Covalent structure of bovine brain calreticulin

Kunie MATSUOKA,* Kazuo SETA,* Yoshio YAMAKAWA,† Tsuneo OKUYAMA,* Tomotaka SHINODA* and Toshiaki ISOBE*‡

*Department of Chemistry, Faculty of Science, Tokyo Metropolitan University, Hachioji-shi, Tokyo 192-03, Japan

and †Department of Biochemistry and Cell Biology, National Institute of Health, Tokyo 162, Japan

The covalent structure of bovine brain calreticulin, a major Ca^{2+} -binding protein in the lumen of the endoplasmic reticulum, was determined by analysis of the purified protein. The protein consisted of 400 amino acids, with an N-linked oligosaccharide attached to the polypeptide chain. The polypeptide sequence determined was compatible with the sequence of calreticulin deduced from cDNA of different sources, with a number of differences presumably due to species-specific amino acid substitutions. The protein retained the C-terminal tetrapeptide, KDEL, involved in retention of proteins resident in the endoplasmic reticulum, whereas the N-terminal signal peptide

predicted from the cDNA sequence had been removed in the purified protein. The bovine brain protein contained a high-mannose type of oligosaccharide attached to Asn¹⁶², which is typical of resident endoplasmic reticulum proteins. The carbohydrate moiety was heterogeneous and had the composition $\text{GlcNAc}_2\text{Man}_{4-9}$, of which $\text{GlcNAc}_2\text{Man}_5$ was the most abundant in the bovine brain preparation. Glycosylation of calreticulin, however, appeared to be a species-specific modification, as Asn¹⁶² is replaced by Asp in the sequences already determined for a number of species. Analysis of the purified protein also identified an intramolecular disulphide bridge between Cys¹²⁰ and Cys¹⁴⁶.

INTRODUCTION

Calreticulin is a Ca^{2+} -binding protein found mainly in the lumen of the endoplasmic reticulum (ER) (Opas et al., 1991; Michalak et al., 1991). It is widely distributed among eukaryotic cells, and is suggested to have a role in Ca^{2+} storage within the ER (Milner et al., 1991). For many reasons, however, the protein is believed to have more diverse, and perhaps very basic, physiological functions [reviewed by Michalak et al. (1992)].

Calreticulin is an acidic protein of M_r 46000, and contains two distinct type of Ca^{2+} -binding site, one high affinity/low capacity and the other low affinity/high capacity (Ostwald and MacLennan, 1974). The amino acid sequence, deduced from its cDNA (Fliegel et al., 1989; Smith and Koch, 1989; McCauliffe et al., 1990a; Murthy et al., 1990), indicates that the calreticulin sequence starts with a hydrophobic N-terminal signal sequence typical of many membrane and secreted proteins and terminates with the C-terminal tetrapeptide sequence KDEL which is involved in the retention of proteins resident in the ER (Munro and Pelham, 1987). The mature protein, comprising 400 amino acids, consists of distinct structural domains, and the different domains are thought to be the Ca^{2+} -binding sites of different affinities as analysed by using mutant recombinant proteins (Baksh and Michalak, 1991).

In spite of extensive structural studies, it is currently unclear whether calreticulin is a glycosylated protein. Although the available sequences of calreticulins from different sources contain one potential site for N-linked oligosaccharide, no carbohydrate has been detected in human or mouse muscle calreticulin or rabbit and chicken liver calreticulin (Waisman et al., 1985; Lewis et al., 1985; McCauliffe et al., 1990a; Milner et al., 1991). However, bovine liver calreticulin appears to be glycosylated, as it binds to concanavalin A-Sepharose and is susceptible to endoglycosidase digestion (Waisman et al., 1985; Lewis et al., 1985). In addition, the rat liver protein has been shown to be glycosylated by a complex hybrid type of oligosaccharide with a terminal galactose residue (Van et al., 1989; Peter et al., 1992).

Because this type of oligosaccharide is very unusual in resident ER protein, this raises the question of not only the structure of calreticulin, but also the more general question concerning the trafficking of proteins between different cellular compartments.

The present paper describes the covalent structure of bovine brain calreticulin determined by analysis of the purified protein. The result confirms the previously reported sequences of calreticulins deduced from cDNA, and provides additional information on the structure of the oligosaccharide and the position of the disulphide bridge.

EXPERIMENTAL

Materials

Bovine brains were obtained fresh from a slaughterhouse. DEAE-cellulofine A-800 was obtained from Seikagaku Kogyo Co. (Tokyo, Japan). Lysyl endopeptidase was purchased from Wako Pure Chemical Industries (Tokyo, Japan), *Staphylococcus aureus* V8 protease from Cooper Biomedical (Malvern, PA, U.S.A.) and glycopeptidase F from Boehringer (Mannheim, Germany). 4-Vinylpyridine was the product of Tokyo Kasei Industries (Tokyo, Japan). Reagents for h.p.l.c. [trifluoroacetic acid and heptafluorobutyric acid (sequanal grades)] were obtained from Wako, and acetonitrile (chromatography grade) was from Merck (Darmstadt, Germany). Water was distilled, passed through a mixed-bed ion-exchange resin, and redistilled before use. Other chemicals were of analytical reagent grade obtained from Wako, unless mentioned otherwise.

Purification of calreticulin

Calreticulin was prepared from bovine brain extracts essentially by our large-scale procedure developed to purify a series of brain acidic proteins (Ishioka et al., 1984). Briefly, 50 brains (20 kg) were homogenized with 4 vol. of 0.1 M potassium phosphate buffer (pH 7.1) containing 2.66 M $(\text{NH}_4)_2\text{SO}_4$ and 1 mM EDTA, and the homogenate was centrifuged at 10000 g for 30 min.

Powdered $(\text{NH}_4)_2\text{SO}_4$ was added to the supernatant to 85% saturation at pH 4.7; the precipitate was collected by centrifugation (10000 *g*), and redissolved in a minimal volume of 0.1 M potassium phosphate buffer (pH 7.1) containing 0.05 M NaCl and 1 mM EDTA. After dialysis against the same buffer, the dialysate was applied to a column (8.5 cm internal diameter $\times$ 23 cm) of DEAE-cellulofine A-800 equilibrated with 0.1 M potassium phosphate buffer (pH 7.1) containing 0.05 M NaCl and 1 mM EDTA and washed with the equilibration buffer. Adsorbed proteins were then eluted stepwise by increasing NaCl concentration in 0.1 M potassium phosphate buffer containing 1 mM EDTA at a linear flow rate of 10.6 cm/h. The NaCl concentration was increased as follows: 0.1 M; 0.15 M; 0.20 M; 0.25 M; 0.30 M; 0.40 M. Calreticulin was obtained in a post-calmodulin fraction eluted with 0.30 M NaCl at a capacity ratio (*k'*) of 3–8. The calreticulin preparation was further purified by high-performance ion-exchange chromatography on TSK-gel DEAE-5PW (Tosoh) eluted isocratically with 25 mM Tris/HCl buffer (pH 7.5) containing 0.29 M NaCl.

Pyridylethylation

Purified calreticulin (about 8 mg) was dialysed against distilled water and lyophilized. It was dissolved in a mixture of 2 M Tris/acetate buffer (pH 7.6; 0.5 ml) and 8 M urea (3 ml), and was added with 50 μ l 2-mercaptoethanol under constant-bubbling N_2 gas. After 1 h at room temperature, three portions of 4-vinylpyridine (50 μ l each) were added to the solution at 10 min intervals. After the final addition, the mixture was allowed to stand for a further 30 min, dialysed against 50 mM NH_4HCO_3 and lyophilized.

Chemical and enzymic cleavage and peptide purification

CNBr cleavage of pyridylethylated calreticulin (1.5 mg) was performed in 70% formic acid (0.45 ml) containing 100 mM CNBr for 24 h at room temperature. The resulting fragments were isolated by directly applying the reaction mixture to a column (4.6 mm internal diameter $\times$ 75 mm) of Phenyl-5PWRP (Tosoh) followed by elution with a linear gradient of acetonitrile (10–60%) in 0.1% trifluoroacetic acid. Subdigestion of three CNBr fragments, CB1, CB3, CB4, with *S. aureus* V8 protease was carried out in 0.1 M NH_4HCO_3 (pH 8.0) at 37 °C for 5 h at an enzyme/substrate ratio of 1:50. Partial acid hydrolysis of CB1 was carried out in 0.01 M HCl at 110 °C for 5 h in an evacuated sealed tube. Digestion of pyridylethylated calreticulin (0.5 mg) with lysyl endopeptidase was performed in 0.5 ml of 0.1 M NH_4HCO_3 (pH 8.0) at 37 °C for 7 h at an enzyme/substrate ratio of 1:100. In each case, the fragments were separated by reversed-phase h.p.l.c. on a column of Vydac C4 (4.6 mm $\times$ 15 cm; Renier Instruments), Lichrosorb RP-8, select B (4.6 mm $\times$ 250 mm; Merck, Darmstadt, Germany) or Shiseido capsule type C18 (4.6 mm $\times$ 25 cm; Shiseido, Tokyo, Japan) eluted with a linear gradient of acetonitrile in 0.1% trifluoroacetic acid. Where necessary, the isolated fragments were further purified by rechromatography using 0.08% heptafluorobutyric acid in place of 0.1% trifluoroacetic acid.

Analysis of amino acid composition and sequence

The amino acid composition was determined on an automated high-performance amino acid analyser (model 800; Jasco, Tokyo, Japan) equipped with a post-column derivatization system with

o-phthalaldehyde (Cunico and Schlanbach, 1983). Peptides and proteins were hydrolysed with 6 M HCl containing 5% (v/v) phenol at 110 °C for 24 h in evacuated sealed tubes.

The N-terminal sequence of calreticulin and the sequence of peptides produced by limited cleavage of calreticulin were determined on an automated sequence analyser (model 477A or 473A; Applied Biosystems) with an on-line h.p.l.c. system for quantitative identification of phenylthiohydantoin derivatives of amino acids. The system was operated essentially according to a protocol provided by the supplier.

Mass spectroscopy

Mass spectra of peptides were obtained with a Finnigan-Mat model TSQ-700 mass spectrometer equipped with an electrospray ionization interface. Peptides were introduced directly on to the interface by an infusion method, or through a reversed-phase C18 capillary column (0.3 mm $\times$ 100 mm) connected in tandem to the interface and the spectrometer. The resulting spectra were analysed by a computer-assisted software package provided by the supplier.

Other analytical techniques

Deglycosylation of purified calreticulin with glycopeptidase F was performed in 80 mM NH_4HCO_3 (pH 8.0) containing 0.2% SDS, 1% 2-mercaptoethanol and 10 mM EDTA at 37 °C for 18 h. Bovine serum transferrin and BSA were used as controls for glycosylated and non-glycosylated protein respectively.

Two-dimensional electrophoresis was carried out by a micro-technique described by Manabe et al. (1982). Samples (1–10 μ l) were separated by first-dimensional isoelectric focusing on a capillary-gel column (1 mm $\times$ 40 mm), followed by second-dimensional electrophoresis on a polyacrylamide gradient (4–17%) slab gel (40 mm $\times$ 40 mm). Two separation systems, one in the presence and the other in the absence of denaturing agents, were employed depending on the purpose of the analysis. Proteins were detected by staining with Coomassie Brilliant Blue R-250.

RESULTS

Characterization of calreticulin from bovine brain

The preparation of bovine brain calreticulin showed a single peak on reversed-phase h.p.l.c. (not shown) and a single spot on two-dimensional electrophoresis at a position corresponding to an isoelectric point (pI) of 4.2, and had an M_r of 59 000 (Figure 1). When the total brain extract was analysed by two-dimensional electrophoresis, calreticulin was mapped at a position adjacent to albumin, smg p25A GDI (Matsui et al., 1990) or neuron-specific enolase (Ishioka et al., 1984). Because the map position of calreticulin relative to these marker proteins remained unchanged in the electrophoresis system without denaturing agents, we deduced that calreticulin is a monomer in its native state. In addition, Western-blot analysis using an antiserum against bovine brain calreticulin after SDS/PAGE of tissue extracts demonstrated a widespread distribution of this protein in various tissues outside the brain, supporting the fact that calreticulin is a constituent of a wide variety of cells including muscle and non-muscle cells (Opas et al., 1991; Michalak et al., 1991; Milner et al., 1991; Burns et al., 1992).

Sequence of the polypeptide chain

The amino acid sequence of calreticulin was determined by analysis of five CNBr-cleaved fragments (CB1–CB5) and 26

Figure 1 Two-dimensional electrophoresis of bovine brain calreticulin

The purified protein (approx. 1 μ g) was separated by first-dimensional electrofocusing on a capillary gel (1 mm $\times$ 40 mm) without denaturing agents, followed by second-dimensional electrophoresis on a polyacrylamide gradient (4–17%) slab gel (40 mm $\times$ 40 mm) containing 0.1% SDS.

Figure 2 Reversed-phase separation of CNBr peptides of S-pyridylethylated calreticulin on Phenyl-5PWRP

The peptides are assigned according to their positions in the final sequence shown in Figure 4. CB2 and CB5 were separated by rechromatography on a Vydac C4 column (4.6 cm $\times$ 15 cm) eluted with a linear gradient of acetonitrile (10–50%) in 0.08% heptafluorobutyric acid. Other experimental details are given in the Experimental section. —, A_{210} ; ····, A_{280} .

fragments produced by lysyl endopeptidase digestion of the pyridylethylated protein, which were separated by reversed-phase h.p.l.c. (Figures 2 and 3). Three large CNBr-cleaved fragments, CB1, CB3 and CB4, were subjected to secondary cleavage with *S. aureus* V8 protease, and the resulting fragments were also analysed. Likewise, one of the peptides derived by dilute acid hydrolysis of CB1 was used to obtain an overlap between two lysyl endopeptidase peptides K9 and K10. Thus more than 60 fragments were analysed to establish the total

Figure 3 Reversed-phase separation of lysyl endopeptidase peptides of S-pyridylethylated calreticulin on Lichrosorb RP-8

The peptides are assigned according to their positions in the final sequence shown in Figure 4. Experimental details are given in the Experimental section. —, A_{220} ; ····, A_{280} .

sequence. A summary of the sequence analysis, including the positions of fragments and the results of sequential Edman degradation, is shown in Figure 4. The Figure indicates that most of the residues were sequenced more than twice or confirmed by amino acid analysis of corresponding relatively small fragments. However, three residues at positions 162, 382 and 392 remained ambiguous even after the analysis described above.

Residue 162 was not identified by conventional sequential degradation of peptide CB3E2 or K16, although the amino acid compositions of both these peptides suggest that Asx (Asp or Asn) is at this position. Because the tripeptide sequence starting this residue, Asx¹⁶²-Asn-Thr, met the consensus for N-linked oligosaccharide, we tentatively assigned a glycosylated asparagine residue at position 162. This was confirmed by subsequent analysis (described below).

Residues 382 and 392 were ambiguous because peptides CB5, K29-30-31 and K32 covering these residues had highly repetitive sequences of aspartic and glutamic acids. To determine these residues, K29-30-31 and K32 were analysed by m.s. In the spectrum obtained for K29-30-31 (Figure 5a), three major ions, $(K29-30-31+2H)^{2+}$, $(K29-30-31+3H)^{3+}$ and $(K29-30-31+4H)^{4+}$, with an average mass value of 1558.0, 1039.0 and 779.5 respectively were observed, from which the mass of K29-30-31 was estimated to be 3114.0. This indicated that the mass of residue 382 was 129.1, namely glutamate, as the remaining portion of this peptide had a calculated mass of 2984.9. This assignment was later confirmed by collision-activated dissociation m.s. of peptide K31 (Figure 5b), which was isolated by further extended analysis of the lysyl endopeptidase digest of calreticulin. The m.s. analysis of K32, on the other hand, resulted in two major ions, $(K32+H)^+$ and $(K32+2H)^{2+}$, with an average mass of 1506.4 and 753.7 respectively (Figure 6), from which the mass of K32 was estimated to be 1505.4. This indicated that the

Figure 4 Amino acid sequence of bovine brain calreticulin and summary of the sequence analysis

The peptides starting with CB and K refer to peptides derived by cleavage with CNBr and lysyl endopeptidase respectively. CB-E refers to peptides from *S. aureus* V8 proteinase digestion of CNBr fragments, and DAH, dilute acid hydrolysis. Arrows indicate the sequences determined by automated Edman degradation. A part of the sequence was established by m.s. as described in the text.

mass of residue 392 was 71.0, namely alanine, as the remaining portion of K32 had a calculated mass of 1434.4.

Figure 4 shows the polypeptide sequence of bovine brain calreticulin. On the basis of this sequence, bovine brain calreticulin is composed of 400 amino acid residues with an M_r of 46379. However, it will be shown below that bovine brain calreticulin is a glycoprotein with heterogeneous carbohydrates attached to the polypeptide chain and that the M_r of the major molecular component is 47597 (average mass value) including the carbohydrate.

Disulphide bridge

The sequence analysis identified three pyridylethylated cysteines in calreticulin at positions 88, 120 and 148. Titration of purified protein with 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) in the presence of 6 M guanidinium chloride (Ellman, 1959), on the other hand, revealed the presence of 1.1 mol of free thiol group/mol of calreticulin. Because the analysis of calreticulin by two-dimensional electrophoresis (Figure 1) as well as by sedimentation equilibrium centrifugation (Waisman et al., 1985)

Figure 5 (a) Electrospray ionization mass spectrum of K29-30-31 and (b) collision-activated dissociation mass spectrum of peptide K31 (DEDEEDEDK)

For (b), predicted fragments of type A_n , B_n and Y_n used to deduce the amino acid sequence are shown above and below the structure of the top of the Figure. Those observed in the spectrum are underlined. Fragment ions of type A_n , B_n and Y_n are labelled in the sequence.

Figure 6 Electrospray ionization mass spectrum of K32

under non-denaturing conditions showed that the protein is a monomer in its native state, we predicted the presence of one intramolecular disulphide bridge.

To localize the disulphide in the sequence, an intact and a reduced form of calreticulin were digested in parallel with lysyl

Figure 7 Peptide maps of lysyl endopeptidase digests of (a) the intact and (b) the reduced form of calreticulin

Purified calreticulin was heat-treated in boiling water for 5 min with or without 1 mM dithiothreitol, digested in parallel with lysyl endopeptidase at 37 °C for 5 h, and analysed by reversed-phase h.p.l.c. on a Lichrosorb RP-8 column eluted with a linear gradient of acetonitrile (5–60%) in 0.1% trifluoroacetic acid. The cystine-containing peptide, K10-S-S-K15, was identified as shown. —, A_{210} ; ····, A_{280} .

Figure 8 Mass spectrum of K10-S-S-K15

The spectrum deconvoluted from the data obtained by electrospray ionization m.s. is shown.

endopeptidase, and the resulting digests were analysed by peptide mapping on a reversed-phase column (Figure 7). In the peptide map of the reduced protein, three peptides, K9, K10 and K15, were identified which contained Cys⁸⁸, Cys¹²⁰ and Cys¹⁴⁸ respectively. In the map of the intact protein, K9 was also identified, but the peaks for the two other peptides, K10 and K15, were not

detected. Instead, we found a peptide, K10-S-S-K15, which was unique in the intact protein and showed a double sequence on N-terminal sequence analysis that could be interpreted as an equimolar mixture of K10 and K15. In addition, this peptide had an average mass of 3959.1 (Figure 8) which coincided with the expected mass of K10 and K15 bridged by a disulphide bond. We concluded therefore that Cys¹²⁰ and Cys¹⁴⁸ formed an intramolecular disulphide bridge in the intact calreticulin.

Oligosaccharide

It has previously been reported that bovine liver calreticulin is a glycoprotein, whereas the equivalent proteins from other species such as human, rabbit, mouse and chick are not glycosylated (Waisman et al., 1985; Lewis et al., 1985; McCauliffe et al., 1990a; Milner et al., 1991). Our preparation of bovine brain calreticulin appeared to have oligosaccharide attached to the polypeptide chain, because the treatment of purified calreticulin with glycopeptidase F increased the mobility of the protein on SDS/PAGE (Figure 9), and the mass of purified calreticulin obtained by electrospray ionization m.s. (Figure 10) was apparently larger (by approx. 1270 mass units) than that estimated from the polypeptide sequence. The mass analysis also suggested heterogeneity in the carbohydrate, because several shoulders were detected in the deconvoluted spectrum.

In the sequence of bovine calreticulin, we found two potential glycosylation sites at positions 162 and 327 that meet the

Figure 9 SDS/PAGE of bovine brain calreticulin, bovine serum transferrin, and BSA before (a) and after (b) digestion with glycopeptidase F

Note that the mobilities of calreticulin and transferrin increased after the glycopeptidase digestion, whereas the mobility of the non-glycosylated protein, BSA, did not. The digestion was performed as described in the Experimental section at a protein/enzyme ratio of 1:225, and a portion of the digest was analysed on 10% polyacrylamide gel containing 0.1% SDS (Coomassie Blue staining).

Figure 10 Mass spectrum of bovine brain calreticulin

The spectrum deconvoluted from the data obtained by electrospray ionization m.s. is shown. The calreticulin preparation gave a broad spectrum with several shoulders, indicating the mass distribution presumably due to the carbohydrate heterogeneity. As the accurate estimate of the mass of each component was difficult because of the insufficient resolution, apparent mass values of the major peak and shoulder are given.

consensus sequence for N-linked oligosaccharide Asn-Xaa-Ser/Thr. However, Asn³²⁷ appeared not to be glycosylated, as the phenylthiohydantoin derivative of this residue was identified quantitatively by sequential Edman degradation. In contrast, Asn¹⁶² was not identified during sequential degradation, indicating that this residue might be the site of glycosylation. In fact, analysis of an acid hydrolysate (1 M HCl; 110 °C; 4 h) of peptide K16 containing Asn¹⁶² on a conventional amino acid analyser detected 2.2 mol of GlcNH₂/mol of peptide. Furthermore, analysis of neutral sugars by g.c./m.s. after trifluoroacetylation of the acid hydrolysate showed that mannose is the only neutral sugar component of the carbohydrate chain (results not shown).

Electrospray ionization m.s. of K16 demonstrated a series of (K16+3 H)³⁺ ions. In fact, the deconvoluted spectrum (Figure 11) showed that K16 is a heterogeneous mixture of six molecular

Figure 11 Mass spectrum of K16

The spectrum deconvoluted from the data obtained by electrospray ionization m.s. is shown.

species with masses 3622.5, 3784.1, 3946.4, 4108.3, 4270.5 and 4430.7. Because the polypeptide portion of K16 was proved to be homogeneous by sequential Edman degradation, and because the mass difference between each species (approx. 162) corresponded to one mannose (hexose) unit, we deduced that the observed heterogeneity was due to heterogeneity of the carbohydrate chain. From the mass of the polypeptide portion of K16 (2567.8), the mass of each carbohydrate chain was estimated as 1054.7, 1216.3, 1378.6, 1540.5, 1702.7 and 1862.9. These values, together with the carbohydrate analysis described above, suggest the carbohydrate composition, GlcNAc₂Man₄₋₉, with GlcNAc₂Man₅ being most abundant in our bovine preparation. On the basis of the core structure of a typical N-linked oligosaccharide of high-mannose type, one of the possible structures of the major carbohydrate chain is predicted in the schematic sequence of calreticulin shown in Figure 12. We assumed that there is no O-linked oligosaccharide attached to the polypeptide chain because all of the OH-containing amino acids were identified during the Edman degradation, and the observed mass difference between the intact and the polypeptide portion of calreticulin (about 1270) could be accounted for by the Asn¹⁶²-linked carbohydrates described above.

DISCUSSION

We report here the covalent structure of bovine brain calreticulin determined by analysis of the purified protein. Although the sequence analysis was hindered by the presence of a highly repetitive sequence of aspartic and glutamic acid at the C-terminus, we were able to obtain the complete sequence by m.s. analysis of the corresponding peptides derived from limited proteolysis of the protein. The final polypeptide sequence confirms the previously reported sequence of calreticulin deduced from cDNA, with small variations compatible with species-specific amino acid substitutions. Alignment of the bovine sequence with the available sequences of human (McCauliffe et al., 1990a), rabbit (Fliegel et al., 1989), rat (Murthy et al., 1990) and mouse (Smith and Koch, 1989) calreticulins and the partial sequences of the *Drosophila* (McCauliffe et al., 1990b) and *Onchocerca* (Unnasch et al., 1988) homologues indicates that the calreticulin structure has been highly conserved during phylogenetic evolution, supporting a previous notion that the protein might have a fundamental physiological function(s) in a wide variety of cells. We note that the C-terminal ER retention signal, KDEL, is found in the purified protein. This suggests that the

Figure 12 Linear model of the covalent structure of bovine brain calreticulin

The model predicts a four-domain structure of calreticulin, and incorporates the positions of cysteine (SH), cystine (S-S) and asparagine-linked high-mannose-type oligosaccharide in the sequence of 400 residues. a, a', a'', b, b' and b'' indicate the triple repeats of two distinct motifs of a proline-rich sequence originally noted by Smith and Koch (1989). The stippled box indicates the position of the putative nuclear localization signal. Polyacidic clusters upstream of the KDEL sequence are indicated by (-). Ca^{2+} illustrated below the sequence of domains II and IV shows that these domains represent the high-affinity and low-affinity Ca^{2+} -binding site respectively (Baksh and Michalak, 1991).

mature protein is associated with the KDEL receptor at the ER membrane. In fact, calreticulin is found mainly in ER, but a portion of the molecule is localized in the nucleus and nuclear envelope as demonstrated by an immunohistochemical study (Opas et al., 1991). Consistent with this observation is the presence of a putative nuclear localization signal PPKIKDPD at residues 187–195 (Fliegel et al., 1989; McCauliffe et al., 1990a). The trafficking mechanism by which the calreticulin molecule is targeted to the nuclear compartment is therefore of interest.

The present study confirms that bovine brain calreticulin is glycosylated by N-linked oligosaccharide of the high-mannose type, which is typical of a resident ER protein. We found that the carbohydrate heterogeneity is due to differences in mannose content from four to nine residues in series, which is produced by different extents of post-translational modification. The oligosaccharide moiety is attached to Asn¹⁶² which is located in the middle of the polypeptide chain (Figure 12). Interestingly, this residue is replaced by asparagine in the available sequences of calreticulins from other sources [aligned in Michalak et al. (1992)] despite the fact that long stretches of sequence surrounding this residue are identical. This would account for the fact that bovine liver calreticulin is glycosylated (Waisman et al., 1985; Lewis et al., 1985) whereas the human and rabbit proteins are not (Waisman et al., 1985; McCauliffe et al., 1990a; Milner et al., 1991). Previous studies (Van et al., 1989; Peter et al., 1992) have shown that rat liver calreticulin is also glycosylated, but by a complex-type oligosaccharide with a terminal galactose residue. Because the rat protein has asparagine at position 162, the only possibility remains that the other potential site for glycosylation, Asn³²⁷, is glycosylated in the rat liver protein, although this residue is not glycosylated in the bovine brain protein. Thus glycosylation of calreticulin appears to be a species-specific and possibly a tissue-specific event, and may not be involved directly in its function.

The three cysteine residues found in bovine brain calreticulin are all conserved among the calreticulin sequences determined to date. Analysis of the bovine protein indicates that Cys¹²⁰ and

Cys¹⁴⁶ are bridged by a disulphide bond, forming a small 25-residue loop in the molecule. Because disulphide, in general, contributes to a great extent to the correct folding and stability of the polypeptide chain, this must be taken into account in structure/function analysis of calreticulin produced by the recombinant technique. On the other hand, Cys⁸⁸ has a free thiol group; this residue appears to be buried inside the molecule as it is not reactive with DTNB on titration of calreticulin dissolved in an aqueous solution without chaotropic reagent.

Figure 12 illustrates a structural model of bovine brain calreticulin. As noted in previous studies (Fliegel et al., 1989; Smith and Koch, 1989), calreticulin is divided into several distinct domains. The C-terminal domain, termed domain IV in Figure 12, is characterized by clusters of acidic amino acids with the ER retention signal at the C-terminus. This domain is believed to be the site of interaction with the KDEL receptor (Lewis and Pelham, 1990), and represents the low-affinity high-capacity Ca^{2+} -binding site of calreticulin (Baksh and Michalak, 1991). Separated from this extremely acidic domain by a moderately hydrophobic segment are internal triple repeat sequences of two distinct motifs, PPKIXDPDAXKPEDWDX and GEWXPPIX-DNPEYK (Smith and Koch, 1989), which represent the high-affinity Ca^{2+} -binding site of calreticulin (Baksh and Michalak, 1991). Domain I, the N-terminal domain comprising approximately half of the entire polypeptide chain, contains the single disulphide bridge and the carbohydrate as demonstrated by the present study. Although domain I appears to be an independent structural unit separated from other portions of the molecule (Smith and Koch, 1989), no activity has been found for it.

A computer-assisted homology search of the current protein sequence databases (PIR and SwissProt databases) suggests that the calreticulin sequence is unique, but shows significant homology to the membrane protein in ER, termed calnexin, which has a putative function in the translocation of nascent polypeptides across the ER membrane (Wada et al., 1991). As previously mentioned by Wada et al. (1991), the homology is extensive towards the C-terminal half of the molecule including domains II, III and IV, and a C-terminal portion of domain I.

Besides the homology to calnexin, recent studies indicate that the calreticulin gene might be co-transcribed with other ER proteins such as protein disulphide isomerase and immunoglobulin heavy-chain binding protein (BiP) (McCauliffe et al., 1992) and that calreticulin forms a tight complex with flavin-containing monooxygenase and a number of synthetic peptides (Guan et al., 1991; Rojiani et al., 1991). Thus the accumulating evidence suggests a role for calreticulin in protein transport, folding and assembly, and the availability of the covalent structure will aid future analyses of this possibility.

REFERENCES

- Baksh, S. and Michalak, M. (1991) *J. Biol. Chem.* **266**, 21458–21465
- Burns, K., Helgason, C. D., Bleackley, R. C. and Michalak, M. (1992) *J. Biol. Chem.* **267**, 19039–19042
- Cunico, R. L. and Schlanbach, T. (1983) *J. Chromatogr.* **266**, 461–470
- Ellman, G. L. (1959) *Arch. Biochem. Biophys.* **82**, 70–77
- Finlay, B. B., Gray, V. and Dedhar, S. (1991) *Biochemistry* **30**, 9859–9866
- Fliegel, L., Burns, K., MacLennan, D. H., Reithmeier, R. A. F. and Michalak, M. (1989) *J. Biol. Chem.* **264**, 21522–21528
- Guan, S., Falick, A. M., Williams, D. E. and Cashman, J. R. (1991) *Biochemistry* **30**, 9892–9900
- Ishioka, N., Isobe, T., Kadoya, T., Okuyama, T. and Nakajima, T. (1984) *J. Biochem. (Tokyo)* **95**, 611–617
- Lewis, M. J. and Pelham, H. R. B. (1990) *Nature (London)* **348**, 162–163
- Lewis, M. J., Mazzarella, R. A. and Green, M. (1985) *J. Biol. Chem.* **260**, 3050–3057
- Manabe, T., Hayama, E. and Okuyama, T. (1982) *Clin. Chem.* **28**, 824–827
- Matsui, Y., Kikuchi, A., Araki, S., Hata, Y., Kondo, J., Teranishi, Y. and Takai, Y. (1990) *Mol. Cell Biol.* **10**, 4116–4122
- McCauliffe, D. P., Lux, F. A., Lieu, T. S., Sanz, I., Hanke, J., Newkirk, M. M., Bachinski, L. L., Ito, Y., Siciliano, M., Reichlin, M., Sontheimer, R. D. and Capra, J. D. (1990a) *J. Clin. Invest.* **85**, 1379–1391
- McCauliffe, D. P., Zappi, E., Lieu, T. S., Michalak, M., Sontheimer, R. D. and Capra, J. D. (1990b) *J. Clin. Invest.* **85**, 1379–1391
- McCauliffe, D. P., Yang, Y. S., Sontheimer, R. D. and Capra, J. D. (1992) *J. Biol. Chem.* **267**, 2557–2562
- Michalak, M., Baksh, S. and Opas, M. (1991) *Exp. Cell Res.* **179**, 91–99
- Michalak, M., Milner, R. E., Burns, K. and Opas, M. (1992) *Biochem. J.* **285**, 681–692
- Milner, R. E., Baksh, S., Shemanko, C., Carpenter, M. R., Smillie, L., Vance, J. E., Opas, M. and Michalak, M. (1991) *J. Biol. Chem.* **266**, 7155–7165
- Munro, S. and Pelham, H. R. B. (1986) *Cell* **46**, 291–300
- Munro, S. and Pelham, H. R. B. (1987) *Cell* **48**, 899–907
- Murthy, K. K., Banville, D., Srikant, C. B., Carrie, F., Holmes, C., Bell, A. and Patel, Y. C. (1990) *Nucleic Acids Res.* **18**, 4933
- Opas, M., Dziak, E., Fliegel, L. and Michalak, M. (1991) *J. Cell Physiol.* **149**, 160–171
- Ostwald, T. J. and MacLennan, D. H. (1974) *J. Biol. Chem.* **249**, 974–979
- Peter, F., Van, P. N. and Soling, H.-D. (1992) *J. Biol. Chem.* **267**, 10631–10637
- Rojiani, M. V., Finlay, B. B., Gray, V. and Dedhar, S. (1991) *Biochemistry* **30**, 9859–9866
- Smith, M. J. and Koch, G. L. E. (1989) *EMBO J.* **8**, 3581–3586
- Unnasch, T. R., Gallin, M. Y., Sobosiay, P. T., Ertmann, K. D. and Greene, B. M. (1988) *J. Clin. Invest.* **82**, 262–269
- Van, P. N., Peter, F. and Soling, H.-D. (1989) *J. Biol. Chem.* **264**, 17494–17501
- Wada, I., Rindress, D., Cameron, P. H., Ou, W.-J., Doherty, II, J. J., Louvard, D., Bell, A. W., Dignard, D. Y. and Bergeron, J. J. M. (1991) *J. Biol. Chem.* **266**, 19599–19610
- Waisman, D., Salimath, B. P. and Anderson, M. I. (1985) *J. Biol. Chem.* **260**, 1654–1660