ELSEVIER

Journal of Chromatography A, 750 (1996) 35–42

JOURNAL OF CHROMATOGRAPHY A

Application of multidimensional affinity high-performance liquid chromatography and electrospray ionization liquid chromatography–mass spectrometry to the characterization of glycosylation in single-chain plasminogen activator Initial results

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, 3500 Deer Creek Road, Palo Alto, CA 94304, USA*
[b]*Berlex Biosciences, 430 Valley Dr., Brisbane, CA 94005, USA*

Abstract

Preliminary results are presented using a combination of affinity chromatography, reversed-phase HPLC and electrospray ionization mass spectrometry to produced peptide maps for N-linked, O-linked and non-glycosylated peptides from an endoproteinase LysC digest of DSPAα1, a recombinant DNA derived glycoprotein. Although the system was used to identify a number of major N-linked structures, notably complex biantennary structures attached to asparagine 362, no O-linked glycopeptides from the possible 4 attachment sites were identified. The system did, however, demonstrate the feasibility of the approach and the applicability of the instrumental system.

Keywords: Glycosylation; Plasminogen activator; Glycoproteins; Proteins

1. Introduction

The growth of the biotechnology industry, and more specifically recombinant DNA derived protein pharmaceuticals, has placed enormous demands on analytical chemistry. Techniques such as protein purification by liquid chromatography and electrophoresis, cDNA sequence analysis, protein and peptide sequencing and peptide mapping by high-performance liquid chromatography–electrospray ionization mass spectrometry (HPLC–ESI-MS) have become invaluable tools for the protein chemist. In spite of the power of these techniques, however, there are a number of areas that still remain formidable challenges. As clarified by the recent FDA discussions on 'Well Characterized Biological Products', amongst these challenges are post-translational modifications, such as glycosylation and phosphorylation. The complexity introduced by protein glycosylation lies in the enormous microheterogeneity introduced to the final product. While the protein amino-acid sequence is maintained, glycosylation can occur at a number of sites and can have a varied and heterogeneous character resulting in a large number of distinct glycoforms. Our work has focused on the characterization of glycosylation patterns in recombinant DNA derived proteins such as

*Corresponding author.

0021-9673/96/$15.00
PII S0021-9673(96)00528-6

DSPAα1 (Desmodus Salivary Plasminogen Activator), a single-chain plasminogen activator.

Affinity chromatography using immobilized lectin columns is becoming a powerful technique for the fractionation of glycopeptides [1,2]. These techniques have largely been used to isolate specific glycopeptides rather than a total fractionation of the heterogeneous glycopeptide pool. Electrospray ionization mass spectrometry has also been used in a number of studies to characterize glycopeptides [3]. However, the combination of these two techniques has only been reported in an off-line coupling [4].

DSPAα1, is a single-chain plasminogen activator with thrombolytic activity. It has a molecular mass of approximately M_r 54 000 with approximately 10% glycosylation distributed across 2 sites of N-linked glycosylation and 4 sites of O-linked glycosylation. There are 14 cysteines, thought to form 7 disulfide bonds by homology with other similar serine proteases. We chose this system firstly due to a genuine analytical need, but secondly as a challenging example to evaluate the strengths and weakness of various analytical approaches.

In the first study of this series [5], we utilized combinations of HPLC, capillary electrophoresis (CE), ESI-MS and matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) to investigate the general character of the protein and the degree of glycosylation. In this study, the use of HPLC–ESI-MS focused on identifying the consistency of the predicted sequence on the basis of the separation and identification of fragments produced upon digestion with proteolytic enzymes. The general identification of glycosylation was addressed through the use of diagonal elution bands in two-dimensional (m/z versus time) plots and the detection of glycosylation specific fragment ions (glycomarkers) through the use of in-source collisionally induced dissociation (CID). From these studies, it was confirmed that the experimentally determined molecular mass of the intact protein is approximately 4000 higher than that predicted from the amino-acid sequence. Furthermore, it was shown from the HPLC–ESI-MS analysis of peptide maps that the majority of the glycosylation was present in two main sites, tentatively identified as the two predicted N-linked sites.

In a second study [6], we further evaluated the specific glycosylation patterns through the comparison of samples before and after stepwise systematic enzymatic removal of the carbohydrates. MALDI-TOF-MS was used to evaluate the degree of N-linked and O-linked glycosylation and sialylation on the intact protein, while HPLC–ESI-MS was used to locate the sites of glycosylation by comparison of fully glycosylated and deglycosylated peptide digests. In these studies, a large number of N-linked structures has been identified, but positive identification of the O-linked structures has remained tentative, at best.

We have also reported on the off-line coupling of CE and MALDI-TOF-MS [7], to initially fractionate classes of glycoforms in the intact protein product for subsequent mass analysis. These studies, in addition to independent carbohydrate analyses have indicated that while there is a substantial amount of N-linked glycosylation (approximately 10%), the O-linked glycoforms are much less abundant (<1%).

In the present work, multidimensional HPLC combining affinity chromatography and reversed-phase chromatography are used in conjunction with ESI-MS to further characterize the glycosylation patterns in DSPAα1. Using automated column switching techniques, lectin-based affinity pre-columns are initially used to enrich and fractionate enzymatic digests of DSPAα1 according to glycosylation type and then the isolated fractions are sequentially separated using reversed-phase HPLC to confirm the identity of previously determined glycoforms as well as identify less abundant forms.

2. Experimental

2.1. HPLC

The analytical system was set up as shown in Fig. 1. The HPLC separations were performed on a Hewlett-Packard 1090 liquid chromatography system equipped with a DR5 ternary solvent delivery system, column oven, autosampler with a 250 μl injection volume and a diode-array detector (Hewlett-Packard, Wilmington, DE, USA).

Column selection was accomplished with a series of 4 Valco 10-port Cheminert valves equipped with microelectroactuators (Valco, Houston, TX, USA).

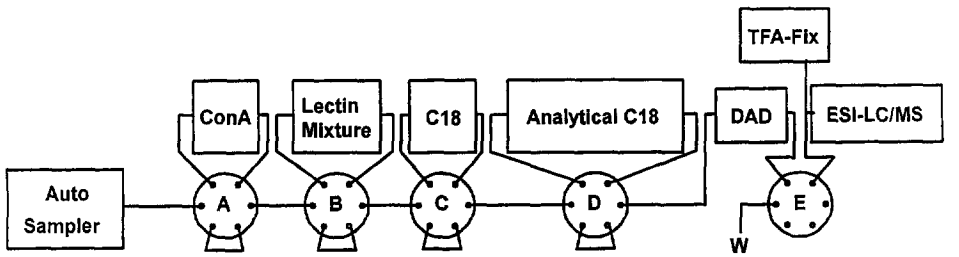

Fig. 1. System schematic: valving scheme for multidimensional HPLC system.

The internal column switching valve in the HP1090LC was used to divert the first 5 min of each peptide map away from the electrospray source. The valves were software controlled from the external event contact closures of the HP1090 LC.

For the multidimensional HPLC operations, the operation sequence is shown in Table 1. Valve A was used to select or isolate a 10×2.1 mm I.D. precolumn (Upchurch Scientific) packed with immobilized Concanavalin A (Con A) (EY Laboratories, San Francisco, CA, USA). Valve B was used to select or isolate a 10×2.1 mm I.D. precolumn (Upchurch Scientific) packed with an equivolume mixture of four lectins [Datura Stramonium Agglutin (DSA), Jacalin (JAC), Peanut Lectin (PNA) and Amarathin (ACA)] obtained from EY Laboratories. The selectivities of immobilized lectins is shown in Table 2. Samples were loaded and washed on the lectin columns in a mobile phase consisting of 100 m*M* Tris, 150 m*M* NaCl, 1 m*M* $MgCl_2$, 1 m*M* $CaCl_2$, pH 7.2 (HP1090 Channel C). Both lectin columns were eluted with a solvent consisting of a mixture of 0.25 *M* of each of the following sugars in the loading buffer; α-lactose, α-D-galactose, methyl-α-mannopyranoside, *n*-acetyl galactose and mannose.

Valve C was used to select or isolate a 10×2.1 mm I.D. 5 μm RP18 precolumn (Upchurch Scientific). Valve D was used to select or isolate the analytical column, a 250×2 mm I.D., 3 μm, 120 Å, YMC ODS-AQ column (YMC Industries, Wilmington, NC, USA). After transferring peptides onto the RP18 precolumn, it was washed with 0.1% TFA (HP1090 Channel A).

A HP1050 quaternary pumping system was used to deliver the eluent flow for the lectin affinity columns and the post-column addition of the electrospray 'TFA Fix' [8,9], consisting of propionic acid–isopropanol (75:25) at 0.1 ml/min.

For peptide mapping, a standard solvent system of 0.1% TFA in water (solvent A) and 0.09% TFA in acetonitrile (solvent B) was used with a linear gradient from 0% to 60% B/90 min at a flow-rate of 0.2 ml/min. The gradient then continued to 90% B in 10 min followed by a 10 min re-equilibration period at 0% B. UV detector signals at 214, 275 and 294 nm were used with a 4 nm bandwidth and the

Table 1
Sequence of multidimensional operation[a]

Operation	Valve A N-Linked Con A Precolumn	Valve B O-Linked Mixed Lectin Precolumn	Valve C RP18 Precolumn	Valve D Analytical column	Valve E (Internal) ESI Divert
Load Sample	+	+	+	–	
Rinse RP18 Precolumn	–	–	+	–	
Peptide Map 1 (non-glycosylated)	–	–	+	+	1st 5 min
Elute N-linked Glycopeptides	+	–	+	–	
Rinse RP18 Precolumn	–	–	+	–	
Peptide Map 2 (N-linked)	–	–	+	+	1st 5 min
Elute O-Linked Glycopeptides	–	+	+	–	
Rinse RP18 Precolumn	–	–	+	–	
Peptide Map 3 (O-Linked)	–	–	+	+	1st 5 min

[a] + = Column in-line; – = column bypass.

Table 2
Selectivities of immobilized lectins

Lectin	Source	Abbreviation	Binding specificity	Hapten sugar for elution
Concanavalin A	Canavalia Ensiformis	Con A	Branched mannoses, αD-Mannose, αGlcNAc	Methyl-D-mannopyranoside
Datura Stramonium Agglutin	Datura Stramonium	DSA	GlcNAc-(Ser/Thr)	GlcNAc
Jacalin	Artocarpus Integrifolia	JAC	Gal-β1-3GalNAc-(Ser/Thr)	D-Galactose
Peanut Lectin	Arachis Hypogea	PNA	Gal-α1-GalcNAC-(Ser/Thr) β-Gal	Lactose
Amaranthin	Amaranthus Caudatus		NeuNAc-Gal-GalNAc	Lactose

reference signals were 450 nm with a 10 nm bandwidth.

2.2. Electrospray 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 API-electrospray source with a high flow nebulizer option. Both the 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 the nebulizing (1.5 l/min) and drying (15 l/min at 300°C) gases.

For peptide mapping, MS data were acquired in scan mode, scanning from m/z 200 to 2000 at an acquisition rate of 1.0 Hz with m/z 0.15 stepsize. Unit mass resolution was maintained for all experiments. Data were filtered in the mass domain with a M_r 0.05 Gaussian mass filter and in the time domain with a 0.05 min Gaussian time filter. To allow detection of CID-generated glycomarker ions m/z 204, 292 and 377, the ESI Source Capillary Exit (CapEx) voltage was set to 300 V for m/z 200–380, and to 100 V for m/z 400–2000. 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.

2.3. Chemicals

HPLC-grade acetonitrile and trifluoroacetic acid (TFA) was obtained from J.T. Baker (Phillipsburg, NJ, USA). Iodoacetic acid, guanidine HCl, dithiothrietol, $MgCl_2$, $CaCl_2$, NaCl and Tris were obtained from Sigma (St. Louis, MO, USA). Endoproteinase Lys C was obtained from Promega (Madison, WI, USA). Immobilized lectins and sugars were obtained from EY Labs. (San Mateo, CA, USA).

2.4. Sample preparation

The endoproteinase LysC digest of DSPAα1 was prepared as follows. 5 nmol DSPAα1 was desalted using a 10K NMWL centrifugal ultrafiltration tube at 3000×g for 20 min. The sample was redesolved and denatured in 6 M guanidine·HCl. The sample was then reduced by addition of 100 μl (9 mg/ml) dithiothreitol in 100 mM ammonium bicarbonate at pH 7.3 and incubated at 37°C for 30 min. Alkylation of the reduced protein was done by a further 30 min incubation with 100 μl iodoacetic acid (26 mg/ml) in ammonium bicarbonate buffer. The reduced and alkylated DSPAα1 was then desalted using the 10K NMWL centrifugal ultrafiltration unit (Millipore, Bedford, MA, USA) for 1 h at 3000×g. The reduced, alkylated, desalted DSPAα1 was then digested with endoproteinase Lys-C (Promega, Madison, WI, USA) in 100 mM ammonium bicarbonate, pH 7.2, at 37°C with an enzyme-to-substrate (mass) ratio of 1:20 for 18 h. The resulting digestion was at a final concentration of 8 pmol/μl. The final digests were

acidified and stored at 5°C until use. 250 μl injections were made at the level of 2 nmol/250 μl.

3. Results and discussion

3.1. Analytical results

The general analytical approach utilized in this system was to inject the sample through a series of precolumns, allowing the specificity of each precolumn to isolate and concentrate a specific fraction of the sample. The isolated fractions were then sequentially eluted and analyzed by reversed-phase HPLC with electrospray ionization mass spectrometric detection. The analytical sequence is shown in Table 1. Initially 250 μl of endoproteinase LysC digest from approximately 2 nmol of the original protein was injected into the system. This exceptionally large amount of sample for typical peptide mapping applications by HPLC–ESI-MS, was used in order to focus the identification and detection on the less abundant glycopeptides in the sample. Using an appropriately buffered mobile phase, this initial sample is pumped through a series of three columns. N-linked glycopeptides are retained on the first immobilized lectin precolumn (concanavalin A); the remaining (unretained) glycopeptides are retained on the second immobilized lectin precolumn (mixed lectins), and, finally, the non-glycosylated peptides are retained on the third precolumn packed with RP18 material. In the subsequent series of operations, each of the precolumns was rinsed, eluted onto the RP precolumn and then onto the analytical column and separated with a standard gradient with detection by ESI-MS.

The affinity media for the immobilized lectin precolumns were chosen to separate, as groups, all N-linked glycopeptides from all O-linked glycopeptides. While the choice of the immobilized concanavalin A is straightforward, since most N-linked glycopeptides should be retained on this phase under appropriate buffer conditions, the affinity phase for the isolation of O-linked glycostructures is more complex. Unfortunately, the O-linked structures lack a common core structure, and without prior knowledge of the specific structures to be isolated, a mixture of different lectins had to be used that would include binding properties for the most common O-linked carbohydrate structures. In retrospect, this strategy was not completely successful and alternate approaches are under consideration.

Using ESI-MS with a dynamically ramped capillary exit (CapEx) potential, it was possible to simultaneously determine the peptide molecular masses and detect any CID fragmentation indicative of glycopeptides. Although this approach is not as sensitive for glycomarker detection as selected ion monitoring, it generally yields acceptable results, especially at higher concentration levels.

Fig. 2 shows the analytical peptide maps produced from this study. Fig. 2a shows the control peptide map of the entire digest without affinity fractionation. Fig. 2b shows the peptide map of the non-glycosylated peptides, i.e. those not retained by either lectin precolumn. Fig. 2c shows the peptide map of the N-linked glycopeptides isolated on the Con A precolumn. Finally, Fig. 2d shows the peptide map of the O-linked glycopeptides retained on the mixed lectin precolumn.

In the control peptide map (Fig. 2a) all the peptide fragments predicted from a theoretical endoproteinase LysC digest of the known amino-acid sequence were found with the exception of those fragments with masses less than M_r=400 and the two predicted sites of N-linked glycosylation (K4 and K21). It is interesting to note that the four predicted sites of O-linked glycosylation (K10, K17, K19 and K22) were found in their non-glycosylated forms, indicating partial occupation of these glycostructures. The resolution of this chromatogram is somewhat poorer than that seen at lower levels in previous studies [6] indicating either band broadening in the transfer from precolumn to analytical column or overloading of the analytical column. Similarly, the peptide map of the non-glycosylated peptides (Fig. 2b) shows the same pattern with the exception of a number of peaks between 28 and 32 min. Inspection of the extracted ion chromatograms of glycomarker fragments, m/z 204, 292 and 377 (not shown) shows the absence of any significant signal.

In the peptide map of the Con A (Fig. 2c) isolated glycopeptides, a series of three major peaks elute between 29 and 32 min. Examination of the spectra of these peaks shows them to be complex

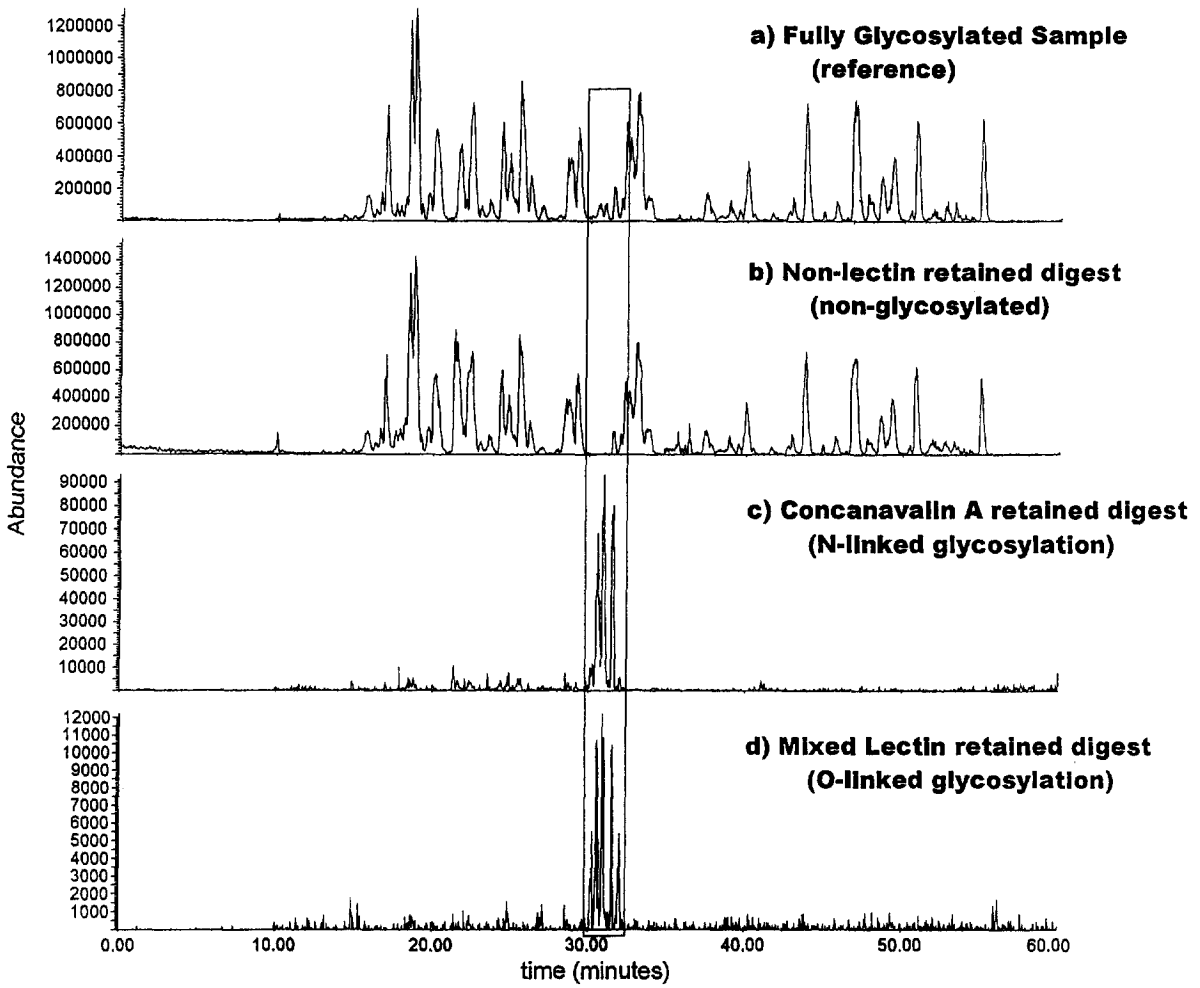

Fig. 2. Electrospray LC–MS results: (a) peptide map reference control (separate run); (b) peptide map of non-lectin precolumn retained peptides (i.e. non-glycosylated); (c) peptide map of Con A precolumn retained peptides (i.e. N-linked glycopeptides); (d) peptide map of mixed lectin precolumn retained peptides (i.e. O-linked glycopeptides). Note especially the changes inside the boxed region.

fucosylated biantennary structures with zero, one or two sialic acid moieties attached to LysC digest fragment K21. By searching for other possible structures with extracted ion chromatograms, it is also possible to tentatively identify several high mannose and triantennary structures as well. Note that the absolute signal abundance is reduced from 1 400 000 for the non-glycosylated peptide map to 90 000 for the N-linked peptide map.

In the peptide map of the mixed lectin column (Fig. 2d), a number of peaks are also found to elute between 28 and 32 min with an absolute abundance of only 12 000. Unfortunately, inspection of the spectra of these peaks shows them to be very similar to those determined in the peptide map of the Con A isolated samples. This result could be attributed to a low level of cross specificity for these lectins between N- and O-linked sugars. Furthermore, no other glycopeptides were identified in this chromatogram either through the use of CID glycomarkers or searching for possible structures with extracted ion studies.

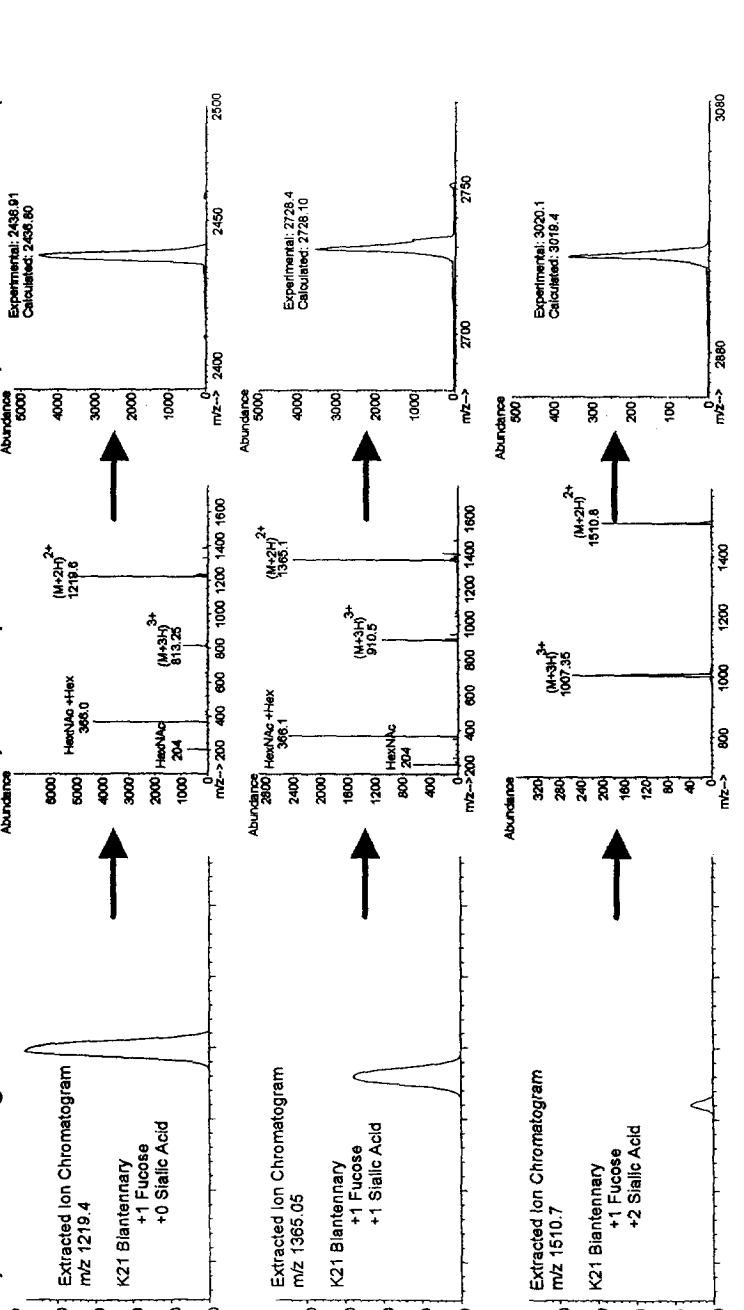

Fig. 3. Identification of specific glycopeptide structures in the Con A precolumn retained peptide map. Data are shown for a family of K21 biantennary structures with 1 fucose and 0, 1 and 2 sialic acids, respectively: (a) extracted ion chromatogram detail (total ion chromatogram shown on bottom for reference); (b) raw electrospray spectra; (c) deconvoluted electrospray spectra.

To demonstrate the utility of electrospray ionization mass spectrometry as a detector in this application, Fig. 3 shows examples of three related glycopeptides that could be found in the Con A retained peptide map. Specifically three biantennary structures from LysC digest fragment K21 with one fucose and zero, one and two sialic acids, respectively. Initially, extracted ion chromatograms can be produced to locate specific structures. The raw electrospray mass spectra can then be obtained which can be deconvoluted to show the mass of the parent molecule.

This study was done on a sample available in only small amounts, and, as such, no studies were done on the reproducibility. However, previous work on both multidimensional HPLC and ESI-MS have shown these techniques to be sufficiently reproducible for accurate qualitative analysis and, if proper procedures are used, quantitative analysis.

4. Conclusions

This work represents one of the first attempts at the combination of affinity chromatography and ESI-MS for the comprehensive characterization of N-linked and O-linked glycosylation in a complex glycoprotein. Although admittedly not completely successful, the work does show a good deal of potential, and the results are instructive and suggestive of possible improvements. The results do demonstrate the feasibility of a multidimensional HPLC system combined with ESI-MS for application to the complex problem of the rapid identification of glycosylation of recombinant DNA derived proteins. In this application, it was possible to reliably identify a number of N-linked glycopeptides present in DSPAα1. Compared to other multidimensional HPLC systems, based on the transfer of specific peaks from one separation stage to another (heart-cutting), an approach such as described here in which retention and elution characteristics can be strictly and reproducibly controlled offers advantages in terms of robustness and automation.

Although no O-linked structures were detected in this study, the approach was useful for the characterization of one of the N-linked glycopeptides in this complex sample. Further work is underway to improve the system, allowing the isolation of the less abundant O-linked glycopeptides.

References

[1] T. Endo, J. Chromatogr. A, 720 (1996) 251–261.
[2] K. Jones, LC–GC, 4 (1991) 32–37.
[3] S.A. Carr, Adv. Drug. Deliv. Rev., 4 (1990) 113.
[4] D.H. Hawke, K.L. Hsi, L.R. Zieske, L. Chen and P.M. Yuan, in J.J. Vilafranca (Ed.), Techniques in Protein Chemistry III, Academic Press, San Diego, 1992, pp. 315–326.
[5] A. Apffel, J. Chakel, S. Udiavar, W.S. Hancock, C. Souders and E. Pungor Jr., J. Chromatogr. A, 717 (1995) 41–60.
[6] A. Apffel, J. Chakel, W.S. Hancock, C. Souders, T. M'Timkulu and E. Pungor Jr., J. Chromatogr. A, 732 (1996) 27.
[7] J.A. Chakel, J.A. Apffel and W.S. Hancock, LC–GC, 13 (1995) 866–876.
[8] A. Apffel, S. Fischer, G. Goldberg, P.C. Goodley and F.E. Kuhlmann, J. Chromatogr. A, 712 (1995) 177–190.
[9] F.E. Kuhlmann, A. Apffel, S.M. Fischer, G. Goldberg and P.C. Goodley, J. Am. Soc. Mass Spectrom., 6 (1995) 1221–1225.