Determination of the Glycosylation Patterns, Disulfide Linkages, and Protein Heterogeneities of Baculovirus-Expressed Mouse Interleukin-3 by Mass Spectrometry†

Thomas P. Knepper,‡ Brian Arbogast,‡ Jolanda Schreurs,§ and Max L. Deinzer*,‡

Department of Agricultural Chemistry, Oregon State University, Corvallis, Oregon 97331-7301, and Department of Protein Chemistry, Chiron Corporation, Emeryville, California 94608-2916

Received March 5, 1992; Revised Manuscript Received September 8, 1992

ABSTRACT: The primary structure of mouse interleukin-3 (IL-3) expressed by recombinant baculovirus-infected silkworm (*Bombyx mori*) larvae was analyzed by subjecting isolated IL-3 derived peptides to liquid secondary ion mass spectrometry. Two species of IL-3 were isolated from the silkworm hemolymph by reverse-phase high-pressure liquid chromatography. The major component has M_r 20–22 × 10^3 as determined by SDS–PAGE. Liquid secondary ion mass spectrometric analysis was carried out on the reduced tryptic and endopeptidase lysyl-C peptides of glycosylated and deglycosylated IL-3. These studies provided evidence that (1) Asn-16 is heterogeneously glycosylated with four different oligosaccharides, (2) Asn-86 is either nonglycosylated or has attached to it one oligosaccharide, (3) the N-glycosylation sites Asn-44 and Asn-51 are not glycosylated, and (4) there is no O-glycosylation. Liquid secondary ion mass spectrometric analysis of the unreduced tryptic peptides provided evidence for disulfide linkages between Cys-140 and Cys-79 or Cys-80 and between Cys-17 and Cys-79 or Cys-80. In comparison to the major component, a minor IL-3 species (M_r 17–19 × 10^3 by SDS–PAGE) isolated from the hemolymph showed no difference with respect to the glycosylation pattern or the disulfide linkages, but it was cleaved between Ala-127 and Ser-128, and only a disulfide linkage between Cys-140 and Cys-79 or Cys-80 held the molecule together. The apparent lower molecular weight of the minor component as determined by SDS–PAGE results from the use of reducing conditions and the loss of the 12 amino acid C-terminal end during the experiment. Molecular weight determinations of IL-3 glycoforms by electrospray mass spectrometry showed a predominant species with m/z 17 423 and several minor glycoforms that could be assigned on the basis of the difference in the masses of three chitobiose core oligosaccharides.

There are many variations in the approaches to glycoconjugate analyses by mass spectrometry (McCloskey, 1990), but the essential features include an analysis of the primary structure of the protein to confirm the sequence and to locate the sites of modification followed by elucidation of the carbohydrate structure. The particular capability of liquid secondary ion mass spectrometry (LSIMS)[1] to analyze complex peptide mixtures has been used in specific molecular mass determinations of peptides in proteolytic digests (Gibson & Biemann, 1984; Biemann et al., 1985; Biemann & Martin, 1987). Information that can be obtained by this procedure includes the molecular masses of both the peptides and glycopeptides and the glycosylation sites (Carr & Roberts, 1986; Carr et al., 1988) and masses of the carbohydrate residues as well as other posttranslational modifications (Carr & Biemann, 1984). The identification of disulfide linkages by mass spectrometry (Morris & Pucci, 1985; Yazdanparast et al., 1986, 1987) is made possible if proteolysis or partial acid hydrolysis of the protein can be achieved without first having to reduce it. These procedures now have been applied to the analysis of a recombinant form of the glycoprotein, interleukin-3.

Interleukin-3 (IL-3) is a T-cell- and mast-cell-derived factor (Rennick et al., 1985; Wodnar-Filipowicz et al., 1989) which supports the division and differentiation of a broad spectrum of hemopoietic cells (Whetton et al., 1989; Ihle & Weinstein, 1986). Its ability to promote the proliferation of progenitor cells and multipotential stem cells makes it a primary regulator of the early stages of hemopoiesis and lymphopoiesis, particularly during the inflammatory process (Miyajima et al., 1988).

Mouse IL-3 has been extensively studied; it is a glycoprotein that consists of 140 amino acids with four potential N-linked glycosylation sites (Miyajima et al., 1987). Structural heterogeneity in IL-3 preparations has been recognized in the N-terminal amino acid sequence as, for example, in WEHI-3-derived IL-3 cells in which the first six amino acids are missing (Ihle et al., 1983; Clark-Lewis & Schrader, 1988) or in the variable glycosylation (Clark-Lewis & Schrader, 1988). These different molecular forms are due to posttranslational modifications and not to the products of multiple genes. Clark-Lewis et al. (1987) also showed by amino acid substitution in synthetic protein preparations that a disulfide linkage forms between Cys-17 and Cys-80.

IL-3 purified from WEHI-3 conditioned cell supernatants has an apparent molecular weight of 28 × 10^3 as shown by SDS–PAGE (Ihle et al., 1982; Clark-Lewis et al., 1984). Further characterization of the structure and properties of this important lymphokine has been facilitated by the availability of large amounts of the protein derived from

† This work was supported by the National Institutes of Health (NIEHS Grants ES0040 and ES00210).

* To whom correspondence should be addressed.

‡ Oregon State University.

§ Chiron Corp.

[1] Abbreviations: IL-3, interleukin-3; LSIMS, liquid secondary ion mass spectrometry; [(M + H)$^+$], protonated molecular ion; RP-HPLC, reversed-phase high-pressure liquid chromatography; DTT, dithiothreitol; DTE, dithioerythritol; 3-NBA, 3-nitrobenzylalcohol; TFA, trifluoroacetic acid.

0006-2960/92/0431-11651$03.00/0

silkworms (*Bombyx mori*) infected with a recombinant baculovirus carrying the mouse IL-3 cDNA (Miyajima et al., 1987). Analysis of IL-3 isolated from silkworm hemolymph by SDS–PAGE showed two predominant bands of approximate molecular weights 20×10^3 and 21×10^3 and several minor species of 17×10^3–19×10^3 (Miyajima et al., 1987; Svoboda et al., 1991). The in vitro biological activity of the mixture of these species was indistinguishable from that of the IL-3 produced by mammalian cells (Miyajima et al., 1987).

The insect-expressed recombinant mouse IL-3, with an apparent molecular weight of 20–21 $\times 10^3$ (SDS–PAGE), was found by ^{252}Cf plasma desorption mass spectrometry (PDMS) to consist of at least two glycoforms with true molecular weights of around 17×10^3 and no modifications at the N-terminal amino acid (Svoboda et al., 1991). Manual Edman degradation in combination with PDMS analysis indicated N-glycosylation at Asn-16 and Asn-86.

The structures of the Asn-linked oligosaccharides also have been determined by liquid secondary ion mass spectrometry (LSIMS) (Hogeland et al., 1992). The oligosaccharide composition was determined to be Man_{2-4}[Fuc]$GlcNAc_2$ and $Man_2GlcNAc_2$. These were believed to be di-, tri-, and tetramannose chitobiose core oligosaccharides, three of which were fucosylated. Further studies have now been conducted on the heterogeneity of glycosylation, the disulfide linkages and the heterogeneity of the protein species isolated from the hemolymph of silkworm larvae infected with the baculovirus recombinant mouse IL-3. The results of these studies are reported here.

EXPERIMENTAL PROCEDURES

Production and Purification of Interleukin-3. The IL-3 was produced by infection of silkworm larvae with a recombinant baculovirus and purified by a modified procedure of Miyajima et al. (1987). Crude silkworm hemolymph containing approximately 0.5% IL-3 was applied to a QAE-Sephadex A-25 column using 50 mM Tris-HCl (pH 8.0) as solvent. Protein determinations throughout the purification employed either the method of Lowry et al. (1951) or were based on the optical density at 280 nm. Protein determinations by SDS–PAGE were carried out according to the procedures of Laemmli (1970). The IL-3 activities in samples and fractions were tested in a colorimetric proliferation assay using MC/9 cells as developed by Mosmann (1983) and described by Miyajima et al. (1987). Fractions containing IL-3 were concentrated with an Amicon filter (YM3). Reverse-phase high-pressure liquid chromatography (RP-HPLC) was used for further purification of IL-3 (see below). For the screening of different species of IL-3 after RP-HPLC separation, mass spectrometric mapping of the tryptic peptides was employed. Approximately 5 μg of protein based on the absorbance at 220 nm of each collected fraction was lyophilized and digested with trypsin. The tryptic peptide mixture was analyzed by LSIMS. The IL-3 fractions containing IL-3 were purified by C18 RP-HPLC. With the exception of the RP-HPLC experiments, all manipulations were done at 4 °C.

Release of Asn-Linked Oligosaccharides from IL-3. IL-3 (35 μg) was dissolved in 17.5 μL of 50 mM NH_4HCO_3 buffer solution (pH 8.7). To this solution was added 3.15 units of N-glycanase (Genzyme) in 12.6 μL of 50% glycerol solution containing 2.5 mM EDTA. The solution was incubated at 37 °C for 18 h. Separation of the deglycosylated IL-3 from the released oligosaccharides was carried out using an Amicon Centricon filtration system. Further purification was carried out by C18 RP-HPLC.

Proteolytic Digestion of Glycosylated IL-3. To 17.5 μg of IL-3 in 20 μL of 50 mM NH_4HCO_3 (pH 8.7) was added 0.5 μL of a solution of either trypsin (from bovine pancreas, TPCK-treated, Sigma) in 10 mM HCl (1 μg/μL) or lysyl-endopeptidase C (Lys-C, Wako) in 50 mM NH_4HCO_3 (1 μg/μL) (1:35 enzyme to substrate ratio, w/w). The solution was incubated at 37 °C for 1 h.

Trypsin Digestion of Deglycosylated IL-3. After the enzymatic deglycosylation of 17.5 μg of IL-3 in 15 μL of 50 mM NH_4HCO_3 buffer solution (pH 8.7), 0.5 μg of trypsin (1:35 enzyme to substrate ratio) was added, and digestion was carried out without further purification.

Cyanogen Bromide Cleavage of IL-3. To 5 μg of IL-3 in 20 μL of 70% formic acid, 0.1 μL of a solution of 5 M cyanogen bromide (BrCN) in acetonitrile (1250-fold molar excess) was added. The reaction was terminated by lyophilization after 4 h at 25 °C.

RP-HPLC. The HPLC system consisted of an Altex Model 322 gradient liquid chromatograph. The detector was a Waters 486 tunable absorbance detector. All RP-HPLC purifications of IL-3 were performed on a Vydac C4 RP column (30 mm). For the isolation of tryptic peptides, only the C18 RP column was used. The solvents consisted of (A) 0.1% aqueous trifluoroacetic acid (TFA) and (B) acetonitrile with 0.08% TFA. Samples were manually collected and lyophilized.

Amino Acid Sequence Analysis. Peptides were subjected to Edman degradation using a model 470A (Applied Biosystems) gas-phase protein sequencer.

Xenon-Liquid Secondary Ion Mass Spectrometry (Xe-LSIMS). LSIMS analyses were carried out on a Kratos MS-50 double focusing mass spectrometer operating at a resolution of 1200–1500 for centroid data collection and at a resolution of 1000 for raw data collection. Xenon gas was used to generate the primary ionizing beam from an Ion-Tech gun operated at 7–8 keV. Ions were accelerated from the ion source at 8 keV. The Kratos postacceleration detector was modified to allow operation at 25 keV. The scan rate was 30 s per decade over high mass ranges (raw data) and 10 s per decade over low mass ranges (centroided data). Peptides were dissolved in 2 μL of 0.1% TFA and were mixed with 2 μL of matrix [either dithiothreitol (DTT)/dithioerythritol (DTE) (5:1) or 3-nitrobenzyl alcohol (3-NBA) on the target]. In some instances in which ions were severely depressed, additional (2%) TFA was added to the DTT/DTE matrix.

Electrospray Mass Spectrometry. Electrospray mass spectra were obtained on a Finnigan Corporation triple quadrupole mass spectrometer (TSQ-700) with an ESI source. IL-3 (23 pmol/μL) was dissolved in methanol/water/acetic acid (50:50:1). The sample solution was introduced into the instrument with a low-pressure syringe pump at a flow rate of 2 μL/min. Scans were taken at a rate of 10 s per decade. Deconvolution of the spectra was accomplished with the BIOMASS software, which is based on a deconvolution algorithm that utilizes a transformation function as described by Mann et al. (1989).

RESULTS

Purification and Characterization of IL-3. Fractions collected from the C4 RP-HPLC column (Figure 1) initially were screened by measuring for bioactivity of IL-3 (Mosmann, 1983; Miyajima et al., 1987). Subsequently, the IL-3-containing fractions were monitored by LSIMS analyses of the tryptic peptides. Fractions 6–8 yielded tryptic peptide mixtures which contained all the protonated molecular ions of the peptides expected from the primary structure of IL-3

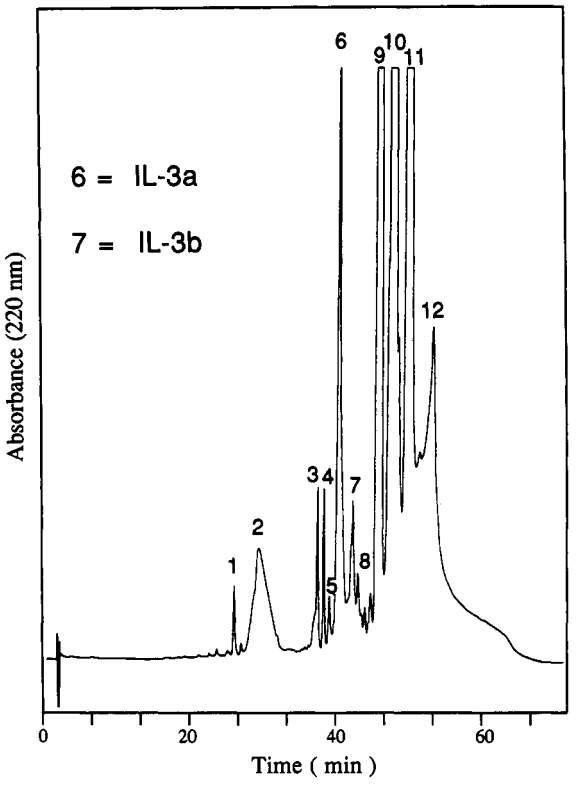

FIGURE 1: Fractionation of IL-3 by C4 RP-HPLC after purification of silkworm hemolymph by anion exchange chromatography and concentration through an Amicon filter. The VYDAC C4 RP column was equilibrated with 0.1% aqueous TFA (solvent A, 80%) and acetonitrile with 0.08% TFA (solvent B, 20%). The elution gradient was as follows: 0–5 min, 80% A; 5–70 min, 80% A to 30% A; 70–75 min, 30% A. Peak numbers indicate the individual collected and pooled fractions.

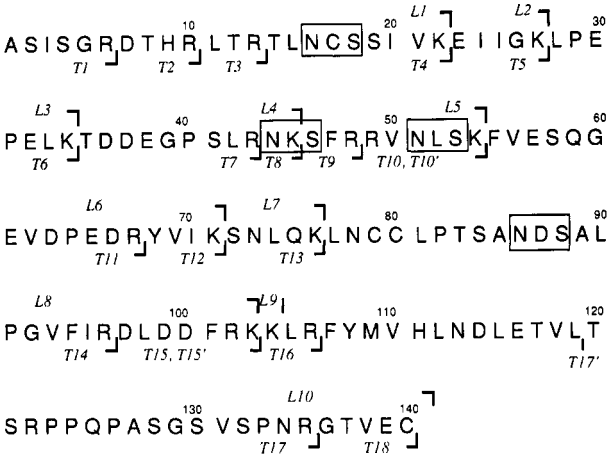

FIGURE 2: Interleukin-3 protein sequence (without leader sequence). Observed tryptic (T) and lysyl endopeptidase-C (L) peptides and potential N-glycosylation sites (indicated by boxes).

(Figure 2). Analysis of the tryptic digests of all other fractions by LSIMS showed the presence of mostly proteinaceous material, but the constituents in these fractions were not identified. Fractions 6–8 were resolved further on a C18 reverse-phase column using a shallow acetonitrile gradient (30–45%). The individual fractions were collected and again screened for the presence of IL-3 by LSIMS analysis of the tryptic peptides. Only fractions 6 (IL-3a) and 7 (IL-3b) contained the expected peptides for IL-3. The bioactivities of the two proteins were similar. Analysis by SDS–PAGE showed that these proteins also were structurally similar to the IL-3 species reported by Miyajima et al. (1987). A total

Table I: Predicted and Experimental Masses of Reduced Tryptic Peptides of Glycosylated and Deglycosylated IL-3a

peptide	sequence	MH^+ [b]	MH^+ in LSIMS spectra[a]	
			deglyco	glyco
T1	1–6	590.3	590.4	590.3
T2	7–10	528.3	sup	528.3
T3	11–13	389.3	389.3	389.4
T4	14–22	964.5	965.5[c]	–
T5	23–27	559.4	559.4	559.4
T6	28–34	825.5	825.5	825.5
T5,6	23–34	1365.8	1365.5[d]	1366.0
T7	35–43	989.5	989.5	989.5
T8	44–45	261.2	nd[e]	nd
T9	46–48	409.2	409.3	409.3
T10	49–54	716.4	716.5	716.5
T10′	50–54	560.3	560.4	560.4
T11	55–67	1506.7	1506.6[d]	1506.9
T12	68–71	522.3	522.4	522.4
T13	72–76	589.3	589.4	589.4
T14	77–96	2092.0	2093.1[c,f]	2091.9[f]
T15	97–102	780.4	780.4	780.4
T15′	97–103	908.0	sup[g]	908.2
T16	104–106	416.3	416.3	416.3
T17	107–135	3214[h]	3214[h]	3214[h]
T17′	107–119	1593.8	1594.0	1594.3
T18	136–140	508.2	nd	nd
T4 [OS-1 at N16][i]		1695[h]	–	1695[h]
T4 [OS-2 at N16][i]		1841[h]	–	1841[h]
T4 [OS-3 at N16][i]		2003[h]	–	2004[d,h]
T4 [OS-4 at N16][i]		2165[h]	–	2164[h]
T14 [OS-2 at N86][i]		2968[h]	–	2969[h]

[a] DTT/DTE/0.1% TFA (1:1:1) was used as a matrix system. [b] Calculated monoisotopic mass (100% relative intensity). [c] One mass unit higher as calculated due to hydrolysis of Asn to Asp during enzymatic deglycosylation. [d] Data not shown. [e] nd, not detected. [f] Asn-86 is not completely glycosylated (see Figure 6). [g] sup, suppressed. [h] Average mass. [i] See Figure 4 for oligosaccharide structure.

of 3 mg of IL-3 was obtained from 13.5 mL of hemolymph with a ratio for IL-3a/IL-3b of approximately 10:1.

Primary Structure and Glycosylation Sites of IL-3a. The tryptic digests of the glycosylated (M_r 20–21 × 10^3 by SDS–PAGE) and of the enzymatically deglycosylated IL-3a (M_r 15 × 10^3) yielded mixtures which could be analyzed directly by LSIMS. Possible disulfide bonds between tryptic fragments were reduced on the target using DTT/DTE as a matrix. In the LSIMS spectra either taken in the mass range of 370–2200 (centroided data) or 1500–4000 (raw data), various ion peaks appeared. Several adjacent digestion sites (Lys, Arg) leading to overlapping peptides were observed in the protein sequence, and these peptides are indicated by T′ (Figure 2).

In the LSIMS spectrum of the tryptic digest of deglycosylated IL-3a, all protonated molecular ions [$(M + H)^+$] corresponding to the expected tryptic peptides except T8 (amino acids 44–45) and the C-terminal peptide T18 (amino acids 136–140) (Table I) were recorded. The ions corresponding to T2 (amino acids 7–10) and T15′ (amino acids 97–103) were suppressed and were detected only under more acidic matrix conditions, i.e., with added (2%) TFA (data not shown). A nonspecific tryptic cleavage at the C-terminal side of Leu-119 reported previously (Svoboda et al., 1991) leads to a peptide (amino acids 107–119) with m/z 1594 (T17′, Table I). The mass accuracy of all mass assignments of the centroided peaks in the spectra was ±0.2 Da (Figure 3). This allowed peptides with a mass difference of 1 Da (T1 and T13; T5 and T10) to be distinguished. The masses of the [$(M + H)^+$] ions of the tryptic peptides T4 (amino acids 14–22) and T14 (amino acids 77–96), which contain the glycosylation sites, Asn-16 and Asn-86, were one mass unit higher than the calculated protonated molecular masses. This is due to the

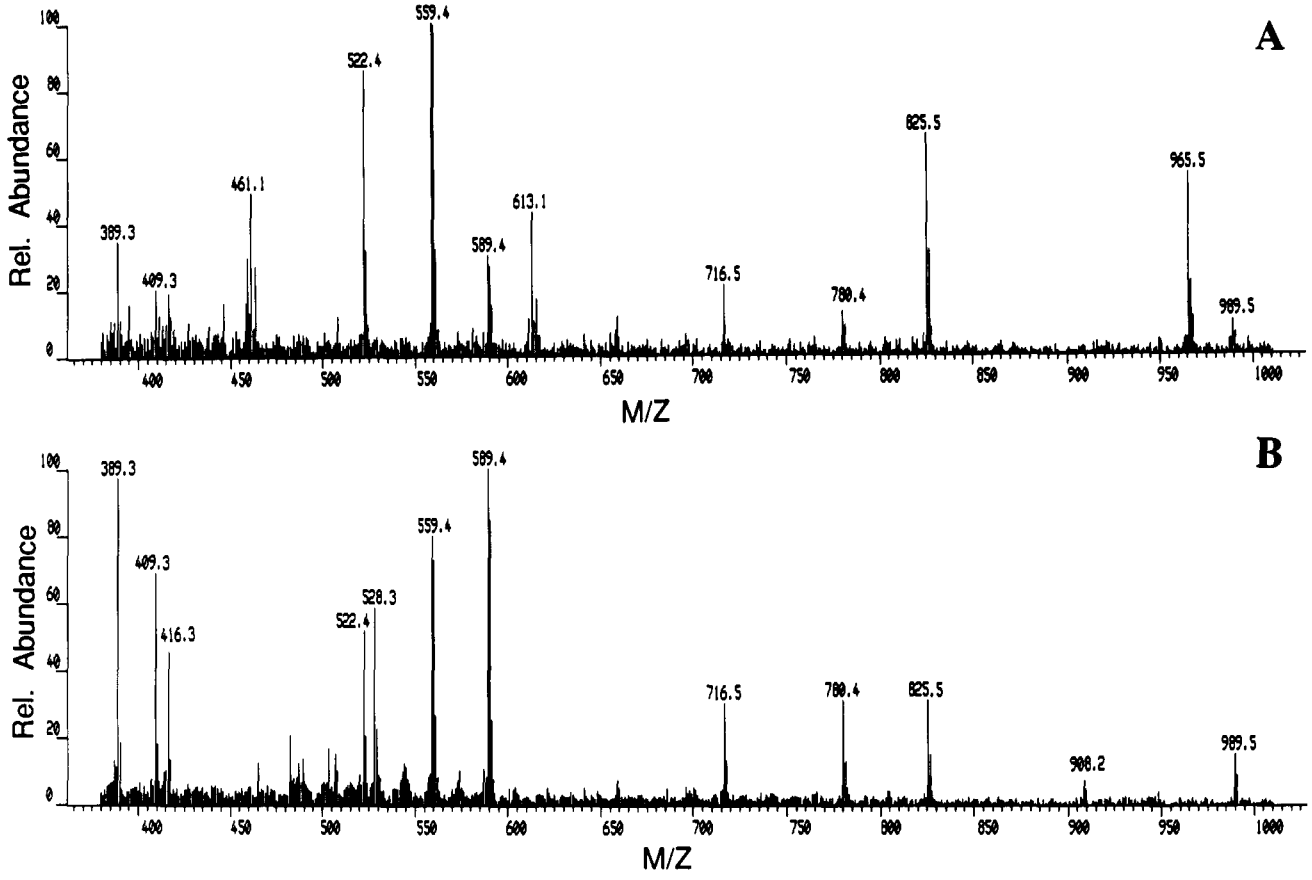

FIGURE 3: LSIMS spectrum (centroided data acquisition) of tryptic digests of reduced deglycosylated (A) and glycosylated (B) IL-3a. The spectra were acquired from the peptide mixture of 0.5 μg (28.7 pmol) of IL-3a in DTT/DTE/0.1% TFA (1:1:1) matrix. The ion peak with m/z 965.5 in spectrum A is the deglycosylated tryptic fragment, T4 (one mass unit higher than predicted due to conversion of Asn to Asp during enzymatic deglycosylation; see also Table I). The remaining peaks are due to the MH^+ of the tryptic fragments, T1–T3 and T5–T18.

Man-Man(1-4)GlcNAc-GlcNAc- (OS-1 , Mw. 731)

Fuc
Man-Man(1-4)GlcNAc-GlcNAc- (OS-2 , Mw. 877)

Fuc
Man
Man(1-4)GlcNAc-GlcNAc- (OS-3 , Mw. 1039)
Man

Fuc
Man
Man_2 Man(1-4)GlcNAc-GlcNAc- (OS-4 , Mw. 1201)

FIGURE 4: Structures and molecular weights of the Asn-linked chitobiose core oligosaccharides of silkworm-expressed recombinant mouse IL-3 (Hogeland et al., 1992).

hydrolysis of asparagine to aspartic acid which occurs during enzymatic deglycosylation (Carr & Roberts, 1986).

Further evidence is found for glycosylation at Asn-16 by comparing the LSIMS spectra of the tryptic digests of the glycosylated IL-3a (Figure 3B) with the deglycosylated IL-3a (Figure 3A) in the mass range of 370–1000 Da. All the protonated ions of the tryptic peptides which could be identified in the spectrum of the deglycosylated IL-3a except T4 were assignable. Glycosylation of Asn-16 should shift the molecular weight of T4 to masses which are higher by the masses of the oligosaccharides attached to the protein (Figure 4). In the LSIMS spectrum of the glycosylated tryptic peptides of IL-3a, ion peaks at m/z 1695, 1841, 2164, and 2969 were observed (Figure 5); these were not present in the LSIMS spectrum of the deglycosylated IL-3 tryptic digest. These m/z values correspond to protonated molecular ions of T4 glycosylated at Asn-16 with oligosaccharides OS-1 (m/z 1695), OS-2 (m/z 1841), or OS-4 (m/z 2164) or of T14 glycosylated at Asn-86 with OS-2 (m/z 2969) (Table I). A small peak with m/z 2004 (Figure 5) suggests that Asn-16 also is glycosylated with OS-3. The glycosylation at Asn-86 appears not to be complete because an ion corresponding to a nonglycosylated T14 is present with m/z 2901.9. However, the majority of T14 is glycosylated as shown by the more intense ion cluster after N-glycanase treatment of the protein (Figure 6A). This ion cluster is shifted 1 Da higher because of the asparagine to aspartic acid conversion.

In previous studies (Hogeland et al., 1992), the carbohydrates were removed from the protein by N-glycanase, derivatized by permethylation and reductive amination, and analyzed by LSIMS. The mass spectral data showed the presence of four chitobiose core oligosaccharides (Figure 4) in the mixture. Since the carbohydrates were permethylated, possible fragmentations such as defucosylation of glycan OS-2 to give OS-1 would have been noted because the resultant glycan fragment ion peak would have appeared 14 mass units lower in the spectrum. This was not observed. Thus, there is very good supportive evidence for the presence of individual glycosylated peptide ions in the LSIMS spectrum (Figure 5).

No ion peaks for T8 and T18 were observed in the LSIMS spectra of the reduced tryptic peptide mixtures from either the glycosylated or the deglycosylated IL-3a preparations. Accordingly, a Lys-C digest of glycosylated IL-3a was analyzed (Figure 2). All predicted fragment ions except L9 (amino acid 104) were present in the spectrum (Table II). The presence of the $[(M + H)^+]$ ions for the fragments L4 (amino acids 35–45) and L10 (amino acids 105–140) indicates that these peptides are not glycosylated. Moreover, there were no ion peaks in the spectrum with masses suggesting

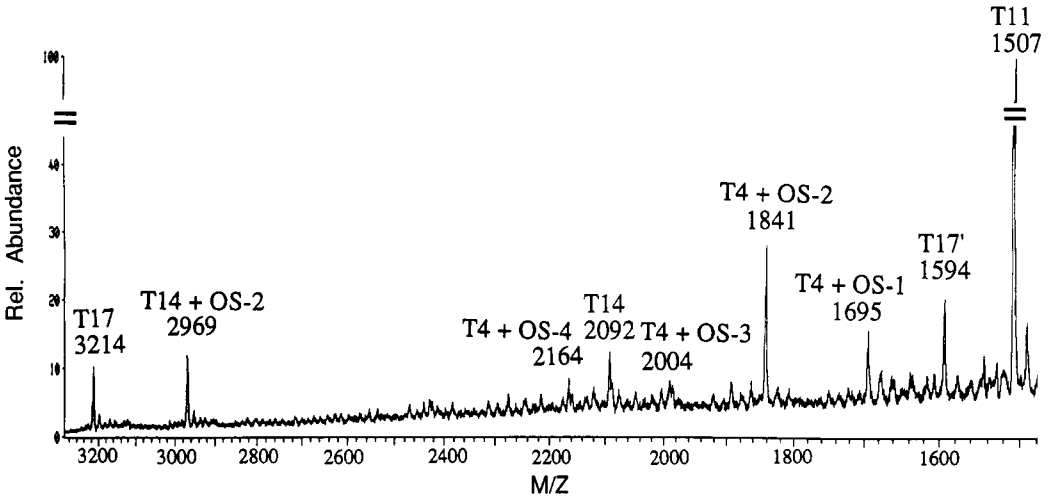

FIGURE 5: LSIMS spectrum (raw data acquisition) of the tryptic peptides and fragments of reduced glycosylated IL-3a, 2 μg (115 pmol) IL-3a, DTT/DTE/0.1% TFA (1:1:1) matrix.

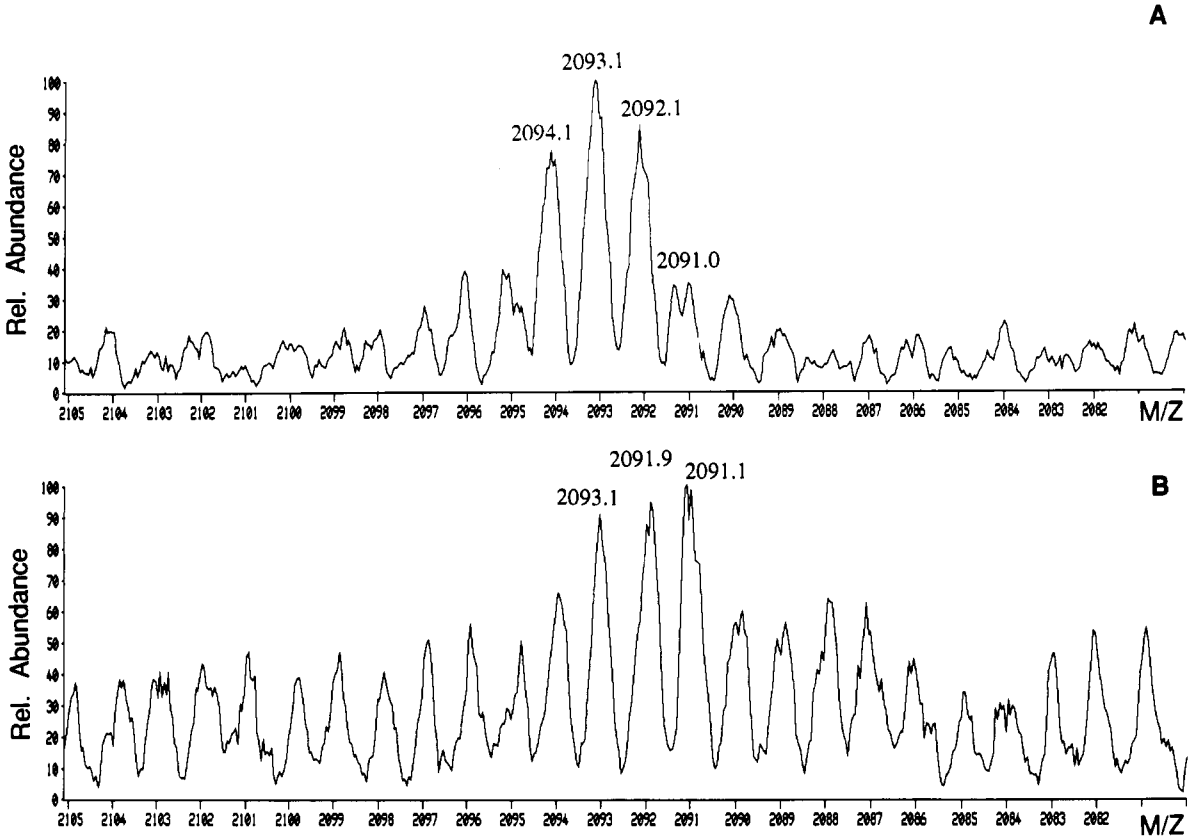

FIGURE 6: Comparison of the LSIMS spectra (raw data acquisition, resolution 2000) of tryptic digests of reduced deglycosylated (A) and glycosylated (B) IL-3a (5 μg, 287 pmol, each) in the mass range 2080–2105 Da; DTT/DTE/0.1% TFA (1:1:1) matrix. The mass (m/z 2092.1) of the deglycosylated tryptic fragment T14 (A) is one unit higher than predicted due to conversion of Asn to Asp during enzymatic deglycosylation with N-glycanase. In comparison, the intensity of the ion cluster of the nonglycosylated T14 (B) is weak (compare S/N ratios) which, in combination with the appearance of intense ions of the glycosylated T14 (see Figure 5), confirms that T14 is mainly glycosylated (see also Table I).

glycosylation at Asn-44, nor were there any for O-glycosylation. It is possible, of course, that such peptides escaped detection because of ion suppression or because of reduced sensitivity at higher masses.

Determination of Disulfide Linkages in IL-3a. Three of the four cysteine residues of IL-3a are located in the tryptic peptides T4 (Cys-17) and T14 (Cys-79 and Cys-80), which also are N-glycosylated. The fourth cysteine is located at the C-terminus (T18) (Cys-140). To determine possible disulfide linkages in the protein, 200 pmol of the unreduced tryptic digest was analyzed by LSIMS using the nonreducing matrix 3-NBA. A protonated molecular ion with m/z 5312 and a less abundant ion with m/z 5166 were recorded (Figure 7), indicating the presence of the heterogeneously glycosylated doubly disulfide-linked tripeptide T4–T14–T18 (amino acids 14–22, 77–96, and 136–140, Table III). Slow reduction of the disulfide bond between T14 and T18 or T4 and T14 on the target was observed after longer measuring times, yielding [(M + H)$^+$] ions with m/z 4808 and 3475, respectively. The reduction of disulfide bonds on the target occurred much faster when glycerol was used as a matrix (data not shown). The glycosylated and disulfide-linked tripeptide was further

Table II: Calculated and Experimental Masses of Reduced Peptides of Glycosylated IL-3a and Lys-C Digestion

peptide	sequence	MH^+ [a] (calcd)	MH^+ in LSIMS spectra[b]
L1 [OS-1 at N16][b]	1–22	3145.8	3145.7
L1 [OS-2 at N16][b]	1–22	3292.0	3292.9
L2	23–27	559.7	559.4
L3	28–34	826.0	825.5
L2, L3	23–34	1366.7	1365.7
L4	35–45	1232.3	1231.6
L5	46–54	1107.3	1106.6
L6	55–71	2011.2	2010.9
L7	72–76	589.7	589.4
L8 [OS-2 at N86][b]	77–103	3857.5	3858.2
L9	104–104	147.2	nd[d]
L10	105–140	3972.5	3972.7

[a] Calculated average mass (100% relative intensity). [b] Average mass; DTT/DTE/0.1% TFA (1:1:1) was used as a matrix system. [c] See Figure 4 for oligosaccharide structure. [d] nd, not detected.

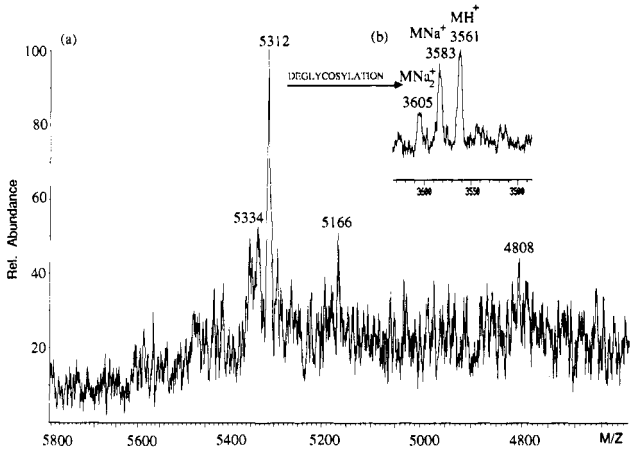

FIGURE 7: (a) LSIMS spectrum (raw data acquisition) of the tryptic digest of unreduced, glycosylated IL-3a in the mass range 4600–5800 Da; 3 μg (172 pmol) of IL-3a in 3-NBA/0.1% TFA (1:1) matrix (for peak assignments see Table IV). (b) LSIMS spectrum acquired from the unreduced deglycosylated IL-3a. The single protonated molecular ion and the sodium adducts are due to the deglycosylated, disulfide-linked tripeptide T4–T14–T18 (see also Table IV).

Table III: Calculated and Experimental Masses of Disulfide-Linked Tryptic Peptides of Glycosylated and Deglycosylated IL-3a

sequence	MH^+ (calcd)[a]	MH^+ in LSIMS spectra[a,b]
(T4)-S-S-(T14)-S-S-(T18)	3559	3561[c]
(T4)-S-S-(T14)-S-S-(T18) [OS-2 at N16 and N86][d]	5312	5312
(T4)-S-S-(T14)-S-S-(T18) [OS-1 at N16 and OS-2 at N86][d]	5166	5166
(T4)-S-S-(T14)-SH [OS-2 at N16 and N86][d,e]	4807	4808
HS-(T14)-S-S-(T18) [OS-2 at N86][d,e]	3475	3475[f]

[a] Average mass. [b] 3-NBA was used as a matrix system. [c] Two mass units higher as calculated due to hydrolysis of Asn to Asp during enzymatic deglycosylation. [d] See Figure 5 for oligosaccharide structures. [e] Partial reduction of disulfide bonds due to Xenon atom bombardment was observed after approximately 60 s measuring time; it occurred even faster in glycerol as matrix. [f] Data not shown.

analyzed by LSIMS after C18 RP-HPLC separation. Reduction of the tripeptide with DTT/DTE yielded the expected protonated molecular ions for glycosylated T4 and the variably glycosylated/nonglycosylated T14 (data not shown). The reduction of the disulfide linkage between Cys-79 or Cys-80 and Cys-140 caused the loss of T18.

In the LSIMS spectrum acquired from the tryptic digest mixture of deglycosylated IL-3a in the nonreducing matrix

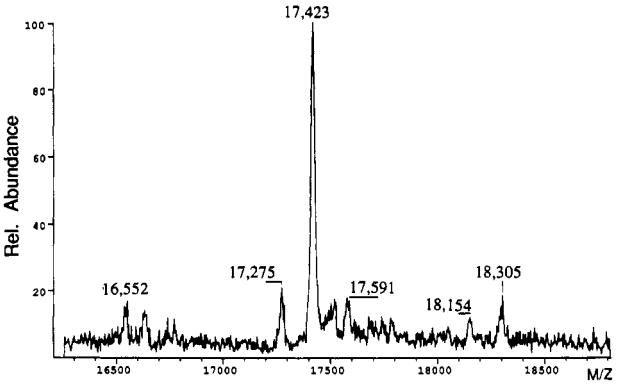

FIGURE 8: Deconvoluted electrospray mass spectrum of heterogeneously glycosylated IL-3a. For conditions, see Experimental Procedures.

Table IV: Calculated and Observed Molecular Weight Distribution of Glycosylated IL-3a As Determined by Electrospray MS

glycosylation pattern	MH^+ (calcd)[a]	MH^+ in electrospray[a,b]	error (%)[c]
[OS-2 at N16]	16 545	16 552	0.04
[OS-1 at N16; OS-2 at N86]	17 275	17 275	
[OS-2 at N16 and N86]	17 421	17 423	
[OS-3 at N16; OS-2 at N86]	17 584	17 591	0.04

[a] Average mass. [b] Deconvoluted spectrum calculated from several multiply charged ions. [c] Mass measurement accuracy by electrospray MS is expected to be within 0.01%. The low intensities of these ions may contribute to the larger errors.

3-NBA, a new peak series with m/z 3561, 3583, and 3605 was observed. These peaks could be assigned to the protonated molecular ion and the monosodium and disodium adducts of the deglycosylated, disulfide-linked tripeptide T4–T14–T18 (Figure 7 and Table III). Tryptic digestion after enzymatic deglycosylation of IL-3a did not lead to scrambling of disulfide linkages, i.e., the formation of a new linkage between Cys-17 and Cys-140. However, HPLC isolation of the double disulfide-linked tryptic tripeptide followed by enzymatic deglycosylation led to scrambling of the disulfide linkages, as shown by the appearance of a new protonated molecular ion. Its mass (m/z 1471) indicated this was peptide T4 linked to T18 through a disulfide bond (data not shown).

Molecular Weight Determination of IL-3a. The fact that IL-3a is heterogeneously glycosylated also was shown by electrospray-MS of the intact glycoprotein (Figure 8). The average molecular weight of the disulfide-linked IL-3 with no carbohydrate residues attached would be 15 669. The most intense peak in this spectrum is observed at m/z 17 423 which corresponds to a predicted IL-3a form glycosylated with OS-2 at two sites. On the basis of the LSIMS results (Tables I and II), these sites must be Asn-16 and Asn-86. The peak with m/z 17 275 is consistent with a protein MH^+ ion bearing OS-1 and OS-2 (Table IV). This also is consistent with the identification of the ion peak with m/z 5166 recorded in the LSIMS spectrum of the disulfide-linked glycosylated tripeptide (Figure 7, Table III). Other ion peaks in the electrospray spectrum (Figure 8) can be assigned on the basis of the different glycoforms (Table IV). The electrospray data generally are consistent with the LSIMS data and previous studies in which the carbohydrates were removed and studied separately (Hogeland et al., 1992). The small peaks in the electrospray spectrum with masses of 18 154 and 18 305 Da could not be assigned.

Determination of Protein Heterogeneities of IL-3. To determine the structural differences between IL-3a and IL-

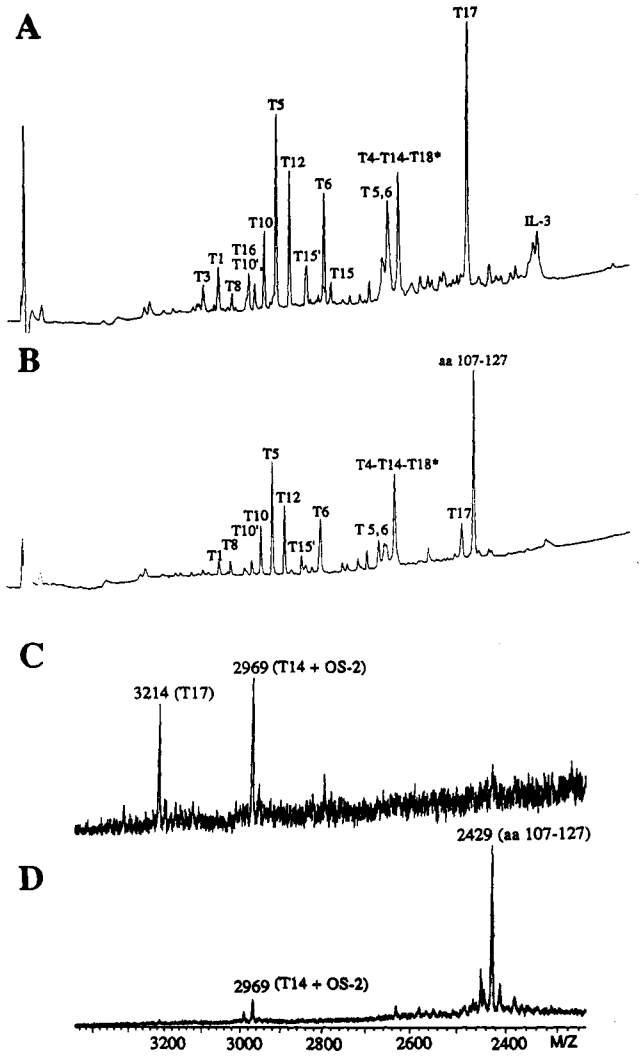

FIGURE 9: HPLC chromatograms of tryptic peptides from glycosylated IL-3a (A) and IL-3b (B). The Vydac C18 RP column was equilibrated with 0.1% aqueous TFA (solvent A, 80%) and acetonitrile with 0.08% TFA (solvent B, 20%). The elution gradient was as follows: 0–5 min, 80% A; 5–60 min, 80% A to 40% A; 60–65 min, 40% A. Peaks are assigned as in Table I; T4–T14–T18* is heterogeneously glycosylated. LSIMS spectra of the tryptic digests of glycosylated, reduced IL-3a (C) and glycosylated, reduced IL-3b (D) in the mass range 2200–3400 Da.

3b which lead to different migration on SDS–PAGE as well as different retention volumes on RP-HPLC columns, tryptic digests of the glycosylated and deglycosylated IL-3b were analyzed by LSIMS, either in the reduced or nonreduced states. No difference was observed between IL-3a and IL-3b either in the glycosylation patterns or in the disulfide linkages. Differences, however, were observed in the LSIMS spectra of the digests. For IL-3b a fairly intense protonated molecular ion with m/z 2429.5 was observed. For IL-3a, the [(M + H)$^+$] ion with m/z 3214 expected for T17 was observed (Figure 9C,D).

The C18 RP-HPLC chromatograms provided further evidence for the structural differences of IL-3a and IL-3b. In the case of IL-3a (Figure 9A), a large late eluting peak was identified by LSIMS analysis as T17. In comparison, the chromatogram of the IL-3b tryptic digest (Figure 9B) shows a much smaller peak for T17. However, there was a new later eluting peak which upon analysis by LSIMS was found to consist of a fragment with mass 2429.5 Da. Gas-phase Edman degradation showed that this was a peptide with an N-terminal sequence Phe-Tyr-Met-Val. The peptide, accordingly, was identified as the amino acid 107–127 sequence.

In order to determine if this peptide resulted from nonspecific trypsin cleavage at the C-terminal end of Ala-127 or if the cleavage had occurred prior to trypsin digestion, IL-3b was chemically treated with cyanogen bromide to cleave the protein after Met-109. LSIMS analysis of this mixture gave peaks at m/z 1986.8, 2014.6, and 2042.5 corresponding to amino acids 100–127 and once and twice formylated peptides at the amino groups (data not shown). This result shows that the cleavage behind Ala-127 in IL-3b occurred prior to treatment of the protein with trypsin. The amino acids 128–140 are connected to the rest of the protein by a disulfide linkage (Figure 10), as shown by the presence of the disulfide-linked tripeptide T4–T14–T18 (amino acids 14–22, 77–96, and 136–140) peaks in the LSIMS spectrum of the tryptic digest (Figure 7). In addition, the LSIMS spectrum of reduced, but nondigested IL-3b showed the presence of a protonated molecular ion with m/z 1294, indicating the presence of the cleaved C-terminal amino acid sequence (amino acids 128–140). The difference in the HPLC retention volumes, thus, can be explained on the basis of a change in the conformation of the two protein species. These results also help to clarify the reason for the additional band at 17–19 $\times$ 10^3 by SDS–PAGE; these experiments normally are carried out under reducing conditions.

DISCUSSION

In the present study, LSIMS analyses in combination with proteolytic digestion techniques were used to determine primary structural features of IL-3. These investigations have led to (1) an efficient mass spectrometric assay for IL-3 isolated from protein preparations, (2) confirmation of the peptides from IL-3 either directly in proteolytic digests or after isolation by HPLC, (3) the location of the glycosylation sites as well as the elucidation of the glycosylation patterns of the N-linked glycans, (4) the location of the disulfide linkages in IL-3, and (5) the identification of adventitious protein heterogeneities.

The most important step in preparing a protein for analysis is its isolation and purification. This requires a detection scheme which most often involves a bioassay technique specific for the protein in question. The disadvantage of the proliferation assay method, like the one Mosmann et al. (1987) used for IL-3, is the dependence on cell number (Miyajima et al., 1987) and the time requirements for the assay, which in this case involved at least 1 day. Fortunately, IL-3 is a small protein, and analysis of the proteolytic digests provided an efficient method for its detection. At least 10 fractions from the supernatants were collected by preparative HPLC and screened in less than an hour. About 30 pmol (0.5 μg) IL-3 was required for a LSIMS spectrum of the tryptic digest, so the method also provided excellent sensitivity.

The mass spectrometric approach for monitoring the isolation of IL-3 was fortunate in another sense; it resulted in the fairly rapid identification of a variant (IL-3b) which otherwise might have escaped detection or at least identification. The presence of the peptide with m/z 2429 led ultimately to the identification of the structural difference between IL-3a and IL-3b (Figure 10). It was shown that IL-3b is cleaved after Ala-127. This structural difference could not have been determined easily by any other means, because the C-terminal amino acids 128–140 were still connected by a disulfide bond to the core protein and the proliferation assay technique showed there was no difference in their bioactivities. Analysis of the mixture of IL-3a and IL-3b by SDS–PAGE led to a misinterpretation that heter-

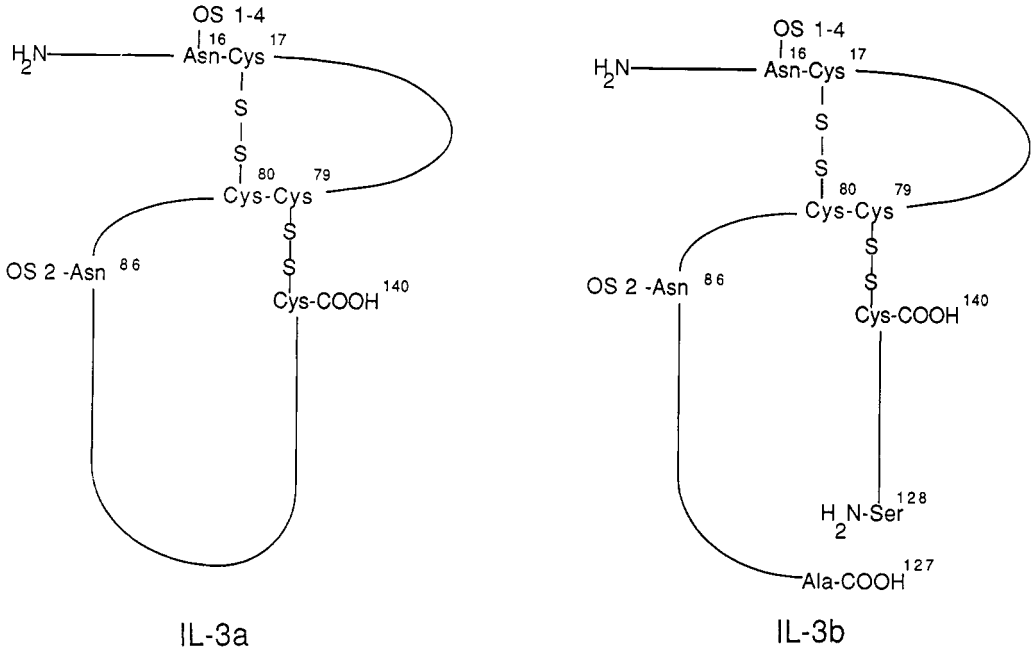

FIGURE 10: Simplified structural representations of IL-3a and IL-3b. IL-3b is distinguished from the main component IL-3a by a cleavage between Ala-127 and Ser-128. The amino acids 128–140 are connected through a disulfide linkage to the cleaved protein. The location of the disulfide linkage between Cys-17 and Cys-80 was proposed by Clark-Lewis and Schrader (1988). IL-3 is N-glycosylated at Asn-16 with oligosaccharides, OS-1 to OS-4 (Figure 4), and Asn-86 is partially glycosylated with OS-2.

ogeneous glycosylation was responsible for the presence of two bands at 20–21 × 10^3 and 17–19 × 10^3 (Miyajima et al., 1987). As is now known, the minor component with a molecular weight of 17–19 × 10^3 identified on the gel results from reduction of IL-3b and the loss of the 13 amino acid C-terminal end. The present study indicates that IL-3b probably is an artifact of the work-up procedure.

Peptide mapping by LSIMS analysis of proteolytic digests of glycosylated and deglycosylated IL-3 showed conclusively that Asn-16 and Asn-86 are glycosylated. In addition, there was no evidence for glycosylation of either Asn-44 or Asn-51. The heterogeneity at each of the glycosylated sites could be determined by the presence of the ion peaks of different glycoforms in the spectra. Evidence indicating that Asn-16 is linked to four different chitobiose core oligosaccharides, i.e., Man_{2-4}[Fuc]$GlcNAc_2$ and $Man_2GlcNAc_2$, and that Asn-86 is mostly glycosylated with Man_2[Fuc]$GlcNAc_2$ was obtained both from LSIMS data on the glycopeptides and from the electrospray data on the intact proteins. These results are corroborated by the identification of the oligosaccharides after treatment of the protein by N-glycanase (Hogeland, et al., 1992). Subsequent experiments in which the oligosaccharides were chromatographically separated and isolated and then identified by mass spectrometry have confirmed their identities (K. E. Hogeland, unpublished results).

The isolation and mass spectrometric analysis of the fully glycosylated tryptic tripeptide T4–T14–T18 (Figure 7) confirmed earlier observations on possible disulfide linkages. The disulfide bonds between T4 and T14 as well as between T14 and T18 could be reduced in situ thus allowing the identification of the peptides that were covalently linked. Clark-Lewis et al. (1987) showed by amino acid substitution that Cys-17 is connected to Cys-80. However, it was concluded that Cys-79 and Cys-140 were most likely present in the reduced state. The results from the present investigation show that all cysteines present form disulfide linkages, and if Cys-17 and Cys-80 are linked as shown by the mutation experiments (Clark-Lewis et al., 1987), then Cys-79 and Cys-140 must be connected. By the present technique alone, however, it is not possible to determine which cysteines are connected, only that both the C-terminal and N-terminal cysteines are connected to either Cys-79 or Cys-80.

Jarvis and Summers (1989) and Kuroda et al. (1990) have reported that insect cells have the necessary enzymes to trim oligosaccharides but apparently not to synthesize complex oligosaccharide chains. Fowl plaque virus hemagluttin that was expressed in *Spodoptera frugiperda* cells using a baculovirus recombinant was found to be largely endo H sensitive (Kuroda et al., 1990). Chromatographic analysis of the oligosaccharide side chains showed the presence of mostly immature $Man_{5-9}GlcNAc_2$. Truncated oligosaccharide cores, $Man_3GlcNAc_2$, and Man_3[Fuc]$GlcNAc_2$ (Kuroda et al., 1990) were also found, indicating that the fucosyltransferases necessary for the trimming of oligosaccharides are present. The studies reported here also show fucosylated oligosaccharides on a protein expressed by a baculovirus recombinant infected silkworm larvae. Fucosylation of the oligosaccharides on glycoproteins from the baculovirus expression system, therefore, may be common regardless of the cells infected.

ACKNOWLEDGMENT

We thank Finnigan Corp. for the electrospray MS analysis and express appreciation to Kathy Rawson and David Barnes for providing the MC/9 cell line and for their help in using the cell culture facilities and to Dr. Michael Glocker for many helpful discussions. This paper is Oregon Agricultural Experiment Station Publication No. 9823.

REFERENCES

Biemann, K., & Martin, S. (1987) *Mass Spectrom. Rev. 6*, 1–76.

Biemann, K., Gibson, B. W., Mathews, W. R., & Pang, H. (1985) *Mass Spectrometry in the Health and Life Sciences* (Burlingame, A. L., & Castagnoli, N., Eds.) Proceedings of an International Symposium, San Francisco, CA, Sept 9–13, 1984.

Carr, S. A., & Biemann, K. (1984) *Methods Enzymol. 106*, 29–58.

Carr, S. A., & Roberts, G. D. (1986) *Anal. Biochem. 157*, 396–406.

Carr, S. A., Roberts, G. D., Jurewicz, A., & Frederick, B. (1988) *Biochimie 70*, 1445–1454.
Clark-Lewis, I., & Schrader, J. W. (1988) *Lymphokines 15*, 1–37.
Clark-Lewis, I., Kent, S. B. H., & Schrader, J. W. (1984) *J. Biol. Chem. 259*, 7488–7494.
Clark-Lewis, I., Lopez, A. F., Vadas, M., Schrader, J. W., Hood, L., & Kent, S. B. H. (1987) in *Molecular Basis for Lymphokine Action* (Webb, D., & Pierce, C., Eds.) Humana Press, Clifton, NJ.
Gibson, B. W., & Biemann, K. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1956–1960.
Hogeland, K. E., Arbogast, B., & Deinzer, M. L. (1992) *J. Am. Soc. Mass Spectrom. 3*, 345–352.
Ihle, J. N., & Weinstein, Y. (1986) *Adv. Immunol. 39*, 1–50.
Ihle, J. N., Keller, J., Henderson, L., Klein, F., & Palaszynski, E. (1982) *J. Immunol. 129*, 2431–2436.
Ihle, J. N., Keller, J., Oroszlan, S., Henderson, L. E., Copeland, T. D., Fitch, F., Prystowsky, M. B., Goldwasser, E., Schrader, J. W., Palaszynski, E., Dy, M., & Lebel, B. (1983) *J. Immunol. 131*, 282–287.
Kuroda, K., Geyer, H., Geyer, R., Doerfler, W., & Klenk, H.-D. (1990) *Virology 174*, 418–429.
Laemmli, U. K. (1970) *Nature 227*, 680–685.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265.
Mann, M., Meng, C. K., & Fenn, J. B. (1989) *Anal. Chem. 61*, 1702–1708.
McCloskey, J. A. (1990) *Methods Enzymol. 193*, 539–738.
Miyajima, A., Schreurs, J., Otsu, K., Kuodo, A., Arai, K.-I., & Maeda, S. (1987) *Gene 58*, 273–281.
Miyajima, A., Yokota, T., Otsuka, T., Miyatake, S., Takeba, Y., Hagiwara, H., Schreurs, J., Kaibuchi, K., Lee, F., Arai, N., Mosmann, T., & Arai, K.-I. (1988) in *Immunology Series, Biological Response Modifiers and Cancer Therapy* (Chiao, J. W., Ed.) Vol. 40, pp 103–148, Marcel Dekker, New York.
Morris, H. R., & Pucci, P. (1985) *Biochem. Biophys. Res. Commun. 126*, 1122–1128.
Mosmann, T. (1983) *J. Immunol. Methods 65*, 55–63.
Rennick, D. M., Lee, F. D., Yolota, T., Arai, K., Cantor, H., & Nabel, G. J. (1985) *J. Immunol. 134*, 910–914.
Svoboda, M., Przybylski, M., Schreurs, J., Miyajima, A., Hogeland, K., & Deinzer, M. (1991) *J. Chromatogr. 562*, 403–419.
Whetton, A. D., & Dexter, T. M. (1986) *Trends Biochem. Sci. 11*, 207–211.
Wodnar-Filipowicz, A., Heusser, C., & Moroni, C. (1989) *Nature 339*, 150–152.
Yazdanparast, R., Andrews, P. C., Smith, D. L., & Dixon, J. E. (1986) *Anal. Biochem. 153*, 348–353.
Yazdanparast, R., Andrews, P. C., Smith, D. L., & Dixon, J. E. (1987) *J. Biol. Chem. 262*, 2507–2513.