Volume 12 Number 19 1984 Nucleic Acids Research

Isolation and nucleotide sequence of a full-length cDNA coding for aldolase B from human liver

G.Paolella, R.Santamaria, P.Izzo, P.Costanzo and F.Salvatore*

Istituto di Scienze Biochimiche, II Facoltà di Medicina e Chirurgia, Università di Napoli, Via S. Pansini 5, 80131 Napoli, Italy

Received 29 June 1984; Revised and Accepted 14 September 1984

ABSTRACT

Two recombinant clones, pA2 and pA3, containing cDNA sequences for human aldolase B have been isolated from a full length human liver cDNA library. The larger one, pA3, has been subcloned in M13 phage and completely sequenced with the chain terminator method. The sequence covers 1,600 nucleotides including the whole coding region (1,050 nucleotides), 67 nucleotides from the 5' non-coding region and the whole 3' non-coding region, 440 nucleotides long, down to the poly-A tail. Comparison with rabbit aldolase A and with a partial sequence of rat aldolase B, shows a homology of about 76% for aldolase A and of about 94% for aldolase B, which indicates that the sequenced cDNA codes for the liver isoenzyme. This is the first complete sequence reported for human aldolase B. The pA3 clone strongly hybridizes to 18S mRNA from human adult liver as expected from the size of the isolated cDNA.

INTRODUCTION

Aldolase is involved in the conversion of fructose-1,6-bisphosphate into glyceraldehyde-3-phosphate and dihydroxyacetone-1-phosphate along the glycolytic pathway. Three isoenzymes are generally expressed in healthy mammals: aldolase A in muscle, B in liver, and C in brain. They have a similar amino acid composition, but are electrophoretically and immunologically different (1). The complete sequence is known from rabbit aldolase A (2), and short fragments of the other isoenzymes have been sequenced from various sources (3-6). Aldolase B, the liver isoenzyme, has a high affinity for fructose-1-phosphate (F1P) and it is mainly involved in the metabolism of dietary fructose; its absence from the liver is associated with a congenital disease, hereditary fructose intolerance (HFI), which is characterized by nausea, vomiting, hypoglycemia following fructose ingestion, and which rapidly evolves to cirrhosis (7).

Aldolase isoenzyme expression in the hepatocyte varies during development; in fact foetal liver, rich in aldolase A and C, lacks aldolase B, which is produced only after birth (8). In chemically induced rat hepatocarcinoma, aldolase A and C are often increased, whereas aldolase B production is

© IRL Press Limited, Oxford, England.

switched off, resulting in a pattern very similar to the foetal one (9). Hence aldolase isoenzymes are a useful system with which to investigate eukaryotic gene regulation.

The isolation of aldolase cDNA clones will allow a detailed study of the mechanisms involved in aldolase isoenzyme expression. In particular it will be possible to isolate genomic clones, to analyze the number of genes coding for each isoenzyme, and to study the modality of expression and tissue specific transcription. Moreover, an aldolase B probe is important in the study of the molecular mechanisms of the aldolase B defect in HFI patients, as well as for the early diagnosis of the disease.

In a preceeding paper we described the cloning of a short cDNA fragment from human liver, homologous to rabbit muscle aldolase (10). Here we report the isolation of a full length cDNA from a human liver library, its complete sequence and its identification as aldolase B. This is the first complete sequence reported for human aldolase B.

MATERIALS AND METHODS

Materials

Restriction enzymes, DNA polymerase, DNA ligase, radionuclides were purchased from Amersham International, plc (U.K.), nitrocellulose filters were from Millipore S.A. (Molsheim, France), deoxy- and dideoxy-nucleotides and the M13 primes from PL-Biochemia, Inc. (Milwaukee, U.S.A.).

Screening of the cDNA library

Clone R38 (10) is an M13 recombinant clone containing a 250bp cDNA fragment from human liver, homologous to rabbit muscle aldolase. The single strand recombinant phage was radioactively labeled ($3x10^8$ cpm/μg), by synthesizing the second strand in the presence of radioactive dATP and an oligonucleotide primer, complementary to M13 in the region located at the 3' of the insert (11).

The library, obtained from Dr. F.E. Baralle (Oxford) (12), was a full length cDNA library (18), cloned in pAT153/Pvu II. The library was plated at a density of 1,000 colonies/plate. The colonies were transferred to nitrocellulose filters and hybridized at 65°C in 6xSSC (1xSSC is 0.15M sodium chloride, 0.015M sodium citrate, pH 7.0), 5x Denhardt's solution, 10 mM EDTA, 0.5% SDS; the filters were then washed in 0.5xSSC, 0.5% SDS, and autoradiographed with Fuji films, using Dupont Cronex intensifying screens.

Restriction site mapping, subcloning and sequencing

Double strand DNA was digested with restriction endonucleases according to the manufacturer's instructions. Sticky end fragments were cloned in

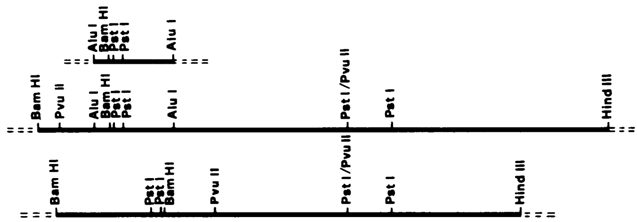

Fig. 1 - Restriction site maps of, from top to bottom, R38, pA3 and pA2 clones.
The continuous line is the insert containing sequences from human aldolase B cDNA; the dotted line is the vector sequence.

M13mp701 and mp8; blunt end fragments were inserted in phosphatase-treated M13mp8, in the Sma I site. The subclones were sequenced with the chain terminator method (13).

Hybridization to liver RNA and genomic DNA

Total liver mRNA, prepared by guanidine hydrochloride extraction and chromatography on oligo-dT cellulose, was separated on formaldehyde agarose gel and transferred to nitrocellulose filters (10). Hybridization was performed as described above.

Human genomic DNA, prepared from peripheral blood nucleated cells, was digested with several restriction endonucleases. The fragments were then separated on 1% agarose gel (10 g of DNA were used for each slot), transferred to nitrocellulose filters and hybridized to nick-translated plasmid DNA (specific activity $2x10^8$ cpm/ g) (14).

RESULTS

The screening of 5,000 colonies from the liver cDNA library led to the identification of two recombinant clones (pA2 and pA3). Restriction site maps,

Fig. 2 - Sequencing strategy of plasmid pA3.
The boxed portion represents the coding region of human aldolase B.
The segments shown are M13 subclones, sequenced by the chain terminator method; the arrows indicate the reading direction.

```
10 20 30 40 50 60 70 80 90 100 110 120
AGCTGCTGCCTCACCCACAGCTTTTGCATATCTAGGAGGACTCTTCTCTCCCAAACTACCTGTCACATGGCCCACCGATTTCCAGCCCTCACCCAGGAGCAGAAGAAGGAGCTCTCAGAA
MetAlaHisArgPheProAlaLeuThrGlnGluGlnLysLysGluLeuSerGlu

130 140 150 160 170 180 190 200 210 220 230 240
ATTGCCCAGAGCATTGTTGCCAATGGAAAAGGGGATCCTGGCTGCAGATGAATCTGTAGGTACCATGGGGAACCGCCTGCAGAGGATCAAGGTGGAAAACACTGAAGAGAACCGCCGGCAG
IleAlaGlnSerIleValAlaAsnGlyLysGlyIleLeuAlaAlaAspGluSerValGlyThrMetGlyAsnArgLeuGlnArgIleLysValGluAsnThrGluGluAsnArgArgGln

250 260 270 280 290 300 310 320 330 340 350 360
TTCCGAGAAATCCTCTTCTCTGTGGACAGTTCCATCAACCAGAGCATCGGGGGTGTGATCCTTTTCCACGAGACCCTCTACCAGAAGGACAGCCAGGGAAAGCTGTTCAGAAACATCCTC
PheArgGluIleLeuPheSerValAspSerSerIleAsnGlnSerIleGlyGlyValIleLeuPheHisGluThrLeuTyrGlnLysAspSerGlnGlyLysLeuPheArgAsnIleLeu

370 380 390 400 410 420 430 440 450 460 470 480
AAGGAAAAAGGGGATCGTGGTGGGAATCAAGTTAGACCAAGGAGGTGCTCCTCTTGCAGGAACAAACAAAGAAACCACCATTCAAGGGCTTGATGGCCTCTCAGAGCGCTGTGCTCAGTAC
LysGluLysGlyIleValValGlyIleLysLeuAspGlnGlyGlyAlaProLeuAlaGlyThrAsnLysGluThrThrIleGlnGlyLeuAspGlyLeuSerGluArgCysAlaGlnTyr

490 500 510 520 530 540 550 560 570 580 590 600
AAGAAAGATGGTGTTGACTTTGGGAAGTGGCGTGCTGTGCTGAGGATTGCCGACCAGTGTCCATCCAGCCTCGCTATCCAGGAAAACGCCAACGCCCTGGCTCGTTACGCCAGCATCTGT
LysLysAspGlyValAspPheGlyLysTrpArgAlaValLeuArgIleAlaAspGlnCysProSerSerLeuAlaIleGlnGluAsnAlaAsnAlaLeuAlaArgTyrAlaSerIleCys

610 620 630 640 650 660 670 680 690 700 710 720
CAGCAGAATGGACTGGTACCTATTGTTGAACCAGAGGTAATTCCTGATGGAGACCATGACCTGGAACACTGCCAGTATGTTACTGAGAAGGTCCTGGCTGCTGTCTACAAGGCCCTAAAT
GlnGlnAsnGlyLeuValProIleValGluProGluValIleProAspGlyAspHisAspLeuGluHisCysGlnTyrValThrGluLysValLeuAlaAlaValTyrLysAlaLeuAsn

730 740 750 760 770 780 790 800 810 820 830 840
GACCATCATGTTTACCTGGAGGGCACCCTGCTAAAGCCCAACATGGTGACTGCTGGACATGCCTGCACCAAGAAGTATACTCCAGAACAAGTAGCTATGGCCACCGTAACAGCTCTCCAC
AspHisHisValTyrLeuGluGlyThrLeuLeuLysProAsnMetValThrAlaGlyHisAlaCysThrLysLysTyrThrProGluGlnValAlaMetAlaThrValThrAlaLeuHis

850 860 870 880 890 900 910 920 930 940 950 960
CGTACTGTTCCTGCAGCTGTTCCTGGCATCTGCTTTTTGTCTGGTGGCATGAGTGAAGAGGATGCCACTCTCAACCTCAATGCTATCAACCTTTGCCCTCTACCAAAGCCCTGGAAACTA
ArgThrValProAlaAlaValProGlyIleCysPheLeuSerGlyGlyMetSerGluGluAspAlaThrLeuAsnLeuAsnAlaIleAsnLeuCysProLeuProLysProTrpLysLeu

970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080
AGTTTCTCTTATGGACGGGCCCTGCAGGCCAGTGCACTGGCTGCCTGGGGTGGCAAGGCTGCAAACAAGGAGGCAACCCAGGAGGCTTTTATGAAGCGGGCCATGGCTAACTGCCAGGCG
SerPheSerTyrGlyArgAlaLeuGlnAlaSerAlaLeuAlaAlaTrpGlyGlyLysAlaAlaAsnLysGluAlaThrGlnGluAlaPheMetLysArgAlaMetAlaAsnCysGlnAla

1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
GCCAAAGGACAGTATGTTCACACGGGTTCTTCTGGGGCTGCTTCCACCCAGTCGCTCTTCACAGCCTGCTATACCTACTAGGGTCCAATGCCCGCCAGCCTAGCTCCAGTGCTTCTAGTA
AlaLysGlyGlnTyrValHisThrGlySerSerGlyAlaAlaSerThrGlnSerLeuPheThrAlaCysTyrThrTyr *

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320
GGAGGGCTGAAAGGGAGCAACTTTTCCTCCAATCCTGGAAATTCGACACAATTAGATTTGAACTGCTGGAAATACAACACATGTTAAATCTTAAGTACAAGGGGGAAAAAATAAATCAGT

1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440
TATTGAAACATAAAAATGAATACCAAGGACCTGATCAAATTTCACACAGCGAGTTTCCTTGCAACACTTTCAGCTCCCCATGCTCCAGAATACCCACCCAAGAAAATAACCGGGATCTAA

1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560
AACAATAATCGGCTCCTCATCCAAAGAACAACTGCTGATTGAAACACCTCATTAGCTGATGTAGAGAAGTGCATCTTATGAAACAGTCTTAGCAGTGGTAGGTTGGGAAGGAGATAGCTG

1570 1580 1590 1600
CAACCAAAAAAGAAATAAATATTCTATAAACCTTCAAAAAAAAAAAAAA
```

obtained from the clones, after digestion with Pst I, Hind III, Bam H1 and Pvu II, are illustrated in Fig. 1. pA3 has the larger insert (1,700 bp), pA2 is only slightly shorter (1,500). Comparing the two maps with the R38 clone it seems that the 5' region in pA2 is in the inverted orientation with respect to the rest of the clone; this was subsequently confirmed by sequence analysis (data not shown).

The sequencing strategy is reported in Fig. 2. A number of M13 subclones were constructed using the Pst I, Bam H1, Hind III sites, as well as the numerous Hinf I, Sau3A I, Hae III, Alu I sites; the subclones were then sequenced with the chain terminator method (13). The complete cDNA nucleotide sequence, reported in Fig. 3, was obtained from the pA3 clone; pA2 was identical, at least in the sequenced fragments. The sequence covers the whole coding sequence of 367 amino acids, starting at a methionine residue in position 67 and ending with a TAG nonsense codon in position 1,159; it is flanked by a long 3' non coding sequence. The amino acid sequence, derived from the translation of the nucleotide sequence, shows an homology of 67% with that of rabbit aldolase A (2). A much higher homology (94%) was found for rat aldolase B (Fig. 4), although it was limited to the available sequence (3). pA3 contains the complete 3' non-coding region (440 nucleotides from the termination codon to the poly-A addition site); the canonical (AATAAA) poly-A addition signal is present 17 bp upstream the poly A. 60 nucleotides have been sequenced from the 5' untranslated region, whose total length is not precisely known, but the size of the mRNA in Northern blotting suggests an estimate of 60-100 nucleotides.

Clone R38 was used also in hybridization experiments, to evaluate the expression of aldolase mRNA in normal adult human liver. Clone R38 hybridizes to a single very abundant 18S RNA species in human liver RNA. When clone pA3 is used as a probe, it hybridizes to an RNA of the same length (Fig. 5).

Clone pA3 was hybridized to human genomic DNA from peripheral blood of healthy individuals (Fig. 6). The digestion with all the enzymes used results in more than one hybridization band of different intensities, at least when non stringent conditions were used (0.5xSSC, 65°C).

Fig. 3 - Complete nucleotide sequence of human aldolase B cDNA.
The coding sequence has been translated to show the primary structure of the protein molecule. Boxed amino acids are the same as in rabbit aldolase A. The star indicates the stop codon and the underlined nucleotides show two canonical sequences for poly-A addition sites.

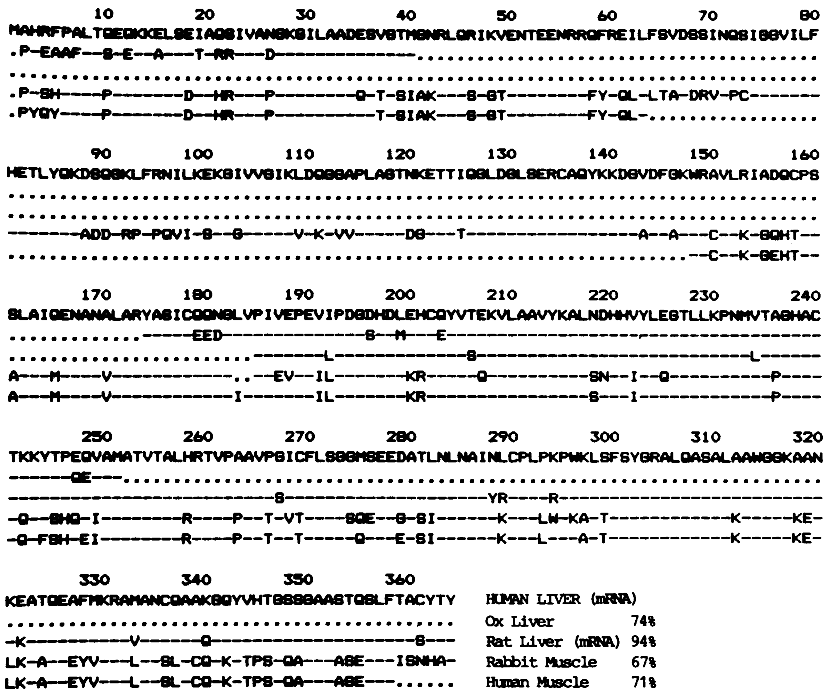

Fig. 4 - Sequence homology between human liver aldolase and four mammalian aldolase amino acid sequences, two from liver and two from muscle. Rabbit muscle aldolase is the only complete sequence known so far, the others are partial sequences obtained from protein or mRNA. Hyphens indicate amino acids identical to human aldolase B, dots correspond to non sequenced regions.

DISCUSSION

Clone R38 was first isolated from a human cDNA library, and was identified by sequence analysis for its homology to amino acids 16-90 of rabbit muscle aldolase. The source of the mRNA (human liver) and the degree of homology (70%) suggested it was not an aldolase A, but, probably, an aldolase B cDNA. This could not be demonstrated, since the complete sequence of aldolase B is not known, either in man or in other mammals. Moreover, the liver RNA used in the construction of the library was from a pool of biopsies containing a significant amount of foetal liver, and foetal liver expresses aldolase isoenzymes A and C.

The first step in the characterization of clone R38 was the study of the expression of the relative mRNA in normal adult human liver. The high level of expression shown by Northern blotting (Fig. 5) is in good agreement with data

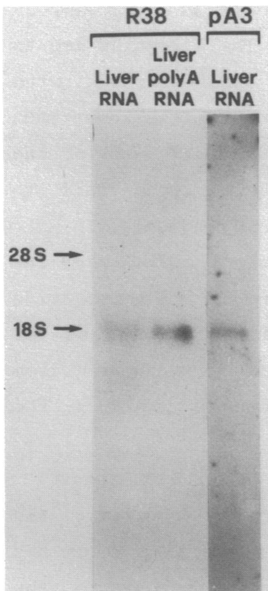

Fig. 5 - Hybridization of R38 and pA3 clones to human liver total RNA and poly-A$^+$RNA. The arrows show the position of 28S and 18S RNA.

from in vitro translation of aldolase B mRNA (15), that indicate its high concentration (>1% of total mRNA) in adult liver.

To establish whether the isolated clone was an aldolase B cDNA, clone R38 was used as a probe to screen a full-length cDNA library from human liver.

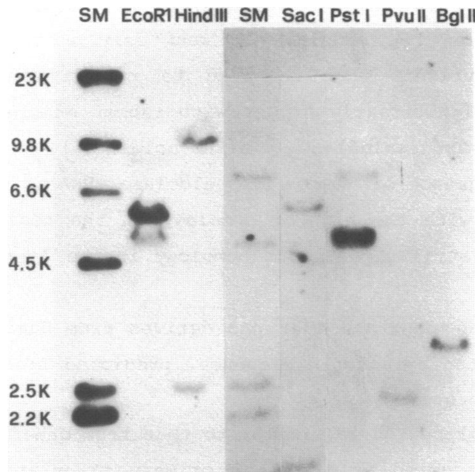

Fig. 6 - Hybridization of nick-translated pA3 to human peripheral blood genomic DNA digested with several restriction enzymes. The size marker (SM) is Hind III digested lambda phage DNA.

Sequencing of larger clones would lead to the identification of the cDNA, as well as to the complete protein sequence, still unknown. The screening led to the isolation of two positive clones, pA2 and pA3, out of 5,000 colonies, with a frequency of 0.04%. This is slightly less than expected from the known mRNA abundance. However, it should be noted that R38 maps near the 5' end of the mRNA and therefore many clones, not quite full-length, could fail to hybridize, so leading to an underestimation of the relative frequency.

The complete cDNA sequence was obtained from the larger clone, pA3. The cDNA is 1,600 nucleotides long and it codes for a protein of 367 amino acids that shows a high degree of homology for rabbit aldolase A as well as for several aldolase A and B fragments (both protein and cDNA sequences) from various sources. A comparison of the different sequences, leads to the following considerations:

a) The homology with rabbit aldolase A (67%) is the same as for R38, and it is uniformly spread over the whole molecule. On the contrary, the comparison with the available sequence of rat aldolase B (3), which has been obtained by translating a partial cDNA, corresponding to the carboxyterminal half of the protein (amino acids 184-364), shows only 11 different amino acids, i.e. 94% homology. Furthermore, a recently described (16) human aldolase B cDNA fragment (amino acids 238-364) is also identical to pA3, with a single amino acid difference in position 7.

b) In the region around the active site (amino acids 215-242) the homology increases to 100% when the protein sequence is compared to fragments from rabbit (4) and ox (5) aldolase B, and only a single valine-leucine substitution is present when compared to rat liver aldolase (3). The differences are significantly higher with rabbit muscle aldolase (6 amino acids) and human muscle aldolase (4 amino acids) (6).

c) The nucleotide sequence of human liver aldolase mRNA is very similar to rat aldolase B mRNA, with the highest homology in the coding sequence (89%), and a lower, but still significant homology in the 3' non coding sequence (3).

All these data demonstrate that pA3 derives from human aldolase B mRNA, and that the translated amino acid sequence, predicted from the cDNA, belongs to the liver enzyme protein.

The aminoterminal region is similar to that from other aldolase proteins. The exact N-terminal amino acid of the primary translation product is not clearly defined, since in the 5'-non coding sequence there is no stop codon in frame. It should be pointed out that there is a single methionine codon in that region, located just before the first homologous amino acid. This

methionine is a good candidate for the N-terminal amino acid, since the data from *in vitro* translation (15) indicate a primary translation product of the same size as the mature protein. This is also in agreement with data from Lebherz et al. (17), who reported, in several aldolase molecules, a blocked N-terminus consisting of a single derivatized methionine. If this methionine is the N-terminus of the protein, then pA3 contains 67 nucleotides of the 5' non coding sequence; its total length is not known, but it should be between 70-100 nucleotides since the messenger is an 18S RNA.

The carboxy terminal end (a tyrosine residue) of human aldolase B, is the same as that of rabbit aldolase A and rat aldolase B. This residue has been demonstrated to be essential for the activity in some aldolase enzymes. Aldolase B mRNA contains a rather long (440 nucleotides) 3' non coding region. A canonical poly-A addition signal (AAUAAA), is present 17 nucleotides before the poly-A, but an additional signal sequence is located in the middle of the 3' non coding sequence, in position 1,310. This additional sequence seems to be skipped during aldolase B mRNA processing in the liver. However, this site may be less frequently used in the liver or it may be specific for other tissues. Further blotting and S1 mapping experiments are currently under way to analyze these possibilities.

Clone pA3 has also been hybridized to total human genomic DNA to elucidate the structure and organization of aldolase genes. The digestion with all the restriction enzymes used, reveals multiple bands: Eco R1, Hind III, Pst I, Pvu II, Sac I all show two strong hybridization bands, Bgl II shows a single strong band; all of them show some additional weaker bands under non stringent conditions. These preliminary results could be explained either by the presence of more than one aldolase B gene per human haploid genoma or of pseudo genes, or by cross-hybridization to aldolase A and C genes.

ACKNOWLEDGEMENTS

We wish to thank Dr. F.E. Baralle for the helpful discussions and for providing the cDNA library.

This work was supported by Consiglio Nazionale delle Ricerche and Ministero Public Education grants, Rome. Special acknowledgment is given to Progetto Finalizzato Ingegneria Genetica and Basi Molecolari delle Malattie Ereditarie (CNR, Rome).

*To whom correspondence should be addressed

REFERENCES

1. Horecker, B.L., Tsolas, O. and Lai, C.Y. (1972), in The Enzymes, Boyer, P.D., Ed., pp. 213-258, Academic Press, New York.
2. Nakai, N., Chang, D. and Lai, C.Y. (1975), Arch.Biochem.Biophys. 166, 347-357.
3. Tsutsumi, K., Mukai, T., Ikada, S., Miyahara, H., Tsutsumi, R., Tanaka, T., Hori, K. and Ishikawa, K. (1983), J.Biol.Chem. 248, 6537-6542.
4. Forcina, B.G. and Perham, R.N. (1971), FEBS Lett. 18, 59-63.
5. Benfield, P.A., Forcina, B.G., Gibbons, I. and Perham, R.N. (1979), Biochem.J. 183, 429-444.
6. Freemont, P.S., Dunbar, B. and Fothergill, L.A. (1984), Arch.Biochem. Biophys. 228, 342-352.
7. Schapira, F., Gregori, C. and Hatzfeld, A. (1977), Clin.Chim.Acta 78,1-7.
8. Hommes, F.A. and Draisma, M.I. (1970), Biochim.Biophys.Acta 222, 251.
9. Matsushima, T., Kawabe, S., Shibuya, M. and Sugimura, T. (1968), Biochem. Biophys.Res.Commun. 30, 565.
10. Costanzo, F., Castagnoli, L., Dente, P., Arcari, P., Smith, M., Costanzo, P., Raugei, G., Izzo, P., Pietropaolo, C., Bougueleret, L., Cimino, F., Salvatore, F. and Cortese, R. (1983), EMBO J. 2, 57-61.
11. Hu, N. and Messing, J. (1982), Gene 17, 271-277.
12. Sharpe, C.R., Sidoli, A., Shelley, C.S., Lucero, M.A., Shoulders, C.C. and Baralle, F.E. (1984), Nuc.Acids Res. 12, 3917-3932.
13. Sanger, F., Nicklen, S. and Coulson, A.R. (1977), Proc.Natl.Acad.Sci., USA 74, 5463.
14. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977), J.Mol.Biol. 113, 237-251.
15. Gregori, C., Besmond, C., Kahn, A. and Dreyfus, J.C. (1982), Biochem.Biophys.Res.Commun. 104, 369-375.
16. Besmond, C., Dreyfus, J.C., Gregori, C., Frain, M., Zakin, M., Sala Trepat, J. and Kahn, A. (1983), Biochem.Biophys.Res.Commun. 117, 601-609.
17. Lebherz, H.G., Bates, O.J. and Bradshaw, R.A. (1984), J.Biol.Chem. 259, 1132-1135.
18. Reid, K.B.M., Bentley, D.R. and Wood, K.Y. (1984), Phyl.Trans.R.Soc.Lond. B, in press.