Articles

Hybrid and Complex Glycans Are Linked to the Conserved N-Glycosylation Site of the Third Eight-Cysteine Domain of LTBP-1 in Insect Cells†

Pauline M. Rudd,‡ A. Kristina Downing,§ Martine Cadene,‖,# David J. Harvey,‡ Mark R. Wormald,‡ Ian Weir,⊥ Raymond A. Dwek,‡ Daniel B. Rifkin,⊥ and Pierre-Emmanuel Gleizes*,⊥,⊗

Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK, Department of Biochemistry and Oxford Centre for Molecular Sciences, University of Oxford, South Parks Road, Oxford OX1 3QT UK, Department of Pharmacology, New York University Medical Center, 550 First Avenue, New York, New York 10016, USA, Department of Cell Biology and Kaplan Cancer Center, New York University Medical Center, 550 First Avenue, New York, New York 10016, USA, and Laboratoire de Biologie Moléculaire Eucaryote du CNRS (UPR 9006) and Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex, France

Received August 5, 1999; Revised Manuscript Received October 29, 1999

ABSTRACT: Covalent association of LTBP-1 (latent TGF-β binding protein-1) to latent TGF-β is mediated by the third eight-cysteine (also referred to as TB) module of LTBP-1, a domain designated as CR3. *Spodoptera frugiperda* (Sf9) cells have proved a suitable cell system in which to study this association and to produce recombinant CR3, and we show here that another lepidopteran cell line, *Trichoplusia ni* TN-5B1−4 (High-Five) cells, allows the recovery of large amounts of functional recombinant CR3. CR3 contains an N-glycosylation site, which is conserved in all forms of LTBP known to date. When we examined the status of this N-glycosylation using MALDI-TOF mass spectrometry and enzymatic analysis, we found that CR3 is one of the rare recombinant peptides modified with complex glycans in insect cells. Sf9 cells mainly processed the fucosylated paucomannosidic structure $(GlcNAc)_2(Mannose)_3Fucose$, although hybrid and complex N-glycosylations were also detected. In High-Five cells, the peptide was found to be modified with a wide variety of hybrid and complex sugars in addition to paucomanosidic oligosaccharides. Most glycans had one or two fucose residues bound through α1,3 and α1,6 linkages to the innermost GlcNAc. On the basis of these results and on the structure of an eight-cysteine domain from fibrillin-1, we present a model of glycosylated CR3 and discuss the role of glycosylation in eight-cysteine domain protein−protein interactions.

Latent TGF-β binding protein-1 (LTBP-1) is the prototypic member of a family of four related proteins, which are potentially involved in the regulation of bioavailability and activity of transforming growth factor-β (TGF-β). LTBP-1 was first characterized as part of the large molecular weight complex containing latent TGF-β1 (*1*, *2*), one of the three forms of TGF-β known in mammals. TGF-β is a cytokine involved in an extensive range of biological processes, including development, wound healing, blood vessel homeostasis, immune system regulation, bone metabolism, fibrosis, and tumor growth (*3*). Unlike many cytokines, TGF-β is secreted as a latent, inactive complex in which the mature growth factor remains linked through noncovalent interactions to the N-terminal part of its propeptide, called LAP (*4*). In the secretion pathway, LAP associates covalently via disulfide bonds with LTBP-1 (Figure 1A), which targets the complex to the extracellular matrix (ECM) after secretion (*5*). Incorporation into the ECM may be a means of storing the cytokine in the intercellular space as well as the initial step in the regulation of the release of the active cytokine from the complex, a process called latent TGF-β activation.

All four LTBPs exhibit a modular organization comprised of stretches of epidermal growth factor (EGF)-like domains interrupted by eight-cysteine (or TB) domains. Similar organizations are found in fibrillin-1 and -2, which are major components of elastic fibers, and whose genes have been

† This work was supported by NIH grant CA 23753 (D.B.R.). The work in the Oxford Glycobiology Institute (GBI) was supported by the Biotechnology and Biochemistry Scientific Research Council, the European Commission (Grant BIO4-CT95-0138), and Oxford GlycoSciences Ltd. A.K.D. thanks the support of the Wellcome Trust.

* Corresponding author: Pierre-Emmanuel Gleizes, Laboratoire de Biologie Moléculaire Eucaryote du CNRS, 118 route de Narbonne, 31062 Toulouse cedex, France. Telephone: 33-561-33-59-26. Fax: 33-561-33-58-86. E-mail:gleizes@ibcg.biotoul.fr.

‡ Department of Biochemistry, University of Oxford.

§ Oxford Centre for Molecular Science, University of Oxford.

‖ Department of Pharmacology, New York University Medical Center.

Present address: Rockefeller University, 1230 York Ave, New York, New York 10021.

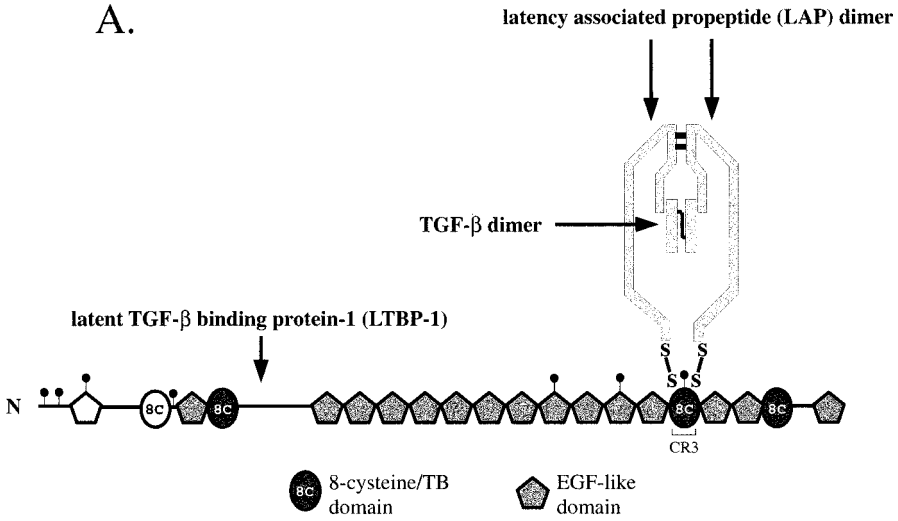
⊥ Department of Cell Biology, New York University Medical Center.

⊗ Laboratoire de Biologie Moléculaire Eucaryote du CNRS.

[1] Abbreviations: TGF-β, transforming growth factor-β; LTBP, latent TGF-β binding protein; MALDI-TOF MS: matrix-assisted laser desorption ionisation time-of-flight mass spectrometry; PCR, polymerase chain reaction; ECM, extracellular matrix; EGF, epidermal growth factor.

10.1021/bi9918285 CCC: $19.00
Published on Web 01/29/2000

A.



B.

```
hLTBP1    CDNVLAPNVTKQECCCTSGAGWGDNCEIFPC
mLTBP1    CDNVLAPNVTKQECCCTSGAGWGDNCEIFPC
rLTBP1    CDNVLAPNVTKQECCCTSGAGWGDNCEIFPC
bLTBP2    CSSLLGRNTTQAECCCTQGTGWGDACDL--C
hLTBP2    CSSVLGRNTTQAECCCTQGATWGDACDL--C
rLTBP2    CSQILGQNSTQAECCSTQGARWGETCDP--C
mLTBP3    CDSVLATNVTQQECCCSLGAGWGDHCEIYPC
hLTBP4    CDNILARNVTWQECCCTVGEGWGSGCRIQQC
```

FIGURE 1: Schematic illustration of the latent TGF-β complex highlighting the interaction of CR3 with LAP. (A) N-glycosylation sites in LTBP-1 are marked by lollipops. The lollipops are designed to show which domains in LTBP-1 are glycosylated, not the exact glycosylation sites. LAP also contains three glycosylation sites at Asn 82, 136, and 176. (B) Partial sequence of the third eight-cysteine domain showing the conserved N-glycosylation site (b, bovine; h, human; m, mouse; r, rat).

linked to congenital pathologies such as Marfan syndrome and congenital contractural arachnodactyly, respectively (*6*). Like fibrillins, LTBPs are secreted and incorporated into the ECM (*5*, *7*−*11*), and LTBP-2 is present in elastin-associated microfibrils (*8*). Together with their role in the regulation of active TGF-β formation, LTBPs may also contribute to the architecture of microfibrils and may target TGF-β to specific regions of the ECM.

Eight-cysteine domains have been found to date in LTBPs and fibrillins. They are characterized by the presence of eight cysteines forming four intramolecular disulfide bonds, which stabilize a globular structure, and their sequences are rather divergent. The function of these domains in fibrillins and LTBPs is poorly understood, but a number of mutations in fibrillin-1 linked to Marfan syndrome are found in these repeats (*12*−*14*). Some evidence points toward a role for these modules in protein−protein interactions. Cell adhesion to fibrillin-1 and -2 is mediated in part by their fourth eight-cysteine domain, which contains an RGD sequence recognized by the integrin $\alpha v\beta 3$ (*15*, *16*). Also, incorporation of LTBP-1 into the ECM requires an N-terminal region consisting of the second eight-cysteine domain adjacent to an EGF-like domain (*17*). Most strikingly, the third eight-cysteine domain (CR3) in LTBP-1 is necessary and sufficient for covalent bonding of LTBP-1 to LAP (*18*, *19*), as illustrated in Figure 1A. The corresponding module in other LTBPs is assumed to play the same role, as LTBP-3 (*20*) and LTBP-4 (*11*) form complexes with latent TGF-β.

The molecular basis for CR3 functional specificity in LTBP-1 is not understood. When not associated with LAP, the eight cysteines in the CR3 domain are engaged in intramolecular disulfide bonds (*19*), as observed in other eight-cysteine domains (*21*, *22*). Thus, a conformational change of the CR3 domain is probably needed for association with LAP. Interestingly, only the third eight-cysteine domain in every LTBP contains an asparagine in a context for N-glycosylation (N-X-S/T) (Figure 1B). Glycosylation is not strictly required for association of CR3 to latent TGF-β, as demonstrated by site-directed mutagenesis (*18*). However, conservation of this N-glycosylation site in various mammalian species (human, mouse, rat, and cow, see Figure 1B) suggests that it may contribute to either the function or the structure of LTBPs. Recently, the structure of an eight-cysteine domain from fibrillin-1 was solved by NMR, providing the first direct insight into the folding of these motifs (*22*). It is not clear from the structure how and why CR3 associates with LAP. More structural and mechanistic data are required to understand this interaction.

We have previously studied the association of CR3 and latent TGF-β in *Spodoptera frugiperda* (Sf9) cells and used these cells to produce the CR3 domain for structural studies (*19*). In Sf9 cells, CR3 associates efficiently with latent TGF-β1 when the two proteins are coexpressed. In addition, we have found that the lepidopteran cell line from *Trichoplusia ni*, TN-5B1-4 or High-Five cells, allows production of large amounts of recombinant CR3 (this work) as well as full-length LTBP-1 (Mazzieri, R., Ushiro, S., Gleizes, P. E., and Rifkin, D. B. unpublished results). With very few exceptions, most recombinant glycoproteins produced in insect cells have been found to be modified with oligomannose type glycans,

whereas endogenous glycoproteins can bear complex sugars. However, few data describing N-glycosylation in High-Five cells are available, which makes it difficult to predict the structure of the oligosaccharides added to CR3 or LTBP-1 produced in this cell line.

In the present study, we have characterized the glycans attached to the conserved N-glycosylation site in CR3 produced in Sf9 and High-Five cells. The CR3 glycoforms and the N-glycans released from CR3 expressed in High-Five cells were analyzed by MALDI-TOF and HPLC. The glycans include hybrid and complex glycans, some of which have not been previously described in these cells. On the basis of protein structural data, the glycan analysis, and the Oxford Glycobiology Institute's oligosaccharide database, we present a model of glycosylated CR3 and discuss the possible roles of glycosylation in eight-cysteine domain protein−protein interactions.

MATERIALS AND METHODS

Expression and Purification of the CR3 Peptide. The region encoding human LTBP-1 CR3 (fragment 3028−3271 in the human LTBP-1S coding sequence) was amplified by polymerase chain reaction (PCR) from human LTBP-1S cDNA and inserted into the baculovirus transfer vector pAcGP67-His, a derivative of pAcGP67 (Pharmingen, San Diego, USA), as previously described (*19*). In pAcGP67-His, the fragment was subcloned in-frame downstream of a sequence encoding a six-histidine tag followed by a thrombin cleavage-site. A recombinant baculovirus was generated with linearized AcNPV DNA using the BaculoGold transfection kit (Pharmingen) according to the manufacturer. For production of the peptide, *T. ni* High-Five cells were seeded in 15 cm dishes at a density of 15×10^6 cells/dish in 401 serum-free medium and infected 20−24 h later. The conditioned medium was collected 60 h after infection and dialyzed against TBS (150 mM NaCl, 20 mM Tris, pH 7.4, 25 °C). The medium was passed through a chromatography column filled with TALON chelating resin (Clontech, Palo Alto, U.S.A.) equilibrated with TBS. After the column was washed with TBS and TBS + 5mM imidazole, the peptide was eluted with TBS containing 50 mM imidazole. A second dialysis against TBS was performed to eliminate imidazole, and the six-histidine tag was removed by digestion with thrombin. The cleaved N-terminal fragment containing the tag was removed from the solution by absorption on TALON chelating resin. The peptide was left with 10 extra amino acids (GSPGLDGMIP) at the N-terminus.

Release and re-N-acetylation of Glycans. A 50 μg sample of protein was dialyzed against 0.1% TFA and lyophilized. Glycans were released by hydrazine at 85 °C and re-N-acetylated using a GlycoPrep 1000 (Oxford GlycoSciences—OGS—Ltd, Abingdon, Oxon., OX14 1RG, UK).

Fluorescent Labeling of the Reducing Terminus with 2-Aminobenzamide (2AB). The pool of released glycans was evaporated to dryness using a vacuum centrifuge. 2AB labeling was carried out by reductive amination using the OGS Ltd. Signal labeling kit (*23*).

Simultaneous Oligosaccharide Sequencing on the Released Glycan Pool. Enzyme digests were performed at 37 °C for 16−24 h in 100 mM citrate/phosphate buffer pH 4.5 containing 0.2 mM zinc acetate and 0.15 M sodium chloride. Conditions for the individual enzymes in the arrays were as follows: almond meal fucosidase (AMF), 3 mU/ml; bovine epididymis α-fucosidase (BEF) (Oxford Glycobiology Institute), 0.2 U/mL/1 mg/mL BSA; *Streptococcus pneumoniae* β-galactosidase (SPG) (OGS Ltd.), 2 U/mL/1 mg/mL bovine serum albumine (BSA); *Streptococcus pneumoniae* hexosaminidase (SPH), 2 U/mL. Sugars were purified from protein and salts prior to matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) by passing through mixed bed resins of $AG3X_4(OH^-)$ (bottom)/$AG50X12(H^+)$ (middle)/C18 (top). Samples were purified from the exoglycosidases before high-performance liquid chromatography (HPLC) analysis by passing through a microcentrigfuge tube inset with a protein-binding filter (Microspin 45 mm CN, Pro-Mem, suppliers R. B. Radley and Co Ltd., Shire Hill, Saffron Walden, Essex). The filter was washed with 15 mL of 5% acetonitrile, and the washings were combined with the remainder of the sample.

Normal Phase Separations of Neutral and Acidic Oligosaccharides. Normal phase separations were performed on a Glycosep-N chromatography column (OGS Ltd.) and structures assigned according to Guile et al. (1996) (*24*).

Matrix-assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS). Protein samples were prepared for MALDI-TOF MS using the dried droplet method. The matrix used was α-cyano-4-hydroxycinnamic acid (Sigma, Saint Louis, USA), which was purified by recrystallization (*25*). To produce the dried droplets, a saturated solution of matrix was prepared in 2:1 aqueous 0.1% TFA:acetonitrile at room temperature. The sample was added to this solution so that the final sample concentration was 1−10 μM. One-half microliter of the solution was placed on the mass spectrometer's probe and allowed to dry. The linear, time-of-flight mass spectrometer was custom built at the Skirball Institute at the New York University Medical Center.

Oligosaccharides were analyzed on a Perceptive Biosystems Voyager Elite time-of-flight mass spectrometer. Samples were prepared by adding the oligosaccharide sample (0.5 μL) in water to the matrix solution (0.5 μL of a saturated solution of 2,5-dihydroxybenzoic acid (2,5-DHB) in acetonitrile) on the mass spectrometer target and allowing it to dry at room temperature. The mixture was re-crystallized from 1 μL of ethanol (*26*).

Molecular Modeling. CR3 was modeled according to Yuan et al. (*22*). Oligosaccharide structures were built using the database of disaccharide linkage conformations based on crystallographic data from saccharides, glycoproteins, and lectins (*27*). The resulting structures were energy minimized to eliminate unfavorable steric interactions, using the modified AMBER force field within the Discover II software (MSI Inc.). The structure of the peptide−glycan linkage was based on the NMR results from glycopeptide studies (*28*). The torsion angles around the Asn Cα−Cβ and Cβ−Cγ bonds were adjusted to eliminate unfavorable steric interactions between the oligosaccharide and the protein surface. This leads to a significant range of allowed values for these two torsion angles.

RESULTS

Production of the CR3 Peptide in High-Five Cells. High-Five cells were infected with a recombinant baculovirus

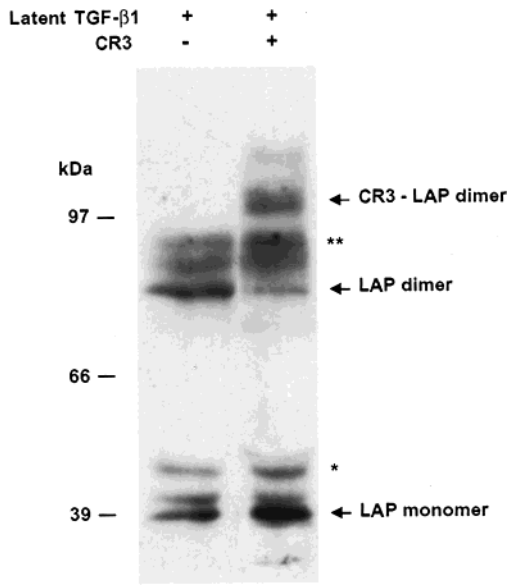


FIGURE 2: CR3 produced in High-Five cells associates with latent TGF-β; Western-blot analysis, using anti-β1LAP monoclonal antibody VB3A9, of conditioned medium from High-Five cells 60 h after infection with a baculovirus encoding latent-TGF-β1 alone or together with CR3. In addition to LAP (monomers and dimers) and the LAP−CR3 complex, unprocessed proTGF-β can be detected due to overexpression (asterisk, monomer; double asterisk, dimer).

encoding CR3. The cDNA was fused to a signal sequence for secretion and to a six-histidine tag for simple purification. Transcription was driven by the polyhedrin promoter. Purification from the conditioned medium yielded 10−15 mg of peptide per 10^9 cells (1 L of medium), which was 5−10-fold higher than in Sf9 cells. After purification, the histidine tag was cleaved with thrombin. As judged by MALDI-TOF MS analysis, the purified peptide was insensitive to alkylating agents under nonreducing conditions, indicating that the eight cysteines were involved in disulfide bonds (not shown). In these cells, as in Sf9 cells, coexpression of CR3 and latent TGF-β1 resulted in the formation of a complex (Figure 2). Therefore, High-Five cells provide an efficient expression system for production of functional CR3.

MALDI-TOF MS Analysis of the CR3 Peptide Produced in Sf9 and in High-Five Cells. The CR3 peptide produced in Sf9 cells or in High-Five cells was analyzed by MALDI-TOF mass spectrometry. The spectra displayed different patterns, depending on the producing cell (Figure 3A vs 3B). The main mass shift for the Sf9-derived peptide corresponded to a paucomannosidic structure containing fucose (H_3N_2F). Minor peaks were consistent with fucosylated hybrid and complex glycans (H_3N_3F, H_3N_4F). These fucosylated residues were all sensitive to PNGase-F, indicating that they were linked α1,6 to the glycan. The CR3 peptide produced in High-Five cells showed a more complex glycosylation pattern (Table 1). The paucomannosidic structures found in CR3 from Sf9 cells were still detectable (Figure 3B, peaks b, and c) and corresponded to major peaks. In addition, higher molecular weight residues were also detected (peaks f−q), with masses consistent with the presence of hybrid glycans such as H_3N_3F (peak f), or complex sugars such as H_3N_4F (peak i) or H_3N_5F (peak l). Most of the fucosylated glycoforms were resistant to treatment with PNGase-F,

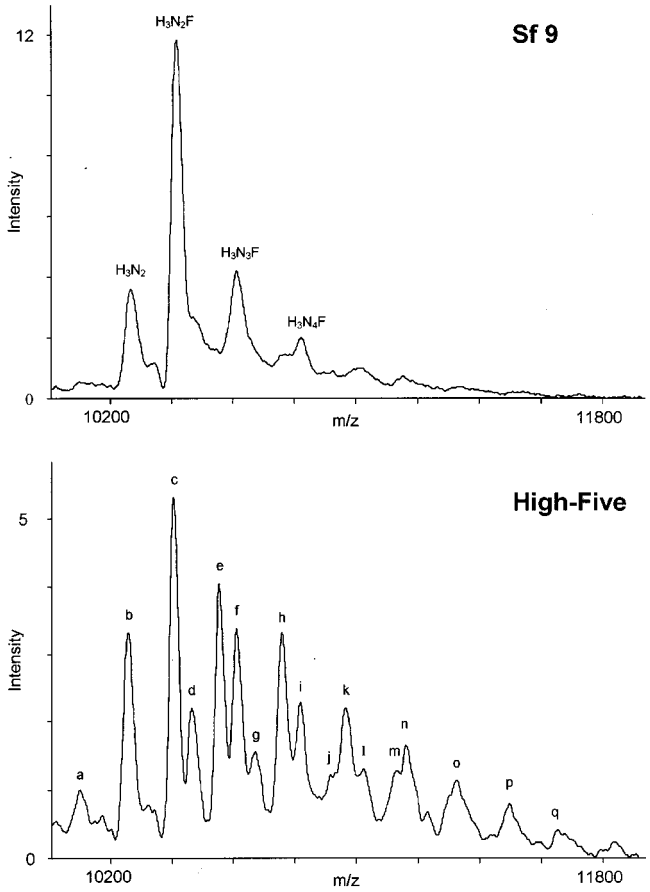


FIGURE 3: MALDI-TOF MS analysis of CR3 produced in Sf9 cells or in High-Five cells. CR3 purified from Sf9 or High-Five cells conditioned medium was directly analyzed by MALDI-TOF MS. The putative composition of the glycans present on CR3 produced in High-Five cells (peaks a−q) is presented in Table 1. H, hexose; N, *N*-acetylhexosamine; F, fucose.

suggesting the presence of α1,3 core fucose residues. Several peaks had masses consistent with the presence of difucosylated glycans (peaks e, h, j, m, and n), consistent with branching of both α1,3- and α1,6- fucose on the innermost GlcNAc.

Identification of the Structure of the Oligosaccharides by HPLC and MALDI-MS. Glycans present on the CR3 peptide produced in High-Five cells were further characterized by HPLC (Figure 4). The glycans were released from the peptide by hydrazine treatment and derivatized with 2-aminobenzamide. The HPLC elution profile was compared to a standard dextran hydrolysate ladder, and a predictive structure was assigned to the major peaks. Preliminary assignments were confirmed by exoglycosidase digestions of the entire glycan pool using enzyme arrays (*29*). As shown in Table 2, hybrid and complex sugars were identified by this technique. Furthermore, these data confirmed the presence of α1,6 and α1,3 core fucose residues as well as glycans with two fucose residues attached to the GlcNAc at the reducing N-terminal. Characterization of the glycans released from CR3 was also performed by MALDI-TOF MS analysis (Table 3). Peaks with compositions corresponding to most of the sugars identified by HPLC were detected. Eleven out of the thirteen glycans observed by MALDI-TOF MS were related to one of CR3 glycopeptides reported in Table 1. The peak J, containing a glycan with a difucosylated core, was analyzed by MALDI-TOF MS using post source decay. The fragment

Table 1: Proposed Structures of the Glycans Born by CR3 Glycopeptides Observed by MALDI-TOF MS Analysis of CR3 Expressed in High-Five Cells

peak designation[a]	calculated glycan mass[b]	theoretical mass	composition[c]
a	733	730.7	H_2N_2[d]
b	892	892.8	H_3N_2[d,e]
c	1040	1039.0	H_3N_2F[d,e]
d	1098	1096.0	H_3N_3[d,e]
e	1189	1185.1	$H_3N_2F_2$[d,e]
f	1245	1242.2	H_3N_3F[d,e]
g	1305	1299.2	H_3N_4[d,e]
h	1391	1388.3	$H_3N_3F_2$[d,e]
i	1447	1445.4	H_3N_4F[d]
j	1556	1550.4	$H_4N_3F_2$[d,e]
k	1598	1591.5	$H_3N_4F_2$[d,e]
		1607.5	H_4N_4F
l	1654	1648.5	H_3N_5F
m	1759	1753.6	$H_4N_4F_2$
n	1801	1794.7	$H_3N_5F_2$
o	1966	1972.8	H_5N_5F
p	2131	2135.0	H_6N_5F
q	2298	2297.1	H_7N_5F

[a] The nomenclature of the peaks refers to Figure 3. [b] The calculated mass corresponds to the difference of mass with the nonglycosylated peptide. [c] H = hexose (mannose or galactose), N = GlcNAc, F = deoxyhexose (fucose). [d] Also detected by HPLC analysis of the released glycan pool (Table 2). [e] Also detected by MALDI-TOF MS analysis of the released glycan pool (Table 3).

ions (data not shown) were those predicted for the structure shown in Table 2. These data are all consistent with the presence of hybrid and complex sugars on recombinant CR3 produced in High-Five cells.

DISCUSSION

High levels of the third eight-cysteine module of LTBP-1 (CR3) were produced in *T. ni* High-Five cells. All eight cysteines were shown to be involved in intramolecular disulfide bonds, indicating correct folding of the peptide. MALDI-TOF MS analysis of CR3 showed that the peptide was posttranslationally modified with N-glycans by both Sf9 and High-Five cells. The molecular masses found for CR3 glycoforms were compatible with the presence of complex glycans in the protein produced in both cell types, High-Five cells processing the larger variety of complex and hybrid oligosaccharides. The presence of complex and hybrid glycans in CR3 from High-Five cells was confirmed by HPLC fractionation and exoglycosidase digestion of the released glycans. Nineteen major species were identified by HPLC analysis of the released glycans. This technique can resolve oligosaccharides that have the same composition, yielding information relating to linkage and arm specificity and, therefore, complements the composition data obtained by MS. We detected the presence of seven different oligosaccharides terminating with *N*-acetylhexosamine and three oligosaccharides terminating with galactose. There was no evidence of sialylation in the HPLC data. The glycans were released by hydrazinolysis optimized for the release and recovery of sialylated glycans. However, although sialylated glycans were recovered from other glycoprotein samples processed at the same time, it remains a possibility that some sialic acid residues were lost during the handling of this sample. MALDI-TOF MS examination of the released glycan pool revealed thirteen structures with distinct masses.

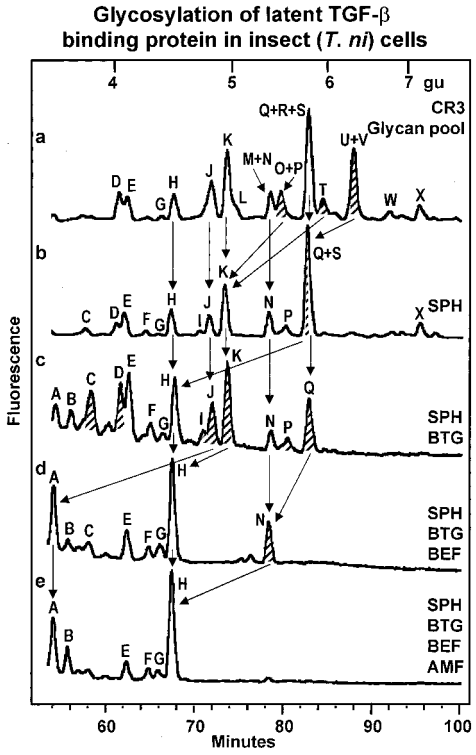


FIGURE 4: Exoglycosidase digestions of glycan pool released from CR3. Glycans released from CR3 by hydrazine treatment were derivatized with 2-aminobenzamide and submitted to sequential digestions with exoglycosidases. The structures of the different glycans detected by this method are listed in Table 2. SPH, *Streptococcus pneumoniae* hexosaminidase; BTG, bovine testes β-galactosidase; BEF, bovine epididymis α-fucosidase; AMF, almond meal fucosidase.

Twelve of these residues had masses expected from sixteen of the glycans identified by HPLC, including the complex sugars. These results, obtained by analyzing the isolated glycans, fitted the data obtained by MALDI-TOF MS analysis of the glycopeptides. MS examination of the glycopeptides also showed peaks with masses corresponding to the presence of oligosaccharides with five to six *N*-acetylhexosamines and five to seven hexoses, indicating that the peptide bore glycans with a higher number of residues than those identified by HPLC. However, these species represented a minor fraction of the total sugars and may have been below HPLC detection threshold.

Most recombinant glycoproteins synthesized in insect cells with baculovirus vectors contain mainly high mannose-type and short-truncated oligosaccharides with core fucosylation (*30*). CR3 is one of the few recombinant proteins reported to be modified with hybrid and complex N-glycosylations. Interleukin-3 produced in *Bombix mori* larvae was found to bear hybrid structures with one terminal GlcNAc, whereas hybrid and complex glycans containing galactose were detected in IgG from High-Five cells (*31*) and interferon-gamma from *Estigmena acrea* (EA-4) cells (*32*). Sialic acid was reported to be present on recombinant plasminogen when produced in several insect cell lines (*33, 34*), and in secreted alkaline phosphatase from *T. ni* larvae (*35*), although the latter result was later questioned by its authors (*30*). In contrast with the situation observed in most recombinant proteins, N-glycosylations observed in endogenous insect glycoproteins, like honey bee venom phospholipase A2 and hyaluronidase (*36, 37*), apolipophorin III from *Locusta*

Table 2: Oligosaccharides Identified By HPLC Analysis and Exoglycosidase Digestions of the Released Glycan Pool

number	peak[a]	assignement	proposed structure	% non-digested peaks		Digests a	b	c	d
1	A	H2N2					3,46	3,46	3,46
2	B	na					3,56	3,56	3,56
3	C	na				3,71	3,71	3,71	
4	D	H2N2F(α1,6)		5,35	3,97	3,97	3,97		
5	E	na		4,29	4,02	4,02	4,02	4,02	
6	F	na				4,18	4,18	4,18	4,18
7	G	H2N2F(α1,3)					4,29	4,29	
8	H	H3N2		0,98	4,40	4,40	4,40	4,40	4,40
9	I	na				4,67	4,67		
10	J	H2N2F2		5,17	4,76	4,76	4,76		
11	K	H3N2F(α1,6)		10,32	4,94	4,94	4,94		
12	L	H3N3		15,59	5,00				
13	M	H3N3F(α1,6)		2,62	5,32				
14	N	H3N2F(α1,3)		2,61	5,32	5,32	5,32	5,32	
15	O	H3N4		4,81	5,43				
16	P	H3N3F		2,41	5,43	5,43	5,43		
17	Q	H3N2F2		6,72	5,74	5,74	5,74		
18	R	H3N3F(α1,3)		6,72	5,74				
19	S	H4N3		6,72	5,74	5,74			
20	T	H3N4F(α1,6)		5,01	5,93				
21	U	H3N3F2		7,28	6,28				
22	V	H3N4F(α1,3)		7,28	6,28				
23	W	H3N4F2		2,50	6,72				
24	X	H4N3F2		3,62	7,12	7,12			

[1] Nomenclature of the peaks refers to figure 4.

DIGESTS
- **a** no-digest
- **b** *Streptococcus pneumoniae* β-N-acetylhexosaminidase
- **c** *Streptococcus pneumoniae* β-N-Acetylhexosaminidase + bovine testes β-galactosidase
- **d** *Streptococcus pneumoniae* β-N-acetylhexosaminidase + bovine testes β-galactosidase + bovine epididymis α-fucosidase
- **e** *Streptococcus pneumoniae* β-N-acetylhexosaminidase + bovine testes β-galactosidase + bovine epididymis α-fucosidase + almond meal α-fucosidase

KEY
- ▲ galactose
- ■ N-acetylglucosamine
- ● mannose
- ★ fucose

migratoria (*38*), or membrane proteins (*39*), include glycans with terminal GlcNAc and galactose residues. This correlates with the identification in insect cells of enzymatic activities involved in the processing of complex sugars, like α-mannosidase I and II (*40*−*45*), *N*-acetylglucos-aminyltransferase I and II activities (*46*, *47*), and *β*-1,4-*N*-acetylgalactosaminyltransferase (*48*). As pointed out, baculovirus infection can perturb these enzymatic activities (*33*, *43*, *47*−*49*), which may explain why most recombinant proteins are modified with high mannose sugars. In addition, our results, as well as those of others (*32*), clearly show that N-glycosylation of the same peptide can vary depending on the insect cell line.

Table 3: Oligosaccharides Identified by MALDI-TOF MS Analysis of the Released Glycans

observed glycan mass[a]	theoretical mass	composition[b]
917.3	917.3	$H_2N_2F^c$
933.3	933.3	$H_3N_2^{c,d}$
1063.3	1063.4	$H_2N_2F_2^{c,d}$
1079.3	1079.4	$H_3N_2F^{c,d}$
1120.4	1120.4	H_2N_3F
1136.4	1136.4	$H_3N_3^{c,d}$
1225.4	1225.4	$H_3N_2F_2^{c,d}$
1282.4	1282.5	$H_3N_3F^{c,d}$
1298.5	1298.5	$H_4N_3^d$
1339.4	1339.5	$H_3N_4^{c,d}$
1428.4	1428.5	$H_3N_3F_2^{c,d}$
1590.5	1590.6	$H_4N_3F_2^{c,d}$
1631.6	1631.6	$H_3N_4F_2^{c,d}$

[a] Monoisotopic mass of the MNa^+ ion. [b] H = hexose (mannose or galactose), N = GlcNAc, F = deoxyhexose (fucose). [c] Also detected by MALDI-TOF MS analysis of CR3 glycopeptides (Table 1). [d] Also detected by HPLC analysis of the released glycan pool (Table 2).

A majority of the glycans in CR3 had a fucose residue linked to the innermost GlcNAc, either through α1,6 or α1,3 linkages. Most notably, some of the major oligosaccharides in High-Five cells were difucosylated, like $H_3N_2F_2$ or $H_3N_3F_2$ (Table 2, peaks Q, and U). Fucosyltransferase activities have previously been characterized in lepidopteran cells. *Mamestrae bressica* cells have been found to contain both α1,3- and α1,6-fucosyltransferase activities (*50*), which also seems to be the case in High-Five cells. In contrast, Sf9 cell extracts contain exclusively α1,6-fucosyltransferase activity (*50*). Indeed, the fucosylated oligosaccharides born by the CR3 peptide produced in Sf9 cells were sensitive to PNGase-F, as expected if only core α1,6 fucose is present. As a practical consequence, Sf9 cells should be used when large quantities of deglycosylated protein are desired, since the presence of α1,3-linked fucose hampers one-step enzymatic deglycosylation of the peptide with PNGase-F. Core α1,3 fucose has been found in insect membrane glycoproteins (*39*), and in honey bee venom (*36*, *37*). Core α1,3 fucose is also present in a wide variety of plant extracts (*51*). It is essential for recognition of the phospholipase A2 from honey bee venom by IgG and IgE (*52*) as well as for the antigenicity of N-linked carbohydrates in plants (*51*), for example, in the Bermuda grass allergen BG60 (*53*). Core α1,3 fucose may thus be an important part of plant and animal allergens (*51*, *52*). Recently, Hsu et al. (*31*) reported the first example of a heterologous protein (IgG produced in High-Five cells) containing this type of sugar. These authors have identified a paucomannosidic oligosaccharide bearing two core fucose resides. Our data extend this result by showing that in addition to paucomannosidic glycans, at least hybrid (and maybe complex) oligosaccharides also contain core α1,3 fucose (Table 2, peaks U, W, and X; see also Table 1, peaks h, j, m, and n). The presence of core α1,3 fucose does not require core α1,6 fucose (Table 2, peak C), although the

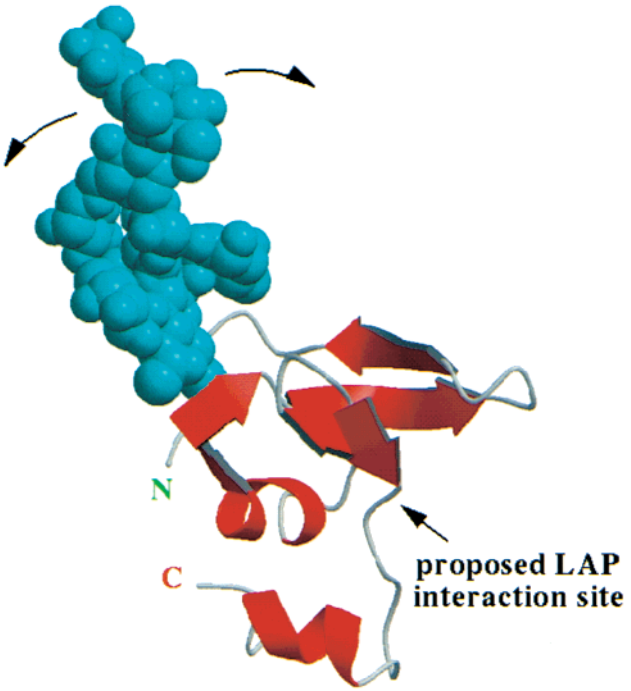


FIGURE 5: Molecular model of one glycoform of CR3 from LTBP-1 with an $H_3N_3F_2$ glycan. This model corresponds to N-glycosylated CR3 containing a di-core-fucosylated biantenary glycan (Table 2, peak U). Conformational flexibility of the carbohydrate moiety may direct ligand binding to the proposed LAP interaction site, or may mediate pairwise domain interactions.

HPLC analysis suggests that most of the oligosaccharides containing α1,3 fucose are difucosylated, as previously seen in other lepidopteran cells (*36*, *50*).

Molecular Modeling. Mutation of the N-glycosylation site in CR3 previously showed that the presence of a glycan is not absolutely required for the association of CR3 and latent TGF-*β*, at least when both proteins are overexpressed (*18*). All sugars can impart a similar range of physical properties to the proteins to which they are attached. In the endoplasmic reticulum (ER), these may include increasing solubility, shielding the protein surface with a network of glycans and organized water, which may prevent aggregation, and orienting anchored glycoproteins in the membrane of the ER. In addition, any sugar may be able to modulate sterically the interactions of the protein if it is located close enough to the binding site (*54*). There may also be roles for specific oligosaccharide motifs, which can impart particular functions to the proteins that contain them. For example, all proteins using calnexin and calreticulin-folding pathways require the Glcα1−3Man motif, turn-over rates may be determined by glycosylation, and recognition of the sialyl Lewis X ligand plays a role in cell adhesion (for review, see reference *55*). The perfect conservation of this site among LTBPs strongly suggests that this posttranslational modification is important to the structure and/or to the function of these proteins. A model of N-glycosylated CR3 containing a di-core-fucosylated biantenary glycan is presented in Figure 5. The size of the carbohydrate with respect to the domain is striking. The moiety is expected to be flexible based on studies of ribonuclease B (*56*) and thus restrict access to a significant region of the surface of the domain. The model suggests several possible roles for N-glycosylation in the formation of the latent TGF-*β* complex. First, a CR3-LAP interaction site has been proposed by Yuan et al. (*22*). It was suggested that a unique insertion of two amino acids (FP) in the CR3 sequence relative to other eight-cysteine domains in LTBP-1 might confer specificity for CR3-LAP interaction. This amino acid insertion is localized to a hydrophobic region of the surface of the CR3 domain and is adjacent to a partially exposed disulfide bond, which could exchange in covalently linked CR3-LAP. Interestingly, the region containing the FP sequence is on the opposite face of the CR3 domain from the glycosylation site. Hence, the mobility of the carbohydrate side chain may be important for guiding the interaction of CR3 with the LAP peptide.

Second, the proximity of the carbohydrate to the N-terminus of the eight-cysteine domain suggests another role for N-glycosylation in LTBP's and the fibrillins: the mediation of pairwise domain interactions. Usually a calcium-binding EGF-like domain precedes an eight-cysteine domain in these proteins. The position and flexibility of the sugar may dictate a specific relative orientation of these tandem modules. It is interesting to consider the pattern of eight-cysteine domain (TB domain) glycosylation in human fibrillin-1 with respect to protein−protein interactions and connective tissue microfibril assembly. TB4 contains an RGD sequence, which binds to $\alpha_v\beta_3$ integrins. Recently, transglutaminase cross-links that stabilize the microfibril have been localized to regions containing TB2 and TB7 (*57*). Only TB domains 5 and 6 are glycosylated in fibrillin-1. Hence, it seems that interaction sites and posttranslational modifications define specific functions for the different eight-cysteine domains in this protein. Since it is expected that carbohydrates will be localized to the surface of the microfibril, it seems likely that N-glycosylation is important for fibrillin assembly. Specifically, the properties of the carbohydrate side chain might limit the number of fibrillin molecules that are incorporated into the microfibril, dictating their final dimensions and also their supramolecular organization (*58*).

CONCLUSION

The data presented in this paper indicate unambiguously that complex glycans are added to recombinant CR3 in insect cells. An interesting feature shown by these results is the high occurrence of difucosylated sugars in High-Five cells as well as the superior ability of this cell line to process hybrid and complex glycans as compared to Sf9 cells. The status of LTBPs and fibrillins N-glycosylation in vertebrate cells remains to be established. The possibility of producing CR3, or any other domain of these proteins, with various N-glycosylation patterns in insect cells may help to elucidate the functions of these posttranslational modifications, which may include mediation of intramolecular and/or intermolecular protein−protein interactions.

ACKNOWLEDGMENT

We are grateful to Dr. Ron Beavis for his help for the analysis of CR3 glycopeptides by MALDI-TOF MS.

REFERENCES

1. Kanzaki, T., Olofsson, A., Moren, A., Wernstedt, C., Hellman, U., Miyazono, K., Claesson-Welsh, L., and Heldin, C.-H. (1990) *Cell 61*, 1051−1061.

2. Tsuji, T., Okada, F., Yamagushi, K., and Nakamura, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8835−8839.
3. Roberts, A. B., and Sporn, M. B. (1990) in *Peptide Growth Factors and Their Receptors I.* (Roberts, A. B., and Sporn, M. B., Eds.) pp 419−472, Springer-Verlag, New York.
4. Munger, J. S., Harpel, J. G., Gleizes, P. E., Mazzieri, R., Nunes, I., and Rifkin, D. B. (1997) *Kidney Int. 51*, 1376−1382.
5. Taipale, J., Miyazono, K., Heldin, C. H., and Keski-Oja, J. (1994) *J. Cell Biol. 124*, 171−181.
6. Ramirez, F. (1996) *Curr. Opin. Genet. Dev. 6*, 309−315.
7. Dallas, S. L., Miyazono, K., Skerry, T. M., Mundy, G. R., and Bonewald, L. F. (1995) *J. Cell Biol. 131*, 539−549.
8. Gibson, M. A., Hatzinikolas, G., Davis, E. C., Baker, E., Sutherland, G. R., and Mecham, R. P. (1995) *Mol. Cell. Biol. 15*, 6932−6942.
9. Hyytiainen, M., Taipale, J., Heldin, C. H., and Keski-Oja, J. (1998) *J. Biol. Chem. 273*, 20669−20676.
10. Moren, A., Olofsson, A., Stenman, G., Sahlin, P., Kanzaki, T., Claesson-Welsh, L., ten Dijke, P., Miyazono, K., and Heldin, C. H. (1994) *J. Biol. Chem. 269*, 32469−32478.
11. Saharinen, J., Taipale, J., Monni, O., and Keski-Oja, J. (1998) *J. Biol. Chem. 273*, 18459−18469.
12. Collod-Beroud, G., Beroud, C., Ades, L., Black, C., Boxer, M., Brock, D. J., Godfrey, M., Hayward, C., Karttunen, L., Milewicz, D., Peltonen, L., Richards, R. I., Wang, M., Junien, C., and Boileau, C. (1997) *Nucleic Acids Res. 25*, 147−150.
13. Dietz, H. C., and Pyeritz, R. E. (1995) *Hum. Mol. Genet. 4*, 1799−1809.
14. Ades, L. C., Haan, E. A., Colley, A. F., and Richard, R. I. (1996) *J. Med. Genet. 33*, 665−671.
15. Pfaff, M., Reinhardt, D. P., Sakai, L. Y., and Timpl, R. (1996) *FEBS Lett. 384*, 247−250.
16. Sakamoto, H., Broekelmann, T., Cheresh, D. A., Ramirez, F., Rosenbloom, J., and Mecham, R. P. (1996) *J. Biol. Chem. 271*, 4916−4922.
17. Nunes, I., Gleizes, P. E., Metz, C. N., and Rifkin, D. B. (1997) *J. Cell Biol. 136*, 1151−1163.
18. Saharinen, J., Taipale, J., and Keski-Oja, J. (1996) *EMBO J. 15*, 245−253.
19. Gleizes, P. E., Beavis, R. C., Mazzieri, R., Shen, B., and Rifkin, D. B. (1996) *J. Biol. Chem. 271*, 29891−29896.
20. Yin, W., Fang, J., Smiley, E., and Bonadio, J. (1998) *Biochim. Biophys. Acta 1383*, 340−350.
21. Reinhardt, D. P., Keene, D. R., Corson, G. M., Poschl, E., Bachinger, H. P., Gambee, J. E., and Sakai, L. Y. (1996) *J. Mol. Biol. 258*, 104−116.
22. Yuan, X., Downing, A. K., Knott, V., and Handford, P. A. (1997) *EMBO J. 16*, 6659−6666.
23. Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M., and Parekh, R. B. (1995) *Anal. Biochem. 230*, 229−238.
24. Guile, G. R., Rudd, P. M., Wing, D. R., Prime, S. B., and Dwek, R. A. (1996) *Anal. Biochem. 240*, 210−226.
25. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) *Anal. Chem. 63*, 1193A-1203A.
26. Harvey, D. J., Rudd, P. M., Bateman, R. H., Bordoli, R. S., Howes, K., Hoyes, J. B., and Vickers, R. G. (1994) *Organic Mass Spectrom. 29*, 753−765.
27. Petrescu, A. J., Petrescu, S. M., Dwek, R. A., and Wormald, M. R. (1999) *Glycobiology 9*, 343−352.
28. Wormald, M. R., Wooten, E. W., Bazzo, R., Edge, C. J., Feinstein, A., Rademacher, T. W., and Dwek, R. A. (1991) *Eur. J. Biochem. 198*, 131−139.
29. Rudd, P. M., Morgan, B. P., Wormald, M. R., Harvey, D. J., van den Berg, C. W., Davis, S. J., Ferguson, M. A., and Dwek, R. A. (1997) *J. Biol. Chem. 272*, 7229−7244.
30. Kulakosky, P. C., Hughes, P. R., and Wood, H. A. (1998) *Glycobiology 8*, 741−745.
31. Hsu, T. A., Takahashi, N., Tsukamoto, Y., Kato, K., Shimada, I., Masuda, K., Whiteley, E. M., Fan, J. Q., Lee, Y. C., and Betenbaugh, M. J. (1997) *J. Biol. Chem. 272*, 9062−9070.
32. Ogonah, O., Freedman, R., Jenkins, N., Patel, K., and Rooney, B. (1996) *Biotechnology 14*, 197−202.
33. Davidson, D. J., and Castellino, F. J. (1991) *Biochemistry 30*, 6165−6174.
34. Davidson, D. J., and Castellino, F. J. (1991) *Biochemistry 30*, 6689−6696.
35. Davis, T. R., and Wood, H. A. (1995) *In Vitro Cell. Dev. Biol. Anim. 31*, 659−663.
36. Kubelka, V., Altmann, F., Staudacher, E., Tretter, V., Marz, L., Hard, K., Kamerling, J. P., and Vliegenthart, J. F. (1993) *Eur. J. Biochem. 213*, 1193−1204.
37. Kubelka, V., Altmann, F., and Marz, L. (1995) *Glycoconj. J. 12*, 77−83.
38. Hard, K., Van Doorn, J. M., Thomas-Oates, J. E., Kamerling, J. P., and Van der Horst, D. J. (1993) *Biochemistry 32*, 766−775.
39. Kubelka, V., Altmann, F., Kornfeld, G., and Marz, L. (1994) *Arch. Biochem. Biophys. 308*, 148−157.
40. Altmann, F., and Marz, L. (1995) *Glycoconj. J. 12*, 150−155.
41. Davis, T. R., Schuler, M. L., Granados, R. R., and Wood, H. A. (1993) *In Vitro Cell. Dev. Biol. Anim. 29A*, 842−846.
42. Foster, J. M., Yudkin, B., Lockyer, A. E., and Roberts, D. B. (1995) *Gene 154*, 183−186.
43. Jarvis, D. L., Bohlmeyer, D. A., Liao, Y. F., Lomax, K. K., Merkle, R. K., Weinkauf, C., and Moremen, K. W. (1997) *Glycobiology 7*, 113−127.
44. Ren, J., Bretthauer, R. K., and Castellino, F. J. (1995) *Biochemistry 34*, 2489−2495.
45. Ren, J., Castellino, F. J., and Bretthauer, R. K. (1997) *Biochem. J. 324*, 951−956.
46. Altmann, F., Kornfeld, G., Dalik, T., Staudacher, E., and Glossl, J. (1993) *Glycobiology 3*, 619−625.
47. Velardo, M. A., Bretthauer, R. K., Boutaud, A., Reinhold, B., Reinhold, V. N., and Castellino, F. J. (1993) *J. Biol. Chem. 268*, 17902−17907.
48. van Die, I., van Tetering, A., Bakker, H., van den Eijnden, D. H., and Joziasse, D. H. (1996) *Glycobiology 6*, 157−164.
49. Davidson, D. J., Bretthauer, R. K., and Castellino, F. J. (1991) *Biochemistry 30*, 9811−9815.
50. Staudacher, E., Kubelka, V., and Marz, L. (1992) *Eur. J. Biochem. 207*, 987−993.
51. Wilson, I. B., Harthill, J. E., Mullin, N. P., Ashford, D. A., and Altmann, F. (1998) *Glycobiology 8*, 651−661.
52. Prenner, C., Mach, L., Glossl, J., and Marz, L. (1992) *Biochem. J. 284*, 377−380.
53. Ohsuga, H., Su, S. N., Takahashi, N., Yang, S. Y., Nakagawa, H., Shimada, I., Arata, Y., and Lee, Y. C. (1996) *J. Biol. Chem. 271*, 26653−26658.
54. Rudd, P. M., Joao, H. C., Coghill, E., Fiten, P., Saunders, M. R., Opdenakker, G., and Dwek, R. A. (1994) *Biochemistry 33*, 17−22.
55. Varki, A. (1993) *Glycobiology 3*, 97−130.
56. Woods, R. J., Edge, C. J., and Dwek, R. A. (1994) *Nat. Struct. Biol. 1*, 499−501.
57. Qian, R. Q., and Glanville, R. W. (1997) *Biochemistry 36*, 15841−15847.
58. Wess, T. J., Purslow, P. P., Sherratt, M. J., Ashworth, J., Shuttleworth, C. A., and Kielty, C. M. (1998) *J. Cell Biol. 141*, 829−837.

BI9918285