FOR THE RECORD

Amino acid sequence of bovine gamma E (IVa) lens crystallin

GREG W. KILBY,^{1,4} MARGARET M. SHEIL,¹ DENIS SHAW,² JOHN J. HARDING,³
AND ROGER J.W. TRUSCOTT¹

¹Australian Cataract Research Foundation, Department of Chemistry, The University of Wollongong, NSW 2522 Australia

²John Curtin School of Medical Research, The Australian National University, Canberra, ACT 0200 Australia

³Nuffield Laboratory of Ophthalmology, Oxford University, Oxford, OX2 6AW, UK

(RECEIVED October 3, 1996; ACCEPTED January 15, 1997)

Abstract: When electrospray ionization mass spectrometry (ES-MS) was used to analyze purified bovine gamma E (γ IVa)-crystallin, it yielded a relative molecular mass (M_r) of $20,955 \pm 5$. This mass is significantly different from that calculated from the published sequence (M_r 20,894) (White HE et al., 1989, *J Mol Biol* 207:217–235). Further, ES-MS analysis of the protein after it had been reduced and carboxymethylated indicated the presence of five cysteine residues, whereas the published sequence contains six (Kilby GW et al., 1995, *Eur Mass Spectrom* 1:203–208). The entire protein sequence of γ E crystallin has therefore been studied via a combination of ES-MS, ES-MS/MS, and Edman amino acid sequencing. The corrected sequence gives an M_r of 20,955.3, which matches that obtained by ES-MS analysis of the purified native protein. The corrected sequence is also in agreement with a recent cDNA sequence obtained for a bovine γ -crystallin by R. Hay (pers. comm.).

Keywords: amino acid sequence; eye; gamma crystallin; lens; mass spectrometry

The soluble proteins of the eye lens, known as crystallins, are comprised of three main classes: α , β , and γ . Crystallin–crystallin interactions are thought to be important for the maintenance of transparency of the lens, and modification of the crystallins is known to occur on ageing (Harding & Dilley, 1976; van den Oetelaar & Hoenders, 1989) and in the development of nuclear cataract (Truscott & Augusteyn, 1977a, 1977b). Due to their accessibility, and the high degree of sequence homology with the human crystallins, bovine lens proteins are often used in studies aimed at characterizing these age-related and cataract modifications. The γ crystallins are monomeric proteins of ~21 kDa that

have a high cysteine content. There are at least six γ crystallin polypeptides in the bovine lens, i.e., γ S (formerly β S), γ A (or γ IVb), γ B (or γ II), γ C (or γ IIIb), γ D (or γ IIIa), and γ E (or γ IVa), and these are products of six genes (reviewed in Harding, 1991). There have been a number of physical and biochemical studies on the bovine γ crystallins, with emphasis on the role of these proteins in crystal packing and intermolecular interactions (Wistow et al., 1983; Chirgadze et al., 1991; Tardieu et al., 1992). The structures of three γ crystallins, γ B (Blundell et al., 1981), γ C (Tardieu et al., 1992), and γ E (White et al., 1989), have been elucidated by X-ray crystallography. The reported sequence of γ E was determined from a combination of partial amino acid sequence information, X-ray data, and extrapolation of the remaining sequence from homologous rodent and human genes (White et al., 1989).

The utility of electrospray ionization mass spectrometry (ES-MS) (Fenn et al., 1990) for the determination of accurate molecular weights of large biomolecules is now well established. Because detection by ES-MS is independent of the nature of the modifying groups, this technique is ideal for the examination of posttranslational modifications to proteins, which may not be detected by traditional methods.

This paper reports the sequence of bovine γ E crystallin, using enzymatic digestion followed by HPLC, ES-MS, and ES-MS/MS, in combination with chemical cleavage and Edman amino acid sequencing of the peptides.

Results and discussion: Bovine γ -crystallins were isolated from a total lens extract by gel permeation chromatography (GPC). Individual γ -crystallin subunits were purified by cation exchange chromatography (Tardieu et al., 1992) and then analyzed by ES-MS.

The relative molecular mass of γ E, as determined by ES-MS ($20,955 \pm 5$), was significantly different from that calculated from the reported sequence (20,894) (White et al., 1989). This is not surprising, because sections of the protein sequence reported by White et al. were determined by homology with proteins from related species. It was assumed, however, that a reasonable portion of the sequence was indeed correct and this was used as a template for comparison with the peptides generated in this work.

Reprint requests to: Roger J.W. Truscott, Australian Cataract Research Foundation, The University of Wollongong, NSW 2522 Australia.

⁴Current address: Research School of Chemistry, The Australian National University, Canberra, ACT 0200 Australia.

Fig. 1. Amino acid sequence of bovine γ E crystallin as determined by Edman amino acid sequencing of various chemical cleavages and enzymatic digests. Corrections to the sequence proposed by White et al. (1989) are shown in outline. Masses of the peptides found by ES-MS are indicated within the arrows. Carboxymethylcysteine residues are represented by C.

Chemical cleavage of γ E with BNPS-skatole will theoretically generate five peptides, which consist of residues 1–42 (Wa), 43–68 (Wb), 69–130 (Wc), 131–156 (Wd), and 157–173 (We), consecutively. The cleavage mixture was separated by HPLC and the peaks (with the exception of the peptide Wc, which was insoluble) were analyzed subsequently by Edman amino acid sequencing. Results obtained in this experiment indicated that residues 1–36 (Wa) of the sequence reported by White et al. (1989) were correct (Fig. 1). ES-MS and ES-MS/MS of tryptic peptides T1+T2 (residues 1–9, GKITF YEDR, 1,128 Da) and T4 (residues 15–31, HYECS SDHSN LQPYF SR, 2,127 Da) also confirmed their respective sections of the first 36 residues of the sequence. Results obtained from the amino acid sequencing of the HPLC peak containing peptide Wb (residues 43–68) indicated possible errors in the sequence, i.e., T $\rightarrow$ Q at residue 51, and C $\rightarrow$ P at residue 53. Subsequent amino acid sequencing and ES-MS/MS analysis of tryptic peptide T6 (residues 37–58, VDSGC WMLYE QPNFQ GPQYF LR, 2,735 Da) confirmed this. The results of the analyses of tryptic peptide T6 also confirmed the sequence of the latter half of Wa (i.e., residues 37–42). Results obtained from amino acid sequencing of the HPLC peak containing peptide Wd (residues 131–156) indicated two possible errors in the sequence, i.e., P $\rightarrow$ T at residue 136, and L $\rightarrow$ H at residue 154. Amino acid sequencing of the HPLC peak containing peptide We (residues 157–173) as the major component also indicated only one possible error in the sequence, i.e., A $\rightarrow$ T at residue 159. The substitutions at residues 154 and 159 were confirmed by ES-MS of tryptic peptide T19 (residues 153–162, YHDWG ATNAR, 1,189 Da) and also by ES-MS/MS, which yielded numerous sequence-specific daughter ions. The relative molecular mass of the protein with the above corrections incorporated is 20,973.3, which is still +18 Da different from the relative molecular mass determined by ES-MS (20,955 $\pm$ 5). Therefore, it seemed likely that there were other remaining errors in the sequence and that they were present in the portion of the polypeptide corresponding to residues 69–130 (Wc). Examination of this section of the published sequence (residues 69–130, MGFSD SVRSC RLIPH TSSHRLRIYE REDYR QGMVE ITEDC SSLQD RFHFS DIHSF HVMEG YW) showed the presence of three methionine residues at positions 69, 101, and 126, respectively. Therefore, a CNBr cleavage (M/) was performed on the intact reduced and carboxymethylated protein in an attempt to cleave this large insoluble peptide into three smaller, and, hope-

fully, more soluble, peptides. Amino acid sequencing of the CNBr cleavage mixture showed that only five peptides were present in the mixture, which indicated one less Met residue than in the sequence reported by White et al. (1989). Analysis of the sequence data acquired on the mixture indicated that the bulk of the remaining sequence errors were, in fact, contained in residues 69–130 (i.e., the insoluble Wc peptide from the BNPS-skatole digest), and that the methionine residue at position 126 of the sequence proposed by White et al. was, in fact, a leucine residue (i.e., M $\rightarrow$ L at residue 126).

The CNBr cleavage mixture was separated via RP-HPLC; however, only one of the reduced and carboxymethylated CNBr peptides yielded an ES-MS spectrum (i.e., residues 70–101, 3,840.2 Da). It was assumed that the other peptides were either not soluble or not present in sufficient amounts for electrospray analysis. Amino acid sequencing of this purified CNBr peptide (residues 70–101) indicated four differences compared to the sequence reported by White et al. (1989), namely F $\rightarrow$ L at residue 71, S $\rightarrow$ N at residue 72, V $\rightarrow$ I at residue 75, and S $\rightarrow$ G at residue 85. These substitutions were also confirmed by ES-MS and ES-MS/MS analysis of tryptic peptides T8 (residues 60–76, GDYPD YQQWM GLNDS IR, 2,057 Da) and T10 (residues 80–88, LIPHT GSHR, 1,016 Da), both of which gave characteristic daughter ion fragments for their respective corrected sequences.

Therefore, the only section of the protein remaining to be examined was residues 102–130 (the latter half of BNPS-skatole peptide Wc), which is encompassed by tryptic peptides T14 (residues 99–114, GQMVE ITEDC SSLHD R, 1,876 Da) and T15 (residues 115–139). Edman amino acid sequencing of HPLC-purified tryptic peptide T14+T15 indicated five possible sequence errors in residues 102–130 (not including the sequence error at residue 136), i.e., Q $\rightarrow$ H at residue 112, D $\rightarrow$ E at residue 119, H $\rightarrow$ N at residue 124, M $\rightarrow$ L at residue 126, and Y $\rightarrow$ W at residue 129. These substitutions were also confirmed by ES-MS/MS analysis of tryptic peptides T14 (residues 99–114, GQMVE ITEDC SSLHD R, 1,876 Da) and T14+T15 (residues 99–139, GQMVE ITEDC SSLHD RFHFS EIHSF NVLEG WVWVLY EMTNY R, 5,065 Da). Tandem mass spectral analysis of T14 also confirmed the corrected sequence, giving numerous characteristic daughter ion fragments.

In conclusion, bovine gamma E crystallin was sequenced via Edman amino acid sequencing of BNPS-skatole (W/) and CNBr cleavages (M/). Tryptic digestion (R/K,\P) coupled to electrospray mass spectrometry and tandem mass spectrometry was also utilized to confirm the sequences of peptides determined by Edman amino acid sequencing. In all, 14 errors in the sequence proposed by White et al. (1989) were demonstrated. The calculated average molecular weight of the corrected polypeptide is 20,955.3 Da. This mass agrees well with results from ES-MS analysis of the native protein, which yielded an average molecular weight of 20,955.2 $\pm$ 5 Da. The corrected sequence is also in agreement with a recent cDNA sequence obtained for a bovine γ -crystallin by R. Hay (pers. comm.).

Materials and methods: Lenses were extracted in 50 mM Tris buffer, pH 7.45, containing 5 mM Na₂ EDTA, 1 mM phenylmethylsulphonylfluoride, 1 mM 1,4-dithiothreitol (DTT), and 0.02% (w/v) sodium azide. The homogenate was centrifuged (12,000 rpm for 30 min) to remove insoluble protein and then dialyzed against Milli-QTM water (five changes). The initial separation of the total bovine lens extract containing the α , β , and γ -crystallins was

performed on a Pharmacia Sephacryl S-300 gel filtration column (1 m $\times$ 25 mm) using 0.05 M phosphate buffer, pH 6.8, containing 0.05% NaN₃ as the eluent. The individual γ -crystallin subunits were purified by FPLC according to the method of Tardieu et al. (1992) on a Pharmacia Mono S HR 10/10 cation exchange column equilibrated with 0.05 M 2-[*N*-morpholino-ethanesulphonic acid] (MES), pH 6.05. A 10% linear salt gradient was employed using 1 M NaCl in 0.05 M MES, pH 6.05. Each fraction was dialyzed extensively against Milli-QTM water (five changes), then lyophilized prior to ES-MS analysis.

Purified γ E-crystallin was reduced and carboxymethylated as follows: the protein was dissolved (5 mg/mL) in 0.1 M Tris buffer, pH 8.3, containing 6 M guanidinium HCl and 1 mM Na₂EDTA. To this mixture, a solution of 100 mM DTT was added to give a final concentration of 2 mM. The resultant mixture was sealed under nitrogen and incubated at 38 °C. After 1 h, a 50-mM solution of iodoacetic acid (neutralized with NaOH) was added to give a final concentration of 5 mM. The mixture was again sealed under nitrogen and incubated for 1 h at 38 °C in the dark, after which the reaction was quenched by the addition of 1% (v/v) β -mercaptoethanol. The mixture was then dialyzed extensively against Milli-QTM water (five changes), during which the reduced and carboxymethylated protein precipitated. The precipitate was collected by centrifugation, dried under vacuum, and then stored at -20 °C until required for ES-MS analysis and enzymatic digestion.

For BNPS-skatole (3-bromo-3-methyl-2-(2-nitrophenylmercapto)-3H-indole) digestion of the reduced and carboxymethylated γ E-crystallin sample, the lyophilized protein (approximately 1 mg) was dissolved in 80 μ L of concentrated formic acid to which 20 μ L of Milli-Q water containing approximately 10% (w/v) cresol was added. Approximately 1 mg of the BNPS-skatole was dissolved in 1 mL 70% (v/v) aqueous formic acid, and 100 μ L of this solution was added to the protein solution, sealed under nitrogen, and incubated at room temperature for 48 h in the dark. One microliter of β -mercaptoethanol was added to the solution, left for 12 h, then diluted with Milli-Q water (approximately 600 μ L), and excess reagent and by-product removed by repeated ($\times$ 3) extraction with ethyl acetate (discarding the organic layer at each extraction), lyophilized, and stored at -20 °C until required.

For CNBr digestion of the reduced and carboxymethylated γ E-crystallin sample, the lyophilized protein (approximately 1 mg) was dissolved in 10 mL of concentrated formic acid to which 1 mL of β -mercaptoethanol was added. The sample was then sealed under nitrogen and incubated at 38 °C for 1 h and then immediately lyophilized. The dried protein was reconstituted in 70 mL of concentrated formic to which 30 mL of Milli-QTM water containing ~10% (w/v) cresol was added.

Approximately 1 mg of CNBr was dissolved in 100 mL concentrated formic acid and diluted to ~70% formic by the addition of Milli-QTM water containing ~10% (w/v) cresol. CNBr solution (100 mL) was added to the protein solution, which was then sealed under nitrogen and incubated at room temperature for at least 24 h in the dark. To remove any traces of cresol, the sample was diluted to 3 mL with Milli-QTM water and extracted three times with ethyl acetate (discarding the organic layer after each extraction). The sample was then lyophilized, reconstituted in 200 mL of 80% (v/v) formic acid, re-lyophilized, and then stored at -20 °C until required.

For tryptic digestion of the reduced and carboxymethylated γ E-crystallin sample, the lyophilized protein was dissolved in 100 mM Tris buffer, pH 8.2, containing 10 mM CaCl₂. Trypsin (bovine

pancreas, TPCK treated type XIII, 1 mg/mL in 0.1 M HCl) was added to the protein at 1% (w/w). The sample was then sealed under nitrogen and incubated at 38 °C for 18 h. The digest mixture was kept at -20 °C until required for HPLC separation.

HPLC was performed on a Waters HPLC system consisting of two model 451 HPLC pumps, a model 680 automated gradient controller, and a model 441 absorbance detector set at 214 nm. An Alltech alltima 5-micron C18 reversed-phase HPLC column (250 $\times$ 4.6 mm) was used and this was equilibrated in 10% aqueous acetonitrile containing 0.1% TFA (solvent A). Peptide separation was achieved by a linear gradient of 0–45% solvent B (0.1% TFA in CH₃CN) over 60 min at a flow rate of 1 mL/min. All peaks collected were stored at -20 °C until required for ES-MS analysis and amino acid sequencing.

Amino acid sequencing was performed on a PE Applied Biosystems ProciseTM 494 protein sequencer. The peptides were adsorbed onto ProBlotTM PVDF membranes and sequenced using standard pulsed liquid blot sequence procedures.

All electrospray mass spectra were acquired on a VG Quattro II mass spectrometer (VG Biotech Ltd., Altonham, Cheshire, UK) equipped with an electrospray ionization source and a quadrupole-hexapole-quadrupole mass analyzer configuration with a 4,000 *m/z* range for singly charged species. The solvent (50% aqueous MeOH) for the spray was delivered via an HP 1090 liquid chromatograph at a flow rate of 8 mL/min. A nitrogen nebulizing gas, flowing concurrent to the stainless steel capillary in the probe at a flow rate of ~10 L/h, assisted in the production of a stable spray. A dry, warm counter-current nitrogen bath gas at a flow rate of ~350 L/h and at atmospheric pressure was employed to assist evaporation of the electrospray solvent. The electrospray probe tip potential was 3.2 kV with 0.2 kV on the pepper-pot counter electrode. Skimmer potentials ranging from 45 to 65 V were used, and the resolution was set at 0.5 Da to distinguish doubly and triply charged ions. The photomultiplier was set at 650 V for the first mass analyzer (MS1). For ES-MS/MS experiments, argon was used as the collision gas and the collision energy varied depending on the peptide analyzed. The photomultiplier was set at 850 V for the second mass analyzer (MS2) and the resolution was set at 1 Da. All mass spectra and tandem mass spectra were acquired in the positive-ion mode by multi-channel analysis (MCA) at a rate of 1 s per *m/z* 100. Typically, data from 10–20 scans were summed to obtain representative spectra. All mass spectra acquired were processed using software supplied by VG Biotech. The mass spectrometer was calibrated using Cs/RbI when analyzing the tryptic peptides and horse heart myoglobin (*M_r* 16,951.5, 0.4 pmol/mL) when analyzing the intact protein. Tryptic peptides were analyzed directly from the HPLC solvent at the concentration in which they eluted from the column. Amino acid sequencing data indicated that the concentrations of the tryptic peptide solutions were on the order of 1 pmol/ μ L.

References

- 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.
- Chirgadze Y, Nevskaya N, Vernoslova E, Nikonov S, Sergeev Y, Brazhnikov E, Fomenkova N, Lunin V, Urzhumtsev A. 1991. Crystal structure of calf eye lens gamma-crystallin IIb at 2.5 Å resolution: Its relation to function. *Exp Eye Res* 53:295–304.
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. 1990. Electrospray ionization—Principles and practice. *Mass Spectrom Rev* 9:37–70.
- Harding JJ. 1991. *Cataract biochemistry, epidemiology and pharmacology*. London: Chapman and Hall.

- Harding JJ, Dilley KJ. 1976. Structural proteins of the mammalian lens: A review with emphasis on changes in development, aging and cataract *Exp Eye Res* 22:1–73.
- Kilby GW, Truscott RJW, Aquilina JA, Sheil MM, Riley ML, Harding JJ. 1995. Electrospray ionisation mass spectrometry of lens crystallins: Verification of, and detection of errors in, protein sequences of bovine γ -crystallins. *Eur Mass Spectrom* 1:203–208.
- Tardieu A, Veretout F, Krop B, Slingsby C. 1992. Protein interactions in the calf lens: Interactions between beta crystallins are repulsive whereas in gamma crystallins they are attractive. *Eur Biophys J* 21:1.
- Truscott RJW, Augusteyn RC. 1977a. Changes in human lens proteins during nuclear cataract formation. *Exp Eye Res* 24:159–170.
- Truscott RJW, Augusteyn RC. 1977b. The state of sulphydryl groups in normal and cataractous human lenses. *Exp Eye Res* 25:139–148.
- van den Oetelaar PJM, Hoenders HJ. 1989. Racemisation of aspartyl residues in proteins from normal and cataractous human lenses: An aging process without involvement of cataract formation. *Exp Eye Res* 48:209–214.
- White HE, Driessen HPC, Slingsby C, Moss DS, Lindley PF. 1989. Packing interactions in the eye lens: Structural analysis, internal symmetry and lattice interactions of bovine γ IVa-crystallin. *J Mol Biol* 207:217–235.
- Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T. 1983. X-ray analysis of the eye lens protein γ -II crystallin at 1.9 Å resolution. *J Mol Biol* 170:175–202.