THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 258, No. 8, Issue of April 25, pp. 5291-5299, 1983
Printed in U.S.A.

Homologous Nucleotide Sequences at the 5′ Termini of Messenger RNAs Synthesized from the Yeast Enolase and Glyceraldehyde-3-phosphate Dehydrogenase Gene Families

THE PRIMARY STRUCTURE OF A THIRD YEAST GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE GENE*

(Received for publication, May 3, 1982)

Janice P. Holland, Laura Labieniec, Candace Swimmer‡, and Michael J. Holland§

From the Department of Biological Chemistry, School of Medicine, University of California, Davis, California 95616

Genomic DNA containing a third yeast glyceraldehyde-3-phosphate dehydrogenase structural gene has been isolated on a bacterial plasmid designated pgap11. The complete nucleotide sequence of this structural gene was determined. The gene contains no intervening sequences, codon usage is highly biased, and the nucleotide sequence of the coding portion of this gene is 90% homologous to the other two glyceraldehyde-3-phosphate dehydrogenase genes (Holland, J. P., and Holland, M. J. (1980) *J. Biol. Chem.* 255, 2596-2605). Based on the extent of nucleotide sequence divergence among the three glyceraldehyde-3-phosphate dehydrogenase genes, it is likely that they arose as a consequence of two duplication events and the gene contained on the hybrid plasmid designated pgap11 is a product of the first duplication event. All three structural genes share extensive nucleotide sequence homology in the 5′-noncoding regions adjacent to the three respective translational initiation codons. The gene contained on pgap11 is not homologous to the others downstream from the respective translational termination codon, however. The 5′ termini of messenger RNAs synthesized from the three glyceraldehyde-3-phosphate dehydrogenase and two yeast enolase genes have been mapped to sites ranging from 36 to 82 nucleotides upstream from the respective translational initiation codons. In each case the 5′ terminus of the mRNA maps to a region of strong nucleotide sequence homology which is shared by all five structural genes. These latter data confirm that all five structural genes are expressed during vegetative cell growth and further support the hypothesis that a portion of the 5′-noncoding flanking region of the yeast glyceraldehyde-3-phosphate dehydrogenase and enolase genes evolved from a common precursor sequence.

The existence of three yeast glyceraldehyde-3-phosphate dehydrogenase structural genes was first suggested from DNA filter-blotting experiments using restriction endonuclease-cleaved genomic DNA and hybridization probes which are complementary to glyceraldehyde-3-phosphate dehydrogenase mRNA sequences (1, 2). Two segments of genomic DNA were subsequently isolated on bacterial plasmids and shown by nucleotide sequencing to contain structural genes which are 95% homologous (3, 4). Here we report the isolation and nucleotide sequence of the third yeast glyceraldehyde-3-phosphate dehydrogenase structural gene.

Glyceraldehyde-3-phosphate dehydrogenase is expressed at high levels in yeast (5, 6) and it is of interest to know if all three structural genes are expressed during vegetative cell growth. There have been numerous reports of multiple forms of the enzyme isolated from yeast but it is unclear if these represent isozymes or forms which have been modified during isolation. The most compelling evidence that at least two of the genes are expressed is based on the fact that a limited number of ambiguities in the primary structure determined for the yeast enzyme (7) can be reconciled if the sequenced protein was a mixture of at least two of the polypeptides encoded by the isolated genes (4). Evidence is presented here that mRNA is synthesized *in vivo* from all three structural genes during vegetative cell growth.

The 5′ termini of the three glyceraldehyde-3-phosphate dehydrogenase mRNAs and the mRNAs synthesized from the two yeast enolase genes (8) have been identified. In all cases the 5′ termini map to a region of nucleotide sequence homology which is shared by all five structural genes. These data support our hypothesis that the 5′-noncoding regions of the glyceraldehyde-3-phosphate dehydrogenase and enolase genes are structurally related (8).

EXPERIMENTAL PROCEDURES[1]

RESULTS

Isolation of pgap11—There are three segments of yeast genomic DNA which are complementary to cDNA hybridization probes synthesized from purified yeast glyceraldehyde-3-phosphate dehydrogenase mRNA (1, 2). Two of these regions of genomic DNA were subsequently isolated and shown to contain glyceraldehyde-3-phosphate dehydrogenase structural genes (3, 4). When yeast genomic DNA is analyzed by

* This research was supported by United States Public Health Service Grant GM 30307 and a grant from the March of Dimes Birth Defects Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address, Department of Microbiology, Health Science Center, State University of New York, Stony Brook, NY 11794.

§ American Heart Association Established Investigator.

[1] Portions of this paper (including "Experimental Procedures," Table IS, and additional references) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 82M-1143, cite authors, and include a check or money order for $2.00 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

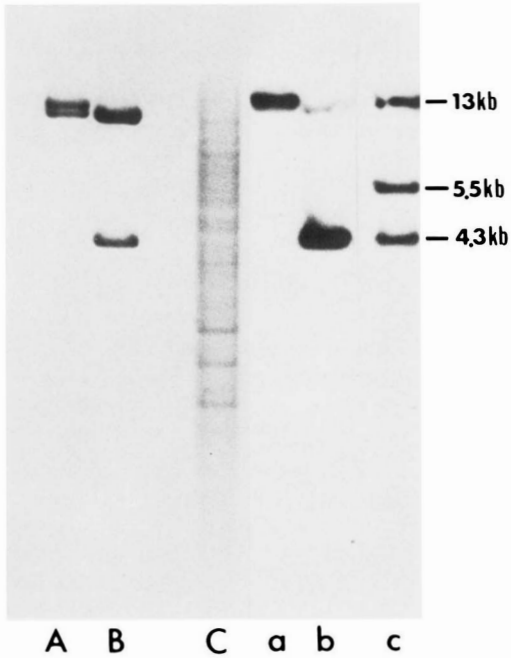


FIG. 1. **DNA filter blot of *Eco*RI-digested p*gap*492, p*gap*11, and yeast DNA.** *Lanes A, B,* and *C* are agarose gels of *Eco*RI-digested p*gap*492, p*gap*11, and yeast DNA (strain F_1), visualized after staining with ethidium bromide. *Lanes a, b,* and *c* are autoradiograms of hybrids formed between the ^{32}P-labeled 2.1-kb *Hin*dIII fragment from p*gap*491 and the DNA shown in *Lanes A, B,* and *C,* respectively, after transfer to nitrocellulose.

DNA filter blotting after *Eco*RI digestion, 13.2-, 5.5-, and 4.3-kb[2] fragments form hybrids with a nick-translated hybridization probe synthesized from the 2.1-kb *Hin*dIII fragment which contains the glyceraldehyde-3-phosphate dehydrogenase gene within the plasmid p*gap*491 (3). The 13.2-kb *Eco*RI fragment has been isolated on a plasmid designated p*gap*492 (2, 4) and corresponds to the structural gene already isolated on the plasmid designated p*gap*49 (1). The 5.5-kb *Eco*RI fragment corresponds to the structural gene previously isolated on the plasmid designated p*gap*63 (4).

The 4.3-kb *Eco*RI fragment was cloned as previously described for the isolation of p*gap*492 (2). *Eco*RI-digested yeast DNA was ligated into the *Eco*RI site of the vector pSF2124. Total plasmid DNA was isolated from the shotgun collection obtained and supercoiled plasmids corresponding in size to pSF2124 plus a 4.5-kb insertion of yeast DNA was isolated by preparative agarose gel electrophoresis. After retransformation of competent *Escherichia coli* with this enriched fraction of plasmids, the collection was screened by colony hybridization using a nick-translated probe synthesized from the 2.1-kb *Hin*dIII fragment isolated from p*gap*491. A hybrid plasmid designated p*gap*11 was identified and analyzed as described below.

Restriction Endonuclease Mapping of pgap11: Comparison with pgap492 and pgap63—In order to confirm that the sequences isolated on p*gap*11 are co-linear with genomic DNA sequences, DNA filter blotting was carried out in parallel with the isolated plasmid and yeast genomic DNA. Fig. 1 illustrates a Southern blot of *Eco*RI-digested p*gap*492 (2, 4), p*gap*11, and yeast DNA. A nick-translated 2.1-kb *Hin*dIII fragment containing the glyceraldehyde-3-phosphate dehydrogenase gene within p*gap*492 was used as the hybridization probe. The 13.2-kb *Eco*RI fragment from p*gap*492 corresponds to the largest hybrid formed with genomic DNA. The 4.3-kb fragment from p*gap*11 corresponds to the smallest hybrid formed in genomic DNA. These data confirm that p*gap*11 contains a segment of DNA which is contiguous with corresponding genomic sequences.

Similar blotting experiments were carried out with genomic DNA isolated from several haploid strains of *Saccharomyces cerevisiae* (data not shown). In some strains, we observed polymorphisms with respect to the location of *Eco*RI and *Hin*dIII restriction endonuclease cleavage sites adjacent to the glyceraldehyde-3-phosphate dehydrogenase genes. Comparison of these blots with those generated with DNA isolated from strain F1 demonstrated that there are three glyceraldehyde-3-phosphate dehydrogenase structural genes per haploid genome in all the strains tested and the three genes corresponded in all cases to those isolated on p*gap*492 (2, 4), p*gap*63 (4), and p*gap*11.

A restriction endonuclease cleavage map of p*gap*11, illustrated in Fig. 2, was generated as described under "Experimental Procedures." The location of the coding sequences in the plasmid and the direction of transcription were determined from the nucleotide sequence described below and are indicated in Fig. 2 by the *shaded region* and *arrow*, respectively. The *Hpa*I cleavage site within the coding portion of the gene in p*gap*11 is homologous to *Hpa*I cleavage sites in the genes contained within p*gap*492 and p*gap*63 (4). The restriction endonuclease cleavage map of p*gap*11 adjacent to the coding sequences is not homologous to the other two structural genes. These data confirm that none of the yesat glyceraldehyde-3-phosphate dehydrogenase genes are tandemly repeated in the yeast genome.

A more detailed restriction endonuclease map of the coding and adjacent noncoding sequences of the glyceraldehyde-3-phosphate dehydrogenase gene within p*gap*11 is shown in Fig. 3. The restriction endonuclease cleavage maps of the corresponding regions of p*gap*491 and p*gap*63 (4) are also shown for comparison. Within the coding regions of the genes, approximately 65% of the restriction endonuclease cleavage sites are present in all three structural genes. These data demonstrate, together with the primary structure described in the following section, that the three structural genes are very homologous. No homology among the genes is observed for restriction endonuclease cleavage sites which map outside of the respective coding sequences.

Primary Structure of the Glyceraldehyde-3-phosphate Dehydrogenase Gene in pgap11—The complete nucleotide sequence of the coding region of the glyceraldehyde-3-phosphate dehydrogenase gene contained within p*gap*11 is shown in Fig.

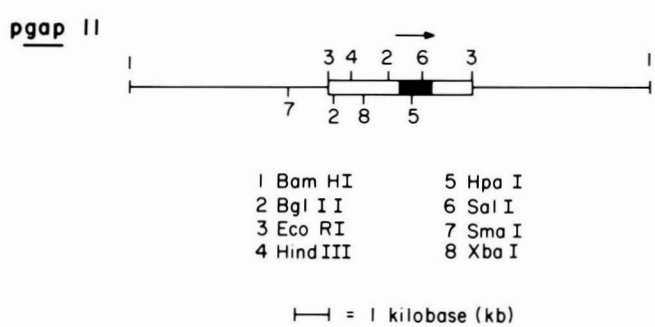


FIG. 2. **Restriction endonuclease cleavage map of p*gap*11.** The *single horizontal lines* indicate the pSF2124 vector portion of the plasmid and the *double horizontal line* indicates the yeast DNA portion of the plasmid. The *shaded region* indicates the location of the coding portion of the glyceraldehyde-3-phosphate dehydrogenase gene in p*gap*11. The *arrow* shows the direction of transcription of the gene. The cleavage map was derived from the data reported in the Miniprint.

[2] The abbreviations used is: kb, kilobase pair.

4 as are the sequences for the genes in p*gap*491 (3) and p*gap*63 (4). The primary structure was determined using the sequencing strategy outlined in Fig. 5. The gene in p*gap*11 contains a single open reading frame without intervening sequences. The primary structure of the polypeptide predicted from this gene is highly homologous to the primary structure determined for yeast glyceraldehyde-3-phosphate dehydrogenase (7), confirming that this plasmid does contain a third structural gene. The primary structure of the polypeptide encoded by the gene within p*gap*11 is 89 and 88% homologous to the polypeptides encoded by the genes within p*gap*63 and p*gap*491, respectively. Based on these data, it is possible that the gene within p*gap*11 was derived from an early duplication event and that the genes contained on p*gap*63 and p*gap*491 arose from a subsequent duplication. Alternatively, the similarity between the genes on p*gap*63 and p*gap*491 could be due to preferential gene conversion between these two structural genes rather than a later duplication event. The possibility of gene conversion among the three glyceraldehyde-3-phosphate dehydrogenase genes is considered under "Discussion."

Codon usage within the gene contained within g*gap*11 follows the same highly biased pattern previously described for those within p*gap*63 and p*gap*491 (3, 4). In the cases of alanine, aspartic acid, isoleucine, serine, threonine, and valine, two codons are used exclusively which contain either C or U in the third position. The remaining 14 amino acids are encoded by a single codon in approximately 98% of the cases. For the six amino acids which are encoded by two codons, alanine and aspartic acid are biased for GCU (81%) and GAC (68%), respectively. The codons for isoleucine, serine, threonine, and valine contain either C or U in the third position approximately 50% of the time. There are 84 positions within the three glyceraldehyde-3-phosphate dehydrogenase structural genes at which all three genes contain an isoleucine, serine, threonine, or valine codon. In 56 of these cases, the third positions of the respective codons are identical. In the remaining 28 cases, the nonidentical third position nucleotide occurred in p*gap*491, p*gap*63, or p*gap*11 in 46, 25, and 29% of the cases, respectively. The distribution of third position changes among the three structural genes will be discussed below.

Within the portions of the three structural genes which encode the NAD-binding domains of the polypeptides (residues 1–147), the gene contained within p*gap*11 differs from those contained on p*gap*491 and p*gap*63 at 69 and 63 nucleotide positions, respectively. Within this same region, the genes contained on p*gap*491 and p*gap*63 differ by 30 nucleotides. Within the portions of the genes which encode the catalytic domains of the polypeptides (residues 148–331), the gene contained on p*gap*11 differs from those contained on p*gap*491 and p*gap*63 by 40 and 38 nucleotides, respectively. Within this region, the genes within p*gap*491 and p*gap*63 differ by 23 nucleotides. These data demonstrate that the observed rate of divergence within the portions of the genes which encode the NAD-binding domains of the polypeptides is higher than that observed within the catalytic domain.

The putative location of amino acid residues which differ among the three yeast glyceraldehyde-3-phosphate dehydrogenase polypeptides in the native enzyme has been ascertained by comparison with the three-dimensional structure determined for lobster glyceraldehyde-3-phosphate dehydrogenase (9, 10). The majority of the amino acids which differ among the three polypeptides are predicted to reside at external regions of the enzyme. Amino acid differences at positions predicted to be within internal regions of the enzyme are chemically conservative amino acid changes. There are a limited number of amino acid substitutions which might alter the structure of the tetrameric enzyme; however, no striking structural alterations are predicted from this analysis.

*The Primary Structure of the 5′- and 3′-noncoding Regions of the Glyceraldehyde-3-phosphate Dehydrogenase Gene in p*gap*11*—The primary strcuture of the regions of the glyceraldehyde-3-phosphate dehydrogenase gene in p*gap*11 which are adjacent to the translational initiation and termination codons is shown in Fig. 6. As observed for the genes contained on p*gap*491 (3) and p*gap*63 (4), the A + T composition of the nocoding regions of the gene on p*gap*11 are 72% for 150 nucleotides upstream from the translational initiation codon and 74% for 120 nucleotides downstream from the translational termination codon. The hexanucleotide TATAAA is located 150 nucleotides upstream from the translational initiation codon in p*gap*11. This hexanucleotide is located 139 and 130 nucleotides upstream from the translational initiation codons in p*gap*491 and p*gap*63, respectively (4). There are two regions of strong nucleotide sequence homology between the 5′-noncoding sequences in p*gap*11 and those in p*gap*491 and p*gap*63. These regions are illustrated in Fig. 7. The sequences between −1 and −38 in p*gap*11 are homologous to the sequences between −1 and −47 in p*gap*491 and those between −1 and −56 in p*gap*63. The region between −39 and −88 in p*gap*11 is also homologous to these same regions in p*gap*491 and p*gap*63. Based on the strong similarities in primary structure between these two regions in p*gap*11, it is likely that they arose by tandem duplication.

Within these same regions of nucleotide sequence homology among the glyceraldehyde-3-phosphate dehydrogenase genes,

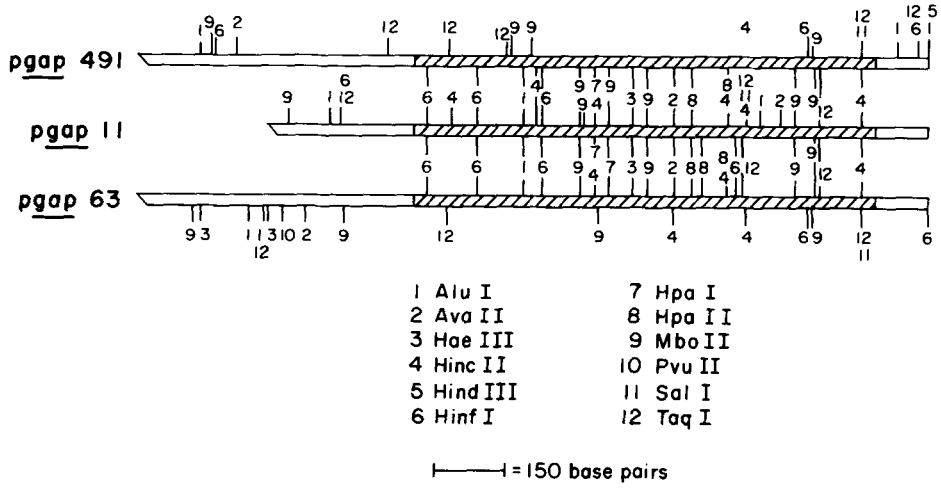


FIG. 3. **Comparison of the restriction endonuclease maps of the three yeast glyceraldehyde-3-phosphate dehydrogenase genes.** A detailed restriction endonuclease map of the portion of p*gap*11 which contains the glyceraldehyde-3-phosphate dehydrogenase gene was derived by partial digestion of *Eco*RI/*Sal*I fragments (1.5 and 2.8 kb) after ^{32}P-labeling of the 5′ termini of *Sal*I-digested p*gap*11 with the indicated restriction endonucleases. The restriction maps of p*gap*491 and p*gap*63 are shown for comparison. The *cross-hatched regions* correspond to the coding portions of the genes.

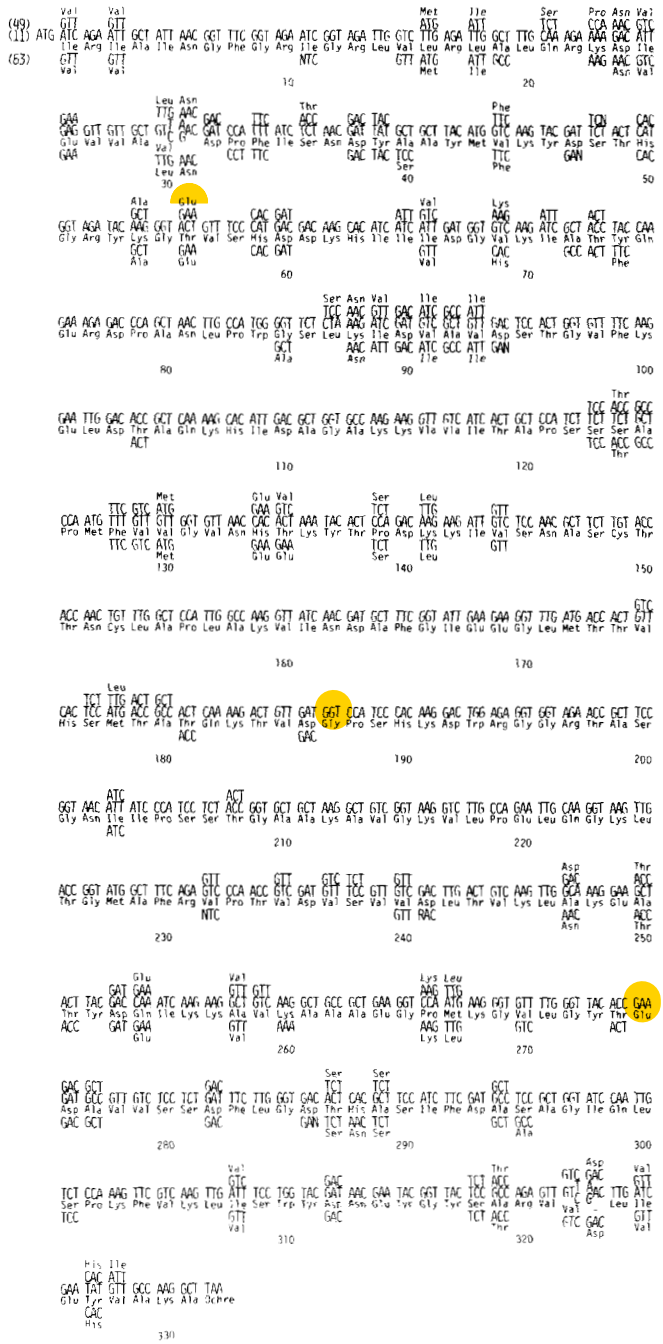


FIG. 4. **Nucleotide sequence of the coding region of the glyceraldehyde-3-phosphate dehydrogenase gene in p*gap*11.** The continuous nucleotide sequence is for the glyceraldehyde-3-phosphate dehydrogenase gene in p*gap*11 extending from the ATG initiation codon to the TAA termination codon. The amino acid sequence predicted from the nucleotide sequence is shown *below* the nucleotide sequence. Codons in p*gap*49 which differ from p*gap*11 are shown *above* the continuous sequence. Codons in p*gap*63 which differ from p*gap*11 are shown *below* the continuous sequence. In positions where the codon in p*gap*49 or p*gap*63 predicts a different amino acid than in p*gap*11, the amino acid is indicated *above* (p*gap*49) or *below* (p*gap*63) the codon.

there are strong homologies with corresponding 5′-noncoding regions of the two yeast enolase genes (8). As shown in Fig. 7, all five genes contain a hexanucleotide homologous to CACACA, 5–15 nucleotides upstream from the respective translational initiation codons. A second region of strong sequence homology is located 22–41 nucleotides upstream from each respective initiation codon. Interestingly, the apparent duplication of sequences in p*gap*11 (nucleotides −39 to −88) begins with the CACACA hexanucleotide and ends with the second homologous region described above. The significance of these homologous regions will be discussed further in the following section.

The 3′-noncoding region of the gene contained in p*gap*11 is not homologous to the corresponding regions in p*gap*491 and p*gap*63, although the latter two genes contain significant homologies within the first 100 nucleotides following the translational termination codon (4).

Mapping the 5′ Termini of the Messenger RNAs Encoded by the Yeast Enolase and Glyceraldehyde-3-phosphate Dehydrogenase Gene Families—In order to determine if the three glyceraldehyde-3-phosphate dehydrogenase structural genes are expressed in yeast, hybridization analyses were carried out with total poly(A)-containing yeast mRNA and hybridization probes containing sequences which are complementary to the 5′-noncoding portions of the three structural genes. Since the coding portions of the structural genes are extremely homologous and cross-hybridize, hybridization probes were prepared which included sequences from the 5′-noncoding regions of the genes. A parallel set of experiments was carried out with the two enolase structural genes. From the analysis, it is possible to map the 5′ terminus of the mRNA and to ascertain if the mRNA is synthesized.

In the case of the yeast enolase genes, both genes are expressed during the vegetative cellular growth; however, the amount of mRNA synthesized from the two genes is highly dependent on the carbon source used to propagate the cells (11). In cells growing logarithmically in medium containing glucose, for example, 95% of the enolase mRNA is derived from the structural gene contained on the plasmid designated p*eno*8, while 5% is encoded by the gene contained on the hybrid plasmid designated p*eno*46 (8, 11). In order to map the 5′ termini of the two enolase mRNAs, hybridization probes were isolated from each plasmid which extended from a common *Hin*fI restriction endonuclease cleavage site located 34 nucleotides downstream from the translational initiation codons of both genes, to sites far upstream from the translational initiation condons. The nucleotide sequences of the two enolase structural genes are identical between the initiation codon and the *Hin*fI cleavage site. The probes were labeled with polynucleotide kinase at the 5′ termini of the *Hin*fI cleavage sites and were then hybridized with total yeast poly(A)-containing mRNA as described under "Experimental Procedures." The poly(A)-containing mRNA was isolated from cells grown in the presence of glucose as carbon source. Under these growth conditions, the mRNA encoded by the gene contained on p*eno*46 comprises approximately 5% of the enolase mRNA in the cell. S1 nuclease digestion of the hybrids formed between the probe derived from the gene on p*eno*46 and this mRNA preparation should reveal at least two resistant hybrids. The major hybrid would be formed between the probe and mRNA synthesized from the gene corresponding to p*eno*8 and should extend from the *Hin*fI site through the

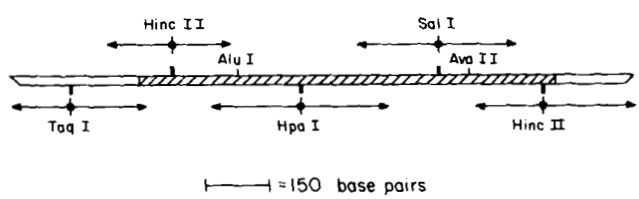


FIG. 5. **Strategy for determining the nucleotide sequence of the glyceraldehyde-3-phosphate dehydrogenase gene in p*gap*11.** The *arrows* indicate the direction and amount of sequence determined from each end-labeled restriction endonuclease fragment.

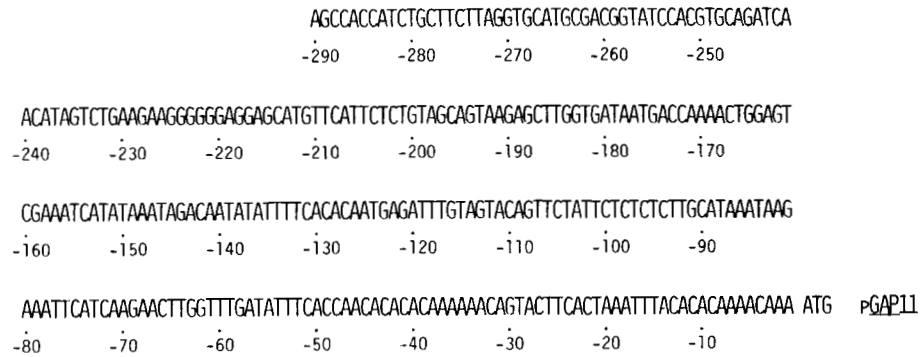


FIG. 6. **Nucleotide sequence of the 5′- and 3′-noncoding regions of the glyceraldehyde-3-phosphate dehydrogenase gene in p*gap*11.** The nucleotide sequence 290 nucleotides upstream from the ATG initiation codon and 126 nucleotides downstream from the TAA termination codon is shown for the gene within p*gap*11.

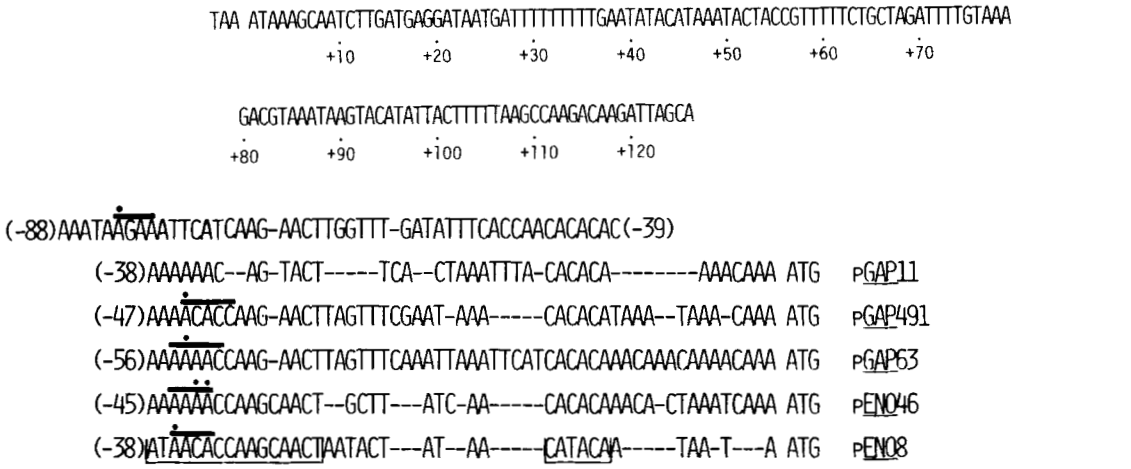
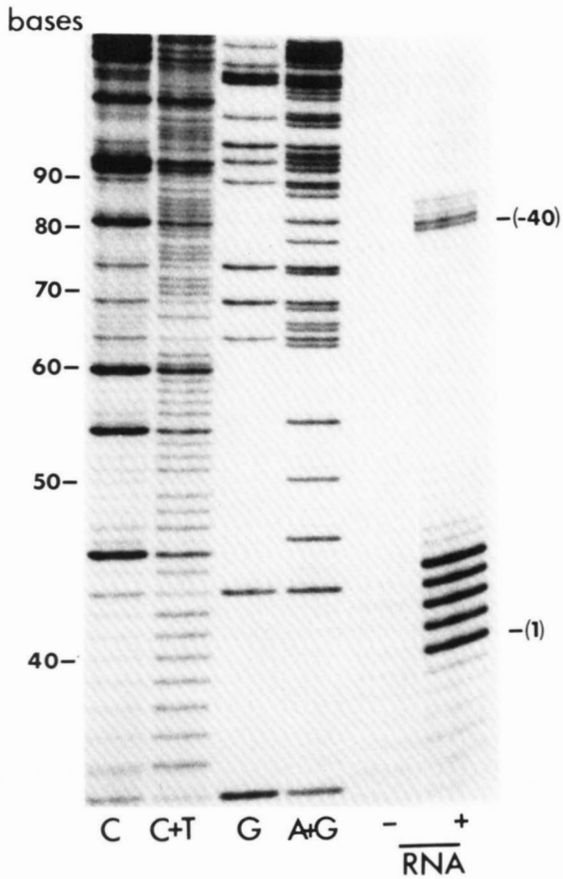


FIG. 7. **Comparison of the 5′-noncoding regions of three yeast glyceraldehyde-3-phosphate dehydrogenase genes and two yeast enolase genes.** The 5′-noncoding regions of p*eno*8, p*eno*46, p*gap*63, p*gap*491, and p*gap*11 have been aligned relative to the ATG initiation codons to maximize homology. Two regions of homology within the 5′-noncoding region of the gene in p*gap*11 have been aligned. Two regions of strong homology among the five genes are indicated by the *brackets* below the p*eno*8 sequence. The mapped positions of the 5′ termini of mRNAs synthesized from the genes are indicated by the *bars above* the sequences. The *solid dots above* the *bars* indicate the location of intense hybrids formed between the genes and mRNA. The *numbers within parentheses* refer to the number of nucleotides upstream from the respective ATG initiation codons.

homologous coding sequences to the last homologous nucleotide which is one nucleotide upstream from the translational initiation codon. The second hybrid would be formed between the probe and mRNA synthesized from the homologous gene corresponding to p*eno*46. As illustrated in Fig. 8, the two expected hybrids are formed. The smaller, more abundant hybrid maps to a position one nucleotide upstream from the initiation codon while a less intense hybrid is present which maps to a position 40 nucleotides upstream from the initiation codon. Neither hybrid is formed in the absence of yeast mRNA. In each case, a series of five hybrids, differing by one nucleotide from each other, is observed. This same pattern is observed for the glyceraldehyde-3-phosphate dehydrogenase-mapping experiments. In the case of the hybrid formed between the probe isolated from the gene in p*eno*46 and the mRNA synthesized from the gene corresponding to p*eno*8, we know from the primary structures of the enolase genes that the last homologous nucleotide is one position upstream from the initiation codon (8). The series of five hybrids extends from three nucleotides upstream from the last homologous nucleotide to one nucleotide downstream from this nucleotide. These data suggest that S1 nuclease trims the hybrids within a few nucleotides of the last base pair in the hybrid. We therefore assume that the error in mapping is ±4–5 nucleotides. Based on these data, it is unlikely that the multiple hybrids reflect heterogeneity at the 5′ termini of the mRNAs.

A similar experiment was carried out with the probe isolated from p*eno*8. As illustrated in Fig. 9, a single S1 nuclease-resistant hybrid is observed which maps 36 nucleotides upstream from the translational initiation codon in p*eno*8.

Hybridization probes were isolated from the three glyceraldehyde-3-phosphate dehydrogenase structural genes which extend from a *Hin*fI restriction endonuclease cleavage site 29 nucleotides downstream from the initiation codon in all three genes to sites far upstream from the initiation codon in each respective gene. Hybridization was carried out as described for the enolase genes utilizing total poly(A)-containing mRNA isolated from cells grown in the presence of glucose. The probes isolated from p*gap*491, p*gap*63, and p*gap*11 formed hybrids which mapped to sites 44, 53, and 82 nucleotides upstream from the initiation codon of each respective gene (Fig. 9). In each case, the hybrid observed extended beyond the initiation codon and includes sequences within the 5′-noncoding region of the gene. Since the 5′-noncoding regions of the three glyceraldehyde-3-phosphate dehydrogenase genes lack sufficient homology to cross-hybridize, we conclude that



the hybrids observed were formed with mRNA synthesized from the gene which corresponds to the hybridization probe. The size of the hybrid formed, therefore, corresponds to the number of nucleotides from the *Hin*fI site to the last nucleotide which is complementary to the 5′ terminus of the mRNA.

The locations of the 5′ termini of the enolase and glyceraldehyde-3-phosphate dehydrogenase mRNAs, calculated from the data presented in Figs. 8 and 9, are shown in Fig. 7. The *solid bars* above each sequence correspond to the end points of the hybrids formed with each probe. The *dots* above the *bars* indicate the most intense of the family of hybrids. The most striking feature of these data is that the 5′ termini of all five mRNAs are located within a region of strong nucleotide sequence homology among the genes. In the case of the gene contained on p*gap*11, the 5′ terminus of the mRNA synthesized from this maps only to one of the duplicated homologous sequences. These data demonstrate that mRNA is synthesized *in vivo* from all three glyceraldehyde-3-phosphate dehydrogenase genes. They also show that the 5′-nontranslated re-

FIG. 8. **Mapping of the 5′ terminus of the mRNA synthesized *in vivo* from the enolase gene corresponding to p*eno*46.** S1 nuclease-resistant hybrids formed between a probe isolated from p*eno*46 which includes the 5′-noncoding portion of the gene (see "Experimental Procedures") and total yeast poly(A)-containing mRNA are shown in the *lane* on the *right* (+*RNA*). The control without added RNA (−*RNA*) is in the *parallel lane*. The number of bases from the 5′ end label is shown on the *right axis*. The number of bases from the ATG initiation codon is shown in *parentheses* on the *left axis*. A series of chemical cleavage reactions were electrophoresed in the *first four lanes* in order to determine the molecular weights of the hybrids.

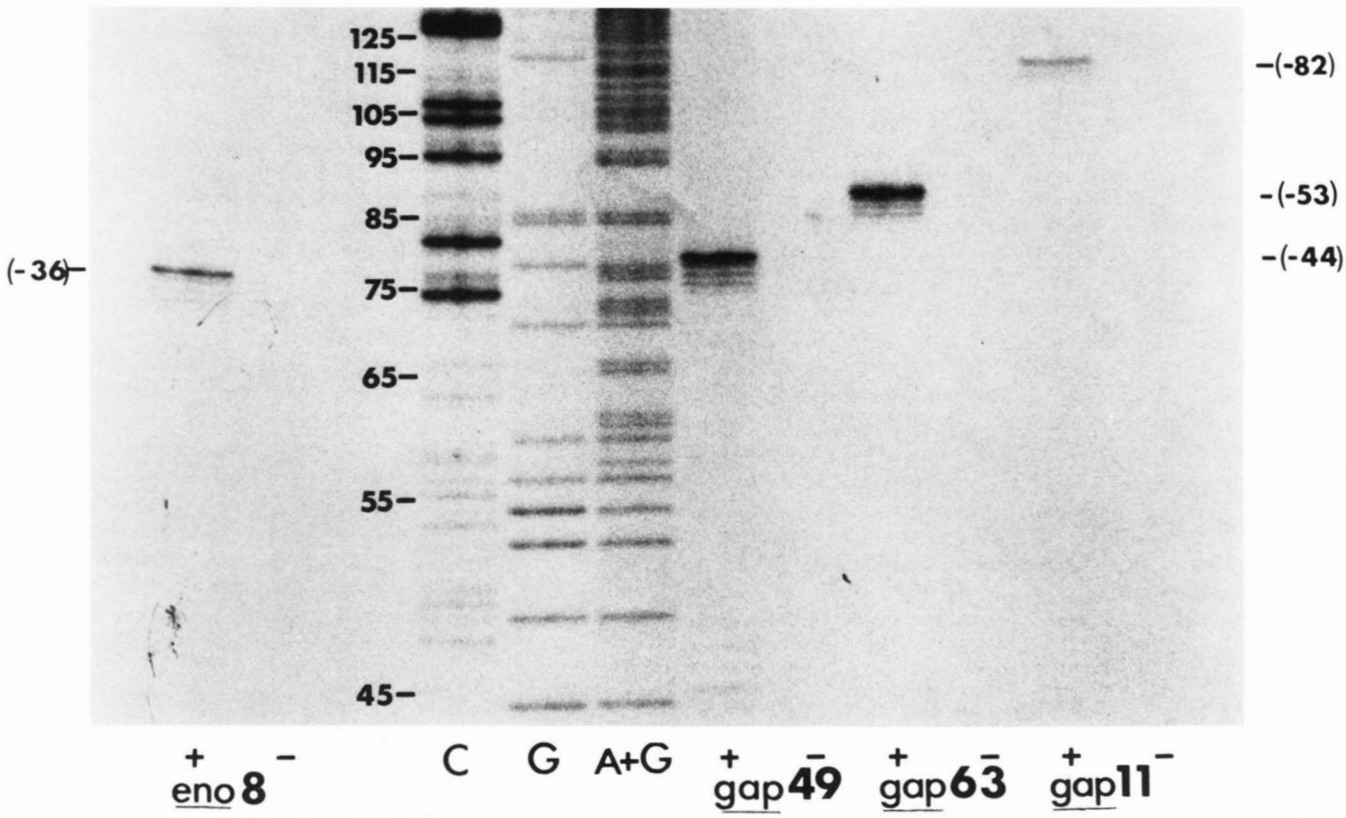


FIG. 9. **Mapping of the 5′ termini of the mRNAs synthesized *in vivo* from the genes contained within p*eno*8, p*gap*49, p*gap*63, and p*gap*11.** S1 nuclease-resistant hybrids formed between ^{32}P-labeled probes (see "Experimental Procedures") containing the 5′-noncoding regions of the genes within p*eno*8, p*gap*49, p*gap*63, and p*gap*11 and total yeast poly(A)-containing RNA are shown in the presence of RNA (+) and in the absence (−) of RNA. The *numbers in parentheses* refer to the number of nucleotides upstream from each respective ATG initiation codon. A series of chemical cleavage reactions were electrophoresed in parallel in order to determine the molecular weights of the hybrids.

gions of the yeast enolase and glyceraldehyde-3-phosphate dehydrogenase mRNAs are homologous.

DISCUSSION

Previous reports from this laboratory (3, 4) and the data presented here confirm that there are three nontandemly repeated glyceraldehyde-3-phosphate dehydrogenase structural genes in yeast. These genes are not alleles since they are present in haploid strains of *S. cerevisiae*. None of the genes contains an intervening sequence (3, 4). The gene contained on the p*gap*11 plasmid is homologous to but not identical with the other two (3, 4) within the coding portions of the genes. Codon usage within the gene on p*gap*11 follows the same highly biased pattern observed for the other glyceraldehyde-3-phosphate dehydrogenase genes (4).

Based on the primary structures of the three glyceraldehyde-3-phosphate dehydrogenase genes, none appears to be a pseudogene. The S1 nuclease-mapping data demonstrate that mRNA is synthesized from all three structural genes. It seems likely that the three polypeptides are also synthesized *in vivo* and that they might give rise to isozymes of the tetrameric enzyme. The most convincing eveidence that there are at least two glyceraldehyde-3-phosphate dehydrogenase polypeptides present in yeast is based on the fact that two ambiguities in the primary structure of the enzyme (7) can be resolved if polypeptides are present in the cell which are encoded by at least two of the isolated genes. Jones and Harris (7) report similar molar yields of serine and threonine at position 36 and high yields of valine and isoleucine at position 328. The gene isolated on p*gap*491 encodes threonine and isoleucine at these positions while serine and valine are encoded at these positions by the genes on p*gap*63 and p*gap*11. The remainder of the primary structure determined by Jones and Harris (7) agrees closely with that predicted from the gene in p*gap*491. Based on these data, one would predict that the polypeptide encoded by the gene in p*gap*491 was the major polypeptide in the preparation which was sequenced. It is also reasonable to conclude that one or both of the polypeptides encoded by the genes in p*gap*63 and p*gap*11 were present in the preparation.

The simplest mechanism for the evolution of three glyceraldehyde-3-phosphate dehydrogenase structural genes in yeast would be to postulate two successive duplication events. The coding portion of the gene contained in p*gap*11 is 90% homologous to the coding portions of the genes contained in p*gap*491 and p*gap*63. The coding portions of the genes in p*gap*491 and p*gap*63 are 95% homologous. If one ignores the possibility of preferential recombination between specific pairs of structural genes, then one would predict that the gene contained in p*gap*11 is a product of the first duplication event while those in p*gap*491 and p*gap*63 are the products of a subsequent duplication. Minimum estimates of the times of these duplication events, calculated from the observed rate of divergence of glyceraldehyde-3-phosphate dehydrogenase (12), are 200 and 100 million years, respectively. These estimates would represent minimum values since concerted evolution of the family of yeast glyceraldehyde-3-phosphate dehydrogenase genes (*i.e.* recombination among the genes) would minimize the observed sequence divergence among the genes. It is quite possible that the yeast genes have evolved in concert since multiple forms of glyceraldehyde-3-phosphate dehydrogenase have been observed in a wide variety of eucaryotic cells, suggesting duplication events which are much earlier than those estimated above.

The distribution of nucleotide sequence changes within specific portions of the coding regions of the three yeast glyceraldehyde-3-phosphate dehydrogenase structural genes was analyzed in order to test the possibility that preferential recombination occurs between specific pairs of structural genes. If the three genes do not undergo recombination or if recombination occurs randomly among the three genes, then the time and order of duplication estimated from the sequence divergence data determined for any statistically significant portion of the three coding regions should be similar. Initially, divergence within the portions of the genes which encode the two functional domains of the polypeptides was examined. Within the portions of the structural genes which encode the NAD-binding domains of the polypeptides (codons 1–147), the gene contained in p*gap*11 differs from those contained in p*gap*491 and p*gap*63 at 69 and 63 nucleotide positions, respectively. The latter two genes differ by 30 nucleotides in this region. The portion of the gene in p*gap*11 which encodes the catalytic domain of the polypeptide (codons 148–331) differs from the genes in p*gap*491 and p*gap*63 at 40 and 38 nucleotide positions, respectively. The genes contained in p*gap*491 and p*gap*63 differ by 23 nucleotides within the regions encoding the catalytic domains. Based on these data, sequences encoding the NAD-binding domain in p*gap*11 are 2.3-fold (69:30) more diverged from the genes in p*gap*491 or p*gap*63 than are the latter genes from each other. In the case of sequences encoding the catalytic domains, the gene in p*gap*11 is 1.7-fold (40:23) more diverged from those in p*gap*491 or p*gap*63 than are the latter genes from each other. In both cases, the data predict that the genes contained in p*gap*491 and p*gap*63 are the products of a second duplication event. The estimated times of the two duplication events, however, are somewhat different if one considers the data from the NAD-binding domain *versus* the catalytic domain. The catalytic domain divergence data predict that the second duplication is closer to the first duplication than do the data from the NAD-binding domain.

A more striking anomaly in the divergence pattern of the three yeast structrual genes is observed if one considers those sequences between codons 144 and 243. This portion of the structural genes encodes the most highly conserved portion of the polypeptides. The amino acid sequences of the yeast, lobster, and pig glyceraldehyde-3-phosphate dehydrogenases are extremely homologous in this region (7, 10). Within this region, there is a single amino acid difference predicted from the sequences of the three yeast genes. The structural genes in p*gap*11 and p*gap*63 predict a methionine residue at position 178 while the gene in p*gap*491 predicts leucine at this position. Within this region, the sequences of the genes in p*gap*491 and p*gap*63 differ by 11 silent third position codon changes and a single first position codon change at codon 178. The gene contained in p*gap*11 differs within this region from the genes in p*gap*491 and p*gap*63 by 13 and 5 nucleotides, respectively. Interestingly, these data predict that the genes in p*gap*11 and p*gap*63 are the products of the second duplication. Since all but one of the nucleotide sequence differences among the three genes in this region are silent third position codon changes, it seems unlikely that the conservation of nucleotide sequence between the genes in p*gap*11 and p*gap*63 in this region of the structural genes is the result of selective pressure at the polypeptide level. An attractive explanation for the anomalous duplication times predicted from the divergence data in the NAD-binding domains *versus* the catalytic domains would be to postulate preferential recombination among the three structural genes within sequences encoding the catalytic domains *versus* the NAD-binding domains. In the case of sequences encoding residues 144–243 of the polypeptides, one would further postulate preferential recombination in this region between the genes in p*gap*11 and p*gap*63.

The 5′-noncoding region of the gene in p*gap*11 is homolo-

gous to the genes contained in p*gap*491 and p*gap*63 within the regions adjacent to the translational initiation codons. This region of homology among the three genes appears to have been tandemly duplicated in p*gap*11. No homology between the 3′-noncoding portion of the gene in p*gap*11 and corresponding regions of the other two genes was observed.

The S1 nuclease-mapping data demonstrate that mRNA is synthesized from all three yeast glyceraldehyde-3-phosphate dehydrogenase genes. These data also show that the 5′ termini of the glyceraldehyde-3-phosphate dehydrogenase genes and the enolase genes map to a region of nucleotide sequence homology which is shared by all five structural genes. Thus, the homologous portions of these two gene families which are adjacent to the initiation codons (8) are present within the 5′-nontranslated regions of the mRNAs synthesized from the genes. While the functional significance of this homologous region of nucleotide sequence is not known, it is likely that these sequences evolved from a common precursor. Since the coding portions of the glyceraldehyde-3-phosphate dehydrogenase gene family is unrelated to the coding regions of the enolase genes, it is likely that the complete genes evolved by a segmental process.

The 5′-noncoding regions of the glyceraldehyde-3-phosphate dehydrogenase and enolase genes have been compared to other yeast genes for which the 5′ terminus of the mRNA synthesized from the gene is known. Within the sequences surrounding those corresponding to the 5′ terminus of the respective mRNA, no homology was found among the genes coding for yeast iso-1-cytochrome *c* (13), yeast TRP5 (14), and those reported here. In contrast, the genes coding for the yeast alcohol dehydrogenases (15, 16) contain regions of significant homology. The region surrounding the sequences coding for the 5′ termini of the glyceraldehyde-3-phosphate dehydrogenase and enolase genes has the general structure: AAAAAACCAAGCAACT where the underlined region indicates the mapped termini of the mRNAs. The corresponding regions in ADC1 is AATATTTCAAGCTATACCAAG CATAC. In ADR2, the sequence of the corresponding region is: AGAATATCAAGCTACA. In both of these genes, the sequence CAAGC is present at or near the mapped 5′ termini of the mRNAs. This sequence is homologous to sequences adjacent to the 5′ termini of the glyceraldehyde-3-phosphate dehydrogenase and enolase genes. Finally, the sequence surrounding the site mapped for the 5′ terminus of the yeast HIS3 gene (17) is: AAAAAATGAGCAGGC. This sequence is also homologous to the corresponding regions of the genes described here but the degree of homology with the CAAGC sequence is not as strong as for the alcohol dehydrogenase genes. Interestingly, the sequence AAAAAAC- - AG-TACT is present in p*gap*11 between nucleotides −26 and −38. Although this sequence is extremely close to the consensus sequence for the genes, it does not correspond to the 5′ terminus of a mRNA detected in the cell. This sequence lacks the CAAG portion of the homology. It is tempting to speculate that the homologous portions of these genes play some role in transcription or translation. Correlation of these sequences with expression of the genes will require further analysis of the expression of genes containing defined alterations within these homologous sequences.

REFERENCES

1. Holland, M. J., and Holland, J. P. (1979) *J. Biol. Chem.* **254,** 5466–5474
2. Holland, M. J., Holland, J. P., and Jackson, K. A. (1979) *Methods Enzymol.* **68,** 498–419
3. Holland, J. P., and Holland, M. J. (1979) *J. Biol. Chem.* **254,** 9839–9845
4. Holland, J. P., and Holland, M. J. (1980) *J. Biol. Chem.* **255,** 2596–2605
5. Krebs, E. G. (1953) *J. Biol. Chem.* **200,** 471–478
6. Holland, M. J., and Holland, J. P. (1978) *Biochemistry* **17,** 4900–4907
7. Jones, G. M. T., and Harris, J. I. (1972) *FEBS Lett.* **22,** 185–189
8. Holland, M. J., Holland, J. P., Thiol, G. P., and Jackson, K. A. (1981) *J. Biol. Chem.* **256,** 1385–1395
9. Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C., Rossmann, M. G., and Harris, J. I. (1975) *J. Biol. Chem.* **250,** 9137–9162.
10. Olsen, K. W., Moras, D., Rossmann, M. G., and Harris, J. I. (1975) *J. Biol. Chem.* **250,** 9313–9321
11. McAlister, L., and Holland, M. J. (1982) *J. Biol. Chem.* **257,** 7181–7188
12. Wilson, A. C., Carlson, S. S., and White, T. (1977) *Annu. Rev. Biochem.* **46,** 573–639
13. Boss, J. M., Gillam, S., Zitomer, R. S., and Smith, M. (1981) *J. Biol. Chem.* **256,** 12958–12961
14. Zalkin, H., and Yanofsky, C. (1982) *J. Biol. Chem.* **257,** 1491–1500
15. Bennetzen, J. L., and Hall, B. D. (1982) *J. Biol. Chem.* **257,** 3018–3025
16. Young, T., Williamson, V., Taguchi, A., Smith, M., Sledziewski, A., Russell, D., Osterman, J., Denis, C., Cox, D., and Beier, D. (1981) in *Genetic Engineering of Microorganisms for Chemicals* (Hollaender, A., ed) pp. 335–361, Plenum Press, New York
17. Struhl, K. (1981) *Proc. Natl. Acad. U. S. A.* **78,** 4461–4465

Additional references are found on p. 5299.

SUPPLEMENTARY MATERIAL

TO

Homologous Nucleotide Sequences at the 5'-Termini of Messenger RNAs Synthesized from the Yeast Enolase and Glyceraldehyde-3-Phosphate Dehydrogenase Gene Families: The Primary Structure of a Third Yeast Glyceraldehyde-3-Phosphate Dehydrogenase Gene

BY

Janice P. Holland, Laura Labieniec, Candice Swimmer, and Michael J. Holland

EXPERIMENTAL PROCEDURES

Materials and Bacterial Strains. [γ-^{32}P]ATP (2000-3000 Ci/mmol) was purchased from New England Nuclear and [α-^{32}P]dCTP (2000-3000 Ci/mmol) was obtained from Amersham Corporation. Restriction endonucleases Alu I, Ava II, BamH I, Hpa II, Sal I, Sma I, Taq I, and Xba I and bacterial alkaline phosphatase were purchased from Bethesda Research Laboratories. Restriction endonucleases Bgl II, Hae III, Hinc II, Hind III, Hinf I, Hpa I, Mbo II and Pvu II were obtained from New England Biolabs. Polynucleotide kinase and DNA ligase (T_4-infected E. coli) were supplied by P-L Biochemicals. DNA polymerase I was purchased from Boehringer Mannheim. Bacterial strains and growth conditions were as described in Holland and Holland (1). Plasmid DNA was prepared according to the procedures of Clewell and Helinski (2) and Colman et al (3). pgap491, an ampicillin resistant recombinant containing a glyceraldehyde-3-phosphate dehydrogenase gene, was isolated as described by Holland and Holland (1). pgap492, a hybrid plasmid composed of the vector, pSF2124 (4) and an Eco RI fragment of yeast DNA containing the same gene, was isolated as outlined in Holland et al (5).

Isolation of pgap11. Ten pools, each of 2000 E. coli RR101 ampicillin resistant transformants containing Eco RI restricted yeast genomic DNA ligated into the Eco RI site of the cloning vector pSF2124, were screened by a combination of Southern blotting and colony filter hybridization techniques as described by Holland et al (5) for the presence of hybrid plasmids containing glyceraldehyde-3-phosphate dehydrogenase genes. The 2.1 kb Hind III fragment from pgap491 (1) was nick translated in the presence of [α-^{32}P]dCTP according to the procedure of Rigby et al (6) and used as a hybridization probe. The hybrid plasmid designated pgap11 was isolated and shown by Southern blotting and restriction mapping described below to contain the third glyceraldehyde-3-phosphate dehydrogenase structural gene on a 4.3 kb Eco RI fragment of genomic DNA.

Southern Blotting Analysis of pgap11. 0.25 µg each of pgap492 (5) and pgap11 and 2.5 µg of yeast DNA were limit digested with the restriction enzyme Eco RI and the cleavage products displayed on a 0.8% agarose slab gel electrophoresed at 60 V for 5 hr. After electrophoresis, DNA was transferred to a nitrocellulose filter by the modified procedure of the Southern technique (8) outlined by Ketner and Kelley (9). The filter was pretreated for 3 h at 65° in 3 X SSC (1 x SSC = 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0) containing Denhardt's solution (10): 0.02% Ficoll, 0.02% polyvinylpyrolidine, and 0.02% bovine serum albumin. The filter was then pretreated for 1 h at 65° with the above solution supplemented with 0.1% sodium dodecyl sulfate and 10 µg/ml depurinated salmon sperm DNA and blotted dry. Hybridization was performed in the latter solution at 65° for 36 h under mineral oil utilizing 2.5 x 10^5 cpm of a [^{32}P]dCTP-labeled 2.1 kb Hind III fragment from pgap491 (1) nick translated according to the procedure of Maniatis et al (9). Following incubation the filter was washed 3 times with chloroform at room temperature to remove the oil and 6 times with 300 ml 6 x SSC at 65° for a total of 90 min. Autoradiography was performed as previously described.

Restriction Endonuclease Cleavage Analysis. A restriction endonuclease cleavage map was established for the pgap11 hybrid plasmid by first digesting 0.5 µg DNA with the following restriction enzymes for 1 h at 37°C in 30 µl reaction volumes under the suppliers' prescribed conditions: Bam HI, Bgl II, Eco RI, Hind III, Hpa I, Sal I, Sma I and Xba I. Digestions were terminated by the addition of 0.2 volumes of a solution containing 0.125 M EDTA, pH 7.8, 25% glycerol and 2% Sarkosyl. A series of digestions with two restriction endonucleases were then analyzed. The results of these digestions are tabulated in Table IS. Based on these data the map shown in figure 2 was determined.

A detailed restriction endonuclease cleavage map involving enzymes with four and five nucleotide recognition sites was also established for pgap11 using the partial digestion procedure of Smith and Birnstiel (7). 25 µg of pgap11 was digested with the restriction enzyme Sal I, dephosphorylated with bacterial alkaline phosphatase according to Smith and Birnstiel (7) and phosphorylated with polynucleotide kinase and [γ-^{32}P]ATP by the method of Maxam and Gilbert (8). Following phenol extraction, three sequential ethanol precipitations were carried out to remove unincorporated [γ-^{32}P]ATP. The end-labeled Sal I linear DNA was digested with Eco RI and the 2.8 kb and 1.5 kb fragments containing the glyceraldehyde-3-phosphate dehydrogenase gene were isolated by preparative agarose gel electrophoresis. A series of partial digestions of these two fragments were carried out (3000 cpm end-labeled DNA per digestion) with Alu I, Ava II, Hae III, Hinc II, Hind III, Hinf I, Hpa II, Mbo II, Tac I and Taq I. The cleavage products were displayed on a 1.5% agarose slab gel (30 x 13.5 cm) and visualized by autoradiography with Kodak X-OMAT R film and a Dupont Lightning Plus intensifying screen for exposure times of 24-48 h. These data and a series of limit digestions with each enzyme were used to generate the map shown in figure 3.

DNA Sequence Analysis. End-labeled 2.8 kb and 1.5 kb Sal I/Eco RI fragments from pgap11, prepared as described above for restriction endonuclease analysis, were chemically modified and cleaved according to the method of Maxam and Gilbert (8). Cleavage products were analyzed by 0.4 mm 20% (or 12%) polyacrylamide gel electrophoresis at 1100 V using a 7, 20 and 36 h (14-16 h) loading schedule. Autoradiography was carried out as described above.

20 µg of the 4.3 kb Eco RI fragment from pgap11 containing the glyceraldehyde-3-phosphate dehydrogenase gene, isolated by sucrose density centrifugation, was digested with the restriction enzyme Taq I, dephosphorylated and phosphorylatnd as previously outlined. Two fragments, 0.75 kb and 0.9 kb containing 5' flanking sequences, were isolated by preparative polyacrylamide gel electrophoresis. These fragments were identified by secondary digestions with the restriction enzyme Hae III. Single end-labeled fragments were eluted from polyacrylamide gels after digestion with Hae III. DNA sequence analysis was carried out according to the procedure of Maxam and Gilbert (8). 25 µg of the 4.3 kb Eco RI fragment was digested with the restriction enzyme Hpa I. Labeling of the blunt-ended fragments was carried out as described above except that the fragments were denatured by boiling for 5 min prior to the polynucleotide kinase reaction. The [^{32}P]-labeled 2.5 kb and 1.8 kb fragments were renatured and cleaved by the restriction endonucleases Bgl II and Sal I, respectively. The resulting 0.7 kb Bgl II/Hpa I and 0.3 kb Hpa I/Sal I fragments were isolated by preparative agarose gel electrophoresis and sequenced as described above. 25 µg of the 4.3 kb Eco RI fragment were digested with the restriction enzyme Hinc II. Blunt-ended fragments were labeled as described above and isolated by preparative agarose gel electrophoresis. A 0.8 kb Hinc II fragment containing 5'-noncoding sequences was cleaved by Hha I and the resulting 0.5 kb end-labeled fragment was sequenced. A 1.5 kb Hinc II fragment containing 3' noncoding sequences was cleaved by Ava II and resulting 1.3 kb Ava II fragment was analyzed. Coding region sequences obtained from the 0.19 and 0.25 kb Hinc II fragment were cleaved respectively by Alu I and Ava II and the resulting 0.16 kb Alu I/Hinc II and 0.17 kb Ava II/Hinc II fragments were sequenced. The overall sequencing strategy is summarized in figure 5.

Mapping of the 5'-Termini of Messenger RNAs Synthesized from the Enolase and Glyceraldehyde-3-Phosphate Dehydrogenase Gene Families. The 5'-termini of mRNAs synthesized from the three glyceraldehyde-3-phosphate dehydrogenase and two enolase genes were determined by the S1 nuclease protection method described by Berk and Sharp (11). In the cases of the three glyceraldehyde-3-phosphate dehydrogenase genes, restriction endonuclease cleavage fragments were isolated from pgap49 (12), pgap63 (12) and pgap11. Each fragment was [^{32}P]-labeled as described above at a Hinf site which is common to all three genes and is located within the coding region, 30 nucleotides downstream from each respective translational initiation codon. In the case of pgap49, a 2.1 kb Hind III fragment containing the gene (12) was digested with Hinf I and the 5' termini were [^{32}P]-labeled as described above. A 0.45 kb Hinf I fragment containing the region of interest was isolated by preparative agarose gel electrophoresis. This fragment was then digested with Ava II and the 0.41 kb fragment containing the desired labeled Hinf terminus was isolated and used in the hybridization. A 3.2 kb Pvu II fragment containing the glyceraldehyde-3-phosphate dehydrogenase gene from pgap63 (12) was isolated, digested with Hinf I, [^{32}P]-labeled and a 0.31 kb Pvu II/Hinf I fragment containing the 5'-terminus of the gene was isolated. This fragment was further digested with Ava II and a 0.26 kb fragment containing the desired labeled Hinf I site was analyzed. In the case of pgap11, the 4.3 kb Eco RI fragment containing the gene was digested with Hinf I, end-labeled and a 0.17 kb Hinf I fragment was analyzed. The enolase genes contained on the hybrid plasmids designated peno8 and peno46 (13) share a common Hinf I cleavage site within the coding region 35 nucleotides downstream from the respective translational initiation codons. This site was end-labeled in order to generate fragments for the S1 nuclease protection experiment. The Sma I/Kpn I fragment which contains the 5'-terminus of the enolase gene in peno8 was isolated as previously described (13), digested with Hinf I, end-labeled as described above and finally, the 1.1 kb Hinf I fragment containing the region of interest was isolated by preparative agarose gel electrophoresis. A 1.6 kb Eco RI fragment containing the 5'-terminus of the gene contained in peno46 was isolated as previously described (13), digested with Hinf I, end-labeled as described above, and a 0.27 kb fragment containing the 5'-terminal sequences was isolated. This fragment was further digested with Hpa II and the 0.23 kb fragment, labeled at the Hinf I site of interest, was isolated and analyzed.

2.5 µg of poly(A)-selected yeast RNA or salmon sperm DNA were hybridized with 0.05 µg of each end-labeled restriction fragment in a 5 µl reaction volume containing 80% deionized formamide, 0.04 M Pipes, pH 6.4, 0.4 M NaCl and 0.001 M EDTA in sealed capillary tubes as described by Casey and Davidson (10). After heating for 10 min at 65°, the samples were incubated at 50° for 8 h. Following hybridization the samples were diluted ten-fold with buffer containing 0.3 M NaCl, 0.03 M sodium acetate, pH 4.6, 0.001 M $ZnSO_4$, 20 µg/ml single stranded salmon sperm DNA and 550 units S1 nuclease and incubated for 1 h at room temperature. After phenol extraction and ethanol precipitation, the sample was suspended in 90% formamide containing 0.05 M Tris-borate, 0.001 M EDTA, pH 8.3 and electrophoresed in a 20% acrylamide gel for 20 h at 1100 V as described for the DNA sequence analysis. Control samples containing salmon sperm DNA instead of poly(A)-selected RNA, were electrophoresed in parallel lanes.

REFERENCES

1. Holland, M.J. and Holland, J.P. (1979) J. Biol. Chem. 254, 5466-5474.
2. Clewell, D.B. and Helinski, D.R. (1970) Biochemistry 9, 4428-4440.
3. Colman, A., Byers, M.J., Primrose, S.B. and Lyons, A. (1978) Eur. J. Biochem. 91, 303-310.
4. So, M., Gill, R. and Falkow, S. (1975) Molec. Gen. Genet. 142, 239-249.
5. Holland, M.J., Holland, J.P. and Jackson, K.A. (1979) Methods Enzmol. 68, 408-418.
6. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
7. Smith, H.O. and Birnstiel, M.L. (1976) Nuc. Acids Res. 3, 2387-2398.
8. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
9. Maniatis, T., Jeffery, A. and Kleid, D.G. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 1184-1188.
10. Casey, J. and Davidson, N. (1977) Nuc. Acids Res. 4, 1539-1552.
11. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
12. Holland, J.P. and Holland, M.J. (1980) J. Biol. Chem. 255, 2596-2605.
13. Holland, M.J., Holland, J.P., Thill, G.P. and Jackson, K.A. (1981) J. Biol. Chem. 256, 1385-1395.

TABLE IS

Molecular Weight of Restriction Endonuclease Cleavage Fragments of pgap11

Restriction Endonuclease	Fragment Molecular Weight (Kb)		
Bgl II	13.9	1.65	
Eco RI	11.2	4.3	
Hind III	15.5		
Hpa I	15.5		
Sal I	15.5		
Xba I	15.5		
Sma I/Bgl II	12.5	1.65	1.35
Sma I/Hind III	13.7	1.8	
Sma I/Xba I	13.3	2.25	
Sma I/Hpa I	11.8	3.7	
Sma I/Sal I	11.5	4.0	
Bgl II/Xba I	13.9	0.9	0.75
Bgl II/Hind III	13.9	1.1	0.55
Bgl II/Hpa I	13.2	1.65	0.7
Bgl II/Sal I	12.9	1.65	1.0
Eco RI/Hind III	11.2	3.6	0.7
Hpa I/Sal I	15.2	0.3	
Eco RI/Sal I	11.2	2.8	1.5

Homologous nucleotide sequences at the 5' termini of messenger RNAs synthesized from the yeast enolase and glyceraldehyde-3-phosphate dehydrogenase gene families. The primary structure of a third yeast glyceraldehyde-3-phosphate dehydrogenase gene.

J P Holland, L Labieniec, C Swimmer and M J Holland

J. Biol. Chem. 1983, 258:5291-5299.
