The Primary Structures of Two Yeast Enolase Genes

HOMOLOGY BETWEEN THE 5' NONCODING FLANKING REGIONS OF YEAST ENOLASE AND GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE GENES*

Michael J. Holland‡, Janice P. Holland, Gregory P. Thill§, and Kimberly A. Jackson

From the Department of Biochemistry, The University of Connecticut Health Center, Farmington, Connecticut 06032

(Received for publication, June 30, 1980, and in revised form, September 4, 1980)

Segments of yeast genomic DNA containing two enolase structural genes have been isolated by subculture cloning procedures using a cDNA hybridization probe synthesized from purified yeast enolase mRNA. Based on restriction endonuclease and transcriptional maps of these two segments of yeast DNA, each hybrid plasmid contains a region of extensive nucleotide sequence homology which forms hybrids with the cDNA probe. The DNA sequences which flank this homologous region in the two hybrid plasmids are nonhomologous indicating that these segments are nontandemly repeated in the yeast genome. The complete nucleotide sequence of the coding as well as the flanking noncoding regions of these genes has been determined. The amino acid sequence predicted from one reading frame of both structural genes is extremely similar to that determined for yeast enolase (Chin, C. C. Q., Brewer, J. M., Eckard, E., and Wold, F. (1981) J. Biol. Chem. 256, 1370-1376), confirming that these isolated structural genes encode yeast enolase. The nucleotide sequences of the coding regions of the genes are approximately 95% homologous, and neither gene contains an intervening sequence. Codon utilization in the enolase genes follows the same biased pattern previously described for two yeast glyceraldehyde-3-phosphate dehydrogenase structural genes (Holland, J. P., and Holland, M. J. (1980) J. Biol. Chem. 255, 2596-2605). DNA blotting analysis confirmed that the isolated segments of yeast DNA are colinear with yeast genomic DNA and that there are two nontandemly repeated enolase genes per haploid yeast genome. The noncoding portions of the two enolase genes adjacent to the initiation and termination codons are approximately 70% homologous and contain sequences thought to be involved in the synthesis and processing of messenger RNA. Finally there are regions of extensive homology between the two enolase structural genes and two yeast glyceraldehyde-3-phosphate dehydrogenase structural genes within the 5' noncoding portions of these glycolytic genes.

Enolase prepared from a wide variety of eucaryotic cells appears to exist as multiple cellular forms (1). These results have led to the hypothesis that the structural gene for enolase

* This research was supported by a grant from The National Institute of General Medical Sciences 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.

‡ American Heart Association Established Investigator.
§ Present address, Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205.

is repeated and that some of these multiple forms of the native dimer are isozymes. Multiple yeast enolases have been resolved by anion exchange chromatography (2-4), gel electrophoresis (5), isoelectric focusing (6), and countercurrent distribution techniques (7, 8). Westhead and McLain (2) originally showed that the multiple forms of yeast enolase resolved by O-(triethylaminoethyl)-Sephadex chromatography arise as a consequence of aging cell extracts, and they suggested that these forms of the enzyme may arise as a consequence of in vitro deamidation of the enzyme. Porcelli et al. (4) have recently shown that the major form of yeast enolase resolved after O-(triethylaminoethyl)-Sephadex chromatography is the precursor to the two minor forms of the enzyme. These latter investigators suggest that the minor forms arise by a process which involves dissociation of the native dimer. Based on these data it is clear that some portion of the heterogeneity observed in preparations of yeast enolase is the result of in vitro modifications. It is not clear, however, if all multiple forms of yeast enolase correspond to artifacts of isolation.

Enolase messenger RNA is one of the most abundant mRNAs in vegetative yeast cells (9). Previous reports from this laboratory have shown that glyceraldehyde-3-phosphate dehydrogenase mRNA, an abundant yeast mRNA, can be obtained in sufficient purity to permit the cloning of its respective structural genes (10). We have used a similar approach to isolate and identify enolase structural genes utilizing a cDNA probe synthesized from partially purified enolase mRNA. We report the isolation and characterization of two nontandemly repeated yeast enolase structural genes. The primary structures of the polypeptides encoded by these two enolase genes differ in 20 out of 436 amino acids. The primary structure of one of these polypeptides is in close agreement to that determined by Chin et al. (17) for the major "A" form of yeast enolase resolved after O-(triethylaminoethyl)-Sephadex chromatography. Evidence is presented that both enolase structural genes are expressed in vegetative yeast cells suggesting that isozymes of yeast enolase do exist in the cell.

The primary structures of the two enolase genes are 95% homologous within their respective coding regions, and the primary structures of the 5' and 3' noncoding regions adjacent to the translational initiation and termination codons are approximately 70% homologous. Comparison of the 5' and 3' noncoding regions of the two enolase genes with two nontandemly repeated yeast glyceraldehyde-3-phosphate dehydrogenase structural genes (11, 12) shows little if any homology among the four genes within the 3' noncoding regions; however, the 5' noncoding regions of all four glycolytic genes are extremely homologous within the first 100 nucleotides adjacent to the translational initiation codons. These latter results suggest that the 5' noncoding regions of enolase and glyceraldehyde-3-phosphate dehydrogenase structural genes may have evolved from a common nucleotide sequence.
EXPERIMENTAL PROCEDURES

The miniprint supplement contains all experimental details.1

RESULTS

Subculture Cloning of pen08 and pen046—Segments of yeast DNA containing sequences complementary to yeast enolase mRNA were isolated by subculture cloning procedures as described in the miniprint supplement. The results of the screening experiments are illustrated in Table I. The subculture cloning procedure used to identify transformants containing yeast enolase structural genes was the same as that previously described for the isolation of a yeast glyceraldehyde-3-phosphate dehydrogenase structural gene (10). A series of cultures was started with a defined number of Escherichia coli transformants containing hybrid plasmids composed of the bacterial vector pSP2124 and randomly sheared segments of yeast genomic DNA. The yeast DNA was joined into the Eco RI restriction endonuclease cleavage site by the A/T joining procedure as previously described (10). Total hybrid plasmid DNA was isolated from each culture, depurinated, and hybridized against a cDNA probe synthesized from partially purified yeast enolase mRNA. Hybridization reactions were carried out in solution with a vast excess of depurinated yeast DNA. Hybridization was carried out to saturation of the probe, and the percentage of the labeled cDNA hybrid plasmid DNA was isolated from each culture, depurinated, and hybridized against a cDNA probe synthesized from partially purified enolase mRNA. Hybridization was carried out to saturation of the probe, and the percentage of the labeled cDNA hybrid plasmid DNA whereas pen046 plasmid DNA hybridized to the same extent as pen08 DNA alone confirming that pen08 contains sequences which are complementary to a segment of the map in Fig. 1. The nucleotide sequences which flank this region of homology are not homologous indicating that the two hybrid plasmids contain different portions of yeast genomic DNA.

In order to establish that the homologous segments of the plasmids contain sequences which hybridize with the cDNA, a series of DNA blotting experiments was carried out. Initially, yeast genomic DNA was digested with the restriction endonucleases Xba I and Xho I, electrophoresed on a 0.8% agarose slab gel, transferred to nitrocellulose filter paper, and hybridized with purified enolase cDNA as described in the miniprint supplement. Two Xba I fragments (12 kb1 and 6.4 kb) and two Xho I fragments (12 kb and 10 kb) formed hybrids with the cDNA, respectively. Since the isolated plasmid DNAs are not cleaved by Xba I or Xho I, we conclude that sequences in pen08 and pen046 which form hybrids with enolase mRNA correspond with sequences in yeast genomic DNA. The fact that the isolated sequences in pen08 and pen046 which form hybrids with enolase cDNA correspond with sequences in yeast genomic DNA

1 The abbreviation used is: kb, kilobase pair.

| Plasmid DNA |
|-------------|
| pen08       |
| pen046      |
| pen08 plus pen046 |

| Transformants/culture | % Hybridization |
|-----------------------|-----------------|
|                       | Negative | Positive |
| pen08                 | 400      | 20       | 41       |
|                        | 100      | 13       | 38       |
|                        | 12       | 14       | 27       |
|                        | 1        | 4        | 22       |
| pen046                |          |          |          |
|                        | 600      | 14       | 40       |
|                        | 41       | 9        | 26       |
|                        | 6        | 13       | 25       |
|                        | 1        | 15       | 27       |

In order to establish that the homologous segments of the plasmids contain sequences which hybridize with the cDNA, a series of DNA blotting experiments was carried out. Initially, yeast genomic DNA was digested with the restriction endonucleases Xba I and Xho I, electrophoresed on a 0.8% agarose slab gel, transferred to nitrocellulose filter paper, and hybridized with purified enolase cDNA as described in the miniprint supplement. Two Xba I fragments (12 kb and 6.4 kb) and two Xho I fragments (12 kb and 10 kb) formed hybrids with the cDNA, respectively. Since the isolated plasmid DNAs are not cleaved by Xba I or Xho I, we conclude that sequences in pen08 and pen046 which form hybrids with enolase mRNA correspond with sequences in yeast genomic DNA.
FIG. 1. Restriction endonuclease cleavage maps of pen08 and pen046. The single horizontal lines indicate pSF2124 vector DNA sequences and the double horizontal lines indicate the yeast DNA portions of the hybrid plasmids. The shaded regions indicate the location of the enolase structural genes in the plasmids, and the arrow indicates the direction of transcription of the enolase structural genes. These cleavage maps were derived from the data reported in the miniprint supplement.

FIG. 2. Hybridization of DNA filter blots of restriction endonuclease cleavage fragments of pen08, pen046, and yeast genomic DNA with an enolase cDNA probe. Lanes A and B are agarose gels of yeast genomic DNA (strain F1) digested with XbaI and XhoI, respectively, visualized after ethidium bromide staining. Lanes a and b are autoradiograms of hybrids formed between 32P-labeled enolase cDNA and the XbaI and XhoI digested yeast DNA, respectively, after transfer to a nitrocellulose filter. Lanes C and D are agarose gels of yeast genomic DNA digested with HindIII and EcoRI, respectively, and lanes c and d are autoradiograms of hybrids formed between enolase cDNA and the DNA fragments shown in lanes C and D, respectively. Lane E, an agarose gel of pen046 digested with HindIII; lane f, an autoradiogram of hybrids formed between 32P-labeled enolase cDNA and HindIII-digested pen046; lane F, an agarose gel of pen08 digested with HindIII; lane g, an autoradiogram of hybrids formed between 32P-labeled enolase cDNA and HindIII-digested pen08; lane G, an autoradiogram of hybrids formed between 32P-labeled enolase cDNA and EcoRI-digested pen046; lane H, an autoradiogram of hybrids formed between 32P-labeled enolase cDNA and EcoRI-digested pen08. Restriction endonuclease cleavage fragments were electrophoresed on agarose gels and transferred to a nitrocellulose filter before hybridization. Enolase cDNA was purified prior to hybridization as described in the miniprint supplement.

FIG. 3. Hybridization of yeast genomic DNA digested with XbaI with an enolase cDNA probe. Genomic yeast DNA isolated from commercially grown strain F1 (lane A), strain F1 grown in YPD media (lane B) and strain D273-10B grown in YPD media (lane C) digested with XbaI, electrophoresed on a 1% agarose slab gel, transferred to a nitrocellulose filter, and hybridized with a 32P-labeled enolase cDNA. The enolase cDNA was purified prior to hybridization as described in the miniprint supplement.
A.

Yeast Enolase Genes

MW(kb)

5.0-

3.0-

1.0-

0.5-

0.2-

12 3 4 5 9 10 11 12 14

B.

pm

pm 8

pm 46

Alu I

Hha I

AvaII

HpaI

Eco RI

HpaII

Hae III

MboI

Hinc II

Tac I

HindIII

Taq I

Hinf I

FIG. 4. Restriction endonuclease cleavage maps of the enolase structural gene portions of peno8 and peno46. A, an autoradiogram of a 1.5% agarose slab gel of partial restriction endonuclease digests of 4.4-kb and 2.0-kb fragments from peno8 and peno46, respectively. Each fragment was 32P-labeled at the 5' terminus of a common BglII cleavage site within each respective enolase structural gene. The 4.4-kb fragment from peno8 extends from the BglII cleavage site to a PvuII cleavage site within a right enolase structural gene (Fig. 1). The 2.0-kb fragment from peno46 extends from the BglII cleavage site to the XbaI cleavage site within the hybrid plasmid. Partial digests of these fragments with the restriction endonucleases indicated were electrophoresed in parallel lanes of the slab gel. The fragments derived from peno8 and peno46 are in the left and right parallel lanes, respectively. B, the cleavage map predicted from the partial digests shown in A as well as the data described in the text. The numbers below the parallel lanes in A correspond to the restriction endonuclease key in B.

supports the conclusion that the yeast DNA sequences in the hybrid plasmids are colinear with genomic DNA sequences.

The DNA blotting analysis shown in Fig. 2 was carried out with DNA isolated from commercially grown baker's yeast (strain F1). Similar DNA blots were carried out with DNA isolated from the haploid Saccharomyces cerevisiae strain D273-10B and strain F1 grown under the same conditions as the haploid strain. As illustrated in Fig. 3, the DNA blots with DNA isolated from haploid and commercial baker's yeast after digestion with the restriction endonuclease Xba I demonstrate a 12-kb and a 6.4-kb fragment which form hybrids with enolase cDNA. These data confirm that there are two nontandemly repeated enolase structural genes per haploid yeast genome. Furthermore, the number and relative location of the two genes in the genome of strain F1 is not influenced by the growth conditions for propagating the cells.

Fig. 4 shows a more detailed restriction endonuclease cleavage map of the homologous regions of peno8 and peno46. The restriction endonuclease cleavage map was generated as described in the miniprint supplement utilizing the partial cleavage method described by Smith and Birnstiel (13). The molecular weights of each of the restriction endonuclease cleavage fragments were determined from limit digests with each of the restriction endonucleases tested. Fig. 4A shows a series of partial digests of two DNA fragments generated by digestion of peno8 with BglII and PvuII (4.4 kb) and by digestion of peno46 with BglII and XbaI (2.0 kb). Both fragments were labeled at the common BglII cleavage site with [α-32P]ATP
and polynucleotide kinase as described in the miniprint supplement. As illustrated by the partial cleavages and the cleavage map shown in Fig. 4, A and B, the DNA sequences surrounding the common Bgl II cleavage site are extremely homologous in the two hybrid plasmids. The region of nucleotide sequence homology between the two plasmids is the approximate size needed to encode yeast enolase. This observation is confirmed by the nucleotide sequence described below. As initially pointed out from the data shown in Fig. 1, the nucleotide sequences which flank the homologous portions of the plasmids are nonhomologous indicating that the genes are nonallelic and are nontandemly repeated in the yeast genome. Based on the restriction endonuclease map (Fig. 4B) approximately 62% of the cleavage sites are in common between the two genes. These data, therefore, suggest that the polypeptides encoded by these two genes have diverged in primary structure.

The Nucleotide Sequence of the Coding Portions of peno8 and peno46—The complete nucleotide sequence of the coding regions of peno8 and peno46 is shown in Fig. 5. The nucleotide sequence was determined by the method of Maxam and Gilbert (14) as described in the miniprint supplement using the strategy outlined in Fig. 6. The continuous nucleotide sequence shown in Fig. 5 corresponds to the enolase gene in peno8. The nucleotide sequence of the coding region of peno46 is identical with that from peno46 except at the positions indicated by the codons above the continuous sequence. At these positions the entire codon in peno8 is indicated when that codon differs from the respective codon in peno46. The amino acid sequence predicted from one reading frame of peno46 is indicated below the nucleotide sequence in Fig. 5. In the case of a codon change in peno8 resulting in a change in the predicted amino acid sequence relative to peno46, the new amino acid is indicated above the peno8 codon. The polypeptides encoded by the two structural genes differ in 20 out of 436 residues. The polypeptide encoded by peno46 agrees in primary structure extremely well with the yeast enolase primary structure determined by Chin et al. (17). Based on these data, it is likely that the enolase protein which was sequenced by Chin et al. is a dimer of the polypeptide encoded by peno46. The relevance of these two gene products to the multiple forms of yeast enolase will be discussed below. These data confirm the identification of two nontandemly repeated yeast enolase structural genes contained within the hybrid plasmids described here. The primary structures of the two enolase genes are 95% homologous within the coding regions. Based on the agreement between the primary structure determined for yeast enolase and those predicted from the gene sequences, we can conclude that neither gene contains intervening sequences. There are a total of 33 silent third position changes of a total of 67 nucleotide differences between the two coding regions.

The codon utilization pattern for the two enolase structural genes is shown in Table II. The observed codon bias is extremely similar to that observed for two yeast glyceraldehyde-3-phosphate dehydrogenase structural genes (12). The major exception to the bias observed in glyceraldehyde-3-phosphate dehydrogenase is the occurrence of the UUA codon for leucine twice in peno8 and three times in peno46. The CGU codon for arginine is present at position 14 in both enolase gene sequences and was not observed in the glyceraldehyde-3-phosphate dehydrogenase structural genes. Cysteine, glutamine, glutamate, histidine, and glycine are encoded by a single codon. Arginine, asparagine, phenylalanine, proline, and tyrosine are also encoded by a single codon with only seven exceptions. Lysine and leucine are encoded by AAG and UUG codons approximately 90% of the time, and finally alanine, aspartate, isoleucine, serine, threonine, and valine are encoded by two codons which contain either U or C in third position. As was the case with the glyceraldehyde-3-phosphate dehydrogenase genes, we observe only 10 to 20% change between U and C in the third positions of the codons for the six amino acids which are encoded by two codons with either C or U in third position. This variation is significantly less than predicted for random change at these third positions,
Yeast Enolase Genes

Fig. 6. Strategy for determining the nucleotide sequences of the enolase structural genes in pen08 and pen046. The arrows indicate the direction and amount of sequence determined from each end-labeled restriction endonuclease fragment.

TABLE II
Table for Codon utilization in pen08 and pen046

The codon utilization in pen08 and pen046 was tabulated from the primary structure shown in Fig. 5. The codon usage pattern for pen08 is shown in parentheses.

| Codon | pen08 Utilization | pen046 Utilization |
|-------|------------------|-------------------|
| ALA   |                   |                   |
| Cys   |                   |                   |
| His   |                   |                   |
| Met   |                   |                   |
| Thr   |                   |                   |

The A+T composition of the first 300 nucleotides beyond the termination codon in both genes is 68%. The nucleotide sequences which are repeated in the 5' noncoding regions of the genes are indicated by the boxed regions in Fig. 7B.

Comparison of the 5' Noncoding Regions of the Two Enolase Genes with Two Yeast Glyceraldehyde-3-Phosphate Dehydrogenase Structural Genes—As described above for

suggesting that divergence of the duplicated genes for enolase relative to each other may be restricted.

Nucleotide Sequence of the 5' and 3' Noncoding Flanking Regions of pen08 and pen046—The nucleotide sequences of the 5' and 3' noncoding regions of pen08 and pen046 are shown in Fig. 7, A and B, respectively. The nucleotide sequences of the two structural genes have been aligned to maximize homology between the two genes. The alignments necessitated leaving gaps in one of the two sequences. Within the 5' noncoding sequences of the genes the A+T composition is approximately 75% for the first 150 nucleotides adjacent to the ATG initiation codon. Approximately 77% of the first 75 nucleotides adjacent to the initiation codon in pen08 are homologous to pen046. In order to maximize homology between pen08 and pen046 within this region, it is necessary to allow a 36-nucleotide deletion in the pen08 sequence. The 25 nucleotides upstream from this deletion in pen08 are greater than 85% homologous to sequences in pen046. The structural gene in pen046 contains a -TATAAA- sequence 140 nucleotides from the initiation codon. The position of these sequences is similar to the -TATAAA-sequences in two yeast glyceraldehyde-3-phosphate dehydrogenase structural genes (12) and the cytochrome C-1 gene from yeast (15). pen08 does not contain -TATAAA- sequences within 175 nucleotides of the initiation codon. As previously described for the yeast glyceraldehyde-3-phosphate dehydrogenase structural genes (12), both enolase genes contain short nucleotide sequences in the 5' noncoding regions which are repeated in noninverted fashion in the 3' noncoding portions of the respective genes. These repetitious sequences are indicated by the boxed regions in Fig. 7, A and B.

The nucleotide sequences adjacent to the termination codons in pen08 and pen046 are also extremely homologous. Both genes are approximately 65% homologous for 130 nucleotides beyond the termination codon and are 62% homologous within the 80 nucleotides surrounding a -AATAA- sequence in both genes. The -AATAA- sequence is homologous to similar sequences located in the 3' noncoding regions of many eucaryotic genes and is usually located approximately 15 to 20 nucleotides from the site of poly A addition to the mRNA. Based on DNA blotting experiments carried out with enolase cDNA probe and restriction endonuclease fragments generated from both plasmid DNAs which contain the 3' noncoding regions, we conclude that enolase mRNA sequences extend to the -AATAA- sequences (data not shown).

The A+T composition of the first 300 nucleotides beyond the termination codon in both genes is 68%. The nucleotide sequences which are repeated in the 5' noncoding regions of the genes are indicated by the boxed regions in Fig. 7B.
Fig. 7. Nucleotide sequences of the 5' and 3' noncoding flanking regions of the enolase structural genes; comparison of the corresponding regions in pen08 and pen046. A, shows the 5' noncoding flanking regions of pen08 and pen046; B, shows the 3' noncoding flanking regions of pen08 and pen046. The solid bars between the pen08 and pen046 sequences indicate homologous regions of the sequences. The boxed nucleotides are repeated in the 5' and 3' noncoding regions of pen08 and pen046, respectively. Two homologous nucleotide sequences in the 5' noncoding region of pen08 are indicated by brackets.
**Yeast Enolase Genes**

The enolase genes and for the yeast glyceraldehyde-3-phosphate dehydrogenase structural genes (12), the noncontiguously repeated genes share extensive nucleotide sequence homology within the first 100 nucleotides upstream from the translational initiation codon. We have analyzed the 5' noncoding regions of the enolase and glyceraldehyde-3-phosphate dehydrogenase genes to determine if these glycolytic genes share homology. As illustrated in Fig. 8 there is extensive nucleotide sequence homology among the enolase and glyceraldehyde-3-phosphate dehydrogenase genes within the 5' noncoding regions. Three regions of extensive nucleotide sequence homology are indicated in Fig. 8. The sequence -CACACA- is homologous among all four genes and is also found at approximately the same position relative to the translational initiation codon in the yeast cytochrome C-1 structural gene (15). Beyond this sequence is a 16-nucleotide segment of DNA which is highly conserved in the four glycolytic genes. We do not find this sequence within the 5' noncoding region of either of the yeast cytochrome c genes (16). A third region of homology is present 30 nucleotides downstream from the initiation codon in the yeast glyceraldehyde-3-phosphate dehydrogenase structural gene contained in pen08. Based on this observation we conclude that the molecular basis for these observations is related to the 3' noncoding regions of the two structural genes. Based on DNA blotting experiments with restriction endonuclease cleavage fragments generated from the 3' noncoding regions of the two structural genes, 32P-labeled enolase cDNA hybridizes with both genes approximately 200 to 300 nucleotides beyond the respective termination codons. The 3' noncoding regions of the two genes are partially homologous; however, they would not be expected to cross-hybridize under stringent hybridization conditions. Since the cDNA is initiated within the 3' noncoding regions of the mRNA and the average size of the cDNA is approximately 600 nucleotides, the 3' noncoding portions of each mRNA are highly represented in the cDNA population. We conclude, therefore, on the basis of the hybridization data shown in Fig. 8 that the mRNA sequences derived from pen046 are not as abundant in the yeast cell as those derived from the gene contained in pen08 plasmid. These data suggest that there are differences in the rates of transcription of the genes and/or the half-lives of the two mRNAs. Since the mRNA encoded by the structural gene contained in pen08 appears to be the most abundant enolase mRNA in the cell and the form of enolase which was sequenced derived from both enolase genes but rather that the protein is probably a dimer of the polypeptide encoded by the structural gene contained in pen046. The presence of two enolase structural genes per haploid yeast genome raises the possibility that true isozymes of enolase exist in the cell. As yet there is no direct evidence demonstrating the presence of the polypeptide encoded by the gene in pen08 in yeast cells, however. Based on the primary structures of the enolase structural genes, both hybrid plasmids contain all of the sequences which would be complementary to cDNA synthesized from enolase mRNA. For this reason it is necessary to consider the fact that pen046 and pen08 do not hybridize to the same extent with the cDNA and that a mixture of the two hybrid plasmid DNAs hybridizes to the same extent as pen08 alone. We believe that the molecular basis for these observations is related to the 3' noncoding regions of the two structural genes. Based on DNA blotting experiments with restriction endonuclease cleavage fragments generated from the 3' noncoding regions of the two structural genes, 32P-labeled enolase cDNA hybridizes with both genes approximately 200 to 300 nucleotides beyond the respective termination codons. The 3' noncoding regions of the two genes are partially homologous; however, they would not be expected to cross-hybridize under stringent hybridization conditions. Since the cDNA is initiated within the 3' noncoding regions of the mRNA and the average size of the cDNA is approximately 600 nucleotides, the 3' noncoding portions of each mRNA are highly represented in the cDNA population. We conclude, therefore, on the basis of the hybridization data shown in Fig. 8 that the mRNA sequences derived from pen046 are not as abundant in the yeast cell as those derived from the gene contained in pen08 plasmid. These data suggest that there are differences in the rates of transcription of the genes and/or the half-lives of the two mRNAs. Since the mRNA encoded by the structural gene contained in pen08 appears to be the most abundant enolase mRNA in the cell and the form of enolase which was sequenced derived from both enolase genes but rather that the protein is probably a dimer of the polypeptide encoded by the gene contained in pen046 it is likely that both enolase genes are expressed during vegetative yeast cell growth. The coding regions of the two enolase structural genes are 95% homologous as are the coding regions of two yeast glyceraldehyde-3-phosphate dehydrogenase structural genes (12) indicating that these two duplicated glycolytic genes have diverged to the same extent. Divergence of the amino acid sequences of the polypeptides encoded by the enolase and glyceraldehyde-3-phosphate dehydrogenase genes is also 5%. The codon usage pattern in the enolase genes is extremely similar to the biased usage pattern previously reported for the two yeast glyceraldehyde-3-phosphate dehydrogenase genes (12). Comparison of the codon usage patterns of all four glycolytic genes reveals that cysteine, glutamine, glutamate, histidine, and glycine are encoded by a single codon while alanine, aspartate, isoleucine, serine, threonine, and valine are.

### DISCUSSION

Two segments of yeast genomic DNA have been isolated by subculture cloning techniques which form hybrids with cDNA synthesized from partially purified yeast enolase messenger RNA. Based on the primary structures of these isolated hybrid plasmids each segment of DNA encodes a polypeptide with an amino acid sequence which is in close agreement with that determined by Chin et al. (17) for a chromatographically homogeneous form of yeast enolase. The primary structure of the polypeptide encoded by one of the isolated hybrid plasmids (pen046) is identical with that determined by Chin et al. (17) in all but 9 residues. Careful examination of all of the nucleic acid and protein sequence data (17) suggests that four or five of these differences in amino acid sequence may in fact correspond to real differences in the primary structures of the enolase polypeptides in the yeast strains from which the protein and the genes were isolated. The primary structures of the enolase polypeptides encoded by the two isolated structural genes differ in 20 of 436 residues. Interestingly, none of the discrepancies between the amino acid sequence of yeast enolase and that predicted from the structural gene contained in pen046 can be reconciled with the polypeptide sequence predicted from the gene contained in pen08. Based on this observation we conclude that the enolase dimer which was sequenced by Chin et al. (17) is not a mixture of polypeptides derived from both enolase genes but rather that the protein is probably a dimer of the polypeptide encoded by the structural gene contained in pen046. The presence of two enolase structural genes per haploid yeast genome raises the possibility that true isozymes of enolase exist in the cell. As yet there is no direct evidence demonstrating the presence of the polypeptide encoded by the gene in pen08 in yeast cells, however. Based on the primary structures of the enolase structural genes, both hybrid plasmids contain all of the sequences which would be complementary to cDNA synthesized from enolase mRNA. For this reason it is necessary to consider the fact that pen046 and pen08 do not hybridize to the same extent with the cDNA and that a mixture of the two hybrid plasmid DNAs hybridizes to the same extent as pen08 alone. We believe that the molecular basis for these observations is related to the 3' noncoding regions of the two structural genes. Based on DNA blotting experiments with restriction endonuclease cleavage fragments generated from the 3' noncoding regions of the two structural genes, 32P-labeled enolase cDNA hybridizes with both genes approximately 200 to 300 nucleotides beyond the respective termination codons. The 3' noncoding regions of the two genes are partially homologous; however, they would not be expected to cross-hybridize under stringent hybridization conditions. Since the cDNA is initiated within the 3' noncoding regions of the mRNA and the average size of the cDNA is approximately 600 nucleotides, the 3' noncoding portions of each mRNA are highly represented in the cDNA population. We conclude, therefore, on the basis of the hybridization data shown in Fig. 8 that the mRNA sequences derived from pen046 are not as abundant in the yeast cell as those derived from the gene contained in pen08 plasmid. These data suggest that there are differences in the rates of transcription of the genes and/or the half-lives of the two mRNAs. Since the mRNA encoded by the structural gene contained in pen08 appears to be the most abundant enolase mRNA in the cell and the form of enolase which was sequenced derived from both enolase genes but rather that the protein is probably a dimer of the polypeptide encoded by the gene contained in pen046 it is likely that both enolase genes are expressed during vegetative yeast cell growth. The coding regions of the two enolase structural genes are 95% homologous as are the coding regions of two yeast glyceraldehyde-3-phosphate dehydrogenase structural genes (12) indicating that these two duplicated glycolytic genes have diverged to the same extent. Divergence of the amino acid sequences of the polypeptides encoded by the enolase and glyceraldehyde-3-phosphate dehydrogenase genes is also 5%. The codon usage pattern in the enolase genes is extremely similar to the biased usage pattern previously reported for the two yeast glyceraldehyde-3-phosphate dehydrogenase genes (12). Comparison of the codon usage patterns of all four glycolytic genes reveals that cysteine, glutamine, glutamate, histidine, and glycine are encoded by a single codon while alanine, aspartate, isoleucine, serine, threonine, and valine are.

---

3 M. J. Holland and J. P. Holland, unpublished observations.
are exclusively encoded by two codons which contain either C or U in third position. Arginine, asparagine, phenylalanine, proline, and tyrosine are encoded by a single codon with only exceptions out of 258 codons. Ninety per cent of the lysines are encoded by AAG, and 95% of the leucines are encoded by UUG. The most abundant codons utilized in the yeast cytochrome c genes are the same as those observed in the glycolytic genes; however, the degree of the bias in the cytochrome c genes is significantly lower (16). The molecular basis for the observed codon bias remains unclear; however, it may be relevant to the efficiency of translation of the glycolytic mRNA. This hypothesis is consistent with the fact that enolase and glyceraldehyde-3-phosphate dehydrogenase are among the most abundant enzymes in yeast cells (2, 9, 18).

Based on the divergence of the nucleotide sequences of the repeated enolase and glyceraldehyde-3-phosphate dehydrogenase genes it is possible to estimate the age of the gene duplication. Utilizing the estimated rate of divergence of glyceraldehyde-3-phosphate dehydrogenase, for example (1% amino acid sequence divergence per 20 million years (19)) one would predict a duplication event which is approximately 100 million years old. This estimate is only valid, however, if the repeated yeast genes evolve independently. In the case of tandemly repeated genes there is considerable evidence that gene repeats do not evolve independently. Recent studies on the yeast ribosomal cistrons suggest that conservation of the primary structures of each cistron within the tandemly repeated array is achieved by unequal crossing over between tandem repeats during mitosis (20). These crossing over events result in conservation of the primary structure of each cistron, and as a consequence the entire gene repetition evolves in concert. Concerted evolution of the tandemly repeated genes which code for the a chains of hemoglobin has also been proposed as an explanation for the fact that the a genes in a single primate species diverge in primary structure at one tenth the rate observed among different primate species (21).

Since the yeast enolase and glyceraldehyde-3-phosphate dehydrogenase genes are not tandemly repeated one might argue that the genes do not recombine during evolution. Alternatively, these genes may recombine by mechanisms which are different from those which occur for tandemly repeated genes. Multiple forms of enolase and glyceraldehyde-3-phosphate dehydrogenase have been resolved from a wide variety of eucaryotic cells (1, 22) suggesting that multiple genes for these glycolytic enzymes are present in these cells. In the case of glyceraldehyde-3-phosphate dehydrogenase the primary structures of the enzymes from yeast, lobster, and pig are 60% homologous (23). If duplicated glyceraldehyde-3-phosphate dehydrogenase genes exist in a common precursor to these eucaryotes, the estimated age of the duplication event based on the primary structures of the two yeast genes is a significant underestimation. Given these considerations, concerted evolution of the yeast genes cannot be ruled out. Concerted evolution of the yeast enolase and glyceraldehyde-3-phosphate dehydrogenase gene repeats would provide an explanation for the fact that divergence at the third position of the codons for the six amino acids which are encoded by two codons is less than 20% of that predicted for random drift at these positions. Such a mechanism would also explain the fact that unusual codons such as the CGU arginine codon at residue 14 in the enolase genes are conserved at the same position in both genes and that the 5' and 3' noncoding flanking regions of the gene repeats are 70% homologous.

The 5' and 3' noncoding flanking regions of the two enolase structural genes are approximately 70% homologous for 100 nucleotides upstream from the translational initiation codon and 280 nucleotides beyond the translational termination codon. Similar homology was observed in the 5' and 3' noncoding flanking regions of two yeast glyceraldehyde-3-phosphate dehydrogenase genes (12) and the genes which encode the ß chains of mouse hemoglobin (24). Maximizing homology between the enolase genes within the noncoding regions eliminates leaving gaps in one of the gene sequences. It is assumed that these gaps arise as a consequence of deletions and/or insertions within one of the gene sequences. Similar gaps are generated when the yeast glyceraldehyde-3-phosphate dehydrogenase gene (12) and mouse ß globin gene sequences are aligned (24). The gene contained in peno46 contains a TATAAA sequence 140 nucleotides upstream from the initiation codon while this hexanucleotide is not present in the 5' noncoding region of the gene contained in peno8. The TATAAA sequence was found in the 5' noncoding regions of both of the yeast glyceraldehyde-3-phosphate dehydrogenase genes (12) and in one of the yeast cytochrome c genes (18). Both enolase genes contain an AATAA sequence 220 to 240 nucleotides beyond the respective termination codons. Finally, short nucleotide sequence repetitions are observed in the 5' and 3' noncoding regions of the two enolase genes. Similar noninverted sequence repetitions were observed in the two yeast glyceraldehyde-3-phosphate dehydrogenase genes (15); however, these latter sequences are not homologous to those found in the enolase genes.

The 5' and 3' noncoding flanking regions of the enolase and glyceraldehyde-3-phosphate dehydrogenase (12) structural genes were compared to determine if homologous sequences were shared by these two glycolytic gene repeats. Little if any significant homology was found within the 3' noncoding regions of the genes; however, the 5' noncoding regions of all four genes are extremely similar for approximately 100 nucleotides upstream from the respective translational initiation codons. As illustrated in Fig. 8, all four genes contain three regions of extensive nucleotide sequence homology. There is a hexanucleotide adjacent to the initiation codon which is homologous in all four genes and is also present at a similar position in the yeast cytochrome c-1 gene (15). Beyond the hexanucleotide is a 16-nucleotide long region which is shared by the four glycolytic genes but is not found in the yeast cytochrome c genes (16). A third region of extensive homology is located approximately 90 nucleotides upstream from the initiation codons in the two enolase genes and the glyceraldehyde-3-phosphate dehydrogenase gene contained in the pgap490 hybrid plasmid (11). This latter sequence is located 30 nucleotides downstream from the TATAAA sequence in the genes contained in the peno46 and pgap491 hybrid plasmids. This sequence is repeated twice in the enolase gene contained in peno8, 49 and 92 nucleotides upstream from the translational initiation codon. The nucleotide sequences of the enolase and glyceraldehyde-3-phosphate dehydrogenase 5' noncoding regions are also homologous within the regions between the sequences described above suggesting that the entire 5' noncoding region of four genes evolved from a common nucleotide sequence. Since enolase and glyceraldehyde-3-phosphate dehydrogenase are nonhomologous proteins, this common nucleotide sequence must have been fused to the coding regions of these two glycolytic genes during evolution. The similarities in nucleotide sequences within the 5' noncoding regions of enolase and glyceraldehyde-3-phosphate dehydrogenase genes are in marked contrast to the fact that the 5' noncoding regions of the coordinately expressed mouse ß and ß globin structural genes are nonhomologous (24, 25). While the structures of the mouse globin genes suggest that conservation of nucleotide sequence is not required for coordinate expression it is tempting to speculate that the homologous 5' noncoding regions of the yeast enolase
and glyceraldehyde-3-phosphate dehydrogenase genes are involved in coordinate expression of these glycolytic genes. This possibility can now be tested by in vitro alteration of these sequences and subsequent analysis of the expression of these altered genes in vivo.

Acknowledgments—We would like to express our appreciation to Drs. Chin, Brewer, and Wold for making available to us their unpublished amino acid sequence of yeast enolase and for helpful discussions. We are indebted to Ms. Laura Labieniec for excellent technical assistance.

REFERENCES

1. Wold, F. (1971) in The Enzymes (Boyer, P. D., ed) 3rd Ed. pp. 499–538, Academic Press, New York
2. Westhead, E. W., and McLain, G. (1964) J. Biol. Chem. 239, 2464–2468
3. Oh, S.-K., Travis, J., and Brewer, J. M. (1973) Biochim. Biophys. Acta 310, 421–429
4. Porcelli, L. J., Jr, Small, E. D., and Brewer, J. M. (1978) Biochem. Biophys. Res. Commun. 82, 316–321
5. Pfleiderer, G., Neufahr-Kreiling, A., Kaplan, R. W., and Fortnagel, P. (1966) Biochem. Z. 346, 289–281
6. Susor, W. A., Kochman, M., and Rutter, W. J. (1969) Science 165, 1260–1262
7. Shanbhag, V., Blomquist, G., Johansson, G., and Hartman, A. (1972) FEBS Lett. 22, 105–108
8. Blomquist, G. (1976) Biochim. Biophys. Acta 420, 81–86
9. Holland, M. J., and Holland, J. P. (1978) Biochemistry 17, 4900–4907
10. Holland, M. J., and Holland, J. P. (1979) J. Biol. Chem. 254, 5466–5474
11. Holland, J. P., and Holland, M. J. (1979) J. Biol. Chem. 254, 9839–9845
12. Holland, J. P., and Holland, M. J. (1980) J. Biol. Chem. 255, 2596–2605
13. Smith, H. O., and Birnstiel, M. L. (1976) Nucleic Acids Res. 3, 2387–2398
14. Maxam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560–564
15. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L., and Hall, B. D. (1979) Cell 16, 753–761
16. Montgomery, D. L., Leung, D. W., Smith, M., Shalit, P., Faye, G., and Hall, B. D. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 541–545
17. Chin, C. C. Q., Brewer, J. M., Eckard, E., and Wold, F. (1981) J. Biol. Chem. 256, 1379–1376
18. Krebs, E. G. (1953) J. Biol. Chem. 200, 471–478
19. Wilson, A. C., Carlson, S. S., and White, J. (1977) Annu. Rev. Biochem. 46, 573–639
20. Petes, T. D. (1980) Cell 19, 765–774
21. Zimmer, E. A., Martin, S. L., Beverly, S. M., Kan, Y. W., and Wilson, A. C. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 2188–2192
22. Leberher, H. G., and Rutter, W. J. (1967) Science 157, 1198–1200
23. Olsen, K. W., Moras, D., Rossmann, M. G., and Harris, J. I. (1975) J. Biol. Chem. 250, 9313–9321
24. Konkel, D. A., Maizel, J. V., Jr., and Leder, P. (1979) Cell 18, 865–873
25. Nishioka, Y., and Leder, P. (1979) Cell 18, 875–882

Additional references are found on p. 1395.
Yeast Enolase Genes

SUPPORTING MATERIAL

to

The Primary Structures of Two Yeast Enolase Genes: Homology between a Protein Substrate of Yeast Enolase and Glyceraldehyde-3-Phosphate Dehydrogenase Genes

By

Michael J. Holland, James P. Holland, Gregory P. Thill and Kimberly A. Levison

EXPERIMENTAL PROCEDURES

Materials

[Data not shown, likely referring to the materials and reagents used in the experiments.]

Cloning Strains of Yeast and plasmids

Subclone of yeast DNA containing structural genes was isolated from a library of yeast chromosomal DNA. Plasmid DNA was isolated from Saccharomyces cerevisiae ATCC 200-55 and used as a source of yeast DNA.

Restriction Enzyme Analysis

Restriction enzyme digestion of yeast DNA was performed as described by Grunberg-Manago et al. (1974) and by Southern et al. (1975).

RESULTS

Restriction Enzyme Patterns of Yeast DNA

Table 1:

| Restriction Enzyme | Fragment Molecular Weight (kDa) |
|---------------------|---------------------------------|
| Sal I               | 2.7, 0.5                         |
| Sal II              | 2.7                             |
| Eco RI              | 2.7, 0.25                       |
| Xho I               | 0.74, 0.74                      |
| Apa I               | 0.74, 0.01, 0.54                 |
| Apa II              | 1.0, 0.74                       |
| Bgl II              | 0.74, 0.72, 0.6, 1.66, 0.9, 0.45 |
| Hind III            | 0.74, 0.72, 0.6, 0.34, 0.26     |
| Hind II             | 0.5, 0.25, 0.34                 |
| Pst I               | 0.5, 0.25                       |
| Pst II              | 0.5, 0.25                       |
| Sac I               | 0.5, 0.25                       |
| Sac II              | 0.5, 0.25                       |
| Sac III             | 0.74, 0.25                      |
| Sac IV              | 0.74, 0.25                      |
| Sac V               | 0.74, 0.25                      |

Restriction Endonuclease Mapping of yeast DNA has been performed by digestion with various restriction enzymes. The resultant restriction fragments were isolated from agarose gels and subcloned into the lambda phage vector. Yeast DNA was isolated from Saccharomyces cerevisiae ATCC 200-55 and used as a source of yeast DNA.

REFERENCES

1. Hollan d, J.P., and Holland, J.P. (1978) Biochemistry 17, 4900-4907.
2. Holland, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
3. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
4. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
5. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
6. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
7. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
8. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
9. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
10. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
11. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
12. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
13. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
14. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
15. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
16. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
17. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
18. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
19. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
20. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
21. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
22. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
23. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
24. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
25. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
26. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.
27. Hollan d, J.P., and Holland, J.P. (1979) Biochemistry 18, 4900-4907.