Vol. 186, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1992 Pages 190-198

cDNA SEQUENCE ANALYSIS OF CP94: RAT LENS FIBER CELL BEADED-FILAMENT STRUCTURAL PROTEIN SHOWS HOMOLOGY TO CYTOKERATINS

Shigeo Masaki* and Tomomasa Watanabe

Department of Biochemistry, Institute for Developmental Research, Aichi Prefecture Colony, Kasugai, Aichi 480-03, JAPAN

Received May 30, 1992

SUMMARY: To study the molecular structure of the gene responsible for a lens fiber cell beaded-filament structural protein of 94kDa (CP94), we isolated its specific cDNA from a rat lens cDNA library by use of anti-mouse CP94 antiserum. The expressed fusion protein kept the epitopes specific against anti-chick CP97 as well as anti-mouse CP94 antibody, and the size was estimated as 190-200kDa, indicating that the cDNA insert of the clone seemed to encode a polypeptide with 80-90kDa in appearance. Northern analysis indicated that CP94 mRNA is expressed only in the lens, and not in the brain, skin, heart, kidney, lung, and liver, and the size was estimated to 2.1-2.3kb. In a lens of inherited microphthalmic mouse, Elo, a trace amount of mRNA with the size closely similar to that of rat mRNA was observed. The entire compiled sequence (1,873bp) showed an open reading frame covering the sequence of 533 amino acids totalling 58,857Da. No sequence homologous to the entire CP94 was found among the entries of any nucleotide and amino acid sequence databases; but with respect to a limited amino acid sequence of N-side region of CP94, a significant homology with cytokeratins was found.

The differentiation of the vertebrate eye lens epithelial cell to a mature fiber cell is a distinctive morphological event that includes loss of organelles and cell elongation, and it is accepted as a useful experimental system for the study of the mechanism of gene expression during morphogenesis (1). The first feature of the event is biochemically characterized by a change in the composition of lens crystallins; and an increase in actin, decreased synthesis of vimentin, tubulin, and the accumulation of membrane intrinsic proteins (MIPs) are similarly observed (1). Also reported was the appearance of a unique cytoskeletal structure, the beaded-chain filament. It is seen only in the elongating lens fiber cells (2), and its structure in the chick mainly contains two lens-specific proteins, referred to as CP97 and CP49 (3, 4).

*To whom correspondence should be addressed. Fax: 0568-88-0829, 0568-92-1335.

0006-291X/92 $4.00

We recently reported the deficiency of a 94-kilodalton (kDa) protein, named CP94, in the non-crystallin fraction from the lens of the inherited microphthalmic mouse, Elo (5). The antibody raised against 94kDa protein reacted with the chick lens CP97 protein as well, and an antibody against CP97 (4) also reacted with the protein. From these, we concluded that the protein corresponds to the chick CP97, and named it 'mouse CP94' in keeping with the designation. Chick CP97 is believed as attachment sites for crystallins and that it is functionally important in lens fiber cell elongation and in keeping its conformation, because they are widely distributed throughout the lens fiber cells and also closely approach the plasma membrane (4). Recently FitzGerald and Casselman (6) showed that the corresponding protein with the same immunoreactivity is widely distributed in the lens of various animal species. As CP94 had not been previously described, we concluded that the protein belongs to the class of lens beaded-filament structural proteins, because antibody raised against mouse CP94 reacted with purified chick CP97, and vice versa (4, 5). Because the molecular structures and the regulation of the genes responsible for this class of the proteins have not been studied, we decided to isolate the specific cDNA encoding CP94 protein in this study.

MATERIALS AND METHODS

The preparation of rabbit antiserum raised against mouse lens fiber cell 94kDa protein (anti-CP94 antibody) and the procedures of SDS-PAGE and Western blot analysis were described previously (5). The rabbit anti-chick CP97 antibody was the gift of Dr. Maisel (4). The method of Huynh et al. (Protoblot® Immunoscreening System, Promega Corp., Madison, WI, ref. 7) was used to screen the lens cDNA expression library in λgt11 (Clontech Lab. Inc., Palo Alto, CA) prepared from young adult SD rat. DNA inserts from the positive clones were subcloned into a pUC18 vector, and the recombinants were further trimmed with various restriction enzymes, exonuclease III, and mung bean nuclease (TOYOBO Corp., Japan) and used for sequence determination of both complementary strands (Sequenase Ver.2.0, USB Corp., Cleveland, OH). DNA oligomers named CPPR1~3 (sequences are shown in Fig. 4) were synthesized by Japan Bioservice Corp. To obtain the 3'-end of CP94 cDNA (3'-RACE, ref. 8), 10μg of rat lens RNA was added to First-Strand cDNA Reaction Mix (containing 15 pmol of Not I-d$(T)_{18}$ primer, Pharmacia LKB Biotech. Inc.), and incubated for 2h at 37℃. The product separated from excess primers was concentrated, and added to the polymerase chain reaction (PCR) cocktail containing 100 pmol of Not I-d$(T)_{18}$ primer, 100 pmol of CPPR3 and 2.5 units of Amplitaq® DNA polymerase (Perkin-Elmer Cetus Corp., Norwalk, CT). DNA amplification was carried out by the program of 40 cycles of 94℃, 1min; 55℃, 2min; 72℃, 3min (8). The product was digested with Sac I and Eco RI, and cloned into the pUC18 vector. To obtain the 5'-end region of CP94 cDNA (5'-RACE), lens RNA was reverse transcribed as described above except for the substitution of 100 pmol of pd$(N)_6$ random primer for Not I-d$(T)_{18}$ primer. The product was concentrated and added to DNA tailing buffer (Bohehringer Mannheim GmbH, Germany). After incubation, the product was added to the PCR cocktail containing 100 pmol of each of Not I-d$(T)_{18}$ primer and CPPR2, and amplified by the program as descrived above. The PCR products were used as the templates for a second PCR reaction contained CPPR1 instead of CPPR2. The second PCR products of 200-

500bp were digested with Eco RV and Eco RI, and cloned into pUC18. RNA was extracted from lenses of SD rats (one day-old, 200 lenses), mice (129/SvJ and microphthalmic 129/SvJ-Elo; one-week-old, 40 lenses from each), and neonatal chicks (five lenses, from local poultry farm), and from various tissues of postnatal SD rats (each two grams) as follows: tissue in 4 M guanidium isothiocyanate was disrupted and centrifuged over CsTFA solution. The RNA precipitate was dissolved in, and an aliquot of it was fractionated on a 1% agarose gel containing 2.2 M formaldehyde, and transferred to Immobilon®-N membrane (0.45μm, Millipore Corp., Bedford, MA). Blots were hybridized to the random primed [^{32}P]-labeled probes (1×10^8 cpm/μg) for 18h. These were finally washed with three changes of 0.2× SSC containing 0.1% SDS at 50℃ (9, 10), and exposed to X-ray film with an screen for 1-4h at -70℃. A chick β-actin gene fragment was purchased from Oncor Inc., Gaithersburg, MD. The sequence data was handled and analyzed with GENETYX programs (Software Develop. Corp., Japan). Homology searches were done on all entries of the nucleotide sequence databases of EMBL-GDB (rel.28.0) and GenBank (rel.69.0), and of protein sequence databanks of SWISS-PROT (rel.19.0) and NBRF-PIR (rel.29.0).

RESULTS AND DISCUSSION

Our antiserum was raised against mouse CP94 (5), but a cDNA library used was made of rat lens because the antibody reacted with a rat lens protein of slightly larger molecular weight (95-96kDa) with the same sensitivity on Western blotting. By screening of the library (~1×10^6), forty phage clones expressing fusion proteins were obtained. All of these contained the same DNA fragment of various lengths, and the one with the longest insert (1.84kbps), named λp94.40, was selected and mainly used for the subsequent analysis. The fusion protein with β-galactosidase and rat CP94 was preliminarily characterized by Western blot analysis following SDS-PAGE (Fig. 1). It was only observed in the lane loaded with the cell infected with λp94.40 induced with IPTG and harvested just before spontaneous lysis. With anti-CP94 antibody used for detection, the apparent molecular weight of the fusion protein was estimated as 190-200kDa, indicating that the cDNA insert of the clone seemed to encode a polypeptide with 80-90kDa (assuming the β-galactosidase part of the fusion protein to be 110kDa) which corresponds to 84-94% of mature rat CP94 sequence. However the insert size was 1.84kbps, and the maximum coding capacity of it was calculated at 68kDa polypeptide. We can not conclusively explain this discrepancy at present, but this result suggests that the configuration of the CP94 molecule is far different from a globular shape. The anti-chick CP97 antibody also reacted with λp94.40 lysate, but bands with lower molecular weights were observed, probably representing degraded forms of the CP94 fusion protein.

Northern blot analysis indicated that the cloned insert λp94.40 hybridized to a mRNA of 2.1-2.3kb length in the lane loaded with RNA from rat lens (Fig. 2). No signal was detected in the lanes of RNAs from brain, skin, heart, kidney, lung, and liver, thus confirming that the expression of the

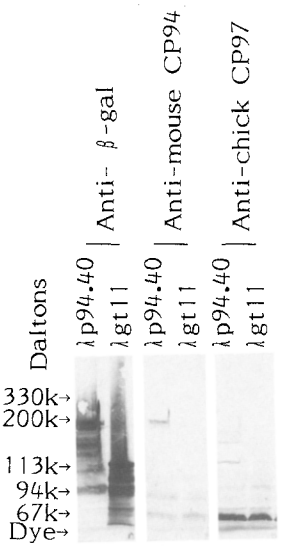

FIG. 1. **Western blotting analysis of the fused protein.** Three-ml cultures of Y1090 strain transfected with λp94.40 or λgt11 were induced to synthesize fusion protein with IPTG and harvested just before spontaneous lysis. The cell bodies were immediately mixed with the SDS-PAGE sample buffer, denatured by heat, and loaded into each lane of an SDS-PAGE gel (1mm, 7.5%). Proteins were then transferred to nitrocellulose and reacted with alkaline phosphatase-conjugated anti-rabbit IgG as the second antibody. Molecular weights of the proteins were estimated from the gel profiles simultaneously achieved; thyroglobulin (330kDa), ferritin half-unit (220kDa; not indicated), myosin heavy chain (200kDa), β-galactosidase (113kDa), phosphorylase b (94kDa), and bovine serum albumin (67kDa).

CP94 is controlled in a lens-specific manner. Next we estimated the homology between rat cDNA insert and the RNAs isolated from the lenses of chicks or mice (Fig. 3). No mRNA species were detected in the lane of chick lens RNA, indicating the relatively lower sequence homology between rat CP94 and chick CP97 mRNAs. While in the lane loaded with the lens RNA isolated from mouse (129/SvJ), an intense band with the size closely similar to that of rat was observed. The trace amount of CP94 transcript with the same size was expressed in the lens of the Elo; nevertheless, the CP94 protein was not detected in the lens of the mouse (5).

The entire sequence of rat CP94 cDNA is shown (Fig. 4). One open reading frame extending to the ^{1602}TAG stop codon was observed following the reading frame of the λgt11 LacZ gene. Since λp94.40 and other selected clones all lacked both 5'- and 3'-end regions of CP94 mRNA, we attempted to learn these sequences by RACE (8). By the 3'-RACE, several recombinants were obtained. Among 18 clones analyzed, the insert sequences were in common until ^{1866}G, but some sequence heterogeneity was observed after nucleotide 1867: CACCATC(A)$_n$ in

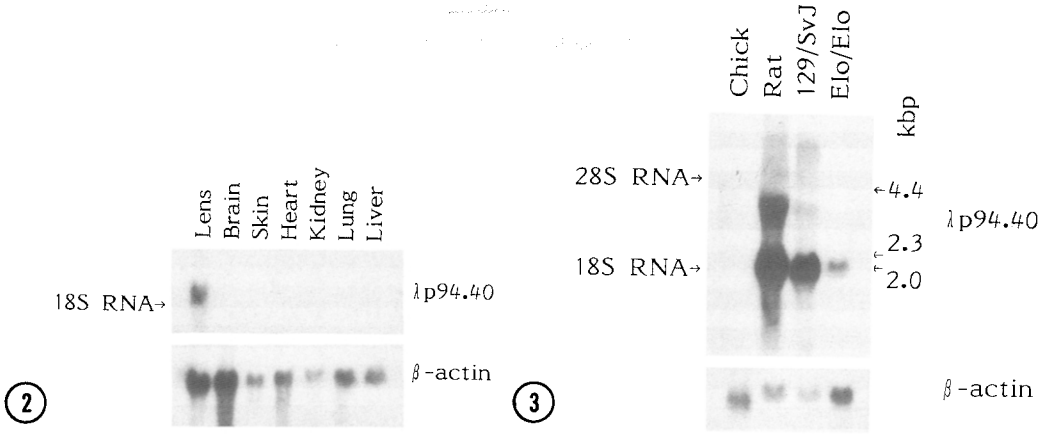

FIG. 2. Northern blot analysis of rat tissues. Ten μg of each RNA prepared from the indicated tissues was resolved on formaldehyde-containing agarose gel (1%), and transferred to a membrane. The blot was first reacted with a probe from λp94.40, then reprobed with a chick β-actin gene fragment (770bp).

FIG. 3. Northern blot analysis of the lenses. Ten μg of each RNA prepared from the indicated lenses was resolved and transferred to a membrane. The blot was first reacted with a λp94.40 probe and then reprobed with β-actin gene fragment. RNA sizes were determined by the estimation from the profiles of the λ/Hind III digest simultaneously electrophoresed. The autoradiogram shown was purposely overexposed in order to clearly demonstrate the faint band in the lane of Elo/Elo lens RNA.

12 clones, $C(A)_n$ in four clones, and $(A)_n$ in two clones. Therefore, a potential polyadenylation signal sequence was identified as ^{1843}ATTAAA, 24-31 nucleotides ahead of the $poly(A)_n$ tail. In an attempt to elongate the 5'-region sequence by 5'-RACE, all of our experiments failed: using oligomers of CPPR1 and CPPR2 for successive nested PCR, we obtained 10 clones. Four recombinants revealed information on the 5'-end portion shown in Fig. 4, but other clones only contained shorter inserts already observed in the insert DNA of λp94.40. Since the size of CP94 mRNA was estimated at 2.1-2.3kb by Northern blot (Fig. 3), and assuming the length of a poly(A) tail as approximately 100 nucleotides, we concluded that the cloned sequence with poly(A) tail is near 2.0kbp in length, and that 85-95% of the entire rat CP94 cDNA was cloned in the study. The amino acid sequence deduced from the nucleotide sequence is also shown (Fig. 4). One open reading frame, from ^{3}TTG to ^{1602}TAG stop codon, covered 533 amino acids reflecting a molecular weight of 58,857Da.

We compared the nucleotide and deduced amino acid sequence data shown above with all entries of nucleotide and protein sequence databanks (11). No sequence homologous to the entire CP94 sequence with the identities of over 48% and 18% in nucleotide and amino acid sequences, respectively, was found in

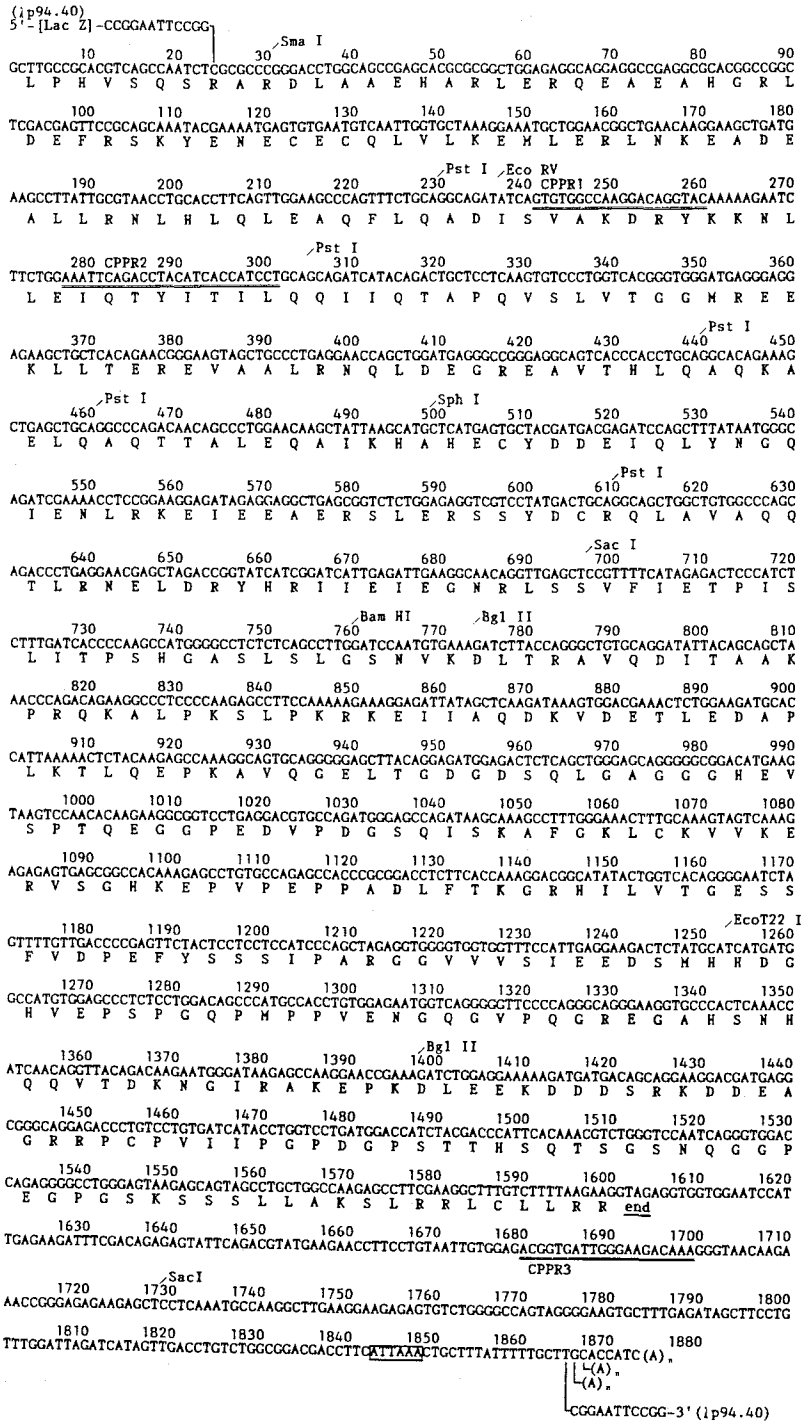

FIG. 4. The cDNA and amino acid sequences of rat CP94. The nucleotide sequence of the coding strand derived from cDNA clone λp94.40 was residues 23-1864. The sequences of 1-22 and 1865-1883 were identified from the recombinants obtained by the RACE. The heterogeneity observed near the 3'-end is shown by branch lines. The cutting sites of restriction enzymes and the positions of the synthetic oligomers (CPPR1, CPPR2, and CPPR3) are also indicated. The TAG stop codon is noted as 'end' and a putative polyadenylation signal is boxed. The sequence will appear in the DDBJ, GenBank, and EMBL-GDB nucleotide sequence databases with the accession no. D10866.

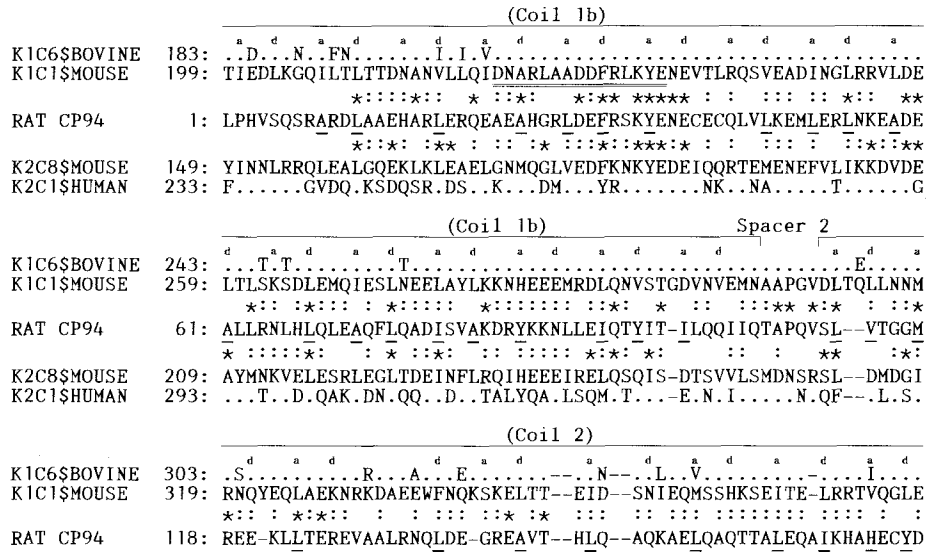

FIG. 5. Alignment and comparison of rat CP94 with cytokeratins. K1C6$BOVINE, bovine cytokeratin type I (class 6B); K1C1$MOUSE, mouse type I (59kDa); K2C8$MOUSE, mouse type II (class 8, EndoA); K2C1$HUMAN, human type II (67kDa). Amino acids are given as single letter symbols, and the residues of K1C6$BOVINE and K2C1$HUMAN identical to K1C1$MOUSE and K2C8$MOUSE, respectively, are shown by ".". The hepted repeat pattern of cytokeratins is noted by a (first position) and d (fourth position), and the structural domains of cytokeratin are indicated above. The residues identical to K1C1$MOUSE and/or K2C8$MOUSE in the RAT CP94 sequence are indicated by "*", and conservative replacements are shown by ":". Hydrophobic amino acids in first and fourth positions of the repeats in the RAT CP94 sequence are denoted by underlines, and a motif in K1C1$MOUSE is shown by doublet line. The cytokeratin data were obtained from SWISS-PROT.

the databases, revealing that the clone was previously uncharacterized. However, in the case of comparisons using the deduced amino acid sequence of the N-side region, some homology with the data on most intermediate filament cytokeratins was detected. The alignment and comparison of the predicted amino acid sequence encoded by the 5'-side part of CP94 cDNA with representative type I and II cytokeratins is shown (Fig. 5). This region of CP94 was calculated to be 19.4% and 21.7% homologous to mouse cytokeratins of type I 59kDa (K1C1$MOUSE in SWISS-PROT) and type II class 8 EndoA (K2C8$MOUSE)

species, respectively. The values were increased to 27.7% and 25.9% when the limited sequence from ^{12}Leu to ^{76}Gln was used for analysis. Using the sequence, the identity to various species of cytokeratin was calculated; bovine type I class 6B (27.7%; K1C6$BOVINE), mouse type I class 10 (29.2%; K1CJ$MOUSE), sheep type I class 8C-1 (30.2%; K1M1$SHEEP), human type I class 18 (28.1%; K1CR$HUMAN), frog type I (26.2%; K1C4$XENLA), human type II 67kDa (22.7%; K2C1$HUMAN), mouse type II 57kDa (24.5%; K2C3$MOUSE), bovine type II class 8 (25.9%), sheep type II (24.2%; K2M1$SHEEP), and frog type II cytokeratins (20.9%; K2C5$XENLA). Furthermore, the motif ^{222}DNARLAADDFRLKYE(NE) seen in the coil 1b region of mouse 59kDa cytokeratin (Fig. 5, K1C1$MOUSE) is highly conserved with limited variation among various type I cytokeratins (12), and nine of the 17 amino acids (52.9%) in the motif were conserved in the corresponding region of the CP94 sequence. Moreover the secondary structure of CP94 predicted by the method of Chou and Fasman (13) revealed the α-helical coiled structure with 7-residue repeat, as is generally observed in the cytokeratin structure (14): in total 51 first (a) and fourth (d) positions of 7-residue repeats of the limited sequence of CP94 from Leu^{12} to Glu^{192} predicted by the alignment with mouse type I 59kDa cytokeratin, 38 positions (74.5%) were occupied by hydrophobic amino acids (Ala, Gly, Ile, Leu, Met, Phe, and Val); e.g., of the Ala used 21 times in the sequence, 7 of them (33.3%) appeared in these specific positions. Similarly 15 of the Leu used 26 times (57.7%) and 7 of the Ile used 10 times (70.0%) occupied these positions. Several researchers (15, 16) previously reported that lens cells were never stained with anti-keratin antibodies, notwithstanding the fact that lens cells are commonly classified as 'epithelial'. The gene identity between CP94 and cytokeratins is low as a whole, and we feel it difficult to suggest a genetic correlation between them at present. However, it is conceivable that CP94 is expressed instead of cytokeratin during lens fiber cell differentiation. In the Elo mouse lens, the size of the CP94 mRNA transcript is apparently normal, but its amount is very small (Fig. 3), and its translation is thoroughly suppressed (5). It is still unclear whether a direct correlation exists there between the abnormal CP94 expression and the mutated locus, Elo, but this unusual gene expression may cause the defective elongation and the following degeneration of the fiber cells observed in the border area of cortex and nucleus in the Elo mouse lens (17, 18).

ACKNOWLEDGMENTS- We wish to thank Prof. H. Maisel, Wayne State Univ. Sch. Med., MI, for the generous gift of anti-chick CP97 antibody and Dr. Y. Yamada for his useful advice. This work was supported in part by Grant no. 90-483 from The Ishida Foundation, Japan.

REFERENCES

1. Harding, J.J. and Crabbe, M.J.C. (1985) In **The Eye,** Vol.1B (Davson, H. ed.) pp217-223, Academic Press, New York.
2. Ireland, M. and Maisel, H. (1983) **Exp. Eye Res.,** 36, 531-536.
3. Ireland, M. and Maisel, H. (1984) **Exp. Eye Res.,** 38, 637-645.
4. Ireland, M. and Maisel, H. (1989) **Lens Eye Toxicity Res.,** 6, 623-638.
5. Masaki, S., Tamai, K., Shoji, S., and Watanabe, T. (1991) **Biochem. Biophys. Res. Commun.,** 179, 1175-1180.
6. FitzGerald, P. G. and Casselman, J. (1991) **Curr. Eye Res.,** 10, 471-478.
7. Huynh, T. V., Young, R. A., and Davis, R.W. (1985) In **DNA Cloning: A practical Approach** Vol. 1 (Glover, D. M. ed.) pp49-78, IRL Press, Oxford.
8. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) **Proc. Natl. Acad. Sci. USA,** 85, 8998-9002.
9. Maniatis, T., Fritsh, E. F., and Sambrook, J. (1989) **Molecular Cloning: A Laboratory Manual** (2nd ed.) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
10. Masaki, S. and Watanabe, T. (1989) **Genomics,** 5, 259-263.
11. Lipman, D. J. and Pearson, W. R. (1985) **Science,** 227, 1435-1441.
12. Chandler, J. S., Calnek, D., and Quaroni, A. (1991) **J. Biol. Chem.,** 266, 11932-11938.
13. Chou, P. Y. and Fasman, G. D. (1978) **Adv. Enzymol.,** 47, 45-148.
14. Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L., and Steven, A. C. (1983) **Nature,** 302, 794-800.
15. Ramaekers, F.C.S., Osborn, M., Schmid, E., Weber, K., Bloemendal, H., and Franke, W. W. (1980) **Exp. Cell Res.,** 127, 309-327.
16. Sun, T. -T., Shih, C., and Green, H. (1979) **Proc. Natl. Acad. Sci. USA,** 76, 2813-2817.
17. Oda, S., Watanabe, T., and Kondo, K. (1980) **Jpn. J. Genetics,** 55, 71-75.
18. Masaki, S. and Watanabe, T. (1988) **Exp. Eye Res.,** 47, 723-735.