ELSEVIER

Journal of Chromatography B, 782 (2002) 405–418

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Use of a lectin affinity selector in the search for unusual glycosylation in proteomics

Li Xiong*, Fred E. Regnier

Department of Chemistry, Purdue University, 1393 Brown Building, West Lafayette, IN 47907, USA

Abstract

The purpose of the work described in this paper was to develop a new approach to the identification of glycoprotein with particular types of glycosylation. The paper demonstrates *N*-glycosylation sites in a glycoproteins can be identified by (1) proteolysis with trypsin, (2) lectin affinity selection, (3) enzymatic deglycosylation with peptide-*N*-glycosidase F (PNGase F) in buffer containing 95% $H_2^{18}O$, which generates deglycosylated peptide pairs separated by 2 or 4 amu, (4) reversed-phase separation of the peptide mixture and MALDI mass analysis, (5) MS–MS sequencing of the ion pairs, and (6) identification of the parent protein through a database search. This process has been tested on the selection of glycopeptides from lactoferrin and mammaglobin, and the identification of the ion pairs of fetuin glycopeptides. Glycosylation sites were identified through PNGase hydrolysis in $H_2^{18}O$. During the process of hydrolyzing the conjugate, Asn is converted to an aspartate residue with the incorporation of ^{18}O. However, PNGase F was observed to incorporate two ^{18}O into the β-carboxyl groups of the Asp residue. This suggests that the hydrolysis is at least partially reversible.

Keywords: Proteomics; Glycosylation; Lectin affinity selector; Glycoproteins

1. Introduction

Glycoproteins and unusual forms of glycosylation can be important indicators of disease. It has been shown in vitro that the level of β(1→6)GlcNAc branched oligosaccharides increases in breast cancer cells along with a loss of contact inhibition and that this increase in branching may be associated with the development of metastatic cells [1]. In another case, up-regulation of the attachment of *N*-acetylgalactosamine terminal residues of glycoproteins has been associated with cancer [2]. Again, it is thought that this unusual form of glycosylation may be elevated in metastatic cells. Similar results have been seen with human breast and colorectal cancer cells grown in vitro and in animal models [3].

This is important because the focus in cancer marker research for many years has been more on the protein portion of markers. Much less effort is being expended on examining the connection between unusual glycosylation and tumorigenesis. This is probably because alterations of glycan structure or concentration are more difficult to notice than similar changes in polypeptides. In fact, it is likely that some of the known cancer markers are glycosylated in unknown ways. Developing methods that examine specific types of glycosylation is the focus of this paper.

It has been shown that fucosylation of the GlcNAc residue attached to asparagine in complex N-linked

*Corresponding author. Tel.: +1-765-494-1648; fax: +1-765-494-0359.

E-mail address: slentz@purdue.edu (L. Xiong).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00671-2

glycoproteins occurs with the formation of α-Fuc(1→6)-β-GlcNAc-Asn- [4]. Although fucosylation of proteins in general has been found to increase in patients with cancer, derivatization of a cancer marker with α-Fuc(1→6)-β-GlcNAc-Asn- has not been reported. The question examined here was whether there could be an unidentified association of this unusual type of cancer-associated glycosylation with a known cancer marker. If so, it strongly suggests this is not a coincidence.

Several experimental protocols were considered. The obvious approach would be to examine glycosylation patterns in all known cancer markers. But glycoproteins can exist in multiple glycoforms and it is possible that the isolated protein is a single glycoform or does not contain a particular type of glycosylation. The question of whether α-Fuc(1→6)-β-GlcNAc-Asn- would be present in only one or all of the glycoforms of a protein complicates this type of search. The idea of examining isolated cancer markers for the presence of α-Fuc(1→6)-β-GlcNAc-Asn- was rejected because of the question of whether all glycoforms would be present. The strategy finally selected was to use a structure specific affinity selector directed at the α-Fuc(1→6)-β-GlcNAc-Asn- moiety and select glycopeptides that bear this group from tryptic digests of serum. Serum is known to contain cancer markers [3]. Selected glycopeptides would then be identified and that information used to search databases to determine whether any of the peptides were derived from a known cancer marker. This approach is unique in that it structurally targets a low probability, unusual event of probable biological significance and makes it possible to identify proteins with which this event is associated.

2. Materials and methods

2.1. Materials

Human sera, human lactoferrin, bovine fetuin, L-1-tosylamide-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin, *Lotus tetragonolobus* agglutinin (LTA), Tris[hydroxy-methyl]aminomethane (Tris base), Tris[hydroxy-methyl]aminomethane hydrochloride (Tris acid), iodoacetic acid, cysteine, dithiothreitol (DTT), *N*-tosyl-L-lysine chloromethyl ketone (TLCK), manganese chloride, calcium chloride, fucose and α-cyano-4-hydroxycinnamic acid were purchased from Sigma (St. Louis, MO, USA). Lichrospher Si 1000 (10 μm, 1000 Å) was obtained from Merck (Darmstadt, Germany). 3-Aminopropyltriethoxy silane, polyacrylic acid (PAA), and dicyclohexyl carbodiimide (DCC) were purchased from Pierce (Rockford, IL, USA). Peptide-*N*-glycosidase F (PNGase F) was obtained from Boehringer-Mannheim (Mannheim, Germany). The peptide human andrenocorticotropic hormone fragment (ACTH) 18–39 and human angiotensin I were purchased from Bachem (Torrance, CA, USA). Enriched $H_2^{18}O$ (95%) was purchased from Isotec (Miamisburg, OH, USA). All commercially available reagents were used directly without purification. HPLC-grade acetonitrile, toluene, dioxane, and dimethyl sulfoxide (DMSO) were purchased from Mallinckrodt Baker (Phillipsburg, NJ, USA).

2.2. Synthesis of lectin columns

The procedure used to immobilize the agglutinin (LTA) is a modification of methods in the literature [5,6]. One gram Lichrospher Si 1000 was activated with 40 ml 6 *M* HCl under N_2 for 6 h at room temperature. After washing with water to neutral pH, the silica was filtered and dried in the oven overnight at 120 °C. The resulting particles were treated with 5% 3-amino-propyl triethoxysilane in 10 ml of sodium-dried toluene for 5 h in 90 °C to produce 3-aminopropyl silane (APS)-derivatized silica (3.44×10^{-4} mol APS/m^2 silica surface). Elemental analysis of the APS product showed a carbon content of 0.48%; hydrogen content of 0.12%; nitrogen of 0.13%.

Polyacrylic acid (0.503 g; M_r 450 000) and *N*-hydroxyl succinimide (1.672 g) were dissolved in 50 ml DMSO, then 6.0 g DCC dissolved in 10 ml DMSO was added and stirred at room temperature for 3 h. The resulting mixture was filtered to get eliminate dicyclohexylurea (DCU), a white precipitate. The filtrate, which contained succinimidyl polyacrylate, was added to the APS-modified silica. After shaking at room temperature for 12 h, the silica was filtered and washed with 50 ml of DMSO, dioxane,

and deionized H_2O, respectively. This step produced a polyacrylate-coated silica with residual *N*-acyloxysuccinimide groups, specified as PAA–NAS silica (7.94×10^{-4} mol succinimidyl/m^2 silica surface). Elemental analysis: C 2.48%; H 0.33%; N 0.43%.

LTA was immobilized on the silica surface in the following way. LTA (100 mg) was dissolved in 20 ml of 0.1 *M* $NaHCO_3$ buffer containing 0.2 *M* fucose and mixed with roughly 1 g of PAA–NHS silica. The reaction was allowed to proceed with vigorous shaking for 12 h, after which the immobilized LTA sorbent was recovered by centrifugation. The remaining active succinimidyl ester sites were blocked with 0.1 *M* ethanolamine (pH 8.0) after agitating for 30 min. The silica particles were then washed with 1 *M* NaCl to remove the uncoupled ligands. Finally the sorbent was washed with loading buffer (0.2 *M* Tris, 0.1 *M* NaCl, 1 m*M* $CaCl_2$, 1 m*M* $MnCl_2$, 0.2% NaN_3, pH 7.5) and packed into a 4.6×50-mm PEEK column. Elemental analysis: C 4.06%; H 0.59%; N 0.94%.

2.3. Proteolysis

One ml of human serum (Sigma S7023) was adjusted to pH 8.0 with 0.1 *M* Tris buffer containing 8 *M* urea before proteolysis. In the case of the pure glycoprotein samples examined, 5 mg of either human lactoferrin or bovine fetuin was dissolved with 0.1 *M* Tris buffer containing 8 *M* urea, pH 8.0. DTT was added to the above protein samples to a final concentration of 10 m*M* to reduce disulfide bonds. After incubation at 50 °C in a water bath for 1 h, iodoacetic acid was added to 20 m*M* in concentration and the samples were incubated in darkness on ice for a further 2 h. The alkylation reaction was then quenched with 10 m*M* cysteine for 30 min at room temperature. After diluting with 0.1 *M* Tris digestion buffer (pH 8.0), buffer containing 10 m*M* $CaCl_2$ was added to a final urea concentration of 0.8 *M* along with TPCK-treated trypsin (mass ratio between trypsin and total protein is 1:50) and incubated at 37 °C for 24 h. Proteolysis was stopped by adding TLCK protease inhibitor at a molar concentration exceeding that of trypsin by 2-fold. Samples were stored in liquid nitrogen until analyzed.

2.4. Chromatography

All chromatographic separations were performed using a Biocad Micro-Analytical Workstation from PE Biosystems (Framingham, MA, USA). Tryptic digested human serum (0.5 ml) or human lactoferrin (0.2 ml) was injected onto the LTA column after equilibrating with the loading buffer containing 0.2 *M* Tris, 0.1 *M* NaCl, 1 m*M* $CaCl_2$, 0.1 *M* $MnCl_2$, pH 7.5. Unbound peptides were eluted with 10 ml of loading buffer. Bound glycopeptides were then eluted from the affinity column with a mobile phase containing 0.2 *M* fucose in a solution of 0.1 *M* Tris, 0.1 *M* NaCl, 1 m*M* $CaCl_2$, and 1 m*M* $MnCl_2$ at pH 7.5 and directly transferred to a 4.6×250-mm Pep-Map C_{18} column (Applied Biosystems). Glycopeptides of human lactoferrin were eluted from the reversed-phase column at 1.0 ml/min in 50 min linear gradient ranging from 5% acetonitrile in 0.1% aqueous TFA (buffer A) to 90% acetonitrile in 0.1% aqueous TFA (buffer B). The separation of human serum glycopeptides were achieved with a 120-min gradient from 5% acetonitrile in 0.1% aqueous TFA (buffer A) to 60% acetonitrile containing 0.1% aqueous TFA (buffer B). Eluted peptides were monitored at 214 nm and fractions were manually collected for MALDI-MS analysis.

2.5. Deglycosylation by N-glycosidase F (PNGase F)

Glycopeptides were speed-dried under high vacuum and reconstituted in 80 μl 100 m*M* sodium phosphate, pH 7.2, containing 25 m*M* EDTA in the present of 95% atom normalized water ($H_2^{18}O$). The reaction mixture was incubated with 2 μl (10 U) of PNGase-F for 24 h at 37 °C.

2.6. MALDI-mass spectrometry

Matrix-assisted laser desorption ionization (MALDI)-mass spectrometry analyses were performed on a Voyager DE-RP BioSpectrometry workstation from PerSeptive Biosystems. The matrix solution was acetonitrile (ACN)–water (50:50) containing 0.1% TFA saturated with α-cyano-4-hydroxycinnamic acid. A 1-μl matrix solution was spotted on top of samples and the solvent allowed to evaporate in air

before placing the sample plate in the mass spectrometer. Glycopeptides were analyzed in the linear, positive ion mode by delayed extraction using an accelerating voltage of 20 kV. Deglycosylated peptides were also analyzed in the reflector mode with delayed extraction. External calibration was achieved using a mixture of standard peptides containing angiotensin I (M_r 1296.68) and ACTH 18–39 (M_r 2465.70).

2.7. ESI-MS analysis

Samples were analyzed by an electrospray ionization (ES) source on a QSTAR hybrid Q-TOF MS (Applied Biosystems/PE Sciex, Foster City, CA, USA). Samples were dissolved in ACN–H_2O (50:50) containing 0.1% formic acid, and infusion injected with a syringe pump flowing at 10–30 μl/min. The QSTAR was operated above 8000 in resolution with a mass accuracy of 10–30 ppm using an external calibration maintained for 24 h. MS–MS sequencing was achieved by selecting the parent ion in the quadruple in the low-resolution mode to get the entire isotope cluster into the collision cell. The typical ion spray voltage used was 5000 V with a collision energy of 30–60 eV to fragment peptides. Nitrogen gas was used for the collision gas and typical pressures in the collision cell during MS/MS were from 4 to 6×10^{-6} Torr. Sequences were obtained from the MS–MS spectrum through manual interpretation and with the MASCOT software (Matrix Sciences, London, UK)

3. Results and discussion

3.1. The analytical strategy

The experimental protocol described here is based on the hypothesis that the simplest way to recognize and identify proteins containing the α-Fuc(1→6)-β-GlcNAc-Asn- moiety would be through some type of affinity selection procedure that targets this specific oligosaccharide. Lectin affinity selection has been demonstrated to be an effective analytical tool to isolate glycoproteins and glycopeptides with unique types of glycosylation [7–11]. Binding specificity is achieved by a highly selective interaction of the carbohydrate moiety with the binding pocket of the lectin, causing lattice formation and precipitation of a glycoprotein or glycopeptide on the lectin surface [12]. *Lotus tetragolonobus* agglutinin (LTA) has been shown with model glycopeptides to have high affinity for fucose-containing glycoconjugates, especially α-Fuc(1→6)-β-GlcNAc-Asn-, and when immobilized on a suitable support can be used in affinity chromatography [13]. An LTA affinity chromatography column was prepared by immobilizing the lectin on a silica-based support through peptide bond formation. This was achieved by preactivating the polyacrylic acid phase on the sorbent surface with *N*-hydroxysuccinimide. Amide bond formation with lysine residues on the surface of the protein occur in the pH range from 8 to 10.

Human serum samples were reduced, alkylated, and trypsin digested. The tryptic digest was then applied to the LTA affinity column for selection of peptides containing α-Fuc(1→6)-β-GlcNAc-Asn. Non-binding peptides were discarded. Those glycopeptides bound to the lectin column were eluted with a buffer containing 0.2 *M* fucose in a solution of 0.1 *M* Tris, 0.1 *M* NaCl, 1 m*M* $CaCl_2$, and 1 m*M* $MnCl_2$ (pH 7). The function of fucose was to act as a biospecific displacer.

It is common that glycopeptides in a tryptic digest may be present in multiple glycoforms. The question being examined here is whether any, or all of these glycoforms are fucosylated. For this reason it is possible that the bound fraction eluted from the affinity column can contain multiple glycoforms of the same peptide. Although glycoforms of a peptide are generally not resolved during reversed-phase chromatography (RPC), those that differ in molecular mass are easily seen during mass spectrometry. The problem is in determining the molecular mass and sequence of the peptide portion of the conjugate needed for identifying the parent protein. The presence of an oligosaccharide can complicate peptide sequencing. A solution to this problem is to enzymatically deglycosylate the conjugate prior to RPC and mass spectrometry.

3.2. Analysis of model proteins

Selectivity of the lectin affinity column was evaluated with lactoferrin, a fucose containing

glycoprotein secreted in human mammary glands. As a member of the human transferrin family, lactoferrin is composed of 711 amino acids, in which there are two possible *N*-glycosylation sites: asparagine 157 and asparagine 498 according to the SWISS-PROT database.

A reversed-phase chromatogram of tryptic-digested lactoferrin presents a complicated profile, as seen in Fig. 1a. But after selection with the LTA affinity column, the complex mixture is simplified to a single broad peak (Fig. 1b), eluting at roughly 68.4% acetonitrile. The fact that the peak is very broad, and shows other partially resolved species is thought to be due to the multiplicity of glycoforms. This is confirmed in the MALDI-MS spectrum (Fig. 2a), showing the major species at mass to charge ratios (m/z) 5046, 5192, 5338, and 5485 associated with this chromatography peak. But the GlcoSuite database suggests there should only be four glycoforms. There are obviously more than that in Fig. 2a. Moreover, some of these glycoforms vary from the structures suggested in GlycoSuite by 42 atomic mass units (amu). This suggests these structures might vary in acetylation. Apparently the glycosyla-

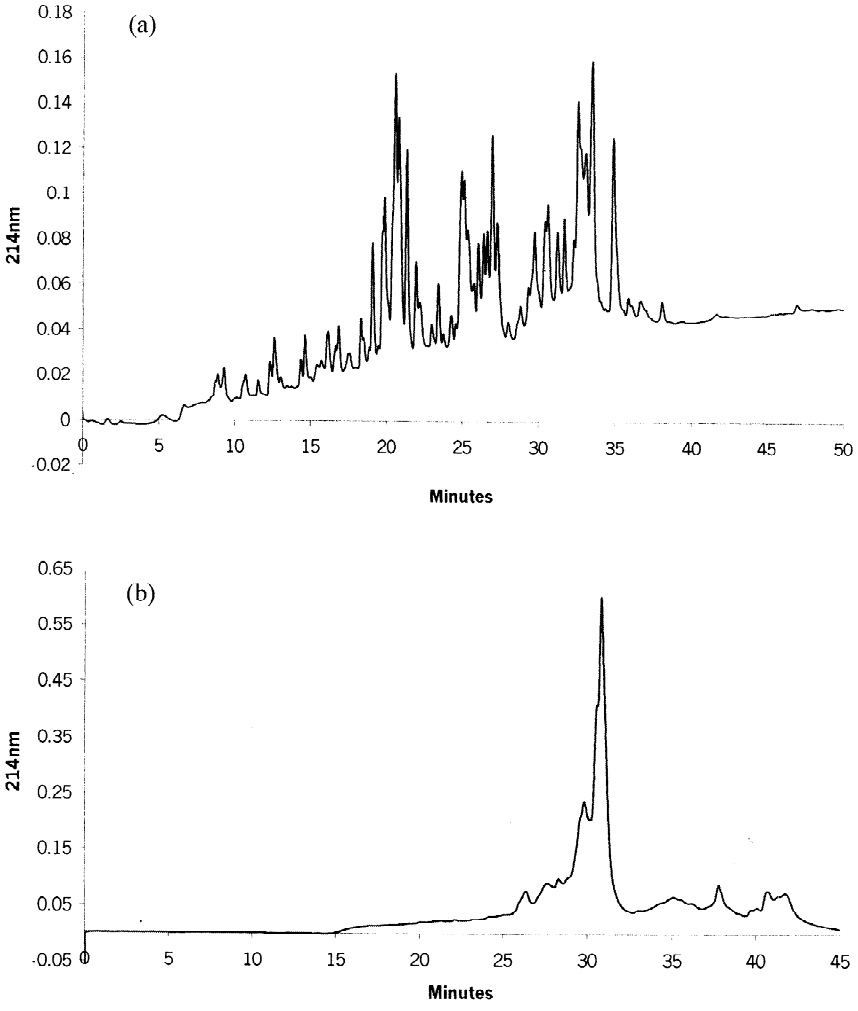

Fig. 1. Reversed-phase chromatograms of lactoferrin tryptic digest: (a) without selection; (b) after selection with an immobilized LTA column and release onto the reversed-phase column. Peptides were eluted from a 4.6×250-mm I.D. C_{18} column with a 50-min linear gradient ranging from 5% ACN in 0.1% aqueous TFA to 90% ACN in 0.1% aqueous TFA at a flow-rate of 1.0 ml/min.

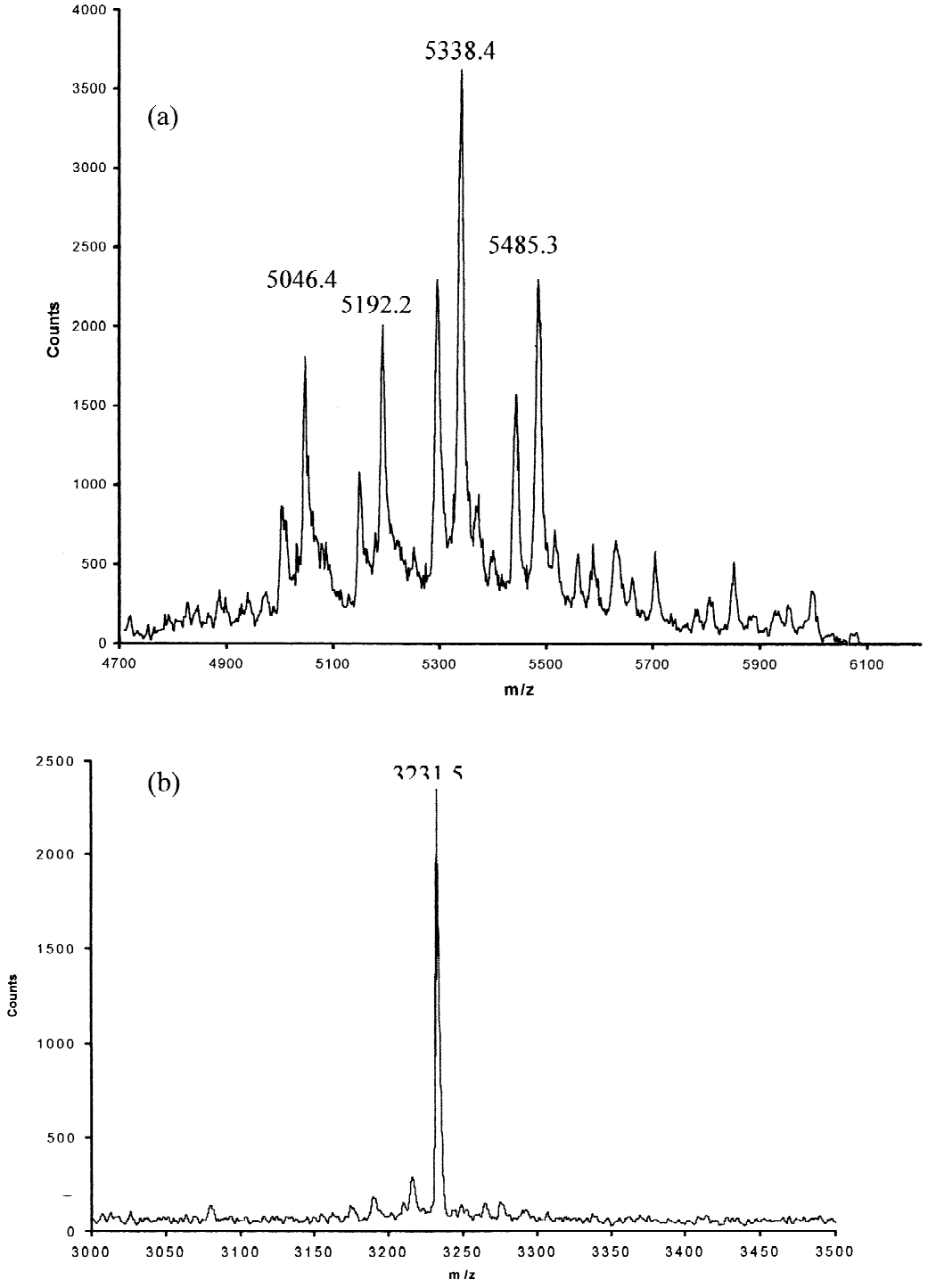

Fig. 2. MALDI-mass spectrum of LTA selected glycopeptide from lactoferrin: (a) before deglycosylation with PNGase F, (b) after deglycosylation.

tion of lactoferrin is still not completely characterized. The problem in proteomics is how to identify the peptides at glycosylation sites when the oligosaccharide portion of a glycopeptide is not completely characterized. Although it is possible to determine the molecular mass of glycopeptides, it is difficult to determine their amino acid sequence from a DNA database unless a specific oligosaccharide of identified molecular mass is known to be associated with the peptide. That is not the case here. Database searches for peptide sequence are greatly facilitated by knowing the molecular mass of a peptide.

One way to eliminate the uncertainty of oligosaccharide molecular mass is to deglycosylate the glycopeptides and to analyze the peptides alone. The peptide *N*-glycosidase F enzyme from *Flavobacterium meningosepticum*, designated PNGase F [14], hydrolyzes the glucosylamine linkage through which N-linked oligosaccharides are coupled to peptides at the consensus sequence Asn-X-Ser/Thr. X is any amino acid except proline [15]. During the course of reaction, an aspartate residue is generated at the original site of glycosylation, a carbohydrate-free peptide is obtained, and an intact oligosaccharide with the di-*N*-acetylchitobiose unit at the reducing end is produced [16,17]. This enzyme hydrolyzes all classes of *N*-glycoproteins, and glycopeptides, including high mannose, hybrid, and complex oligosaccharides [18]. Because Asp is 1 Da larger than Asn in molecular mass, peptides derived from PNGase F hydrolysis will be 1 Da higher in molecular mass than predicted by a DNA database in which the amino acid is present as Asn.

Assuming that the species in Fig. 2a represent glycoforms in which the oligosaccharide is coupled to a peptide through an *N*-aspartamide linkage, the entire fraction was subjected to treatment with PNGase F to deglycosylate any glycopeptides in the fraction. Subsequent to PNGase F treatment the relatively complex spectrum in Fig. 2a collapsed to that of a single peptide with a molecular mass of 3231.5538 as in Fig. 2b. According to DNA databases, lactoferrin should yield a tryptic peptide, ranging between amino acids 142 and 171 in the parent protein, with the sequence TAGWNVPIGTLRPFL**N**WTGPPEPIEAAVAR. The theoretical molecular mass of this peptide is 3230.711. The asparagine residue identified with the bold N is the position of glycosylation and was actually an aspartate residue in the analyzed peptide.

Based on the fact that PNGase F converts asparagine to aspartate during the course of cleaving a glycopeptide conjugate, it is expected that the molecular mass of the selected peptide would increase by 1 Da after hydrolysis with PNGase F and that the mass of the peptide should be $m/z=3231.6957$. The relative error between the theoretical and experimentally derived molecular masses vary by −0.00439%. A search of the GycoSuite database based on mass shows that there are four oligosaccharide structures attached to N-157 of lactoferrin (Fig. 3). They all have a fucose α1–6 linked to the inner most GlcNAc that is attached to asparagine. But it is important to note that more than four glycoforms were seen in Fig. 2a. Do they all contain fucose in an α1–6 linkage? There is no chemical data that unequivocally establishes the presence of the fucose α1–6 linkage. This is the nature of affinity chromatography, immunological assays, and DNA hybridization-based identification. Non-specific binding could cause non-fucose containing peptides to bind,

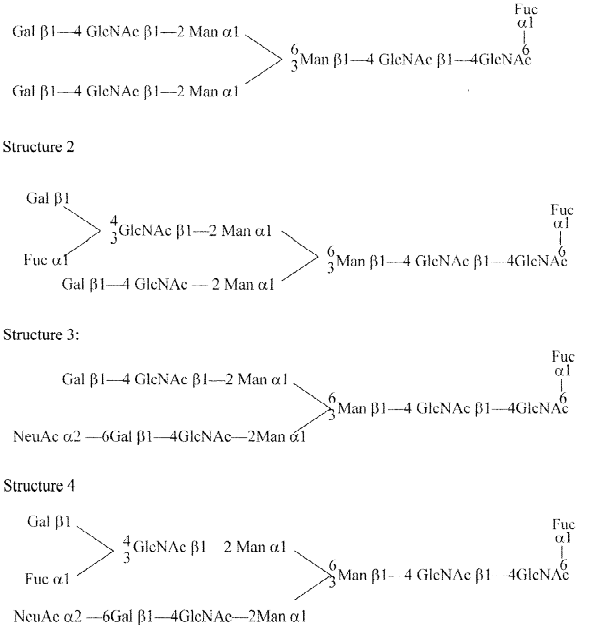

Fig. 3. Four different glycan structures 1–4, corresponding to four glycoforms in (a), from left to right.

but it is unlikely that four non-fucose containing peptides differing from those in Fig. 3 by 42 amu would be encountered.

An issue in using PNGase for structure elucidation is how to preserve information at the site of glycosylation for later identification during mass spectrometry. This has been done in the past by using ^{18}O labeling during PNGase F hydrolysis in 40% $H_2^{18}O$ [19,20]. When PNGase F cleaves Asn-linked carbohydrates at the β-aspartylglucosylamine linkage, ^{18}O is incorporated at β-carboxyl groups during conversion to aspartate. Subsequent proteolysis did not alter the pattern of ^{18}O labeling when examined by FAB mass spectrometry [19,20].

The efficacy of this approach with fucosylated peptides was examined in these studies using the model glycoprotein bovine fetuin. Fetuin, also called asialofetuin and α_2-HS-glycoprotein, is a fucosylated glycoprotein [20] with 359 amino acids containing N-linked glycosylation sites at Asn 99, 156, and 176 and several O-linked glycosylation sites. One of the glycopeptides, selected by affinity chromatography with the LTA column is seen in Fig. 4a. After deglycosylation with PNGase F in an appropriate buffer containing 95% $H_2^{18}O$, roughly 86% of the β-carboxyl groups of aspartic acid formed during hydrolysis contained ^{18}O (Fig. 4b). A dominant ion arising from ^{18}O enrichment is apparent in the spectrum. MS–MS analysis of the peptide found the sequence to be KLCPDCPLLAPLNDSR (Fig. 5) based on the conversion of asparagine to aspartate during hydrolysis, the presence of ^{18}O in the newly formed carboxyl, and the fact that the cysteine residue had been alkylated with iodoacetic acid. The parent ion was a doubly charged ion of m/z 937.5, i.e., the monoisotopic mass was 1874. In the spectrum, b ions (N-terminal sequence ions) are of higher relative abundance than y ions (C-terminal sequence ions). Daughter ions containing the newly formed aspartate residue labeled with ^{16}O and ^{18}O (designated Asp*) should appear as ion pairs $(M+1)^+$ and $(M+3)^+$. These ions are designated $(M+1)^+$ and $(M+3)^+$ based on M being the molecular mass of the original asparagine contain peptide prior to hydrolysis. The increased mass in the aspartate residue of the peptide allows the position of carbohydrate attachment to be determined easily.

3.3. The search for a fucosylated cancer marker

Human serum is a very complicated biological fluid that is thought to contain 10 000 or more proteins. Tryptic digestion of human serum would theoretically generate more than one million peptides alone, even without miscleavages [21]. Targeted analyses through affinity selection with the lectin affinity column should be a great asset in reducing sample complexity.

A pooled human serum sample was reduced, alkylated, tryptic digested, and the digest passed through the LTA lectin column. Again the selected glycopeptides were displaced with 0.2 *M* fucose in 0.1 *M* Tris buffer (pH 7.5), and transported directly to the reversed-phase chromatography (RPC) column (Fig. 6). Chromatograms of the tryptic digest of serum and that of glycopeptides selected with the LTA column show that affinity selection substantially simplifies the mixture (Fig. 7a,b). Glycopeptides were eluted with a linear solvent gradient ranging from 5 to 95% acetonitrile in an aqueous solution containing 0.1% trifluoroacetic acid.

Enzymatic deglycosylation of the glycopeptides was carried out in 100 m*M* phosphate buffer containing ^{18}O-labeled water during the course of identifying the peptides involved and their sites of glycosylation. It was unexpected that some of the peptides incorporated 2 mol of ^{18}O (Fig. 7) into the newly formed aspartate residue (Asp*). Although this has no impact on the utility of ^{18}O-labeling in identifying glycosylation sites, it is an important fact to know in spectral interpretation. Incorporation of a second mole of ^{18}O during PNGase hydrolysis requires that the reaction be reversible, at least to the extent that covalent bond formation with the product occurs reversibly. This behavior mimics trypsin, which is known to incorporate 2 mol of ^{18}O into peptides during hydrolysis [22].

A reasonable hypothesis for this phenomenon would be that PNGase F reversibly associates with the newly formed aspartate residue and that resonance structures are formed that allow further exchange of ^{18}O from $H_2^{18}O$ into the carboxyl group. This process can occur many times and result in almost complete labeling with a second mole of ^{18}O (Fig. 8). The fact that incorporation of a second mole

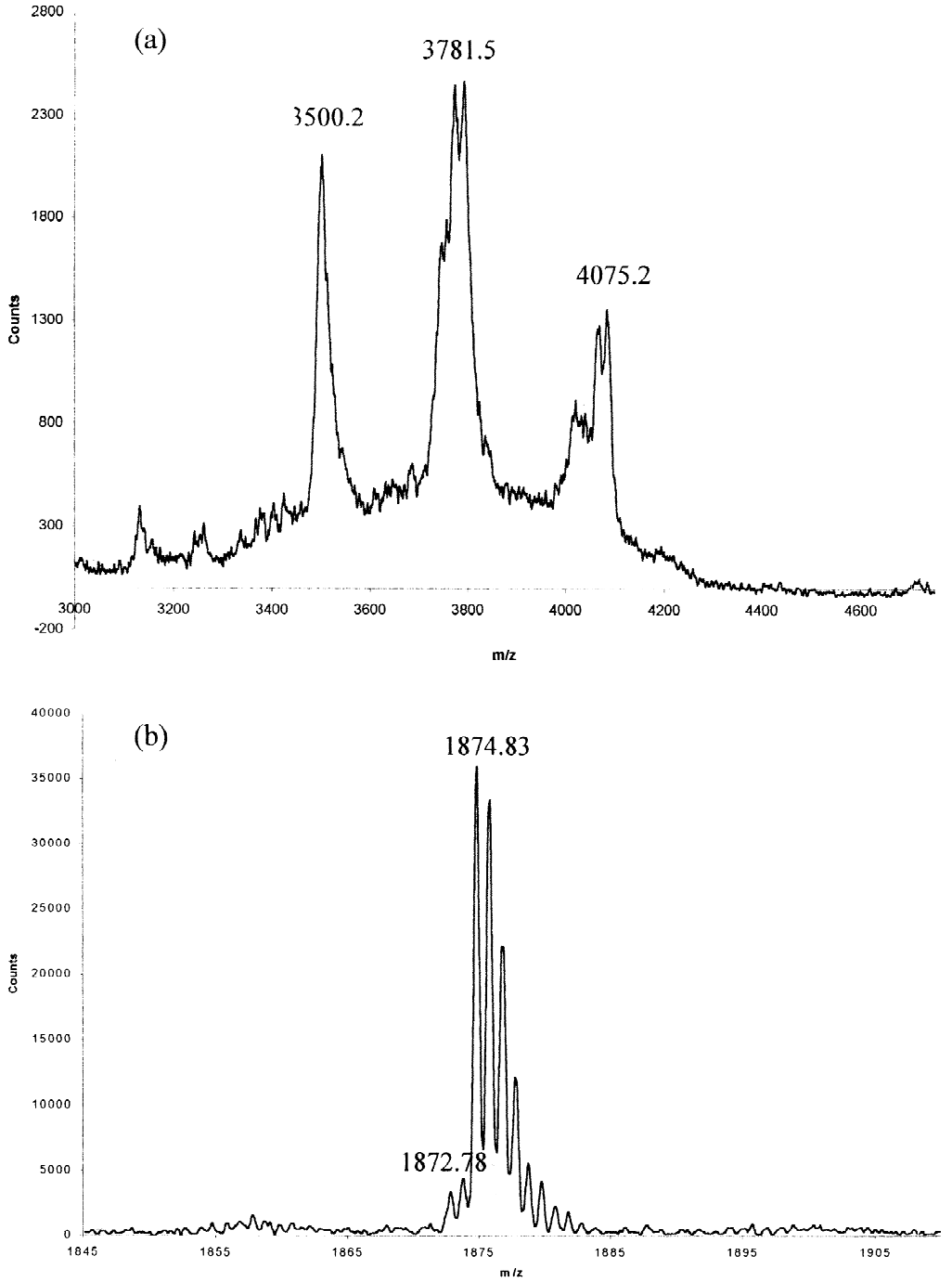

Fig. 4. MALDI-mass spectrum of LTA selected glycopeptide from fetuin: (a) before deglycosylation, (b) after deglycosylation. MALDI with PNGase F in the presence of 95% $H_2^{18}O$.

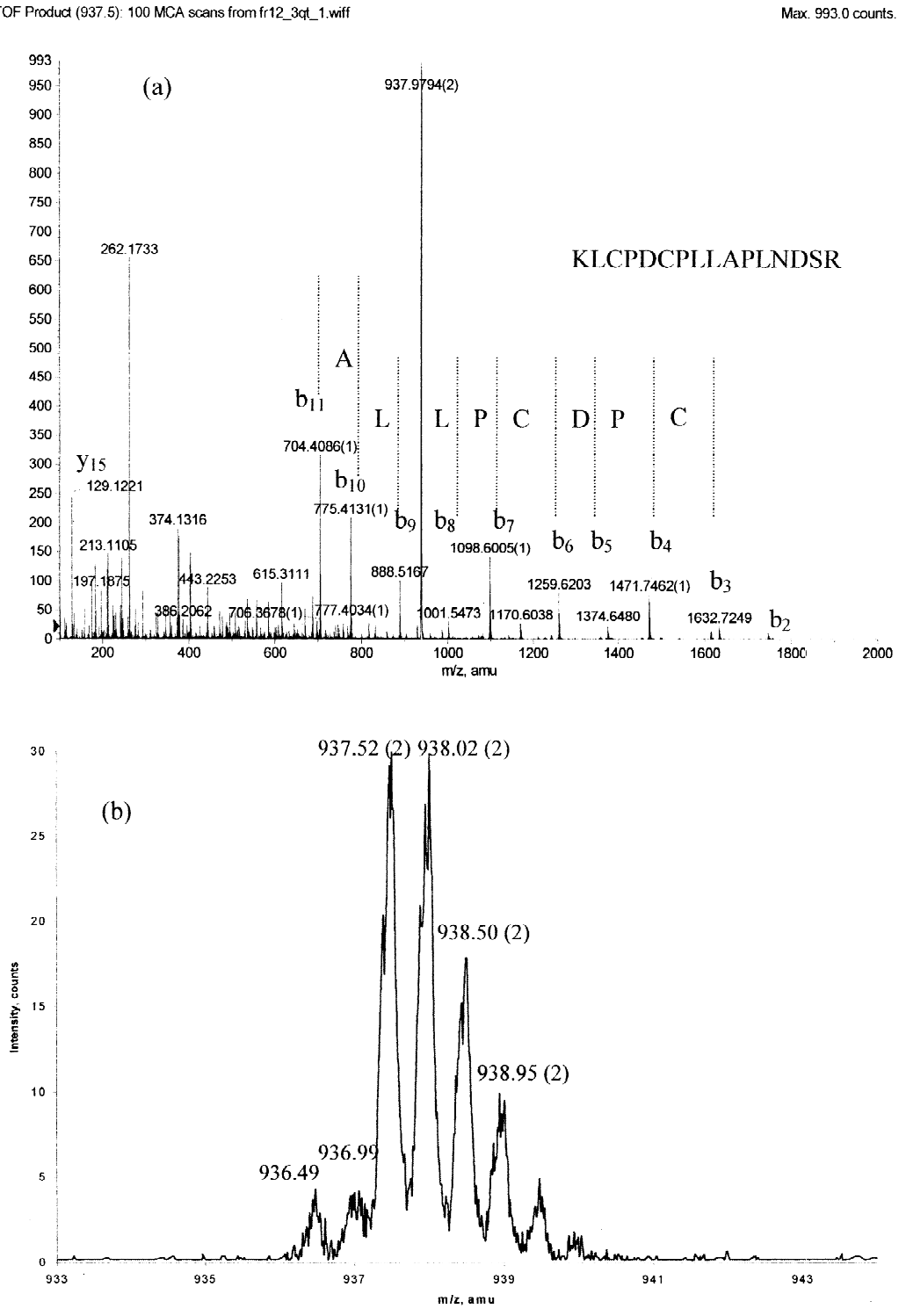

Fig. 5. Mass spectra for the deglycosylated, ^{18}O-labeled fetuin peptide: (a) MS–MS spectrum from the Q-TOF; (b) TOF mass spectrum for the deglycosylated fetuin peptide.

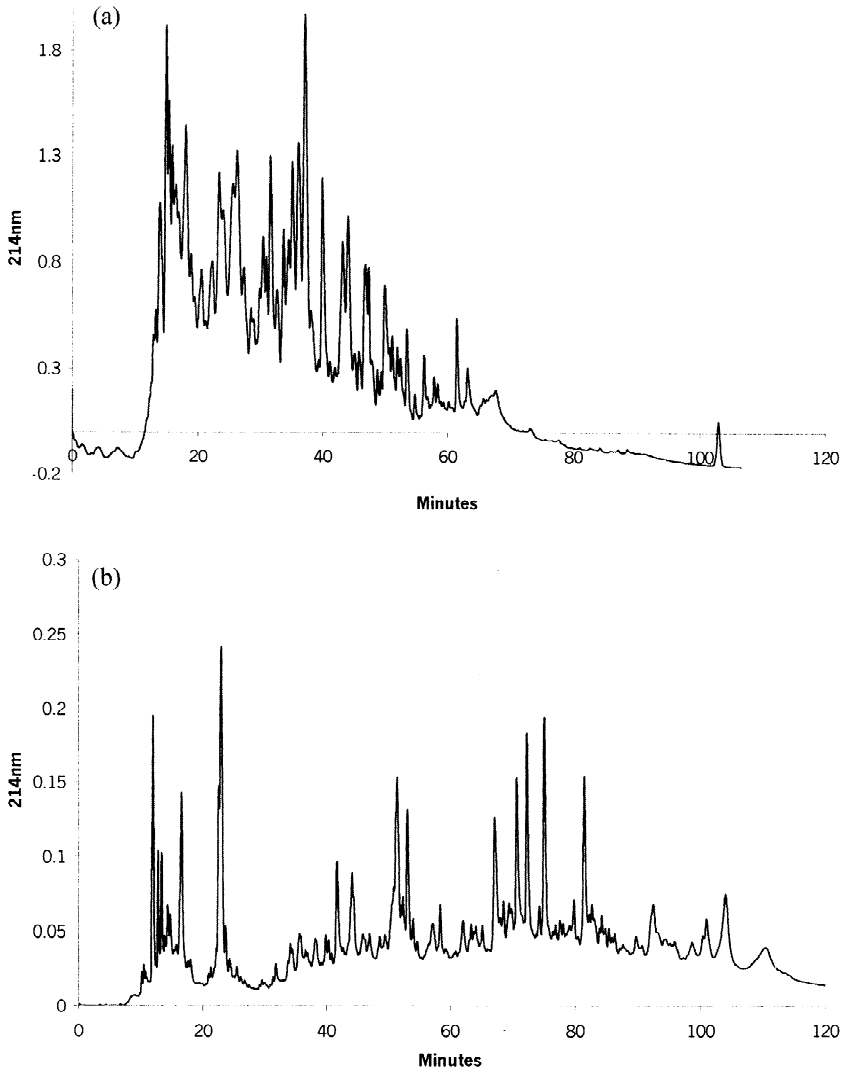

Fig. 6. Reversed-phase chromatograms of human serum: (a) total tryptic digest, (b) after LTA lectin selection and PNGase F deglycosylation. Peptides were eluted from a 4.6×250-mm I.D. C_{18} column with a 120-min linear gradient ranging from 5% ACN in 0.1% aqueous TFA to 60% ACN in 0.1% aqueous TFA at a flow-rate of 1.0 ml/min.

of ^{18}O varies between peptides means that the reversibility of PNGase F is structure specific. The above mechanism is consistent with the theory that an acyl-enzyme intermediate is formed in the case of PNGase catalyzed deglycosylation, although the exact mechanism is still unknown [23,24]. Under acidic conditions in $H_2^{16}O$, the ^{18}O label can be exchanged to ^{16}O because of acid-catalyzed non-enzymatic back exchange. However, in the above experiments, PNGase F was quenched with 0.1% TFA and the sample in liquid N_2 after deglycosylation.

One of the selected peptides identified was from human mammaglobin (hMAM) (Fig. 9), a recently identified protein of the uteroglobin family [25] that is reported in the SWISS-PROT database (accession number O75556). This protein was also reported to contain an *N*-glycosylation site with the NQS sequence. Mass spectra of the peptide show an ion at $[M+H]^+$ (m/z=1467.8) and an amino acid sequence

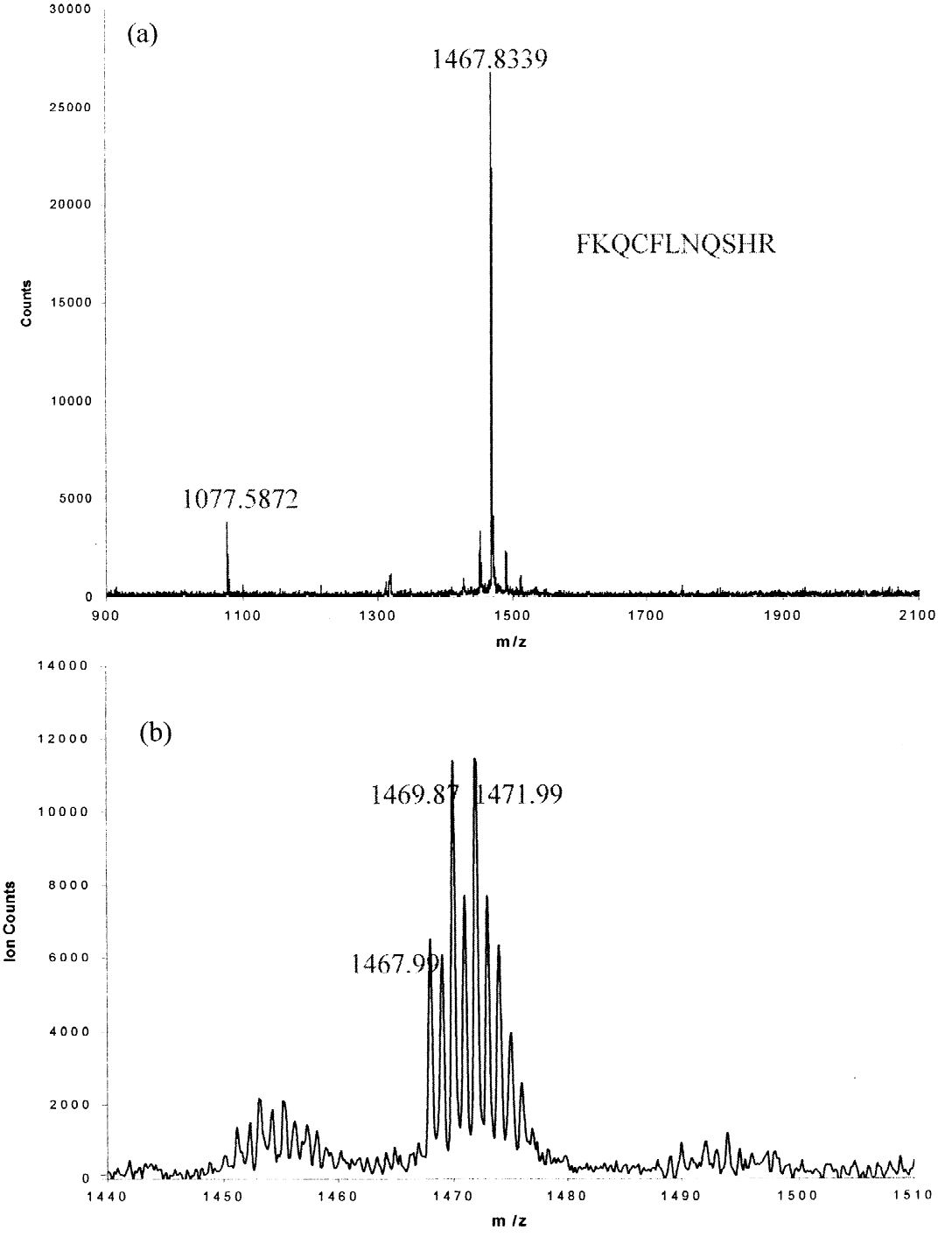

Fig. 7. Mass spectra of a deglycosylated peptide from a reversed-phase fraction of the LTA selected peptides in human serum: (a) deglycosylation was achieved in an $H_2^{16}O$ buffer; (b) hydrolysis was executed in 95% $H_2^{18}O$.

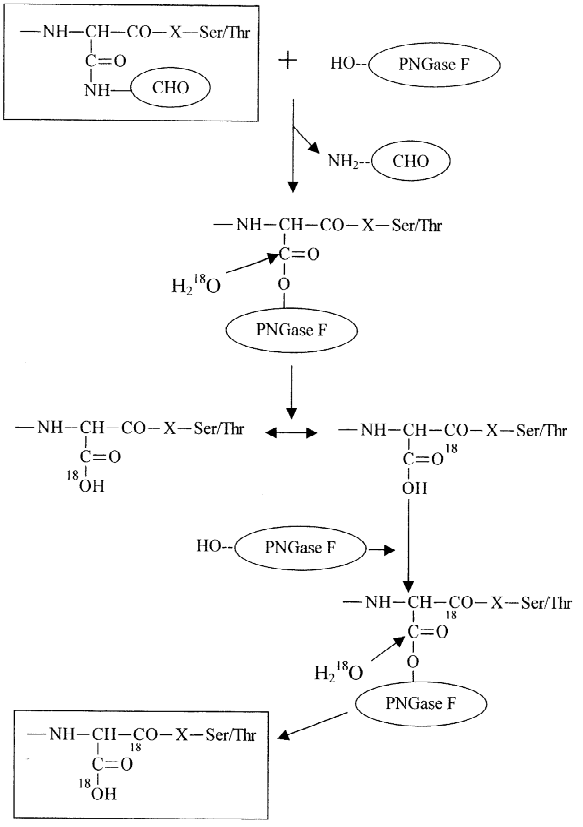

Fig. 8. Potential deglycosylation mechanism for peptide *N*-glycosidase F.

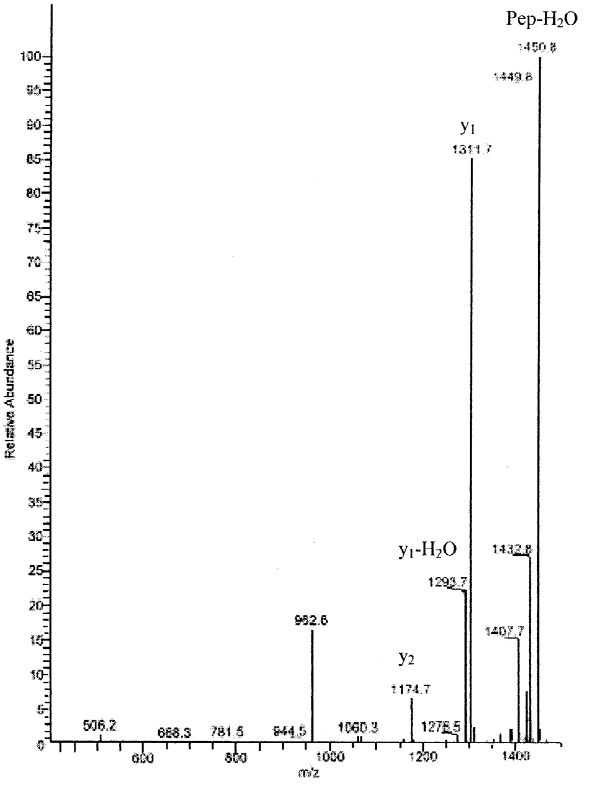

Fig. 9. MS–MS spectrum of $[M+H]^+$ peptide at m/z 1467.8 obtained with LCQ. The peptide has the sequence FKQCFLNQSHR and is from human mammoglobin.

of FKQCFLDQSHR after considering deamidation at asparagine. This corresponds to the sequence in mammaglobin ranging from N-terminal amino acids 62 to 72. Although the hMAM parent of this peptide was obviously glycosylated with the α-Fuc(1→6)-β-GlcNAc- moiety on asparagine-68, none of the reports on hMAM in the scientific literature indicate that it is fucosylated. This either means that another glycoform was used to characterize hMAM or that characterization of hMAM glycosylation is incomplete.

Studies with reversed transcription-polymerase chain reaction (RT-PCR) show that human mammaglobin (hMAM) mRNA is overexpressed in breast tumor [26,27] and that it is frequently up regulated in the peripheral blood of breast cancer patients [28]. The possibility that hMAM is a superior marker for clinical detection of occult metastases in breast carcinoma has been proposed based on auxiliary lymph node expression [29]. Having serum cancer markers is of critical importance because current histological detection is not sensitive enough to monitor metastatic cancer cells. Roughly 30% of lymph-node negative breast cancer patients relapse within 5 years [30,31]. As a serum marker, mammaglobin has potential for accessing therapeutic response or recurrence.

3.4. Other fucosylated proteins in human serum

As the chromatogram in Fig. 6b indicates, the LTA column binds many peptides occurring in human serum. Presumably most are fucosylated, but some may be non-specifically bound. It is not known how many of these peptides may also be cancer markers. However, recent studies on dogs with lymphosarcoma indicate that the concentration of at

least 15 fucosylated proteins changed in serum and that chemotherapy impacts the concentration of these proteins. Further studies are needed to elucidate the relationship between protein expression and post-translational modification in cancer markers and to further access the importance of glycosylation in metastasis.

4. Conclusion

Based on the experiments above with model proteins, it is concluded that fucose-containing glycopeptides can be selected from tryptic digests of complicated mixtures by lectin affinity chromatography. Multidimensional chromatography, in combination with mass spectrometry simplifies the analytical process of identifying these proteins and provides peptide sequence after deglycosylation. Through the use of ^{18}O labeling, it is possible to determine the glycosylation site(s) in these peptides. The fact that PNGase can incorporation 2 mol of ^{18}O into peptides during deglycosylation suggests the reaction is at least partially reversible.

It is further concluded based on affinity chromatography that derivatization of proteins with α-Fuc(1→6)-β-GlcNAc-Asn occurs in the case of at least one known cancer marker. Because this type of glycosylation is structurally unique and targetable, it may provide an easy way to search for cancer markers in the future.

Acknowledgements

The authors greatly acknowledge financial support from the National Institute of Health (grant numbers GM 59995 & GM 35421).

References

[1] P.J. Seberger, W.G. Chaney, Glycobiology 9 (1999) 235.
[2] S.A. Brooks, A.J. Leathem, Lancet 338 (1991) 71.
[3] M.V. Dwek, H.A. Ross, A.J. Leathem, Proteomics 1 (2001) 756.
[4] T. Haselhorst, T. Weimar, T. Peters, J. Am. Chem. Soc. 123 (2001) 10705.
[5] L. Tan, Ph.D. Thesis, Purdue University, Lafayette, IN, 1998, p. 78.
[6] M. Geng, J. Ji, F.E. Regnier, J. Chromatogr. A 870 (2000) 295.
[7] M. Geng, X. Zhang, M. Bina, F. Regnier, J. Chromatogr. B 752 (2001) 293.
[8] G.H. Carpenter, C.L. Pankhurst, G.B. Proctor, Electrophoresis 20 (1999) 2124.
[9] K. Yamamoto, Y. Konami, T. Osawa, T. Ivimura, J. Biochem. 111 (1992) 436.
[10] I. Katnik, J. Jadach, H. Krotkiewski, J. Gerber, Glycosylation Dis. 1 (1994) 97.
[11] S. Stelck, A. Robitzki, E. Willbold, P.G. Layer, Glycobiology 9 (1999) 1171.
[12] J.H. Naismith, C. Emmerich, J. Habash, S. Harrop, J.R. Helliwell, N. Hunter, J. Raftery, A. Kalb, J. Yariv, Biol. Crystallogr. D50 (1994) 847.
[13] L. Yan, P. Wilkins, G. Alvarez-Manilla, S.I. Do, D.F. Smith, R.D. Cummings, Glycoconjugate 14 (1997) 45.
[14] F.H. Plummer Jr., J.H. Elder, S. Alexander, A.W. Phelan, A.L. Tarentino, J. Biol. Chem. 259 (1984) 10700.
[15] E. Bause, H. Hettkamp, FEBS. Lett. 108 (1979) 341.
[16] A.L. Tarentino, C.M. Gomez, T.H. Plummer Jr., Biochemistry 24 (1985) 4665.
[17] R. Kornfeld, S. Kornfeld, Annu. Rev. Biochem. 54 (1985) 631.
[18] F. Maley, R.B. Trimble, A.L. Tarentino, T.H. Plummer Jr., Anal. Biochem. 180 (1989) 195.
[19] B. Küster, M. Mann, Anal. Chem. 263 (1999) 1431.
[20] C. W Sutton, J.A. O'Neill, J.S. Cottrell, Anal. Biochem. 218 (1) (1994) 34–46.
[21] X. Zhang, Ph.D. Thesis, Purdue University, Lafayette, IN, 2001, p. 42.
[22] M. Schnolzer, P. Jedrzejewski, W.D. Lehamann, Electrophoresis 17 (1996) 945.
[23] V. Tretter, F. Altmann, L. Marz, Eur. J. Biochem. 199 (1991) 647.
[24] S. Gosslin, K. Payie, T. Viswanatha, Glycobiology 3 (1993) 449.
[25] M.A. Watson, T.P. Fleming, Cancer Res. 54 (1994) 4598.
[26] B. Suchy, F. Austrup, G. Driesel, C. Eder, I. Kusiak, P. Uciechowski, H.J. Grill, M. Giesing, Cancer Lett. 158 (2000) 171.
[27] E. Leygue, L. Snell, H. Dotzlaw, K. Hole, S. Troup, T. Hiller-Jichcock, L.C. Murphy, P.H. Watson, J. Pathol. 189 (1999) 28.
[28] O. Zach, H. Kasparu, O. Krieger, W. Hehenwarter, M. Girschikofsky, D. Lutz, J. Clin. Oncol. 17 (1999) 2015.
[29] A. Marchetti, F. Bwttita, G. Bertacca, K. Zavaglia, G. Bevilacqua, D. Angelucci, P. Viacava, A. Naccarato, A. Bonadio, F. Barassi, L. Filicioni, S. Salvatore, F. Mucilli, J. Pathol. 195 (2001) 186.
[30] S. Hellman, J. Clin. Oncol. 12 (1994) 2229.
[31] S. Noguchi, T. Aihara, K. Montomura, H. Inaji, S. Imaoka, H. Koyama, Am. J. Pathol. 148 (1996) 649.