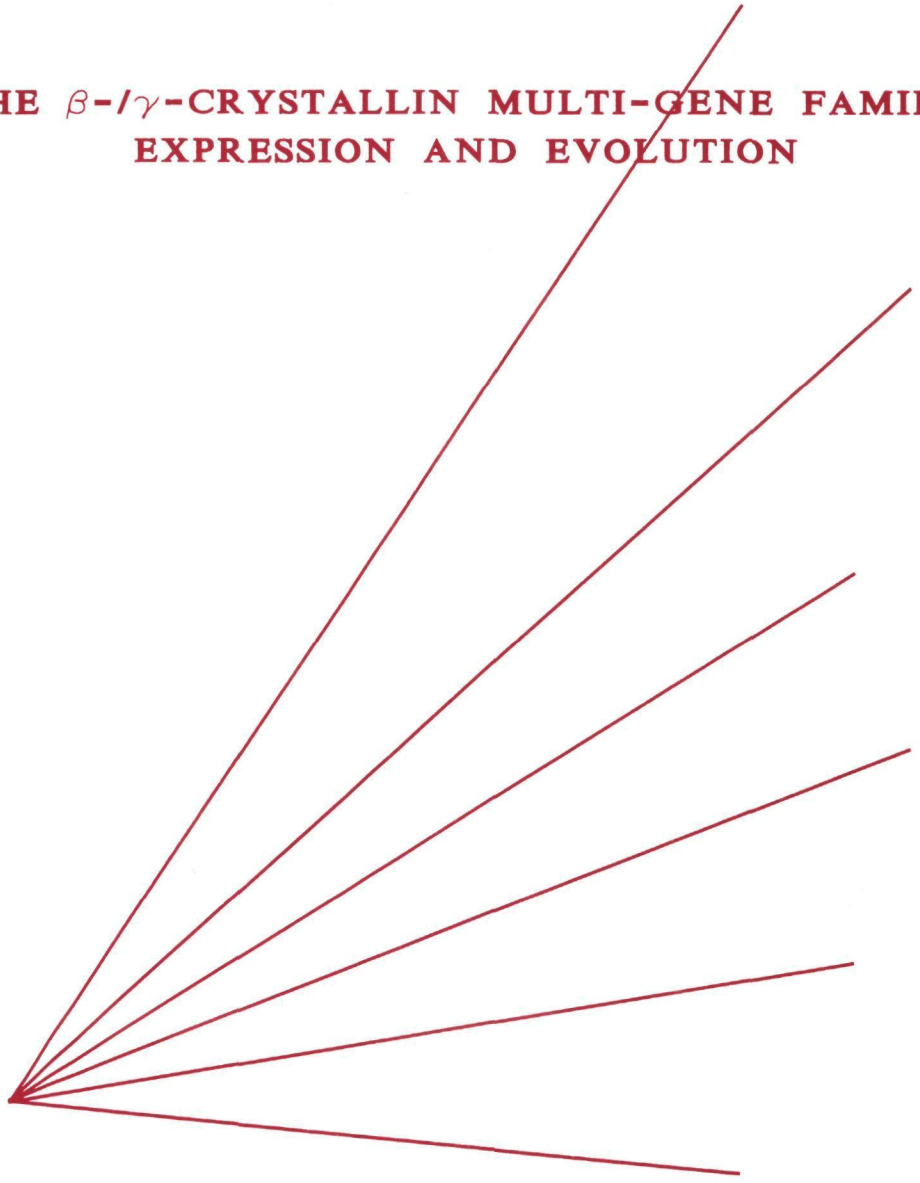
THE β-/γ-CRYSTALLIN MULTI-GENE FAMILY: EXPRESSION AND EVOLUTION



HENK AARTS

THE β-/γ-CRYSTALLIN MULTI-GENE FAMILY: EXPRESSION AND EVOLUTION

THE β-/γ-CRYSTALLIN MULTI-GENE FAMILY: EXPRESSION AND EVOLUTION

Een wetenschappelijke proeve op het gebied van de
WISKUNDE EN NATUURWETENSCHAPPEN

PROEFSCHRIFT
ter verkrijging van de graad van doctor
aan de Katholieke Universiteit te Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
vrijdag 26 mei 1989
des namiddags te 1.30 uur precies

door

HENRICUS JOZEF MARIA AARTS

geboren op 15 augustus 1957
te Budel

Promotor: Prof. Dr. J.G.G. Schoenmakers

Co-referent: Dr. N.H. Lubsen

ter nagedachtenis aan mijn vader

voor mijn moeder,

Anita, Marloes en Iris

Mijn dank gaat uit naar een ieder, zeker niet allen hier genoemd, die een bijdrage heeft geleverd aan de totstandkoming van dit proefschrift. Met name Gijsbert, Irene, Frank, Joost en Emiel, die in het kader van hun doctoraal stage, een belangrijke bijdrage hebben geleverd aan het onderzoek. Marja en Eldine dank ik voor hun uitstekende hulp bij de uitvoering van een groot aantal experimenten. Zeker wil ik ook alle (oud-) leden van de crystalline groep, alsmede alle medewerkers van de afdeling Moleculaire Biologie, danken voor de plezierige en leerzame samenwerking.

Nogmaals, hartelijk dank

Henk

The investigations described in this thesis were carried out under the auspices of the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

CONTENTS

INTRODUCTION

Lens development

Lens development is a complex process of closely coordinated morphological and molecular changes (for review see Piatigorsky, 1981). Its first morphological manifestation is the induction of the presumptive lens ectoderm cells to form the lens placode. The lens placode invaginates and eventually forms the lens vesicle. The cells located at the anterior site of this vesicle differentiate into cuboidal shaped lens epithelial cells while the posterior cells elongate to form the primary fiber cells. Subsequent growth of the lens occurs through differentiation of epithelial cells into secondary fiber cells at the equator. The newly formed fiber cells thus overlay the pre-existing ones and the oldest cells comprise the core region while the newly formed fiber cells comprise the cortex region of the lens (Papaconstantinou, 1967).

During the terminal differentiation of the primary and secondary fiber cells the cell organelles, which are a potential source of light scattering, are lost. With the loss of these organelles these fiber cells also presumably lose the capacity for de novo RNA and protein synthesis.

At the molecular level lens development and lens cell differentiation is characterized by the synthesis of a set of lens specific proteins. The most conspicuous of these are the few (10 to 20) protein species which form the bulk of the water soluble lens protein fraction. These proteins are collectively known as the crystallins.

The crystallins

The same three crystallin protein families, the α-, β- and γ-crystallins, are found in all vertebrate lenses examined (see Clayton, 1974; Harding and Dilley, 1976; Bloemendal, 1981, 1982; Piatigorsky, 1984). Some species contain additional crystallins, the so-called taxon-specific crystallins. It has been recently discovered that these taxon-specific crystallins are closely related or even identical to enzymes, which apparently have been recruited as structural proteins by the lens (Stapel et al., 1985; Carper et al., 1987; Huang et al., 1987; Wistow et al., 1987; Wistow and Piatigorsky, 1987; Hendriks et al.,

1988; Piatigorsky et al., 1988; Watanabe et al., 1988; Wistow and Piatigorsky, 1988). The α-, β-, and γ-crystallins are probably also descendants of housekeeping genes as the α-crystallins are related to the small heat-shock proteins (Ingolia and Craig, 1982; De Jong and Hendriks, 1986; Wistow and Piatigorsky, 1988) while the β- and γ-crystallins sequences show some similarity to the sequence of the Myxococcus xanthus S-protein (Wistow et al., 1985) and the myc-gene (Crabbe, 1985).

The structure and evolution of the crystallins and their genes have been extensively reviewed (Argos and Siezen, 1983; Schoenmakers et al., 1984; Piatigorsky, 1984; Bloemendal, 1985; Slingsby, 1985; De Jong and Hendriks, 1986; Lubsen et al., 1988; Slingsby et al., 1988) and will only be briefly summarized below.

The α-crystallins

The large aggregates (with an average molecular weight of approximately 800 kD) found in the soluble protein fraction of the lens are mainly composed of α-crystallins (Spector and Katz, 1965; Van der Ouderaa et al., 1973, 1974; Bloemendal, 1982). The primary α-crystallins are encoded by two single copy genes, the αA- and the αB-crystallin gene (Quax-Jeuken et al., 1985a; Van den Heuvel et al., 1985). Both genes possess the same intron/exon organization and the sequences of their protein products are approximately 57% similar which suggests that both genes are derived from a common progenitor. Within the rat, mouse and hamster lens approximately 10% of all αA-crystallin transcripts are alternatively spliced and contain 69 additional nucleotides yielding a 23 amino acid residues longer translation product, the so-called αA^{ins} (Cohen et al., 1978; De Jong et al., 1980; King and Piatigorsky, 1983; Van den Heuvel et al., 1985; Wistow, 1985). Additional diversity within the α-crystallins is generated by post-translational modification (Bloemendal, 1977, 1982; Zigler and Goosey, 1981; Ringers et al., 1982a; Harding and Crabbe, 1984; Thomson and Augusteyn, 1985; Spector et al., 1985; Voorter et al., 1986, 1987; Chiesa et al., 1987a,b).

The β-crystallins

The β-crystallins, with molecular weights between 23 and 33 kD, are present within the lens as oligomers varying from dimers to octamers (Berbers et al., 1982; Ramaekers et al., 1982). This rather heterogeneous class of proteins consists of at least six primary gene products divided into two subclasses, namely the 'acidic' (e.g. βA3/A1, βA2, and βA4) and 'basic' β-crystallins (e.g. βB1, βB2, and βB3). The β-crystallin amino acid sequences show some similarity to those of the γ-crystallins and prediction of the three dimensional structure of the β-crystallins by computer graphics techniques showed that these proteins have the same tertiary structure as that deduced for γ-crystallins from X-ray crystallographic studies (Driessen et al., 1980, 1981; Blundell et al., 1981, Wistow et al., 1983). The β- and γ-crystallins are built up out of four "Greek key" motifs organized into two globular domains. In addition the β-crystallins contain N- and C-terminal extensions which are probably involved in the oligomerization of these proteins (Wistow et al., 1981; Berbers et al., 1983).

The β-crystallin genes have six exons (Inana et al., 1983; Den Dunnen et al., 1986a; Hogg et al., 1986; this thesis). Each of the functional protein entities is encoded by a separate exon, except for the last motif and the C-terminal extension which are both encoded by the last exon. In the βA3/A1-crystallin gene the N-terminal extension is encoded by the first two exons (Quax-Jeuken et al., 1984; Hogg et al., 1986; Peterson and Piatigorsky, 1986), in the βB1-crystallin gene the first exons is non-coding (Den Dunnen et al., 1986a). Initial studies on the chromosomal location of the β-crystallin genes suggested that these genes might be dispersed: the rat βB1-crystallin gene was found not to be closely linked to another β-crystallin gene (Moormann et al., 1984), while the human βA3/A1- and βB2-crystallin genes were shown to be located on different chromosomes (Law et al., 1986; Sparkes et al., 1986; Hogg et al., 1987). However, in the study reported in this thesis close linkage was observed between the 'basic' βB2- and βB3-crystallin genes (Schoenmakers et al., 1986; Aarts et al., 1987) in man as well as rat, showing that some clustering of β-crystallin genes exists.

The γ-crystallins

The third class of crystallins, the γ-crystallins, consists of monomeric proteins with molecular weights of approximately 20 kD (Bjork, 1970; Croft, 1972). As mentioned above, the γ-crystallins have the same tertiary structure as the β-crystallins except that they lack N- and C-terminal extensions. The γ-crystallins are also divided into two subfamilies. The first family contains only a single member, namely βs, a protein traditionally classified as a monomeric β-crystallin but now known to be much closer related to the γ-crystallins (Quax-Jeuken et al., 1985b). βs (γS in the new crystallin nomenclature) is found in all vertebrate lenses examined thus far (Van Dam, 1966; Croft, 1973; Kabasawa et al., 1977; McDevitt and Croft, 1977; Bindels et al., 1981; Zigler et al., 1981, 1986; Harding and Crabbe, 1984). The second γ-crystallin protein family is formed by a homogeneous group of proteins with closely similar sequences. The exact number of these primary γ-crystallin gene products varies between species. For example, the rat lens contains six abundant γ-crystallins (γA through γF; Siezen et al., 1988) while only two (γC and γD) have been found in the human lens (Zigler et al., 1986; Siezen et al., 1987; see also below). This protein family is absent in bird and reptile lenses.

The shared four motif/two domain protein folding pattern of the β- and γ-crystallins suggests that these genes probably arose by successive duplication of a common ancestor 'one-motif' coding unit to yield the present four motif/two domain protein (for review see Lubsen et al., 1988). However, the structure of γ-crystallin genes differs from that outlined above for the β-crystallin genes: in the γ-crystallin genes two motifs are encoded by one exon. Hence, the γ-crystallin genes lack the intra-domain introns present in the β-crystallin genes (Moormann et al., 1983; Lok et al., 1984; Den Dunnen et al., 1985a, 1986b; Meakin et al., 1985, 1987). In contrast to the β-crystallin genes, all γ-crystallin genes are located on the same chromosome in rat or man (Willard et al., 1985; Den Dunnen et al., 1985b, 1987; Shiloh et al., 1986). Within the rat genome five γ-crystallin genes (γA through γE) are closely linked while the sixth gene (γF) is located at least 25 kb away (probably upstream; Quinlan et al., 1988) from this five gene cluster (Moormann et al., 1985). The genomic organization of the human γ-crystallin genes is the same as that of the rat genes, except that the human γ-crystallin gene family contains an additional γ-crystallin

gene fragment (R. Brakenhoff, unpubl. res.). Two of the human genes, γE and γF, contain an in frame stop codon and are thought to be inactive in the human lens (Meakin et al., 1985, 1987).

Developmental and species-specific regulation of crystallin synthesis

The optical properties of the vertebrate lenses vary widely: some species, such as man or birds, have soft flat lenses and can accommodate while other species (e.g. rodents or fish) have hard, almost spherical lenses and cannot accommodate (see De Jong, 1981; Maisel, 1985). One of the determinants of the optical properties of the lens is the amount of crystallins present. This can vary between 20% (in soft lenses) and 60% (in hard lenses) of the wet weight of the lens. The refractive index gradient across the lens necessary for focussing of the light on the retina is established by varying the crystallin concentration (Philipson, 1969; Sun et al., 1984). Hence the crystallin concentration is always higher in the core of the lens than in the cortex of the lens. The crystallin gradient is not a simple one as the contribution of each crystallin protein to the total amount of crystallin also varies across the lens. Apparently each crystallin species has been molded during evolution to suit a particular region within the lens. The γ-crystallins, for example, are thought to be particulary adapted to a high protein/low water environment and are found primarily in the lens core (Slingsby and Croft, 1973; Slingsby and Miller, 1983; Siezen et al., 1988). In contrast, one of the β-crystallins, βB2-crystallin, is usually found to be most prominent in the cortical region of the lens (Carper et al., 1985; McFall-Ngai et al., 1985). As no protein turn-over has been observed in the normal lens, the required crystallin gradient of the lens must be established by the differential synthesis of the members of the crystallin families. Additional diversity in the crystallin composition of the various regions of the lens is caused by post-translational modification of the primary crystallin products. The specific optical properties of a lens are thus determined by the developmental regulation of the level of expression of each of the crystallin genes and by the manner and extent of subsequent post-translational modification. These processes must be adapted to the optical needs of a specific species and may therefore vary between species. Such species-specific differences in crystallin synthesis and modification are illustrated below

through a comparison of the protein patterns found in the human lens (high water content) with that in the rat lens (low water content).

The differential distribution of the human crystallins

The water-soluble proteins of the human lens are usually divided into three size fractions, the high molecular weight (HMW), the intermediate molecular weight (IMW) and the low molecular weight (LMW) fraction.

HMW-fraction

The HMW fraction is mainly composed of the α-crystallins, αA and αB. In a one year old human lens the α-crystallins comprise about 34% of all water soluble proteins and are found in the outer as well as in the inner layers (Thomson and Augusteyn, 1985). The loss of nuclear α-crystallin and the disappearance of higher molecular weight aggregates which both occur approximately in the fourth decade of life are correlated with a decreased ability of the cytoplasm to redistribute after releasing or increasing the zonular stress on the capsule and contribute to the loss of accommodative ability (Roy and Spector, 1976; McFall- Ngai et al., 1985).

IMW-fraction

The intermediate molecular weight fraction constitutes approximately 35% of the water-soluble proteins of the human lens (Thomson and Augusteyn, 1985). This fraction contains the β-crystallins as well as a minor group of very basic proteins (VBP). These VBPs are probably enzymes (Jedziniak et al., 1985). The IMW proteins can be separated into three distinct peaks on a Sephadex G-200 column. The proteins in these peaks are mostly built up out of identical subunits and differences are only observed in the distribution of several minor polypeptides (Zigler et al., 1980a, b; Thomson and Augusteyn, 1985). As in other species, the human β-crystallins are very heterogeneous. Their molecular weight ranges from 43 to 26 kD and their isoelectric points from pH 5.5 to 7.5 (Jedziniak et al., 1978; Ringers et al., 1982a). One of the problems in comparing the human lens protein patterns with those of bovine or rat lenses is that the orthologous relationships between the human β-crystallins and those of other species have not yet been well established. The largest human β-crystallin, the 43 kD protein, is

unique to the human lens (Ringers et al., 1982a; Zigler et al., 1980a). On SDS polyacrylamide gel the 43 kD protein is sometimes seen, depending on the age of the lens, as a broad diffuse band and sometimes as a group of numerous distinct bands (Ringers et al., 1982a). The 43 kD polypeptide cross-reacts with bovine β-crystallin antisera and is probably formed by cross-linking of two smaller polypeptides (Spector et al., 1975; Zigler et al., 1980a). Even though the 43 kD polypeptide is only present in very low amounts it could turn out to be, because of its suggested association with the lens cell plasma membrane as an intrinsic protein, a functionally very important protein (Spector et al., 1979). The more prominent proteins within the IMW-fraction are found in the lower molecular weight regions. One of these proteins, the 29 kD β-crystallin, constitutes approximately 10% of the amount of lens crystallins at 8 months of fetal age. Its amount decreases steadily thereafter to 3.3% during the first postnatal year and to 0.3% after 5 years of age. Upon aging the 29 kD protein is lost from deep cortical and nuclear fibers (Alcala et al., 1988). The amount of another β-crystallin, the 27 kD protein, increases steadily from 3.5% at 8 months of fetal development to 7% at year 5. Later on it decreases to 1.2% (Alcala et al., 1988). The 27 kD polypeptide is present in all regions of the human lens but at increasing age its concentration decreases markedly in the deep cortical and nuclear fibers. The size, spatial and temporal distribution of this protein suggested that the 27 kD protein could be orthologous to bovine βB2-crystallin. However Western blot immunoassays of the human water-soluble fraction using a βB2-crystallin specific antibody revealed that cross-reactivity was only found with polypeptides with higher mobilities on SDS-polyacryl-amide gels (Alcala et al., 1988). The human βB2-crystallin proteins are synthesized in the cortical cells of the lens throughout life and a significant decrease in the amount of βB2-crystallin is only found in the older nuclear fibers (McFall-Ngai et al., 1985). Two βB2-crystallin fragments are present within the older human lens while only one is found in the young lens (McFall-Ngai et al., 1986). By the use of N- and C- terminal specific antisera Takemoto et al. (1987) showed that approximately 4 kD is cleaved from the N-terminal region of the 26 kD βB2-crystallin. The heat stability, an important property of the βB2-crystallin protein, is maintained after cleavage (McFall-Ngai et al., 1986). It was suggested by McFall-Ngai et al. (1986) that the βB2-crystallin protein serves as

a constitutive heat-shock protein and acts to stabilize the other more heat labile crystallins.

LMW proteins

Three distinct populations of monomeric crystallins with molecular weights of 19, 21 and 24 kD respectively are present in the human lens (Zigler et al., 1981, 1986; Slingsby and Miller, 1983; Siezen et al., 1987). The 24 kD crystallin population corresponds to the bovine βs-(γS)-crystallin (Zigler et al., 1986). The other two populations were characterized by Siezen et al. (1987). They showed that the 21 kD protein is encoded by the γC-crystallin gene and the 19 kD protein by the γD-crystallin gene. The γC- and γD-crystallins comprise together approximately 90% of all γ-crystallins in the 2 months old human lens. The γD-crystallin is orthologous to the rat γD-crystallin cryo-protein and also exhibits phase separation.

Coghlan and Augusteyn (1977) reported an age related decline of the γ-crystallins within the human lens. These findings were confirmed by Thomson and Augusteyn (1985) who demonstrated that the decline in γ-crystallins starts approximately around birth. This decline is not, as proposed by these authors, an abrupt process but happens gradually (Zigler et al., 1985; McFall-Ngai et al., 1985). The amount of βs-crystallin starts to increase simultaneously with the decrease in the amount of the γ-crystallins.

As found in the other crystallin classes, an increase in the heterogeneity of the LMW crystallin class is observed when the human lens becomes older (Zigler et al., 1985; Hoenders and Bloemendal, 1981; Ringers et al., 1982a, b). The three major proteins of the LMW protein class all undergo modifications and become more acidic (Zigler et al., 1986). Upon aging a fourth class of LMW proteins appears. These 10 kD proteins are probably breakdown products of the α- and/or γ-crystallins (Roy and Spector, 1978; Hoenders and Bloemendal, 1981; Takemoto et al., 1983).

The data summarized above suggest that post-translational modification of crystallins plays an important role in the human lens. In addition, marked changes in crystallin composition, even loss of crystallins, are seen upon aging. In the rat lens the crystallin composition also changes with age but a loss of crystallins has not been reported. Rat lens crystallins are also less subject to post-tran-

slational modification.

The differential distribution of the rat crystallins

A smooth gradient of lens proteins, increasing from 27% in the outer layers to 57% in the inner layers is present within the lens of a 210 days old rat (Siezen et al., 1988). Dehydration and differential synthesis of the lens proteins are two factors which are involved in the establishment of this gradient. The differential synthesis of the lens proteins is nicely reflected by the age related differences observed in the gel filtration patterns obtained after separating the water-soluble proteins of the rat lens by Sephadex G-200 column chromatography (Ramaekers et al., 1982; Uchiumi et al., 1983). The relative amount of α-crystallin remains constant during age, while the low molecular weight β-crystallin (βL) fraction increases and the high molecular weight β-crystallin (βH) and γ-fractions decrease with increasing age. The decrease in the relative amount of βH and the increase in that of βL is strongly correlated with a shift from the synthesis of βB1-crystallin protein to that of other β-crystallins (Ramaekers et al., 1982; Uchiumi et al., 1983; Carper et al., 1985). It was shown that βB1-crystallin is involved in the formation of βH (Siezen et al., 1986).

The age dependent variations within the LMW proteins have been studied by Siezen et al. 1988. They showed that the overall concentration of γ-crystallins decreases after birth while the amount of βs-crystallin increases. They further showed that the γ-crystallin pool present in the rat lens nucleus is mainly composed of the γE, γD and γF cryo-proteins. Only trace amounts of γE and γF crystallins could be detected in the outer layers of 60 days old rat lenses, whereas the other γ-crystallins (γA, γB, and γC) are present in considerable amounts. Hence, the cryo-proteins (γD, γE, γF) are mainly synthesized early during development while the synthesis of the non-cryo-proteins (γA, γB, γC) increases within the older lens. The γD-crystallin, however, remains the predominant γ-crystallin. No post-translational modification of the γ-crystallins is observed within the rat lens, this in contrast to the extensive post-translational changes found in these proteins in the human lens.

The most obvious species-specific difference between the human and rat crystallin patterns is found in the LMW fraction, the γ-crystal-

lins. The human lens contains fewer γ-crystallin species at a lower concentration than the rat lens. This difference is thought to be correlated with the difference in water content between the human and the rat lens (Summers et al., 1986). The synthetic patterns of the β-crystallins are more difficult to compare between rat and man, since, as already mentioned above, the orthologous relationships are not yet fully known at the protein level.

From the lens protein patterns it is very difficult to derive and compare the pattern of gene activity between various species as the rate of accumulation rather than the rate of synthesis of a protein is measured. By determining the developmental changes in the levels of the various crystallin mRNAs, a much more direct measure of the developmental regulation of the activity of various crystallin genes is obtained. Such studies are reported in this thesis.

A brief outline of this thesis

The lack of knowledge about the genomic organization of the (rat or human) β-crystallin genes led us to the study of the organization of the βB2- and βB3-crystallin genes in the human and rat genomes (Chapter 2). The developmental pattern of expression of these genes was established in the rat (Chapter 4). Using the newly elucidated human βB2- and βB3-, and the rat βA3/A1-crystallin sequences together with the sequences previously reported in the literature, we have attempted to provide a first approximation to the functional importance of the β-crystallins by comparing the extent of change in these proteins during evolution in various vertebrate lineages (Chapter 8).

The observation that the human lens contains four functional γ-crystallin genes but only two abundant γ-crystallin proteins, led us to determine the level of expression of the six human γ-crystallin genes in the human lens (Chapter 5). The lower γ-crystallin content of the human lens as compared to the rat lens further suggested that the γ-crystallins might be functionally less important in the human lens than in the rat lens. This hypothesis was tested by comparing the rate of evolution, a possible measure of the extent of functional constraint, of these proteins within the primate and rodent lineages (Chapter 7).

REFERENCES

Aarts, H.J.M., Den Dunnen, J.T., Lubsen, N.H. and Schoenmakers, J.G.G. (1987). Gene 59, 127-135.

Alcala, J., Katar, M., Rudner, G. and Maisel, H. (1988). Curr. Eye Res. 7, 353-359.

Argos, P. and Siezen, R.J. (1983). Eur. J. Biochem. 131, 143-148.

Berbers, G.A.M., Boerman, O.C., Bloemendal, H. and De Jong, W.W.(1982). Eur. J. Biochem. 128, 495-502.

Berbers, G.A.M., Brans, A.M.M., Hoekman, W.A., Slingsby, C., Bloemendal, H. and De Jong, W.W. (1983). Biochim. Biophys. Acta, 748, 213-219.

Bindels, J.G., Koppers, A. and Hoenders, H. (1981). Exp. Eye Res. 33, 333-343.

Bjork, I. (1970). Exp. Eye Res. 9, 152-157.

Bloemendal, H. (1977). Science 197, 127-138.

Bloemendal, H. (1981). In: Molecular and cellular biology of the eye lens (ed. H. Bloemendal). John Wiley, New York, pp 1-47.

Bloemendal, H. (1982). CRC Critical Reviews in Biochemistry 12, 1-39.

Bloemendal, H. (1985). Exp. Eye Res. 41, 429-448.

Blundell, T.L., Lindley, P.F., Miller, L., Moss, D.S., Slingsby, C., Tickle, I.J., Turnell, W.G. and Wistow, G.J. (1981). Nature 289, 771-777. Carper, D., Russell, P., Shinohara, T. and Kinoshita, J.H. (1985). Exp. Eye Res. 40, 83-94.

Carper, D., Nishimura, C., Shinohara, T., Dietzchold, B., Wistow, G., Craft, C., Kador, P. and Kinoshita, J.H. (1987). FEBS Lett. 220, 209-313.

Chiesa, R., Gawinowicz-Kolks, M.A., Kleiman, N.J., and Spector, A. (1987). Biochem. Biophys. Res. Comm. 144, 1340-1347.

Chiesa, R., Gawinowicz-Kolks, M.A., and Spector, A. (1987). J. Biol. Chem. 262, 1438-1441.

Clayton, R.M. (1974). In: The Eye (eds. H. Davson and L.T. Graham). Academic Press, New York, pp 221-278.

Coghlan, S. and Augusteyn, R. (1977). Exp. Eye Res. 25, 603-611.

Cohen, L.H., Westerhuis, L.W., De Jong, W.W. and Bloemendal, H. (1978). Eur. J. Biochem. 89, 259-266.

Crabbe, M.J.C. (1985). FEBS Lett. 181, 157-159.

Croft, L.R. (1972). Biochem. J. 128, 961-970.

Croft, L.R. (1973). Ciba Foundation Symposium, vol. 19. Elsevier, Amsterdam, pp 207-224.

De Jong, W.W., Cohen, L.H., Leunissen, J.A.M. and Zweers, A. (1980). Biochem. Biophys. Res. Comm. 96, 648-655.

De Jong, W.W. (1981). In: Molecular and cellular biology of the eye lens (ed. H. Bloemendal). John Wiley, New York, pp 221-278.

De Jong, W.W. and Hendriks, W. (1986). J. Mol. Evol. 24, 121-129.

Den Dunnen, J.T., Moormann, R.J.M., Cremers, F.P.M. and Schoenmakers, J.G.G. (1985a). Gene 38, 197-204.

Den Dunnen, J.T., Jongbloed, R.J.E., Geurts van Kessel, A.H.M. and Schoenmakers, J.G.G. (1985b). Hum. Genet. 70, 217- 221.

Den Dunnen, J.T., Moormann, R.J.M., Lubsen, N.H. and Schoenmakers, J.G.G. (1986a). Proc. Natl. Acad. Sci. USA 83, 2855-2859.

Den Dunnen, J.T., Lubsen, N.H. and Schoenmakers, J.G.G. (1986b). J. Molec. Biol. 189, 37-46.

Den Dunnen, J.T., Szpirer, J., Levan, G., Islam, Q. and Schoenmakers, J.G.G. (1987). Exp. Eye Res. 45, 747-750.

Driessen, H.P.C., Herbrink, P., Bloemendal, H. and De Jong, W.W. (1980). Exp. Eye Res. 31, 243-246.

Driessen, H.P.C., Herbrink, P., Bloemendal, H. and De Jong, W.W. (1981). Eur. J. Biochem. 121, 83-91.
Harding, J.J. and Dilley, K.J. (1976). Exp. Eye Res. 22, 1-74.
Harding, J.J. and Crabbe, M.J.C. (1984) In: The Eye (eds. H. Davson and L.T. Graham). Academic Press. pp 207-492.
Hendriks, W., Mulders, J.W.M., Bibby, M.A., Slingsby, C., Bloemendal, H. and De Jong, W.W. (1988). Proc. Natl. Acad. Sci. USA 85, 7114-711.
Hoenders, H.J. and Bloemendal, H. (1981). In: Molecular and cellular biology of the eye lens (ed H. Bloemendal). John Wiley, New York, pp 279-326.
Hogg, D., Tsui, L.-C., Gorin, M. and Breitman, M.L. (1986). J. Biol. Chem. 261, 12420-12427.
Hogg, D., Gorin, M.B., Heinzmann, C., Zollman, S., Mohandas, T., Klisak, I., Sparkes, R.S., Breitman, M., Tsui, L.-C. and Horwitz, J. (1987). Curr. Eye Res. 6, 1335-1342.
Huang, Q.C., Russell, P., Stone, S.H. and Zigler J.R., Jr. (1987). Curr. Eye Res. 6, 725-732.
Inana, G., Piatigorsky, J., Norman, B., Slingsby, C. and Blundell, T. (1983). Nature 302, 310-315.
Ingolia, T.D. and Craig, E.A. (1982). Proc. Natl. Acad. Sci. USA 79, 2360-2364.
Jedziniak, J.A., Baram, H. and Chylack, L.T., Jr. (1978). Exp. Eye Res. 26, 377-388.
Jedziniak, J., Arredondo, L., Meys, M. and Miller, J. (1985). Invest. Ophth. 26, 50.
Kabasawa, I., Tsunematsu, Y., Barber, G.W. and Kinoshita, J.H. (1977). Exp. Eye Res. 24, 437-448.
King, C.R. and Piatigorsky, J. (1983). Cell 32,707-712.
Law, M.L., Cai, G.-Y., Hartz, J., Kao, F.-T., Hogg, D., Breitman, M.L. and Tsui, L.-C. (1986). Cytogenet. Cell Genet. 42, 202-207.
Lok, S., Tsui, L.-C., Shinohara, T., Piatigorsky, J., Gold, R. and Breitman, M.L. (1984). Nucl. Acids Res. 12, 4517- 4529.
Lubsen, N.H., Aarts, H.J.M. and Schoenmakers, J.G.G. (1988). Prog. Biophys. Molec. Biol. 51, 47-76.
Maisel, H. (1985). The Ocular Lens. Marcel Dekker, New York.
McDevitt, D.S. and Croft, L.R. (1977). Exp. Eye Res. 25, 473-481.
McFall-Ngai, M.J., Ding, L.-L., Takemoto, L.J. and Horwitz, J. (1985). Exp. Eye Res. 41, 745-758.
McFall-Ngai, M., Horwitz, J., Ding, L.-L. and Lacey, L. (1986). Curr. Eye Res. 5, 387-394.
Meakin, S.O., Breitman, M.L. and Tsui, L.-C. (1985). Mol. Cell. Biol. 5, 1408-1414.
Meakin, S.O., Du, R.P., Tsui, L.-C. and Breitman, M.L. (1987). Mol. Cell. Biol. 7, 2671-2679.
Moormann, R.J.M., Den Dunnen, J.T., Mulleners, L., Andreoli, P., Bloemendal, H. and Schoenmakers, J.G.G. (1983). J. Molec. Biol. 171, 353-368.
Moormann, R.J.M., Jongbloed, R. and Schoenmakers, J.G.G. (1984). Gene 29, 1-9.
Moormann, R.J.M., Den Dunnen, J.T., Heuyerjans, J., Jongbloed, R.J.E., Van Leen, R.W., Lubsen, N.H. and Schoenmakers, J.G.G. (1985). J. Molec. Biol. 182, 419-430.
Papaconstantinou, J. (1967). Science 156, 338-346.
Peterson, C.A. and Piatigorsky, J. (1986). Gene 45, 139-147.
Philipson, B. (1969). Invest. Ophth. 8, 258-270.

Piatigorsky, J. (1981). Differentiation 19, 134-153.
Piatigorsky, J. (1984). Cell 38, 620-621.
Piatigorsky, J., O'Brien, W.E., Norman, B.L., Kalumuck, K., Wistow, G.J., Borras, T., Nickerson, T.M. and Wawrousek, E.F. (1988). Proc. Natl. Acad. Sci. USA 85, 3479- 3483.
Quax-Jeuken, Y., Janssen, G., Quax, W., Van den Heuvel, R. and Bloemendal, H. (1984). J. Molec. Biol. 180, 457-472.
Quax-Jeuken, Y., Quax, W., Van Rens, G., Meera Khan, P. and Bloemendal, H. (1985a). Proc. Natl. Acad. Sci. USA 82, 5819-5823.
Quax-Jeuken, Y., Driessen, H., Leunissen, J., Quax, W., De Jong, W. and Bloemendal, H. (1985b). EMBO J. 4, 2597-2602.
Quinlan, P., Oda, S., Breitman, M.L. and Tsui, L.-C. (1988). Genes Dev., in press.
Ramaekers, F., Dodemont, H., Vorstenbosch, P. and Bloemendal, H. (1982). Eur. J. Biochem. 128, 503-508.
Ringers, P.J., Hoenders, H.J. and Bloemendal, H. (1982a). Exp. Eye Res. 34, 815-823.
Ringers, P.J., Hoenders, H.J. and Bloemendal, H. (1982b). Exp. Eye Res. 34, 204-211.
Roy, D. and Spector, A. (1976). Proc. Natl. Acad. Sci. USA 73, 3484-3487.
Roy, D. and Spector, A. (1978). Exp. Eye Res. 26, 445-452.
Schoenmakers, J.G.G., Den Dunnen, J.T., Moormann, R.J.M., Jongbloed, R., Van Leen, R.W. and Lubsen, N.H. (1984). In: Human Cataract Formation. Pitman, London. pp 208-218.
Schoenmakers, J.G.G., Aarts, H.J.M., Van Leen, R.J., Den Dunnen, J.T. and Lubsen, N.H. (1986). In: Proc. Int. Soc. Eye Res. (ed. S. Iwata), vol IV, Nagoya, p 44.
Shiloh, Y., Donlon, T., Bruns, G., Breitman, M.L. and Tsui, L.-C. (1986). Hum. Genet. 73, 17-19.
Siezen, R.J., Anello, R.D. and Thomson, J.A. (1986). Exp. Eye Res. 43, 293-303.
Siezen, R.J., Thomson, J.A., Kaplan, E.D. and Benedek, G.B. (1987). Proc. Natl. Acad. Sci. USA 84, 6088-6092.
Siezen, R.J., Wu, E., Kaplan, E.D., Thomson, J.A. and Benedek, G.B. (1988). J. Molec. Biol. 199, 475-490.
Slingsby, C. and Croft, L.R. (1973). Exp. Eye Res. 17, 369-376.
Slingsby, C. and Miller, L.R. (1983). Exp. Eye Res. 37, 517-530.
Slingsby, C. (1985). TIBS 10, 281-284.
Slingsby, C., Driessen, H.P.C., Mahadevan, D., Bax, B. and Blundell, T.L. (1988). Exp. Eye Res. 46, 375-403.
Sparkes, R.S., Mohandes, T., Heinzmann C., Gorin, M.B., Zollman, S. and Horwitz, J. (1986). Hum. Genet. 74, 133-136.
Spector, A. and Katz, E. (1965). J. Biol. Chem. 240, 1979-1985.
Spector, A., Roy, D. and Stauffer, J. (1975). Exp. Eye Res. 21, 9-24.
Spector, A., Garner, M.H., Garner, W.H., Roy, D., Farnsworth, P. and Shyne, S. (1979). Science 204, 1323-1326.
Spector, A., Chiesa, R,. Stredy, J., and Garner, W. (1985). Proc. Natl. Acad. Sci. USA 82, 4712-4716.
Stapel, S.O., Zweers, A., Dodemont, H.J., Kan, J.H. and De Jong, W.W. (1985). Eur. J. Biochem. 147, 129-136.
Summers, L. J., Slingsby, C., Blundell, T.L., Den Dunnen, J.T., Moormann, R.J.M. and Schoenmakers, J.G.G. (1986). Exp. Eye Res. 43, 77-92.
Sun, S.-T., Tanaka, T., Nishio, I., Peetermans, J., Maizel, J.V., Jr. and Piatigorsky, J. (1984). Proc. Natl. Acad. Sci. USA 81,

785-787.
Takemoto, L., Hansen, J. and Horwitz, J. (1983). Curr. Eye Res. 2, 443-449.
Takemoto, L., Takemoto, D., Brown, G., Takahana, M., Smith, J. and Horwitz, J. (1987). Exp. Eye Res. 45, 385-392.
Thomson, J.A. and Augusteyn, R.C. (1985). Exp. Eye Res. 40, 393-410.
Uchiumi, T., Kimura, S. and Ogata, K. (1983). Exp. Eye Res. 36, 125-134.
Van Dam, A.F.M. (1966). Exp. Eye Res. 5, 255-266.
Van den Heuvel, R., Hendriks, W., Quax, W. and Bloemendal, H. (1985). J. Molec. Biol. 185, 273-284.
Van der Ouderaa, F.J., De Jong, W.W. and Bloemendal, H. (1973). Eur. J. Biochem. 39, 207-222.
Van der Ouderaa, F.J., De Jong, W.W., Helderink, A. and Bloemendal, H. (1974). Eur. J. Biochem. 49, 157-168.
Voorter, C.E., Mulder, J.W., Bloemendal, H. and De Jong, W.W. (1986). Eur. J. Biochem. 160, 203-210.
Voorter, C.E., Roersma, E.S., Bloemendal, H. and De Jong, W.W. (1987). FEBS Lett. 221, 249-252.
Watanabe, K., Fujii, Y., Nakayama, K., Ohkubo, H., Kuramitzu, S., Kagamiyama, H., Nakanishi, S. and Hayaishi, O. (1988). Proc. Natl. Acad. Sci. USA 85, 11-15.
Willard, H.F., Meakin, S.O., Tsui, L.-C. and Breitman, M.L. (1985). Somat. Cell. Molec. Genet. 11, 511-516.
Wistow, G., Slingsby, C., Blundell, T., Driesen, H., De Jong, W.W. and Bloemendal, H. (1981). FEBS Lett. 133, 9-16.
Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. and Blundell, T. (1983). J. Molec. Biol. 170, 175-202.
Wistow, G. (1985). FEBS Lett. 181, 1-6.
Wistow, G., Summers, L. and Blundell, T. (1985). Nature 316, 771-773.
Wistow, G. and Piatigorsky, J. (1987). Science 236, 1554-1556.
Wistow, G., Mulders, J.W.M. and De Jong, W.W. (1987). Nature 326, 622-624.
Wistow, G. and Piatigorsky, J. (1988). Ann. Rev. Biochem. 57, 479-504.
Zigler, J.S., Jr., Horwitz, J. and Kinoshita, J.H. (1980a). Exp. Eye Res. 31, 41-55.
Zigler, J.S., Jr., Carper, D., Russell, P. and Kinoshita, J.H. (1980b). Exp. Eye Res. 31, 398-397.
Zigler, J.S., Jr., Horwitz, J. and Kinoshita, J.H. (1981). Exp. Eye Res. 32, 21-30.
Zigler, J.S., Jr. and Goosey, J. (1981). TIBS 6, 133-136.
Zigler, J.S., Jr., Russell, P., Takemoto, L., Swab, S., Hansen, J., Horwitz, J. and Kinoshita, J. (1985). Invest. Opth. 26, 525-531.
Zigler, J.S., Jr., Russell, P., Horwitz, J., Reddy, V.N. and Kinoshita, J.H. (1986). Curr. Eye Res. 5, 395.

CHAPTER 2

LINKAGE BETWEEN THE ßB2- AND ßB3 CRYSTALLIN GENES IN MAN AND RAT: A REMNANT OF AN ANCIENT ß-CRYSTALLIN GENE CLUSTER

Henk J.M. Aarts, Johan T. Den Dunnen, Nicolette H. Lubsen

and John G.G. Schoenmakers

Gene (1987) 59, 127–135

LINKAGE BETWEEN THE βB2 AND βB3 CRYSTALLIN GENES IN MAN AND RAT: A REMNANT OF AN ANCIENT β-CRYSTALLIN GENE CLUSTER

SUMMARY

Human and rat genomic clones containing βB2- and/or βB3-crystallin sequences have been isolated and characterized. Both in the human and the rat genome the single-copy βB3-crystallin gene is linked to a βB2-crystallin gene. In both species the linked genes, separated by 20 kb in the human and 11 kb in the rat genome, are oriented head to tail with respect to transcription. A single copy of the βB2-crystallin gene is present in the rat genome, in the human genome two copies of this gene are found. The second human copy could as yet not be linked to the βB2/βB3-crystallin gene cluster.

INTRODUCTION

About 90% of the water-soluble protein of the vertebrate eye lens consists of structural proteins, the so-called crystallins. In mammals, these proteins are divided into three immunologically distinct families of closely related polypeptides, the α-, β- and γ-crystallins (Piatigors-ky, 1981; Bloemendal, 1982). The α-crystallin oligomer is built up out of two primary gene products, αA and αB, both of which are encoded by a single gene (Quax-Jeuken et al., 1985a; Van den Heuvel et al., 1985). The β- and γ-crystallins are a mixture of about (depending on the species) six primary gene products (Piatigorsky, 1984; Driessen et al., 1981; Berbers et al., 1984; Shinohara et al., 1982; Ramaekers et al., 1982; Hejtmancik et al., 1983; Berbers et al., 1982; Slingsby and Croft, 1978) and are encoded by gene families. It has been shown that the β- and γ-crystallins are related and derived from a common ancestral sequence (Den Dunnen et al., 1985a; 1986; Inana et al., 1983). The β- and γ-crystallin gene families are thus members of the same super family. To trace the course of evolution of this super family the structure and chromosomal organization of its constituent genes must be determined. The organization of the γ-crystallin gene family has been well studied in rat and man (Den Dunnen et al., 1985b; Moormann et al., 1985; Meakin et al., 1985). In both species the γ-crystallin genes are closely linked. In contrast, little is known about the organization of the β-crystallin gene family. We have previously isolated a 60-kb regi-

on from the rat genome containing the βB1 gene. Since no other β-crystallin gene could be detected within this region, we suggested that the β-crystallin gene family might be dispersed (Moormann et al., 1984). However, we now show that two other β-crystallin genes, the βB2 and the βB3 genes, are closely linked within the rat genome. The organization of the human genomic copies of these genes is slightly more complex: we found that the human genome contains two copies of the βB2-crystallin gene. One of these two genes, which we have denoted the βB2-1 gene, is linked to the βB3-crystallin gene. The location of the second gene, the βB2-2 gene, is unknown.

MATERIALS AND METHODS

Isolation of genomic clones

A rat genomic cosmid library (Moormann et al., 1984) was screened with the nick-translated insert of the βB3 cDNA clone pRLβB3-2 (Den Dunnen et al., 1985a). Positive colonies were picked and purified by replating and rescreening.

A human library was constructed by cloning size-fractionated (7-21 kb) EcoRI-digested human genomic DNA in the EcoRI site of the vector Charon 4A. 200,000 plaques were screened with the nick-translated inserts of the βB2 cDNA clone pBLβBp (Quax-Jeuken et al., 1984) and the βB3 cDNA clone pRLβB3-2 (Den Dunnen et al., 1985a).

800,000 plaques of a second human genomic library, constructed by Dr. P. Leder and co-workers and consisting of a partial MboI digest of human DNA cloned in the BamHI site of the vector Charon 28, were also screened with the same probes. Positive plaques were purified by successive replating and screening. Hybridization conditions were as described in Moormann et al. (1984).

Isolation and digestion of DNA

Human chromosomal DNA was isolated from whole blood as described in Bell et al. (1981). Rat chromosomal DNA was isolated from liver as described in Moormann et al. (1983).

Phage DNA and cosmid DNA were generally prepared with the use of standard methods as described by Maniatis et al. (1982). Restriction-enzyme digestions were carried out under the conditions described by the supplier. Analysis of restriction digests by Southern blotting was performed as described in Moormann et al. (1984).

Probes

The PstI-excised inserts of the cDNA clones pRLβB1-3 (rat βB1), pRLβB3-2 (rat βB3) (Den Dunnen et al., 1985a), pBLβBp (calf βB2), pBLβ-A(1+3) (calf βA3/A1) and pBLβs (calf βs) (Quax-Jeuken et al., 1984) were labeled with ^{32}P by nick-translation and were used as probe in the various filter-hybridization experiments. M13mp subclones of pRLβB3-2 and pBLβBp were used to determine the transcriptional orientation of the βB2 and the βB3 genes.

RESULTS AND DISCUSSION

Isolation and characterization of the rat genomic clones

From a previously described rat genomic cosmid library (Moormann

et al., 1984) four clones were isolated which hybridized with the rat βB3 cDNA clone. The physical map of the insert of two of these clones, pRcosβ-1 and pRcosβ-3, is shown in Figure 1. These two clones were found to contain two regions which hybridized with the βB3 cDNA clone: the contiguous 3.0-kb BamHI/HindIII and the 4.8-kb HindIII fragments (see Figure 1) hybridized strongly while a weak hybridization signal was found in the 5.6-kb region located between a BglII and a XhoI site. Since these two regions are separated by 11-kb, we suspected that the weakly hybridizing region might harbor a second, different, β-crystallin gene. Restriction digests were therefore hybridized with four other cDNA clones, namely the rat βB1, the calf βB2, the calf βA3/A1 and the calf βs cDNA clone (Den Dunnen et al., 1985a; Quax-Jeuken et al., 1984). Only the calf βB2 cDNA clone hybridized strongly to the 5.6-kb BglII/XhoI fragment of pRcosβ-1 and pRcosβ-3, which suggests that the genomic region cloned in pRcosβ-1 and pRcosβ-3 contains the βB3-crystallin gene as well as a sequence fairly homologous to the calf βB2-crystallin sequence.

To show that the cloned region does indeed contain the rat βB2 and βB3 genes and that its organization corresponds to that of the genome, genomic DNA was digested with various restriction enzymes and the digests were hybridized with either the βB2 or the βB3 cDNA clone. As shown in Figure 2, the various genomic fragments co-migrate with the cloned fragments (except when the cloned fragments contain part of the vector molecule). No other genomic fragments than those also present in clone pRcosβ-3 are found. Thus, clone pRcosβ-3 (or pRcosβ-1) contains all of the βB2 and βB3 sequences present in the rat genome. The direction of transcription of these two genes was established by hybridization with the 5', middle and 3' fragments of the βB3 cDNA clone and the middle and 3' regions of the βB2 cDNA clone. The two genes are oriented head to tail as shown in Figure 1.

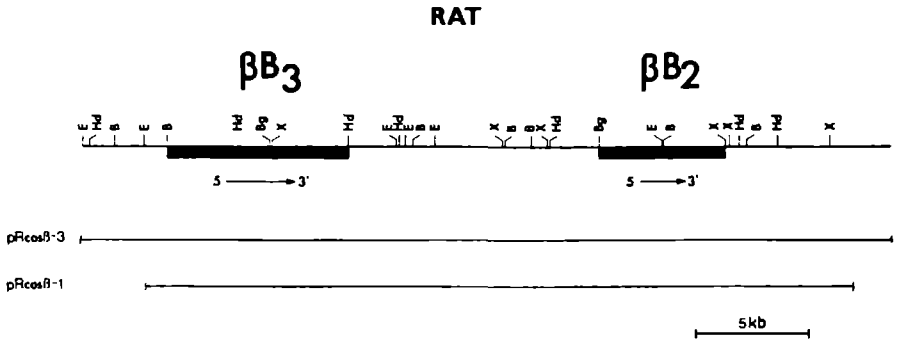


Figure 1: Physical map of the rat genomic region containing the βB2- and βB3-crystallin genes. The thin lines below the physical map of the rat genomic region represent the inserts of the cosmid clones pRcosβ-1 and pRcosβ-3. Black bars correspond to the regions that hybridize with the rat βB3 (pRLb3-2) or the calf βB2 (pBLβBp) cDNA clone. Arrows denote the transcriptional orientation. Restriction sites are indicated by: B=BamHI, Bg=BglII, E=EcoRI, Hd=HindIII, X=XhoI.

Isolation and characterization of the human genomic clones

To isolate the human genomic copies of the βB2- and βB3-crystallin genes, we initially cloned the 13-kb genomic EcoRI fragment hybridizing with the rat or calf βB3-crystallin sequences and the 14-kb and 11-kb EcoRI fragments hybridizing with the calf βB2-crystallin sequences (see also Figure 3). Since the genomic EcoRI hybridization pattern obtained with the βB2-crystallin probe contains an additional 9.0-kb fragment, and since a preliminary characterization of the genomic βB3-crystallin clones showed that the cloned EcoRI fragment lacked the 5' end of the gene, additional genomic clones were isolated from a bank containing a partial MboI digest of human genomic DNA. The genomic regions contained in the six clones selected for further study are shown in Figure 4. The approximate location of the βB2-crystallin and the βB3-crystallin coding sequences was established by hybridizing various restriction digests with the appropriate cDNA probes. In addition the direction of transcription was determined by hybridizing with the 5', "middle", and 3' regions of the rat βB3-crystallin cDNA clone and the "middle" and 3' regions of the calf βB2-crystallin cDNA clone (this clone lacks the 5' region). The results of these studies are summarized in Figure 4. The cloned fragments containing the βB3-crystallin sequences represent all

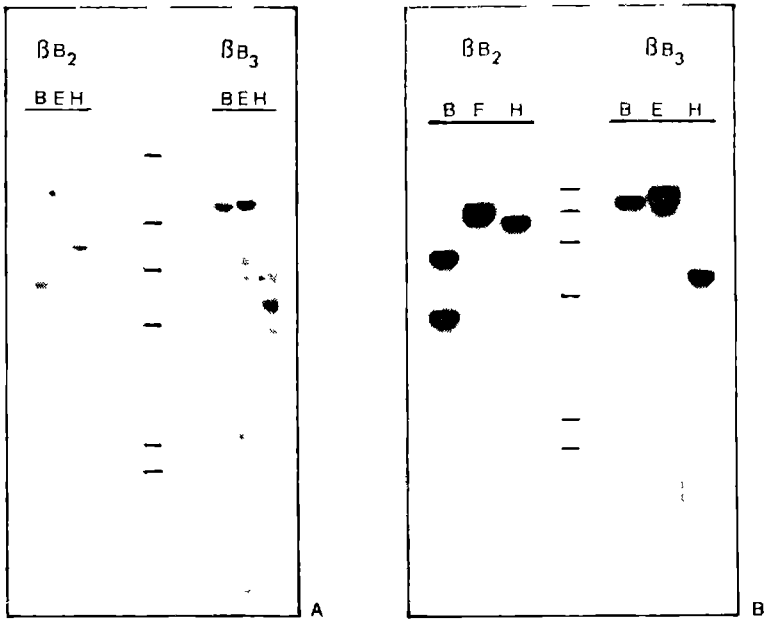


Figure 2: Hybridization pattern of rat genomic and cloned DNA coding for βB2- and βB3-crystallins. Southern blot hybridization analysis of various restriction digests of rat genomic (panel A) and pRcosβ-3 (panel B) DNAs. 10 µg of rat liver DNA or 1 µg of pRcosβ-3 DNA were digested with BamH1 (B), EcoRI (E) or HindIII (H). Electrophoresis of restricted DNA was carried out in 0.7% agarose gels using 1xE (40 mM Tris, 20 mM NaAc, 2 mM EDTA, pH 7.8) buffer in panel A and 0.5xTBE (50 mM Tris, 50 mM H_3BO_3, 1.25 mM EDTA, pH 8.2) in panel B. The DNA was blotted onto nitrocellulose and hybridized with the nick-translated ^{32}P-labeled insert of the βB2 cDNA clone pBLβBp or with the nick-translated ^{32}P-labeled insert of the βB3 cDNA clone pRLβB3-2, as described by in Moormann et al. (1985). The blots were washed twice with 2xSSC/0.1% SDS at 42 °C for 20 min, and once with 0.5xSSC/0.1% SDS at 50 °C for 20 min. DNA size-markers are specified in kb and were obtained by HindIII digestion of phage λ DNA. Weakly hybridizing bands (panel A) in the chromosomal EcoRI (βB2, ± 16-kb) and HindIII digest (βB3, ± 6.6-kb) are indicated by asterisks. The difference between the βB2 hybridization patterns of the EcoRI-digested genomic and the EcoRI-digested cloned DNA is due to the fact that only a part (10.1-kb) of the largest genomic EcoRI fragment, which hybridizes with the βB2 cDNA clone, is present in clone pRcosβ-3 (see Figure 1). The pRcosβ-3 EcoRI fragment also contains 0.5-kb of vector sequences. The clone thus contains an EcoRI fragment which has approximately the same size as the second genomic or cloned EcoRI fragment.

the βB3-sequences present in genomic DNA (see for example Figure 3). We therefore conclude that a single copy of the βB3-crystallin gene is present in the human genome.

The presence of two βB2-crystallin genes in the human genome

As shown in Figure 4 we obtained four genomic clones containing βB2-crystallin coding sequences. One of these, λHeB2-4, contains only the 14-kb genomic EcoRI fragment. A second, λHmB2-13, also contains the 14-kb EcoRI fragment but extends further up and downstream. The restriction map of the 3' region of this clone overlaps that of a third clone, λHmB2-40 (see Figure 4).

Together these three clones define a genomic region of about 28-kb containing one βB2-crystallin gene (the βB2-1 gene). The fourth genomic clone containing βB2-crystallin sequences, λHeB2-3, contains an 11-kb EcoRI fragment. The restriction map of this fragment does not correspond to that of the three clones mentioned above. We have considered the possibility that the insert of this clone represents a cloning artefact. However, the genomic hybridization pattern contained fragments corresponding to those present in this clone (data not shown, see also Figure 3). We therefore concluded that the insert of this clone contained (the 5' part of) a second human βB2-crystallin gene, which we have denoted the βB2-2 gene. The parts of the coding region of these two genes that have been sequenced thus far indicate that these two genes are more than 90% homologous (see Figure 5a). The high sequence homology extends into the intronic regions (Figure 5b), the divergence of the intronic regions sequenced thus far is 10%. Since the estimated rate of divergence of silent regions in the human lineage is 1% in 1.4 million years (Perler et al., 1980), the duplication of the βB2-crystallin gene must have occurred at least 14 million years ago. The data presented above show that the human genome contains two copies of the βB2-crystallin gene while the rat genome has only one. It is curious that the human γ-crystallin gene cluster also contains one more gene copy than the corresponding gene cluster of the rat (Den Dunnen et al., 1985b; Moormann et al., 1985; Meakin et al., 1985). From a comparison of the human and the rat γ-crystallin gene clusters it is apparent that both are orthologous, yet at least two of the genes which are active in the rat have become pseudogenes in man. Hence, although the human geno-

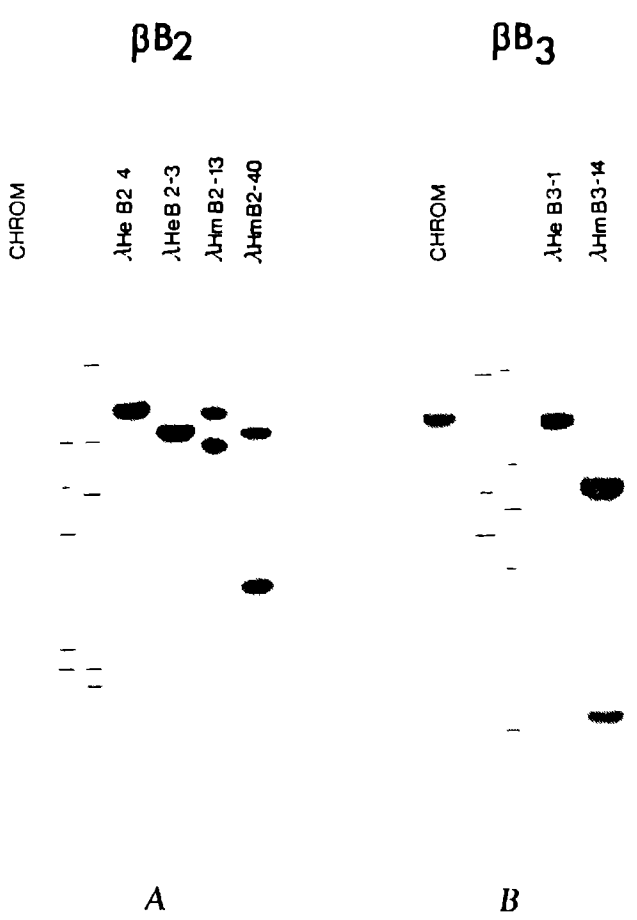


Figure 3: Hybridization pattern of human genomic and cloned DNA coding for βB2- and βB3-crystallins. **Panel A:** 10 µg of EcoRI-digested human genomic and 1 µg of EcoRI-digested λHeB2-3, λHeB2-4, λHmB2-13 or λHmB2-40 DNA were hybridized with the βB2 cDNA clone. Note that both the genomic clones λHeB2-3 and λHmB2-40 contain an 11-kb EcoRI fragment. The genomic 11-kb EcoRI band must thus be a doublet. The black dot indicates the very weakly hybridizing genomic 9-kb EcoRI band. The lower hybridizing EcoRI fragments of λHmB2-13 and λHmB2-40 also contain vector sequences. **Panel B:** 10 µg of EcoRI-digested human genomic and 1 µg of EcoRI-digested λHeB3-1, and 1 µg of EcoRI+HpaI-digested λHmB3-14 DNA were hybridized with the βB3 cDNA clone. Electrophoresis, blotting and hybridization were performed as described in Figure 2. DNA size-markers are as in Figure 2.

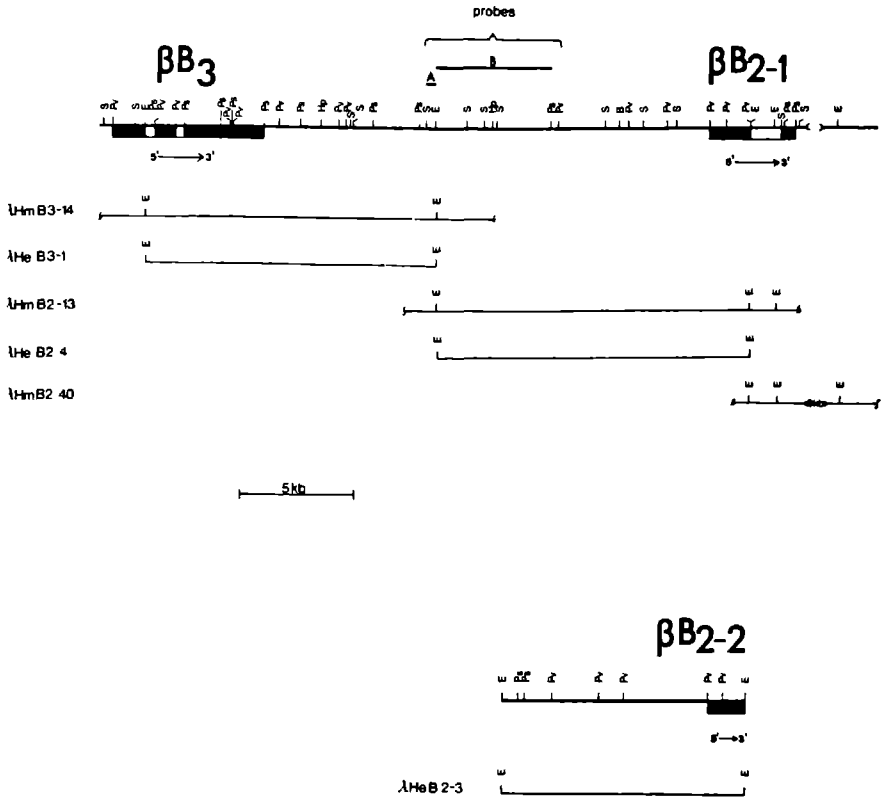


Figure 4: Physical map of the human genomic region containing the βB2- and βB3-crystallin genes. The thin lines below the physical map of the human genomic region represents the inserts of the phage λ clones λHmB3-14, λHeB3-1, λHmB2-13, λHeB2-4, λHmB2-40 and λHeB2-3. The junctions between insert and vector of the partial MboI phage λ clones are indicated by /. <8kb> indicates that the distance between these points is 8-kb. Black bars correspond to the regions that hybridize with the rat βB3 (pRLb3-2) or the calf βB2 (pBLβBp) cDNA clone. Open bars represent restriction fragments that do not hybridize with the cDNA probes and presumably contain intronic regions only. Arrows denote the transcriptional orientation. Restriction sites are indicated by: B=BamHI, E=EcoRI, Hp=HpaI, Ps=PstI, Pv=PvuII and S=SstI. The fragments which were used as probe (probe A and B) in the hybridization experiments described in RESULTS are indicated.

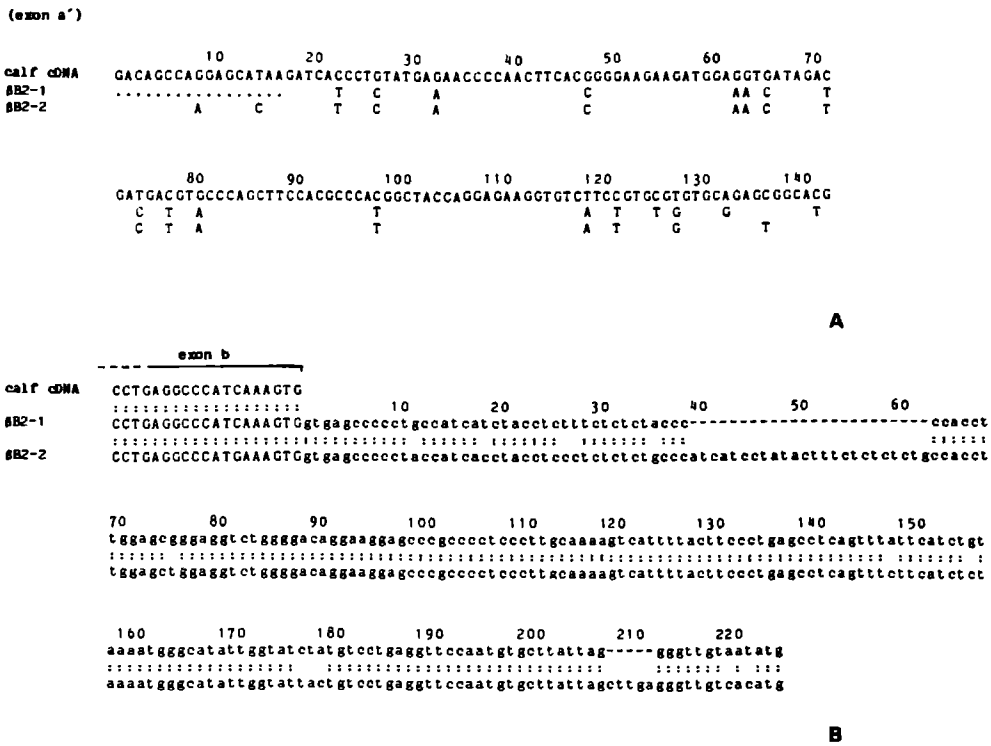


Figure 5: Comparison of exonic and intronic sequences of the human βB2-1 and βB2-2 genes. **(a)**: Nucleotide sequences of exon a' (motif III) of the human βB2-1 and βB2-2 genes compared with the sequence of the calf cDNA (H.J.M. A., unpublished results). The entire sequence of the calf cDNA is shown, whereas for the human genes only the differences with the calf sequence are indicated. Sequences which are not determined are indicated by a dotted line. The numbering starts from the first nucleotide of exon a'. **(b)**: Comparison of the intronic sequences flanking exon b (motif II) of the human βB2-1 and βB2-2 genes. The sequence are aligned to optimize homology. In the upper left, the 3' part of exon b of both human genes are compared with the calf sequence. The numbering starts from the first nucleotide (g) of the intron. Identical nucleotides are indicated with colons. Deletions are indicated by dashed lines.

me contains more γ-crystallin gene copies than the rat genome, it actually has fewer active γ-crystallin genes. Whether the same is true for the βB2-crystallin genes is unknown: in Northern blots of human lens RNA only a single βB2-crystallin RNA is detected, but this species

could be encoded by either one or both of the genes. Studies to determine whether both copies of the βB2-crystallin genes are expressed in the human lens are presently under way.

The cloned fragments containing the human βB2-crystallin sequences do not comprise all the genomic fragments: for example, our clones do not contain the 9-kb genomic EcoRI fragment (see Figure 3). We assume that this genomic fragment contains the 3' region of the βB2-2-crystallin gene, which is not present in our clones.

Linkage of the βB2-crystallin and βB3-crystallin genes

We have shown above that in the rat genome the βB2-crystallin and the βB3-crystallin genes are linked. To determine whether this is the case in the human genome also, we have isolated a 0.45-kb EcoRI/SstI fragment (probe A, see Figure 4) containing only single-copy sequences from the 3' region of the genomic βB3-crystallin clone λHeB3-1. This fragment hybridized specifically to the 2.0-kb HpaI/EcoRI fragment of the genomic βB2-crystallin clone λHmB2-13 (see Figure 6). This fragment contains 1.4-kb of the utmost 5' region of the insert (the HpaI site is located in the vector). Conversely, a 5.2-kb EcoRI/PstI fragment from the 5' region of the genomic βB2-crystallin clone λHeB2-4 (probe B, see Figure 4) hybridized specifically to the 3' EcoRI/HpaI fragment of the genomic βB3-crystallin clone λHmB3-14 (see Figure 6). This experiment shows that the 3' end of the insert of λHmB3-14 and the 5' end of the insert of λHe2-13 contain the same sequences. Hence the βB3-crystallin gene and the βB2-1-crystallin gene must be linked and are arranged as shown in Figure 4. We have as yet not been able to determine the location of the second βB2-crystallin gene, the βB2-2 gene, with respect to the βB3- and βB2-1-crystallin two gene cluster. Both in the human genome and in the rat genome (RESULTS AND DISCUSSION) the single-copy βB3-crystallin gene is linked to a βB2-crystallin gene. Thus at least some of the β-crystallin genes are clustered. Others must, however, be dispersed. The best evidence for dispersion of the β-crystallin genes comes from chromosome mapping studies in man. These have shown that a gene for an acidic β-crystallin, the βA3/A1-crystallin gene, is located on chromosome 17 (Law et al., 1986), while βB2-crystallin sequences were mapped to chromosome 22 (De Jong and Hendriks, 1986). A comparison of the (calf) β-crystallin sequences suggests that separation into the acidic (such as βA3/A1) and basic β-crystallins (e.g. βB1, βB2, and

βB3) was an early event in the evolution of the β-crystallin protein family (Quax-Jeuken et al., 1985b). Dispersion of the proto-βA- and the proto-βB-crystallin genes could have preceded the duplications of the proto-βB-crystallin gene, which yielded the present three βB-crystallin genes. In the phylogenetic tree of the calf β-crystallins, the βB2-

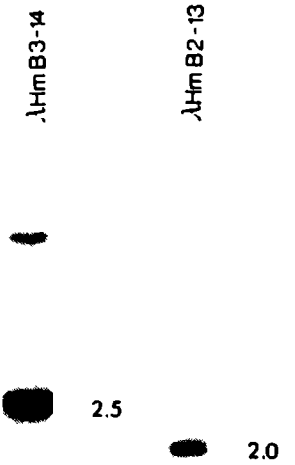


Figure 6: Linkage of the human βB3 and the βB2-1-crystallin gene. 0.5 μg of λHmB3-14 DNA and 0.5 μg of λHmB2-13 were double digested with HpaI+EcoRI. After electrophoresis on a 0.7% agarose gel the DNA was blotted onto nitrocellulose. The filter containing the λHmB3-14 digest was hybridized with the ^{32}P-labeled probe B (see Figure 4) and the filter containing the λHmB2-13 digest was hybridized with ^{32}P-labeled probe A (see Figure 4). The sizes (in kb) of the hybridizing fragments are indicated.

crystallin branches off before the βB1- and βB3-crystallins. Hence, the close linkage found here between the βB2- and the βB3-crystallin genes is unlikely to be the result of a late gene duplication. Rather, it is probably a remnant of an ancestral cluster of the βB-crystallin genes including the βB1-crystallin gene. The determination of the chromosomal location of the βB1-crystallin gene should show whether the distance between this gene and the βB2/βB3-crystallin gene cluster has become too large to be detected in cloning experiments such as reported here or elsewhere (Moormann et al., 1984) or whether this gene has jumped to a different chromosome.

CONCLUSIONS

(1) The βB2- and βB3-crystallin genes are clustered in man and rat. This cluster is probably a remnant of an ancestral βB-crystallin gene cluster.

(2) In the rat genome the βB2-crystallin gene is a single-copy gene. In man two copies are found. The duplication of the ancestral βB2-crystallin gene in the primate lineage must have occurred at least 14 millions years ago.

ACKNOWLEDGEMENT: We thank R. Segers for help with the rat genomic DNA digests. This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

REFERENCES

Bell, C.I., Karam, J.H. and Rutter, W.J.: Polymorphic DNA region adjacent to the 5' end of the human insulin gene. Proc. Natl. Acad. Sci. U.S.A. 78 (1981) 5759-5763.

Berbers, G.A.M., Hoekman, W.A., Bloemendal, H., De Jong, W.W., Kleinschmidt, T. and Braunitzer, G.: Homology between the primary structures of the major bovine β-crystallin chains. Eur. J. Biochem. 139 (1984) 467-479.

Berbers, G.A.M., Boerman, O.C., Bloemendal, H. and De Jong, W.W.: Primary gene products of bovine β-crystallin and reassociation behavior of its aggregates. Eur. J. Biochem. 128 (1982) 495-502.

Bloemendal, H.: Lens proteins. CRC Critical Rev. Biochem. 12 (1982) 1-38.

De Jong, W.W. and Hendriks, W.: The eye lens crystallins: Ambiguity as evolutionary strategy. J. Mol. Evol. 24 (1986) 121-129.

Den Dunnen, J.T., Moormann, R.J.M. and Schoenmakers, J.G.G.: Rat lens β-crystallins are internally duplicated and homologous to γ-crystallins. Biochim. Biophys. Acta 824 (1985a) 295-303.

Den Dunnen, J.T., Moormann, R.J.M., Creemers, F.P.M. and Schoenmakers, J.G.G.: Two human γ-crystallin genes are linked and riddled with Alu-repeats. Gene 38 (1985b) 197-204.

Den Dunnen, J.T., Moorman, R.J.M., Lubsen, N.H. and Schoenmakers, J.G.G.: Intron insertions and deletions in the β-/γ-crystallin gene family: The rat βB1-gene. Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 2855-2859.

Driessen, H.P.C., Herbrink, P., Bloemendal, H. and De Jong, W.W.: Primary structure of the bovine β-crystallin Bp chain: Internal duplication and homology with γ-crystallin. Eur. J. Biochem. 121 (1981) 83-91.

Hejtmancik, J.F. and Piatigorsky, J.: Diversity of β-crystallin mRNAs of the chicken lens: Hybridization analysis with cDNA clones. J. Biol. Chem 258 (1983) 3382-3387.

Inana, G., Piatigorsky, J., Norman, B., Slingsby, C. and Blundell, T.: Gene and protein structure of a β-crystallin polypeptide in murine lens: relationship of exons and structural motifs. Nature 302 (1983) 310-315.

Law, M.L., Cai, G.-Y., Hartz, J., Kao, F.-T., Hogg, D., Breitman, M.L. and Tsui, L.-C.: Localization of a β-crystallin gene, the huβA3/A1

(gene symbol, CRYB1), to the long arm of human chromosome-17. Cytogenet. Cell Genet. 42 (1986) 202-207.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning, A laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Meakin, S.O., Breitman, M.L. and Tsui, L.-C.: Structural and evolutionary relationships among five members of the human γ-crystallin gene family. Mol. Cell. Biol. 5 (1985) 1408-1414.

Moormann, R.J.M., Den Dunnen, J.T., Mulleners, L., Andreoli, P.M., Bloemendal, H. and Schoenmakers, J.G.G.: Strict colinearity of genetic and protein folding domains in an intragenically duplicated rat lens γ-crystallin gene. J. Mol. Biol. 171 (1983) 353-368.

Moormann, R.J.M., Jongbloed, R. and Schoenmakers, J.G.G.: Isolation and characterization of β- and γ-crystallin genes from rat genomic cosmid libraries. Gene 29 (1984) 1-9.

Moormann, R.J.M., Den Dunnen, J.T., Heuyerjans, J., Jongbloed, R.J.E., Van Leen, R.W., Lubsen, N.H. and Schoenmakers, J.G.G.: Characterization of the rat γ-crystallin gene family and its expression in the eye lens. J. Mol. Biol. 182 (1985) 419-430.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J.: The evolution of genes: The chicken preproinsulin gene. Cell 20 (1980) 555-566.

Piatigorsky, J.: Lens differentiation in vertebrates: A review of cellular and molecular features. Differentiation 19 (1981) 134-153.

Piatigorsky, J.: Lens crystallins and their gene families. Cell 38 (1984) 620-621.

Quax-Jeuken, Y., Janssen, G., Quax, W., Van den Heuvel, R. and Bloemendal, H.: Bovine β-crystallin complementary DNA clones. Alternating proline/alanine sequence of βB1 subunit originates from a repetitive DNA sequence. J. Mol. Biol. 180 (1984) 457-472.

Quax-Jeuken, Y., Quax, W., Van Rens, G., Meera Khan, P. and Bloemendal, H.: Complete structure of the αB-crystallin gene: Conservation of the exon/intron distribution in the two nonlinked α-crystallin genes. Proc. Natl. Acad. Sci. U.S.A. 82 (1985a) 5819-5823.

Quax-Jeuken, Y., Driessen, H., Leunissen, J., Quax, W., De Jong, W. and Bloemendal, H.: βs-crystallin: Structure and evolution of a distinct member of the βγ-superfamily. EMBO J. 4 (1985b) 2597-2602.

Ramaekers, F., Dodemont, H., Vorstenbosch, P. and Bloemendal, H.: Classification of rat lens crystallins and identification of proteins encoded by rat lens mRNA. Eur. J. Biochem. 128 (1982) 503-508.

Shinohara, T., Robinson, E.A., Apella, E. and Piatigorsky, J.: Multiple γ-crystallin mRNAs of the mouse lens: Fractionation of mRNAs by cDNA cloning. Proc. Natl. Acad. Sci. U.S.A. 79 (1982) 2783-2787.

Slingsby, C. and Croft, L.R.: Structural studies on calf lens γ-crystallin fraction IV: a comparison of the cysteine-containing tryptic peptides with the corresponding amino acid sequence of the γ-crystallin fraction II. Exp. Eye Res. 26 (1978) 291-304.

Van den Heuvel, R., Hendriks, W., Quax, W. and Bloemendal, H.: Complete structure of the αA-crystallin gene: Reflection of an evolutionary history by means of exon shuffling. J. Mol. Biol. 185 (1985) 273-284.

CHAPTER 3

THE STRUCTURE AND SEQUENCE OF THE RAT βB2- AND βB3-CRYSTALLIN GENES

The rat is known to have at least three basic β-crystallin genes, the βB1, βB2 and βB3 genes. The complete structure of one of these genes, the βB1-crystallin gene, was elucidated by Den Dunnen et al. (1986a). They showed that the rat βB1-crystallin gene is 13.6 kb long and contains 6 exons. Each exon, except the first and last one, encodes a single functional entity of the βB1 protein. Exon 6 contains the coding sequence of motif IV as well as the coding sequence of the C-terminal extension of the protein. The first exon of the βB1-crystallin gene, 27 bp long, contains only non-coding sequences. It is separated by 1.36 kb of DNA from the exon (exon 2) which contains the ATG translation initiation codon. In contrast, in the human βA3/A1-crystallin gene, which also contains six exons, a translation initiation site is found within the first two exons. Hence this gene does not contain non-coding exons.

To investigate whether the 5' ends of the rat βB2- and βB3-crystallin genes are specified by a coding or non-coding exon we have determined the exact location and sequence of the first and second exons of these genes.

Mapping of the 5' end of the rat βB3-crystallin gene

We have previously reported the isolation of a genomic clone, pRcosβ-3, containing both the βB2- and βB3-crystallin gene sequence (see Aarts et al., 1987). The physical map of this clone is given in Figure 1. Hybridization with the rat βB3-crystallin cDNA probe (Den Dunnen et al., 1985a) had revealed that the βB3-crystallin gene is located on the contiguous 3.0 kb BamHI-HindIII and 4.8 kb HindIII fragments (Aarts et al., 1987). To map the 5' end of this gene we subcloned the 5.6 kb EcoRI-BglII fragment, containing the 5' part of the hybridizing region. A restriction map of this region is given in Figure 1. The approximate location of the exon which contains the ATG translation initiation codon was determined through hybridization with the rat βB3 cDNA probe. This region was subcloned in M13 vectors and sequenced. The sequence of the 5' end and upstream sequences of this exon are given in Figure 2. The exact 5' end of the exon was determined in a S1-nuclease

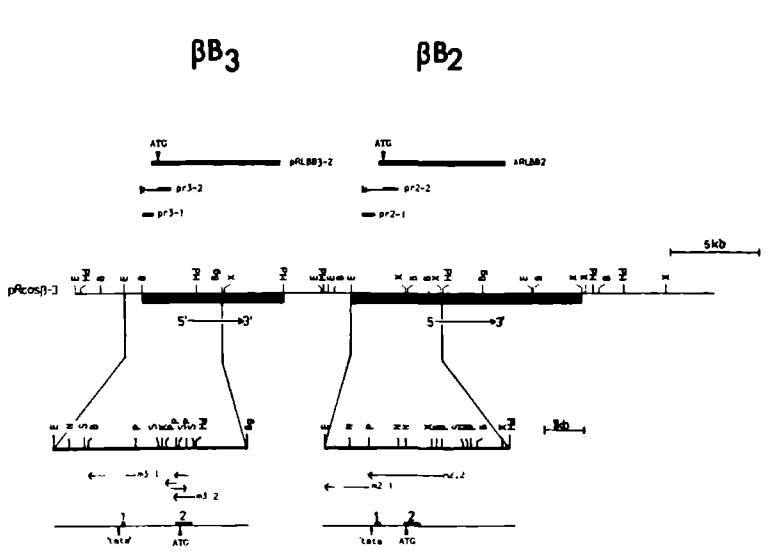


Figure 1: Physical map of pRcosβ-3 and the strategy to identify the 5' end of the βB3- and βB2-crystallin genes. The restriction map of pRcosβ-3 (see also Aarts et al., 1987) is given in the middle. The black bars correspond to the regions that hybridize with the rat βB3 (pRLβB3-2; Den Dunnen et al., 1985a) or the rat βB2 (λRLβB2; Chapter 4) cDNA clone. The cDNAs are depicted as solid bars at the top of the figure. The primers used are indicated as small black bars, pr3-2 (5'-CGTGCTGCTCTGCCAT-3') is complementary to the first 16 coding nucleotides of the cDNA clone pRLβB3-2, pr2-2 (5'-CTGGTGGTCTGAGGCCATG-3') to the last non-coding and the first 18 coding nucleotides of the cDNA clone λRLβB2; pr3-1 (5'-CTGGCCTCCAGGACC-3') and pr2-1 (5'-GACCAGATNGCCAGCCCCNGCGAG-3') are complementary to the 5' sequences of the βB3- and βB2-crystallin RNA respectively. pr3-1 and pr2-1 were used to determine the exact location of the first exons of the two genes within the subcloned regions. A fine map of these regions is given under the physical map of pRcosβ-3. The locations of exon 1 and exon 2 of both genes are given on the bottom line of the figure. The positions of the TATA boxes ("tata") and ATG initiation codons are indicated. The horizontal arrows indicate the extent and direction of sequencing of fragments subcloned in M13mp vectors. Restriction sites are indicated by: B, BamHI; Bg, BglII, E, EcoRI; Hd, HindIII; K, KpnI; N, NcoI; P, PstI; S, SstI; X, XhoI.

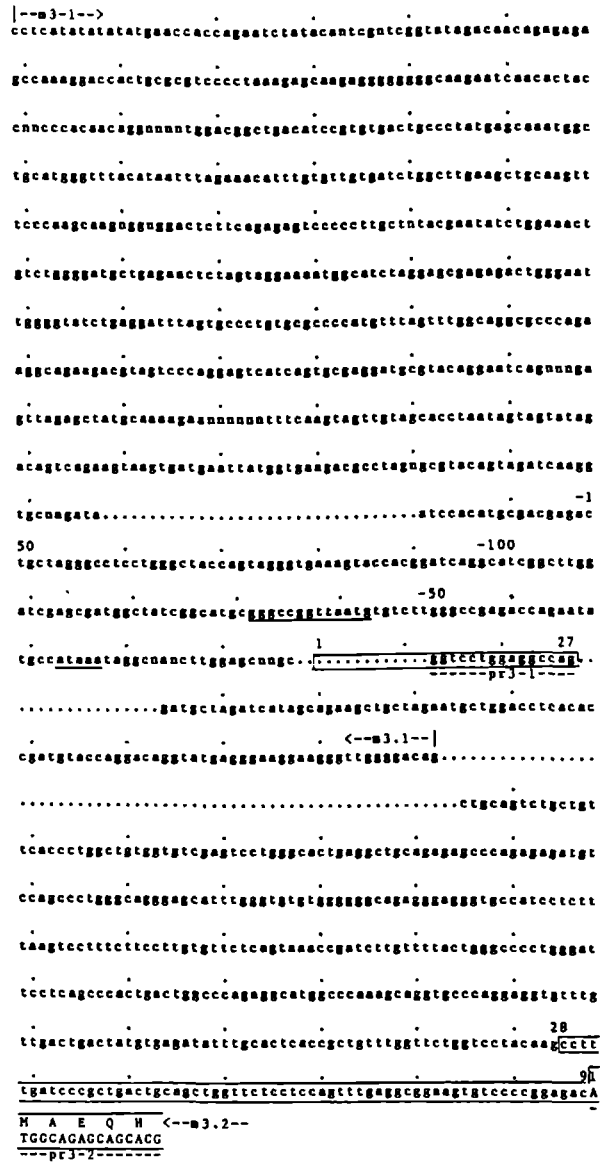


Figure 2: Sequences of the rat βB3-crystallin gene. The M13-clones indicated correspond to the clones depicted in Figure 1. The TATA-box, the G- and T-box (see also Figures 8 and 9) are underlined. The sequences complementary to the primers used are indicated. Exonic sequences are boxed. The beginning of the coding sequence is given in capitals. It was taken from Den Dunnen et al. (1985a) and is a cDNA sequence. The deduced amino acid sequence is given above the DNA sequence.

protection experiment and was found to be located near or at the aag sequence found 63 nucleotides upstream from the ATG (Figures 2 and 3). In front of this sequence no consensus promoter sequence required for the expression of eukaryotic genes is found and in analogy with the rat βB1-crystallin gene we suspected that this gene might contain a 5' non-coding exon. We therefore determined the length of the 5' non-coding region of the βB3 RNA by extending a primer complementary to the first 16 coding nucleotides of the βB3-crystallin mRNA (pr3-2; see Figure 1 and 2). Two extended products, 112 and 117 nucleotides long, were found (see Figure 3). These products extend past the 5' end of the ATG-containing exon, as mapped in the S1-nuclease experiments, by 25-30 nucleotides, suggesting that the rat βB3-crystallin gene does contain a 25-30 bp long 5' non-coding exon. Part of the sequence of this exon was

Figure 3: S1-nuclease and primer extension analysis of rat βB3-crystallin mRNA. The S1-nuclease reaction was carried out as described by Moormann et al. (1985) and the primer extension reaction as described by Geliebter et al. (1986). The ^{32}P labeled probe used in the S1-nuclease protection experiment was synthesized using the M13 subclone m3.2 as template and pr3-2 as primer. Primer pr3-2 was also used in the primer extension reaction and in the sequence reaction of m3.2. The S1-nuclease protected fragments (lane S1), the primer extension products (lane PE) and the sequence ladder of m3.2 (see Figure 1) were analysed on a 6% polyacrylamide sequencing gel.

determined by extending primer pr3-2 with reverse transcriptase in the presence of dideoxynucleotides (see Figure 4). Identity with the βB3-crystallin gene sequence is maintained up to the predicted splice site. The remainder of the RNA sequence should correspond to the sequence of the first exon. An oligonucleotide (pr3-1) complementary to this part of the RNA sequence was synthesized and used as probe to determine the exact location of the first exon of the βB3-crystallin gene in the genomic clone pRcosβ-3. Hybridization with pr3-1 was exclusively found with a 1.2 kb BamHI-PstI fragment (Figure 1). This fragment was subcloned in the vector M13mp18 (clone m3.1; see Figure 1), and the 5' gene flanking sequences were determined by using pr3-1 as primer in the sequence analysis. A TATA sequence was found approximately 30 nucleotides upstream from exon 1. The sequence of the rat βB3-crystallin gene, as far as it is now known, is given in Figure 2.

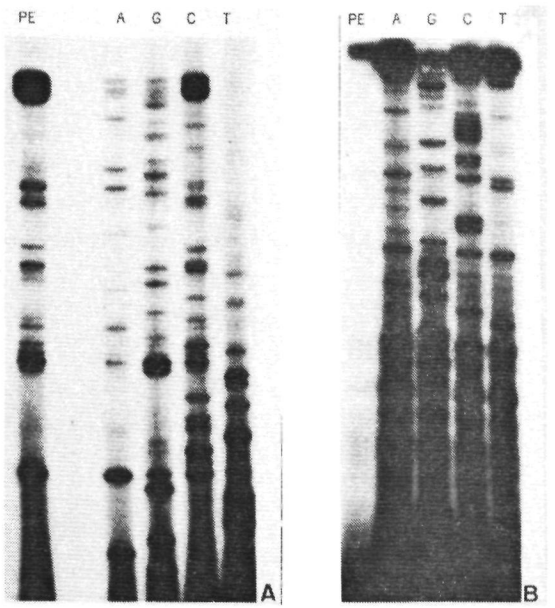


Figure 4: Sequence ladders of the 5' end of the βB3- and βB2-crystallin RNAs. Primer pr3-2 **(A)** and pr2-2 **(B)** were extended on total rat lens RNA under dideoxy-sequencing conditions as described by Geliebter et al. (1986). The primer extension products were analysed on a 9% polyacrylamide sequencing gel. The autoradiograms of these gels are shown. PE = primer extension reaction in the absence of dideoxynucleotides.

Mapping of the 5' end of the rat βB2-crystallin gene

Within pRcosβ-3, the βB2-crystallin gene was mapped to the 5.6 kb BglII-XhoI fragment by using the calf βB2 cDNA clone as probe (Aarts et al., 1987; see Figure 1). However, this probe only contains the 3' part of the coding sequence. When the full length rat βB2-crystallin cDNA clone (Chapter 4) was used as probe, the 8.0 kb EcoRI-BglII fragment located in front of the 5.6 kb BglII-XhoI fragment also appeared to contain βB2-crystallin sequences. The 5' part of this fragment, a 5.2 kb EcoRI-HindIII fragment, which was expected to contain the 5' end of the βB2-crystallin gene, was subcloned in the vector pUC19. A more detailed map of this region is given in Figure 1. The exon containing the ATG initiation codon was mapped around the second NcoI site within the 2.0 kb PstI fragment (the βB2-crystallin gene contains a NcoI site around its translation initiation codon). The sequence of the 5' end and upstream sequences of this exon were determined by using an oligonucleotide complementary to the last non-coding and first 18 coding nucleotides of the βB2-crystallin mRNA (pr2-2) as primer in a sequence analysis of the M13mp8 clone m2.2 (see Figure 1). To determine the utmost 5' sequences of the βB2-crystallin mRNA the same primer was extended on total rat lens RNA with reverse transcriptase in the presence of dideoxynucleotides. As shown in Figure 5, the cag splice acceptor consensus sequence marks the point at which the homology between the determined RNA and DNA sequences disappears. Therefore the g found at position 28 is probably the first nucleotide of the second exon of the βB2-crystallin gene (see also Figure 6).

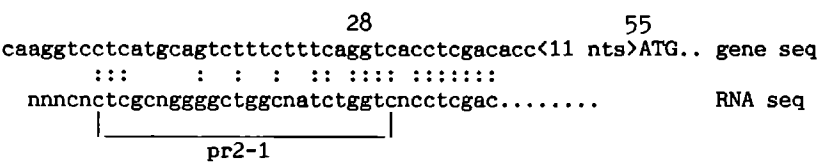


Figure 5: Comparison of the sequence of the 5' end and upstream region of the second exon of the rat βB2-crystallin gene (gene seq) and RNA sequence (RNA seq). Identical nucleotides are indicated with a colon. "n" indicates bases of uncertain identity. The numbering is as in Figure 6. Primer pr2-1 is complementary to the region indicated.

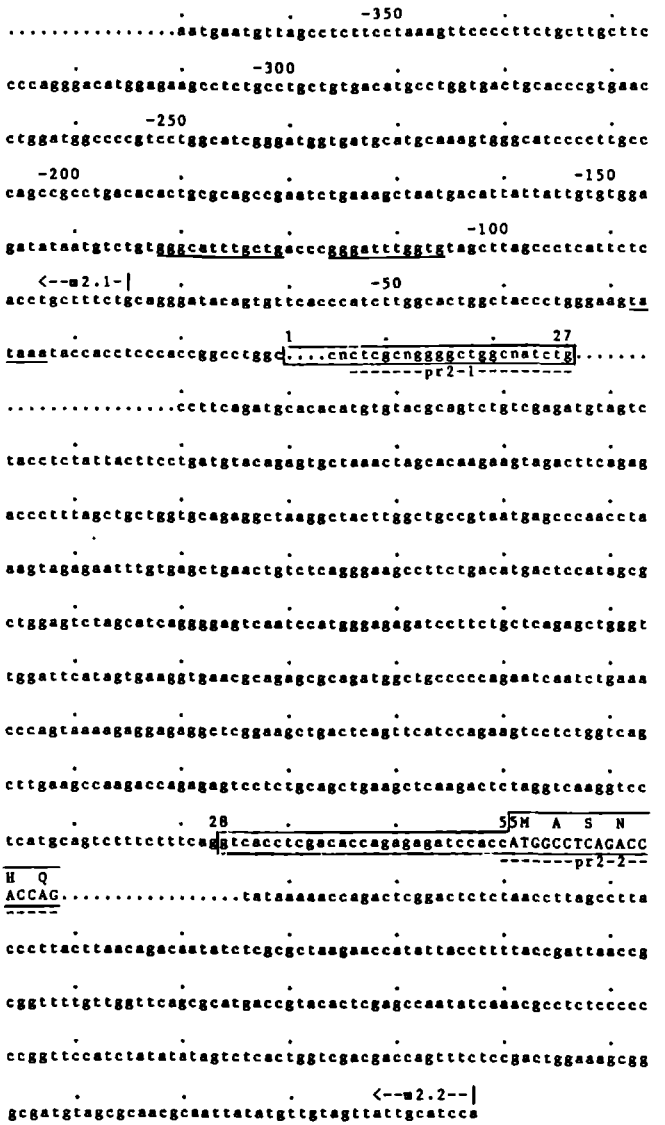


Figure 6: Sequences of the rat βB2-crystallin gene. The M13-clones indicated correspond to the clones depicted in Figure 1. The TATA-box, the G-and T-boxes (see also Figures 8 and 9) are underlined. The sequences complementary to the primers used are indicated. Exonic sequences are boxed. The beginning of the coding sequence is given in capitals. It was taken from Chapter 4 and is a cDNA sequence. The deduced amino acid sequence is given above the DNA sequence.

To identify the location of the 5' end of the βB2-crystallin gene an oligonucleotide complementary to the 5' region of the βB2-crystallin mRNA (pr2-1; Figures 1 and 6) was used. Hybridization with this oligonucleotide was exclusively found with the 2.0 kb PstI fragment of pRcosβ-3. Sequence analysis of the corresponding clone (m2.2; see above and Figure 1) using pr2-1 as primer revealed that the first exon of the βB2-crystallin gene is located approximately 80 bp downstream from the PstI site. A TATA box was found approximately 30 nucleotides upstream from exon 1. To determine the sequences further upstream the contiguous 1.1 kb EcoR1-PstI fragment was subcloned in the vector M13mp18 and sequenced. The complete sequence of the rat βB2-crystallin gene, as far as it is now known, is given in Figure 6.

Comparison of the 5' regions of the rat βB-crystallin genes

The structure of the 5' ends of the rat βB-crystallin genes is outlined in Figure 7. Each rat βB-crystallin gene contains a 5' non-coding exon. The length (25-30 bp) of this exon has remained constant during evolution but only traces of sequence resemblance are found between these regions (see Figure 7). The total length of the 5' non-coding region is variable as the ATG-translation initiation codons of the βB1-, βB2- and βB3-crystallin genes are located 11, 27 and 63 nucleotides, respectively, within the second exons.

βB1 ataacaccagcctgtccaccaacacag <intron 1.4 kb> < 11 bp > ATG

βB2 cnctcgcngggctggcnatctg <intron 0.8 kb> <-- 27 bp -> ATG

βB3 nnnnnnggtcctggaggccag <intron 1.2 kb> <--- 63 bp ---> ATG

Figure 7: Schematic representation of the first and second exon of the βB1- (Den Dunnen et al. 1986a), βB2- and βB3-crystallin genes. The exons are underlined. The lengths of the first intron and of the non-coding region of the second exon are shown. The sequence of the first exon or part thereof is given. The translation initiation site is indicated.

When the 5' flanking regions of the βB1-, βB2- and βB3-crystallin genes are compared many short regions of sequence resemblance are found (Figure 8). Within these regions a conserved sequence element was noted. This element is composed of a G-box (GGGC) followed, after a region variable in length, by a T-box (A/GTTT$(N)_n$TG with n=1, 2 or 3; see Figure 9). Similar elements are also present within the 5' flanking

sequences of the human or rat γ-crystallin genes (Den Dunnen et al., 1985b, 1986b; Meakin et al., 1985, 1987) and the human βA3/A1-crystallin gene (Hogg et al., 1986; see also Figure 9), but not within the 5' flanking sequences of the chicken δ- (Borras et al. 1985) or hamster α-crystallin genes (Van den Heuvel et al., 1985; Quax-Jeuken et al., 1985). As shown in Figures 8 and 9 this conserved element is present twice in the 5' flanking region of the βB1- and βB2-crystallin genes. The distance found between the G- and T-boxes to the TATA-box is variable. These length differences are mainly due to the insertion of a different number of ssDNA type elements (consensus sequence CAGAG),

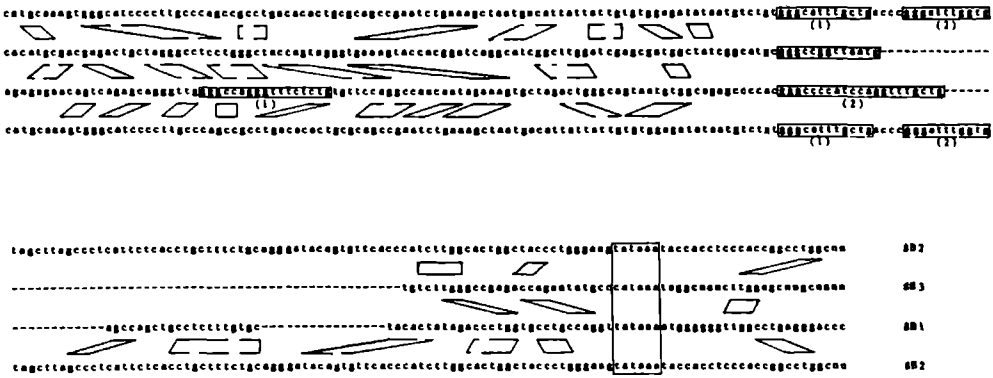


Figure 8: Comparison of the 5' flanking sequences of the rat βB1- (Den Dunnen et al., 1986a), βB2- and βB3-crystallin genes. Regions (>2 nucleotides long) of sequence identity are connected. Gaps were introduced to align the sequences from the TATA box up to the (first) conserved element (see text).

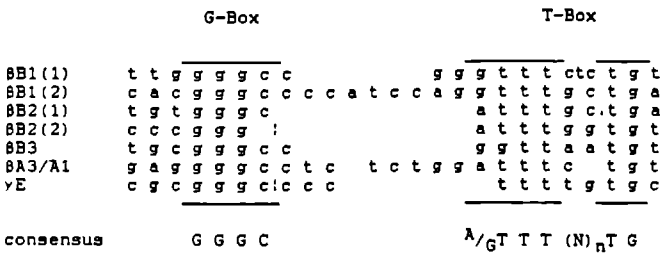


Figure 9: Alignment of the conserved element(s) present in the 5' flanking sequences of the rat βB1-, βB2- and βB3-, the human βA3/A1- and rat γE-crystallin gene. The deduced consensus sequence of the G- and T-box is given below the crystallin gene sequences.

which are also present within the 5' flanking sequences of the human βA3/A1- (Hogg et al., 1986) and human or rat γ-crystallin genes (Den Dunnen et al., 1985b, 1986b; Meakin et al., 1985, 1987). The role of the G- and T-boxes in the expression of the β-crystallin genes remains to be determined. Preliminary experiments showed that at least the G-box is involved in the tissue specific expression of the γ-crystallin genes (R. Peek, pers. comm.).

REFERENCES

Aarts, H.J.M., Den Dunnen, J.T., Lubsen, N.H., and Schoenmakers, J.G.G. (1987). Linkage between the βB2 and βB3 crystallin genes in man and rat: a remnant of an ancient β-crystallin gene cluster. Gene 59, 127-135.

Borras, T., Nickerson, J.M., Chepelinsky, A.B., and Piatigorsky, J. (1985). Structural and functional evidence for differential promoter activity of the two linked δ-crystallin genes in the chicken. EMBO J. 4, 442-452.

Den Dunnen, J.T., Moormann, R.J.M., and Schoenmakers, J.G.G. (1985a). Rat lens β-crystallins are internally duplicated and homologous to γ-crystallins. Biochim. Biophys. Acta 824, 295-303.

Den Dunnen, J.T., Moormann, R.J.M., Cremers, F.P.M., and Schoenmakers, J.G.G. (1985b). Two human γ-crystallin genes are linked and riddled with Alu-repeats. Gene 38, 197-204.

Den Dunnen, J.T., Moormann, R.J.M., Lubsen, N.H., and Schoenmakers, J.G.G. (1986a). Intron insertions and deletions in the β-/γ-crystallin gene family: The rat βB1 gene. Proc. Natl. Acad. Sci. U.S.A. 83, 2855-2859.

Den Dunnen, J.T., Moormann, R.J.M., Lubsen, N.H., and Schoenmakers, J.G.G. (1986b). Concerted and divergent evolution within the rat γ-crystallin gene family. J. Mol. Biol. 189, 37-46.

Geliebter, J., Zeff, R.A., Melvold, R.W., and Nathenson, S.G. (1986). Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis: Kbm9 and Kbm6. Proc. Natl. Acad. Sci. U.S.A. 83, 3371-3375.

Hogg, D., Tsui, L.-C., Gorin, M., and Breitman, M.L. (1986). Characterization of the human β-crystallin gene HuβA3/A1 reveals ancestral relationships among the βγ-crystallin superfamily. J. Biol. Chem. 261, 12420-12427.

Meakin, S.O., Breitman, M.L., and Tsui, L.-C. (1985). Structural and evolutionary relationships among five members of the human γ-crystallin gene family. Mol. Cell. Biol. 5, 1408-1414.

Meakin, S.O., Du, P.R., Tsui, L.-C., and Breitman, M.L. (1987). γ-crystallins of the human eye lens: Expression analysis of five members of the gene family. Mol. Cell. Biol. 7, 2671-2679.

Moormann, R.J.M., Den Dunnen, J.T., Heuyerjans, J., Jongbloed, R.J.E., Van Leen, R.W., Lubsen, N.H., and Schoenmakers, J.G.G. (1985). Characterization of the rat γ-crystallin gene family and its expression in the eye lens. J. Mol. Biol. 182, 419-430.

Quax-Jeuken, Y., Quax, W., Van Rens, G., Meera Khan, P., and Bloemendal, H. (1985). Complete structure of the αB-crystallin gene: Conservation of the exon/intron distribution in the two nonlinked α-crystallin genes. Proc. Natl. Acad. Sci. U.S.A. 82, 5819-5823.

Van den Heuvel, R., Hendriks, W., Quax, W., and Bloemendal, H. (1985). Complete structure of the α-crystallin gene. Reflection of an evolutionary history by means of exon shuffling. J. Mol. Biol. 185, 273-284.

CHAPTER 4

CRYSTALLIN GENE EXPRESSION DURING RAT LENS DEVELOPMENT

Henk J.M. Aarts, Nicolette H. Lubsen

and John G.G. Schoenmakers

Submitted for publication

SUMMARY

The analysis of the developmental pattern of expression of the αB-, βB1-, βB2-, βB3-, βA3/A1-, and βs-crystallin genes during fetal and postnatal development of the rat shows that the differential regulation of crystallin synthesis relies on differential gene shutdown rather than differential gene activation, that is all crystallin genes are active during early development but turn off at different stages. The only two exceptions to this rule are the αB- and βs-crystallin genes. The αB-crystallin gene transcript becomes first detectable at 18 days of fetal development, while the βs-crystallin gene appears to be active only in the postnatal period.

The possible contribution of post-transcriptional processes to the control of the synthesis of the crystallins was investigated by determining the absolute numbers of the αA-, αB-, βB1-, βB2-, βB3-, βA3/A1, βs-, and γ-crystallin gene transcripts present in the lens. Comparison of these RNA data with the in vitro protein synthetic pattern showed that the αB- and βB2-crystallin RNAs are relatively poorly translated. Examination of the known (hamster) αB-crystallin sequence and elucidation of the (rat) βB2-crystallin sequence yielded no evidence for aberrant codon usage. The possibility that protein synthesis is relatively poorly initiated on these two mRNAs is suggested by the fact that these are the only crystallin mRNAs in which the translation initiation codon is preceded by CCACC.

INTRODUCTION

The abundant water soluble proteins of the mammalian lens are encoded by three gene families, the α-, β- and γ-crystallin gene families (for review see [1, 2]). The protein products of these three gene families are not uniformly distributed throughout the lens and, as there is no cell death or protein turn-over within the lens, the crystallin concentration gradient must be laid down by differential expression of the various members of these three gene families during development of the lens. The exact mechanisms by which this differential synthesis is achieved have not yet been elucidated, but the regulation of transcription of the crystallin genes probably plays a major role. It has been shown, for instance, that the (chicken) β-crystallin [3, 4] and the (rat or mouse) γ-crystallin genes [5, 6] are differentially expressed during lens development. We show here that in general differential regulation of crystallin synthesis in the rat lens relies on differential shutdown rather than differential activation. To determine whether post-transcriptional regulation of crystallin synthesis

contributes to the spatial distribution of these proteins in the lens, it is necessary to establish the absolute levels of crystallin mRNA and to compare these levels with the protein synthetic pattern. We have therefore measured the amount of αA-, αB-, βB1-, βB3-, βA3/A1-, βs- and γ-crystallin RNA during development of the rat lens. These results, combined with the results from in vitro translation studies, make it likely that the αB- and βB2-crystallin mRNAs are poorly translated. Elucidation of the sequence of βB2-crystallin RNA showed no significant difference in codon usage between the basic β-crystallin mRNAs, nor does the codon use of the αB-crystallin mRNA differ from that of the αA-crystallin mRNA. The αB- and βB2-crystallin mRNAs do have the same sequence surrounding the translation initiation codon, suggesting that a reduced efficiency of initiation may be the cause of the poor use of these two mRNAs.

MATERIALS AND METHODS

Northern, Southern and Dot blotting

Total lens RNA was isolated from rat embryos and rats of different ages as described in [5]. Aliquots of 1.0 µg of each stage were loaded on a 1% agarose-formaldehyde gel and run in 3.1% formaldehyde, 25 mM phosphate (pH 7.0). After electrophoresis the gel was soaked in 20xSSC for 5 min and blotted to nitrocellulose (Schleicher & Schull). Dot blots were made by spotting aliquots of 1 µg and 0.2 µg of total lens RNA from each developmental stage on nitrocellulose filters. Serial dilutions of PstI-digested DNA of the appropriate cDNA clones (see below) were run on a 1% agarose gel in 0.5 x TBE. After electrophoresis the DNA was blotted to nitrocellulose. The nick-translated PstI-inserts of the following crystallin cDNA clones were used as probe: pRLαA-1 (rat αA [7]); pBLαA-1 (bovine αA [8]); pBLβA(1+3) (bovine βA3/A1 [10]); pBlαB-1 (bovine αB [8]); pRLβB1-3 (rat βB1 [9]); pBLβBp (bovine βB2 [10]); pRLβB3-1 (rat βB3 [9]); pBLbs (bovine βs [11]); pRLγ-2 (rat γC [12]); pRLγ-3 (rat γE [12]). Hybridization conditions were as described by [13].

Quantitation of the mRNAs

The relative quantities of the mRNAs were determined as described by [5]. The amount of total lens RNA dotted on nitrocellulose or transferred from an agarose gel was checked by hybridizing the RNA filters with a ribosomal DNA probe. The differences in signal found between the samples were used to correct the amounts of total lens RNA used. The autoradiograms of the Southern blots containing the serial dilutions of the PstI-digested DNA of the cDNA clones, which were hybridized together with the RNA blots, were used to generate a standard for quantitation of the corresponding mRNAs in the newborn or one-month old (for βs

mRNA) lens. The standard (S) was determined using the equation:

$$S = \frac{a \times N}{Mr} \times \frac{b}{c}$$

in which a represents the amount (g) of insert DNA of the cDNA clone, N Avogadro's number, Mr the molecular weight (of the hybridizing part) of the probe (see Table 1), b the ratio between "RNA" signal and the signal of the corresponding insert of the cDNA clone and c the amount (mg) of total lens RNA. The amount of transcripts of a particular crystallin gene per lens was determined using the following equation:

$$\#\ \text{transcripts per lens} = S \times Tn \times Rn/100$$

in which S represents the calculated standard, Tn the total amount of RNA per lens at time tn as determined by [5] and Rn the relative amount of that crystallin RNA at time tn (see Figure 2).

In vitro translation of lens RNA

The in vitro translation of lens RNA extracted from newborn rats in a wheat germ extract was performed as recommended by the supplier (BRL).

Isolation and sequencing of the rat βB2 cDNA clone λRLβB2

A rat lens λgt11-cDNA library was screened with the nick-translated PstI-insert of the calf βB2 cDNA clone pBLβBp [10]. Positive plaques were purified by successive rounds of replating and screening. Hybridization conditions were as described in [13]. DNA from the positive plaques was isolated, digested and analysed by gel electrophoresis. M13mp subclones of the inserts were used to determine the nucleotide sequence of the rat βB2-crystallin RNA according to the dideoxy chain termination method [14].

RESULTS

Relative amounts of crystallin mRNA during fetal and postnatal development of the rat eye lens

To determine the pattern of expression of different crystallin genes, RNA was extracted from lenses isolated from rats at various developmental stages, Northern or dot blotted and hybridized with the appropriate cDNA probe. An example of these studies is shown in Figure 1. In this experiment RNA isolated from lenses of various fetal and postnatal developmental stages was Northern blotted and the blots were hybridized successively with a probe for the αB, βB2, βB3, βA3/A1 and βs gene transcripts. From the variation in hybridization intensity it is apparent that these five crystallin genes are differentially expressed during lens development. To quantitate the level of the transcripts

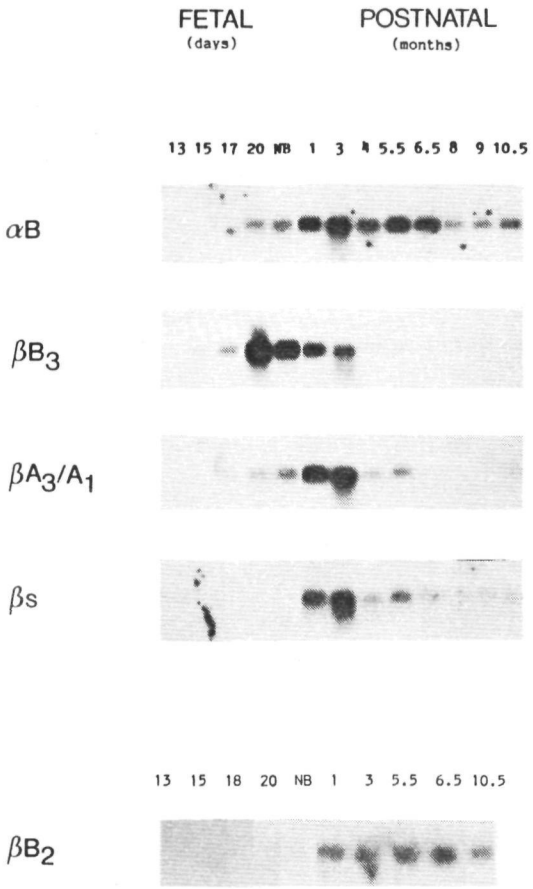


Figure 1: Northern blot analyses of total lens RNA isolated from embryonic, neonatal and postnatal rats. Autoradiograms of a Northern blot hybridized successively with the αB-, βB3-, βA3/A1- and βs-crystallin probe and an autoradiogram of a Northern blot hybridized with the βB2-crystallin probe are shown (for details see Materials and Methods). Each lane contained 1 µg of total lens RNA except lane 8 (4 months), which contained only 0.5 µg. The age of the rats from which the lenses were isolated is indicated above each lane. Fetal age is shown in days, NB indicates newborn rats and postnatal age is shown in months. The sequences used as probe are indicated on the left. Only the relevant parts of the autoradiograms are shown. For nomenclature of the β-crystallins, see legend to Table 1.

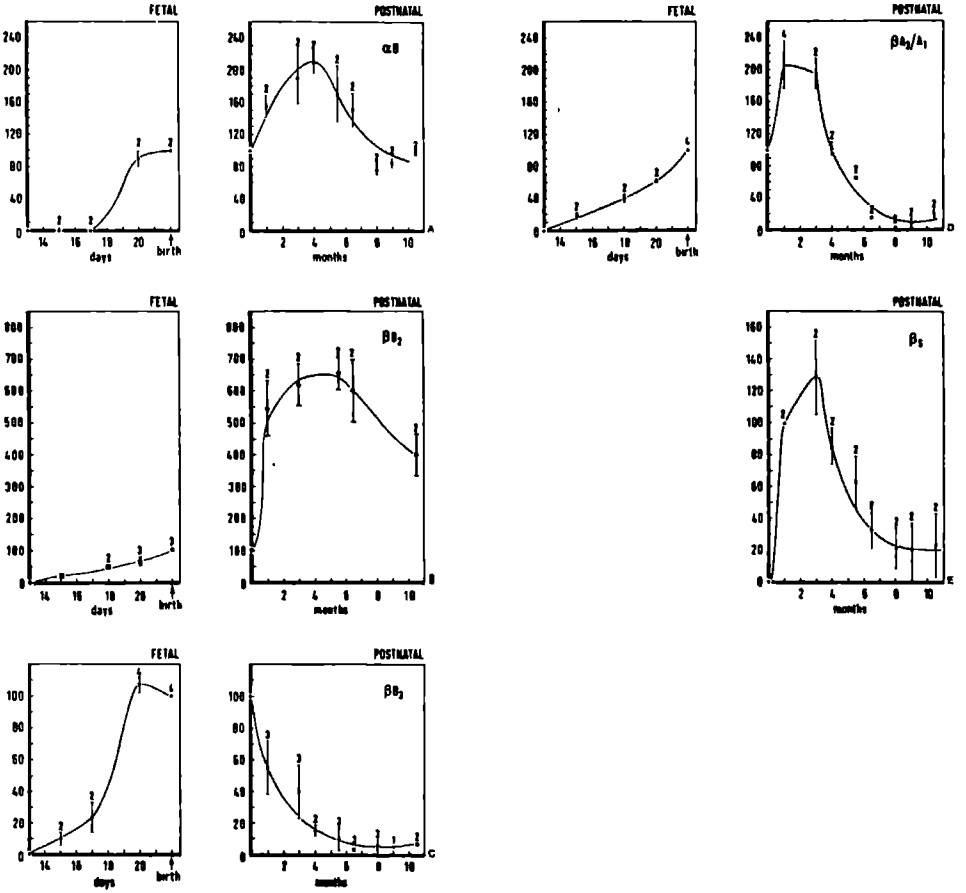


Figure 2: The level of various crystallin RNAs during development. The relative amounts of αB- (a), βB2- (b), βB3- (c), βA3/A1- (d) and βs-crystallin mRNA (e) in total lens RNA isolated from embryos and rats of different ages were determined by scanning multiple exposures of autoradiograms of Northern and dot blots with a densitometer (see Materials and Methods). The values are expressed as percentage of the values obtained from RNA isolated from lenses of newborn or 1-month old (for βs) rats. The standard deviation is shown by the bars, the numbers above these bars indicate the number of independent experiments performed.

of these five genes at different developmental stages, the amount of hybridization to Northern blots such as shown in Figure 1, or to dot blots hybridized with the same probes, was measured by densitometry (see Materials and Methods). The results of these experiments are summarized in Figure 2. Transcripts from the αB gene were first detectable during late embryogenesis and reached their maximal concentration around 4 months after birth (Figure 2a). The βB2, βB3 and βA3/A1 genes are all expressed during early fetal development, but the developmental stage at which the maximal concentration of RNA is found differs between these three β-crystallin genes. The βB3 gene is clearly an "early" β-crystallin gene and its maximal expression is already reached during late embryogenesis (Figure 2c). In contrast, the βB2 gene is a "late" β-crystallin gene. The concentration of the βB2 transcripts increases rapidly after birth and reaches a maximum between 4 and 5 months after birth (Figure 2b). The βA3/A1 gene is intermediate; the peak in the concentration of the transcripts of this gene is found between 2 and 4 months after birth (Figure 2d). The pattern of expression of the βs gene is exceptional in that transcripts of this gene cannot be detected during embryogenesis or in neonatal rats. After this silent period, however, the transcripts from this gene accumulate rapidly during the first month after birth and reach their maximal concentration shortly thereafter, namely between 2 and 3 months after birth (Figure 2e). βs RNA is still detectable, although in low amounts, in lenses of mature rats (10.5 months of age). The αB and βB2 RNAs are also abundantly present in lenses of mature rats. In contrast, expression of the βB3 and βA3/A1 genes has ceased at this stage.

Crystallin mRNA quantities in the newborn and postnatal rat lens

The data given in Figure 2 represent only relative levels. To assess the real number of transcripts per lens, the hybridization signals obtained from Northern or dot blots were compared with the signals obtained from known amounts of the corresponding cDNA clone on Southern blots (see Materials and Methods). In addition, we have extended previous studies [5] and have determined the number of transcripts of the αA, βB1 and the six γ-crystallin genes in neonatal rat lenses and at various times during postnatal development. The data obtained are

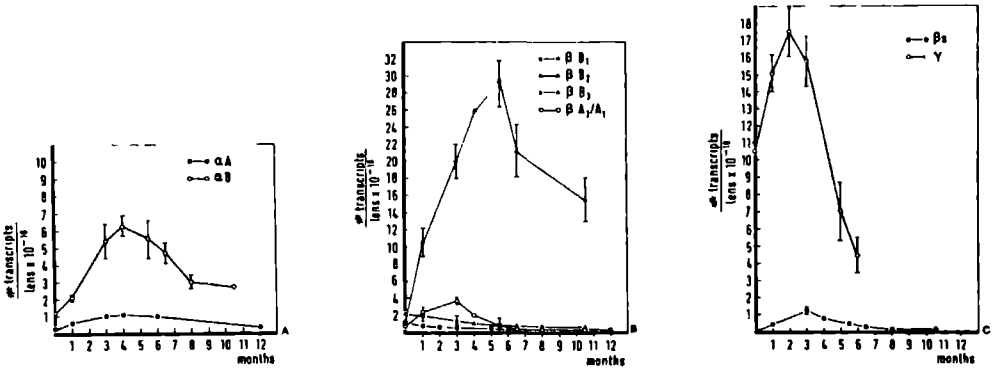


Figure 3: The number of crystallin transcripts in the rat lens. The total amounts of αA-, αB- (a), βB1-, βB2-, βB3-, βA3/A1- (b), βs- and γ-crystallin mRNA (c) in the neonatal and postnatal rat lens were determined as described in Materials and Methods. The amount of total lens RNA during postnatal development and the relative amounts of the αA-, βB1- and γ-crystallin mRNA were taken from [5]. Standard deviations are shown. The amount of αA-crystallin mRNA was determined using the rat αA probe while the amount of αB-crystallin mRNA was measured using a calf probe. To correct for possible differences due to the use of a probe from a different species the Northern and Southern blots were also hybridized with the calf αA probe. Under our washing conditions the rat αA probe yielded a signal 15 % higher than that of the calf αA probe. The amount of αB-crystallin transcripts in rat lens was therefore calculated from the signal obtained multiplied by 1.15. The amounts of the βB2, βA3/A1 and βs transcripts were determined by using calf probes, and therefore may be underestimates.

presented in Figure 3 (note that, due to growth of the lens, the time at which the maximal number of transcripts per lens is reached may differ from the time at which the maximal concentration is reached). With respect to the transcripts which code for the α-crystallin subunits it is clear that the postnatal lens contains significantly more αB RNA then αA RNA: the ratio between the number of αB and αA transcripts varies from 3 to 5. Within the β-crystallin gene family even greater

differences in expression are found. The predominant β-crystallin RNA species in the postnatal lens are the transcripts of the βB2 gene. In the mature lens more than 90% of the β-crystallin RNA is βB2 RNA (Figure 3b). Our study of the expression of the β-crystallin genes in the rat lens is, of course, not complete: we still lack probes for the βA2 and βA4 gene transcripts. However, the results of in vitro translation studies of calf lens RNA showed that these genes are probably only expressed during fetal life [8] and the lack of data on the transcripts of these genes should therefore not affect our conclusions with respect to the β-crystallin RNA composition of postnatal lenses.

As mentioned in the previous section, βs RNA is not detectable in fetal lenses. The number of transcripts of this gene rises rapidly after birth to a maximum of 1.2×10^{10} at 3 months of age. Finally, the γ-crystallin gene family contributes about 1×10^{11} RNA molecules to the RNA population of the newborn lens (Figure 3c). Comparable numbers of γ-crystallin mRNAs were found in a study of the murine lens [6].

The in vitro translation pattern of rat lens RNA

Comparison of the absolute amount of crystallin RNA as determined above with published protein patterns shows some marked discrepancies. For example, the α-crystallin protein complex contains 3 times as much αA- as αB-crystallin [15-17], yet we find that there is between 3 and 5 times as much αB- as αA-crystallin RNA. Similarly, βB2-crystallin protein is not detectable in neonatal rodent lenses [17, 18] while βB1-crystallin is present. Nevertheless, newborn lenses contain as much βB2 RNA as βB1 RNA. To determine whether the inherent translatability of the αB- or βB2-crystallin mRNA differs from that of αA- or βB1-crystallin mRNA, rat lens RNA extracted from 1 week old rats was translated in a wheat-germ extract and the products were separated by 2D gel electrophoresis. As shown in Figure 4b, the RNA extracted from 1 week old rats did direct detectable synthesis of βB1- but not of βB2-crystallin. In agreement with previous data [10], this lens RNA also directed considerably less synthesis of αB- compared to αA-crystallin protein.

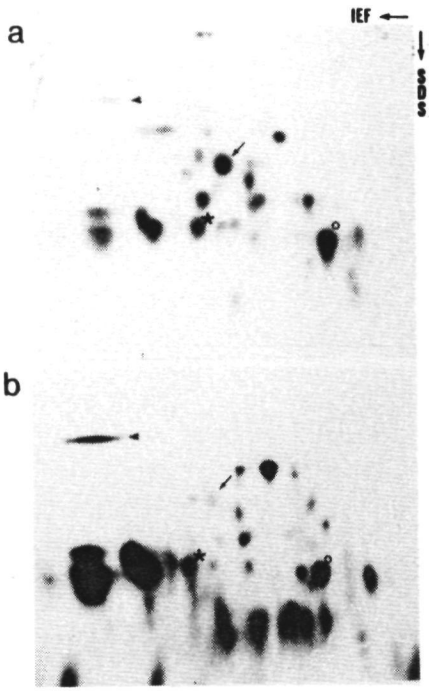


Figure 4: In vitro translation of lens RNA isolated from 1 week old rats. The [^{35}S]methionine labeled in vitro translation products were mixed with the unlabeled water-soluble proteins isolated from three months old rats and separated on a two-dimensional gel. **a** shows the Coomassie Brilliant Blue stained gels, **b** the autoradiograms of this gel. The βB1 primary protein product is indicated by an arrowhead, the βB2 protein by an arrow. The αA and αB primary protein products are indicated by a (o) and (*) respectively.

Possible sequence determinants of the translatability of αB- and βB2-crystallin mRNA

As the lower translation efficiency of the αB and βB2 mRNA deduced from the in vivo data can be mimicked in a wheat germ extract, it is likely to be the result of some feature of the RNA sequence itself. Codon usage is one of the sequence parameters that can affect translatability of a mRNA [19, 20]. We therefore compared the codon usage of the hamster αB- [8] and rat or hamster αA-crystallin genes [7, 21] but found no significant difference. To determine whether the codon usage of the rat βB2 gene is the same as that of its close relatives, the βB1 and βB3 genes, a βB2 cDNA clone was isolated and sequenced. The sequence shows an open reading frame encoding 205 amino acids starting with a ATG initiation codon at position 19 and ending with a TAA termination codon at position 634 (see Figure 5). The codon usage of this mRNA does not differ markedly from that of the βB1- and βB3-crystallin mRNAs.

Hence the relatively poor translation of the αB- and βB2-crystallin mRNAs is unlikely to be due to differences in codon usage.

The relative translation efficiency of a mRNA is also dependent on the presence of secondary structure in the 5' non-coding region [22] and on the sequence around the translation initiation codon [23, 24]. The 5' non-coding regions of the crystallin transcripts are short (30-40 nt) and cannot form stable stem-loops, hence this region of the mRNA is unlikely to be responsible for the poor translation of the αB or βB2

```
                    M  A  S  N  H  Q  T  Q  A  G  K  P  Q  P
gacaccagagagtccaccATGGCCTCAGACCACCAGACACAGGCGGGCAAGCCCCAGCCC
                                                          60
 L  N  P  K  I  I  I  F  E  Q  E  N  F  Q  G  H  S  H  E  L
CTTAACCCTAAGATCATCATCTTCGAACAGGAGAACTTCCAGGGCCATTCCCATGAGCTC
                                                         120
 R  E  P  C  P  N  L  K  E  T  G  M  E  K  A  G  S  V  L  V
AGAGAGCCCTGCCCCAACCTGAAGGAGACTGGGATGGAGAAGGCGGGCTCCGTCCTGGTG
                                                         180
 Q  A  G  P  W  V  G  Y  E  Q  A  K  C  K  G  E  Q  F  V  F
CAGGCTGGGCCGTGGGTGGGCTACGAGCAGGCTAAGTGCAAGGGAGAGCAGTTTGTGTTT
                                                         240
 E  K  G  E  Y  H  Q  W  D  S  W  T  S  S  R  S  T  H  S  L
GAGAAGGGCGAGTACCACCAGTGGGACTCCTGGACCAGCAGCCGGAGCACGCACTCCCTC
                                                         300
 S  S  L  R  P  I  K  V  D  S  Q  E  H  K  I  I  L  Y  E  N
AGTTCTCTGAGACCCATCAAAGTGGACAGCCAGGAGCACAAGATCATCTTATATGAGAAC
                                                         360
 P  N  F  T  G  K  K  M  D  E  I  V  D  D  V  P  S  F  H  A
CCCAACTTTACGGGCAAGAAGATGGATGAGATTGTCGATGACGTACCCAGCTTTCACGCT
                                                         420
 H  G  Y  Q  E  K  V  S  S  V  A  V  Q  S  G  T  W  V  G  Y
CACGGATACCAAGAGAAAGTGTCTTCCGTGCGCGTGCAGAGCGGCACGTGGGTTGGTTAT
                                                         480
 Q  Y  P  G  Y  L  G  L  Q  Y  L  L  E  K  G  D  Y  K  D  N
CAGTACCCTGGCTACCTCGGGCTGCAGTATCTGCTGGAGAAGGGGGATTACAAGGACAAC
                                                         540
 S  D  F  G  A  P  L  P  Q  V  Q  S  V  R  R  I  R  D  M  Q
AGCGACTTCGGCGCTCCCCTACCCCAGGTGCAATCTGTGCGTCGAATCCGTGACATGCAG
                                                         600
 W  Q  H  R  G  A  F  H  P  S  T  *
TGGCAGCACCGGGGCGCCTTCCACCCCTCCACGTAAagccctgcccgtccctccttcccca
                                                         660
ggggctggccatcccccagggtcctcctgacatccagagtgAATAAAgtgtgacttgca
                                                         720
acttgaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
                         750
```

Figure 5: Nucleotide and deduced amino acid sequence of the rat βB2 cDNA clone λβB2. The derived amino acid sequence is given above the nucleotide sequence. The coding sequence and the poly (A) addition signal are shown in capital letters.

mRNA. Inspection of the translation initiation region of the crystallin mRNAs shows that the same five nucleotides (CCACC) precede the AUG of the αB- and the βB2-crystallin RNA (Table 1). As these are the only two crystallin mRNAs that have this sequence motif, this observation points to the possibility that these two mRNAs have a lower efficiency of translation initiation than their close relatives. Curiously, only the αB and βB2 RNAs match the Kozak consensus initiation sequence perfectly ([23] see Table 1). Hence this consensus sequence may well not be optimal for crystallin mRNAs. The sequence compilation shown in Table 1 suggests that the C at position -4 in the consensus sequence should be replaced by a purine. Site directed mutagenesis experiments are required to show that such a change would indeed improve the translation efficiency of the αB- and βB2-crystallin mRNAs.

TABLE 1

I	II	III	IV
αA	hamster	agaagaacaugg	[21]
αB	hamster	uagccaccaugg	[8]
βB1	rat	cagaaaccaugu	[9]
βB2	rat	agtccaccaugg	Fig. 5
βB3	rat	ccggagacaugg	[9]
βA3	mouse	accagaugg	[25]
βA1	mouse	ccaccaagaugg	[25]
βs	calf	caccaaacaugu	[11]
γA	rat	caacaacgaugg	[26]
γB	rat	ucagcgagaugg	[26]
γC	rat	agccagccaugg	[26]
γD	rat	caacagccaugg	[26]
γE	rat	caacagccaugg	[26]
γF	rat	caccagccaugg	[26]
	consensus	ccaccaugg	[23]

Table 1: Comparison of the sequences flanking the AUG initiation codon of the various crystallin mRNAs. I, mRNA; II, species; III, sequence; IV, reference(s). The sequence region compared with the consensus sequence is underlined. Note that we have used the nomenclature suggested by [27] for the β-crystallins and that suggested by [28] for the γ-crystallins.

DISCUSSION

Cytological studies of the expression of the crystallins during the morphogenesis of the rat lens have shown that αA RNA and α-crystallin protein can be detected in the lens vesicle, before formation of

the primary fiber cells [29, 30]. Our results show that the αB transcripts can only be detected at a much later stage of development, i.e. after formation of the primary fiber cells. The α-crystallin synthesized during early development of the rat lens must thus consist of αA chains only. In the calf, more αB than αA RNA is detected in the central epithelial region, while in the equatorial region more αA than αB RNA is found (as assayed by in vitro translation; for review see [1]). These studies suggested that, at least in the calf, the αB-crystallin gene is the first crystallin gene to be activated during lens cell differentiation. In the light of our results, it would be worthwhile to reexamine this question and to determine whether the late accumulation of αB RNA with respect to αA RNA is a property of the lens fiber cells only, or whether the αB gene is also inactive in early embryonic lens epithelial cells.

Transcriptional activation of all β-crystallin genes (with the exception of the βs gene) apparently occurs during early lens embryogenesis and the distinction between "early" (e.g. βB3) and "late" (e.g. βB2) genes within this family appears to be the result of a differential shutdown of the expression of these genes. The same phenomenon has been observed in a study of the expression of the six γ-crystallin genes during rat lens development [5]. The evolutionary relationship between these two gene families might thus well be reflected not only by their common protein structure but also by a common mechanism of gene activation. Experiments are in progress to show whether the tissue specific activation of the expression of the β-and γ-crystallin genes is indeed steered by the same cellular factor(s).

The chromosomal organization of the basic β-crystallin genes has been partially elucidated in the rat. We have shown that the "early" βB3 gene is closely linked to the "late" βB2 gene [31], while the other "early" basic β-crystallin gene, the βB1 gene, is located at least 30 kb away from these genes (and may even be not linked to these genes at all [13]). Thus the developmental expression pattern of the basic β-crystallin genes does not correlate with their chromosomal organization but may rather reflect the more distant evolutionary relationship between the βB2 gene on the one hand and the βB1 and βB3 genes on the other hand [11].

Evidence for regulation of crystallin synthesis at the post-transcriptional level is somewhat incidental and derives mainly from studies

of the chicken lens. For example, in the one-day-old chick lens not all of the β-crystallin RNA is found associated with polysomes [32]. Also, the synthesis of (the bird and reptile specific) δ-crystallin in the chicken epithelial lens cells decreases during embryonic development even though the amount of δ-crystallin RNA remains the same [33, 34]. The discrepancies noted here between our RNA data and previous protein data (or the synthetic pattern) are quantitative rather than qualitative. Hence we find no evidence for the occurrence of post-transcriptional regulation at the translational level in the rat lens fiber cells, which synthesize the bulk of the crystallin RNA and protein. Whether post-transcriptional regulation occurs at other levels, for example changes in mRNA stability could be involved in the differential decrease in the level of the expression of the crystallin genes, is at present under investigation.

ACKNOWLEDGEMENTS We thank Marja Timmermans for help with the lens RNA isolations, Jacques Jansen for donating the rat lens cDNA library, and Kees van Roosendaal and Rob van Leen for performing the *in vitro* translations. This investigation was carried out under the auspices of the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (NWO).

REFERENCES

1. Bloemendal, H. (1985) Exp. Eye Res. 41, 429-448.
2. Wistow, G. & Piatigorsky, J. (1988) Ann. Rev. Biochem. 57, 479-504.
3. Ostrer, H., Beebe, D.C. & Piatigorsky, J. (1981) Dev. Biol. 86, 403-408.
4. Hejtmancik, J.F., Beebe, D.C., Ostrer, H., & Piatigorsky, J. (1985) Dev. Biol. 109, 72-81.
5. Van Leen, R.W., Van Roozendaal, K.E.P., Lubsen, N.H. & Schoenmakers, J.G.G. (1987) Dev. Biol. 120, 457-464.
6. Murer-Orlando, M., Paterson, R.C., Lok, S., Tsui, L.-C. & Breitman, M.L. (1987) Dev. Biol. 119, 260-267.
7. Moormann, R.J.M., Van der Velden, H.M.W., Dodemont, H.J., Andreoli, P.M., Bloemendal, H. & Schoenmakers, J.G.G. (1981) Nucl. Acids Res. 9, 4813-4822.
8. Quax-Jeuken, Y., Quax, W., Van Rens, G., Meera Khan, P. & Bloemendal, H. (1985) Proc. Natl. Acad. Sci. USA 82, 5819-5823.
9. Den Dunnen, J.T., Moormann, R.J.M. & Schoenmakers, J.G.G. (1985) Biochim. Biophys. Acta 824, 295-303.
10. Quax-Jeuken, Y., Janssen, G., Quax, W., Van den Heuvel, R. & Bloemendal, H. (1984) J. Mol. Biol. 180, 457-472.
11. Quax-Jeuken, Y., Driessen, H., Leunissen, J., Quax, W., De Jong, W. & Bloemendal, H. (1985) EMBO J. 4, 2597-2602.
12. Moormann, R.J.M., Den Dunnen, J.T., Bloemendal, H. & Schoenmakers,

J.G.G. (1982) Proc. Natl. Acad. Sci. USA 79, 6876-6880.
13. Moormann, R.J.M., Jongbloed, R. and Schoenmakers, J.G.G. (1984) Gene 29, 1-9.
14. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
15. Delcour, J. & Papaconstantinou, J. (1974) Biochem. Biophys. Res. Comm. 57, 134-141.
16. Uchiumi, T., Kimura, S. & Ogata, K. (1983) Exp. Eye Res. 36, 125-134.
17. Carper, D., Russel, P., Shinohara, T. & Kinoshita, J.H. (1985) Exp. Eye Res. 40, 85-94.
18. Carper, D., Smith-Gill, S.J. & Kinoshita, J.H. (1986) Dev. Biol. 113, 104-109.
19. Bulmer, M. (1987) Nature, 325, 728-730.
20. Sharp, P.M. & Li, W.-H. (1987) Nucl. Acids Res. 15, 1281-1295.
21. Van den Heuvel, R., Hendriks, W., Quax, W. & Bloemendal, H. (1985) J. Mol. Biol. 185, 273-284.
22. Pelletier, J. & Sonenberg, N. (1985) Cell 40, 515-526.
23. Kozak, M. (1986) Cell 44, 283-292.
24. Cavener, D.R. (1987) Nucl. Acids Res. 15, 1353-1361.
25. Peterson, C.A. & Piatigorsky, J. (1986). Gene 45, 139-147.
26. Den Dunnen, J.T., Moormann, R.J.M., Lubsen, N.H. & Schoenmakers, J.G.G., (1986) J. Mol. Biol. 189, 37-46.
27. Berbers, G.A.M., Hoekman, W.A., Bloemendal, H., De Jong, W.W., Kleinschmidt, T. & Braunitzer, G. (1984) Eur. J. Biochem. 139, 467-479.
28. Aarts, H.J.M., Den Dunnen, J.T., Leunissen, J., Lubsen, N.H. & Schoenmakers, J.G.G. (1988) J. Mol. Evol. 27, 163-172.
29. McAvoy, J.W. (1978) J. Embryol. Exp. Morph. 45, 271-281.
30. Van Leen, R.W., Breuer, M.L., Lubsen, N.H. & Schoenmakers, J.G.G. (1987) Dev. Biol. 123, 338-345.
31. Aarts, H.J.M., Den Dunnen, J.T., Lubsen, N.H. & Schoenmakers, J.G.G. (1987) Gene 59, 127-135.
32. Thomson, I., Wilkinson, C.E., Jackson, J.F., De Pomerai, D.I., Clayton, R.M., Truman, D.E.S. & Williamson, R. (1978) Dev. Biol. 65, 372-382.
33. Beebe, D.C. & Piatigorsky, J. (1977) Dev. Biol. 59, 174-182.
34. Beebe, D.C. & Piatigorsky, J. (1981) Dev. Biol. 84, 96-101.

CHAPTER 5

THE HUMAN γ-CRYSTALLIN GENES: A FAMILY ON ITS WAY TO EXTINCTION?

THE HUMAN γ-CRYSTALLIN GENES: A FAMILY ON ITS WAY TO EXTINCTION?

ABSTRACT

In the human lens transcripts from the γ-crystallin gene family could be detected from the γC, γD and pseudoγE genes, no transcript from the γA gene could be found. The presence of transcripts from the γB-crystallin gene could not be shown beyond any doubt. Within the human fetal lens the onset of the γ-crystallin transcription during development is delayed relative to that of the αA-, βB2- and βB3-crystallin genes. The composition of the γ-crystallin mRNA pool remains the same till after birth, then a differential shut off is observed, leaving only the γD transcripts at 10 years of age. Our data, together with previously published data, suggests that the predominant synthesis of only the γC and γD proteins is due to poor transcription of the γA- and probably also of the γB crystallin gene.

INTRODUCTION

The optical properties of the vertebrate lens are thought to be due to the abundant presence of lens-specific water-soluble structural proteins, the crystallins. The short-range order interaction between the crystallins is held responsible for the transparency of the lens (Benedek, 1971; Delaye and Tardieu, 1983), while the complex gradient of the crystallins, which is laid down during the growth and development of the lens, determines the curve of the refraction index in the lens (Philipson, 1969; Sun et al., 1984). The major mammalian crystallins are encoded by three gene families, the α-, the β- and the γ-crystallin gene families (for review see Piatigorsky, 1984; Bloemendal, 1985; Lubsen et al., 1988). From the distribution of the crystallins within the lens (Ramaekers et al., 1982; Uchiumi et al., 1983; Thomson and Augusteyn, 1985) it is to be expected that the expression of the members of these three gene families is developmentally regulated. Indeed, studies of the developmental pattern of expression of these three gene families in the rat (Van Leen et al., 1987a; Chapter 4) and of the β-crystallin gene family in the chicken (Ostrer et al., 1981; Hejtmancik et al., 1985) have shown that each stage of development of the lens is characterized by a specific profile of crystallin gene expression. This profile of expression might be in part species specific, as the optical properties of the lenses of the diverse vertebrate species differ widely (De Jong, 1981; Maisel, 1985). In this respect, the expression of the γ-crystallins is of special interest, since the presence of these proteins appears to correlate with a low water con-

tent of the lens (Summers et al., 1984). It is therefore of importance to compare the structure and expression of this gene family between two mammalian species which have a marked difference in the water content of the lens, e.g. between rat (low water content) and man (high water content). The similarities and differences in the expression of the γ-crystallin family between these two species may reveal the underlaying generality in the mechanisms regulating the developmental expression of these genes and may show how these are adapted to the specific needs of the various species.

The structural organization of the γ-crystallin gene family in rat and man has been almost completely characterized. The rat contains six γ-crystallin genes, of which five are closely linked (Moormann et al., 1985). All these rat genes are functional (Moormann et al., 1985), and are expressed during fetal development. They are differentially shut off during postnatal development (Van Leen et al., 1987a). From the human genome six γ-crystallin genes have been characterized thus far (Den Dunnen et al., 1985a; Meakin et al., 1985, 1987). These genes are orthologous to the six rat genes and are organized in the same manner (Aarts et al., 1988). Of the six human genes known, two are pseudogenes, as they contain an in frame stop codon (Meakin et al., 1985). Meakin et al. (1987) have shown that three of the other human genes have an active promotor. We show here that the fourth gene, the γB gene, is also active. We further show that, as in the rat (Van Leen, 1987a), the members of the human γ-crystallin gene family are differentially expressed during development. Finally, our data suggest that different molecular mechanisms are responsible for the functional suppression of four of the six human γ-crystallin genes, leaving only the γC- and γD-crystallin genes as producers of abundant crystallins.

MATERIALS AND METHODS.

RNA isolation, Northern blotting and probes

Lenses were homogenized in 1 ml lysis buffer (0.1 M Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA, 0.2% SDS; see Maniatis et al., 1982). The homogenate was extracted twice with phenol/chloroform/isoamylalcohol (25:24:1 v:v:v). The nucleic acids were precipitated with ethanol by centrifugation. The pellet was dissolved in water and the precipitation step was repeated. For Northern blot analysis lens RNA was separated on a 1% agarose-formaldehyde gel. After electrophoresis the gel was soaked in 20xSSC for 5 min and the RNA was blotted to nitrocellulose (Schleicher & Schull).

The construction and synthesis of the αA-, αB-, βB1- and γ-crystallin SP6-probes has been described by Van Leen et al. (1987b). The

βB2-, βB3-, βA3/A1- and βs-crystallin probes were obtained by radiolabeling the inserts of the corresponding cDNA clones (Quax-Jeuken et al., 1984; Den Dunnen et al., 1985b) with α-^{32}P-dATP through nick-translation.

RNAse H treatment

Reaction conditions were as described by Van Leen et al. (1987a).

Primer extension and S1-nuclease mapping

Annealing of the oligonucleotide to the RNA and the primer extension reaction were performed according to Geliebter et al. (1986). The oligonucleotide used, prγex12 (5'-GAAGGTGATCTT(C/T)CCCAT- 3', 18 mer), has been described earlier by Van Leen et al. (1986) and is complementary to the last 9 nucleotides of the coding region of the first exons and the first 9 nucleotides of the second exons of the human γA-, γB-, γC- and γF-crystallin genes, and with one mismatch (at nucleotide 7 of the second exon) complementary to the same region of the human γD- and γE-crystallin genes.

M13mp subclones of the γB-, γC-, γD- and γE-crystallin genes were constructed and the continuously labeled extension products from these clones were used as probe in a S1-nuclease protection assay to detect the corresponding crystallin transcripts in 9 weeks old fetal lens RNA. End-labeled fragments of the γA- and γE-crystallin genes were also used as probe in the S1-nuclease protection assays. S1-nuclease reactions were carried out as described (Moormann et al., 1985).

RESULTS

The transcriptional activity of the human γ-crystallin genes

Of the six human γ-crystallin genes, two contain an in frame stop codon and cannot give rise to a functional protein. The other four genes, the γA, γB, γC and γD genes, encode a conserved amino acid sequence (Aarts et al., 1988) and are flanked by the signals known to be required for the expression of eukaryotic genes. Yet in human lens only the protein products from the γC and γD genes could be detected (Siezen et al., 1987). With repect to the γB gene it was therefore suggested that this gene was inactivated by the insertion of an Alu repeat directly upstream of the gene (Den Dunnen et al., 1989). To obtain direct evidence for the activity of the γA-, γB-, γC-, and γD-crystallin genes in the human lens, the presence or absence of transcripts of each of these genes in fetal human lens RNA was assayed by S1-nuclease mapping. We also looked for transcripts of the pseudoγE-crystallin gene as this gene has an apparently normal TATA box region (although its promoter has been shown to be inactive in chicken lens epithelial cells; Meakin et al., 1987). The probes used in these experiments are outlined in Figure 1, and consisted either of continuously labeled M13 probes or of end-labeled restriction fragments. The probes corresponding to the

transcripts of the γB-, γC-, and γD-crystallin gene all yielded the expected protected fragment from the second exon (see Figure 2a), showing that transcripts from these genes are abundant in fetal human lenses. However S1-nuclease protection experiments carried out with a M13 probe containing part of the third exon (the more variable part of the gene) of the human γB- crystallin gene yielded only a very weak signal (R. Brakenhoff; pers. comm.), suggesting that there are few γB-crystallin transcripts present in the human fetal lens. It is possible, because of the high sequence similarity between the second exon of the γB and the second exon of the γC gene, that the fragment obtained with the 5' γB-crystallin probe (see Figures 1 and 2a) is protected by γC instead of γB transcripts. We did not detect a transcript from the γA-crystallin gene (Figure 2b). However, our failure to detect a transcript of this gene may be due to the relative insensitivity of the S1-nuclease protection assay. Meakin et al. (1987) have provided evidence that this gene does have an active promotor. Unexpectedly, we also found a protected fragment using a second exon probe for the pseudoγE-crystallin gene. As cross-hybridization of the second exon probe with transcripts from the γC and γD gene cannot be completely excluded we repeated the S1-nuclease experiment using an end-labeled fragment containing part of the third exon of this gene. A fragment of 260 nucleotides was obtained. A fragment of this size is expected if the splice acceptor site of the third exon has been properly used (Figure 2b). Further mapping of the γE transcript is required to show that this gene indeed yields a properly initiated and spliced transcript.

The differential expression of the γ-crystallin genes during development

The sequence similarity between the γ-crystallin genes is too high to allow discrimination between the transcripts from individual genes in Northern blotting experiments. As the amount of human lens material is very limited, we used primer extension rather than S1-nuclease mapping experiments to detect changes in the composition of the γ-crystallin RNA pool during development. The primer used is complementary to the first 18 nucleotides of the translated region of all γ-crystallin genes, i.e. to the last 9 nucleotides of the first exon and the first 9 nucleotides of the second exon. Since the 5' non-translated regions of the genes differ in length, the primer extended products from the tran-

scripts of each of the genes also differ in length. In these experiments a complex pattern is obtained (see Figure 3a). This complexity is most probably caused by the occurrence of multiple transcription initiation sites similar to what is observed in rat lenses (Van Leen et al., 1986). From the putative transcription initiation sites of the human genes (Figure 3b), assigned by analogy to the rat genes (Van Leen et al., 1986), we suggest that the large primer extension products

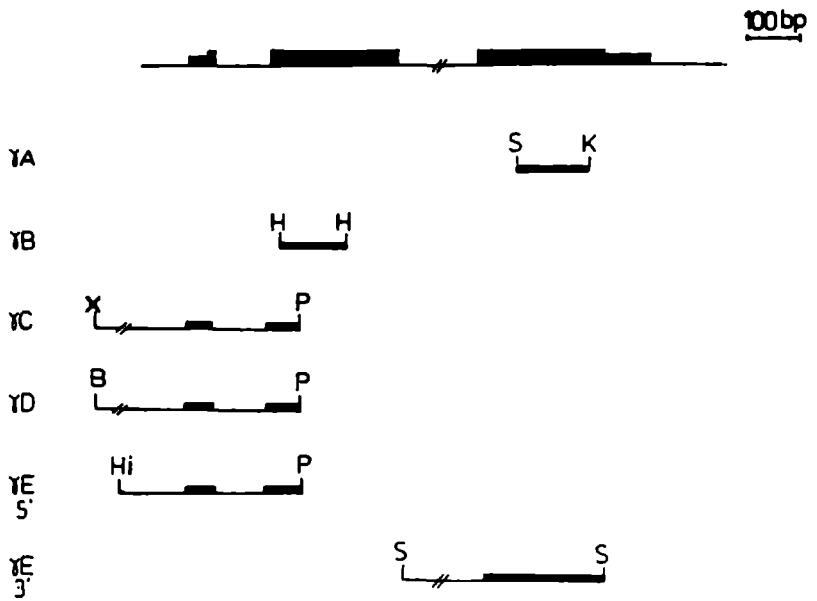


Figure 1: Schematic representation of the gene fragments which were used as probe in the S1-nuclease protection analysis. Restriction fragments from the γB, γC, γD and 5' region of the γE gene were subcloned in M13mp vectors and used as templates for the preparation of uniformily labeled probes. These probes only detect second exon sequences. The γA and 3' γE fragments were end-labeled. The probes were used in the S1-nuclease protection experiments of which the results are shown in Figure 2. Restriction sites are indicated as: B=BamHI; H=HaeIII; Hi= HindIII; K=KpnI; P=PstI; S=SstI; X=XbaI

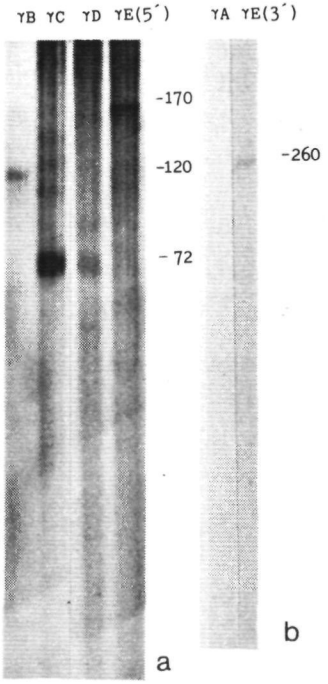


Figure 2: S1-nuclease analysis of the γ-crystallin transcripts isolated from 9 weeks old human fetal lenses. The mRNAs were analysed using continuously labeled M13 probes **(a)** or end-labeled restriction fragments **(b)** (see Figure 1). The products were analysed on a 9% polyacrylamide sequencing gel. The lenghts (in nucleotides) of the S1-nuclease protected fragments are indicated.

(around 65 nt) derive from the transcript of the γD-crystallin gene while the smaller products (49-55 nt) probably represent transcripts of the other active γ-crystallin genes.

As shown in Figure 3a, no changes in the primer extension pattern are seen with age up to birth. Between 2 days and 22 months after birth, a loss of the smaller primer extension products is found, and at 10 years of age only the 65 nt product presumably deriving from the γD-crystallin gene remains. Hence in the human, as well as in the rat lens all γ-crystallin genes are active during fetal development but differential shut off occurs after birth. However, it is the γD-crystallin gene that remains active in man and the γB-crystallin gene that remains active in the rat (Van Leen et al., 1987a).

The onset of the transcription of the human γ-crystallin genes is somewhat delayed relative to that of the αA- and βB-crystallin genes. At the earliest stage examined, 6 weeks of fetal development, no transcripts from the γ-crystallin genes could be detected, although transcripts from the αA-, βB2- and βB3-crystallin genes are already present (Figure 4a). The expression of the αB-, βA3/A1-, βB1- and βs-crystallin genes was not followed in time. Transcripts of these genes are all present in the 9 weeks old fetal human lens (Figure 4d, e, f, g). We

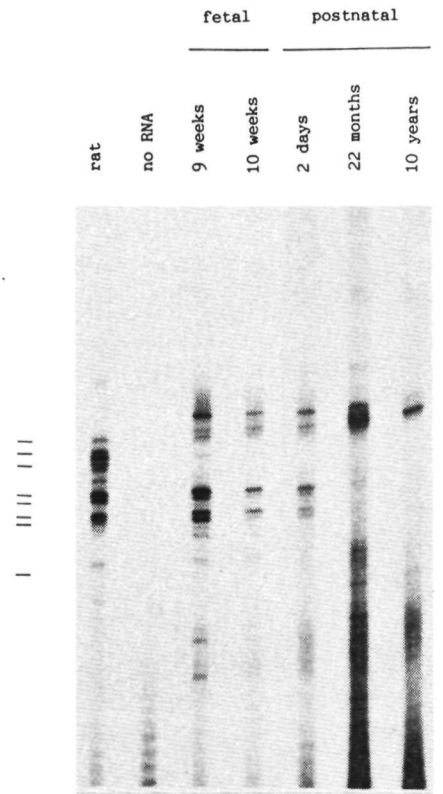


Figure 3: (a): γ-crystallin primer extension patterns obtained with RNA from human lenses at various developmental stages. The source of the RNA is indicated at the top of the figure, where fetal and postnatal refer to human lenses from the corresponding developmental stage. Rat lens RNA was isolated from newborn rats. Oligonucleotide prγex12 (see Materials and Methods) was used as primer. The primer extension products were analysed on a 9% polyacrylamide gel. Exposure times were 16 hours for the part containing the rat and human fetal lens RNA primer extension products and 72 hours for the part containing the human postnatal lens RNA primer extension products. The bars at the left represent the rat extension products, from top to bottom 63, 60, 58, 54, 53, 51, 49 and 43 nucleotides long (Van Leen et al., 1986).

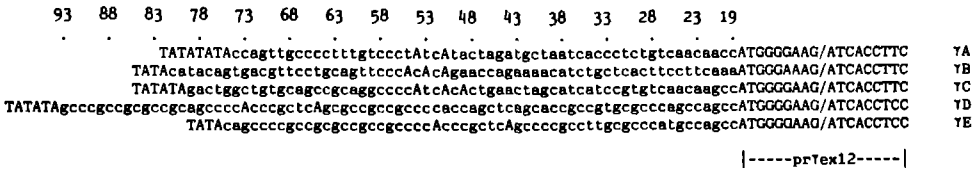


Figure 3: (b): Probable transcription initiation sites of five human γ-crystallin genes. The TATA box, the coding region complementary to the oligonucleotide prγex12 and the probable transcription initiation sites are given in capital characters. The transcription initition sites were assigned by analogy to the orthologous rat genes (Van Leen et al., 1986). The numbering starts with 19 from the first nucleotide at the 5' end of the region complementary to prγex12.

noted a broadening of the γ-crystallin RNA band from lenses between 12 and 13 weeks and therefore we examined the mobility of the γ-crystallin RNA more carefully. These experiments showed that lenses of 13 weeks and older contained two size classes of γ-crystallin RNA (Figure 4b). In view of the primer extension experiments reported above, which showed that the composition of the γ-crystallin RNA pool does not change markedly during this time, we suspected that the higher mobility class represented γ-crystallin transcripts with a shorter poly-A tail. To show that this is the case, lens RNA was hybridized with oligo-dT, treated with RNAse H, Northern blotted and hybridized with a γ-crystallin probe. After this treatment, only a single size class was detected (Figure 4c) and we therefore conclude that a class of γ-crystallin RNA with a shorter poly-A tail appears around 13 weeks of fetal development.

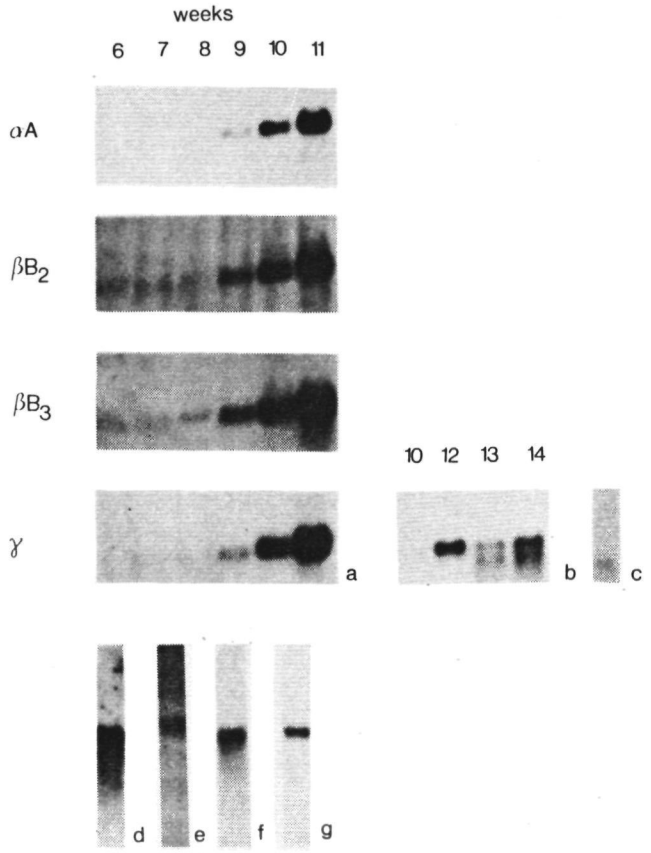


Figure 4: Northern blot analysis of the crystallin transcripts in human lenses. **(a):** Autoradiograms of Northern blots containing RNA isolated from human fetal lenses of different ages. The blot was hybridized successively with an αA-, a βB2-, a βB3- and a γ-crystallin probe. Each panel contains one quarter of the total RNA content of one lens. **(b):** Autoradiogram of a Northern blot containing RNA isolated from 10, 12, 13 and 14 weeks old fetal lenses hybridized with the γ-crystallin probe. **(c):** Analysis of the 3' end of the human γ-crystallin transcripts. One µg of 14 weeks old human fetal total lens RNA was hybridized with oligo-dT and digested with RNase H. The RNA was separated on a 1% agarose-formaldehyde gel, blotted to nitrocellulose and hybridized with the γ-crystallin probe. **(d, e, f, g):** Autoradiograms of a Northern blot containing 2 µg of 9 weeks old human fetal total lens RNA hybridized with an αB- **(d)**, a βB1- **(e)**, a βA3/A1- **(f)** or a βs crystallin probe **(g)**.

DISCUSSION

Evolutionary analysis has suggested that the present six membered γ-crystallin gene family already existed in the common ancestor of man and rat. However, when we compare the expression of this orthologous set of genes between man and rat, important differences emerge. First of all, only four of the human γ-crystallin genes can give rise to a functional protein product (Den Dunnen et al., 1985a; Meakin et al., 1985, 1987), whereas in rat all six genes are active (Moormann et al., 1985). Secondly, in two months old human lenses the γC- and γD-crystallin proteins comprise at least 90% of the γ-crystallins (Siezen et al., 1987), while in the rat all genes contribute significantly to the γ-crystallin protein pool (Siezen et al., 1988). Our data show that the observed differences in the level of expression among the human γ-crystallin genes is likely to be due mainly to transcriptional effects. The very low amount of the presumed protein product of the γA-crystallin gene detected in the human lens (Russell et al., 1987) is probably the result of a low rate of transcription as we could not detect a transcript of this gene. Similarly the γB-crystallin gene yields only little transcript.

The secondary loss of γ-crystallin in higher mammals with lenses with a high water content can be attributed, according to our data, to at least two distinct molecular mechanisms: inactivation of the coding region in case of the human psuedoγE and psuedoγF genes and poor transcription (or rapid mRNA degradation) in case of the human γA and (probably also) the human γB gene. Clearly, during evolution of the primate lineage the expression of four of the six γ-crystallin genes has been shut-off or turned down, resulting in a lower abundance of γ-crystallin in the human lens.

REFERENCES

Aarts, H.J.M., Den Dunnen, J.T., Leunissen, J., Lubsen, N.H., and Schoenmakers, J.G.G. (1988). The γ-crystallin gene family: sequence and evolutionary patterns. J. Mol. Evol. 27, 163-172.

Bloemendal, H. (1985). Lens research: from protein to gene. Exp. Eye Res. 41, 429-448.

Benedek, G.B. (1971). Theory of transparency of the eye. Appl. Opt. 10, 459-473.

Delaye, M., and Tardieu, A. (1983). Short-range order of crystallin proteins accounts for eye lens transparency. Nature 302, 415-417.

Den Dunnen, J.T., Moormann, R.J.M., Cremers, F.P.M., and Schoenmakers, J.G.G. (1985a). Two human γ-crystallin genes are linked and ridd-

led with Alu-repeats. Gene 38, 197-204.
Den Dunnen, J.T., Moormann, R.J.M., and Schoenmakers, J.G.G. (1985b). Rat lens β-crystallin genes are internally duplicated and homologous to γ-crystallins. Biochim. Biophys. Acta 824, 295-303.
Den Dunnen J.T., Van Neck, J.W., Cremers, F.P.M., Lubsen, N.H., and Schoenmakers, J.G.G. (1989). Nucleotide sequence of the rat γ-crystallin gene region and comparison with an orthologous human region. Gene, in press.
De Jong, W.W. (1981). Evolution of lens and crystallins. In: Bloemendal, H. (ed). Molecular and cellular biology of the eye lens. John Wiley and Sons, New York, pp 221-278.
Geliebter, J., Zeff, R.A., Melvold, R.W., and Nathenson, S.G. (1986). Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis: K^{bm9} and K^{bm6}. Proc. Natl. Acad. Sci. U.S.A. 83, 3371-3375.
Hejtmancik, J.F., Beebe, D.C., Ostrer, H., and Piatigorsky, J. (1985). δ- and β-crystallin mRNA levels in the embryonic and posthatched chicken lens: temporal and spatial changes during development. Dev. Biol. 109, 72-81.
Lubsen, N.H., Aarts, H.J.M., and Schoenmakers, J.G.G. (1988). The evolution of lenticular proteins: the β- and γ-crystallin super gene family. Prog. Biophys. Molec. Biol. 51, 47-76.
Maisel, H. (1985). The Ocular Lens. Marcel Dekker, New York.
Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning, a laboratory Manual. Cold Spring Harbor Laboratory.
Meakin, S.O., Breitman, M.L., and Tsui, L.-C. (1985). Structural and evolutionary relationships among five members of the human γ-crystallin gene family. Mol. Cell. Biol. 5, 1408-1414.
Meakin, S.O., Du, P.R., Tsui, L.-C., and Breitman, M.L. (1987). γ-crystallins of the human eye lens: expression analysis of five members of the gene family. Mol. Cell. Biol. 7, 2671-2679.
Moormann, R.J.M., Den Dunnen, J.T., Heuyerjans, J., Jongbloed, R.J.E., Van Leen, R.W., Lubsen, N.H., and Schoenmakers, J.G.G. (1985). Characterization of the rat γ-crystallin gene family and its expression in the eye lens. J. Mol. Biol. 182, 419-430.
Ostrer, H., Beebe, D.C., and Piatigorsky, J. (1981). β-crystallin mRNAs: differential distribution in the developing chicken lens. Dev. Biol. 86, 403-408.
Philipson, B. (1969). Distribution of protein within the normal rat lens. Invest. Opht. 8, 258-270.
Piatigorsky, J. (1984). Lens crystallins and their gene families. Cell 38, 620-621.
Quax-Jeuken, Y., Janssen, G., Quax, W., Van den Heuvel, R., and Bloemendal, H. (1984). Bovine β-crystallin complementary DNA clones. Alternating proline/alanine sequence of the βB1 subunit originates from a repetitive DNA sequence. J. Mol. Biol. 180, 457-472.
Ramaekers, F., Dodemont, H., Vorstenbosch, P., and Bloemendal, H. (1982). Classification of rat lens crystallins and identification of proteins encoded by rat lens mRNA. Eur. J. Biochem. 128, 503-508.
Russell, P., Meakin, S.O., Hohman, T.C., Tsui, L.-C., and Breitman, M.L. (1987). Relationship between proteins encoded by three human γ-crystallin genes and distinct polypeptides in the eye lens. Mol. Cell. Biol. 7, 3320-3323.
Siezen, R.J., Thomson, J.A., Kaplan, E.D., and Benedek, G.B. (1987).

Human lens γ-crystallins: Isolation, identification, and characterization of the expressed gene products. Proc. Natl. Acad. Sci. U.S.A. 84, 6088-6092.

Siezen, R.J., Wu, E., Kaplan, E.D., Thomson, J.A., and Benedek, G.B. (1988). Rat lens γ-crystallins. Characterization of the six gene products and their spatial and temporal distribution resulting from differential synthesis. J. Mol. Biol. 199, 475-490.

Sun, S.-T., Tanaka, T., Nishio, I., Peetermans, J., Maizel, J.V. jr., and Piatigorsky, J. (1984). Direct observation of δ-crystallin accumulation by laser light scattering spectroscopy in the chicken embryo lens. Proc. Natl. Acad. Sci. U.S.A. 81, 785-787.

Summers, L., Slingsby, C., White, H., Narebor, M., Moss, D., Miller, L., Mahadevan, D., Lindley, P., Driessen, H., and Blundell, T. (1984). The molecular structure and interactions of bovine and human γ-crystallins. In: Nugent, J., and Whelan, J. (eds). Human Cataract Formation. Pitman, London. CIBA Foundation symposia 106, pp 219-230.

Thomson, J.A., and Augusteyn, R.C., (1985). Ontogeny of human lens crystallins. Exp. Eye Res. 40, 393-410.

Uchiumi, T., Kimura, S., and Ogata, K. (1983). Protein composition of the rat lens and their age-related changes observed with two dimensional polyacrylamide gel electrophoresis. Exp. Eye Res. 36, 125-134.

Van Leen, R.W., Kastrop, P.M.M., Van Roozendaal, K.E.P., and Schoenmakers, J.G.G. (1986). Sequence divergence and selection of cap sites in the rat γ-crystallin gene family. Eur. J. Biochem. 157, 203-208.

Van Leen, R. W., Van Roozendaal, K.E.P., Lubsen, N.H., and Schoenmakers, J.G.G. (1987a). Differential expression of crystallin genes during development of the rat eye lens. Dev. Biol. 120, 457-464.

Van Leen, R.W., Breuer, M.L., Lubsen, N.H., and Schoenmakers, J.G.G. (1987b). Developmental expression of crystallin genes: in situ hybridization reveals a differential localization of specific mRNAs. Dev. Biol. 123, 338-345.

CHAPTER 6

THE HUMAN βB2-2-CRYSTALLIN GENE IS A PSEUDOGENE

Henk J.M. Aarts, Frank H.J. Schuren, Nicolette H. Lubsen

and John G.G. Schoenmakers.

Submitted for publication

THE HUMAN βB2-2-CRYSTALLIN GENE IS A PSEUDOGENE.

SUMMARY

Comparison of the partial sequences of the human βB2-1- and βB2-2-crystallin genes with orthologous rat or calf sequences shows that the fourth exon of the human βB2-2 gene contains an one triplet deletion and a mutated 5' splice acceptor site. Whereas transcripts of the βB2-1-crystallin gene could be readily detected, no transcripts from the βB2-2-crystallin gene could be found in the 22 months old human lens. These data suggest that the human βB2-2-crystallin gene is a pseudogene.

INTRODUCTION

The abundant mammalian lens specific proteins are encoded by three (α, β and γ) crystallin gene families. The structure and genomic organization of the members of the β- and γ-crystallin gene families have been extensively studied both in rat and man (for review see Lubsen et al., 1988). The genomic organization of the γ-crystallin family appears to be the same in these two species except for an additional γ-crystallin gene fragment present within the human genome (R. Brakenhoff, pers. comm.). Characterization of human and rat genomic clones containing βB-crystallin sequences revealed that, in contrast to the rat, the human genome contains two βB2-crystallin genes (Aarts et al., 1987). These βB2 sequences must have been derived from a very recent duplication of the βB2-crystallin gene in the human lineage, as the intronic sequences available from these genes are almost 90% identical (Aarts et al., 1987).

To investigate whether both human genes could yield a functional βB2 protein product we examined part of the coding sequence of each gene. In addition efforts were made to detect transcripts of these two genes within the human lens.

MATERIALS AND METHODS

RNA isolation, Northern blotting and probes

Lenses were homogenized in 1 ml lysis buffer (0.1 M Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA, 0.2% SDS; see Maniatis et al., 1982). The homogenate was extracted twice with phenol/chloroform/isoamylalcohol (25:24:1 v:v:v), the nucleic acids were precipitated with ethanol and collected by centrifugation. The pellet was dissolved in water and the precipitation step was repeated. For Northern blot analysis lens RNA was separated on a 1% agarose-formaldehyde gel. After electrophoresis the gel was soaked in 20xSSC for 5 min and the RNA was blotted to nitrocellulose (Schleicher & Schull). The βB2- and βB3-crystallin probes

were obtained by radiolabeling the inserts of the corresponding rat cDNA clones (Chapter 4; Den Dunnen et al., 1985) with α-^{32}P-dATP through nick-translation. Hybridization and washing conditions were as described previously (Chapter 4).

Oligonucleotides, primer extension and S1-nuclease mapping

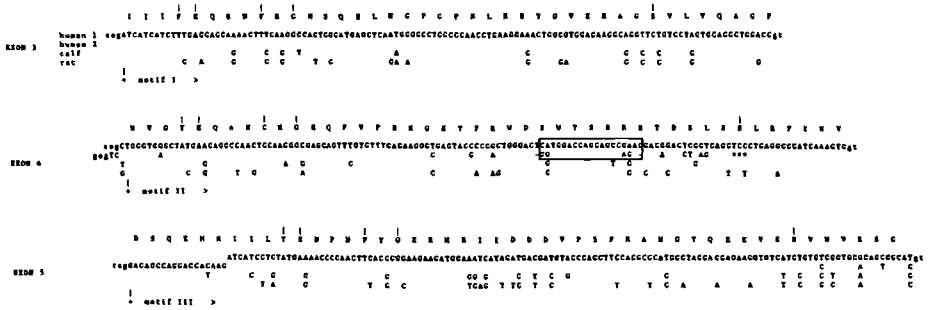
The following two oligonucleotides were used:
prB2-1: 5'-CTTCGGCTGCTGGTCCATG-3'; prB2-2: 5'-CTCTGGCTGCTGGTCCACC-3'
Oligonucleotide prB2-1 is complementary to a region (see also figure 1a) of the putative 4th exon of the human βB2-1-crystallin gene. Oligonucleotide prB2-2 is complementary to the same region of the human βB2-2-crystallin gene. Annealing of the oligonucleotide to the RNA and the sub-sequent primer extension reaction was performed according to Geliebter et al. (1986).

Sequence analysis

Restriction fragments of the human genomic clones λHeB2-3 and λHeB2-4 (Aarts et al., 1987) were used to construct M13mp clones containing βB2-crystallin sequences. All sequences given in this report were determined by the dideoxy chain termination method (Sanger et al., 1977).

RESULTS AND DISCUSSION

The partial sequence of the human βB2-1- and βB2-2-crystallin genes is given in Figure 1a. A high overall sequence correspondence is noticed when both sequences are compared with orthologous rat or calf sequences. However, when the extent of change of the coding sequences of the human genes relative to the calf or rat sequences is calculated, that of the βB2-2-crystallin gene is almost twice as high as that of the βB2-1-crystallin gene (Figure 1b). A possible explanation for this finding could be the inability of the βB2-2-crystallin gene to yield a functional βB2 protein, which would result in a relaxation of the selective pressure on its sequence. A further pointer to the fact that this gene is probably a pseudogene is the nature of the differences found between the sequences of both human βB2-crystallin genes. The most striking differences are found in the sequences of the putative 4th exons (see Figure 1a). Compared to the βB2-1- (or the calf; Hogg et al., 1987; and the rat βB2-; Chapter 4) crystallin sequence, the βB2-2-crystallin gene lacks the triplet encoding the serine residue involved in maintaining the characteristic "Greek key" protein fold. In addition the βB2-2-crystallin gene contains a mutation at the 5' splice acceptor site. The effect of the one amino acid deletion on the protein structure is not clear. However, the mutation at the splice acceptor site is



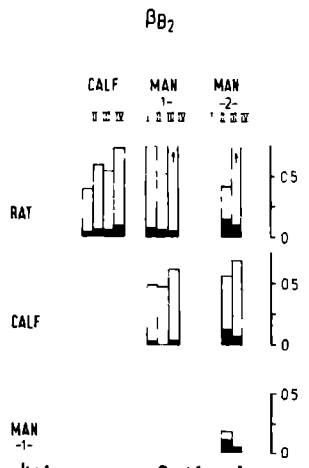


Figure 1: (a): Comparison of the human, calf and rat βB2-crystallin sequences. The sequence of exon 3, 4, and 5 of the human βB2-1-crystallin gene (human 1) is given completely, except for the beginning of exon 5 where the sequence is from the human βB2-2 gene (human 2). For the other sequences only the nucleotides that differ are shown. Identical nucleotides are indicated by a dash. The deletion of the triplet TCC in the human βB2-2-crystallin gene (human 2) is indicated by three asterisks. The numbering of the exons is by analogy with that of the rat βB1-crystallin gene (Den Dunnen et al., 1986). The sequences of the putative fifth exons of the human genes were taken from Aarts et al. (1987), the calf sequence from Hogg et al. (1987) and the rat sequence from Aarts (unpubl. res.; Chapter 4). The deduced amino acid sequence of the human βB2-1 crystallin is given above the nucleotide sequences. The protein motifs are indicated. The residues involved in the characteristic "Greek key" folding of the protein are indicated with a (|). The oligonucleotides used in the primer extension reactions as shown in figure 3 are complementary to the regions within the open box.

(b): Extent of synonymous and nonsynonymous substitutions in paralogous human βB2-1 (-1-), βB2-2 (-2-) and orthologous human, calf and rat βB2-crystallin genes. The extent of substitution was calculated according to Miyata and Yasunaga (1980) and corrected for back mutations by the method of Jukes and Cantor (1969). Open bars indicate the extent of synonymous substitution, filled bars the extent of nonsynonymous substitution. Arrows show where the fraction of nucleotides synonymously substituted is greater than one. Roman numerals indicate motif regions I to IV and correspond to the four motif coding regions. The first two nucleotides of the 4th exon of the human βB2-2-crystallin gene, the triplet tcc or tct of the 4th exon of the human βB2-1, the calf and the rat βB2-crystallin gene and the first 18 nucleotides of the 5th exon of the human, calf and rat βB2-crystallin sequences were excluded from the comparisons.

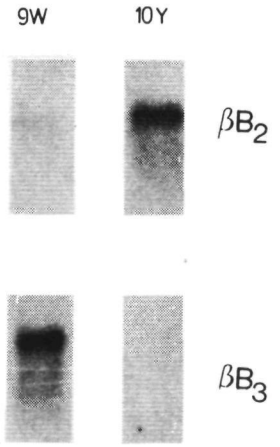


Figure 2: Autoradiograms of a Northern blot containing 9 week old fetal and 10 year old postnatal human total lens RNA hybridized successively with a βB2- and βB3-crystallin probe. 2 µg of total lens RNA was Northern blotted and hybridized with the nick translated insert of the rat cDNA clones as described in Materials and Methods.

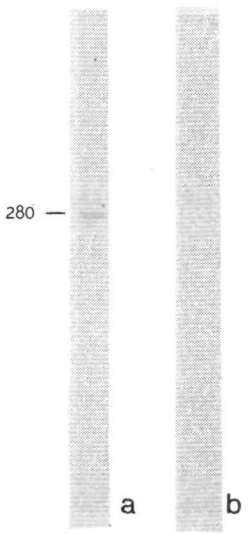


Figure 3: Primer extension on RNA isolated from 22 months old human lenses. The βB2-1- or βB2-2-crystallin gene specific 19mer oligonucleotides were extended on lens RNA as described in Materials and Methods. The primer extension products were analysed on a 6% polyacrylamide sequencing gel. The length in nucleotides of the primer extension product is indicated.

predicted to shift the splice site two nucleotides upstream, which would result in a shift in the reading frame. Hence from this region of the coding sequences we would predict that only the βB2-1-crystallin

gene can yield a functional protein product.

The βB2-crystallin transcripts are mainly found during postnatal development of the human lens (Figure 2). We therefore used RNA isolated from a 22 months old lens in experiments designed to detect a transcript from the βB2-2-crystallin gene. The high sequence similarity of the two human βB2-crystallin genes makes it impossible to discriminate between the two human βB2-crystallin transcripts on Northern- or dot-blots. We therefore assayed for the transcripts from the βB2-1- or βB2-2-crystallin genes by primer extension. The oligonucleotides used in these experiments are indicated in Figure 1a, and are expected to be specific either for the βB2-1- or the βB2-2-crystallin gene. A primer extension product was found only from the βB2-1-crystallin gene (Figure 3). Hence, if the βB2-2-crystallin gene is transcribed, the concentration of its transcript must be much lower than that of the βB2-1-crystallin gene RNA.

From these results we conclude that the human βB2-2-crystallin gene is inactive by three criteria: no transcript is detected, it contains a mutated splice site and its rate of change is much higher than that of the βB2-1-crystallin gene. It is not known whether the βB2-2-crystallin gene is a complete gene, as we have as yet not succeeded in cloning the first exonic regions of either one of the two human βB2-crystallin genes.

REFERENCES

Aarts, H.J.M., Den Dunnen, J.T., Lubsen, N.H., and Schoenmakers, J.G.G. (1987). Linkage between the βB2 and βB3 crystallin genes in man and rat: a remnant of an ancient β-crystallin gene cluster. Gene 59, 127-135.

Den Dunnen, J.T., Moormann, R.J.M., and Schoenmakers, J.G.G. (1985). Rat lens β-crystallins are internally duplicated and homologous to γ-crystallins. Biochim. Biophys. Acta 824, 295-303.

Den Dunnen, J.T., Moormann, R.J.M., Lubsen, N.H., and Schoenmakers, J.G.G. (1986). Intron insertions and deletions in the β-/γ-crystallin gene family: The rat βB1-gene. Proc. Natl. Acad. Sci. USA 83, 2855-2859.

Geliebter, J., Zeff, R.A., Melvold, R.W., and Nathenson, S.G. (1986). Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis: K^{bm9} and K^{bm6}. Proc. Natl. Acad. Sci. USA 83, 3371-3375.

Hogg, D., Gorin, M.B., Heinzmann, C., Zollman, S., Mohandas, T., Klinsak, I., Sparkes, R.S., Breitman, M., Tsui, L.-C., and Horwitz, J. (1987). Nucleotide sequence for the cDNA of the bovine βB2 crystallin and assignment of the orthologous human locus to chromosome

22. Curr. Eye Res. 6, 1335-1342.
Jukes, T.H., and Cantor, C.R. (1969). Evolution of protein molecules. In: Munro, H.N. (ed) Mammalian protein metabolism. Academic Press, New York, pp. 21-123.
Lubsen, N.H., Aarts, H.J.M., and Schoenmakers, J.G.G. (1988). The evolution of lenticular proteins: the β- and γ-crystallin supergene family. Prog. Biophys. Molec. Biol. 51, 47-76.
Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning, A laboratory Manual. Cold Spring Harbor Laboratory.
Miyata, T., and Yasunaga, T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J. Mol. Evol. 16, 23-36.
Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

CHAPTER 7

THE γ-CRYSTALLIN GENE FAMILIES: SEQUENCE AND EVOLUTIONARY PATTERNS

Henk J.M. Aarts, Johan T. Den Dunnen, Jack Leunissen,

Nicolette H. Lubsen and John G.G. Schoenmakers.

J Mol Evol (1988) 27:163–172

THE γ-CRYSTALLIN GENE FAMILIES: SEQUENCE AND EVOLUTIONARY PATTERNS

SUMMARY

The γ-crystallin proteins consist of two topologically equivalent domains, each built up out of two similar motifs. They are encoded by a gene family, which already contained five members before the divergence of rodents and primates. A further gene duplication took place in each lineage. To analyze the pattern of evolution within this gene family, the coding sequences of six human genes, six rat genes, and four mouse genes were compared. Between species, a uniform rate of evolution of all regions of the protein is seen. The ratio of synonymous to nonsynonymous substitution in the human/rat or human/mouse comparison is much lower than the ratio when rat and mouse are compared indicating that the γ-crystallin proteins are better conserved in the rodent lineage. Within species, the regions encoding the two external motifs I and III of the protein show a greater extent of nonsynonymous substitution than the regions encoding the two internal protein motifs II and IV. The low extent of synonymous substitution between the second exons (encoding motifs I and II) of the rat γ-crystallin genes suggests the frequent occurrence of gene conversion. In contrast, a high extent of synonymous substitution is found in exon 3 (encoding motifs III and IV) of the rat genes. The same phenomenon is seen within the human gene family. The frequencies of occurrence of the various dinucleotides deviate less from those predicted from the frequencies of occurrence of each individual nucleotide in the second exons than in the third exons. The sequences of the third exons are significantly depleted in CpG, ApA and GpT and enriched in CpT and GpA.

INTRODUCTION

The γ-crystallins are lens-specific proteins that account for up to 40% of the soluble proteins of the mammalian lens. The exact function of the γ-crystallins is unknown, although their high concentration and short-range packing order within the lens is generally believed to be essential in the maintenance of lens transparency (Delaye and Tardieu 1983).

The γ-crystallins are encoded by a gene family whose expression is correlated with differentiation of the lens fiber cells (for review see Bloemendal 1981; Piatigorsky 1984). In the rat this family has six members, five of which are located in a head-to-tail orientation within 50 kb of DNA (Moormann et al. 1985). The sixth gene is located on the same chromosome, i.e., chromosome 9 (Den Dunnen et al. 1987) but at least 23 kb away from the five gene cluster (Moormann et al. 1985). All of the rat genes are functional. Expression of the rat genes is switched on most probably simultaneously during the early developmental

stages followed by a differential switch-off in the postnatal and later stages of development of the lens (Van Leen et al. 1987). From humans, six γ-crystallin genes have been characterized (Den Dunnen et al. 1985a; Meakin et al. 1985, 1987). Two of these genes contain an in-frame termination codon at identical positions in their second exon and must thus be pseudogenes. The human genome is thought to contain a seventh γ-crystallin gene, but this gene has not yet been isolated. In the human genome all γ-crystallin genes are located on chromosome 2, region q33-36 (Den Dunnen et al. 1985b; Willard et al. 1985; Shiloh et al. 1986) and at least four of the human γ-crystallin genes are closely linked (Den Dunnen et al. 1985a; Meakin et al. 1985). The mouse genome also contains multiple γ-crystallin genes. The (partial) sequence of four of these has been established (Breitman et al. 1984; Lok et al. 1984, 1985) but their exact genomic organization is unknown.

The γ-crystallins are monomeric basic proteins (Mr 21,000), relatively rich in sulfhydryl groups (for review, see Bloemendal 1981; Piatigorsky 1984). The tertiary structure of the protein is highly symmetrical: the building blocks are four homologous protein motifs (for review, see Lindley et al. 1985). These four motifs are organized into two similar domains that are connected by a short connecting peptide three or four amino acid residues in length. The two domains are held together by the interaction between protein motifs II and IV, which are on the inside of the protein. Protein motifs I and III face the outside of the protein. Protein motifs I and II are encoded by the second exon of the gene (the first exon contains only nine bases of the coding region), while the third exon of the gene encodes the connecting peptide as well as motifs III and IV (Moormann et al. 1983; Den Dunnen et al. 1985a, 1986; Meakin et al. 1985, 1987).

We show here that within a species the sequence of the "internal" motifs II and IV is better conserved than that of the "outside" motifs I and III, but that all four motif sequences are equally variable between human and rat. We further show that (frequent) gene conversion between the γ-crystallin genes is confined to the second exon. In addition the second and third exons of the γ-crystallin gene differ in (di)nucleotide frequencies.

METHODS

All the γ-crystallin gene sequences have been published. The six rat sequences can be found in Moormann et al. (1983) and Den Dunnen et al. (1986), the four mouse sequences in Breitman et al. (1984) and Lok

et al. (1984, 1985), the six human sequences in Meakin et al. (1985, 1987) and Den Dunnen et al. (1985a), and the calf cDNA sequence in Bhat and Spector (1984). For sequence comparison only the amino acid coding regions of the second exon (243 bp) and the third exon (273 bp) were used. These were subdivided into the regions encoding each of the four protein motifs. The triplets encoding the additional amino acid found in the connecting peptide of the human γ1-2, the rat γ1-2 and the mouse γ3 gene were excluded from the comparison. At the site of the one base pair deletion in the third exon of the human pseudo-1 gene a G (the most common nucleotide at that site) was substituted.

The number and extent of synonymous and nonsynonymous substitution was calculated according to Miyata and Yasunaga (1980). These were corrected for back mutation by the method of Jukes and Cantor (1969).

Nucleotide sequences were aligned and nucleotide and dinucleotide frequencies were calculated using the computer programs of Staden (1982).

RESULTS AND DISCUSSION

Conservation of the γ-crystallin gene cluster

The 17 known mammalian γ-crystallin coding sequences are shown in Fig. 1. The orthologous relationships between these sequences were established by calculating the number of synonymous and nonsynonymous base substitutions (see Table 1). From these numbers a parsimony phylogenetic tree was constructed (Fig. 2). Where possible, orthology was confirmed by comparing the flanking and intronic sequences. A comparison of the flanking sequences also showed that the mouse γ2 gene shares a direct ancestor with the rat γ4-1 rather than with the rat γ3-1 gene. Fig. 2 suggests that a five member γ-crystallin gene family was already present in the common ancestor of human and rats. The chromosomal organization of this ancestral γ-crystallin gene cluster has been preserved: when the orthologous genes are aligned, the chromosomal organization of the rat and human genes (to the extent that the latter is known) can be maintained (see Fig. 3). The analysis of the coding sequence further suggests that in both the human and the rodent lineage a duplication of the last gene of the cluster took place to yield the present human pseudo-1 and -2 or the rat γ3-1 and γ4-1 genes, although the sequence similarity of the large introns of the two rat genes is much less than expected. Since the human pseudo-2 and rat γ3-1 or the

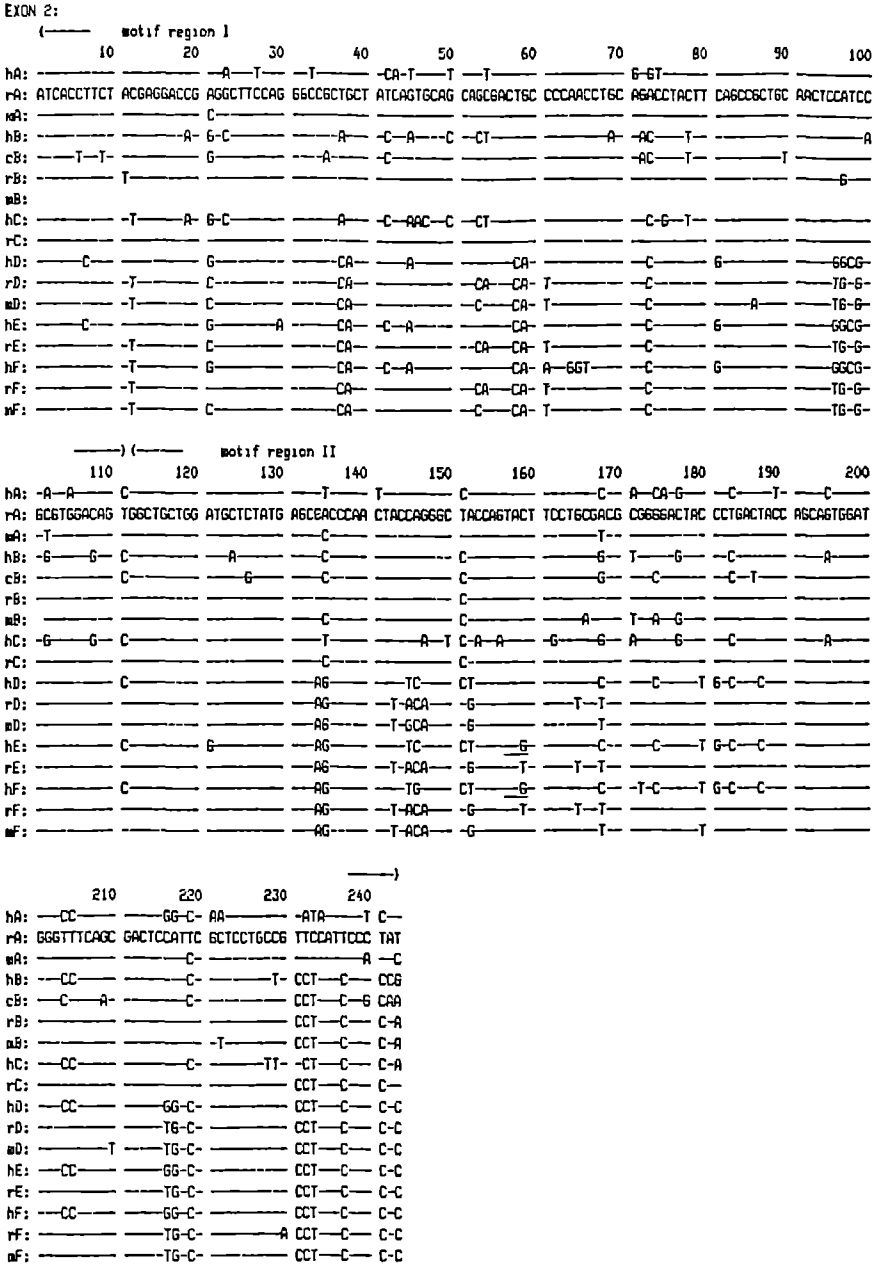
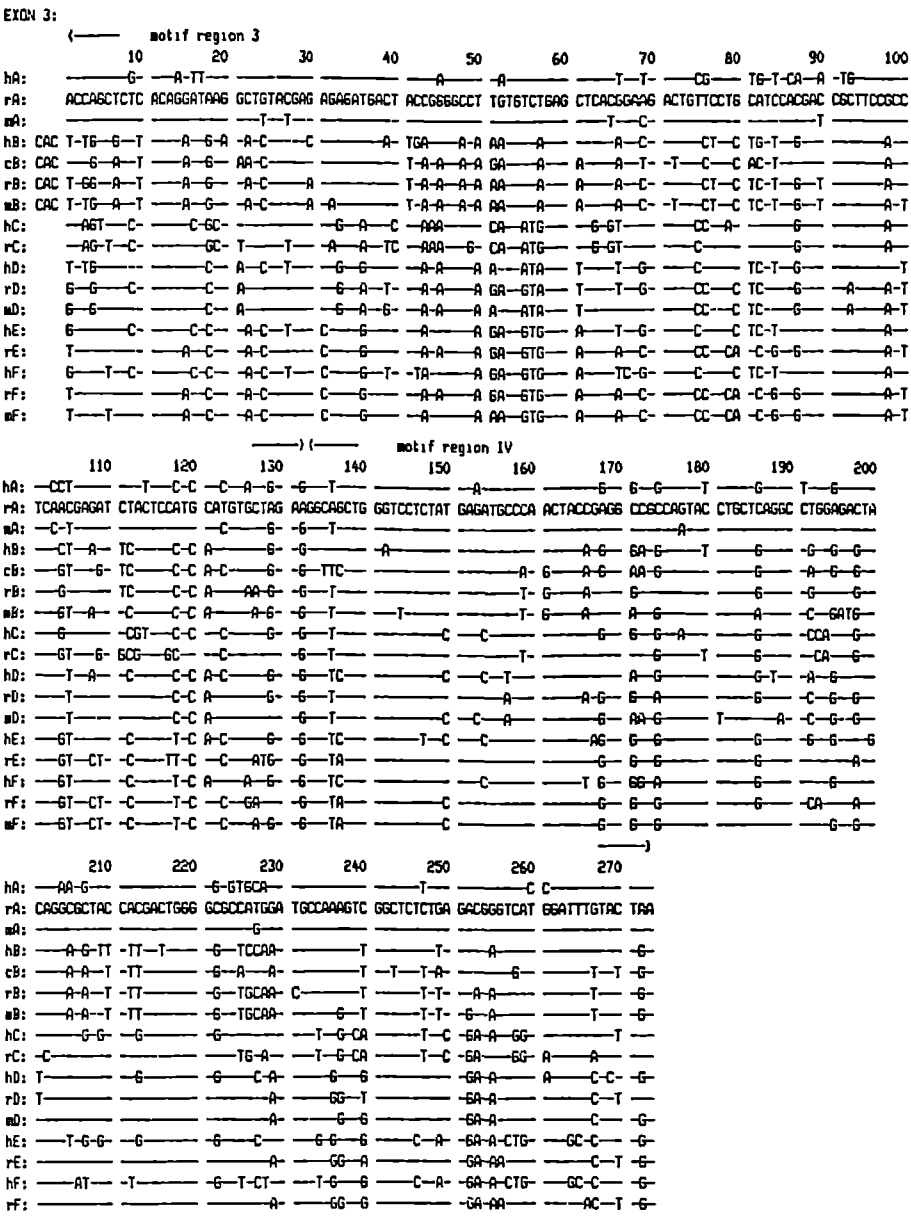


Fig. 1. Alignment of the γ-crystallin coding sequences. The sequences are indicated as h (human), r (rat) and m (mouse) according to the nomenclature given in Fig. 2, bottom row. The sequence of the rat γA



gene is shown entirely; for the other genes only differences are specified. The stop codons in the human gE and gF genes are underlined. The four motif regions are indicated.

human pseudo-1 and rat γ4-1 occupy equivalent chromosomal positions, we have paired these genes in the following discussion. It is evident from Figs. 2 and 3 that the nomenclature of the γ-crystallin genes is not standardized. To avoid confusion in the comparison of genes from various species we have introduced a new nomenclature as indicated in Fig. 3 (bottom line) and we will, for example, refer below to the rat γA and mouse γA sequences rather than to the rat γ1-1 and the mouse γ4 sequences.

The rate of evolution of the γ-crystallin sequences between species

The extent of synonymous and nonsynonymous substitution between the coding regions of orthologous γ-crystallin genes is shown in Table 2. The ratio of silent to nonsynonymous substitution is much higher in the rat/mouse than in the human/rodent comparison, suggesting that the selective pressure on the protein sequence is higher in the rodent lineage than in the primate lineage. There may be a general increase of the selective pressure on crystallins in rodents as the αA-crystallin (protein) is also better conserved in the rodent lineage than in the primate lineage (De Jong 1981). With respect to the γ-crystallins, this may be correlated with the difference in water content of the lens: it has been postulated that the main function of the γ-crystallins is the ordering of the lens water molecules (for review, see Lindley et al. 1985) and the low water content of the rat lens may increase the functional constraints on the γ-crystallin protein.

We have also compared the extent of synonymous and of nonsynonymous substitution between the regions encoding each of the four motifs of orthologous genes from rat, mouse, and human. When we compare the (functional) genes between human and rat, all four motif regions show about the same extent of (synonymous or nonsynonymous) change (see Fig. 4, bottom). This observation suggests that the selective pressure is the same on the "internal" and the "external" motifs of the protein. The high selective pressure on the external regions may be a consequence of the close short-range interactions between the crystallins (Delaye and Tardieu 1983).

The two human pseudogenes have diverged significantly further from the rodent genes in motif regions III and IV than the functional genes have (Fig. 4, middle and bottom row). This drift is presumably due to the loss of selective pressure on the sequence of these pseudogenes.

Table 1

	hG5	h12	h21	hG4	hp2	hp1	r11	r12	r21	r22	r31	r41	m4	m3	m1	m2	cII
Exon 2																	
hG5		22	21 5	18 5	18 5	19 5	19 5	21 5	21 5	26 5	27 5	26 5	20 5	—	26 5	26 5	23 5
h12	19		13	16 5	16 5	17 5	20 5	19	17 5	22 5	23 5	23 5	19 5	—	22 5	22 5	15 5
h21	21 5	6		20 5	20 5	19 5	22	22	22	22 5	23 5	23 5	22	—	22 5	22 5	22 5
hG4	27 5	24 5	26 5		2	3	15 5	13	13	14	15	15	19	—	13	11	15
hp2	29 5	26 5	28 5	2		3	17 5	15	15	16	16	16	17	—	15	13	15
hp1	31 5	28 5	30 5	7	7		18 5	16	16	15	15	15	18	—	13	12	16
r11	17 5	15 5	18	21 5	23 5	25 5		2	3	13 5	14 5	13 5	7	—	12 5	12 5	19
r12	17 5	13	15	18	20	22	7		1	11	12	11	8	—	10	10	17
r21	14 5	11 5	14	18	20	22	4	3		11	12	11	8	—	10	10	16
r22	27 5	22 5	24 5	12	14	16	17 5	16	16		1	2	11	—	3	3	21 5
r31	27 5	22 5	24 5	12	14	16	17 5	16	16	0		1	12	—	4	4	23
r41	28 5	23 5	25 5	13	15	17	18 5	17	17	1	1		13	—	5	5	23
m4	17 5	15 5	20	18	24	26	0	7	4	18	18	18 5		—	10	10	18
m3	—	—	—	—	—	—	—	—	—	—	—	—	—		—	—	—
m1	28 5	23 5	25 5	13	15	17	18 5	17	17	2	2	3	19	—		2	21
m2	27 5	22 5	23 5	12	14	16	17 5	16	16	0	0	1	18	—	2		21
cII	18 5	11 5	13	21	23	25	9	6	5	19	19	20	9	—	20	19	
Exon 3																	
hG5		31 5	36	34	32 5	21 5	29	36	40 5	33	33 5	33	29 5	38	36	35 5	37
h12	35 5		41 5	28 5	30 5	30 5	34 5	16	46 5	33	31 5	32	37	16	35	29 5	24
h21	35	40 5		30 5	29 5	37 5	34 5	39	26	25 5	24	23	32 5	40	29	26	41 5
hG4	35	36 5	23 5		19	27 5	33	33 5	37	22	21 5	19	29 5	28 5	19 5	19 5	34 5
hp2	39 5	39 5	32 5	25 5		13	36	36 5	41 5	26 5	25	18 5	33	37	29	22	35 5
hp1	42 5	39 5	36 5	27 5	10		37 5	35 5	38 5	26	25 5	26	37	38	32 5	23 5	39
r11	20	33	28	26	32	32 5		33 5	33 5	31	25	25 5	9	37 5	27	25	39 5
r12	34	16	39	30 5	33 5	32 5	33 5		44	37	34 5	34 5	36	16	40 5	36 5	27 5
r21	33 5	41 5	12	35	38 5	38 5	27 5	37		34 5	30 5	33 5	32	44	36	33 5	48 5
r22	33	30	29 5	16	26	26	23	24	30 5		22 5	22	30 5	35	17	24 5	37 5
r31	38 5	34 5	36	26 5	23	27 5	32	24 5	37 5	17 5		4	23 5	33	24 5	5	35
r41	41	37	37	29	29 5	30	33 5	27 5	37 5	20	3		39	32	20	7	35 5
m4	18 5	35	30 5	28 5	34	34	4	33	29	25 5	31 5	20		36	30	23 5	40 5
m3	38	20	43	32 5	38	36	37 5	4	39	26	27	30	37		33 5	32	25 5
m1	35	30	31	17 5	31	31 5	23	26 5	32	5	23 5	25	25	28 5		19 5	39 5
m2	38 5	34 5	36	26 5	27	27 5	32	24 5	37 5	17 5	0	3	31 5	27	23 5		37 5
cII	35	21	35 5	29 5	31 5	31	29 5	9 5	33 5	21 5	24	26 5	29 5	13 5	23 5	23 5	

Table 1. Synonymous and nonsynonumous substitution between the γ- crystallin genes. The number of synonymous and nonsynonymous substitutions between the two exons of the various γ-crystallin genes was calculated as detailed in Methods. The numbers shown are not corrected for back mutation. The number of synonymous substitution is shown in the upper right corner, while the number of nonsynonymous substitutions is shown in the lower left corner. The various species are indicated by h (human), r (rat), m (mouse) or c (calf). The nomenclature of the genes is slightly abbreviated from the full nomenclature shown in Figs. 2 and 3.

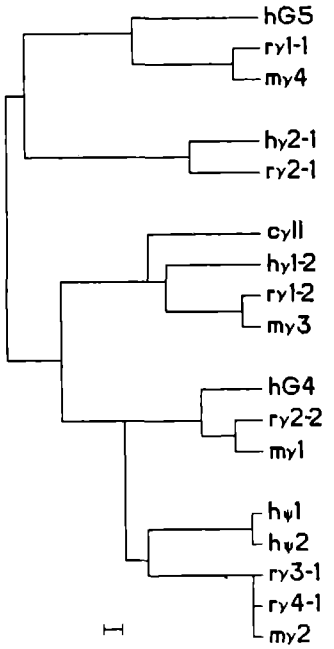


Fig. 2. Parsimony evolutionary tree of the γ-crystallin genes. The tree was constructed from a difference matrix based upon the extent of non-synonymous substitution in motif region III (except for the mouse γ2/ratγ3-1 divergence, see text). This method does not allow the unambiguous placement of the human G4/rat γ2-2/ mouse γ1 genes. For these genes we have followed the path suggested by the analysis of shared mutations within the rat γ-crystallin gene family (Den Dunnen et al. 1986; see also Fig. 5). The various species are indicated as in Table 1 and each γ-crystallin sequence is identified according to the nomenclature used by the original author (for references, see Methods). The length of the branches shows the extent of nonsynonymous substitution, the bar corresponds to 1%.

Finally, the rat-mouse comparison (Fig. 4, top row) shows that the extent of synonymous substitution increases going from motif region I to motif region IV. The significance of this observation is, in view of the large standard deviation, questionable, but it may represent a conservative effect of gene conversion involving motif regions I and II in rat and mouse (see below).

The extent of changes between γ-crystallin genes within species

The results presented above suggested that the functional constraints upon the four motif regions of the protein are about the same. However, when the extent of synonymous and nonsynonymous changes between the various γ-crystallin genes within a species is compared (see Fig. 5), it is immediately obvious that the four γ-crystallin motif regions have not evolved at the same rate. As we have argued previously (Den Dunnen et al. 1986), the virtual sequence identity between the first two motif regions of the rat γA, γB and γC genes as well as be-

tween the first two motif regions of the rat γD, γE, and γF genes suggests that these regions evolve in concert. The data presented here further suggest that at an earlier time genetic information was also exchanged between the first two motif regions of the rat γA, γB, and γC group and those of the rat γD, γE, and γF group. Traces of concerted evolution can also be discerned within the human γ-crystallin gene family. Most obvious is the sequence identity, again between the first

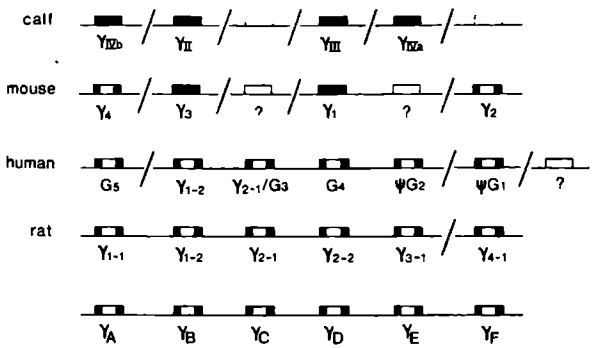


Fig. 3. Alignment of orthologous γ-crystallin genes. The various γ-crystallin sequences from human, calf, and mouse are aligned with their orthologs from the rat. The chromosomal organization of the rat gene cluster has been maintained. The various species are indicated on the left and each γ-crystallin sequence is identified as in Fig. 1. Partially filled rectangles indicate sequenced genes, filled rectangles represent cDNA sequences, while the gray rectangles correspond to the (partially) determined γ-crystallin protein sequences from the calf (Slingsby and Croft 1978). γ-crystallin sequences identified only as hybridizing genomic fragments are shown as white bars. Stippled bars indicate γ-crystallin genes expected to exist but as yet unidentified. The sequences are shown in their chromosomal organization (where known) on an uninterrupted horizontal line; slashes mark at present unlinked sequences. The nomenclature used in this paper is shown on the bottom line.

two motif regions only, between the γD, γE, and γF genes. Considering the stability of the chromosomal organization of the γ-crystallin gene cluster (see Fig. 3), the concerted evolution of this part of the γ-crystallin genes is probably due to gene conversion rather than unequal recombination.

In motif regions III and IV the fraction of synonymous substituted nucleotides often is close to or larger than 1. Hence these regions do not evolve in concert. The sequence divergence in these regions is

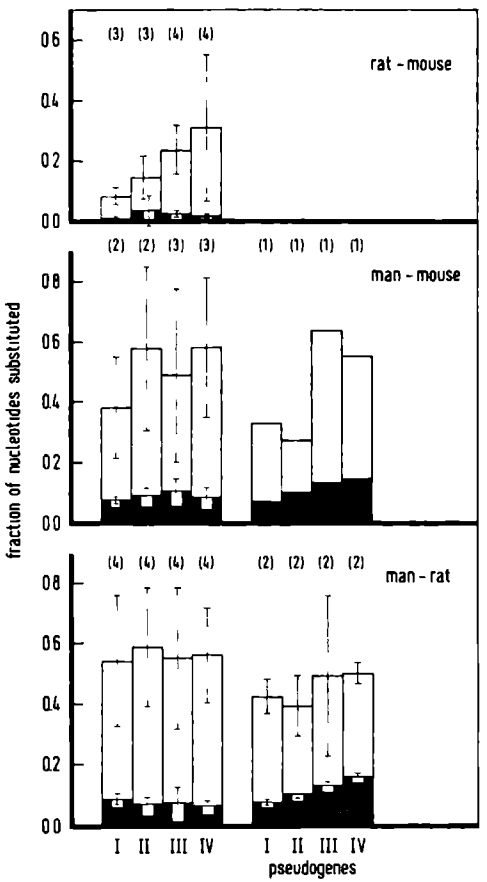


Fig. 4. Extent of synonymous and nonsynonymous substitution in orthologous γ-crystallin genes. The extent of substitution between orthologous gene pairs from the species indicated was calculated as described in Methods. In the case of the human genes the extent was calculated separately for the functional and for the pseudogenes. Open bars indicate the extent of synonymous substitution; filled bars that of nonsynonymous substitution. The standard deviation is shown. The four motif regions are indicated by the roman numerals I-IV; the arabic numerals between brackets above the bars show the number of orthologous gene pairs compared. Note that from the mouse γB sequence only motif regions III and IV were used.

probably due to random mutation rather than slippage (Tautz et al. 1986), since mutations are spread throughout and are often shared by evolutionary closely related or orthologous genes (see Fig. 1). Motif region IV shows a lower extent of nonsynonymous substitution than motif region III, suggesting that the sequence of motif region IV has been subject to a higher selective pressure than that of motif region

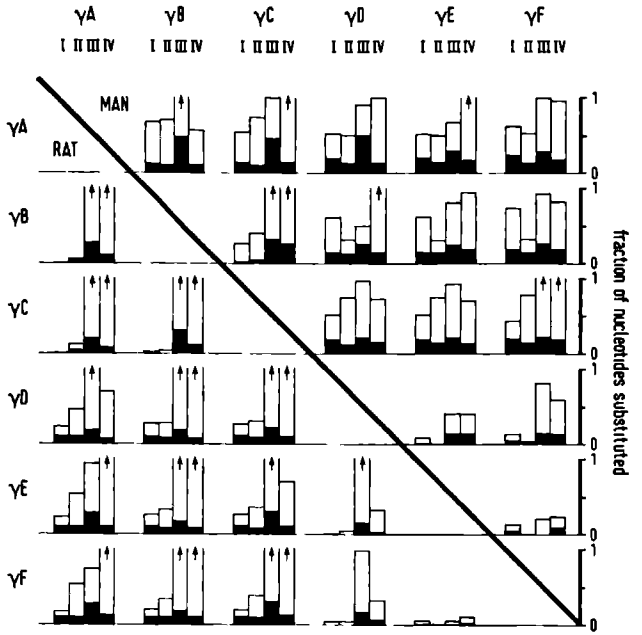


Fig. 5. Extent of synonymous and nonsynonymous substitution in the γ-crystallin genes within a species. The comparison between the human genes is shown in the upper right corner; that between the rat genes in the lower left corner. The extent of substitution was calculated as described in Methods. Open bars indicate the extent of synonymous substitution; filled bars that of nonsynonymous substitution. Arrows show where the extent of synonymous substitution is greater than one. Roman numerals indicate motif regions I-IV, respectively, while the genes are named as in Fig. 2, bottom row.

Table 2

	Synonymous substitution K_s	Nonsynonymous substitution K_a	K_a/K_s	n
Human/rat	0.549 ± 0.131	0.078 ± 0.014	7.0	4
Human (pseudo)/rat	0.501 ± 0.059	0.117 ± 0.012	4.3	2
Human/mouse	0.495 ± 0.187	0.084 ± 0.009	5.9	3
Human (pseudo)/mouse	0.443	0.114	3.9	1
Rat/mouse	0.209 ± 0.083	0.013 ± 0.004	16.1	3

Table 2. The rate of synonymous and nonsynonymous substitution in the γ-crystallin genes. The extent of synonymous (Ks) and nonsynonymous (Ka) substitution was calculated for the coding sequence contained in the second and third exons of the γ-crystallin genes as detailed in Methods. When the comparison involved the mouse γB sequence, the coding region of the third exon only was used. The standard deviation is given. The extent was calculated separately for the human functional and pseudo-genes. n indicates the number of orthologous gene pairs compared.

Table 3

	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG
Exon 2																
hA	−14	+06	−01	+07	+23	−12	+11	−20	−07	−01	−17	+12	−14	+08	+03	+03
hB	−11	08	−01	+18	+30	+04	+11	−41	+09	+02	−18	+07	−28	+04	+03	−18
hC	−11	+01	−08	+19	+26	0	+11	−36	+04	−05	−06	−04	−18	+06	0	+13
hD	−13	−05	+01	+19	+36	−24	+09	−23	+05	+04	−15	+06	−26	+27	+02	−01
hE	−12	−04	+01	+16	+33	−24	+09	−21	−03	+04	−11	+10	−17	+24	−03	−04
hF	−13	−07	−01	+22	+34	−24	+14	−26	+03	+06	−18	+10	−24	−28	+01	−04
rA	−07	−15	+01	+18	+35	−12	+08	29	−03	+16	−25	+14	−24	+12	+14	−01
rB	−04	−08	−08	+21	+34	−11	+10	−30	−08	+10	−19	+14	−20	+11	+14	−03
rC	−08	13	−03	+22	+33	−10	+14	−34	+01	+10	−23	+15	24	−16	+10	−01
rD	−11	−06	−09	+26	+36	−24	+25	−40	09	+13	−20	+17	−15	+18	+01	−03
rE	−08	−06	−10	+26	+31	−21	+26	−39	−06	+10	−20	+17	−16	+19	+01	−03
rF	−09	−06	−12	+26	+32	18	+28	45	−07	+12	−23	+20	−15	+13	+04	−01
Exon 3																
hA	−17	01	−02	+20	+43	−01	−08	−32	−13	+17	−10	+03	−13	−14	+18	+10
hB	−09	+23	29	+16	+42	−11	+14	−44	−03	+03	−12	+10	−32	−14	+28	−19
hC	06	−06	−08	+23	−30	+12	+10	−52	−05	−05	−12	+19	−18	0	+08	+11
hD	−23	+27	−10	+08	+60	−11	+02	−50	−10	+06	24	+27	29	−21	−34	+13
hE	−15	+10	10	−16	+58	−04	+11	−63	−13	+11	−27	+26	−13	−16	+26	+19
hF	−26	+14	+03	+10	+69	−04	+03	−68	−05	+11	−31	+22	−37	−20	+26	+32
rA	−22	+04	−02	+21	+44	−12	0	−30	−02	+11	−22	+10	−20	0	+22	0
rB	−04	−24	−20	+01	+42	−15	+07	−34	−06	0	−24	+27	−36	−08	−38	+07
rC	−23	01	+03	+22	+43	−10	−01	−30	−04	−05	−15	+21	−14	+18	+10	−12
rD	14	+09	−04	+10	+38	+08	+04	−48	+07	+04	−31	+18	−30	−19	+32	+18
rE	−10	+08	−10	+13	+37	−04	+12	−44	+05	+12	−36	+17	−35	−14	+35	+16
rF	−11	+06	−04	+10	+35	−03	+11	−42	+04	+15	−40	+19	−30	−16	+14	+14

Table 3. Dinucleotide frequencies of the γ-crystallin sequences. The expected dinucleotide frequency was calculated from the frequency of each individual nucleotide (see Table 4). The difference between observed and expected dinucleotides/100 bp is shown; the number is underlined when the difference between expected and observed was significant with $p < 0.05$ as calculated by the chi-square test (McClelland and Ivarie 1982). The dinucleotide frequencies are shown separately for the coding regions of the second and third exons. The genes are indicated as in Fig. 1.

Table 4

	T	C	A	G	G+C
Exon 2					
hA	21.8	33.3	21.6	23.5	56.8
hB	17.7	37.0	20.6	24.7	61.7
hC	18.9	34.6	21.8	24.7	59.3
hD	16.0	39.5	17.7	26.7	66.2
hE	15.6	39.5	17.7	27.2	66.7
hF	17.3	37.4	17.3	28.0	65.4
rA	21.0	34.6	19.3	25.1	59.7
rB	19.3	36.2	18.9	25.5	61.7
rC	19.3	36.6	18.9	25.1	61.7
rD	21.0	36.2	18.1	24.7	60.9
rE	21.4	35.8	18.1	24.7	60.5
rF	21.4	35.4	18.9	24.3	59.7
Exon 3					
hA	23.1	26.0	21.6	29.3	55.3
hB	23.8	22.7	24.5	28.9	51.6
hC	17.9	28.2	21.2	32.6	60.8
hD	21.2	27.1	22.3	29.3	56.4
hE	18.3	28.6	19.4	33.7	62.3
hF	22.0	27.1	20.5	30.4	57.5
rA	20.1	29.3	22.7	27.8	57.1
rB	23.4	22.7	26.0	27.8	50.5
rC	21.2	24.9	22.7	31.1	56.0
rD	20.9	25.6	23.1	30.4	56.0
rE	19.8	27.8	23.4	28.9	56.7
rF	18.3	28.9	24.2	28.6	57.5

Table 4. Frequencies of each of the four nucleotides in the γ- crystallin sequences. The nucleotide frequencies are shown separately for the coding regions of the second and third exons. The genes are indicated as in Fig. 1.

III. Within the rat gene family, recent gene conversions mask any difference between motif regions I and II in their rate of change. From the comparison of the human genes, it is apparent that the "external" motif region I has a higher extent of nonsynonymous changes than the "internal" motif region II. The differences in the extent of change of the various motifs within the human lineage indicate that, as expected from the protein structure, motifs I and III or II and IV are topologically equivalent.

The stronger conservation of motif regions II and IV within a species is in apparent contradiction with the equal conservation of all motif regions between species. As pointed out earlier (Den Dunnen et

al. 1986), the larger differences between motifs I or III of the various γ-crystallin proteins may have resulted from a period of radiation directly after gene duplication as also suggested for other gene families (see Czelusniak et al. 1982; Li 1983).

The effect of gene conversion on the nucleotide sequence

The data presented above show that the domain comprised by motif I and II which is encoded by the second exon is functionally and topologically equivalent to that consisting of motif III and IV and encoded by the third exon. The γ-crystallin genes thus provide a unique opportunity to determine whether the sequence composition of an exon is affected by the mode of evolution of that exon. In 3 out of the 12 genes the dinucleotide frequencies in the second exon do not differ significantly from those predicted from the frequencies of occurrence of the individual nucleotides; in half of the genes a significant lack of CpG or excess of CpT is noted (see Table 3). Deviation from the predicted dinucleotide frequencies is much larger in the third exon of the genes. All genes have an excess of CpT in this region and most are depleted in CpG. Noteworthy is also the trend towards a significant lack of ApA and GpT and an excess of GpA.

The general underrepresentation of CpG in vertebrate DNA (Josse et al. 1961; McClelland and Ivarie 1982) could be a consequence of increased deamination of the methylated C which is found almost exclusively in CpG. If this were the case, then a significant elevation of the resulting dinucleotides TpG and CpA would be expected. A consistent excess of these dinucleotides is indeed found in the coding region of the γ-crystallin genes (see Table 3). In the case of the second, but not the third, exon sequences the excess of TpG and CpA entirely compensates for the lack of CpG. The methylcytosine deamination hypothesis does, however, not explain the high CpT or the low GpT content. The source of the CpT excess/GpT lack is not clear. It is unlikely that it is simply caused by GpT -> CpT mutation since the third exons (which have the largest deviation) also have an increased G and decreased C content relative to the second exons (see Table 4). Furthermore, a large part of the ancestral rodent/human γ-crystallin gene sequence can be derived and going from the ancestral γ-crystallin sequence to the present sequences, GpT mutations are seldom observed and they never yield CpT but usually ApT (in accordance with the prevalence of trans-

itions over transversions; Li et al. 1985). Mutations to CpT are virtually exclusively from CpC (data not shown). It must, however, be noted that this ancestral sequence already has a CpT excess and a GpT (as well as a CpG) depletion, suggesting that the processes that led to the bias in dinucleotide frequencies preceded the radiation of the γ-crystallin genes.

The CpG "avoidance" in the vertebrate DNA sequence has led to a debate over its possible cause (e.g., see Grantham 1985) and correlations with methylation state (e.g., Max 1985) or genome compartments (Bernardi and Bernardi 1986) have been proposed. The analysis of the sequence variation within the γ-crystallin genes suggests that gene conversion should be taken into account as one of the factors that influences the dinucleotide frequencies in vertebrate DNA.

ACKNOWLEDGEMENT. We thank Drs. Breitman and Tsui for providing their sequence data prior to publication and Dr. W. de Jong for critical discussion. This research was carried out under the auspices of the Netherlands Foundation for Chemical Research (SON), with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

REFERENCES

Bernardi G, Bernardi G (1986) Compositional constraints and genome evolution. J Mol Evol 24:1-11

Bhat SP, Spector A (1984) Complete nucleotide sequence of a cDNA derived from calf lens γ-crystallin mRNA: presence of sequences complementary to Alu-I like DNA. DNA 3:287-295

Bloemendal H (1981) The lens proteins. In: Bloemendal H (ed) Molecular and cellular biology of the eye lens. John Wiley & Sons, New York, pp 1-48

Breitman ML, Lok S, Wistow G, Piatigorsky J, Treton JA, Gold RJM, Tsui L-C (1984) γ-crystallin family of the mouse lens: structural and evolutionary relationships. Proc Natl Acad Sci USA 81:7762-7766

Czelusniak J, Goodman M, Hewett-Emmett D, Weiss ML, Venta PJ, Tashian RE (1982) Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature 298:297-300

De Jong WW (1981) Evolution of lens and crystallins. In: Bloemendal H (ed) Molecular and cellular biology of the eye lens. John Wiley & Sons, New York, pp 221-278

Delaye M, Tardieu A (1983) Short-range order of crystallin proteins accounts for eye lens transparency. Nature 302:415-417

Den Dunnen JT, Moormann RJM, Cremers FPM, Schoenmakers JGG (1985a) Two human γ-crystallin genes are linked and riddled with Alu-repeats. Gene 38:197-204

Den Dunnen JT, Jongbloed RJE, Geurts van Kessel AHM, Schoenmakers JGG (1985b) Human lens γ-crystallin gene are located in the p12-qter region of chromosome 2. Hum Genet 70:217-221

Den Dunnen JT, Moormann RJM, Lubsen NH, Schoenmakers JGG (1986) Concerted and divergent evolution within the rat γ-crystallin gene family. J Mol Biol 189:37-46

Den Dunnen JT, Szpirer J, Levan G, Islam Q, Schoenmakers JGG (1987) All six rat γ-crystallin genes are located on chromosome 9. Exp Eye Res 45:747-750

Grantham R (1985) CG doublet difficulties in vertebrate DNA. Nature 313:437

Josse J, Kaiser AD, Kornberg A (1962) Enzymatic synthesis of deoxyribonucleic acid VIII. Frequencies of nearest neighbour base sequences in deoxyribonucleic acid. J Biol Chem 236:864-875

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21-123

Li W-H (1983) Evolution of duplicate genes and psuedogenes. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland MA, pp 1-94

Li W-H, Luo C-C, Wu C-I (1985) Evolution of DNA sequences. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum Press, New York, pp 1-94

Lindley PF, Narebor ME, Summers LJ, Wistow GJ (1985) The structure of lens proteins. In: Maisel H (ed) The ocular lens. Marcel Dekker, New York, pp 123-167

Lok S, Tsui L-C, Shinohara T, Piatigorsky J, Gold R, Breitman ML (1984) Analysis of the mouse γ-crystallin gene family: assignment of multiple cDNAs to discrete genomic sequences and characterization of a representative gene. Nucl Acids Res 12:4517-4529

Lok S, Breitman ML, Chepelinsky AB, Piatigorsky J, Gold RJM, Tsui L-C (1985) Lens-specific promotor activity of a mouse γ-crystallin gene. Mol Cell Biol 5:2221-2230

Max EE (1985) CG doublet difficulties in vertebrate DNA. Nature 313:4-37-438

McClelland M, Ivarie R (1982) Asymmetrical distribution of CpG in an `average' mammalian gene. Nucl Acids Res 10:7865-7877

Meakin SO, Breitman ML, Tsui L-C (1985) Structural and evolutionary relationships among five members of the human γ-crystallin gene family. Mol Cell Biol 5:1408-1414

Meakin SO, Du RP, Tsui L-C, Breitman ML (1987) γ-crystallins of the human eye lens: expression analysis of five members of the gene family. Mol Cell Biol 7:2671-2679

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23-36

Moormann RJM, den Dunnen JT, Mulleners L, Andreoli PM, Bloemendal H, Schoenmakers JGG (1983) Strict colinearity of genetic and protein folding domains in an intragenically duplicated rat lens γ-crystallin gene. J Mol Biol 171:353-368

Moormann RJM, den Dunnen JT, Heuyerjans J, Jongbloed RJE, van Leen RW, Lubsen NH, Schoenmakers JGG (1985) Characterization of the rat γ-crystallin gene family and its expression in the eye lens. J Mol Biol 182:419-430

Piatigorsky J (1984) Lens crystallins and their gene families. Cell 38:620-621

Shiloh Y, Donlon T, Bruns G, Breitman ML, Tsui, L-C (1986) Assignment of the human γ-crystallin gene cluster (CRYG) to the long arm of

chromosome 2, region q33-q36. Hum Genet 73:17-19
Slingsby C, Croft LR (1978) Structural studies on calf lens γ-crystallin fraction IV: a comparison of the cysteine containing tryptic peptides with the corresponding amino acid sequence of the γ-crystallin fraction II. Exp Eye Res 37:291-304
Staden R (1982) An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucl Acids Res 10:29-51-2961
Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652-656
Van Leen RW, van Roozendaal KEP, Lubsen NH, Schoenmakers JGG (1987) Differential expression of crystallin genes during development of the rat eye lens. Dev Biol 120:457-464
Willard HF, Meakin SO, Tsui L-C, Breitman ML (1985) Assignment of human gamma crystallin multigene family to chromosome 2. Somat Cell Mol Genet 11:511-516

CHAPTER 8

DIFFERENT EVOLUTION RATES WITHIN THE LENS-SPECIFIC ß-CRYSTALLIN GENE FAMILY

Henk J.M. Aarts, Eldine H.M. Jacobs,

Gijsbert van Willigen, Nicolette H. Lubsen

and John G.G. Schoenmakers

J Mol Evol, in press.

DIFFERENT EVOLUTION RATES WITHIN THE LENS-SPECIFIC β-CRYSTALLIN GENE FAMILY.

SUMMARY

We have determined the sequence of a rat βA3/A1-crystallin cDNA clone and the (partial) sequence of the human βB3-crystallin gene. Calculation of the ratio of silent to nonsynonymous substitution between orthologous βA3/A1-, βB3- and other β- and γ-crystallin sequences revealed that the region encoding the two globular domains of the βA3/A1-crystallin sequence is the best conserved during evolution, much better than the corresponding region of the βB1-, βB3- or the γ-crystallin sequences, and even better (at least in the rodent/frog comparison) than the well conserved αA-crystallin sequence. Remarkably, the rate of change of the βA3/A1-crystallin coding sequence does not differ in the rodent and primate lineages, in contrast with previous findings concerning the evolution rates of the αA- or γ-crystallin sequences in these two lineages. Comparison of the regions which encode the four motifs of the β-crystallin between orthologous mammalian sequences showed that the extent of nonsynonymous substitution in each of these four homologous motif regions is the same. However when the orthologous β-crystallin genes of more distantly related species (mammals vs chicken or frog) are compared the extent of nonsynonymous substitution is higher in the regions encoding the external motifs I and III than in the regions encoding the internal motif II and IV. This phenomenon is also observed when paralogous members of the β-/γ-crystallin super-gene family are compared.

INTRODUCTION

The crystallins are the major water-soluble structural proteins of the vertebrate lens. These proteins are divided into four major families, the α-, β-, γ-, and δ-crystallins. All vertebrate lenses contain α- and β-crystallins. The γ-crystallins are absent in bird and reptile lenses which contain δ-crystallins instead (for reviews; see Bloemendal 1985; Wistow and Piatigorsky 1988).

Both the β- and the γ-crystallin proteins are built up out of four homologous "Greek key" motifs organized into two domains (Blundell et al. 1981; Wistow et al. 1981, 1983; Inana et al. 1983; see Fig. 1). The protein structure of the β-crystallins differs markedly only from the γ-crystallins in the presence of a N- and sometimes a C-terminal extension (Wistow et al. 1981). The similarity in protein structure between the β- and γ-crystallins suggested that these proteins might have evolved from a common ancestral four motif/two domain protein. However, elucidation of the gene structures has shown that the evolutionary relation between these two families is likely to be more distant. Wit-

hin the β-crystallin gene, each of the four homologous protein motifs is encoded by a separate exon (Inana et al. 1983; Den Dunnen et al. 1986a; Hogg et al. 1986; see also Fig. 2), while in the γ-crystallin

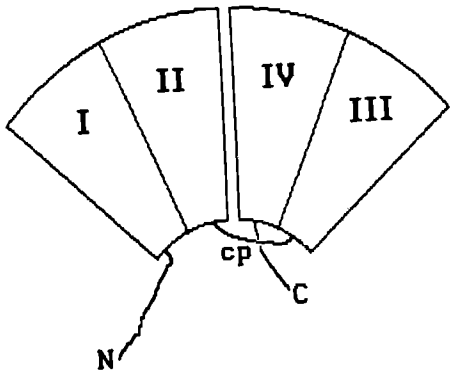


Fig. 1. Schematic representation of the three dimensional structure of a β-crystallin protein. The four motifs are indicated with the corresponding roman numerals, the connecting peptide is shown as c.p.. The N- and C-terminal extensions are indicated by N and C respectively.

genes two motifs share an exon (Moormann et al. 1983; Lok et al. 1984; Den Dunnen et al. 1986b). Hence the β- and γ-crystallin gene families probably diverged at the level of a two motif/one domain ancestral gene as outlined in Fig. 2. Further successive gene duplications within the β- and γ-crystallin gene branches would then have led to the two gene families found today. The first duplication of the ancestral β-crystallin gene probably gave rise to the progenitors of the present acidic (βA3/A1, βA2, and βA4) and basic (βB1, βB2, and βB3) β-crystallins (Berbers et al. 1984). These two groups differ in the presence of C-terminal extensions (absent in the acidic group) and in small insertions/deletions in the protein motifs (Berbers et al. 1984). The sequence data presently available indicate that the formation of the β-crystallin gene family may have preceded the vertebrate radiation (Berbers et al. 1984; Quax-Jeuken et al. 1985; see also below). In contrast, separate bursts of duplication of the γ-crystallin genes probably occurred within the amphibian and mammalian lineages (Quax-Jeuken et al. 1985).

The β- and γ-crystallin genes appear to have followed parallel evolutionary paths. It is therefore of interest to compare the rate and mode of evolution of the β- and γ-crystallin genes. We have previously

shown that all functional γ-crystallin genes evolve at the same rate and, at least in part, in concert (Aarts et al. 1988). Using the published sequence data and the two additional β-crystallin sequences reported here, we show that the β-crystallin genes evolve independently and at different rates. The lowest rate was found for the βA3/A1-crystallin. This protein is, in contrast to the αA- and γ-crystallins (De Jong 1981; Aarts et al. 1988), as well conserved in the primate as in the rodent lineage.

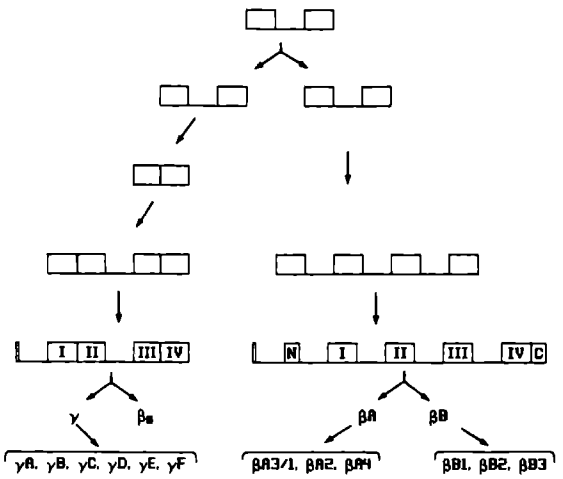


Fig. 2. Schematic representation of the evolutionary history of the β-/γ-crystallin gene family. The presumptive ancestral motif coding region is indicated by an open box. The structures of the present γ- and β-crystallin genes (structure given is for the βB1-crystallin gene; Den Dunnen et al. 1986a, the βA3/A1-crystallin gene differs in that an additional translation initiation site is found in the first exon; Hogg et al. 1986; Peterson and Piatigorsky 1986) are shown. The motif coding regions are indicated with roman numerals, the regions encoding the N- and C-terminal extension with N and C respectively. The first exon of the γ- and β-crystallin gene is indicated by a small box. The known members of the present γ- and β-crystallin gene families are given at the bottom of the figure.

MATERIALS AND METHODS

cDNA isolation

Lens RNA was prepared from 5 weeks old rats as described by Van Leen et al. (1987). cDNA was synthesized using oligo (dT) as primer according to the method of Lapeyre and Almaric (1985). EcoRI linkers were ligated to the blunt-ended cDNA, cleaved with EcoRI and the resulting molecules were cloned in the EcoRI site of the vector λgt11. The phage DNA molecules were packaged and plated on Y1090R cells. The λgt11-cDNA library was screened with the nick-translated PstI-insert of the calf βA3/A1-crystallin cDNA clone pBLβA(1+3) (Quax-Jeuken et al. 1984). Positive plaques were purified by successive rounds of replating and screening. Hybridization conditions were as described in Moormann et al. (1984). DNA from the positive plaques was isolated, digested and analysed by gel electrophoresis.

Nucleotide sequencing

M13mp subclones of the inserts of rat βA3/A1 λgt11 clones were used to determine the nucleotide sequence according to the dideoxy chain termination method (Sanger et al. 1977). The partial sequence of the human βB3-crystallin gene was determined by sequencing M13mp subclones of the genomic clone λHeB3-1 (Aarts et al. 1987).

Sequence comparisons

The rat βA3/A1-crystallin cDNA sequence and the partial sequence of the human βB3-crystallin gene are given in this report. The rat βB1- and βB3-crystallin sequences can be found in Den Dunnen et al. (1985a), the mouse βA3/A1 sequence in Inana et al. (1983), the calf βA3/A1 sequence in Quax-Jeuken et al. (1984), the human βA3/A1 sequence in Hogg et al. (1986), the chicken βA3/A1 sequence in Peterson and Piatigorsky (1986), the frog βA3/A1 sequence in Luchin et al. (1987), the calf βB1 sequence in Quax-Jeuken et al. (1984) and the chicken βB1 sequence in Hejtmancik et al. (1986). The rat αA sequence is given in Moormann et al. (1981), the mouse αA sequence in King et al. (1982) and the frog αA in Tomarev et al. (1983). The mammalian γ-crystallin sequences are compiled in Aarts et al. (1988), the frog γ-crystallin sequences are reported by Tomarev et al. (1984). For sequence comparison, only the regions encoding the globular domains of the proteins were used. Triplet insertions/deletions were excluded from the comparisons between the members of the rat β-/γ-crystallin super-gene family (see Fig. 6b). The extent of synonymous and nonsynonymous substitution was calculated according to Miyata and Yasunaga (1980) and was corrected for back mutation by the method of Jukes and Cantor (1969).

Nomenclature

For the β-crystallins the nomenclature proposed by Berbers et al. (1984) is used, for the γ-crystallins we have followed the nomenclature suggested by Aarts et al. (1988).

RESULTS

The analysis of the rate of evolution of the β-crystallins has, up to now, been precluded by the paucity of information about orthologous sequences from different species. We have therefore established the sequence of two β-crystallins, of which the orthologous sequence in another mammalian species is known.

The sequence of the rat βA3/A1- and human βB3-crystallins

To establish the sequence of the rat βA3/A1-crystallin, a rat lens cDNA bank was screened with the calf βA3/A1 cDNA clone and the inserts of the positive clones were sequenced. This sequence is shown in Fig. 3. As the cDNA clones were incomplete at their 5' end, the sequence lacks the N-terminal extension and the first six codons of the first motif sequence. The cDNA clones do contain the complete 3' non-coding region. It has been previously noted that the 3' non-coding region of the βA3/A1-crystallin gene is relatively well conserved: the human and the mouse sequences were found to be 83% similar (Hogg et al. 1986; compared to the 65% sequence similarity found in general between human and murine intronic or gene flanking sequences, Soares et al. 1985; Den Dunnen et al. 1988). The much closer related rat and mouse sequences, however, differ by almost the same extent: alignment of these non-coding regions shows 88% sequence similarity (Fig. 3). The conservation of this region is thus much less obvious in the rodent lineage.

To obtain the coding sequence of the human βB3-crystallin gene, the previously described genomic clone of this gene was further subcloned (see Materials and Methods). As we failed to obtain subclones from the 5' region of the gene, we could only sequence the exons encoding motifs II, III, IV and the C-terminal extension of the protein (Fig. 4). Our sequence data show that the human βB3 gene has the same structure as the rat βB1- (Den Dunnen et al. 1986a) and the human βA3/A1-crystallin genes (Hogg et al. 1986) in that each of the four motifs of the protein is encoded by a separate exon (see also Fig. 2). The 3' non-coding region of the βB3-crystallin gene is less well conserved than the 3' non-coding region of the βA3/A1-crystallin gene. The rat sequence is 65 bp shorter than the human sequence and aligns best, albeit with only 63% sequence similarity, with the 3' region of the human non-coding sequence (see Fig. 4).

In both the rat βA3/A1- and the human βB3-crystallin sequence all amino acid residues known to be involved in the "Greek key" fold of the motifs are conserved, hence these two crystallins conform to the general β- and γ-crystallin folding pattern.

The rate of evolution of the β-crystallins

To calculate the rate of evolution of the region encoding the four

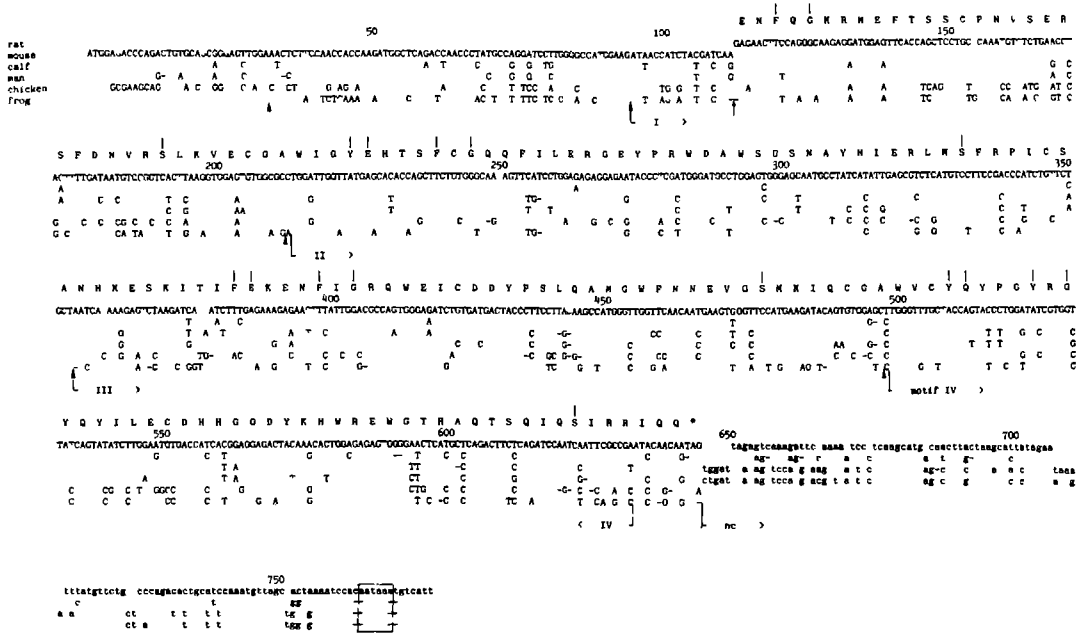


Fig. 3. Nucleotide and deduced amino acid sequence of the rat β-crystallin cDNA λRLβA3/A1-crystallin, and comparison with the mouse, calf, man, chicken and frog βA3/A1-crystallin sequences. The deduced rat amino acid sequence is shown above the nucleotide sequences. The poly-A addition signal is boxed, the termination codon is shown by an asterisk. The residues involved in the characteristic 'Greek key' folding of the proteins are indicated by a (|). The motif coding regions are indicated with the corresponding roman numeral. The arrow indicates the position of the (truncated) 5' end of the rat βA3/A1-crystallin cDNA in λRLβA3/A1. The arrowheads show the exact location of the introns as determined in the human gene (Hogg et al. 1986). Only the rat βA3/A1-crystallin sequence is shown completely. For the other sequences the nucleotides that differ are shown and the identical nucleotides are indicated by a dash. Gaps were introduced to allow maximal alignment of the 3' untranslated regions (nc). The 3' untranslated regions of the chicken and frog sequences are not shown.

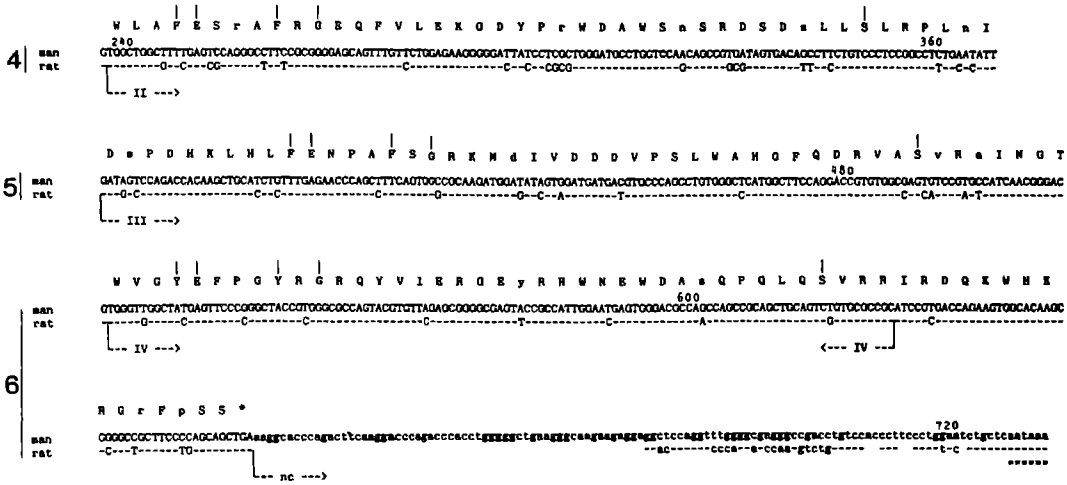


Fig. 4. Nucleotide and deduced amino acid sequence of exon 4, 5 and 6 of the human βB3-crystallin gene. Capital letters indicate coding regions, lower case letters the 3' non-coding region. The intron/exon junctions of the human sequence were deduced from a comparison with the rat βB3-crystallin cDNA sequence (Den Dunnen et al. 1985a). The numbering of the exons is by analogy with that of the rat βB1- (Den Dunnen et al. 1986a) and the human βA3/A1-crystallin gene (Hogg et al. 1986). The human sequence is shown completely, for the rat sequence only the nucleotides that differ are shown. Identical nucleotides are indicated by a dash. Gaps were introduced in the rat sequence to allow maximal alignment of the 3' untranslated regions (nc). The derived human amino acid sequence is given above the nucleotide sequences. The residues involved in the characteristic 'Greek key' folding of the protein are indicated by a (|), the residues which differ from the deduced rat amino acid sequence (Den Dunnen et al. 1985a) are given in lower case letters. The termination codon is indicated by an asterisk and the poly-A addition signal is underlined. The motif coding regions II, III and IV are indicated.

motif/two domains of the βA3/A1- and the βB3-crystallin protein, the rat βA3/A1- or the human βB3-crystallin sequences were aligned with orthologous sequences from other species (see Figs. 3 and 4). The extent of synonymous and nonsynymous substitution between the various species was then calculated according to Miyata and Yasunaga (1980). For comparison, the extent of synonymous and nonsynonymous substitution between the known βB1-, αA- and γ-crystallin sequences was calculated in the same manner. As shown in Table 1, the βA3/A1-crystallin sequence is by far the least variable β-crystallin sequence. It is even better conserved than αA-crystallin as the extent of nonsynonymous substitution in the βA3/A1-crystallin sequence between rat and frog is only 0.072, while the extent in αA-crystallin is 0.135. This protein also appears to be better conserved than for instance the very well conserved Na,K-ATPase β-subunit or the acetylcholine receptor γ-subunit (Li et al. 1987). The extent of change in the γ-crystallins between mouse and rat is almost as low as that in βA3/A1-crystallin, but in general the γ-crystallins are more variable than the αA- or β-crystallins (note that the mammalian and frog γ-crystallins are probably paralogous, see Quax-Jeuken et al. 1985 and Discussion).

One measure of the relative rate of change in a protein that is independent of an evolutionary timescale is the ratio of the extent of synonymous and nonsynonymous substitution, Ks/Ka. As shown in Table 1, this ratio is higher and more uniform for the βA3/A1-crystallin sequence than for the βB1- or the γ-crystallins, which again points to a lower rate of evolution of the βA3/A1-crystallin protein. Most remarkable is the fact that for the βA3/A1-crystallin the ratio in the rat-man comparison equals that of the rat-mouse comparison, showing equal conservation in both lineages, while both the αA- and the γ- crystallins are less well conserved in the primate lineage than in the rodent lineage (De Jong 1981, Aarts et al. 1988; see also Table 1). The data are insufficient to show whether the equal conservation in the rodent and primate lineages holds true for all β-crystallins, but the low Ks/Ka found in the rat vs man comparison of the βB3-crystallin sequence suggests that this sequence could well be also more variable in the primate lineage than in the rodent lineage. Note that the low Ks/Ka values found in the mammal vs chicken comparison of the βB1-crystallin sequence are a consequence of anomalously low Ks values and are probably not significant.

Table 1

gene	comparison	synonymous substitution (Ks)	nonsynonymous substitution (Ka)	Ks/Ka
βA3/A1	rat vs mouse	0.228	0.011	20.7
	rat vs calf	0.705	0.019	37.1
	rat vs man	0.576	0.027	21.3
	rat vs chicken	>1	0.068	--
	rat vs frog	>1	0.072	--
	mouse vs calf	0.663	0.021	31.6
	mouse vs man	0.653	0.033	19.5
	mouse vs chicken	>1	0.075	--
	mouse vs frog	>1	0.078	--
	calf vs man	0.472	0.032	14.8
	calf vs chicken	>1	0.071	--
	calf vs frog	>1	0.079	--
	man vs chicken	>1	0.089	--
	man vs frog	>1	0.086	--
	chicken vs frog	>1	0.067	--
βB1	rat vs calf	0.673	0.034	19.8
	rat vs chicken	0.950	0.127	7.5
	calf vs chicken	0.782	0.114	6.9
βB3	rat vs man (motif II,III,IV)	0.561	0.058	9.6
αA	rat vs mouse	0.260	0.000	>>
	rat vs frog	>1	0.135	--
	mouse vs frog	>1	0.132	--
γ	rat vs mouse	0.209	0.013	16.1
	rat vs calf	0.602	0.051	11.8
	rat vs man	0.549	0.078	7.0
	rat vs frog	>1	0.243	--
	mouse vs calf	0.628	0.060	10.5
	mouse vs man	0.495	0.084	5.9
	mouse vs frog	>1	0.243	--
	calf vs man	0.438	0.075	5.9
	calf vs frog	>1	0.286	--
	man vs frog	>1	0.224	--

Table 1: The extent of synonymous (Ks) and nonsynonymous substitution (Ka) and the ratios of synonymous and nonsynonymous substitutions in the coding region of the αA-crystallin gene, and region encoding the globular domains of the βA3/A1-, βB1-, βB3- and γ-crystallin genes. The extent of synonymous and nonsynonymous substitution was calculated as indicated in Materials and Methods. The data for the γ-crystallin genes were taken in part from Aarts et al. (1988). Sequence references can be found in Materials and Methods.

The rate of change of the β-crystallin motif sequences.

The γ-crystallin proteins and the core of the β-crystallin proteins are built up out of four motifs, two of which are internal and two of which are external (see Fig. 1). We have previously shown that the motifs of the γ-crystallin show the same rate of change when orthologous γ-crystallins were compared. However, when paralogous γ-crystallins were compared, the sequences of the two internal motifs, motifs II and IV, were found to be less variable than the external motifs I and III (Aarts et al. 1988). The γ-crystallins are symmetrical proteins and motif I is topologically equivalent to motif III (Blundell et al. 1981). Due to the presence of the N- and C-terminal extensions the β-crystallins are not symmetrical. This asymmetry could influence the rate of change of motif I or III with respect to the rate of change of motifs II and IV. We therefore calculated separately the extent of synonymous and nonsynonymous substitution in the sequences encoding each of the four motifs of the protein core of β-crystallins. The results for the βA3/A1-crystallin sequences are shown in Fig. 5. No significant difference between the extent of change of each motif is seen when mammalian sequences are compared. However, when mammalian sequences are compared with those of more distantly related species, e.g. frog and chicken, the extent of nonsynonymous substitution is always lower in motifs II and IV, the two internal motifs of the protein, than in motifs I and III, the two outside protein motifs. This substitution pattern is not unique for the βA3/A1-crystallin gene as the same phenomenon is also seen for the βB1-crystallin gene (see Fig. 5): comparison of the rat or calf sequences with the chicken sequence also shows a better conservation of motifs II and IV than of motifs I and III.

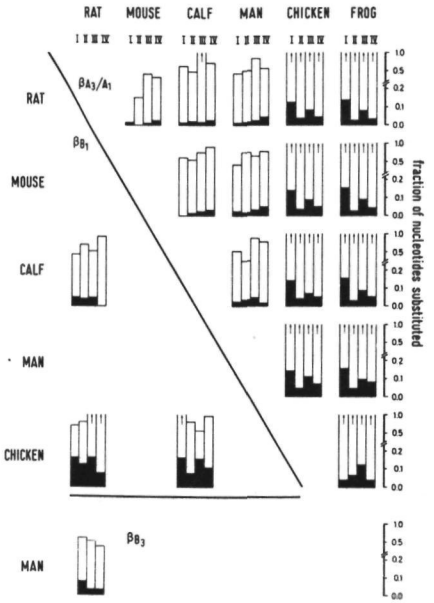
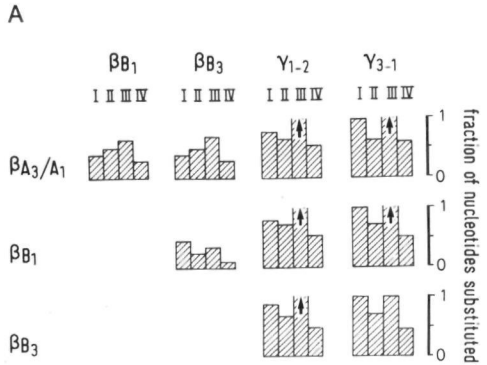


Fig. 5. Extent of synonymous and nonsynonymous substitution in orthologous βA3/A1-, βB1- and βB3-crystallin genes. The comparison between the βB1-crystallin genes is shown in the lower left corner, that between the βA3/A1-crystallin genes in the upper right and that between the rat and human βB3-crystallin sequences under the horizontal line. The extent of substitution was calculated as described in Materials and Methods. Open bars indicate the extent of synonymous substitution, filled bars that of nonsynonymous substitution. Arrows show where the fraction of nucleotides synonymously substituted is greater than one. Roman numerals indicate motif regions I-IV, and correspond to the motif regions indicated in Figs. 3 and 4. The first 18 nucleotides of motif region I were excluded from the βA3/A1 comparisons. The sequence references are given in Materials and Methods. Note the change in scale of the y-axis.



B

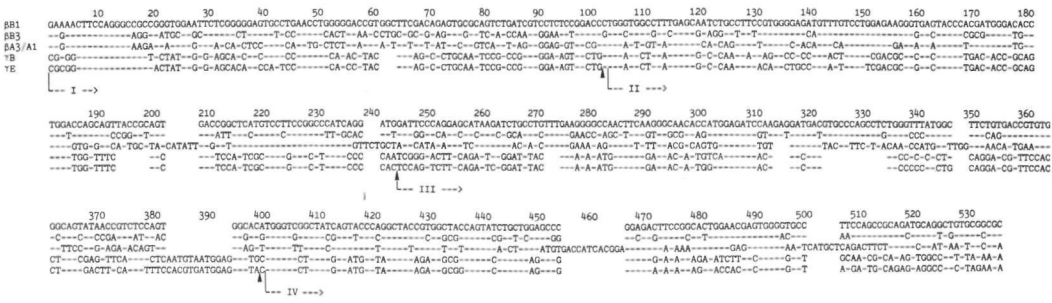


Fig. 6. (a): Extent of nonsynonymous substitution in the rat β- and γ-crystallin genes. The extent of substitution between paralogous gene pairs as indicated was calculated as described in Materials and Methods. The four motif coding regions are indicated by the roman numerals I-IV and correspond to the motif coding regions indicated in Fig. 6b. Arrows show when the fraction of nonsynonymous substitutions is greater than one. The first 18 nucleotides of motif region I, insertions and the triplet (cac) encoding the additional amino acid found in the connecting peptide of the rat γB sequence were excluded from the comparisons.

(b): Comparison of the rat βB1-, βB3-, βA3/A1-, γB- and γE-crystallin nucleotide sequences encoding the globular domains of the corresponding polypeptides. Only the rat βB1-crystallin sequence is shown completely. For the other sequences the nucleotides that differ are shown and the identical nucleotides are indicated by a dash. Gaps were introduced to allow maximal alignment. The four motif coding regions are indicated. Arrowheads show the locations of the intradomain (β-crystallin genes) and the interdomain (β- and γ-crystallin genes) introns.

Finally, a comparison of the human and rat sequences of a second member of the basic β-crystallin gene family, the βB3 gene, shows no difference between the extent of substitution between motifs III and IV, but an unexpectedly high extent of nonsynonymous substitutions within motif II is observed. Thus in general there is no detectable difference between the rate of change of the internal motifs II and IV and that of the external motifs I and III when orthologous mammalian β-crystallins are compared.

When orthologous β-crystallin proteins from distantly related species are compared, the internal motifs II and IV of the protein are found to be better conserved than the external motifs I and III (Fig. 5). One might then expect that a higher sequence similarity in motif regions II and IV would also be seen when paralogous β-crystallin sequences of a single species are compared (Fig. 6a). This is indeed found when the βB1- and βB3-crystallin sequences are compared, but not when the βA3/A1-crystallin sequence is compared with either one of the two basic β-crystallin sequences. In these cases an unexpected low extent of substitution is seen in motif I. Finally, between the β-crystallins and the γ-crystallins motifs II and IV are again best conserved, while motif III is the most variable (Fig. 6a).

DISCUSSION

During our sequence comparison of the various rat β-crystallin coding sequences we encountered no evidence for the occurrence of concerted evolution within this gene family, that is we found no sequence stretches with an unexpectedly high similarity. The lack of concerted evolution may be a consequence of the genomic organization of the β-crystallin genes. Only the βB2- and βB3-crystallin genes are closely linked within the (human or rat) genome (Aarts et al. 1987). The location of the βB1-crystallin gene is unknown, but it must be located at least 30 kb away from the βB2-/βB3-crystallin gene cluster (Moormann et al. 1984). The βA3/A1-crystallin gene is not linked to the βB2-/βB3-crystallin gene cluster as it is located on another chromosome (in man: βA3/A1 on chromosome 17; Law et al. 1986 vs βB2 on chromosome 22; Hogg et al. 1987). The γ-crystallin genes, which do show concerted evolution, are closely linked (Moormann et al. 1985; Den Dunnen et al. 1985b; Meakin et al. 1985).

If we assume that the β-crystallin genes indeed evolve indepen-

dently, then the divergence between the coding sequences may be used to estimate the time at which these genes diverged. From the overall extent of nonsynonymous substitution between the rat β-crystallins (e.g. Fig. 5) and the data presented in Table 1 we suggest that the split between acidic and basic β-crystallin genes already existed at the time of the vertebrate radiation, and that the divergence within the basic β-crystallin group probably occurred before the amphibian split. Isolation and characterization of the basic β-crystallins from these species is required to confirm this suggestion. Within the γ-crystallin gene family gene conversion masks the evolutionary path of motif regions I and II (Aarts et al. 1988), and therefore the time of divergence of the members of this gene family was estimated from the extent of nonsynonymous substitution within the motif coding regions III and IV (see Aarts et al. 1988) and the data presented in Table 1. Our calculation suggests, assuming that the artiodactyl lineage separated from the rodent /primate lineage 80 myr ago (Dayhoff 1972), that the mammalian γ-crystallin genes diverged 180-240 myr ago. This implies that the divergence of these genes took place before the mammalian but after the vertebrate radiation, in agreement with a previous suggestion that the frog and mammalian γ-crystallin genes are paralogous (Quax-Jeuken et al. 1985).

In our analysis of the rates of changes of the regions encoding each of the six protein domains of a β-crystallin (e.g. the N- and C-terminal extensions and the four motifs of the protein core) we have only included the regions encoding the four motifs. The few available data show that the N-terminal extensions change faster than the core regions. In the case of the βA3/A1-crystallin sequence, this change is a gradual one (Peterson and Piatigorsky 1986; see also Fig. 3), while in the case of the N-terminal extension of the βB1-crystallin clear evidence for rapid change due to insertion of simple sequence elements, perhaps by a "slippage" mechanism (Tautz et al. 1986), has been presented (Quax-Jeuken et al. 1984; Den Dunnen et al. 1985a). We had expected that the presence of N-terminal and C-terminal extensions on the β-crystallin protein would influence the rate of change of the neighboring motifs of the protein core. However, the overall trend discernable from the analysis of the extent of change of each of the four motifs of the β-crystallin proteins is the same as that found previously within the γ-crystallin proteins, which lack the extensions, namely a stronger conservation of motifs II and IV than of motifs I and III (Aarts et al.

1988). The only difference is that motif III was found to be most variable within the γ-crystallin gene family, while at least between orthologous βA3/A1-crystallins motif I appears to show the greatest variation. To what extent the differences in the rate of change of each of the four motifs are a consequence of differences in functional constraint is not known. A close-range interaction between the lens structural proteins is thought to be required for the establishment and maintenance of the transparency of the lens (Benedek 1971; Delaye and Tardieu 1983). Hence both the oligomeric β- and the monomeric γ-crystallins are likely to be involved in close protein-protein contacts. Which (parts) of the motifs make these contacts is, however, unknown. Additional protein contacts must be encountered by the β-crystallins as they are present in the lens as homo- and hetero-oligomers. The better conservation of motif III in the βA3/A1-crystallin may point to a close interaction of this motif in the β-crystallin oligomer, in agreement with the finding that surface hydrophobic patches are located in this motif (Luchin et al. 1987).

To the extent that one may equate a slow rate of evolution with functional constraint, our consideration of the rate of evolution of the crystallins suggests that the βA3/A1-crystallin protein plays a crucial role in all vertebrate lenses, while the requirements for the basic β-crystallins or the γ-crystallins are somewhat more flexible. The γ-crystallins can even be dispensed with as they are absent in chicken and reptile lenses.

ACKNOWLEDGEMENT We thank Jack Leunissen for his help with the computer analysis and Jacques Janssen for his assistance in the preparation of the cDNA library. This research was carried out under the auspices of the Netherlands Foundation for Chemical Research (SON), with financial aid from the Netherlands Organization for the Advancement of pure Research (NWO).

Aarts HJM, Den Dunnen JT, Lubsen NH, Schoenmakers JGG (1987) Linkage between the βB2 and βB3 crystallin genes in man and rat: a remnant of an ancient β-crystallin gene cluster. Gene 59:127-135

Aarts HJM, Den Dunnen JT, Leunissen J, Lubsen NH, Schoenmakers JGG (1988) The γ-crystallin gene families: Sequence and evolutionary patterns. J Mol Evol 27:163-172

Benedek GB (1971) Theory of transparency of the eye. Appl Opt 10:459-473

Berbers GAM, Hoekman WA, Bloemendal H, De Jong WW, Kleinschmidt T, Braunitzer G (1984) Homology between the primary structures of the major bovine β-crystallin chains. Eur J Biochem 139:467-479

Bloemendal H (1985) Lens research: from protein to gene. Exp Eye Res 41:429-448

Blundell T, Lindley P, Miller L, Moss D, Slingsby C, Tickle I, Turnell B, Wistow G (1981) The molecular structure and stability of the eye lens: X- ray analysis of γ-crystallin II. Nature 289:771-777

Dayhoff MO (1972) Atlas of protein sequences and structure, vol 5. National Biomedical Research Foundation, Silver Spring MD

De Jong WW (1981) Evolution of lens and crystallins. In: Bloemendal H (ed) Molecular and cellular biology of the eye lens. John Wiley and Sons, New York, pp 221-278

Delaye M, Tardieu A (1983) Short-range order of crystallin proteins accounts for eye lens transparancy. Nature 302:415-417

Den Dunnen JT, Moormann RJM, Schoenmakers JGG (1985a) Rat lens β-crystallins are internally duplicated and homologous to γ-crystallins. Biochim Biophys Acta 824:295-303

Den Dunnen JT, Moormann RJM, Cremers FPM, Schoenmakers JGG (1985b) Two human γ-crystallin genes are linked and riddled with alu-repeats. Gene 38:197-204

Den Dunnen JT, Moormann RJM, Lubsen NH, Schoenmakers JGG (1986a) Intron insertions and deletions in the β-/γ-crystallin gene family: The rat βB1-gene. Proc Natl Acad Sci USA 83:2855-2859

Den Dunnen JT, Moormann RJM, Lubsen NH, Schoenmakers JGG (1986b) Concerted and divergent evolution within the rat γ-crystallin gene family. J Mol Biol 189:37-46

Den Dunnen JT, Van Neck JW, Cremers FPM, Lubsen NH, Schoenmakers JGG (1988) Nucleotide sequence of the rat γ-crystallin gene region and comparison with an orthologous human region. Gene in press

Hejtmancik JF, Thompson MA, Wistow G, Piatigorsky J (1986) cDNA and deduced protein sequence for the βB1-crystallin polypeptide of the chicken lens. Conservation of the PAPA sequence. J Biol Chem 261: 982-987

Hogg D, Tsui L-C, Gorin M, Breitman ML (1986) Characterization of the human β-crystallin gene HuβA3/A1 reveals ancestral relationships among the βγ-crystallin superfamily. J Biol Chem 261:12420-12427

Hogg D, Gorin MB, Heinzmann C, Zollman S, Mohandas T, Klisak I, Sparkes RS, Breitman M, Tsui L-C, Horwitz J (1987) Nucleotide sequence for the cDNA of the bovine βB2 crystallin and assignment of the orthologous human locus to chromosome 22. Curr Eye Res 6:1335-1342

Inana G, Piatigorsky J, Norman B, Slingsby C, Blundell T (1983) Gene and protein structure of the β-crystallin polypeptide in murine lens: relationship of exons and structural motifs. Nature 302:310-315

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN

(ed) Mammalian protein metabolism. Academic Press, New York, pp 21-123

King CR, Shinohara T, Piatigorsky J (1982) αA-crystallin messenger RNA of the mouse lens: more noncoding than coding sequences. Science 215:985-987

Lapeyre B, Almaric F (1985) A powerful method for the preparation of cDNA libraries: isolation of cDNA encoding a 100-kDal nucleolar protein. Gene 37:215-220

Law ML, Cai C-Y, Hartz J, Kao F-T, Hogg D, Breitman ML, Tsui L-C (1986) Localization of a β-crystallin gene, the huβA3/1 (gene symbol, CRYB1), to the long arm of human chromosome-17. Cytogenet Cell Genet 42:202-207

Li W-H, Tanimura M, Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J Mol Evol 25:330-342

Lok S, Tsui L-C, Shinohara T, Piatigorsky J, Gold R, Breitman ML (1984) Analysis of the mouse γ-crystallin gene family: assignment of multiple cDNAs to discrete genomic sequences and characterization of a representative gene. Nucl Acids Res 12:4517-4529

Luchin SV, Zinovieva RD, Tomarev SI, Dolgilevich SM, Gause Jr GG, Bax B, Driessen H, Blundell TL (1987) Frog lens βA1-crystallin: the nucleotide sequence of the cloned cDNA and computer graphics modelling of the three-dimensional structure. Biochim Biophys Acta 916:163-171

Meakin SO, Breitman ML, Tsui L-C (1985) Structural and evolutionary relationships among five members of the human γ-crystallin gene family. Mol Cell Biol 5:1408-1414

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23-36

Moormann RJM, Van der Velden HWM, Dodemont HJ, Andreoli PM, Bloemendal H, Schoenmakers JGG (1981) An unusually long non-coding region in the rat lens α-crystallin messenger RNA. Nucl Acids Res 9:4813-4822

Moormann RJM, Den Dunnen JT, Mulleners L, Andreoli PM, Bloemendal H, Schoenmakers JGG (1983) Strict colinearity of genetic and protein folding domains in an intragenically duplicated rat lens γ-crystallin gene. J Mol Biol 171:353-368

Moormann RJM, Jongbloed R, Schoenmakers JGG (1984) Isolation and characterization of β- and γ-crystallin genes from rat genomic cosmid libraries. Gene 29:1-9

Moormann RJM, Den Dunnen JT, Heuyerjans J, Jongbloed RJE, Van Leen RW, Lubsen NH, Schoenmakers JGG (1985) Characterization of the rat γ-crystallin gene family and its expression in the eye lens. J Mol Biol 182, 419-430

Peterson CA, Piatigorsky J (1986) Preferential conservation of the globular domains of the βA3/A1-crystallin polypeptide of the chicken eye lens. Gene 45:139-147

Quax-Jeuken Y, Janssen G, Quax W, van den Heuvel R, Bloemendal H (1984) Bovine β-crystallin complementary DNA clones. Alternating proline/alanine sequence of the βB1 subunit originates from a repetitive DNA sequence. J Mol Biol 180:457-472

Quax-Jeuken Y, Driessen H, Leunissen J, Quax W, De Jong W, Bloemendal H (1985) βs-crystallin: Structure and evolution of a distinct member of the β/γ-superfamily. EMBO J 4:2597-2602

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463-5467

Soares MB, Schon E, Henderson A, Karathanasis SK, Cata R, Zeitlin S, Chirgwin J, Efstratiadis A (1985) RNA-mediated gene duplication: the rat preproinsulin gene is a functional retroposon. Mol Cell Biol 5:2090-2103

Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652-656

Tomarev SI, Zinovieva RD, Dolgilevich SM, Krayev AS, Skryabin KG, Gause Jr GG (1983) The absence of the long 3'-non-translated region in mRNA coding for eye lens αA2-crystallin of the frog (Rana temporaria). FEBS Lett 162:47-51

Tomarev SI, Zinovieva RD, Chalovka P, Krayev AS, Skryabin KG, Gausse Jr GG (1984) Multiple genes coding for the frog eye lens γ-crystallin. Gene 27:301-308

Van Leen RW, Van Roozendaal KEP, Lubsen NH, Schoenmakers JGG (1987) Differential expression of crystallin genes during development of the rat eye lens. Devel Biol 120:457-464

Wistow GJ, Slingsby C, Blundell TL, Driessen HPC, De Jong WW, Bloemendal H (1981) Eye lens proteins: the three dimensional structure of β-crystallin predicted from monomeric γ-crystallin. FEBS Lett 133:9-16

Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T (1983) X-ray analysis of the lens protein γ-II crystallin at 1.9 A resolution. J Mol Biol 170:175-202

Wistow GJ, Piatigorsky J (1988) Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. Ann Rev Biochem 57:479-504

SUMMARY AND CONCLUSIONS
SAMENVATTING EN CONCLUSIES

The lens-specific β-crystallin protein family is composed of a group of rather heterogenous proteins. To get further insight in the structure, sequence, evolution and developmental pattern of expression of the genes encoding this protein family, we first of all set out to isolate and characterize human and rat genomic clones containing β-crystallin sequences (Chapter 2). These studies revealed that in both species the βB3-crystallin gene is closely linked to βB2-crystallin sequences. The rat βB2-crystallin gene is a single copy gene, whereas the human genome contains two copies of the βB2-crystallin gene. The second human βB2-crystallin gene, the βB2-2 gene, could not be linked to the βB3/βB2-1-crystallin two gene cluster. Sequence analysis of the βB2-2 gene showed that it has a three base pair deletion in the 4th exon and a mutation at the 5' splice site of the same exon. No transcripts of the βB2-2 gene were detected in a 22 months old human lens. The human βB2-2 gene is probably a pseudogene (Chapter 6).

Both the rat βB2- and the βB3-crystallin gene contain, as does the rat βB1-crystallin gene, a small 5' non-coding exon, of about 30 bp in length. A short conserved region was identified in the 5' flanking sequences of these genes (Chapter 3).

The differential distribution of the crystallin proteins within the lens already suggested a gene specific regulation of the expression of the crystallin genes. Differences in the temporal expression were found both between and within the crystallin gene families and were most marked within the family of β-crystallin genes (Chapter 4). Expression of all members of this gene family was detected during early embryogenesis but the βB1- and βB3-crystallin genes reached their maximal level of expression around birth while the concentration of the βB2-crystallin transcripts continued to increase till 4 months after birth. The β-crystallin genes show the same mode of regulation as the γ-crystallin genes in that the differential expression pattern is a consequence of differential shut-down rather than differential activation. The only exception is the βs gene, of which transcripts could be detected only after birth. The βs gene is the only crystallin gene which appears to be active exclusively in the postnatal lens.

A comparison of the absolute amounts of crystallin RNA with the crystallin protein pattern indicated that the βB2- as well as the αB-

crystallin transcripts are poorly translated in the rat eye lens. The sequences preceding the translation initiation codon of these two transcripts are the only crystallin transcripts which match perfectly with the consensus sequences as proposed by Kozak (1986; Cell 44, 283-292). Hence this consensus sequence may well not be optimal for crystallin mRNAs.

As a result of our studies of the structure and expression of the β-crystallin genes, the (partial) sequences of a number of β-crystallin coding regions from man as well as rat became available. Together with other available crystallin sequence data these were used to analyse the patterns of evolution within and between the β- and the related γ-crystallin gene families (Chapter 7 and 8). Sequence comparison of the β-crystallin sequences showed no evidence for the occurrence of non-reciprocal genetic exchange between the members of this family. The evolutionary distance between these genes is large and the ancestors of the 'basic' and 'acidic' β-crystallins probably already existed before the vertebrate radiation. The divergence of the present 'basic' crystallins was estimated to have occurred 450 myrs ago. The mammalian γ-crystallin gene family is younger and probably arose some 180-240 myrs ago. Gene conversion has been a frequent event during the evolution of this gene family.

The pattern of the evolution of the β- and the γ-crystallin proteins is very much alike. In both families the four homologous protein motifs are equally well conserved when relatively closely related species are compared (e.g. rat and man). Over a longer evolutionary distance a better conservation of the internal motifs II and IV is seen. Of all known crystallin sequences, the βA3/A1 sequence is best conserved. This sequence is also equally well conserved in the rodent and in the primate lineage, while both the αA- and the γ-crystallins diverge faster in the primate lineage than in the rodent lineage.

Our finding that the γ-crystallin genes are better conserved within the rodent than within the primate lineage suggested that the γ-crystallins play a less important role in the function of the human lens, possibly due to the higher water content of the human lens. The difference in the function of the γ-crystallin genes between rat and man is further shown by the fact that the rat lens contains six abundant γ-crystallins while the orthologous six genes in man only encode

two abundant γ-crystallin (γC and γD) proteins. It was demonstrated by others that two of the human γ-crystallin genes are pseudogenes. To determine the transcriptional activity of the four remaining genes, the composition of the human fetal lens γ-crystallin RNA pool was analysed by S1-nuclease mapping and primer extension experiments (Chapter 5). Besides the (expected) transcripts of the γB- and γC-crystallin gene, transcripts of the pseudo γE-crystallin gene were also present. No transcript of the γA-crystallin gene could be found and the results obtained for the γB-crystallin gene could not be unambigeously interpreted. The low amounts of γA- and γB-crystallins within the human lens is probably due to the poor transcription of the corresponding genes. As in the rat, the complete spectrum of human γ-crystallin gene transcripts is present during early development, although the onset of the γ-crystallin gene expression relative to that of the αA- and β-crystallin genes is somewhat delayed. Differential shut-off is found during later development, with the γD-crystallin gene being the last gene active.

Van de drie crystalline eiwitfamilies die voorkomen in de ooglens van zoogdieren is de β-crystalline familie het meest heterogeen van samenstelling. Om meer inzicht te krijgen in de structuur, de nucleotidevolgorde, de evolutionaire verwantschap en de ontwikkelings-afhankelijke expressie van de genen die coderen voor deze eiwitfamilie werden in eerste instantie genomische fragmenten welke humane of ratte β-crystalline sequenties bevatten gekloneerd en in kaart gebracht (hoofdstuk 2). Uit dit onderzoek bleek dat de βB3- en βB2-crystalline genen nauw gekoppeld zijn, zowel binnen het humane als binnen het ratte genoom. In de rat blijkt het βB2-crystalline gen een uniek gen te zijn. In het humane genoom, echter, komen als gevolg van een recente duplicatie twee copiën voor van dit gen. De ligging van de tweede copie, het humane βB2-2-crystalline gen, ten opzichte van het βB3- en het βB2-1-crystalline gen kon aan de hand van de geisoleerde klonen niet worden bepaald. Wel is bekend dat beide βB2- genen op hetzelfde chromosoom zijn gelegen. Er zijn sterke aanwijzingen dat het humane βB2-2-crystalline gen een pseudo-gen is en derhalve geen aanleiding kan geven tot de productie van een functioneel eiwit. Het bevat namelijk naast een deletie van drie basenparen in het vierde exon ook nog een mutatie in de 5' splice site van dit exon. Tevens konden met behulp van een specifieke primer in primer-verlengingsexperimenten geen transcripten van dit gen worden aangetoond in humaan lens RNA (hoofdstuk 6).

Van het ratte βB2- en het βB3-crystalline gen werden de 5' uiteinden en de 5' flankerende sequenties bepaald (hoofdstuk 3). Beide genen bleken evenals het ratte βB1-crystalline gen een niet-coderend 5' exon te bevatten. Deze 5' niet-coderende exonen zijn bijna identiek qua grootte maar verschillend qua basenvolgorde. In de 5' flankerende gebieden van deze genen werden wel TATA-boxen aangetroffen maar geen sequenties die enigzins aan de CCAAT-box consensus voldeden. Wel bleek in alle drie genen een geconserveerd gebied aanwezig te zijn. Dit gebied, met de consensus sequentie $GGGC(N)_nPuTTT(N)_nTG$, komt in het 5' flankerende gebied van het βB3-crystalline gen een keer en in het 5' flankerende gebied van het βB1- of het βB2-crystalline gen twee keer voor. Hetzelfde gebied blijkt ook voor te komen in het 5' flankerende gebied van het βA3/A1-crystalline gen en wordt tevens in de 5' flankerende gebieden van de γ-crystalline genen aangetroffen. Mogelijk speelt

deze geconserveerde sequentie een rol bij de lens-specifieke expressie van deze genen.

De differentiële verdeling en synthese van de crystallines in de ratte ooglens suggereren een ontwikkelingsspecifieke expressie van de individuele crystalline genen. Om na te gaan of de expressie van deze genen onder transcriptionele en/of translationele controle staat werden de niveaus van de αB-, de βB2-, de βB3-, de βA3/A1- en de βs-crystalline transcripten in de ratte ooglens tijdens de prenatale en postnatale ontwikkeling door middel van Northern- en dotblot analyse gemeten. Uit de resultaten van dit onderzoek bleek dat er niet alleen verschillen bestaan in de ontwikkelingsafhankelijke expressie tussen de vertegenwoordigers van de verschillende genfamilies maar ook tussen de verschillende leden van een genfamilie. Zo konden transcripten van het αA-crystalline gen al in een vroeg stadium tijdens de prenatale ontwikkelling worden aangetoond terwijl de transcripten van het αB-crystalline gen pas laat tijdens de embryogenese detecteerbaar zijn. Binnen de β-crystalline genfamilie blijken alle genen gelijktijdig en vroeg in de ontwikkeling te worden aangeschakeld maar in verschillende stadia tijdens de ontwikkeling te worden afgeschakeld (hoofdstuk 4). Bijvoorbeeld, van de twee gekoppelde β-crystalline genen, het βB2- en het βB3-crystalline gen, bereikt het βB3-crystalline gen zijn maximale expressie reeds vroeg tijdens de ontwikkeling (geboorte), maximale expressie van het βB2-crystalline gen wordt daarentegen pas laat (4 maanden) in de ontwikkeling gevonden. Een van de crystalline genen, het βs-crystalline gen, is uitzonderlijk: van dit gen worden de transcripten pas na de geboorte aangetroffen, terwijl de andere leden van de β-/γ-crystalline genfamilie op een veel vroeger stadium in de ontwikkeling worden aangeschakeld.

Tijdens de postnatale ontwikkeling komen er in de rattelens ongeveer 3 tot 5 maal meer αB- dan αA-crystalline transcripten voor. Deze gegevens staan haaks op de in de literatuur beschikbare eiwitgegevens waaruit blijkt dat in de rattelens het αA-crystalline in een hogere concentratie voorkomt dan het αB-crystalline. De meest voor de hand liggende verklaring voor deze tegenstelling is een inefficiënte translatie van de αB-crystalline transcripten in de rattelens. Ook binnen de β-crystalline genfamilie komt de verhouding van de hoeveelheid RNA niet overeen met de verhouding van de hoeveelheid eiwit. Op basis hiervan zou een inefficiënte translatie ook kunnen worden verondersteld voor de

βB2-crystalline transcripten. In dit verband is het opmerkelijk dat de αB- en de βB2-crystalline mRNA's de enige twee crystalline transcripten zijn waarvan de nucleotidevolgorde die vooraf gaat aan het translatie initiatiecodon, en waarvan verondersteld wordt dat deze mede de efficientie van translatie bepaalt, volledig identiek is aan de "consensus" sequentie zoals die door Kozak (1986; Cell 44, 283-292) hiervoor is voorgesteld. Beide transcripten bezitten de "consensus" C op positie -4 waar de andere crystalline transcripten een purine bevatten. De aanwezigheid van een purine op positie -4 is misschien een vereiste voor een optimale translatie van crystalline transcripten.

De uit onze studies verkregen β-crystalline gensequenties werden gebruikt tesamen met andere beschikbare crystalline sequenties om de evolutie van β- en γ-crystalline genen en genfamilies te bestuderen (hoofdstuk 7 en 8). Zowel de β- als de γ-crystallines zijn opgebouwd uit vier homologe motiefstructuren. Twee motieven (I en III) zijn gelegen aan de buitenzijde, de twee andere (II en IV) aan de binnenzijde van het eiwit. Het is dan ook aannemelijk dat de β- en γ-crystalline genen ontstaan zijn uit opeenvolgende duplicaties van dezelfde 'een motief' eenheid en derhalve behoren tot dezelfde supergenfamilie. Binnen de humane en ratte β-crystalline genen evenals binnen de γ-crystalline genfamilie zijn de aan de buitenkant gelegen motieven over het algemeen het minst geconserveerd en bepalen vermoedelijk het specifieke karakter van het eiwit. Ook wanneer orthologe β-crystalline genen van ver verwante dieren (bv. mens en kip) met elkaar worden vergeleken blijken de interne motiefen (II en IV) beter geconserveerd te zijn. Worden echter orthologe β- of γ-sequenties van verschillende zoogdieren met elkaar vergeleken dan zijn er geen significante verschillen waarneembaar in het percentage niet-synonieme substituties tussen de vier motief-coderende regios.

De evolutionaire afstand tussen de β-crystalline genen is vrij groot. De voorlopers van de 'basische' en 'zure' β-crystalline genen bestonden waarschijnlijk al voor de radiatie van de vertebraten, terwijl geschat wordt dat de huidige 'basische' β-crystalline genen zo'n 450 miljoen jaar geleden gedivergeerd zijn. De γ-crystalline genfamilie is jonger en ontstond zo'n 180-240 miljoen jaar geleden.

De geringe mate van synonieme substituties in het tweede exon (coderend voor motief I en II) van de ratte γ-crystalline genen sugge-

reert een frequent optreden van genconversie. Ondanks de overeenkomstige evolutionaire herkomst met de γ-crystalline genen hebben we uit onze gegevens geen aanwijzingen gevonden die erop wijzen dat ook in de β-crystalline genfamilie de genen of gedeelten ervan co-evolueren middels genconversie.

Van alle crystallines blijkt het βA3/A1-crystalline het best geconserveerd te zijn. De ratios synonieme tot niet-synonieme substituties zijn voor het βA3/A1-crystalline gen nagenoeg gelijk in de humane en de knaagdierlijn hetgeen erop duidt dat de selektiedruk gelijk is in deze twee lijnen. De γ-crystalline genen zijn echter, net als de αA-crystalline sequenties beter geconserveerd in de knaagdierlijn dan in de humane lijn. Dit suggereert dat de γ-crystallines een minder belangrijke rol spelen in de humane lens, vermoedelijk als gevolg van het hogere watergehalte van de humane lens.

Het verschil in functie van de ratte en humane γ-crystalline genen wordt ondersteund door het feit dat in de rattelens alle zes genproducten voorkomen terwijl in de humane lens slechts twee γ-crystallines (γC en γD) in redelijke mate aanwezig zijn. Het was al bekend dat twee van de humane γ-crystalline genen (γE en γF) pseudo-genen zijn. Om een beter inzicht te krijgen in de expressie van de humane γ-crystalline genfamilie werd de samenstelling van de γ-crystalline mRNA pool in de humane lens door middel van Northern-blot analyse, S1-nuclease beschermings- en primer-verlengingsexperimenten nader geanalyseerd (hoofdstuk 5). Transcripten van de γ-crystalline genen konden voor het eerst worden waargenomen in 7 weken oude foetale lenzen. De transcripten waren afkomstig van zowel het γC-, het γD- (hiervan zijn ook de translatie producten aangetoond in de lens) als het pseudo γE-crystalline gen. Van het γA-crystalline gen konden geen transcripten worden gedetecteerd. De resultaten verkregen voor het γB gen konden niet eenduidig worden geinterpreteerd maar wijzen erop dat transcripten van dit gen niet in hoge concentraties aanwezig zijn. De primer-verlengingsexperimenten lieten tevens zien dat de 'aktieve' vertegenwoordigers van de γ- crystalline genfamilie differentiëel tot expressie komen. Zoals dit reeds eerder voor de ratte γ-crystalline genen is gevonden, worden ook alle humane genen op hetzelfde tijdstip van ontwikkeling aangeschakeld. Het tijdstip van afschakeling is echter voor ieder gen verschillend. De transcripten van het γD gen blijven het langst aantoonbaar, dit in tegenstelling tot de rat waar het γB-crystalline gen het langst aktief is.

De γ-crystalline eiwitfractie in de humane lens is voornamelijk samengesteld uit de producten van het γC- en het γD-crystalline gen, hetgeen in overeenstemming is met onze resultaten. De lage concentratie van de producten van de γA- en γB-crystalline genen wordt waarschijnlijk veroorzaakt door een inefficiënte transcriptie van deze genen.

CURRICULUM VITAE

Henk Aarts werd geboren op 15 augustus 1957 te Budel als dertiende kind van Nel van Meijl en Ties Aarts.

Na het behalen van het MAVO (1973), HAVO (1975) en Atheneum-B (1977) examen begon hij in 1977 aan zijn studie Biologie te Nijmegen. In december 1983 werd deze studie afgesloten met het behalen van het doctoraal examen met als hoofdvak Genetica (Dr. T. Hulsebos; Prof. Dr. W. Hennig) en bijvakken Botanie (Dr. J. Schrauwen; Prof. Dr. H.F. Linskens) en Biochemie (Dr. W. Quint; Dr. A. Berns; Prof. Dr. H. Bloemendal). Vanaf 1 januari 1984 tot en met 31 december 1987 was hij in dienst van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (NWO) en werkzaam op de afdeling Moleculaire Biologie aan de Katholieke Universiteit van Nijmegen. Onder leiding van Dr. N.H. Lubsen en Prof. Dr. J.G.G. Schoenmakers werd het in dit proefschrift beschreven onderzoek verricht.

Vanaf 1 juli 1988 is hij werkzaam als wetenschappelijk medewerker op het Centrum voor Onderzoek en Voorlichting voor de Pluimveehouderij "Het Spelderholt".

Op 25 mei 1984 trouwde Henk met Anita Hendrikx en ze hebben nu samen twee kinderen, Marloes en Iris.

Stellingen

1. De aanwezigheid van een vroegtijdig stopcodon in het humane gammaE-crystalline gen en het niet waarnemen van aktiviteit van diens promoter in primaire kippenlens epitheel cellen, rechtvaardigt niet de conclusie van Meakin et al. (1987) dat dit gen niet aktief is in de humane lens.

   Meakin, S.O., Du, R.P., Tsui, L.-C. and Breitman, M.L. (1987). Mol. Cell. Biol. 7, 2671-2679.

   Dit proefschrift.

2. De eerste twee exonen van de β-crystalline genen zijn onstaan uit hetzelfde (oer-)exon.

   Dit proefschrift.

3. In het onderzoek zoals dit is uitgevoerd door Murer-Orlando et al. (1987) en Treton et al. (1988) worden de vier muizen gamma-crystalline cDNA (sub-)probes ten onrechte als gen-specifieke probes gebruikt.

   Murer-Orlando, M., Paterson, R.C., Lok, S., Tsui, L.-C. and Breitman, M.L. (1987). Dev. Biol. 119, 260-267.

   Treton, J.A., Jacquemin, E. and Courtois, Y. (1988). Exp. Eye Res. 46, 405-413.

4. Thomson en Augusteyn (1985) gebruiken ten onrechte de bovine nomenclatuur voor de humane β-crystallines.

   Thomson, J.A. and Augusteyn, R.C. (1985). Exp. Eye Res. 40, 393-410.

5. De conclusie van Wolfe et al. (1989) dat de mate van synonieme substituties gelijk is voor gekoppelde genen is niet in overeenstemming met alle getoonde data.

   Wolfe, K.H., Sharp, P.M., Li, W.-H. (1989). Nature 337, 283-285.

6. Jost et al. (1986) hadden in hun artikel waarin het bewijs wordt geleverd dat een monoclonaal antilichaam tegen LPS bescherming kan bieden tegen infectie met Leptospira interrogans, ook hun eerdere publicatie, waaruit het tegendeel blijkt, moeten aanhalen.

Jost, B.H., Adler, B. and Faine, S. (1986). J Med Microbiol. 22, 269-275.

Adler, B. and Faine, S. (1982). J. Med Microbiol. 15, 259-262.

7. Het doden van een transgene muis, voordat gezorgd is voor nakomelingen, is vergelijkbaar met het doden van de kip met de gouden eieren.

Kondoh, H., Katoh, K., Takahashi, Y., Fujisawa, H., Yokoyama, M., Kimura, S., Katsuki, M., Saito, M., Nomuro, T., Hiramoto, Y. Okada, T.S. (1987). Dev. Biol. 120, 177-185.

8. EV (Endogene-Virale) genen bij vertebraten kunnen het beste geïdentificeerd worden aan de hand van de flankerende gastheer-sequenties.

9. Een "ingebouwde" geregelde-drieweg katalysator voor landbouwhuisdieren zou een oplossing kunnen zijn voor de overbemesting.

10. Ook een recessief gen kan in het onderzoek dominant zijn.

11. Het schrijven van een proefschrift zou aanmerkelijk versneld worden indien de mogelijkheid van synonieme substituties in de taal even groot zou zijn als die in de erfelijke informatie.

Nijmegen, 26 mei 1989
Henk Aarts