ELSEVIER

Journal of Chromatography A, 732 (1996) 27–42

JOURNAL OF CHROMATOGRAPHY A

Application of high-performance liquid chromatography–electrospray ionization mass spectrometry and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry in combination with selective enzymatic modifications in the characterization of glycosylation patterns in single-chain plasminogen activator

Alex Apffel[a,*], John A. Chakel[a], William S. Hancock[a], Carrie Souders[b], Thabiso M'Timkulu[b], Erno Pungor Jr.[b]

[a] *Biomeasurements Group, Hewlett-Packard Laboratories, MS 26U/R6, 3500 Deer Creek Road, Palo Alto, CA 94304, USA*

[b] *Berlex Biosciences, Brisbane, CA 94005, USA*

Received 6 July 1995; revised 27 October 1995; accepted 27 October 1995

Abstract

The application of high-performance liquid chromatography (HPLC), electrospray ionization mass spectrometry (ESI–MS) and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and selective enzymatic deglycosylation treatments is demonstrated in the analysis of glycosylation patterns in recombinant *Desmodus* salivary plasminogen activator, a heterogeneous glycoprotein. The sample was initially digested with a proteolytic enzyme (endoproteinase Lys-C) and then further treated with either PNGase F to remove N-linked carbohydrates or a combination of neuraminidase and O-glycosidase to remove sialic acid and O-linked carbohydrates. By comparison of the LC–ESI-MS peptide maps for the fully glycosylated and deglycosylated samples, it was possible to unambiguously identify the sites of N-linked glycosylation as well a number of N-linked glycopeptides. The O-linked glycopeptides, which are present at a low level (<1%), were not detected prior to the deglycosylation, nor could changes in peptide elution in the map following deglycosylation be correlated with potential O-linked glycosylation sites.

Keywords: Glycosylation; Pharmaceutical analysis; Glycoproteins; Enzymes

1. Introduction

In the case of recombinant protein pharmaceuticals, in addition to primary sequence determination, it has become necessary to determine secondary and tertiary structure as well as post-translational modifications such as glycosylation, phosphorylation and deamidation. This information is needed at a number of different points in the pharmaceutical life cycle. During a development phase, it is critical to understand the parameters that control the efficacy, adsorption and metabolic fate of the pharmaceutical under development. In the regulatory approval process, there are requirements for the demonstration of

*Corresponding author.

0021-9673/96/$15.00
SSDI 0021-9673(95)01231-1

product authenticity and stability. In terms of production and quality control, it is important to monitor product variability in order to control changes due to production modifications or genetic drift in the recombinant process.

A new glycoprotein DSPAα1, expressed by rDNA technology in Chinese hamster ovary cells (CHO), is a serine protease derived from *Desmodus Rotundus* (vampire bat) salivary glands [1–3] which plays a role in clot lysis. The potential applications of DSPAα1 are in the areas treatment of myocardial and cerebral infarctions, deep-vein thrombosis, lung embolism and peripheral occlusive diseases. It consists of 441 amino acids with a predicted (non-glycosylated) average molecular mass of 49 508 with six potential sites of glycosylation, four O-linked and 2 N-linked. From initial carbohydrate studies, it is known that the O-linked structures account for $<1\%$ of the total glycosylation while the N-linked structures present are largely high-mannose type (34.5%) and complex type (65%). The proposed sites of glycosylation are based on predicted consensus amino acid sequences and in the case of O-linked the sequences are not well understood (see later). In addition it is not clear if all of the six sites are indeed glycosylated or if a given site is only partially occupied [4].

The difficulty encountered in analytical problems steadily becomes more challenging as biotechnology produces more complex molecules, such as antibodies which combine high molecular mass and well as substantial post-translational modifications. Amongst the most significant developments with respect to analytical techniques and this problem, have been the advent of multidimensional or hyphenated techniques such as multidimensional high-performance liquid chromatography (HPLC) [5], on-line HPLC–electrospray ionization mass spectrometry (LC–ESI-MS) [6], capillary electrophoresis (CE)–ESI-MS [7] and off-line HPLC or CE–matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS).

The key to these combinations is the ability to approach the sample from substantially orthogonal and independent directions. In the case of LC–ESI-MS, using reversed-phase HPLC, the combination juxtaposes hydrophobicity with mass/charge ratio, which are only loosely correlated. Similarly, through the off-line coupling of HPLC and MALDI-TOF, the orthogonality of the data can be exploited. Furthermore, ESI-MS and MALDI-TOF-MS have been found to be highly complementary in terms of their ability to analyze different sample types. In spite of the impressive analytical power of the instrumental techniques described above, most of the studies have been limited to the characterization of known glycoproteins, in which the mass spectrometric studies were supported by knowledge of the full carbohydrate structures. A major point of these studies, therefore, is to explore what are the limitations of the combination of liquid-phase analysis with mass spectrometry for the determination of unknown structures.

In previous studies of DSPAα1 we explored the characterization through a combined use of capillary electrophoresis, MALDI-TOF-MS and LC–ESI-MS, exploiting the multidimensional information produced by the analysis of the intact protein and enzymatic digestions. Since the primary sequence is known, it was possible to identify the degree and heterogeneity of glycosylation by comparing the non-glycosylated molecular mass based on the expected protein sequence and expected peak width with the experimentally determined values by MALDI-TOF. The next step in analysis of a rDNA-derived protein is the examination of an enzyme digest. However, the complexities of the non-fractionated enzymatic digest mixture can overwhelm the MALDI-TOF in terms of selectivity and resolution for detection of all components. Conversely, when LC–ESI-MS is used to analyze the multiple-charged spectra produced from the heterogeneous mixture of glycoforms of the intact protein, the technique yields complex and ambiguous data which are difficult to interpret. When applied to the analysis of peptide maps, however, LC–ESI-MS can be used to confirm the primary sequence [8], as well as localize sites of glycosylation, either through the use of collisionally induced dissociation (CID) to produce fragment ions indicative of specific sugar structures [9–11] or through the identification of diagonal bands in two-dimensional time vs. m/z plots [12] or by extraction of predicted glycosylation forms from the mass spectral data [13,14]. The characterization of post-

translational modifications such as glycosylation, however, is not as well established. Classical approaches such as selective enzymatic cleavage combined with gel electrophoresis or anion chromatography [15] can be used to characterize the carbohydrates themselves, but are not capable of localization of individual attachment sites.

While both MALDI-TOF-MS and LC–ESI-MS have been used to identify the general glycosylation patterns, there are still limitations and further selectivity is required for the differentiation of N-linked and O-linked glycosylation. In the current work, DSPAα1 is examined, both intact and in enzymatic digests, after stepwise removal of specific sugar structures. Specifically, the samples were analyzed by MALDI-TOF-MS as: (1) the intact fully glycosylated protein; (2) after removal of N-linked sugars by a specific glycosidase (PNGase F); and (3) after removal of sialic acid by neuraminidase and removal of O-linked sugars by another specific enzyme, O-glycosidase. These same samples were then digested with endoproteinase Lys-C and the resulting peptide maps analyzed with LC–ESI-MS. Comparison of the multidimensional results produced by these techniques produced a more complete understanding of the complex glycosylation pattern of DSPAα1 which is consistent with both predicted patterns based on CHO cell protein production and separately conducted carbohydrate analyses performed on the sample.

The following work should be considered as a second part of an earlier work. In the first part, the characterization of the glycoprotein was approached using combinations of HPLC, HPCE, LC–ESI-MS and MALDI-TOF. In this work, we have extended this to include selective enzymatic deglycosylation to aid in the characterization of the glycosylation patterns. Clearly, on a glycoprotein of this complexity, no single technique is capable of generating a complete and full picture of the sample. The techniques described here are used in conjunction with a range of other techniques. There are two main goals to this work; firstly, the analytical focus is on the techniques used for this characterization and, in particular the complementary nature of electrospray LC–MS and MALDI-TOF-MS; and secondly, research is presented in the characterization of a recombinant glycoprotein that has not yet been fully characterized.

2. Experimental

2.1. HPLC

The HPLC separation was performed on a Hewlett-Packard 1090 liquid chromatography system equipped with DR5 ternary solvent delivery system, diode-array UV–Vis detector (DAD), autosampler and heated column compartment (Hewlett-Packard, Wilmington, DE, USA). All HPLC separations were done using a YMC (Wilmington, NC, USA) 5-μm particle, 200 Å pore size ODS-AQ reversed-phase column 250×2.1 mm I.D. A standard solvent system of 0.1% TFA in H_2O (solvent A) and 0.09% TFA in acetonitrile (solvent B), was used with a flow-rate of 0.2 ml/min. The gradient for the separation was constructed as 0–60% B in 60 min. The column temperature was maintained at 60°C throughout the separation.

2.2. ESI-MS

Mass spectrometry was done on a Hewlett-Packard 5989B quadrupole mass spectrometer equipped with extended mass range, high-energy dynode detector (HED) and a Hewlett-Packard 59987A atmospheric-pressure chemical-ionization (API) electrospray source with high-flow nebulizer option. Both HPLC and MS were controlled by the HP Chemstation software allowing simultaneous instrument control, data acquisition and data analysis. The high-flow nebulizer was operated in a standard manner with N_2 as nebulizing (1.5 l/min) and drying (15 l/min at 300°C) gases.

To counteract the signal-suppressing effects of trifluoroacetic acid on electrospray LC–MS, a previously reported method [16], referred to as the 'TFA Fix' was employed. The 'TFA Fix' consisted of post-column addition of 75% propionic acid, 25% isopropanol at a flow-rate of 100 μl/min. The TFA Fix was delivered using an HP 1050 HPLC pump and was mixed with a zero dead volume tee into the column effluent after the DAD detector and after the

column-switching valve. Column effluent was diverted from MS for the first 10 min of the chromatogram, during which time excess reagents and unretained components eluted.

For peptide mapping, MS data were acquired in scan mode, scanning from M_r 300 to 1600 at an acquisition rate of 1.0 Hz at 0.15-u stepsize. Data were filtered in the mass domain with a 0.05-amu gaussian mass filter and in the time domain with a 0.05-min gaussian time filter. Unit mass resolution was maintained for all experiments, allowing unambiguous identification of +1 and +2 charge states. Unit resolution also allows high-charge states to be identified as >+2. For masses encountered in the tryptic map of the samples analyzed (<10 000), mass accuracy was <0.5 amu. The fragment identification was done with the aid of HP G1048A protein and peptide analysis software, a software utility which assigns predicted fragments from a given sequence and digest with peak spectral characteristics.

For the in-source collisionally induced dissociation (CID) method for detecting fragments indicative of glycopeptides a scan acquisition was performed with dynamic CapEx (capillary exit potential) voltage to allow simultaneous acquisition of glycomarkers and intact peptide spectra. In this experiment, CapEx was held at 300 V for the scan range 60–380 amu and then changed to 100 V for the scan range 400–2000 amu. Data were filtered in the time domain with a 0.05-min gaussian time filter.

2.3. MALDI-TOF-MS

Mass spectra were generated with a Hewlett-Packard G2025A MALDI-TOF system. This system utilizes a N_2 laser (337 nm) for the desorption/ionization event, 28 kV of ion acceleration and a linear 1.0-m time-of-flight analyzer for mass analysis. Spectra were acquired at laser powers slightly above the ionization threshold using a matrix consisting of 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid, HP P.N. G2055A).

Intact protein: DSPAα1 (0.55 mg/ml) was mixed 1:1 with a 100 m*M* sinapinic acid matrix solution and 1 μl (ca. 5 pmol total) was deposited onto the probe tip and dried. Data were collected at a 100-MHz sampling rate and a total of 100 laser shots were summed. The mass scale was calibrated using a protein standard mixture (HP P.N. G2053A) as an external calibrant.

2.4. Chemicals and reagents

HPLC-grade acetonitrile and trifluoroacetic acid (TFA), as well as EDTA were obtained from J.T. Baker (Phillipsburg, NJ, USA). Distilled, deionized Milli-Q water (Millipore, Bedford, MA, USA) was used. Urea was obtained from GIBCO (Gaithersburg, MD, USA), DL-dithiothreitol (DTT), iodoacetic acid, bicine, NaOH, and $CaCl_2$ were obtained from Sigma (St. Louis, MO, USA). The enzyme endoproteinase Lys-C (Wako BioProducts, Richmond, VA, USA) was used for mapping. Recombinant DSPAα1 (from Berlex Biosciences, Richmond, CA, USA) was purified from CHO cells and the starting concentration was 0.5 mg/ml.

2.5. Methods

2.5.1. Sample preparation

DSPAα1 (5 mg) was denatured in 6 *M* urea and titrated with 5 *M* NaOH to pH 8.3. The sample was then reduced using 32 m*M* DTT (molar excess to DSPAα1 1:300) by incubating at 55°C for 3 h. Alkylation of the reduced protein was done by a further 30-min incubation with 70 m*M* iodoacetic acid. The reduced and alkylated DSPAα1 was then desalted using size-exclusion chromatography.

2.5.2. Lys-C digestion

The reduced, alkylated, desalted DSPAα1 was titrated to pH 9 and then digested with endoproteinase Lys-C in 200 m*M* bicine at 37°C with an enzyme-to-substrate (mass) ratio of 1:100 for 18 h in the presence of 20 m*M* DTT, 0.5 m*M* EDTA, and 8.5 m*M* $CaCl_2$. The digestion was quenched by titrating the sample to pH 1.5

2.5.3. Carbohydrate analysis

Initial characterization of the glycan structures were done by independent means. The glycan structures of DSPAα1 were characterized by first cleaving the oligosaccharides by chemical means. Carbohydrates sequencing of purified oligosaccharides was performed using a cocktail array of exoenzymes

followed by separation on a sequencing gel using fluorophore-assisted carbohydrate electrophoresis (FACE) (Glyko, Novato, CA, USA).

2.5.4. Enzymatic deglycosylation

Both the endoproteinase Lys-C digest of the reduced and alkylated DSPAα1 and the intact reduced and alkylated DSPAα1 were further treated for enzymatic deglycosylation with identical protocols.

Removal of N-linked carbohydrates was accomplished with PNGase F (New England BioLabs, Beverly, MA, USA). An amount of 20 nmol of DSPAα1 was digested with 62 500 units of the enzyme in 50 m*M* sodium phosphate buffer, pH 7.5 at 37°C for 2 h.

Removal of O-linked carbohydrates was accomplished by O-glycosidase (Glyko) following removal of the sialic acid by NANase III (Glyko), a recombinant preparation of neuraminidase. An amount of 20 nmol of DSPAα1 was first digested in 50 m*M* sodium phosphate buffer, pH 5.5 at 37°C in the presence of 200 mU of NANase III for 2 h. This mixture was further digested in the presence of 40 mU O-glycosidase for 6 h.

3. Results and discussion

3.1. Analytical strategy

The analytical strategy employed in this study is focused on the utilization of specific chemical and enzymatic treatment to the glycoprotein followed by analysis by both MALDI-TOF-MS and LC–ESI-MS. By employing these two powerful techniques, a large amount of complementary data is generated which can subsequently be re-integrated into a single cohesive picture of the analytical problem at hand. Our focus has been to exploit the strengths of the individual techniques to accentuate their complementary nature. Specifically, we have found MALDI-TOF-MS well suited the analysis of simple mixtures of large peptides and proteins, while LC–ESI-MS is best suited to the separation and analysis of complex mixtures of small- to medium-size peptides. Note, that this strategy is not exclusive; in many cases, MALDI-TOF-MS has been found useful for mixtures of smaller fragments, particularly in light of on-probe chemical modifications [17]. Similarly, LC–ESI-MS can be utilized with larger proteins that do not contain multiple post-translational modifications, although for all but the simplest mixtures, it is usually necessary to use a chromatographic separation technique as well.

MALDI-TOF-MS is a rapid and extremely sensitive technique which can be used to obtain molecular mass data to confirm identity or as a scouting technique prior to Edman sequencing. Even in the cases of complex, heterogeneous samples, MALDI-TOF is more accurate, faster and more sensitive than classical methods such as SDS-PAGE.

LC–ESI-MS, as mentioned in the introduction has gained rapid acceptance in the application of peptide mapping. In the analysis of glycopeptides, there are several approaches to utilizing LC–ESI-MS. Carr et al. [9] demonstrated the use of collisionally induced dissociation (CID) utilized 'in-source' in a single quadrupole MS instrument or in the collision cell of a MS–MS instrument. In either case, by monitoring the specific masses (principally 147, $[dHex]^+$; 204, $[HexNAc^+]$; 292, $[NeuAc^+]$; and 366, $[Hex+HexNAc^+]$) under elevated collision energies, signals could be generated that correspond to peptide fragments which are glycosylated. Unfortunately, it is not possible to utilize CID methods to differentiate between N-linked and O-linked sugars.

A second approach which can be used for glycopeptide analysis by LC–ESI-MS utilizes the detection of diagonal bands on two-dimensional plots (time vs. m/z) developed by Ling et al. [12] These diagonal bands result from the fact that the microheterogeneity of glycopeptides leads to increasing degree of glycosylation (mass) with decreasing hydrophobicity (or retention). Although this can be a powerful technique, its utility is somewhat dependent on the complexity and heterogeneity present in the original glycoprotein. In cases of extreme heterogeneity, the total amount of peptide present is distributed over a large number of different glycoforms and it may be difficult to distinguish any of the individual species from noise.

Finally, if the sequence of the protein is known, fragments and their glycoforms can be predicted from knowledge of the structures and hence the mass of expected peptides. In the absence of other information, the individual sites of glycosylation can

be predicted based on the minimum consensus sequence required for glycosylation: Asn–X–(Ser/Thr) for the N-linked [18] and (Ser/Thr)–X–X–Pro for the O-linked [19] structures. The mass spectral data can then be interrogated searching for the presence of the predicted ions. Once again, in the case of low levels of glycosylation or extreme microheterogeneity, the individual species may be below the detection limit.

It is also possible to process the sample with various endoglycosidases such as peptide-N-glycosidase F (removes most N-glycans), O-glycosidase (removes O-glycans) or exoglycosidases such as neuraminidase (to remove sialic acid). If these enzymes are used sequentially, it is possible to identify changes in a peptide map due to the removal of a specific type of glycan and specifically to differentiate between N-linked and O-linked glycopeptides

3.2. Analytical results

3.2.1. MALDI-TOF-MS of DSPAα1

As a first step in this process, DSPAα1 was analyzed by MALDI-TOF-MS in four forms: (a) fully glycosylated; (b) treated with PNGase F; (c) treated with neuraminidase; and (d) treated with neuraminidase and O-glycosidase. The results are shown in Fig. 1. From these data it can be seen that the intact, reduced and alkylated protein shows a M_r of 53 407 compared to the predicted non-glycosylated mass of 49 508. The expected mass accuracy is $\approx$0.1% based on an external mass calibration. Thus there is a difference of 3899 which can be presumed due to glycosylation. It is also interesting to note that the peak width is broad ($\approx 10\times$) compared to a non-glycosylated protein standard. This is presumed to be due to heterogeneity. Fig. 1b shows the same sample after treatment with neuraminidase and Fig. 1c after treatment with neuraminidase and O-glycosidase. After removal of the sialic acid by neuraminidase digestion, there is a small, but significant drop in the average M_r, while the peak profile is partially resolved as three major groups. Since O-linked glycosylation (Fig. 1c) is expected to account for $<1\%$ of the total glycosylation, it is not surprising that the mass shift appears insignificant under these resolution conditions with respect to the asialo sample. In Fig. 1d, after treatment with PNGase F, there is a peak at 49 985, with a noticeable drop in molecular mass of approximately 3500. There is also a peak at approximately 46 700 which may be due to a clipped form since the observed mass of the heavier part is close to the expected mass of the non-glycosylated form. No signal was seen at 46 700 with the native protein preparation.

Fig. 1e (inset) shows the corresponding SDS-PAGE analysis of these same samples, which is entirely consistent with the MALDI-TOF-MS data. Note that the difference between the intact, fully glycosylated sample (SDS-PAGE lane 4, and MALDI-TOF-MS, Fig. 1a) and the intact neuraminidase treated sample (SDS-PAGE lane 5, and MALDI-TOF-MS, Fig. 1b) or the intact neuraminidase/O-glycosidase treated sample (SDS-PAGE lane 6, and MALDI-TOF-MS, Fig. 1c) show very little difference, implying little O-glycosylation. Comparing the intact, fully glycosylated sample (SDS-PAGE lane 4, and MALDI-TOF-MS, Fig. 1a) and the intact PNGase F treated sample (SDS-PAGE lane 7, and MALDI-TOF-MS, Fig. 1d) shows a marked reduction in molecular mass in both the SDS-PAGE and the MALDI-TOF-MS, although the SDS-PAGE does not show the resolution to distinguish two peaks seen in the MALDI-TOF-MS. It is not clear what the species present at the lower mass in the MALDI-TOF of PNGase treated material is. One of the limitations of the techniques discussed here is that it is not possible to make absolute assignments of this nature. However, the resolution of the SDS-PAGE is such that it is consistent with the MALDI-TOF-MS spectral data. The lower resolution of the SDS-PAGE is apparent in the inability to note any difference in the mobility with the neuraminidase treated sample. Another feature of the MS approach is that MALDI-TOF-MS gives accurate values for samples such as glycoproteins which usually exhibit non-ideal behaviour in gel electrophoretic approaches (see Table 1). In this case, the estimated M_r values of 43 000 and 41 000 for the asialo form are not consistent with the cDNA-derived mass of 49 508 and a value of 8% (w/w) for the carbohydrate content of the glycoprotein. By comparison, however, the MALDI-TOF-MS experiment was completed in a few minutes while sub-picomole amounts of sample were

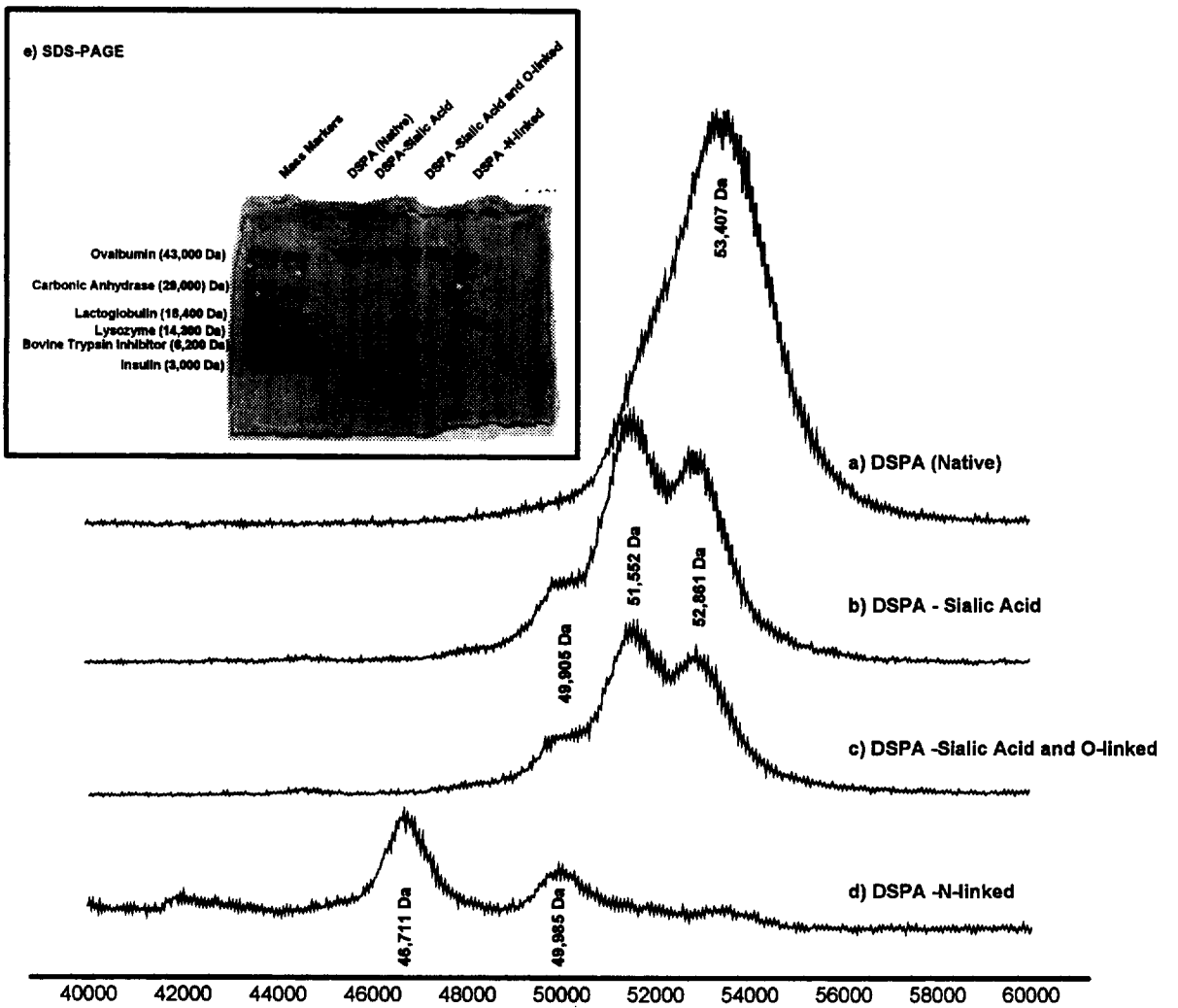

Fig. 1. MALDI-TOF-MS data: (a) DSPAα1 native protein; (b) after neuraminidase treatment; (c) after neuraminidase and O-glycosidase treatment; (d) after PNGase F treatment; (e) SDS-PAGE.

Table 1
Comparison of MALDI-TOF-MS and SDS-PAGE

Sample	M_r		
	Expected	MALDI-TOF-MS	SDS-PAGE
AA sequence:	49 508		
Glycoprotein[a]	54 000	53 407	43 300
-N-linked	49 850[b]	49 985	41 000

[a] Calculated from 8% carbohydrate content.
[b] Calculated from 10% w/w O-linked vs. N-linked.

loaded and accurate masses were obtained. It should be noted that even these rapid and sensitive measurements are relatively poor compared to a more well-behaved non-glycosylated protein. In spite of the difficulties of analyzing heterogeneous glycoproteins by MALDI-TOF-MS [20], it was useful in this context as a rapid screening technique.

3.3. LC–ESI-MS of DSPAα1 digests

Having screened the sample for successful deglycosylation with MALDI-TOF-MS, the Lys-C

digests of the fully glycosylated, N-linkage deglycosylated and O-linkage deglycosylated DSPAα1 were analyzed by LC–ESI-MS. The three resulting mass chromatograms are shown in Fig. 2. Initially, the digest fragments predicted from the theoretical digestion of the known sequence were identified in the Lys-C digest of the fully glycosylated sample (Fig. 2a). The predicted and identified fragments are listed in Table 2. Since, in the dynamic CapEx method, the CID potential is relatively high for masses below 400, peptides with molecular masses less than 400 were not observed intact (fragments K6, K13, K14 and K18). In addition, fragments K2 (8–29), K3 (30–82) and K22 (364–419) were not identified. Interestingly, K22 contains one of the O-linked carbohydrates and thus its mass should be shifted to an unknown higher mass. Peptides K2 and K3 may have undergone unidentified secondary cleavage as both contain multiple basic sites. Since K2 contains the usual N and C termini and four arginines, and K4 contains N and C termini and three arginines) both peptides should fall within the mass range of the instrument due to multiple charging in the electrospray spectra. The use of Lys-C is advantageous over other enzymes, such as trypsin, as the number of fragments is significantly reduced (24 vs. 48 in the case of trypsin). Also the larger peptides (K3, K4 and K22) contain several internal basic residues which facilitate accurate mass measurements. Also, Lys-C digestion results in no more than one predicted glycosylation site per peptide.

Similarly, the LC–ESI-MS analysis of the two

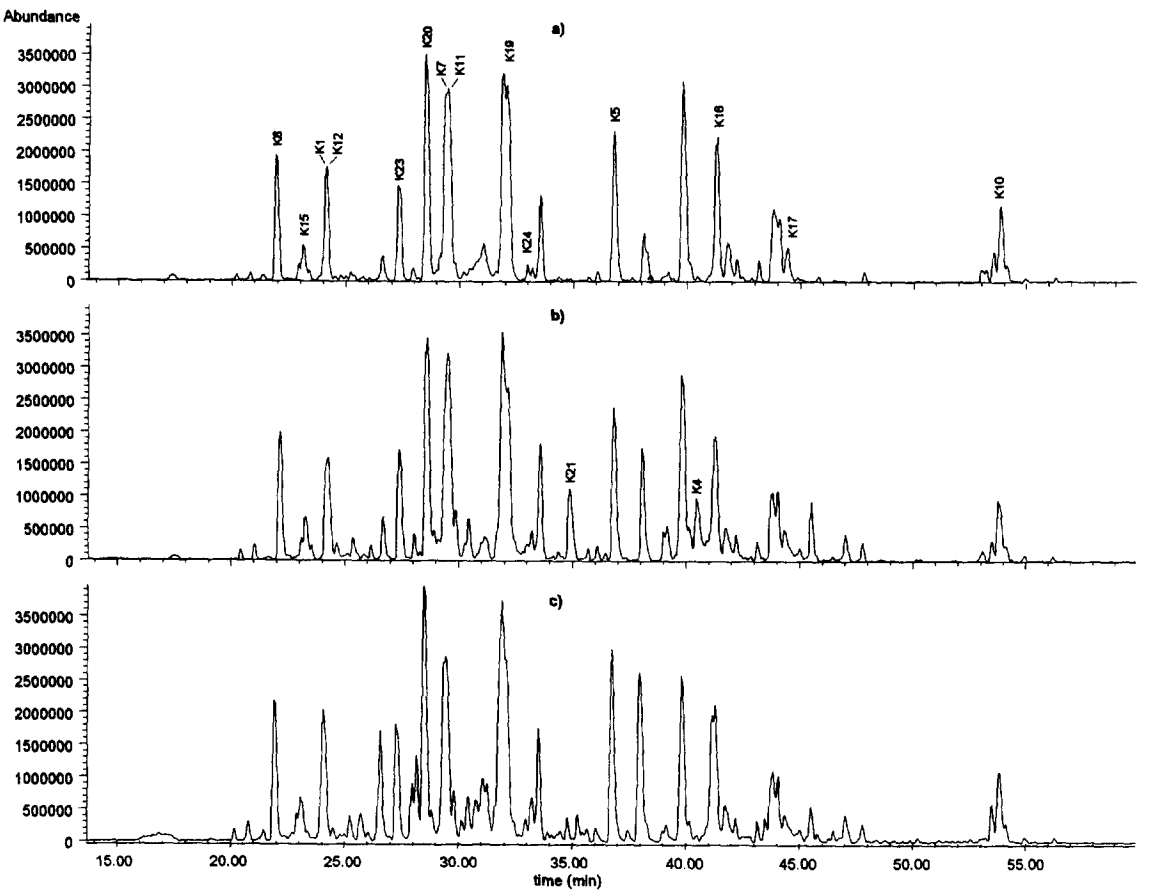

Fig. 2. ESI mass chromatograms – total-ion chromatogram for DSPAα1 (reduced and alkylated) endoproteinase Lys-C digest. Identified fragments are labeled according to Table 2. (a) Fully glycosylated; (b) after PNGase F treatment; (c) after neuraminidase and O-glycosidase treatment.

Table 2
Predicted and found fragments

Fragment	Residues	Retention time (min)	M_r (predicted)	Sequence	Glycosylation
K1	1–7	24.1	768.4	AYGVA CK	
K2	8–29	NF[a]	2808.4	DEITQ MTYRR QESWL RPEVR SK	
K3	30–82	NF	6367.6	RVEHC QCDRG QARCH TVPVN SCSEP RCFNG GTCWQ AVYFS DFVCQ CPAGY TGK	
K4	83–152	40.5	8162.7	RCEVD TRATC YEGQG VTYRG TWSTA ESRVE CINWN SSLLT RRTYN GRMPD AFNLG LGNHN YCRNP NGAPK	N-linked at N117
K5	153–159	36.79	965.5	PWCYV IK	
K6	160–162	NA[b]	274.2	AGK	
K7	163–175	29.3	1488.6	FTSES CSVPV CSK	
K8	176–182	21.94	805.4	ATCGL RK	
K9	183–184	NA	309.2	YK	
K10	185–248	53.8	7295.5	EPQLH STGGL FTDIT SHPWQ AAIFA QNRRS SGERF LCGGI LISSC WVLTA AHCFQ ESYLP DQLK	O-linked at T199
K11	249–258	29.48	1189.7	VVLGR TYRVK	
K12	259–267	24.15	1063.5	PGEEE QTFK	
K13	268–269	NA	245.2	VK	
K14	270–270	NA	146.1	K	
K15	271–275	23.40	658.4	YIVHK	
K16	276–292	41.33	2026.9	EFDDD TYNND IALLQ LK	
K17	293–330	44.4	4314.9	SDSPQ CAQES DSVRA ICLPE ANLQL PDWTE CELSG YGK	O-linked at S293
K18	331–332	NA	283.2	HK	
K19	333–343	32.1	1271.6	SSSPF YSEQL K	O-linked at S333
K20	344–358	28.51	1756.9	EGHVR LYPSS RCAPK	
K21	359–363	34.98	667.4	FLFNK	N-linked at N362
K22	364–419	NF	6084.6	TVTNN MLCAG DTRSG EIYPN VHDAC QGDSG GPLVC MNDNH MTLLG IISWG VGCGE K	O-linked at G393
K23	420–427	27.31	877.5	DVPGV YTK	
K24	428–441	33.0	1730.9	VTNYL GWIRD NMHL	

[a] Fragment not detected.
[b] Fragment below mass range.

deglycosylated Lys-C digests (Fig. 2b and c) have been examined for predicted (non-glycosylated) digest fragments. The most striking identification is that in Fig. 2b at approximately 35.3 min a strong peak emerges in the N-linkage deglycosylated sample corresponding to the K21 fragment containing N_{362}, one of the expected N-linkage sites. Similarly, at 40.5 min, a peak is present in the N-linkage deglycosylated sample (but absent in the other two samples) corresponding to the K4 fragment which contains N_{117}, the other expected site of N-linkage. Comparing Fig. 2a and c does show some differences, but at the present, these structures have not been determined. Specifically (as shown in Table 2), K4 and K21 are possible sites of N-linked glycosylation, while K10, K17, K19 and K22 are possible sites of O-linked glycosylation. Considering the Lys-C peptide maps of the fully glycosylated, N-linkage deglycosylated and O-linkage deglycosylated samples, all the potential glycopeptides were detected, with the exception of K22. The observation of the proposed O-lined peptides at the non-glycosylated mass is consistent with the low levels of O-glycosylation observed in this protein.

Experimentally, very comparable results have been obtained from reduced and alkylated DSPAα1 samples in which the protein was deglycosylated instead of the peptide mixture. For consistency's sake, the data discussed here were first digested with Lys-C and then deglycosylated.

3.4. LC–ESI-MS of DSPAα1 digests with in-source collisionally induced dissociation

As a second step, the glycomarker ions produced by elevated collision energy at m/z<400 were extracted for the same three samples as analyzed in Fig. 2. The reconstructed mass chromatograms resulting from the monitored glycomarker ions are shown in Fig. 3 (the TIC determined under low collision energy for the fully glycosylated sample is shown for reference). The most significant observation is that there is almost no signal for the sample treated with PNGase F (Fig. 3b) for m/z 204 and 366. This is as one would expect since the N-linked glycosylation is the most abundant type in this sample. Comparing the fully glycosylated (Fig. 3a) with the sample treated with neuraminidase and O-glycosidase (Fig. 3c) not much difference for the m/z 204 (HexNAc) or m/z 366 (Hex+HexNAc) is seen, as might be expected since the O-glycosidase corresponds to <1% of the total glycosylation. However, the desialylation results in nearly complete removal of m/z 292 (NeuAc). The sharp peak eluting at approximately 24 mins in the fully glycosylated sample shows an intense m/z 292 signal and is

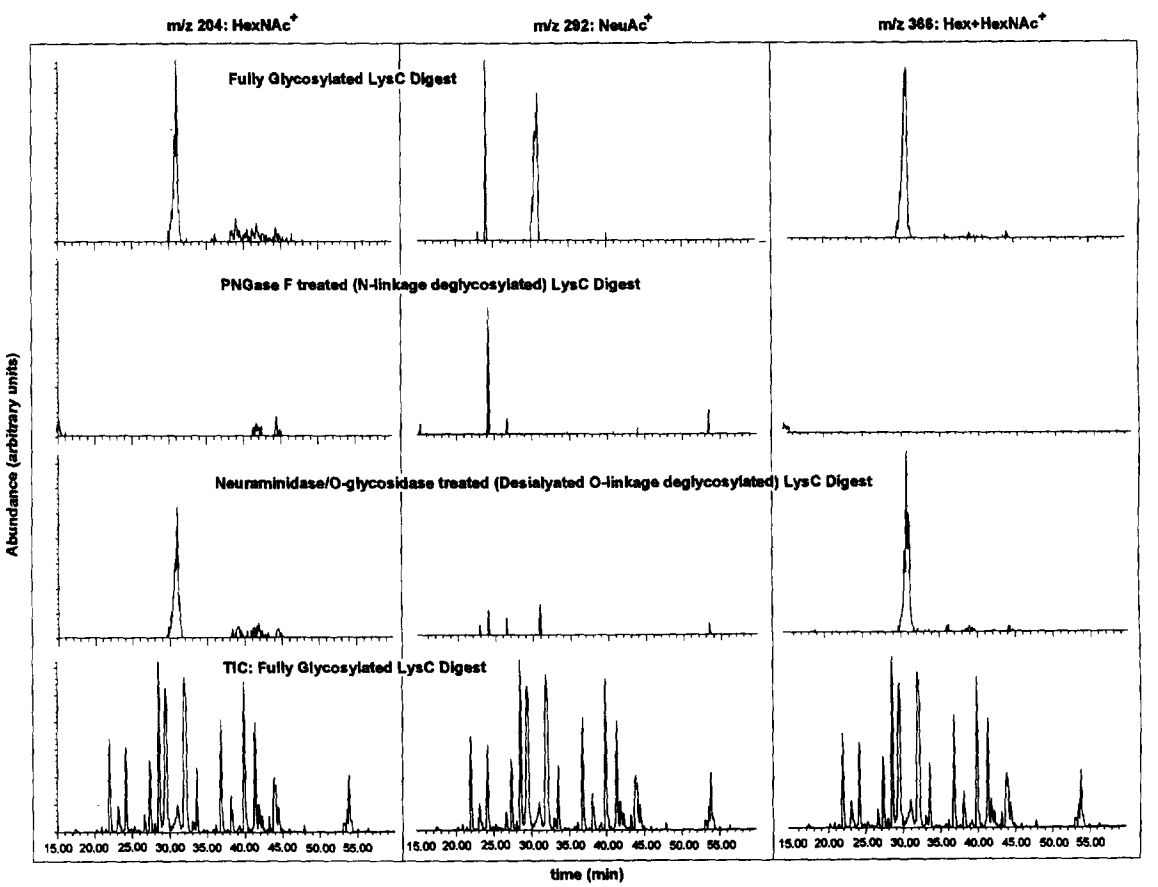

Fig. 3. CID extracted ion chromatograms, LysC digest. Left panel: m/z 204, $[HexNAc]^+$; center panel: m/z 292, $[NeuAc]^+$; right panel: m/z 366, $[Hex+HexNAc]^+$. (a) Fully glycosylated; (b) after PNGase F treatment; (c) after neuraminidase and O-glycosidase treatment; (d) total-ion current for fully glycosylated sample (for comparison).

reduced after treatment with neuraminidase, however, examining the spectra of the peak reveals it to be a partially resolved mixture of two non-glycosylated peptides, K1 and K12, In addition to the m/z 292, the peak at 24.3 mins includes 769.4 $[K1+1H]^{+1}$ and 1064.6 $[K12+1H]^{+1}$ and 532.8 $[K12+2H]^{+2}$. It appears not to be a CID glycomarker of sialic acid since it is present even at low collision energy thus the origin of the sialic acid signal is not clear. Although this might suggest non-specific cleavage, until the identity of this peak is determined, it is not possible to characterize this further.

Comparison of the high collision energy data with the low collision energy data from the fully glycosylated sample (Fig. 3a and d) shows a large broad unresolved peak in the CID extracted-ion current data from 31 to 33 min with a low-abundance corresponding peak in the scan data. It has been our experience that glycopeptides typically show 5–10 times lower abundance than the corresponding non-glycosylated peptide. This is in combination with the fact that the microheterogeneity results in a large number of relatively low intensity peaks. Since this peak disappears in both the CID extracted-ion current high collision energy data (Fig. 3b) and the scan low collision energy data (Fig. 2b), it can be concluded that this region is due to heterogeneous glycopeptides due to N-linked glycosylation.

In our previous work [4], the determination of glycosylation through the use of CID was conducted with a high, fixed CapEx value and selected-ion monitoring (SIM) acquisition. Similar results can be obtained as shown in the present work in scan acquisition with a dynamically ramped CapEx voltage. There is a ramp rate limitation of approximately 10 V/amu which accounts for the gap in mass range of 380–400 V during which the CapEx voltage is intermediate between 300 and 100 V. As would be expected, this method is not as sensitive as selected-ion monitoring, but allows acquisition of both glycomarker ions and intact peptide spectra in a single run.

3.5. Identification of glycosylation

Since the primary sequence and source of DSPAα1 are known, it is possible to predict possible glycopeptides that may be present in the Lys-C digestion. DSPAα1 is a rDNA-derived protein expressed in Chinese hamster ovary (CHO) cells. As such, it is possible to predict common patterns of glycosylation seen in these cells. Using the masses of the predicted glycopeptides in different charge states, one can extract ions from the mass chromatograms to identify actual glycopeptides. Due to the high degree of heterogeneity in the glycosylation, each of these ions will be present at rather low levels. Nonetheless, assignments can be made with relatively good confidence based on the following factors. (1) Since one of the characteristics of the electrospray ionization of peptides and proteins is the formation of multiple-charged ions [21] the coelution of multiple-charge states for a single parent ion is an important indication of peak identity and homogeneity. (2) The relative elution position of a specific glycoform compared to other glycoforms in a closely related family. For example, if two species vary only by the presence or absence of a sialic acid, one would expect a small and predictable effect on retention. (3) The extracted-ion chromatograms can also be compared to the presence of glycomarker ions. This is one advantage of the dynamically ramped CID method described above, in that both the peptides and the glycomarkers can be examined in a single separation. (4) Finally, the data from a glycosylated sample can be compared with the enzymatically deglycosylated sample to verify glycopeptide assignments. It is helpful to exploit all of these data (sometimes iteratively) to unravel as complete a picture of the sample as possible.

It should be noted that in such a highly microheterogeneous sample (whereas the identification of a specific glycoform can be made with varying degrees of confidence if the specific glycopeptide is present at sufficient levels), the failure to detect a specific form can only imply that the glycopeptide is below a specific level.

For the sake of illustration, the analysis of the glycoforms of fragment K21 will be examined. Table 3 shows a list of the glycoforms of fragment K21 (359–419) that were identified by the method above. As an example, Fig. 4a shows the extracted-ion profiles of the biantennary complex glycoforms that contain one fucose and variable sialic acid. Note that in the total-ion chromatogram, this area consists of a

Table 3
Glycoforms of K21 identified.

Type	Nr. sialic acid	Nr. fucose	Retention time	Charge states
High mannose				
$Man_3GlcNAc_2$	0	0	30.40	+2, +3, +4
$Man_4GlcNAc_2$	0	0	29.8	+2
$Man_5GlcNAc_2$	0	0	27.2	+2, +3
$Man_6GlcNAc_2$	0	0	26.8	+3, +4
$Man_7GlcNAc_2$	0	0	26.6	+3, +4
$Man_8GlcNAc_2$	0	0	26.1	+2, +3, +4, +5
Complex				
Biantennary	0	0	32.8	+5
	1	0	32.4	+4
	2	0	32.3	+3
	0	1	31.3	+2, +3
	1	1	31.15	+2, +3
	2	1	30.98	+2, +3
	0	2	32.3	+3, +4
	1	2	32.3	+3, +4, +5
	2	2	31.8	+3, +4, +5,
Triantennary	0	0	33.65	+2, +3, +4
	0	1	30.88	+2
	1	1	30.68	+2, +3
	2	1	30.55	+2, +3
	3	1	30.35	+3
	2	2	33.2	+5
	3	2	32.2	+3, +4, +5,
Tetraantennary	0	0	33.6	+2, +3
	1	0	31.0	+3, +4, +5
	0	1	30.7	+2, +3
	1	1	30.52	+3
	2	1	30.38	+3
	3	1	30.25	+3
	0	2	33.65	+3, +4, +5
	2	2	33.1	+4
	3	2	32.2	+4

broad unresolved zone. Fig. 4b shows the extracted-ion chromatograms for this zone for the m/z 292 (NeuAc) and m/z 366 (Hex+HexNAc) marker ions. Thus, having predicted the presence of the glycoforms, ion chromatograms could be extracted. For each of these peaks at least two charge states were present in the spectra. These closely related glycopeptides elute adjacent to each other and as the degree of sialylation increases, the retention decreases, as would be expected. Also note that the peak cluster elutes between 30 and 32 min, significantly earlier than the K21 fragment in the PNGase F treated sample (see Fig. 2b). The glycomarker ions m/z 292 and 366 are consistent with the presence of sialylated N-linked glycopeptides. When these same ions are extracted from the same analysis of the O-glycosidase treated sample, these peaks are present in slightly lower abundance, but for the PNGase F treated sample, the peaks are totally absent (not shown). Also note the strong m/z 292 signal at approximately 24 min indicates the presence of sialic acid; however, the spectra of this peak show only two, unresolved, non-glycosylated peptides, K1 and K12.

In some cases, it is not possible to unambiguously identify a glycoform from the mass spectrum alone.

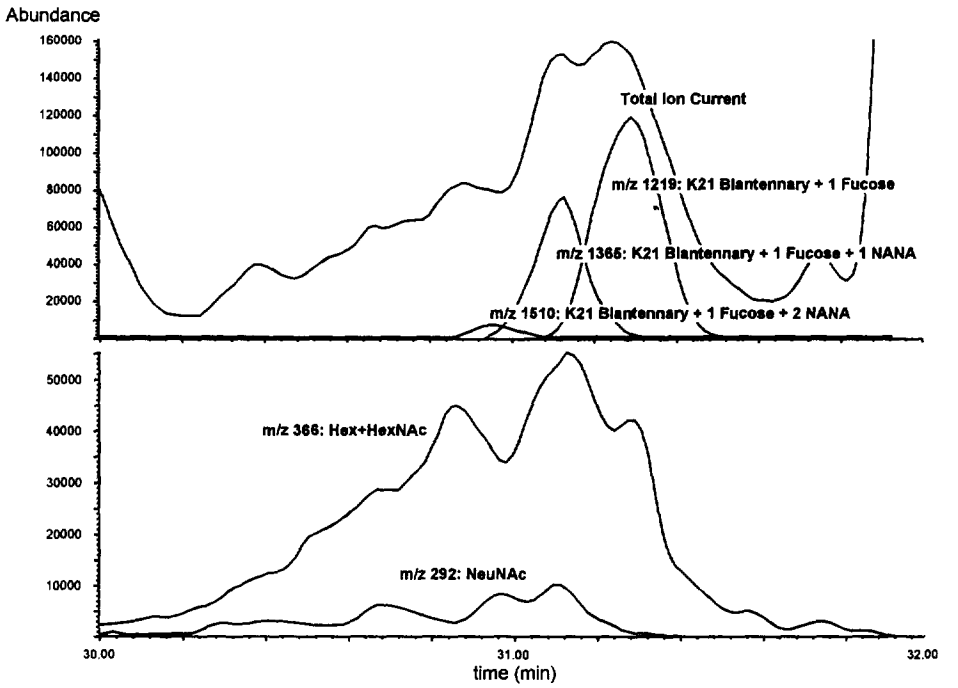

Fig. 4. Identification of specific K21 glycopeptides. Top panel: total-ion current and three extracted-ion chromatograms. The extracted-ion chromatograms are for the $[M+H]^{1+}$ ions as stated and were extracted with −0.3, +0.5 amu mass windows. Bottom panel: *m/z* 292 and 366 CID glycomarkers.

In such cases, a tentative identification can be made based on the mass spectrum and the chromatographic behavior. For example, in the case of the triantennary structure with one fucose and variable sialylation, the identification of 0 sialic acid species is based on the identification of a +2 charge isotope envelope. In the case of the 3 sialic acid species, the identification is based on the presence of the +3 charge state ion (for which the isotopes cannot be resolved, implying charge state >+2) and the fact that the peak elutes slightly earlier in the chromatograms than the +0, +1 and +2 form.

The chromatographic behaviour often plays a critical role in confirming the identity of glycopeptide. Referring to the retention times in Table 3, several general observations can be made. Increasing degrees of sialylation results in decreasing retention. This effect is more pronounced in the larger glycoforms. For example for the K21 biantennary structure with 1 fucose, the doubly sialylated form elutes 0.17 min earlier than the singly sialylated form, whereas in the K21 tetraantennary glycoform with 2 fucoses, the triply sialylated form elutes 0.9 min earlier than the doubly sialylated form.

In the case of the Lys-C fragment K4, while the intact deglycosylated peptide was found in the PNGase F treated sample, it was not possible to determine any of the expected glycopeptides with any confidence. There are a number of possible explanations for this. It is not clear what the relative abundances of the K4-glycopeptides are in the sample. It is possible that the levels are very low and spread across a large number of glycoforms. The K4 peptide predicted molecular mass is 8162.7 and thus, if residue N_{117} is indeed glycosylated, then the mass of the peptide will be larger by 1000–4000. Thus it is possible that the multiple-charged ion envelope will fall above the 2000 mass range of the instrument. This limitation could be addressed by a proteolytic subdigestion. It should be noted that in some cases, the assignment of specific carbohydrate structures to glycopeptides is tentative. In the ab-

sence of further sub-digestions or corroborating data, the masses alone may be consistent with more than one oligosaccharide structure. For example, some lactosamine-type structures may show the same mass as oligo-mannosidic structures.

Note from Fig. 2a, that there are several major peaks which are not predicted from the primary sequence. It is not unusual to find some contamination and/or non-specific chymotryptic cleavages either from the Lys-C digestion or from subsequent deglycosylation. However, some of these peaks can be tentatively identified as K21 glycopeptides. Although the molecular mass does not correspond to common high-mannose or complex glycosylation patterns, it is likely based on the similarity of the retention and molecular mass to the other identified glycopeptides, that the peak at 32.0 min, with a molecular mass of 2235 is a hybrid form K21 glycopeptide.

So far, the method of hunting for possible glycopeptides has not been successful in positively identifying either any O-linked glycopeptides. This may not be surprising since they are present in substantially lower amounts (<1%) compared to the N-linked. Three of the four possible O-linked peptides were identified in their non-glycosylated forms. Unfortunately, there are no O-linked specific ions that can be used with CID to differentiate between O- and N-linked. However, comparing the CID chromatograms shown in Fig. 3b and c indicates that after treatment with PNGase F, the detection of these glycoforms will require enrichment of the sample.

Finally, it is possible to exploit the stepwise enzymatic removal of glycosylation in conjunction with two-dimensional (m/z vs. time) plots to confirm identification of glycosylation patterns. Fig. 5 shows the two-dimensional plots for the Lys-C digest of DSPAα1 (a), the PNGase F treated Lys-C digest of DSPAα1 (b), and the neuraminidase/O-glycosidase treated Lys-C digest of DSPAα1 (c). Since these data were acquired using the dynamically ramped CapEx voltage method, the areas of glycosylation can quickly be identified in these plots by the presence of CID produced glycomarker signals at m/z 204 (HexNAc), 292 (NeuAc) and 366 (Hex+HexNAc). Note that there are also other signals below m/z 400, which may include fragmentation of the peptide backbone itself, and which will be investigated further in the near future. Additionally, the time windows identified by these signals will also show the characteristic diagonal bands in the two-dimensional plots. These bands result from the fact that while additional glycosylation increases the mass of a specific fragment, it reduces its hydrophobicity, so that the heavier glycopeptides elute slightly earlier. Comparing the fully glycosylated Lys-C digest with the PNGase F treated sample reveals that both the marker signals and the diagonal bands that are present on the non-glycosylated signal at around 30–32 min, are completely missing from the sample in which the N-linked sugars have been removed. In the neuraminidase/O-glycosidase treated sample, on the other hand, very little difference can be detected with respect to the non-deglycosylated sample with the exception of the absence to the m/z 292 signal due to the sialic acid. Thus, the use of the two-dimensional plots in conjunction with the enzymatic deglycosylation and CID produced marker signals further confirms that the glycosylation pattern is primarily N-linked with little O-linked character. Note that in the two-dimensional plot of the O-linkage deglycosylated sample, there are a series of diagonally rising signals in m/z range 400–1000 and time range 40–55 min. These peaks do not show up in either the MS TIC or the UV chromatograms and are probably due to surfactants or detergents introduced in the deglycosylation step. Unlike glycopeptides, for these relatively hydrophobic detergents, the ladder slopes in the opposite direction, as mass increases so does the hydrophobicity and consequently, retention.

4. Conclusions

In the current work, the utilization of stepwise enzymatic deglycosylation in combination with MALDI-TOF-MS and LC–ESI-MS has been demonstrated for the characterization of a complex glycoprotein, DSPAα1. It was found that MALDI-TOF-MS was useful in confirming the enzymatic deglycosylation steps and for generating an estimate of the degree and type of glycosylation on the intact protein. LC–ESI-MS was used to characterize the peptide maps of the fully glycosylated, PNGase F treated and neuraminidase/O-glycosidase treated

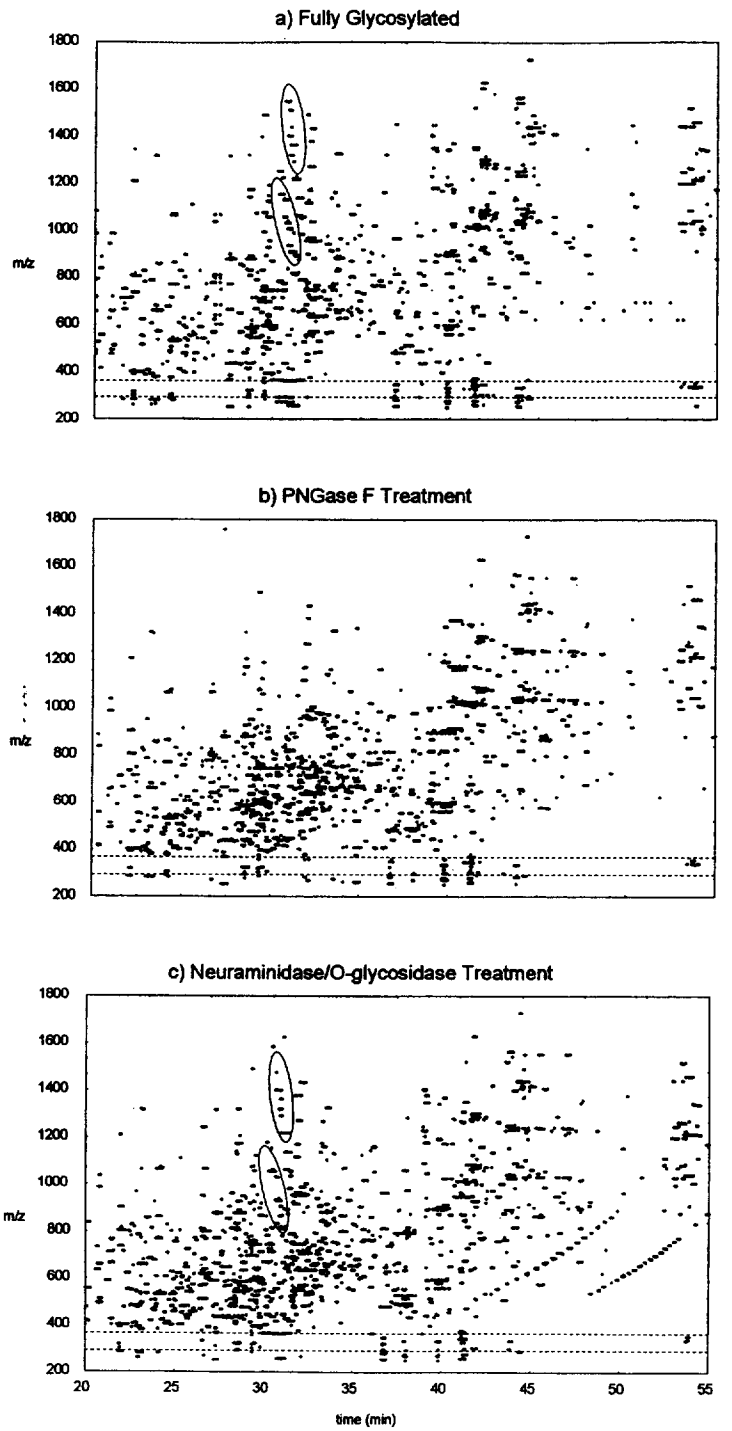

Fig. 5. Two-dimensional plots of LysC digest of DSPA: (a) fully glycosylated; (b) after PNGase F treatment; (c) after neuraminidase and O-glycosidase treatment; K21 glycopeptides are circled.

DSPAα1. The combination of the enzymatic deglycosylation with LC–ESI-MS methods utilizing collisionally induced dissociation and two-dimensional (m/z vs. time) plots to identify areas of glycosylation proved very powerful. In addition to confirming identifications, the use of the enzymatic deglycosylation allowed the differentiation of N-linked and O-linked glycosylation. In DSPAα1, it was found that the glycosylation is primarily N-linked.

In future work, we plan on exploring the application of affinity chromatography ESI-MS and affinity MS to this problem. In addition, this application will be used as a model for the use of multivariate statistical techniques to characterize highly multidimensional data sets.

References

[1] J. Kraetschmar, B. Haendler, G. Langer, W. Boidol, P. Bringman, A. Alagon, P. Donner and W.D. Schleuning, Gene, 105 (1991) 229.
[2] W. Witt, B. Maass, B. Baldus, M. Hildebrand, P. Donner and W.D. Scheuning, Circulation, 90 (1994) 421.
[3] W. Witt, B. Baldus, P. Bringmann, L. Cashion, P. Donner and W.D. Scheuning, Blood, 79 (1992) 1213.
[4] A. Apffel, J.A. Chakel, S. Udiavar, W.S. Hancock, C. Souders and E. Pungor, J. Chromatogr. A, 717 (1995) 41.
[5] K. Johansen and P.O. Edholm, J. Chromatogr., 506 (1990) 471.
[6] A.P. Bruins, T.R. Covey and J.D. Henion, Anal. Chem., 59 (1987) 2624.
[7] S.A. Hofstadler, J.H. Waho, J.E. Bruce and R.D. Smith, J. Am. Chem. Soc., 115 (1993) 6983.
[8] C.J. Barinaga, C.G. Edmonds, H.R. Udseth and R.D. Smith, Rapid Commun. Mass Spectrom., 3 (1989) 160.
[9] S.A. Carr, M.J. Huddleston and M.F. Bean, Protein Sci., 2 (1993) 183.
[10] M.J. Huddleston, M.F. Bean and S.A. Carr, Anal. Chem., 65 (1993) 877.
[11] J.J. Conboy and J.D. Henion, J. Am. Soc. Mass Spectrom., 3 (1992) 804.
[12] V. Ling, A.W. Guzzetta, E. Canova-Davis, J.T. Stults, W.S. Hancock, T.R. Covey and B.I. Shushan, Anal. Chem., 63 (1991) 2909.
[13] A.W. Guzzetta and W.S. Hancock, Recent Advances in Tryptic Mapping, CRC Series in Biotechnology, CRC Press, Boca Raton, FL, 1994, in press.
[14] A.W. Guzzetta, L.J. Basa, W.S. Hancock, B.A. Keyt and W.F. Bennet, Anal. Chem., 65 (1993) 2953.
[15] M.R. Hardy and R.R. Townsend, Proc. Natl. Acad. Sci., 85 (1988) 3289.
[16] A. Apffel, S. Fischer, G. Goldberg, P.C. Goodley and F.E. Kuhlmann, J. Chromatogr. A, 712 (1995) 177.
[17] D. Petterson, G. Tarr, S. Martin, F. Regnier, Proc. 43rd ASMS, Atlanta, GA, May 21–26, 1995.
[18] R.D. Marshall, Ann. Rev. Biochem., 41 (1972) 673.
[19] J.P. Aubert, G. Biserte and M.-H. Loucheux-Lefebre, Arch. Biochem. Biophys., 175 (1976) 410.
[20] Y. Yang and R. Orlando, Proc. 43rd ASMS, Atlanta, GA, May 21–26, 1995.
[21] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong and C.M. Whitehouse, Science, 246 (1989) 64.