Homologous Nucleotide Sequences at the 5' Termi 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 98% 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

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

* Portions of this paper (including “Experimental Procedures,” Table 18, 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, 9550 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.

† Portions of this paper (including “Experimental Procedures,” Table 18, 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, 9550 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.

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

§ American Heart Association Established Investigator.
DNA filter blotting after EcoRI digestion, 13.2-, 5.5-, and 4.3-kb fragments form hybrids with a nick-translated hybridization probe synthesized from the 2.1-kb HindIII fragment which contains the glyceraldehyde-3-phosphate dehydrogenase gene within the plasmid pgap491 (3). The 13.2-kb EcoRI fragment has been isolated on a plasmid designated pgap492 (2, 4) and corresponds to the structural gene already isolated on the plasmid designated pgap49 (1). The 5.5-kb EcoRI fragment corresponds to the structural gene previously isolated on the plasmid designated pgap63 (4).

The 4.3-kb EcoRI fragment was cloned as previously described for the isolation of pgap492 (2). EcoRI-digested yeast DNA was ligated into the EcoRI site of vector pSF2124. Total plasmid DNA was isolated from the shuttle vector preparation 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 HindIII fragment isolated from pgap491. A hybrid plasmid designated pgap11 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 pgap11 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 EcoRI-digested pgap492 (2, 4), pgap11, and yeast DNA. A nick-translated 2.1-kb HindIII fragment containing the glyceraldehyde-3-phosphate dehydrogenase gene within pgap492 was used as the hybridization probe. The 13.2-kb EcoRI fragment from pgap492 corresponds to the largest hybrid formed with genomic DNA. The 4.3-kb fragment from pgap11 corresponds to the smallest hybrid formed in genomic DNA. These data confirm that pgap11 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 EcoRI and HindIII 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 pgap492 (2, 4), pgap63 (4), and pgap11.

A restriction endonuclease cleavage map of pgap11, 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 HpaI cleavage site within the coding portion of the gene in pgap11 is homologous to HpaI cleavage sites in the genes contained within pgap492 and pgap63 (4). The restriction endonuclease cleavage map of pgap11 adjacent to the coding sequences is not homologous to the other two structural genes. These data confirm that none of the yeast 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 pgap11 is shown in Fig. 3. The restriction endonuclease cleavage maps of the corresponding regions of pgap491 and pgap63 (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 pgap11 is shown in Fig. 2.
as are the sequences for the genes in pgap491 (3) and pgap63 (4). The primary structure was determined using the sequencing strategy outlined in Fig. 5. The gene in pgap11 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 pgap491 is 89 and 88% homologous to the polypeptides encoded by the genes within pgap63 and pgap491, respectively. Based on these data, it is possible that the gene within pgap491 was derived from an early duplication event and that the genes contained on pgap63 and pgap491 arose from a subsequent duplication. Alternatively, the similarity between the genes on pgap63 and pgap491 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 pgap11 follows the same highly biased pattern previously described for those within pgap63 and pgap491 (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 pgap491, pgap63, or pgap11 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 pgap11 differs from those contained on pgap491 and pgap63 at 69 and 63 nucleotide positions, respectively. Within this same region, the genes contained on pgap491 and pgap63 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 pgap11 differs from those contained on pgap491 and pgap63 by 40 and 38 nucleotides, respectively. Within this region, the genes within pgap491 and pgap63 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 pgap11—The primary structure of the regions of the glyceraldehyde-3-phosphate dehydrogenase gene in pgap11 which are adjacent to the translational initiation and termination codons is shown in Fig. 6. As observed for the genes contained on pgap491 (3) and pgap63 (4), the A + T composition of the nongcoding regions of the gene on pgap41 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 pgap11. This hexanucleotide is located 139 and 130 nucleotides upstream from the translational initiation codons in pgap491 and pgap63, respectively (4). There are two regions of strong nucleotide sequence homology between the 5'-noncoding sequences in pgap41 and those in pgap491 and pgap63. These regions are illustrated in Fig. 7. The sequences between -1 and -38 in pgap41 are homologous to the sequences between -1 and -47 in pgap41 and those between -1 and -56 in pgap63. The region between -39 and -88 in pgap41 is also homologous to these same regions in pgap491 and pgap63. Based on the strong similarities in primary structure between these two regions in pgap41, it is likely that they arose by tandem duplication.

Within these same regions of nucleotide sequence homology among the glyceraldehyde-3-phosphate dehydrogenase genes,
there are strong homologies with corresponding 5'-noncoding regions of the two yeast enolase genes (8). As shown in Fig. 5, all five genes contain a hexanucleotide homologous to CATACA, 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 pgap11 (nucleotides -39 to -88) begins with the CATACA 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 pgap11 is not homologous to the corresponding regions in pgap49 and pgap63, although the latter two genes contain significant homologies within the first 100 nucleotides following the translational termination codon (4). Mapped the 5' Termini of the Messenger RNAs Encoded by the Yeast Enolase and Glyceroldehyde-3-phosphate Dehydrogenase Gene Families—In order to determine if the three glyceroldehyde-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 peno8, while 5% is encoded by the gene contained on the hybrid plasmid designated peno46 (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 Hinfi restriction endonuclease cleavage site located 34 nucleotides downstream from the translational initiation codon of both genes, to sites far upstream from the translational initiation codons. The nucleotide sequence of the two enolase structural genes are identical between the initiation codon and the Hinfi cleavage site. The probes were labeled with polymucleotide kinase at the 5' termini of the Hinfi restriction endonuclease 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 peno46 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 peno46 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 peno8 and should extend from the Hinfi site through the appropriate restriction endonuclease fragment.
Enolase and Glyceraldehyde-3-phosphate Dehydrogenase Gene Expression

Fig. 6. Nucleotide sequence of the 5'- and 3'-noncoding regions of the glyceraldehyde-3-phosphate dehydrogenase gene in pgap11. 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 pgap11.

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 peno8, peno46, pgup63, pgup491, and pgap11 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 pgap11 have been aligned. Two regions of strong homology among the five genes are indicated by the brackets below the peno8 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 peno46. 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 peno46 and the mRNA synthesized from the gene corresponding to peno8, 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 peno8. As illustrated in Fig. 9, a single S1 nuclease-resistant hybrid is observed which maps 36 nucleotides upstream from the translational initiation codon in peno8.
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 HinfI 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 pgap11, the 5' terminus of the mRNA synthesized from this map 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 peno14. S1 nuclease-resistant hybrids formed between a probe isolated from peno46 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 peno8, pgap49, pgap63, and pgap11. S1 nuclease-resistant hybrids formed between 32P-labeled probes (see "Experimental Procedures") containing the 5' noncoding regions of the genes within peno8, pgap49, pgap63, and pgap11 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 non tandemly 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 pgap11 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 pgap11 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 evidence 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 pgap491 encodes threonine and isoleucine at these positions while serine and valine are encoded at these positions by the genes on pgap63 and pgap11. The remainder of the primary structure determined by Jones and Harris (7) agrees closely with that predicted from the gene in pgap491. Based on these data, one would predict that the polypeptide encoded by the gene in pgap491 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 pgap63 and pgap11 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 pgap11 is 90% homologous to the coding portions of the genes contained in pgap491 and pgap63. The coding portions of the genes in pgap491 and pgap63 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 pgap11 is a product of the first duplication event while those in pgap491 and pgap63 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 pgap11 differs from those contained in pgap491 and pgap63 at 59 and 63 nucleotide positions, respectively. The latter two genes differ by 30 nucleotides in this region. The portion of the gene in pgap11 which encodes the catalytic domain of the polypeptide (codons 148-331) differs from the genes in pgap491 and pgap63 at 40 and 38 nucleotide positions, respectively. The genes contained in pgap491 and pgap63 differ by 23 nucleotides within the regions encoding the catalytic domains. Based on these data, sequences encoding the NAD-binding domain in pgap11 are 2.3-fold (69:30) more diverged from the genes in pgap491 or pgap63 than are the latter genes from each other. In the case of sequences encoding the catalytic domains, the gene in pgap11 is 1.7-fold (40:23) more diverged from those in pgap491 or pgap63 than are the latter genes from each other. In both cases, the data predict that the genes contained in pgap491 and pgap63 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 structural 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 pgap11 and pgap63 predict a methionine residue at position 178 while the gene in pgap491 predicts leucine at this position. Within this region, the sequences of the genes in pgap491 and pgap63 differ by 11 silent third position codon changes and a single first position codon change at codon 178. The gene contained in pgap11 differs within this region from the genes in pgap491 and pgap63 by 13 and 5 nucleotides, respectively. Interestingly, these data predict that the genes in pgap11 and pgap63 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 pgap11 and pgap63 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 pgap11 and pgap63.

The 5'-noncoding region of the gene in pgap11 is homolo-
gous to the genes contained in pgap491 and pgap65 within the regions adjacent to the translational initiation codons. This region of homology among the three genes appears to have been tandemly duplicated in pgap11. No homology between the 3' non-coding portion of the gene in pgap11 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' non-coding 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: AAAAAACCAAGCA where the underlined region indicates the mapped termini of the mRNAs. The corresponding regions in ADC1 is AATATTCAAGCTATACCAAGCATAC. 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 AAAAAA--AG-TACT is present in pgap11 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. 88, 498-499
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., Rossman, M. G., and Harris, J. I. (1975) J. Biol. Chem. 250, 9137-9162.
10. Olsen, K. W., Moras, D., Rossman, 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. Zaikin, H., and Yanofsky, C. (1982) J. Biol. Chem. 257, 1491-1500
15. Benetzen, J.-L., and Hall, B. D. (1985) J. Biol. Chem. 257, 3018-3025
16. Young, T., Williamson, V., Taguchi, A., Smithe, M., Sledziewski, A., Russell, D., Osterman, J., Denis, C., Cox, D., and Beier, D. (1981) in Genetic Engineering of Microorganisms for Chemicals (Hollander, 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.
Enolase and Glyceraldehyde-3-phosphate Dehydrogenase Gene Expression

SALUTARY MATERIAL

To understand nucleotide sequences at the 5'-terminal of Messenger RNA molecules, we have developed an approach based on affinity chromatography using Detergent-solubilized Detergent-sensitive gene families, specifically enolase and glyceraldehyde-3-phosphate dehydrogenase genes.

BY

Julie F. Holland, Laura Lai, George D. Sambrook, and Michael J. Holland

EXPERIMENTAL PROCEDURES

Materials and Chemicals. (Phosphate-buffered saline (PBS) was purchased from New England Nuclear. Restriction endonucleases Pst I, Ava II, Bam HI, Sal I, Bcl I, and Bgl II were obtained from New England Nuclear. Recombinant DNA and cloning equipment were obtained from Amersham. DNA polymers were purchased from Pascall. Photographs were taken by D. W. T. Holland.)

Cell cultures were grown by usual methods. The hybridoma cells were grown in Medium 199 and were maintained in a humidified atmosphere at 37°C in a 95% air and 5% CO2 mixture.

Preparation of plasmid DNA. The hybridoma cells were grown in Medium 199 and were maintained in a humidified atmosphere at 37°C in a 95% air and 5% CO2 mixture. The cells were harvested by centrifugation at 600 g for 10 min at 4°C. The cell pellets were lysed in 100 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM MgCl2) containing 0.5% Triton X-100 by gentle agitation on a vortex mixer. The lysates were then centrifuged at 10,000 g for 30 min at 4°C. The supernatants were then used as the source of DNA.

Preparation of RNA. The hybridoma cells were grown in Medium 199 and were maintained in a humidified atmosphere at 37°C in a 95% air and 5% CO2 mixture. The cells were harvested by centrifugation at 600 g for 10 min at 4°C. The cell pellets were lysed in 100 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM MgCl2) containing 0.5% Triton X-100 by gentle agitation on a vortex mixer. The lysates were then centrifuged at 10,000 g for 30 min at 4°C. The supernatants were then used as the source of RNA.

Preparation of Protein. The hybridoma cells were grown in Medium 199 and were maintained in a humidified atmosphere at 37°C in a 95% air and 5% CO2 mixture. The cells were harvested by centrifugation at 600 g for 10 min at 4°C. The cell pellets were lysed in 100 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM MgCl2) containing 0.5% Triton X-100 by gentle agitation on a vortex mixer. The lysates were then centrifuged at 10,000 g for 30 min at 4°C. The supernatants were then used as the source of protein.

Preparation of Precipitated Protein. The hybridoma cells were grown in Medium 199 and were maintained in a humidified atmosphere at 37°C in a 95% air and 5% CO2 mixture. The cells were harvested by centrifugation at 600 g for 10 min at 4°C. The cell pellets were lysed in 100 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM MgCl2) containing 0.5% Triton X-100 by gentle agitation on a vortex mixer. The lysates were then centrifuged at 10,000 g for 30 min at 4°C. The supernatants were then used as the source of precipitated protein.

Preparation of Antisera. The hybridoma cells were grown in Medium 199 and were maintained in a humidified atmosphere at 37°C in a 95% air and 5% CO2 mixture. The cells were harvested by centrifugation at 600 g for 10 min at 4°C. The cell pellets were lysed in 100 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM MgCl2) containing 0.5% Triton X-100 by gentle agitation on a vortex mixer. The lysates were then centrifuged at 10,000 g for 30 min at 4°C. The supernatants were then used as the source of antisera.

Preparation of Affinity-purified Antibodies. The hybridoma cells were grown in Medium 199 and were maintained in a humidified atmosphere at 37°C in a 95% air and 5% CO2 mixture. The cells were harvested by centrifugation at 600 g for 10 min at 4°C. The cell pellets were lysed in 100 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM MgCl2) containing 0.5% Triton X-100 by gentle agitation on a vortex mixer. The lysates were then centrifuged at 10,000 g for 30 min at 4°C. The supernatants were then used as the source of affinity-purified antibodies.

Preparation of Cell lysates. The hybridoma cells were grown in Medium 199 and were maintained in a humidified atmosphere at 37°C in a 95% air and 5% CO2 mixture. The cells were harvested by centrifugation at 600 g for 10 min at 4°C. The cell pellets were lysed in 100 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM MgCl2) containing 0.5% Triton X-100 by gentle agitation on a vortex mixer. The lysates were then centrifuged at 10,000 g for 30 min at 4°C. The supernatants were then used as the source of cell lysates.

TABLE 1

Restriction Enzyme Map of Hybridoma DNA Fragments

| Restriction Enzyme | Fragment | Molecular Weight (kb) |
|---------------------|----------|----------------------|
| Pst I               | II       | 0.85                  |
| Pst I               | III      | 0.65                  |
| Pst I               | IV       | 0.45                  |

REFERENCES

1. Holland, M. E. and Holland, J. F. (1979). J. Biol. Chem. 254, 6043-6046.

2. Holland, M. E. and Holland, J. F. (1981). Mol. Cell. Biol. 1, 179-187.

3. Holland, M. E. and Holland, J. F. (1983). Cell. 30, 511-518.

4. Holland, M. E. and Holland, J. F. (1986). Mol. Cell. Biol. 6, 177-187.

5. Holland, M. E. and Holland, J. F. (1989). Mol. Cell. Biol. 9, 1089-1099.

6. Holland, M. E. and Holland, J. F. (1990). Mol. Cell. Biol. 10, 1249-1257.

7. Holland, M. E. and Holland, J. F. (1991). Mol. Cell. Biol. 11, 1249-1257.

8. Holland, M. E. and Holland, J. F. (1992). Mol. Cell. Biol. 12, 1249-1257.

9. Holland, M. E. and Holland, J. F. (1993). Mol. Cell. Biol. 13, 1249-1257.

10. Holland, M. E. and Holland, J. F. (1994). Mol. Cell. Biol. 14, 1249-1257.

11. Holland, M. E. and Holland, J. F. (1995). Mol. Cell. Biol. 15, 1249-1257.

12. Holland, M. E. and Holland, J. F. (1996). Mol. Cell. Biol. 16, 1249-1257.

13. Holland, M. E. and Holland, J. F. (1997). Mol. Cell. Biol. 17, 1249-1257.
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.

Access the most updated version of this article at http://www.jbc.org/content/258/8/5291

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/258/8/5291.full.html#ref-list-1