The nucleic acid and deduced protein sequence of cDNA clones for δ-crystallin of the chicken lens

John M. Nickerson and Joram Piatigorsky

Laboratory of Molecular and Developmental Biology, National Eye Institute, National Institutes of Health, Bethesda, MD 20205, USA

Received 10 June 1981

1. INTRODUCTION

The vertebrate lens is a transparent cellular tissue filled with structural proteins called α-, β-, γ- and δ-crystallins [1]. In contrast to the other crystallins, δ-crystallin is present only in birds and reptiles [4]. Native δ-crystallin is a tetrameric protein containing ≥2 extremely similar polypeptides with M_r 50 000 and 48 000 [3,4]. Interestingly, the ratio of synthesis of the δ-crystallin polypeptides depends upon the intralenticular concentration of ions [5] – a finding which may have significance in understanding the reduction in crystallin synthesis in osmotic cataracts [6]. δ-Crystallin is also particularly interesting from a developmental perspective since it is the first and principal crystallin synthesized in the embryonic chicken lens, and its rate of synthesis is modulated in the differentiating lens cells [7].

To study the structure and expression of δ-crystallin, we have constructed a cDNA clone, pδCr2, using lens mRNA from outbred chickens [8]. Since there was no amino acid sequence data available for δ-crystallin it was not possible to establish directly the identity of pδCr2 by nucleotide sequencing. Thus, pδCr2 was identified indirectly as a δ-crystallin clone by hybrid-mediated arrest of translation, abundance of mRNA protecting pδCr2 from S1 nuclease digestion, and R-loop analysis [8]. Considering that putative δ-crystallin cDNAs were used to identify the δ-crystallin genes and their mRNAs in [9–12], it is important to establish unequivocally the identity of these cDNAs. We have taken advantage of the report describing the amino acid compositions of several tryptic peptides of δ-crystallin [13] to identify directly δ-crystallin cDNA clones and to provide the first nucleic acid and protein sequence data for δ-crystallin. Rather than sequencing pδCr2 constructed earlier from outbred chickens [8], we have sequenced cDNAs derived from the inbred chicken from which we have cloned the putative δ-crystallin genes [9,10].

2. MATERIALS AND METHODS

cDNAs derived from cytoplasmic, polyadenylated RNAs of 15-day-old chicken embryos (Spafas gs^-, Norwich CT) were cloned in the bacterial plasmid pBR322 by the GC-tailing procedure in [8]. DNA fragments resulting from digestion of the cDNA clones with *Eco*RI, *Pst*I and *Sau*3A were subcloned in M13 suquencing vectors. *Pst*I fragments were cloned in M13 *mp2/Pst* [14] and the *Sau*3A and *Eco*RI fragments were cloned in M13 *mp7* [15]. One fragment, AB4, double-digested with *Hpa*II and *Sau*3A, was cloned in M13 *mp9*. M13 *mp2/Pst* was the generous gift of Drs D. Bentley and T.H. Rabbitts (MRC, Laboratory of Molecular Biology, Cambridge UK). M13 *mp9* was a kind gift of Drs J. Messing and J. Vieira, (Department of Plant Pathology, University of Minnesota, St. Paul MN). Single-stranded DNAs from recombinant bacteriophage were sequenced by the di-deoxy method [16–18]. Each fragment was sequenced at least twice.

3. RESULTS AND DISCUSSION

Two plasmids, pδCr118 and pδCr520, hybridizing strongly to the cDNA insert of pδCr2 [8] and to the putative δ-crystallin gene fragment pδCr4 [10],

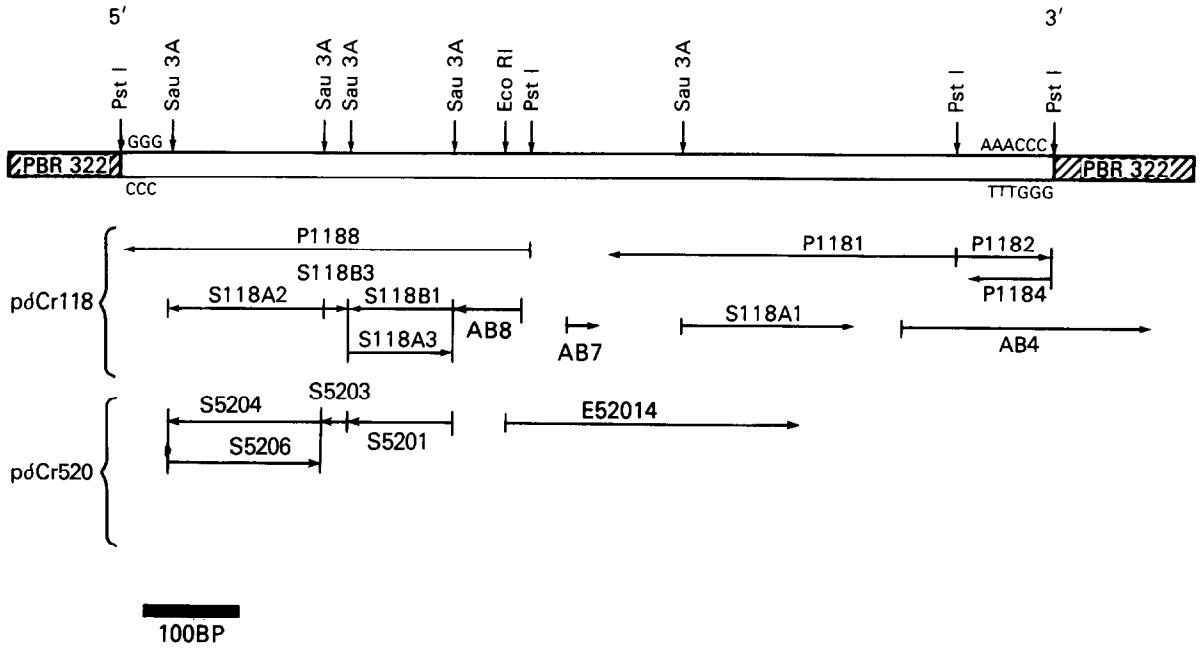

Fig.1. Sequencing strategy for pδCr118 and pδCr520. The subclones used in assembling the sequence are identified on the figure. All of the M13 subclones were independently isolated from the original inserts and not from larger M13 subclones (i.e., S118A2 was isolated directly from pδCr118 and not from P1188).

were sequenced using the strategy shown in fig.1. The common fragments in pδCr118 and pδCr520 contained identical sequences, suggesting that these 2 cDNA clones were derived from the same mRNA. The nucleotide sequence (fig.2) is 846 residues long, excluding poly(A) and poly(C); 469 nucleotides correspond to the coding region (see below), and 377 residues correspond to the 3′-untranslated region. A poly(A) sequence addition signal [19], AATAAA, starts at residue 817 and a stretch of A's begins at residue 847. The A's are followed by 18–20 C's and pBR322 sequence (not shown).

Only one reading frame is open in the 5′-half of the sequence. No stop codons appear in this reading frame until nucleotide 467; this is followed by 7 additional stop codons. The protein sequence encoded by the sequence given in fig.2 contains 155 amino acids (M_r 16 884), specifying the carboxy-terminal third of a δ-crystallin polypeptide.

This protein sequence provides compelling evidence that the cDNAs were derived from a δ-crystallin mRNA. The leucine content is 17%, which is unusually high for a structural protein and is characteristic of δ-crystallin [2,3]. Moreover, it does not contain cysteine, which is very low in or absent from δ-crystallin [2,3]. Most important, 5 of the tryptic peptides have identical amino acid compositions to the experimentally determined compositions of selected tryptic peptides of chicken δ-crystallin [13]. These tryptic peptides have been designated E, O, M, H and B in fig.2 to be consistent with the nomenclature used in [13]. The probability of obtaining these peptide amino acid compositions by random chance is negligible.

We used the computer program in [20] to assign secondary structural properties to regions of this δ-crystallin sequence. This is of interest since, in contrast to other crystallins, δ-crystallin is ~80% α-helical [21,22]. Five regions including 76 residues appear α-helical, 1 region composed of 4 residues appears as β-pleated sheet and 1 region of 3 residues appears as random coil (fig.3). No secondary structure was given to the remaining 71 amino acids since this program assigns secondary structure to a residue only when the probability of the most likely structure is >5-times the probability of the next most likely structure.

We searched for homologies of the partial se-

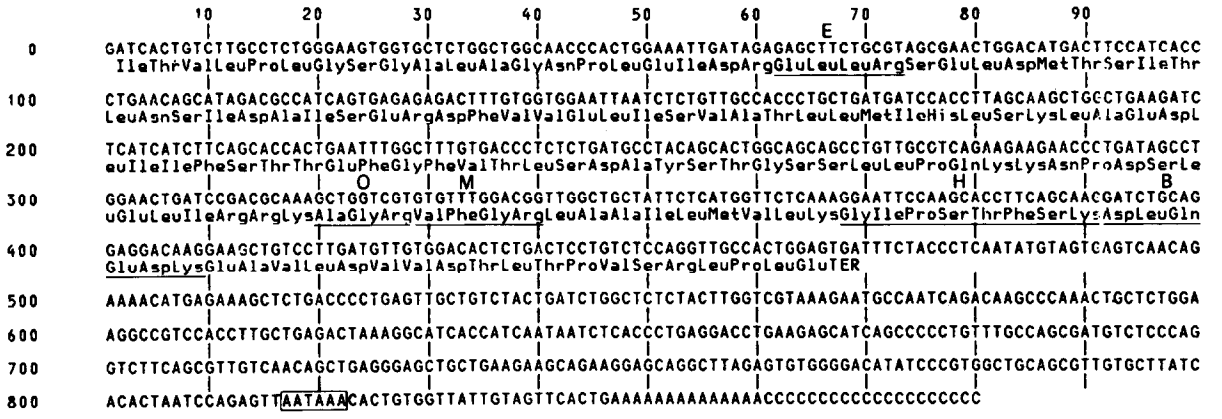

Fig.2. Partial nucleotide sequence and translation of δ-crystallin mRNA. The poly(A) addition site is boxed in the figure. Several of the predicted tryptic peptides from this sequence are underlined and identified by the letters E, O, M, H and B. These tryptic peptides have the same amino acid composition as authentic, similarly identified, tryptic peptides from δ-crystallin experimentally determined in [13].

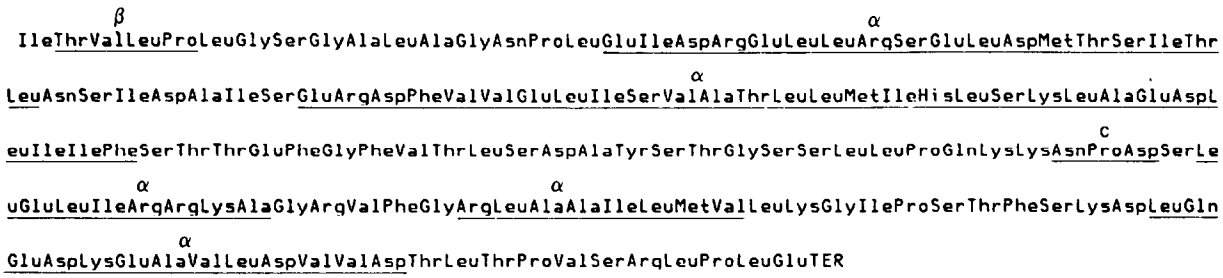

Fig.3. Possible regions of secondary structure in the δ-crystallin protein sequence. A secondary structure prediction program was used [20]. Regions α-helix (α) and β-pleated sheet (β) and random coil (C) are identified.

quence of δ-crystallin to α, β- and γ-crystallins of the bovine lens by a dot matrix computer program [23] and to 1667 polypeptides in the Dayhoff database [24]. No homologies were detected.

These data demonstrate directly that these cDNA clones encode δ-crystallin and provide the first nucleic acid and protein sequences for δ-crystallin. These sequences provide a basis for comparison with other δ-crystallin sequences as they become available to identify and characterize the δ-crystallin genes, mRNAs and proteins.

ACKNOWLEDGEMENTS

We thank Drs J.F. Hejtmancik, G. Inana, R.E. Jones, C.R. King, H. Ostrer, T. Shinohara and G. Vogeli for helpful discussions throughout the course of this work, Mrs B. Norman for help with the computer printouts and Mrs Florine Carpenter for excellent secretarial assistance.

REFERENCES

[1] Harding, J.J. and Dilley, K.J. (1976) Exp. Eye Res. 22, 1–73.

[2] De Jong, W.W. (1981) in: Molecular and Cellular Biology of the Eye Lens (Bloemendal, H. ed), pp. 221–278, Wiley New York.

[3] Piatigorsky, J., Zelenka, P. and Simpson, R.T. (1974) Exp. Eye Res. 18, 435–446.

[4] Reszelbach, R., Shinohara, T. and Piatigorsky, J. (1977) Exp. Eye Res. 25, 583–593.

[5] Shinohara, T. and Piatigorsky, J. (1977) Nature 270, 406–411.
[6] Piatigorsky, J. (1980) Curr. Topi. Eye Res. 3, 1–39.
[7] Piatigorsky, J. (1981) Differentiation 19, 134–153.
[8] Bhat, S.P. and Piatigorsky, J. (1979) Proc. Natl. Acad. Sci. USA 76, 3299–3303.
[9] Shinohara, T. and Piatigorsky, J. (1977) Nature 270, 406–411.
[10] Jones, R.E., Bhat, S.P., Sullivan, M. and Piatigorsky, J. (1980) Proc. Natl. Acad. Sci. USA 77, 5879–5883.
11] Yasuda, K., Kondoh, H., Okada, T.S., Nakajima, N. and Shimura, Y. (1982) Nucleic Acids Res. 10, 2879–2891.
[12] Bower, K.J., Errington, L.H., Wainwright, N.R., Sime, C., Morris, S. and Clayton, R.M. (1981) Biochem. J. 201, 339–344.
[13] De Jong, W.W., Stapel, S.O. and Zweers, A. (1981) Comp. Biochem. Physiol. 69B, 593–598.
[14] Bentley, D.L. and Rabbitts, T.H. (1980) Nature 288, 730–733.
[15] Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids. Res. 9, 309–321.
[16] Gardner, R.C., Howarth, A.J., Hahn, P., Brown-Leudi, M., Sheperd, R.J. and Messing, J. (1981) Nucleic Acids Res. 9, 2871–2888.
[17] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[18] Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) Nucleic Acids Res. 8, 1731–1743.
[19] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[20] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[21] Kuck, J.F.R., East, E. and Yu, N.T. (1976) Exp. Eye Res. 23, 9–14.
[22] Piatigorsky, J., Horwitz, J. and Simpson, R.T. (1977) Biochim. Biophys. Acta 490, 279–289.
[23] Maizel, J.V. and Lenk, R.P. (1981) Proc. Natl. Acad. Sci. USA 78, 7665–7669.
[24] Dayhoff, M.O., Hunt, L.T., Barker, W.C., Orcutt, D.C., Yeh, L.-S., Chen, H.R., George, D.G., Blomquist, M.C., Fredrickson, J.A. and Johnson, G.C. (1981) Protein Sequence Database, National Biomedical Research Foundation, Washington DC.