Volume 15 Number 12 1987 Nucleic Acids Research

Human mRNA polyadenylate binding protein: evolutionary conservation of a nucleic acid binding motif

Thierry Grange, Cezar Martins de Sa[1], Jacqueline Oddos and Raymond Pictet

Institut Jacques Monod du CNRS, Unité INSERM 257 Université Paris VII, Tour 43−2 Place Jussieu, 75251 Paris Cedex 05 and [1]Institut Jacques Monod, Laboratoire de Biochimie de la Différenciation, France

Received March 30, 1987; Revised and Accepted May 22, 1987 Accession no. Y00345

ABSTRACT

We have isolated a full length cDNA (cDNA) coding for the human poly(A) binding protein. The cDNA derived 73 kd basic translation product has the same Mr, isoelectric point and peptidic map as the poly(A) binding protein. DNA sequence analysis reveals a 70,244 dalton protein. The N terminal part, highly homologous to the yeast poly(A) binding protein, is sufficient for poly(A) binding activity. This domain consists of a four-fold repeated unit of approximately 80 amino acids present in other nucleic acid binding proteins. In the C terminal part there is, as in the yeast protein, a sequence of approximately 150 amino acids, rich in proline, alanine and glutamine which together account for 48 % of the residues. A 2,9 kb mRNA corresponding to this cDNA has been detected in several vertebrate cell types and in Drosophila melanogaster at every developmental stage including oogenesis.

INTRODUCTION

The long track of riboadenylic acid, poly(A), added post-transcriptionaly to the end of most, but remarquably not all, eucaryotic mRNAs is a major modification, up to 25 % of the molecular weight, whose function, however, is not completely understood (for a review see 1). Stabilization of mRNA or facilitation of translation are the best documented hypotheses for its mode of action (2). From yeast to higher eucaryotes, the poly(A) tail has consistently been found associated with tight binding proteins in both the cytoplasm and the nucleus. However, the variety of the methods and stringencies of binding conditions used to isolate and define the associated polypeptides has resulted in an accumulation of conflicting results (3).

The most generally accepted data indicates the existence of a 70 to 78,000 molecular weight protein (3,4). Some experiments suggest that this protein plays a role in the translation process. It is associated with polyribosomal mRNP but not with free, translationaly repressed mRNP (5,6). In reticulocyte lysates, the addition of poly(A) inhibits more strongly the translation of naked mRNA than that of polyribosomal mRNP (7). This inhibition can be overcomed by further addition

of purified poly(A) binding protein, suggesting that the poly(A) acts by sequestering the poly(A) binding protein (M.F. Grossi de Sa et al., manuscript in preparation). This protein could also play a role in discriminating between different mRNAs (8) in agreement with the demonstration that poly(A) inhibits the in vitro translation of some viral poly(A)+ mRNAs less than it inhibits the cellular mRNAs (9). These questions should become clearer with more knowledge of the structure of this protein and the regulation of its synthesis.

We report here the isolation of a full length cDNA coding for the human poly(A) binding protein and the determination of its complete nucleotide sequence. The protein shows a high degree of homology with the recently published yeast poly(A) binding protein sequence (10,11). The conserved region covers a domain which includes the poly(A) binding function. This cDNA probe allowed us to analyse the pattern of accumulation of the corresponding transcript during development of both Drosophila melanogaster and Xenopus laevis.

MATERIAL AND METHODS

- Construction of the cDNA librairies

The rat hepatoma library was constructed from poly(A)+ mRNA (a gift from G. Beck, IBMC, Strasbourg) purified from HTC cells (12) treated for 12 hours with 10^{-6} M dexamethasone. cDNA was synthesized using current procedures (13) and cloned into the Pst I site of pBR322 using dG.dC tailing (13). Differential screening was performed with cDNA probes synthesized from mRNA purified from HTC cells stimulated or not with dexamethasone. A family of cDNA clones corresponding to a unique mRNA whose concentration was increased in the dexamethasone treated cells was selected and further characterized by in vitro translation of the hybrid-selected mRNA (13). Since the dexamethasone inducibility of the related mRNA could not be demonstrated in other experiments, other unknown parameters must be the cause of the observed differential signal.

The human melanoma library was constructed from poly(A)+ mRNA (a gift from J. Axelrod, Weizman Institut, Rehovot) purified from Bowes cells (14) using a modification (15) of the procedure of Cooke et al. (16). First strand cDNA was synthesized as described (13). After alkali treatment, it was chromatographed on Bio-Gel P-60 (Biorad). An homopolymeric d(A) tail, estimated to be 40 residues long, was then added with terminal deoxynucleotidyl transferase (PL. Biochemicals) (16). Oligo dT primed second strand synthesis was performed with AMV reverse transcriptase (Anglian Biotech) under conditions identical to the first strand synthesis.

After incubation for one hour at 42°C, the reaction volume was diluted with 2

volumes of 200 mM Hepes pH 6.9 ; 10 mM MgCl2 ; 140 mM KCl and the Klenow fragment of E. coli DNA polymerase I (BRL) was added (50 U/µg of first strand cDNA). The reaction was allowed to proceed for 16 hours at 15°C. The double stranded cDNA thus obtained was then treated with EcoRI methylase (Biolabs) and after addition of EcoRI linkers was cloned into the EcoRI site of λ gt10 as described (17). Part of this cDNA library was screened with a 1.6 kb rat cDNA isolated as described above.

- Blot hybridization

In situ hybridization and probe synthesis were performed using current procedures (13). RNAs were fractionated by electrophoresis on 1.2 % agarose formaldehyde gels (13) and electroblotted onto Zeta probe membranes (Biorad) as recommended by the manufacturer. Hybridization was carried out in 50 % formamide /5xSSC (SSC is 0.15 M NaCl, 15 mM tri-sodium citrate, pH 7.0) at 42°C (13). Southern blots of genomic DNA (18) were hybridized in 50 % formamide/5xSSC at 42°C (13).

- In vitro transcription and translation

The 2.9 kb cDNA was recovered from a λgt10 recombinant phage as an EcoRI fragment by partial EcoRI digestion and subcloned into the EcoRI site of $pGEM_1$ (Promega Biotec). After linearization by BamH I (which cuts 3' to the inserted cDNA), the corresponding RNA was synthesized in vitro from the T7 promoter with T7 RNA polymerase (Promega Biotec) following the recommendations of the manufacturer. After DNAse (Promega Biotec) treatment, the synthesized RNA was translated in a rabbit reticulocytes lysate (19) at a concentration of 10 µg/ml in the presence of 50 MBq/ml ^{35}S methionine (30 TBq/mmole). The shortened translation product was obtained after transcription of the same plasmid digested by Tth 111 I which cuts the cDNA at nucleotide 1792.

- Peptide mapping analysis

Duck 15 S globin mRNP used as the source of poly(A) binding protein were obtained from EDTA-dissociated duck reticulocyte polysomes by isokinetic sucrose gradient centrifugation as described by Vincent et al. (6). 15 µl of the translation mixture (corresponding to 2 x 10^5 TCA precipitable cpm) was mixed with 30 µl of a preparation of duck erythroblast 15 S mRNP (corresponding to 30 µg of protein). The mixture was separated on bidimensional gels as described (20) and, after Coomassie Blue staining, the spot corresponding to the previously described poly(A) binding protein (6) was cut out of the gel. When the same amount of lysate is loaded alone onto the bidimensional gel, this spot is hardly detectable by Coomassie Blue staining and can be estimated to constitute less than 5 % of the total amount of the purified duck protein. The protein in the

gel slice was submitted to partial proteolytic cleavage in the stacking gel of a 16 % acrylamide-SDS gel as described (21). We used either 1 µg of chymotrypsine or 50 ng of Staphylococcus aureus V8 protease.

- Poly(A) binding assay

The translation product of the full length and the shortened Hu73 RNA (2 x 10^5 TCA precipitable cpm, 15 to 20 µl of lysate) was diluted in 500 µl of a buffer containing 10 mM triethanol-amine/HCl pH 7.4, 10 mM EDTA, 5 mM 2-mercaptoethanol and assayed as follows : the sample was recycled 4 times through a pre-equilibrated poly(A) sepharose 4B column (100 mg resin, Pharmacia). The unbound fraction was saved and the column washed with 4 ml of buffer. Bound material was then eluted with the same buffer by increasing the concentration of KCl in a stepwise fashion. At each step 1 ml of eluate was saved and the column further washed with 2 ml. The material still retained at 1 M KCl was eluted with 1 ml of 50 mM Tris-HCl (pH 7.6), 2.0 M LiCl, 1.0 M urea (22). Eluates were TCA precipitated (10 % v/v final concentration), washed with acetone and half of the material analysed by SDS-PAGE and autoradiographed.

- Nucleotide sequencing

The full length cDNA was subcloned after partial EcoRI digestion in the EcoRI site of pTZ19 (Pharmacia) in both orientations. Unidirectional deletions were introduced after cleavage by BamHI and SphI by exonuclease III followed by nuclease S1 treatment (23). The resulting clones were sequenced by the dideoxy method (24). Whenever necessary (unresolved uncertainties or sequences obtained from only one strand) some regions were sequenced by the chemical method (25).

Computer work was performed using the CITI2 facilities.

RESULTS

Isolation of a family of cDNA clones corresponding to a basic protein of 73 kd

In the course of screening cDNA clones in libraries constructed from rat hepatoma (HTC ; 12) and human melanoma (Bowes ; 14) cells mRNA, we fortuitously isolated cDNAs corresponding to a protein whose properties were coincident with those of the poly(A) binding protein (see Material and Methods).

The translation product of the mRNA selected by hybridization with one of these cDNAs migrates on bidimensional gels (20) as a polypeptide with a molecular weight of 73 kd and with a isoelectric point above 8 (data not shown). In duck erythroblasts, a protein of the same molecular weight (Mr) and isoelectric point, has been characterized and shown to bind strongly to the 3' poly(A) tail of polyribosomal mRNA (6).

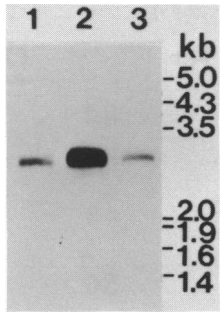


Figure 1 - Size and homology between human, rat and duck mRNAs coding for the 73 kd protein
Poly(A)+ RNA (2.5 µg) from rat HTC cells (lane 1), Bowes melanoma (lane 2) and duck erythroblast (lane 3), were fractionated on gels, electroblotted onto Zeta probe membrane and hybridized with the 2.9 kb human cDNA (Hu 73), labelled by nick-translation. After hybridization, the blot was washed with a final wash of 30 minutes in 0.1 x SSC at 65°C.

Northern blot analysis shows that these cDNAs correspond to a 2.9 kb mRNA present not only in rat hepatoma and human melanoma cell lines but also in duck erythroblasts (Figure 1). The stringency of the hybridization conditions allows us to predict that these cDNAs correspond to a highly conserved protein. The poly(A) binding protein has been described in many cell types (3) and is highly conserved among vertebrates (26).

The association of these characteristics, molecular weight, isoelectric point and evolutionary conservation, motivated us to verify that the protein coded by these cDNAs was indeed the poly(A) binding protein. To date, no antibody against the vertebrate poly(A) binding protein is available. Thus in order to clearly identify the nature of this cDNA, we used a full or nearly full length human (2.9 kilobases) cDNA, thereafter called Hu73, to synthesize, from the in vitro transcription product, an amount of protein sufficient for a direct identification.

The 73 kd in vitro translation product corresponding to the human Hu 73 cDNA has the same MW, pI and peptide map as the duck poly(A) binding protein and possesses affinity for poly(A)

This cDNA was subcloned in PGEM1 (Promega Biotec), a plasmid vector allowing in vitro transcription of the inserted DNA. RNAs corresponding to each strand of the cDNA were synthesized in vitro and translated in a rabbit reticulocytes lysate. As expected, only one of these RNAs gave rise to detectable products. The largest and major polypeptide has a Mr of 73 kd (Figure 2).

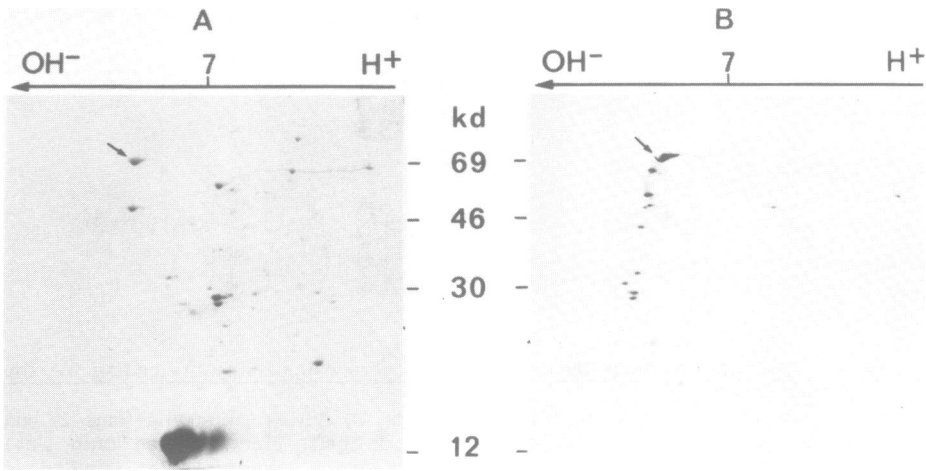


Figure 2 - Duck poly(A) binding protein and the 73 kd translation product have the same mobility on bidimensional gel 2 µl of the ^{35}S methionine labelled translation product (3 x 10^4 TCA precipitated cpm) of the sense strand RNA synthesized from the Hu 73 cDNA was mixed with 4 µl of duck 15 S mRNP (4 µg of protein) and separated on bidimensional gels (20). The amount of poly(A) binding protein contributed by the lysate is negligible (see Material and Methods). The arrow shows the position of the intact 73 kd poly(A) binding protein (A) and the 73 kd translation product (B). A. Coomassie Blue staining of the gel.
B. Autoradiography of the same gel.

The large number of polypeptides resulting from the in vitro translation is a particularity of the RNA derived from the Hu73 cDNA. The parallel translation of the tyrosine aminotransferase RNA (8) synthesized under the same conditions gives a unique polypeptide corresponding to the full length protein. A kinetic analysis of the translation of the Hu 73 RNA shows that the smallest products appear before the largest ones suggesting that they are not due to proteolytic degradation of the 73 kd polypeptide (not shown). These smaller polypeptides may result from initiation at internal methionines. Multiple initiations have already been shown to occur with pSP6 synthesized RNA (37). The sequence shown in Figure 5B reveals a polypeptide with a molecular weight smaller than estimated on gels and with a large number of methionines. This precludes the establishment of significant correlations between peptide molecular weight and methionine location. The discrepancy between the number of peptides and of methionines as well as the difference in signal intensity may result from differences in the quality of the various methionines to serve as initiator (38) or of secondary structures of the RNA.

The 73 kd translation product from Hu 73 cDNA was compared with the poly(A) binding protein of duck erythroblasts. A mixture of the translation products and of the duck 15 S mRNP proteins (6) was separated on bidimensional gels (20). The position of the 73 kd translation product (detected by autoradiography) is coincident with the position of the duck poly(A) binding protein (detected by Coomassie Blue staining) (Figure 2). In order to compare the peptidic

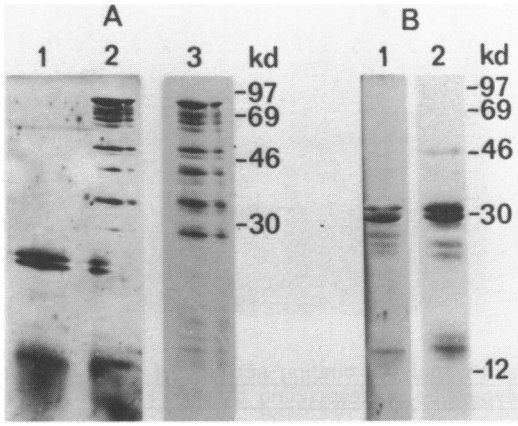


Figure 3 - Comparative peptide mapping analysis of the duck poly(A) binding protein with the 73 kd translation product
A mixture of the two proteins purified on bidimensional gels was subjected to partial proteolytic cleavage as described in Material and Methods. Silver staining (39) allowed the revelation of the proteolytic products of the duck poly(A) binding protein and autoradiography, the revelation of the products of the in vitro synthesized 73 kd polypeptide.
Panel A - Peptide mapping with chymotrypsin
- Lane 1 : chymotrypsin alone - silver staining
- Lane 2 : mixture of the 73 kd translation product and the duck poly(A) binding protein cleaved by chymotrypsin - silver staining
- Lane 3 : autoradiography of lane 2,
Panel B - Peptide mapping with staphylococcus aureus V8 protease
- Lane 1 : same mixture as in lane A2 but cleaved by V8 protease - silver staining (The amount of V8 protease used does not give detectable bands)
- Lane 2 autoradiography of lane 1.

map of these two proteins, the corresponding spot was cut out of the gel and the mixture was submitted to partial proteolytic cleavage (21). Figure 3 shows that the cleavage patterns given by chymotrypsin (3A) and V8 protease (3B), respectively, are nearly identical for the unlabelled and the labelled protein. Subtle differences can be attributed to the different origins of the two polypeptides (duck and human) and to the different modes of detection (autoradiography and silver staining).

Poly(A) binding protein purified from duck erythroblast cytosol on poly(A) sepharose column (22) gave similar results (not shown). The identification of the cloned Hu 73 cDNA is further supported by the affinity of the 73 kd translation product for the poly(A). The translation products were passed through a poly(A) sepharose column, and the retained material was eluted with buffer of increasing ionic strength. Gel electrophoresis analysis of the eluted fractions

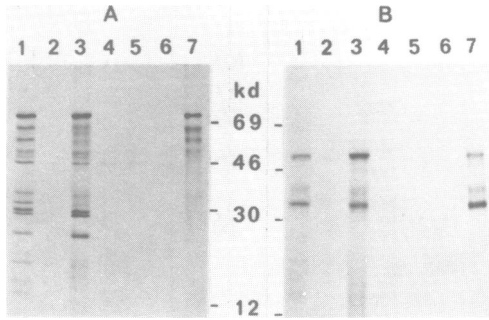


Figure 4 - Assay of the poly(A) binding activity of the 73 kd polypeptide and of a truncated 50 kd polypeptide corresponding to the N terminal

Material equivalent to 2 x 10^5 cpm of the translation products of each, the full size and the shortened poly(A) binding protein mRNAs synthesized, respectively, from the full size and the shortened cDNAs (for the location of the cleavage site, see Figure 5B), were loaded onto, and eluted from, a poly(A) sepharose column. The eluted fractions were analyzed by gel electrophoresis and autoradiography as described in Material and Methods. The large number of polypeptides resulting from the in vitro translation may result from initiation at internal methionines (see legend to Figure 2).
A) Products of the full-length Hu 73 RNA and B) products of the 3'-cleaved Hu 73 RNA. In both A and B, the numbers correspond to the material washed at different steps as follows : 1 : Crude translation product ; 2 : unbound material ; 3-4-5- and 6 : material washed with 250 mM, 500 mM, 750 mM and 1 M KCl, respectively ; 7 : 2 M LiCl-1 M urea eluted material.

(Figure 4A) demonstrates that the 73 kd polypeptide has the same affinity for poly(A) as the cytosol poly(A) binding protein (22). The smaller products with a size ranging from 50 to 70 kd also possess a comparable affinity for poly(A).

Nucleotide sequence of the Hu 73 cDNA

Both strands of the 2.9 kb Hu 73 cDNA were sequenced (Figure 5). The sequence length, 2848 nucleotides, poly(A) tail excluded, is in good agreement with the estimated size of the mRNA (Figure 1). The analysis of this sequence reveals a long open reading frame coding for a protein of 633 amino acids with a calculated molecular weight of 70244 d. The molecular weight of the protein is thus slightly lower than estimated by SDS polyacrylamide gel electrophoresis, a result probably due to its high proline content (see below).

The cDNA region corresponding to the 5' and 3' untranslated region of the human mRNA are at least 502 nucleotides and 447 nucleotides long, respectively. The 3' untranslated region has two polyadenylation sequences. The first one close to the termination codon is, in fact, a doublet but if used it does not give rise to an important proportion of the mRNA in the tested cells (see Figure 1).

The examination of the primary structure of the protein reveals several

noteworthy features, particularly the presence of two distinct regions of unequal length. The comparison of the sequence with itself using the Diagon program of Staden (27) shows the existence, in the N terminal two-thirds of the molecule, of a four-fold repeated unit of approximately 80 amino acids (Figure 6A). The alignment of these units requires the introduction of only two small gaps of one or two amino acids in the units three and four, respectively. The region of the greatest homology is found, in every unit, between the fourtieth and sixty-third amino acids where, accepting three equivalences (two Phe = Tyr and one Arg = Lys), eleven out of twenty-four amino acids are identical. The four units have roughly the same length on their N terminal side, but on the C terminal side the last three cover a sequence about fifteen amino acids longer than the first one.

There are differences in amino acid composition between the N terminal part defined by the domain containing the four units and the C terminal part. The N terminal part contains a higher percentage of both acidic and basic amino acid (14 % and 16 % respectively) than the C terminal (4 % et 8 % respectively). The C terminal third is characterized by a 168 amino acid proline-rich domain (amino acids 390 to 557) which directly follows the last of the four repeats. In addition to proline, which represents 18 % of the amino acids, this region is rich in alanine (17 %) and glutamine (13 %) and markedly depleted in acidic amino acids. With the exception of a stretch of seven consecutive alanines, there is apparently no regular pattern of amino acid distribution.

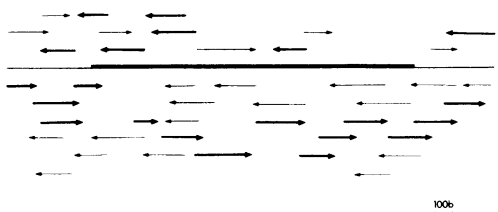
The domain corresponding to the four-fold repeated unit has the same affinity for poly(A) as the intact protein

The yeast poly(A) binding protein gene has recently been cloned (10,11). Figure 6B shows the protein sequence comparison diagram between the yeast and human protein. It clearly underlines the high degree of homology which exist at the level of the N terminal region and the conservation of the number, size and arrangement of the repeated units. The N terminal region of the yeast poly(A) binding protein is believed to carry the poly(A) binding domain (11). We thus directly tested the affinity for poly (A) of a shortened, *in vitro* synthesized product of 430 amino acids corresponding to the four repeated units (see arrow in Figure 5B). This product is retained on the poly(A) sepharose column and eluted under the same conditions as the complete 73 kd polypeptide (Figure 4B). This is also true for a shorter 33 kd product of the *in vitro* synthesis, which according to its size, can accomodate at most three complete repeated units (Figure 4B).

Detection of poly(A) binding protein gene transcription in Xenopus laevis and Drosophila melanogaster

The conservation between yeast and human of the region containing the

A



B

```
...CCTTCTCCCCGGCGGTTAGTGCTGAGAGTGCGGAGTGTGTGCTCCGGGCTCGGAACACACATTTATTATTAAAAAATCCAAAAAAAATCTAAAAAAATCTTTTAAAAAA  109

CCCCAAAAAAATTTACAAAAAAATCCGCGTCTCCCCCGCCGGAGACTTTTATTTTTTTTCTTCCTCTTTTATAAAATAACCCGGTGAAGCAGCCGAGACCGACCCGCCCGCCC  221

GCGGCCCCGCAGCAGCTCCAAGAAGGAACCAAGAGACCGAGGCCTTCCCGCTGCCCGGACCCGACACCGCCACCCTCGCTCCCCGCCGGCAGCCGGCAGCCAGCGGCAGTGG  333

ATCGACCCCGTTCTGCGGCCGTTGAGTAGTTTTCAATTCCGGTTGATTTTTGTCCCTCTGCGCTTGCTCCCCGCTCCCCTCCCCCCGGCTCCGGCCCCCAGCCCCGGCACTC  445

GCTCTCCTCCTCTCACGGAAAGGTCGCGGCCTGTGCCCTGCGGGCAGCCGTGCCGAG   ATG AAC CCC AGT GCC CCC AGC TAC CCC ATG GCC TCG CTC   541
                                                              Met Asn Pro Ser Ala Pro Ser Tyr Pro Met Ala Ser Leu

   TAC GTG GGG GAC CTC CAC CCC GAC GTG ACC GAG GCG ATG CTC TAC GAG AAG TTC AGC CCG GCC GGG CCC ATC CTC TCC ATC CGG   625
14 Tyr Val Gly Asp Leu His Pro Asp Val Thr Glu Ala Met Leu Tyr Glu Lys Phe Ser Pro Ala Gly Pro Ile Leu Ser Ile Arg

   GTC TGC AGG GAC ATG ATC ACC CGC CGC TCC TTG GGC TAC GCG TAT GTG AAC TTC CAG CAG CCG GCG GAC GCG GAG CGT GCT TTG   709
42 Val Cys Arg Asp Met Ile Thr Arg Arg Ser Leu Gly Tyr Ala Tyr Val Asn Phe Gln Gln Pro Ala Asp Ala Glu Arg Ala Leu

   GAC ACC ATG AAT TTT GAT GTT ATA AAG GGC AAG CCA GTA CGC ATC ATG TGG TCT CAG CGT GAT CCA TCA CTT CGC AAA AGT GGA   793
70 Asp Thr Met Asn Phe Asp Val Ile Lys Gly Lys Pro Val Arg Ile Met Trp Ser Gln Arg Asp Pro Ser Leu Arg Lys Ser Gly

   GTA GGC AAC ATA TTC ATT AAA AAT CTG GAC AAA TCC ATT GAT AAT AAA GCA CTG TAT GAT ACA TTT TCT GCT TTT GGT AAC ATC   877
98 Val Gly Asn Ile Phe Ile Lys Asn Leu Asp Lys Ser Ile Asp Asn Lys Ala Leu Tyr Asp Thr Phe Ser Ala Phe Gly Asn Ile

   CTT TCA TGT AAG GTG GTT TGT GAT GAA AAT GGT TCC AAG GGC TAC GGA TTT GTA CAC TTT GAG ACG CAG GAA GCA GCT GAA AGA   961
126 Leu Ser Cys Lys Val Val Cys Asp Glu Asn Gly Ser Lys Gly Tyr Gly Phe Val His Phe Glu Thr Gln Glu Ala Ala Glu Arg

   GCT ATT GAA AAA ATG AAT GGA ATG CTC CTA AAT GAT CGC AAA GTA TTT GTT GGA CGA TTT AAG TCT CGT AAA GAA CGA GAA GCT  1045
154 Ala Ile Glu Lys Met Asn Gly Met Leu Leu Asn Asp Arg Lys Val Phe Val Gly Arg Phe Lys Ser Arg Lys Glu Arg Glu Ala

   GAA CTT GGA GCT AGG GCA AAA GAA TTC ACC AAT GTT TAC ATC AAG AAT TTT GGA GAA GAC ATG GAT GAT GAG CGC CTT AAG GAT  1129
182 Glu Leu Gly Ala Arg Ala Lys Glu Phe Thr Asn Val Tyr Ile Lys Asn Phe Gly Glu Asp Met Asp Asp Glu Arg Leu Lys Asp

   CTC TTT GGG CCT GCC TTA AGT GTG AAA GTA ATG ACT GAT GAA AGT GGA AAA TCC AAA GGA TTT GGA TTT GTA AGC TTT GAA AGG  1213
210 Leu Phe Gly Pro Ala Leu Ser Val Lys Val Met Thr Asp Glu Ser Gly Lys Ser Lys Gly Phe Gly Phe Val Ser Phe Glu Arg

   CAT GAA GAT GCA CAG AAA GCT GTG GAT GAG ATG AAC GGA AAG GAG CTC AAT GGA AAA CAA ATT TAT GTT GGT CGA GCT CAG AAA  1297
238 His Glu Asp Ala Gln Lys Ala Val Asp Glu Met Asn Gly Lys Glu Leu Asn Gly Lys Gln Ile Tyr Val Gly Arg Ala Gln Lys

   AAG GTG GAA CGG CAG ACG GAA CTT AAG CGC AAA TTT GAA CAG ATG AAA CAA GAT AGG ATC ACC AGA TAC CAG GGT GTT AAT CTT  1381
266 Lys Val Glu Arg Gln Thr Glu Leu Lys Arg Lys Phe Glu Gln Met Lys Gln Asp Arg Ile Thr Arg Tyr Gln Gly Val Asn Leu

   TAT GTG AAA AAT CTT GAT GAT GGT ATT GAT GAA GAA CGT CTC CGG AAA GAG TTT TCT CCA TTT GGT ACA ATC ACT AGT GCA AAG  1465
294 Tyr Val Lys Asn Leu Asp Asp Gly Ile Asp Asp Glu Arg Leu Arg Lys Glu Phe Ser Pro Phe Gly Thr Ile Thr Ser Ala Lys

   GTT ATG ATG GAG GGT GGT CGC AGC AAA GGG TTT GGT TTT GTA TGT TTC TCC TCC CCA GAA GAA GCC ACT AAA GCA GTT ACA GAA  1549
322 Val Met Met Glu Gly Gly Arg Ser Lys Gly Phe Gly Phe Val Cys Phe Ser Ser Pro Glu Glu Ala Thr Lys Ala Val Thr Glu

   ATG AAC GGT AGA ATT GTG GCC ACA AAG CCA TTG TAT GTA GCT TTA GCT CAG CGC AAA GAA GAG CGC CAG GCT CAC CTC ACT AAC  1633
350 Met Asn Gly Arg Ile Val Ala Thr Lys Pro Leu Tyr Val Ala Leu Ala Gln Arg Lys Glu Glu Arg Gln Ala His Leu Thr Asn

   CAG TAT ATG CAG AGA ATG GCA AGT GTA CGA GCT GTT CCC AAC CCT GTA ATC AAC CCC TAC CAG CCA GCA CCT CCT TCA GGT TAC  1717
378 Gln Tyr Met Gln Arg Met Ala Ser Val Arg Ala Val Pro Asn Pro Val Ile Asn Pro Tyr Gln Pro Ala Pro Pro Ser Gly Tyr

   TTC ATG GCA GCT ATC CCA CAG ACT CAG AAC CGT GCT GCA TAC TAT CCT CCT AGC CAA GTT GCT CAA CTA AGA CCA AGT CCT CGC  1801
406 Phe Met Ala Ala Ile Pro Gln Thr Gln Asn Arg Ala Ala Tyr Tyr Pro Pro Ser Gln Val Ala Gln Leu Arg Pro Ser Pro Arg

   TGG ACT GCT CAG GGT GCC AGA CCT CAT CCA TTC CAA AAT ATG CCC GGT GCT ATC CGC CCA GCT GCT CCT AGA CCA CCA TTT AGT  1885
434 Trp Thr Ala Gln Gly Ala Arg Pro His Pro Phe Gln Asn Met Pro Gly Ala Ile Arg Pro Ala Ala Pro Arg Pro Pro Phe Ser

   ACT ATG AGA CCA GCT TCT TCA CAG GTT CCA CGA GTC ATG TCA ACA CAG CGT GTT GCT AAC ACA TCA ACA CAG ACA ATG GGT CCA  1969
462 Thr Met Arg Pro Ala Ser Ser Gln Val Pro Arg Val Met Ser Thr Gln Arg Val Ala Asn Thr Ser Thr Gln Thr Met Gly Pro

   CGT CCT GCA GCT GCA GCC GCT GCA GCT ACT CCT GCT GTC GGC ACC GTT CCA CAG TAT AAA TAT GCT GCA GGA GTT CGC AAT CCT  2053
490 Arg Pro Ala Ala Ala Ala Ala Ala Ala Thr Pro Ala Val Arg Thr Val Pro Gln Tyr Lys Tyr Ala Ala Gly Val Arg Asn Pro

   CAG CAA CAT CTT AAT GCA CAG CCA CAA GTT ACA ATG CAA CAG CCT GCT GTT CAT GTA CAA GGT CAG GAA CCT TTG ACT GCT TCC  2137
518 Gln Gln His Leu Asn Ala Gln Pro Gln Val Thr Met Gln Gln Pro Ala Val His Val Gln Gly Gln Glu Pro Leu Thr Ala Ser

   ATG TTG GCA TCT GCC CCT CCT CAA GAG CAA AAG CAA ATG TTG GGT GAA CGG CTG TTT CCT CTT ATT CAA GCC ATG CAC CCT ACT  2221
546 Met Leu Ala Ser Ala Pro Pro Gln Glu Gln Lys Gln Met Leu Gly Glu Arg Leu Phe Pro Leu Ile Gln Ala Met His Pro Thr

   CTT GCT GGT AAA ATC ACT GGC ATG TTG TTG GAG ATT GAT AAT TCA GAA CTT CTT CAT ATG CTC GAG TCT CCA GAG TCA CTC CGT  2305
574 Leu Ala Gly Lys Ile Thr Gly Met Leu Leu Glu Ile Asp Asn Ser Glu Leu Leu His Met Leu Glu Ser Pro Glu Ser Leu Arg

   TCT AAG GTT GAT GAA GCT GTA GCT GTA CTA CAA GCC CAC CAA GCT AAA GAG GCT GCC CAG AAA GCA GTT AAC AGT GCC ACC GGT  2389
602 Ser Lys Val Asp Glu Ala Val Ala Val Leu Gln Ala His Gln Ala Lys Glu Ala Ala Gln Lys Ala Val Asn Ser Ala Thr Gly

   GTT CCA ACT GTT TAAAATTGATCAGGGACCATGAAAAGAAACTTGTGCTTCACCGAAGAAAAATATCTAAACATCGAAAAACTTAAATATTATGGAAAAAAAACATTG 2497
630 Val Pro Thr Val

   CAAAATATAAAATAAATAAAAAAAGGAAAGGAAACTTTGAACCTTATGTACCGAGCAAATGCCAGGTCTAGCAAACATAATGCTAGTCCTAGATTACTTATTGATTTAAAAA 2609

   CAAAAAAACACAAAAAATAGTAAAATATAAAAACAAATTAATGTTTTATAGACCCTGGGAAAAAGAATTTTCAGCAAAGTACAAAAATTTAAAGCATTCCTTTCTTTAATTT 2721

   IGTAATTCTTTACTGTGGAATAGCTCAGAATGTCAGTTCTGTTTTAAGTAACAGAATTGATAACTGAGCAAGGAAACGTAATTTGGATTATAAAATTCTTGCTTTAATAAAA 2833

   ATTCCTTAAACAGTGAAAAAAAAAAAAAAAAAA                                                                              2866
```

repeated units is also reflected at the nucleotide level. There is 67 % homology between nucleotides 1174-1665 of Hu 73 cDNA and nucleotides 1626-2122 of the yeast gene (11) and 61 % homology between nucleotides 533-1092 of Hu 73 cDNA and nucleotides 973-1535 of the yeast gene. We thus expected that this same region would be conserved in other eucaryotes. We used a cDNA fragment covering this region to detect, on Northern blots, the corresponding RNA in two species for which much developmental data on transcription are available : Drosophila melanogaster and Xenopus laevis. Hybridizable RNAs were detected in both species at every developmental stages tested including during oogenesis. The results obtained with Xenopus laevis will be published in a separate paper. In Drosophila melanogaster (Figure 7), an mRNA, approximately 2.9 kb long, is detectable from the very early embryonic stage (0 to 2 hours) before transcription starts. Thus it accumulates in the oocyte, in agreement with the larger abundance of this mRNA in the females than in the males. Smaller RNAs are also detected, but further investigation using fully homologous probe is required to determine the true nature of these signals.

Analysis of the genomic complexity of poly(A) binding protein coding sequence

The human Hu73 cDNA was used to analyse on a Southern blot of human genomic DNA the complexity of the corresponding sequence.This reveals a rather complex pattern (Figure 8) indicating that either there are many introns in the poly(A) binding protein gene or that other related genes or pseudo-genes are present in the genome. Analysis of a genomic library which should allow to distinguish between this two hypothesis is currently under way.

Figure 5 - Nucleotide sequence of the human poly(A) binding protein cDNA and the deduced amino acid sequence of the protein

5A : Sequencing strategy

The cloned cDNA is oriented in the 5' to 3' direction relative to the mRNA. The thick part of the line indicates the coding region. Arrows indicate length, direction and strand of DNA sequences analysed. The thick and thin arrows indicate, respectively, that coding and non coding strands were sequenced. Arrows above and below the line show, respectively, the sequences obtained by the procedures of Maxam and Gilbert (25) and Sanger et al. (24). The bar (100b) represents the length of 100 bases.

5B : Sequence of the coding strand of the human poly(A) binding protein cDNA

Nucleotides are numbered from the first base directly following the poly(dT) stretch (represented by three dots), generated during the synthesis of the double stranded cDNA. The predicted amino acid sequence is shown below the nucleotide sequence. Numbering on the left of the figure is for the amino acids and on the right for the nucleotides. The sequences AATAAA in the 3' non coding region are underlined (———). The amino acid sequences corresponding to the regions forming the four domains, as described in the text, are underlined (- - -) as well as an adenosine rich nucleotide sequence in the 5' untranslated region (------). The vertical arrows indicate the cleavage site (Tth111 I) used to synthesize the shortened polypeptide (330 amino acids) analysed in Figure 4.

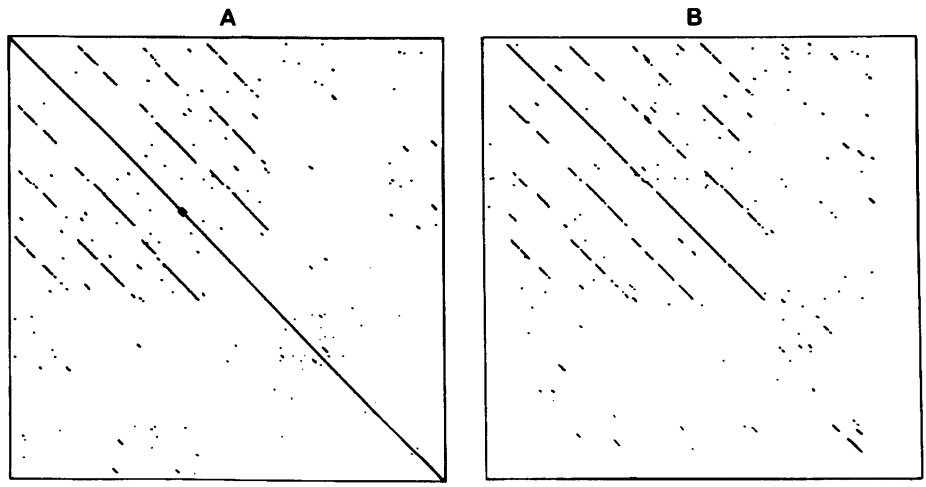


Figure 6 - Homology between the deduced amino acid sequence of yeast and human poly(A) binding protein

A - For internal homology search, the sequence of the human poly(A) binding protein is compared with its own sequence using the Diagon of Staden program (27) with a score of 130 (proportional algorithm) and a span length of 11 amino acids. This comparison reveals a 4 times repeat in the first two-thirds of the protein.

B - The sequence of the human poly(A) binding protein is compared using the same conditions with the recently published yeast poly(A) binding protein (11). The human and yeast sequences are on the ordinate and abscissa, respectively. In order to keep the same scale on both coordinates a blank space equivalent to 56 amino acids has been added at the C terminal end of the yeast sequence. The striking conservation of the N terminal two third of the molecule, including the repeated units, appears clearly.

DISCUSSION

In higher eucaryotes, as in yeast (11), the most prevalent view supports the existence of two poly(A) binding proteins : a high Mr 70-78 kd and a low Mr 55-60 kd protein. However, their nature, their respective location in the cell and their relationship with one another are not clear (6,28,29,30,31). Our data on the human mRNA, in agreement with, but obtained independently from the yeast results, prove without ambiguity that the high Mr poly(A) binding protein is the basic 73 kd protein previously described (6,10) and is not related to the 73 kd acidic heat shock gene product (32).

Sachs et al. (11) have shown that in yeast the two high and low Mr poly(A) binding proteins are encoded by a single gene and a single mRNA. The primary translation product which corresponds to the intact high Mr protein gives rise by proteolysis to the low Mr 53 kd protein corresponding to the N terminal

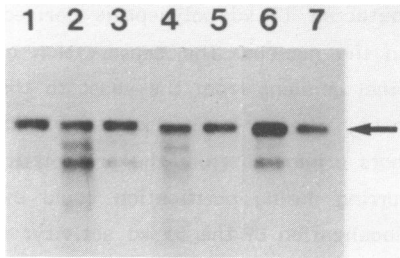


Figure 7 - Accumulation pattern analysis of the mRNA homologous to the human poly(A) binding protein cDNA during Drosophila melanogaster development

The Northern blot of Drosophila melanogaster RNA was kindly provided by A. Vincent (Centre de Biochimie et Génétique cellulaire, Toulouse). The presence of the correct amount of RNA in each lane was verified by screening with DNA sequences corresponding to unmodulated RNAs (A. Vincent, personal communication). This blot was probed with a Hu 73 cDNA fragment extending from nucleotides 627 to 1713. Hybridization was performed in 50 % formamide/5xSSC at 37°C and the final wash was performed in 2xSSC at 55°C. 1 : Embryos from 0 to 12 h post fertilization ; 2 : embryos from 0 to 2 h ; 3 : embryos from 2 to 4 h ; 4 : third instar stage ; 5 : climbing larvae ; 6 : adult female ; 7 : adult male.
Each lane contains 2.5 µg of poly(A)+ mRNA. The arrow marks the position of the mRNA which is approximately 2.9 kb long as determined relatively to a DNA molecular weight marker.

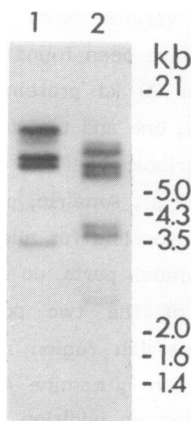


Figure 8 - Genomic organization of the human poly(A) binding protein coding sequence 20 µg of human genomic DNA digested by either EcoRI (lane 1), or HindIII (lane 2) were fractionated on a 1 % agarose gel and blotted onto Zeta probe membrane as described (18). Hybridization was carried out with the full length Hu 73 cDNA labelled by nick-translation. After hybridization the blot was washed at high stringency (0,1xSSC at 65°C).

and to a previously undetected 17 kd polypeptide derived from the C terminal, both exclusively found in the nucleus. The conservation of the same features in both the N and C terminal domains from the yeast to the human poly(A) binding protein, suggests that, in higher eucaryotes also, the 55 kd poly(A) binding protein observed by several authors originates from the proteolytic cleavage of the 73 kd protein. Proteolysis occurring during purification could explain the contradictory results obtained on the localization of the 55 kd activity.

In humans, the high complexity revealed by the Southern blot is compatible with the existence of more than one gene. The unique or largely dominant 2.9 kb mRNA detected by Northern blot analysis and the fact that we have not observed any heterogeneity among 17 cloned cDNA analyzed by restriction site mapping favor however the existence of a single poly(A) binding protein gene coding for a single mRNA.

The determination of the human poly(A) binding protein sequence allows the comparison between two evolutionary distant homologous proteins and should give interesting insight into the regions involved in some specific functions. Conserved features are most likely to be important. The most striking conservation is located in the region which we show in the human protein to contain the poly(A) binding activity (Figure 4B). Excluding the 22 additional amino acids present on the N terminal side of the yeast protein, the homology extends to the whole region containing the four units (Figure 6B). In addition, the respective position and the size of the various units are also maintained. Sequences homologous to the repeated unit have been found in two other nucleic acid binding proteins, the U1 RNA associated 70 kd protein and the rat helix destabilizing protein which contain, respectively, one and two such units (36). Furthermore the C terminal part of the nucleolar preribosomal RNA binding protein, the nucleolin (33) contains 4 similar consecutive units (F. Amalric, personal communication). This unit is thus likely to be a sequence responsible for nucleic acid binding activity.

In contrast the C terminal parts do not have any primary sequence homology, but the comparison of the two poly(A) binding protein sequences reveals the presence of a proline-rich region of about 150 amino acids. This segment, arbitrarily located between glutamine 405 and 540 in yeast and proline 390 and glutamine 557 in human, is, in addition to proline, rich in glutamine and alanine.The three residues together represent about 40 % of the total amino acids in yeast and close to 50 % in human. In yeast, the C terminal part, representing about 30 % of the poly(A) binding protein, seems to constitute an independent domain as it can exist in the nucleus as a 17 kd proteolytic cleavage product even though in this case, it still remains associated with the poly(A)-protein complex

(11). This region is thus likely to be involved in a role distinct from poly(A) binding and related to the function of the protein.

Apart from the structural information, we also present some preliminary data on the expression of the poly(A) binding protein mRNA. We could detect an homologous RNA in all tissues tested as well as during the development of species like Drosophila melanogaster and Xenopus laevis. In these two species, this RNA is present at all stages tested. It starts to accumulate in the oocytes and thus belongs to the class of stable maternal RNAs. The ubiquitous presence and the highly conserved structure suggests that this protein plays a fundamental role in some post-transcriptional event of the poly(A)+ RNA. Experimental data indicate that its mode of action could be at least in part, at the translational level (5,6,7 ; Grossi de Sa et al., in preparation).

A poly(A) rich stretch of 50 nucleotides has been identified in the 5' untranslated region of the yeast poly(A) binding protein mRNA (10,11) and Sachs et al. (11) have shown that the poly(A) binding protein itself can bind to that sequence. In the 5' untranslated region of the human mRNA, there is such an adenosine rich sequence of 60 nucleotides, composed of 6 groups of 6 to 8 adenosine residues separated by 3 to 5 uridine and cytosine residues (see Figure 5). The conservation of this sequence between such evolutionary distant species suggests that it has an important function. This sequence could allow the regulation of the translational activity of the poly(A) binding protein messenger RNA by its own translation product. There are some data supporting this hypothesis. The poly(A) binding protein synthesis is increased early after serum stimulation and this increase is inhibited by actinomycin D (34). However this serum effect is not at the transcriptional level since no increase of the corresponding mRNA was detected either by in vitro translation of cellular mRNAs (5) or by Northern blot analysis with the Hu73 cDNA (results not shown). One can hypothesize that an excess of poly(A) binding protein, following, for example, the inhibition of poly(A)+ mRNA appearance leads to the binding of this protein at the 5' end of its own mRNA, thus inhibiting its own synthesis. Stimulation of poly(A)+ mRNA appearance would, on the other hand, titrate away the poly(A) binding protein from the 5' end of its own mRNA, thus leading to a reactivation of its translation. Such a mechanism is compatible with several observations. Schwartz and Darnell (35) have shown that newly synthesized poly(A) binding protein is associated with the poly(A) tail of newly synthesized mRNA. Furthermore, Maundrell et al. (22) found that in duck reticulocytes there is a repression of the poly(A) binding protein mRNA translation concomitant to the physiological shutdown of mRNA synthesis.

AKNOWLEDGEMENTS

We thank G. Beck and J. Axelrod for the gifts of poly(A)+ RNA, A. Vincent for kindly providing us with his Northern blots, O. Akhayat, S. Ben Tahar and A. Vincent for helpful discussions, C. Mugnier for computer assistance, R. Schwartzmann for the photographs and C. Sagot for typing the manuscript.

This work was supported by the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Association pour la Recherche sur le Cancer (ARC) and the Ligue Nationale Française contre le Cancer. T.G. was supported in part by a fellowship from the Fondation pour la Recherche Médicale. C. M. de S. was supported by a fellowship from the Brazilian Conselho Nacional de Desenvolvimento Cientifico e Technologico (CNPq).

REFERENCES

1. Brawerman G. (1981) CRC. Critical Rev. of Biochem. 10 : 1-38
2. Drummond D.R., Armstrong J. and Colman A. (1985) Nucl. Acids Res. 13 : 7375-7394
3. Preobrazhensky A.A. and Spirin A.S. (1978) Prog. Nucl. Acid Res. Mol. Biol. 21 : 1-38
4. Blobel G. (1973) Proc. Natl. Acad. Sci. USA 70 : 924-928
5. van Venrooij W.J., van Eckelen C.A.G., Jansen R.T.P. and Princen J.M.G. (1977) Nature 270 : 189-191
6. Vincent A., Goldenberg S. and Scherrer K. (1981) Eur. J. Biochem. 114 : 179-193
7. Jacobson A. and Favreau M. (1983) Nucl. Acids Res. 11 : 6353-6368
8. Littauer U.Z. and Soreq H. (1982) Prog. Nucl. Acid Res. Mol. Biol. 27 : 53-83
9. Bablanian R., and Banerjee A.K. (1986) Proc. Natl. Acad. Sci. USA 83 : 1290-1294
10. Adam S.A., Nakagawa T., Swanson M.S., Woodruff T.K. and Dreyfuss G. (1986) Mol. Cell. Biol. 6 : 2932-2943
11. Sachs A.B., Bond M.W. and Kornberg R.D. (1986) Cell 45 : 827-835
12. Thompson E.B., Tomkins G.M. and Curran J.F. (1966) Proc. Natl. Acad. Sci. USA 56 : 296-306
13. Maniatis T., Fritsch E.F. and Sambrook J. (1982) Molecular clining : a laboratory manual, Cold Spring Harbor Laboratory New York (USA)
14. Reich E. (1975) in Proteases and Biological control (Eds. Reich E., D.B. Rifkin and E. Shaw) pp. 333-341 (Cold Spring Harbor Laboratory, New York)
15. Grange T., Guénet C., Dietrich J.B., Chasserot S., Fromont M., Befort N., Jami J., Beck G. and Pictet P. (1985) J. Mol. Biol. 184 : 347-350
16. Cooke N.E., Coit D., Weiner R.I., Baxter J.D. and Martial J.A. (1980) J. Biol. Chem. 255 : 6502-6510
17. Huynh T.V., Young R.A. and Davis R.W. (1985) (Ed. Glover D.) pp. 49-78 (IRL, Oxford).
18. Rigaud G., Grange T. and Pictet R. (1987) Nucl. Acids Res. 15 : 857.
19. Jackson R.J. and Hunt T. (1983) Methods Enzymol. 96 : 50-74
20. O'Farrell P.Z., Goodman H.M. and O'Farrell P.H. (1977) Cell 12 : 1133-1142
21. Cleveland D.W., Fischer S.G., Kirschner M.W. and Laemmli U.K.(1977) J. Biol. Chem. 252 : 1102-1106
22. Maundrell K., Imaizumi-Scherrer M.T., Maxwell E.S., Civelli O. and Scherrer K. (1983) J. Biol. Chem. 258 : 1387-1390

23. Henikoff S. (1984) Gene 28 : 351-359
24. Sanger F., Nicklen S. and Coulson A.R. (1977) Proc. Natl. Acad. Sci. USA 74 : 5463-5467
25. Maxam A.M. and Gilbert W. (1980) Meth. Enzymol. 65 : 499-560
26. Standart N., Vincent A. and Scherrer K. (1981) FEBS Letters 135 : 56-60
27. Staden R. (1982) Nucl. Acids Res. 10 : 2951-2961
28. Jeffery W.R. (1977) J. Biol. Chem. 252 : 3525-3532
29. Kish V.M. and Pederson T. (1976) J. Biol. Chem. 251 : 5888-5894
30. Sachs A.B. and Kornberg R.D. (1985) Mol. Cell. Biol. 5, 1993-1996
31. Setyono B. and Greenberg J.R. (1981) Cell 24 : 775-783
32. Schonfeider M., Horsch A. and Schmid H.P. (1985) Proc. Natl. Acad. Sci. USA 82 : 6884-6888
33. Lapeyre B., Bourbon H. and Amalric F. (1987) Proc. Natl. Acad. Sci. USA - 84 : 1472-1476
34. Thomas G. and Thomas G. (1986) J. Cell. Biol. 103 : 2137-2144
35. Schwartz H. and Darnell J.E. (1976) J. Mol. Biol. 104 : 833-851
36. Theissen H., Etzerodt M., Reuter R., Schneider C., Lottspeich F., Argos P., Luhrmann R. and Philipson L. (1986) EMBO J. 5 : 3209-3217
37. Parker P.J., Coussens L., Totty N., Rhee L., Young S., Chen E., Stabel S., Waterfield M.D. and Ullrich A. (1986) Science 233 : 853-859
38. Kozak M. (1986) Cell 44 : 283-292
39. Wray W., Boulikas T., Wray V.P. and Hancock R. (1981) Anal. Biochem. 118 : 197-203