Vol. 118, No. 3, 1984 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 14, 1984 Pages 767-773

COMPLETE NUCLEOTIDE SEQUENCE OF THE MESSENGER RNA CODING FOR CHICKEN MUSCLE GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE

F. Panabières, M. Piechaczyk, B. Rainer*, Ch. Dani, Ph.Fort, S. Riaad, L.Marty, J.L.Imbach*, Ph. Jeanteur and J.M. Blanchard

Laboratoire de Biologie Moléculaire,
Université des Sciences et Techniques du Languedoc,
Place E. Bataillon, 34060 Montpellier, France

Laboratoire de Biochimie, Centre Paul Lamarque,
BP 5054, 34033, Montpellier, France

*Laboratoire de chimie Bio-Organique,
Université des Sciences et Techniques du Languedoc,
Place E. Bataillon, 34060 Montpellier, France

Received November 21, 1983

The complete nucleotide sequence for chicken glyceraldehyde-3-phosphate dehydrogenase mRNA has been determined, thereby extending the longest such sequence previously reported (Dugaiczyk et al. Biochemistry, 1983, 22, 1605-1613) by 27 nucleotides. The complete mRNA with the exclusion of poly(A) is 1284 nucleotides long and contains 56 nucleotides of 5' non coding sequence and 229 nucleotides of 3' non coding region. Knowledge of the complete sequence allows us to propose secondary structures models which may be of biological significance.

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) consists of four identical subunits whose amino acid sequence has been determined for a wide variety of organisms as different as Bacillus stearothermophilus and man. This enzyme catalyzes the oxidative phosphorylation of D glyceraldehyde-3-phosphate into 1,3-bis phosphoglycerate and therefore plays a very important role in the metabolism of most cells.

Examination of GAPDH mRNA in different tissues from chicken indicated the existence of a single electrophoretic band (1) with approximately the same abundancy. This observation is not in accord with the abundance of the enzyme which is much higher in muscle cells where it constitutes a considerable fraction of total proteins (2). This discrepancy could be accounted for by subtle differences in mRNA structure overlooked by the electrophoretic analysis as well as by differences in gene activity and/or mRNA maturation. As a first step towards understanding the regulation of chicken GAPDH gene expression, we have established the complete sequence of the GAPDH mRNA from chicken leg muscle using a combination of various strategies. This sequence now includes the 27 nucleotides at the 5' end

0006-291X/84 $1.50

which were missing in the nearly complete sequence reported by Dugaiczyk et al.(3). These additional nucleotides have allowed us to propose secondary structure models which might be of physiological significance.

MATERIALS AND METHODS

a) Isolation and fractionation of RNA.

RNA was isolated from 6 months old chicken leg muscle using a guanidinium thiocyanate extraction procedure communicated to us by G. Cathala. Briefly, pieces of frozen muscle were homogenized in 9 volumes of 100mM Tris-HCl pH 7.5 (25°C), 5M GuSCN, 10mMEDTA, 1.4M β-mercapto-ethanol. The homogenate is then adjusted to 3M LiCl and 6M urea with a five-fold dilution and allowed to stand overnight at 4°C. RNA is recovered by centrifugation for 30 min. at 10,000 rpm (Sorvall HB 4 rotor), solubilized in 1 mM EDTA, 0.1% SDS and 10mM Tris-HCl pH 7.5 (25°C) and then extracted successively with one volume of phenol and one volume of chloroform. After ethanol precipitation, Poly(A) containing RNA was isolated by two cycles of oligo(dT) cellulose chromatography (4). Enrichment in GAPDH mRNA was obtained by centrifugation for 20 hours at 40,000 rpm (Kontron TST 41 rotor) on a 15-30 % (w/v) sucrose gradient in 10 mM Tris-HCl pH 7.4 (25°C), 1 mM EDTA, 0.1 % SDS. Fractions corresponding to 14S were pooled, ethanol precipitated twice and stored in water at - 20°C.

b) Plasmid DNA preparation and labeling.

Recombinant plasmid pGPD-1 was a generous gift from Dr. A.R. McLeod (1) and was prepared following the alkaline method described by McMaster (5). The 850 bp insert was purified by a centrifugation step of 22 hours at 25,000 rpm (SW27 rotor) on a 5-20 % (w/v) sucrose gradient in 10 mM Tris-HCl pH 8 (25°C), 100 mM NaCl, 2 mM EDTA. DNA fragments were labeled at their 5' ends with (γ ^{32}P) ATP and polynucleotide kinase (Boehringer) after treatment with intestine alkaline phosphatase (Boehringer). Labeled fragments were purified by polyacrylamide gel electrophoresis and eluted by diffusion overnight at 37°C from crushed gel pieces in 0.5 M ammonium acetate, 10 mM magnesium acetate, 0.1 % SDS and 0.1 mM EDTA.

c) Oligodeoxynucleotide synthesis.

Two synthetic oligonucleotides were synthesized by the phosphotriester method (6) on a solid-phase polystyrene resin (Bachem, California) and purified by high performance liquid chromatography through a Waters C18 column (7). Oligo I is: 5' CTCATGGTTGACACCCAT 3' and Oligo II is: 5' CACCACTTGGACTTTGCC 3'.

d) Primer extension of labeled oligo II with reverse transcriptase.

10 µg of partially purified GAPDH mRNA was hybridized with 100 ng ^{32}P labeled oligo II (5.10^6 cpm/ µg) for 10 min at 65°C in 25 µl of 50 mM Tris-HCl pH 8.6 (25°C), 6 mM $MgCl_2$, and submitted to reverse transcription for one hour at 42°C in the presence of 0.4 mM each of the four deoxynucleotide triphosphates, 5 mM dithiothreitol and 20 units of reverse transcriptase (Life Sciences Inc.). The reaction products were denatured for 15 min at 100°C in 10 mM NaOH, 50 % formamide and layered on a 10 % polyacrylamide sequencing gel (8). Primer-extended cDNA was located by autoradiography, eluted from the gel and ethanol precipitated as above.

e) Sequence analysis.

Oligo II extended cDNA and DNA restriction fragments were sequenced using the Maxam and Gilbert procedure (9).

Sequencing of the reverse transcripts primed with oligo I or oligo II with the base specific chain termination method of Sanger et al. (10) was carried out with the following modifications. 4 µg RNA and 100 ng primer (oligo I or oligo II) were allowed to react for 5 min at 65°C in 10 µl 50 mM Tris-HCl pH 8.6 (25°C), 50 mM NaCl, 6 mM $MgCl_2$ and 5 mM dithiothreitol (R buffer). The volume was adjusted to 16 µl by adding 1 µl 10 X R buffer and

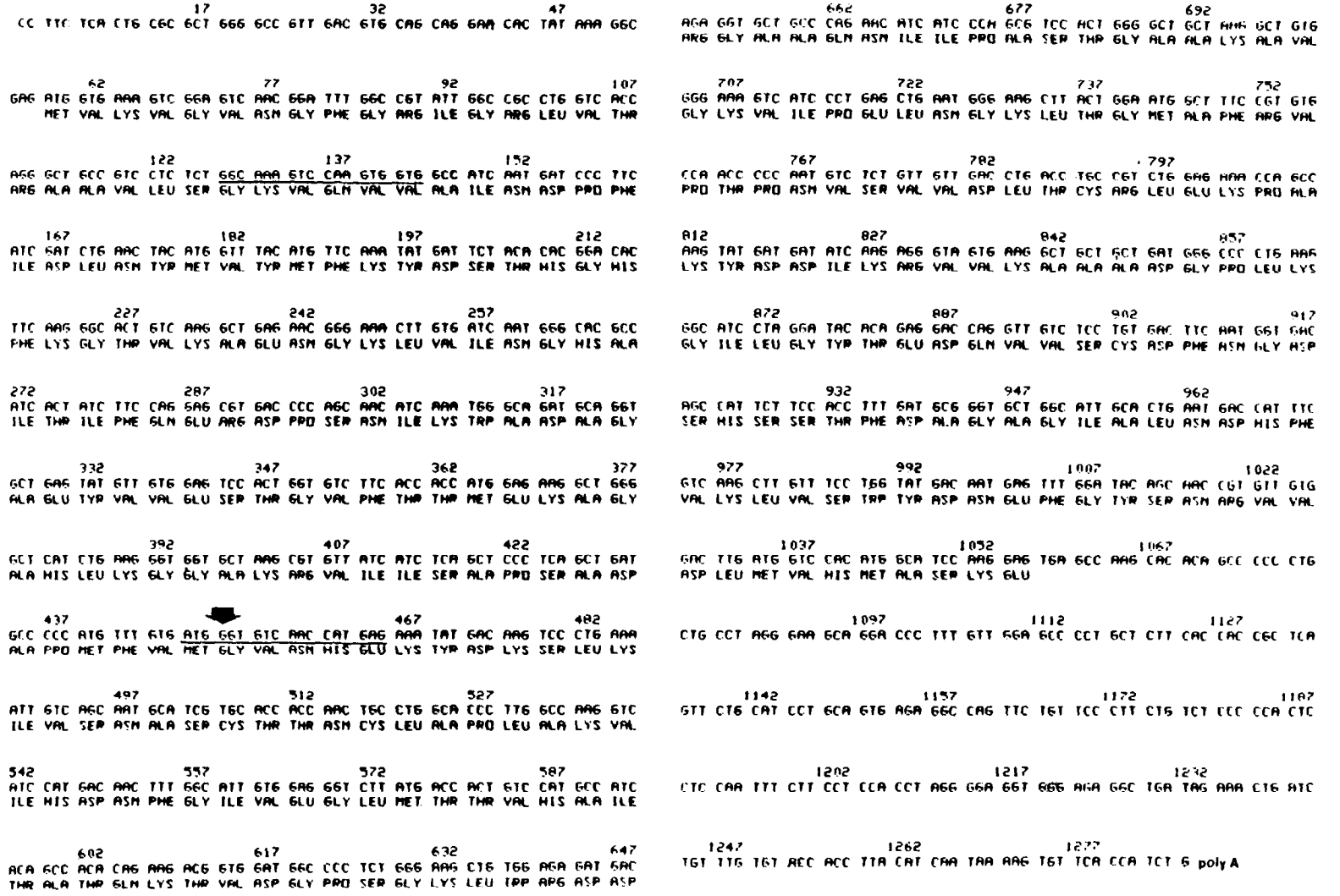
5 µl of a mixture containing 0.5 mM dTTP, dGTP and dATP. Finally this mixture was used to dissolve 40 µCi of dried (α^{32}P) dCTP (1,000 Ci/mMole) and dispatched into 3 µl aliquotes in 4 separate vials containing 1 µl reverse transcriptase (5 units) and 1 µl of: 0.5 mM ddTTP for T specific reaction, 0.125 mM ddGTP for G specific reaction, 0.25 mM ddATP for A specific reaction and 0.01 mM ddCTP for C specific reaction. The reactions were kept for 15 min at 42°C before the addition of 3 µl of a 0.5 mM solution of each four dNTP containing 5 units reverse transcriptase. Incubation was for an additional 15 min at 42°C. Reactions were terminated by adding 8 µl formamide, 10 mM EDTA, 0.01 % xylene cyanol and 0.01 % bromophenol blue. The products were separated on 0.4 mm thick, 5 and 6 % polyacrylamide gels containing 8 M urea which were dried before autoradiography (8).

RESULTS AND DISCUSSION

The starting material was the partial cDNA (pGPD-1) clone utilized by McLeod (1) and kindly donated to us. The 850 bp insert was excised by EcoRI, digested with several restriction enzymes and the resulting map was used to establish the sequencing strategy. Double-stranded DNA fragments were generated and labeled with polynucleotide kinase and (γ^{32}P)ATP and then cleaved by secondary digestion with appropriate restriction enzymes to obtain singly labeled DNA subfragments which were purified through polyacrylamide gels. Eluted fragments were sequenced according to Maxam and Gilbert (9). In order to confirm this sequence, several types of restriction fragments were subcloned in M13 mp8 and mp9 and sequenced by the chain terminators method (10)(data not shown).

The partial protein sequence thus derived (Figure 1) allowed unambiguous identification of GAPDH by comparison to known sequences from different sources (11). It turned out that the first triplet in the clone encoded a glycine residue immediately following a methionine which appeared to be conserved in all GAPDH sequenced so far. We therefore designed an 18 mer oligonucleotide (oligo I) corresponding to the 15 nucleotides at the 5' end of the cDNA clone plus the three residues corresponding to this methionine residue. In order to extend the sequence of GAPDH mRNA beyond the 5' limit of the pGPD-1 clone, this oligonucleotide was used to prime specifically the synthesis of cDNA on partially purified GAPDH mRNA (12-14). We used 2',3'-dideoxy nucleoside triphosphates as specific chain terminators for reverse transcriptase (10) to derive the sequence of the next 330 nucleotides.

The occurrence of premature termination at several sites as well as the difficulty to unambiguously read the sequence beyond 330 nucleotides prompted us to synthesize a second oligodeoxynucleotide (oligo II) to complete the sequence down to the 5' terminus of the mRNA. This sequence was obtained by the chain-terminators method used above and confirmed by chemical sequencing of the largest discrete cDNA produced by run-off reverse transcription primed by 5' ^{32}P labeled oligo II (not shown). The size


17 32 47
CC TTC TCA CTG CGC GCT GGG GCC GTT GAC GTG CAG CAG GAA CAC TAT AAA GGC
62 77 92 107
GAG ATG GTG AAA GTC GGA GTC AAC GGA TTT GGC CGT ATT GGC CGC CTG GTC ACC
MET VAL LYS VAL GLY VAL ASN GLY PHE GLY ARG ILE GLY ARG LEU VAL THR
122 137 152
AGG GCT GCC GTC CTC TCT GGC AAA GTC CAA GTG GTG GCC ATC AAT GAT CCC TTC
ARG ALA ALA VAL LEU SER GLY LYS VAL GLN VAL VAL ALA ILE ASN ASP PRO PHE
167 182 197 212
ATC GAT CTG AAC TAC ATG GTT TAC ATG TTC AAA TAT GAT TCT ACA CAC GGA CAC
ILE ASP LEU ASN TYR MET VAL TYR MET PHE LYS TYR ASP SER THR HIS GLY HIS
227 242 257
TTC AAG GGC ACT GTC AAG GCT GAG AAC GGG AAA CTT GTG ATC AAT GGG CAC GCC
PHE LYS GLY THR VAL LYS ALA GLU ASN GLY LYS LEU VAL ILE ASN GLY HIS ALA
272 287 302 317
ATC ACT ATC TTC CAG GAG CGT GAC CCC AGC AAC ATC AAA TGG GCA GAT GCA GGT
ILE THR ILE PHE GLN GLU ARG ASP PRO SER ASN ILE LYS TRP ALA ASP ALA GLY
332 347 362 377
GCT GAG TAT GTT GTG GAG TCC ACT GGT GTC TTC ACC ACC ATG GAG AAG GCT GGG
ALA GLU TYR VAL VAL GLU SER THR GLY VAL PHE THR THR MET GLU LYS ALA GLY
392 407 422
GCT CAT CTG AAG GGT GCT AAG CGT GTT ATC ATC TCA GCT CCC TCA GCT GAT
ALA HIS LEU LYS GLY ALA LYS ARG VAL ILE ILE SER ALA PRO SER ALA ASP
437 467 482
GCC CCC ATG TTT GTG ATG GGT GTC AAC CAT GAG AAA TAT GAC AAG TCC CTG AAA
ALA PRO MET PHE VAL MET GLY VAL ASN HIS GLU LYS TYR ASP LYS SER LEU LYS
497 512 527
ATT GTC AGC AAT GCA TCG TGC ACC ACC AAC TGC CTG GCA CCC TTG GCC AAG GTC
ILE VAL SER ASN ALA SER CYS THR THR ASN CYS LEU ALA PRO LEU ALA LYS VAL
542 557 572 587
ATC CAT GAC AAC TTT GGC ATT GTG GAG GGT CTT ATG ACC ACT GTC CAT GCC ATC
ILE HIS ASP ASN PHE GLY ILE VAL GLU GLY LEU MET THR THR VAL HIS ALA ILE
602 617 632 647
ACA GCC ACA CAG AAG ACG GTG GAT GGC CCC TCT GGG AAG CTG TGG AGA GAT GAC
THR ALA THR GLN LYS THR VAL ASP GLY PRO SER GLY LYS LEU TRP ARG ASP ASP
662 677 692
AGA GGT GCT GCC CAG AAC ATC ATC CCA GCG TCC ACT GGG GCT GCT AAG GCT GTG
ARG GLY ALA ALA GLN ASN ILE ILE PRO ALA SER THR GLY ALA ALA LYS ALA VAL
707 722 737 752
GGG AAA GTC ATC CCT GAG CTG AAT GGG AAG CTT ACT GGA ATG GCT TTC CGT GTG
GLY LYS VAL ILE PRO GLU LEU ASN GLY LYS LEU THR GLY MET ALA PHE ARG VAL
767 782 797
CCA ACC CCC AAT GTC TCT GTT GTT GAC CTG ACC TGC CGT CTG GAG AAA CCA GCC
PRO THR PRO ASN VAL SER VAL VAL ASP LEU THR CYS ARG LEU GLU LYS PRO ALA
812 827 842 857
AAG TAT GAT GAT ATC AAG AGG GTA GTG AAG GCT GCT GCT GAT GGG CCC CTG AAG
LYS TYR ASP ASP ILE LYS ARG VAL VAL LYS ALA ALA ALA ASP GLY PRO LEU LYS
872 887 902 917
GGC ATC CTA GGA TAC ACA GAG GAC CAG GTT GTC TCC TGT GAC TTC AAT GGT GAC
GLY ILE LEU GLY TYR THR GLU ASP GLN VAL VAL SER CYS ASP PHE ASN GLY ASP
932 947 962
AGC CAT TCT TCC ACC TTT GAT GCG GGT GCT GGC ATT GCA CTG AAT GAC CAT TTC
SER HIS SER SER THR PHE ASP ALA GLY ALA GLY ILE ALA LEU ASN ASP HIS PHE
977 992 1007 1022
GTC AAG CTT GTT TCC TGG TAT GAC AAT GAG TTT GGA TAC AGC AAC CGT GTT GTG
VAL LYS LEU VAL SER TRP TYR ASP ASN GLU PHE GLY TYR SER ASN ARG VAL VAL
1037 1052 1067
GAC TTG ATG GTC CAC ATG GCA TCC AAG GAG TGA GCC AAG CAC ACA GCC CCC CTG
ASP LEU MET VAL HIS MET ALA SER LYS GLU
1097 1112 1127
CTG CCT AGG GAA GCA GGA CCC TTT GTT GGA GCC CCT GCT CTT CAC CAC CGC TCA
1142 1157 1172 1187
GTT CTG CAT CCT GCA GTG AGA GGC CAG TTC TGT TCC CTT CTG TCT CCC CCA CTC
1202 1217 1232
CTC CAA TTT CTT CCT CCA CCT AGG GGA GGT GGG AGA GGC TGA TAG AAA CTG ATC
1247 1262 1277
TGT TTG TGT ACC ACC TTA CAT CAA TAA AAG TGT TCA CCA TCT G poly A

of this cDNA (140 $\pm$ 5 nucleotides) corresponds to what should be expected from the size of the mRNA (about 1300 nucleotides without poly(A)) as determined in denaturing gels by reference to DNA restriction fragments used as markers (not shown).

The complete nucleotide sequence is shown in Figure 1 along with the protein sequence it encodes. It contains 56 non-coding nucleotides at the 5' end, 999 coding nucleotides corresponding to 333 amino-acids, a termination codon (UGA), and 230 non-coding nucleotides at the 3' end. While this work was being completed, two partial sequences have been published (3,15) which are in full agreement with our sequence as far as the coding portion of the mRNA is concerned. However, several minor differences located in the 3' non-coding region, probably due to individual polymorphism, can be noticed.

Computer analysis (16) of possible base-pairings involving the now complete 5' region of the mRNA has suggested several possibilities for the secondary structure of this molecule. A plausible model is represented in Figure 2a, where the AUG initiation codon is bulging out of a quite stable stem in an easily available configuration. 14 nucleotides upstream from the 3' end, the 3' non-coding region exhibits the AAUAAAA which contains the canonical sequence AAUAAA supposed to function as a signal for poly-adenylation. Large portions of this 3' terminal sequence are able to form base pairs and could for example lead to the structure shown in Figure 2b where the UGA stop codon is close to a very stable double-stranded region. This structure is quite similar to that already proposed by Domdey et al. (15) and it is interesting to note that all the differences observed with the two published partial sequences (3,15) reside in the 3' non-coding region and are conservative with regard to this hypothetical secondary structure. Another interesting feature of the 3' non-coding region is its ability to interact with the 5' end thereby generating a rather peculiar configuration which could regulate the access of ribosomes (Figure 2c).

The aminoacid sequence of chicken GAPDH derived from the nucleotide sequence is in agreement with its extreme evolutionary conservation. For example, the pig sequence shows only 26 differences out of a total of 333 aminoacids and 19 of these differences can be accounted for by single nucleotide substitutions. Moreover, all the essential amino acids thought to be important for catalysis such as Cys-149, His-176 or Lys-183, are

---

Figure 1 : Complete nucleotide sequence of chicken leg muscle GAPDH mRNA and its translated protein sequence.
The arrow indicates the 5' limit of the cDNA pGPD-1 clone. Underlined sequences are those which served to design the two oligonucleotides used as primers for reverse transcription

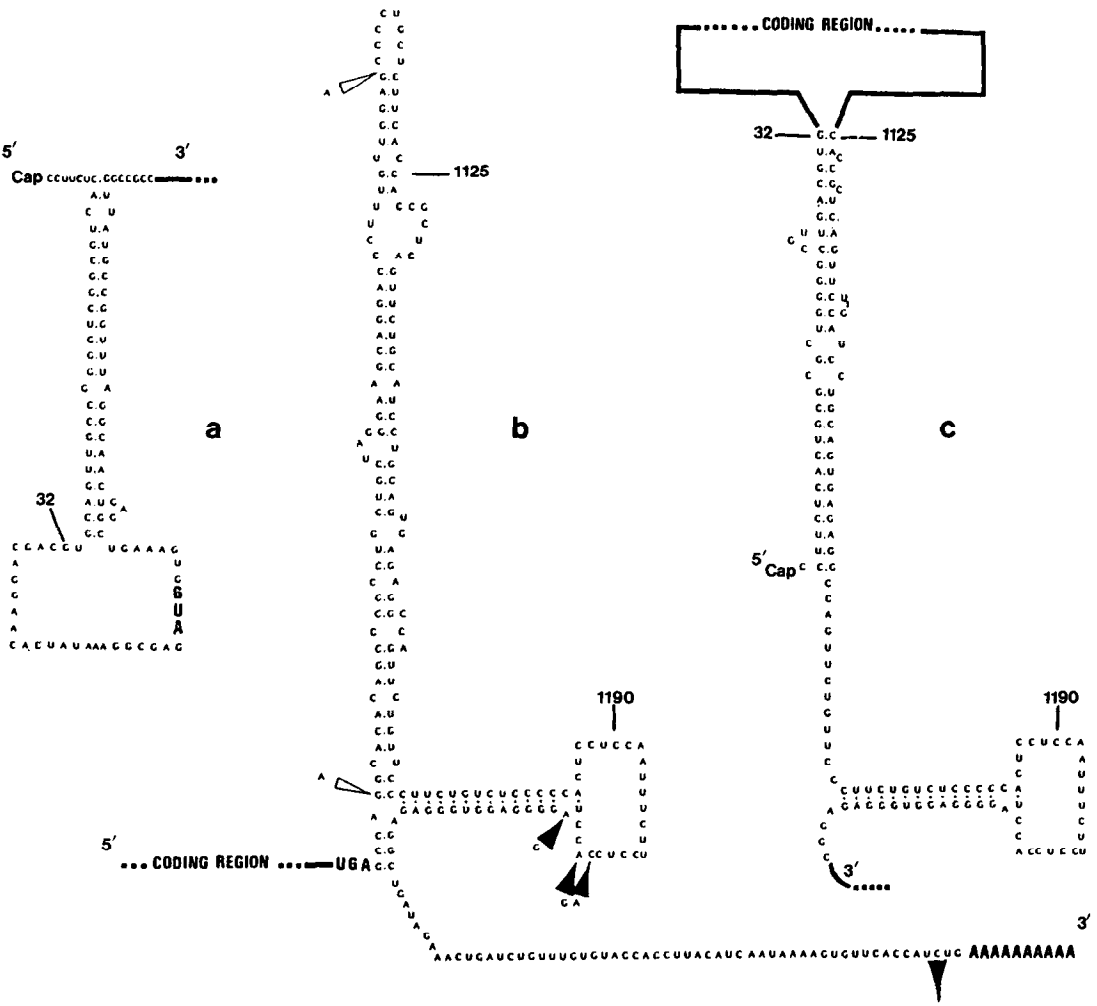


Figure 2 : A secondary structure model for the 5' (a) and 3' (b) noncoding regions of chicken GAPDH mRNA and their possible interaction (c).

The secondary structure models were obtained by computer search for dyad symmetry regions within the 200 nucleotides at the 5' end (a) and at the 3' end (b) or pairing capabilities between them (c) using the SEQ computer program (16) made available by Intelligenetics (Palo Alto). Open triangles refer to differences observed between this work and that of Dugaiczyk et al. (15) and closed triangles between this work and that of Domdey et al. (14). The inverted arrow close to the poly(A) in (b) indicates an additional C residue in our sequence.

conserved except for the replacement of this latter aminoacid by Arg in Bacillus stearothermophilus (17) and Thermus aquaticus (18). However, the overall nucleotide homology between the different vertebrate mRNA coding for GAPDH is much less than that observed at the level of protein sequence. For example, if we compare the nucleotide sequence corresponding to the two known functional domains of GAPDH (19), the NAD binding site (amino acids 1-147) and the catalytic site (amino acids 148-333), there is a 36 % homology

between yeast and chicken for the first 400 nucleotides which increases to 50 % for the other part of the coding region. Moreover, the two oligonucleotides used here have failed to prime reverse transcription on GAPDH mRNA from CHO as well as HeLa cells, even in non-stringent conditions (20°C). This indicates a high degree of sequence divergence at least at these two loci.

ACKNOWLEDGEMENTS

We are grateful to A.R. McLeod for his generous gift of pGPD-1. We thank G. Cathala for helpful discussions, M. Fize for typing the manuscript, Ch. Aymard and A. Vie for excellent technical assistance. This work was supported by grants from the CNRS, INSERM DGRST, Ligue Contre le Cancer, Caisse Régionale d' Assurance Maladie and Fondation pour la Recherche Medicale.

REFERENCES

1) McLeod, A.R.M. (1982) Eur. J. Biochem. 119, 353-358.

2) Shonk, C.E. and Boxer, G.E. (1964) Cancer Res. 24, 709-719.

3) Dugaiczyk, A., Haron, J.A., Stone, E.M., Dennison, O.E., Rothblum, K.N. and Schwartz, R.J. (1983) Biochemistry 22, 1605-1613.

4) Bantle, J.A., Maxwell, I.M. and Halm, W.E.(1976). Anal. Biochem. 72, 413-427.

5) McMaster, G.K., Samulski, R.J., Stein, J.L. and Stein, G.S. (1980) Anal. Biochem. 109, 47-54.

6) Ito, K., Ike, Y., Ikuta, S. and Itakura, K. (1982) Nucl. Ac. Res. 10, 1755-1769.

7) McFarland, G. and Borer, P. (1979) Nucl. Ac. Res. 7, 1067-1080.

8) Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.

9) Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564.

10) Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.

11) Dayhoff,M.O.(1978) in Atlas of Protein Sequence (Dayhoff ed., National Biomedical Rsearch Foundation, Silver Spring, Md.) 5, 62-63.

12) Hamlyn, P.H., Brownlee, G.G., Chen-ChiCheng, Grait, M.J. and Milstein, C. (1978) Cell 15, 1067-1075.

13) Akusjärvi, G. and Petterson, V. (1979) Cell 16, 841-850.

14) Hagenbüchle, O., Bovery, R. and Young, R.A. (1980) Cell 21, 179-187.

15) Domdey, H., Wiebauer, K., Klapthor, H. and Arnold, H.H. (1983) Eur. J. Biochem. 131, 129-135.

16) Brutlag, D.L., Clayton, J., Friedland, P. and Kedes, L.(1982) Nucl. Ac. Res., 10, 279-294.

17) Walker, J.E., Carne, A.F., Runswick, M.J., Bridgen, J. and Harris, J.L. (1080) Eur. J. Biochem., 108, 549-565.

18) Hocking, J.D. and Harris, J.L. (1980) Eur. J. Biochem. 108, 567-579.

19) Olsen, K.W., Moras, D., Rossman, M.G. and Harris, J.I. (1975) J. Biol. Chem., 250, 9313-9321.