Human, yeast and hybrid 3-phosphoglycerate kinase gene expression in yeast

Christina Y.Chen and Ronald A.Hitzeman

Department of Cell Genetics, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, USA

Received August 12, 1986; Revised and Accepted December 17, 1986

ABSTRACT

When the gene for yeast 3-phosphoglycerate kinase (PGK) is present on a high copy number plasmid in *Saccharomyces cerevisiae*, 30–40 percent of yeast protein is produced as PGK. However, when the structural part of this gene is replaced by as many as twenty different heterologous genes, production of gene products is greatly reduced—usually by more than 20 fold. This decrease in protein production is accompanied by large decreases in the steady-state levels of mRNA. However, in contrast to these coding sequences, replacement of the yeast PGK structural gene with a human PGK cDNA has little effect on the steady-state mRNA level in yeast. PGK is a two-domain enzyme and its 3-dimensional structure is highly conserved among species. These observations and others have led us to propose that the PGK protein itself might influence its own mRNA levels (Chen *et al.*, *Nucleic Acids Res.* 12, pp. 8951–8969, 1984). In addition, data is presented here which suggests that the human PGK mRNA is less efficiently translated than the yeast PGK mRNA. Two different mechanisms of controlling gene expression are indicated. Both mechanisms appear to be independent of gene copy number.

INTRODUCTION

The 5'- and 3'-flanking DNA sequences of the highly expressed yeast 3-phosphoglycerate kinase (PGK) gene can be used to express heterologous gene products in yeast (1–3). However, these modified expression systems on high copy number plasmids do not express levels of heterologous gene products at the high level obtained for yeast PGK by the original plasmid. We have extensively studied this phenomenon and have shown that decreased protein product levels appear to reflect changes in the steady-state levels of mRNA and are not due to reductions in plasmid copy numbers (1). Recent results by Mellor *et al.* (4) support these conclusions.

We have also shown that disruption of the PGK coding sequence by premature stop codons, by substitution of a heterologous gene, by in-frame deletions of PGK sequence, or by insertions of heterologous or homologous DNA sequences (in-frame or out-of-frame) leads to decreased steady-state levels of mRNA (1). However, the same inserts (some over 1000 bp) can be

inserted after the stop codon of the structural gene and be included within the mRNA without any effect on the steady-state levels of mRNA (1). The combined data suggested that essentially the entire structural portion of the gene encoding PGK is necessary to preserve high steady-state levels of mRNA. Evidence was also presented that the normal PGK expression system was unable to correct several of these defective systems when present on the same plasmid, suggesting that the wild-type PGK protein does not act in trans to correct these defective systems. Codon bias differences of heterologous genes did not appear to be responsible for this lowering of steady-state mRNA levels, since this lowering occurred with or without the translation of such sequences (1).

Three models have been proposed to explain these decreases in mRNA (1). The first model suggests that there is a large internal promoter sequence (or sequences) within the PGK coding sequence which affects transcription levels in conjunction with the 5'-flanking sequence. This may be similar to eukaryotic enhancer sequences (5) or to internal promoters found in tRNA genes (6). The second model suggests that a large primary sequence in the mRNA results in stabilization of the mRNA due to secondary or tertiary interactions. The third model (protein-feedback) suggests that the information is in the PGK protein itself. Within the protein structure there may be a domain or domains which interact directly with the mRNA or translation machinery to prevent mRNA degradation during translation. This paper describes experiments which were done based on the possibility of this third model.

Some data that are supportive of a protein-feedback model were obtained several years ago by Losson and Lacroute (7). They found that nonsense mutations throughout the URA3 gene in the chromosome resulted in reduction of steady-state levels of mRNA. Zitomer *et al.* (8) observed a similar behavior for the yeast cytochrome C gene. However, Losson and Lacroute (7) took their analysis two steps further. They showed that one stop codon near the beginning of the gene reduced the half-life of the URA3 mRNA from 10 minutes to 2 minutes, which explained the reduction in the steady-state level of mRNA. Second, they showed that the suppression of this stop codon during translation reversed the effect on the steady-state mRNA level, suggesting that the effect occurs during translation. They suggested that the large untranslated region in such mutants leads to mRNA degradation in the cytoplasm, perhaps due to the lack of ribosome protection of this part of the mRNA. We have observed the same phenomenon with nonsense mutants at

two different locations in the PGK gene, carried on a high copy number plasmid (1,9). However, the half-life of the PGK mRNA is at a higher level of about 80 minutes (9). Furthermore, removal of normally translated DNA sequence after such a premature stop codon to the normal stop codon does not significantly improve the steady-state level of mRNA (9). We have also shown that large untranslated regions inserted at the end of the complete structural gene for PGK do not reduce mRNA half-lives or expression (1). Perhaps these results suggest a relationship between steady-state levels of mRNA and protein integrity.

To further test the protein-feedback model, we decided to express the human cDNA for 3-phosphoglycerate kinase (hPGK) (10,11) using the yeast PGK gene 5'- and 3'-flanking sequences. The hPGK gene encodes an enzyme that has 65 percent homology at the amino acid level with yeast PGK (yPGK) (12) and much more similarity at the three-dimensional level (15).

MATERIALS AND METHODS

Materials

Materials have all been previously described (1).

Strains, Plasmids and Growth Conditions

Yeast strain 20B-12 (α trp1 pep4-3) (20) was used for most experiments. Plasmid YEp9T (Fig. 1) has been previously described (1). YNB+CAA (1) broth and plates were used for transformation and for all growth experiments to maintain the plasmids using Trp^+ selective pressure. All other strains and growth conditions have been described elsewhere (1).

Methods

All methods have been previously described (1,2); however, a pulse-chase ^{35}S -methionine labeling technique was used here to obtain Fig. 5 instead of a normal labeling as previously used (2).

RESULTS

Construction of Expression Units Producing yPGK, hPGK, and Hybrid Proteins

For these studies, we used the yeast high copy number plasmid, YEp9T (Fig. 1), which replicates in both *E. coli* and yeast (1). Inserts b-g were placed between the EcoRI and HindIII restriction sites of YEp9T in the same orientation. By filling in the EcoRI restriction end of YEp9T and ClaI restriction ends of b-g with nucleotides using *E. coli* DNA polymerase I, two fragment ligations were done to construct these plasmids. All the plasmids thus formed have a copy number of 20-30 based on "Southern" analysis (16)

Figure 1. PGK gene expression unit constructions and the plasmid used. Plasmid YE9T has been previously described (1). Expression units are shown as straight line drawings with 5'-flanking DNAs or promoter regions to the left, followed by structural genes as solid bars (yPGK) or open bars (hPGK), and 3' flanking sequences as straight lines on the right. IFN-α1 cDNA inserts at the *Cla*I restriction site have been previously described (1). They are transcribed but are not translated. Transcription starts at -36 nucleotides (before ATG) and ends (polyadenylation) about 90 nucleotides past the translation stop (12). The hybrid genes were made as described in text and Fig. 2.

using chromosomal unit a as the standard of one (data not shown).

Expression unit a is the normal yPGK expression unit present as one copy on chromosome III (17,18). Units b and c are this same unit, but on the high copy plasmid. We have previously shown that placement of the interferon IFN-α1 cDNA within the *Cla*I restriction site, shortly after the translation stop of yPGK, has no effect on expression of yPGK (1). It does, of course, increase the size of the mRNA correspondingly.

Expression unit d was made using M13 oligonucleotide-directed mutagenesis (19) of the human PGK cDNA (11). An *Eco*RI restriction recognition site was placed immediately before the ATG of the hPGK cDNA.

The 5'-flanking sequence of the yPGK gene had been previously modified with XbaI/EcoRI sites immediately before the ATG (TCTAGAATTCATG) (2). This EcoRI restriction site was used to join with the modified hPGK cDNA. An XbaI site was made at the end of the hPGK cDNA changing ATTTAGT to ATCTAGA. The change of ATT to ATC maintains the isoleucine as the last amino acid of hPGK (10). The end of the yPGK structural gene (12) was also changed from AATAAA to TCTAGA (an XbaI site) so that the 3'-flanking sequence of yPGK can be ligated to the end of the modified hPGK cDNA to make unit d. This change at the stop codon is not present in yeast PGK expression units a-c. Expression unit e was made by inserting IFN- α 1 into the unique ClaI site of unit d. This construction makes possible a direct comparison of the levels of mRNAs produced by units c and e, using IFN- α 1 cDNA as the common probe. A direct comparison of units b and d is difficult due to very little homology at the DNA and mRNA level for hPGK versus yPGK. Sequences in common at 5' and 3' ends of the mRNAs are very short and very A-T rich.

Expression units f and g contain hybrid genes. They were constructed using synthetic DNA such that the junction is in the third nucleotide of a serine codon in yeast and human PGKs as shown in Fig. 2. This region is highly conserved as to primary amino acid sequence and is very close to the "hinge" region (13) between the two domains, as shown in Fig. 2. The first hybrid, yhPGK, was made using a unique KpnI restriction site in yeast sequence (location marked with X in Fig. 2) and synthetic DNA to the NcoI site in human sequence (marked Y). The second hybrid, hyPGK, was made using synthetic DNA from the NcoI site of human to the HpaII restriction site of yeast PGK (marked Z).

Protein Production by the Different Expression Units

The constructions described and shown in Fig. 1 were placed in yeast 20B-12 (α trp1 pep4-3) (20) using standard yeast transformation techniques (21). The yeast were grown on Trp⁺ selective media to the same optical density. Total SDS-soluble proteins were compared by SDS-polyacrylamide gel electrophoresis (22). Fig. 3 shows the result of such an analysis. In lanes 1 and 2 the yPGK gene products from the chromosome and a high copy number plasmid are compared. The difference in migration is due to overloading of yPGK in lane 2 versus lane 1. About 2 percent of the protein is PGK (marked by dot) in lane 1 versus 30-40 percent in lane 2 (quantitation by gel scanning). Lane 3 shows only 5 percent of the protein as hPGK. Note the difference in migration of hPGK versus yPGK protein. Lanes 4 and 5 are the yhPGK and hyPGK hybrids. Their production levels are

Domain A

Cooperative Protein Homology For Yeast, Human, and Equine PGK

Y 1 ser leu ser ser lys leu ser val gln asp leu asp leu lys asp lys
H 1 ser leu ser asn lys leu thr leu asp lys leu asp val lys gly lys
E 1 ser leu ser asn lys leu thr leu asp lys leu asn val lys gly lys

Y 17 arg val phe ile arg val asp phe asn val pro leu asp gly lys lys
H 17 arg val val met arg val asp phe asn val pro met lys asn asn gln
E 17 arg val val met arg val asp phe asn val pro met lys asn asn gln

Y 33 ile thr ser asn gln arg ile val ala ala leu pro thr ile lys
H 33 ile thr asn asn gln arg ile lys ala ala val pro ser ile lys
E 33 ile thr asn asn gln arg ile lys ala ala val pro ser ile lys

Y 48 tyr val leu glu his his pro arg tyr val val leu ala ser his leu
H 48 phe cys leu asp asn gly ala lys ser val val leu met ser his leu
E 48 phe cys leu asp asp gly ala lys ser val val leu met ser his leu

Y 64 gly arg pro asn gly glu arg asn glu lys tyr ser leu ala pro
H 64 gly arg pro asp gly val pro met pro¹⁵⁵ lys tyr ser leu glu pro
E 64 gly arg pro asp val gly pro met pro¹⁵⁵ lys tyr ser leu gln pro

Y 78 val ala lys glu leu gln ser leu leu gly lys asp val thr phe leu
H 80 val ala val glu leu lys ser leu leu gly lys asp val leu phe leu
E 80 val ala val glu leu lys ser leu leu gly lys asp val leu phe leu

Y 95 asn asp cys val gly pro glu val glu ala ala val lys ala ser ala
H 96 lys asp cys val gly pro glu val glu lys ala cys ala asn pro ala
E 96 lys asp cys val gly pro glu val glu lys ala cys ala asp pro ala

Y 111 pro gly ser val ile leu leu glu asn leu arg tyr his ile glu glu
H 112 ala gly ser val ile leu leu glu asn leu arg phe his val glu glu
E 112 ala gly ser val ile leu leu glu asn leu arg phe his val glu glu

Y 126 glu gly ser arg lys val asp gly gln lys val lys ala ser lys
H 128 glu gly¹⁷⁵ gly lys asp ala ser gly asn lys val lys ala glu pro
E 128 glu gly¹⁷⁵ gly lys asp ala ser gly asn lys val lys ala glu pro

Y 142 glu asp val gln lys phe arg his glu leu ser ser leu ala asp val
H 144 ala lys ile glu ala phe arg ala ser leu ser lys leu gly asp val
E 144 ala lys ile glu thr phe arg ala ser leu ser lys leu gly asp val

Y 158 tyr ile asn asp ala phe gly thr ala his arg ala his ser ser met
H 160 tyr val asn asp ala phe gly thr ala his arg ala his ser ser met
E 160 tyr val asn asp ala phe gly thr ala his arg ala his ser ser met

Y 174 val gly phe asp leu pro gln arg ala ala gly phe leu leu glu lys
H 176 val gly val asn leu pro gln lys ala gly phe leu met lys lys
E 176 val gly val asn leu pro gln lys ala gly phe leu met lys lys

Y 190 glu leu lys tyr phe gly lys ala leu glu asn pro thr arg pro phe
H 192 glu leu asn tyr phe ala lys ala leu glu ser pro glu arg pro phe
E 192 glu leu asn tyr phe ala lys ala leu glu ser pro glu arg pro phe

Domain B

Y 206 leu ala ile leu gly gly ala lys val ala asp lys ile gln leu ile
H 208 leu ala ile leu gly gly ala lys val ala asp lys ile gln leu ile
E 208 leu ala ile leu gly gly ala lys val ala asp lys ile gln leu ile

Y 222 asp asn leu leu asp lys val asp ser ile ile ile gly gly gly met
H 224 asn asn met leu asp lys val asp ser ile ile gly gly gly met
E 224 asn asn met leu asp lys val asp glu met ile ile gly gly gly met

Y 238 ala phe thr phe lys lys val leu glu asn thr glu ile gly asp ser
H 240 ala phe thr phe leu lys val leu asn asn met glu ile gly thr ser
E 240 ala phe thr phe leu lys val leu asn asn met glu ile gly thr ser

Y 254 ile phe asp lys ala gly ala glu ile val pro lys leu met glu lys
H 256 leu phe asp glu glu gly ala lys ile val lys asp leu met ser lys
E 256 leu phe asp glu glu gly ala lys ile val lys asp leu met ser lys

Y 270 ala lys ala lys gly val glu val val leu pro val asp phe ile ile
H 272 ala glu lys asn gly val lys ile thr leu pro val asp phe val thr
E 272 ala glu lys asn gly val lys ile thr leu pro val asp phe val thr

Y 286 ala asp ala phe ser ala asp ala asn thr lys thr val thr asp lys
H 288 ala asp lys phe asp glu asn ala lys thr gly gln ala thr val ala
E 288 ala asp lys phe asp glu his ala lys thr gly gln ala thr val ala

Y 302 glu gly ile pro ala gly trp gln gly leu asp asn gly pro glu ser
H 304 ser gly ile pro ala gly trp met gly leu asp cys gly pro glu ser
E 304 ser gly ile pro ala gly trp met gly leu asp cys gly thr glu ser

Y 318 arg lys leu phe ala ala thr val ala lys ala lys thr ile val trp
H 320 ser lys lys tyr ala glu ala val thr arg ala lys gln ile val trp
E 320 ser lys lys tyr ala glu ala val ala arg ala lys gln ile val trp

Y 334 asn gly pro pro gly val phe glu phe glu lys phe ala ala gly thr
H 336 asp gly pro val gly val phe glu trp glu ala phe ala arg gly thr
E 336 asn gly pro val gly val phe glu trp glu ala phe ala arg gly thr

Y 350 lys ala leu leu asp glu val val lys ser ser ala ala gly asn thr
H 352 lys ala leu met asp glu val val lys ala thr ser arg gly cys ile
E 352 lys ala leu met asp glu val val lys ala thr ser arg gly cys ile

Y 366 val ile ile gly gly gly asp thr ala thr val ala lys lys tyr gly
H 368 thr ile ile gly gly gly asp thr ala thr cys cys ala lys trp asn
E 368 thr ile ile gly gly gly asp thr ala thr cys cys ala lys trp asn

Y 382 val thr asp lys ile ser his val ser thr gly gly gly ala ser leu
H 384 thr glu asp lys val ser his val ser thr gly gly gly ala ser leu
E 384 thr glu asp lys val ser his val ser thr gly gly gly ala ser leu

Y 398 glu leu leu glu gly lys glu leu pro gly val ala phe leu ser glu
H 400 glu leu leu glu gly lys val leu pro gly val asp ala leu ser asn
E 400 glu leu leu glu gly lys val leu pro gly val asp ala leu ser asn

Y 414 lys lys
H 416 ile
E 416 val

Human/Yeast PGK Homologies

35% at DNA level
65% at protein level

Watson et al. EMBO J. 1:1635,1982.

essentially identical at 10 percent of the cell protein. By comparison with expression unit b, protein levels of all hPGK containing expression units are significantly decreased. Differences in migration of the various PGKs are due to differences in amino acid compositions (15). The above results are summarized in Table 1.

mRNA Production by the Different Expression Units

Fig. 4 is the result of "Northern" (23) analyses of mRNA levels resulting from the various expression units. Lanes 1-4 compare mRNA produced by units c and e on high copy number plasmids in yeast. IFN- α 1 DNA was used as a common probe. Lanes 1-4 were also analyzed with yPGK DNA as a probe to show that loading of mRNA is essentially identical (see part B of lanes 1-4). Loading is measured by the intensity of the 1500 nucleotide chromosomal mRNA marked. Only lane 1 shows a reduction in this 1500 band and is therefore underloaded. Based on these results and other "Northern" (not shown), we conclude that the hPGK mRNA produced by unit e is about 70 percent of that produced by unit c, quantitation done by removal of bands and scintillation counting. Therefore in contrast to other heterologous genes we have expressed using the same expression system (producing ≤ 10 percent of the mRNA produced by c; ref. 1), this gene (hPGK) produces 70 percent of the normal steady-state level of mRNA produced by unit c.

Lanes 5-8 show that d, e, and f produce mRNAs at the same level using hPGK DNA as a common probe. Therefore insertion of IFN- α 1 makes no difference in mRNA levels. The protein levels for the insertions were also the same (data not shown). Furthermore the results show that the construction of a hybrid PGK gene does not reduce steady-state levels of mRNA. The second hybrid, unit g, also produces mRNA at this same level (data not shown). Comparison of lanes 9 and 10 show that there is no

Figure 2. Amino acid sequence of three PGK enzymes and the structure of yeast PGK. The primary amino acid sequences of yeast (Y, 12), human (H, 35,10), and horse or equine (E, 14) PGKs are compared. The slash through the lysine shows that this amino acid by protein sequencing (35) was not found by DNA sequencing (10). The asterisks show identity. Below is the outline of the X-ray crystallographic results of Watson *et al.* (13) showing the two domain structure of yeast PGK. Note the two domains are connected by an α -helical structure represented by a spring-like structure. Its location in primary sequence is shown above. The two domains are thought to approach one another to bring the substrates (approximate location of binding shown for 3-PGA, 3-phosphoglycerates, and Mg-ATP) together after binding (13). Portions of the two domains may move 10-20Å with respect to each other. An arrow shows the directions of this movement. X, Y, and Z are corresponding restriction sites in the DNA coding for these amino acids (see text). The hybrid junction used to make yhPGK and hyPGK is shown (the third nucleotide of the serine codon).

Figure 3. SDS-polyacrylamide gel of proteins produced in yeast. Yeast were transformed using standard procedures (21). Yeast containing the expression units designated were grown to an A_{660} of 1 selectively and extracts prepared as previously described (1). Lanes 1-5 were loaded with .06 absorbance unit and lanes 8-12 with .6 unit each. The two 10 percent polyacrylamide gels (lanes 1-6 and 7-12) were run as previously described (1) and stained with Coomassie blue dye. Standards (1) are marked with K representing kilodaltons. PGK proteins are marked with dots. Expression unit a is chromosomal, while b-g are carried by YEp9T (high copy number plasmid), except for b and d in lanes 11 and 12, which are carried on the centromere (CEN) plasmid YCp50 (30).

reduction of mRNA associated with the insertion of IFN- α 1 DNA into unit b to make c. All these results are summarized in Table 1.

Nature of the Protein Product Produced

As shown in Table 1, all the PGKs produced have enzymatic activities. The relative activities correspond fairly well with percent of protein by gel scan, except for expression units a and b. The relative activities suggest that only 17 percent of cell protein should be PGK by gel scan of b, instead of 30-40 percent. However, we know that a large portion of yPGK produced by unit b (about 10 to 20 percent of the cell protein) is present

Table 1. PGK Expression Levels

Expression Unit	Gene Expressed	Copy Number	Relative Steady State mRNA Level	In Extract Units/mg Protein	Relative Activity	Percent of Protein by Gel Scan
a	Chromosomal PGK	1	1X	.110	1.0X	2
b	Yeast PGK	20-30	10X	.940	8.6X	30-40
d	Human PGK	20-30	7X	.380	3.5X	5
f	Yeast-Human Hybrid	20-30	7X	.820	7.6X	10
g	Human-Yeast Hybrid	20-30	7X	.820	7.6X	10
a	Chromosomal PGK	1	1X			2.0
b	yPGK-CEN	3	3X			4.0
d	hPGK-CEN	3	2X			0.6

Table 1. PGK expression levels. Genes expressed are either chromosomal (1 copy), on a high copy number plasmid (20-30 copies), or on a centromere plasmid (3 copies per cell). Relative steady-state mRNA levels were obtained from Fig. 4 results using band removal followed by scintillation counting. Glass bead extracts for PGK enzymatic assays were made at a cell density of one (A_{660}) and assayed as previously described (1). Specific activities of extracts are in micromoles of 3-phosphoglycerate phosphorylated per min. at 30°C per mg of extract protein. All activities have a standard deviation of about 10 percent. The percent of protein by gel scan was obtained from a LKB2022 Ultrascan Laser Densitometer at 633 m μ .

as an insoluble form which is solubilized by SDS-MSH extraction. The normal specific activity of this purified protein (soluble without SDS added, ref. 15) and the above results suggest an explanation for the difference in activity versus gel scan results.

The hPGK protein level is 6-8 fold less than yPGK by gel scans. This is not due to decreased mRNA level (Table 1). Furthermore, the protein produced is active and complements a yeast that is pgk⁻. This pgk⁻trp1⁻ yeast (18,24) does not produce any mRNA or PGK cross-reacting protein (results not shown). The purified hPGK and yPGK from these strains have identical specific activities and very similar kinetic constants (15).

Both hybrid PGK proteins are also enzymatically active and complement the same pgk⁻trp1⁻ yeast when present on a plasmid, allowing the yeast to grow on glucose. A comparison of the properties of these hybrid proteins, containing portions of yPGK and hPGK, is presented in another paper (15).

Figure 4. Analysis of PGK specific mRNAs. "Northern" analyses were done as previously described (1) with total RNA preparations isolated from cells (at Abs₆₆₀ of 1.0) containing the various expression units designated. Sizes of the mRNAs are given in nucleotides, sizes being previously determined (1). The ³²P-labeled (36) DNA probes used were the 560 bp *Eco*RI IFN- α 1 cDNA (37), the 540 bp *Eco*RI to *Nco*I fragment of hPGK (see Fig. 1), and the 150 bp *Eco*RI to *Bgl*II fragment of yPGK (12). Lanes 1-4 are an autoradiogram of a "Northern" blot (A) first probed with IFN- α 1 and then re-probed and re-exposed (B) with yPGK probe to obtain the lower mRNAs (the second exposure was 5 times longer). Lanes 5-8 are an autoradiogram of a "Northern" blot probed with hPGK DNA. The lower bands are the same bands but with 5 times the development time. Lanes 9 and 10 were obtained using yPGK as a probe.

Codon Bias, Context, and Protein Stability Effects on Expression

Looking at Table 1, the 6-8 fold difference in the relative amounts of hPGK versus yPGK cannot be due to mRNA levels. It is possible that hPGK is

Figure 5. Comparison of hPGK and yPGK protein stabilities *in vivo*. Cultures of yeast containing plasmid-borne expression units b and d were grown in minimal media without methionine (2) to A_{660} of 1.0 at 30°C. ^{35}S -methionine (200 μCi) was added to 3.5 ml of each culture with a labelling time of 1h. The cultures were then centrifuged and washed two times with equal volumes of the same media without radioactive methionine. The cultures were resuspended in equal volumes of media containing cold methionine and growth continued removing 0.5 ml samples at the indicated times after washing. Extracts of these were made as previously described (2), loading 2×10^6 cpm for each time point on a 10 percent polyacrylamide SDS gel (21) (except lanes 6 and 12 with one-half the counts). During this growth period the culture went from an A_{660} of 1.5 to 8.0. After drying the gels an autoradiogram was made by exposure to film for 16h. The positions of hPGK, yPGK, and protein X are designated. yPGK and hPGK are of similar intensities due to their difference in methionine content of 3 as compared to 13 residues per molecule, respectively.

less stable in yeast. We have ruled out this possibility using pulse-chase experiments with ^{35}S -methionine. As seen in Fig. 5, after 10 hours of chase with cold methionine, we see no difference in the relative stability of these two proteins by comparison with protein X. Therefore we conclude that the human mRNA is less efficiently translated than the yPGK mRNA. Since both hybrid gene protein products are present in approximately the same amounts (whether yeast or human gene sequence is first), it seems

unlikely that this defect is in translational initiation, but may be in the translational rate of protein elongation.

If the hPGK mRNA is defective in supporting normal protein synthesis, the mRNA could be defective due to structural constraints, codon context (effects of adjacent codons on an intermediate codon), or codon usage. Codon context effects (e.g, tRNA-codon or tRNA-tRNA interactions, 25) as well as structural constraints (e.g., secondary structure) should be independent of mRNA concentration in the cell, while codon usage effects could be related to mRNA concentration. Codon usage may affect translation to a greater extent at a high concentration of nonpreferred codons or mRNA (hPGK mRNA from the 2 μ plasmid is 18 percent of total yeast mRNA) if certain aminoacyl-tRNAs are being used at a faster rate than they can be regenerated. It has been previously shown that there is a strong correlation between the abundance of yeast tRNAs and the occurrence of the respective codons in protein genes and that some of these tRNAs are present at extremely low levels (38). All other mRNAs containing these codons would also have their translation reduced. Therefore if hPGK and yPGK mRNA concentrations are reduced to the same lower level in the cell, the relative amount of protein product produced by each might change.

To test this possibility, the following experiments were done. Expression unit d in Fig. 1 was put on an integrating plasmid, YIp5 (27); however, transformants analyzed contained plasmids of copy numbers of ≥ 10 copies per cell by "Southern" analysis (data not shown). We suspect that the hPGK cDNA contains a DNA sequence which can act as an origin of replication in yeast. Human sequences which behave like this in yeast have been previously described (28).

As an alternative way to obtain lower copy number, we placed expression units b and d on separate plasmids containing a centromere (29). The plasmid YCp50 (30) contains the centromere from chromosome IV, the ARS1 origin of replication, and the yeast URA3 gene as a selectable marker. When yeast strain TE411 (trp1⁻ ura3⁻) (31) is transformed with either of these two plasmids, the two plasmids maintain an average copy number of 3 per cell, due to the centromere. This copy number was determined by Southern analysis using chromosomal PGK (unit a, Fig. 1) as a standard of 1. Other characteristics of these expression systems are summarized in Table 1. Relative steady-state levels of mRNA correspond with copy numbers. Percents of protein (PGKs designated with dots) were determined using gel scans of lanes 8-12 (Fig. 3). The chromosomal contribution (lane

8) to the yPGK produced by the centromere (CEN) plasmid (lane 11) has been subtracted. The hPGK produced by hPGK-CEN (lane 12) is separate on the gel from chromosomal yPGK. Again there is a 6-fold difference in the reduced production levels of hPGK and yPGK. Although yeast is capable of producing a level of hPGK equivalent to that seen in lane 9 (at 20 copies of unit d per cell), it produces only one-sixth this level in lane 12 (at 3 copies of unit d per cell). Thus a lower concentration of mRNA for hPGK does not improve its protein producing ability.

CONCLUSIONS AND DISCUSSION

As many as twenty different heterologous genes with yeast PGK 5'- and 3'-flanking DNA sequences demonstrate an expression defect due to a lowering of steady-state levels of mRNA (1,9). Partly as a result of this, these systems do not produce protein at the level produced from the yeast PGK gene. Our earlier data suggested that the defect is most likely due to a loss of expressional information within the structural gene for yPGK, which is not present in the heterologous genes. We also showed that many changes throughout the yPGK gene (without addition of a foreign gene) result in the loss of this information (1).

One of the models we have suggested to explain this phenomenon is that the yPGK protein itself may contain information for the stabilization of its mRNA during translation. Since the primary and especially the three-dimensional structures of PGK enzymes are so highly conserved among species, we decided to express the cDNA for human PGK (10,11) using the 5'- and 3'-flanking sequences from the yeast PGK gene. If this conservation of protein structure also retains the structure for such a hypothetical interaction, this human gene may retain such function. We found that the expression of the hPGK cDNA is almost normal with respect to steady-state levels of mRNA, unlike many other human genes tried. Therefore the hPGK cDNA apparently contains the information which is necessary to maintain high steady-state levels of mRNA. This is the primary result reported here. We think it supports the protein-feedback model. Furthermore, we have found that gene fusions at the 5' end of yPGK (e.g. human serum albumin and human epidermal growth factor) which retain all of the yPGK protein structure retain high steady-state levels of mRNA (data not shown). This is in stark contrast to similar fusions where part of the PGK gene sequence is deleted and steady-state levels of mRNA are greatly diminished (1,9).

The codon usage of the hPGK gene (10,32) is very similar to the codon

Human IFN- α 1				Human IFN- α 1				Human IFN- α 1				Human IFN- α 1			
Human PGK		Yeast PGK		Human PGK		Yeast PGK		Human PGK		Yeast PGK		Human PGK		Yeast PGK	
4	10	1	•UUU/phe	4	7	16	UCU/ser	1	2	0	•UAU/tyr	4	4	1	UGU/cys
4	6	18	UUC/phe	4	4	6	UCC/ser	3	2	7	UAC/tyr	1	3	0	•UGC/cys
2	1	5	UUA/leu	2	1	2	•UCA/ser	1	0	1	UAA/OC	0	0	0	UGA/OP
4	7	36	UUG/leu	0	2	0	•UCG/ser	0	1	0	UAG/AM	2	4	2	UGG/trp
<hr/>				<hr/>				<hr/>				<hr/>			
0	5	0	•CUU/leu	3	8	0	•CCU/pro	2	2	1	•CAU/his	0	0	3	•CGU/arg
8	7	0	•CUC/leu	2	2	0	•CCC/pro	1	3	7	CAC/his	0	1	0	•CGC/arg
1	3	0	•CUA/leu	1	7	17	CCA/pro	2	1	8	CAA/gln	1	2	0	•CGA/arg
7	15	0	•CUG/leu	0	0	0	•CCG/pro	8	5	0	•CAG/gln	0	4	0	•CGG/arg
<hr/>				<hr/>				<hr/>				<hr/>			
0	8	9	AUU/ile	2	10	10	ACU/thr	2	11	1	•AAU/asn	0	2	0	•AGU/ser
7	7	14	AUC/ile	4	4	8	ACC/thr	4	12	13	AAC/asn	3	8	2	•AGC/ser
0	4	0	•AUA/ile	3	1	0	•ACA/thr	4	16	2	•AAA/lys	6	2	10	AGA/arg
7	14	4	AUG/met	0	2	0	•ACG/thr	4	26	40	AAG/lys	5	2	0	•AGG/arg
<hr/>				<hr/>				<hr/>				<hr/>			
1	8	16	GUU/val	4	23	32	GCU/ala	5	10	8	GAU/asp	0	12	35	GGU/gly
2	15	22	GUC/val	3	16	10	GCC/ala	6	13	18	GAC/asp	1	11	1	•GGC/gly
0	4	0	•GUA/val	2	2	1	•GCA/ala	6	10	28	GAA/glu	2	8	0	•GGA/gly
4	12	0	•GUG/val	0	0	0	•GCG/ala	9	17	1	•GAG/glu	0	9	1	•GGG/gly

HUMAN IFN- α 1	19K	• Least preferred codons in yeast
HUMAN PGK	45K	
YEAST PGK	45K	* Less preferred codons in yeast

Figure 6. Codon usage comparison of three genes. The sequences of human IFN- α 1 cDNA (33), the human PGK gene (10), and the yeast PGK gene (12) have been previously determined. The least and less preferred codons in yeast are based on a prior tabulation of codon usage for high and low expressed genes in yeast (26). The sizes of the three proteins are given in kilodaltons (K).

usage of other mammalian genes we have expressed (1,2) and is quite different from that of yeast (26). A thorough comparison is shown in Fig. 6, with IFN- α 1 codon usage as the example of another human gene. Thus it appears that codon bias is not a primary factor in the decrease of steady-state levels of mRNA observed for other human heterologous genes (such as IFN- α 1, which exhibits a 10-fold decrease in mRNA; ref. 1), since the hPGK mRNA levels remain relatively high, even with this same codon usage. Previously published data also support this conclusion (1). Furthermore we have found that some synthetic heterologous genes with preferred yeast codon bias do not express well in yeast (data not shown). Human codon usage contributes at most to a small 30 percent or less drop in the hPGK mRNA.

There is a 6–8 fold lower production of hPGK protein from similar mRNA levels. The possibility that the hPGK protein is unstable in the cell is unlikely due to the pulse-chase experiments (Fig. 5), which show that the

proteins are degraded in vivo at the same rate. Another possibility is that translation of the hPGK mRNA is not properly initiated. However, this possibility is strongly addressed by data concerning the two hybrid genes. Whether the hPGK DNA sequence is first or second, mRNA levels are the same and protein levels are essentially identical. It is interesting that the level of chimeric protein is twice that of hPGK. This suggests that getting rid of either half of hPGK DNA sequence gets rid of half the protein production problem. Thus, the problem is probably not in translational initiation but more likely in the translational elongation rate, with both halves of hPGK contributing essentially equally.

Reducing the plasmid copy number and the level of hPGK mRNA did not change its translational efficiency. The ratios of hPGK to yPGK mRNA and protein remained about the same. Therefore aminoacyl-tRNAs are probably not being used at a faster rate than they can be generated at the higher level of hPGK mRNA. However, it is possible that the cell is limiting for certain aminoacyl-tRNAs at both concentrations of the hPGK mRNA and that this limitation normally affects the rate of translation of all mRNAs, depending on their content of the corresponding codons. One possible mechanism for this occurring would be the competition between abundant and rare aminoacyl-tRNAs for the ribosomal binding site as a rate-limiting step for translation. In vitro translation studies using a yeast system (34) which compares various mRNAs may better address differences in these rates of translation.

In terms of evolutionary diversity, it is interesting that an enzyme which apparently functions by extensive interaction of two substrate binding lobes (13,14) can still function well when one lobe is from yeast and the other from human. Not only do these lobes need to bind substrate but they need to approach one another in order for substrates to interact. The specific activities and rate constants of these hybrid proteins are very close to those of hPGK and yPGK (15). Important structure-function relationships may be obtained by further study of these hybrids.

If the PGK protein regulates the steady-state level of its own mRNA and translatability of the mRNA is decreased, a protein-dependent mRNA protection should begin to fail. This failure does not appear to be a linear event since a 6-8 fold decrease in translation only results in a 30 percent drop in mRNA level. More dramatic decreases in translation should result in even greater decreases in mRNA levels. It is disturbing that there is not a linear relationship between mRNA levels and protein levels;

however, the quality of the protein may be more important than the quantity. This non-linearity suggests that both human and yeast PGK mRNAs have evolved to be very stable (second model). However the breakdown in this stability is very sensitive to slight changes throughout the yPGK mRNA, and due to the great difference between yPGK and hPGK mRNA sequences, one might expect that different stabilizing structures may have evolved in these mRNAs from distinctly different organisms. Nevertheless, the two hybrid genes which have no sequence in common retain this stability and the proteins produced retain enzymatic activity. To further test the possibility that the protein may contain information to stabilize its own mRNA, the authors are trying to make missense mutations which may cause a drop in mRNA levels. With such missense mutations, it may be possible to obtain temperature-sensitive revertants which show a temperature-sensitive effect on mRNA levels.

We also suggest that the function of the hybrids and the close structural similarities between hPGK and yPGK may have resulted from more than a conservation of enzymatic mechanisms. These similarities may have resulted from a conservation of structural characteristics associated with the maintenance of high steady-state levels of mRNA. If indeed such a novel mechanism exists, we visualize it occurring during translation with a certain protein folding occurring to somehow protect the mRNA from nuclease attack. Domains completely different from normal enzymatic domains may be associated with such a process.

ACKNOWLEDGMENTS

The authors wish to thank Arthur D. Riggs and Herman de Boer for critical reviews of the manuscript, Dennis Henner for the suggestion that certain aminoacyl-tRNAs may be a limiting factor at both hPGK mRNA concentrations, and Jeanne Arch for typing the manuscript. The authors wish to thank Ronald W. Davis for plasmid YCp50 and Robert Elder for YIp5. The human cDNA was kindly supplied by Arthur Riggs (see refs. 11 and 15).

References and Notes

1. Chen, C.Y., Oppermann, H. and Hitzeman, R.A. (1984) Nucleic Acids Res. **12**, 8951-8970.
2. Hitzeman, R.A., Chen, C.Y., Hagie, F.E., Patzer, E.J., Liu, C., Estell, D.A., Miller, J.V., Yaffe, A., Kleid, D.G., Levinson, A.D., and Oppermann, H. (1983) Nucleic Acids Res. **11**, 2745-2763.
3. Mellor, J., Dobson, M.J., Roberts, N.A., Tuite, M.F., Emtage, U.S., White, S., Lowe, P.A., Patel, T., Kingsman, A.J., and Kingsman, S.M. (1983) Gene **24**, 1-14.

4. Mellor, J., Dobson, M.J., Roberts, N.A., Kingsman, A.J., and Kingsman, S.M. (1985) Gene 33, 215-226.
5. Gluzman, Y. and Shenk, T., Eds., Enhancers and Eukaryotic Gene Expression (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983).
6. Sentenac, A. and Hall, B. (1982) In The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression, J.N. Strathern, E.W. Jones, J.R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), pp. 561-606.
7. Losson, R. and Lacroute, F. (1979) Proc. Natl. Acad. Sci. USA 76, 5134-5137.
8. Zitomer, R.S., Montgomery, D.L., Nichols, O.L., Hall, B.D. (1979) Proc. Natl. Acad. Sci. USA 76, 3627-3631.
9. Chen, C.Y. and Hitzeman, R.A. (1987) In Biological Research on Industrial Yeasts: Fundamental Aspects, G.G. Stewart, ed., Vol. 2, CRC Press, in press.
10. Michelson, A.M., Markham, A.F., and Orkin, S.H. (1983) Proc. Natl. Acad. Sci. USA 80, 472-476.
11. Singer-Sam, J., Simmer, R.L., Keith, D.H., Shively, L., Teplitz, M., Itakura, K., Gartler, S.M., and Riggs, A.D. (1983) Proc. Natl. Acad. Sci. USA 80, 802-806.
12. Hitzeman, R.A., Hagie, F.E., Hayflick, J.S., Chen, C.Y., Seeburg, P.H., and Derynck, R. (1982) Nucleic Acids Res. 10, 7791-7808.
13. Watson, J.C., Walker, N.P.C., Shaw, P.J., Bryant, T., Wendell, P.L., Fothergill, L.A., Perkins, R.E., Conroy, S.C., Dobson, M.J., Tuite, M.F., Kingsman, A.J., and Kingsman, S.M. (1982) EMBO J. 1, 1635-1640.
14. Banks, R.D., Blake, C.C.F., Evans, P.R., Haser, R., Rice, D.W., Hardy, G.W., Merritt, M., and Phillips, A.W. (1979) Nature (London) 279, 773-777.
15. Mas, M.T., Chen, C.Y., Hitzeman, R.A., and Riggs, A.D. (1986) Science 233, 788-790.
16. Southern, E.M. (1975) J. Mol. Biol. 99, 503-517.
17. Hitzeman, R.A., Chinault, A.C., Kingsman, A.J., and Carbon, J. (1979) In Eucaryotic Gene Regulation, Vol. 14, R. Axel and T. Maniatis, Eds. (Academic Press, New York), pp. 57-68.
18. Hitzeman, R.A., Clark, L., and Carbon, J. (1980) J. Biol. Chem. 255, 12073-12080.
19. Zoller, M.J. and Smith, M. (1982) Nucleic Acids Res. 10, 6487-6500.
20. Jones, E. (1976) Genetics 85, 23-33.
21. Hinnen, A., Hicks, J.B., and Fink, F.R. (1978) Proc. Natl. Acad. Sci. USA 75, 1929-1933.
22. Laemmli, U.K. (1977) Nature (London) 227, 680-685.
23. Dobner, P.R., Kawasaki, E.S., Yu, L.Y., and Bancroft, F.C. (1981) Proc. Natl. Acad. Sci. USA 78, 2230-2234.
24. Lam, K. and Marmur, J. (1977) J. Bacteriol. 130, 746-749.
25. de Boer, H.A. and Kastelein, R.A. (1986) In Maximizing Gene Expression (Biotechnology Series, vol. 9) W. Reznikoff and L. Gold, Eds. (Butterworths, Stoneham, Mass.) in press.
26. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026-3031.
27. Struhl, K., Stinchcomb, D.T., Scherer, S., and Davis, R.W. (1979) Proc. Natl. Acad. Sci. USA 76, 1035-1039.
28. Stinchcomb, D.T., Thomas, M., Kelly, J., Selker, E., and Davis, R.W. (1980) Proc. Natl. Acad. Sci. USA 77, 4559-4563.
29. Clarke, L. and Carbon, J. (1980) Nature (London) 287, 504-509.
30. Johnston, M. and Davis, R.W. (1984) Mol. Cell Biol. 4, 1440-1448.
31. Etcheverry, M.T. (1985) In Genome Rearrangement (Alan R. Liss, Inc.), pp. 221-232.

32. Michelson, A.M., Blake, C.C.F., Evans, S.T., and Orkin, S.H. (1985) Proc. Natl. Acad. Sci. USA **82**, 6965-6969.
33. Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., McCandliss, R., Lawn, R.W., Seeburg, P.H., Ullrich, A., Yelverton, E., and Gray, P.W. (1981) Nature (London) **290**, 20-26.
34. Gasior, E., Herrera, F., Sadnik, I., McLaughlin, C.S., and Moldave, K. (1979) J. Biol. Chem. **254**, 3965-3969.
35. Huang, I., Welch, C.D., and Yoshida, A. (1980) J. Biol. Chem. **255**, 6412-6420.
36. Taylor, J.M., Illmensee, R., and Summers, S. (1976) Biochem. Biophys. Acta **442**, 324-330.
37. Hitzeman, R.A., Hagie, F.E., Levine, H.L., Goeddel, D.V., Ammerer, G., and Hall, B.D. (1981) Nature (London) **293**, 717-722.
38. Ikemura, T. (1982) J. Mol. Biol. **158**, 573-597.