Volume 10 Number 23 1982 Nucleic Acids Research

The primary structure of the *Saccharomyces cerevisiae* gene for 3-phosphoglycerate kinase

Ronald A.Hitzeman, Frank E.Hagie, Joel S.Hayflick, Christina Y.Chen, Peter H.Seeburg, and Rik Derynck

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

Received 5 August 1982; Revised and Accepted 26 October 1982

ABSTRACT

The DNA sequence of the gene for the yeast glycolytic enzyme, 3-phosphoglycerate kinase (PGK), has been obtained by sequencing part of a 3.1 kbp HindIII fragment obtained from the yeast genome. The structural gene sequence corresponds to a reading frame of 1251 bp coding for 416 amino acids with no intervening DNA sequences. The amino acid sequence is approximately 65 percent homologous with human and horse PGK protein sequences and is in general agreement with the published protein sequence for yeast PGK. As for other highly expressed structural genes in yeast, the coding sequence is highly codon biased with 95 percent of the amino acids coded for by a select 25 codons (out of 61 possible). Besides structural DNA sequence, 291 bp of 5'-flanking sequence and 286 bp of 3'-flanking sequence were determined. Transcription starts 36 nucleotides upstream from the translational start and stops 86-93 nucleotides downstream from the translational stop. These results suggest a non-polyadenylated mRNA length of 1373 to 1380 nucleotides, which is consistent with the observed length of 1500 nucleotides for polyadenylated PGK mRNA. A sequence TATATATAAA is found at 145 nucleotides upstream from the translational start. This sequence resembles the TATAAA box that is possibly associated with RNA polymerase II binding.

INTRODUCTION

Glycolytic genes from yeast are of great interest due to their combined high expression of 65 percent or more of the soluble protein in yeast (1). Two of these enzymes, 3-phosphoglycerate kinase (PGK) and glyceraldehyde-3-phosphate dehydrogenase, individually constitute 4 to 10 percent of the soluble protein depending on growth conditions (1, 2). The high level of glyceraldehyde-3-phosphate dehydrogenase may result in part from the presence of three copies of the gene (3); while PGK is present as only one copy per haploid cell (4, 5). Furthermore, high expression appears to correlate with the high levels of mRNA from these genes (6).

The gene for PGK has been previously cloned from the yeast genome on a plasmid by an immunological screening technique (5). Early characterization of this plasmid suggested that it contained the PGK structural gene on a 3.1

0305-1048/82/1023-7791 $2.00/0

kbp HindIII fragment. A DNA insertion at the EcoRI site (see Fig. 1) in this fragment was shown to affect gene expression adversely (5). We have further characterized the 3.1 kbp HindIII fragment by sequencing 1828 bp from a PvuI site to a HindIII site. This DNA fragment contains the EcoRI site within the structural gene as well as 5'- and 3'-flanking sequences for the PGK structural gene.

Recent studies of yeast gene expression have focused on comparisons of the DNA sequences of 5'- and 3'-flanking regions of isolated genes. All genes apparently have a TATAAA-like sequence at various distances upstream (5') from the translational start (3, 7-10). This sequence is thought to be associated with RNA polymerase II positioning and binding in vivo (11, 12) and is usually about 30 nucleotides upstream from the translational start in most multicellular eukaryotes (13). However, in the lower eukaryote yeast this distance is quite variable (11). Yeast genes also demonstrate variable numbers of transcription starts for the same gene with as many as seven for iso-1-cytochrome c (11) and two for alcohol dehydrogenase I (ADH-1) (14). In one gene system, this variation has been shown to be involved with the regulation of gene expression. The SUC2 gene produces a 1.8 kb transcript for constitutive levels of an intracellular invertase; while a larger 1.9 kb transcript, which includes the secretion pre-sequence of invertase, is derepressed to give the extracellular form of this enzyme (16).

The untranslated 3'-flanking regions of yeast genes are thought to be associated with at least two functions: the termination of transcription and polyadenylation of the 3' end of the transcript. A deletion of 38 bp in the 3'-flanking sequence of the CYC1 gene (17) results in 5-10 percent of the wild type level of iso-1-cytochrome c and only 10 percent of the normal level of mRNA. However, the mRNA in this mutant is terminating in other regions 3' to the structural gene. By comparison of the sequence defined by this deletion with other 3'-flanking gene sequences, Zaret and Sherman (17) have suggested a consensus sequence for transcription termination and polyadenylation for most yeast genes. Bennetzen and Hall (14) have suggested another consensus sequence thought to be involved with these same functions.

We now present the sequence for the PGK structural gene and flanking regions and compare it with those of other yeast genes to possibly identify the factors involved in high expression observed for glycolytic genes. We also determine the presence and relevance of some consensus sequences thought to be involved with gene expression in yeast. Furthermore, we compare the

yeast PGK amino acid sequence with amino acid sequence determined for PGKs from higher eukaryotes.

MATERIALS AND METHODS

Materials

Restriction endonucleases, T4 DNA ligase, and polynucleotide kinase were purchased from BRL or New England Biolabs and used essentially as recommended by manufacturer. *E. coli* DNA polymerase I (large fragment) was from Boehringer and S1 nuclease from Miles. ATP and deoxyribonucleotide triphosphates were obtained from PL Biochemicals; while ^{32}P-labeled nucleotides were obtained from Amersham Radiochemicals. Glass beads (0.45-0.50 mm) were purchased from B. Braun Melsungen AG. The oligonucleotide primer (5'-ATTTGTTGTAAA-3') was synthesized by conventional means (18).

Strains, plasmids, and growth conditions

E. coli K-12 strain 294 (*endA thr*$^-$ *hsr*$^-$ *hsm*$_K^+$) (19) was used for bacterial transformations. *S. cerevisiae* strains 20B-12 (*α trp1 pep4-3*) (20) and GM3C-2 (*a leu2-3 leu2-112 trp1-1 his4-519 cyc1-1 cyp3-1*) (11) were used for yeast transformations. Five plasmids were used, some of which have been previously described: pB1 (5), YEp13 (21), YRp7' (22, 23), PGK-YEp13, and pFRM31. Plasmid pB1 contains the 3.1 kbp PGK *Hin*dIII fragment, isolated by immunological screening (5), in the *Hin*dIII site of pBR322. Plasmid YEp13 contains the yeast 2μ origin of replication and the yeast *LEU2* gene for complementation of the double *leu2* mutation in GM3C-2. YRp7' contains the 1.4 kbp *TRP1* *Eco*RI fragment (23) in pBR322 in the opposite orientation as compared to YRp7 (22). This fragment contains the *ars1* origin of replication and the *TRP1* gene for complementation of the *trp1* mutation in either 20B-12 or GM3C-2. PGK-YEp13 contains the 3.1 kbp PGK *Hin*dIII fragment from pB1 substituted for the small *Hin*dIII fragment in YEp13 (orientation of PGK is with transcription toward adjacent DNA from yeast 2μ plasmid). Plasmid pFRM31 was constructed by inserting the 3.1 kbp PGK *Hin*dIII fragment into the pBR322 *Hin*dIII site of pFRD7 (which is YRp7' with the *Eco*RI sites removed by filling in *Eco*RI ends using large fragment of *E. coli* DNA polymerase I).

LB medium for *E. coli* growth was as described by Miller (24) with the addition of 20 μg/ml ampicillin (Sigma) for selection of plasmid transformants. Yeast were grown on the following media: YEPD (nonselective) contained 1 percent yeast extract, 2 percent peptone and 2 percent glucose with or without 3 percent Difco agar. YNB+CAA (used for Trp$^+$ selection)

contained 6.7 grams of Difco yeast nitrogen base (without amino acids) (YNB), 10 mg of adenine, 10 mg of uracil, 5 grams Difco casamino acids (CAA), 20 grams glucose and with or without 30 grams agar per liter. YNB-leu (leucine absent for Leu^+ selection) contained the same components as YNB+CAA; however, with 20 ml of -leu drop-out mix replacing the CAA (-leu drop-out mix contained these amino acids per 100 ml of H_2O: 0.2g arg, 0.1g his, 0.6g ile, 0.4g lys, 0.1g met, 0.6g phe, 0.5g thr, and 0.4g trp). The transformed yeast were always grown in media for selective maintenance of the plasmid.

Plasmid DNA preparation and transformations

Purification of plasmid DNAs from E. coli (25) and transformation of E. coli (26) were done in accordance with previously described procedures. E. coli plasmid identification was as described by Birmboim and Doly (27). Transformation of yeast was done as previously described (28).

DNA sequence determination

DNA sequencing of the phosphoglycerate kinase gene and flanking regions was done both by the chemical degradation method (29), using 5' ^{32}P-labelled restriction fragments, and by the dideoxy-method (30). For the latter approach the restriction fragments containing Sau3A ends (Fig. 1), were subcloned into the BamHI site of the single stranded phage M13 mp8 (31). The dideoxy-chain termination procedure was used as described using a synthetic phage-specific primer.

mRNA analysis

Preparation of yeast total RNA was done as previously described (32) growing the yeast under selective conditions for plasmid retention to an A_{660} of 1.0. Total yeast RNA was then denatured and electrophoresed on a 1.0 percent agarose gel containing 6 percent formaldehyde and 1X MOPS buffer (20 mM MOPS, 5 mM NaOAc, 1 mM EDTA at pH 7.0) (33). The gel was stained with ethidium bromide using the yeast or E. coli ribosomal RNAs (34) and HindIII cut phage λ DNA (35) as size standards. The gel was transferred to nitrocellulose and hybridized (36) with the EcoRI-BglII PGK-termination DNA fragment (from pB1) that was ^{32}P-labelled using the calf thymus primer method (37).

The initiation of transcription of the PGK gene was determined using cDNA extension (38) from a specific 5' ^{32}P-labelled DNA primer, 5' ATTTGTTGTAAA 3', complementary to nucleotides -10 to -21 (Fig. 2). Polyadenylated mRNA was prepared from yeast pFRM31/20B-12 (39). The extension product was sized on a 15 percent polyacrylamide-7M urea gel

alongside a sequencing ladder.

The polyadenylation site was mapped with the S1 mapping procedure (40), using the conditions of Mantei et al. (41). The BglII-HindIII PGK-terminator fragment from pB1 (5), 3' labelled at the BglII site using the Klenow fragment of E. coli DNA polymerase I and α^{32}P-dCTP, was hybridized to 10 µg mRNA from pFRM31/20B-12 followed by S1 nuclease digestion. The reaction products were run on a 6 percent polyacrylamide-7M urea gel. The ^{32}P-labelled HpaII fragments of pBR322 (42) were used as size markers.

RESULTS

Primary structure of the yeast PGK gene and protein

The 3.1 kb HindIII fragment, carrying the PGK gene, was isolated from plasmid pB1 (5). The recognition sites for several restriction endonucleases were mapped using standard techniques. The resulting preliminary restriction map served as a basis for the DNA sequence determination. Both the Maxam-Gilbert and dideoxy method were used to determine the sequence of the PvuI-HindIII segment by the strategy shown in Fig. 1. The arrows illustrate direction and scope of the sequencing as well as the method employed. Most regions were determined by sequence analysis in both directions (both strands) or using both methods.

The DNA sequence of the PvuI-HindIII fragment (1828 bp) is shown in Fig. 2. The largest open reading frame is 1248 nucleotides long, thus coding for a protein of 416 amino acids, with a calculated molecular weight of 44,746 daltons for the unmodified polypeptide. This value agrees well with the previous estimations of 45,000 to 50,000 daltons (43). In analogy with other yeast and higher eukaryotic proteins, it is very likely that the NH_2-terminal methionine is removed. The adjacent serine is then probably

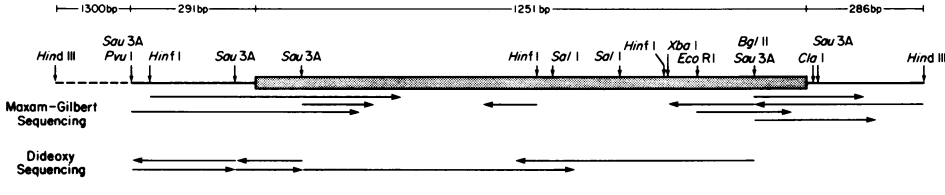


Figure 1. DNA sequencing strategy for the PGK gene. The 3.1 kbp HindIII fragment from plasmid pB1 (5) was sequenced from the PvuI site to the HindIII site (1828 bp) using two sequencing techniques for strands, distances, and directions as indicated by the arrows. The structural gene is shown as the bar region. Restriction sites are to scale except from HindIII to PvuI on the left which was not sequenced.

modified since automated Edman-degradation (44) of the native protein from yeast did not result in the release of terminal amino acids (43). This was similarly observed for the human (45) and horse (46) PGKs.

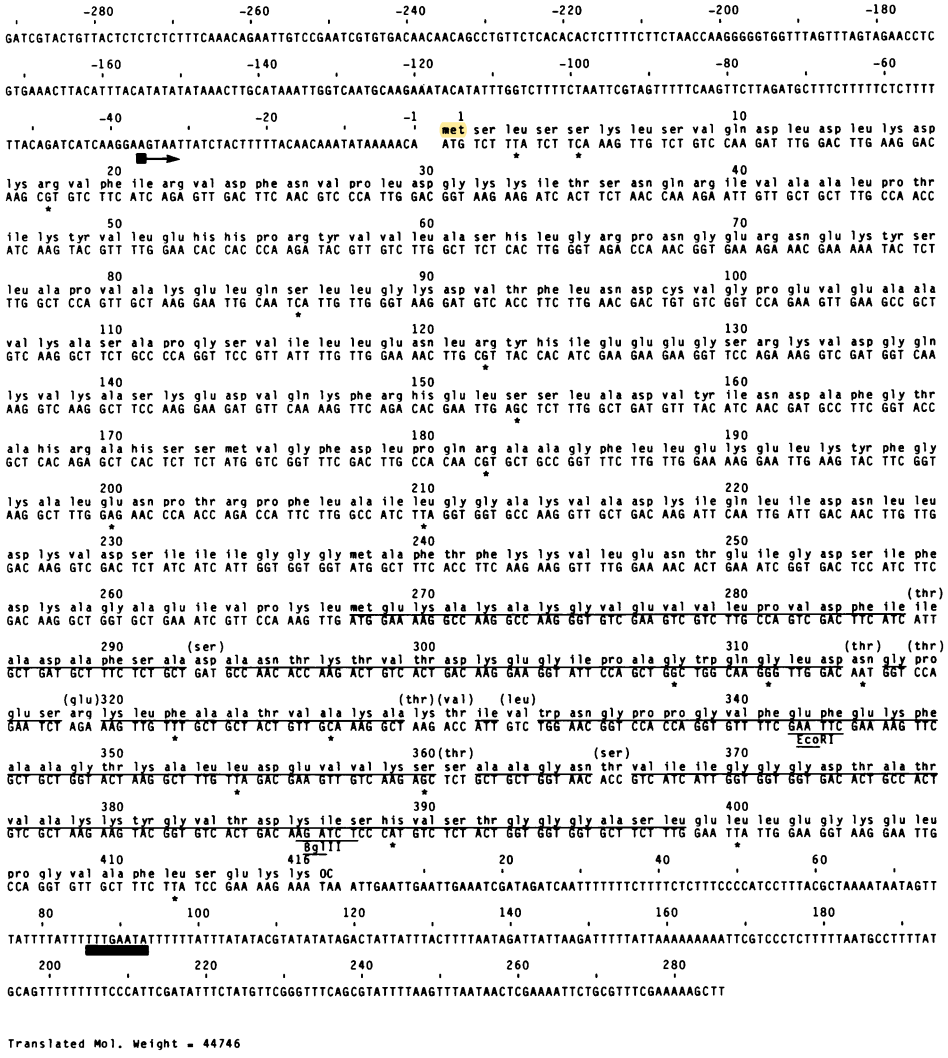
Recently, Dobson et al. (47) have reported the partial amino acid sequence (residues 270-400) of the CN2 cyanogen bromide fragment (amino acids 270 to 419) from yeast PGK. Their sequence corresponds well with the amino acids 268 to 398 of our deduced polypeptide sequence, as shown by the underlined residues in Fig. 2. However, assuming the removal of the terminal methionine, there is a difference of three amino acids in the numbering. Their numbering corresponds to the analogous residues in the human PGK protein (45). Our protein sequence, as deduced from the DNA sequence, differs in 10 out of the 130 residues, with their directly determined protein sequence. A single base change in the codons can account for seven of these differences (differences designated by amino acids in parentheses in Fig. 2). They can therefore be explained by some variation depending on the yeast strain. Although such differences have been seen for other yeast genes, it seems unlikely that strain variations would account for the three additional amino acid differences, which correspond to codon changes of two or three nucleotides.

PGK gene codon bias

Recently, Bennetzen and Hall (48) have tabulated the codon usage of seven yeast genes and noted that seven codons were never used (CUC, leu; CUG, leu; CCG, pro; CGU, arg; CGC, arg; CGA, arg; CGG, arg) and that highly expressed genes use fewer codons than genes that are expressed at much lower levels. Furthermore, for the two highly expressed genes, alcohol dehydrogenase I (ADH-1) and glyceraldehyde-3-phosphate dehydrogenase, 96 percent of the amino acids are coded for by 25 of the 61 possible coding triplets. The ADH-1 gene uses only 33 of 61 possible codons and 30 of these are used more than once (14). The same preference is observed in the yeast enolase genes (49).

Fig. 3 shows that a similar codon bias is present in the yeast PGK gene. Although more different codons (38 of 61) are used than in the genes for ADH-1, glyceraldehyde-3-phosphate dehydrogenase, and enolase; a very strong preference for 25 select codons, accounting for 95 percent of the codon capacity, is apparent. Therefore PGK codon usage is similarly restricted as for the other highly expressed genes, in contrast with the less biased genes which demonstrate low level expression in yeast. As shown by asterisks in Fig. 2, non-preferred codons are randomly spaced throughout

the sequence. A similarly random spacing of the less favored codons is observed in the ADH-1 gene (14). These less preferred codons in the PGK gene probably do not grossly affect or restrict higher expression levels,



Translated Mol. Weight = 44746

Figure 2. DNA and protein sequences of the yeast PGK gene. The sense strand of DNA sequence from PvuI to HindIII is shown from 5' to 3'. The translated amino acid sequence is shown with amino acids underlined from 268 through 398 which agree with published protein sequence. Amino acids in parentheses designate differences in published protein sequence (47). Asterisks designate the use of nonpreferred codons for highly expressed yeast genes. The transcription start is shown by the arrow and the polyadenylation region by the bar region at the end of the gene.

since insertion of the gene on a yeast multicopy plasmid leads to very elevated PGK expression levels (see below).

However, the codon usage in the PGK gene is not consistent with one of the codon bias rules of Bennetzen and Hall (48), since one of the seven codons that are never used in the genes which they compared is used in PGK. This exception is the CGU codon for arginine which is used three times (positions 18, 122 and 182; Fig. 2). The CGU codon for arginine is also used in the yeast enolase genes (49). Another one of these seven codons, the CCG codon for proline, has been observed in the yeast TRP5 gene (50).

Homology with the human and equine PGKs

The protein sequence for two other eukaryotic PGKs, the human (45) and horse (46) enzymes, has been elucidated. However, no corresponding DNA sequence is available. The horse and human PGK have a remarkable homology with only 14 amino acid differences, which probably resulted from single base changes in the gene. Another difference in the human PGK was the insertion of an extra lysine between amino acids 38 and 39 of the horse sequence. This results in a length of 416 amino acids for horse PGK versus a length of 417 amino acids for human PGK.

The yeast PGK protein sequence, which contains 415 amino acids (assuming the removal of the NH_2-terminal methionine), is compared with the human and equine PGKs in Fig. 4. The sequences are aligned for maximal homology.

Codon Usage:

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

Figure 3. Codon usage for the PGK gene. Amino acids for each codon are abbreviated with the numbers indicating frequency of use in the gene. OC, AM, and OP refer to the translational stops ochre, amber, and opal, respectively.

One difference is the presence of an additional amino acid at the COOH-terminus of yeast PGK. In support of this sequence, Markland et al. (43) have reported that the COOH-terminal amino acid for yeast PGK is lysine

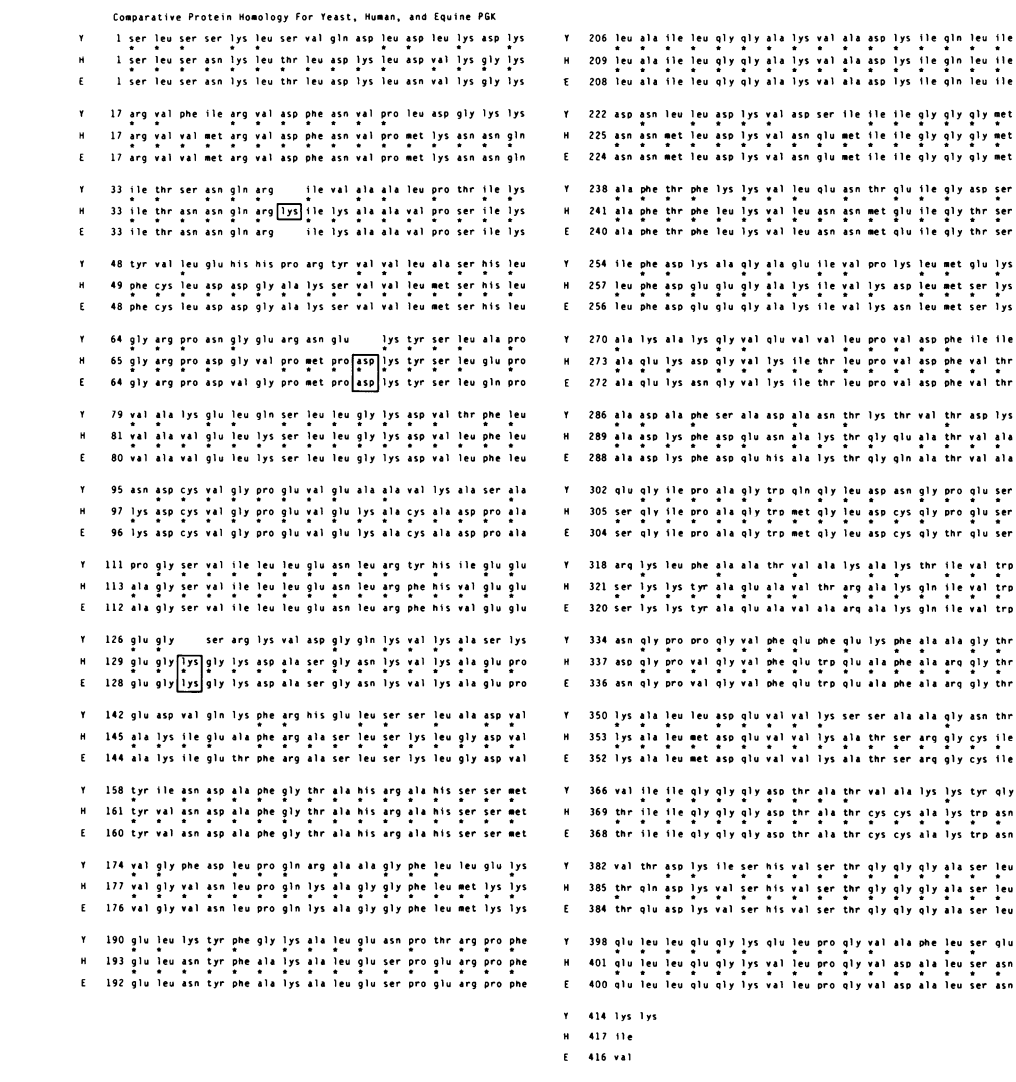


Figure 4. Comparative protein homologies for yeast, human, and equine PGKs. Y, H, and E refer to yeast, human, and equine (horse), respectively. Sequences are lined up for greatest homology with squares showing the insertion of an amino acid in one sequence with respect to another. However, the insertions with respect to yeast sequence occur within regions 69 to 72 and 128 to 132 of yeast sequence; thus, the amino acid insertion shown is arbitrary. Asterisks indicate homology.

(see Fig. 2). Other major changes are the insertion of an amino acid in the human and horse sequence in the region of positions 69 to 72 and 128 to 132 of the yeast PGK gene. An additional amino acid insertion in human PGK with respect to yeast PGK occurs between residues 38 and 39 of the yeast sequence. Interestingly, this is identical to the lysine insertion in human PGK with respect to horse PGK. The yeast PGK has about 65 percent homology with both the human and horse sequence. Although this homology is spread all over the sequence, it is clear that some regions, e.g. residues 160 to 175 and 203 to 221, are strongly conserved. This might implicate an important role of such segments for the enzymatic function.

The PGK mRNA

Polyadenylated or total RNA from S. cerevisiae GM3C-2 containing either YEp13 or PGK-YEp13 was sized on a formaldehyde-agarose gel and "Northern" hybridization (36) was performed using a PGK-gene specific probe. YEp13 (21) is a high copy number, 2μ origin based plasmid; while PGK-YEp13 contains the PGK HindIII fragment between the two HindIII sites of YEp13. It should further be noted that these haploid yeast contain one copy of this PGK HindIII fragment in chromosome III (4) which is also producing mRNA. As shown in Fig. 5 a single mRNA band of about 1500 bases long was observed. The size of the PGK mRNA from PGK-YEp13/GM3C-2 is identical to the chromosomal PGK mRNA, isolated from YEp13/GM3C-2. The intensity of the former PGK mRNA is obviously much stronger (10-20X), since the gene and the flanking regions are situated on a high copy number plasmid. These levels

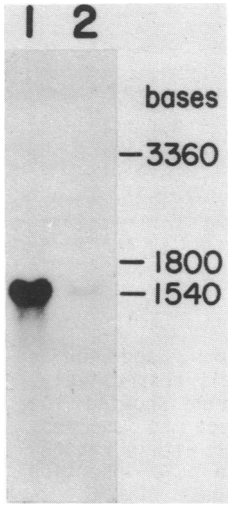


Figure 5. PGK gene mRNA identification. Lane 1 is total RNA from yeast strain GM3C-2 containing plasmid PGK-YEp13. Lane 2 is total RNA from yeast strain GM3C-2 with plasmid YEp13 (no PGK gene present on the second plasmid). The mRNAs with PGK sequence were visualized by hybridization with ^{32}P-labeled EcoRI/BglII fragment from the PGK gene (see Fig. 1) followed by autoradiography. Size standards are designated (see Methods).

of mRNA correspond well with PGK-YEp13 expression of PGK at about 20 percent of the total protein as compared to 1 percent of the total protein by the chromosomal copy of the gene in the other yeast strain (data not shown). These results strongly suggest that all control signals needed for transcription of the PGK gene are contained within the 3.1 kb HindIII fragment, situated on PGK-YEp13.

Transcriptional initiation

The start of transcription of the PGK gene was determined by specifically primed cDNA synthesis on the mRNA from yeast pFRM31/20B-12. Plasmid pFRM31 contains the 3.1 kbp PGK HindIII fragment in the HindIII site of the pBR322 portion of pFRD7 (see Methods), which contains the yeast ars1 chromosomal origin of replication and the yeast TRP1 gene (22, 23). The extension started from an oligonucleotide complementary to position -10 to -21 (Fig. 2) and gave a distinct product of 26 base pairs long (Fig. 6). The initiation of transcription can therefore be located at -36 (Fig. 2). It is possible that the observed 5' end is the result of post-transcriptional processing (15) and that the transcription start is further upstream. No evidence for such 5' terminal mRNA processing in yeast is available. This initiation of transcription takes place at an AG-duplet, which is also observed for the two transcripts of yeast ADH-1 (14). This seems, however,

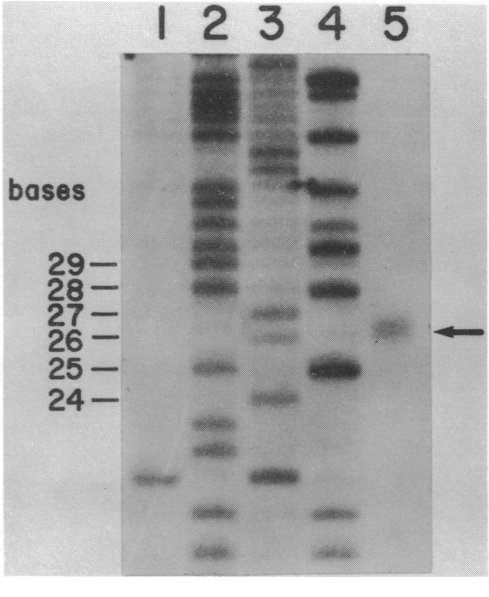


Figure 6. Determination of the transcription initiation point of the PGK gene. A discrete product was obtained from specifically primed cDNA synthesis. This extension product, shown with an arrow in lane 5, is sized on a 15 percent polyacrylamide-7M urea gel alongside an unrelated sequencing ladder, obtained by the chemical degradation method (lanes 1-4). The length of the sequencing reaction products is shown on the left.

not to be the case in other genes (11,50). Only one discrete cDNA extension product is observed, implicating a single transcription start. This is in clear contrast to an apparent multiplicity of initiation points for several other yeast genes such as TRP5 (50), ADH-1 (14) and CYC1 (11). The mechanism behind this phenomenon is unclear.

Transcription termination and polyadenylation site

The 286 bp region following the stop codon of the yeast PGK gene has also been sequenced (Fig. 2). We used the S1-mapping procedure (40, 41) to determine the polyadenylation site of the mRNA. The BglII-HindIII PGK-terminator fragment, 3' ^{32}P-labelled at the BglII site served as hybridization probe to mRNA from pFRM31/20B-12. As shown in Fig. 7, there is a heterogeneity in the length of the DNA fragments, protected from S1-nuclease activity, ranging from 180 to 187 nucleotides. This implies that the polyadenylation site of the PGK mRNA is located at 86 to 93 nucleotides behind the stop codon (Fig. 2). The heterogeneity in the length of the protected DNA fragments might be due to an overdigestion by S1-nuclease or may result from a slight heterogeneity in the polyadenylation site. Multiple polyadenylation sites have been seen for several other eukaryotic mRNAs, including the yeast ADH-1 mRNA (14) and the yeast actin mRNA (52). Again steady state levels of mRNA were used so it cannot be concluded that this defines the actual transcription termination site for PGK mRNA. Transcription may proceed further followed by processing before polyadenylation.

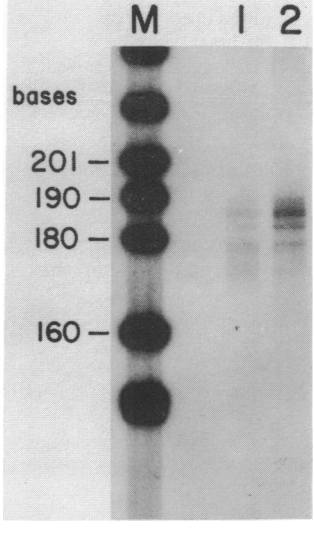


Figure 7. S1-nuclease mapping of the polyadenylation site of the PGK mRNA. The RNA-DNA hybrids were treated with 50 units S1 nuclease for 30 min (lane 1) or 100 units S1 nuclease for 1 hr (lane 2) and then sized on a 6 percent polyacrylamide-7M urea gel. The lane M shows the ^{32}P-labelled HpaII fragments of pBR322 with their sizes indicated on the left.

DISCUSSION

PGK coding region and protein sequence

The PGK coding sequence is 1248 bp corresponding to 416 amino acids. Like other highly expressed genes from yeast, this gene is highly codon biased using only 25 codons out of 61 for 95 percent of the amino acids. Some less biased codons are also used; however, this does not appear to limit expression, since the PGK gene expresses up to 20 percent of the total protein when present on a high copy number plasmid. The amino acid sequence is identical with 120 out of 130 previously published residues (47).

The entire amino acid sequence of yeast PGK, as determined from DNA sequence, is 65 percent homologous with horse and human PGKs. Previous x-ray crystallographic data demonstrate two major domains in both the horse (46) and yeast (53, 54) enzymes. Furthermore, such data suggest that the mechanism of PGK enzyme activity is the contact between these two domains to bring the domains binding the two substrates, ATP and 3-phosphoglycerate, together (46). For horse PGK, ATP binds to the carboxyl domain (residues 190 to 403) and 3-phosphoglycerate binds to the amino terminal domain (residues 1-185) while the rest of the carboxyl end interacts with the amino terminal domain. Two links between these domains are residues 186-189 and residues 404-406. Yeast PGK appears to be very similar. Our protein sequence shows the greatest homology from residues 133 through 413 encompassing the two links between domains. The most severe differences are codon insertions in human and horse PGK sequences from residues 1 through 128 with respect to the yeast sequence. Twenty-four of 25 amino acids which are thought to be involved in substrate binding in the horse enzyme (46) are identical in the yeast enzyme. Homology between the yeast and horse enzymes both coincides and falls in between α-helical and β-sheet structures in horse PGK (46); suggesting regions between these structural components are as important to enzyme function as the defined structural components themselves.

PGK gene 5'-flanking sequence

The Goldberg-Hogness box or TATAAA-consensus sequence is generally found 25 to 30 nucleotides upstream from the RNA start in higher eukaryotes or eukaryotic viruses and seems to be important in the positioning of the transcription start (for review see refs. 55 and 56). A TATAAA-sequence is here found 145 bp in front of the PGK-gene translational start (Fig. 2) (109 basepairs in front of the mRNA start at -36). However, although a TATAAA-sequence is present in front of the transcribed sequence of yeast genes, the

distance from the transcriptional initiation is longer and more variable than in higher eukaryotes (11,14,50,52). No unambiguous evidence for the functional importance of this sequence in yeast is as yet available.

In between the TATAAA-sequence and the transcription starts of the ADH-1 gene is a very pyrimidine-rich cluster (14) and it has been proposed that this might serve as a signal for high level transcription (9). A similar C-T rich stretch is present in the 5' flanking region of the PGK gene at -70 to -52. This feature cannot, however, be clearly distinguished in front of the yeast genes for the other glycolytic enzymes, enolase (49) and glyceraldehyde-3-phosphate dehydrogenase (3).

Another potentially important sequence is located at -48 to -28. As noticed by Dobson et al. (47), who recently reported the 5' flanking sequence of the PGK gene (largely corresponding to our sequence) the above-mentioned region has a similarity with the corresponding sequence for the enolase (49) and glyceraldehyde-3-phosphate dehydrogenase (3) genes. Since there is, however, no clear homology with the ADH-1 gene (14), it is doubtful if this is a typical feature of genes for the glycolytic enzymes in yeast.

The sequence 5'-PuCACACA-3' precedes by 4 to 15 residues the initiation codon of several yeast genes like CYC1 (8), histone H2B1 (57), glyceraldehyde-3-phosphate dehydrogenase (3), enolase (49), TRP1 (23) and TRP5 (50). Although there is no evidence for a specific function, it has been suggested that this sequence might play a role in the initiation of translation (50). A variation of such a sequence is present in front of the initiation codon of the ADH-1 gene as 5'-PuCAATCAA-3' (-15 to -22) (14). A similar, but not identical, sequence (5'-PuCAACAA-3') is present 10 residues prior to the PGK gene (Fig. 2). However, the expression of human leukocyte interferon D cDNA in a hybrid yeast expression system demonstrates that up to 33 nucleotides of 5'-sequence adjacent (from yeast alcohol dehydrogenase I gene) to the ATG can be deleted without eliminating expression; suggesting that this sequence containing similarities among different yeast genes is not critical to expression (66).

A current widely accepted view of eukaryotic translation initiation is that the 40S ribosomal subunit binds at or near the 5' end of the mRNA and moves along the mRNA until it encounters the first AUG (58-60). The efficiency of ribosome binding might be enhanced by a complementarity with the 3' end of the 18S ribosomal RNA. Such a complementarity has been noticed for several eukaryotic mRNAs (61, 62). Similarly, the sequence 5' TAATTATC 3' at -34 to -27, just behind the transcription initiation site

of the yeast PGK gene, has a clear complementarity to the 3' end of the yeast 18S ribosomal RNA 3' AUUACUAG 5' (63). A similar complementarity has been observed for several other yeast mRNAs, as for TRP5, by Zalkin and Yanofsky (50).

The efficiency of the ribosome binding to the mRNA is likely to be influenced by the sequence immediately preceding the ATG-codon as proposed by Kozak (60). From in vitro ribosome binding studies (60) and from in vivo studies with various sequences in front of the start codon (R.D., manuscript in preparation), it can be concluded that the position of an A at -3 is very favorable for translation initiation. This feature is observed in the yeast PGK mRNA and similarly in almost all other known yeast mRNAs (64). One exception to this apparent rule for yeast is a revertant of iso-1-cytochrome c which has a T at -3 and expresses the same level of enzyme as the wild type gene with an A at -3 (65). Furthermore, the relatively high expression of leukocyte interferon D, by a constructed hybrid expression system with a T residue at -3, also suggests that the absence of A at -3 might not grossly affect translation (66).

PGK gene 3'-flanking sequence

On the basis of the determined transcription start and of the location of the polyadenylation site (86 to 93 nucleotides from the translation stop), it can be concluded that the size of the PGK transcript is about 1380 bases. This is compatible with the observed length of 1500 nucleotides of the polyadenylated PGK mRNA, since polyadenylation adds about 50-100 bases to the length (67). The mRNA size, considered in combination with the size of the PGK coding sequence and protein, suggests that there are no introns within the PGK coding region which is consistent with other yeast glycolytic genes (3,14,49).

The polyadenylation site of the mRNAs of higher eukaryotes is preceded 20 to 26 nucleotides by a 5' AAUAAA 3' (68). Although some yeast mRNAs (7,17,52,69) also contain this sequence, most others (17) clearly lack this signal sequence. Bennetzen and Hall (48) recently proposed that a related consensus sequence 5' TAAATAA$^{A}_{G}$ 3' might play a role in the transcription termination in yeast. This sequence or a variant of it precedes the polyadenylation site by 28 to 34 nucleotides. However the comparison of the 3' terminal sequences of yeast mRNAs by these authors shows that a relatively wide variation is present, which makes its significance less convincing. A possibly related sequence in PGK is present at 64 to 71 nucleotides behind the stop codon.

Alternatively, Zaret and Sherman (17) proposed a sequence 5' TAG...TA(T)GT...TTT 3' as being important for transcription termination. This sequence or a variant precedes the polyadenylation site of most yeast mRNAs at variable distance. At 58 to 88 nucleotides behind the stop codon, a related sequence can be recognized.

Much attention is being focused on the regulation of transcription and expression in yeast. As illustrated above with the determined structure of the yeast PGK gene and mRNA, some homologies and relations in the sequences upstream and downstream of the coding regions can be observed. However, the expected functional importance of these sequences is mainly based on structural comparisons and speculations. It is obvious that much more research is needed, which will undoubtedly lead to more insight into the structural requirements for these processes in yeast.

ACKNOWLEDGMENTS

The authors wish to thank Mark Vasser for the synthesis of the oligonucleotide primer which is complementary to a portion of the PGK gene. We also wish to thank Dr. John Carbon for kindly making us aware of a C deletion (at codon 407) in our DNA sequence prior to publication.

REFERENCES

1. Scopes, R.K. (1973) In The Enzymes, 8, 3rd Ed., P.D. Boyer, ed., pp. 335-351, Academic Press, New York.
2. Hommes, F.A. (1966) Arch. Biochem. Biophys. 114, 231-233.
3. Holland, J.P., and Holland, M.J. (1980) J. Biol. Chem. 255, 2596-2605.
4. Hitzeman, R.A., Chinault, A.C., Kingman, A.J., and Carbon, J.A. (1979) in ICN-UCLA Symposium on Molecular and Cellular Biology, Maniatis, T. and Fox, C.F. Eds., Vol. 14, pp. 57-68, Academic Press, New York.
5. Hitzeman, R.A., Clarke, L., and Carbon, J. (1980) J. Biol. Chem. 255, 12073.
6. Holland, M.J., and Holland, J.P. (1978) Biochemistry 17, 4900-4907.
7. Holland, J.P., and Holland, M.J. (1979) J. Biol. Chem. 254, 9839-9845.
8. Smith, M., Leung, D.W., Gillam, S., Astell, C.R., Montgomery, D.L., and Hall, B.D. (1979) Cell 16, 753-761.
9. Montgomery, D.L., Leung, D.W., Smith, M., Shalit, P., Faye, G., and Hall, B.D. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 541-545.
10. Gallwitz, D., and Sures, I. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 2546-2550.
11. Faye, G., Leung, D.W., Tatchell, K., Hall, B.D., and Smith, M. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 2258-2262.
12. Grosschedl, R., and Birnstiel, M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1432-1436.
13. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
14. Bennetzen, J.L., and Hall, B.D. (1982) J. Biol. Chem. 257, 3018-3025.
15. Ziff, E.B., and Evans, R.M. (1978) Cell 15, 1463-1475.
16. Carlson, M., and Botstein, D. (1982) Cell 28, 145-154.

17. Zaret, K.S., and Sherman, F. (1982) Cell 28, 563-573.
18. Crea, R., Kraszewski, A., Hirose, T., and Itakura, K. (1978) Proc. Natl. Acad. Sci. USA 75, 5765-5769.
19. Bachman, K., Ptashne, M., and Gilbert, W. (1976) Proc. Natl. Acad. Sci. USA 73, 4174-4178.
20. Jones, E. (1976) Genetics 85, 23-33.
21. Broach, J.R., Strathern, J.N., and Hicks, J.B. (1979) Gene 8, 121-133.
22. Stinchcomb, D.T., Struhl, K., and Davis, R.W. (1979). Nature 282, 39-43.
23. Tschumper, G., and Carbon, J. (1980) Gene 10, 157-166.
24. Miller, J.H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
25. Clewell, D.B. (1972) J. Bacteriol. 110, 667-676.
26. Hershfield, V., Boyer, H.W., Yanofsky, C., Lovett, M.A., and Helsinki, D.R. (1976) Proc. Natl. Acad. Sci. USA 71, 3455-3459.
27. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
28. Hinnen, A., Hicks, J.B., and Fink, F.R. (1978) Proc. Natl. Acad. Sci. USA 75, 1929-1933.
29. Maxam, A.M., and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
30. Smith, A.J.H. (1980) Meth. Enzymol. 65, 499-560.
31. Messing, J., Crea, R., and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
32. Zitomer, R.S., and Hall, B.D. (1976) J. Biol. Chem. 251, 6320-6326.
33. Dobner, P.R., Kawasaki, E.S., Yu, L.Y., and Bancroft, F.C. (1981) Proc. Natl. Acad. Sci. USA 78, 2230-2234.
34. Warner, J.R. (1981) In The Molecular Biology of the Yeast Saccharomyces, J.N. Strathern, E.W. Jones, and J.R. Broach, eds., Vol. 2, in press, Cold Spring Harbor Laboratory, New York.
35. Szybalski, E.H., and Szybalski, W. (1979) Gene 7, 217-270.
36. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5201-5205.
37. Taylor, J.M., Illmensee, R., and Summers, S. (1976) Biochim. Biophys. Acta 442, 324-330.
38. Derynck, R., Leung, D.W., Gray, P.W., and Goeddel, D.V. (1982) Nucleic Acids Res. 10, 3605-3615.
39. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1408-1412.
40. Berk, A.J., and Sharp, P.A. (1978) Proc. Natl. Acad. Sci. USA 75, 1274-1278.
41. Mantei, N., Schwarzstein, M., Streuli, M., Panem, S., Nagata, S., and Weissmann, C. (1980) Gene 10, 1-10.
42. Sutcliffe, J.G. (1978) Nucleic Acids Res. 5, 2721-2728.
43. Markland, F.S., Bacharach, A.D.E., Weber, B.H., O'Grady, T.C., Saunders, G.C., and Umemura, N. (1975) J. Biol. Chem. 250, 1301-1310.
44. Edman, P., and Begg, G. (1967) Eur. J. Biochem. 1, 80-91.
45. Huang, I., Welch, C.D., and Yoshida, A. (1980) J. Biol. Chem. 255, 6412-6420.
46. 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.
47. Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J., Kingsman, S.M., Perkins, R.E., Canzoy, S.C., Dunbar, B. and Fothergill, L.A. (1982) Nucleic Acids Res. 10, 2625-2637.
48. Bennetzen, J.L., and Hall, B.D. (1982) J. Biol. Chem. 257, 3026-3031.
49. Holland, M.J., Holland, J.P., Thill, G.P., and Jackson, R.A. (1981) J. Biol. Chem. 256, 1385-1395.
50. Zalkin, H., and Yanofsky, C. (1982) J. Biol. Chem. 257, 1491-1500.

51. Levy, W.P., Rubinstein, M., Shively, J., Del Valle, U., Lai, C-Y., Moschera, J., Brink, L., Gerber, L., Stein, S.and Pestka, S. (1981) Proc. Natl. Acad. Sci. USA 78, 6186-6190.
52. Gallwitz, D., Purin, F. and Seidel, R. Nucleic Acids Res. 9, 6339-6350 (1981).
53. Wendell, P.L., Bryant, T.N., and Watson, H.C. (1972) Nature New Biol. 240, 134-138.
54. Bryant, T.N., Watson, H.C., and Wendell, P.L. (1974) Nature 247, 14-17.
55. Braetnach, R., and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
56. Darnell, J.E. Jr. (1982) Nature 297, 365-371.
57. Wallis, J.W., Hereford, L., and Grunstein, M. (1980) Cell 22, 799-805.
58. Kozak, M. (1978) Cell 15, 1109-1123.
59. Kozak, M. (1981) Current Topics in Microbiology and Immunology 93, 81-123.
60. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
61. Shine, J., and Dalgarno, L. (1979) Biochem. J. 141, 609-615.
62. Hagenbuchle, O., Santer, M., Argetsinger-Steitz, J., and Mans, R.J. (1978) Cell 13, 551-563.
63. Rubstov, P.M., Mushkhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G., and Bayev, A.A. (1980) Nucleic Acids Res. 8, 5779-5794.
64. Ammerer, G., Hitzeman, R., Hagie, F., Barta, A., and Hall, B.D. (1981) In Recombinant DNA, Proceedings of the Third Cleveland Symposium on Macromolecules, A.G. Walton, ed., pp. 185-197, Amsterdam, Netherlands, Elsevier Scientific Publishing Co.
65. Sherman, F., Stewart, J.W., and Schweingruber, A.M. (1980) Cell 20, 215-222.
66. Hitzeman, R.A., Hagie, F.E., Levine, H.L., Goeddel, D.V., Ammerer, G., and Hall, B.D. (1981) Nature 293, 717-722.
67. McLaughlin, C.S., Warner, J.R., Edmonds, M., Nakazato, H., and Vaughan, M.H. (1973) J. Biol. Chem. 248, 1466-1471.
68. Proudfoot, N.J., Chang, C.C., and Brownlee, G.G. (1976) Progr. Nucl. Acid Res. Mol. Biol. 19, 123-134.
69. Astell, C.R., Ahlstrom-Jonasson, L., Smith, M., Tatchell, K., Nasmyth, K.A., and Hall, B.D. (1981) Cell 27, 15-23.