

Sequence Analysis of the Cloned mRNA Coding for Glyceraldehyde-3-Phosphate Dehydrogenase from Chicken Heart Muscle

Horst DOMDEY, Karin WIEBAUER, Heidi KLAPTHOR, and Hans Henning ARNOLD

Department of Toxicology, University of Hamburg, Medical School

(Received September 23/November 15, 1982) – EJB 6033

Using a cloned cDNA (pGAP30) the nucleotide sequence for chicken glyceraldehyde-3-phosphate dehydrogenase mRNA has been determined. The cDNA insert contains 1051 nucleotides representing the amino acid coding sequence, with the exception of 49 NH₂-terminal amino acids, and includes the entire 3'-noncoding region. Sequence information on the missing 5' terminus of the mRNA, not represented in the clone pGAP30, was obtained by extension of the cDNA using an 85-nucleotide-long internal fragment as a primer. Thus the sequence of 310 amino acids of chicken glyceraldehyde-3-phosphate dehydrogenase representing 93% of the complete primary structure could be derived. The coding portion exhibits non-random utilization of synonymous codons with a strong bias for codons with G or C at the third position. The non-coding region contains several octanucleotides which are repeated and shows a potentially stable stem-and-loop structure located towards the end of the mRNA. Hypothetical functional implications of the putative secondary structure are discussed.

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) catalyzes the oxidation and phosphorylation of D-glyceraldehyde 3-phosphate to 1,3-bisphosphoglycerate [1]. Therefore, the enzyme plays a crucial role in glycolysis and gluconeogenesis in most cells, particularly in muscle cells where it constitutes a considerable fraction of total cellular protein. We have utilized chick embryonic cardiac muscle as a model system to investigate molecular aspects of cell differentiation and early embryonic gene functions. In order to relate the synthesis of muscle proteins during muscle cell differentiation to alterations in gene activities, we have isolated and characterized the mRNAs coding for contractile proteins, myosin light and heavy-chain polypeptides [2,3] and more recently for GAPDH, another abundant protein in heart muscle cells. To gain further insights into the organization and regulation of these genes, we have cloned the mRNAs in bacterial vectors [2,4,5].

The GAPDH enzyme is composed of four identical subunits [6] whose amino acid sequence has been determined for a wide variety of organisms such as *Bacillus stearothermophilus* [7], yeast [8], lobster [9], pig [10] and humans [11]. In yeast, multiple forms of GAPDH polypeptides have been identified and isolated [12–14] but an apparently unique amino acid sequence was established [8]. More recently the DNA sequences of two active GAPDH structural genes and evidence for a third gene per haploid yeast genome have been presented [15]. Examination of the GAPDH enzyme and its corresponding mRNA in different tissues from chicken indicated the existence of a single molecular mRNA species coding for a single GAPDH polypeptide chain [16]. Here, we describe the nucleic acid sequence determination of cloned

GAPDH cDNA from chicken heart muscle and the amino acid sequence of the enzyme as predicted by the nucleotide sequence. The structure is compared with the known amino acid sequences of GAPDH from other organisms to establish the degree of divergence during evolution. To the best of our knowledge, the nucleotide sequence data on chicken GAPDH mRNA constitutes the first report on the primary structure of the GAPDH coding sequence from vertebrates.

EXPERIMENTAL PROCEDURES

Preparation of Cloned Glyceraldehyde-3-Phosphate Dehydrogenase DNA for Sequence Analysis

The construction of recombinant plasmids pGAP30 and pGAP36 bearing chicken GAPDH double-stranded cDNA inserts was described previously [5]. Large preparations of pGAP30 plasmid DNA were performed by methods described by Holland and McCarthy [17]. To generate fragments suitable for sequence determination, appropriate restriction enzymes were used for digestion as recommended by the supplier (Bethesda Research Laboratories).

DNA Sequence Analysis

Nucleotide sequence determinations were performed as described in detail by Maxam and Gilbert [18]. The insert of the recombinant plasmid pGAP30 was completely sequenced. DNA fragments were labeled at their 5' ends with [γ -³²P]ATP and polynucleotide kinase (Boehringer) after treatment with calf intestine alkaline phosphatase (Boehringer). Labeled fragments were purified by polyacrylamide gel electrophoresis and eluted by crushing the gel pieces and shaking them overnight at 37 °C in 0.5 M ammonium acetate, 10 mM magnesium acetate, 0.1% sodium dodecyl sulfate and 0.1 mM EDTA.

Abbreviations. GAPDH, glyceraldehyde-3-phosphate dehydrogenase; Pipes, 1,4-piperazine diethanesulfonic acid.

Enzymes. Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12); restriction endonucleases *EcoRI* (EC 3.1.23.13), *HindIII* (EC 3.1.23.21), *HinfI* (EC 3.1.23.22), *Sau3A* (EC 3.1.23.27), *RsaI* (EC 3.1.23.101); reverse transcriptase (EC 2.7.7.49); polynucleotide kinase (EC 2.7.1.78).

For each singly labeled fragment the following four sequence reactions were performed: Guanine with dimethyl-sulfate; guanine + adenine with formic acid; cytosine + thymine with hydrazine; cytosine with hydrazine plus NaCl. Sequence samples were resolved on 22%, 10% and 6% polyacrylamide sequencing gels. The gels were 400 mm long and 0.3 mm thick.

Primer Extension

Primer extension was performed essentially as described by Hagenbüchle et al. [19]. DNA from pGAP30 plasmid was digested with *Hinf*I restriction endonuclease and the resulting fragments were radiolabeled at their 5' ends with polynucleotide kinase. A 780-nucleotide fragment containing the 5'-terminal portion of GAPDH mRNA was isolated by gel electrophoresis and cleaved with restriction endonuclease *Sau*3A yielding a singly labeled fragment of 85 nucleotides which was used as specific primer.

³²P-labeled primer (10–20 pmol) was hybridized with 40 µg of polyadenylated RNA from heart muscle in 50 µl of 80% deionized formamide, 0.4 M NaCl, 0.01 M Pipes pH 6.5 and 0.1% sodium dodecyl sulfate for 60 min at 50°C. The reaction was terminated by adding 200 µl of ethanol at –20°C. The precipitate was washed twice with cold 70% ethanol and dried. The total ethanol precipitate was used in a 50-µl reaction mix containing 50 mM Tris pH 8.3, 10 mM MgCl₂, 0.5 mM dithiothreitol, 1 mM each of the four deoxynucleotide triphosphates and 20 units of reverse transcriptase (Life Sciences, Inc.). The cDNA reaction mix was incubated for 1 h at 42°C. The reaction products were precipitated with ethanol, denatured in 0.1 M NaOH for 10 min at 68°C and layered on a 6% polyacrylamide gel in Tris/borate/EDTA buffer. Primer-extended cDNA was located by autoradiography, eluted from the gel and sequenced as described above.

Secondary Structure Analysis of 3'-Noncoding Sequence

Determination of potentially base-paired regions in the 3'-noncoding sequence of GAPDH mRNA was performed by computer analysis on a VAX 11/780 using the program of Zucker and Steigler [20]. Values for optimal binding energies were taken from Henco [21].

RESULTS

The construction and characterization of cDNA recombinant plasmids containing structural gene sequences of GAPDH from chicken heart muscle has previously been reported [5]. One of these clones, designated pGAP30, contains an insert of 1051 base pairs plus poly(A) stretches of various length at both ends of the inserted DNA. The identity of this plasmid has now been fully confirmed by direct sequence analysis of the entire cloned DNA.

Based on the restriction map of pGAP30 DNA, which we had established earlier [5], the strategy for sequence determination was developed. Double-stranded DNA fragments of defined length were generated with restriction endonucleases as indicated in Fig. 1. The fragments were labeled on both 5' ends with polynucleotide kinase and [γ -³²P]ATP and cleaved by secondary digestion with appropriate restriction enzymes to obtain singly labeled DNA subfragments which were then purified on polyacrylamide gels. Fragments eluted from the gel were sequenced according to Maxam and Gilbert [18]. The internal 320-base-pair fragment obtained

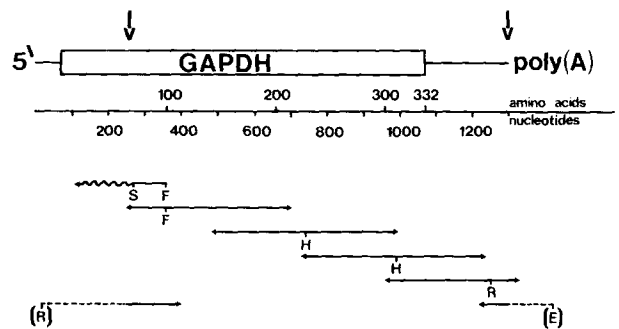


Fig. 1. Sequencing strategy for chicken heart muscle glyceraldehyde-3-phosphate dehydrogenase mRNA. The horizontal box delineates the coding region. The lines extending from the horizontal rectangle represent the length of the noncoding regions of the mRNA. Poly(A) is present at the 3' end in addition to the noncoding sequence. The thick vertical arrows indicate the portion of GAPDH contained in the cDNA recombinant clone pGAP30 [5]. The direction and extent of each sequencing run, starting at the position of the restriction endonuclease sites noted, are indicated by the horizontal arrows. S = *Sau*III; F = *Hinf*I; H = *Hind*III; R = *Rsa*I; E = *Eco*RI. Restriction sites given in parenthesis are located within the PBR322 flanking DNA of clone pGAP30, broken lines stand for PBR322 sequences. The wavy horizontal line represents the sequence determined by 'primer extension' analysis

by digestion with *Hind*III endonuclease (see Fig. 1) was strand-separated and both strands were sequenced separately. The nucleotide sequence of most of the GAPDH mRNA thus derived is shown in Fig. 2. To verify the final sequence, overlapping fragments of the same or opposite polarity, as shown in Fig. 1, were analyzed repeatedly. As revealed by the amino acid sequence predicted from the open reading frame, the cDNA insert of pGAP30 represents coding information for 274 amino acids (numbers 60–332, according to the numbering of the porcine enzyme [10]). The coding information is followed by an UGA termination codon and 228 nucleotides constituting the 3'-noncoding sequence. Additionally, the clone contains about 30 adenosine residues at the left end and a variable number of adenosines ranging from 50 to 120 at the right end which corresponds to the 3' end of the mRNA. The variation in length of poly(A) stretches is presumably due to a recombination-independent deletion of homopolymers since it was observed within single clones even after transformation of *rec*⁻ *Escherichia coli* with purified pGAP30 DNA at very low concentrations (unpublished results). To obtain additional information on the 5' portion of the mRNA that was not represented in the clone, we synthesized the corresponding cDNA using primer extension of cloned DNA. For this experiment pGAP30 DNA was digested with *Hinf*I and *Sau*3A to yield a fragment of 85 nucleotides from the left end of the insert (see Fig. 1). After 5'-end radio-labeling, the fragment served as specific primer for cDNA synthesis using total mRNA of chicken hearts as template. As shown in Fig. 3, the predominant product made in this reaction was the maximally extended cDNA of 344 (± 5) nucleotides which we further subjected to sequence analysis. From the observed length of the primer-extended cDNA we calculated the actual size of chicken GAPDH mRNA to be 1196 (± 5) nucleotides leaving space for 68 (± 5) nucleotides of 5'-noncoding sequence in front of the AUG start codon. The relatively low radioactivity introduced into the primer together with a rather ineffective synthesis of cDNA under the conditions described (only 10–20% of the primer

| | | | | | | | | | | | | | | | | | | | |
|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--|------------|------------|------------|---|------------|------------|------------|------------|------------|
| | | UCU Ser | GGC Gly | AAA Lys | GUC Val | CAG Gln | GUG Val | GUG Val | GCC Ala | 190 GGG Gly | AAG Lys | CUG Leu | UGG Trp | AGA Arg | GAU Asp | GGC Gly | AGA Arg | GGU Gly | GCU Ala |
| 30 AUC Ile | AAU Asn | GAU Asp | CCC Pro | UUC Phe | AUC Ile | GAU Asp | CUG Leu | AAC Asn | UAC Tyr | 200 GCC Ala | CAG Gln | AAC Asn | AUC Ile | AUC Ile | CCA Pro | GCG Ala | UCC Ser | ACU Thr | GGG Gly |
| 40 AUG Met | GUU Val | UAC Tyr | AUG Met | UUC Phe | AAA Lys | UAU Tyr | GAU Asp | UCU Ser | ACA Thr | 210 GCU Ala | GCU Ala | AAG Lys | GCU Ala | GUG Val | GGG Gly | AAA Lys | GUC Val | AUC Ile | CCU Pro |
| 50 CAC His | GGA Gly | CAC His | UUC Phe | AAG Lys | GGC Gly | ACU Thr | GUC Val | AAG Lys | GCU Ala | 220 GAG Glu | CUG Leu | AAU Asn | GGG Gly | AAG Lys | CUU Leu | ACU Thr | GGA Gly | AUG Met | GCU Ala |
| 60 GAG Glu | AAC Asn | GGG Gly | AAA Lys | CUU Leu | GUG Val | AUC Ile | AAU Asn | GGG Gly | CAC His | 230 UUC Phe | CGU Arg | GUG Val | CCA Pro | ACC Thr | CCC Pro | AAU Asn | GUC Val | UCU Ser | GUU Val |
| 70 GCC Ala | AUC Ile | ACU Thr | AUC Ile | UUC Phe | CAG Gln | GAG Glu | CGU Arg | GAC Asp | CCC Pro | 240 GUU Val | GAC Asp | CUG Leu | ACC Thr | UGC Cys | CGU Arg | CUG Leu | GAG Glu | AAA Lys | CCA Pro |
| 80 AGC Ser | AAC Asn | AUC Ile | AAA Lys | UGG Trp | GCA Ala | GAU Asp | GCA Ala | GGU Gly | GCU Ala | 250 GCC Ala | AAG Lys | UAU Thr | GAU Asp | GAU Asp | AUC Ile | AAG Lys | AGG Arg | GUA Val | GUG Val |
| 90 GAG Glu | UAU Tyr | GUU Val | GUG Val | GAG Glu | UCC Ser | ACU Thr | GGU Gly | GUC Val | UUC Phe | 260 AAG Lys | GCU Ala | GCU Ala | GCU Ala | GAU Asp | GGG Gly | CCC Pro | CUG Leu | AAG Lys | GGC Gly |
| 100 ACC Thr | ACC Thr | AUG Met | GAG Glu | AAG Lys | GCU Ala | GGG Gly | GCU Ala | CAU His | CUG Leu | 270 AUC Ile | CUA Leu | GGA Gly | UAC Tyr | ACA Thr | GAG Glu | GAC Asp | CAG Gln | GUU Val | GUC Val |
| 110 AAG Lys | GGU Gly | GGU Gly | GCU Ala | AAG Lys | CGU Arg | GUU Val | AUC Ile | AUC Ile | UCA Ser | 280 UCC Ser | UGU Cys | GAC Asp | UUC Phe | AAU Asn | GGU Gly | GAC Asp | AGC Ser | CAU His | UCC Ser |
| 120 GCU Ala | CCC Pro | UCA Ser | GCC Ala | GAU Asp | GCC Ala | CCC Pro | AUG Met | UUU Phe | GUG Val | 290 UCC Ser | ACC Thr | UUU Phe | GAU Asp | GCG Ala | GGU Gly | GCU Ala | GGC Gly | AUU Ile | GCA Ala |
| 130 AUG Met | GGU Gly | GUC Val | AAC Asn | CAU His | GAG Glu | AAA Lys | UAU Tyr | GAC Asp | AAG Lys | 300 CUG Leu | AAU Asn | GAC Asp | CAU His | UUC Phe | GUC Val | AAG Lys | CUU Leu | GUU Val | UCA Ser |
| 140 UCC Ser | CUG Leu | AAA Lys | AUU Ile | GUC Val | AGC Ser | AAU Asn | GCA Ala | UCG Ser | UGC Cys | 310 UGG Trp | UAU Tyr | GAC Asp | AAU Asn | GAG Glu | UUU Phe | GGA Gly | UAC Tyr | AGC Ser | AAC Asn |
| 150 ACC Thr | ACC Thr | AAC Asn | UGC Cys | CUG Leu | GCA Ala | CCC Pro | UUG Leu | GCC Ala | AAG Lys | 320 CGU Arg | GUU Val | GUG Val | GAC Asp | UUG Leu | AUG Met | GUC Val | CAC His | AUG Met | GCA Ala |
| 160 GUC Val | AUC Ile | CAU His | GAC Asp | AAC Asn | UUU Phe | GGC Gly | AUU Ile | GUG Val | GAG Glu | 330 UCC Ser | AAG Lys | GAG Glu | UGA | GCCAGGCACACAGCC <u>CCUUGCU</u> GCCUAGGGAA | | | | | |
| 170 GGU Gly | CUU Leu | AUG Met | ACC Thr | ACU Thr | GUC Val | CAU His | GCC Ala | AUC Ile | ACA Thr | GCAGGACCCUUGUUGGAG <u>CCUUGCU</u> CUUCACCACCGCU <u>CAGUUCU</u> GCAUCCUGCA | | | | | | | | | |
| 180 GCC Ala | ACA Thr | CAG Gln | AAG Lys | ACG Thr | GUG Val | GAU Asp | GGC Gly | CCC Pro | UCU Ser | GUGAGAGGG <u>CAGUUCU</u> GUUC <u>CCUUCU</u> GUCUCC <u>CCACU</u> CCUCCA <u>UUUCU</u> CCUCCAGC | | | | | | | | | |
| | | | | | | | | | | CUGGGGGAGGUGGGAGGGCUGAUAGA <u>AAACUGAUCU</u> GUUGUGUACCACCU <u>UACAUC</u> | | | | | | | | | |
| | | | | | | | | | | <u>AAUAAAGU</u> UUCACCAUCUGA _n | | | | | | | | | |

Fig. 2. Nucleotide sequence of chicken heart muscle GAPDH mRNA and the encoded amino acid sequence. Numbering of the amino acids was done according to the published primary structure of porcine GAPDH [10]. The amino acids different between the two enzymes are underlined. The conserved sequence A-A-U-A-A-A in the 3'-noncoding region as well as repeated octanucleotides are indicated by boxes. Hyphens representing phosphodiester bonds have been omitted

strand was extended) allowed the analysis of only an additional 114 nucleotides beyond the cloned sequence.

In summary, we have determined 310 amino acids of the chick GAPDH enzyme representing 93% of the total primary structure of 322 amino acids.

DISCUSSION

To compare the amino acid sequence of chicken GAPDH with the known structures of GAPDH from yeast [8], pig [10] and human [11], 310 amino acids of each enzyme were aligned to maximize homology as shown in Fig. 4. The chicken enzyme

contains 332 amino acids, the same number as the pig GAPDH, but differs from the enzyme in yeast, which is one amino acid shorter, and from the human enzyme, which contains 334 amino acids. It can be seen that the degree of conservation of the amino acid sequences of GAPDH during evolution is relatively high within the vertebrate kingdom, whereas the primary structure of the yeast enzyme differs considerably. The highest percentage of conserved amino acids in comparison to chicken GAPDH is exhibited by the porcine enzyme containing 92.6% of sequence homology, followed by the human enzyme with 87%. Only 66.5% of the sequence from yeast GAPDH are identical with the chicken enzyme. Most of the substitutions can be regarded as con-

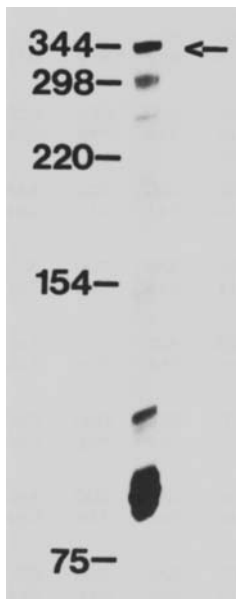


Fig. 3. Autoradiograph of primer-extended cDNA of the 5' end portion of GAPDH mRNA separated on a 6% polyacrylamide sequencing gel. The isolation of the priming DNA fragment and the reaction conditions for cDNA synthesis are described under Experimental Procedures. The maximally extended primer cDNA of 344 nucleotides is indicated by the arrow. Less intensive, shorter bands are presumably premature termination products. The radioactivity at the bottom of the gel represents the unreacted primer fragment (*Hinf*I - *Sau*3A fragment of pGAP30)

servative and could have resulted from single base changes. The amino acids Ser-148, Cys-149, Thr-150, His-176, Lys-183 and Arg-231, which are considered to be essential for the enzymatic activity of GAPDH [11,22], are present in chicken GAPDH at the same position as in all other GAPDH enzymes sequenced so far.

Chicken GAPDH mRNA reveals a non-random utilization of synonymous codons as previously shown for other mRNAs [22-26], suggesting that codons for the same amino acid are not neutral. Fig. 5 shows marked differences in the frequency with which certain codons are used. In particular, the amino acids Leu, Ile, Glu and Arg show a highly significant preference in codon use. Similar preference but to a lesser extent can be observed for Pro, Ala and Gln, whereas codons for His, Tyr, Asn, Asp and Gly are used almost equally. In general, there is an underutilization of codons ending in A or U and a preference for codons ending in G or C, although certain amino acids, e.g. Ala or Arg, behave the opposite way. The bias in codon utilization appears to correlate with the mRNA *per se*, rather than with the organism or tissue, since another distinct mRNAs in the same cell shows a different pattern of preference, as we have shown for the myosin light chain mRNA of chicken cardiac tissue (unpublished results). On the other hand, yeast GAPDH mRNA reveals yet another combination of predominant codons [15].

In the chicken GAPDH mRNA the 3'-noncoding region between the UGA stop codon and the beginning of the poly(A) tract is 228 nucleotides long. The putative signal for polyadenylation, the A-A-U-A-A-A sequence is located 15 nu-

| | | | | | |
|---|---|-----------------|-----------------|-----------------------------|---------------------------------|
| | 22 | | 30 | | 40 |
| Chicken | Ser - Gly - Lys - Val - Gln - Val - Val - Ala - Ile - Asn - Asp - Pro - Phe - Ile - Asp - Leu - Asn - Tyr - Met - | | | | |
| Yeast | Arg - Pro - Asn | | Ser | Asn | Asn Asp Ala |
| Pig | | Asp - Ile | | | His |
| Human | | Asp - Ile | | | His |
| | | | 50 | | 60 |
| Val - Tyr - Met - Phe - Lys - Tyr - Asp - Ser - Thr - His - Gly - His - Phe - Lys - Gly - Thr - Val - Lys - Ala - Glu - Asn - Gly-Lys | | | | | |
| Ala | | | Arg - Tyr - Ala | Glu | Ser - His - Asp - Asp - Lys-His |
| | Gln | | Lys | His | Asp |
| | Glu | | Lys | His | Asp |
| | | 70 | | 80 | |
| Leu - Val - Ile - Asn - Gly - His - Ala - Ile - Thr - Ile - Phe - Gln - Glu - Arg - Asp - Pro - Ser - Asn - Ile - Lys - Trp - Ala - | | | | | |
| Ile - Ile - Val - Asp | Lys - Lys | Ala - Thr - Tyr | | Ala | Leu - Pro Ser - |
| | Asp | Lys | | Ala | Gly |
| | Asp | Lys | Glu | Glu | Gly |
| | | 90 | | 100 | |
| Asp - Ala - Gly - Ala - Glu - Tyr - Val - Val-Glu - Ser - Thr - Gly - Val - Phe - Tyr - Thr - Met - Glu - Lys - Ala - Gly - | | | | | |
| Ser - Gly - Asp - Ser - Val - Ile - Ala - Ile-Asp | | | | Lys - Glu - Leu - Asp - Thr | Gln - |
| | Thr - Ala | | | | |
| | Thr - Ala | | | | |
| | | 110 | | 120 | |
| Ala - His - Leu - Lys - Gly - Gly - Ala - Lys - Arg - Val - Ile - Ile - Ser - Ala - Pro - Ser - Ala - Asp - Ala - Pro - Met - | | | | | |
| Lys | Ile - Asp - Ala | Lys | Val | Thr | Ser - Thr |

Fig. 4

Ile - Val

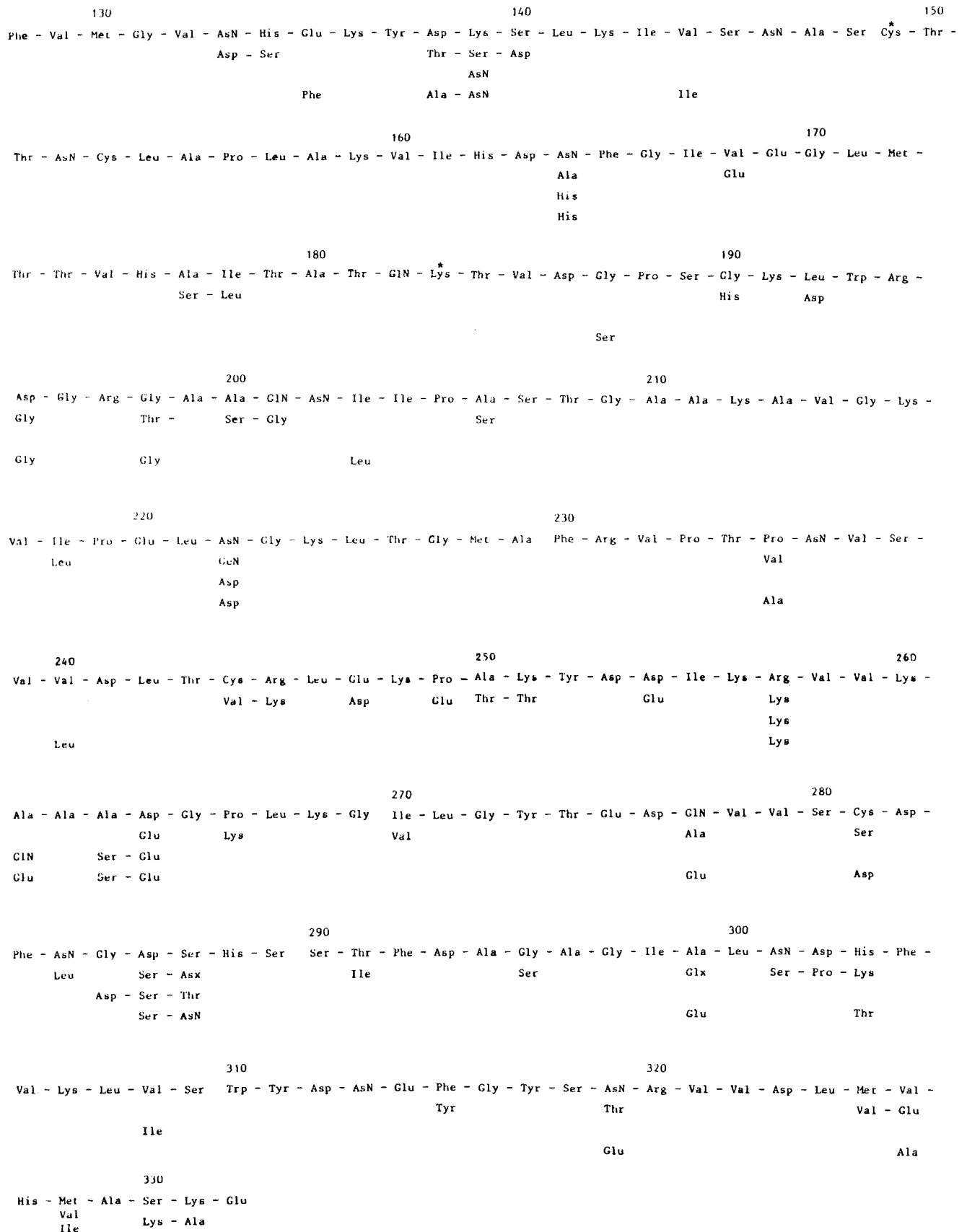


Fig. 4. Alignment of the amino acid sequence of GAPDH from chicken, yeast, pig and human. The alignment begins with amino acid 22 of the porcine enzyme since the first 21 amino acids of chicken GAPDH have not been determined. To facilitate comparison the numbering of the pig GAPDH was used for all four enzymes shown. Regions of heterogeneity are indicated. Amino acids identical for the enzyme of all four organisms are shown only for the chicken GAPDH

| | | | | | | | |
|------------|---|------------|---|-----|---|------------|--|
| <u>Phe</u> | <ul style="list-style-type: none"> U U U 4 U U C 9 U U A 0 U U G 2 | <u>Ser</u> | <ul style="list-style-type: none"> U C U 4 U C C 7 U C A 3 U C G 1 | Tyr | <ul style="list-style-type: none"> U A U 5 U A C 4 U A A - U A G - | Cys | <ul style="list-style-type: none"> U G U 1 U G C 3 U G A - U G G - |
| <u>Leu</u> | <ul style="list-style-type: none"> C U U 4 C U C 0 C U A 1 C U G 10 | <u>Pro</u> | <ul style="list-style-type: none"> C C U 1 C C C 8 C C A 3 C C G 0 | His | <ul style="list-style-type: none"> C A U 6 C A C 6 C A A 0 C A G 5 | <u>Arg</u> | <ul style="list-style-type: none"> C G U 5 C G C 0 C G A 0 C G G 0 |
| <u>Ile</u> | <ul style="list-style-type: none"> A U U 3 A U C 15 A U A 0 A U G 0 | <u>Thr</u> | <ul style="list-style-type: none"> A C U 7 A C C 8 A C A 5 A C G 1 | Asn | <ul style="list-style-type: none"> A A U 8 A A C 8 A A A 8 A A G 19 | <u>Ser</u> | <ul style="list-style-type: none"> A G U 0 A G C 4 A G A 2 A G G 1 |
| <u>Met</u> | <ul style="list-style-type: none"> A U G 0 | <u>Ala</u> | <ul style="list-style-type: none"> G C U 16 G C C 9 G C A 6 G C G 2 | Asp | <ul style="list-style-type: none"> G A U 11 G A C 10 G A A 0 G A G 12 | <u>Gly</u> | <ul style="list-style-type: none"> G G U 9 G G C 7 G G A 5 G G G 8 |

Fig. 5. Codon usage for synonymous amino acids in GAPDH mRNA from chicken heart muscle. The frequency of use of each codon, as revealed by the sequenced part of GAPDH mRNA from chicken heart, is indicated. Amino acids showing a highly significant preference are underlined with solid lines, marginal preference is indicated by dotted lines

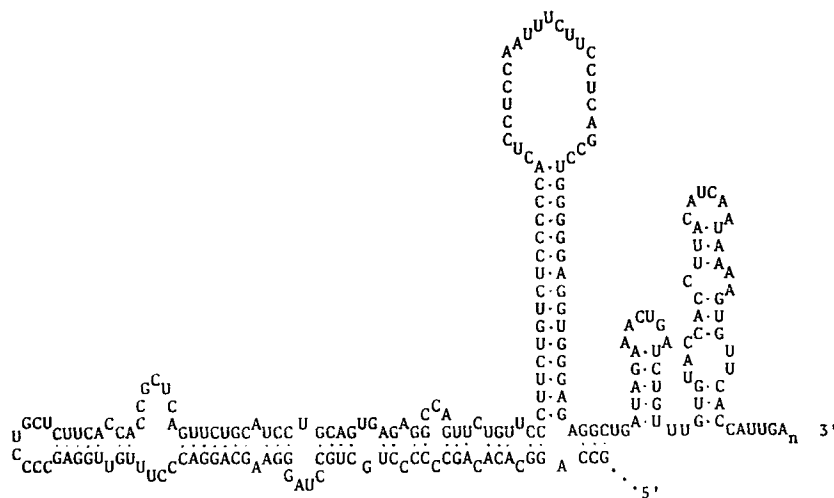


Fig. 6. Possible secondary structure of the 3'-noncoding sequence of chicken GAPDH mRNA. Maximal base-pairing of the 228 nucleotides of the 3' trailing sequence was determined by computer analysis as described under Experimental Procedures. No correction was made for single nucleotide loops or bubbles

cleotides upstream from the 3' end, as expected from most known mRNA sequences [23]. The potential secondary structure of the 3'-trailing sequence, as shown in Fig. 6, was determined by computer analysis. Large portions of the sequence are able to form base pairs, thus leading to a rather compact shape of the trailer. It is particularly interesting to note that a very stable loop structure is formed by a stem of 15 base pairs and 23 single-stranded nucleotides in the loop. In the secondary structure, drawn by the computer, this large helical stem is located very close to the UGA stop codon of the mRNA. The arrangement could suggest a hypothetical role in translation, since it might support termination, e.g. the ribosomes running off the mRNA. An alternative function of the loop could be envisaged in termination of transcription assuming that the 3' terminus of the mRNA coincides with the end of the primary transcript. As has been shown for the attenuation of bacterial operons encoding amino acid bio-

synthetic enzymes [27] or the premature termination of late SV40 transcription [28], potential stem-end-loop configurations of the transcript seem to be involved in termination of transcription.

Another interesting feature of the 3'-noncoding region is the occurrence of several repeated oligonucleotide sequences, of which the two longest are octanucleotides. Although the function of these repeats remains obscure, one could speculate that they might be important for the formation of spontaneous deletions in noncoding regions, which are subject to high mutation rates compared to coding regions of genes. The involvement of similar short repeats for deletion mutations has recently been documented in bacteria [29].

We would like to thank B. Förtisch (Max Planck Institute of Biochemistry, Martinsried, FRG) for the computer analysis of the secondary structure. H. H. Arnold is indebted to Dr H. Diggelmann

and Dr G. Fey (*Institut Suisse de Recherche Experimentales sur le Cancer*, Lausanne) for the hospitality at their institute, where most of the sequencing work was conducted. We also thank Dr H. Marquardt for continuous support and help in preparing the manuscript. This work was supported by *Deutsche Forschungsgemeinschaft* grant Ar 115/5. Part of the sequence has been published previously [5].

REFERENCES

1. Harris, J. J. & Waters, M. (1976) in *The Enzymes* (Boyer, P. D., ed.) vol. 13, pp. 1–49, Academic Press, New York, San Francisco, London.
2. Jakowlew, S. B., Khandekar, P. L., Datta, K., Arnold, H. H., Narula, S. K. & Siddiqui, M. A. Q. (1982) *J. Mol. Biol.* **156**, 673–682.
3. Arnold, H. H. & Siddiqui, M. A. Q. (1979) *Biochemistry*, **18**, 647–654.
4. Arnold, H. H. & Siddiqui, M. A. Q. (1979) *Biochemistry*, **18**, 5641–5647.
5. Arnold, H. H., Domdey, H., Wiebauer, K., Datta, K. & Siddiqui, M. A. Q. (1982) *J. Biol. Chem.* **257**, 9872–9877.
6. Harris, J. H. & Perham, R. N. (1968) *Biochemistry*, **7**, 4011–4022.
7. Walker, J. E., Carne, A. F., Runswick, M. J., Bridgen, J. & Harris, J. I. (1980) *Eur. J. Biochem.* **108**, 549–565.
8. Jones, G. M. T. & Harris, J. I. (1972) *FEBS Lett.* **22**, 185–189.
9. Davidson, B. E., Sajgo, M., Noller, H. F. & Harris, J. I. (1967) *Nature (Lond.)* **216**, 1181.
10. Harris, J. I. & Perham, R. N. (1968) *Nature (Lond.)* **219**, 1025–1028.
11. Nowak, K., Wolny, M. & Banás, T. (1981) *FEBS Lett.* **134**, 143–146.
12. Lebercz, H. G. & Rutter, W. J. (1967) *Science (Wash. DC)* **157**, 1198–1200.
13. Kirschner, K. & Voigt, B. (1968) *Hoppe-Seyler's Z. Physiol. Chem.* **349**, 632–644.
14. Holland, M. J. & Westhead, E. W. (1973) *Biochemistry*, **12**, 2276–2281.
15. Holland, J. P. & Holland, M. J. (1980) *J. Biol. Chem.* **255**, 2596–2605.
16. McLeod, A. R. (1982) *Eur. J. Biochem.* **119**, 353–358.
17. Holland, L. J. & McCarthy, B. J. (1980) *Biochemistry*, **19**, 2965–2976.
18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
19. Hagenbüchle, O., Boverly, R. & Young, R. A. (1980) *Cell*, **21**, 179–187.
20. Zuker, M. & Stiegler, P. (1982) *Nucleic Acid Res.* **9**, 133–148.
21. Henco, K. (1979), Ph.D. Thesis, Darmstadt, FRG.
22. Park, J. H., Aguello, C. F. & Mathew, E. (1966) *J. Biol. Chem.* **242**, 769.
23. Proudfoot, N. Y. & Brownlee, G. G. (1976) *Nature (Lond.)* **263**, 211–214.
24. Efstradiadis, A., Kafatos, F. C. & Maniatis, T. (1977) *Cell*, **10**, 571–585.
25. Suzuki, Y. & Brown, D. D. (1972) *J. Mol. Biol.* **63**, 409.
26. Hampe, A., Therwath, A., Soriano, P. & Galibert, F. (1982) *Gene*, **14**, 11–21.
27. Stroynowski, I. & Yanofsky, C. (1982) *Nature (Lond.)* **298**, 34–38.
28. Skolnik-David, H., Hay, N. & Aloni, Y. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 2743–2747.
29. Albertini, A. M., Hofer, M., Calos, M. P. & Miller, J. H. (1982) *Cell*, **29**, 319–328.

H. H. Arnold* and H. Klapthor, Institut für Toxicologie der Universität Hamburg, Grindelallee 117, D-2000 Hamburg 13, Federal Republic of Germany

K. Wiebauer, Department of Biochemistry, University of Michigan, Ann Arbor, Michigan, USA 48104

H. Domdey, Department of Chemistry, University of California, Alcalá Park, San Diego, California, USA 92110

* To whom requests for reprints should be addressed.