Volume 9 Number 19 1981 Nucleic Acids Research

An unusually long non-coding region in rat lens α-crystallin messenger RNA

Rob J.M.Moormann*, Huub M.W.van der Velden*, Huub J.Dodemont†, Peter M.Andreoli*, Hans Bloemendal† and John G.G.Schoenmakers*

*Department of Molecular Biology, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, and †Department of Biochemistry, University of Nijmegen, Geert Grooteplein 21, 6525 EZ Nijmegen, The Netherlands

Received 5 August 1981

ABSTRACT

Most of the mRNA sequence coding for the αA_2 chain of the ocular lens protein α-crystallin from rat, has been determined by sequencing cloned DNA copies of this mRNA. The 892-base pair cDNA sequence encompasses all but 52 N-terminal amino acids of the αA_2 chain. It lacks the sequence characteristic for the 22 extra amino acids inserted in the αA_2-like chain, named αA^{Ins}. A stretch of 583 nucleotides, representing more than 50% of the entire mRNA sequence, is located 3' wards of the αA_2 coding sequence. It contains the characteristic AAUAAA signal involved in poly(A)-addition and represents an unexpectedly long non-coding region. Examination of the total cytoplasmic poly(A)RNA of rat lens by filter-hybridization and subsequent translation of the electrophoretically separated mRNA fractions shows that the αA_2 chain is encoded by mRNA species which are distinct from the αA^{Ins}-encoding mRNA. No evidence is obtained for an extensive size heterogeneity in the 3' untranslated regions of these two different rat lens mRNAs.

INTRODUCTION

The structural protein of the ocular lens, α-crystallin, has been studied in great detail with emphasis on subunit composition, biosynthesis and structure (1,2). It is the predominant biosynthetic product in the adult vertebrate lens and represents a convenient example of controlled gene expression in terminally differentiated cells. The high molecular weight protein is composed of two types of primary gene products, named αA_2 (mol. wt. 19,832) and αB_2 (mol.wt. 20,070), which show about 60% of homology suggesting they originate from one ancestral gene. The synthesis of αB_2 (175 amino acids) is directed by a message of the expected size, namely 10S mRNA. The αA_2 chains (173 amino acids) are encoded, however, by a rather long 14S mRNA. Though the latter mRNA of about 1300 nucleotides is theoretically large enough to encode two polypeptides of that size and Chen and Spector (3) recently put forward that the αA_2 mRNA of calf lens is of bicistronic character, we disfavour such a conclusion.

In view of the primary structure of many eukaryotic mRNAs determined in

recent years, the length of the 5' and 3' non-coding regions can vary to a large extent (4-6). Our findings that the αA_2 and αB_2 chains are synthesized on mRNAs with a striking difference in length might be explained on the basis of the observed variance in non-coding regions. Although such a region of about 750 nucleotides is unusually long for a eukaryotic mRNA, its occurrence is apparently not limited to rat lens since 14S mRNA coding for an αA_2 chain of about 175 amino acids only has also been demonstrated in other species, like chicken (7) and calf (8). Another point of interest is that in rat lens an additional αA_2 chain has been detected, named αA^{Ins}, with 22 extra amino acids inserted in an internal position of the molecule. Its synthesis is also directed by a 14S mRNA (9,10). In how far this insertion reflects a distortion of the splicing mechanism (11) in rat lens, has still to be ascertained.

To address these questions and as a basis for further research we have cloned the rat 14S mRNA as a series of recombinant DNA plasmids and have determined the nucleotide sequence of these clones. Our conclusion from the sequence data is that the clones represent αA_2 sequences rather than αA^{Ins} sequences. No coding function can be assigned to the 3' stretch of about 590 nucleotides which apparently represents an extremely long non-coding part of this crystallin mRNA in rat.

MATERIALS AND METHODS

Isolation of recombinant plasmid DNA

The construction of bacterial plasmids containing sequences homologous to rat lens crystallin mRNAs inserted into the PstI-site of pBR322 by oligo-dG-dC tailing and the subsequent amplification by transformation of E.coli χ1776 has been described elsewhere (12). After characterization of recombinants by a positive hybridization assay (12) appropriate candidates were taken for transformation of the EK1 host E.coli HB101. All manipulations were carried out in a Category II containment laboratory as recommended by the Netherlands Committee *Ad Hoc* on Recombinant DNA Research. Recombinant plasmid DNA was isolated by the cleared lysate procedure of Clewell and Helinski (13) and further purified by CsCl density gradient centrifugation.

Isolation, gel electrophoresis and hybridization of RNA

Total cytoplasmic poly(A)RNA from lenses of 4-day old rats were isolated essentially as described by Palmiter (14).

Electrophoresis of cytoplasmic poly(A)RNA (20 µg) was performed on a

vertical 35 cm long 1.5% agarose gel in the presence of 10 mM CH_3HgOH (15). For blot-hybridization the RNA was transferred to nitrocellulose paper according the procedure of Thomas (16) and hybridized with ^{32}P-labelled, nick-translated DNA probes from α-crystallin-specific cDNA clones using the conditions described by Wahl *et al.* (17).

Translation of RNA and analysis of translation products

After electrophoresis of RNA, one half of the gel was used for blotting and hybridization to the α-crystallin-specific cDNA clone pRLαA-1. The other half was used for translation experiments. The regions of the gel which hybridized to ^{32}P-labelled pRLαA-1 DNA were sliced into 1-mm segments. RNA was isolated from the gel segments by homogenizing each slice in 0.3 ml of 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5 mM EDTA, 10 mM DTT, 0.2% SDS containing 10 µg of *E.coli* tRNA and extracting overnight with buffer-saturated phenol, followed by re-extraction of the water phase with phenol and chloroform-isoamylalcohol (24:1). After two successive ethanol precipitations the RNA pellet was dissolved in 5 µl water. Aliquots of 1 µl were assayed in a nuclease-treated rabbit reticulocyte lysate (18). The translation products were analysed on a 13% polyacrylamide-SDS gel. Autoradiography was performed at -80°C after processing according to Bonner and Laskey (19).

Restriction endonuclease analysis and DNA sequencing

Restriction digests of plasmid DNA were electrophoresed in 2% agarose horizontal slab gels containing 40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.8. The isolation and purification of restriction fragments and the 5' end labelling with T4 polynucleotide kinase was carried out as described previously (20). DNA sequencing was performed according to the method of Maxam and Gilbert (21). Fragments were processed through the G, G+A, C+T, C and A>C specific, limited chemical cleavage reactions and applied to 0.4 mm thin 6%, 8% and 20% polyacrylamide gels (22).

RESULTS AND DISCUSSION

We recently reported the construction of recombinant plasmids containing sequences homologous to several rat lens crystallin mRNAs (12). Plasmid DNA of the various crystallin-specific clones selected were analysed for insert length and further characterized by hybrid-selected translation.

Two representative clones, pRLαA-1 and pRLαA-3, which directed the synthesis of αA_2 chains and concomitantly the synthesis of αA^{Ins} chains, were

selected for sequencing according to the method of Maxam and Gilbert (21). The restriction map of each insert relative to the flanking pBR322 sequences, of which only the nearest *Hin*fI sites are indicated, is shown in Fig. 1. The positions of the various restriction enzyme cleavage sites found in each map already suggest that both DNA inserts represent overlapping sequences. pRLαA-1 carries a more extended 5' region whereas pRLαA-3 has a more extended region at its 3' end. The strategy used in determining the nucleotide sequence of each insert is included in Fig. 1. To ensure the reliability of the final sequence, most of the sequence operations were carried out at least twice. Furthermore, the nucleotide sequence obtained has been determined independently from two different clones which appear to have a very large

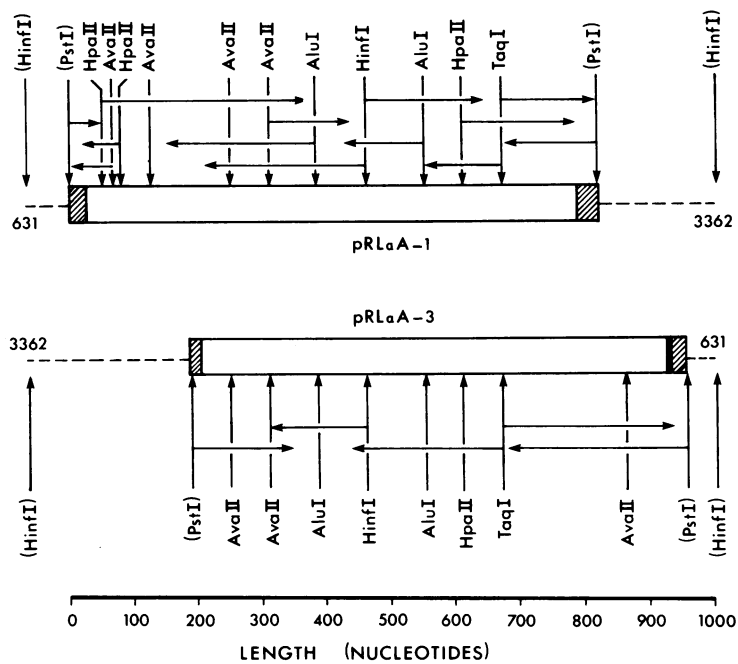

Figure 1
Restriction map of pRLαA-1 and pRLαA-3 showing the restriction sites used for sequence analysis of the rat lens αA_2-crystallin mRNA. The horizontal arrows are indicative for the orientation and approximate length of sequences obtained. The cleavage coordinates of the *Hin*fI-sites neighbouring the *Pst*I-site of pBR322 are indicated (28). Note that the cDNAs are inserted in a reversed orientation in the vector. The open boxes represent mRNA sequences, the hatched boxes are the poly(dG-dC)-tracks, the black box in pRLαA-3 represents a stretch of seven A-residues which is probably a remnant of the poly(A)-track of the mRNA template used in cDNA synthesis

part of their final DNA sequence in common. The combined nucleotide sequence of the αA_2-crystallin mRNA as deduced from the sequence data of each clone is given in Fig. 2. The regions flanking the insert in pRLαA-1 at the 5' and 3' side, extending to the plasmid *Hin*fI sites have the expected nucleotide sequence surrounding the *Pst*I site of insertion in the parent plasmid pBR322 (data not shown). Consequently, generation of two new *Pst*I sites by the host after insertion of DNA into the original *Pst*I site has occurred correctly. The length of the GC-linkers was found to be 30 base pairs at the 5' end and 29 base pairs at the 3' end of the mRNA sequence. The size of the crystallin mRNA sequence in pRLαA-1 as deduced from the sequence data is 759 base pairs. Its sequence is deprived of the hexanucleotide sequence AATAAA, considered to be the signal involved in polyadenylation (23,24), which suggests this DNA insert lacks a complete 3' region of the αA_2 mRNA sequence.

The nucleotide sequence of the crystallin-specific DNA in the second clone, pRLαA-3, is 720 nucleotides long. By comparing this sequence with that of pRLαA-1, it appears that a sequence of 580 base pairs is identical in both inserts. From this observation we conclude that both clones contain a DNA copy which originates from the same mRNA template. pRLαA-3 contains in addition 126 nucleotides at its 3' end. This sequence encompasses the characteristic AATAAA sequence which is found 21 nucleotides upstream of the site of poly(A) addition.

The amino acid sequence of the rat lens αA_2 polypeptide chain has been elucidated earlier (7). The nucleotide sequence from position 2 - 365, as given in Fig. 2, exactly fits the amino acid sequence in that it represents the sequence ranging from residue 53 to 173, *i.e.* the C-terminal serine residue of the αA_2 chain. Another conclusion which could be drawn is that the clones analysed represent an αA_2-mRNA sequence and not an αA^{Ins}-mRNA sequence. αA^{Ins}, which is encoded also by a 14S mRNA, has 22 extra amino acids inserted between amino acid residue 63 (Glu) an 64 (Val) of the αA_2 chain (7). The DNA sequence run of this particular region, shown in Fig. 3, fits the αA_2 amino acid sequence only.

In our first study on lens mRNA we estimated that the 14S mRNA encoding the αA_2 chain might theoretically comprise a coding region sufficiently long to direct the synthesis of two products of that size (8), although we gained experimental evidence which made the assumption of a bicistronic eukaryotic message very unlikely (25). In contrast, Chen and Spector (3) reported on the basis of product analysis after translation of 14S mRNA in a cell-free system that the message of calf lens might be bicistronic. The

```
                              20                                  40
TC TTC CGC ACA GTG TTG GAC TCC GGC ATC TCT GAG GTC CGA TCT GAC CGG GAC AAG
   Phe Arg Thr Val Leu Asp Ser Gly Ile Ser Glu Val Arg Ser Asp Arg Asp Lys

  60                          80                                 100
TTT GTC ATC TTC TTG GAT GTG AAG CAC TTC TCT CCT GAG GAC CTC ACC GTG AAG GTA
Phe Val Ile Phe Leu Asp Val Lys His Phe Ser Pro Glu Asp Leu Thr Val Lys Val

     120                         140          → pRLαA-3      160
CTG GAA GAT TTC GTG GAG ATC CAT GGC AAA CAC↓AAC GAG AGG CAG GAT GAC CAT GGC
Leu Glu Asp Phe Val Glu Ile His Gly Lys His Asn Glu Arg Gln Asp Asp His Gly

         180                         200                         220
TAC ATT TCC CGT GAA TTT CAC CGT CGC TAC CGT CTG CCT TCC AAT GTG GAC CAG TCC
Tyr Ile Ser Arg Glu Phe His Arg Arg Tyr Arg Leu Pro Ser Asn Val Asp Gln Ser

             240                         260                         280
GCC CTC TCC TGC TCC TTG TCT GCG GAT GGC ATG CTG ACC TTC TCT GGC CCC AAG GTC
Ala Leu Ser Cys Ser Leu Ser Ala Asp Gly Met Leu Thr Phe Ser Gly Pro Lys Val

                   300                         320                         340
CAG TCT GGC TTG GAT GCT GGC CAC AGC GAG AGG GCC ATT CCC GTG TCA CGG GAG GAG
Gln Ser Gly Leu Asp Ala Gly His Ser Glu Arg Ala Ile Pro Val Ser Arg Glu Glu

                       360                         380                         400
AAG CCC AGC TCG GCA CCC TCG TCC TGA GC AGGCCTCGCC TTGGTTGTCC CCTGATGCCC CTG
Lys Pro Ser Ser Ala Pro Ser Ser

                 420                         440                         460
ATCCATC TGCCCAGGGG CCACAGCAAA GAGTCTGCCT TCCTGACTTC TTTTCTTTCT CTTTGTTTCC T

       480                         500                         520                         540
TTCCACTTT CTCAGAGGGC TGAGGATTTG AGAGAGTGGC TTAAAGAGCT TGGGGGGTCT TGGCCTGAGA

                     560                         580                         600
TGGCTGCGGG TTCAGGGTGA CCCAGGCTCA ACACCAGCCG GTCAGAGGGA ATGATGGCAT TGAACTCTT

           620                         640                         660
A AGATTTCCTG TCCTCCTGGA AAGTGGCATC GAGCTCTGCC AAAGGCAGAG TGAATGGTGG CTAACCA

680                         700                         720                         740
ACC CCAAGAGCCC TCTGCCAAGC CCCTGGATGG CAGCCTCCCA CCCCCTTTGC CCACACTTAC CGCAG

  pRLαA-1 ← 760                         780                         800
GCGTA TATGCTGGGC↓TCCAACAGTC CGCTTCTCTC ATGCCCTCTT CCTGTGACTT TCTCTACTAT GTA

    820                         840                         860                         880
GTATCGC TCCTGGGGAC CCTGATCACC CATGAGAATG GGCCCCTGGC AGAC[AATAAA] GAGCAGGTGA

CAAGCAAAAA AA
```

Figure 2
The combined nucleotide sequence of the rat lens αA_2-crystallin cDNA inserts in pRLαA-1 and pRLαA-3 and its deduced amino acid sequence. Numbers above each line refer to nucleotide positions. The boxed region specifies the nucleotide sequence involved in poly(A)-addition. The nucleotide sequence of pRLαA-1 ranges from 1-759, that of pRLαA-3 from 146-892

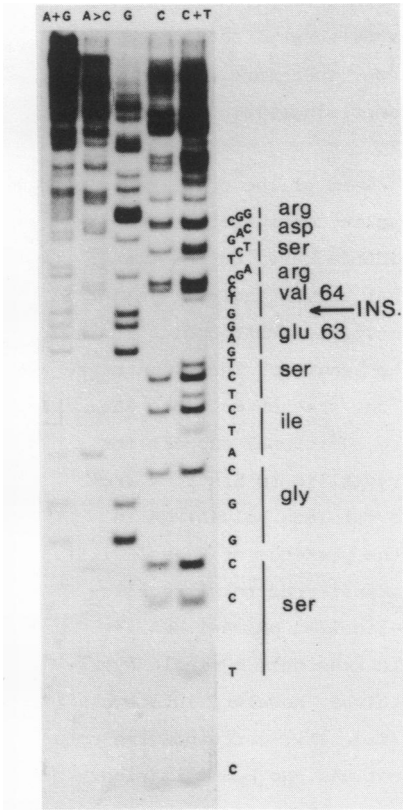

Figure 3
Autoradiogram of a sequencing gel showing the region around the amino acid residues 63(Glu) and 64(Val) of the αA_2-crystallin. The deduced amino acid sequence fits the known amino acid sequence of αA_2 from rat lens and lacks the insertion of 22 extra amino acid residues between the residues 63-64 which is characteristic for the αA^{Ins} amino acid sequence (7).

nucleotide sequence established here for 14S mRNA argues strongly against a dual coding function. The coding frame given for the αA_2-mRNA is the only reading frame which is capable of directing a polypeptide significantly long to be considered. In the other two possible reading frames several termination codons are present which are randomly distributed, giving rise to very short polypeptides only.

Akusjärvi *et al.* (26) recently demonstrated that the non-coding part of the adenovirus hexon-mRNA discloses a novel adenovirus gene. In view with these findings we considered also the possibility that the 5' part of the DNA sequence, as demonstrated, codes for αA_2 chains and that the 3' part codes (in part) for a crystallin-like protein of similar size as that of αA_2. The length of the region ranging from the C-terminal end of the αA_2 coding sequence to the start of the poly(A)-tail comprises about 550 nucleotides and could theoretically code for a protein of such length. However,

in all three possible reading frames only 2 and 3 ATG triplets are present and in each case these ATG triplets are followed by several stop codons in the same phase at distances of 40 triplets or less. Therefore, we conclude that the 3' region of the αA_2 message, though comprising over 50% of the total mRNA sequence, has no coding function.

The significance of untranslated regions 3' ward of the protein coding sequence in eukaryotic mRNAs remains unclear despite considerable speculations. Kronenberg *et al.* (27) have presented evidence based on the translational capability of truncated globin mRNAs that argues against a major role for the 3' untranslated region. Moreover, multiple mRNAs coding for a single protein, and which differ primarily in the length of the 3' untranslated regions have been identified (4). To find out whether multiple αA_2-crystallin mRNAs exist in rat lens and to acquire additional support for our conclusion that the mRNA coding for αA^{Ins}-crystallin is distinct from the αA_2 mRNA, we have analysed total cytoplasmic rat lens poly(A)RNA in more detail. Poly(A)RNA was electrophoresed in the presence of CH_3HgOH (15). Thereafter, the RNA on one half of the gel was transferred to nitrocellulose paper and subsequently hybridized with ^{32}P-labelled pRLαA-1 DNA. As noted before (*cf.*12), after short electrophoretic runs only a single hybridizing band was apparent. This band could be resolved, however, into two distinct bands after prolonged electrophoresis (Fig. 4A), corresponding to RNA species of approximately 1300 and 1250 nucleotides. No hybridization occurred in the other regions of the gel, indicating that size heterogeneity of the α-crystallin-specific mRNA is very unlikely. On the other hand, hybridization with pRLαA-1 DNA does not discriminate between αA_2 and αA^{Ins}-specific sequences as these two proteins are very homologous in their amino acid sequence. To find out whether the two RNA bands represent two different αA_2 mRNAs which differ in length or that these bands represent αA_2 and αA^{Ins} mRNA species, translation studies have been carried out. Of the second half of the agarose gel the region comprising the two RNA species was sliced into 2-mm fractions and the RNA contained in each fraction was translated in a rabbit reticulocyte lysate system. From the results presented in Fig. 4B, it is clear that the fraction(s) representing the 1250 nucleotide-long RNA were capable of synthesizing αA_2 polypeptide chains only, whereas synthesis of αA^{Ins} polypeptides is only apparent in the fraction containing the 1300 nucleotide-long RNA. These data strongly suggest that the mRNAs coding for αA^{Ins} and αA_2-crystallin are distinctly different and can be well separated from each other. No evidence is obtained for the presence of αA_2 mRNA species

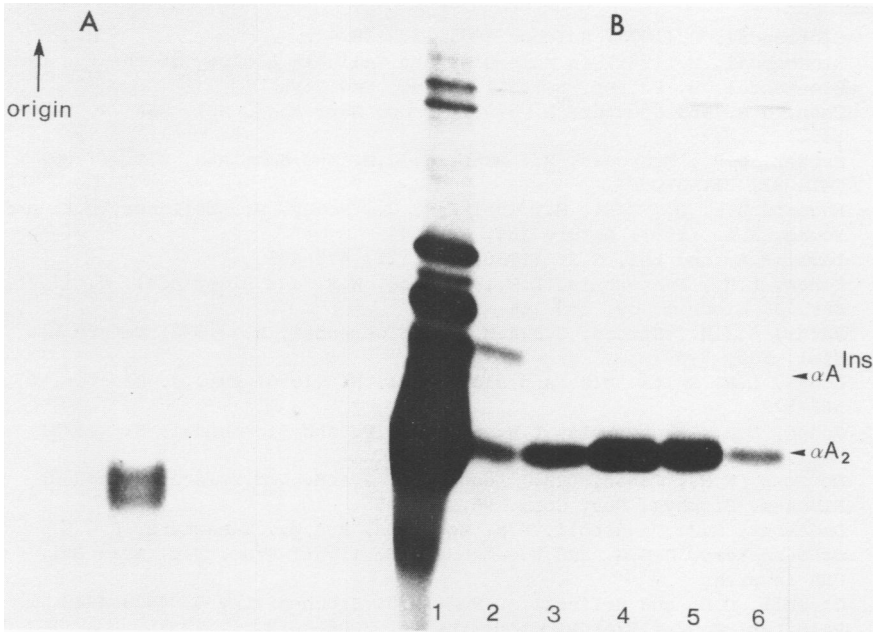

Figure 4

A. Hybridization of total cytoplasmic rat lens poly(A)RNA with ^{32}P-labelled pRLαA-1 DNA after electrophoresis of RNA in the presence of methyl mercuri and transfer to nitrocellulose paper.

B. Translation of the RNA contained in the various gel segments after electrophoresis of rat lens poly(A)RNA. The region of the gel containing the doublet band as vizualised under A. was sliced into 2-mm, segments. RNA was extracted from each segment as described under Materials and methods and translated in a nuclease-treated reticulocyte lysate. Lane 1 represents the translation products from the unfractionated poly(A)RNA, lane 2-6 show the polypeptides encoded by the RNA in the gelsegments. The length of the RNA decreases from left to right.

which have truncated a very large part of their 3' sequence although a slight variance in the length of the non-coding region or the poly(A)-tail cannot be discriminated by our analysis. Further studies at the chromosomal level are needed to find out whether αA^{Ins} and αA_2 mRNA originate from separate genes.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization of Pure Research (ZWO).

REFERENCES

1. Bloemendal, H.(1977) Science 197, 127-138
2. Bloemendal, H.(1981) in Molecular and Cellular Biology of the Eye Lens, Bloemendal, H. Ed. pp. 5.1-5.12 Wiley, New York
3. Chen, J.H. and Spector, A. (1977) Proc. Nat. Acad. Sci. USA 74, 5448-5452
4. Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schminke, R.T. (1980) Cell 22, 361-370
5. Hagenbüchle, O., Tosi, M., Schibler, U., Bovey, R., Wellauer, P.K. and Young, R.A. (1981) Nature 289, 643-646
6. Lamb, R.A. and Lai, C.J. (1980) Cell 21, 475-485
7. Cohen, L.H., Westerhuis, L.W., de Jong, W.W. and Bloemendal, H. (1978) Eur. J. Biochem. 89, 259-266
8. Berns, A.J.M., Strous, G.J.A.M. and Bloemendal, H. (1972) Nature New Biol. 236, 7-9
9. Cohen, L.H. Smits, D.P. and Bloemendal, H. (1976) Eur. J. Biochem. 67, 563-572
10. Cohen, L.H., Westerhuis, L.H.,Smits, D.P. and Bloemendal, H. (1978) Eur. J. Biochem. 89, 251-258
11. de Jong, W.W., Cohen, L.H., Leunissen, J.A.M. and Zweers, A. (1980) Biochem. Biophys. Res. Comm. 96, 648-655
12. Dodemont, H.J., Andreoli, P.M. Moormann, R.J.M., Ramaekers, F.C.S., Schoenmakers, J.G.G. and Bloemendal, H. (1981) Proc. Nat. Acad Sci. USA in press
13. Clewell, D.D. and Helinski, D.R. (1970) Biochemistry 9, 4428-4440
14. Palmiter, R.D. (1974) Biochemistry 13, 3606-3615
15. Bailey, J.M. and Davidson, N. (1976) Anal. Biochem. 70, 75-85
16. Thomas, P. (1980) Proc. Nat. Acad. Sci. USA 77, 5201-5205
17. Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Nat. Acad. Sci. USA 76, 3683-3687
18. Pelham, H.R.B. and Jackson, R.J. (1976) Eur. J. Biochem. 67, 247-253
19. Bonner, W.J. and Laskey, R.A. (1974) Eur. J. Biochem. 46, 83-88
20. van Wezenbeek, P.M.G., Hulsebos, T.J.M. and Schoenmakers, J.G.G. (1980) Gene 11, 129-148
21. Maxam, A. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564
22. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110
23. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214
24. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260
25. Favre, A., Bertazzoni, U., Berns, A.J.M. and Bloemendal, H. (1974) Biochem. Biophys. Res. Comm. 56, 273-280
26. Akusjärvi, G. Zabielski, J., Perricaudet, M. and Petterson, U. (1981) Nucleic Acids Res. 9, 1-17
27. Kronenberg, H.M., Roberts, B.E. and Efstratiadis, A. (1979) Nucleic Acids Res. 6, 153-166
28. Sutcliffe, J.G. (1978) Nucleic Acids Res. 5, 2721-2728