ELSEVIER

Journal of Chromatography A, 870 (2000) 295–313

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Signature-peptide approach to detecting proteins in complex mixtures

Minghui Geng, Junyan Ji, Fred E. Regnier*

Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA

Abstract

The objective of the work presented in this paper was to test the concept that tryptic peptides may be used as analytical surrogates of the protein from which they were derived. Proteins in complex mixtures were digested with trypsin and classes of peptide fragments selected by affinity chromatography, lectin columns were used in this case. Affinity selected peptide mixtures were directly transferred to a high-resolution reversed-phase chromatography column and further resolved into fractions that were collected and subjected to matrix-assisted laser desorption ionization (MALDI) mass spectrometry. The presence of specific proteins was determined by identification of signature peptides in the mass spectra. Data are also presented that suggest proteins may be quantified as their signature peptides by using isotopically labeled internal standards. Isotope ratios of peptides were determined by MALDI mass spectrometry and used to determine the concentration of a peptide relative to that of the labeled internal standard. Peptides in tryptic digests were labeled by acetylation with acetyl *N*-hydroxysuccinimide while internal standard peptides were labeled with the trideuteroacetylated analogue. Advantages of this approach are that (i) it is easier to separate peptides than proteins, (ii) native structure of the protein does not have to be maintained during the analysis, (iii) structural variants do not interfere and (iv) putative proteins suggested from DNA databases can be recognized by using a signature peptide probe.

Keywords: Signature-peptide approach; Proteins; Peptides

1. Introduction

Proteins in complex mixtures are generally detected by some type of fractionation [1] or immunological assay technique [2]. The advantages of immunological assay methods are their sensitivity, specificity for certain structural features of antigens, low cost, and simplicity of execution. Immunological assays are generally restricted to the determination of single protein analytes. This means it is necessary to conduct multiple assays when it is necessary to determine small numbers of proteins in a sample. Hormone–receptor association [3], enzyme–inhibitor binding [4], DNA–protein binding [5] and lectin–glycoprotein association [6] are other types of bioaffinity that have been exploited in protein identification, but are not as widely used as immunorecognition. Although not biospecific, immobilized metal affinity chromatography (IMAC) is yet another affinity method that recognizes a specific structural element of polypeptides [7].

The fractionation approach to protein identification in mixtures is often more lengthy because analytes must be purified sufficiently to allow a detector to recognize specific features of the protein. Properties ranging from chemical reactivity to spectral characteristics and molecular mass have been exploited for detection [8–10]. Higher degrees of purification are required to eliminate interfering substances as the

*Corresponding author. Tel.: +1-317-494-1648; fax: +1-317-494-0359.

0021-9673/00/$ – see front matter
PII: S0021-9673(99)00951-6

detection mode becomes less specific. Since no single purification mode can resolve thousands of proteins, multidimensional fractionation procedures must be used with complex mixtures. Ideally, the various separation modes constituting the multidimensional method should be orthogonal in selectivity. The two-dimensional (2D) gel electrophoresis method of O'Farrel [11] is a good example. The first dimension exploits isoelectric focusing while the second is based on molecular size discrimination. At the limit, 6000 or more proteins can be resolved. 2D gel electrophoresis is now widely used in proteomics where it is the objective to identify thousands of proteins in complex biological extracts [12,13]. The most definitive way to identify proteins in gels is by mass spectral analysis of peptides obtained from a tryptic digest of the excised spot [14]. Identification is greatly facilitated when the peptide molecular mass can be correlated with tryptic cleavage fragments predicted from a genomic database. Proteins can also be identified by their separation characteristics alone in some cases. The great advantage of this technique is that large numbers of proteins can be identified simultaneously. The disadvantages of 2D electrophoresis followed by tryptic mapping are (1) it is very slow and requires a large number of either manual or robotic manipulations (2) charged isoforms are resolved whereas uncharged variants in which no new charge is introduced are not (3) proteins must be soluble to be examined and (4) quantification by staining is poor.

Multidimensional chromatography is faster, easier to automate, and couples more readily to MS than 2D gel electrophoresis. One of the more attractive features of chromatographic systems is that they allow many dimensions of analysis to be coupled by analyte transfer between dimensions through automated valve switching. A recent [15] report of an integrated six dimensional analytical system in which serum hemoglobin was purified and sequenced automatically in <2 h is an example. Subsequent to purification on an immunoaffinity column, hemoglobin was desorbed into an ion-exchange column for buffer exchange and then tryptic digested by passage through an immobilized trypsin column. Peptides eluting from the immobilized enzyme column were concentrated and desalted on a small, low-surface-area reversed-phase liquid chromatography (RPLC) column and then transferred to an analytical RPLC column where they were separated and introduced into a mass spectrometer through an electrospray interface. Identification at the primary structure level was achieved by a combination of chromatographic properties and multidimensional mass spectrometry of the tryptic peptides. The ability of the immunosorbent to rapidly select the desired analyte for analysis was a great asset to this analysis. Size-exclusion or ion-exchange chromatography coupled to reversed-phase chromatography are other examples of multidimensional systems, albeit of lower selectivity than those using immunosorbents [16].

Although the methods described above are highly selective and widely used, they have some attributes that limit their efficacy. One is the necessity for proteins to be soluble before they can be analyzed. This can be a serious limitation in the case of membrane and structural proteins that are sparingly soluble. A second is that it is desirable or even necessary in some cases for the protein analyte to be of native structure during at least part of the analysis. This is a limitation because it restricts the sample preparation protocol. Native macromolecular structures are notoriously more difficult to analyze than small molecules. The necessity for post separation proteolysis, as in the 2D gel approach, is another limitation. Large numbers of fractions must be subjected to a 24 h tryptic digestion protocol in the analysis of a single sample when many proteins are being identified. This slow tryptic digestion step is necessary because the mass of intact proteins is far less useful in searching DNA databases than that of signature peptides derived from the protein. And finally, pure proteins are a prerequisite for antibody preparation in all the immunorecognition methods. In addition to being lengthy, laborious, and costly, many antigens have never been purified. This is particularly true of proteins predicted by genomic data alone. Purification is complicated by the fact that one does not know the degree to which a protein is expressed, whether it is part of a multisubunit complex, or if it is post translationally modified. All of these issues have a major impact on antigen purification.

This paper is based on the premise that through affinity selection of tryptic peptides unique to a protein analyte, some of the limitations noted above

can be circumvented. These analytical surrogates will be identified in the discussion below as 'signature peptides'.

2. Materials and methods

2.1. Materials

Human serotransferrin, human serum, *N*-tosyl-L-phenylalanine chloromethyl ketone (TPCK)-treated trypsin, concanavalin A (Con A), *Bandeiraea simplicifolia* (BS-II) lectin, tris(hydroxymethyl)aminomethane (Tris base), iodoacetic acid, tris(hydroxymethyl)aminomethane hydrochloride (Tris acid), cysteine, dithiothreitol (DTT), *N*-tosyl-L-lysyl chloromethyl ketone (TLCK), and *N*-acetyl-D-glucosamine were purchased from Sigma (St. Louis, MO, USA). Nuclear extract from calf thymus was provided by Professor M. Bina (Department of Chemistry, Purdue University, W. Lafayette, IN, USA). LiChrospher Si 1000 (10 μm, 1000 Å) was obtained from Merck (Darmstadt, Germany). 3,5-Dimethoxy-4-hydroxy-cinnamic acid (sinipinic acid), 3-aminopropyltriethoxysilane, polyacrylic acid (PAA), and dicyclohexyl carbodiimide (DCC), d_3-C^1 acetic anhydride were purchased from Aldrich (Milwaukee, WI, USA). Methyl-α-D-mannopyranoside was obtained from Calbiochem (La Jolla, CA, USA). Toluene, 4-dioxane and dimethylsulfoxide (DMSO) were purchased from Fisher Scientific (Fair Lawn, NJ, USA). *N*-Hydroxyl succinimide (NHS) and high-performance liquid chromatography (HPLC)-grade trifluoroacetic acid (TFA) were purchased from Pierce (Rockford, IL, USA). HPLC-grade water and acetonitrile (ACN) were purchased from EM science (Gibbstown, NJ, USA). All reagents used directly without further purification.

2.2. Synthesis of lectin column

A 1-g of LiChrospher Si 1000 was activated for 5 h at room temperature by addition of 40 ml 6 *M* HCl. The silica particles were then filtered and washed to neutrality with deionized water after which they were dried initially for 2 h at 105°C and then at 215°C overnight. Silica particles thus treated were reacted with 0.5% 3-aminopropyltriethoxysilane in 10 ml toluene for 24 h at 105°C to produce 3-aminopropylsilane derivatized silica (APS silica). Polyacrylic acid (0.503 g; M_r 450 000), *N*-hydroxysuccinamide (1.672 g), and dicyclohexyl carbodiimide (6.0 g) were dissolved into 40 ml DMSO and shaken for 3 h at room temperature to activate the polymer. The reaction mixture was filtered and the activated polymer harvested in the supernatant. Acrylate polymer was grafted to silica particles by adding the APS silica described above to the activated acrylate polymer containing supernatant. Following a 12-h reaction at room temperature, the particles were filtered and washed sequentially with 50 ml DMSO, 50 ml dioxane and 50 ml deionized water. This procedure produces a polyacrylate coated silica with residual *N*-acyloxysuccinamide activated groups, specified as NAS-PAA silica. NAS-PAA silica (0.5 g) was added to 10 ml of 0.1 *M* $NaHCO_3$ (pH 7.5) containing 0.2 *M* methyl-α-D-mannopyranoside and 200 mg Con A. The reaction was allowed to proceed with shaking for 12 h at room temperature after which the immobilized Con A sorbent was isolated by centrifugation and was washed with 0.1 *M* Tris buffer (pH 7.5). The sorbent was stored in 0.1 *M* Tris buffer (pH 7.5) with 0.2 *M* NaCl until use.

NAS-PAA silica (0.3 g) was added to 10 ml of 0.1 *M* $NaHCO_3$ buffer (pH 7.5) containing 0.2 *M* *N*-acetyl-D-glucosamine and 20 mg BS-II lectin. The reaction was allowed to proceed with shaking for 12 h at room temperature after which the immobilized lectin containing particles were isolated by centrifugation, washed with 0.1 *M* (pH 7.5) Tris buffer, and packed into a stainless steel column (50×4.6 mm) using the wash buffer and a high-pressure pump from Shandon Southern Instruments (Sewickley, PA, USA). Affinity columns were washed by 0.1 *M* Tris (pH 7.5) with 0.2 *M* NaCl before use.

2.3. Proteolysis

Human serotransferrin (5 mg), nuclear extract from bovine cells, or human serum were reduced and alkylated in the same way by adding to 1 ml 0.2 *M* Tris buffer (pH 8.5) containing 8 *M* urea and 10 m*M* DTT. After a 2-h incubated at 37°C, iodoacetic acid was added to a final concentration of 20 m*M* and incubated in darkness on ice for a further 2 h.

Cysteine was then added to the reaction mixture to a final concentration of 40 m*M* and the reaction allowed to proceed at room temperature for 30 min. After dilution with 0.2 *M* Tris buffer to a final urea concentration of 3 *M*, TPCK-treated trypsin (2%, w/w, of enzyme to that of the protein) was added and incubated for 24 h at 37°C. Digestion was stopped by adding TLCK in a slight molar excess over that of trypsin.

2.4. Chromatography

All chromatographic steps were performed using an Integral microanalytical workstation from PE Biosystems (Framingham, MA, USA). Tryptic digested human serotransferrin (0.1 ml) was injected onto the Con A affinity column that had been equilibrated with a loading buffer containing 1 m*M* $CaCl_2$, 1 m*M* $MgCl_2$, 0.2 *M* NaCl and 0.1 *M* Tris–HCl (pH 7.5). The Con A column was eluted sequentially at 1 ml/min with two column volumes of loading buffer and then 0.2 *M* methyl-α-D-mannopyranoside in 0.1 *M* Tris (pH 6.0). Analytes displaced from the affinity column with 0.2 *M* methyl-α-D-mannopyranoside were directed to a 250×4.6 mm Peptide C_{18} (PE Biosystems) analytical reversed-phase HPLC column, which had been equilibrated for 5 min at 1.0 ml/min with 5% ACN containing 0.1% aqueous TFA. The glycopeptides were then eluted at 1.0 ml/min in a 35-min linear gradient to 50% ACN in 0.1% aqueous TFA. Eluted peptides were monitored at 220 nm and fractions manually collected for matrix-assisted laser desorption ioniation time-of-flight (MALDI-TOF) analysis.

Tryptic digested human serum (0.2 ml) was injected on the Con A and reversed-phase HPLC column using conditions similar to those used with human serotransferrin with the following exceptions. The reversed-phase column was washed for 10 min at 1 ml/min with 10% ACN containing 0.1% aqueous TFA and the glycopeptides were eluted at 1 ml/min with a 120-min linear gradient to 70% ACN containing 0.1% aqueous TFA.

Nuclear extract (0.1 ml) was injected onto the BS-II column which had been equilibrated with loading buffer, 0.2 *M* NaCl with 0.1 *M* Tris (pH 7.5). After sample loading the BS-II column was washed with 20 column volumes of loading buffer and then eluted with 0.2 *M* *N*-acetyl-D-glucosamine in the loading buffer. Glycopeptides and glycoproteins eluted from the BS-II column were transferred to a reversed-phase column, which had been equilibrated for 5 min at 1 ml/min with 5% ACN containing 0.1% aqueous TFA. The glycoproteins were then eluted at 1 ml/min with a 25-min linear gradient to 35% ACN containing 0.1% aqueous TFA. The glycopeptides were eluted at 1 ml/min with a 35-min linear gradient to 50% ACN containing 0.1% aqueous TFA.

2.5. Synthesis of d_3-C^1 N-acetoxysuccinamide[1]

A solution of 4.0 g (34.8 mmol) of *N*-hydroxysuccinimide in 10.7 g (105 mmol) of d_3-C^1 acetic anhydride was stirred at room temperature. After 10 min, white crystals began to deposit. The liquid phase was allowed to evaporate and the crystalline residue extracted with hexane which is allowed to dry in vacuum. The yield of the substances was 5.43 g (100%), m.p. 133–134°C.

2.6. Acetylation reaction with the peptides

A 3-fold molar excess of *N*-acetoxysuccinamide and d_3-C^1 *N*-acetoxysuccinamide was added individually to the two equal aliquots of 1 mg/ml peptide solution in phosphate buffer at pH 7.5, respectively. The reaction was carried at room temperature. After stirring for about 4–5 h, equal aliquots of the two samples were mixed and purified on a C_{18} column. The collected fraction were then subjected to MALDI-MS.

2.7. MALDI-TOF-MS

MALDI-TOF-MS was performed using a Voyager DE-RP BioSpectrometry workstation (PE Biosystems). Samples were prepared by mixing a 1-μl aliquot with 1 μl of matrix solution. The matrix solution for glycopeptides was prepared by saturating a water–ACN (50:50, v/v), 3% TFA solution with sinipinic acid. A 1-μl sample volume was spotted into a well of the MALDI sample plate and allowed to air-dry before being placed in the mass spectrometer. All peptides were analyzed in the linear, positive ion mode by delayed extraction using an accelerating

voltage of 20 kV unless otherwise noted. External calibration was achieved using a standard 'calibration 2' mixture from PE Biosystems.

The matrix for acetylated peptides was a solution of 3% TFA, ACN–water (50:50) solution saturated with α-cyano-4-hydroxycinnamic acid. Peptide quantitation was performed on MALDI-TOF-MS in the reflector mode as described above. Ten spectra were collected from each sample spot and the peak intensities averaged for each spot. A linear equation was deduced from the ion current intensity ratio of the deuterium-labeled and the unlabeled acetylated peptides versus the ratio of the amount of these two peptides.

3. Results and discussion

3.1. Analytical strategy

The work reported here is based on the proposition that signature peptides generated by tryptic digestion of sample proteins may be selected from complex mixtures and be used as analytical surrogates for the protein from which they were derived. The rationale for this approach is that (i) it will be easier to separate and identify signature peptides than intact proteins in many cases, (ii) the requisite isolation of proteins for reagent preparation and identification can be precluded by synthesizing signature peptides identified in protein and DNA databases, and (iii) it is easier to tryptic digest all proteins in a single reaction than to isolate and digest each individually as in the 2D electrophoretic approach. A five-step protocol was used for production of signature peptides. The first step was to select a sample from a particular compartment of organelle. Simple methods, such as centrifugal fractionation of organelles, greatly enrich a sample in the components being examined. The second step embodied reduction and alkylation of all proteins in the sample. In some cases the alkylating agent can be affinity labeled to facilitate subsequent selection of cysteine containing peptides. The third step was tryptic digestion of all polypeptides in the reduced and alkylated sample. A few to more than a hundred peptides will be generated from each protein, depending on solubility and ease of digestion. Although data are not presented, it was found that trypsin will partially digest leather and by so doing generates signature peptides. This potentially offers an avenue to the analysis of insoluble proteins. The enormous complexity of the sample produced by proteolysis was reduced in a third step by using affinity chromatography methods to select peptides with unique structural features. Affinity selected peptides were then fractionated by high-resolution RPLC in a fourth step. And finally, target peptides from RPLC fractions were identified by MALDI-TOF-MS mass in the fifth step.

The analytical strategy employed in this study focused on the ability of Con A lectin columns to select glycopeptides from tryptic digests, RPLC to further fractionate the selected peptides, and MALDI-TOF-MS to identify specific peptides in RPLC fractions. Lectin columns have been widely used to purify glycopeptides, generally for the purpose of studying the oligosaccharide portion of the conjugate. When this is the object it important to fractionate as many of the glycoforms as possible, either with serial lectin columns, anion-exchange chromatography, or capillary electrophoresis. The focus of this work, in contrast, was on the peptide portion of the glycoconjugate. Any glycoform containing the signature peptide backbone is appropriate for protein identification. Con A has high affinity for N-type hybrid and high-mannose oligosaccharides, slightly lower affinity for complex di-antenary oligosaccharides, and virtually no affinity for complex N-type tri- and tetra-antenary oligosaccharides. Most of the N-type glycoproteins contain glycoforms that are recognized by Con A. Thus, a Con A column is ideal for selecting glycopeptides from digests of N-type glycoproteins.

3.2. Compartmentalization

Protein(s) of interest often residue in a particular compartment in a cell or organism. The act of first isolating the compartment within which the protein is contained can produce a very substantial simplification of the sample. One system chosen for this study was glycoproteins in bovine cellular nuclei.

Glycoproteins in the nuclei of mammalian cells are uniquely different to those found in the cytosol [17,18]. Higher animal cells reversibly O-glycosylate some nuclear proteins with a single *N*-acetyl glucos-

amine (O-GlcNAc) at a specific serine or threonine residue. It is thought that this O-GlcNAc glycosylation is associated with transcription factors and is part of a control process; thus it is necessary to have enzymes for both glycosylate and deglycosylate in the same compartment. It was the objective in this study to gain a rough idea of the number of these glycoproteins in the nuclei of bovine pancreas cells.

Subsequent to the isolation of nuclei by centrifugation, histones were selectively removed and O-glycosylated proteins isolated as a group by chromatography on a *Bandeiraea simplicifolia* (BS-II) lectin affinity column. This lectin has been found to be specific for *N*-acetyl glucosamine [19]. A silica based BS-II column was synthesized and coupled with a switching valve to a reversed-phase column. This two-dimensional chromatographic system was used to concentrate and purify glycoproteins from nuclei. Reversed-phase chromatography (Fig. 1) and 2D gel electrophoresis (data not shown) of the protein fraction eluted from the lectin column by *N*-acetyl-D-glucosamine confirm the presence of some 25–35 major components in the sample. More components may be present but below the limits of

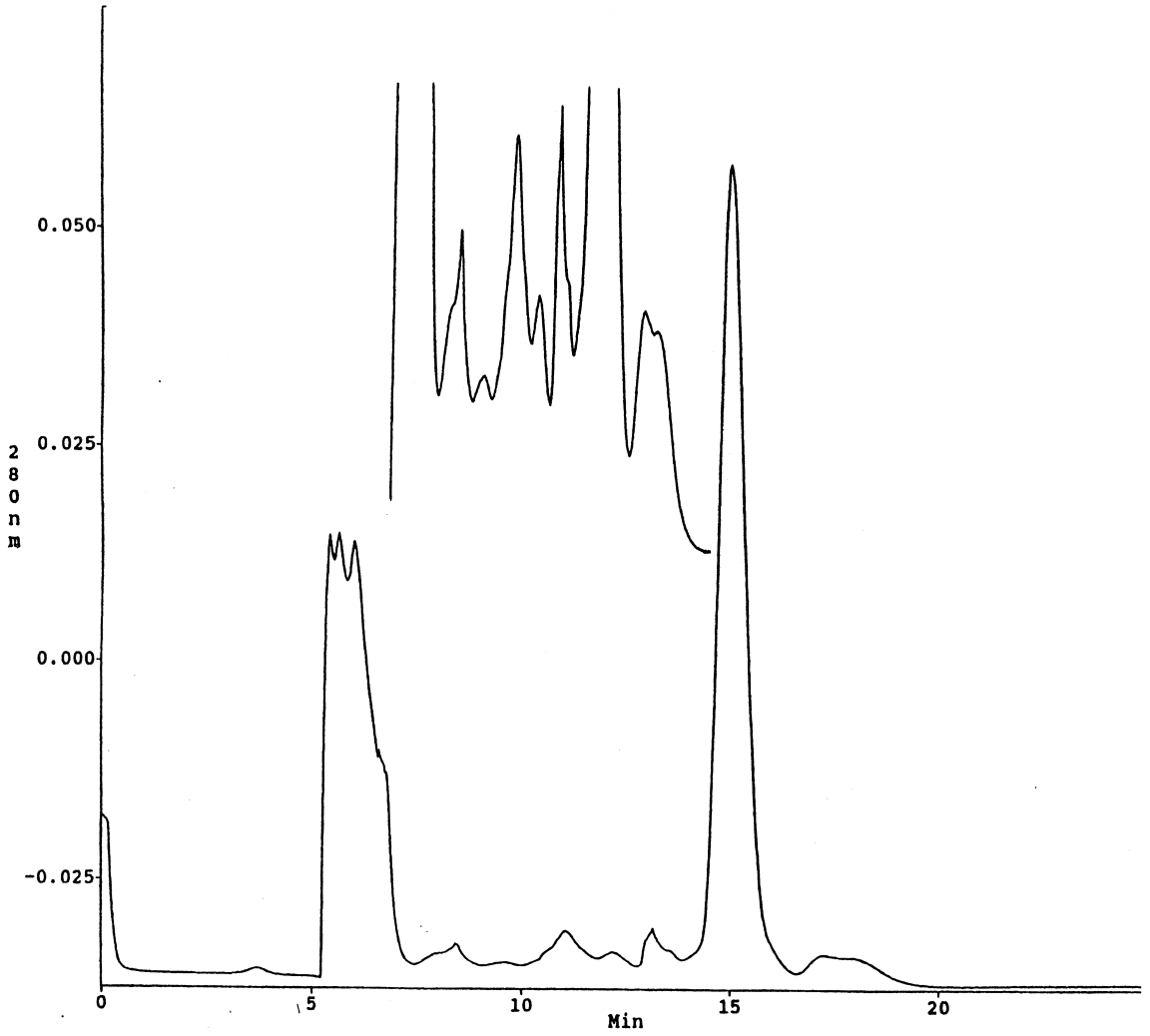

Fig. 1. Reversed-phase chromatogram of proteins isolated from bovine nuclei by chromatography on a *Bandeiraea simplicifolia* (BS-II) lectin affinity column. Elution was achieved by a 0.20 *M* solution of *N*-acetylglucosamine.

detection. Considering that some 20 000 proteins may be expressed in mammalian cells, this is much simpler than anticipated. The results of this study show that compartmentalization and affinity selection of specific proteins from a cell can greatly reduce the number of proteins in a sample.

When the protein sample used for glycoprotein analysis was reduced, alkylated with iodoacetamide, and trypsin digested before chromatography on the (BS-II) lectin affinity column, the reversed-phase chromatogram (Fig. 2) of the glycopeptides captured by the affinity column again shows unexpected simplicity. Mass spectra of selected peaks (Fig. 3) indicate a relatively low degree of complexity in fractions collected from the reversed-phase column. No attempt was made to identify these peptides by either database searches or multidimensional MS.

3.3. Signature peptide selection from serotransferrin

Serotransferrin, i.e. transferrin from serum, was

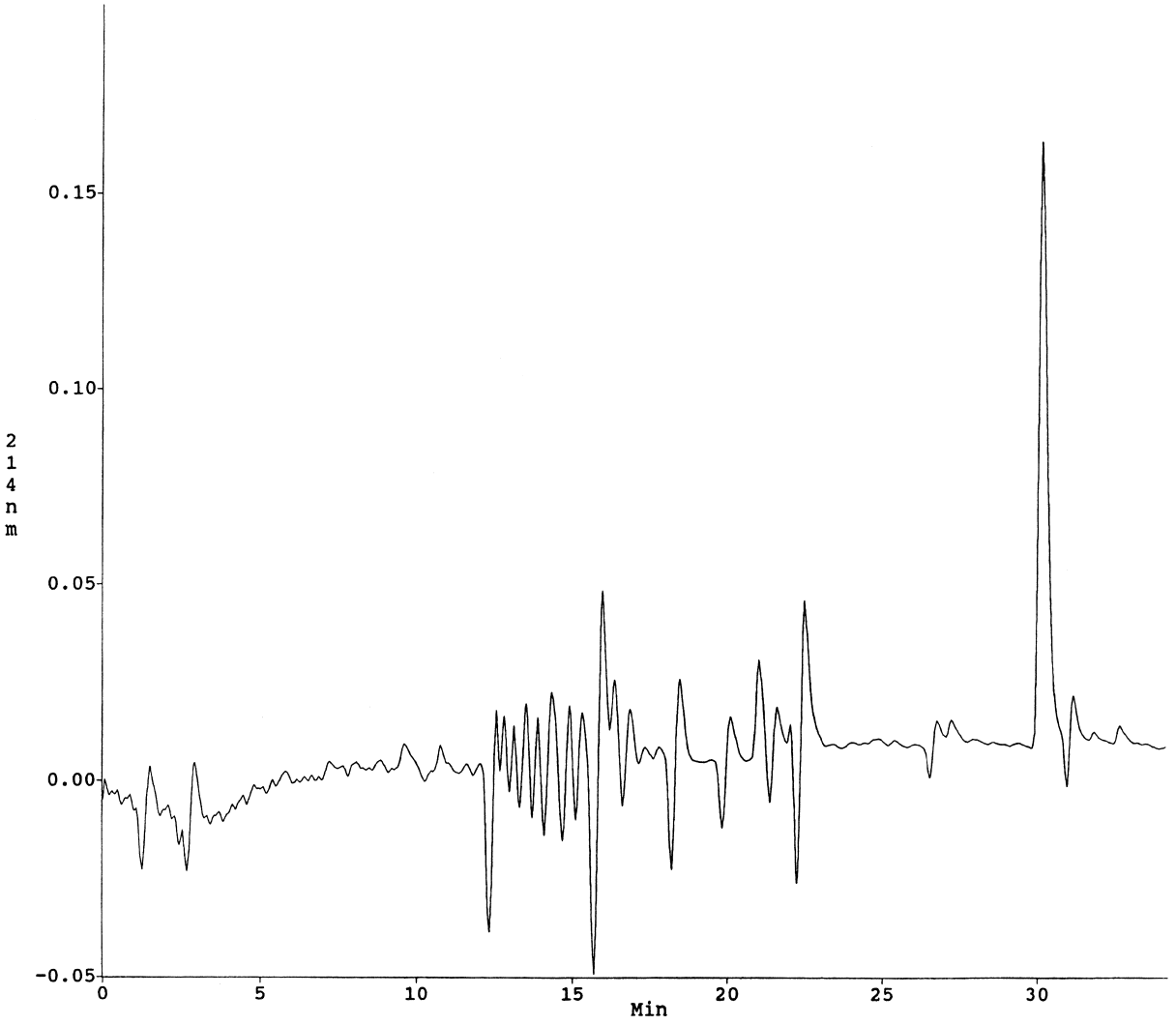

Fig. 2. Reversed-phase chromatogram of tryptic digested glycopeptides isolated from bovine nuclei by chromatography on a *Banderiaeea simplicifolia* (BS-II) lectin affinity column. Elution was achieved by a 0.20 *M* solution of *N*-acetylglucosamine.

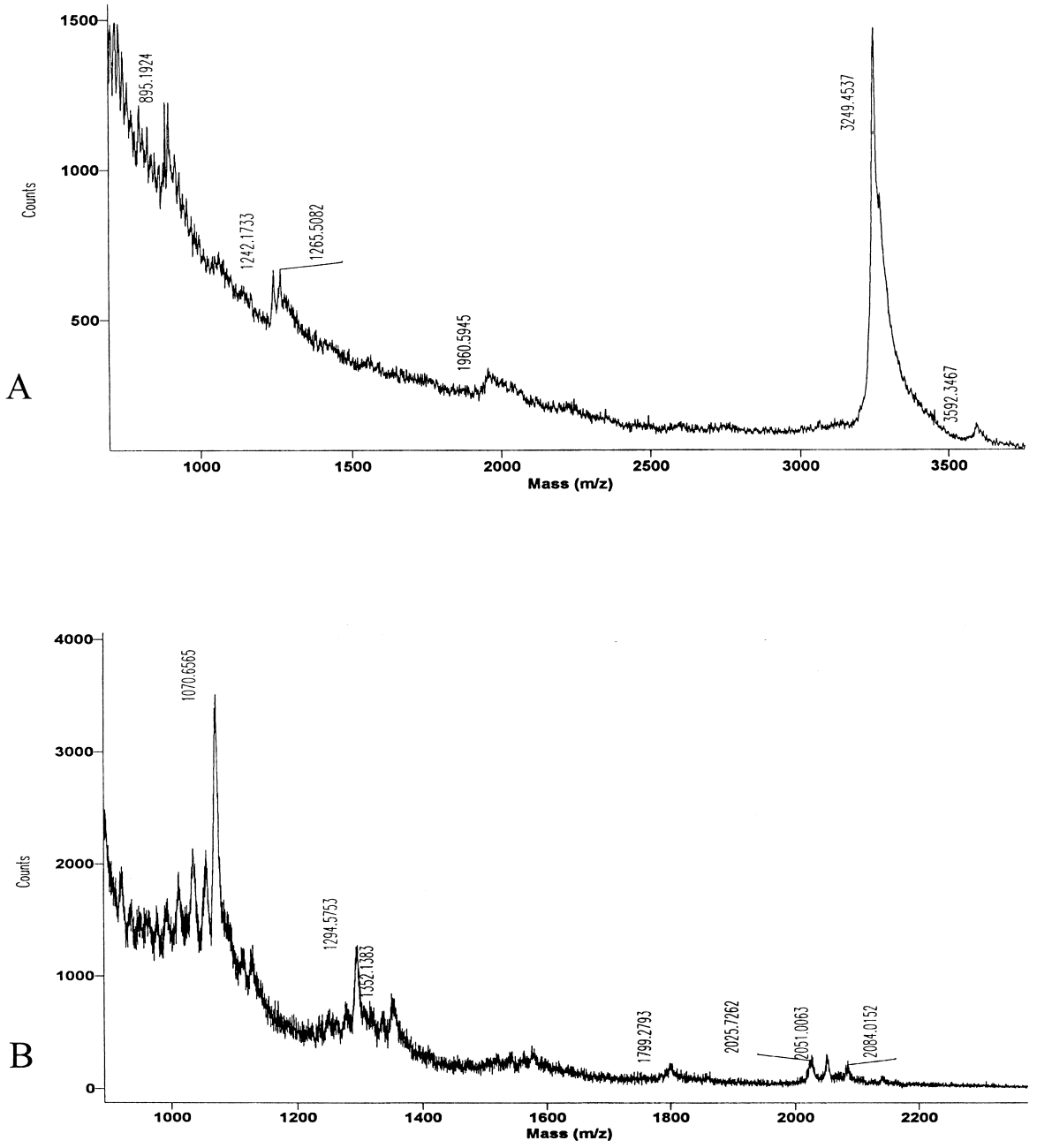

Fig. 3. (A–D) Mass spectra of glycopeptides fractions collected from the reversed-phase column.

chosen as a model protein to examine affinity selection of affinity peptides. Human serotransferrin is a glycoprotein of M_r 80 000 containing 679 amino acid residues. Potential sites for *N*-glycosylation are found in the sequence at residues Asn_{413} and Asn_{611}. The reversed-phase chromatogram of a tryptic digest (Fig. 4a) is seen to be substantially reduced in complexity when non-glycosylated peptides are first removed with a concanavalin A affinity chromatography column (Fig. 4b). The peptides glycosylated at

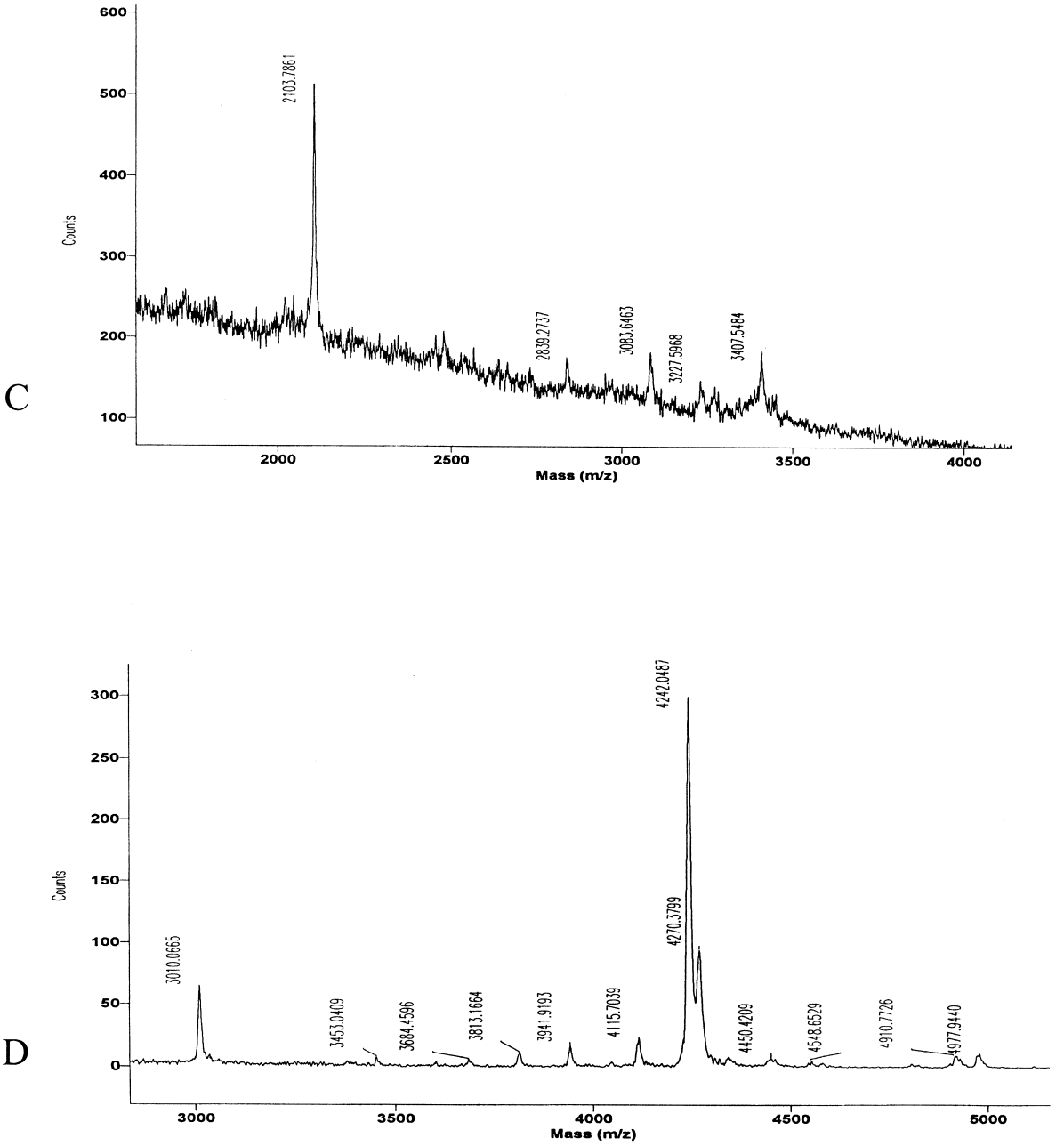

Fig. 3 (*continued*).

residues Asn_{413} and Asn_{611} eluted at 27.5 and 33.4% of solvent B, respectively. MALDI-MS of the two major components from Fig. 4b are seen in Fig. 5a and b, respectively. Although the chromatographic peaks appear to be homogeneous, MALDI-TOF-MS indicates considerable heterogeneity within the two fractions. This is as expected. It is known that there is often substantial heterogeneity in the oligosaccharide portion of a glycopeptide. The stationary phase of the reversed-phase column interacts almost

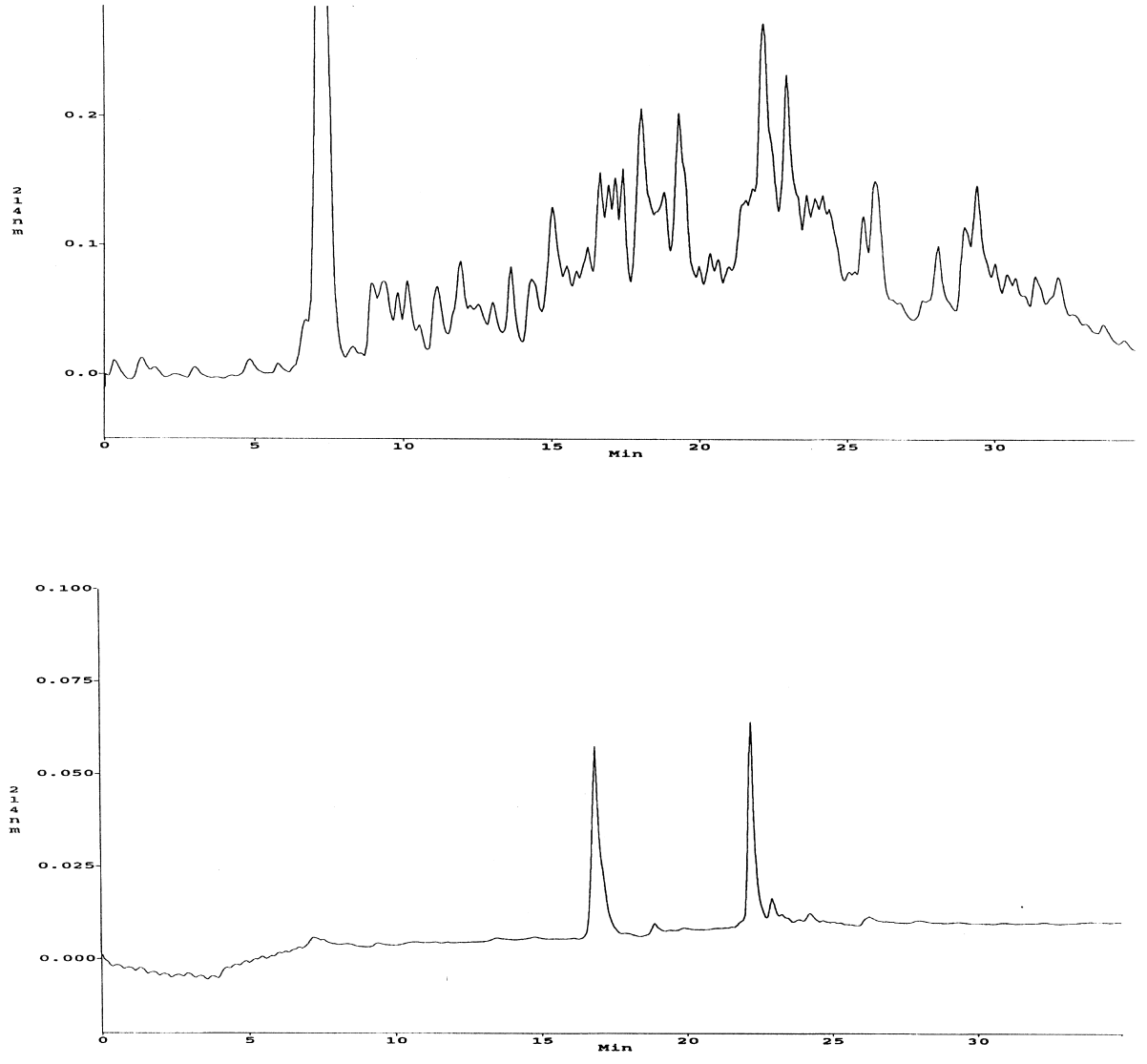

Fig. 4. Reversed-phase chromatographic result: (a) peptide map of human serotransferrin; (b) two human serotransferrin glycopeptides isolated from a Con A column.

exclusively with the peptide region of glycopeptides, essentially ignoring the oligosaccharide portion. This means that glycopeptides which are polymorphic in the oligosaccharide part of the molecule will produce a single chromatographic peak, albeit slightly broader than that of a single species. On the other hand, MALDI-TOF-MS discriminates on the basis of mass and detects all species that differ in mass without regard to structure. Used together, these two methods produce a high degree of structural selectivity.

3.4. Identification of serotransferrin signature peptides from serum

Based on the solvent composition known to elute the serotransferrin glycopeptides and their mass spectra, an experiment was undertaken to identify

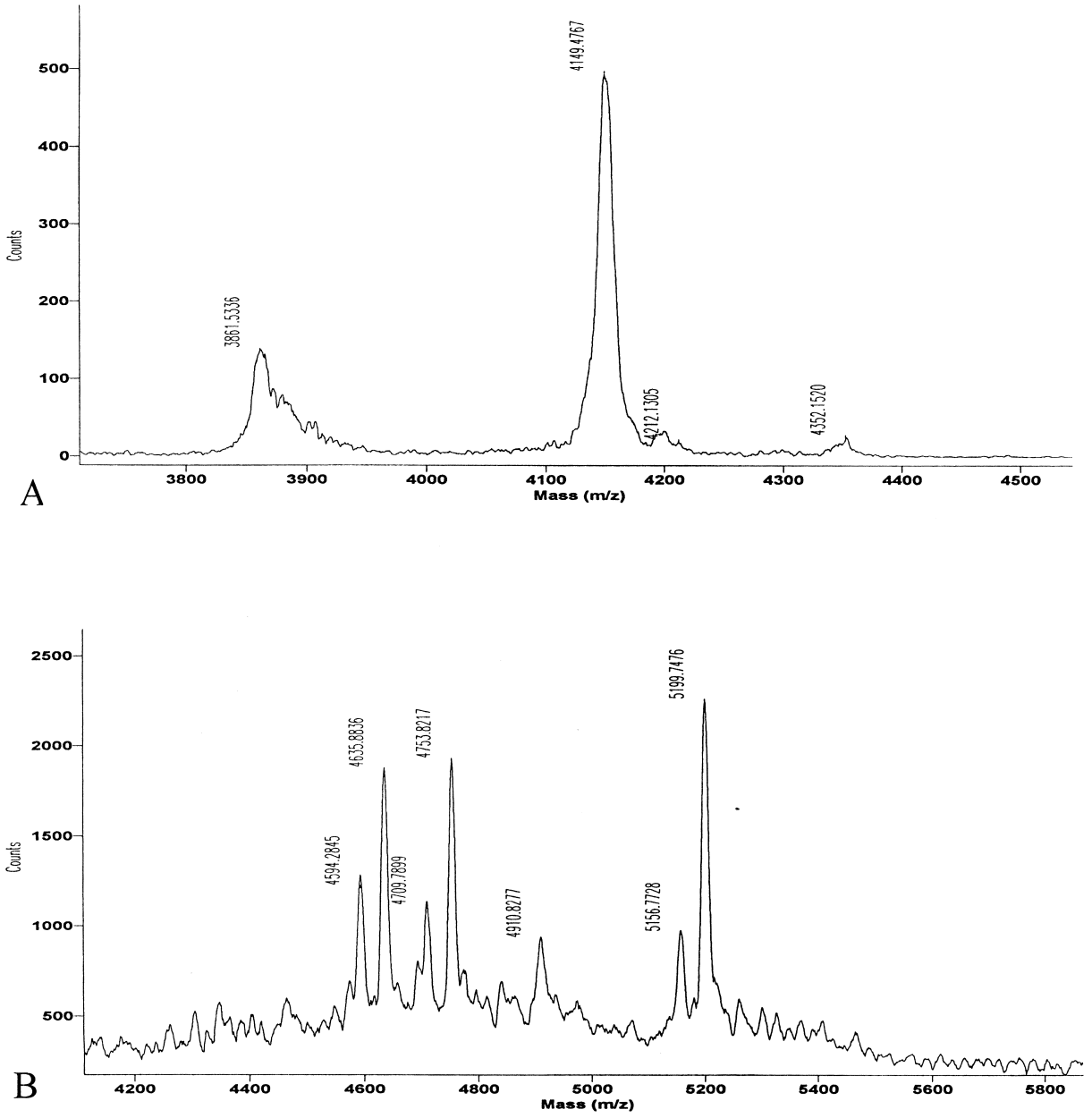

Fig. 5. MALDI-TOF mass spectrum result: (a) mass spectrum of the first glycopeptide from human serotransferrin; (b) mass spectrum of the second glycopeptide from human serotransferrin.

these signature peptides in a tryptic digest of human serum proteins. Chromatograms in Fig. 6a and b show the enormous complexity of the glycopeptide mixture selected from a tryptic digest of human serum by a Con A affinity chromatography column. Fractions eluting between 27 and 28% and between 33 and 34% were collected from the reversed-phase column and their mass spectra compared with that of human serotransferrin. Although extremely complex, mass spectra (Fig. 5a and b) obtained from fractions corresponding in chromatographic properties to the serotransferrin glycopeptides reveal the presence of

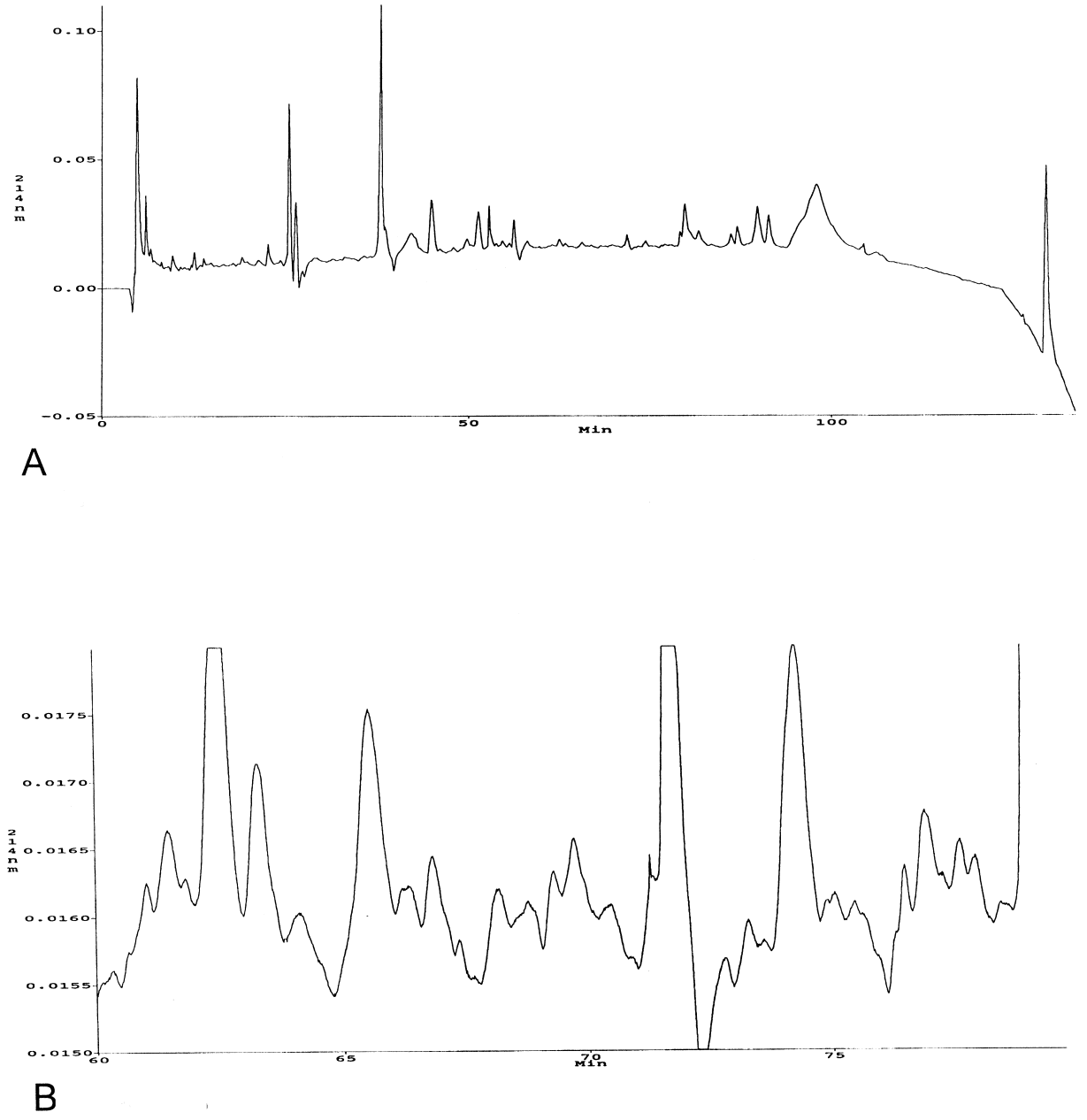

Fig. 6. (a) Reversed-phase chromatogram of glycopeptides isolated from human serum; (b) expanded reversed-phase chromatogram of glycopeptides isolated from human serum.

these signature peptides in the serum sample. Fig. 7a shows masses at 3861, 4153 and 4213 u, matching the glycopeptide peaks from Fig. 5a. Mass error was typically <4 u using external calibration. Because of the relatively lower amount of the human transferrin in an individual's serum, higher laser power was used to generate the spectra than that in pure human transferrin. Therefore, peak intensity were lower and spectral resolution were lower. In order to increase signal to noise ratio, all the spectra were smoothed

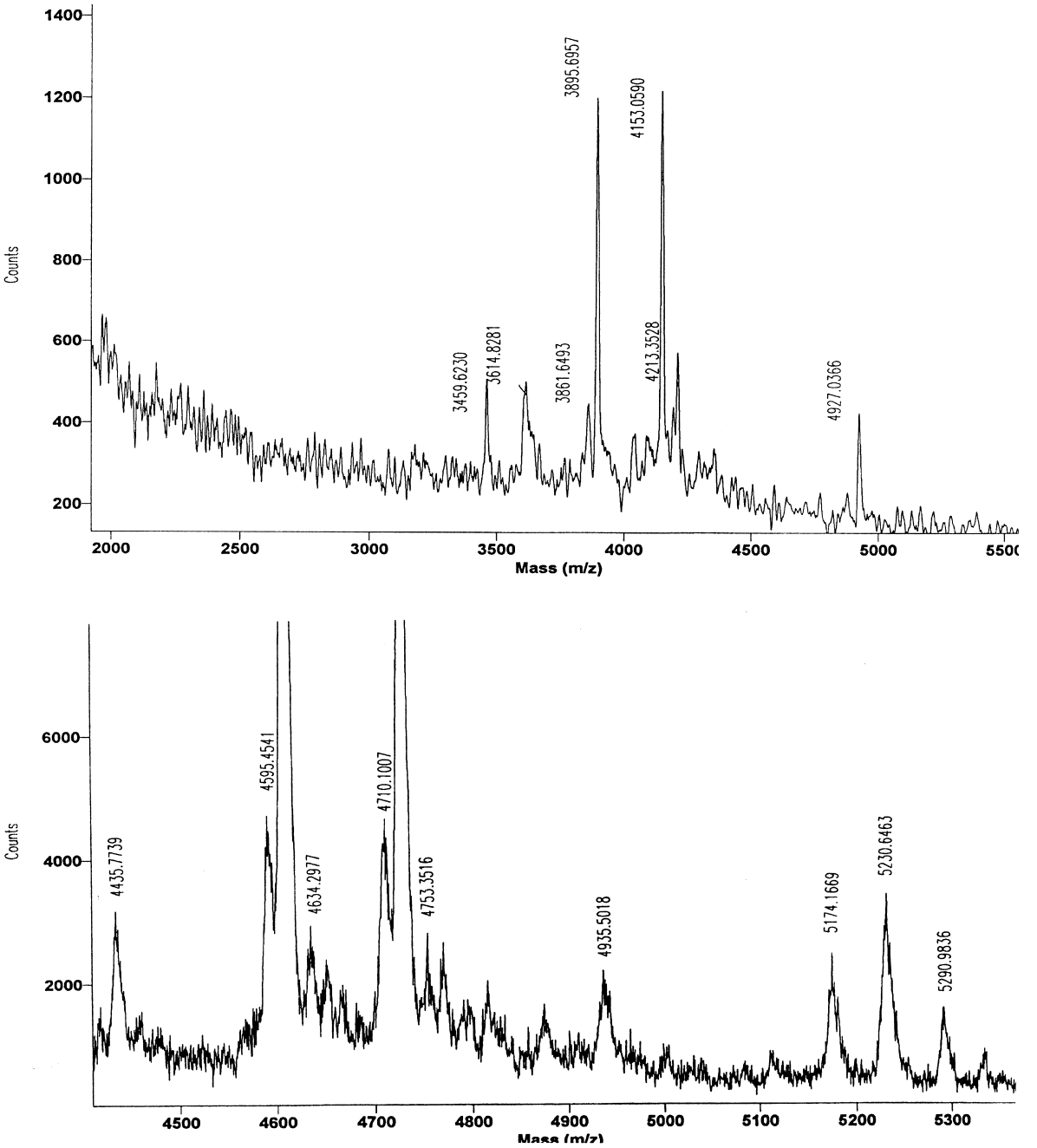

Fig. 7. Mass spectrum of glycopeptides isolated from human serum. (a) Fraction containing the first glycopeptide from human serotransferrin; (b) fraction containing the second glycopeptide from human serotransferrin.

by a 19-point averaging process. This caused the mass error to be a little higher. Glycoforms at 3459, 3614 and 3895 u were either absent or ion suppressed sufficiently so that they could not be seen. We also checked the fraction from 25 to 27% and from 29 to 31%, there was no more than one peak

matching glycopeptide peaks from Fig. 5a. It demonstrated that the matching of these peaks were not coincident. Fig. 7b shows that 4595, 4634, 4710 and 4753 matched the glycopeptides peaks from Fig. 5b. Again, fractions from 31 to 33% and 34 to 36% were checked and no matching was found. The fact that the spectra are not identical in relative intensities to the standards can be explained by possible reasons: differences in glycosylation ratio between the reference protein and that in the serum sample of an individual; inter-run variations in MALDI spectra resulting from difference in MALDI ionization.

Although not examined, other modes of selection are also potentially possible. A variety of lectins are available that allow the selection of specific types of post-translational modification on the basis of oligosaccharide structure. Antibodies would be another way to select for specific types of post-translational modification such as phosphorylation. Antibodies have also been used to select dinitrophenyl derivatized amino acids, such as tryptophan. Alkylation of cysteine with a biotinylated form of maleimide has been suggested as another way to select cysteine-containing peptides with avidin. Perhaps double selection by a combination of these affinity methods will give even higher degrees of selectivity.

3.5. Isotopically labeled internal standard quantification

One of the issues with the signature peptide approach is how to quantitate the protein being identified. Because tryptic digests of samples containing many proteins are enormously complex, the mixture generally will not be resolved into individual components by reversed-phase chromatography. Simple absorbance monitoring is precluded. This will even be true with affinity selected samples as was seen in Figs. 2 and 6. Fig. 6a and b shows that there can be so many components in reversed-phase chromatograms of affinity selected samples that quantification of any particular peptide is impossible. The next avenue to quantification would be to use peak height in the MALDI-TOF spectrum. Unfortunately, MALDI-TOF is not very quantitative. A better method is needed.

Internal standards are frequently used in quantitation. The internal standard method of quantification is based on the concept that the concentration of an analyte in a complex mixture of substances may be determined by adding a known amount of a very similar, but distinguishable substance to the solution and determining the concentration of analyte relative to a known concentration of the internal standard. Assuming that the relative molar response of the detection system for these two substances ($\Re/R$) can be determined, then $A = \Lambda[\Re/R]\Delta$. The term A is the instrument response to analyte, Λ is instrument response to the internal standard, R is specific molar response to analyte, $\Re$ is specific molar response to the internal standard, and Δ is the relative concentration of analyte to that of the internal standard. It is important that these substances are as similar as possible in chemical properties so they will behave the same way in all the steps of the analysis. In view of the fact that the last step of the analytical protocol used to identify signature peptides is MS, isotopic labeling of either the internal standard or the analyte would be the best way to produce an internal standard. Chromatographic systems are generally not able to resolve isotopic forms of an analyte whereas isotopically labeled species are easily resolved by MS. Behavioral equivalency in all stages except MS is critical. The question is how to easily create isotopically labeled internal standards of peptides in mixtures.

This may be done in two ways. One is through the synthesis of peptides in which one of the amino acids is labeled. The second is by derivatizing peptides with an isotopically labeled reagent. Although it is more lengthy, the second route was chosen because it can also be used to create internal standards of unknown structures. Work in progress will show that this is critical in proteomic studies where the object is to identify unknown proteins in regulatory flux.

Signature peptides generated by trypsin digestion have a primary amino group at their amino-terminus in all cases except those in which the peptide originated from the blocked amino-terminus of a protein. The specificity of trypsin cleavage dictates that the C-terminus of signature peptides will have either a lysine or arginine (except the C-terminal peptide from the protein) and that in rare cases there may also be a lysine or arginine adjacent to the C-terminus. Primary amino groups of peptides were acylated with *N*-acetoxysuccinamide (NAS). When

analyzed by MALDI-MS in the positive ion mode, it is seen (Fig. 8) that a peptide with five amino groups can be quantitatively derivatized with this reaction. Internal standard peptides are acetylated with trideutero-NAS. This means that peptides in samples containing both the native and deuterated internal standard species would appear in the mass spectrum as a doublet (Fig. 9a). The presence of a carboxyl group in all tryptic peptides allows them to be analyzed by MALDI-TOF-MS in the negative ion mode. It was found that the ϵ-amino group of all lysines can be derivatized in addition to the amino-terminus of the peptide, as expected. Arginine residues are not acetylated. This means that 3 u would be added for each lysine when using trideutero-NAS. The number of lysines in a peptide is revealed by the mass shift. (Multiple basic amino acids occasionally occur at the C-terminus with trypsin.) It is also possible to differentiate between peptides in which the only basic amino acid is lysine, or arginine, or a combination of the two. Peptides in which the only basic amino acid is lysine have no positive charge after acetylation. No spectra will be produced in the positive ion mode of ion acceleration unless a cationizing agent is added to the peptide. Actually, the peptide in this case picks up sodium and potassium ions from the matrix in the MALDI source, causing an increase in mass equivalent to that of sodium or potassium. Because the mass of these two ions is different, they appear in the spectrum as a

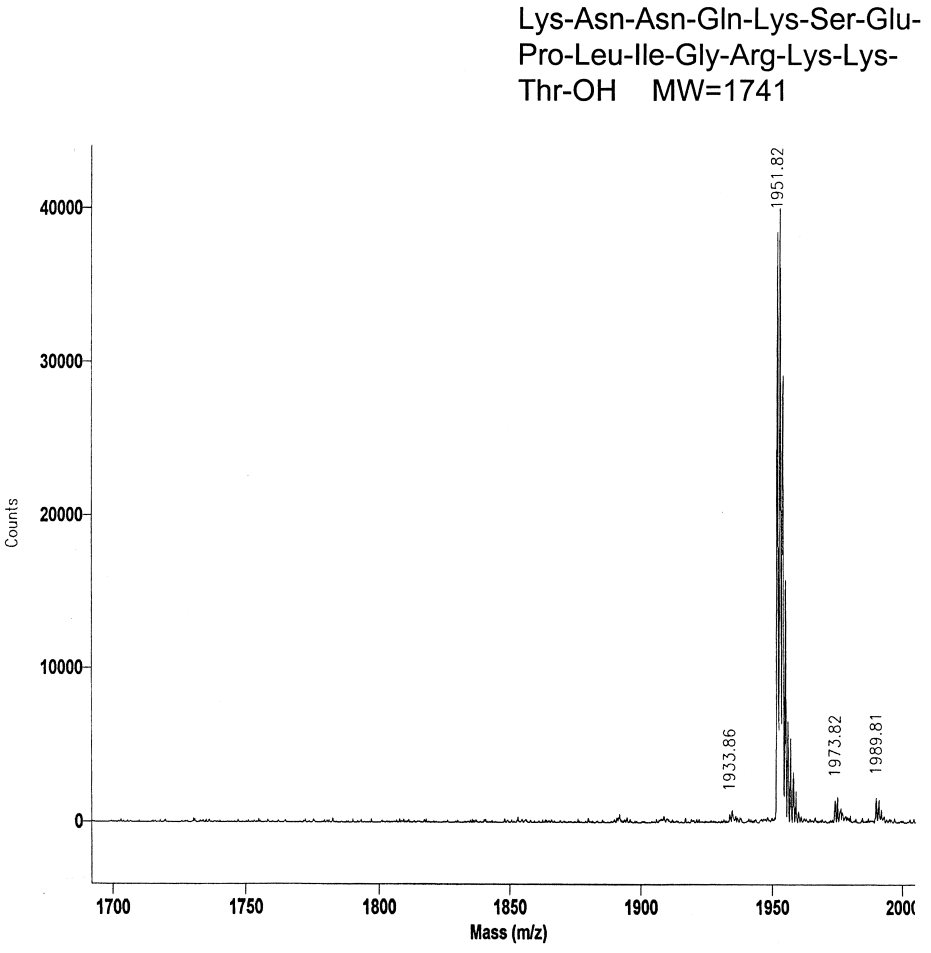

Fig. 8. MALDI-mass spectrum of a deuterium labeled peptide containing four lysines. M_r = molecular mass.

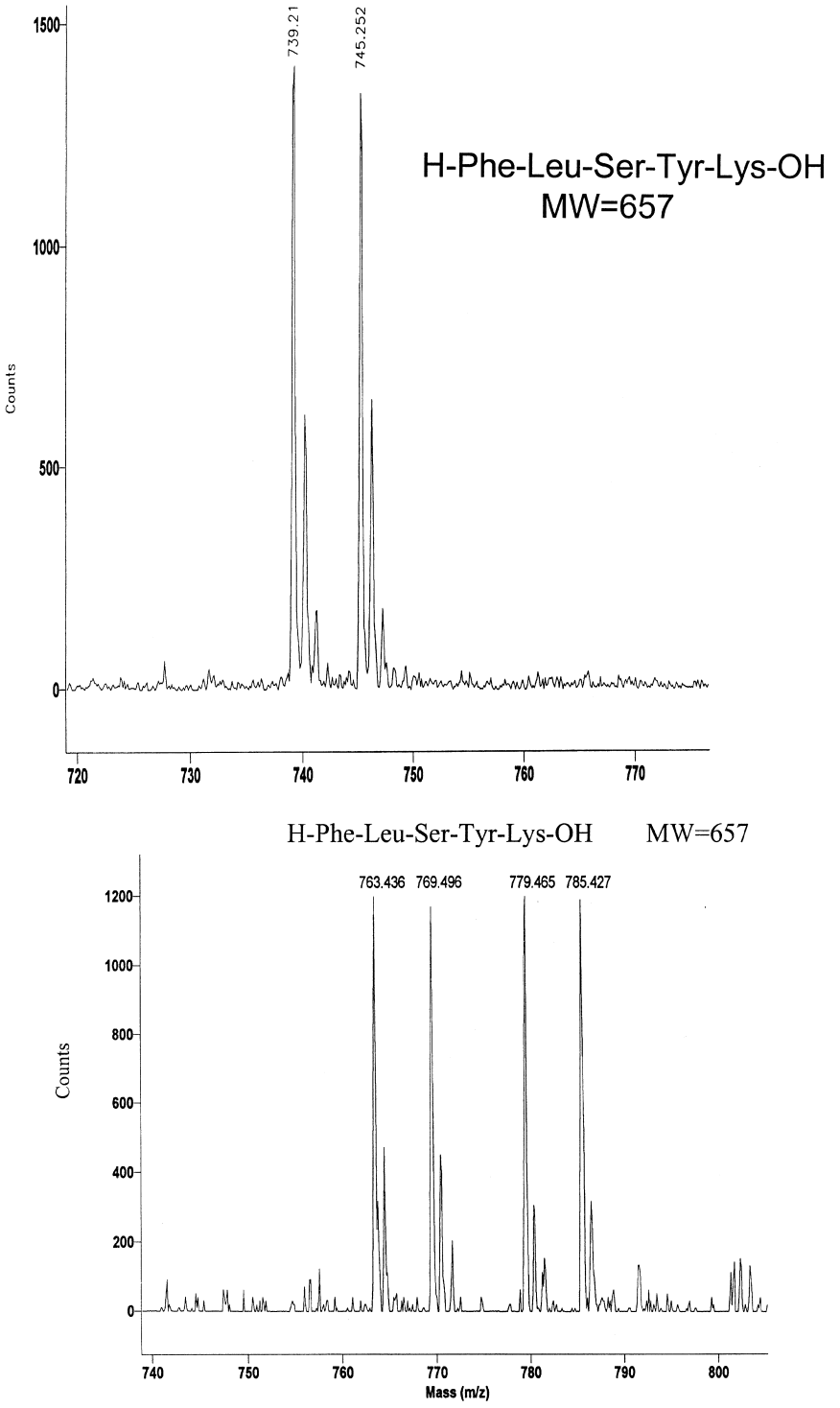

Fig. 9. (a) MALDI-TOF mass spectrum of the labeled and unlabeled peptide in negative mode detection. (b) MALDI-TOF mass spectrum of a lysine containing peptide detected in positive mode.

double. When coupled with the fact that the lysine peptide described above in Fig. 9a is also deuterated, the mass spectrum of this peptide in the positive ion mode of acceleration will show four peaks (Fig. 9b).

The mass spectrum for any peptide in a sample containing an isotopically labeled internal standard will appear as at least a doublet. The simplest case would be the one where (i) trideutero-NAS was used as the labeling agent, (ii) the C-terminus was arginine, and (iii) there were no other basic amino acids in the peptide. Spectra in this case show a doublet in which the two peaks are separated by 3 u (Fig. 10a). With one lysine the doublet peaks were separated by 6 u (Fig. 10b) and with two lysine by 9 u. For each lysine that is added the difference in mass between the experimental and control would increase an additional 3 u. Quantification of the relative amounts of both lysine and arginine containing peptides using MALDI-TOF and isotopically labeled internal standards was studied. A linear equation was deduced from the ion current intensity ratio of deuterium-labeled and unlabeled acetylated peptides versus the known ratio of the amount of these two peptides. The equation of the arginine-containing peptide (TAGPLR) was $y = 0.9509x - 0.3148$ ($R^2 = 0.9846$) while that for a lysine-con-

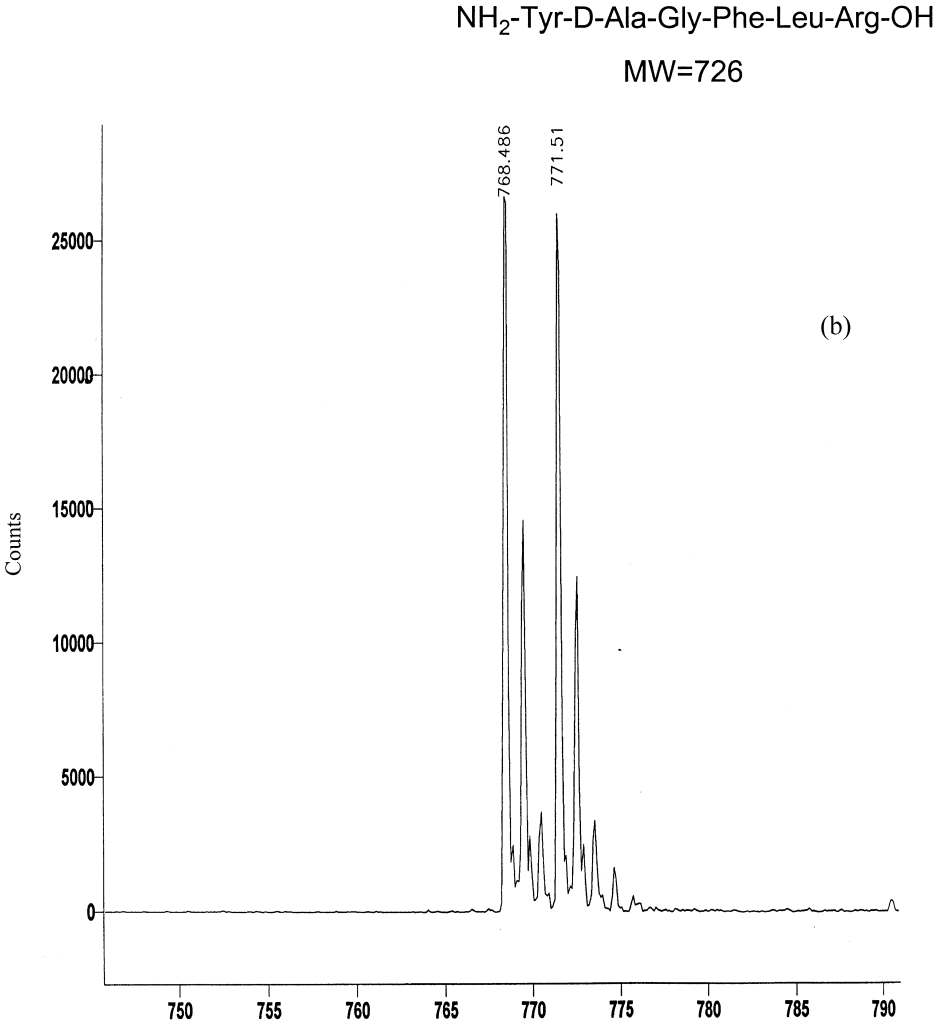

Fig. 10. (a) MALDI mass spectrum of a peptide which contains arginine; (b) MALDI mass spectrum of a peptide which contains a lysine.

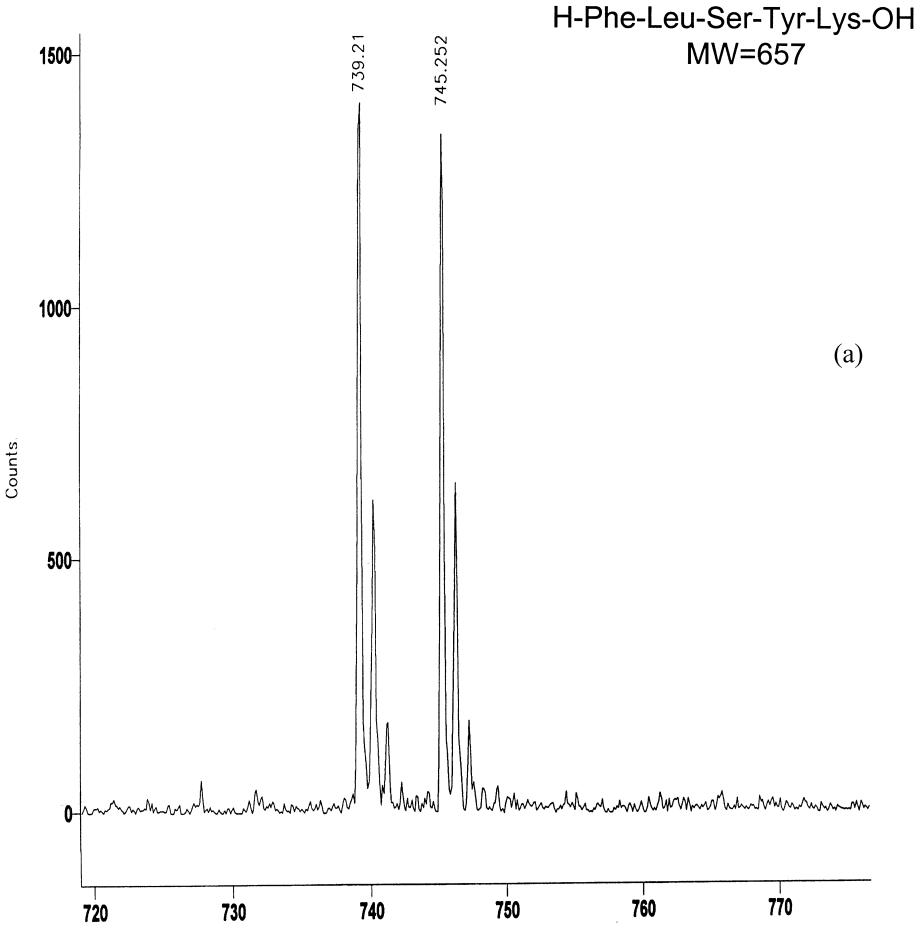

Fig. 10 (*continued*).

taining peptide (FLSYK) was $y = 0.9492x + 0.4112$ ($R^2 = 0.9937$). The term y stands for the intensity ratio of the deuterium-labeled to unlabeled acetylated peptides and x stands for the relative amount of these two peptides.

These results strongly suggest that a method in which internal standard peptides are created by isotopic labeling and ratios of native to internal standard species quantified by MS will be useful in determining the relative concentration of signature peptides.

4. Conclusions

It is concluded that signature peptides derived from tryptic digests of complex protein mixtures can be used as analytical surrogates, at least in the case of glycoproteins. Even in the case of samples with the complexity of human serum, the multidimensional analytical approach of affinity chromatography, reversed-phase chromatography and mass spectrometry has sufficient resolution to identify single signature peptide species. Because the whole protein is not needed for analysis, this strategy is particularly suited to the identification of proteins of limited solubility or that are suggested from DNA data bases but have never been isolated.

It may also be concluded that isotopically labeled internal standard analysis provides a useful method for the quantification of peptides. There is a strong possibility that when coupled with signature peptide

derived from proteins, these combined methods will provide a powerful new method for the quantification of multiple proteins in complex mixtures.

Acknowledgements

The authors gratefully acknowledge financial support from National Institute of Health grant number 25431.

References

[1] D.H. Leaback, H.K. Robinson, Biochem. Biophys. Res. Commun. 67 (1975) 248.
[2] D.R. Davis, E.A. Padlan, S. Sheriff, Annu. Rev. Biochem. 59 (1990) 439–473.
[3] B. Desbuquois, P. Cuatrecasa, Annu. Rev. Med. 24 (1973) 233.
[4] M.A. Navia, B.M. Mckeever, Proc. Natl. Acad. Sci. USA 86 (1989) 7–11.
[5] P.J. Rosenfeld, T.J.J. Kelly, Biol. Chem. 261 (1996) 1398.
[6] D.T. Fu, H.V. Halbeek, Anal. Biochem. 206 (1992) 53–63.
[7] J. Porath, I. Carlsson, I. Olsson, G. Belfrage, Nature 258 (1992) 598–599.
[8] L. Dou, J. Mazzeo, Biochromatography 5 (1990) 704.
[9] H.P. Sievert, S.-L. Wu, J. Chromatogr. 499 (1990) 221.
[10] F. Hillenkamp, M. Kaoras, R.C. Beavis, B.T. Chait, Anal. Chem. 63 (1991) 1193A.
[11] P.H. O'Farrel, J. Biol. Chem. 250 (1975) 4007–4021.
[12] L. Anderson, in: Presented at the 2nd International Conference on proteomics, Coronado, CA, 1998.
[13] P. Roepstorff, Curr. Opin. Biotech. 8 (1997) 6–13.
[14] J.R. Yates, J. Mass Spec. 33 (1998) 1–19.
[15] F. Hsieh, H. Wang, J. Mark, S.A. Martin, F.E. Regnier, Anal. Chem. 68 (1996) 455.
[16] W. Kopaciewicz, F.E. Regnier, Anal. Biochem. 129 (1983) 472.
[17] G. Hart, Annu. Rev. Biochem. 66 (1997) 315–335.
[18] R. Tjian, Cell 55 (1988) 125–133.
[19] S. Ebisu, Carbohydr. Res. 61 (1978) 129.