Complete structure of the αB-crystallin gene: Conservation of the exon–intron distribution in the two nonlinked α-crystallin genes

(crystallins/3' conservation/chromosome mapping)

YVONNE QUAX-JEUKEN*, WIM QUAX*, GEERT VAN RENS*, P. MEERA KHAN†, AND HANS BLOEMENDAL*‡

*Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands; and †Department of Human Genetics, Sylvius Laboratories, University of Leiden, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands

Communicated by L. N. M. Duyvens, April 25, 1985

ABSTRACT We isolated bovine complementary DNA clones for the αA- and αB-crystallin subunits. The αB cDNA clone was used to isolate an αB-crystallin gene. This gene, derived from hamster, occurs as a single copy in the genome and is 3.2 kilobases long. The coding sequences are spread on three exons with a total length of 709 nucleotides. The exon–intron distribution of the hamster αB-crystallin gene is similar to that of the αA-crystallin gene except for the 69 nucleotides that specify the 23 "insert" residues of the αAαin chain by means of differential splicing. The 3' noncoding region of the αB mRNA (140 bases), which is short compared with the αA mRNA (520 bases), shows a remarkable homology between calf and hamster. Both α-crystallin cDNA clones have been used to assign the chromosomal location of the corresponding human genes with the aid of somatic cell hybrids. It is shown that the single-copies αA- and αB-crystallin genes are located on different chromosomes.

Among the structural proteins of the mammalian lens, α-crystallin is a major component with a molecular weight of ~800,000, which is composed of acidic and basic polypeptides. Of these subunits, only two are primary gene products, named αA2 and αB2, varying in ratio and in rate of synthesis during differentiation of the lens cell. All other α-crystallin subunits arise from αA2 and αB2 by post-translational modification (1–3). The hamster αA-crystallin gene, whose structure has been elucidated recently (4), consists of four exons. Exon 2, which encodes the so-called insert peptide, characteristic for some rodents (5, 6), arises by differential splicing (4–7).

To obtain structural evidence for the different expression patterns of the two α-crystallin genes, we isolated αA- and αB-crystallin cDNA clones from a cDNA bank of bovine lens mRNAs (8). Previously, αA-crystallin cDNA clones had been isolated for a few other species (9–11, 9). Quite surprisingly, however, an αB-specific cDNA clone had never been found in spite of the fact that several research groups pursued studies on molecular cloning of the crystallin genes. We have now also isolated and characterized by sequence analysis the hamster αB-crystallin gene and performed comparative studies with the hamster αA-crystallin gene (4). Moreover, we investigated the chromosomal location of the α-crystallin genes to address the question of whether these genes are linked on the genome, like the δ- and γ-crystallin families; such linkage might have implications for the regulation of expression.

MATERIALS AND METHODS

Isolation and Identification of Bovine αA- and αB-Crystallin cDNA Clones. Isolation and cloning of bovine lens mRNAs, the screening of the recombinant clones in a hybridization selection assay, and identification of the translational products by one- and two-dimensional gel electrophoresis was done as described (8). This resulted in the isolation of the αA- and αB-crystallin clones pBLA-1 and pBLA-1. Construction and Screening of a Hamster Genomic Library. A partial gene library from a Syrian gold hamster was constructed. Sau3A restriction fragments ranging from 15 to 20 kilobases (kb) were cloned in Charon 28 as vector (12). Screening was done with the probes pBLA-1 and pBLA-1. Hybridization was carried out at 42°C for 16 hr in 50% formamide/5× NaCl/Cit/1× Denhardt's solution/20 mM sodium phosphate, pH 6.8/5 mM EDTA/single-stranded salmon sperm DNA (100 μg/ml) (1× NaCl/Cit = 0.15 M NaCl/0.015 M Na citrate; 1× Denhardt's solution = 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone). Washings were at 42°C for 3 hr with the hybridization solution at 55°C for 15 min in 2× NaCl/Cit/0.1% NaDODSO4. Positive plaques were purified, and characterization was done by restriction enzyme mapping using established methods.

DNA Sequence Analysis. Two EcoRI fragments, derived from clone HA4A2, of 4.6 kb and 4.2 kb containing the part that hybridizes to our cDNA probe, were subcloned into a suitable pBR322 vector. The inserts of these subclones were digested with the enzymes Sau3A, Hae III, and Hpa II, and the resulting fragments were ligated into M13 mp8, mp9, mp10, or mp11 (13). The chain-termination method (14) was used to sequence the M13 inserts. Analysis was on 40-cm 6% sequencing gels. The gel readings were recorded, edited, and compared by the programs of Staden (15). Search for homology and optimal alignment was done by the program DIAGON (16).

Cell Hybrids and Genetic Analysis. The methods for production, isolation, and propagation of the interspecific somatic cell hybrids have been reported (17, 18). Previous reports described the PG/Me hybrids (19–21), the a3G hybrid (22), the a3Y hybrids (23), and the E36 series of hybrids (24) included in the present study. The hybrids were investigated for identifying the individual human chromosomes with G-banding, Giemsa-11, and/or Q-banding techniques (25). A list of 37 chromosome-specific enzyme markers used to screen to hybrids, including the references of the methods used, was reported elsewhere (24).

DNA Preparation and Southern Hybridization Analysis. Total DNA of the cell lines was isolated as described (26). Each DNA sample (10 μg) was digested with appropriate restriction enzymes, electrophoresed on 0.8% agarose gels, blotted onto nitrocellulose, and hybridized under the condi-

Abbreviations: kb, kilobase(s); bp, base pair(s).
1To whom reprint requests should be addressed.

5819
Identification of αA- and αB-Crystallin cDNA Clones. A bovine cDNA library constructed in pBR322 was screened for αA- and αB-crystallin cDNAs by a hybridization selection assay (8). The polypeptides synthesized in vitro were analyzed by one-dimensional (data not shown) and two-dimensional NaDodSO4/polyacrylamide gel electrophoresis and fluorography. Fig. 1 shows the identification of an αA- and αB-crystallin cDNA clone by two-dimensional gel coelectrophoresis with the soluble protein fraction of bovine lens. Fig. 1A presents one of these gels stained with Coomassie blue showing the migration of the bovine crystallins used as markers. Fig. 1C and D show the products synthesized in vitro from the mRNAs selected by the clones pBLαA-1 and pBLαB-1. The inserts of these α-crystallin clones were 1000 base pairs (bp) for αA and 700 bp for αB. Of the clones tested, 15% contained α-crystallin specific cDNA (αA and αB clones were found at a ratio of 4:1).

Isolation of the Hamster αB-Crystallin Gene. A hamster genomic library constructed in the λ phage Charon 28 (27) was screened for the αB-crystallin gene by using the nick-translated cDNA clone pBLαB-1 as hybridization probe. After screening 5 × 10^7 recombinants, one positive clone was identified. For restriction enzyme and blotting analysis, two Pst I fragments (150 and 500 bp long) were isolated from pBLαB-1 and used as probes. It was found that the coding sequences of the αB-crystallin gene were located within two EcoRI fragments and that the gene is ~3.2 kb long flanked by 5.8 kb 5′ upstream and 4.6 kb 3′ downstream sequences (Fig. 2). To allow a more detailed study of the gene, we made subclones of the two EcoRI fragments (4.6 and 4.2 kb) in pBR322, which were used for subsequent sequence analysis.

Southern blot analysis of hamster genomic DNA with the insert of pBLαB-1 as probe revealed the same hybridizing bands as the blotting analysis of the two EcoRI subclones (data not shown). These findings clearly indicate that all hamster genomic αB-crystallin sequences are located within these two EcoRI fragments and that αB-crystallin is encoded by a single-copy gene.

Nucleotide Sequence of the αB-Crystallin Gene. To define the exon/intron organization of the gene and to examine the structural basis of αB-crystallin expression, we determined the sequence of the coding parts and flanking sequences (Fig. 3). A comparison of the gene sequences with the reported bovine and human αB-crystallin amino acid sequences (28, 29) allowed us to locate the different exons. As shown in Figs. 2 and 3, the hamster αB gene consists of 3 exons and 2 introns. The first exon contains the coding information for amino acids 1–67 as well as 41 bp of the 5′ noncoding region. The second exon has a length of 123 bp and encodes residues 68–108 of the αB chain. The third and last exon is 340 bp long and contains the coding sequence for amino acids 109–175 and the information for the 140 bp of the 3′ noncoding region. The two introns of the αB-crystallin gene are 0.9 kb and 1.6 kb long. The translated exon sequences are almost identical with the known amino acid sequence of the αB2 subunit of calf and man (28, 29); the few substitutions are indicated in Fig. 3. Residues 39–41 apparently have a high mutation rate because they differ in all species examined.

In accordance with the general GT–AG border nucleotide rule, both introns begin with the dinucleotide GT and end with AG (30). More extensive consensus sequences have been proposed for the splice sites of eukaryotic genes that are transcribed by RNA polymerase II (31). The donor 5′-G/T-A-G/G-T-G-T-3′ and especially the acceptor 5′-C/T-H-N/C-T-A-G/G-3′ consensus sequence are reasonably well matched by the exon–intron junctions of the αB-crystallin gene. These exon–intron borders were compared with the corresponding regions of the hamster αA-crystallin gene (4). Since the products of the αA- and αB-crystallin genes αA2

Fig. 1. Identification of clones containing α-crystallin cDNA sequences by hybridization selection and two-dimensional NaDodSO4/polyacrylamide gel electrophoresis. (A) Coomassie blue-stained gel showing the crystallins from calf lens cortex. (B) Schematic representation of α-crystallin polypeptides of bovine lens. αA1 and αB1 are post-translationally modified products of αA2 and αB2. (C and D) Fluorographs of gels showing the in vitro translation products of mRNAs hybridized to the specific plasmids pBLαA-1 and pBLαB-1, respectively. Identification was accomplished by matching the fluorograph with the cold pattern of the α-crystallins run on the same gel. IEF, isoelectric focusing.

Fig. 2. Physical map of the αB-crystallin gene. (a) The sites for some restriction enzymes in the HA-αB insert are shown. B, BamH; E, EcoRI; H, HindIII; HII, HindIII; HIII, HindIII; HII, HincII; X, Xho I; X, Xba I; S, Sac I. (b) Shaded bars represent the position and size of αB exons, which are numbered with roman numerals. Note: There is an additional Sac I site 0.2 kb upstream from the Xba I site.
and αB2 have 56% sequence homology (28), they are supposed to be derived from a common ancestral gene by a duplication event, which is thought to have taken place >500 million years ago. As can be seen in Fig. 4, all exon-intron junctions are exactly located at the same sites in both genes, thereby yielding further evidence for the evolutionary relationship of these two genes. 

Flanking Regions. The 5' and 3' flanking regions were also sequenced because these regions may be important in the regulation of gene expression. At positions 159–165 in Fig. 3, there is a sequence "TATATAA" that obeys the consensus promoter sequence for euakaryotic genes (Goldberg–Hogness box). The transcription initiation site or the cap site could tentatively be assigned to the adenine residue at position 189, which is not preceded by the consensus cytosine, but rather by guanine. As for the αA-crystallin gene of hamster (4) and mouse (7) and the chicken αB-crystallin gene (32), no consensus CCAAT-box could be detected. The 3' end of the hamster αB-crystallin gene was determined by comparing the nucleotide sequence of the 3' untranslated region of the gene with the sequence of the bovine cDNA clone pBLaB-1, which contains the complete 3' noncoding region. The polyadenylation signal AATAAA is located 23 nucleotides upstream from the poly(A) addition site. Comparison of the 3' noncoding regions of the bovine cDNA clone and the hamster αB-crystallin gene clone (Fig. 5) revealed that these regions are equal in length and have >90% sequence homology.

Mapping of the αA- and αB-Crystallin Genes on Human Chromosomes. The linkage relationship between the two α-crystallin genes in the genome was studied by investigating the chromosomal location of the αA- and αB-crystallin genes by using somatic cell hybrids (rodent–human). We used a panel of five mouse–human hybrids and a panel of nine

Fig. 3. Sequence of the hamster αB-crystallin gene. The sequences of all coding regions and flanking nucleotides are shown. Introns are marked by arrows and the amino acid sequence encoded by the exons is given in the single-letter code. Lengths of intron fragments that are not sequenced are indicated. TATA box points to the transcriptional control signal. Arrow at position 485 indicates the poly(A) addition site, which is found by comparison with the bovine αB-crystallin cDNA sequence (Fig. 5). Asterisks and dots mark amino acid substitutions between hamster αB-crystallin and the known amino acid sequences of calf (28) and human (29), respectively.

Fig. 4. Exon-intron junctions of the αB-crystallin gene compared with the corresponding junctions of the hamster αA-crystallin gene (4). Nucleotides around the splice sites are compared to the consensus donor (Left) and acceptor (Right) sequence (30). Upper and lower lines represent hamster αB- and hamster αA-crystallin sequences, respectively. Exon sequences are in capital letters; amino acids bordering the introns are given above and below the corresponding codons. The insert sequence of the αA<sup>100</sup> is not included. The "ag" and "gt" intron consensus sequences are always present, the other nucleotides are slightly different from the consensus.

Fig. 5. Comparison of 3' nontranslated regions of the hamster αB-crystallin gene and bovine αB cDNA clone pBLaB. The 3' noncoding sequences of hamster and calf are aligned. Common residues are indicated by an asterisk. From the stop codon up to the poly(A) addition site, the length of the 3' noncoding sequence is identical and homology is >90%.
Table 1. Distribution of human chromosomes in cell hybrids

<table>
<thead>
<tr>
<th>Lane no. (Fig. 6)</th>
<th>Clone</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>X</th>
<th>αA</th>
<th>αB</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>aY 14-2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>aY 13-1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>aG 5-2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>E36 33.7-2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>E36 10 CB 1-2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>E36 10 CB 22B2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>E36 78-3-13-2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discordance αA: 3 4 4 4 5 8 3 4 5 5 1 2 3 4 3 2 4 7 3 4 0 6 2 αB: 1 4 4 4 5 6 5 2 7 5 3 4 3 4 3 0 2 5 5 4 2 6 2

Cell lines H–P are Chinese hamster–human somatic hybrids. +, Presence of human chromosome containing αA and αB sequences. The lowest discordance is with chromosome 21 for αA and with chromosome 16 for αB. Cell line O is omitted from this table, because the precise chromosome content of this line could not be identified.

Chinese hamster–human hybrids. The cryopreserved cell lines, which were screened for the human chromosomes and chromosome-specific markers during previous studies (19, 22–24), were brought back to culture, and the cell pellets of each hybrid line were divided into two portions. One part was used to verify the human chromosome content by retesting for chromosome-specific enzyme markers (Table 1). The other part was used to isolate total DNA, which had been digested with an appropriate restriction enzyme and, after electrophoresis and blotting, was hybridized to the nick-translated insert of pBLaA-1 and pBLaB-1. For the αA-crystallin probe, it was found that in the case of hamster–human cell hybrids, the restriction enzyme BamHI gave a good resolution of the hamster and human bands. The 4.0-kb human BamHI fragment is found in lanes H, I, L, M, N, and P (Fig. 6 Upper). Comparison of these results with the chromosomal information of the hybrids (Table 1) indicated that the αA-crystallin gene is in chromosome 21. The use of a panel of five mouse–human cell lines also was in concert with the assignment of the αA-crystallin gene to chromosome 21 (data not shown). The αB probe gave a good resolution between the human and hamster bands with Sac I. The human 10-kb Sac I fragment is found in lanes H–P (Fig. 6 Lower). Comparison of these results with Table 1 and calculation of a discordance score indicates that the αB-crystallin gene is presumably in chromosome 16. Some doubt remains about the hybridization signal in the cell line in lane L. Only a very faint band with a different migration behavior could be detected after hybridizing with the αB-crystallin probe. Screening of another panel of cell lines (mouse–human) did not give conclusive results. From these results, however, it is clear that both α-crystallin genes segregate with different chromosomes, excluding the possibility of a close physical linkage of these genes on the same chromosome.

**DISCUSSION**

The highly conserved slowly evolving crystallins have demonstrated their usefulness for the study of molecular evolution (34). The α-crystallins have the lowest mutation rate among crystallins, only 3 amino acid replacements per 100 residues in 100 million years.

At the DNA level, the αB-crystallin is also highly conserved, surprisingly not only for the protein-encoding nucleotides. Alignment of the exon sequences of the hamster αB-crystallin gene with the sequence of the bovine αB-crystallin cDNA revealed a high degree of homology in the 3' noncoding region (Fig. 5)—only 3 single base deletions and 15 substitutions in a sequence of 140 bp. Assuming an equal rate of mutation after divergence and a neutral mutation rate for 3' noncoding regions, which was estimated to be 0.7% per million years (35), this degree of difference would imply that this divergence took place = 0.7 × 10^7 years ago. However, the last common ancestor of hamster and celf existed probably 6 × 10^7 years ago. Therefore, considerable selective constraint must have maintained this sequence homology between the 3' untranslated regions after species divergence. On the contrary, the unusually long 3' noncoding part of the αA-
crystallin mRNA (520 bases for the hamster) does not show this evolutionary sequence conservation (4), since there is already 20% difference in this region comparing the closely related species hamster and mouse. Homologous 3' untranslated regions have been reported for members of several multigene families encoding cytoskeletal proteins, actins (36), tubulins (37), and also for the intermediate filament protein vimentin (38). It has been suggested that these conserved 3' noncoding regions might be necessary for the secondary structure of the RNA or for protein–nucleic acid interactions.

Comparative studies of the hamster aA- and aB-crystallin genes revealed almost no homology in the noncoding parts. Only consensus sequences like the "TATA" box and the poly(A) addition signal are conserved in addition to several short stretches of homology. A consensus "CCAAAT" box could be detected neither in the aA-crystallin gene nor in the aB-crystallin gene.

Initial experiments revealed a potentially important DNA sequence for the expression of the murine aA-crystallin gene between −85 and −400 bp from the cap site (39). The sequence of the aB-crystallin gene is determined up to −140 bp from the TATA box. Therefore, it might be possible that stretches of sequences that are involved in the regulation of transcription and in the tissue-specific expression of the a-crystallin genes are situated upstream from the sequence, as depicted in Fig. 3. Further experiments with a suitable expression system are needed to study the promoter regions of the a-crystallin genes in more detail. Another interesting feature of the a-crystallins is their homology with the small heat shock proteins (hsp) of Drosophila (40), Caenorhabditis, and soybean (41, 42). Alignment of the protein sequences of hamster aB-crystallin and Drosophila hsp26 showed that the homologous parts of these proteins are restricted to the second and third exon of the aB-crystallin gene (data not shown). The homology starts at residue 68 of the aB chain, the 5' end of the second exon. These confined stretches of homology were also found for the hamster aA-crystallin gene (4). It may be suggested that the greatest part of the a-crystallin genes has originated by duplication from ancestral heat shock genes. Combination of different exons might have taken place after this duplication. These suppositions would be in agreement with the hypothesis that new proteins may arise by combining exons encoding separate structural and functional domains, as proposed by Gilbert (43). The possible function of the heat shock protein-like domain in eye lens proteins is an intriguing object of future studies.

Another aim of our studies was to address the question of how the expression of the a-crystallin genes is regulated on the genome. Since aA- and aB-crystallin together form higher aggregates, it might be supposed that their expression is regulated coordinately. Linkage of the a-crystallin genes on the genome might be essential in that respect. Several crystallin families are linked on the genome. The δ-crystallin genes are only 4.2 kb apart on the chicken genome (44), and the y-crystallin genes are linked on the human chromosome 2 (33). Our results clearly show that the a-crystallin genes are not located on the same chromosome, although additional experiments are needed to verify the assignment of the aB-crystallin gene. The aA-crystallin gene is assigned to chromosome 21. In this respect, the observed trisomy of chromosome 21 in patients with Down syndrome, who often suffer from juvenile cataract, is striking and may point to a correlation between cataract and abnormal crystallin expression.

We thank Dr. W. de Jong for his comments and Dr. G. Berbers for the two-dimensional gel electrophoresis. The present investigations have been carried out in part under the auspices of The Netherlands Foundation for Chemical Research (SON) and with financial aid of The Netherlands Organization for the Advancement of Pure Research (ZWO).