The primary structure of the 3-phosphoglycerate kinase (PGK) gene from *Kluyveromyces lactis*

Alain Fournier, Reinhard Fleer, Patrice Yeh and Jean-François Mayaux

Rhône Poulenc Santé, Biotechnology Department, 13 Quai Jules Guesde, 94403 Vitry Cedex, France

Submitted December 5, 1989

EMBL accession no. X17654

The 3-phosphoglycerate kinase (*PGK*) gene from the industrially important dairy yeast *Kluyveromyces lactis* was isolated from a partial genomic library derived from wild-type strain CBS2359 using the *Saccharomyces cerevisiae* *PGK* gene (1) as a hybridization probe. The entire gene is contained on a *Hind*III fragment which is 4 kb in size and able to support growth of a *pgk* null-mutant of *S. cerevisiae* on glucose when harbored on a 2 μ based multi-copy vector. We report here the complete nucleotide sequence of the *K. lactis* *PGK* structural gene which is identical in size (416 aa) and highly homologous (82% on the nucleotide level and 83% on the deduced amino acid level) to its *S. cerevisiae* counterpart (2). Compared to the *PGK* genes from humans (3), horse (4), *Aspergillus nidulans* (5), and *Penicillium chrysogenum* (6) homologies on the amino acid levels are 65%, 65%, 66%, and 67%, respectively. As has been shown for *S. cerevisiae*, the *K. lactis* *PGK* gene exhibits a strong codon bias (CAI = 0.81) (7). The codons preferred by *K. lactis* are quasi identical to those found in the corresponding *S. cerevisiae*

gene. Even though the correlation between codon bias and level of protein production (8) is likely to resemble that of *S. cerevisiae* its closer examination in *K. lactis* will require the compilation of further sequence information which to date remains rather limited.

REFERENCES

1. Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J. and Kingsman, S.M. (1982) *Nucl. Acids Res.* **10**, 2625–2637.
2. Hitzeman, R.A., Hagie, F.E., Hayflick, J.S., Chen, C.Y., Seeburg, P.H. and Derynck, R. (1982) *Nucl. Acids Res.* **10**, 7791–7808.
3. Banks, R.D., Blake, C.C.F., Evans, P.R., Haser, R., Rice, D.W., Hardy, G.W., Merret, M. and Phillips, A.W. (1979) *Nature* **279**, 773–777.
4. Huang, I.-Y., Welch, C.D. and Yoshida, A. (1980) *J. Biol. Chem.* **255**, 6412–6420.
5. Clements, J.M. and Roberts, C.F. (1986) *Gene* **44**, 97–105.
6. Van Solingen, P., Muurling, H., Koekman, B. and Van den Berg, J. (1988) *Nucl. Acids Res.* **16**, 11823.
7. Sharp, P.M. and Li, W.H. (1987) *Nucl. Acids Res.* **15**, 1281–1295.
8. Ikemura, T. (1985) *Mol. Biol. Evol.* **2**, 13–34.

```

ATGCTTTTGAAGTCAAAATTGACTGTCAAGGATTGGATGTCAGTGGTAAAGAGTGTTCATCAGAGTTGATTCAACGTCCCATGGAC   90
M S L S S K L T V K D L D V T G K R V F I R V D F N V P L D   30
GGTAAGAAGATCACTTCCAACCAAGAATTGTAGCAGCTCTGCTACTATTCAATACGTTTGGAAAAGAAGCCAAAGGCCATTGTCTTG   180
G K K I T S N Q R I V A A L P T I Q Y V L E K K P K A I V L   60
GCCTCTCATTGGGTAGACCAACGGTGAAGTTAACGACAAATACTCATTGGCTCCAGTTGCCGATGAATTGTCTAGACTATTGCAAAAG   270
A S H L G R P N G E V N D K Y S L A P V A D E L S R L L Q K   90
CCAGTTACTTCTTTCATGACTGTGTTGGTGAAGAAGTCACCAATGCCGTTAACACGCTAAGGACGGTGAAGTTTCTATTGGAAAAC   360
P V T F L H D C V G E E V T N A V N N A K D G E V F L L E N   120
TTGAGATTCCACATCGAAGAAGAAGGTTCCAGAAAGGTGGATGGTAACAGGTTAAGGCTGACAAGGCTGCTGTTACCAAAATTCAGAGAA   450
L R F H I E E E G S R K V D G N K V K A D K A A V T K F R E   150
CAATTGAGTTCTTTGGCCGATGTTACGTTAACGATGCCCTCGGTACCGCTCAGAGCTCAGTCTTCCATTGTCGGTTTCGACTTGCCA   540
Q L S S L A D V Y V N D A F G T A H R A H S S I V G F D L P   180
AACAGAGCTGCCGTTTCTTGTGTCAAGGAATTGCAATACCTTGTCTAAGGCTTGGAAAACCAACAGACCATCTTGGCTATCTTG   630
N R A A G F L L S K E L Q Y F A K A L E N P T R P F L A I L   210
GGTGGTCTAAGGTTGACAGACAAGATTCAATTGATCGACAATTGTTGGACAAGGTCGATTCTTGTATCATGGTGGTGGTATGGCTTTC   720
G G A K V A D K I Q L I D N L L D K V D S L I I G G G M A F   240
ACCTTCAAGAAGTCTTGGAAAACACTGAAATCGGTGACTCCATCTACGATGCTGCTGGTGGCGAATTGGTTCCAAAGTTGGTCGAAAAG   810
T F K K C V L E N T E I G D S I Y D A A G A E L V P K L V E K   270
GCTAAGAAGAACAACGTCAGATCGTCTTGCCAACCTGACTTCGTCATTGGTGACAAGTTCTCTGCTGATGCTAACACAAAGGTTGTCCAC   900
A K K N N V K I V L P T D F V I G D K F S A D A N T K V V T   300
GACAAGGAAGGTTATCCAAAGCGGCTGGCAAGGTTGGACAACGGTCCAGATCTAGAAAAGGCTTTCGCTGCCACTGTTGCTGAAGCTAAG   990
D K E G I P S G W Q G L D N G P E S R K A F A A T V A E A K   330
ACTATTGCTGGAACGGTCCACAGGTGCTTCGAATTGCTCCTTTCGCTAAGGTCACCAAGCTCTATTGGACGCCGTTGTCGCCTCC   1080
T I V W N G P P G V F E F A P F A K G T E A L L D A V V A S   360
TCTCAAGCTGCTAACAAGCTCATCGGTGGTGGTGACACCGTACCGTCCGCAAGAAGTACGGTGTGCTCGACAAGATCTCTCACGTC   1170
S Q A G N N V I I G G G D T A T V A K K Y G V V D K I S H V   390
TCCACTGGTGGTGGTCTCTTGGAAATGTTGGAAGGTAAGAATGGCTGGTGTACATTTCTTGTCCAACAAGCAATAA   1251
S T G G G A S L E L L E G K E L P G V T F L S N K Q *   416

```