Expression, purification and characterization of the structure and disulfide linkages of insulin-like growth factor binding protein-4

D Chelius, M A Baldwin1, X Lu and E M Spencer

Department of Growth and Development, California Pacific Medical Center Research Institute, San Francisco, California 94114, USA
1Mass Spectrometry Facility, Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143–0446, USA

Abstract

Insulin-like growth factor binding protein-4 (IGFBP-4), like the other five IGFBPs, is a critical regulator of the activity of insulin-like growth factor (IGF)-I and IGF-II. However IGFBP-4 seems to be the only IGFBP with no potential to enhance the mitogenic actions of the IGFs. IGFBP-1 to -3 and -5 each contain 18 conserved cysteine residues, IGFBP-6 lacks two of the twelve N-terminal cysteines, while IGFBP-4 has two additional cysteines in the central region. A plasmid was constructed to express rat IGFBP-4 as a thioredoxin fusion protein that included a hexahistidine sequence to permit affinity purification. The fusion protein was expressed in E.coli, purified using nickel-chelate affinity chromatography and cleaved by tobacco etch virus (TEV) protease to produce mature rat IGFBP-4 with an additional glycine residue at the N-terminus. Final purification was achieved by further nickel affinity chromatography and reverse phase HPLC. The isoelectric points of the recombinant IGFBP-4 were the same as those of the non-glycosylated isoforms of IGFBP-4 in rat serum. The binding affinities of the recombinant protein and IGFBP-4 secreted by rat cells to IGF-I were compared using a newly developed binding assay. No significant difference could be detected, consistent with proper folding of the recombinant protein. This indicates that glycosylation of IGFBP-4 does not affect its binding to IGF-I. Using mass spectrometry and tandem mass spectrometry no differences between authentic and recombinant IGFBP-4 could be detected. Eight of the ten disulfide linkages have been determined, including linkages of conserved cysteine residues not previously identified in other IGFBPs. Numbering the cysteine residues sequentially from the N-terminus only the disulfide connectivity of C1, C2, C5 and C6 could not be determined. However, C1 is not linked to C2 and C5 is not linked to C6. The established linkages were C3 to C8, C4 to C7, C9 to C11 and C10 to C12. The two cysteines in the non-conserved mid-region unique to IGFBP-4 (C13 and C14) are linked together. Linkage of the C-terminal cysteine residues is identical to that of IGFBP-2, -5 and -6 (C15 to C16, C17 to C18 and C19 to C20). The central flexible core of IGFBP-4, containing two additional cysteines may contribute to its unique biological action.

Introduction

The insulin-like growth factors (IGFs) modulate an extraordinarily wide variety of biological actions in addition to mediating the growth promoting effect of growth hormone (Jones & Clemmons 1995). IGFs stimulate proliferation and differentiation of numerous cell types and promote anabolic functions in numerous tissues, such as increased production of protein, RNA and DNA (Sara & Hall 1990, Guidice 1992, Cohick & Clemmons 1993). They also play important roles in the prevention of apoptosis (Williams et al. 1990), in the regulation of pituitary growth hormone (GH) secretion and in facilitating the actions of other hormones and factors (Tannenbaum et al. 1983). The actions of IGFs are modulated by IGF-binding proteins (IGFBPs), a family of six structurally related proteins showing high sequence homology in the N- and C-terminal thirds of the molecule (Jones & Clemmons 1995). The central regions of these proteins have no homology. The core molecular weights of the IGFBPs range from 22.8 kDa for IGFBP-6 up to 31.3 kDa for IGFBP-2. Whereas IGFBP-1 and -2 are not glycosylated, IGFBP-3 and -4 are N-glycosylated and IGFBP-5 and -6 are O-glycosylated (Neumann et al. 1998). IGFBP-4 is generally considered to be the only one of these binding proteins not to potentiate the actions of IGF. The inhibitory effects of IGFBP-4 have been shown under a wide variety of experimental conditions, including IGF-I and IGF-II-stimulated DNA synthesis.
(Mohan et al. 1989, Rechler 1993). The liver is the primary source if IGFBP-4 in the rat, but it is also expressed in other tissues such as smooth and striated muscle, lung, liver and to a lesser degree by the heart. IGFBP-4 is one of the most abundant IGFBPs synthesized by osteoblasts, acting in a paracrine manner to inhibit IGF-mediated osteoblast proliferation (Mohan et al. 1989). It is regulated by several key osteotrophic agents, such as parathyroid hormone, 1,25-dihydroxyvitamin D3, and bone morphogenic protein-7 (LaTour et al. 1990, Rosen et al. 1994).

Mohan et al. (1989) and Culouscou & Shoyab (1991) first isolated IGFBP-4 on the basis of its activity to inhibit cell growth. The cloning and isolation of both rat and human IGFBP-4 cDNAs was carried out by Shimasaki et al. (1990), the cDNAs showing very high homology to each other. IGFBP-4 is expressed as a precursor protein. Removal of the propeptide sequence from rat IGFBP-4 yields the mature 233-residue protein (237 residues for human). Both rat and human IGFBP-4 have a single asparagine-linked glycosylation site and can be isolated from various cells in both the glycosylated (28 kDa) and non-glycosylated (25.7 kDa) form, that typically migrates at 24 kDa on SDS-PAGE.

Most IGFBPs contain eighteen highly conserved cysteines, IGFBP-6 lacks two cysteines whereas IGFBP-4 has an additional two cysteines in the central region (Shimasaki et al. 1991a). The linkage of these cysteines plays an important role in the correct folding of the proteins and consequently affects their biological activity. A complete disulfide linkage map has been reported only for IGFBP-6 (Neumann & Bach 1999), the protein with the lowest number of cysteines. The linkage of the six C-terminal cysteines of IGFBP-2 and -6 are identical (Forbes et al. 1998, Neumann & Bach 1999), underscoring the importance of these in producing an active IGF binding domain. It has been demonstrated that the N-terminus of IGFBP-6 differs from IGFBP-1, suggesting that this part of the protein influences the different preference of IGFBPs for the IGFs (Neumann & Bach 1999). IGFBP-6, like IGFBP-2 and -5 has a higher affinity for IGF-II whereas IGFBP-1, like the other two IGFBPs, shows no such preference between IGF-I and -II (Jones & Clemmons 1995).

In this study a new expression and purification system for rat IGFBP-4 was used to resolve all but two disulfide linkages, using reversed phase HPLC of proteolytic digests, electrospray ionization mass spectrometry and tandem mass spectrometry (Smith et al. 1990). No difference in the disulfide linkages between recombinant and authentic IGFBP-4 (Shimasaki et al. 1991b) could be detected, indicating the correct folding of our expressed IGFBP-4. This work confirms previously reported homologies and presents new information on disulfide pairings in the N-terminal and mid-region.

**Materials and Methods**

**Construction of the pIGFBP-4 expression plasmid**

cDNA was obtained by PCR using primer 1 TTC GAAACCTTGTATTCACCGGCAAGCCA TCCACTG (5′ primer) and primer 2 GCTCAGCTCA CTCTTGGAGCTGTACG (3′ primer) (nucleotides encoding IGFBP-4 are shown in bold). RNA extracted from rat liver tissue with Trizol reagent (Gibco) was used as a template. Additionally the codons of a unique protease cleavage site were incorporated before the IGFBP-4 cDNA encoding for the TEV protease recognition site (ENLYFQG) (nucleotides encoding for the protease recognition site are underlined). The PCR product was cloned into pCR-II vector (Novagen) and positive plasmids were identified by restriction enzyme mapping and DNA sequencing. A positive plasmid was digested with NspV and BamHI and ligated into the multiple cloning site of pET31a vector (Novagen) downstream of the thioredoxin and hexahistidine sequence, yielding the 6438 bp IGFBP-4 thioredoxin fusion protein expression plasmid pIGFBP-4.

**Expression and purification of IGFBP-4**

*E. coli* BL21 cells (Promega) were used to express the IGFBP-4 fusion protein from pIGFBP-4. Cultures (4 x 300 ml) were grown at 37 °C on a shaking incubator in LB medium (1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl) supplemented with ampicillin, 50 µg/ml. When cultures reached an optical density of 1.0 at 560 nm, isopropyl thiogalactoside (IPTG) was added to a final concentration of 1 mM, and incubation was continued for a further 5 h. Cells were harvested by centrifugation (8670 × g, 10 min, 4 °C), and suspended and resuspended in 60 ml of buffer A (50 mM Tris; pH 8.0, 10 mM MgCl2, 10 mM imidazole and 0.1% Triton X-100). The suspension was frozen in acetone/dry ice and thawed in warm water. Lysozyme (1 mg) was added and the freeze-thaw procedure was repeated twice. The resulting mixture was treated with RNaseI and DNaseI (final concentration of 1 µg/ml each) for 10 min at 4 °C, then the cells were disrupted by sonication. Insoluble fragments were removed by centrifugation at 48 400 × g for 20 min at 4 °C, and phenylmethylsulfonic fluoride (PMSF) was added to the supernatant to a final concentration of 1 mM. The supernatant was than loaded onto a 5 ml Ni column (Promega). The column was washed with buffer A containing 30 mM imidazole to remove weakly bound material, then bound proteins were eluted with 500 mM imidazole in buffer A. Fractions containing protein were combined (total volume 20 ml) and immediately dialyzed against 2 l buffer B (50 mM Tris, pH 8.0) for 1-5 h. The fusion protein was cleaved by overnight incubation with 2000 U TEV protease (Gibco) at room temperature. The
solution was dialyzed against 21 buffer B for 1 h before loading onto the 5 ml Ni column. IGFBP-4 was eluted with 40 mM imidazole in buffer B. Fractions containing IGFBP-4 were further purified by reversed phase HPLC. Samples were applied to a C-4 column (Vydac) and the bound material was eluted with a linear gradient of 20–40% acetonitrile over 100 min at a flow rate of 4 ml/min. All HPLC solutions contained 0.1% trifluoroacetic acid (TFA). Fractions containing IGFBP-4 were frozen in acetone/dry ice and lyophilized. Purified IGFBP-4 could be stored in a desiccator at −20 °C for several months.

**Preparation of anti-rat IGFBP-4**

Approximately 0.7 ml immunogen was prepared by adding 0.35 ml (0.5 mg) rIGFBP-4 to 0.35 ml of the Freund’s complete adjuvant. The suspension was homogenized for 3 min giving a milky solution, then injected subcutaneously into the backs of 2-3 kg New Zealand White rabbits. Animals were boosted after 4 weeks and sera were tested for antibodies 2 weeks later by EIA and RIA (Chelius & Spencer 1998).

**Preparation of [125I]IGFBP-4**

Six micrograms of purified recombinant IGFBP-4 was iodinated using the chloramine T method (Spencer et al. 1993). [125I]IGFBP-4 was separated from free [125I] using a G-75 Sephadex (Pharmacia) column. The specific activity of the radiolabeled IGFBP-4 varied between 100 and 200 μCi/μg protein. Aliquots of radiolabeled IGFBP-4 were stored at −20 °C and could be used for up to 2 months for RIA.

**IGF-I binding assay for IGFBP-4**

To compare to the affinity of recombinant IGFBP-4 and authentic IGFBP-4 to IGF-I, a new assay was developed. Purified IGFBP-4 (2.5 ng) was incubated with [125I]-IGF-I (50 000–70 000 c.p.m.) in the presence or absence of unlabeled IGF-I (200–0.78 ng) in a total volume of 500 μl assay buffer (50 mM Tris, 50 mM NaCl, 0.05% TWEEN-20, 0.1% EDTA, 1% BSA; pH 8.0). After incubation at room temperature for 4 h, 200 μl primary IGFBP-4 antibody (0.2 ml, 1:4000 dilution) was added and samples were incubated at 4 °C for 16–24 h. Bound IGF-I:IGFBP-4 complex was separated from free [125I]IGF-I by immunoprecipitation, using 0.1 ml gamma globulin (Antibody Inc., Davis) (1:200 dilution), 0.1 ml goat anti-rabbit antibody (Antibody Inc., Davis) and 0.7 ml PEG-8000 (9-14%) solution. Samples were incubated for 15 min at 4 °C then centrifuged for 30 min at 3000 × g and 4 °C. The immunoprecipitants were counted in a Berthold 12-well gamma counter. A binding curve was calculated using a one-site competition model (GraphPad). The maximal [125I]IGF-I tracer binding (as [125I]IGF-I: IGFBP-4 complex) to antiserum under the assay conditions described in this study was 50%. [125I]IGFBP-4 tracer binding to antiserum under the same conditions was also 50%, suggesting that essentially all [125I]IGF formed a complex with IGFBP-4 and could be precipitated using IGFBP-4 antiserum. The non-specific binding of radiolabeled IGF-I measured in the absence of IGFBP-4 was less than 1%. Addition of increasing concentrations of purified ‘cold’ IGF-I inhibited the binding of [125I]IGF-I to IGFBP-4 in a dose-dependent manner.

**2-D Western ligand blot analysis of IGFBP-4**

The method for two-dimensional PAGE was described by Weber et al. (1999) to separate IGFBP-4 under non-reducing conditions. Immobilized pH gradient gels with a pH range of 4–10 were prepared as previously described for isoelectric focusing in the first dimension. This was carried out at 200 V for 30 min (20 °C), followed by 6 h at 3000 V in a cooled horizontal electrophoresis unit (Multiphor II, Pharmacia). For separation in the second dimension (SDS-PAGE), the immobilized pH gradient gel strips were placed on a horizontal SDS-PAGE gradient gel (10–15%). After protein transfer (45 min, 200 V, 10 °C) the gel strips were removed and electrophoresis was continued at 800 V. The IGFBP-4 isoforms were identified according to Weber et al. (1999) by Western ligand blot analysis using [125I]IGF-II as a ligand.

**Digestion of IGFBP-4 with proteases and separation of fragments**

IGFBP-4 (40 μg) was digested with 1 μg l-1-tosylamido-2-phenylethylchloromethyl ketone–treated trypsin (Promega) in 1 ml buffer (50 mM Tris; pH 8.0), 1 mM CaCl₂ for 3 h at 37 °C. Fragments were separated by reverse phase HPLC on a C-4 column (Vydac), and eluted with a gradient of 10–50% acetonitrile/0.1% TFA at a flow rate of 1 ml/min over 100 min. Detection was by UV absorbance at 215 nm. Fractions were collected manually and vacuum dried (SpeedVac). HPLC fractions containing disulfide-linked peptides were identified by electrospray mass spectrometry (ESI-MS) as described below. For further digestion fractions were resuspended with 1 μg Staphylococcus aureus V8 protease (endoprotease Glu-C, Promega), thermolysin (Boehringer-Mannheim) or papain (Boehringer-Mannheim) in 200 μl buffer (50 mM NH₄HCO₃; pH 7.8) or with 1 μg Nα-p-tosyl-l-lysine chloromethyl ketone–treated chymotrypsin ( Worthington) in 200 μl buffer (50 mM Tris; pH 8.1, 1 mM CaCl₂) or with pepsin (Boehringer-Mannheim) in 200 μl 0.001 N HCl and incubated for 3 h at 37 °C.

**Electrospray mass spectrometry (ESI-MS)**

Electrospray mass spectrometry (ESI-MS) was carried out using an orthogonal acceleration time-of-flight mass...
spectrometer (oa-ToFMS) (PE Biosystems, Framingham, MA, USA) equipped with an ESI source and an ion mirror (reflector). The mass scale was calibrated using the doubly and singly charged ions of Gramicidin S (Sigma). Most analyses were carried out on dried HPLC fractions collected offline. These were taken up in 1:1 acetonitrile/water containing 1% acetic acid and injected into the ESI source via a syringe pump (Harvard Apparatus, Holliston, MA, USA) at 1 µl/min and spectra were accumulated for 1–3 min. Each molecular species gave one or more peak clusters arising from different charge states. The spectra were analyzed, either manually by determination of the charge states of individual ions from the peak spacing, or by deconvolution using the software provided by the mass spectrometer manufacturer. For the largest proteolytic fragments, the isotopic components at adjacent masses were not fully resolved by the oa-ToFMS. In such cases the mass of each peak representing an average of the distribution of all isotopes present was compared with values calculated from average atomic masses. For lighter ions that were clearly resolved, the masses recorded were those of the first isotope peak containing $^{12}$C, $^{14}$N, etc., corresponding to the monoisotopic molecular mass. All masses presented here are monoisotopic unless stated otherwise.

**Tandem mass spectrometry**

This was carried out using a PE Sciex QSTAR mass spectrometer (Concorde, Ontario, Canada) of QqoaToF geometry, which combined quadrupole and oa-ToF analyzers. The mass scale was calibrated with Glu-Fibrinogen peptide (Sigma). Samples as described above were continuously injected at a flow rate of ∼50 nL/min into a home-made coaxial micro-ESI device, using a sheath flow of 75% water, 18% ethanol, 7% n-propanol, 0.1% formic acid at 100 nL/min. The ions selected for collision-induced dissociation (CID) and tandem mass spectrometry (MS) were selectively transmitted through the quadrupole mass analyzer and subjected to collisions with N$_2$ gas in the collision cell. The extent of fragmentation was optimized by adjustment of the gas pressure and collision energy and the resulting fragment ions were separated in the qa-ToF. All spectra were analyzed manually using similar methods to those described above. The high-resolution capability of this instrument allowed all isotopes to be separated, thus the charge states could be determined from the peak spacing and all ion masses were recorded as monoisotopic values.

**Results**

*Cloning, expression and purification of IGFBP-4*

The expression plasmid pIGFBP-4 was constructed to express a novel fusion protein consisting of the mature form of rat IGFBP-4 fused at its N-terminus with thiorredoxin, a hexahistidine tag and a TEV protease cleavage site. DNA sequencing confirmed the published sequence of rat IGFBP-4 (Shimasaki et al. 1990).

The IGFBP-4 fusion protein could be expressed to about 2-5% of the total protein in the lysate after IPTG induction. Following purification using a nickel-chelating affinity chromatography column, the fusion protein was cleaved, yielding the mature IGFBP-4 with an additional glycyl residue at the N-terminus. The cleavage mixture containing undigested fusion protein, the fusion peptide and TEV protease were separated from IGFBP-4 by a second nickel-chelating chromatography step. Final purification was achieved by reverse phase HPLC. Fractions from each stage of purification were analyzed by SDS-PAGE (Fig. 1). In a typical purification 4–5 mg purified IGFBP-4 could be isolated from 1·21 mg.

**IGF-I binding assay for IGFBP-4**

To determine the binding affinity of recombinant IGFBP-4 from *E. coli* and authentic IGFBP-4 isolated from rat serum (Shimasaki et al. 1991b), competitive binding experiments with [125I]IGF-I were performed in the presence of various concentrations of unlabeled IGF-I. Recombinant and authentic IGFBP-4 were equipotent in their affinities for IGF-I (Fig. 2). The IC$_{50}$ value for recombinant IGFBP-4 was $2\cdot4 \times 10^{-10}$ M and the IC$_{50}$ value for IGFBP-4 from rat serum was $2\cdot5 \times 10^{-10}$ M. These values compare very well with
2-D Western ligand blot analysis of IGFBP-4

Recombinant IGFBP-4 from E. coli, IGFBP-4 isolated from rat serum and rat serum were also compared by two dimensional gel electrophoresis to ensure proper folding (Fig. 3). As expected the 28 kDa glycosylated forms of IGFBP-4 were not identified (Fig. 3). As expected the 28 kDa glycosylated forms of IGFBP-4 were not identified (Fig. 3). However the 25·7 kDa forms exhibited identical isoelectric and molecular weight migration patterns when compared with IGFBP-4 from rat serum. The isoelectric points of the four isoforms were at 5·8, 6·2, 6·3 and 6·5. The isoform with an isoelectric point of 6·3 was dominant in both sources of IGFBP-4, and the relative amounts of each isoform were also very similar.

Digression of IGFBP-4 with proteases and separation of fragments and ESI-MS

Trypsin digestion was used to generate disulfide-linked peptides from both recombinant IGFBP-4 from E. coli and authentic rat IGFBP-4. The predicted tryptic peptides are illustrated in Fig. 4; the numbering of which is based on the assumption that the arginine—proline and lysine—proline bonds would not be digested. The fragments were fractionated by C-4 reverse phase HPLC using a gradient of acetonitrile in 0·1% TFA. The UV chromatogram of recombinant IGFBP-4 (Fig. 5) was virtually identical to that from authentic IGFBP-4 (Shimasaki et al. 1991b), and ESI-MS analysis of the collected fractions confirmed this (data not shown). None of the short peptides found in the HPLC flow-through contained cysteine residues, and in general the smallest predicted peptides of one to four amino acids were not detected. However, fragments containing all 20 cysteine residues were identified in the HPLC fractions by their unique molecular masses, and the detected peptides accounted for more than 80% of the total amino acids of IGFBP-4. The observed and calculated molecular weights for the various cysteine-containing peptides are listed in Table 1, and are discussed in more detail below.

Fraction 1 gave a mass spectrum that showed three related species of masses 3584, 3642 and 3770, the first of which was calculated to correspond to the combination of peptides 22 (C17), 24 (C18, C19) and 28 (C20) linked together by disulfide bonds. The higher mass components resulted from incomplete digestion of peptides 23 (GK) and 25 (K). Clearly C17 and C20 are linked to C18 and C19, but the order could not be distinguished directly. However, C18 and C19 within peptide 24 are connected via a tryptophan residue, which should be susceptible to digestion with chymotrypsin. It was calculated that the combination of peptide 20 plus C18W and peptide 28 plus C19DR should give two species of molecular weights 1613·8 and 2046·0, whereas the alternative pairing should give 1612·8 and 2047·0 respectively. After digestion with chymotrypsin the mixture was lyophilized to remove the ammonium bicarbonate buffer then analyzed directly by ESI-MS, to give the spectrum shown in Fig. 6. Analysis of the spectrum revealed two major species of molecular weights 1613·9 and 2046·0, labeled as A and B respectively. This identified the linkage as C17 to C18 and C19 to C20.

Fractions 2 and 3 were poorly resolved and both showed multiple species. A typical mass spectrum is shown in Fig. 7 for fraction 2. Deconvolution of the individual peptide signals identified a lower mass component of monoisotopic mass 3727·4, designated C, and two higher mass components, jointly designated as D. The mass of C is equivalent to peptide 17 (C15) linked via a disulfide to peptide 20 (C16). Fraction 3 showed a higher proportion of this species, plus an additional component of molecular weight 4523 Da that closely matches the same combination plus peptide 21 undigested by trypsin. These two observations confirm a disulfide linkage between C15 and C16. The higher mass species D are discussed below.

Fraction 4 showed a 6217 Da species, calculated to correspond to the linkage of peptides 5 (C10), 8 (C13), 9 (C15) and 10 (C14), with either peptides 8 and 9, or 9 and 10 not separated by the tryptic digestion. This combination containing four cysteine residues was further digested with endoproteinase Glu-C (V8 protease). Two major components were identified by ESI-MS (data not shown); a species of molecular weight 1781·8 corresponding to peptide 5 linked via a disulfide to the N-terminus of peptide 8 (TLMHGQGVCS12TE), and a second species of 2436·2 corresponding to peptide 9 without the N-terminal DE, (SEHPNSSKENPC12SAHDR) linked via. The isoelectric points of the four isoforms were at 5·8, 6·2, 6·3 and 6·5. The isoform with an isoelectric point of 6·3 was dominant in both sources of IGFBP-4, and the relative amounts of each isoform were also very similar.

Figure 2 Competitive protein binding assay between [125I]IGF-I and unlabeled IGF-I for binding to IGFBP-4. Recombinant IGFBP-4 (■) or authentic IGFBP-4 (Shimasaki et al. 1991b) (▲) were incubated with [125I]IGF-I and increasing concentrations of unlabeled IGF-I for 4 h.
a disulfide to peptide 10. Therefore C\textsuperscript{10} is linked to C\textsuperscript{12} and C\textsuperscript{13} is linked to C\textsuperscript{14}.

Thus 10 cysteine residues from the C-terminal portion of IGFBP-4 had been accounted for. It was probable that the other 10 cysteines were contained within the higher mass components D from fraction 2 (Fig. 7). This showed two separate molecular species with average molecular masses of 5837.5 and 5855.5. The latter, which revealed sufficient isotopic resolution to suggest a monoisotopic mass of 5851.7, corresponds to a combination of peptides 1 (C\textsuperscript{1}, C\textsuperscript{2}), 3 (C\textsuperscript{3}, C\textsuperscript{4}), 4 (C\textsuperscript{5}, C\textsuperscript{6}, C\textsuperscript{7}, C\textsuperscript{8}, C\textsuperscript{9}) and 6 (C\textsuperscript{11}) linked through five disulfide bonds. An equivalent species of mass 5796 was observed in the digest of authentic rat IGFBP-4, corresponding to the same combination without the additional glycine at the N-terminus of recombinant IGFBP-4. Minor peaks at 5837.5 (recombinant IGFBP-4) and 5784 (for IGFBP-4 from rat cells) suggested the same species but with one H\textsubscript{2}O less, consistent with the same peptide combination as above, but with peptides 3 and 4 not separated by tryptic digestion. The existence of this complex assembly of peptides containing the N-terminal cysteine residues shows that C\textsuperscript{1} cannot be linked to C\textsuperscript{2} and C\textsuperscript{3} cannot be linked to C\textsuperscript{4}. Similarly C\textsuperscript{1} and C\textsuperscript{2} cannot be linked to C\textsuperscript{3} and C\textsuperscript{4} by two disulfide bonds. Thus, at least one cysteine from peptide 1 and one from peptide 3 must be linked to cysteine residues within peptide 4.

**Tandem mass spectrometry**

Tandem mass spectrometry with CID was carried out on this complex of peptides, selected in the quadrupole section of the Qq\textsubscript{oa}TOF tandem mass spectrometer as the 7+ charge state (m/z 837.5), shown in Fig. 8. No fragmentation was identified that involved cleavage within the cysteine-linked core, identified in Fig. 9 as a hexagon. Series of singly and doubly charged fragment ions corresponding to C-terminal fragments of peptides 1, 4 and 6, denoted as 1-, 4- and 6\textsubscript{y\textsubscript{n}} ions respectively, were identified. Some b ions with charge retention on the N-terminal fragment were also detected, but both peptides 4 and 6 terminate in arginine, favoring the retention of positive charges at their C-termini. The upper panel of Fig. 8 highlights a series of doubly charged ions and clearly shows that peptides 4 and 6 are linked together through C\textsuperscript{9}.
and C11. In other CID experiments carried out at lower collision energies, ions of higher charge states corresponding to losses of residues from the N-termini of peptides 1 and 4 were observed, including successive y ions corresponding to losses of the amino acids GDEAIH (data not shown). However, there were no other fragments that revealed further details of the cysteine linkages.

Tandem mass spectrometry was unaffected by the poor separation of this HPLC fraction as the precursor ion mass was selected with high precision in the quadrupole section of the tandem mass spectrometer. However, for further proteolysis it was necessary to repurify the components using a slower HPLC gradient. Aliquots were then sub-digested with either thermolysin, papain or pepsin, and analyzed by ESI-MS. Each of these enzymes gave rise to closely related smaller species that were lacking the C-terminus of peptide 4 linked to peptide 6, which confirmed the linkage of C9 and C11. For example, papain gave a single species of molecular weight 4045, equivalent to the combination of peptides 1, 3 and the N-terminus of 4 (EPGC5GC6C7ATC8ALG). Thermolysin gave three separate species of molecular masses 3673, 3418 and 3305. The largest of these is the same as the species from digestion with papain except for the removal of GDEA from the N-terminus of peptide 1, and the middle and lowest mass species lack VR and LVR respectively from the C-terminus of peptide 3. The major species from pepsin digestion of mass 3348 was the same as

![Figure 4](image-url)

**Figure 4** Predicted peptides from trypsin digestion of rat IGFBP-4, based on the numbering of the wild type protein (18). The additional N-terminal glycine is present only in the recombinant protein, for which it is numbered as residue 0. Peptide numbering assumes that Arg-Pro (residues 18/19 and 142/143) and Lys-Pro (67/68) bonds are not cleaved by trypsin. Sequential peptide numbers 1–28 and superscripted cysteine numbers 1–20 identify the peptides and cysteine residues in the text. Residue numbers are given for those peptides that contain cysteine residues.

![Figure 5](image-url)

**Figure 5** Reverse phase HPLC of a tryptic digest of IGFBP-4. Recombinant IGFBP-4 (40 μg) was digested with trypsin and applied to a C-4 reverse-phase HPLC column using a gradient of 10–50% acetonitrile, 0.1% trifluoroacetic acid over 100 min. Fractions were collected manually and analysed by ESI-MS. Selected fractions were further digested with chymotrypsin, S. aureus V8 protease (endoproteinase Glu-C), papain or pepsin.
Table 1  ESI-MS analysis of IGFBP-4 proteolytic fragments in the major HPLC fractions from tryptic digestion of recombinant IGFBP-4. Observed masses are compared with calculated (Calc.) masses based on the peptide sequences shown in Fig. 4. In all cases the same peptides were observed for authentic rat IGFBP-4 (Shimasaki et al. 1991b) except that N-terminal peptides in fraction 2 lacked one glycine residue. In most instances the isotopes were resolved and values given are monoisotopic; if the isotopes were unresolved the masses are identified as average (Av). Although ions were observed in a variety of multiple charge states, all measured \(m/z\) values are given as molecular masses of neutrals.

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Peptides</th>
<th>Cysteines</th>
<th>Obs. Mass</th>
<th>Calc. Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22, 24–25, 28</td>
<td>17, 18–19, 20</td>
<td>3584·5</td>
<td>3584·6</td>
</tr>
<tr>
<td></td>
<td>22–23, 24, 28</td>
<td>17, 18–19, 20</td>
<td>3641·5</td>
<td>3641·6</td>
</tr>
<tr>
<td></td>
<td>22–23, 24–25, 28</td>
<td>17, 18–19, 20</td>
<td>3769·0</td>
<td>3769·7</td>
</tr>
<tr>
<td>2</td>
<td>1, 3–4, 6</td>
<td>1–2, 3–4, 5–6–7–8–9, 11</td>
<td>5837·5 (Av)</td>
<td>5738·8 (Av)</td>
</tr>
<tr>
<td></td>
<td>1, 3, 4, 6</td>
<td>1–2, 3–4, 5–6–7–8–9, 11</td>
<td>5851·7</td>
<td>5852·5</td>
</tr>
<tr>
<td>3</td>
<td>17, 20</td>
<td>15, 16</td>
<td>3727·4</td>
<td>3727·8</td>
</tr>
<tr>
<td></td>
<td>17, 20–21</td>
<td>15, 16</td>
<td>4522·0</td>
<td>4522·1</td>
</tr>
<tr>
<td>4</td>
<td>5, 8–9, 10 or 5, 8, 9–10</td>
<td>10, 12–13, 14 or 10, 12, 13–14</td>
<td>6219·0 (Av)</td>
<td>6219·7 (Av)</td>
</tr>
</tbody>
</table>

Figure 6  ESI-MS of the peptide mixture derived from digestion of fraction 1 with chymotrypsin. The mass spectrum shows ion series for two separate molecular species A and B, each with two to four charges. These were deconvoluted separately to give well-resolved isotope patterns, from which the monoisotopic molecular weights were obtained.
the 3305·3 component from thermolysin but without the glycine residue at the C-terminus of truncated peptide 4. As before, tandem mass spectra showed predominant y ion series, confirming sequences such as EELVR for species in which peptide 3 had not been truncated by digestion (data not shown). A unique doubly charged ion of m/z 996·0 was derived from the species of molecular weights 4045·7 and 3673·5. This corresponds to a neutral fragment of molecular weight 1990·0, which is the calculated mass of intact peptide 3 linked through two disulfide bonds to the C-terminus of truncated peptide 4 (C7ATC8ALG), formed by the cleavage of the bond between C6 and C7. The presence of this ion was unaffected by differences in the N-terminus of peptide 1, but it was absent from species in which the C-terminus of peptide 3 was truncated, providing confirmation of this assignment (Fig. 10). Another ion observed in several spectra at m/z 565·2 (singly charged) is likely due to TC8ALG plus an additional cysteine, most probably C3 from the C-terminus of peptide 3, but no related ions could be identified to confirm this. This ion was absent from the tandem spectrum of the pepsin-derived species lacking the glycine residue at the C-terminus of the truncated peptide 4. These fragmentations are summarized in Fig. 11. Although no further cysteine interconnections were revealed by tandem mass spectrometry, it is clear that C1 and C2 are linked to C5 and C6; similarly C3 and C4 are linked to C7 and C8, probably in reverse order. At the present time the complete ordering of these connections remains unknown.

Discussion

We have described the construction of a plasmid vector designed to produce a rat IGFBP-4 fusion protein at high expression levels in E. coli. Up to 5 mg of IGFBP-4 could be purified with this method from 1·2 l culture medium.
The large number of disulfide bonds present in the IGFBPs makes the high level expression and purification very difficult. Previously not more than 0·2 mg/l IGFBP-4 could be purified from culture media (Honda et al. 1996). The method described here provides the highest yield for any IGFBP-4 expression systems reported so far. The results of several assays comparing recombinant and authentic rat IGFBP-4 confirmed proper folding and biological activity of the recombinant material. No free sulfhydryls were found by Ellman’s reagent (data not shown) nor by mass spectroscopy.

IGFBP-4 has one N-linked glycosylation site at asparagine 104. Authentic rat IGFBP-4 was a mixture of about 20% glycosylated and 80% non-glycosylated forms. If the glycosylation of IGFBP-4 affected IGF-I binding, the differences in the IC\textsubscript{50} values of the two IGFBP-4 preparations should have been detectable. Hence this data suggests that the glycosylation state of IGFBP-4 has no effect on its affinity to IGF-I and, therefore the glycosylation site of IGFBP-4 is in a region of the protein that is not involved in IGF binding.

The isoelectric points of authentic rat IGFBP-4 were determined by two-dimensional gel electrophoresis, in which the protein separated into seven isoforms, three glycosylated and four non-glycosylated. We compared the recombinant material with authentic rat IGFBP-4 and with IGFBP-4 in rat serum. The recombinant material predictably lacked the three glycosylated isoforms, but showed the four non-glycosylated forms. The isoelectric points of the four non-glycosylated forms were 5·8, 6·2, 6·3 and 6·5 and were the same for the recombinant and the material from serum. The isoelectric points were also similar to human IGFBP-4 (Weber et al. 1999). We conclude that the isoforms of mature IGFBP-4 are not

Figure 8 Tandem mass spectrum of the 7+ charge state ion at mass/charge (m/z) 837·5, representing the higher mass component D in Fig. 7 (5855 Da) from fraction 2. The lower panel shows predominant singly charged y and b ions from cleavages outside the core region contained within the hexagon in Fig. 9. Doubly charged ions highlighted in the upper panel correspond to peptide 6 linked to the y ions of peptide 4, showing that C\textsuperscript{9} and C\textsuperscript{11} are linked.
related to amino acid substitutions, because the recombinant IGFBP-4 was expressed from a single clone. The isoforms would seem to be chemical or post-translational modifications, which occur in E.coli and rat cells in the same way. Minor modifications such as oxidation of methionine, deamidations and acetylations are potential causes of such differences. However, except for a minor proportion of methionine sulfoxide, these were not observed by mass spectrometry. It is unlikely that such a modification exert different biological function based on our binding data.

The correct formation of disulfide bonds is essential for proper folding of the protein and consequently for the biological action. No difference between recombinant IGFBP-4 and IGFBP-4 purified from rat serum could be detected, suggesting that the recombinant material is folded the same way and therefore should have the same biological activity.

It has been shown previously that even the highly conserved cysteines in the N-terminal region of the IGFBPs do not necessarily have the same disulfide bond linkage. Neumann & Bach (1999) demonstrated by mass spectrometry that the N-terminal linkages involving C1 to C4 of IGFBP-1 are different from those of IGFBP-6. However, the C-terminal linkages of the six conserved cysteines are the same for at least IGFBP-2 and -6. In this study we determined eight of 10 disulfide linkage of IGFBP-4 including linkages not previously identified and narrowed the choices of the remaining two, as summarized in Fig. 12. The C-terminal cysteines of IGFBP-4 are linked in the same manner as in IGFBP-2 and -6 and, therefore, are probably the same for all six IGFBPs. Spencer & Chan (1996) suggest that this part of the protein plays an important role in binding to IGFs. Forbes et al. (1996) showed that deletion of 63 amino acids at the C-terminus of IGFBP-2 results in an up to 80-fold reduction in its ability to bind IGF-I compared with the wild-type IGFBP-2. It has also been shown that removal of 24 or 48 amino acids at the C-terminus of IGFBP-2 alters the IGF-II preference over IGF-I. Both IGFBP-2 and -6 share a preference for IGF-II and have the same C-terminal disulfide linkages, leading to the suggestion that a different disulfide linkage (such as those in the highly conserved C18WC19V region) could conceivably influence IGF-II binding preference. In this report we have shown that the three C-terminal disulfide linkages of IGFBP-4 are the same as in IGFBP-2 and -6. However, IGFBP-4 does not share the same IGF-II preference, thus the highly conserved three C-terminal disulfide linkages either have no effect on the IGF binding preference, or they are not the sole determinants.

IGFBP-4 contains the largest number of cysteines of all the IGFBPs. It has two unique cysteines in the central region of the protein at residues 110 and 117 (C113 and C114) which we have shown are linked to each other. This unique linkage in an otherwise flexible region of IGFBP-4 may play an important part in the distinctive biological behavior of this binding protein, i.e. purely inhibitory and lack of cell surface association.

![Figure 9](https://example.com/figure9.png)

**Figure 9** Summary of the collision-induced dissociation (CID) of the cluster of N-terminal peptides illustrated in Fig. 7. Vertical lines between amino acids indicate bonds that were seen to break. Ions corresponding to the N-terminal portion of a peptide are classified as b whereas C-terminal ions are classified as y. The single number or group of numbers given as a prefix indicates the peptides involved in an ion, whereas the subscripted suffix indicates the site of cleavage, e.g. 6–4y7–132+ refers to a series of ions containing intact peptide 6 linked to C-terminal fragments of peptide 4 cleaved at the 7th–13th amino acids from the C-terminus. The superscript 2+ indicates that these ions are doubly charged, thus they appear in the mass spectrum at mass/charge (m/z) values equivalent to half the actual mass.
Kalus et al. (1998) recently reported a component of the IGF binding site in the N-terminus of IGFBP-5, containing C9 to C12. The disulfide linkage pattern in this region of the protein is conserved between IGFBPs 1, 3 and 6 (Hashimoto et al. 1997, Neumann & Bach 1999). Here we have demonstrated that the disulfide linkages of IGFBP-4 are identical in this region.

Of the six N-terminal linkages of IGFBP-4 two could not be fully determined, however, our data indicate that the cysteines in this region are probably the same as in IGFBP-1, but not IGFBP-6. In IGFBP-6 C1 is linked to C2 and C3 is linked to C4, whereas in IGFBP-1 and IGFBP-4 C1 is not linked to C2 and C3 is not linked to C4. We confirm the conclusion of Neumann & Bach (1999) that this N-terminal region might be important for binding to IGF-II, as IGFBP-6 has the highest affinity for IGF-II and a unique disulfide linkage in this part of the protein. The conserved sequence C5GC6C7 makes it difficult or impossible to determine the complete pattern of N-terminal disulfide linkages by proteolysis alone, because there is no suitable cleavage site in this region. In fact, the N-terminal disulfide linkage of IGFBP-1 and
IGFBP-3 could not be determined by this approach (Neumann & Bach 1999, Hashimoto et al. 1997). However, tandem mass spectrometry has provided new information concerning some of these linkages in IGFBP-4, suggesting that C5 is linked to C6 and C8 is linked to C7. It appears that C1 and C2 are linked to C5 and C6, but the order of these linkages could not be determined. These findings are summarized in Fig. 12, which distinguishes between confirmation of previous reports (dotted lines) and new findings reported here (continuous lines).

To understand the biological action of the IGFs and their binding proteins, it will be essential to determine the three-dimensional structure of the IGFBPs. This study presents an expression and purification system yielding sufficient quantities of IGFBP-4 for intensive structural research by NMR spectroscopy and crystallography as well as disulfide bond pairings that will be useful in interpreting 3-D data. Ultimately the information on the 3-D structure could be extremely valuable for drug design aimed at modulating the actions of IGFs and the direct actions of IGFBPs.

In summary, we have developed an expression and purification system giving a high yield of correctly folded IGFBP-4. We have determined eight of 10 disulfide linkages in this protein, which has a very high disulfide bond ratio per amino acid. We have shown that the three C-terminal disulfide bonds are the same as for IGFBP-2 and IGFBP-6. In addition, we have confirmed that the disulfide linkage in one component of the IGF binding site (containing C9 to C11) is likely to be the same for all IGFBPs and that the two cysteines unique to IGFBP-4 (C13 and C14) are linked together.

Acknowledgements

We thank Dr S Shimasaki for the IGFBP-4 purified from rat serum and Gerald Spott and Matthias Weber for the preparation of the two-dimensional gel electrophoreses. The work was supported by grant GM 27345. Mass spectrometry was carried out in the University of California, San Francisco Mass Spectrometry Facility supported by NIH NCRR grant RR 01614.

References


Neumann GM & Bach LA 1999 The N-terminal disulfide linkages of human insulin-like growth factor-binding protein-6 (hIGFBP-6) and hIGFBP-1 are different as determined by mass spectrometry. Journal of Biological Chemistry 274 14587–14594.


www.endocrinology.org


Received in final form 21 July 2000
Accepted 27 September 2000