thought to contain a string of five contiguous non-preferred codons. This region was found to be difficult to sequence and was shown to have three extra bases when compared with the published sequence [(1983) J. Biol. Chem. 258, 2193–2201]. The revised sequence demonstrates that the yeast pyruvate kinase gene does not have a cluster of non-preferred codons, and that it therefore is not an example of the class of genes which possibly exhibit translational control by the presence of non-preferred codons.

Pyruvate kinase gene; Nucleotide sequence; Translational control; Non-preferred codon; (*Saccharomyces cerevisiae*)

1. INTRODUCTION

Pyruvate kinase has been extensively studied because of its abundance in most organisms and because of its importance in glycolysis for controlling the flux from fructose 1,6-bisphosphate through to pyruvate. The enzyme catalyses the essentially irreversible conversion of phospho*enol*-pyruvate to pyruvate by the addition of a proton and the loss of a phospho group, which is transferred to ADP. The high resolution crystal structure of mammalian muscle pyruvate kinase has been determined and correlated with the protein sequence [1]. Amino acid sequences deduced from DNA sequences are also available for all four mammalian isoenzymes [2–6], and for pyruvate kinase from chicken [7] and from *Saccharomyces cerevisiae* [8].

A notable feature of the yeast pyruvate kinase DNA sequence previously reported [8] is that it has an unusual arrangement of five consecutive non-preferred codons (as defined by Bennetzen and Hall [9]) located approximately four-fifths of the way through the coding region. This provided indirect evidence for the possibility that the gene may possess a translational pause in this region, thus enabling the multidomain structure of the enzyme to fold correctly during translation [10]. Experiments were devised by the Glasgow group to test this hypothesis by site-directed mutagenesis to replace the non-preferred codons with the synonymous preferred codons. Three oligonucleotides were synthesised for this purpose, but they failed to give the appropriate mutants, and their T_m values seemed to be lower than theoretically expected [11]. By coincidence, site-directed mutagenesis experiments involving the same region were being done by the Edinburgh group to study intersubunit contacts and allosteric control of pyruvate kinase. It was decided to re-examine the DNA sequence in

Correspondence address: A.J.P. Brown, Institute of Genetics, University of Glasgow, Church Street, Glasgow G11 5JS, Scotland

+*Present address:* Genetics Unit, Glaxo Group Research Ltd, Greenford Road, Greenford, Middlesex UB6 0HE, England

The nucleotide sequence presented here will appear in the EMBL/GenBank database under the accession no. X14400

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

this region, and eventually to re-determine the entire sequence.

2. EXPERIMENTAL

The plasmid pAYE4(34) (fig.1), which corresponds to the *S. cerevisiae PYK1* gene cloned into the multicopy vector pJDB207 [12], was generously provided by R. Cafferkey, A. Goddey and B. Carter formerly of G.D. Searle. Suitable restriction fragments were subcloned into phage M13 for sequence determination by the dideoxy method. The oligonucleotides used by the Glasgow group were synthesised by V. Math, A.J.P. Brown and J.R. Coggins (Departments of Biochemistry and Genetics, University of Glasgow); those used by the Edinburgh group were from Oswel DNA Service, University of Edinburgh.

Site-directed mutagenesis was done by the Glasgow group in an attempt to replace the contiguous string of five non-preferred codons in *PYK1* in the proposed 'pause region' [10] with the synonymous preferred codons. The procedures of Winter et al. [13] were used. Following mutagenesis, phages were screened by plaque hydridisation with the ^{32}P-end labelled mutagenic oligonucleotide as a probe. Filters were serially washed at increasing temperatures, and autoradiographed between each wash [14]. The expected T_m value for the wild-type *PYK1* sequence was 37°C, whereas that for the mutant was 74°C. In fact, the plaques displayed a range of T_m values intermediate between these two extremes. Several putative mutant phages were sequenced with the use of a 14-mer primer of the sequence 5′-TGTACTCCAAAGCC-3′ which hybridises 16 nucleotides upstream from the 'pause region'. To our surprise the sequences neither contained the published wild-type nor the expected modified sequences. Instead, all contained three extra C-residues interspersed with the wild-type 'pause' sequence (fig.2). Sequencing of the original plasmid pAYE4(34) in this region established that the sequence differences did not arise during the multiple subcloning steps during the mutagenesis experiments. The Edinburgh group independently confirmed the presence of the extra C-residues with a different sequencing primer.

3. RESULTS AND DISCUSSION

3.1. *Nucleotide sequence of* PYK1 *gene*

The strategy used to sequence the *PYK1* gene is summarised in fig.1. A contiguous stretch of 1791 bases was sequenced which includes an open reading frame of 1500 bases encoding the enzyme pyruvate kinase (fig.2). The sequence is identical to the previously published sequence [8] except for three extra C-residues located between codons 379 and 386. These sequence changes have the consequence that the five non-preferred codons are converted to six preferred codons.

Both the Glasgow and Edinburgh groups found that the results of DNA sequencing reactions in this region were unusually difficult to interpret (fig.3). This was true with the use of different primers and

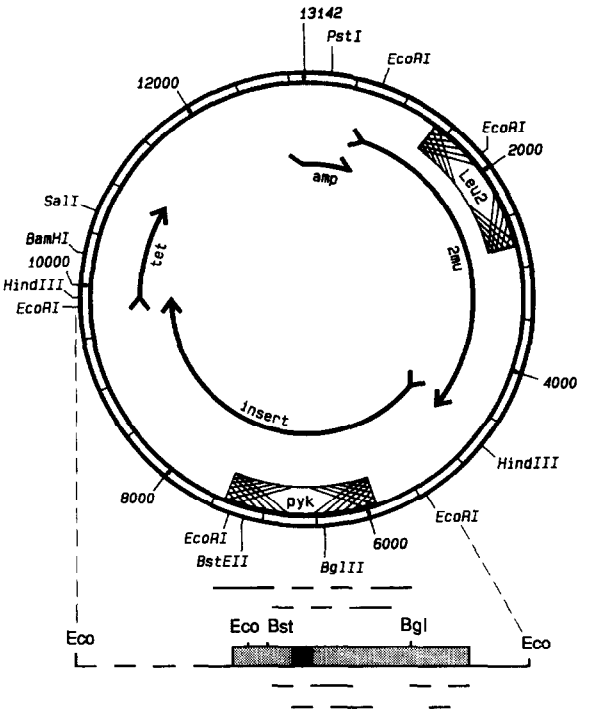

Fig.1. The yeast/*E. coli* shuttle plasmid pAYE4(34) with a 7 kbp insert encoding the yeast *PYK1* gene. The two *Eco*RI restriction fragments covering the entire *PYK1* coding sequence were subcloned individually into pAT153 and subsequently into M13mp19 in two orientations to allow sequencing of both strands. In addition, the *Bgl*II/*Eco*RI fragment containing the sequence differences was cloned into M13mp18 and M13mp19 to allow for detailed analysis of the region. The contig diagram illustrates the sequencing strategy. Universal sequencing primers were used, as well as oligonucleotides synthesised for use in the mutagenesis studies. The shaded region shows where the three extra 'C'-residues are located. This region of the DNA has been sequenced extensively on both strands using either the Klenow fragment of DNA polymerase or Sequenase.

with different DNA polymerases. It seems likely that these residues were missed with the sequencing conditions used by Burke et al. [8]. An additional and different line of evidence in support of this suggestion comes from the fact that the revised sequence is more similar to pyruvate kinase sequences from other organisms (fig.4).

3.2. *Translational pauses specified by non-preferred codons*

It is clear from these results that pyruvate kinase no longer provides strong indirect evidence to support a possible link between the rate of mRNA

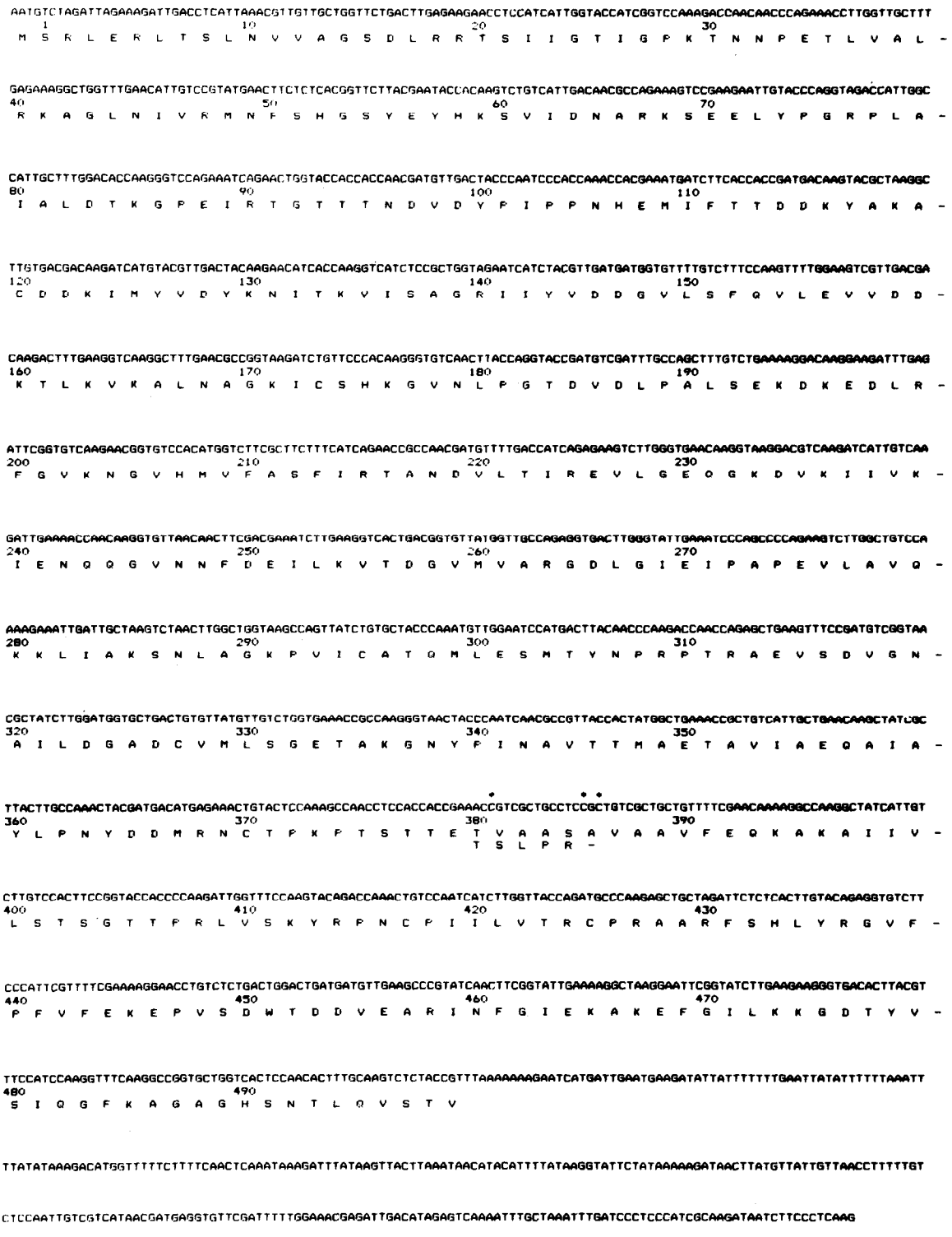

Fig.2. Nucleotide sequence and deduced amino acid sequence of *PYK1*. The three extra 'C'-residues found in this study are indicated by the asterisks, and are discussed in the text.

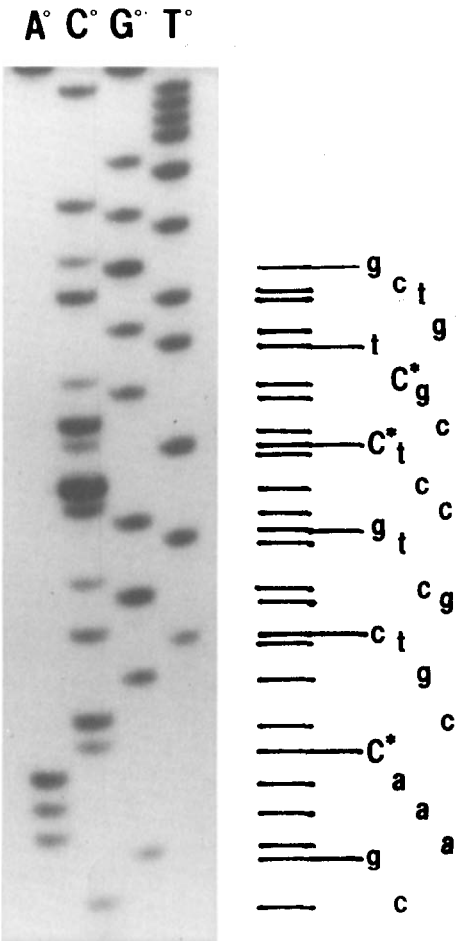

Fig.3. The results of dideoxy sequencing reactions with the Klenow fragment of DNA polymerase in the region of the sequence differences. The three 'C'-residues not reported in the orginal sequence data are indicated by the asterisks, and appear fainter than other C-residues in the same region. The use of dITP instead of dGTP had no effect on this observation. Sequenase was only effective for sequencing this region when 7-deaza-dGTP was substituted for dGTP in the reaction mixes.

translation and protein folding in vivo. However, this does not mean that the hypothesis is no longer tenable. The codon usage patterns in several other yeast genes reveal significant contiguous strings of non-preferred codons; these include *ARO1*, *GAL4* and *TRP3* [10] as well as *FAS1*, *HIS4* and *TRP5* (fig.5). In some genes the strings of non-preferred codons seem to be preferentially located close to interdomain regions (*ARO1, FAS1*), whereas in others no such correlations exist (*GAL4, HIS4, TRP5*). In the latter cases it is possible that the strings of non-preferred codons promote intradomain folding, but there is no evidence to support this. It remains attractive to suggest that putative translational pauses within the *ARO1* and *FAS1* genes temporally separate the synthesis, and hence the folding, of their multiple structural domains in vivo. As far as we know there is no direct evidence yet that either invalidates or supports the hypothesis. The Glasgow group is attempting to provide such evidence by mutating the string of ten non-preferred codons in *TRP3* [10].

PYK

```
                    370                 380                 390
                                      * * * * *
Yeast [8]           ... C T P K P T S T T E T S L P R - V A A V F E Q K A K A ...
Yeast - this report ... C T P K P T S T T E T V A A S A V A A V F E Q K A K A ...
Chicken M [7]       ... H S V H H R E P A D A M A A G A V E A S F K C L A A A ...
Cat M1 [1]          ... G S S H S T D L M E A M A M G S V E A S Y K C L A A A ...
Rat M1 [2]          ... A S S Q S T D P L E A M A M G S V E A S Y K C L A A A ...
Rat M2 [2]          ... L A P I T S D P T E A A A V G A V E A S F K C C S G A ...
Rat L & R [3-5]     ... A A P L S R D P T E V T A I G A V E A S F K C C A A A ...
Human L [6]         ... A A P L S R D P T E V T A I G A V E A A F K C C A A A ...
```

Fig.4. Pyruvate kinase amino acid sequences in the region of the yeast sequence differences (indicated by the asterisks). The yeast sequences are compared with those from chicken muscle [7], cat skeletal muscle isoenzyme M1 [1], rat skeletal muscle isoenzyme M1 [2], rat kidney isoenzyme M2 [2], rat liver and erythrocyte isoenzymes L and R [3–5], and human liver isoenzyme L [6]. Residues that are identical in four or more sequences are boxed.

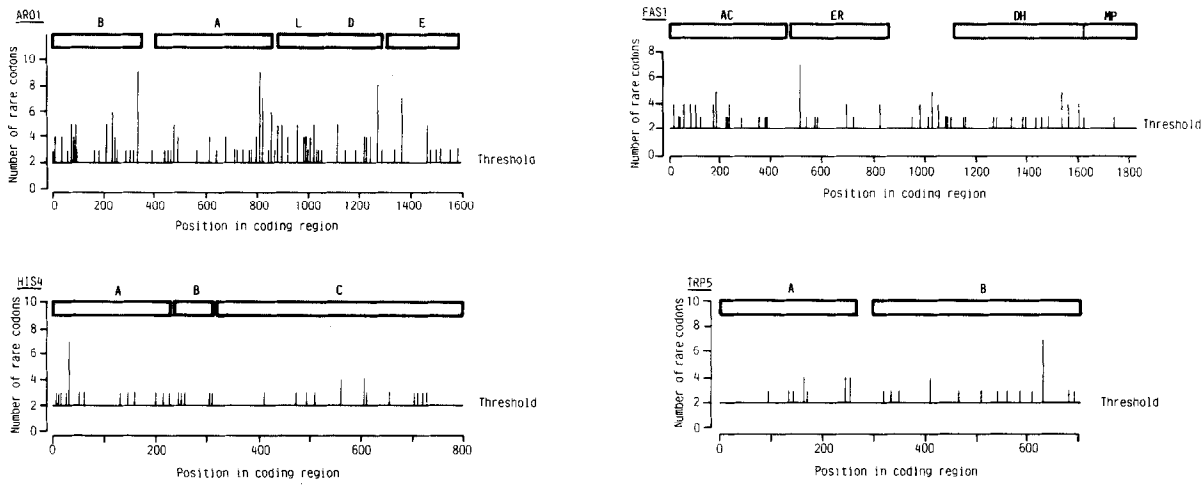

Fig.5. The distribution of non-preferred codons in the *ARO1, FAS1, HIS4* and *TRP5* genes from *S. cerevisiae*. The unique codons for methionine and tryptophan are considered to be neutral in this analysis; they do not stop a string of non-preferred codons, nor contribute to it. The number of consecutive non-preferred codons in a string is plotted against the location of the string within the coding region. The domain organisation of each coding region is presented above each graph. *ARO1* (diagram adapted from [10]; sequence published in [15]). *FAS1:* AC, acetyl transferase; ER, enoyl reductase; DH, dehydratase; MP, malonyl palmityl transferase [16]. *HIS4:* A, phosphoribosyl-AMP cycolhydrolase; B, phosphoribosyl-ATP pyrophosphohydrolase; C, histidinol dehydrogenase [17]. *TRP5*: A and B, alpha and beta domains of tryptophan synthase [18].

Acknowledgements: We are grateful to Dr V. Math and Professor J.R. Coggins for their help in synthesising the oligonucleotides used by the Glasgow laboratory. We thank the Science and Engineering Research Council for financial support, and the Wellcome Trust for funding the Oswel DNA Service. L.A.F.-G. is the recipient of a Wellcome Trust University Award.

REFERENCES

[1] Muirhead, H., Clayden, D.A., Barford, D., Lorimer, C.G., Fothergill-Gilmore, L.A., Schiltz, E. and Schmitt, W. (1986) EMBO J. 5, 475–481.
[2] Noguchi, T., Inoue, H. and Tanaka, T. (1986) J. Biol. Chem. 261, 13807–13812.
[3] Inoue, H., Noguchi, T. and Tanaka, T. (1986) Eur. J. Biochem. 154, 465–469.
[4] Lone, Y.-C., Simon, M.-P., Kahn, A. and Marie, J. (1986) FEBS Lett. 195, 97–100.
[5] Noguchi, T., Yamada, K., Inoue, H., Matsuda, T. and Tanaka, T. (1987) J. Biol. Chem. 262, 14366–14371.
[6] Tani, K., Fujii, H., Nagata, S. and Miwa, S. (1988) Proc. Natl. Acad. Sci. USA 85, 1792–1795.
[7] Lonberg, N. and Gilbert, W. (1985) Cell 40, 81–90.
[8] Burke, R.L., Tekamp-Olson, P. and Najarian, R. (1983) J. Biol. Chem. 258, 2193–2201.
[9] Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 11927–11937.
[10] Purvis, I.J., Bettany, A.J.E., Santiago, T.C., Coggins, J.R., Duncan, K., Eason, R. and Brown, A.J.P. (1987) J. Mol. Biol. 193, 413–417.
[11] Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature 246, 40–41.
[12] Parent, S.A., Fenimore, C.M. and Bostian, K.A. (1985) Yeast 1, 83–138.
[13] Winter, G., Fersht, A.R., Wilkinson, A.J., Zoller, M. and Smith, M. (1982) Nature 299, 756–758.
[14] Wallace, R.B., Johnson, P.F., Tanaka, S., Schold, M., Itakura, K. and Abelson, J. (1980) Science 209, 1396–1400.
[15] Duncan, K., Edwards, R.M. and Coggins, J.R. (1987) Biochem. J. 246, 375–386.
[16] Schweizer, M., Roberts, L.M., Holte, H.-J., Takabayashi, K., Hollerer, E., Hoffman, B., Muller, G., Kottig, H. and Schweizer, E. (1986) Mol. Gen. Genet. 203, 479–486.
[17] Donahue, T.F., Farabaugh, P.J. and Fink, G.R. (1982) Gene 18, 47–59.
[18] Zalkin, H. and Yanofsky, C. (1982) J. Biol. Chem. 257, 1491–1500.