THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 254, No. 19, Issue of October 10, pp. 9839-9845, 1979
Printed in U.S.A.

The Primary Structure of a Glyceraldehyde-3-phosphate Dehydrogenase Gene from *Saccharomyces cerevisiae**

(Received for publication, April 24, 1979, and in revised form, June 26, 1979)

Janice P. Holland and Michael J. Holland‡

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

The complete nucleotide sequence of the coding, as well as the flanking noncoding regions, of a yeast glyceraldehyde-3-phosphate dehydrogenase gene was determined. Both the 5′ and 3′ noncoding sequences are extremely AT-rich and regions of partial dyad symmetry are present immediately adjacent to the 5′ and 3′ ends of the translated portion of the gene. The sequence AAUAAA is present in the 3′ noncoding region of this gene and is a part of an extensive region of dyad symmetry which is structurally related to the 3′-terminal portion of both procaryotic mRNAs, as well as some eukaryotic mRNAs. The coding region of this gene does not contain intervening sequences. Establishment of the primary structure of this glyceraldehyde-3-phosphate dehydrogenase gene provides a basis for further studies involving *in vitro* mutation of the gene and subsequent analysis of gene expression *in vivo*.

The reaction mechanism and the structure of glyceraldehyde-3-phosphate dehydrogenase have been extensively investigated. The primary structures of the lobster (1), pig (2), yeast (3), and *Bacillus stearothermophilus* (4) glyceraldehyde-3-phosphate dehydrogenases have been determined as have the structures of the lobster (5) and *B. stearothermophilus* enzymes from crystallographic studies (4). The primary structures of the enzymes investigated to date are extremely homologous within particular domains of the protein and show 51 to 70% overall sequence homology. Variation in the amino acid sequences outside of the extremely homologous domains is in most cases restricted to a limited number of amino acid residues. Crystallographic studies (5) and investigations of the active sites within the tetrameric enzyme with sulfhydryl reagents (6) suggest that an active enzyme molecule exists as an asymmetric pair of symmetric dimers rather than as a fully symmetric tetrameric molecule.

Glyceraldehyde-3-phosphate dehydrogenase is a ubiquitous enzyme in nature and is present in high concentration in cells which are specialized for the process of glycolysis, such as skeletal muscle and yeast. A striking illustration of its potential abundance in a cell is the observation that this protein can comprise as much as 5% of the dry weight of commercial baker's yeast (7). Several lines of evidence also suggest that expression of the glyceraldehyde-3-phosphate dehydrogenase gene in yeast is under transcriptional control (8–10). It appears likely that this gene contains strong signals for the initiation of transcription.

Given the extensive information on the structure of glyceraldehyde-3-phosphate dehydrogenase and the possibility that this gene contains strong transcriptional control signals, investigation of the glyceraldehyde-3-phosphate dehydrogenase gene should yield useful information about the relationship of DNA structure to transcriptional control. Recently developed technology for the construction of mutants *in vitro* (11, 12) and methods for the reintroduction of cloned segments of yeast DNA into yeast cells provide the necessary tools for such a study (13).

We have previously described the isolation and identification of yeast glyceraldehyde-3-phosphate dehydrogenase mRNA (10). Utilizing a complementary DNA probe synthesized from this mRNA we isolated and characterized a glyceraldehyde-3-phosphate dehydrogenase structural gene using molecular cloning techniques (14). We report here the complete nucleotide sequence of this gene, including the translated and flanking, nontranslated sequences. The sequence reported establishes the organization of the coding and noncoding segments of the gene. Regions of extensive dyad symmetry in the flanking sequences of the gene are discussed as are structures which may be relevant to the initiation and termination of transcription.

EXPERIMENTAL PROCEDURES

Materials—[γ-^{32}P]ATP (3000 Ci/mmol) was purchased from New England Nuclear. Restriction endonucleases *Alu* I, *Ava* II, *Hin*dIII, *Hin*fI, *Hpa* I, *Hpa* II, *Sal* I, and *Taq* I were obtained from Bethesda Research Laboratories; *Hae* III and *Hha* I were from New England Biolabs. Bacterial alkaline phosphatase was supplied by Boehringer Mannheim and polynucleotide kinase was purchased from P-L Biochemicals. Plasmid DNA was prepared according to the procedure of Clewell and Helinski (15). p*gap*491, an ampicillin-resistant recombinant plasmid containing the glyceraldehyde-3-phosphate dehydrogenase gene, was isolated as previously described (14).

Restriction Endonuclease Digestion—Digestions involving *Hha* I, *Alu* I, and *Hin*dIII were performed at 37°C in 50 mM NaCl, 6 mM Tris (pH 7.4), 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol, and 0.1 mg/ml of gelatin. *Hpa* I and *Hpa* II digestions were carried out at 37°C in 6 mM KCl, 10 mM Tris (pH 7.4), 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.1 mg/ml of gelatin. *Hae* III digestions were identical to *Hpa* I with the exception that NaCl was used. *Ava* II digestions were performed at 37°C in 30 mM NaCl, 80 mM Tris (pH 7.4), 10 mM $MgCl_2$, and 6 mM dithiothreitol. *Sal* I digestions were carried out in 150 mM NaCl, 8 mM Tris (pH 7.6), 6 mM $MgCl_2$, 0.2 mM Na_2EDTA, 50 μg/ml of bovine serum albumin. Digestions with *Taq* I were performed at 65°C in buffer containing 100 mM NaCl, 10 mM Tris (pH 8.3), 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol.

Restriction Site Mapping and 5′-terminal Labeling of Restriction Fragments—The partial digestion procedure of Smith and Birnstiel (16) was employed for mapping restriction sites relative to the ^{32}P-labeled *Sal* I site in the glyceraldehyde-3-phosphate dehydrogenase gene. Fifty micrograms of *Sal* I linear p*gap*491 DNA were dephosphorylated with bacterial alkaline phosphatase according to Smith

* This research was supported by a grant from The National Institute of General Medical Sciences and by a Basil O'Connor research grant from the National Foundation-March of Dimes. 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.

and Birnstiel (16) and phosphorylated with polynucleotide kinase and [γ-^{32}P]ATP by the method of Maxam and Gilbert (17). Following a phenol extraction, unincorporated [γ-^{32}P]ATP was removed by Sephadex G-50 gel filtration. The ^{32}P-labeled *Sal* linear DNA was ethanol-precipitated, resuspended, and digested with *Hin*dIII. ^{32}P-end-labeled DNA fragments (2.0 and 0.14 kb[1]) were isolated by preparative agarose gel electrophoresis as previously described (14). A series of partial digestions were carried out with *Hha* I, *Alu* I, *Hpa* I, Hae III, *Ava* II, *Hpa* II, *Taq* I, and *Hin*fI and the cleavage patterns were visualized by autoradiography with Kodak X-Omat R film and a Dupont Lightning Plus intensifying screen. Both the 2.0- and 0.14-kb fragments were also utilized in the DNA sequence analysis to be discussed.

DNA Sequence Analysis—The 2.1 kb *Hin*dIII fragment of p*gap*491 was prepared by sucrose density gradient centrifugation. All sequence strategies were carried out with this fragment with the exception of the analysis to either side of the *Sal* I restriction site which involved the isolation of ^{32}P-labeled *Sal* linear fragment as described above. Labeling of restriction fragments was identical to that described for the partial digestion mapping procedure with the exception of the blunt-ended *Hpa* I/*Hin*dIII fragments which were denatured prior to labeling and renatured before the secondary *Ava* II cleavage. The sequence analyses were performed according to the method of Maxam and Gilbert (17) employing 0.4-mm 20% acrylamide gel electrophoresis at 1100 V with a 7-, 20-, and 36-h loading schedule. Autoradiography was carried out as described above. The following restriction fragments isolated by preparative agarose gel electrophoresis (or acrylamide gel electrophoresis in the case of the *Taq* I/*Hin*fI fragments under 100 nucleotides in size) were sequenced: 2.0-kb [^{32}P]*Sal* I/*Hin*dIII; 139-ntd [^{32}P]*Sal* I/*Hin*dIII; 2.0-kb [^{32}P]*Hin*dIII/*Sal* I; 139-ntd [^{32}P]*Hin*dIII/*Sal* I; 403-ntd [^{32}P]*Taq* I/*Hae* III; 287-ntd [^{32}P]*Hpa* II/*Sal* I; 405-ntd [^{32}P]*Ava* II/*Sal* I; 324-ntd [^{32}P]*Ava* II/*Alu* I; 163-ntd [^{32}P]*Hpa* I/*Ava* II; 0.75-kb [^{32}P]*Hpa* I/*Ava* II; 272-ntd [^{32}P]*Taq* I/*Hae* III; 57-ntd [^{32}P]*Taq* I/*Hin*fI; 65-ntd [^{32}P]*Taq* I/*Hin*fI; 44-ntd [^{32}P]*Taq* I/*Hin*fI; 0.38-kb [^{32}P]*Hin*fI/*Hha* I; 0.54-kb [^{32}P]*Taq* I/*Hae* III.

Transcription Mapping—A transcription map of the 2.1-kb *Hin*dIII fragment containing a glyceraldehyde 3-phosphate dehydrogenase gene was established by digesting this fragment with the restriction endonuclease *Hin*fI. Electrophoresis of the resulting fragments was carried out in a 1.5% agarose slab gel in the presence of ethidium bromide (0.5 μg/ml). DNA was transferred to a nitrocellulose filter by the modified procedure of the Southern (18) technique outlined by Ketner and Kelly (19). Filter hybridization was carried out under oil for 15 h at 37°C in 50% formamide, 4 × SSC (0.6 M sodium chloride, 0.06 M sodium citrate, pH 7.0), 13 mM sodium phosphate, 1 mM EDTA, and 0.5% SDS with glyceraldehyde 3-phosphate mRNA prepared according to Holland and Holland (10) which was end-labeled to a specific activity of greater than 10^7 cpm/μg by alkali treatment (20) and phosphorylation with [γ-^{32}P]ATP and polynucleotide kinase (17). Following a chloroform wash at room temperature and four washes at 65°C with 4 × SSC for a total of 90 min, the filter was subjected to autoradiography.

Containment—All the experiments reported here were carried out in accordance with the NIH *Guidelines for Research involving Recombinant DNA Molecules*.[2]

RESULTS AND DISCUSSION

Nucleotide Sequence of a Yeast Glyceraldehyde-3-phosphate Dehydrogenase Gene—A yeast glyceraldehyde-3-phosphate dehydrogenase structural gene was isolated by subculture cloning as previously described (14). A portion of the cloned segments of yeast DNA, which contains the gene, was subcloned from the original hybrid plasmid in order to facilitate isolation of DNA for nucleotide sequence analysis. The gene in this subclone (p*gap*491) is located within a 2.1-kb *Hin*dIII restriction endonuclease cleavage fragment which is easily resolved from the remainder of the hybrid plasmid DNA by sucrose density gradient centrifugation. A restriction endonuclease cleavage map of this segment of DNA was determined for 10 restriction endonucleases by the method described by Smith and Birnstiel (16). Fig. 1 illustrates the partial cleavages of a 2.0-kb fragment which extends from a *Sal* I cleavage site located within the extreme 3′ end of the coding region of the gene (14) to the *Hin*dIII cleavage site. The *Sal* I cleavage site was labeled at the 5′ end with ^{32}P for these experiments (17). The restriction endonuclease cleavage map predicted from these data is diagrammed in Fig. 1*b*. The order of the restriction endonuclease cleavage sites shown in Fig. 1 was verified in a series of partial cleavages of two fragments, labeled at the two *Hin*dIII termini of the fragment

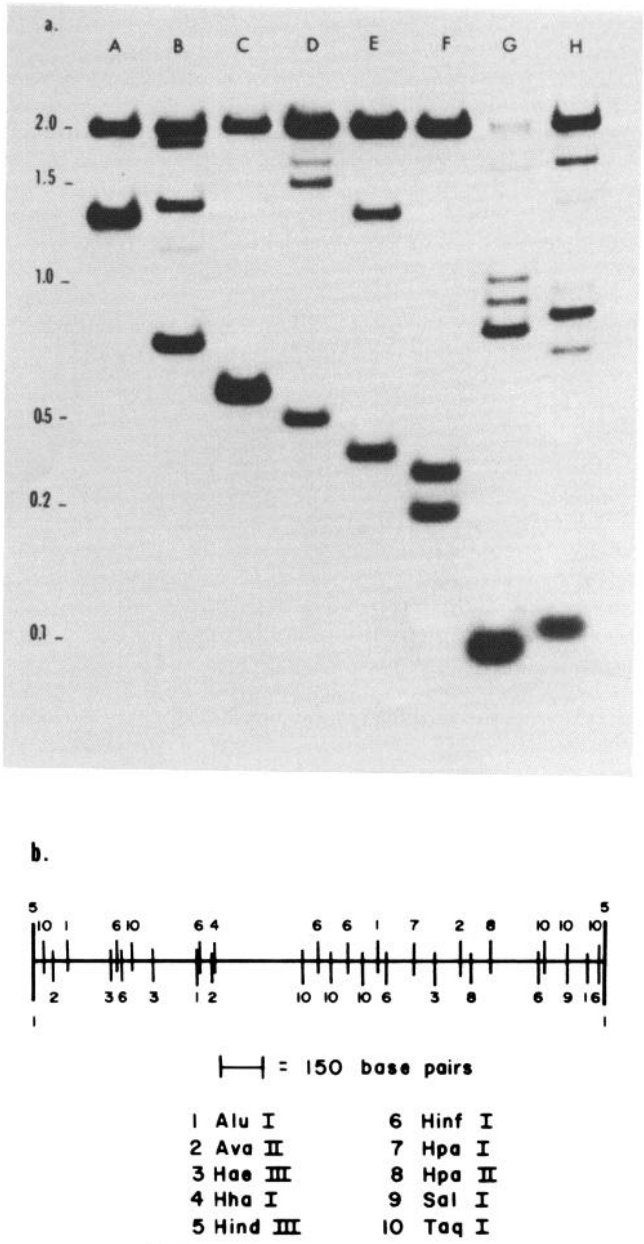


FIG. 1. **Restriction endonuclease cleavage map of the 2.1-kb *Hin*dIII fragment of p*gap*491.** *Panel a* is an autoradiogram of a 1.5% agarose slab gel of partial digests of the 2.0-kb *Sal* I/*Hin*dIII fragment which was labeled with ^{32}P at the 5′ end of the *Sal* I cleavage site. Partial digests were carried out with: *A, Hha* I; *B, Alu* I; *C, Hpa* I; *D, Hae* III; *E, Ava* II; *F, Hpa* II; *G, Taq* I; and *H, Hin*fI. *Panel b* shows the cleavage map which was predicted from the partial digests shown in *Panel a*. Cleavage sites to the right of the *Sal* I site were mapped as described in the text.

[1] The abbreviations used are: kb, kilobase pairs; ntd, nucleotide; SDS, sodium dodecyl sulfate.

[2] *Guidelines for Research Involving Recombinant DNA Molecules* (1979) National Institutes of Health, Bethesda.

and generated, after labeling, by cleavage with *Hpa* I (data not shown). These latter data were used to determine the cleavage sites in the 0.14-kb *Hin*dIII/*Sal* I fragment on the extreme right of the map shown in Fig. 1. These cleavages also positioned a *Hin*fI cleavage site at the extreme left of the map which was not detected in the partial cleavages shown in Fig. 1. Cleavage of the 2.0-kb [^{32}P]*Sal* I/*Hin*dIII fragment with *Alu* I yields a partial digestion product at 1.2 kb (Fig. 1) which is also seen in limit digests of the fragment but always in submolar amounts. The nucleotide sequence of this region of the DNA was determined and no *Alu* I cleavage site was found. It is likely that this cleavage is a result of a specific contaminating endonuclease in the *Alu* I preparation. The molecular weights of the fragments generated by these 10 restriction endonucleases were determined from limit digests of the 2.1-kb *Hin*dIII fragment (data not shown).

The nucleotide sequence of this glyceraldehyde-3-phosphate dehydrogenase structural gene, as well as the sequences which flank the gene, were determined using the chemical cleavage method of Maxam and Gilbert (17). In order to maximize resolution, 0.4-mm thick polyacrylamide gels were employed. From 80 to 140 nucleotides could be accurately determined with each labeled fragment. The nucleotide sequence of the 3′ noncoding sequences was determined from the analysis of both strands of the DNA as was the sequence of coding regions of the gene in which the amino acid sequence of the protein was not determined. The nucleotide sequence of the 5′ noncoding region could not be verified by analysis of both DNA strands due to the paucity of restriction endonuclease cleavage sites in this region of the DNA; however, the sequence reported was verified in three separate sequencing experiments from the *Hin*fI cleavage site at position 47 and the *Taq* I cleavage site at position −24. We find in agreement with others that the most frequent error encountered is a distinction between the C and C + T cleavage. We find that this error can be minimized by keeping the NaCl concentration in the chemical cleavage reactions at precisely that recommended by Maxam and Gilbert (17).

The complete nucleotide sequence of the glyceraldehyde-3-phosphate dehydrogenase gene, as well as 150 nucleotides adjacent to the 5′ end of the coding region and 112 nucleotides at the 3′ noncoding portion of the gene, is shown in Fig. 2*a*. The strategy utilized for establishing this sequence is also illustrated in Fig. 2*b*.

The Amino Acid Sequence Predicted from the Coding Sequences—The amino acid sequence of yeast glyceraldehyde-3-phosphate dehydrogenase was determined by Jones and Harris (3). Two hundred eighty residues in the protein were experimentally determined, while 51 residues were predicted using the amino acid composition data for specific peptides and maximizing conservation of amino acid sequence of the peptide relative to the previously determined sequences of the pig and the lobster muscle enzymes. The amino acid sequence predicted from the nucleotide sequence of this gene agrees with that experimentally determined by Jones and Harris in all but 2 residues. Residues 6 and 133 were determined to be aspartate while the nucleotide sequence predicts asparagine in these positions. Of the 51 amino acid residues which were predicted but not experimentally determined by Jones and Harris, 40 were in agreement with those predicted from the nucleotide sequence. Four amino acid residues are predicted from the nucleotide sequence which are not found in the primary structure of glyceraldehyde-3-phosphate dehydrogenases isolated from different species or are not compatible with the predicted three-dimensional structure of the yeast enzyme (based on the known structure of the lobster muscle enzyme (21)). These residues are: 6, asparagine instead of aspartate; 37, asparagine instead of leucine; 87, serine (not among the acceptable alternative residues at this position); 135, valine instead of glutamate. In each of these positions the nucleotide sequence was determined from either both strands of the DNA or the experimentally determined nucleotides were purines in an easily read portion of the sequencing gel. It is worth noting at this point that the glyceraldehyde-3-phosphate dehydrogenase gene is nontandemly repeated in yeast (14) and that the amino acid sequence predicted from the coding region of one gene may not be identical to those predicted from the other genes. If, for example, the protein which was sequenced by Jones and Harris was a mixture of polypeptides derived from different genes, the discepancy in amino acid sequence noted above may arise from differences in the protein sequence encoded by the different genes. Clarification of this point will require further studies on the nucleotide sequence and expression of the yeast glyceraldehyde-3-phosphate dehydrogenase genes.

Utilization of codons in this gene is clearly not random as indicated in Table I. Such a selection bias has been observed in a number of eukaryotic mRNAs (22–25); however, the degree of codon selectivity in this gene is particularly unusual. Ten amino acids are encoded by a single codon (arginine, asparagine, cysteine, glutamine, glycine, histidine, leucine, phenylalanine, proline, and tyrosine) and five are encoded by two codons (alanine, isoleucine, serine, threonine, and valine). The third position of these latter codons is always C or T. In the cases of glutamate and lysine two codons are utilized; however, the usage is heavily biased to one of the two codons. Since differences in codon utilization are observed among genes isolated from the same organism (*e.g.* rabbit α-globin (22) and rabbit β-globin (23)), it is difficult to predict from these data the basis for nonrandom codon utilization. It is worth noting, however, that glyceraldehyde-3-phosphate dehydrogenase is synthesized in yeast cells in large quantity. The relative abundance of the mRNA which codes for the protein could account for the synthetic levels of the protein; however, it is possible that the codon utilization is biased toward those with high intracellular levels of corresponding tRNAs and that this may result in a higher rate of translation of the mRNA. The possibility of control at this level has been suggested by Lodish (26). Demonstration of such a control mechanism for the glyceraldehyde-3-phosphate dehydrogenase mRNA will require further investigation.

Location of the 5′ and 3′ Noncoding Sequences of Glyceraldehyde-3-phosphate Dehydrogenase Messenger RNA—Since the nucleotide sequence of glyceraldehyde-3-phosphate dehydrogenase mRNA has not been determined, a transcription map of the coding and noncoding sequences in the gene was carried out in order to locate the sequences in the cloned segment of DNA which are complementary to glyceraldehyde 3-phosphate dehydrogenase mRNA. A radioactive hybridization probe was generated by labeling the 5′-termini of partially hydrolized glyceraldehyde-3-phosphate dehydrogenase mRNA with [^{32}P]ATP and polynucleotide kinase (17, 20). The 2.1-kb *Hin*dIII fragment was digested with *Hin*fI and the digests were electrophoresed on an agarose slab gel. DNA was transferred to a nitrocellulose filter and hybridized with the labeled probe described above. As illustrated in Fig. 3, the B fragment from the *Hin*fI digest represents the sequences which overlap the 5′ noncoding region of the gene. The B fragment (*Hin*fI) contains 32 nucleotides from the coding region of the gene, as well as approximately 500 nucleotides which flank the 5′ end of the gene. On comparison of the amount of hybridization to the B fragment to that for the 141

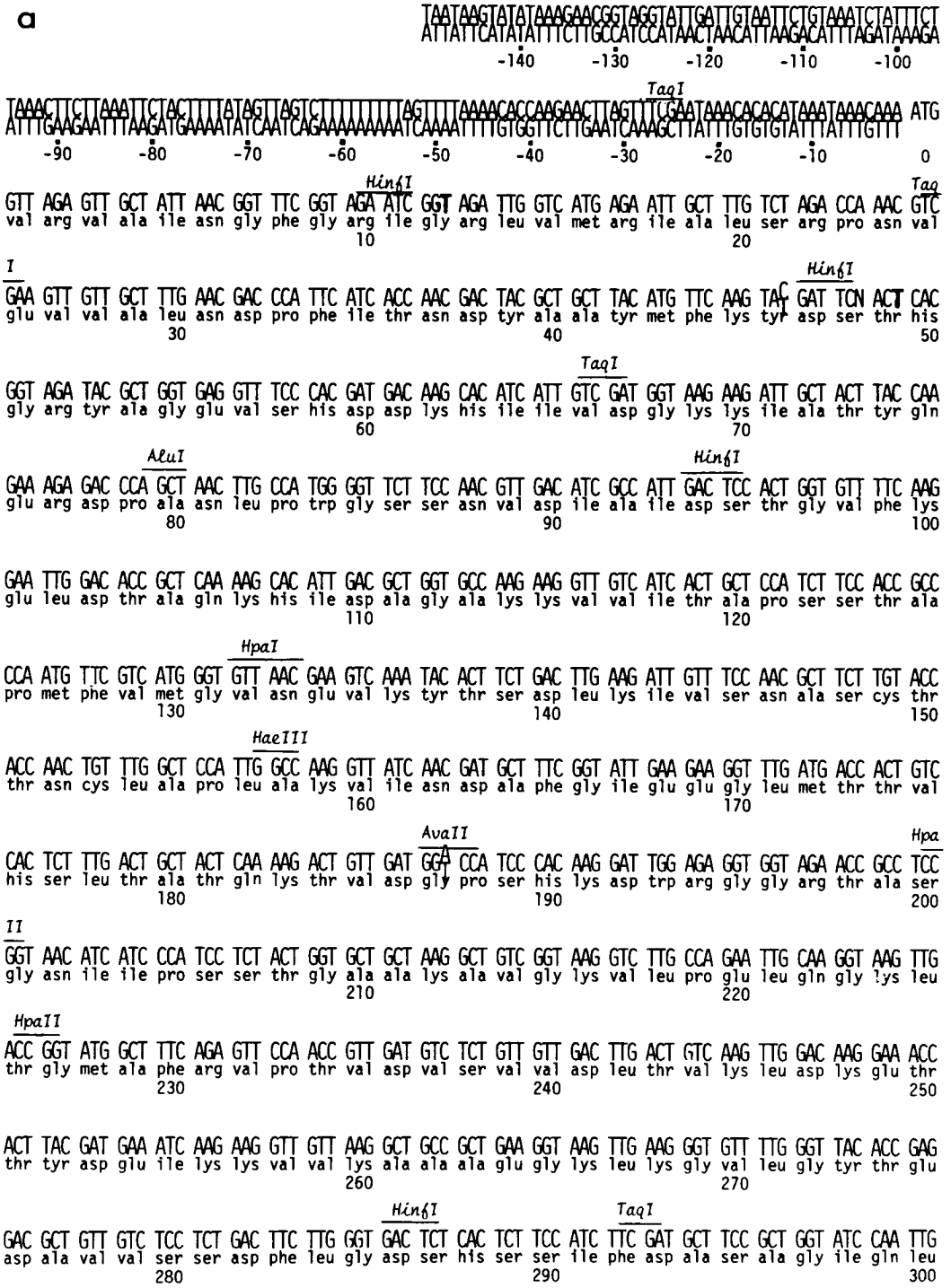
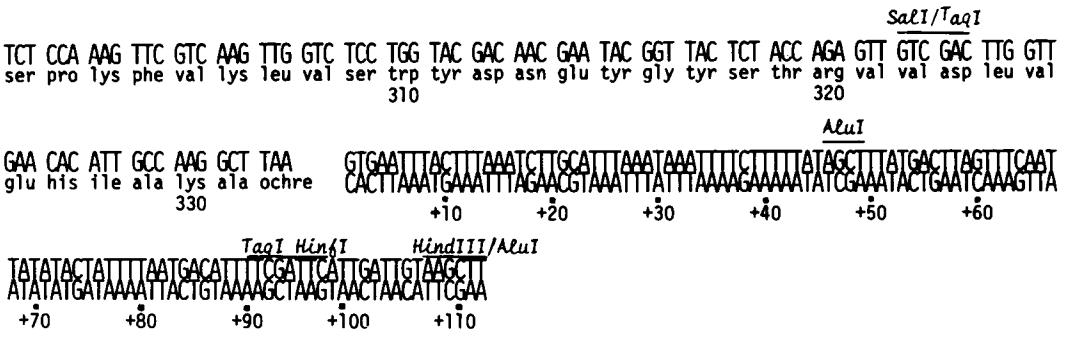


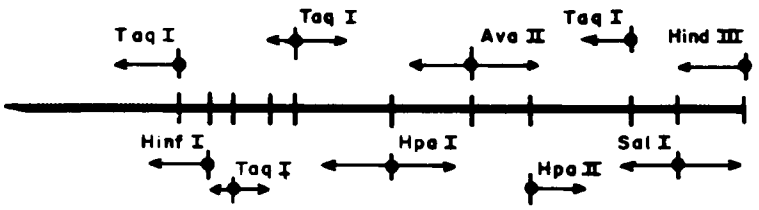
FIG. 2. **Nucleotide sequence of a yeast glyceraldehyde-3-phosphate dehydrogenase gene.** *a*, all of the nucleotide sequences were determined by the method of Maxam and Gilbert (17); *b*, the strategy for sequencing the gene. The 5′ end of each indicated restriction fragment was labeled with ^{32}P and the fragment was isolated as described under "Experimental Procedures." The direction of the *arrows* indicates the direction of sequence determination and the length of the *arrows* corresponds to the amount of sequence determined in each experiment. *Numbers outside* the coding sequence refer to nucleotides; *numbers inside* the coding region refer to amino acid residues. The amino acid sequence shown is for one reading frame of the nucleotide sequence.

nucleotide *Hin*fI G fragment which comes entirely from the coding portion of the gene, we conclude that the mRNA sequences do not extend beyond the nucleotide sequence which was determined (Fig. 2). Based on the relative intensity of hybridization, it is unlikely that the glyceraldehyde-3-phosphate dehydrogenase mRNA contains more than 50 to 100 nucleotides of noncoding sequence at the 5′ end of the gene.

Similar experiments were carried out to locate the noncoding sequences at the 3′ terminus of the gene which are complementary to glyceraldehyde-3-phosphate dehydrogenase mRNA. A small amount of hybridization to the 0.13-kb *Sal* I/*Hin*dIII cleavage fragment (Fig. 1), which contains the 3′ terminus of the gene (data not shown), was detected. In parallel filter blots the extent of hybridization to the 0.13-kb



b



⊢⊣ = 100 base pairs

FIG. 2*b*

TABLE I

Codon utilization in a glyceraldehyde-3-phosphate dehydrogenase gene

Ala	GCA 0	Arg	CGA 0	Asn	AAT 0	Asp	GAT 9
	T 25		T 0		C 12		C 16
	C 7		C 0				
	G 0		G 0				
			AGA 11				
			G 0				
Cys	TGT 2	Gln	CAA 5	Glu	GAA 12	Gly	GGA 0
	C 0		G 0		G 2		T 25
							C 0
							G 0
His	CAT 0	Ileu	ATA 0	Leu	TTA 0	Lys	AAA 1
	C 8		T 9		G 21		G 25
			C 11		CTA 0		
					T 0		
					C 0		
					G 0		
Met	AUG 6	Phe	TTT 0	Pro	CCA 12	Ser	TCA 0
			C 10		T 0		T 13
					C 0		C 12
					G 0		G 0
							AGT 0
							C 0
Thr	ACA 0	Trp	TGG 3	Tyr	TAT 0	Val	GTA 0
	T 12				C 10		T 22
	C 12						C 15
	G 0						G 0

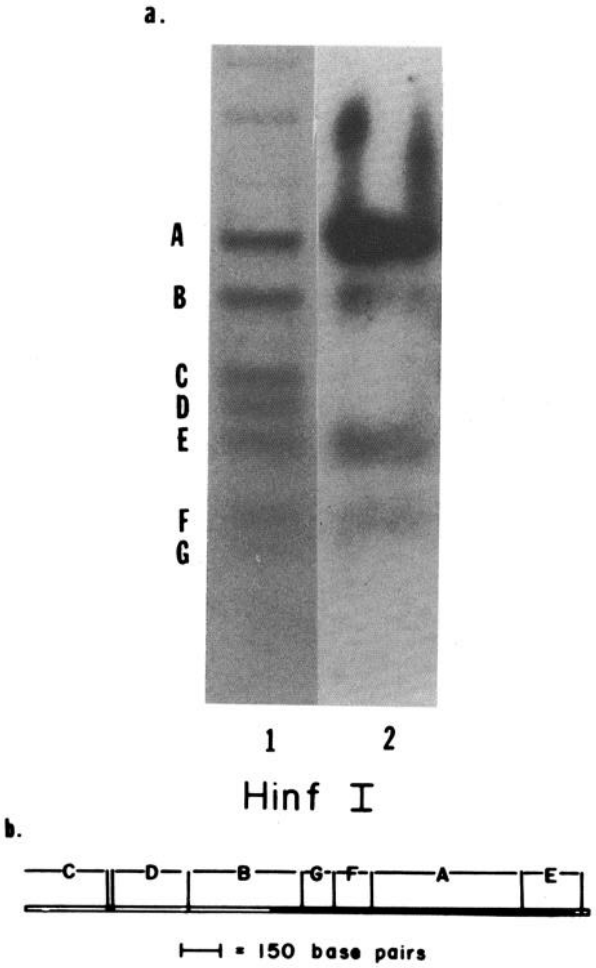
Sal I/*Hin*dIII fragment was less than half that observed for the *Hinf*B fragment described above. We, therefore, conclude that the 3′ noncoding sequences in glyceraldehyde-3-phosphate dehydrogenase mRNA do not extend beyond the nucleotide sequence which was experimentally determined (Fig. 2).

The 5′ Noncoding Flanking Sequences—The 5′-noncoding sequences of the glyceraldehyde-3-phosphate dehydrogenase gene are extremely AT-rich. The A + T composition of the 150 nucleotides which are immediately adjacent to the AUG initiation codon is 78% as compared to 53% for the coding region of the gene. An extensive region of partial dyad symmetry exists in this region as illustrated in Fig. 4*a*. Of the 24-base-paired nucleotides in this structure, 21 are AT pairs. The structure illustrated in Fig. 3*a* is only one of many possible structures in this segment of DNA and, in the absence of functional data, the significance of such structures cannot be addressed. It is worth noting, however, that such structures are not found adjacent to the 5′ ends of the 5 S rRNA genes (see a review of 5 S rDNA 5′ termini in Korn and Brown, (27)).

A second feature of this sequence which will be of interest in further studies on the function of this gene is the sequence which precedes the AT-rich region described above. Within nulceotides −113 through −72 there are no regions of dyad symmetry; however, the noncoding strand of the gene is purine-rich, extending through the run of 9 adenylate residues to nucleotide −48. This segment of the gene contains a series of AAA and TTT triplets interspersed with AAG, GAA, and AGA triplets as illustrated in Fig. 4*b*.

In agreement with other eukaryotic sequences there are no AUG codons in the coding strand preceding the initiation codon (28). In contrast to procaryotic genes we do not find a sequence with the structure TATRATG which has been described in a number of prokaryotic genes as a transcriptional start signal (29).

The 3′ Noncoding Flanking Sequences—As described for the sequence which flanks the 5′ end of the gene, the sequence adjacent to the 3′ end is also AT-rich (79%). Analysis of the nucleotide sequence adjacent to the 3′ end also reveals an extensive region of dyad symmetry which is illustrated in Fig. 4*c*. The AAUAAA sequence present in this sequence is a base-paired component of the palindrome and a T-rich sequence follows the dyad symmetry. The dinucleotide GC follows the T-rich sequence and is within the distance (10 to 20 nucleotides) from AAUAAA observed for other eucaryotic mRNAs as the site for poly(A) addition (see review in Proudfoot *et al.* (30)). The nucleotide sequence described above is similar to those described for the termini of prokaryotic mRNAs (31) and for eukaryotic 5 S rDNA (27) with the exception that the



dyad symmetry is not composed of GC pairs but rather AT pairs. If in fact the palindromic structure is an important feature of this gene, it is worth noting that the size of the paired region (14 base pairs) can account for the lower stability of the AT base pairs relative to GC pairs. A similar AT-paired structure has been described for the 3′-terminal sequence of the mouse L chain immunoglobulin mRNA (32).

An analysis of the nucleotide sequence of this glyceraldehyde-3-phosphate dehydrogenase gene does not in itself answer the obvious questions about the relationship between gene structure and function; however, alteration of specific sequences of isolated genes *in vitro* and reintroduction of these altered genes in to the cell will allow one to probe this question in a more satisfying manner. The existence of nontandemly repeated glyceraldehyde-3-phosphate dehydrogenase genes in yeast will provide through comparative sequence analysis the opportunity to study the relevance of such palindromic structures which flank the coding region, as well as the function of the multiple genes in the cell.

FIG. 3. **Mapping of glyceraldehyde-3-phosphate dehydrogenase messenger RNA sequences within the 2.1-kb *Hin*dIII fragment isolated from p*gap*491.** The 2.1-kb *Hin*dIII fragment isolated from p*gap*491 was digested with *Hin*fI, electrophoresed on a 1.5% agarose slab gel, transferred to a nitrocellulose membrane and hybridized to a ^{32}P-labeled glyceraldehyde 3-phosphate dehydrogenase mRNA probe as described under "Experimental Procedures." *Panel a: Lane 1, Hin*fI cleavage fragments visualized with ethidium bromide after agarose gel electrophoresis; *Lane 2*, an autoradiogram of the hybridization of the *Hin*fI fragments, after transfer to nitrocellulose paper, with ^{32}P-labeled mRNA. *Panel b:* a transcription map of the 2.1-kb *Hin*dIII fragment indicating the location of sequences within the DNA fragment which are complementary to glyceraldehyde-3-phosphate dehydrogenase mRNA (*shaded portion* of the map). These data were obtained with the hybridization results shown in *Panel a* and those discussed in the text.

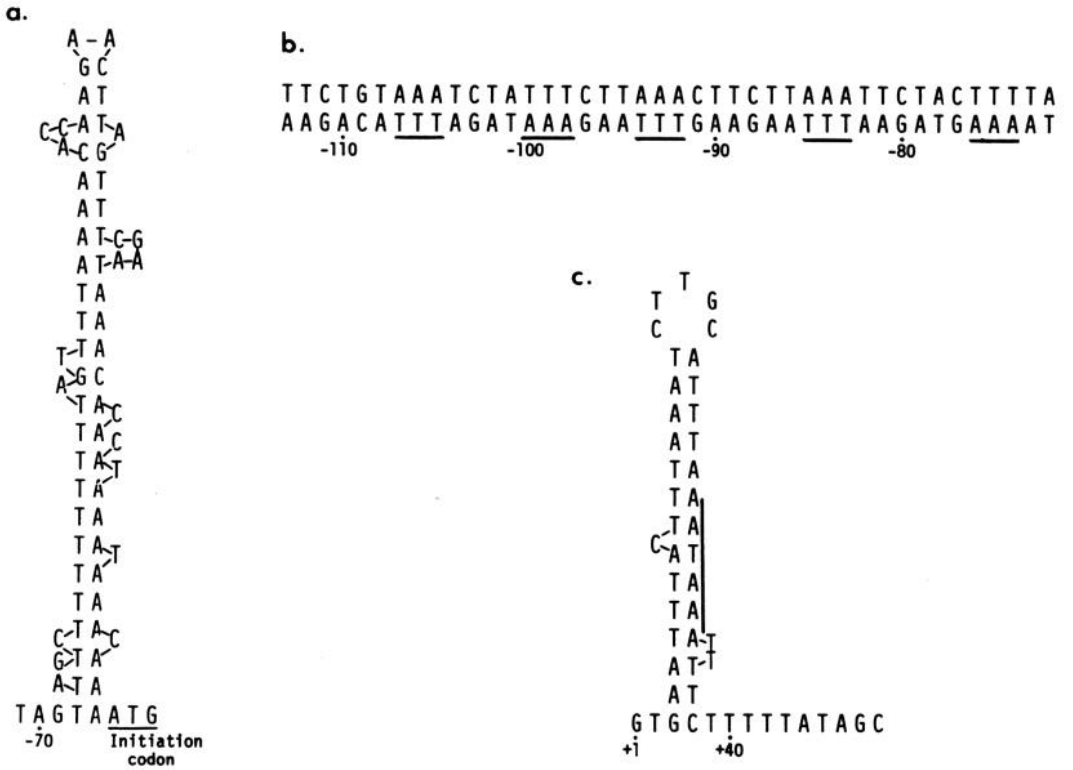


FIG. 4. **Sequence features in the noncoding 3′ and 5′ flanking regions of the glyceraldehyde-3-phosphate dehydrogenase gene.** *Panel a,* region of extensive dyad symmetry which preceeds the AUG initiation codon; *Panel b,* unique sequence which preceeds the region shown in *Panel a; Panel c,* region of dyad symmetry adjacent to the 3′ end of the coding portion of the gene. The *heavy line* indicates the location of the AAUAAA sequence in the structure.

Acknowledgments—The advice of Dr. B. Thimmapaya with regard to DNA sequence analysis is gratefully appreciated.

REFERENCES

1. Davidson, B. E., Sajgo, M., Noller, H. F., and Harris, J. I. (1967) *Nature* **216,** 1181–1185
2. Harris, J. I., and Perham, R. N. (1968) *Nature* **219,** 1025–1028
3. Jones, G. M. T., and Harris, J. I. (1972) *FEBS Lett.* **22,** 185–189
4. Biesecker, G., Harris, J. I., Thierry, J. C., Walker, J. E., and Wonacott, A. J. (1977) *Nature* **266,** 328–333
5. Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C., and Rossmann, M. G. (1975) *J. Biol. Chem.* **250,** 9137–9162
6. Bernard, S. A., and MacQuarrie, R. A. (1973) *J. Mol. Biol.* **74,** 73–78
7. Krebs, E. G. (1953) *J. Biol. Chem.* **200,** 471–478
8. Hommes, F. A. (1966) *Arch. Biochem. Biophys.* **1,** 231–233
9. Maitra, P. K., and Lobo, Z. (1971) *J. Biol. Chem.* **246,** 475–488
10. Holland, M. J., and Holland, J. P. (1978) *Biochemistry* **17,** 4900–4907
11. Shenk, T. E., Carbon, J., and Berg, P. (1976) *J. Virol.* **18,** 664–671
12. Shortle, D., and Nathans, D. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 2170–2174
13. Hinnen, A., Hicks, J. B., and Fink, G. R. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 1929–1933
14. Holland, M. J., and Holland, J. P. (1979) *J. Biol. Chem.* **254,** 5456–5474
15. Clewell, D. B., and Helinski, D. R. (1970) *Biochemistry* **9,** 4428–4440
16. Smith, H. O., and Bernstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2398
17. Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 560–564
18. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
19. Ketner, G., and Kelly, T. J., Jr. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73,** 1102–1106
20. Maizels, N. (1976) *Cell* **9,** 431–438
21. Olsen, K. W., Moras, D., Rossmann, M. G., and Harris, J. I. (1975) *J. Biol. Chem.* **250,** 9313–9321
22. Heindell, H. C., Liu, A., Paddock, G. V., Studnicka, G. M., and Salser, W. A. (1978) *Cell* **15,** 43–54
23. Efstratiadis, A., Kafatos, F. C., and Maniatis, T. (1977) *Cell* **10,** 571–585
24. Milstein, C., Brownlee, G. G., Cheng, C. C., Hamlyn, P. H., Proudfoot, N. J., and Rabbits, T. H. (1976) *Mossbacher Colloquium* (Melchers, F., and Rajewsky, K., eds) pp. 75–85, Springer Verlag, New York
25. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tigcher, E., Rutter, W. J., and Goodman, H. M. (1977) *Science* **196,** 1313–1319
26. Lodish, H. F. (1976) *Annu. Rev. Biochem.* **45,** 39–72
27. Korn, L. J., and Brown, D. D. (1978) *Cell* **15,** 1145–1156
28. Kozak, M. (1978) *Cell* **15,** 1109–1123
29. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. U. S. A.* **72,** 784–788
30. Proudfoot, N. J., Cheng, C. C., and Brownlee, G. G. (1976) *Prog. Nucleic Acid Res. Mol. Biol.* **19,** 123–134
31. Gilbert, W. (1976) *RNA Polymerase* (Losick, R., and Chamberlin, M. eds) pp. 193–205, Cold Spring Harbor Press, Cold Spring Harbor, New York
32. Hamlyn, P. H., Brownlee, G. G., Cheng, C.-C., Gait, M. J., and Milstein, C. (1978) *Cell* **15,** 1067–1075

The primary structure of a glyceraldehyde-3-phosphate dehydrogenase gene from Saccharomyces cerevisiae.

J P Holland and M J Holland

J. Biol. Chem. 1979, 254:9839-9845.