

## REFLECTIONS ON THE INTERNAL PRIMARY, SECONDARY AND TERTIARY STRUCTURE HOMOMOLOGY OF THE EYE LENS PROTEINS $\alpha$ -, $\beta$ - AND $\gamma$ -CRYSTALLIN

Roland J. SIEZEN

*Department of Physical Biochemistry, John Curtin School of Medical Research, Australian National University, Canberra, ACT 2601, Australia*

Received 31 July 1981

### 1. Introduction

Crystallins are the water-soluble proteins of the eye lens. Their extremely high concentration in the lens fibre cells (200–600 mg/ml [1,2]) produces the uniformly high refractive index necessary for focussing visible light on the retina, yet leaves the lens completely transparent. The dense gel-like packing of these crystallins raises the question as to whether any long-range ordering of proteins exists in the lens and whether lens transparency depends on protein conformation, intermolecular interactions and long-range ordering [3–5]. The crystallins are uniquely stable, with little or no degradation occurring throughout the life of the individual. The oldest crystallins are found in the lens nucleus, the younger proteins in the lens cortex or periphery. In certain types of cataract (lens opacity), and particularly in human senile nuclear cataract, the irregularities in the refractive index which induce light scattering (and decrease transparency) are associated with an increased degradation, cross-linking (disulfide and non-disulfide) and aggregation of crystallins [3–12].

In mammals three classes of crystallins are distinguished:  $\alpha$ ,  $\beta$  and  $\gamma$ ; in bovine lens the relative amounts are about 40%, 40% and 20%, respectively [13]. Both  $\alpha$ - and  $\beta$ -crystallins are oligomeric proteins, whereas  $\gamma$ -crystallins are monomeric. In each class a number of related polypeptide chains are observed, which are either synthesized as different gene products or derived from each other by post-translational deamidation or

degradation [2,14–21]. The amino acid sequences have been determined of the predominant polypeptides in the bovine lens: the  $\alpha$ -crystallin A<sub>2</sub> and B<sub>2</sub> chains of  $M_r$  19 830 and  $M_r$  20 070 [22,23], the  $\beta$ -crystallin B<sub>p</sub> chain of  $M_r$  23 150 [24] and the  $\gamma$ -crystallin fraction II chain of  $M_r$  19 870 [25]\*. Raman spectroscopy of whole lenses suggests that the secondary structure of lens proteins is nearly exclusively anti-parallel  $\beta$ -sheet [5], and far-UV circular dichroism spectra of individual  $\alpha$ -,  $\beta$ - and  $\gamma$ -crystallins have also been interpreted in terms of high  $\beta$ -sheet and little or no  $\alpha$ -helix contribution [26–28].

Recently, the first three-dimensional structure of a lens protein, bovine  $\gamma$ -crystallin, has been determined by Blundell et al. [29], and a slightly revised primary structure of  $M_r$  21 000 best fits their electron density map. They suggest that the great stability of  $\gamma$ -crystallin lies in its high internal symmetry (4 three-dimensionally similar folding motifs) and in the arrangement of charged and non-polar groups on the molecular surface, while on the other hand, the position of several sulfhydryl groups would also allow for both intra- and intermolecular crosslinking, leading to aggregation [29]. Driessen et al. [24] detected a sequence homology between  $\beta$ - and  $\gamma$ -crystallin and suggested that their three-dimensional structures might be similar. Here, this hypothesis is considered in terms of  $\beta$ -crystallin internal homology, secondary structure, domains, charge stabilization, sulfhydryl groups, intermolecular interactions and aging. Although there is no apparent sequence homology between  $\alpha$ A- (or  $\alpha$ B-) and  $\beta$ - or  $\gamma$ -crystallin, similar principles of repetitive folding units, domains, hydrophobic cores and charge stabilization are proposed for  $\alpha$ -crystallin, and discussed in relation to molecular aging processes.

\* For the sake of simplicity and to avoid confusion with the nomenclature for folding units (I–IV) the  $\alpha$ A<sub>2</sub>-,  $\alpha$ B<sub>2</sub>-,  $\beta$ B<sub>p</sub>- and  $\gamma$ -crystallin fraction II will be referred to only as  $\alpha$ A-,  $\alpha$ B-,  $\beta$ - and  $\gamma$ -crystallin in this article

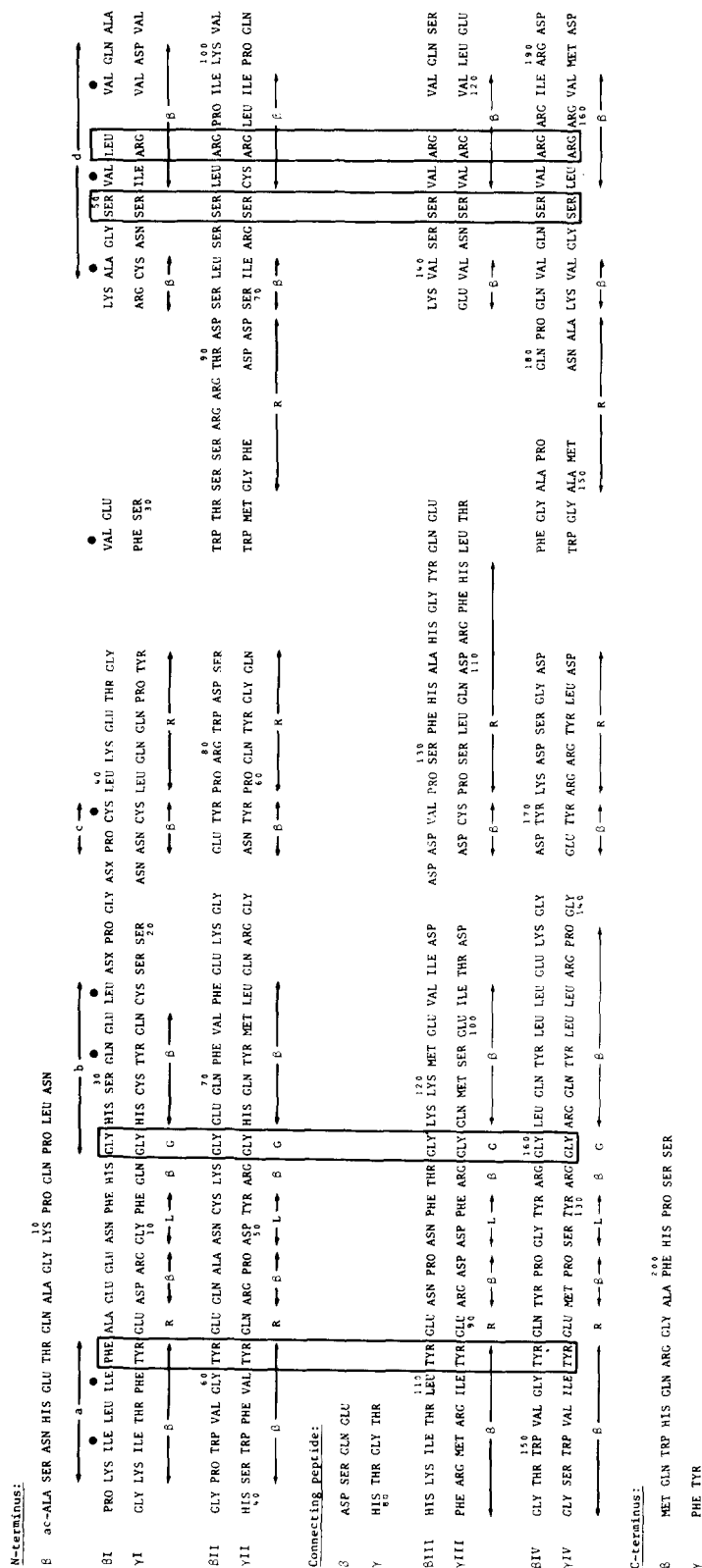


Fig.1. Optimal alignment of bovine  $\beta$ -crystallin and  $\gamma$ -crystallin sequences. The numbering of  $\gamma$ -crystallin is originally that of Croft [25], while inserts and replacements necessitated by the electron density map have been given subscripts A,B,C and D [29]. Both sequences are divided into four folding units (I–IV), each with four strands (a–d). The secondary structure refers to that of  $\gamma$ -crystallin, as indicated by the torsion angles in the three-dimensional structure:  $\beta$ ,  $\beta$ -sheet; R, righthanded helix generally distorted from  $\alpha$ -helix; L, lefthanded helix; G, glycine conformation ( $\phi \approx 110^\circ$ ,  $\psi \approx -110^\circ$ ) [29]. Topologically equivalent residues are in columns: the columns corresponding to the four invariant residues in all four folding units of  $\gamma$ -crystallin are boxed; the columns beneath the dots (•) correspond to the hydrophobic positions which contribute to the cores of the domains. The connecting peptide links the N- and C-terminal domains.

## 2. $\gamma$ -Crystallin

To facilitate the comparison of  $\alpha$ -,  $\beta$ - and  $\gamma$ -crystallin structures a brief summary of the three-dimensional structure of  $\gamma$ -crystallin is required (see [29] for details).

The protein molecule is ellipsoidal and consists of an N-terminal and a C-terminal domain, which are related by an approximate diad axis such that about 80  $\alpha$ -carbon atoms in each domain are topologically equivalent. The amino acid sequence is divided into four folding units  $\gamma$ I– $\gamma$ IV, or 'Greek key' motifs [30], of about 40 residues each (fig.1), which show only distant sequence homology: only 18–36% of the residues are identical (depending on which pair of folding units is compared) whereas another 30–45% are conservatively varied (table 1). However, all four units show very close three-dimensional homology: folding units  $\gamma$ I and  $\gamma$ III in different domains are most similar, as are  $\gamma$ II and  $\gamma$ IV (this pair has by far the highest sequence homology), whereas folding units within one domain ( $\gamma$ I– $\gamma$ II;  $\gamma$ III– $\gamma$ IV) resemble each other less. This type of structure is assumed to have evolved by two sequential gene duplications. Each domain comprises a sandwich of two righthanded, four-stranded anti-parallel  $\beta$ -sheets, each sheet being composed of three major strands (fig.1a,b,d) from one folding unit and one minor strand (c) from the other folding unit in the same domain. About 50% of the amino acids participate in forming  $\beta$ -sheets, while another 25% of the residues which connect the strands are in irregular righthanded (20%) or lefthanded helices (5%) as indicated by the torsion angles (fig.1).

Only four residues are found to be invariant at topologically equivalent positions in all four folding units (boxed in fig.1): three of these are  $\gamma$ Tyr 6,  $\gamma$ Ser 34 and  $\gamma$ Arg 36 which are all situated around conservatively varied  $\gamma$ Phe 11 ( $\gamma$ Tyr 51) which itself is in a  $\beta$ -turn, exposed, and at a lattice contact point forming a hydrophobic intermolecular contact. The fourth invariant is  $\gamma$ Gly 13 which has torsion angles disallowed for other residues. Other structurally important hydrophobic residues are in  $\gamma$ -chain positions 3,5,16,18,23,29,32,35 and 37 (and also the topologically equivalent positions in columns beneath the residues listed) which form the hydrophobic core between  $\beta$ -sheets in each folding unit, and the hydrophobic residues of motifs  $\gamma$ II and  $\gamma$ IV which form contacts between N- and C-terminal domains:  $\gamma$ Phe 43A,  $\gamma$ Met 55B,  $\gamma$ Val 125A,  $\gamma$ Leu 136 (subscripts A,B,C and D

were introduced by Blundell et al. [29] to denote revised or inserted residues). Invariance is apparently not required of these residues, but they must be predominantly hydrophobic.

The surface of  $\gamma$ -crystallin appears to be stabilized by charged and polar side chains organized in pairs or chains of ionic and/or polar interactions: e.g., Glu 7–Arg 31–Asp 8, Glu 122–Arg 85–Glu 100–Arg 87 and Glu 141–Arg 143–Asp 93–Arg 95–Asp 92–Arg 91. Intermolecular charge interactions are also observed in the crystal (e.g., Arg 144–Asp 103, Glu 90–Arg 91) and thus similar interactions may contribute to protein supramolecular organization in the lens. Interestingly, the charge stabilization is not symmetrical, i.e. most interactions cited for the C-terminal domain do not occur at equivalent positions in the N-terminal domain.

The six free sulfhydryl groups are also asymmetrically distributed. Only  $\gamma$ Cys 23 and  $\gamma$ Cys 105 are in equivalent positions in motifs  $\gamma$ I and  $\gamma$ III, and the remaining four cysteines are in the N-terminal domain, with three cysteines (18,23 and 74) clustered together such that they could form an intramolecular disulfide bridge, without affecting the overall conformation dramatically. However, four cysteines (15,18,23 and 105) on the surface of the molecule could also form intermolecular S–S bonds upon oxidation, and thus contribute to covalent crosslinking of lens proteins as observed in aging and cataract.

## 3. $\beta$ -Crystallin

The optimal alignment of the  $\beta$ -crystallin sequence of 204 amino acids with the revised  $\gamma$ -crystallin sequence of 174 amino acids [29] is shown in fig.1. Not only is the  $\beta$ -chain internally duplicated [24], it now becomes evident that most of the characteristics of the  $\gamma$ -crystallin three-dimensional structure are also manifested by  $\beta$ -crystallin. Not taking into account the N- and C-terminal extensions of 15 and 13 residues, the rest of the  $\beta$ -chain aligns perfectly with the four folding units  $\gamma$ I– $\gamma$ IV, albeit than an additional four residues ( $\beta$ Gly 36,  $\beta$ Arg 88,  $\beta$ Arg 89,  $\beta$ Asp 126) are inserted in segments between  $\beta$ -sheet strands where the four  $\gamma$ -crystallin folding units themselves already vary in length. When each motif is examined separately, the sequence homology between  $\beta$ - and  $\gamma$ -crystallin units is generally greater than between the  $\gamma$ -crystallin units, the exception being the  $\gamma$ II– $\gamma$ IV pair (table 1). Thus some 73–77% of the residues are

Table 1  
Relatedness of  $\beta$ - and  $\gamma$ -crystallin folding units

Unit	Total residues	Identical		Related <sup>b</sup>		Different		
		(%)	(%)	SBC	MBC	(%)	Gap <sup>a</sup>	Other (%)
$\gamma$ I $\gamma$ III	39 41	9 (21)	12	1	(31)	3 1	16	(48)
$\gamma$ II $\gamma$ IV	44 44	16 (36)	14	6	(45)	0 0	8	(18)
$\gamma$ I $\gamma$ II	39 44	8 (18)	9	4	(30)	5 0	18	(52)
$\gamma$ III $\gamma$ IV	41 44	10 (21)	13	3	(34)	6 3	12	(45)
$\beta$ I $\beta$ III	40 42	13 (30)	9	4	(30)	3 1	13	(40)
$\beta$ II $\beta$ IV	46 44	14 (30)	16	4	(43)	0 2	10	(26)
$\beta$ I $\beta$ II	40 46	4 (9)	15	5	(43)	7 1	15	(49)
$\beta$ III $\beta$ IV	42 44	8 (17)	11	5	(34)	5 3	15	(49)
$\gamma$ I $\beta$ I	39 40	9 (23)	15	5	(50)	1 0	10	(28)
$\gamma$ II $\beta$ II	44 46	13 (28)	17	4	(46)	2 0	10	(26)
$\gamma$ III $\beta$ III	41 42	14 (33)	15	3	(43)	1 0	9	(24)
$\gamma$ IV $\beta$ IV	44 44	22 (50)	10	2	(27)	0 0	10	(23)

<sup>a</sup> Each number represents a single residue deletion in the corresponding chain

<sup>b</sup> A relatedness matrix was used which is based on an evolutionary distance of 250 accepted point mutations (fig.84 of [31]). For the sake of simplicity, matrix element values  $\geq 0$  were taken to signify relatedness (SBC = single base change in the triplet DNA codon; MBC = multiple base change). The related residues fall into groups of chemically and/or conformationally similar amino acids that tend to replace each other more frequently than random chance would predict. This increased frequency of replacement arises principally by natural selection and only secondarily by constraints of the genetic code, and reflects the similarity of functions of residues in their interactions with each other in formation of secondary and tertiary structure in proteins

either identical or conservatively varied (and most of these are single base changes), while a full 50% is identical in the  $\beta$ IV- $\gamma$ IV pair. The internal sequence homology of the  $\beta$ -chain is much the same as  $\gamma$ -crystallin (table 1), with the pairs  $\beta$ I- $\beta$ III and  $\beta$ II- $\beta$ IV showing more homology (30% identical, 30-43% related) than pairs  $\beta$ I- $\beta$ II and  $\beta$ III- $\beta$ IV within a domain (9-17% identical, 34-43% related). In an evolutionary sense, this confirms that gene duplication in formation of domains preceded the divergence of  $\beta$ - and  $\gamma$ -crystallin.

The four invariant residues of  $\gamma$ I- $\gamma$ IV ( $\gamma$ Tyr 6,  $\gamma$ Gly 13,  $\gamma$ Ser 34,  $\gamma$ Arg 36; boxed in fig.1) are identical in  $\beta$ II- $\beta$ IV; in  $\beta$ I the tyrosine is very conservatively

replaced by  $\beta$ Phe 21, and  $\beta$ Leu 52 replaces arginine. The latter change may not be detrimental to local interactions since it is only the aliphatic side chain part of  $\gamma$ Arg 36 which interacts with the aromatic ring of  $\gamma$ Phe 11 [29] and this aliphatic part is still present in  $\beta$ Leu 52 to interact with  $\beta$ Phe 26. No extra invariant positions are found in  $\beta$ -crystallin relative to  $\gamma$ -crystallin. The residues contributing to the core of the domains are also largely hydrophobic in  $\beta$ -crystallin: 31 out of 36 residues in  $\beta$ -chain positions 18, 20, 31, 33, 39, 45, 48, 51 and 53 (and equivalent positions), which correspond to the earlier mentioned hydrophobic core positions in  $\gamma$ -crystallin. Likewise, the hydrophobic contacts between domains can be formed by  $\beta$ Val 59,  $\beta$ Val 72,  $\beta$ Val 151 and  $\beta$ Leu 164 which are in corresponding positions in the  $\beta$ II and  $\beta$ IV motifs.

The evidence presented above clearly points to a very close three-dimensional homology between  $\beta$ - and  $\gamma$ -crystallin. Therefore, it seems fairly safe to predict that most, if not all, of the secondary structure illustrated for  $\gamma$ -crystallin in fig.1 should also apply to  $\beta$ -crystallin, i.e. ~50% anti-parallel  $\beta$ -sheet and about 25% irregular connecting helices. It should be kept in mind that these percentages do not include the extra N- and C-terminal pieces of  $\beta$ -crystallin, which amount to 14% of the total residues, for which no secondary or tertiary structure can be predicted.

Although the hydrophobic positions are conserved in  $\beta$ - and  $\gamma$ -crystallin, the charged and sulfhydryl groups differ considerably. There are only two free sulfhydryl groups in the  $\beta$ -chain, and only  $\beta$ Cys 39 corresponds to an equivalent position in  $\gamma$ -crystallin ( $\gamma$ Cys 23). The other,  $\beta$ Cys 66 (corresponding to  $\gamma$ Tyr 51), would be sufficiently close to  $\beta$ Cys 39 to form an intramolecular disulfide bridge (cf. fig.5a of [29]). Both sulfhydryl groups would be accessible to solvent and could therefore also form intermolecular S-S bonds in oxidizing conditions. This notion takes on extra importance in view of the fact that  $\beta$ Cys 66 would be at a lattice contact point (as is  $\gamma$ Tyr 51). Surprisingly, most of the surface ionic pair and chain interactions cited for  $\gamma$ -crystallin, both intra- and intermolecular, would not occur in  $\beta$ -crystallin because the corresponding residues are not charged: compare, for instance,  $\gamma$ 90-95 with  $\beta$ 112-117 (fig.1). Admittedly, the  $\beta$ -crystallin three-dimensional structure could be stabilized by ionic interactions at alternative surface positions, but one would need to construct a three-dimensional model to verify this. On the other



Fig.2. Internal homology of  $\alpha$ -crystallin A- and B-chains. Both sequences are divided into six folding units (I–VI), consisting of three homologous pairs of different lengths, which have been optimally aligned. Column beneath dots (•) correspond to hydrophobic positions, whereas columns under triangles (▼) correspond to charged/polar positions. The proposed subdivision of folding units into secondary structure strands a,b,c and d is based on the positions of gaps, inserts and proline residues (structure breaking). The extra inserted loop (residues 154–167) in unit VI of the normal  $\alpha$ A- and  $\alpha$ B-chain is enclosed in a dashed box; the position of a 22-residue insert in elongated rodent  $\alpha$ A-chain is indicated by (\*). The main bonds susceptible to degradation in vivo are shown as (=).

hand, differences in the balance of polar and non-polar surface residues may be responsible for the different properties of  $\beta$ -crystallins: for instance, their ability to form different-sized oligomers [32–34]. Also, large numbers of both lysine and glutamine residues occur in  $\beta$ -crystallins (relative to  $\alpha$ - and  $\gamma$ -crystallins), and if these are located on the molecule surface it could explain why  $\beta$ -crystallins are particularly susceptible to formation of  $\gamma$ -glutamyl- $\epsilon$ -lysine crosslinks in vitro by the action of lens transglutaminase [35]. These non-disulfide covalent crosslinks have been found in a polymer fraction of human cataract tissue [35].

#### 4. $\alpha$ -Crystallin

No sequence homology, however distant, has been detected between bovine  $\alpha$ -crystallin chains (A and B) and chains of the  $\beta/\gamma$ -crystallin class. Nevertheless, the principle features of the secondary and tertiary structure of the latter class (internal repeats,  $\beta$ -sheets, domains, conserved hydrophobic positions, ionic stabilization) may well be present in other lens proteins, since it is now clear that close three-dimensional structure similarity can be achieved with very little sequence homology.

As mentioned earlier, spectral studies indicate that  $\alpha$ -crystallin has secondary structure similar to  $\beta$ - and  $\gamma$ -crystallin [26–28]. In addition, a statistical analysis revealed an internal duplication in the first 60 residues of the A-chain [36]. A more extensive internal sequence homology is now proposed for the  $\alpha$ A-chain of 173 residues and the  $\alpha$ B-chain of 175 residues, using as the main criterion the conservation of hydrophobic positions (fig.2). The homologous units are divided into three pairs which differ in length:  $\alpha$ I– $\alpha$ II,  $\alpha$ III– $\alpha$ IV and  $\alpha$ V– $\alpha$ VI. For an optimal alignment about three gaps have been introduced in each pair, which is not uncommon in distant relationships [37], and an extra loop of 14 residues in unit  $\alpha$ VI (which will be discussed later), leaving only two connecting dipeptides (Ser 62–Glu 63 and Leu 85–Thr 86 in the  $\alpha$ A-chain). The gaps may occur in connecting loops, as in  $\beta$ - and  $\gamma$ -crystallin, or in some cases may even be due to sequencing errors.

The sequence homology within each pair (table 2) is at least as high as the  $\beta$ - or  $\gamma$ -crystallin internal homology (table 1), particularly for the  $\alpha$ A-chain pairs where 59–75% of the residues are either identical or related. The 57% overall homology of  $\alpha$ A- and  $\alpha$ B-

chain [23] can be subdivided by comparing each of the six  $\alpha$ A-chain units with the corresponding  $\alpha$ B-chain unit (table 2). Thus, the  $\alpha$ V units are found to be most homologous (28 out of 29 residues are identical or single base changes), and  $\alpha$ II and  $\alpha$ VI units least homologous because of several gaps.

Taking the internal sequence homology of  $\alpha$ -crystallin chains one step further, the three homologous pairs can be aligned such that most hydrophobic residues are in equivalent positions. It is proposed that  $\alpha$ A-chain residues 1,3,9,10,14,22,27 and 31, and equivalent residues in columns under dots in fig.2, are essential for hydrophobic contacts. In the  $\alpha$ A-chain, 34 out of 40 of these residues are hydrophobic, the others being either small uncharged residues or gaps. In this alignment no residues are invariant in all six units, but the tyrosines in units I, II and V coincide, and  $\alpha$ A-Pro 19 (or  $\alpha$ B-Pro 20) is only very conservatively varied to Gln in units II, IV or VI, suggesting a further functional conservation. Moreover, many

Table 2  
Relatedness of  $\alpha$ -crystallin A- and B-chain folding units

Unit	Total residues	Identical (%)	Related <sup>b</sup>			Different		
			SBC <sup>c</sup>	MBC	(%)	Gap <sup>a</sup>	Other	(%)
$\alpha$ A-I $\alpha$ A-II	31 30	8 (25)	12	4	(50)	1 2	5	(25)
$\alpha$ A-III $\alpha$ A-IV	21 20	5 (23)	7	1	(36)	1 2	6	(41)
$\alpha$ A-V $\alpha$ A-VI	29 24	7 (24)	10	3	(45)	0 5	4	(31)
$\alpha$ B-I $\alpha$ B-II	32 33	5 (14)	14	3	(49)	3 2	8	(37)
$\alpha$ B-III $\alpha$ B-IV	21 20	4 (18)	9	1	(45)	1 2	5	(36)
$\alpha$ B-V $\alpha$ B-VI	29 22	6 (21)	8	4	(41)	0 7	4	(38)
$\alpha$ A-I $\alpha$ B-I	31 32	23 (70)	5	0	(15)	2 1	2	(15)
$\alpha$ A-II $\alpha$ B-II	30 33	12 (35)	14	0	(41)	4 1	3	(24)
$\alpha$ A-III $\alpha$ B-III	21 21	11 (52)	7	0	(33)	0 0	3	(14)
$\alpha$ A-IV $\alpha$ B-IV	20 20	13 (65)	4	1	(25)	0 0	2	(10)
$\alpha$ A-V $\alpha$ B-V	29 29	18 (62)	10	0	(34)	0 0	1	(3)
$\alpha$ A-VI $\alpha$ B-VI	24 22	8 (31)	8	2	(38)	2 4	2	(31)

<sup>a,b</sup> As in table 1

charged and polar residues now coincide and at least four positions seem to qualify as preferentially charged/polar positions (columns under triangles). These could be required for surface charge stabilizing interactions, as could three segments consisting solely of charged and polar residues: Asp-Arg-Asp-Lys ( $\alpha$ A-III unit), Lys-His-Asn-Glu-Arg-Gln-Asp-Asp ( $\alpha$ A-IV unit) and Arg-Glu-Glu-Lys ( $\alpha$ A-VI unit extra loop), and corresponding segments in the  $\alpha$ B-chain.

There is a distinct possibility therefore that all six  $\alpha$ -chain units are folded in a similar fashion, though not necessarily the same folding as in the  $\beta/\gamma$ -chain units. On the other hand, in analogy with the  $\beta/\gamma$ -chain units, the positioning of gaps, inserts and proline residues suggests a subdivision of these units into four strands (fig.2a-d), parts of which could fold into anti-parallel  $\beta$ -sheet. Strand 'a' would be lacking in units  $\alpha$ V and  $\alpha$ VI, as would strand 'd' in units  $\alpha$ III and  $\alpha$ IV. The shorter 'c' strands of units  $\alpha$ III and  $\alpha$ IV could also be interpreted as connecting peptides (6 and 5 residues, respectively) rather than segments essential for  $\beta$ -sheet secondary structure. Insertions or deletions of various lengths up to 6 residues are encountered in the  $\beta/\gamma$ -crystallin folding units mainly between strands 'c' and 'd' (fig.1). In  $\alpha$ -crystallin a similar variability is evident between strands 'c' and 'd', and it is proposed that an extra loop of 14 residues is inserted here in the  $\alpha$ VI unit in both A- and B-chain (enclosed in dashed boxes in fig.2). This notion is prompted not only by the necessity for optimal alignment of units  $\alpha$ V and  $\alpha$ VI, but also because the segment  $\alpha$ A 156-163 has earlier been proposed as an external loop since it is very accessible to proteolytic enzymes *in vitro* [38-39]. In rodents an elongated  $\alpha$ A-chain has been identified which has an insert of 22 residues between positions Glu 63 and Val 64 of the normal  $\alpha$ A-chain [40]. This fits in nicely with the proposed model for  $\alpha$ A-chain subdivision, as the insert would be exactly in the connection between  $\alpha$ A-II and  $\alpha$ A-III units (\*) in fig.2).

Three domains of different size could exist in  $\alpha$ -crystallin chains if each homologous pair of folding units forms a separate domain. From an evolutionary and a symmetry viewpoint, it is more likely that two domains exist: each would consist of three folding units, one from each pair. The combination illustrated in fig.3 with units  $\alpha$ I,  $\alpha$ IV and  $\alpha$ V in domain 1 and  $\alpha$ II,  $\alpha$ III and  $\alpha$ VI in domain 2, has the lowest number (3) of crossovers between domains.

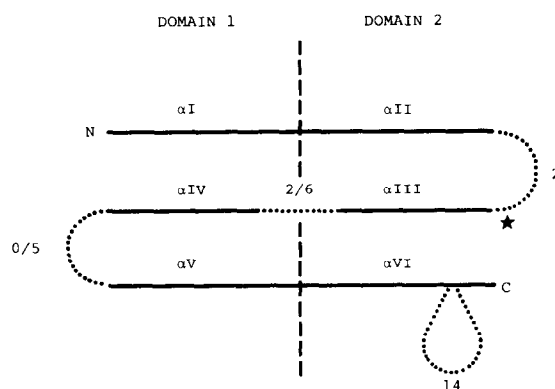


Fig.3. Hypothetical two-domain structure of  $\alpha$ -crystallin. Solid lines indicate the six folding units (I-VI). Dotted lines indicate additional connecting and inserted peptides and their lengths in number of residues. The position of a 22-residue insert in rodent  $\alpha$ A-chain is indicated by (\*).

With respect to age-related disulfide crosslinking of  $\alpha$ -crystallin, the solvent accessibility of the single cysteine  $\alpha$ A-Cys 131 (in the middle of strand 'd' in  $\alpha$ A-V) cannot be predicted from this model alone, but earlier modification studies indicate it is exposed [41] and thus presumably capable of forming intermolecular S-S bonds in an oxidizing milieu. The  $\alpha$ A-Cys 131 occurs in all 41 mammals investigated so far, but only 5 mammals, including man, have a second sulfhydryl residue:  $\alpha$ A-Cys 142 [19,42]. When the  $\alpha$ -crystallin three-dimensional structure is eventually elucidated, it will be interesting to ascertain whether these two SH-groups are sufficiently close together to form an intramolecular S-S bond, as in  $\beta$ -crystallin, and whether the position of  $\alpha$ A-Cys 142 bears any relationship to the higher sensitivity of human lenses to formation of cataract.

Specific age-related degradation occurs *in vivo* on the N-terminal side of  $\alpha$ A-Glu 102,  $\alpha$ A-Ala 152,  $\alpha$ A-Ser 169 and  $\alpha$ B-Ala 171 [18] (bonds indicated by = in fig.2). Only the latter two residues were shown to be accessible to proteolysis *in vitro* [39], and as they occupy positions topologically equivalent to  $\alpha$ A-Cys 131, their surface exposure agrees with that of the sulfhydryl group. The  $\alpha$ A-Glu 102 is in the middle of a chain of charged and polar residues,  $\alpha$ A 99-107, and is thus presumably also near the molecule surface. If the short 'c' strands of units  $\alpha$ A-III and  $\alpha$ A-IV are merely connecting peptides, rather than essential parts of these units, the cleavage at  $\alpha$ A-Glu 102 would represent the separation of the first four  $\alpha$ A units (I-IV) from the last two units (V-VI). The  $\alpha$ A-Ala

152 is in the 'c'-strand of unit  $\alpha VI$ ; a homology study of the  $\alpha A$ -chain of 41 mammals has identified this particular 'c'-strand to be the most variable region of the whole  $\alpha A$ -chain [42], and thus presumably located at the surface. However, as postulated earlier [39], the inaccessibility of  $\alpha A$ -Glu 102 and Ala 152 in native  $\alpha$ -crystallin to proteolysis in vitro probably means that these residues are initially protected, perhaps by the extra loop of 14 residues, and only become exposed as a result of age-related changes in secondary and/or tertiary structure.

Finally, it must be stressed that slightly different alignments of the six  $\alpha$ -crystallin folding units are possible which are marginally less acceptable on the basis of sequence homology than that postulated in fig.2, but most would agree with the main predictions of  $\alpha$ -crystallin three-dimensional structure presented here.

### Acknowledgement

I thank Denis Shaw for invaluable advice and criticism in the preparation of this manuscript.

### References

- [1] Philipson, B. (1969) *Invest. Ophthalmol.* 8, 258–270.
- [2] Van Kleef, F. S. M. (1975) Thesis, University of Nijmegen.
- [3] Trokel, S. (1962) *Invest. Ophthalmol.* 1, 493–501.
- [4] Benedek, G. B. (1971) *Appl. Optics* 10, 459–473.
- [5] Schachar, R. A. and Solin, S. A. (1975) *Invest. Ophthalmol.* 14, 380–396.
- [6] Truscott, R. J. W. and Augusteyn, R. C. (1977) *Exp. Eye Res.* 25, 139–148.
- [7] Van Haard, P. M. M., Kramps, J. A., Hoenders, H. J. and Wollensak, J. (1978) in: *Interdisciplinary Topics in Gerontology* (Von Hahn, H. P. ed) vol. 13, pp. 212–224, Karger, Basel.
- [8] Anderson, E. I. and Spector, A. (1978) *Exp. Eye Res.* 26, 407–417.
- [9] Spector, A. and Roy, D. (1978) *Proc. Natl. Acad. Sci. USA* 75, 3244–3246.
- [10] Spector, A., Garner, M. H., Garner, W. H., Roy, D., Farnsworth, P. and Shyne, S. (1979) *Science* 204, 1323–1326.
- [11] Siezen, R. J., Bindels, J. G. and Hoenders, H. J. (1979) *Exp. Eye Res.* 28, 551–567.
- [12] Bettelheim, F. A. (1979) *Exp. Eye Res.* 28, 189–197.
- [13] Van Kamp, G. J. and Hoenders, H. J. (1973) *Exp. Eye Res.* 17, 417–426.
- [14] Croft, L. R. (1973) in: *The Human Lens in Relation to Cataract*, CIBA Foundation Symposium 19, Elsevier/Excerpta Medica, Amsterdam, New York.
- [15] Slingsby, C. and Croft, L. R. (1973) *Exp. Eye Res.* 17, 369–376.
- [16] Kabasawa, I., Kinoshita, J. H. and Barber, G. W. (1974) *Exp. Eye Res.* 18, 457–466.
- [17] Herbrink, P., Van Westreenen, H. and Bloemendal, H. (1975) *Exp. Eye Res.* 20, 541–548.
- [18] Van Kleef, F. S. M., De Jong, W. W. and Hoenders, H. J. (1975) *Nature* 258, 264–266.
- [19] Kramps, J. A., De Jong, W. W., Wollensak, J. and Hoenders, H. J. (1978) *Biochim. Biophys. Acta* 533, 487–495.
- [20] Zigler, J. S. (1978) *Exp. Eye Res.* 26, 537–546.
- [21] Mostafapour, M. K. and Reddy, V. N. (1980) *Invest. Ophthalmol.* 19, 1053–1058.
- [22] Van der Ouderaa, F. J., De Jong, W. W. and Bloemendal, H. (1973) *Eur. J. Biochem.* 39, 207–222.
- [23] Van der Ouderaa, F. J., De Jong, W. W., Hilderink, A. and Bloemendal, H. (1974) *Eur. J. Biochem.* 49, 157–168.
- [24] Driessen, H. P. C., Herbrink, P., Bloemendal, H. and De Jong, W. W. (1980) *Exp. Eye Res.* 31, 243–246.
- [25] Croft, L. (1972) *Biochem. J.* 128, 961–970.
- [26] Horwitz, J. (1976) *Exp. Eye Res.* 23, 471–481.
- [27] Horwitz, J., Kabasawa, I. and Kinoshita, J. H. (1977) *Exp. Eye Res.* 25, 199–208.
- [28] Chiou, S.-H., Azari, P., Himmel, M. E. and Squire, P. G. (1979) *Int. J. Pept. Prot. Res.* 13, 409–417.
- [29] Blundell, T., Lindley, P., Miller, L., Moss, D., Slingsby, C., Tickle, I., Turnell, B. and Wistow, G. (1981) *Nature* 289, 771–777.
- [30] Richardson, J. S. (1977) *Nature* 268, 495–500.
- [31] Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C. (1978) in: *Atlas of Protein Sequence and Structure* (Dayhoff, M. O. ed) vol. 5, suppl. 3, pp. 345–352, National Biomedical Foundation, Washington DC.
- [32] Asselbergs, F. A. M., Koopmans, M., Van Venrooij, W. J. and Bloemendal, H. (1979) *Exp. Eye Res.* 28, 223–228.
- [33] Zigler, J. S. jr, Horwitz, J. and Kinoshita, J. H. (1980) *Exp. Eye Res.* 31, 41–55.
- [34] Mostafapour, M. K. and Reddy, V. N. (1978) *Invest. Ophthalmol.* 17, 660–666.
- [35] Lorand, L., Hsu, L. K. H., Siefring, G. E. jr and Rafferty, N. S. (1981) *Proc. Natl. Acad. Sci. USA* 78, 1356–1360.
- [36] Barker, W. C., Ketcham, L. K. and Dayhoff, M. O. (1978) *J. Mol. Evol.* 10, 265–281.
- [37] Dayhoff, M. O. and Barker, W. C. (1972) in: *Atlas of Protein Sequence and Structure* (Dayhoff, M. O. ed) vol. 5, pp. 41–45, National Biomedical Foundation, Washington DC.
- [38] Siezen, R. J. and Hoenders, H. J. (1977) *FEBS Lett.* 80, 75–80.
- [39] Siezen, R. J. and Hoenders, H. J. (1979) *Eur. J. Biochem.* 96, 431–440.
- [40] Cohen, L. H., Westerhuis, L. W., De Jong, W. W. and Bloemendal, H. (1978) *Eur. J. Biochem.* 89, 259–266.
- [41] Siezen, R. J., Coenders, F. G. M. and Hoenders, H. J. (1978) *Biochim. Biophys. Acta* 537, 456–465.
- [42] De Jong, W. W., Zweers, A. and Goodman, M. (1980) *Prot. Biol. Fluids* 28, 161–164.