The complete sequence of human lens γs-crystallin

Jean B. SMITH,*§ Zhijing YANG,* Peiping LIN,* Zafar ZAIDI,† Atiya ABBASI† and Paul RUSSELL‡

Department of Medicinal Chemistry, Purdue University, *West Lafayette, IN 47907, U.S.A., †H. E. J. Research Institute of Chemistry, University of Karachi, Karachi-75270, Pakistan, and ‡National Eye Institute, National Institutes of Health, Bethesda, MD 20892, U.S.A.

The complete sequence of human γs-crystallin has been determined and confirmed using a combination of MS methods, peptide sequencing and cDNA sequencing. Regions 21–35 and 102–107, which were previously assumed to be the same as the bovine sequence, differ from the bovine sequence at residues 22, 28, 31 and 104. An additional six residues were also found to be different from the original sequence determined for Pakastani lenses. Whether these differences represent errors in the original sequence or two different sequences among human lens crystallins is not yet known.

INTRODUCTION

The sequence for human lens γs-crystallin, formerly called βs-crystallin, published in [1], included two regions, residues 21–35 and 102–107, that were assumed to be the same as the bovine sequence. Because γs-crystallin is the major γ-crystallin of the human lens [2] and is a principal component of the soluble high molecular mass proteins [3] that are associated with aging and cataract [4], it is important that the sequence of γs-cystallin be known in its entirety. Using γs-crystallin isolated from non-cataractous lenses from donors in the U.S.A., we have determined that the sequences of the two unknown regions differ from the bovine sequence. We have also found an additional six residues that do not agree with the published sequence.

EXPERIMENTAL

Lenses obtained from The National Disease Research Interchange (Philadelphia, PA, U.S.A.) were removed within 24 h post-mortem and shipped on solid CO_2. They were stored at −80 °C until analysis. The γs-crystallins were isolated from 12 clear donor lenses, ranging in age from 12 to 57 years. Two pairs of lenses were used in the cDNA sequencing; 8 lenses were separately analysed by mass spectrometric methods. None of the donors were diabetic or uremic nor had they been taking any drugs associated with cataractogenesis.

Isolation of proteins

Each lens was homogenized in a buffer of 0.5 M NaCl/50 mM Tris-HCl/1 mM EDTA, pH 7.4, for 1 h and centrifuged at 15000 *g* for 1 h. The water-soluble crystallins were fractionated into α-, β- and γ-crystallins by gel filtration chromatography using a 2.5 × 85 cm column of Sephadex G-200 (Pharmacia, Piscataway, NJ, U.S.A.) with a flow rate of 6–7 ml/h. The UV absorbance at 280 nm was measured for each 2 ml fraction with a DU-40 spectrophotometer (Beckman, Irvine, CA, U.S.A.). Fractions corresponding to α-, β- and γ-crystallins were pooled. The γ-crystallins were further fractionated by reverse-phase HPLC on a C-4 column (0.46 × 15 cm, 300 Å, Vydac, Hesperia, CA, U.S.A.) with a gradient HPLC system (Rainin, Woburn, MA, U.S.A.). Solvent A was 0.1 % trifluoroacetic acid (TFA) in water, and solvent B was 0.1 % TFA in acetonitrile. The gradient was 0–30 % B in 10 min and 30–60 % B at 65 min. UV absorbance of the eluent was monitored at 280 nm with a Spectro-Monitor 3100 UV detector (Milton Roy, Riviera Beach, FL, U.S.A.). The peak which eluted at 38–39 % acetonitrile (Figure 1) was nearly pure γs-crystallin as judged by electrospray ionization mass spectrometry (ESI-MS) spectra and peptide mapping results.

Proteolytic digestion and chemical cleavage

Portions of the HPLC-isolated protein corresponding to approximately 200 mg were redissolved in a buffer of 0.1 M Tris/HCl, pH 8.2, and digested for 4 h with either trypsin (Worthington, Freehold, NJ, U.S.A.) or Asp-N (Boehringer Mannheim, Indianapolis, IN, U.S.A.) at an enzyme:substrate ratio of 1:50. Trypsin cleaves C-terminally to Arg and Lys; endoproteinase Asp-N cleaves N-terminally to Asp and sometimes Glu. Another portion (200 mg) of the γs fraction was dissolved in 50 ml of 70 % formic acid, a crystal of CNBr was added and the solution was incubated in the dark at 25 °C for 18 h. Cyanogen bromide cleaves C-terminally to Met, forming the lactone. The sample was diluted with H_2O, dried, re-solubilized in 0.1 M Tris/6 M guanidine hydrochloride/0.1 M dithiothreitol, pH 8.5. The peptides resulting from the enzymic digestions or the chemical cleavage were separated by reverse-phase HPLC using a Vydac C-18 column and a linear gradient of 1–40 % CH_3CN over

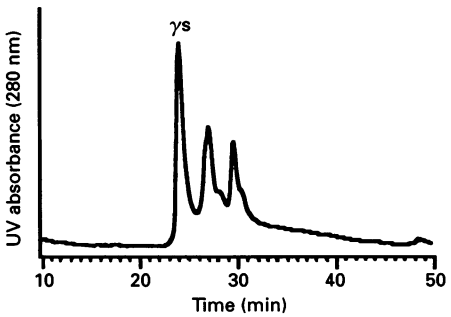

Figure 1 Reverse-phase HPLC chromatogram of the fractionation of human γ-crystallins, previously isolated by gel filtration

The first peak is nearly pure γs-crystallin.

Abbrevations used: TFA, trifluoroacetic acid; FAB, fast atom bombardment; ESI, electrospray ionization.

§ To whom correspondence should be addressed.

60 min. The fractions were collected and the molecular masses of the peptides present in each fraction were determined by fast atom bombardment mass spectrometry (FAB-MS) or ESI-MS.

FAB-MS and ESI-MS analyses

ESI-MS is capable of determining the molecular mass of the intact protein; both ESI-MS and FAB-MS can be used to determine the molecular masses of peptides. ESI-MS analyses were performed on a Kratos MS-25 fitted with an ESI source. The solvent was water:methanol (50:50) with 3% acetic acid. Mass accuracy was determined by analysis of myoglobin under the same conditions that were used to analyse the sample. The precision (approx. 0.01%) was 2–3 u at a molecular mass of 20000 Da. Analyses by FAB-MS were performed on a Kratos MS-50 (Kratos Analytical, Manchester, U.K.). Peptide samples (1 μl of a 1 nmol/ml solution in 0.1% TFA) were mixed with a glycerol:thioglycerol (1:1) matrix. The instrument was calibrated with CsI; mass accuracy was within 0.3 u.

Peptide sequencing

Peptide sequencing was provided by Dr. Zafar Zaidi's laboratory at the H.E.J. Research Institute of Chemistry in Karachi, Pakistan or by the Purdue Laboratory for Macromolecular Structure. Both laboratories used an automated Edman degradation method [5] with an on-line HPLC analyser.

cDNA Sequencing

The human lens cDNA was made from mRNA extracted from two pairs of lenses aged 12 and 48 years. The cDNA was used with the Taq DyeDeoxy Terminator Sequencing kit (Applied Biosystems Inc., Forster City, CA, U.S.A.) to generate the sequence for the γs-crystallin. The manufacturer's protocol was followed exactly. The oligonucleotide primers used for the PCR reaction were: 5′-GAAGACAAAAACCTTCAAGGCCGC and 5′-GGAACCAAAATTACTTTCT for the region upstream of the nucleotides encoding amino acid 16 and 5′-GGAAGGGC-AGTCTTCCGTGGTCTC for the region downstream of the nucleotides encoding amino acid 106. The sequence was read using a 370A DNA Sequencer (Applied Biosystems).

RESULTS

The difference between the molecular mass of the intact protein determined by ESI-MS (20918 Da) and the molecular mass calculated from the published sequence (20891 Da) indicated that either the sequence was different or that our samples included a modification not present in the samples analysed in Pakistan. Analysis of the molecular masses of the peptides produced by enzymic digestion or chemical cleavage confirmed major portions of the published sequence, including residues 7–18, 59–100 and 119–177. A variety of techniques were used to obtain the sequence of the remaining portions. One approach, illustrated in Figure 2, was to digest an unknown peptide with another enzyme such as

```
   CGCTATGACT GTGATTGCGA CTGTGCAGAT TTCCACACAT ACCTAAGTCG CTGCAACTCC
19   R  Y  D   C  D  C   D  C  A  D   F  H  T   Y  L  S   R  C  N  S

   ATTAAAGTGG AAGGAGGCAC CTGGGCTGTT TATGAAAGGC CCAACTTTGC TGGGTACATG
39   I  K  V   E  G  G   T  W  A  V   Y  E  R   P  N  F   A  G  Y  M

   TACATCTTAC CACAGGGAGA GTACCCTGAA TACCAGCGTT GGATGGGCCT CAACGACCGC
59   Y  I  L   P  Q  G   E  Y  P  E   Y  Q  R   W  M  G   L  N  D  R

   CTCAGCTCCT GCAGAGCTGT TCATCTGCCT AGTGGAGGCC AGTATAAGAT TCAGATCTTT
79   L  S  S   C  R  A   V  H  L  P   S  G  G   Q  Y  K   I  Q  I  F

   GAGAAAGGGG ATTTTAGTGG TCAG
99   E  K  G   D  F  S   G  Q
```

Figure 3 The nucleotide sequence and deduced amino acid sequence of human γs-crystallin from amino acids 19 to 106

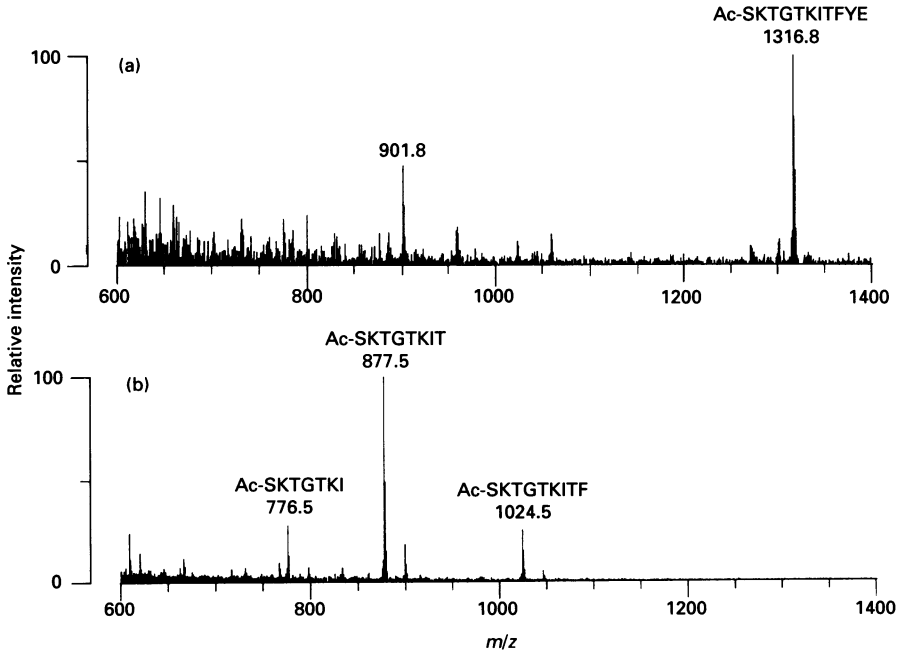

Figure 2 FAB mass spectra illustrating the confirmation of MH^+ 1316 as γs-crystallin peptide 1-11

(**a**) Before and (**b**) after digestion of the peptide with a mixture of carboxypeptidases B and Y.

Table 1 Revisions to the published sequence of human lens γs-crystallin

Revision	Method of confirmation
Thr-3 instead of Ala-3	FAB-MS analysis of peptide 1–11 after carboxypeptidase digestion
Arg-19 instead of His-19	cDNA deduced sequence
Cys-22 instead of Ser-22	cDNA deduced sequence
Asp-28 instead of Glu-28	cDNA deduced sequence
Thr-31 instead of Met-31	cDNA deduced sequence
Lys-40 instead of Arg-40	Amino acid sequencing of residues 36–40
Arg-51 instead of Thr-51	FAB-MS analysis of peptide 50–76 after carboxypeptidase digestion
Ser-104 instead of Asn-104	Amino acid sequencing of residues 102–106
Tyr-108 instead of Arg-108	FAB-MS analysis of peptide 101–124 after pepsin digestion
Asp-113 instead of Asn-113	FAB-MS analysis of peptide 113–118 after carboxypeptidase digestion

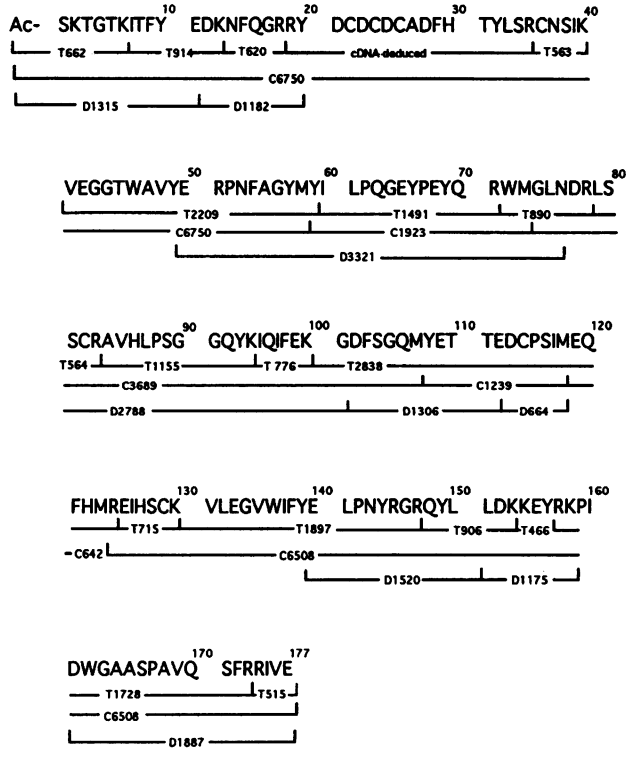

Figure 4 The complete sequence of human γs-crystallin

Brackets underneath the sequence indicate the peptides found in tryptic (T), Asp-N (D) and CNBr (C) digests. Region 19–35 was determined by cDNA sequencing.

carboxypeptidase or pepsin. The peptide in Figure 2, MH^+ 1316, was treated with a mixture of carboxypeptidases B and Y to sequentially remove amino acids from the C-terminus. The presence of peaks at MH^+ 1024 (loss of YE), MH^+ 877 (loss of FYE) and MH^+ 776 (loss of TFYE), confirmed that this peptide was from the N-terminus and that there was a difference of +30 u present in residues 1–7. The molecular masses of the fragment ion found in the FAB mass spectrum of peptide MH^+ 1316 (results not shown) indicated that the difference was at residue 2 or 3. The molecular mass change of 30 fits substitution of Thr for Ala at residue 3. No other single substitution for Lys or Ala has a change of 30. Similar analyses showed that residue 51 was Arg instead of Thr, residue 108 was Tyr instead of Arg and residue 113 was Asp instead of Asn. The change at residue 113 was also indicated by Asp-N cleavage before 113. The changes at residue 40, Lys instead of Arg, and residue 104, Ser instead of Asn, were established by amino acid sequencing of peptides 36–40 and 102–106. Peptides corresponding to residues 19–35 were not found in either the tryptic or Asp-N digests. Using the sequences of known peptides as probes, cDNA sequencing showed that region 19–35 contained four changes, Arg-19 instead of His-19, Cys-22 instead of Ser-22, Asp-28 instead of Glu-28, and Thr-31 instead of Met-31. Continued cDNA sequencing also confirmed the changes at residues 40, 51 and 104 (Figure 3). The revisions to the sequence and the methods of confirmation are listed in Table 1. The complete sequence consistent with the present results is given in Figure 4.

DISCUSSION

This determination of the sequence of human γs-crystallin illustrates the advantages of using several techniques. Although *de novo* sequencing can be done by mass spectrometric techniques, ESI-MS and FAB-MS are most valuable for confirming and/or correcting previously determined sequences [6]. The published sequence of human γs-crystallin had been determined by protein sequencing. ESI-MS and FAB-MS showed that the protein isolated from lenses from U.S.A. donors did not agree with the published sequence. Mass spectrometric methods alone determined changes at four residues, a combination of MS data and peptide sequencing data specified two additional changes, and four more changes were determined by cDNA sequencing. Although any one technique should, theoretically, be capable of determining the correct sequence, there are conditions when one technique has advantages over another and the most efficient procedure to obtain a correct sequence includes the complementary use of several techniques. For example, peptide/protein sequencing requires relatively pure peptides, cannot be performed for N-terminally blocked peptides and does not indicate post-translationally modified peptides. MS can determine the molecular masses of peptides in a mixture and confirm sequenator results. On the other hand, sequencing by interpreting mass spectral fragmentation patterns is difficult when co-eluting peptides have similar molecular masses. For example, unknown peptide 36–40 (MH^+ 564) and peptide 79–83 (MH^+ 565) eluted in the same fraction. The masses were too close to obtain easily interpreted tandem MS fragmentation patterns. In this case, the data from peptide sequencing showed the presence of two residues at each cycle. One amino acid of each cycle fitted the known peptide 79–83. The other amino acid at each cycle could then be attributed to the unknown peptide, 36–40.

The peptide for residues 19–35 was particularly elusive in mass spectral analysis and could not be identified in the HPLC chromatogram. This peptide may not have behaved well in the HPLC fractionation or it may not have given a good MS response because of the high incidence of cysteinyl residues, Cys-22, Cys-24 and Cys-26. Cysteine-containing peptides are notoriously bad responders in FAB-MS analysis because of their high affinity for the glycerol/thioglycerol matrix. Because the sequences for regions on either side of the unknown portion provided the necessary probes, cDNA sequencing could determine the sequence of peptide 19–35. Errors can also occur if

one relies on cDNA sequencing alone. With PCR, misincorporation of nucleotides by the polymerase is a known source for inaccuracies in the data. Mispriming and primer–dimer artifacts also can cause flaws in the sequence. Band compression, although alleviated somewhat with the DyeDeoxy Terminator method, can cause misreading of the sequence. The deduced protein is actually the consensus sequence that is obtained after a series of reactions. Identity between this deduced protein and the sequence information obtained by MS and peptide sequencing indicates a high level of confidence in the proposed γs-crystallin sequence.

The disagreement between the sequence determined for γs-crystallin from the lenses from U.S.A. donors and that determined for γs-crystallin from Pakistani lenses raises the question of whether there were errors in the initial determination or whether variation in the sequence between different donors exists. In our analysis of γs-crystallins from eight U.S.A. donors, we found no evidence of more than one correct sequence.

Received 9 August 1994/16 November 1994; accepted 23 November 1994

This work was supported by Grant EY07609 from the National Institutes of Health.

REFERENCES

1 Zarina, S., Abbasi, A. and Zaidi, Z. H. (1992) Biochem. J. **287**, 375–381
2 Miesbauer, L. R., Zhou, X., Yang, Z., Yang, Z., Sun, Y., Smith, D. L. and Smith, J. B. (1994) J. Biol. Chem. **269**, 12494–12502
3 Yang, Z., Chamorro, M., Smith, D. L. and Smith, J. B. (1994) Curr. Eye Res. **13**, 415–421
4 Jedziniak, J. A., Kinoshita, J. H., Yates, E. M. and Benedek, G. B. (1975) Exp. Eye Res. **20**, 367–369
5 Edman, P. and Begg, G. (1967) Eur. J. Biochem. **1**, 80–91
6 Gibson, B. W. and Biemann, K. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 1956-1960