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*

(Received for publication, December 24, 1998, and in revised form, February 22, 1999)

Gregory M. Neumann‡ and Leon A. Bach§

From the ‡Department of Biochemistry, Latrobe University, Bundoora, Victoria 3083 and the §Department of Medicine, University of Melbourne, Austin & Repatriation Medical Centre, Heidelberg, Victoria 3084, Australia

The actions of insulin-like growth factors (IGFs) are modulated by a family of six high affinity binding proteins (IGFBPs 1–6). IGFBP-6 differs from other IGFBPs in having the highest affinity for IGF-II and in binding IGF-I with 20–100-fold lower affinity. IGFBPs 1–5 contain 18 conserved cysteines, but human IGFBP-6 lacks 2 of the 12 N-terminal cysteines. The complete disulfide linkages of IGFBP-6 were determined using electrospray ionization mass spectrometry of purified tryp tic peptide complexes digested with combinations of trypsin, thermolysin, and endoproteinase Glu-C. Numbering IGFBP-6 cysteines sequentially from the N terminus, the first three disulfide linkages are Cys1-Cys2, Cys3-Cys4, and Cys5-Cys6. The next two linkages are Cys7-Cys8 and Cys9-Cys10, which are analogous to those previously determined for IGFBP-3 and IGFBP-5. The C-terminal linkages are Cys11-Cys12, Cys13-Cys14, and Cys15-Cys16, analogous to those previously determined for IGFBP-2. Disulfide linkages of IGFBP-1 were partially determined and show that Cys1 is not linked to Cys2 and Cys3 is not linked to Cys4. Analogous with IGFBP-3, IGFBP-5, and IGFBP-6, Cys1-Cys2 and Cys9-Cys10. Cys12 of IGFBP-1 are also disulfide-linked. The N-terminal linkages of IGFBP-6 differ significantly from those of IGFBP-1 (and, by implication, the other IGFBPs), which could contribute to the distinctive IGF binding properties of IGFBP-6.

The insulin-like growth factor (IGF)1 system plays an important role in normal physiology (1). Dysregulation of the system has also been implicated in many diseases including malignancy, atherosclerosis, and the development of diabetic complications. IGF-I and IGF-II promote proliferation and differentiation of many cell types. More recently, IGFBPs have been ascribed a potent role as anti-apoptotic agents. The actions of IGFBPs are modulated by a family of six structurally related high affinity IGF-binding proteins (IGFBPs 1–6) (1–3). The N- and C-terminal domains of IGFBPs 1–6 share significant sequence homology, whereas the central regions linking these domains are not homologous with each other. Under differing circumstances, IGFBPs may inhibit or enhance IGF actions. The IGFBPs differ in their regulation, relative IGF binding affinities, susceptibility to proteolysis by specific proteases, and sites of synthesis. The IGFBPs therefore constitute a flexible system for the regulation of IGF activity.

IGFBP-6 is an O-linked glycoprotein, which differs from IGFBPs 1–5 in a number of significant respects. Of the IGFBPs, IGFBP-6 has the highest binding affinity for IGF-II (4). IGFBP-6 also has the highest specificity for IGF-II, exhibiting a 20–100-fold higher binding affinity for IGF-II than IGF-I, whereas IGFBPs 1–5 bind to IGF-II with equal or slightly higher affinity than IGF-I (4–7). A potentially significant structural difference is that, whereas human and rat IGFBPs 1–5 have 12 homologous N-terminal and 6 homologous C-terminal cysteines, 2 of the N-terminal cysteines are not present in human and rat IGFBP-6 and a short sequence containing another 2 N-terminal cysteines is missing from rat IGFBP-6 (8).

Disulfide linkages of cysteines are important for the correct folding and maintenance of the three-dimensional structure of many proteins. The N-terminal cysteines of the IGFBPs are disulfide-linked to each other, as are the C-terminal cysteines, with no disulfide linkages between the N- and C-terminal domains (2). To date, a complete set of disulfide linkages has not been reported for any of the IGFBPs. Recently, the three C-terminal disulfide linkages of bovine IGFBP-2 were determined (9). The C-terminal disulfide linkages of human IGFBP-6 were also partially determined (10) and are consistent with those of IGFBP-2. We recently suggested that the N-terminal region of IGFBP-6 consists of two disulfide-linked domains, the first including the six most N-terminal cysteines and the second including the next four cysteines in the sequence (10). A recent structural study of IGFBP-5 has shown that the second of these domains contains the high affinity IGF binding site (11); the two disulfide linkages within this domain are the same as those determined for the homologous cysteines of IGFBP-3 (12). The other N-terminal domain of IGFBP-5, containing the eight most N-terminal cysteines, is thought to form a separate domain that does not interact with the high affinity binding site of IGFBP-5 nor strongly influence IGF binding to this site (11).

IGFBPs 1–5 contain a GCGCC motif in their N-terminal sequences, which is replaced by GCAEA in human IGFBP-6 (8). Therefore, unless the two adjacent cysteines in IGFBPs 1–5 are disulfide-linked, which is unlikely on structural grounds, the disulfide linkages of IGFBP-6 in this domain must differ from those of the other IGFBPs. A chimera of the N-terminal and non-conserved middle regions of IGFBP-6 with the C-terminal region of IGFBP-5 retains preferential binding affinity for
IGF-II over IGF-I (13). Since the non-conserved middle regions of IGFBPs are not thought to be directly involved in high affinity IGF binding, this suggests that the determinants of IGF-II binding preference are located in the N-terminal region. In order to further characterize the structural differences between IGFBP-6 and IGFBPs 1–5, we completely solved the disulfide linkages of human IGFBP-6 using electrospray ionization mass spectrometry (14) and reverse-phase HPLC of proteolytic digests. Since none of the N-terminal disulfide linkages and only a few other linkages of some of IGFBPs 1–5 are known, we also partially solved the disulfide linkages of IGFBP-1 to provide a basis for comparison. We found that at least two of the N-terminal disulfide linkages of IGFBP-1 are different from those of IGFBP-6.

**EXPERIMENTAL PROCEDURES**

**Expression and Purification of Recombinant Human IGFBP-6**—Recombinant glycoylated human IGFBP-6 (hIGFBP-6) was previously expressed in Chinese hamster ovary cells and purified by IGF-II affinity chromatography and reverse-phase FPLC, but was contaminated with endogenously synthesized IGFBP-4, which copurified with hIGFBP-6, necessitating lectin chromatography for removal of the IGFBP-4 (10). For this reason, human embryonic kidney 293 cells, which synthesize only small amounts of IGFBP-2 (15), were stably transfected with phBG6-E3, the eukaryotic expression vector encoding human IGFBP-6 (16), using LipofectAMINE. Genetin-resistant clones were selected, and the colony producing the highest levels of hIGFBP-6 were recloned and expanded for production of recombinant protein.

hIGFBP-6 was purified from conditioned medium by IGF-II affinity chromatography followed by reverse-phase medium-pressure chromatography (ProRPC 5/10 on an FPLC system, Amersham Pharmacia Biotech) using a 16–40% acetonitrile, 0.1% trifluoroacetic acid gradient (17). The identity and purity of hIGFBP-6 were confirmed by N-terminal amino acid sequencing.

**Purification of Human IGFBP-1**—Human IGFBP-1 (hIGFBP-1) was purified from amniotic fluid by acid gel filtration (Bio-Gel P-30, Bio-Rad), IGF-II affinity chromatography and reverse-phase FPLC, but was contaminated with endogenously synthesized IGFBP-4 (10, 18) on a Perkin-Elmer Sciex API-300 triple quadrupole mass spectrometer with micro-ion spray ion source, calibrated to an accuracy of ±0.01% using singly charged poly(propylene glycol) ions. MS/MS spectra (Q3 scans) were obtained using nitrogen collision gas (4 millitorr pressure, 20.7 cm cell length) and optimized collision energies of 32–64 eV. Signal-averaged raw mass spectra were analyzed manually and transformed to a true mass scale using the PE-Sciex BioMultiview program Biospec Reconstruct. Peptide sequences inferred from observed masses were confirmed by analysis of MS/MS spectra manually and using the BioMultiview programs Predict Sequence and Peptide Fragments.

**Amino Acid Sequencing**—N-terminal amino acid sequencing was performed by sequential Edman degradation using a Hewlett-Packard G1005A automated protein sequencing system, calibrated with phenylthiohydantoin-amino acid standards prior to each sequencing run.

**RESULTS**

**Expression and Purification of hIGFBP-6 in 293 Cells**—Approximately 300 μg of recombinant human IGFBP-6 was purified from 500 ml of medium conditioned by stably transfected 293 human embryonic kidney cells. N-terminal Edman sequencing showed that ~80% of the protein commenced at Arg28 with the sequence RC/P/G/C/GQG (numbering based on the sequence of the IGFBP-6 precursor protein, SWISS-PROT accession number P24592; Fig. 1). The remainder of the protein...
Molecular masses of major observed components were determined by ESMS of disulfide-linked tryptic peptide complexes in HPLC fractions 1, 2, and 3 (Fig. 2), with or without additional digestion (as indicated under “protease”) by endoprotease Glu-C (Glu-C), thermolysin (Therm), chymotrypsin (Chym), or a mixture of chymotrypsin and thermolysin (Ch&Th), as described under “Experimental Procedures.” Sequence assignments of tryptic peptide complexes were determined previously by ESMS and Edman sequencing, as was the linkage C11-C12 (10). Sequence assignments of fragments with masses less than 2300 Da from additional digestion were confirmed by MS/MS fragmentation analysis as in Fig. 3 (other data not shown). ESMS (observed) masses averaged from several charge states (typically 2+ to 4+) are compared with calculated average molecular masses above 2000 Da, but with monoisotopic (m) masses below 2000 Da due to the lower observed charge states (1+ to 2+). Uncertainties are mean deviations plus calibration uncertainty. Calculated masses assume disulfide linkage of all cysteines (C), which are 1 molecular masses above 2000 Da, but with monoisotopic (m) masses below 2000 Da due to the lower observed charge states (1+ to 2+). 

By default, and as previously determined (10), the remaining three fractions account for 14 of the 16 cysteines in hIGFBP-6. Three terminal fragments analyzed in the present study. In the results that follow, cysteines are numbered 1–16 in hIGFBP-6 and 1–18 in hIGFBP-1 sequentially from the N terminus (Fig. 1).

**TABLE I**

| HPLC no./protease | Disulfide-linked peptide complex or digestion product | Molecular mass | Observed | Calculated |
|-------------------|-----------------------------------------------------|---------------|---------|-----------|
|                   |                                                     | Da            |         |           |
| 2/none            | RC³Pqc²G—GC³PGGC³V—GC³AaEGC⁶LR                     | 3633.6 ± 0.6  | 3633.9  |
| 2/Glu-C           | RC³Pqc²GqGqVAGC³PqGC³VEE                            | 2031.7 ± 0.7  | 2032.2  |
| 2/Glu-C           | DGGSPAEGC³AEAEGC³LR                                  | 1619.0 ± 0.5  | 1618.6 (m) |
| 2/Therm           | RC³Pqc²GGQ                                      | 774.6 ± 0.5   | 774.3 (m) |
| 2/Therm           | VQAGC³PGGC³VEE—AEGC³AEAEGC³LR                   | 2876.6 ± 0.7  | 2877.1  |
| 3/none            | EGGEC³GVYTPNC³APLQC³H—DEAPLR, linked to C¹⁹LPAR       | 3679.8 ± 0.6  | 3680.1  |
| 3/Ch&Th           | EGGEC³GVY, linked to LQC³HPKDDDEAP                   | 2230.8 ± 0.7  | 2231.4  |
| 3/Ch&Th           | TPNC³APG, linked to C¹⁰LPAR                          | 1214.8 ± 0.5  | 1214.6 (m) |
| 1/none            | QC¹¹R, linked to RGC¹¹WC¹⁵VDR, linked to SLPGSPDGNSGSSC¹⁶PTGSS | 3186.4 ± 0.5  | 3186.5  |
| 1/Chym            | QC¹¹R, linked to RGC¹¹W                              | 1020.7 ± 0.5  | 1020.4 (m) |
| 1/Chym            | C¹⁵VDR, linked to SLPGSPDGNSGSSC¹⁶PTGSS               | 2182.8 ± 0.6  | 2183.3  |

**Deduced hIGFBP-6 linkages**

|                         | N-terminal: C¹—C², C³—C², C⁵—C², C⁶—C⁵, C⁶—C¹⁰ |
|-------------------------|------------------------------------------------------|
|                         | C-terminal: C¹—C² (see legend), C¹—C¹⁴, C⁵—C¹⁶       |

**Purification of Disulfide-linked Tryptic Fragments of hIGFBP-6**—Disulfide-linked tryptic fragments of hIGFBP-6 were produced by tryptic digestion of 40 µg of purified hIGFBP-6 followed by C18 reverse-phase HPLC fractionation, which yielded a HPLC profile (Fig. 2) similar to that following a previous digestion using the same procedure on purified hIGFBP-6 expressed in Chinese hamster ovary cells (10). ESMS analysis of all major and some minor peak fractions confirmed that the three most intense peaks (Fig. 2, peaks 1–3) correspond to the same three hIGFBP-6 disulfide-linked tryptic fragments of hIGFBP-6 that were previously identified by ESMS and Edman sequencing (10) and that other major and minor peptide fragments were also previously identified. As reported previously (10), attempts to release disulfide-linked peptides by reduction with 2-mercaptoethanol in concentrations compatible with electrospray ionization (less than a few percent, v/v) were unsuccessful at room temperature and resulted in peptide degradation and/or loss at elevated temperatures. Therefore, assigned sequences for all fragments with masses below 2300 Da were confirmed by MS/MS fragmentation analysis as described under “Experimental Procedures.”

Fractions 1, 2, and 3 were further digested and reanalyzed by ESMS, as described below. The peptide complexes in these three fractions account for 14 of the 16 cysteines in hIGFBP-6. By default, and as previously determined (10), the remaining two cysteines (Cys¹³ and Cys¹² in the sequences below) are disulfide-linked, and connect glycosylated (140–165)NPGTST-TSPQPNASAVQDTEMGCP¹¹RR to (181–193)GAQTLYVPNC¹²DHR. A number of glycoforms of this complex, with masses corresponding to carbohydrate contents of 3–7 N-acetylhexosamine-hexose disaccharides and 0–5 sialic acids, were identified as minor components of peak fraction 1 (Fig. 2) and nearby fractions (data not shown), thereby confirming that Cys¹¹ is indeed linked to Cys¹² in hIGFBP-6 expressed in 293 cells.

**ESMS of Proteolytic Digests of hIGFBP-6 Disulfide-linked Tryptic Fragments**—The three disulfide-linked tryptic complexes purified as HPLC fractions 1, 2, and 3 (Fig. 2) were subjected to further digestion by *S. aureus* V8 protease (endoprotease Glu-C), thermolysin, chymotrypsin, or a mixture of chymotrypsin and thermolysin, as detailed in Table I. Samples of fraction 2 (Fig. 2), containing the 3634 Da N-terminal tryptic fragment (28–65)RC¹¹PQC³GGVAGC³PqGC³VEEEDGG-SPAEGC³AaEaEgC³6LR with three intramolecular disulfide linkages, were digested with endoprotease Glu-C and reanalyzed by ESMS (Table I). This revealed digestion into two complementary fragments with masses of 1619 and 2032 Da (Table I), as shown in Fig. 3A. Minor fragments were present with extra or missing Glu (E), consistent with alternate Glu-C cleavages in the sequence EEE²⁶ (Fig. 3A), and minor fragments due to additional cleavage at one or more Glu residues (E) in the sequence EGCAE¹² were also observed at low mass (not shown). The identity of the 1619 Da fragment as fragment (49–65)DGGSSPAEGC³AaEaEgC³6LR with the disulfide linkage Cys¹⁻Cys⁶, was confirmed by MS/MS collisional fragmentation (spectrum not shown) of the doubly protonated (m/z 811) ion, which generated the following b-type (N-terminal) and complementary y-type (C-terminal) sequence ions: b₂—b₁₀, y₁⁻y₁₃ (singly charged), and y₁₃—y₁₅ (doubly charged). The ions b₁₀—b₁₄ and y₁₃—y₁₅ were completely absent, as expected due to the presence of the disulfide linkage bridging the cysteines at positions 9 and 15 in the fragment sequence (21).

The identity of the 2031 Da fragment as fragment (28–48)RC¹¹PQC³GGVAGC³PqGC³VEE with two intramolecular disulfide linkages, was confirmed and the disulfide linkage arrangement also determined by MS/MS collisional fragmentation of the doubly protonated (m/z 1017) ion (Fig. 3B). Prominent b-type (N-terminal) fragment ions b₁₀−b₁₄ and b₁₇−b₂₁ were
observed, while the ions b_{13}–b_{16} and b_{2}–b_{4} were conspicuously absent, as expected due to the presence of bridging disulfide linkages (21). The key observation is the presence of prominent b_{5}–b_{12} ions, which would not be observed unless the two disulfide linkages are sequential (do not cross over), as shown schematically in Fig. 3B. The linkage arrangement Cys1-Cys2 was independently determined by ESMS analysis of thermolysin digests of the 3634-Da N-terminal tryptic fragment, which were found to contain the complementary fragments (28–35)RC1PGCGQGVQAGCPGCVEEE and (36–65)VQAGCPGCVEEEDGGSPARCGCAYEGCLR as major components (Table I). Combining this result with that from Glu-C digestion (Fig. 3A) independently implies the same disulfide linkages as those determined by MS/MS (Fig. 3B). It is concluded that the three most N-terminal disulfide linkages of hIGFBP-6 are Cys1-Cys2, Cys3-Cys4, and Cys5-Cys6.

Samples of fraction 3 (Fig. 2), containing the 3680-Da tryptic fragment of hIGFBP-6, were digested with endoproteinase Glu-C and analyzed by ESMS as described under “Experimental Procedures.” Upper spectrum (A), mass spectrum (transform) showing two major peptide fragments with deduced disulfide linkages (see below) as indicated. Lower spectrum (B), MS/MS fragmentation spectrum (48-eV collision energy) of the [M+2H]^2+ ion (m/z 1017) of the RCPGCQGQGVQAGCPGCGVEEVE fragment (with two intramolecular disulfide linkages) shown in spectrum A. As indicated schematically, the two disulfide linkages are sequential (do not cross over), as determined from the presence of prominent b-type (N-terminal) fragment ions b_{5}–b_{12} and supported by the presence of b_{17}–b_{21}, but absence of b_{13}–b_{16} and b_{2}–b_{4}. Above m/z 1300 (not shown), only weak signals due to singly charged b_{17}–b_{19} fragment ions were observed.

![Figure 3](image-url)  
**FIG. 3.** Electrospray ionization mass spectra of an endoproteinase Glu-C digest of the 3.6-kDa N-terminal tryptic fragment of hIGFBP-6. Aliquots (50 pmol) of fraction 2 (Fig. 2), containing the purified 3634-Da N-terminal tryptic fragment RCPGCQGQGVQAGCPGCVEEEDGGSPARCGCAYEGCLR with three intramolecular disulfide linkages, were digested with endoproteinase Glu-C and analyzed by ESMS as described under “Experimental Procedures.” Upper spectrum (A), mass spectrum (transform) showing two major peptide fragments with deduced disulfide linkages (see below) as indicated. Lower spectrum (B), MS/MS fragmentation spectrum (48-eV collision energy) of the [M+2H]^2+ ion (m/z 1017) of the RCPGCQGQGVQAGCPGCGVEEVE fragment (with two intramolecular disulfide linkages) shown in spectrum A. As indicated schematically, the two disulfide linkages are sequential (do not cross over), as determined from the presence of prominent b-type (N-terminal) fragment ions b_{5}–b_{12} and supported by the presence of b_{17}–b_{21}, but absence of b_{13}–b_{16} and b_{2}–b_{4}. Above m/z 1300 (not shown), only weak signals due to singly charged b_{17}–b_{19} fragment ions were observed.

![Figure 4](image-url)  
**FIG. 4.** Structural features of hIGFBP-6. Numbers indicate cysteines involved in disulfide linkages (numbering based on the proIGFBP-6 sequence, SWISS-PROT accession number P24592, Fig. 1). The conserved N-terminal and C-terminal regions of human IGFBP-6 are shaded, and the non-conserved mid-region is unshaded. Cysteines are indicated by vertical lines and disulfide linkages by horizontal lines. Cysteines contained in the sequence AGCPGCGCVE, which is not present in the rat IGFBP-6 sequence (8), are indicated by crosses. Previously determined O-glycosylation sites (10) are indicated by arrows.

![Figure 5](image-url)  
**FIG. 5.** Reverse-phase HPLC fractionation of a proteolytic digest of hIGFBP-1. hIGFBP-1 (20 μg) was digested with a mixture of trypsin and *S. aureus* V8 protease (endoproteinase Glu-C/Asp-C) as described under “Experimental Procedures” and subjected to C18 reverse-phase HPLC as in Fig. 2. Upper trace, A_{215}. Lower trace, A_{280}. Peptides were identified by ESMS (Table II, Fig. 6) and the 5275-Da N-terminal disulfide-linked peptide complex in peak fraction 16 additionally analyzed by Edman sequencing (Table III).
Table II
ESMS analysis of hIGFBP-1 proteolytic fragments and deduced disulfide linkages

HPLC fractions (as defined in Fig. 5) from reverse-phase HPLC of hIGFBP-1 digested with a mixture of trypsin and S. aureus V8 protease (endoproteinase Glu-CAsp-C) were analyzed by ESMS as described under “Experimental Procedures,” yielding observed masses as indicated. Uncertainties are mean deviations plus calibration uncertainty. Calculated molecular masses based on the hIGFBP-1 sequence (Fig. 1) are monoisotopic below 2000 Da and average (av) above 2000 Da. Assigned sequences for all fragments with masses below 2000 Da were confirmed by MS/MS fragmentation analysis (data not shown) as in Fig. 3. The sequence assignment for the 5275-Da fraction 16 fragment (see Fig. 6 for mass spectrum) was confirmed by Edman sequencing (Table III). Disulfide-linked peptides (indicated by * in Table II) identified in HPLC fractions 4, 13, 14, and 16 (indicated in bold) account for all cysteines except C5, C6, C7, and C8. Digestion of the 5275-Da fraction 16 complex by thermolysin (Th) resulted in the generation of two complementary digestion products (indicated in bold) with fragments still disulfide-linked (data not shown). Digestion of the 5275-Da fraction 16 complex with a mixture of thermolysin and chymotrypsin resulted in the generation of two complementary digestion products (indicated in bold) with fragments still disulfide-linked (data not shown). Digestion of the 5275-Da fraction 16 complex by thermolysin (Th) resulted in an 18-Da mass increase of the complex (or with chymotrypsin at protease-to-peptide molar ratios that were slightly (2–3-fold) higher than those used with other digests in order to obtain complete digestion. ESMS analysis showed that two major fragments with masses of 1021 and 2183 Da were generated (Table I); these correspond to (200–202)QC13R linked to (209–213)RGPC14W and (214–217)C15-VDR linked to (221–239)SLPGSPDGNGSSSC16PTGSS, as confirmed by MS/MS analysis (data not shown). These results clearly differentiate between the two possible linkage arrangements within the C-terminal tryptic complex, indicating that the disulfide linkages are Cys13–Cys14 and Cys15–Cys16. Since Cys14 is connected to Cys13 (Table I), all three C-terminal disulfide linkages are therefore sequential. The complete disulfide linkages determined for hIGFBP-6 are summarized at the foot of Table II and shown in Fig. 4.

ESMS and Edman Sequencing of Proteolytic Fragments of hIGFBP-1—hIGFBP-1 was digested with a mixture of trypsin and S. aureus V8 protease and subjected to C18 reverse-phase HPLC fractionation as shown in Fig. 5. All major and most minor HPLC peak fractions (Fig. 5) were analyzed by ESMS (Table II) and initially identified by comparison of observed masses with those expected on the basis of the hIGFBP-1 sequence (Fig. 1) and the protease specificities of trypsin (Arg-

| HPLC fraction | Sequential peptide fragment or disulfide-linked peptide complex | Observed | Calculated |
|---------------|---------------------------------------------------------------|----------|------------|
| 16 (Th)       | (1–12)APWQC1APC2SAEK*                                        | 5275.0 ± 0.6 | 5275.2 (av) |
|               | (13–25)LALCPVVSASC2SE*                                         |           |            |
|               | (29–52)SAGC3GC4C5PMC6ALPLGAAC6GVATAR*                          |           |            |
|               | (56–60)GLSC6R*                                                 |           |            |
| 16            | (1–12)APWQC1APC2SAEK*                                        | 3771.2 ± 0.6 | 3771.5 (av) |
|               | (13–25)LALCPVVSASC2SE*                                         |           |            |
|               | (29–41)SAGC3GC4C5PMC6ALP6*                                     |           |            |
|               | (56–60)GLSC6R*                                                 |           |            |

Deduced hIGFBP-1 linkage information

C5 not linked to C6; C7 not linked to C8; one or both of C1, C2, and one or both of C3, C4 linked to two or four of C5, C6, C7, C8; C9-C11 (see legend); C10-C11. By default, C12-C14 (or C13); C15-C17 (or C16).

Disulfide Linkages of IGFBP-6

HPLC fraction (as defined in Fig. 4) from reverse-phase HPLC of hIGFBP-1 digested with a mixture of trypsin and S. aureus V8 protease (endoproteinase Glu-CAsp-C) were analyzed by ESMS as described under “Experimental Procedures,” yielding observed masses as indicated. Uncertainties are mean deviations plus calibration uncertainty. Calculated molecular masses based on the hIGFBP-1 sequence (Fig. 1) are monoisotopic below 2000 Da and average (av) above 2000 Da. Assigned sequences for all fragments with masses below 2000 Da were confirmed by MS/MS fragmentation analysis (data not shown) as in Fig. 3. The sequence assignment for the 5275-Da fraction 16 fragment (see Fig. 6 for mass spectrum) was confirmed by Edman sequencing (Table III). Disulfide-linked peptides (indicated by *) identified in HPLC fractions 4, 13, 14, and 16 (indicated in bold) account for all cysteines except C5, C6, C7, and C8. Digestion of the 5275-Da fraction 16 complex by thermolysin (Th) resulted in a major 3771-Da fragment (as indicated), which corresponds to removal of (42–52)LGAAC6GVATAR linked to (56–60)GLSC6R, demonstrating that C6 is connected to C7. Deduced disulfide linkage information for hIGFBP-1 is summarized at the foot of the table. All cysteines (C) are numbered from 1 to 18 sequentially (Fig. 1).
Cys3, and Cys4 cannot all be interconnected. Given the result Fraction 13 contained the fragment (146–154)KWKEPCys13RIE removal of WK148 and/or IE154 were identified as minor components in Fraction 14, and additional related fragments due to further fragments: (1–12)APWQC1APC2SAEK, (13–25)LALC3PPVSA-N-terminal disulfide-linked complex consisting of the following components of fractions 10 and 11 (data not shown).

II) contained peptides that were identified as oxidized (methi- identified. For example, fractions 10 and 15 (not listed in Table major and minor peak fractions analyzed were successfully expected to contain all four cysteines (Table II) was not observed. All other expected fragments containing more than four amino acids were recovered (Table II), and all fragments in all major and minor peak fractions analyzed were successfully identified. For example, fractions 10 and 15 (not listed in Table II) contained peptides that were identified as oxidized (methi-}

The identified disulfide-linked peptides (Table II) account for all hIGFBP-1 cysteines except the four C-terminal cysteines, Cys15 to Cys18. No peptide complexes containing any of these cysteines were identified (Table II), and all fragments in all major and minor peak fractions analyzed were successfully identified. For example, fractions 10 and 15 (not listed in Table II) contained peptides that were identified as oxidized (methi-}

Fig. 6. Electrospray ionization mass spectrum of the N-terminal disulfide-linked fragment of hIGFBP-1. Fraction 16 was reverse-phase HPLC of proteolysed hIGFBP-1 (Fig. 5) was analyzed by ESMS as described under “Experimental Procedures.” Upper spectrum, unprocessed mass spectrum from 2 pmol of peptide (in 2 μl of 50% acetonitrile, 0.1% formic acid), showing ions with (left to right) 6, 5, 4, and 3 positive charges. Lower spectrum, mass transform from m/z to a true mass scale, showing the single major 5275-Da N-terminal disulfide-linked fragment (identified in Tables II and III) and adjacent minor peaks corresponding to oxygen, sodium, potassium, and ferrous ion adducts.

The IGFBPs are important modulators of IGF actions (1–3). Although the actions of IGFBPs have been the subject of intense scrutiny for over a decade, only limited information regarding their structure is available. The results of this study demonstrate for the first time the complete disulfide linkages of an IGFBP. In addition, the disulfide linkages of IGFBP-1 have been partially characterized and compared with those of IGFBP-6, revealing that the N-terminal disulfide linkages of these IGFBPs differ significantly. In contrast, the C-terminal disulfide linkages of bovine IGFBP-2 (9) and human IGFBP-6 are identical, and the disulfide linkages of IGFBP-5 within its putative high affinity IGF binding domain (11) are identical to those in the corresponding domains of IGFBP-6, IGFBP-1 (both shown in the present study) and IGFBP-3 (12).

A distinctive feature of IGFBP-6 compared with IGFBPs 1–5 is its high binding affinity for IGF-II and its marked preferential binding of IGF-II relative to IGF-1. A chimera of the N-terminal and middle regions of IGFBP-6 with the C-terminal region of IGFBP-5 retains the binding preference for IGF-II (13). Since the non-conserved middle regions of the IGFBPs are not thought to be directly involved in IGF binding, the structural determinants of the IGF-II binding preference are therefore likely to reside in the N terminus. It may be of structural significance that rat IGFBP-6 shares the IGF-II binding preference with human IGFBP-6 (22) despite lacking the AGC-PGcC^3VE peptide segment present in human IGFBP-6.
Results of Edman sequencing of the 5275-Da N-terminal disulfide-linked peptide complex (HPLC fraction 16, Fig. 5) isolated as described in Fig. 5 by reverse-phase HPLC of hIGFBP-1 digested with trypsin and \( \text{S. aureus} \) V8 protease. The 5275-Da complex was initially identified as consisting of the four disulfide-linked peptides shown in Table II on the basis of the hIGFBP-1 sequence (Fig. 1), the specificities of trypsin (Arg-C/Lys-C) and \( \text{S. aureus} \) V8 protease (Glu-C/Aeg-C), and the ESMS mass (Table II). Corresponding amino acid signals expected in each Edman cycle are compared with the observed signals from 16 cycles of Edman sequencing carried out as described under "Experimental Procedures." Multiple signals in respective Edman cycles are arranged vertically according to the assigned sequences, therefore some expected or observed signals are repeated for ease of comparison. Dashes (-) represent cysteine (not expected to be observed under Edman sequencing conditions) or expected amino acid signal changes that were not observed. Edman observed signals that are underlined were equivocal due to carryover of the same signal from a previous cycle. C-terminal amino acid washout, or weak signals from serine or in cycles following proline. The Edman sequencing data confirm the ESMS-based identification (Table II).

| Peptide complex assigned from ESMS mass | Expected Edman sequencing result | Observed Edman sequencing result |
|----------------------------------------|---------------------------------|---------------------------------|
| APWQCAPCSAE                            | 1 5 10 15                       | 1 5 10 15                       |
| LACCPVSACSE                             | APWQ-AP-SA                  | APWQ-AP-SA                    |
| SAGGGCCMPCALPGLAGACVATAR                | SAG-G-PM-ALPLGA              | SAG-G-EM-DLP-A                |
| GLS-R                                  | GL-S-R                        | GL-S-R                        |

(8). Notably, the two N-terminal cysteines (Cys\(^3\) and Cys\(^6\)) within this segment are disulfide-linked in human IGFBP-6. It is therefore likely that the other N-terminal linkages in rat IGFBP-6 are the same as those in human IGFBP-6.

The N-terminal disulfide linkages of IGFBP-6 differ substantially from those of IGFBP-1, and, by implication, the other IGFBPs. The N-terminal disulfide linkages of IGFBP-6 are the sequential linkages Cys\(^1\)-Cys\(^2\), Cys\(^3\)-Cys\(^4\), and Cys\(^5\)-Cys\(^6\), whereas in IGFBP-1, Cys\(^1\) is not linked to Cys\(^2\) and Cys\(^3\) is not linked to Cys\(^4\). Although these observations suggest that a different local fold in the N-terminal region is likely, it is possible that this is not the case. For example, the presence of proline-containing sequences in the N-terminal regions of IGFBP-6 could result in a similar fold to that stabilized by the disulfide linkages in IGFBP-6.

Human and rat IGFBP-6 lack the adjacent cysteine (CC) pair that is part of the GC\(^6\)GC\(^6\)C\(^7\) motif present in the N-terminal domains of IGFBP-1–5 (8). Although it is possible that these two adjacent cysteines in IGFBP-6–5 are disulfide-linked, this is unlikely since disulfide linkage of adjacent cysteines requires a relatively rare cis peptide bond between the cysteines (23). Additionally, linkage of these adjacent cysteines would serve no obvious structural purpose, which seems inconsistent with their being conserved in all IGFBPs other than IGFBP-6. It is therefore probable that these cysteines are disulfide-linked to other N-terminal cysteines in IGFBP-6–5.

While further investigation is clearly needed to determine the N-terminal disulfide linkages of IGFBP-1 (and IGFBP-2–5), it is unlikely that this can be accomplished by peptide mapping techniques relying on proteolysis alone, due to lack of suitable cleavage sites in the conserved sequence GC\(^6\)GC\(^6\)C\(^7\) and elsewhere. Indeed, the N-terminal disulfide linkages of IGFBP-3 could not be determined by this approach (12). Appropriate cleavage sites could conceivably be introduced by site-directed mutagenesis. Peptide mapping methods could also be coupled with selective mutagenesis of cysteines in order to determine disulfide linkages one at a time, provided correct formation of remaining disulfide linkages was not affected.

The high affinity IGF binding site of IGFBP-5 has recently been localized to the sequence containing Cys\(^9\) to Cys\(^12\) of that protein (equivalent to Cys\(^7\) to Cys\(^10\) of IGFBP-6) (11)). Since IGFBPs –1, –3, and –6 share the same disulfide linkages in this region as IGFBP-5 and there is substantial homology between the IGFBPs of critical hydrophobic amino acids involved in IGF binding, it is likely that this region confers high IGF binding affinity to all IGFBPs. It was suggested that the N-terminal domain of IGFBP-5 does not significantly interact with the high affinity binding site of IGFBP-5 nor influence IGF binding by this site (11). With IGFBP-6, however, the substantially different disulfide linkages of the N-terminal domain compared with those of IGFBP-1 (and presumably the other IGFBPs) could result in the stabilization of a different N-terminal domain structure which may be responsible for or contribute to the distinctive IGF binding characteristics of IGFBP-6. An alternate possibility is that the structures of the N-terminal domains of IGFBP-6 and other IGFBPs are folded in a similar way that is stabilized by quite different disulfide linkages. In that case, assuming the remainder of the structure also to be similar in IGFBPs 1–6, the IGF-II binding preference of IGFBP-6 could depend more on differences in the primary sequence compared with other IGFBPs than to differences in three dimensional structure, although both are likely to be influential.

It has been claimed that the C-terminal domain of the IGFBPs is also involved in IGF binding (11, 24). Although many studies have suggested that the C-terminal domains of IGFBPs do not independently bind IGFs (11, 25, 26), the presence of this domain substantially increases IGF binding affinity, possibly by stabilizing the IGF-IGFBP complex once the IGF has bound. In contrast, two studies have shown that C-terminal fragments of IGFBP-2 bind IGFs with low affinity (27, 28). Interestingly, deletion of the sequence containing the two C-terminal disulfide linkages eliminated the 12-fold IGF-II binding preference of bovine IGFBP-2 over IGF-1 (9). IGFBP-6 and IGFBP-2 share a preference for IGF-II and have the same three C-terminal disulfide linkages. It has not been demonstrated that these linkages are different in the other IGFBPs, which lack substantial IGF-II binding preference; but, if present, such a difference in the C-terminal disulfide linkages (such as those connected to the conserved CWCV sequence) could conceivably influence IGF-II binding specificity. However, the sequences of IGFBPs 1–6 are homologous with respect to C-terminal cysteines, suggesting that their disulfide linkages are likely to be the same. Unfortunately, the peptide complex containing the four C-terminal cysteines of IGFBP-1, which would provide considerable information on this matter, was not recovered in the present study. In any case, it would be difficult to reconcile the IGF-II binding preference of the IGFBP-6/IGFBP-5 chimera mentioned above (13) with a C-terminal determinant of IGF-II binding preference.

In conclusion, we have determined all of the disulfide linkages of human IGFBP-6. As suggested previously (10), the N-terminal linkages fall into two domains comprising Cys\(^1\) to Cys\(^6\) (the N-terminal domain) and Cys\(^7\) to Cys\(^10\) (the IGF binding domain), respectively. The six C-terminal cysteines of IGFBP-6 form three disulfide bonds, and there are no linkages between the N- and C-terminal regions. The N-terminal linkages differ significantly from those of IGFBP-1 and therefore,
by inference, those of the other IGFBPs. In contrast, the remaining disulfide linkages of IGFBP-6 are the same as those reported in other IGFBPs; all of these linkages are therefore likely to be identical in IGFBPs 1–6. Knowledge of the disulfide linkages of IGFBP-6 will facilitate further structural studies of this protein as well as aiding in the rational design of mutants for structure-function correlations.

Acknowledgments—We thank Joe Marinaro for assistance in expression and purification of IGFBP-6 and Rosemary Condron for performing Edman N-terminal sequencing.

REFERENCES

1. Jones, J. I., and Clemmons, D. R. (1995) Endocr. Rev. 16, 3–34
2. Rechler, M. M. (1993) Vitam. Horm. 47, 1–114
3. Bach, L. A., and Rechler, M. M. (1995) Diabetes Rev. 3, 38–61
4. Bach, L. A., Hsieh, S., Sakano, K., Fujiwara, H., Perdue, J. F., and Rechler, M. M. (1993) J. Biol. Chem. 268, 9246–9254
5. Martin, J. J., Willetts, K. E., and Baxter, R. C. (1990) J. Biol. Chem. 265, 4124–4130
6. Forbes, B., Ballard, F. J., and Wallace, J. C. (1990) J. Endocrinol. 126, 497–506
7. Boghiani, M., Lassarre, C., Zapf, J., Povoa, G., and Binoux, M. (1991) J. Clin. Endocrinol. Metab. 73, 658–666
8. Shimazaki, S., Gao, L., Shimokawa, M., and Ling, N. (1991) Mol. Endocrinol. 5, 938–948
9. Forbes, B. E., Turner, D., Hodge, S. J., McNeil, K. A., Forsberg, G., and Wallace, J. C. (1998) J. Biol. Chem. 273, 4647–4652
10. Neumann, G. M., Marinaro, J. A., and Bach, L. A. (1988) Biochemistry 37, 6572–6585
11. Kalus, W., Zweckstetter, M., Renner, C., Sanchez, Y., Georgescu, J., Crol, M., Demuth, D., Schumacher, R., Dony, C., Lang, K., and Holak, T. A. (1998) EMBO J. 17, 6558–6572
12. Hashimoto, R., Ono, M., Fujiwara, H., Higashihashi, N., Yoshida, M., Enjoj-Klimura, T., and Sakano, K. (1997) J. Biol. Chem. 272, 27936–27942
13. Twigg, S. M., Kiefer, M. C., Zapf, J., and Baxter, R. C. (1998) J. Biol. Chem. 273, 26791–26798
14. Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J., and Udseth, H. R. (1990) Anal. Chem. 62, 882–889
15. Boisclair, Y. R., Yang, W. Y.-H., Stewart, J. M., and Rechler, M. M. (1994) Growth Regul. 4, 136–146
16. Bach, L. A., Hsieh, S., Brown, A. L., and Rechler, M. M. (1994) Endocrinology 135, 2168–2176
17. Bach, L. A., Thetakura, N. R., and Rechler, M. M. (1992) Biochem. Biophys. Res. Commun. 186, 301–307
18. Neumann, G. M., Condron, R., and Povoa, G. M. (1996) Biochim. Biophys. Acta 1298, 223–240
19. Zapf, J., Kiefer, M., Merryweather, J., Masiarz, F., Bauer, D., Born, W., Fischer, J. A., and Froesch, E. R. (1990) J. Biol. Chem. 265, 14892–14898
20. Kato, M., Ishizaki, A., Hellman, U., Wernstedt, C., Kyogoku, M., Miyazono, K., Heldin, C.-H., and Funah, K. (1995) J. Biol. Chem. 270, 12373–12379
21. Bean, M. F., and Carr, S. A. (1992) Anal. Biochem. 201, 216–226
22. Bach, L. A., Tseng, L. Y.-H., Swartz, J. E., and Rechler, M. M. (1993) Endocrinology 133, 990–995
23. Zhang, R. M., and Snyder, G. H. (1989) J. Biol. Chem. 264, 18472–18479
24. Spencer, E. M., and Chan, K. (1995) Prog. Growth Factor Res. 6, 209–214
25. Lalou, C., Lassarre, C., and Binoux, M. (1996) Endocrinology 137, 3206–3212
26. Qin, X. Z., Strong, D. D., Baylink, D. J., and Mohan, S. (1998) J. Biol. Chem. 273, 23509–23516
27. Wang, J.-F., Hampton, B., Mehlmam, T., Burgess, W. H., and Rechler, M. M. (1988) Biochem. Biophys. Res. Commun. 157, 718–726
28. Ho, P. J., and Baxter, R. C. (1997) Endocrinology 138, 3811–3818