Vol. 148, No. 1, 1987 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 14, 1987 Pages 31-37

NUCLEOTIDE SEQUENCE OF A BOVINE LENS αA-CRYSTALLIN cDNA

Regine E. Hay and J. Mark Petrash*

Department of Ophthalmology
Emory University School of Medicine
Atlanta, Georgia 30322

Received August 24, 1987

We have determined the nucleotide sequence of a bovine lens αA2-crystallin cDNA clone, designated pBLαA2-1. The 793 bp cDNA insert contains coding information for the entire 173 amino acid αA2-crystallin polypeptide, as well as nontranslated sequences located both upstream and downstream from the coding region. The coding sequences contained in pBLαA2-1 are at least 89% homologous with the corresponding sequences from other mammalian αA-crystallin genes, and are 78% homologous to the frog αA-crystallin coding region. In contrast, the downstream nontranslated sequences of the mammalian αA-crystallin transcripts show much greater sequence divergence, with the bovine sequences averaging 47% homology with the corresponding sequences from other mammalian species.

Eye lenses of vertebrate species contain a high concentration of soluble proteins called crystallins. Mammalian crystallins may be grouped according to immunological and structural characteristics into three classes, α, β, and γ (1). Each class may be further subdivided into multiple structurally related polypeptides. The α-crystallins derive from two primary gene products, αA2- and αB2-crystallins, which are synthesized as M_r 20,000 polypeptides (2). Recent studies have shown that the αA2- and αB2-crystallins may be modified post-translationally by a cAMP-dependent phosphorylation reaction, resulting in the production of αA1- and αB1-crystallins (3-5). The α-crystallins are characterized as slowly evolving proteins (6,7). Based on comparisons of the αA-crystallin amino acid sequence for a variety of mammalian species, the average evolutionary rate is one amino acid substitution per 100 residues per 40 million years (6,8). Indeed, the bovine αA2- and αB2-crystallins are about 57% homologous even though the genes coding for these proteins probably duplicated and diverged more than 450 million years ago (9). Expression of αA-crystallin in

*Corresponding Author.

0006-291X/87 $1.50

the lens involves several interesting processes. The 14S mRNA encoding bovine (10) and rodent lens αA-crystallin is approximately 1200 nucleotides in length, more than two times the size required for translation of the M_r 20,000 polypeptide (11,12). In some rodents, the αA-crystallin mRNA transcript undergoes differential splicing, a process that provides templates for the synthesis of both the normal αA- as well as the αA^{ins}- polypeptide (11,13,14). The tissue-specific and developmentally-regulated transcription of the murine αA2-crystallin gene appears to be controlled by regulatory elements located upstream from the transcriptional start site (15,16).

Genomic or cDNA αA-crystallin clones have been reported for mouse (12), rat (11), hamster (17), human (18), chicken (19), and frog (Rana temporaria) (20). While the synthesis and posttranslational metabolism of the bovine αA-crystallins have been extensively studied at the protein level, their corresponding cDNA or genomic sequence have not been reported. In this communication, we present the nucleotide sequence of a bovine lens αA2-crystallin cDNA clone. Portions of this work have been presented in abstract form (21).

MATERIALS AND METHODS

Materials. All restriction enzymes were obtained from either Bethesda Research Laboratories or Boehringer Mannheim Biochemicals. Reverse transcriptase and 7-deaza-dGTP used for DNA sequencing were obtained from Promega Biotech and Boehringer Mannheim Biochemicals, respectively. Oligodeoxynucleotide sequencing primers were prepared with an Applied Biosystems Model 381A DNA synthesizer using reagents purchased from the manufacturer.

Isolation and Nucleotide Sequencing of cDNA Clones. αA-crystallin cDNA clones were isolated from a total bovine cDNA library using a synthetic oligonucleotide hybridization probe as described previously (22). Based on restriction mapping and blot hybridization analysis, the cDNA clone designated pBLαA2-1 was selected as the most informative among the 9 αA-crystallin cDNA clones initially isolated. For nucleotide sequence analysis by the dideoxy chain termination method (23), the cDNA insert from pBLαA2-1 was transferred into bacteriophage M13, mp18 and mp19. Initially, sequencing reactions were primed with the 30-mer hybridization probe used for library screening. Subsequent sequence determinations were carried out using primers constructed to anneal at unique sites within the cDNA sequence. In some cases, restriction fragments of pBLαA2-1 were subcloned into the multicloning site of M13 and were sequenced from the universal primer binding site. Sequence determinations in G+C rich areas required the use of reverse transcriptase (24) and 7-deaza-dGTP (25). Sequencing data were analyzed using the NUMSEQ (26) and NUCALN programs from a software package made available by David Mount of the University of Arizona, Tuscon. The NUCALN program produces optimal alignments of two nucleic acid sequences based on the algorithm described by Wilbur and Lipman (27).

RESULTS AND DISCUSSION

The cDNA insert contained in pBLαA2-1 was sequenced using the strategy given in Figure 1. The 793 bp cDNA insert contains 519 bp of sequence encoding the αA2-crystallin polypeptide, as well as apparently untranslated sequences located both upstream and downstream from the coding region (Figure 2). Translation of the sequence from the initiation codon commencing at nucleotide #31 reveals an open reading frame encoding a polypeptide containing 173 amino acids. The predicted amino acid sequence from this reading frame is identical to that determined previously by direct sequencing of the bovine αA polypeptide (28), confirming that the cDNA in pBLαA2-1 is derived from αA2-crystallin mRNA. The open reading frame is terminated by a stop codon found immediately following the carboxy-terminal serine residue. Two additional termination codons are also found 150 and 195 bp further downstream. Examination of the cDNA sequence of pBLαA2-1 reveals that multiple termination codons are encountered when the cDNA sequence is translated in the two alternative reading frames (not shown). Northern blot analysis of bovine lens mRNA using pBLαA2-1 as hybridization probe indicated that the αA2-crystallin mRNA transcript is approximately 1200 nucleotides in length (22). Therefore, we estimate that the 793 bp cDNA insert contained in pBLαA2-1 is approximately 400 bp short of the corresponding full length mRNA sequence. By analogy with the rat (11) and mouse (12) αA-crystallin

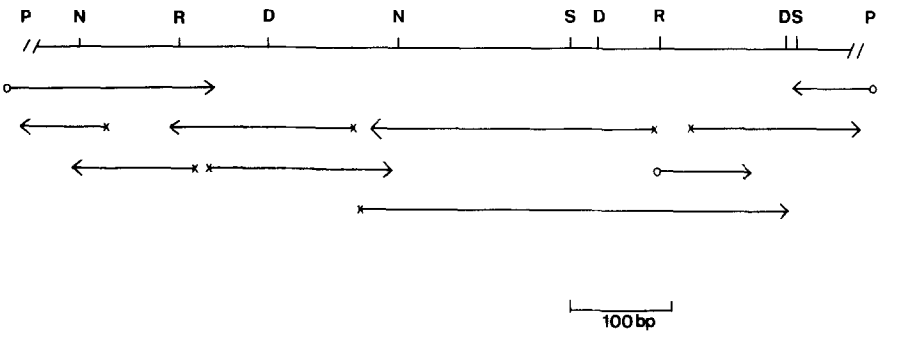

Figure 1. Partial restriction map and sequencing strategy for pBLαA2-1. Restriction sites shown are P=Pst I, N=Nco I, D=Dde I, R=Rsa, and S=Sma I. Arrows indicate location and orientation of sequencing reactions primed by synthetic oligonucleotides "*" or by universal primers "o" used on insert fragments cloned into M13. Oligo d(G)-d(C) homopolymer tails are indicated as "//".

```
          10          20          30          40          50          60
CCC CGG GTG CCC ACA GAG CCG CTG CCC ACC ATG GAT ATC GCC ATT CAG CAC CCC TGG TTC
                                    MET Asp Ile Ala Ile Gln His Pro Trp Phe
                                     1                                      10

          70          80          90          100         110         120
AAA CGC ACC CTG GGC CCC TTC TAC CCC AGC CGG CTG TTC GAC CAG TTC TTC GGC GAG GGC
Lys Arg Thr Leu Gly Pro Phe Tyr Pro Ser Arg Leu Phe Asp Gln Phe Phe Gly Glu Gly
                                    20                                      30

          130         140         150         160         170         180
CTC TTC GAG TAC GAC CTG CTG CCC TTC CTG TCC TCC ACC ATC AGC CCC TAC TAC CGC CAG
Leu Phe Glu Tyr Asp Leu Leu Pro Phe Leu Ser Ser Thr Ile Ser Pro Tyr Tyr Arg Gln
                                    40                                      50

          190         200         210         220         230         240
TCC CTC TTC CGC ACC GTG CTG GAC TCC GGC ATC TCT GAG GTC CGA TCC GAC CGG GAC AAG
Ser Leu Phe Arg Thr Val Leu Asp Ser Gly Ile Ser Glu Val Arg Ser Asp Arg Asp Lys
                                    60                                      70

          250         260         270         280         290         300
TTT GTC ATC TTC CTG GAT GTG AAG CAC TTC TCT CCC GAG GAC CTG ACG GTG AAG GTG CAG
Phe Val Ile Phe Leu Asp Val Lys His Phe Ser Pro Glu Asp Leu Thr Val Lys Val Gln
                                    80                                      90

          310         320         330         340         350         360
GAG GAC TTC GTG GAG ATC CAC GGC AAG CAC AAC GAG CGG CAG GAT GAC CAT GGC TAC ATC
Glu Asp Phe Val Glu Ile His Gly Lys His Asn Glu Arg Gln Asp Asp His Gly Tyr Ile
                                   100                                     110

          370         380         390         400         410         420
TCC CGC GAG TTC CAC CGC CGC TAC CGC CTG CCT TCC AAC GTG GAC CAG TCC GCA CTC TCC
Ser Arg Glu Phe His Arg Arg Tyr Arg Leu Pro Ser Asn Val Asp Gln Ser Ala Leu Ser
                                   120                                     130

          430         440         450         460         470         480
TGC TCC CTG TCC GCT GAT GGC ATG CTG ACC TTC TCT GGC CCC AAG ATC CCA TCT GGC GTG
Cys Ser Leu Ser Ala Asp Gly Met Leu Thr Phe Ser Gly Pro Lys Ile Pro Ser Gly Val
                                   140                                     150

          490         500         510         520         530         540
GAC GCC GGC CAC AGC GAG CGG GCC ATC CCC GTG TCC CGG GAG GAG AAG CCC AGC TCT GCG
Asp Ala Gly His Ser Glu Arg Ala Ile Pro Val Ser Arg Glu Glu Lys Pro Ser Ser Ala
                                   160                                     170

          550         560         570         580         590         600
CCC TCG TCC TAA GCT CGG CCT TGG CCT CGG CTG CCA CCC GCT GCG GCC CCC GTA CCC ATC
Pro Ser Ser TER

          610         620         630         640         650         660
CAT CTG GGG GAC CCT AGA AAG TGG GGC ATC CAT CTC CCT CTG CTT CCC CCC TTT CCA GTT

          670         680         690         700         710         720
CCT TTT CCT CTT CTT CGA GGG CTT GAG GGT TTG AGA GAG TAG CCG GGA GGC CCA GGG CCA
                                                TER

          730         740         750         760         770         780
GCC TGT GGT GCA AAG ACC TCA GAG TGA CCC GGG TCC CAA CAC CAG CCC CTC GGG GGA AGT
                                TER

          790
GAC CAC TGT CGG A
```

Figure 2. Nucleotide and predicted amino acid sequence for the cDNA insert contained in pBLαA2-1. The numbers above the nucleotide sequence refer to the position within the cDNA insert. Not shown are approximately 30 oligo d(G) and d(C) residues located at either end of the cloned cDNA. Translation of the cDNA sequence begins with the first ATG codon at nucleotide #31. This initiation codon is contained with a consensus translational initiation sequence (PuXXATGG) characteristic of many eucaryotic mRNA transcripts (30). Numbers below the amino acid sequence refer to the position within the polypeptide. Termination signals are indicated by TER. The binding site for the 30-mer hybridization probe is underlined.

Table 1. Nucleotide and amino acid sequence homologies of αA-crystallins from different species. The nucleotide and predicted amino acid sequences from pBLαA2-1 were aligned with the corresponding sequences from mouse (12,13), hamster (17), rat (11), and frog (20) αA-crystallin genes. Sequences corresponding to the coding region and 3'NTR of the hamster αA-crystallin were derived entirely from genomic DNA sequences (17). "*" Comparisons with rat αA-crystallin were limited to sequences encoding amino acids 53-173 as well as the entire 3'NTR (11). "**" Comparisons with frog αA-crystallin sequences were limited to sequences encoding amino acids 25-173 and the 130 bp 3'NTR (20).

SPECIES COMPARISON	NUCLEOTIDE SEQUENCE HOMOLOGY			AMINO ACID TOTAL
	TOTAL SEQUENCE	CODING SEQUENCE	3'NTR SEQUENCE	
Cow vs Mouse	76%	89%	53%	96%
Cow vs Hamster	73%	90%	41%	96%
Cow vs Rat*	74%	90%	48%	96%
Cow vs Frog**	67%	78%	29%	86%
Mouse vs Hamster	85%	94%	69%	100%
Mouse vs Rat*	87%	95%	82%	100%
Hamster vs Rat*	81%	93%	73%	100%

mRNA transcripts, the bovine αA-crystallin mRNA sequences that are not represented in pBLαA2-1 probably do not serve a coding function.

Sequence homologies between bovine αA-crystallin and the corresponding sequences in the rat (11), mouse (12,13), hamster (17), and frog (20) systems are compiled in Table 1. The extensive amino acid sequence homology between the bovine and rodent αA-crystallins (96%) as well as between the bovine and frog crystallins (86%) is indicative of the slow mutation rate for this protein. This evolutionary conservation is reflected in the nucleotide sequences for the coding region of these respective genes. The homology is 89-95% among the mammalian crystallin genes and 78% between the bovine and frog genes. Although these homologies are exceptionally high, they are not as high as the amino acid sequence homologies. Therefore, most of the point mutations that have accumulated in the coding sequences of the αA-crystallin genes are silent. Among the amino acid replacements, many are single base substitutions which result in

chemically conservative changes that probably would not significantly alter the structure and properties of the corresponding αA-crystallin polypeptide.

In contrast to that observed in the coding region of the αA-crystallin genes, comparatively less homology is found in the nontranslated region located downstream from the coding sequences (3'NTR), particularly between more distantly related species (Table 1). The bovine sequences from this region average about 47% homology when compared with the mouse, rat, and hamster sequences. When the rodent αA-crystallin genes are compared, the 3'NTRs are at least 69% identical. Thus, it appears that the nucleotide sequences contained in the mammalian 3'NTRs have diverged at a much greater rate than have the corresponding coding sequences. No significant homology in the 3'NTR could be identified between frog and either bovine or rodent sequences.

While the size of the 3' NTR has been conserved among the mammalian species studied, the frog αA-crystallin mRNA is unique in that its 3' NTR is only 130 nucleotides in length (20). Tomarev and coworkers have suggested that a splice site present in the frog αA-crystallin gene may have been lost in mammals, resulting in the unusually long 3' NTR characteristic of the mammalian αA-crystallin mRNA transcript. The high degree of sequence divergence observed in the 3'NTR of mammals would suggest that such a mutational event would be plausible. No specific role for the 3'NTR in mammalian systems has been established. However, its size conservation may have some bearing on the translational efficiency of the αA-crystallin transcript by virtue of secondary structural considerations. In addition, the relatively long 3'NTR may play some role in posttranscriptional regulation of αA mRNA by interacting with factors which influence mRNA processing, transport, or stabilization. Such control mechanisms are being investigated in other systems (29) and remain to be studied in the αA-crystallin gene system.

ACKNOWLEDGEMENTS

This work was supported by research grants from the National Eye Institute (R23 EY05639 to JMP; REH is supported by NEI Training Grant T32 EY07092). Additional support comes from a departmental core center grant (1P30 EY06360) and a grant from Research to Prevent Blindness, Inc. We thank Dr. Robert Church for oligonucleotide synthesis and many valuable discussions.

REFERENCES

1. Harding, J. J., and Crabbe, M. J. C. (1984) In The Eye (3rd ed.) (Ed. Davson, H.) pp. 207-492, Academic Press, New York.
2. Harding, J. J., and Dilley, K. J. (1976) Exp. Eye Res. 22, 1-73.
3. Voorter, C. E. M., Mulders, J. W. M., Bloemendal, H., and de Jong, W. W. (1986) Eur. J. Biochem. 160, 203-210.
4. Chiesa, R., Gawinowicz-Kolks, M. A., and Spector A. (1987) J. Biol. Chem. 262, 1438-1441.
5. Chiesa, R., Gawinowicz-Kolks, M. A., Kleinman, N. J., and Spector, A. (1987) Biochem. Biophys. Res. Commun. 144, 1340-1347.
6. de Jong, W. W., van Amelsvoort, J. M., van der Ouderaa, F. J., and Bloemendal H. (1973) Nature 246, 233-235.
7. Kramps, J. A., de Man, B. M., and de Jong, W. W. (1977) FEBS Letts. 74, 82-84.
8. de Jong, W. W., van der Ouderaa F. J., Versteeg, M., Groenewoud, G., van Amelsvoort, J. M., and Bloemendal H. (1975) Eur. J. Biochem. 53, 237-242.
9. van der Ouderaa, J., de Jong, W. W., Hilderink, A., and Bloemendal, H. (1974) Eur. J. Biochem. 49, 157-168.
10. Berns, A., Janssen, P., and Bloemendal H. (1974) Biochem. Biophys. Res. Commun. 59, 1157-1164.
11. Moormann, R. J. M., van der Velden, H. M. W., Dodemont, H. J., Andreoli, P. M., Bloemendal, H., and Schoenmakers, J. G. G. (1981) Nucl. Acids Res. 9, 4813-4822.
12. King, C. R., Shinohara, T., and Piatigorsky, J. (1982) Science 215, 985-987.
13. King, C. R., and Piatigorsky, J. (1983) Cell 32, 707-712.
14. King, C. R., and Piatigorsky, J. (1984) J. Biol. Chem. 259, 1822-1826.
15. Overbeek, P.A., Chepelinsky, A.B., Khillan, J.S., Piatigorsky, J., and Westphal, H. (1985) Proc. Natl. Acad. Sci. 82, 7815-7819.
16. Chepelinsky, A.B., Sommer, B., and Piatigorsky, J. (1987) Mol. Cell. Biol. 7, 1807-1814.
17. van den Heuvel, R., Hendriks, W., Quax, W., and Bloemendal, H. (1985) J. Mol. Biol. 185, 273-284.
18. McDevitt, D. S., Hawkins, J. W., Jaworski, C. J., and Piatigorsky, J. (1986) Exp. Eye Res. 43, 285-291.
19. Yasuda, K. and Okada, T. S. (1986) In Oxford Surveys on Eukaryotic Genes Vol. 3 (Ed. MacLean, N.) pp. 183-209.
20. Tomarev, S. I., Zinovieva, R. D., Dolgilevich, S. M., Krayev, A. S., Skryabin, K. G., and Gause, G. G., Jr. (1983) FEBS LETTS 162, 47-51.
21. Hay, R. E., Petrash, J. M., Church, R. L., and Woods, W. (1987) Invest. Ophthalmol. Vis. Sci. 28(3), 85a.
22. Petrash, J. M., Church, R. L., and Hay, R. E. (1987) Lens Res. (In press).
23. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 5463-5467.
24. Pfeffer, D., and Mierendorf, R. (1986) In Promega Notes, User Newsletter, Promega Biotec, Madison, WI 53711.
25. Barr, P. J., Thayer, R. M., Laybourn, P., Najarian, R. C., Seela, F., and Tolan, D. R. (1986) BioTechniques 4, 428-432.
26. Fristensky, B., Lis, J., and Wu, R. (1982) Nucl. Acids Res. 10, 6451-6463.
27. Wilbur, W. J., and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 726-730.
28. van der Ouderaa, F. J., de Jong, W. W., and Bloemendal, H. (1973) Eur. J. Biochem. 39, 207-222.
29. Yaffe, D., Nudel, V., Mayer, Y., and Neuman, S. (1985) Nucl. Acids Res. 13, 3723-3737.
30. Kozak, M. (1981) Nucl. Acids Res. 9, 5233-5252.