

STRUCTURE DETERMINATION OF GLYCOPROTEINS  
BY MASS SPECTROMETRY

by

Chenhui Zeng

Submitted to the Department of Chemistry in  
Partial Fulfillment of the requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 1996

© 1996 Massachusetts Institute of Technology  
All Rights Reserved

Signature of Author .....

Department of Chemistry  
May 1, 1996

Certified by .....

Professor Klaus Biemann  
Thesis Supervisor

Accepted by .....

Professor Dietmar Seyferth  
Chairman, Departmental Committee on Graduate Studies  
MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY

JUN 12 1996

Science

This doctoral thesis has been examined by Committee of the Department of Chemistry as follows:

Professor Daniel S. Kemp .....

.....

Chairman

Professor Klaus Biemann .....

..  
Thesis Supervisor

Professor Peter T. Lansbury .....

.....  


*To My Parents*

# STRUCTURE DETERMINATION OF GLYCOPROTEINS BY MASS SPECTROMETRY

by

Chenhui Zeng

Submitted to the Department of Chemistry on May 1, 1996  
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry.

## ABSTRACT

The relevant background information about the glycoprotein, enzymatic digestion, and the new matrix-assisted laser desorption/ionization time-of flight mass spectrometry (MALDI-TOF MS) ionization technique is reviewed. This method allows the detection of a complex mixture over a wide mass range (500-200,000 Da) using picomole to femtomole amounts of sample. Therefore, it is now possible to determine the structure of a glycoprotein using relatively little material and provide precise information about the glycosylation distribution. In this thesis, two glycoproteins were investigated by the combination of various enzyme digestions and MALDI-TOF MS.

The extent of *N*-glycosylation of yeast external invertase at each of the 14 potential sites was determined by MALDI MS. The size of the oligosaccharide, the degree of glycosylation and the distribution of the oligosaccharides at each individual potential glycosylation site were characterized. This information goes far beyond previously published data and sometimes differs from that. During this study, the amino acid sequence originally derived from the DNA sequence of the gene coding for invertase was also verified and it was found that this protein when expressed from *SUC2* gene might be created as more than one sequence which differ by a few amino acid substitutions ( $A^{58} \leftrightarrow T$ ,  $N^{65} \rightarrow H$ , and  $V^{412} \leftrightarrow A$ ).

In addition, an unknown glycoprotein that is secreted by the human axilla, originally named apocrine secretion odor binding protein (ASOB2) was studied. This protein has a blocked N-terminus and thus is not suitable for Edman sequencing. The partial sequences of some peptides generated by Lys-C digestion were deduced from MALDI-post-source decay spectra. By individually matching the two most reliable sequences with the protein database, it was revealed that ASOB2 is a known glycoprotein, human apolipoprotein D (apo D). Detailed mass spectral data confirmed that the amino acid sequences are completely identical. However, the type of glycosylation of axillary apo D was found to be different from that reported recently for human plasma apo D. It was also shown that the odor producing component, 3-methyl-2-hexenoic acid, carried by this protein is not covalently bound.

Thesis supervisor: Klaus Biemann  
Title: Professor of Chemistry

## Table of Contents

	Page
List of Figures .....	7
List of Tables .....	12
List of Abbreviations Used .....	13
I. Introduction .....	15
A. Background of Glycoprotein Studies .....	15
B. Methodology of Glycoprotein Determination .....	20
1. Identification of the glycopeptides in the chromatogram .....	22
2. Assignment of attachment sites and oligosaccharide populations .....	24
3. Determination of the degree of occupancy at each glycosylation site .....	26
4. Sequencing of oligosaccharides .....	27
5. Confirm the sequence of the deglycosylated peptide .....	30
C. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight mass spectrometry ..	37
1. Principles of matrix-assisted laser desorption/ionization .....	39
2. Time-of-Flight Mass Spectrometer .....	44
D. Post-source Decay Mass Spectra .....	50
II. Determination of <i>N</i> -linked glycosylation in Yeast external invertase .....	53
A. Introduction .....	53
B. Results and Discussion .....	56
C. Analysis of an Endo H Contaminated Sample .....	147
D. Conclusion .....	150
E. Materials and Methods .....	167
III. Identification of a Human Axillary Odor Carrying Protein .....	173
A. Background .....	173
B. Materials and Methods .....	176
C. Results and Discussion .....	179
D. Conclusion .....	209
IV. References .....	212
V. Appendices .....	226

VI. Biographical Note .....	Page 228
VII. Publications, Abstracts and presentations .....	229
VIII. Acknowledgements .....	231

## List of Figures

	Page
Chapter I.	
Figure I.1. Determination of a complex type carbohydrate residue linkage by sequential exoglycosidase digestion .....	28
Figure I.2. Schematic diagram of the JEOL HX110/HX110 four-sector mass spectrometer .....	34
Figure I.3. Nomenclature and definition of the fragment ions formed by CID from protonated peptides .....	36
Figure I.4. The general processes of matrix-assisted laser desorption/ionization .....	43
Figure I.5. Schematics of (a) VESTEC VT2000, and (b) PerSeptive Voyager-Elite MALDI time-of-flight mass spectrometer .....	45
Chapter II.	
Figure II.1. MALDI-TOF mass spectrum of intact external invertase .....	57
Figure II.2. HPLC trace of the trypsin digest of intact external invertase .....	59
Figure II.3. HPLC trace of the Lys-C digest of intact external invertase .....	60
Figure II.4. HPLC trace of the Asp-N digest of intact external invertase .....	61
Figure II.5. MALDI spectra of HPLC fraction T6 of the trypsin digest of external invertase before and after PNGase F deglycosylation .....	64
Figure II.6. MALDI spectra of HPLC fraction T7 of the trypsin digest of external invertase before and after PNGase F deglycosylation .....	67
Figure II.7. MALDI spectra of HPLC fraction T8 of the trypsin digest of external invertase before and after deglycosylation with PNGase F and Endo H .....	70
Figure II.8. MALDI-PSD spectrum of peptide $(M+H)^+ = m/z$ 1113.3 .....	74
Figure II.9. MALDI spectrum of the 1:1 mixture of fractions T7 + T8, and the spectrum obtained after PNGase F deglycosylation .....	77

Figure II.10.	MALDI spectra of HPLC fraction T9 of the trypsin digest of external invertase before and after deglycosylation with PNGase F and Endo H . . . . .	79
Figure II.11.	HPLC trace of the chymotryptic subdigest of 1:1 mixture of fraction T8 an T9 . . . . .	83
Figure II.12.	MALDI spectra of fractions CT3, CT4, and the 1:1 mixture of CT3 and CT4 of the chymotryptic subdigest . . . . .	85
Figure II.13.	MALDI-PSD spectrum of peptide (M+H) <sup>+</sup> = <i>m/z</i> 767.2 . . . . .	86
Figure II.14.	MALDI spectra of fraction CT6 of the chymotryptic subdigest before and after PNGase F deglycosylation . . . . .	88
Figure II.15.	MALDI-PSD spectrum of peptide (M+H) <sup>+</sup> = <i>m/z</i> 978.1 . . . . .	89
Figure II.16.	MALDI spectra of fraction CT9 of the chymotryptic subdigest before and after PNGase F deglycosylation . . . . .	90
Figure II.17.	MALDI spectra of fraction CT10 of the chymotryptic subdigest before and after PNGase F deglycosylation . . . . .	91
Figure II.18.	MALDI-PSD spectrum of peptide (M+H) <sup>+</sup> = <i>m/z</i> 2438.3 . . . . .	92
Figure II.19.	MALDI spectra of fraction CT11 of the chymotryptic subdigest . . . . .	94
Figure II.20.	MALDI spectrum of fraction CT11 of the chymotryptic subdigest after PNGase F deglycosylation, and MALDI-PSD spectrum of peptide (M+H) <sup>+</sup> = <i>m/z</i> 1626 and (M+Na) <sup>+</sup> = 1648 . . . . .	95
Figure II.21.	MALDI spectra of HPLC fraction T10 of the trypsin digest of external invertase before and after PNGase F deglycosylation . . . . .	98
Figure II.22.	MALDI spectra of HPLC fraction T15 . . . . .	100
Figure II.23.	MALDI spectra of T15 after PNGase F and Endo H deglycosylation . . . . .	101
Figure II.24.	MALDI spectra of HPLC fraction T16, T17, and T18 . . . . .	104
Figure II.25.	MALDI spectra of T16, T17, and T18 after PNGase F and Endo H deglycosylation . . . . .	106



Figure II.26.	MALDI spectrum of HPLC fraction L6 of the Lys-C digest of external invertase .....	109
Figure II.27.	MALDI spectrum of HPLC fraction L7 of the Lys-C digest of external invertase .....	111
Figure II.28.	MALDI spectrum of HPLC fraction L8 of the Lys-C digest of external invertase .....	112
Figure II.29.	MALDI spectra of HPLC fraction L12 of the Lys-C digest of external invertase before and after deglycosylation with PNGase F and Endo H .....	113
Figure II.30.	MALDI spectra of HPLC fraction L13 of the Lys-C digest of external invertase before and after deglycosylation with PNGase F and Endo H .....	115
Figure II.31.	MALDI spectrum of HPLC fraction L14 of the Lys-C digest of external invertase .....	116
Figure II.32.	MALDI spectra of mixture of L12, L13, and L14 before and after Endo H deglycosylation .....	118
Figure II.33.	MALDI spectra of HPLC fraction L16 of the Lys-C digest of external invertase before and after deglycosylation with PNGase F and Endo H .....	121
Figure II.34.	HPLC trace of the Glu-C subdigest of fraction L16 from the Lys-C digest .....	123
Figure II.35.	MALDI spectra of fraction G4 of the Glu-C subdigest of L16 before and after deglycosylation with PNGase F and Endo H .....	124
Figure II.36.	MALDI spectra of fraction G5 of the Glu-C subdigest of L16 before and after deglycosylation with PNGase F and Endo H .....	125
Figure II.37.	MALDI spectra of fraction G7 of the Glu-C subdigest of L16 before and after deglycosylation with PNGase F and Endo H .....	127
Figure II.38.	MALDI spectra of fraction G11 of the Glu-C subdigest of L16 before and after deglycosylation with PNGase F and Endo H .....	128
Figure II.39.	MALDI spectra of fraction L18 of the Lys-C digest of external invertase before and after Endo H deglycosylation .....	130

Figure II.40.	MALDI spectrum of HPLC fraction L19 of the Lys-C digest of external invertase and the HPLC trace of the chymotryptic subdigest of L19 . . . . .	132
Figure II.41.	MALDI spectrum of fraction 18 of the chymotryptic subdigest and the PSD spectra of peptide (M+H) <sup>+</sup> = <i>m/z</i> 1239.7 and 1269.7 . . . . .	133
Figure II.42.	HPLC trace of the Arg-C subdigest of fraction L19 from the Lys-C digest . . . . .	136
Figure II.43.	MALDI spectrum of fraction R6 of the Arg-C subdigest and the PSD spectra of peptide (M+H) <sup>+</sup> = <i>m/z</i> 1565.7 and 1593.8 . . . . .	137
Figure II.44.	MALDI spectra of fraction A4 of the Asp-N digest before and after PNGase F deglycosylation . . . . .	139
Figure II.45.	MALDI-PSD spectrum of the PNGase F peptide (M+H) <sup>+</sup> = <i>m/z</i> 818.8 . . . . .	140
Figure II.46.	MALDI-DE spectra of fraction A5 and fractions A4 and A5 after PNGase F deglycosylation . . . . .	142
Figure II.47.	MALDI spectra of fraction A11 of the Asp-N digest before and after PNGase F deglycosylation . . . . .	143
Figure II.48.	MALDI spectra of fraction A12 of the Asp-N digest before and after Endo H deglycosylation . . . . .	144
Figure II.49.	MALDI spectra of fraction A18 of the Asp-N digest before and after Endo H deglycosylation . . . . .	146
Figure II.50.	MALDI spectra of the Endo H contaminated external invertase and the same sample analyzed three years later . . . . .	149
Figure II.51.	Amino acid sequence of invertase . . . . .	160

### Chapter III.

Figure III.1.	MALDI spectra of intact ASOB2 and after PNGase F deglycosylation . . . . .	181
Figure III.2.	HPLC trace of the Lys-C digest of ASOB2 . . . . .	182
Figure III.3.	(a) PSD spectrum of peptide (M+H) <sup>+</sup> = 1671.9 . . . . .	183
	(b) PSD spectrum of peptide (M+H) <sup>+</sup> = 1306.6 . . . . .	184

Figure III.4.	Amino acid sequence of apoD .....	187
Figure III.5.	PSD spectrum of peptide (M+H) <sup>+</sup> = 783.5 .....	188
Figure III.6.	PSD spectrum of peptide (M+H) <sup>+</sup> = 1230.8 .....	189
Figure III.7.	PSD spectrum of peptide (M+H) <sup>+</sup> = 1357.7 .....	190
Figure III.8.	PSD spectrum of peptide (M+H) <sup>+</sup> = 1601.7 .....	191
Figure III.9.	PSD spectrum of peptide (M+H) <sup>+</sup> = 1676.9 .....	193
Figure III.10.	MALDI spectra of fraction 12 of the Lys-C digest before and after sequential deglycosylation with neuraminidase, galactosidase, and PNGase F .....	194
Figure III.11.	PSD spectrum of peptide (M+H) <sup>+</sup> = 1035.1 .....	199
Figure III.12.	MALDI spectra of fraction 13 of the Lys-C digest before and after PNGase F deglycosylation .....	200
Figure III.13.	PSD spectrum of peptide (M+H) <sup>+</sup> = 872.0 .....	201
Figure III.14.	HPLC trace of the Glu-C digest of ASOB2 .....	203
Figure III.15.	MALDI spectrum of fraction 4 of the Glu-C digest and PSD spectrum of peptide (M+H) <sup>+</sup> = <i>m/z</i> 1724.4 .....	204
Figure III.16.	MALDI spectra of fraction 4 of the Glu-C digest before and after DTT treatment .....	205
Figure III.17.	MALDI spectra of refraction 5 of fraction 12 (from the Lys-C digest) before and after DTT treatment .....	207

## List of Tables

	Page
Chapter I.	
Table I.1.	Comparison of the "soft" ionization techniques for biomolecules . . . . . 38
Table I.2.	Summery of the laser source adopted in MALDI mass spectrometry . . . . . 40
Table I.3.	Summery of the commonly used matrices in MALDI mass spectrometry . . . 41
Chapter II.	
Table II.1.	Observed sequons in the HPLC fractions of the tryptic digest . . . . . 151
Table II.2.	Observed sequons in the HPLC fractions of the Lys-C digest . . . . . 152
Table II.3.	Observed sequons in the HPLC fractions of the Asp-N digest . . . . . 153
Table II.4.	Summary of Results from MALDI MS compared to the published data . . . . 154
Table II.5.	Peptides found in tryptic digest which do not contain a sequon . . . . . 157
Table II.6.	Peptides found in Lys-C digest which do not contain a sequon . . . . . 158
Table II.5.	Peptides found in Asp-N digest which do not contain a sequon . . . . . 159
Chapter III.	
Table III.1.	Summary of the characterization of the carbohydrate moieties at Asn <sup>45</sup> and Asn <sup>78</sup> . . . . . 197

### List of Abbreviations frequently used

ACN	acetonitrile
ASOB1	apocrine secretion odor binding protein 1
ASOB2	apocrine secretion odor binding protein 2
CID	collision induced dissociation
$\alpha$ -CHCA	$\alpha$ -cyano-4-hydroxycinnamic acid
Da	dalton (mass unit)
DE	delayed extraction
Dol-P-P-	dolichol pyrophosphate
DHB	2,5-dihydroxybenzoic acid
Endo H	endo- $\beta$ - <i>N</i> -acetylglucosaminidase
ER	endoplasmic reticulum
ESI	electrospray ionization
eV	electron volt
FAB	fast atom bombardment
Fuc	fucose
Gal	galactose
GalNAc	<i>N</i> -acetylgalactosamine
GC/MS	gas chromatography combined with mass spectrometry
Glc	glucose
GlcNAc	<i>N</i> -acetylglucosamine
Glu-C	endoprotease Glu-C
h	hour
HPAE-PAD	high-pH anion exchange (HPAE) chromatography in conjunction with pulsed amperometric detection
HPLC	rp-HPLC, reverse phase high performance liquid chromatograph(y)
Hz	hertz( $s^{-1}$ )
Lys-C	endoprotease Lys-C

LSIMS	liquid secondary ion mass spectrometry (same process as FAB)
MALDI	matrix assisted laser desorption/ionization
Man	mannose
MW	molecular weight
MS	mass spectrometry
MS/MS	mass spectrometry/mass spectrometry or tandem mass spectrometry
<i>m/z</i>	mass-to-charge ratio
3M2H	3-methyl-2-hexenoic acid
NeuAc	<i>N</i> -acetyl neuraminic acid or sialic acid
Hex	Hexose
HexNAc	<i>N</i> -acetylhexosamine
PBS	phosphate-buffered saline
PD	plasma desorption
PIC	Phenylisocyanate
PITC	Phenylisothiocyanate
PNGase F	peptide: <i>N</i> -glycosidase F
PSD	post-source decay
SA	Sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid)
TFA	trifluoroacetic acid
TOF	time-of-flight
Tris-HCl	Tris(hydroxymethyl)aminomethane-HCl
V	volt

## I. Introduction

### A. Background of Glycoprotein Studies

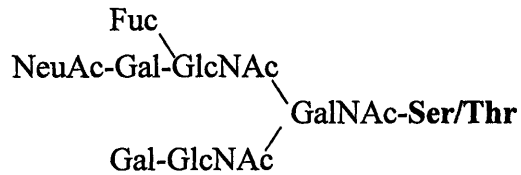
The biological significance of the carbohydrate moieties of glycoproteins has only recently become recognized. To understand the biological functions of oligosaccharide units attached to a protein, glycosylation analysis of a glycoprotein may provide important clues to the role of the oligosaccharides and their biosynthetic pathway. It is well known that oligosaccharides are commonly linked to a protein by *O*-linked or *N*-linked glycosylation. In the former case the oligosaccharide is bound to the hydroxyl group of serine or threonine, while *N*-linked oligosaccharides are attached to the amide nitrogen of an asparagine in a specific environment, Asn-Xxx-Ser/Thr (where Xxx can be any amino acid except proline). This amino acid triplet is often referred to as a "sequon". These potential glycosylation sites can be recognized in a protein sequence, but it is not possible to predict whether a particular glycosylation site is actually occupied, nor can one predict the composition and structure of the oligosaccharides. This information can only be obtained experimentally.

Generally, the structures of *O*-linked and *N*-linked oligosaccharides are different. The common *O*-linked oligosaccharides are usually short and more variable than *N*-linked structures.

Fuc-Ser/Thr

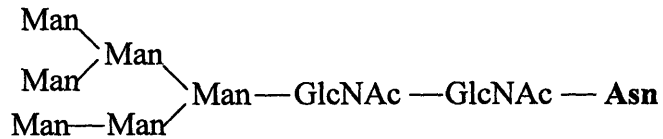
GlcNAc-Ser/Thr

Gal-GalNAc-Ser/Thr

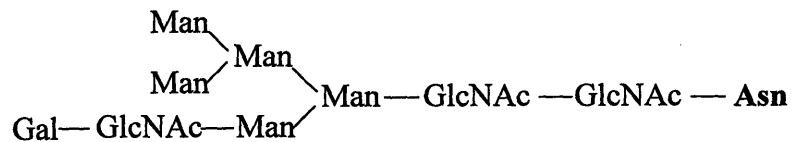


In contrast, the *N*-linked oligosaccharides always contain a common pentasaccharide core region consisting of two *N*-acetylglucosamine residues and three mannose residues, and the extended sugar chains are divided into three major categories

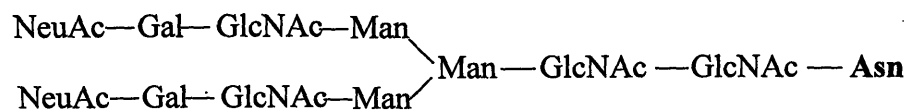
High mannose:



Hybrid:



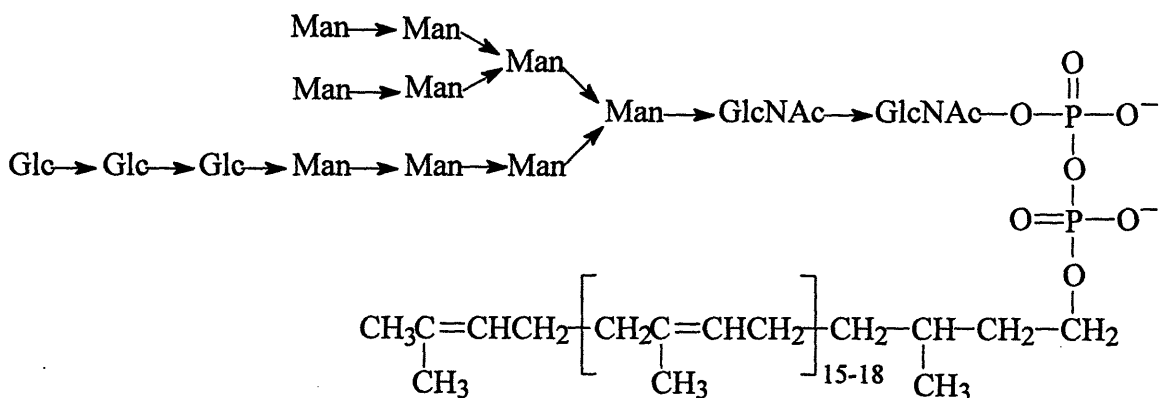
Complex:





The complexity and diversity of the carbohydrate units of glycoproteins indicate that they must be functionally important, because nature will not construct complex patterns if simple forms are sufficient. Even though some clues concerning the biological roles of oligosaccharide units are emerging, a definitive explanation of all this intricacy and diversity is not yet possible. An excellent review (Ajit Varki, 1993) with 1049 references to the study of the biological roles of oligosaccharides moieties in glycoproteins demonstrates the contributions of carbohydrate residues to protein folding, protein stabilization, as well as to their synthesis and positioning within a cell. This thesis adds to the study of glycoproteins by using mass spectrometry, a technique that has recently become important in the study of biomolecules.

Along with the investigation of their biological functions, the study of glycoprotein biosynthesis is equally important. It is believed that an *N*-linked glycoprotein is synthesized beginning with the assembling an oligosaccharide precursor, Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-P-P-Dol.



Structure of dolichol pyrophosphoryl oligosaccharide (Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-P-P-Dol)

Its carbohydrate portion is then transferred to the nascent protein followed by trimming the three outer glucose and one mannose in the rough endoplasmic reticulum (ER). The initial glycosylation product,  $\text{Man}_8\text{GlcNAc}_2\text{-Asn}(\text{protein})$ , is then transported to Golgi vesicles where the oligosaccharides elongation takes place (Kukuruzinska et al., 1987). In animal cells, glycosylation process can either yield the small size (usually less than 10 mannose) of high mannose type oligosaccharides or produce the complex type oligosaccharide species with a common  $\text{Man}_3\text{GlcNAc}_2$  core (Hubbard and Ivatt, 1981). Unlike mammalian cells, yeasts only synthesize high mannose oligosaccharides with two distinct sizes: a  $\text{Man}_{8-15}\text{GlcNAc}_2$  class and a class of very large oligosaccharides (Nakajima and Ballou, 1975; Chen et al., 1982; Trimble and Maley, 1977; Lehle et al., 1979; Ziegler et al., 1988). Byrd et al. (1982) reported that the smallest oligosaccharide chain,  $\text{Man}_8\text{GlcNAc}$ , is an essential biosynthetic intermediate. This result has been widely accepted as a biosynthetic pathway for yeast glycoproteins (Kukuruzinska et al, 1987; Ziegler and Trimble 1991). Two new hypotheses for the yeast glycoproteins biosynthesis process are proposed in the conclusion section of Chapter II.

A complete structure determination of glycoprotein has to deal with both the protein and the carbohydrate aspects. These two different structural problems in addition to inherent difficulties of carbohydrate analysis (multiple linkage sites; two linkage types,  $\alpha$  vs.  $\beta$ ; branching; and diversity of sugar derivatives, e.g., phosphates, sulfates, esters, amides) complicates the structure determination of glycoproteins which is thus more difficult than that of other large biomolecules. In general, the structure determination of a glycoprotein involves first the protein sequencing and then the glycosylation analysis at several different levels of detail: a) the occupancy of each

potential site and the type of glycosylation, e.g. *O*-linked or *N*-linked, and if *N*-linked further detail (high mannose, complex, or hybrid oligosaccharide); b) the composition of sugars; c) the size of the oligosaccharide species, and the percentage and distribution at each individual glycosylation site; d) the linkage of carbohydrates and their conformation.

Two of the most powerful techniques for carbohydrate structure elucidation are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. Both techniques have been widely used successfully, and are in some way complementary. Mass spectrometry can provide the accurate molecular weight of intact glycoproteins/glycopeptides, carbohydrate residues, and non-carbohydrate substituents and, under favorable circumstances, produce sequence specific fragment ions (Carr et al., 1991; Schindler et al., 1995; Carr et al., 1993; Carr et al., 1990; Huberty et al., 1993; Dell et al., 1994). In contrast, <sup>1</sup>H NMR provides the information about linkage position and anomeric configurations (Halbeek, 1994; Trimble and Atkinson, 1986; Trimble and Atkinson, 1992). The work described in this thesis utilizes various mass spectrometric strategies in conjunction with enzymatic digestions to determine the location, distribution and structure of the carbohydrate moieties at the *N*-glycosylation sites of external invertase of yeast. The second part (Chapter III) describes the identification of another glycoprotein, apolipoprotein D, including its glycosylation pattern.

## **B. Methodology for the Determination of the Structure of Glycoproteins**

Traditionally, glycoproteins obtained from natural sources are identified by phenol-sulphuric acid assay (Dubois 1956), and the sugar composition is determined by acid hydrolysis followed by chromatographic analysis (Spiro, 1966) or analysis of derivatized monosaccharide residues by MS (Kochetkov and Chizhov, 1966; 1972). Further investigation extends to the location of glycosylation, degree of glycosylation, distribution of oligosaccharides at each sequon, the primary sequence of oligosaccharides and proteins, and the conformational information of the sugar moiety. All these information can now be obtained by NMR and MS (Laine, 1990) combined with enzymatic (e.g. protease, endo- and exoglycosidase) digestion.

When the site of glycosylation is not the primary objective, determination usually begins with releasing glycans. In most NMR studies the carbohydrate residues are first released from protein because sequence analysis needs a single glycan component and the glycopeptide usually has multiple glycans. Furthermore, the peptide moiety may complicate the interpretation of the NMR spectrum. Techniques available for glycan release include enzymatic digestion with PNGase F, Endo H, or O-glycanase (Tarentino et al., 1989; Tarentino and Plummer, Jr. 1994) and chemical cleavage by hydrazinolysis (Patel and Parekh, 1994) and  $\beta$ -elimination (reductive alkaline degradation) (Lee and Scoocca, 1972; Spiro, 1972; Linsley et al, 1994). The released glycans are subsequently separated from the protein and the glycan pool fractionated by different chromatographic techniques including HPAE-PAD (Hardy and Townsend, 1994), HPLC (Baenziger 1994), electrophoresis (Jackson 1994; Linhardt, 1994), gel filtration (Spiro, 1966), and

lectin/antibody affinity (Cummings, 1994). The carbohydrate residues are then analyzed by NMR. Analysis of the NMR parameters such as chemical shifts, integration, coupling constants, nuclear Overhauser effect (NOE), and spin-lattice relaxation times ( $T_1$ ) together with 2-D NMR information will provide sufficient information to deduce the structures of the individual carbohydrates (Gorin, 1981; Bock and Thorgersen, 1982; Bock, and Pedersen, 1983; Sweeley and Nunez, 1985; Halbeek, 1994). Even though HPAE-PAD (Spellman, 1990; Hardy and Townsend, 1994) and other techniques (such as MS, discussed later) can also provide some information of the glycan sequence, NMR is the single most powerful and reliable approach for the determination of linkage and conformation of carbohydrates. However, this technique needs a relatively large amount of homogeneous sample (normally  $\mu\text{mol}$  level) and the data interpretation requires considerable expertise. Obviously, this strategy does not provide any information about the connections between the glycans and the glycosylation sites in the glycoprotein.

For the whole structure determination of a glycoprotein, the amino acid sequence has to be known first either from DNA sequence or by mass spectrometry or Edman degradation. In order to determine the site(s) of glycosylation, the intact glycoprotein is first cleaved by a protease followed by chromatographic separation in order to provide peptides preferably containing a single glycosylation site. The proteolytic peptides are then analyzed by mass spectrometry including the following steps.

1. Identification of the glycopeptides in the chromatogram.
2. Assignment of attachment sites and oligosaccharide populations.
3. Determination of the degree of occupancy at each glycosylation site.

4. Sequencing of oligosaccharides.

5. Confirm the sequence of the deglycosylated peptide.

The molecular weight of the glycopeptides and the deglycosylated peptides are determined by mass spectrometry which provides direct information for glycoprotein structure elucidation. During the 1980's, mass spectrometry has emerged as a powerful method for the analysis of large biomolecules and was summarized in an entire volume of *Methods of Enzymology* (McCloskey, 1990). Electrospray ionization (ESI) MS (Meng et al., 1988; Fenn et al., 1989; Loo et al., 1989; Hemling et al., 1990; Smith et al., 1990) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Karas and Hillenkamp, 1988, 1989; Hillenkamp and Karas, 1991; Siegel et al., 1991; Chait and Kent, 1992; Huberty et al., 1993) have been employed successfully for the analysis of glycopeptides and to structurally characterize carbohydrates. Due to their overwhelming advantages in sensitivity, mass range, and ease of data acquisition MALDI and ESI have now gradually replaced fast atom bombardment (FAB) MS (Barber et al., 1981), the favorite technique of the 80's.

#### 1. *Identification of the glycopeptides in the chromatogram*

Localization of glycopeptides in a chromatogram can be accomplished by monitoring for specific diagnostic sugar oxonium ions with online HPLC/ESIMS scanning in the selected ion monitoring (SIM) mode (Huddleston et al., 1991; Carr et al., 1993; Schindler et al., 1995). In this experiment, abundant fragment ions are produced by increasing the collision excitation potential in the electrospray ionization source or in the second collision region of a triple quadrupole mass spectrometer (Carr et al., 1993). The instrument is then set to monitor one or more of the following

carbohydrate specific ions (product ions formed from the glycopeptide precursor ions):  $m/z$  366 (hexosyl *N*-acetylhexosamine, [Hex-HexNAc]<sup>+</sup>),  $m/z$  292 (*N*-acetylneuraminic acid or sialic acid, [NeuAc]<sup>+</sup>),  $m/z$  204 (*N*-acetylhexosamine, [HexNAc]<sup>+</sup>),  $m/z$  163 (Hexose [Hex]<sup>+</sup>), and  $m/z$  147 (deoxy-hexose, [dHex]<sup>+</sup>). Since  $m/z$  204 can be formed from both *N*- and *O*-linked glycopeptide, it is often chosen as the ion to be monitored (Carr et al., 1993). The marker ions detected in the SIM chromatogram indicate the presence of glycopeptides (either *N*- or *O*-linked). The *N*- and *O*-linked glycopeptides can be distinguished by repeating the same experiment after removing the *N*-linked carbohydrates by PNFase F.

Localizing *N*-linked glycopeptides in a chromatogram is usually simpler by using MALDI-TOF MS, because this technique generally produces singly charged ions and the  $m/z$  values observed can be directly related to the molecular weights of the components of a more complex heterogeneous glycopeptide. Most *N*-linked glycopeptides carry oligosaccharide chains of various size at a glycosylation site. These oligosaccharide moieties differ by successive terminal sugar residues at the non-reducing end and thus generate a series of (M+H)<sup>+</sup> ions that differ in the mass by an additional sugar residue: 132 = pentose, 146 = deoxy-hexose, 162 = hexose, 203 = *N*-acetylhexosamine, and 291 = *N*-acetylneuraminic acid (sialic acid). The presence of such a signal pattern indicates the presence of a glycopeptide in a chromatographic fraction and simultaneously reveals the distribution of oligosaccharides. This approach was applied throughout this work and is further discussed in Chapter II. *O*-linked glycopeptides can be essentially localized using a similar approach assisted by *O*-glycanase and/or chemical cleavage. The less heterogeneous or even homogeneous short *O*-linked oligosaccharides may make it more difficult to recognize these

glycopeptides because their mass spectra will display a less obvious pattern or even only a single peak.

## 2. *Assignment of attachment sites and oligosaccharide populations.*

The methods for locating a particular occupied glycosylation sites and deducing the population of the oligosaccharides at each site have been described in the literature (Carr et al., 1993; Huberty et al, 1993). One of these methods was used during this study. Its basic strategy is to remove the glycans by endo glycosidase digestion (PNGase F, Endo H, *O*-glycanase) or chemical cleavage (hydrazinolysis,  $\beta$ -elimination) followed by determination of the molecular weight of the deglycosylated peptide by mass spectrometry. The molecular weight of the deglycosylated peptide is then used for its identification by matching it with those predicted from the known amino acid sequence of the protein and the specificity of the protease used . The sequence of the matching peptide indicates the glycosylation site candidate. In the rare case where the protein sequence is unknown, the amino acid sequence of the deglycosylated peptide must first be determined as discussed in Section I.B.5.

It is well known that PNGase F catalyzes the hydrolysis of the glycosidic bond between asparagine and the first *N*-acetylglucosamine (GlcNAc) converting the glycosylated asparagine to aspartic acid. On the other hand, with the exceptions noted below, Endo H cleaves the bond between the first and the second *N*-acetylglucosamine in the core region of the oligosaccharide leaving the first GlcNAc attached to asparagine (Tarentino and Maley, 1974a; Tarentino et al., 1985; Tarentino



et al., 1989; Maley et al., 1989). Therefore, the resulting deglycosylated peptide is one dalton heavier in mass than the non-glycosylated peptide when PNGase F was used, and 203 Da heavier if Endo H was used. Therefore, by comparing the mass difference between the peptide deglycosylated by PNGase F and that deglycosylated by Endo H, one can deduce the number of occupied sequons in the glycopeptide. For instance, if two PNGase F and Endo H deglycosylated peptides have a mass difference of 404 Da ( $203 \times 2 - 2$ ) the glycopeptide must contain two sequons. However, this method is not universal because it holds only for N-glycosylated peptides and certain oligosaccharides such as complex oligosaccharides and those oligosaccharides that have less than three mannose residues can not be cleaved by Endo H. In addition, the oligosaccharides without additional mannose attached to the core  $\alpha$ -1,6-mannose (the upper arm of the N-oligosaccharide structures shown previously) are very difficult to hydrolyze with Endo H (Tarentino et al., 1989; Maley et al., 1989).

The mass of the carbohydrate moiety in a glycopeptide is an important factor of its structure, and it can be easily obtained by subtracting the mass of the deglycosylated peptide from the mass of the glycopeptide. Huberty et al. (1993) derived the oligosaccharide composition from its mass by the equation

$$C_m = \sum n_i m_i \quad (1)$$

where,  $C_m$  is the mass of the carbohydrate portion of a glycopeptide,  $m$  is the mass of a monosaccharide residue ( $i$ ), and  $n$  is the number of each monosaccharide residue present. Taking

into account the knowledge of the core region structure,  $\text{Man}_3\text{GlcNAc}_2$ , in *N*-linked glycopeptides and the structures which have been reported in the literature, the potential composition of a carbohydrate residue can be derived using the equation described above. For the *O*-linked oligosaccharides, the protocol described above should work as well. In addition, as the *O*-linked carbohydrates usually consist of only one to four sugar residues (Darnell et al., 1990), it is not difficult to calculate their composition.

As discussed above, the distribution of oligosaccharides at each glycosylation site can be easily derived from a MALDI spectrum while it is more difficult to obtain this information by ESI, especially those "wide range glycosylation", high mannose type glycoproteins produced by yeasts (for the analysis of "wide range glycosylation", high mannose type glycoproteins see Chapter II).

### 3. *Determination of the degree of occupancy at each glycosylation site*

The determination of the degree of occupancy at each glycosylation site requires a quantitative method. The mass spectrometric methods used for large, polar molecules are not well suited for quantitative analysis and, therefore, theoretically can not provide this kind of information. However, the methods described below allow a rough estimate of the fraction to which a particular glycosylation site is occupied. After proteolytic digestion, the partially occupied glycosylation site will be recovered both as a peptide and a glycopeptide containing an identical amino acid sequence. Usually, they both can be detected in adjacent HPLC fractions based on the mass(es) difference of the associated carbohydrate portion. The oligosaccharide occupancy at this particular site can be

assessed by combining these fractions, removing the oligosaccharide portion with Endo H, and then comparing the relative signal intensities between the carbohydrate-free peptide and the Endo H peptide (contains the first GlcNAc); or acquiring the mass spectra before and after removing the carbohydrates with PNGase F, observing the changes of relative abundance between the sugar-free peptide signal and the reference peak (usually is the strongest peak) in the spectrum before and after PNGase F deglycosylation, . These methods have been used in the study of invertase and further described in Chapter II. Quantitative analysis of HPLC or SDS-PAGE fractions of the free peptide and the glycopeptide can also provide the information of the glycosylation site occupancy (Parekh, 1994). A more accurate method would be to determine the amino acid sequence by Edman degradation on the mixture of the two peptides (the carbohydrate-free peptide and the PNGase F deglycosylated peptide) and quantitatively compare the asparagine/aspartic acid ratio at the sequon.

#### *4. Sequencing of oligosaccharides*

As discussed earlier, NMR spectroscopy is the best single technique for the determination of the structure of a homogeneous oligosaccharide. For the analysis of heterogeneous glycopeptides at a picomole level, a more sensitive method with the ability to handle complex mixtures is required. Such methods have been reported, including sequential digestion with exoglycosidases (Kobata, 1979; Ohta et al., 1991; Jacob and Scudder, 1994; Schindler et al., 1995) and periodate oxidation followed by reduction and permethylation (Angel et al., 1987; Angel and Nilsson, 1990; Linsley et al., 1994). The resulting products are detected and analyzed by a variety of mass spectrometric techniques.

A typical example of sequential exoglycosidase digestion illustrated in Figure I.1. The selected exoglycosidases should have the following features: cleaving only non-reducing terminal monosaccharide(s), specificity for configuration, specificity for anomericity, and specificity for

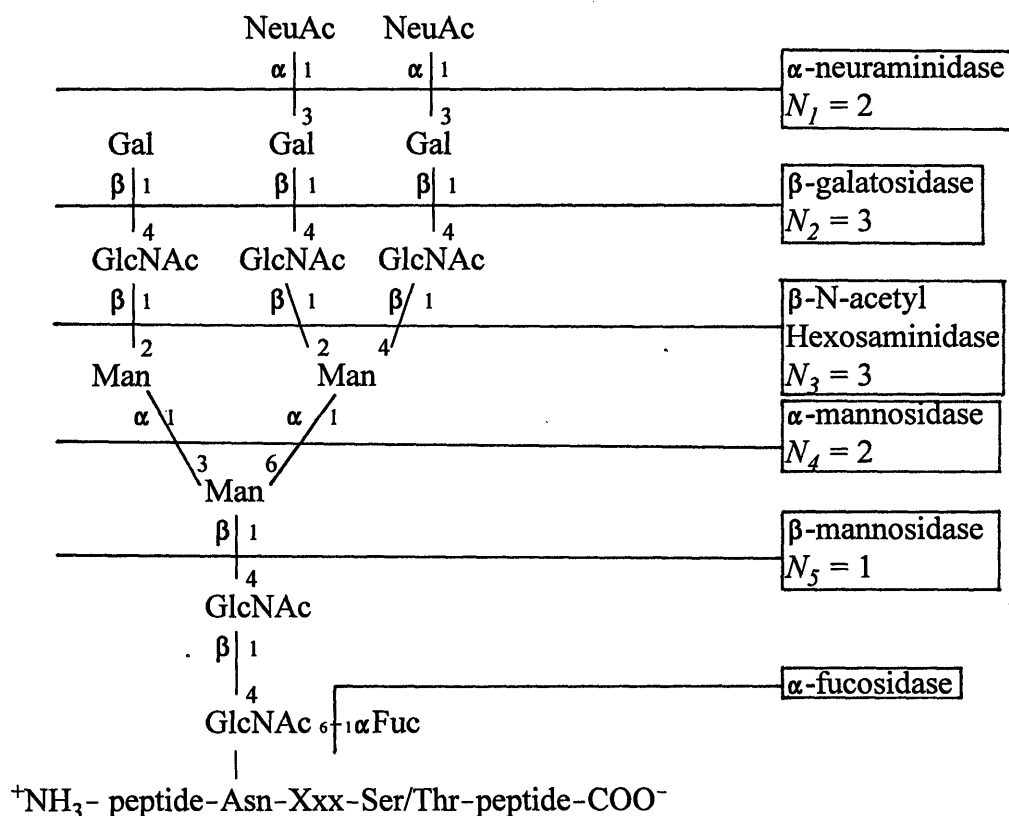


Figure I.1. Determination of a complex type carbohydrate residue linkage by sequential exoglycosidase digestion.

different monosaccharides and, preferably, some linkage specificity. The basic strategy of these experiments involves the sequential removal of the outer (non-reducing end) sugar(s) by one of the

selected specific exoglycosidases at a time, followed by molecular weight measurements of the oligosaccharide chain(s) by MS or gel filtration and other techniques such as HPAE. The mass difference ( $\Delta m$ ) between the newly formed shorter oligosaccharide and its precursor determined by MS can be used to calculate the monosaccharide mass ( $M_m$ ) of non-reducing terminal sugar(s) by the equation

$$M_m = \Delta m / N_n \quad (2)$$

where,  $N$  is the number of the monosaccharide(s) being removed by a specific exoglycosidase; and  $n$  is the step in the exoglycosidase digestion sequence. The calculated  $M_m$  identifies the non-reducing terminal sugar residue (the mass of monosaccharide residues have been listed earlier). As illustrated in Figure 1, the largest value of  $N$  in equation (2) could also indicate the branch information of the oligosaccharide (e.g. biantennary or triantennary, et al.) (Schindler et al. 1995). The advantage of this method is that one can determine the glycan sequence without liberating the oligosaccharide from the peptide. However, it can only work on the complex type glycosylation, and it could give the wrong information for a branch in which two or more of the same monosaccharides are linked in line with the same type of linkage.

Periodate oxidation, a well-known reaction in carbohydrate chemistry, cleaves the adjacent diol C-C bond in a monosaccharide residue and converts the hydroxyl groups to aldehydes. The aldehydes are subsequently reduced by NaBD<sub>4</sub> and permethylated by MeI. The products will cause certain mass changes which indicate the possible linkage(s) (Angel, 1987; Angel and Nilsson,

1990). Additional structural information of the periodate oxidation products can be obtained by analyzing the fragments produced in normal FAB MS and/or CID MS/MS (Linsley et al., 1994; Reinhold et al., 1994). Some aspects of the oligosaccharide sequence can be also determined directly by collision-induced-dissociation (CID) FAB mass spectrometry (Domon and Costello, 1988; Laine, 1990; Gillece-Castro and Burlingame, 1990; Orlando et al., 1990). More recently it has been reported that MALDI-PSD can be used for the sequence determination of the carbohydrate moieties released from glycopeptides (Rouse et al., 1995, Talbo and Mann, 1996). Details of the MALDI-PSD technique will be described in Section I.D.

##### *5. Confirm the sequence of the deglycosylated peptide*

The determination of the amino acid sequence of peptides is another part of glycoprotein analysis. In general, the amino acid sequence of the glycoprotein will be known either from the corresponding DNA sequence or via amino acid sequence analysis before determining the sites and nature of the attached glycans. The two most successful approaches for peptide sequencing are the conventional Edman degradation (Edman and Begg, 1967) and more recently, tandem mass spectrometry utilizing low energy CID (Hunt et al., 1986) or high energy CID (Biemann, 1986). Even though the latter two have many advantages that the Edman method lacks, the two different methods often complement each other. In this section the use of mass spectrometry alone and used in conjunction with chemical or enzymatic processes for amino acid sequence determination will be discussed briefly.

The "soft" ionization techniques such as FAB, PD, MALDI, and ESI were mainly developed for the determination of thermolabile and none-volatile biomolecules. The protonated molecules produced by these methods usually do not have the excess energy required for fragmentation in the ion source, hence the resulting mass spectra often lack structural information. The attempts to overcome this problem are basically divided into two strategies: (1) prior ionization fragmentation: the "fragments" produced in this way are actually the "ragged end" peptides sequentially truncated by a modified Edman degradation and/or exopeptidase digestion; (2) post ionization fragmentation: ions extracted from the ion source can be dissociated by transferring sufficient energy to a protonated peptide via collision with an gas atom, a method known as collision-induced-dissociation (CID). Post ionization fragments can also be produced by the decay of the metastable ions during the free flight in a time-of-flight instrument and this fragmentation can now be detected (see Section I.D discussing post-source decay).

The "prior ionization fragmentation" can be accomplished by a modified Edman degradation method (Chait et al., 1993), termed "chemical ladder sequencing". By adding a small amount (about 5%) of phenylisocyanate (PIC) to prevent further removal of the new N-terminal amino acid in each reaction cycle, a sequence-defining concatenated set of product peptides are produced after a certain number of cycles. The MALDI-TOF MS of the resulting mixture consists of a set of  $(M+H)^+$  ions of N-phenylaminocarbonyl-peptides spaced by the mass of consecutive amino acid residues ("ladder"), and the peptide sequence can be read directly from this spectrum. A variation of this strategy for generating such "ladder" involves addition of equal aliquot of the starting peptide to each cycle instead of using an additional chain termination reagent (Bartlet-Jones et al., 1994). In this

method, a volatile degradation reagent, trifluoroethylisothiocyanate (TFEITC), instead of the original Edman reagent phenylisothiocyanate (PITC) is used. Since all the reagents used are volatile, they can be easily removed in vacuo and providing a clear spectrum consisting of the  $(M+H)^+$  ions of shorter and shorter peptides.

The "prior ionization fragmentation" can also be generated by exopeptidases (Bradley et al., 1982; Chait et al., 1986; Wagner and Fraser, 1987; Schaer et al., 1991), resulting in "enzymatic ladder sequencing". The strategy is to partially remove amino acids either from the N-terminus by an aminopeptidase or from the C-terminus by a carboxypeptidase and measuring the mass spectrum at varying time intervals. The biggest problem of this method is that the available exopeptidases (either amino- or carboxypeptidases) cleave peptide bonds at a different rate depending on the amino acid involved. Therefore, there may be no signal for a certain amino acid in the sequence when a fast-cleaving amino acid is followed by a slow-cleaving one, no matter how careful the time-dependent studies are carried out.

The advantages of these "prior ionization fragmentation" methods are: (i) simple interpretation of the mass spectra, (ii) can sequence a relatively large peptide (e.g. 3000-5000 Da) than tandem mass spectrometry can do. Their major problems are: first, no side chain information can be provided so that the amino acids like leucine and isoleucine can not be identified; second, neither the modified Edman degradation (chemical ladder sequencing) nor partial exopeptidase digestion (enzymatic ladder sequencing) alone can sequence the whole peptide since the further the truncation goes the weaker the signals become. Therefore, they are often used to confirm the identity



of proteolytic peptides from a protein of known sequence rather than to sequence an unknown peptide.

The "post ionization fragmentation" is achieved by converting some of the kinetic energy of the molecule ions into vibrational energy. This can be accomplished by introducing a neutral gas into a chamber to cause the collision with the molecule ions passing through, and detection of the resulting fragments by tandem mass spectrometry (CID MS/MS) (Biemann, 1990a).

In a FAB CID experiment, the peptide molecules are ionized in the FAB source, to form protonated peptide ions  $(M+H)^+$  which are accelerated (usually to 10 kV) and selected by the first double focusing sector instrument (MS1). The resolution in MS1 is set to allow only the  $C^{12}$  isotope component of this protonated peptide ion  $(M+H)^+$  (precursor ion) to pass through a collision cell located in the field free region between MS1 and MS2 (see Figure I.2). The collision cell is held at a potential of a few kV (usually 3 kV) above ground and filled with an inert gas (He, Ar, or Xe) at a pressure of about  $10^{-3}$  torr which just suffices to decrease the abundance of the precursor ion by two thirds or more. The  $(M+H)^+$  ion (precursor ion) traveling with a kinetic energy of 10 keV will collide with the inert gas atom causing transfer of some of this energy into vibrational energy to break the backbone peptide bonds and/or the side chain bonds. Those fragments which retain the charge (product ions) can be analyzed in MS2 using a linked scan (Jennings and Mason, 1983) where the ratio of the magnetic and electric fields (B/E) are kept constant. This scan function is used with a grounded collision cell, because it keeps the fragment ions which have a lower kinetic energy

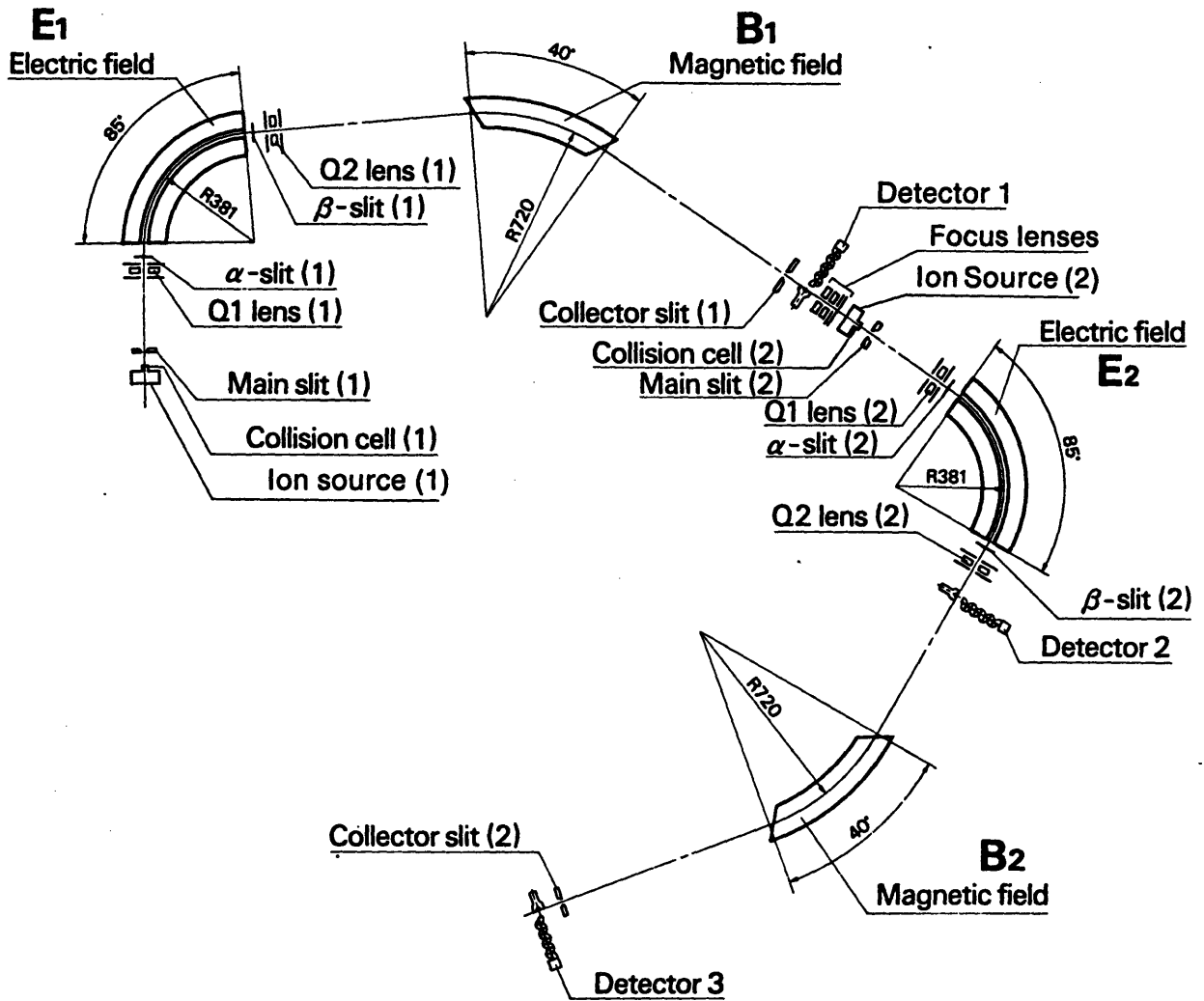
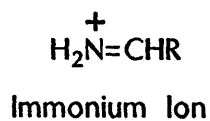
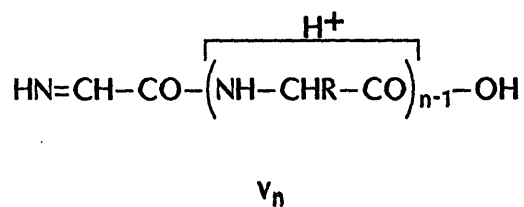
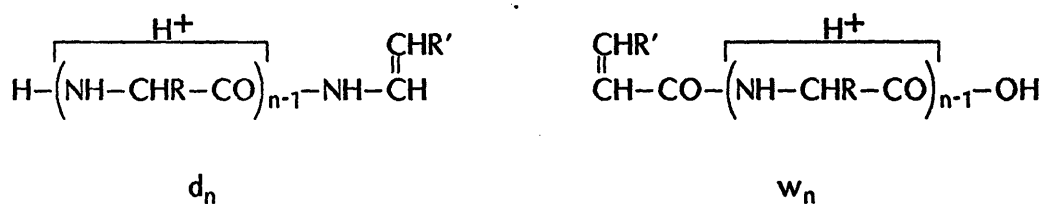
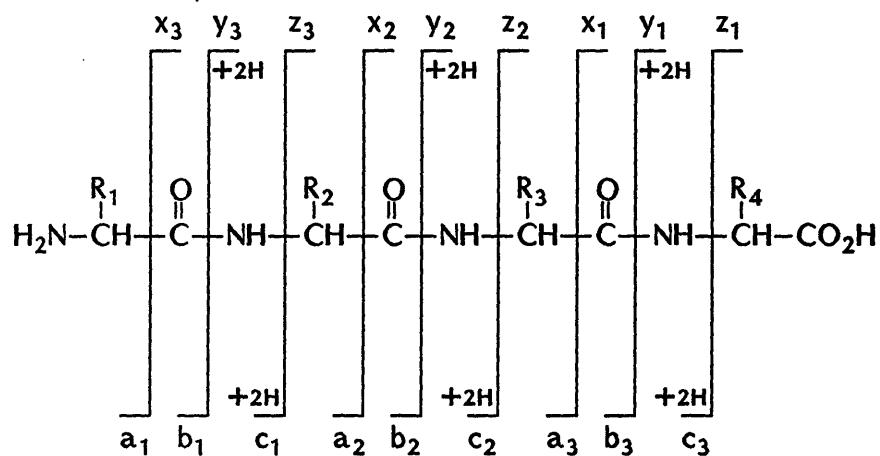


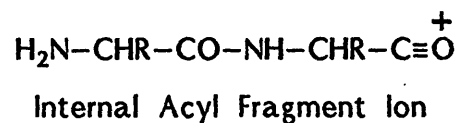
Figure I.2. Schematic diagram of the JEOL HX110/HX110 four-sector mass spectrometry.

in focus. For a collision cell held at a potential of several kV above ground the ratio of B/E must be continuously varied during the scan (Sato et al., 1987).

The fragments produced by CID have been clearly described in detail (Papayannopoulos, 1987; Johnson, 1988; Biemann, 1988; 1990b; 1992) and are shown in Figure I.3 where the  $a_n$ ,  $b_n$ , and  $c_n$  and the  $x_n$ ,  $y_n$ , and  $z_n$  indicate the N-terminal fragment ions and the C-terminal fragment ions, respectively. They are formed by cleaving a single bond in the peptide backbone. Note that the  $c_n$  and  $y_n$  ions are retaining the original protonating hydrogen after the cleavage and transferred an hydrogen from the other side of the neutral fragment species. The immonium ions of single amino acids and the internal fragment ions are dissociated by breaking two peptide bonds; these ions are defined by the single amino acid code. The immonium ions in a CID spectrum indicate the amino acids present in the peptide, and the internal ions are usually used to support the interpretation of the CID spectrum. The other N-terminal ions and C-terminal ions defined as  $d_n$ , and  $w_n$ , respectively, are formed by the cleavage on both the backbone and of the  $\beta$ - $\gamma$  bond of the adjacent side chain. They are useful in differentiating the two isomeric amino acids leucine and isoleucine. Similar to the fragmentation processes leading to  $w_n$  ions, the  $v_n$  ions are the C-terminal fragments produced by breaking the peptide bond and loss of the entire adjacent side chain (cleavage of the  $\alpha$ - $\beta$  bond). The interpretation of CID spectra will be further discussed together with PSD spectra in Section I.D.



(denoted by single letter code)



(denoted by single letter code)

Figure I.3. Nomenclature and definition of the fragment ions formed by CID from protonated peptides.

### C. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry

The success of mass spectrometry in solving biochemical and biological problems relies on the "soft" ionization techniques developed since the late 1970's, based either on desorption or on spray ionization. In the former, gas-phase ions are produced by applying a strong electric field to the sample (field desorption, FD; Beckey, 1977); by bombardment of the sample with ions (secondary ion mass spectrometry, SIMS; Benninghoven and Sichtermann, 1978); by impact of the analyte molecules with the particles produced from  $^{252}\text{Cf}$  fission (plasma desorption, PD; Macfarlane et al., 1976); or by irradiating the sample with photons from a laser irradiation (laser desorption/ionization, LDI; Posthumus et al., 1978). With the exception of PD, none of these were useful for large organic compounds.

Addition of a liquid matrix (e.g. glycerol) and using fast (kV) atoms Barber et al. (1981) was able to overcome the radiation damage to the analyte (fast atom bombardment, FAB). Later it was shown in Burlingame's group (Aberth et al., 1982) that ions also can be used (liquid secondary ion mass spectrometry, LSIMS). Both atom and ion bombardment of analytes dissolved in a liquid matrix are now widely used and even though in the work described here a  $\text{Cs}^+$  ion beam was applied we still use the term FAB in deference to the pioneering original work by Barber.

More recently, LDI was adopted and improved by embedding the sample in a large molar excess of a matrix absorbing strongly at the irradiating wavelength. In this case the laser beam does not desorb the analyte directly but does initiate a more controllable efficient energy transfer from

matrix to the analyte molecules which are consequently desorbed and ionized (matrix-assisted laser desorption/ionization, MALDI; Karas and Hillenkamp, 1988). For the spray techniques, the gas-phase ions are produced by spraying a dilute sample solution through a needle like nozzle placed either in a heating area (thermospray ionization, TSI; Blakely et al., 1980) or in a strong electrostatic field (electrospray ionization, ESI; Meng et al., 1988). A comparison of these techniques (Table I.1), shows that to a certain extent they are complementary.

Table I.1. Comparison of the "soft" ionization techniques for biomolecules.

	Accuracy (%)	Sensitivity (pmole)	Time required <sup>a</sup> (minutes)	Mass range <sup>b</sup> (dalton)	Major ions <sup>c</sup> produced
FD	0.05-0.01 <sup>d</sup>	500-1,000	5-10	100-3,000	M <sup>+</sup>
FAB/LSIMS	0.05-0.01 <sup>d</sup>	100-1,000	2-6	300-10,000	(M+H) <sup>+</sup>
PD	0.50-0.10	50-1,000	10-1,200	100-25,000	(M+H) <sup>+</sup>
MALDI	0.30-0.10	0.1-50	4-8	500-250,000	(M+H) <sup>+</sup>
TSI	0.10-0.05	10-100	5-15	100-1000	(M+H) <sup>+</sup>
ESI	0.10-0.05	0.1-50	5-15	100-100,000	(M+nH) <sup>n+</sup>

<sup>a</sup> The time required for acquire a spectrum, not including the sample preparation.

<sup>b</sup> The practical useful mass range.

<sup>c</sup> When the instrument is operated on the positive mode.

<sup>d</sup> The accuracy when the instrument is operating at low resolution.

MALDI-TOF MS has been used extensively in the work to be presented in the next two chapters.

The methodology is discussed in the following.

## 1. *Principles of matrix-assisted laser desorption/ionization*

Matrix-assisted laser desorption/ionization, one of the most important improvements in ionization techniques, was developed in the late 80's in Hillenkamp's laboratory (Münster, Germany) and has become a powerful tool for the analysis of nonvolatile, thermolabile, large biomolecules. It has a detection limit from femtomoles to low picomole (Hillenkamp et al., 1991; Stahl et al., 1991; Spengler and Cotter, 1990; Beavis and Chait, 1990; Karas et al. 1990). The use of lasers as an energy source for the ionization of small bio-organic molecules for the analytical application had been explored since the late 1960's (Vastola and Pirone, 1968; Vastola et al., 1970; Mumma and Vastola, 1972; Posthumus et al., 1978). Although a variety of lasers with different wavelength and pulse widths had been employed to combine with every available type of mass spectrometer (see Table I.2 for commonly used laser sources), these attempts to analyze large molecules were not quite successful. They failed because energy transfer required for resonant excitation of the analyte molecules also puts energy into photodissociation channels during resonant desorption; in nonresonant desorption experiments the required irradiances are very close to the onset of plasma generation which also destroys large molecules (Hillenkamp et al., 1991). This problem was finally overcome when the matrix technique was introduced by Hillenkamp's group (Karas et al., 1987). It was demonstrated that "matrix-assisted" laser desorption/ionization can generate gas-phase molecular ions of proteins up to 60 kDa (Karas and Hillenkamp, 1988) and shortly thereafter a protein signal at ca. 300 kDa was detected (Karas et al., 1990). Two of the key elements which led to this success are the use of a pulsed UV laser (frequency-quadrupled Nd:YAG laser with a wavelength of 266 nm) and a suitable UV absorbing matrix (nicotinic acid). In principle, the use

of a very short time (1-100 ns) pulse laser avoids thermal decomposition and is also ideally suited for mass analysis in a time-of-flight mass spectrometer which allows the detection of very large

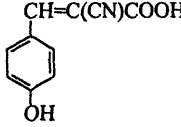
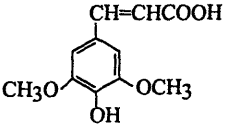
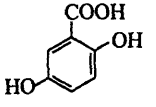
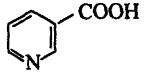
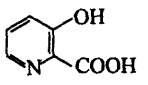
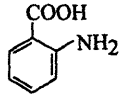
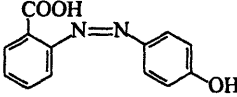
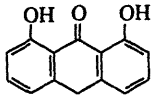
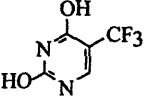
Table I.2. Summary of the laser source adopted in MALDI mass spectrometry.

Laser types	Wavelength
	<u>uv (nm)</u>
N <sub>2</sub>	337
4 $\omega$ -Nd:YAG	266
3 $\omega$ -Nd:YAG	355
Excimer	193
	248
	308
	351
	266
Excimer pumped dye	581
	<u>IR (<math>\mu</math>m)</u>
Er:YAG	2.94
TEA-CO <sub>2</sub>	10.6

molecules. The large molar excess of matrix (>1000:1) serves to absorb most of the energy, forming a plume of desorbed matrix and analyte molecules (Hillenkamp et al. 1991), in which the latter retain only a small fraction of the total energy. The commonly used matrices are shown in Table I.3, and each of these matrices are: (i) a small solid organic compounds of low volatility; (ii) soluble in the solvents used for dissolving analytes (for proteins/glycoproteins, the solvent usually is water or a mixture of water and acetonitrile or ethanol); (iii) strongly absorbing at the laser wavelength used for desorption; (iv) compounds containing OH and/or NH group(s) capable of transferring a proton



Table I.3. Summary of the commonly used matrices in MALDI mass spectrometry.

Name	Abbreviation	Structure	Used at	Applications
$\alpha$ -cyano-4-hydroxy-cinnamic acid	$\alpha$ -CHCA		266 nm 337 nm 2.94 $\mu$ m	proteins/peptides glycoproteins/ glycopeptides
sinapinic acid	SA		266 nm 337 nm 2.94 $\mu$ m	proteins/peptides
2,5-dihydroxybenzoic acid	DHB		337 nm 2.94 $\mu$ m	proteins/peptides glycoproteins/ glycopeptides
nicotinic acid			266 nm 337 nm	proteins
3-hydroxypicolinic acid	HPA		266 nm 337 nm	oligonucleotides
2-aminobenzoic acid	ABA		337 nm	oligonucleotides
2-(4-hydroxyphenylazo)-benzoic acid	HABA		266 nm 337 nm	proteins glycoproteins synthetic polymers
dithranol			337 nm	nonpolar synthetic polymers
succinic acid		$(\text{CH}_2\text{COOH})_2$	2.94 $\mu$ m	proteins/peptides oligonucleotides
5-(trifluoromethyl)-uracil	TFMU		337 nm 2.94 $\mu$ m	proteins/peptides

to the analyte to form a  $(M+H)^+$  ion during ionization. However, even if an organic compound has all these properties it is not necessary a good matrix for MALDI. Sometimes even structural isomers are not equally suited matrices. For example, 2,5-dihydroxybenzoic acid (DHB), is a much better matrix than the other isomers (Ehring et al., 1992). Another problem is that some peptides can be detected in one special matrix (such as DHB) but not in another (such as  $\alpha$ -CHCA), and *vice versa*. Great efforts have made to find useful matrices (Beavis and Chait, 1989; Strupat et al., 1991; Ehring et al., 1992; Beavis et al., 1992; Juhasz et al., 1993; Fitzgerald et al., 1993). Also certain mixtures of matrices have advantages. For example, a mixture of DHB and 2-hydroxy-5-methoxybenzoic acid (2H5MB; also named 5-methoxysalicylic acid, 5MSA) significantly enhances both resolution and sensitivity, especially at high mass ( $> 50$  kDa) (Karas et al., 1992, Billeci and Stults, 1993). When DHB was mixed with a monosaccharide such as glucose or fructose the resolution increased for MALDI in a Fourier Transform Mass Spectrometer (FTMS) (Köster et al., 1992). It was found that the mixture of fucose/DHB results in the best coverage for peptide mapping (Billeci and Stults, 1993), and also reduced 10-fold the abundance of sodium adduct. As discussed earlier, MALDI mass spectrometry generally is not very suitable for quantitative analysis. However, it was reported that the use of a trimatrix, DHB/fucose/2H5MB, can produce the highest quality spectra (Billeci and Stults, 1993) and significantly enhance the shot-to-shot reproducibility, allowing quantitative analysis by MALDI MS (Gusev et al., 1995).

Although there is no doubt that the major role of the matrix is in assisting the laser desorption/ionization, up to date, the mechanism(s) of desorption/ionization of large molecules stimulated by the laser excited matrix molecules is not yet completely understood. Figure I.4

illustrates the general features of MALDI: when a matrix is exposed to a laser irradiance of proper wavelength it will either couple to electronic states during UV irradiation (electronic excitation) or excite to rovibrational states during IR irradiation (rovibronic excitation). These energized matrix molecules quickly transfer sufficient energy to the analyte molecules to cause subsequential desorption and ionization. Another effect of the high sample dilution is the "matrix isolation" of the individual analyte molecules which prevents their association to complexes of a mass too large to be desorbed (Hillenkamp et al., 1991).

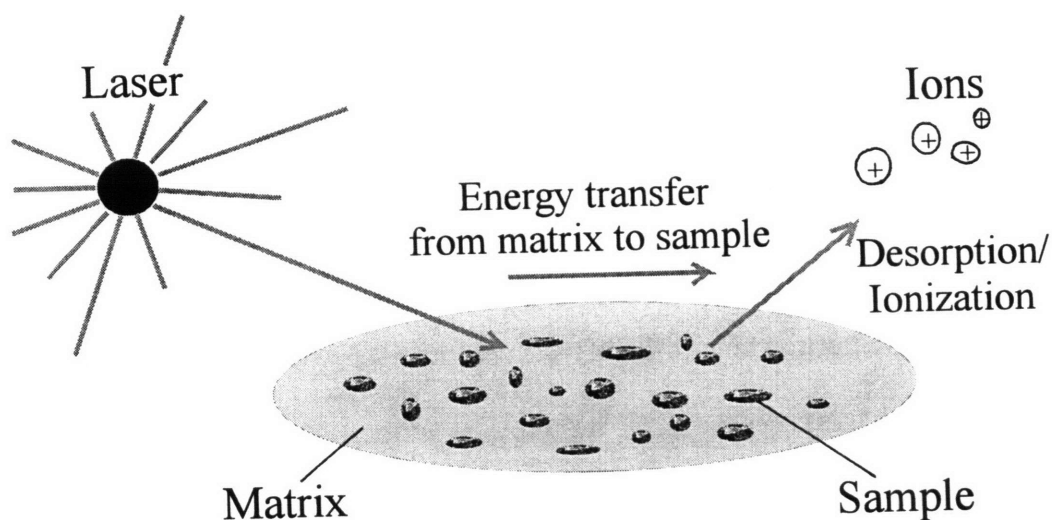


Figure I.4. The general processes of matrix-assisted laser desorption/ionization.

The basic desorption/ionization processes described above (see Figure I.4) include three steps: (i) the matrix absorbs energy from laser irradiation; (ii) energy transfer; (iii) desorption and ionization. The processes of energy transfer from the energized matrix molecules to the analyte molecules leading to desorption and ionization is not well understood. A reasonable and acceptable kinetic model has been proposed for the desorption process (Vertes et al., 1990). This model describes how the analyte molecules desorb via a rapid sublimation and is supported by numerical simulations of the sublimation kinetics based on realistic experimental parameters. On the other hand, the ionization procedure is not easily described by a simple mechanism. In fact, many mechanistic models have been proposed by others (Ehring et al., 1992; Beavis and Chait, 1992; Gimon et al., 1992; Chiarelli et al., 1993; Karas et al., 1995; Liao and Allison, 1995).

## *2. Time-of-flight mass spectrometer*

The principles of time-of-flight mass spectrometry have been summarized recently (Cotter, 1994; Guilhaus, 1995). Figures I.5a and I.5b show the schematic of the linear Vestec VT2000 and PerSeptive Voyager-Elite reflectron time-of-flight mass spectrometers which were used in this study. The VT2000 time-of-flight mass spectrometer was the first MALDI-TOF commercial instrument which is equipped with a Nd:YAG laser ( $\lambda=266$  nm, 9 nsec pulse width), a N<sub>2</sub> laser ( $\lambda=377$ nm, 3 nsec pulse width), and an Er-YAG laser ( $\lambda=2.94$   $\mu$ m, 120 nsec pulse width). This variety of laser sources attached to the VT2000 expands the capability for the study of various materials (e.g. proteins, carbohydrates, DNA/RNA, lipids, and synthetic polymers) by using the different matrix

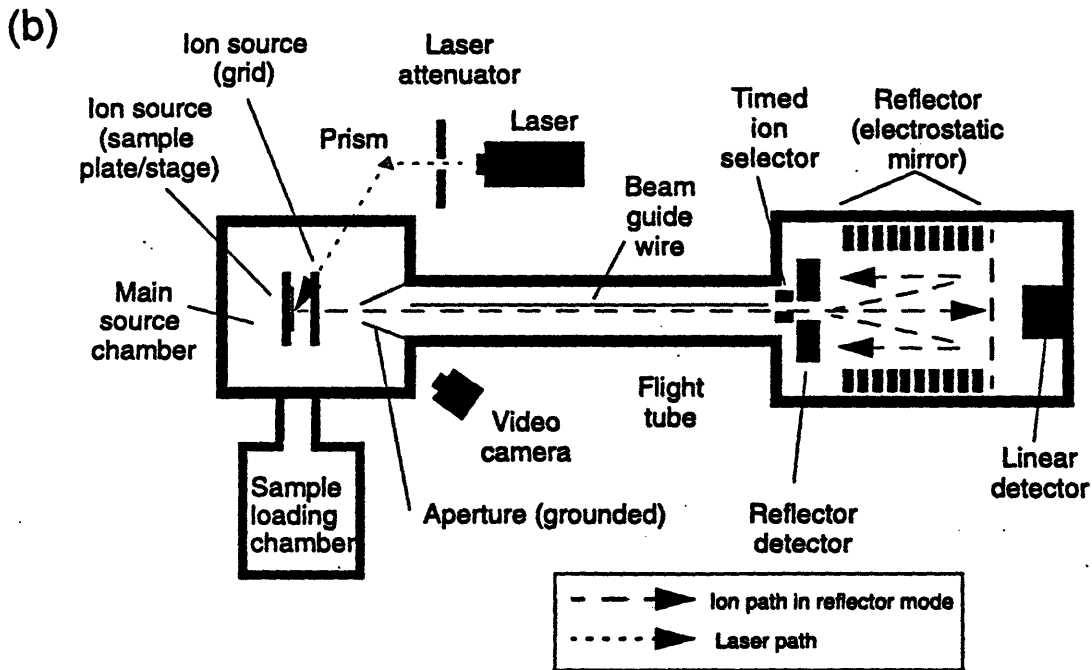
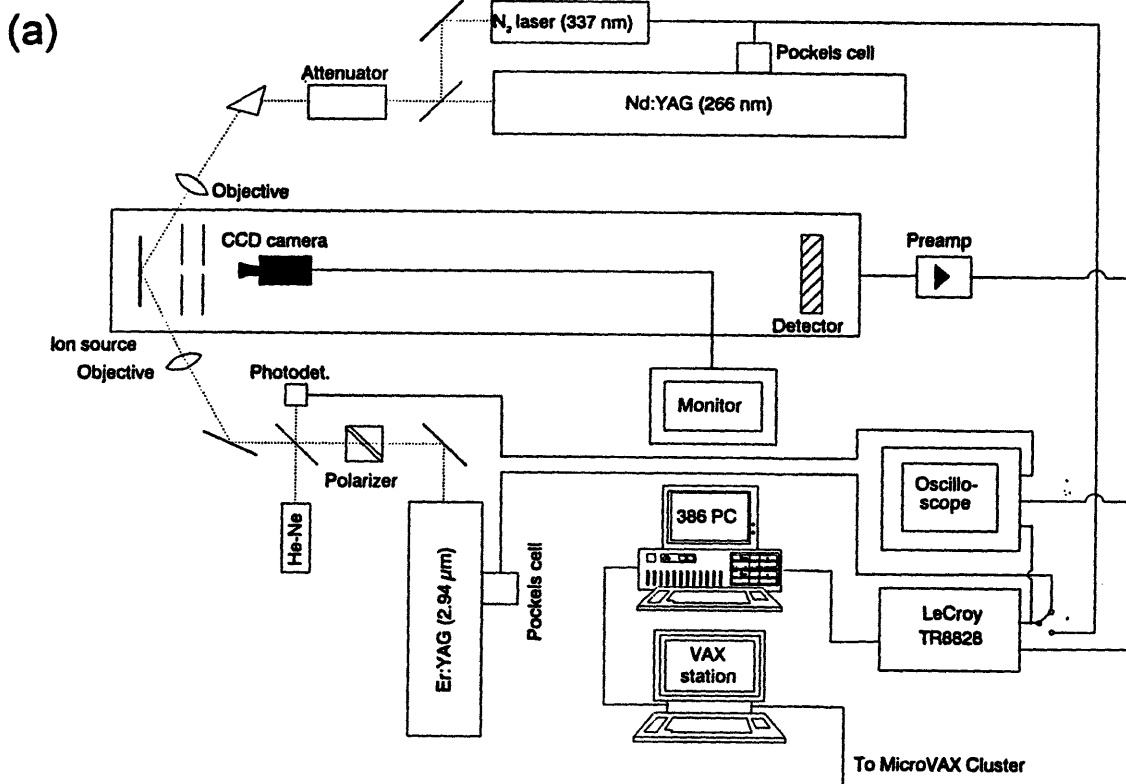


Figure I.5. Schematics of (a) VESTEC VT2000, and (b) PerSeptive Voyager-Elite MALDI time-of-flight mass spectrometer.

molecules. In contrast, the Voyager Elite reflectron mass spectrometer is equipped only a N<sub>2</sub> laser, but it can be used both in the linear and reflector mode and it is able to operate in a delayed extraction (DE) mode. An important feature of both the VT2000 and the Voyager Elite is a guide wire (see Figure I.5) running through the center of the flight tube. The voltage applied to the guide wire minimizes the dispersion effect from the ion source, refocuses ions on the detector, and thus increases the sensitivity. The potential applied to the guide wire in the VT2000 can be switched to +100 V<sup>a</sup> for a very short time (normally remains at -100 V) at the start of acquisition (synchronously with the trigger signal) to deflect the very abundant low mass ions from the matrix which may saturate the detector. Instead of using the guide wire, the Voyager Elite instrument achieves this by turning on the high voltage on the detector after the low mass ions have arrived. Eliminating the saturation of the detector by low mass ions greatly increases the sensitivity for detection of larger ions of interest.

In a typical linear time-of-flight mass spectrometer as shown in Figure I.5a, ions are generated in an ion source, extracted and accelerated by a static electric-field over a short distance between the planar electrodes and then transverse a field-free flight tube to the detector placed at the end of the tube. The only difference in configuration between linear and reflectron TOF mass spectrometer is an additional electrostatic mirror stage at the end of the flight tube (see Figure I.5b). This electric mirror filters out the neutral molecules and corrects time dispersion and, therefore, provides better resolution and higher mass accuracy than the linear mode. A reflectron instrument

---

<sup>a</sup>The instrument conditions discussed here refer to the positive ion mode since most proteins and peptides form a high abundance (M+H)<sup>+</sup> ions.

also provides an opportunity to measure the post-source decay (PSD) metastable ions which is described below. In a time-of-flight mass spectrometer, the packet of ions obtains a fixed kinetic energy

$$1/2mv^2 = zeEl \quad (4)$$

where  $m$  is the mass of the ion ( $m = \text{kg}$ ,  $1.66 \times 10^{-27} \text{ kg/Da}$ );  $v$  is its velocity (meter/second);  $z$  is the number of charges;  $e$  is the charge on an electron ( $e = 1.60 \times 10^{-19} \text{ coulomb}$ );  $E$  is the strength of the electric field ( $E = V/\text{meter}$ ) at a certain fixed voltage ( $V = \text{volt}$ ); and  $l$  is the distance ( $l = \text{meter}$ ) between the electrodes (from the sample probe to the accelerating grid). Because  $E = V/l$  or  $V = El$ , equation (4) can be re-written as

$$1/2mv^2 = zeV \text{ or } v = [2zeV/m]^{1/2} \quad (5)$$

Therefore, when the ions pass through the length of the field free drift region ( $L$ ), their time-of-flight ( $t$ ) measured at the detector will be

$$t = [m/2zeV]^{1/2}L$$

$$\text{or } t = (m/z)^{1/2} \cdot L/(2eV)^{1/2} \quad (6)$$

In equation (6),  $L$  = the length of the flight tube (in the case of our instruments  $L$  is two meters),  $e = 1.60 \times 10^{-19} \text{ coulomb}$ , and  $V = \text{volt}$  (usually 25-30 kV). Therefore,  $L/(2eV)^{1/2}$  can be considered as a constant, which means that the arrival time ( $t = \text{second}$ ) for an ion ( $m = \text{kg}$ ) carrying a charge ( $z$ ) at the detector is proportional to  $(m/z)^{1/2}$  (note: this mass-to-charge ratio is presented as the x-axis in a conventional mass spectrum). Since  $z$  for the  $(M+H)^+$  ion is always one,  $t$  is a measure of  $m$ . Thus, the accuracy of the flight-time measurement directly corresponds to the accuracy of the resulting mass. To measure the flight-times precisely, theoretically, all the ions should start drifting exactly at the same time (extraction time) with the same initial velocity. Practically, the mass

accuracy and mass resolution are always limited by small differences in measuring the flight-times of the same mass due to the initial energy distribution, positions and time of formation of the ions. The use of a very short time (nsec) pulse laser or applying a pulsed extraction electrostatic field (reviewed by Cotter, 1994) in the ion source minimizes these problems. Because MALDI produces ions at nearly identical position (surface of sample probe) and uses a short laser pulse, it is ideally suited for a TOF mass spectrometer. The problem of shot-to-shot reproducibility in MALDI experiments also affects the mass accuracy when using external calibration, and it is thus better to use internal calibration. The ions' initial velocity distribution and the differences in extraction time can cause errors in mass measurement as well as in a relatively low mass resolution in a linear TOF mass spectrometer because the resulting broad peaks makes it more difficult to define the peak centroid. These accuracy errors and resolution limitations can be greatly reduced by using the reflectron configuration (Hillenkamp et al., 1991; Cotter, 1994; Guilhaus, 1995) and/or delayed extraction technique (Vestal et al., 1995). As mentioned above, appropriately arranging the mirror electrode voltages in a reflectron configuration can significantly correct the initial velocity distribution which leads to increase in mass resolution for all the stable ions arriving at the detector. Delayed extraction originally named "time-lag energy focusing" was developed quite early by Wiley and McLaren (1955). In this experiment, the ions are produced in a field-free region followed by applying a fast pulse accelerating field at a predetermined delay time. Applied to MALDI-TOF MS, the mass resolution ( $\Delta m/m$ ) in the mass spectrum is increased from a few hundred (linear mode) to a few thousand (delayed extraction in linear mode) (Colby et al., 1994; Brown and Lennon, 1995; Whittall and Li, 1995; Vestal et al., 1995). Another practical limitation of the mass resolution and mass accuracy is due to metastable fragmentation of the ions originally formed. Because the



fragment ions have a lower kinetic energy they have a slower velocity when emerging from the electrostatic mirror and the resolution is decreased, especially in the higher mass range (Karas et al., 1995; Cotter, 1994). Thus, while the energy focusing occurring in the ion mirror can separate the small fragments from their precursor ions (see the following section). For large molecules the difference in kinetic energy are usually not only not large enough to permit resolving the large fragments such as  $[MH-H_2O]^+$  or  $[MH-NH_3]^+$  from their precursor  $(M+H)^+$  but also results in more broadening their precursor ion peak than it is observed in the linear mode. Karas et al. (1995) observed that this metastable fragmentation is matrix dependent and an appropriately selected matrix can significantly reduce the metastable fragmentation.

#### D. Post-source decay Mass Spectra

As discussed in the above section, when it is the purpose to determine the molecular weight of a peptide from the  $m/z$  value of the corresponding  $(M+H)^+$  ion, any fragmentation in the mass spectrometer is undesirable. However, fragmentation is required in order to obtain structural information. The MALDI process and the characteristics of a reflectron TOF mass spectrometer make it possible to essentially "switch" from the recording of intact  $(M+H)^+$  ions to fragment ions. In general, three kinds of ions are produced in MALDI source: unstable ions, stable ions, and metastable ions depending on their life time. If an ion is unstable it decomposes promptly (life time of nsec or shorter) in the ion source before acceleration, while ions that arrive at the detector completely intact are stable ions (life time of msec or longer). However, any ion that dissociates after acceleration but before arrival at the detector is considered a "metastable" ion. Recently the fragmentation of these metastable ions termed "post-source decay" has been exploited to generate a fragment ion spectrum in a reflectron time-of-flight mass spectrometer (Spengler et al., 1992a, and 1992b; Kaufmann et al., 1994; Rouse et al., 1995a and 1995b; Hines et al., 1995). This methodology combines the information content of a CID spectrum with the high sensitivity (pmol level) of a MALDI-TOF system.

The basic principle of the PSD technique is to focus all the fragments by stepwise decreasing the reflector potential. Because the decomposition of the fully accelerated metastable ion  $m_1^+$  to a fragment ion  $m_2^+$  ( $m_1^+ \rightarrow m_2^+ + m_n^0$ ) occurs in the field-free region, both the precursor ion, the fragment ions, and the neutrals ( $m_n^0$ ) retain (nearly) the same velocity ( $v_1$  of the precursor ion). They

all arrive simultaneously at the detector placed at the end of the flight tube and thus appear to have the  $m/z$  value of the ion originally formed in the ion source. The peak will be slightly broadened because the fragment ions will have a slight velocity spread caused by the fragmentation process. In a reflectron TOF mass spectrometer these conditions will apply when the mirror voltage is set to zero, whereas, if the reflector potential is equal (or slightly higher) to the accelerating potential (a regular reflectron experiment), the precursor ion will be well focused on the second detector (see Figure I.5b). However, the kinetic energy ( $1/2mv^2$ ) loss in  $m_2$  will be proportional to the mass loss. *e.g.* if  $m_1$  is  $m/z$  1000 and  $m_2$  is  $m/z$  500 then the kinetic energy for  $m_2$  will be half energy of that of  $m_1$ . Therefore, fragment ions appear poorly focused and also result in an incorrect mass measurement (Yu et al., 1993). On the other hand, Tang et al. (1988) found that when stepping the mirror potential from zero to full accelerating voltage all fragment ions formed from the precursor ion will be focussed correctly on the second detector. In order to record the PSD spectrum representing fragment ions derived only from a single precursor,  $m_1$ , the latter can be selected by using a timed ion selector which allows only that precursor ion and its fragment ions to enter the ion mirror. The timed ion selector in the Voyager-Elite mass spectrometer (see Figure I.5b) is placed before the reflector in the flight tube. An appropriate electrostatic field is applied perpendicular to the flight path to deflect the ions derived from other precursors, and is timed to turn off for a very short time segment to only allow the selected precursor ions to pass through. This feature is analogous to the precursor ion selection by MS-1 in a MS/MS system and thus allows the recording of PSD spectra of a component in a mixture. Presently this "precursor resolution" is about  $\pm 0.1\%$  of the mass of the precursor ion.

The fragment ion spectra of peptides in PSD have been compared with those obtained by low and high energy FAB CID (Rouse et al., 1995), and also with the PSD-CID analysis (Carr et al., 1995). It was found that the PSD spectra are similar to low energy FAB CID spectra while the PSD-CID spectra are similar to high energy FAB CID spectra. Basically, PSD fragment ions are very abundant in  $b_n$  and  $y_n$  ions as well as internal acyl ( $b$ - $y$  ions) and immonium ions (refer to Figure I.3 for the peptide fragments nomenclature), and usually less abundant  $a_n$  ( $b_n - 28$ ) and  $z_n$  ( $y_n - 17$ ) ions. In PSD spectra, abundant signals of  $b_{n-17}$  ( $\text{NH}_3$ ),  $a_{n-17}$ ,  $b_{n-1}+18$  ( $\text{H}_2\text{O}$ ), and internal acyl ion - 28 (CO) ions are also often observed. In contrast, low energy FAB CID usually produces fewer internal ions and almost no immonium ions (Rouse et al., 1995). The biggest difference between PSD and high energy CID is the absence of fragment ions due to additional side-chain cleavage ( $w_n$ ,  $v_n$ , and  $d_n$  ions) observed in the latter case (both high energy FAB CID and PSD-CID spectra). As discussed in Section I.B.5 these ions are particularly important in distinguishing the isomeric amino acids leucine and isoleucine. The major problem with PSD is that the quality of the spectrum depends on the particular peptide features such as the amino acid composition, sequence, or even its three-dimensional structure. This problem appeared when PSD was applied in this work, as it was found that in the mass range about  $m/z$  900-1600 (which usually is the optimal size for PSD analysis) some peptides produced a satisfactory PSD spectrum while others give a poor PSD spectrum or even none at all. For this reason, the PSD-CID methodology which has not yet been used much, may be very promising for peptide sequencing and similar problems.

## II. Determination of *N*-linked glycosylation in Yeast external invertase

### A. Introduction

Invertase (EC 3.2.1.26,  $\beta$ -D-fructofuranoside fructohydrolase) from *Saccharomyces cerevisiae* is an enzyme that catalyzes the hydrolysis of sucrose into fructose and glucose. This enzyme contains 14 potential *N*-glycosylation sites (sequons) (Taussig and Carlson, 1983) and exists in a glycosylated form (external invertase) with 50% of its mass as polymannan (Neumann and Lampen, 1967; Gascon et al., 1968) and a non-glycosylated internal form (Gascon and Lampen, 1968). Since yeast can provide an alternative system for protein glycosylation that is similar to mammalian systems (Kukuruzinska et al., 1987), external invertase has long been used as an ideal model for the study of the function of oligosaccharides in glycoproteins (Tarentino et al., 1974b; Trimble and Maley, 1977; Chu et al., 1978; Brown et al., 1979; Takegawa et al., 1990; Reddy et al., 1990) and for studies on glycoprotein synthesis (Esmon et al., 1981; Novick et al., 1981; Byrd et al., 1982; Trimble et al., 1983). External invertase was therefore selected in this thesis as a model for the study of glycoprotein structure by MALDI MS and related techniques.

The amino acid sequence of invertase was initially established from the corresponding DNA sequence (Taussig and Carlson, 1983). It was later verified by mass spectrometry and one amino acid was corrected (Reddy et al., 1988). The same work revealed that 13 of 14 potential sequons are either fully or partially glycosylated, which agrees with the average value of about 9-10 oligosaccharides per subunit (Trimble and Maley, 1977). The glycosylation at four sequons had

been reported to be long chain oligosaccharides ( $\text{Man}_{>50}\text{GlcNAc}_2$ ) and the other nine sequons to contain short chain oligosaccharides ( $\text{Man}_{8-14}\text{GlcNAc}_2$ ), while the remaining sequon is not glycosylated at all (Reddy et al., 1988; Ziegler et al., 1988). The conformational structure of the short chain oligosaccharides ( $\text{Man}_{8-13}$ ) isolated from invertase has been assigned by high-field one-dimensional  $^1\text{H}$  NMR spectroscopy combined with  $\alpha$ -1,2-linkage-specific mannosidase digestion (Trimble and Atkinson, 1986). It was suggested that  $\text{Man}_8\text{GlcNAc}_2$  is the smallest species and the only trimming intermediate during the biosynthesis of this glycoprotein (Byrd et al., 1982; Trimble and Atkinson, 1986).

Although invertase has been studied widely, its complete structure has not yet been characterized precisely. In fact, different average molecular weights of external invertase were reported, with estimates ranging from 94 kDa to 135 kDa (Gascon and Lampen, 1968; Trimble and Maley, 1977; Ziegler et al., 1988; Takegawa et al., 1991). The size of the oligosaccharide at each glycosylation site was determined only approximately (Ziegler et al., 1988) and no glycosylation distribution information has yet been shown. Mass spectrometry provides a precise way to determine the molecular weight with relatively high sensitivity and resolution. A method which combines enzymatic digestion and fast atom bombardment mass spectrometry (FAB MS) has been successfully employed for determining the glycosylation pattern for a glycoprotein (Carr et al., 1990). More recently, matrix-assisted laser desorption ionization time-of-flight mass spectrometry has been developed in Hillenkamp's laboratory (Karas et al., 1989; Hillenkamp et al., 1991) to generate molecular weight information up to 200 kDa and beyond. This technique provides a significantly higher sensitivity than FAB MS, and thus has become the method of choice for the

analysis of glycoproteins (Siegel et al., 1991; Huberty et al., 1993).

In this study, the average molecular weight and the complex heterogeneous glycosylation profile of intact yeast external invertase was determined by MALDI-TOF MS. The distribution of the oligosaccharides at each glycosylation site was established by proteolytic digestion in conjunction with MALDI-TOF MS. The degree of glycosylation at each sequon was also estimated. Although our results generally agree with the published data (Ziegler et al., 1988), some differences have been found and much more detailed information has been established.

By comparing the mass of proteolytic peptides determined from MALDI-TOF MS with those of peptides predicted by the COMPOST program (Papayannopoulos and Biemann, 1991) and by using the post-source decay (PSD) technique (Spengler et al., 1992a; 1992b), the amino acid sequence originally derived from the DNA sequence has been confirmed but three of the amino acids in the sequence were found to be partially or fully replaced by other amino acids.

## **B. Results and Discussion**

The basic strategy for the study of the glycosylation of external invertase required the isolation of each individual sequon in the form of a glycopeptide generated by proteolytic digestion and HPLC separation. The resulting fractions were examined by MALDI-TOF MS to identify those containing glycopeptides. These were then characterized by endoglycosidase deglycosylation experiments, in some cases confirmed by the PSD spectrum of the resulting peptide. To generate peptides containing only one sequon, one could, theoretically, select an enzyme for the primary digestion and subsequently employ one or more enzyme sub-digestions as necessary to achieve this goal. In reality, however, it often happened that some peptides and/or glycopeptides could not be observed after the initial enzyme digestion. This was possibly because of their irreversible adsorption to the HPLC column, or due to their low ionization efficiency by MALDI, or because the signal from these peptides might have been suppressed by other peptides present in the same fraction. Therefore, three proteolytic enzymes were selected for the primary digestion in three series of parallel experiments to obtain complementary data for a complete result. Details will be described below.

Prior to the enzymatic digestion, the average molecular weight of the intact external invertase was determined to be 96 k Da by MALDI-TOF MS. The signal consisted of a broad peak with a nearly symmetrical shape covering the mass range from 77 to 122 k Da (Figure II.1). The molecular weight of wholly deglycosylated external invertase resulting from PNGase F digestion was found to be 58,696 Da, in good agreement with the values previously reported (Taussig and Carlson, 1983;



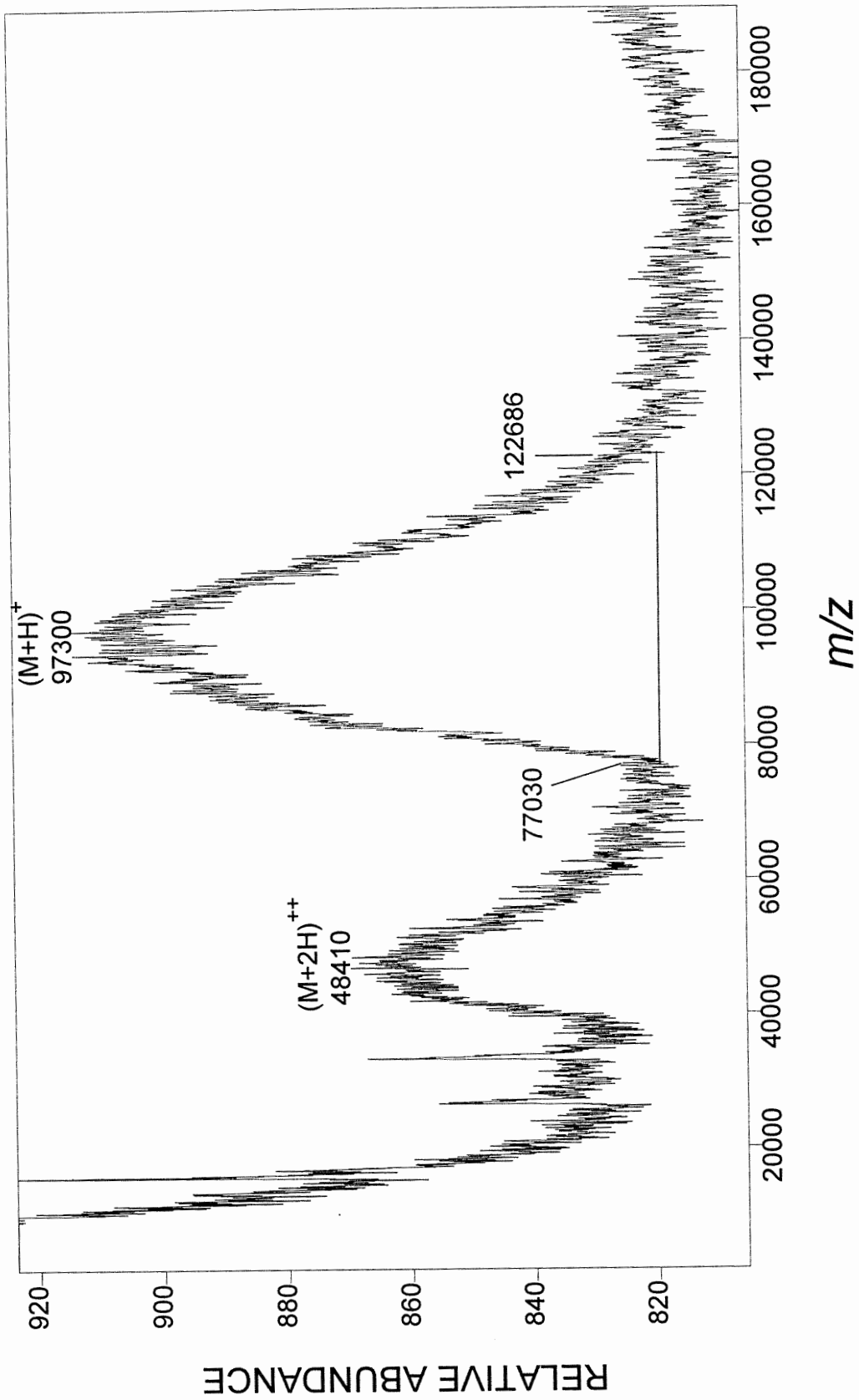


Figure II.1. MALDI-TOF mass spectrum of intact yeast external invertase.

Reddy et al., 1988).

In this work, three separate parallel experiments were carried out to generate the complementary information. Intact external invertase was digested with trypsin, Lys-C, and Asp-N. Each of these digests were fractionated by HPLC. The HPLC traces, indicating the fractions collected, are shown in Figures II.2, II.3, and II.4, respectively. It should be noted that the proteolytic glycopeptides are usually eluted as a broad peak with the more heavily glycosylated glycoforms eluting first (Hemling et al., 1990). Therefore, some of the glycopeptides (especially the "wide range glycosylated" ones) encompassed two or three HPLC fractions. The molecular weights of the peptides in the digests detected by MALDI-TOF MS were matched with the molecular weights predicted from the published amino acid sequence deduced from the DNA sequence (Taussig and Carlson, 1983). The predicted peptide fragments resulting from a specific enzyme digestion were generated by the computer program "COMPOST" (Papayannopoulos and Biemann, 1991), which helps to quickly identify the peptides. For the identification of the glycopeptides, the glycans were first removed by treatment with either PNGase F or Endo H and the masses of the deglycosylated peptide were then matched with those predicted by "COMPOST" to identify the original glycopeptide and locate the unique sequon.

As discussed in Chapter I, glycoproteins often contain glycans of various sizes at each sequon. In the case of external invertase (a high mannose-type *N*-linked glycoprotein), the signals due to different lengths of glycan chains at a sequon will be spaced apart by one mannose (162 Da). Therefore, one can expect to observe the signal for a glycopeptide in a mass spectrum as a series of

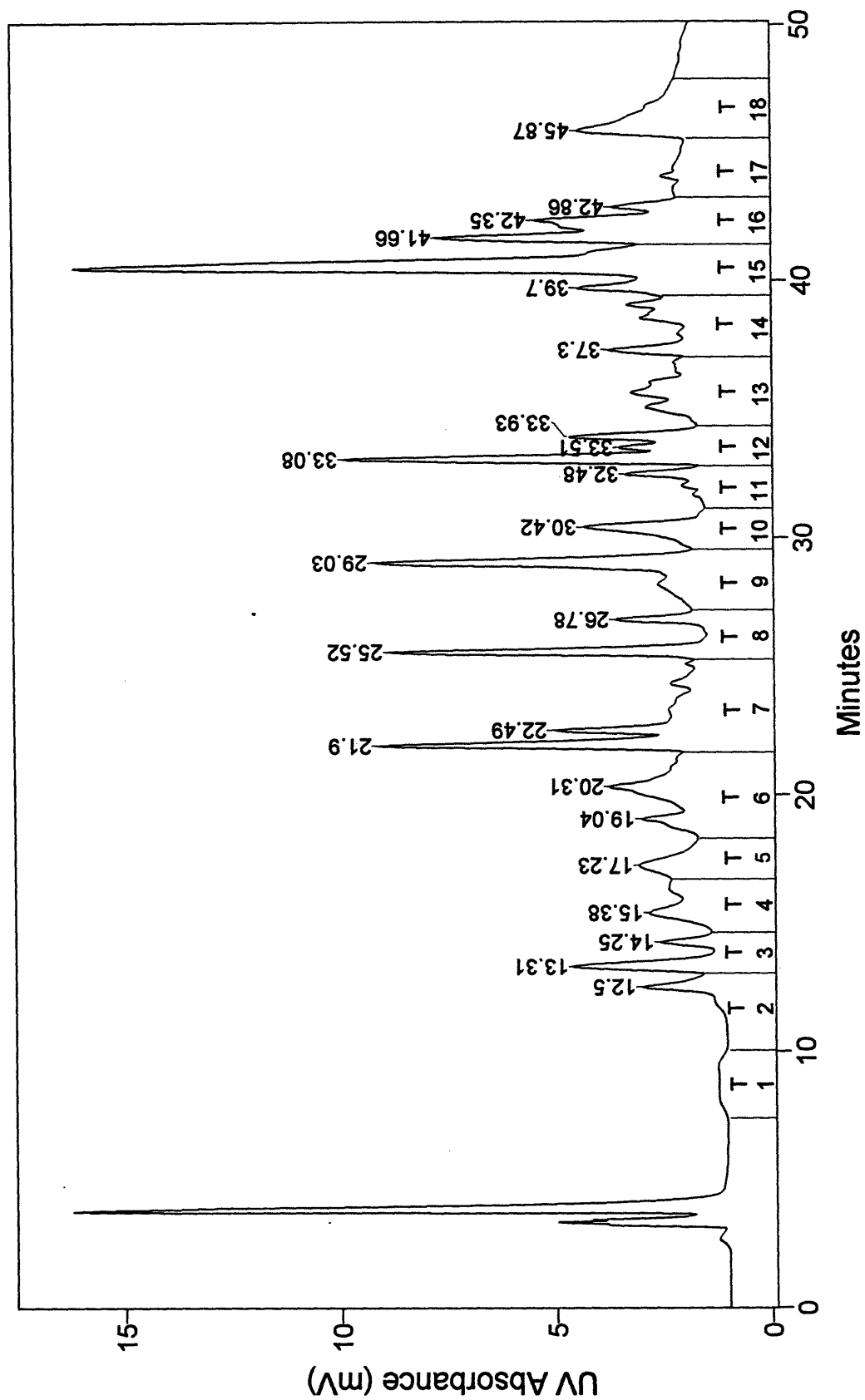


Figure II.2. HPLC trace of the tryptic digest of external invertase. The fractions collected are marked. Numbers printed vertically are retention times in minutes.

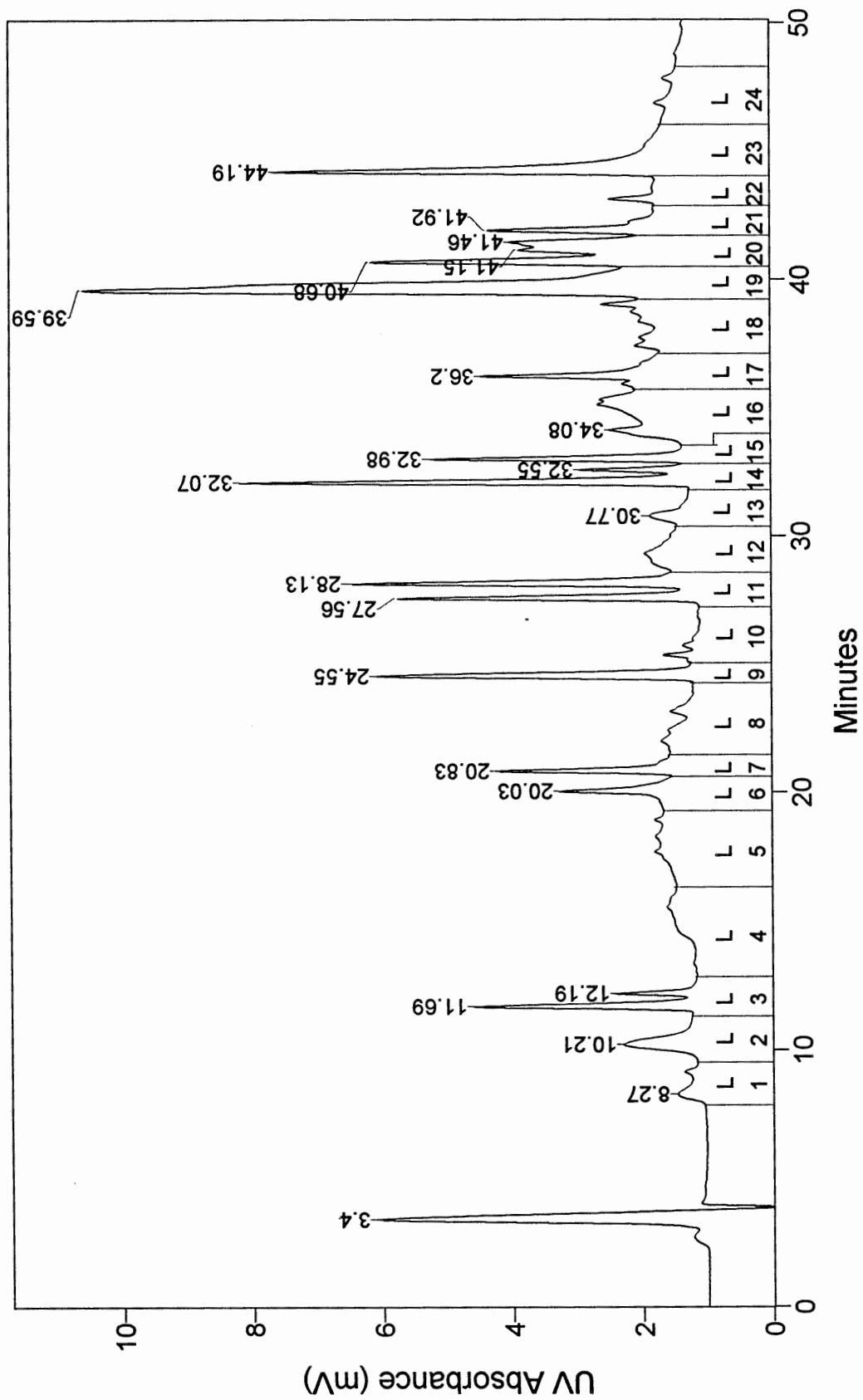


Figure II.3. HPLC trace of the Lys-C digest of external invertase. The fractions collected are marked.

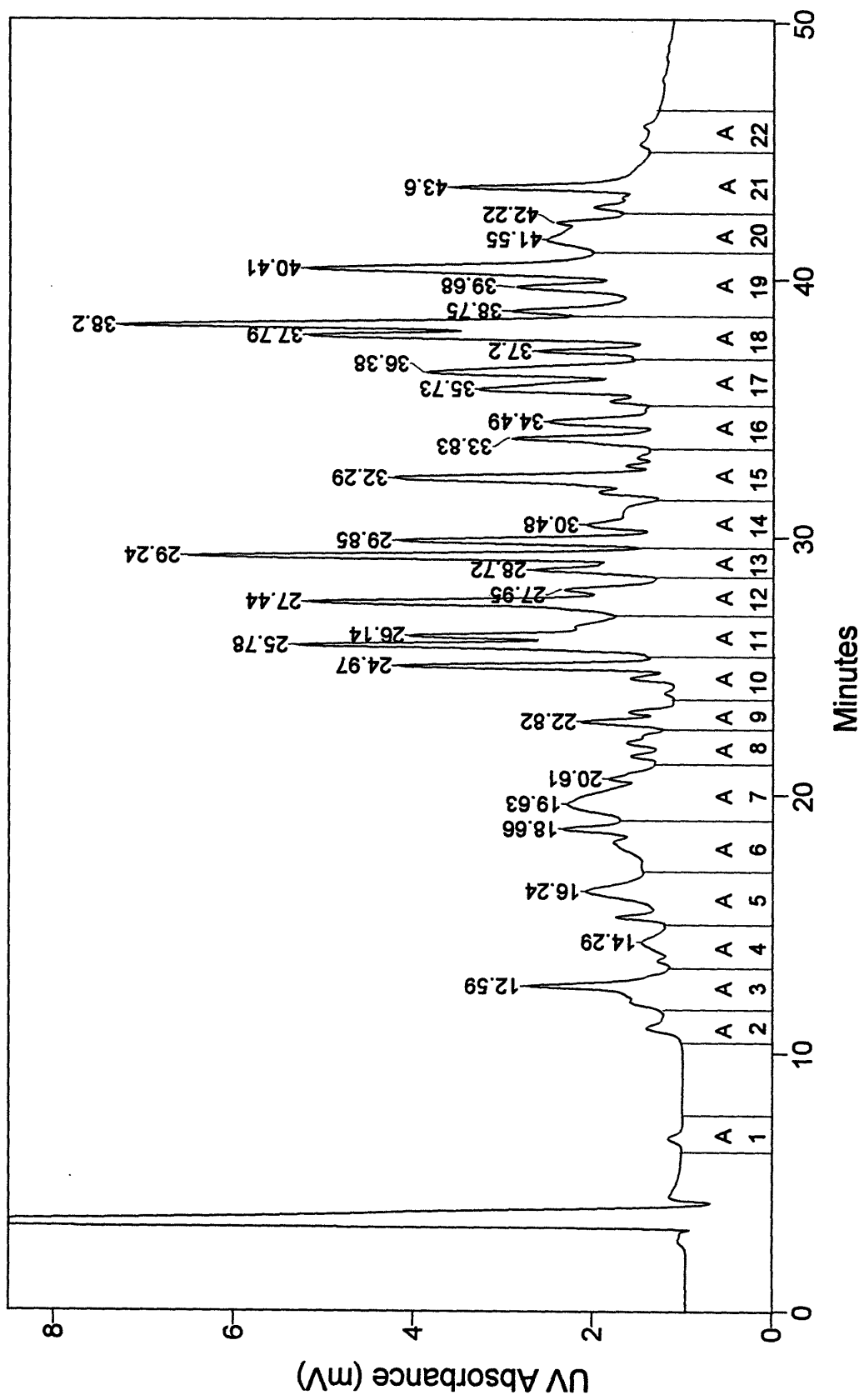


Figure II.4. HPLC trace of the Asp-N digest of external invertase. The fractions collected are marked.

peaks differing by 162 Da, defining the length of the various polymannan oligosaccharides. The distribution and relative abundance of the components of this series will be referred to as "glycopatterns" and this term will be used throughout this thesis. The size of the oligosaccharide portion of an observed glycopeptide can be identified indirectly by detecting the mass of the corresponding deglycosylated peptide resulting from endoglycosidase PNGase F and/or Endo H digestion.

In order to locate the glycosylation site of a glycopeptide, the mass of the deglycosylated peptide was matched to the predicted peptide containing one or more glycosylation site(s). It is well known that PNGase F hydrolyses the glycosidic bond between asparagine and the first *N*-acetylglucosamine (GlcNAc), thus converting glycosylated asparagine to aspartic acid (Plummer et al., 1984) and producing a peptide that is one dalton heavier than the unglycosylated peptide. In contrast, Endo H cleaves the bond between the first and second *N*-acetylglucosamine in the core region of the oligosaccharide leaving the first GlcNAc attached to asparagine (i.e. an Asn-glycan is converted to Asn-GlcNAc) (Tarentino and Maley, 1974). The resulting deglycosylated peptide is therefore 203 Da heavier than the corresponding non-glycosylated peptide. Based on this knowledge, one can infer the number of sequons in the unknown glycopeptide by calculating the mass difference between the peptide deglycosylated by PNGase F digestion and the peptide deglycosylated by Endo H digestion. In the case of external invertase, this calculation is still necessary to identify a glycopeptide containing two or more sequons even though the amino acid sequence and the positions of the sequons are known. When two or more sequons were found to remain in one peptide fragment, a second endoprotease digestion was required to isolate these

sequons for further analysis. When the size of the oligosaccharide chain was found to be different from the published data (Ziegler et al., 1988) or the mass assignment was outside the accuracy of MALDI, another enzyme digestion or MALDI-PSD MS were carried out to verify the result and to identify the modified amino acid that caused the mass difference. The determination of the occupancy of the 14 sequons of external invertase and the size range of the attached glycans are described in the following.

#### 1. *Trypsin Digestion of native intact protein:*

First, intact external invertase was digested with trypsin followed by HPLC fractionation, and 18 HPLC fractions were collected (Figure II.2). Each fraction was screened by MALDI-TOF mass spectrometry to search for glycopatterns. It was found that fractions T6, T7, T8, T9, T10, T15, T16, T17, and T18 exhibited one or more glycopattern(s).

Figure II.5a represents the MALDI spectrum of fraction T6 which exhibited a signal at  $m/z$  549.7 that was identified to be tryptic peptide 507-510 which does not contain a sequon, together with two glycopatterns in the higher mass region. The major glycopattern consisting of 54 peaks spaced about 162 Da apart in the mass range of 2,800-12,000 Da (shown expanded in Figure II.5b), indicates a "wide range glycosylation" (this term will be used throughout this thesis to denote the presence of oligosaccharide chains ranging from a few to up to 50+ mannoses). To identify these

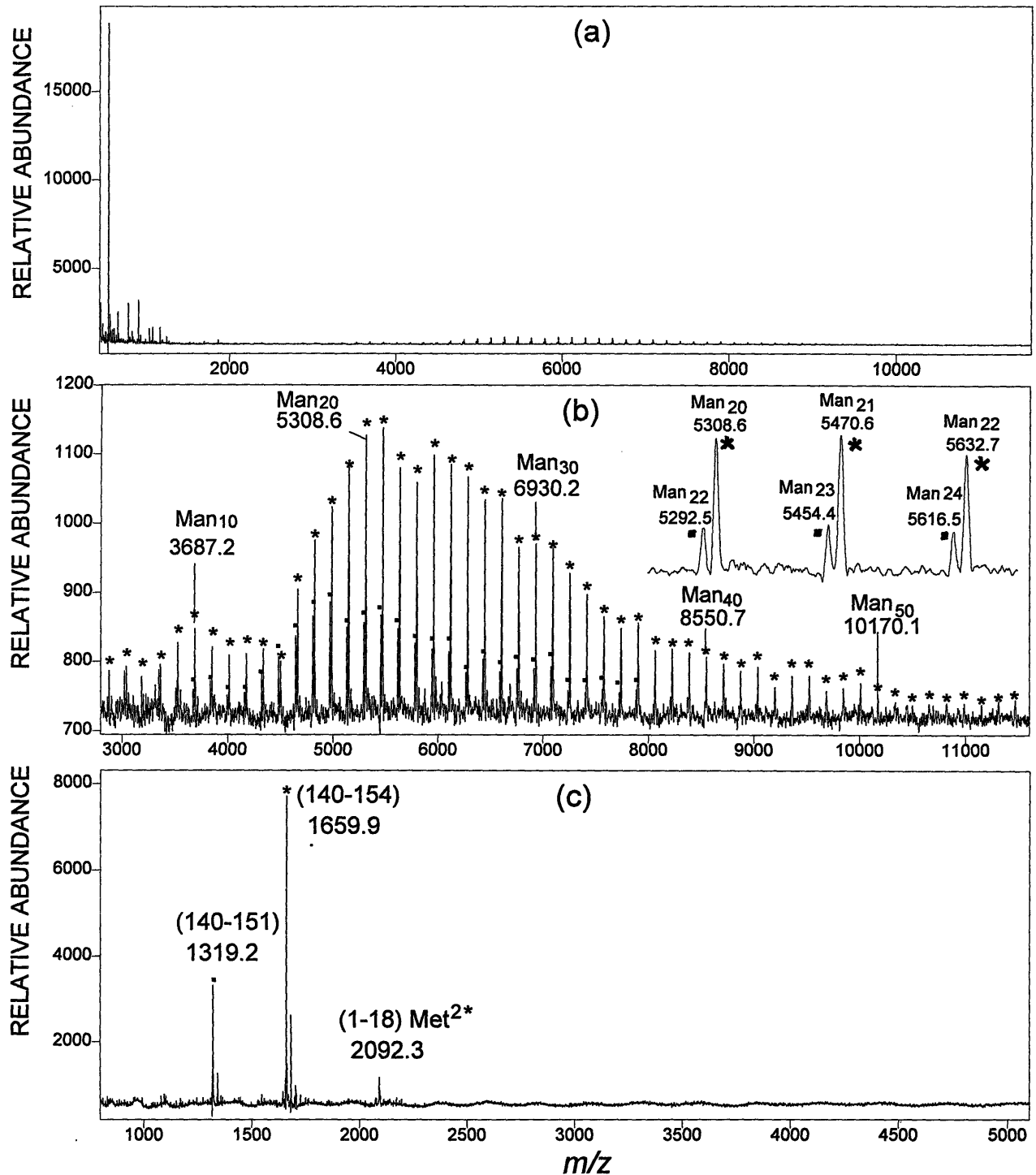


Figure II.5. (a) MALDI spectrum of HPLC fraction T6 of the trypsin digest of external invertase. (b) The region  $m/z$  2,800-12,000 of (a) expanded in the Y-axis; inset: region  $m/z$  5,200-5,700 expanded in the X-axis to show the presence of two glycopeptides (see text). (c) MALDI spectrum of T6 after deglycosylation with PNGase F. Met<sup>2\*</sup> indicates the oxidized Met at position 2.



glycopeptide(s), fraction T6 was digested with PNGase F to remove the oligosaccharide chains. The resulting MALDI-TOF MS showed that these multiplets collapsed to three peaks (Figure II.5c). The weak signal at  $m/z$  2092.3<sup>b</sup>, must originate from a glycopeptide the signals of which were too weak to be detected. However, its mass was 16 Da heavier than the predicted tryptic peptide 1-18 (which contains sequon 1), indicating an artifact in which Met<sup>2</sup> is oxidized to its sulfoxide. The peaks at  $m/z$  1319.2 and 1659.9 (Figure II.5c) correspond to the tryptic peptides 140-151(calc. 1319.5)<sup>c</sup> and 140-154 (calc. 1659.8), respectively. These peptides both encompass sequon 7 (Asn<sup>146</sup>-Ser-Thr). The former was produced by cleavage at Arg<sup>151</sup> and the latter at Lys<sup>154</sup>. In addition, when fraction T6 was digested with Endo H, the corresponding deglycosylated peptide signals were observed at  $m/z$  1520.9 and 1861.1 (spectrum not shown) which also matched the N-acetylglucosamine (GlcNAc) containing peptides 140-151 (calc. 1521.7) and 140-154 (calc. 1862.1), respectively. The difference of about 202 Da between the PNGase F peptides and the relative Endo H peptides indicated the presence of one glycosylation site (sequon 7) (see Chapter I). Again, the low intensity peak at  $m/z$  2293.3 was found after Endo H deglycosylation. This mass was about 202 Da heavier than the PNGase F peptide (2092.3 Da) confirming the presence of glycopeptide 1-18 in which Met<sup>2</sup> is oxidized as discussed above.

A detailed inspection of the MALDI spectrum of fraction T6 (expanded in Figure II.5b) revealed that the series of peaks of the major glycopattern (labeled with the symbol "★") belong to

---

<sup>b</sup>All the  $m/z$  and mass values reported in this thesis refer to the protonated (M+H)<sup>+</sup> ions.

<sup>c</sup>The calculated mass values of the PNGase F deglycosylated peptides correspond to that of the sequence in which the Asn of the sequon is replaced by Asp.

peptide 140-154 carrying an oligosaccharide  $\text{Man}_{5-58}\text{GlcNAc}_2$ . For example, the peak at  $m/z$  5308.6 corresponds to peptide 140-154 (1658.86 Da) bearing  $\text{Man}_{20}\text{GlcNAc}_2$  (3649.18 Da)<sup>d</sup> and the next peak at  $m/z$  5470.6 corresponds to peptide 140-154 with  $\text{Man}_{21}\text{GlcNAc}_2$  (3811.32 Da) etc. The peaks which belong to the minor glycopattern (as labeled with the symbol "•") turned out to be due to the peptide 140-151 (1318.48 Da) with the same glycosylation at sequon 7. It should be pointed out that the 16 Da doublets formed by the major and the minor glycopattern (see the insert in Figure II.5b) is due to the fact that the difference in mass between the smaller peptide 140-151 and the larger peptide 140-154 is 340.38 Da, while the mass for two mannose residues is 324.28 Da, causing the 16 Da differences. For example, the peak at  $m/z$  5292.3 is the peptide 140-151 glycosylated with  $\text{Man}_{22}\text{GlcNAc}_2$  (calc. 5291.9) while the peak at  $m/z$  5308.6 is the peptide 140-154 with  $\text{Man}_{20}\text{GlcNAc}_2$  (calc. 5308.0). As expected, the profiles of these two glycopatterns are practically identical if one takes the mass differences of the two peptides into account.

The MALDI mass spectrum shown in Figure II.6a was obtained from fraction T7 (see Figure II.2). The four peaks at  $m/z$  780.0, 1041.3, 1065.4, and 1119.4 in the low mass region (insert in Figure II.6a) were identified as peptides 347-355 (calc. mass 779.9), 431-438 (calc. mass 1041.1), 439-447 (calc. mass 1065.2), and 448-456 (calc. mass 1119.2), respectively. There was also a complex glycopattern in the high mass region (Figure II.6b). The major set of 48 peaks spaced about 162 Da apart in the mass range 3,000-11,000 Da indicates another wide range type glycosylated peptide (Figure II.6b). In order to identify the corresponding peptide, PNGase F and Endo H

---

<sup>d</sup>The composition of the oligosaccharide chain is based on the knowledge that in yeast glycosylation produces only high mannose glycan.

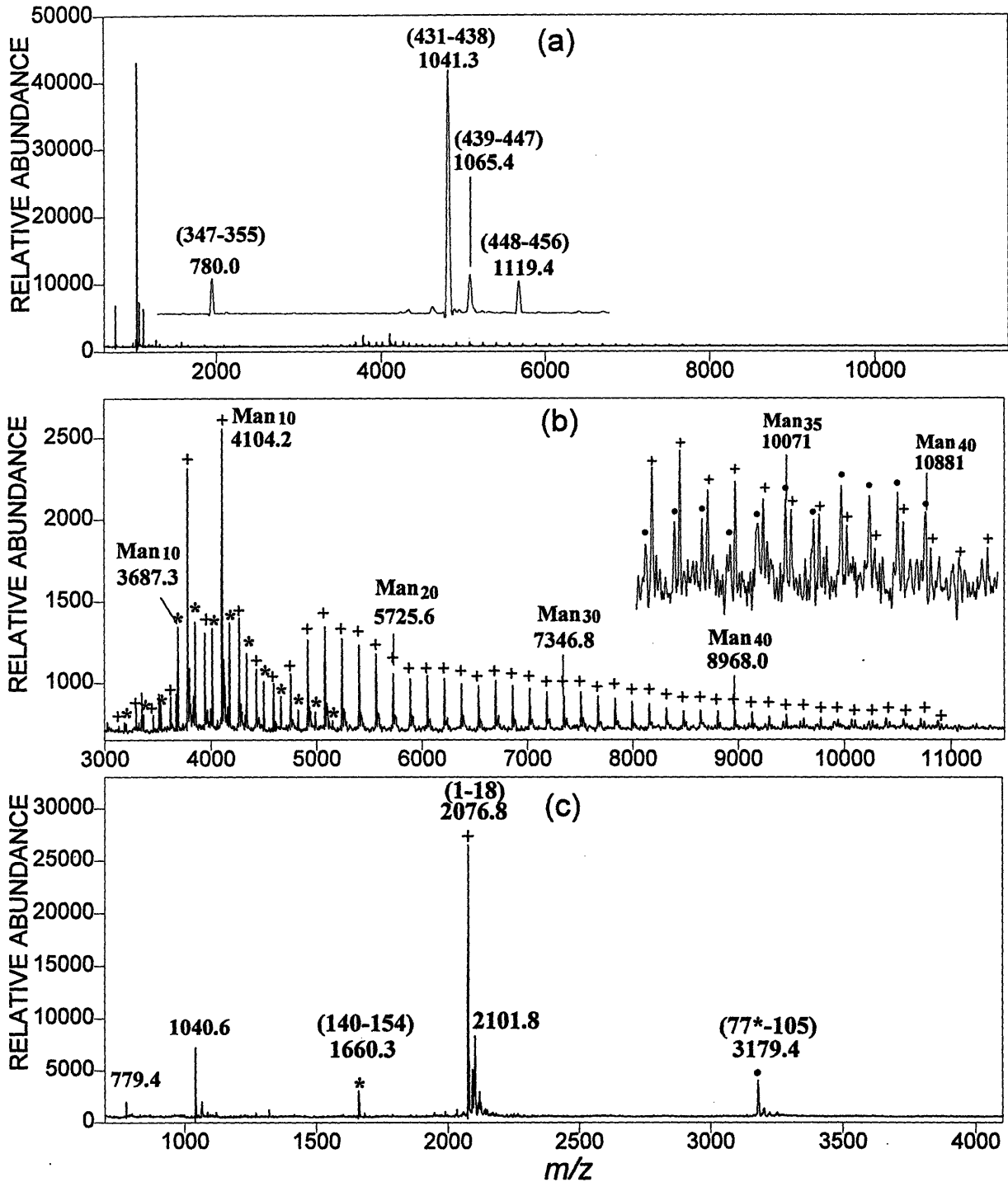


Figure II.6. (a) MALDI spectrum of HPLC fraction T7 of the trypsin digest. (b) The high mass region of (a) expanded. (c) MALDI spectrum of T7 after deglycosylation with PNGase F.

deglycosylations were carried out on fraction T7. The MALDI mass spectrum of the former reaction is shown in Figure II.6c. The major signal at  $m/z$  2076.7 perfectly matches the deglycosylated peptide 1-18 (calc. 2076.3) containing sequon 1 (Asn<sup>4</sup>-Glu-Thr). The minor signal at  $m/z$  1660.3 fits the peptide 140-154 (calc. 1659.8) containing sequon 7, and another minor signal at  $m/z$  3179.4 could not be assigned to any tryptic peptide at this point. The MALDI spectrum (not shown) of Endo H deglycosylation of fraction T7, showed a major peak at  $m/z$  2278.7 corresponding to peptide 1-18 with one GlcNAc (calc. 2278.5) and the minor peak at  $m/z$  1862.4 was the GlcNAc attached peptide 140-154. The signal at  $m/z$  3785.2 was 606 Da ( $3 \times 202$  Da) heavier than the corresponding PNGase F peptide  $(M+H)^+ = m/z$  3179.4 indicating the presence of three glycosylated sequons. However, none of the possible tryptic peptides of invertase containing three or more sequons did fit this mass (within  $\pm 0.3\%$  error). This result can only be explained as being due to a minor component of peptide 77-105 (which contains sequons 3, 4, 5, and 6 at Asn<sup>78</sup>-Asp-Ser, Asn<sup>92</sup>-Asn-Thr, Asn<sup>93</sup>-Thr-Ser, and Asn<sup>99</sup>-Asp-Thr) where Arg<sup>77</sup> is replaced by Phe (calc. 3179.3 for PNGase F digested peptide). Further experiments are required to prove this suggestion.

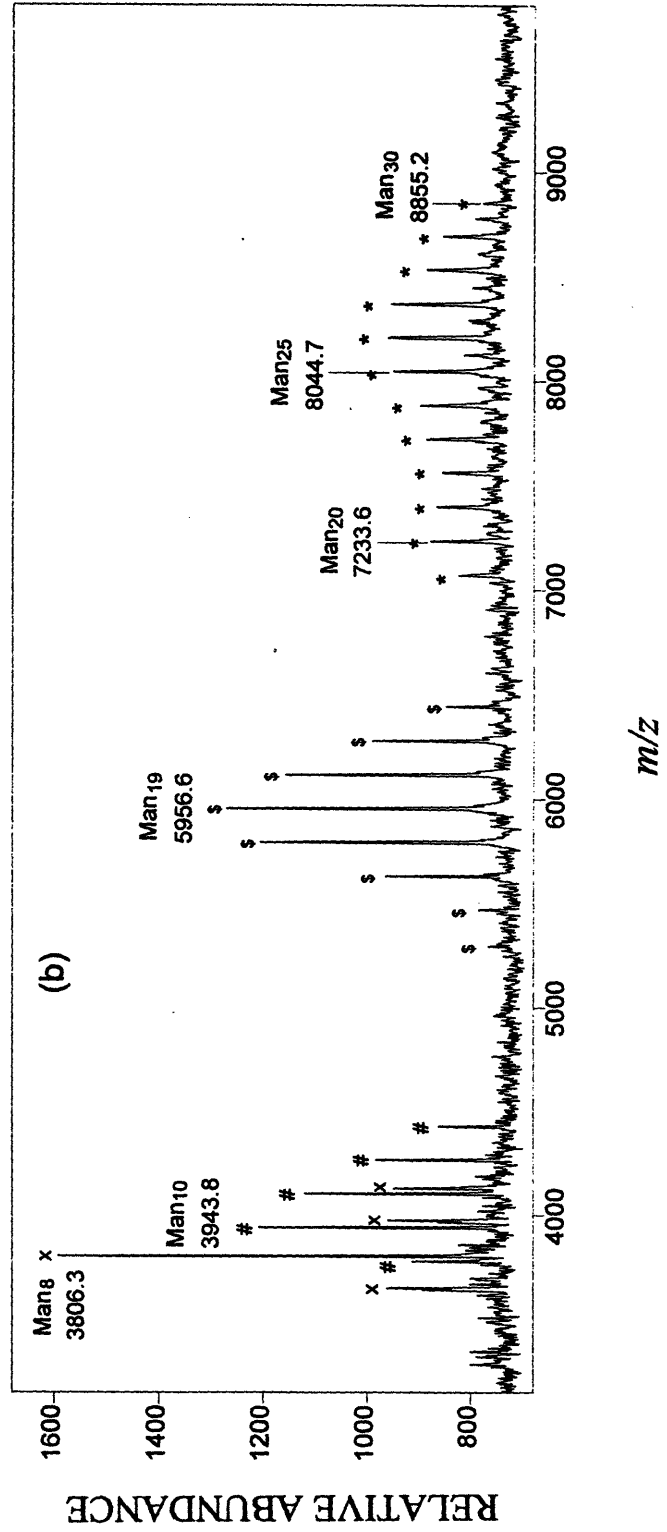
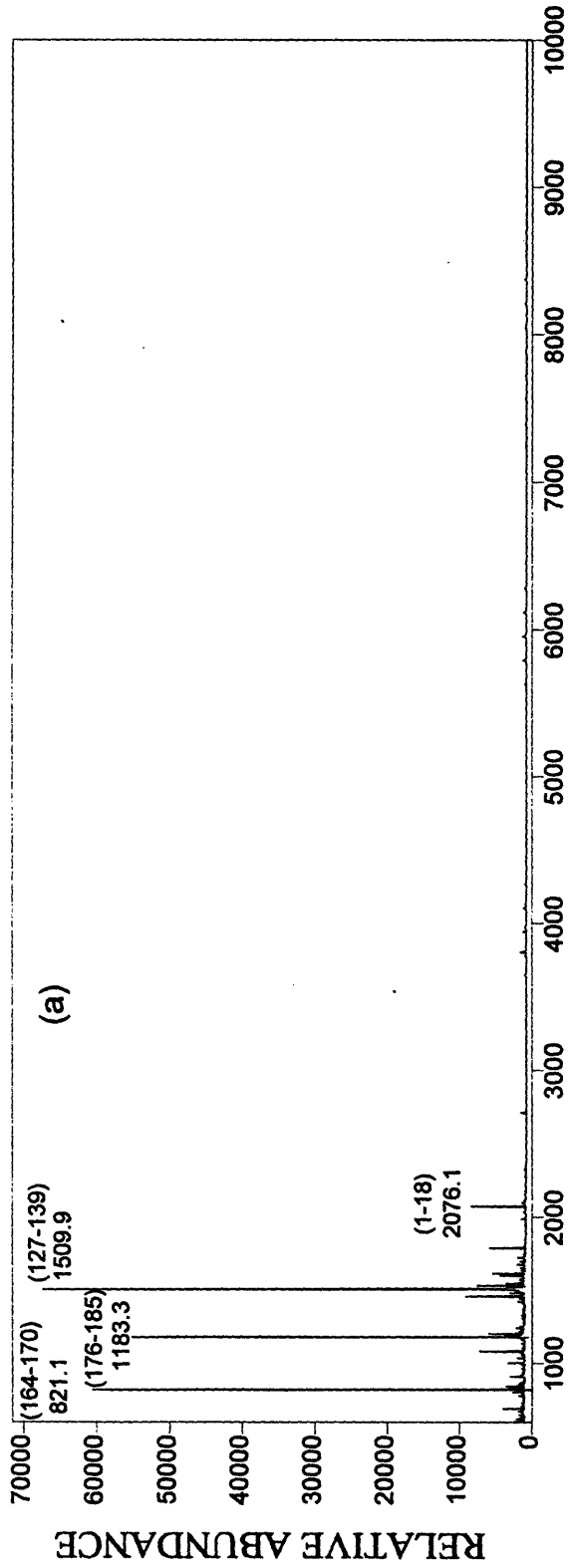
The analysis of the expanded spectrum (Figure II.6b) reveals that the major glycopattern (labeled "+") corresponds to peptide 1-18 (containing sequon 1 at Asn<sup>4</sup>) with two GlcNAc and 4 to 52 mannoses. The two major components at  $m/z$  3779.8 and 4104.2 correspond to Man<sub>8</sub>GlcNAc<sub>2</sub> (calc. 3788.8) and Man<sub>10</sub>GlcNAc<sub>2</sub> (calc. 4103.1). Another less abundant glycopattern in the mass range of 3,100-5,400 (labeled "★") is due to peptide 140-154 bearing Man<sub>7-19</sub>GlcNAc<sub>2</sub>. This represents the short mannose chain portion of glycopeptide 140-154, the major component of which eluted in fraction T6 as discussed above. In the mass range of 9,200-10,920 Da (insert in Figure

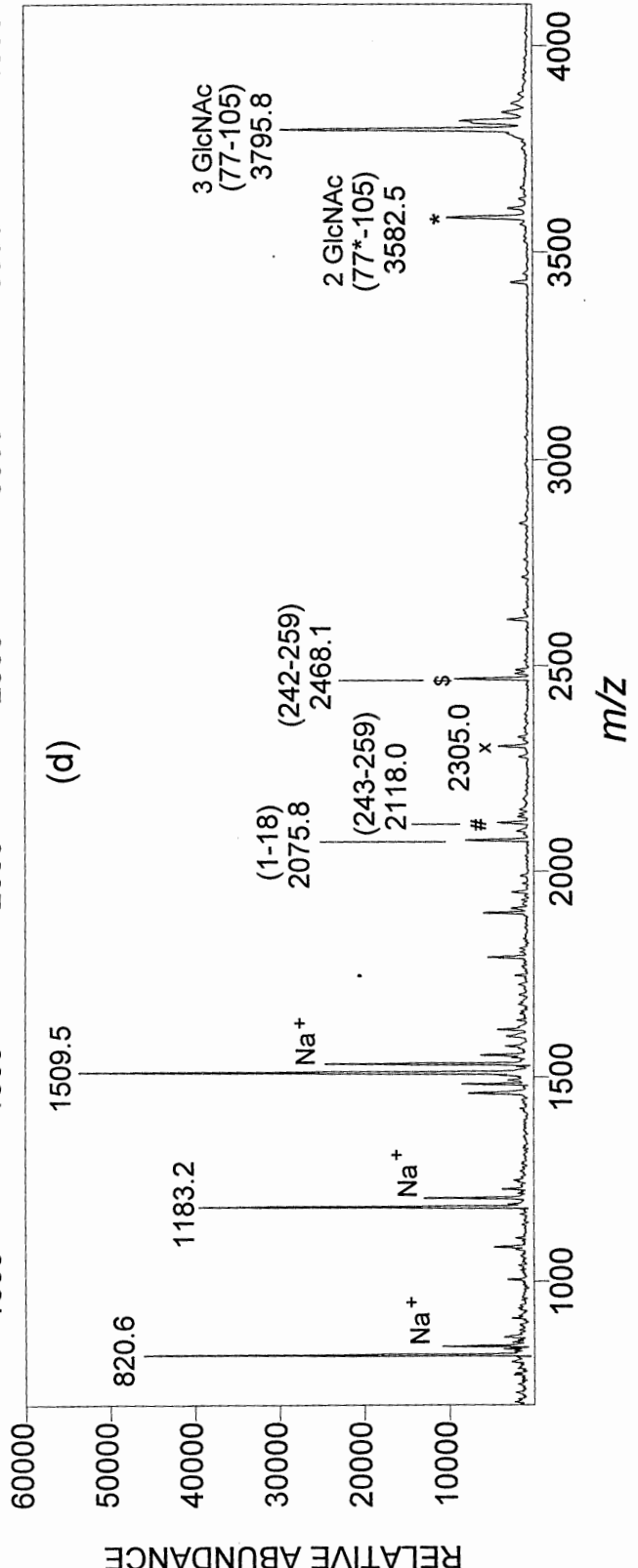
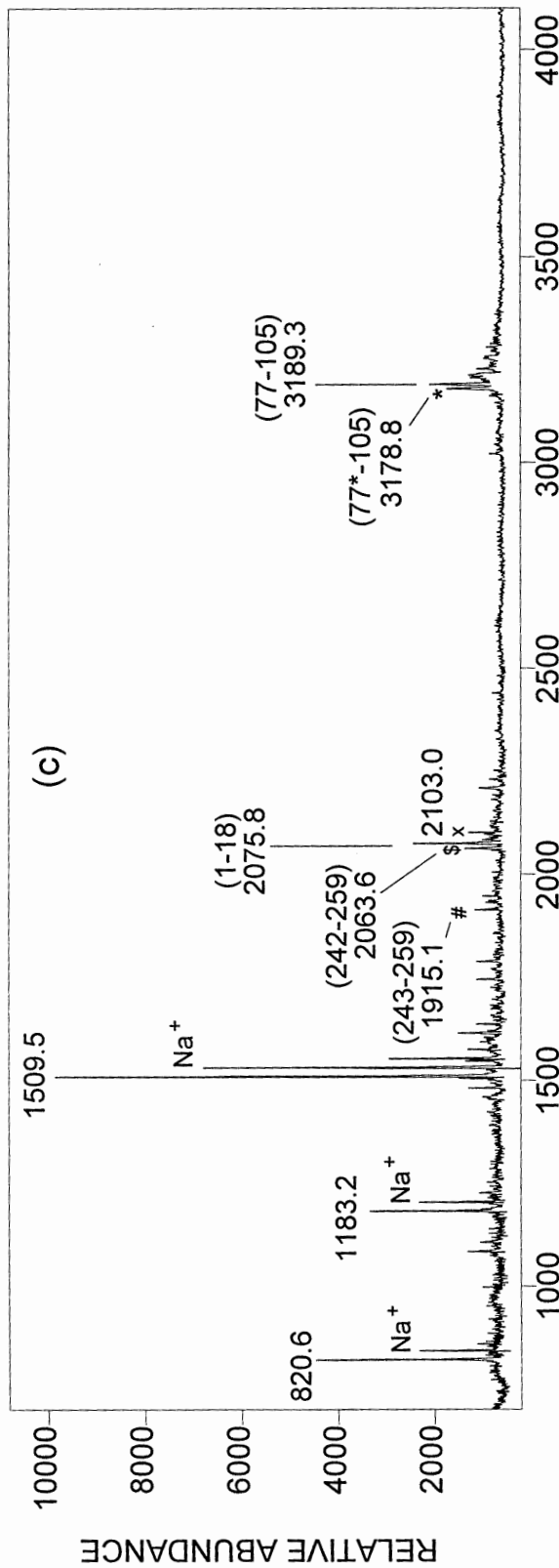
II.6b), there is yet another glycopattern, also spaced 160 Da apart (peaks marked "•"). The  $m/z$  values in this pattern corresponded to the peptide above mentioned PNGase F deglycosylated  $(M+H)^+ = 3179.4$  bearing  $\text{Man}_{31-40}\text{GlcNAc}_6$  (three occupied glycosylation sites).

The MALDI spectrum of HPLC fraction T8 of the tryptic digest of external invertase was relatively complex (Figure II.7a and 7b). The strong signals in the low mass region correspond to three peptides not containing a sequon. The peaks at  $m/z$  821.1 and 1183.4 matched the tryptic peptide 164-170 (calc. 821.0) and peptide 176-185 (calc. 1183.3) while the peak at  $m/z$  1509.9 was identified from its PSD spectrum to be peptide 127-139, a non-specific cleavage product. Another singlet peak at  $m/z$  2076.1 represents the free (unglycosylated) peptide 1-18 (calc. 2075.3). Expansion of the mass range  $m/z$  3,500-9,800 (Figure II.7b) revealed three or more glycopatterns. To identify these glycopatterns, fraction T8 was treated with PNGase F and Endo H, respectively, to reveal the corresponding deglycosylated peptide (Figures II.7.c and II.7d).

Most of the major peaks in both spectra are due to the unglycosylated peptides already identified (Figure II.7a). The PNGase F deglycosylated peptide  $(M+H)^+ = m/z$  3189.3 matched the peptide 77-105 (calc. 3188.3 for three-, or 3189.3 for four occupied sequons). By comparing the corresponding Endo H peptide  $(M+H)^+ = 3795.8$  (Figure II.7d) that was 606 Da ( $3 \times 202$  Da) heavier than the PNGase F peptide, one can conclude that peptide 77-105 (contains sequons 3, 4, 5, and 6) was glycosylated with three-, not four, oligosaccharide chains. Since the main component of this glycopeptide was found in fraction T9 (see below), the minor component containing longer oligosaccharide chains was expected to be detected in fraction T8 (as described earlier). However,

Figure II.7 (two pages). (a) MALDI spectrum of HPLC fraction T8 of the tryptic digest. (b) The high mass region of (a) expanded in the Y-axis. (c) and (d) The MALDI spectra of fraction T8 after PNGase F and Endo H deglycosylation, respectively.







none of the glycopatterns from fraction T8 (Figure II.7b) matched this peptide with three high mannose oligosaccharide chains. The corresponding glycopeptide would be expected to appear in the mass range of  $m/z$  1,1000-12,000 (e.g. calc. 10,241.6 for  $\text{Man}_{36}\text{GlcNAc}_6$  and 11,700.8 for  $\text{Man}_{45}\text{GlcNAc}_6$ ), but they may be too weak to be detected. Correlation of the PNGase F peptide  $(\text{M}+\text{H})^+ = m/z$  3178.8 with the Endo H peptide  $(\text{M}+\text{H})^+ = m/z$  3582.5 indicated that in peptide 77\*-105 (calc. 3178.3, where Arg<sup>77</sup> was replaced with Phe) that two of the four sequons are occupied (the same peptide glycosylated with three oligosaccharide chains has been found in fraction T7 as discussed above). Its corresponding glycosylated peptide (two oligosaccharide chains which total  $\text{Man}_{19-31}\text{GlcNAc}_4$ ) was found as a glycopattern in the mass range of 7,000-9,000 Da in Figure II.7b (labeled "★").

Another PNGase F peptide  $(\text{M}+\text{H})^+ = m/z$  2063.6 (Figure II.7c) did not match any predicted tryptic peptide and thus must be due to non-specific cleavage. From the corresponding Endo H peptide  $(\text{M}+\text{H})^+ = m/z$  2468.1 (Figure II.7d), one can conclude that this peptide contains two glycans. Using the "COMPOST" algorithm and including this information, peptide 242-259 (which contains sequons 8 and 9 at Asn<sup>247</sup>-Gly-Thr and Asn<sup>256</sup>-Gln-Ser) was the only one that fit all requirements. It must have been formed by cleaving at the C-terminal side of Tyr<sup>241</sup> (a chymotryptic specific site) and at the C-terminal side of Arg<sup>259</sup> (a typical tryptic cleavage). To confirm this result, PNGase F peptide  $(\text{M}+\text{H})^+ = 2063.6$  was isolated from the others by HPLC and further digested with chymotrypsin. A smaller chymotryptic peptide  $(\text{M}+\text{H})^+ = m/z$  1113.5 was obtained which matched the chymotryptic peptide 242-251. The PSD spectrum of this ion confirmed the sequence, 242FVGSFDGTHF251 (Figure II.8). Note the underlined three amino acids indicate the position of

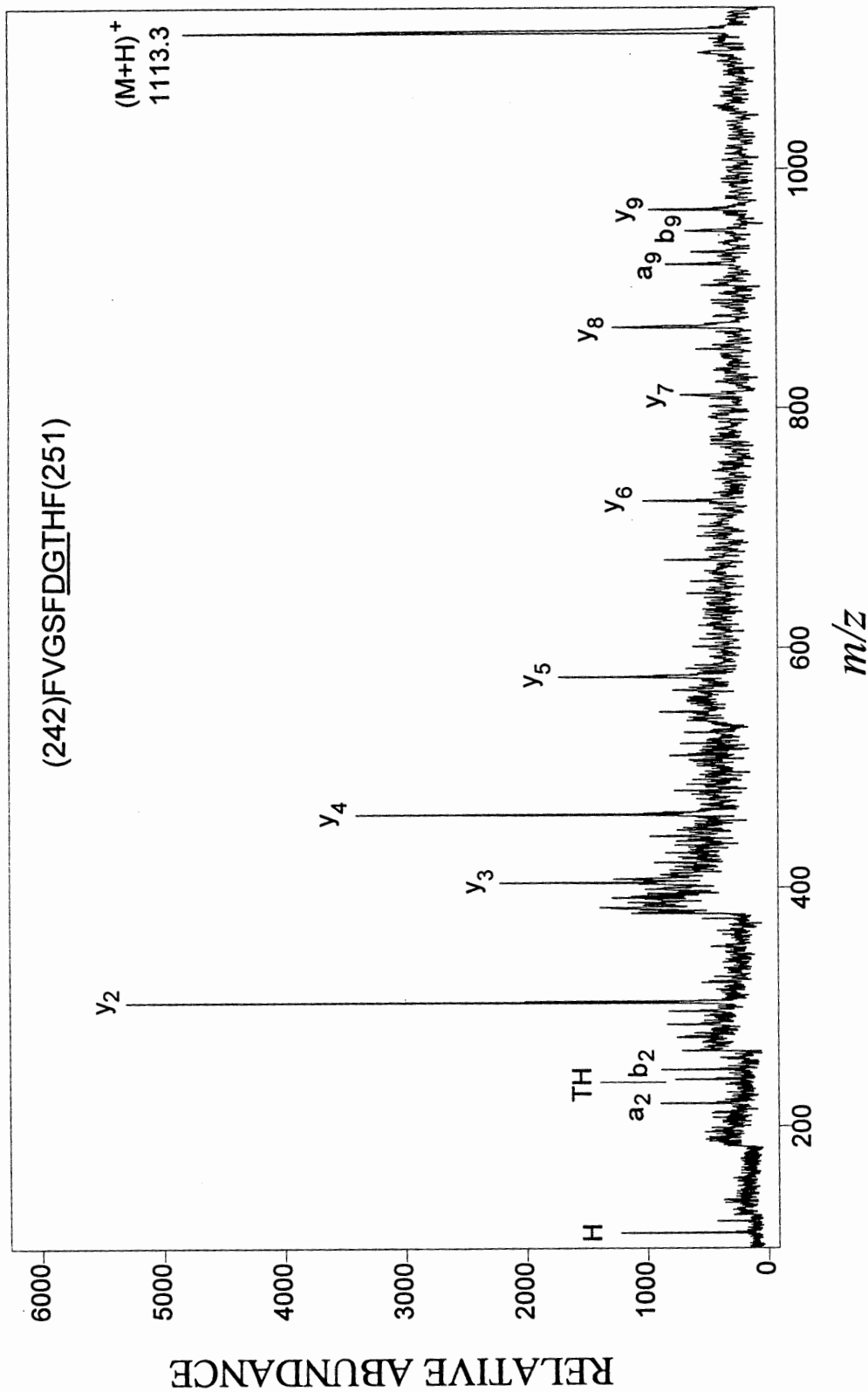


Figure II.8. PSD spectrum of peptide (M+H)<sup>+</sup> = m/z 1113.3. The sequence shown is that of peptide 242-251 which contains sequon 8 (underlined). It should be noted that the glycan has been removed with PNGase F, which converted N<sup>247</sup> to D<sup>247</sup>.

the original sequon with the Asn<sup>247</sup> replaced by Asp after PNGase F digestion (this notation is used throughout this thesis). Therefore, the glycopattern observed from 5,306 to 6,604 Da in Figure II.7b (labeled "\$") must be this peptide (which contains sequons 8 and 9) bearing a total of Man<sub>15</sub>-<sub>23</sub>GlcNAc<sub>4</sub>.

In Figure II.7c, and II.7d, the PNGase F peptide (M+H)<sup>+</sup> = *m/z* 1915.1 and the corresponding Endo H peptide (M+H)<sup>+</sup> = *m/z* 2118.0 were found to be peptide 243-259 (calc. 1915.0 for PNGase F deglycosylated peptide) with one occupied sequon. Like the peptide 242-259 discussed above, this peptide also encompassed sequons 8 and 9, but was produced by cleaving at the C-terminal side of Phe<sup>242</sup> instead of Tyr<sup>241</sup> (which produced peptide 242-259) and it contains only one glycan in its two sequons. The multiple signals at *m/z* 3780.4, 3943.8, 4105.3, 4267.7, and 4429.9 (labeled "#" in Figure II.7b) were matched to this peptide with Man<sub>9-13</sub>GlcNAc<sub>2</sub>.

The glycopattern of *m/z* 3643.8, 3806.3, 3968.1, and 4130.9 (labeled "x" in Figure II.7b) correspond to PNGase F peptide (M+H)<sup>+</sup> = *m/z* 2103.0 (Figure II.7c) with Man<sub>7-10</sub>GlcNAc<sub>2</sub>. The Endo H peptide (M+H)<sup>+</sup> = *m/z* 2305.0 (Figure II.7d) confirmed the presence of a single glycan in this peptide. However, this peptide matched neither a tryptic nor a chymotryptic fragment of invertase. It may be produced by non-specific cleavage or from a contaminant of the sample.

It is important to note that the observation of the unglycosylated peptide 1-18 (see Figure II.7a) indicates that sequon 1 at Asn<sup>4</sup> is only partially occupied. Up to this point, peptide 1-18 was mainly found in fraction T7 as a wide range glycosylated glycopeptide (as discussed above). The

degree of occupancy of sequon 1 was estimated by combining fractions T7 and T8 in 1:1 ratio and acquiring a few MALDI spectra before and after PNGase F deglycosylation (Figure II.9a and II.9b, respectively). Before PNGase F digestion, the relative abundance (averaged over 10 spectra) for the unglycosylated peptide  $(M+H)^+ = 2076.5$  was 6.1% of the reference peak at  $m/z$  1041.1 (peptide 431-438 which does not contain a sequon). After PNGase F deglycosylation, the peptide  $(M+H)^+ = m/z$  2076.5<sup>c</sup> became the strongest peak (see Figure II.9b) and the averaged relative abundance was 130.2% of the reference peak at  $m/z$  1041.1 (which remains 100%). Thus the degree of the occupancy can be derived by a simple calculation using the equation

$$P_G = 100 - (100a'_m a_0 / a_m a'_0) \quad (\text{II-1})$$

where  $P_G$  is the percentage of the glycosylated peptide;  $a_m$  and  $a'_m$  are the observed relative abundances of the reference peaks before and after PNGase F digestion, respectively ( $a_m$  usually is the strongest peak of a peptide which does not contain a sequon in the spectrum of the mixture before deglycosylation);  $a_0$  and  $a'_0$  are the observed relative abundances of the non-glycosylated peptide and the deglycosylated peptide after PNGase F treatment, respectively. In the case of sequon 1 discussed above, the observed relative abundances  $a_0$  and  $a'_0$  of the peptide (before and after PNGase F digestion) at  $m/z$  2076.5 are 6.1 and 130.2, respectively, if the reference peak at  $m/z$  1041.5 remains 100 for both  $a_m$  and  $a'_m$ , (as labeled in Figure II.9). Therefore, the glycosylation

---

<sup>c</sup>The resolution of MALDI-TOF MS is not sufficient to separate the unglycosylated peptide and the PNGase F deglycosylated peptide because they differ in mass only by one dalton. Therefore, signal at  $m/z$  2077 represent both these peptides and the signal intensity is the sum of the two peptides.

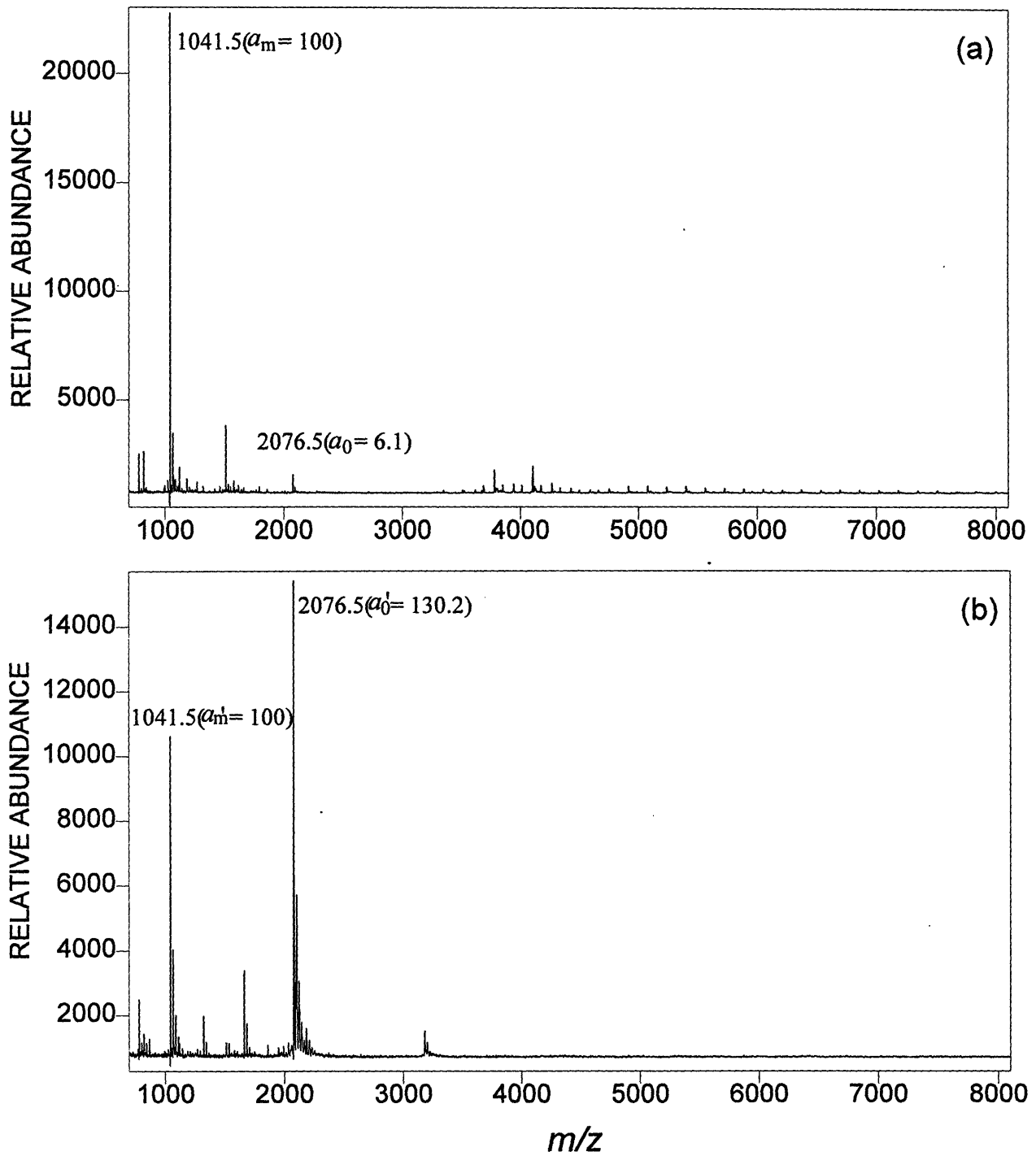
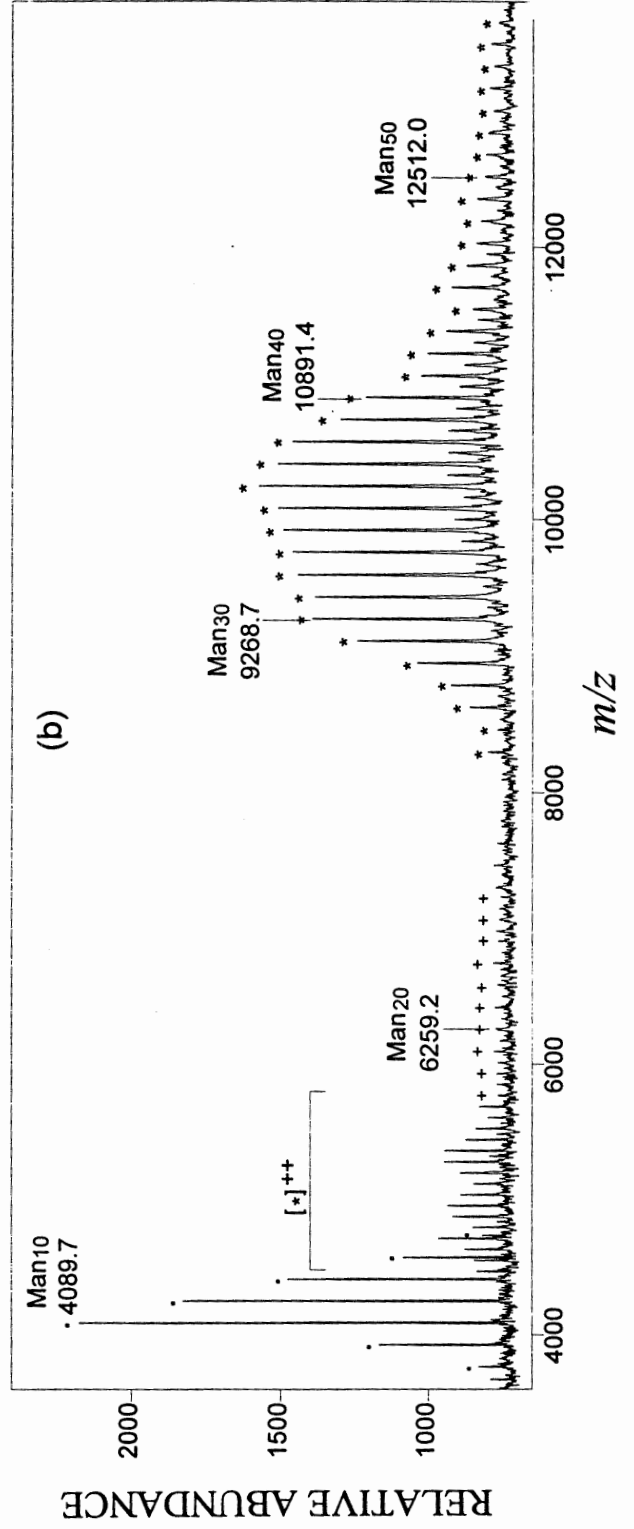
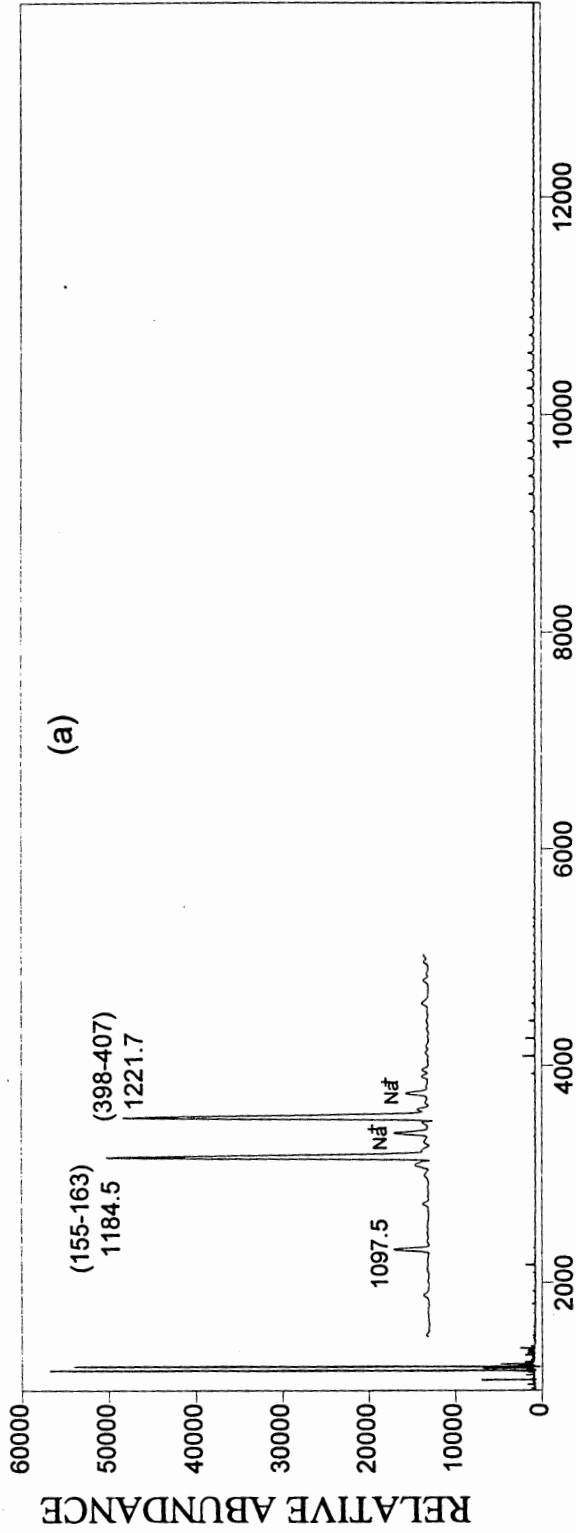


Figure II.9. (a) MALDI spectrum of the 1:1 mixture of T7 and T8. (b) MALDI spectrum of this mixture after PNGase F deglycosylation.

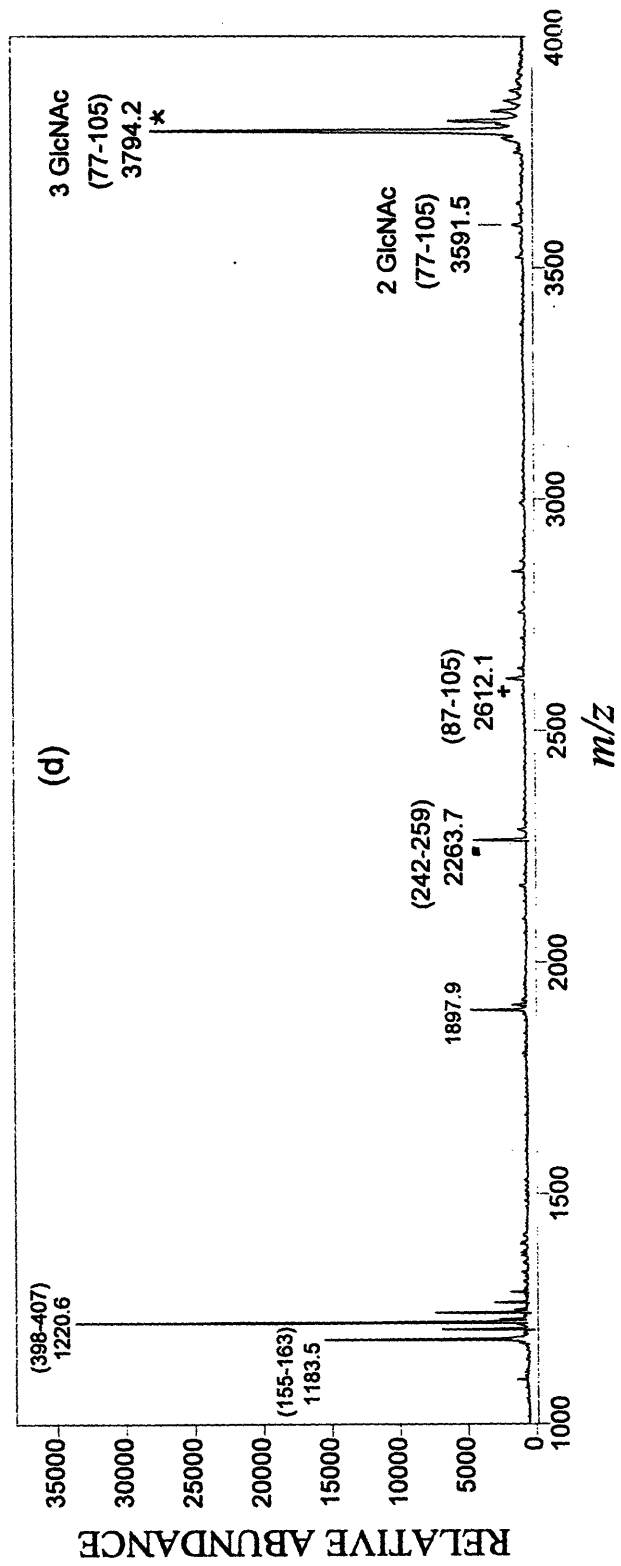
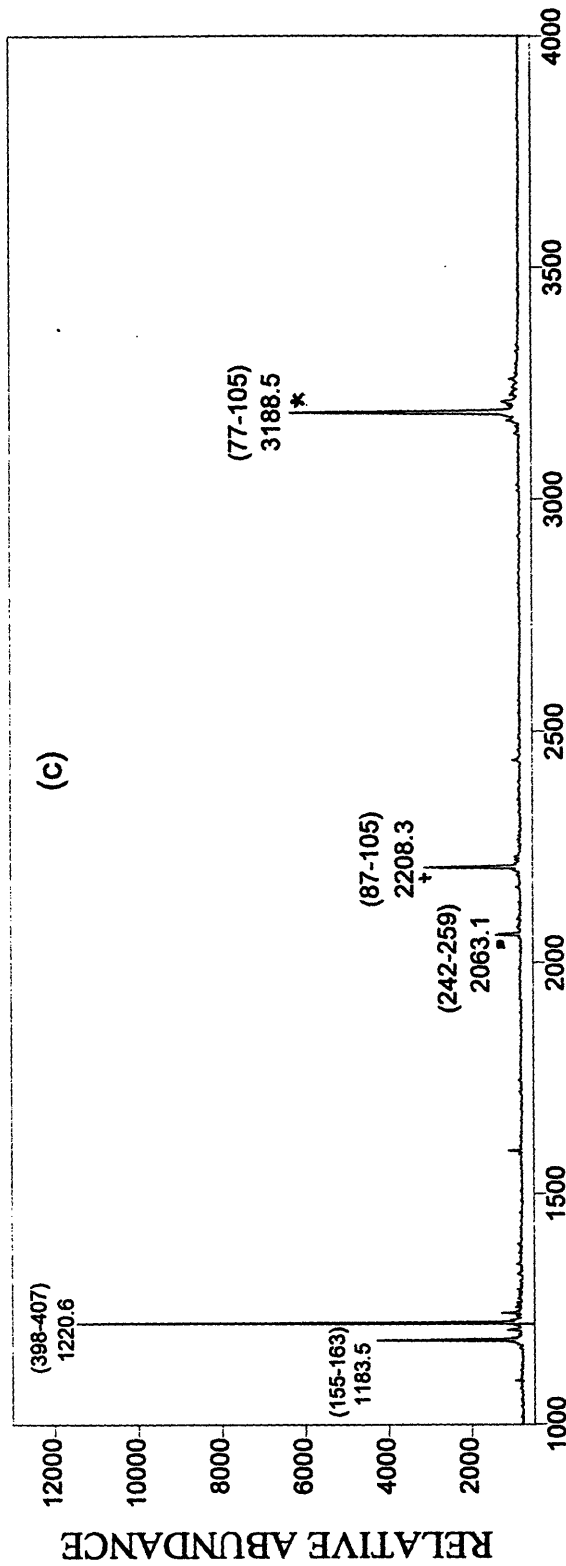
percentage ( $P_G$ ) at sequon 1 is calculated to be 95%. As discussed in Chapter I, the degree of glycosylation deduced in this way is only a rough estimated value. The determination of the occupancy would be simpler if one could compare the averaged abundance of  $(M+H)^+$  of the free peptide with that of the Endo H deglycosylated peptide. However, this method would not be suitable for a highly occupied glycosylation site such as sequon 1, because the signal for the Endo H deglycosylated peptide would be high but that of the free peptide would be lost in the base line noise.

The MALDI spectrum of fraction T9 (see Figure II.2) exhibited three singlet signals in the low mass region (Figure II.10a). The peaks at  $m/z$  1184.5 and  $m/z$  1221.7 were identified as the tryptic peptides 155-163 (calc. 1184.3) and 398-407 (calc. 1221.3), respectively, while the small one at  $m/z$  1097.5 remains unidentified. As the expanded high mass region shown in Figure II.10b indicates, the multiplet signals could represent two or more glycopeptides. After PNGase F digestion, these multiplets collapsed to single peaks at  $m/z$  2063.1, 2208.3, and 3188.5 (Figure II.10c) which had already been found and discussed in fraction T8 as peptides 242-259 (calc. 2062.2 for one occupied sequon or 2063.2 for two), 87-105 (calc. 2208.4 for two occupied sequons), and 77-105 (calc. 3188.4 for three occupied sequons), respectively. The corresponding Endo H deglycosylated peptides were detected at  $m/z$  2263.7, 2612.1, and 3794.2 (Figure II.10d) which confirmed that peptide 242-259 (including sequons 8 and 9) has only one occupied sequon; peptide 86-105 (including sequons 4, 5, and 6) has two occupied sequons; and peptide 77-105 (including sequons 3, 4, 5, and 6) has three occupied sequons. The signal at  $m/z$  1897.9 which appeared after Endo H digestion could not be identified, and is most likely due to a contaminant generated during the Endo H deglycosylation experiment. Based on the information obtained from PNGase F and

Figure II.10 (two pages). (a) MALDI spectrum of fraction T9 of the tryptic digest. (b) The high mass region of (a) expanded in the Y-axis. (c) and (d) The MALDI spectra of fraction T9 after PNGase F and Endo H deglycosylation, respectively.







Endo H deglycosylation, one can assign the multiplet at  $m/z$  3766.8, 3927.3, 4089.7, 4251.7, 4413.9, 4576.1, and 4738.1 (labeled "•") in Figure II.10b to the signals of peptide 242-259 with  $\text{Man}_{8-14}\text{GlcNAc}_2$ , while the glycopattern of the weak signals between 5,400-7,800 Da (labeled "+") corresponds to peptide 86-105. The glycopattern observed in the region from  $m/z$  8,294.6 to 13,808.3 (labeled "★") corresponds to peptide 77-105 with three carbohydrate chains containing a total of  $\text{Man}_{24-58}\text{GlcNAc}_6$ . It should be pointed out that in between these glycopattern signals (labeled "★") another glycopattern rose weakly in the region of  $m/z$  9,800-12,000 (see Figure II.10b), the signals of the latter are about 82 Da heavier or lighter than the former signals which usually give a hint to that the latter has the same peptide fragment but with one more or one less occupied sequon than the former<sup>f</sup>. Because a weak signal at  $m/z$  3591.5 was found in Figure II.10d (which has about 202 Da lighter than 3794.2, peptide 77-105 with three occupied sequons) the latter is in fact one occupied sequon less than the former, therefore, it corresponds to peptide 77-105 with  $\text{Man}_{35-49}\text{GlcNAc}_4$ . Since sequon 5 has been reported not to be glycosylated (Reddy et al, 1988), at most three of the four sequons in peptide 77-105 were expected to be glycosylated which has been the case so far. In order to characterize each of the individual glycosylation sites contained in peptide 77-105 and to confirm that sequon 5 is indeed not glycosylated, the combined fractions T8 and T9 were further digested with chymotrypsin followed by HPLC separation. Thirteen HPLC fractions (Figure II.11) were collected and analyzed by MALDI MS. Their MALDI spectra indicated that fractions CT3, CT4, CT6, CT9, CT10, and CT11 contain glycopeptide(s).

---

<sup>f</sup>There is 82 Da difference in between  $\text{Man}_2$  (324.28 Da) and  $\text{GlcNAc}_2$  (406.38 Da). Therefore, if peptide A containing  $\text{Man}_{20}\text{GlcNAc}_4$  (two occupied sequon), it will be 82 Da heavier than the same peptide bearing  $\text{Man}_{22}\text{GlcNAc}_2$  (one occupied sequon), and vice versa.

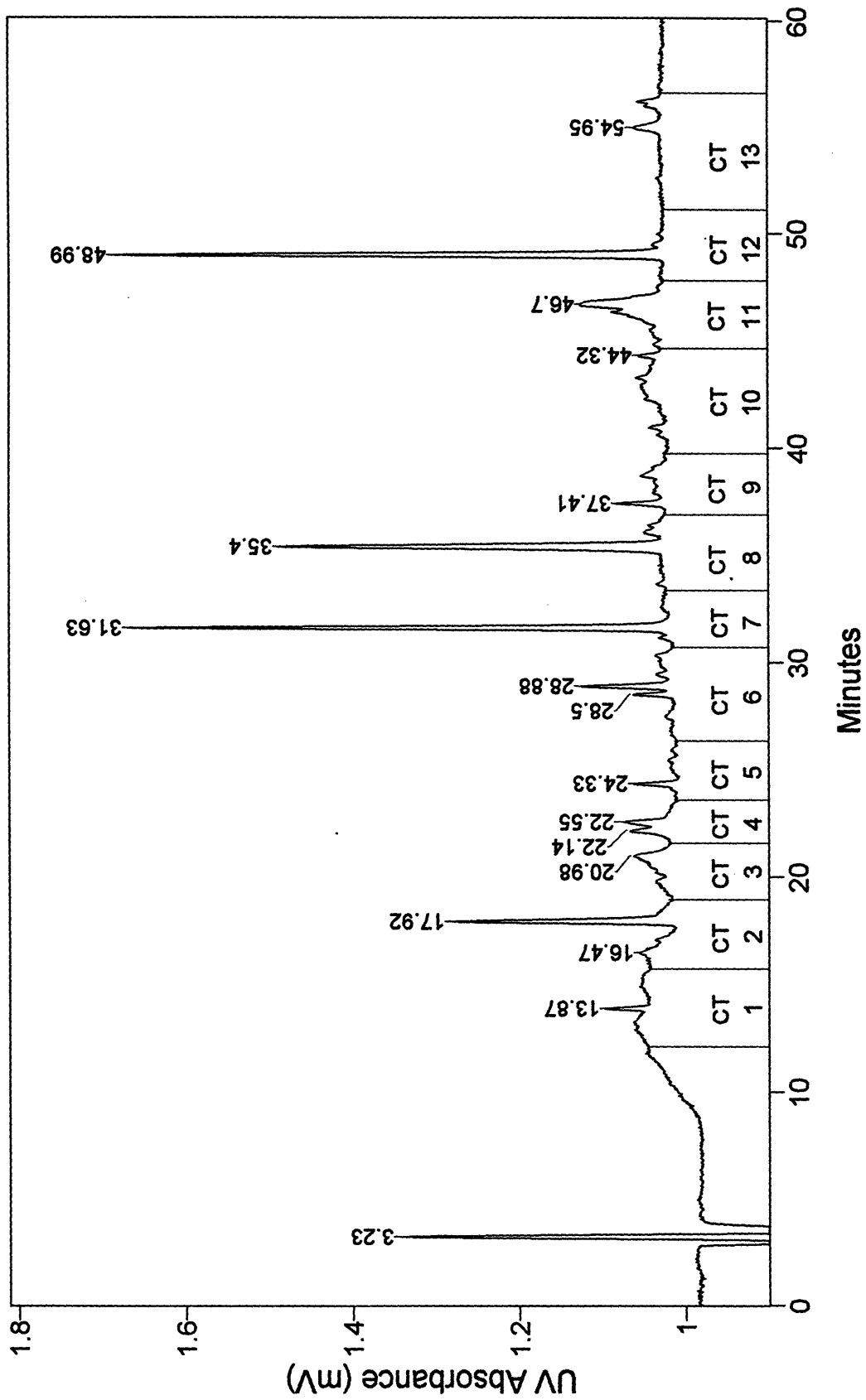


Figure II.11. HPLC trace of the chymotryptic subdigest of the mixture of fractions T8 and T9 (from the tryptic digest). The fractions collected are marked.

The MALDI spectrum of fraction CT3 (Figure II.12a) exhibited two glycopatterns, one a series of ten peaks from  $m/z$  2535.6 to 3996.0 (labeled "•") and the other a multiplet from  $m/z$  3282.3 to 5713.2 (labeled "★"). After removal of the glycan with PNGase F, the glycopatterns collapsed to two peaks (data not shown). The strong signal at  $m/z$  767.4 matches peptide 77-83 (calc. 767.8) which contains sequon 3, while another weak signal at  $m/z$  831.3 fits to peptide 99-105 (calc. 831.8) which contains sequon 6. The signal at  $m/z$  831.3 was weak due to most of glycans at sequon 6 have not been removed (further discussion in the conclusion section). On the other hand, Figure II.12b, the MALDI spectrum of fraction CT4, exhibits a glycopattern at  $m/z$  2147.2-3282.8. The corresponding PNGase F peptide  $(M+H)^+ = m/z$  767.2 fits peptide 77-83 which has been found in fraction CT3. However, the spectrum of fraction CT3 exhibited peaks at  $m/z$  3282.3 to 5714.5 (labeled "★" in Figure II.12a) corresponding to peptide-Man<sub>13-27</sub>GlcNAc<sub>2</sub>, while the spectrum of fraction CT4 presenting a series of peaks at  $m/z$  2147.2 to 3282.8 (labeled "★" in Figure II.12b) fit to peptide-Man<sub>6-13</sub>GlcNAc<sub>2</sub>. Thus peptide 77-83 with the longer glycan chains eluted in fractions CT3 while the one with the shorter glycan chains was collected in fraction CT4. In order to obtain the complete glycosylation profile of sequon 3, fractions CT3 and CT4 were combined and the MALDI MS was acquired (Figure II.12c). Two glycopatterns were observed as expected, where the signals labeled "★" correspond to the peptide 77-83 with Man<sub>4-36</sub>GlcNAc<sub>2</sub> and the peaks labeled "•" correspond to peptide 99-105. The MALDI spectrum of this mixture after PNGase F deglycosylation is shown in Figure II.13a and the sequence of the PNGase F deglycosylated peptide  $(M+H)^+ = m/z$  767.2 was identified to be <sup>77</sup>RDDSGAF<sup>83</sup> from its PSD spectrum (Figure II.13b).

The MALDI spectrum of fraction CT6 exhibited a glycopattern consisting of at least 14

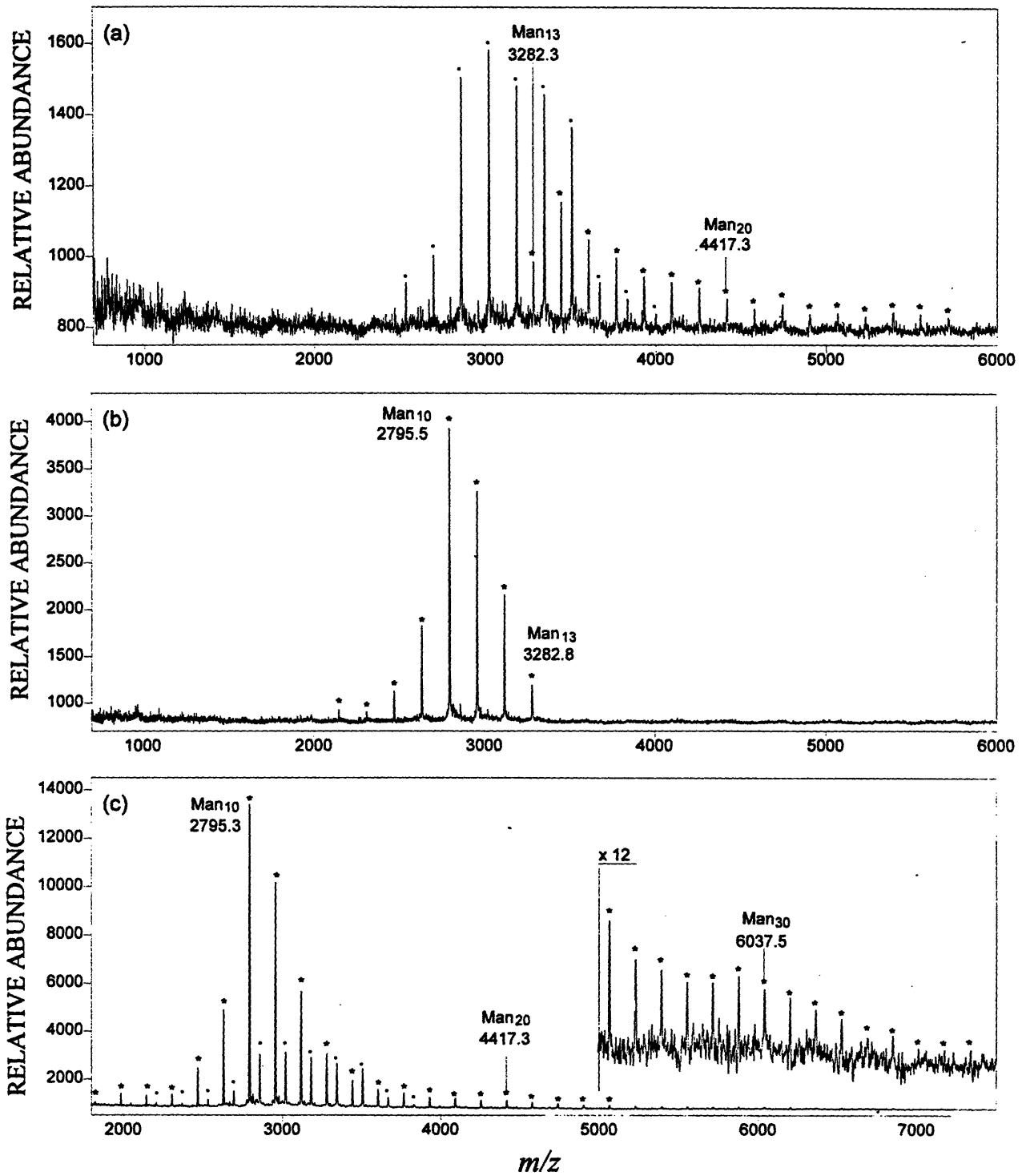


Figure II.12. MALDI spectra of (a) fraction CT3 and (b) fraction CT4 of the chymotryptic subdigest of the combination of fractions T8 and T9 (from the tryptic digest). (c) The MALDI spectrum of the 1:1 mixture of CT3 and CT4.

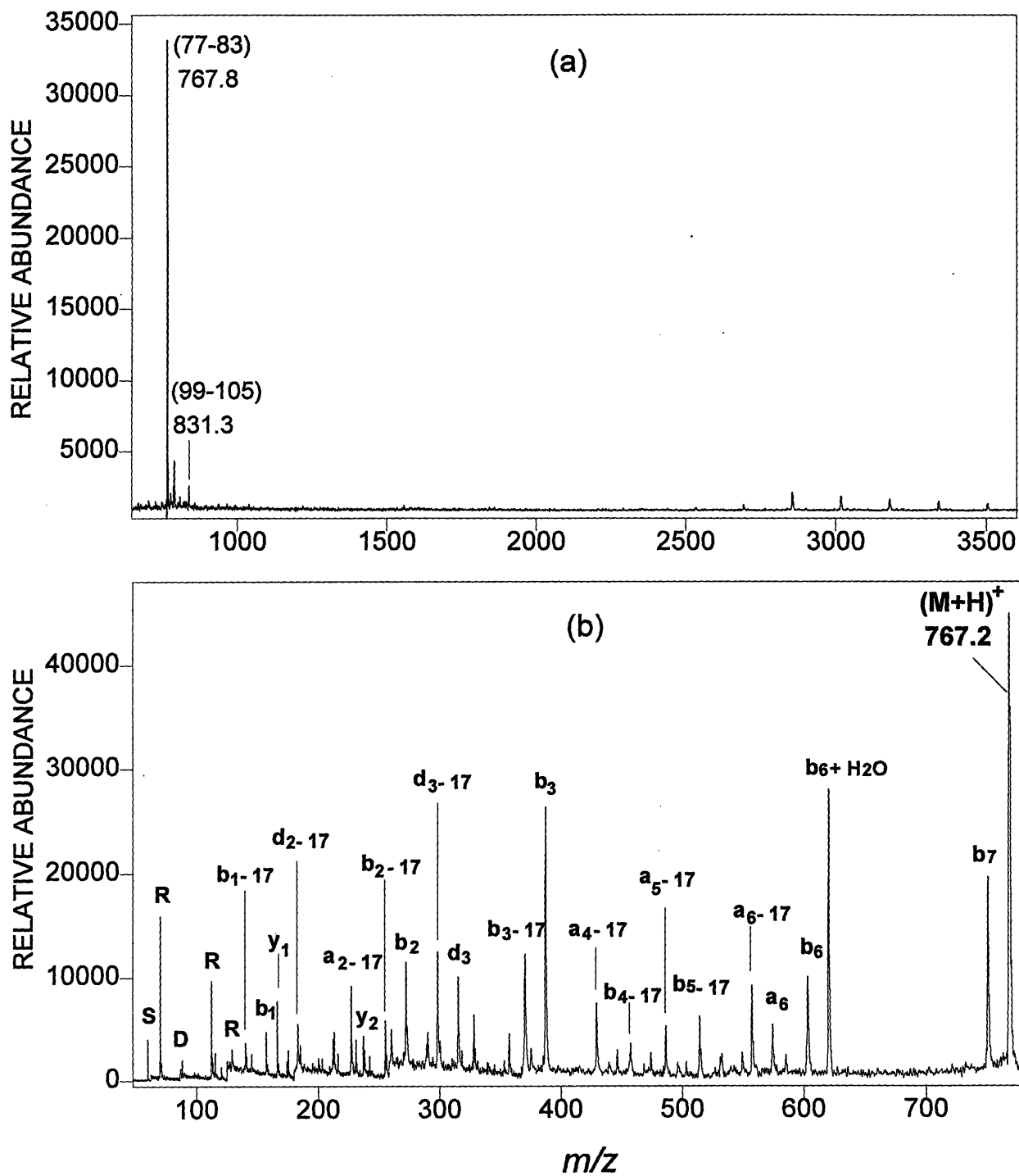


Figure II.13. (a) The MALDI spectrum of the mixture of fractions CT3 and CT4 after PNGase F deglycosylation. (b) MALDI-PSD spectrum of the PNGase F peptide  $(M+H)^+ = 767.8$ . The sequence of this peptide is identified to be  $^{77}RDDSGAF^{83}$ . The very unusual  $d_3$ ,  $d_3-17$ , and  $d_2-17$  ions together with the abundant  $a_n-17$  and  $b_n-17$  ions must somehow be related to the N-terminal RDD sequence.

peaks (Figure II.14a) which collapsed to a single peak at  $m/z$  978.6 (Figure II.14b) upon PNGase F treatment. Analysis of the PSD spectrum (Figure II.15) of this ion confirmed that it is the expected chymotryptic peptide  ${}^{98}\text{FDDTIDPR}^{105}$  (calc. 979.0). Thus the glycopattern in Figure II.14a represents peptide 98-105 (which contains sequon 6) bearing  $\text{Man}_{5,23}\text{GlcNAc}_2$ .

The MALDI spectrum of fraction CT9 exhibits a weak glycopattern at  $m/z$  3366.4, 3529.4, 3691.8, 3853.4, and 4015.6 (Figure II.16a). These signals collapsed to at  $m/z$  1478.5 after PNGase F deglycosylation (Figure II.16b) which matches peptide 84-97 (calc. 1479.6) including sequons 4 and 5 ( $\text{Asn}^{92}\text{-Asn}^{93}\text{-Thr-Ser}$ ). The signal for the sodiated peptide  $(\text{M}+\text{Na})^+ = 1501.1$  is unusually strong even after desalting by HPLC. In addition, the glycopattern shown in Figure II.16a fits only to the sodium adduct of peptide 84-97 with  $\text{Man}_{9,13}\text{GlcNAc}_2$ . The signal of this peptide was too weak to obtain a PSD spectrum to confirm this assignment and to determine which sequon has been occupied.

Figure II.17a shows the MALDI spectrum of fraction CT10. The glycopattern present at  $m/z$  6,000-8,200 (marked "•") collapsed to a peak at  $m/z$  2438.3 after PNGase F deglycosylation (Figure II.17b) which corresponds to an incompletely digested chymotryptic peptide 84-105 bearing  $\text{Man}_{17,30}\text{GlcNAc}_4$ . This peptide (calc. 2439.6) encompasses sequons 4, 5, and 6 ( $\text{Asn}^{92}\text{-Asn}^{93}\text{-Thr-Ser}$  and  $\text{Asn}^{99}\text{-Asp-Thr}$ ). The PSD spectrum of peptide  $(\text{M}+\text{H})^+ = 2438.3$  shown in Figure II.18 exhibits only a few fragment ions. Even though they are barely sufficient to confirm the identity of this peptide, it did not permit the assignment of occupied sequons since the mass difference between Asp and Asn is only one dalton. The mass accuracy in the PSD mode is about 0.01-0.1% but depends

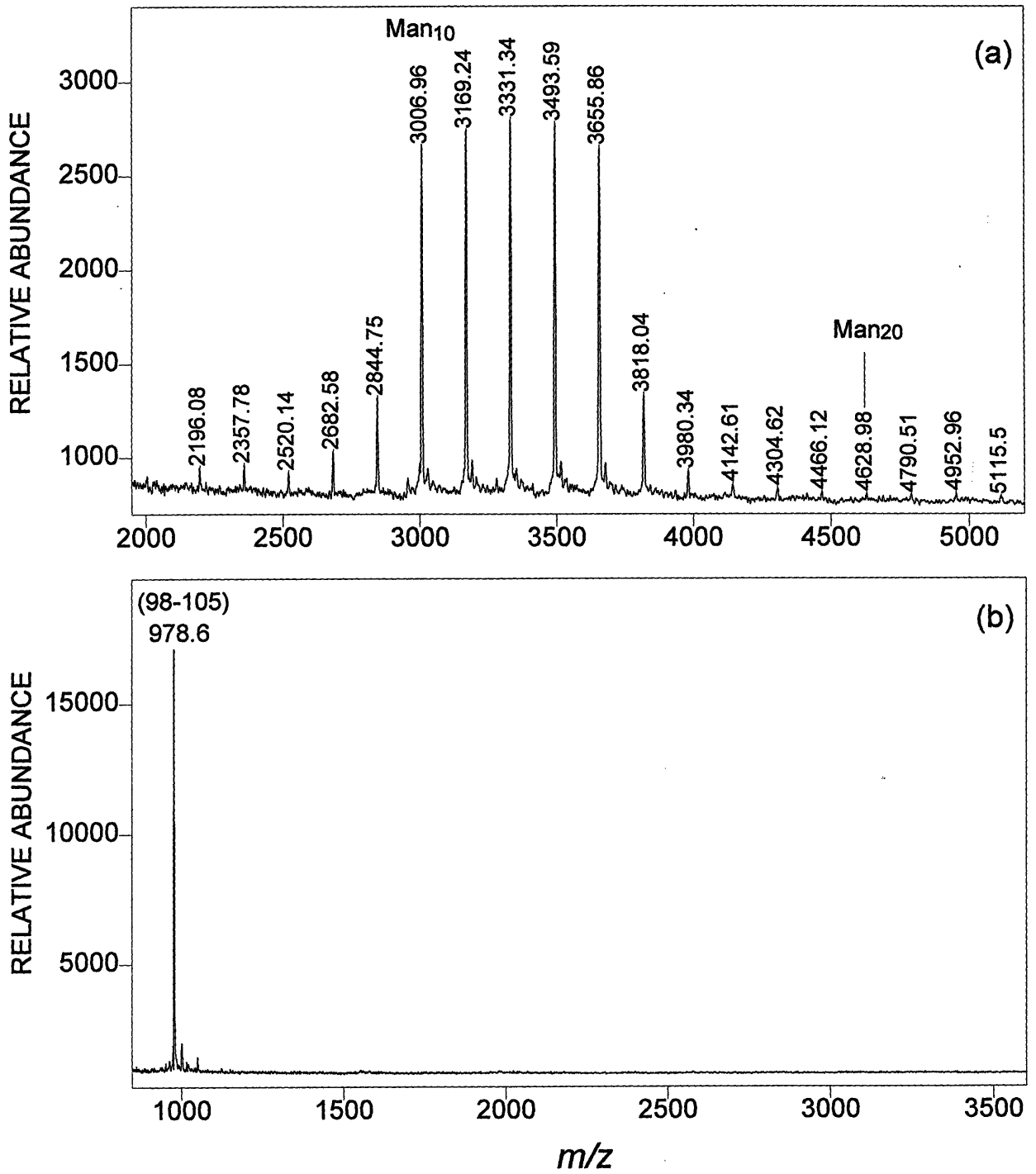


Figure II.14. MALDI spectrum of (a) fraction CT6 of the chymotryptic subdigest of the 1:1 mixture of fractions T8 and T9 (from the tryptic digest), and (b) the corresponding PNGase F deglycosylated peptide.



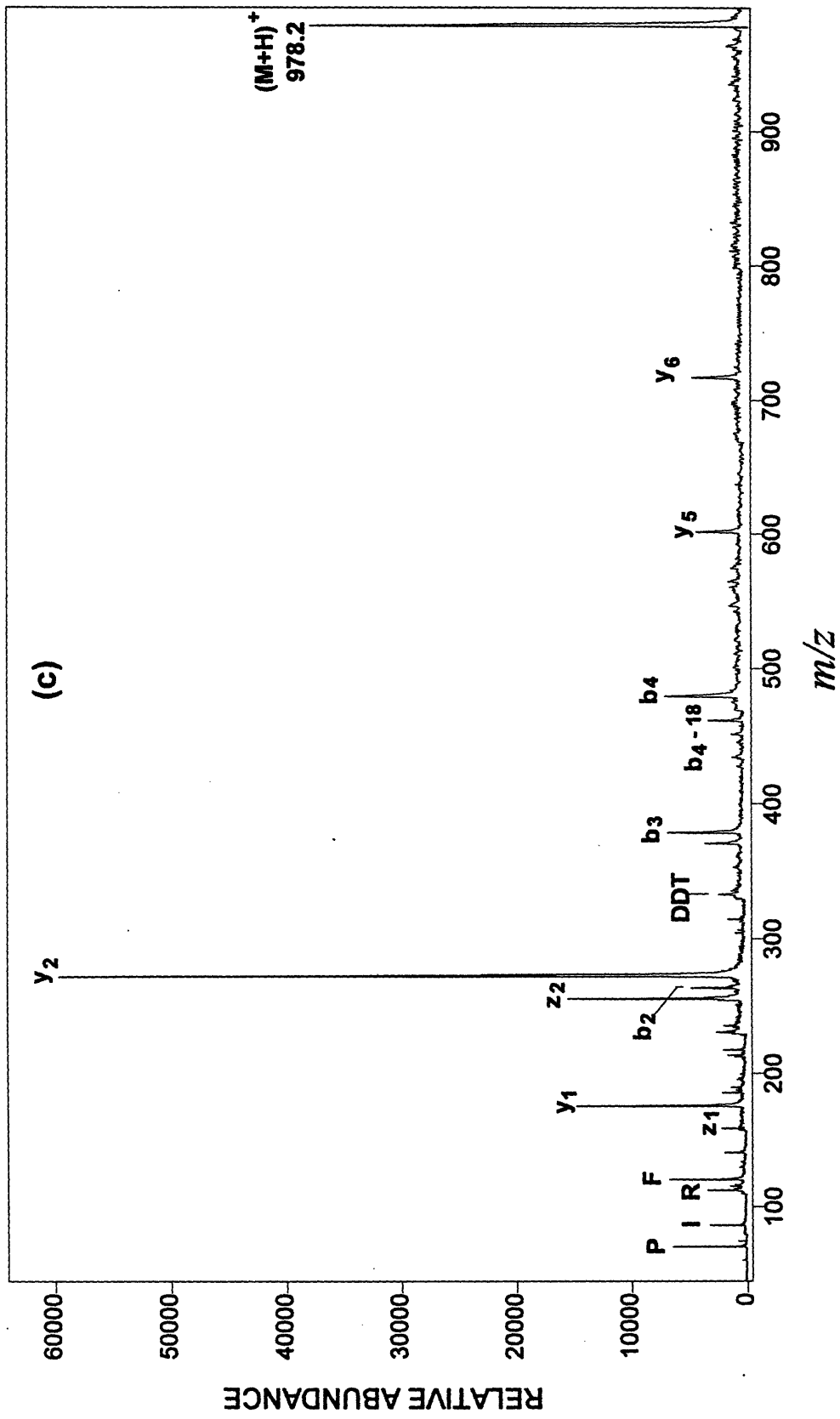


Figure II.15. MALDI-PSD spectrum of the PNGase F peptide (M+H)<sup>+</sup> = 978.2 which confirms the sequence <sup>98</sup>FDDIIDPR<sup>105</sup>.

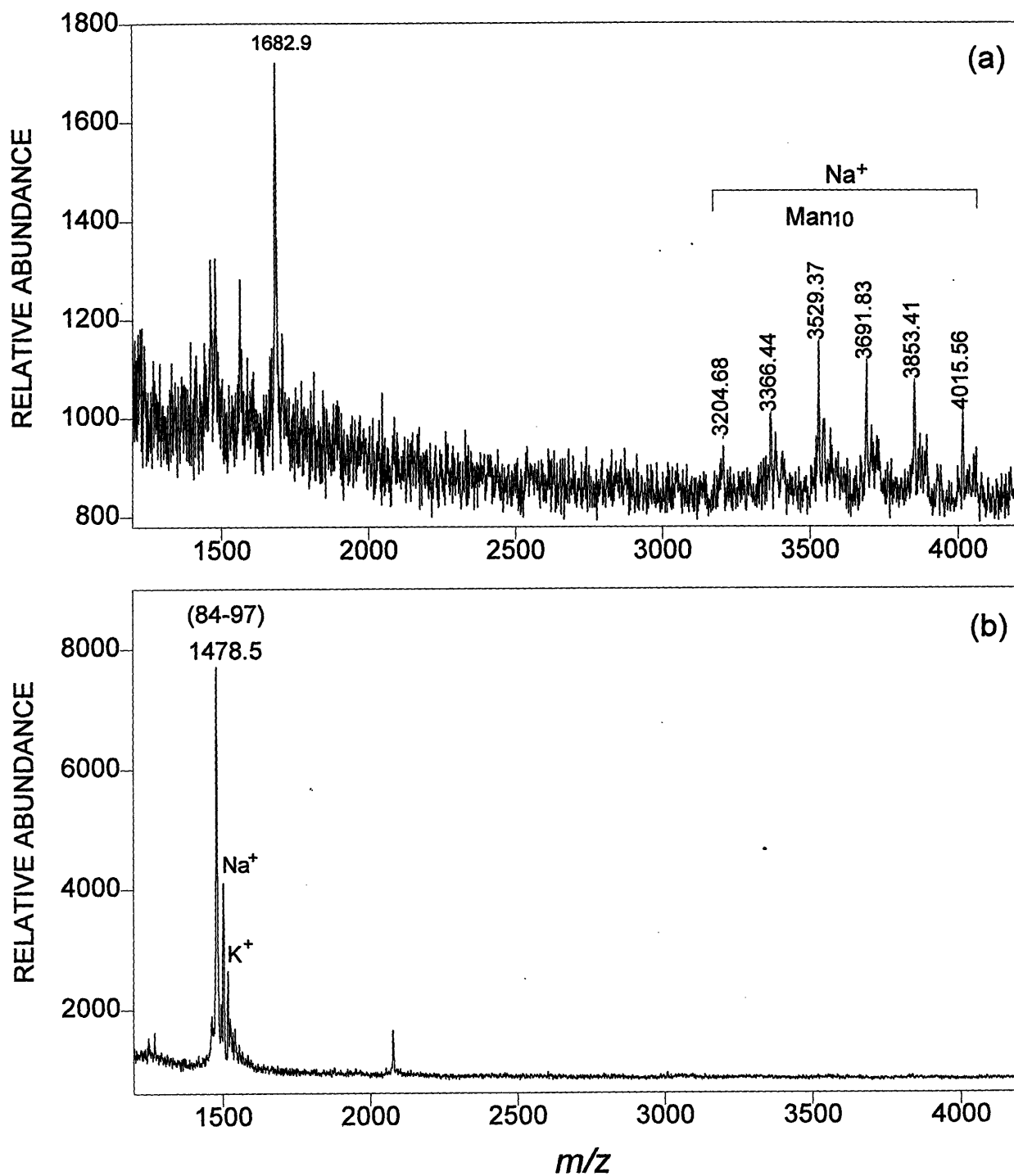


Figure II.16. MALDI spectrum of (a) fraction CT9 of the chymotryptic subdigest of the 1:1 mixture of fractions T8 and T9 (from the tryptic digest), and (b) the corresponding PNGase F deglycosylated peptide.

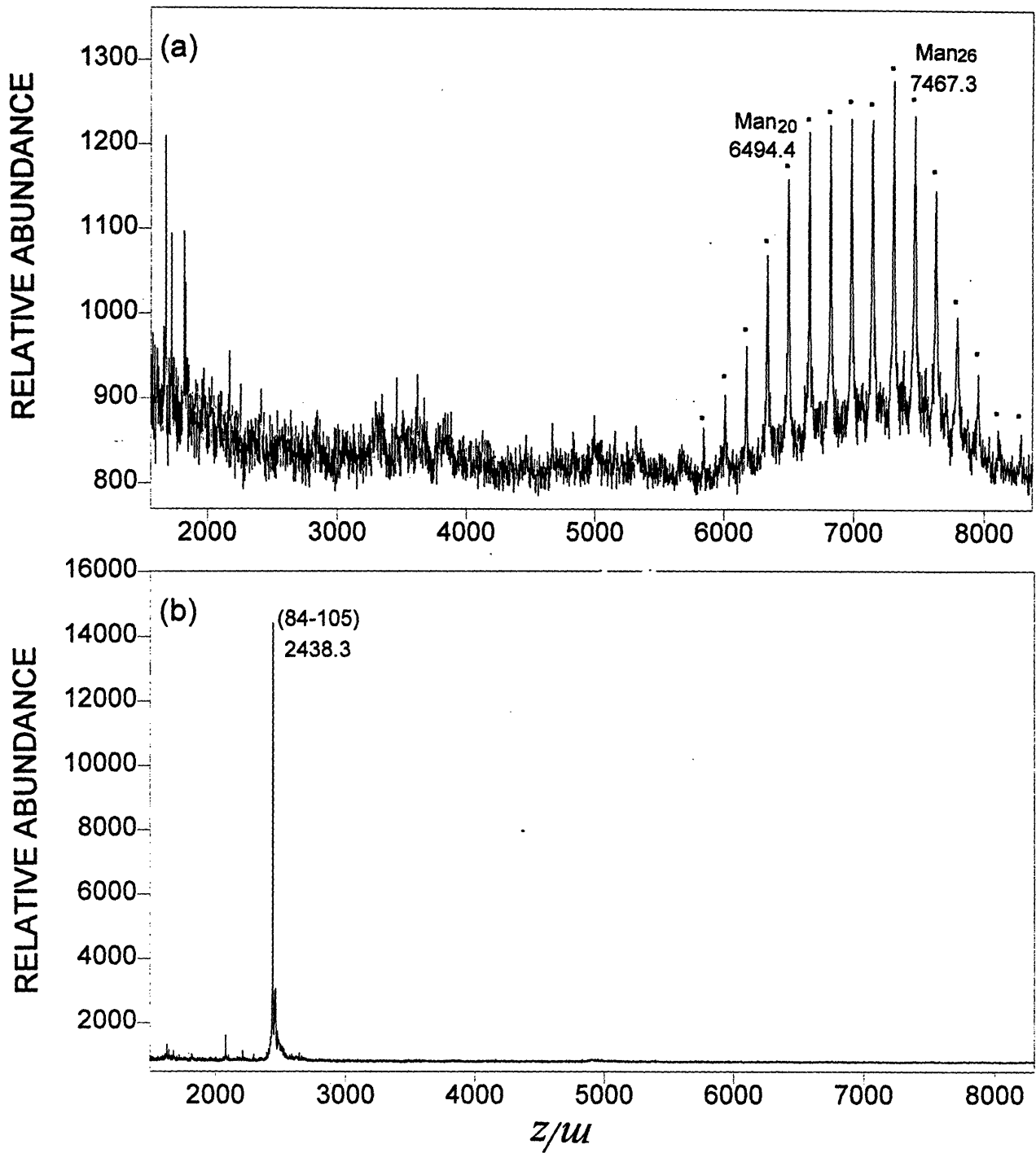


Figure II.17. MALDI spectrum of (a) fraction CT10 of the chymotryptic subdigest of the 1:1 mixture of fractions T8 and T9 (from the tryptic digest), and (b) the corresponding PNGase F deglycosylated peptide.

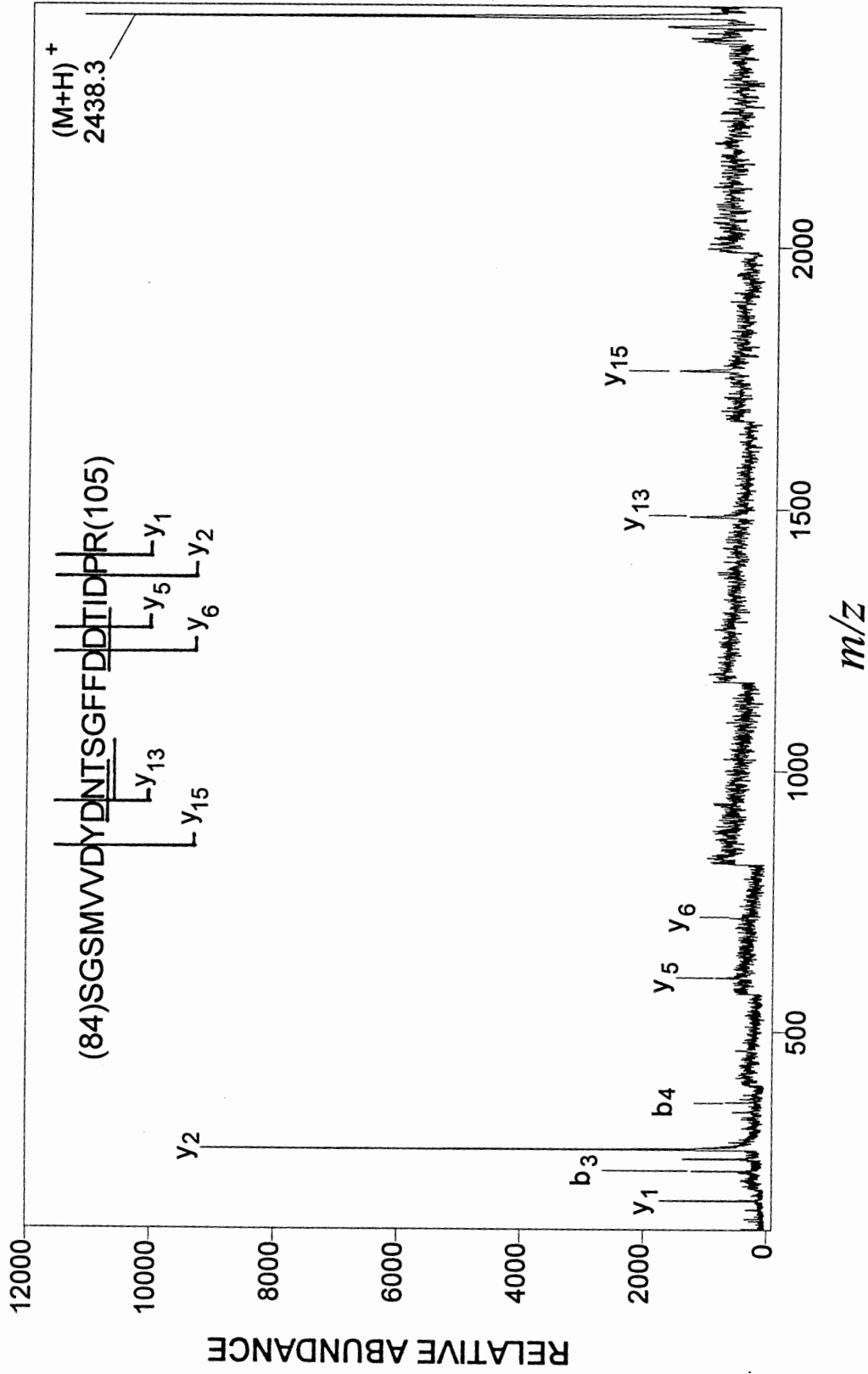


Figure II.18. MALDI-PSD spectrum of the PNGase F peptide (M+H)<sup>+</sup> = *m/z* 2438.3. The sequence shown is that of peptide 84-105 which contains sequons 4, 5, and 6 (underlined).

upon the size of the fragment and the intensity of the signal. In this case the significant fragment ions  $y_{13}$  and  $y_{15}$  are large ( $m/z$  1485.4 and 1763.1) and the signals are weak so the expected error should be no better than 0.1% (which is equivalent to  $\pm 1.5$  to 1.8 Da), insufficient for the differentiation between Asp and Asn.

The MALDI spectrum of fraction CT11 is shown in Figure II.19a. In the low mass region, the peak at  $m/z$  1525.0 matches the tryptic peptide 152-163 (calc. 1524.7) while the peak at  $m/z$  1626.0 fits peptide 84-98 (calc. 1625.8) with sequons 4 and 5 at Asn<sup>92</sup> and Asn<sup>93</sup> unoccupied. Another abundant peak at  $m/z$  1828.7 is just 203 Da heavier than the peptide 84-98 (the peak at  $m/z$  1626.0) which may indicate an additional GlcNAc residue attached to this peptide. This is an unexpected finding which can only be explained by some Endo H activity being present in the system.

A weak and highly cationized (with Na<sup>+</sup> and K<sup>+</sup>, respectively) glycopattern was observed in the high mass region (Figure II.19b). This fraction was then treated with PNGase F followed by HPLC desalting. The resulting PNGase F deglycosylated peptide exhibited a cluster in the MALDI spectrum consisting of the proton adduct ion at  $m/z$  1626.5, the sodium adduct ion at  $m/z$  1648.9, and the potassium adduct ion  $m/z$  1665.2 (Figure II.20a) all of which match peptide 84-98 (calc. 1626.7 for the protonated ion) encompassing sequons 4 and 5. This indicates that the unusual highly cationized glycopattern shown in Figure II.19b corresponds to peptide 84-98 with a total of Man<sub>3</sub>-<sub>15</sub>GlcNAc<sub>2</sub>. It must be noted that the signal at  $m/z$  1828.7 (Figure II.19a) disappeared after PNGase F deglycosylation (Figure II.20a) which confirmed the GlcNAc attachment based on the personal

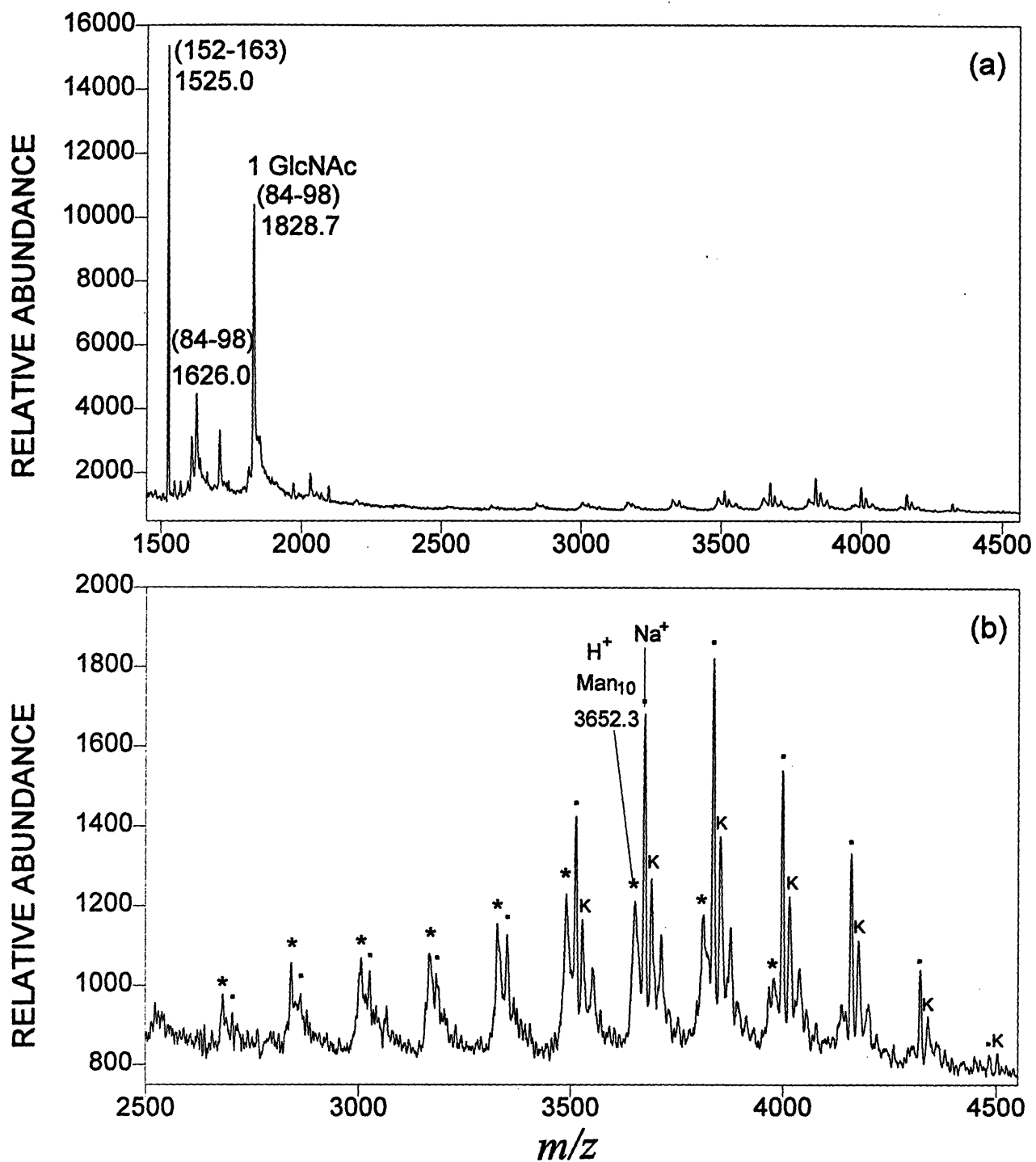


Figure II.19. (a) MALDI spectrum of fraction CT 11 of the chymotryptic subdigest of the 1:1 mixture of fractions T8 and T9 (from the tryptic digest). (b) The expanded region of (a), to show the glycopattern ("★" indicates the  $(M+H)^+$  ions; "•" indicates the  $(M+Na)^+$  ions; and "K" indicates the  $(M+K)^+$  ions).

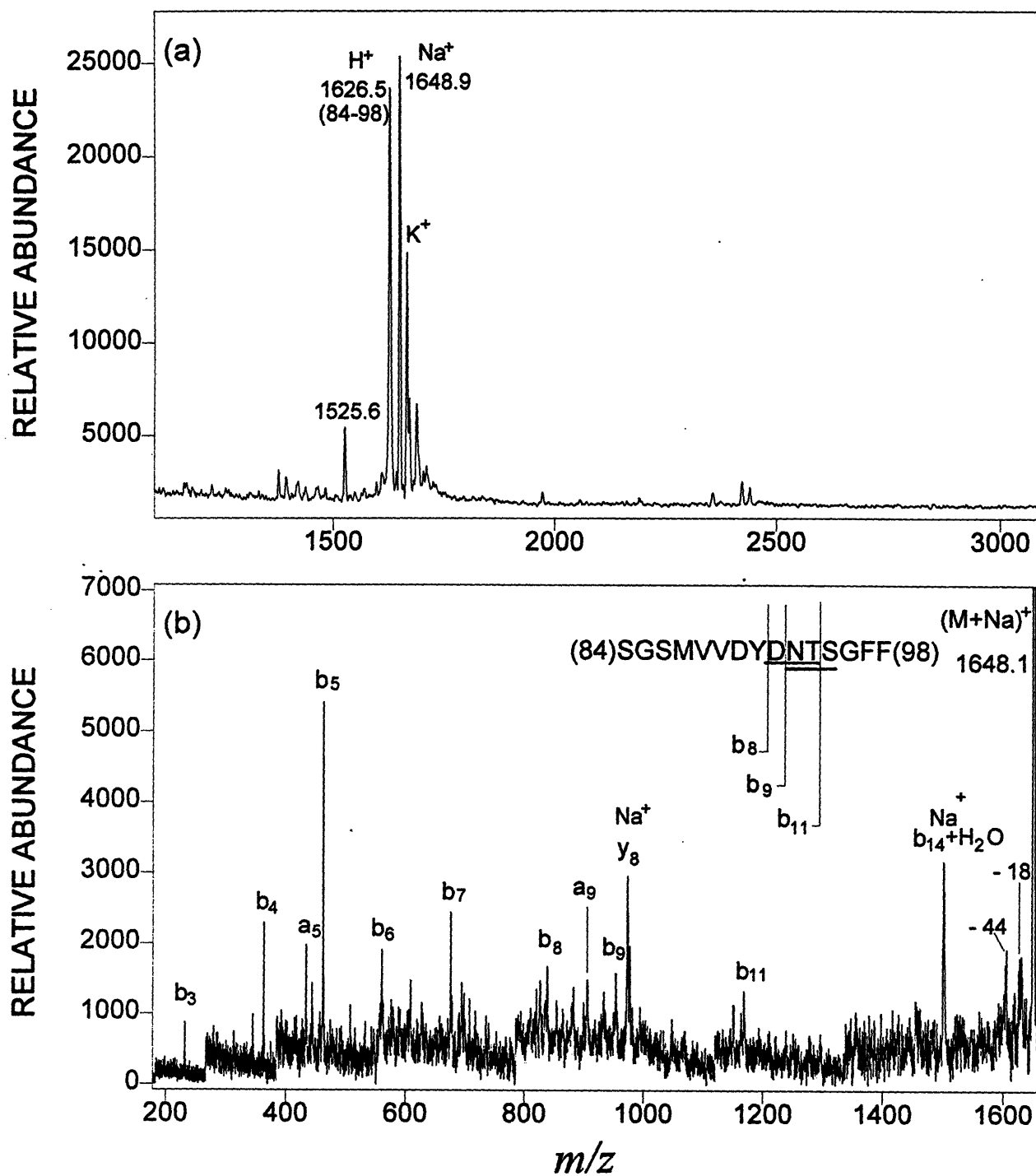


Figure II.20. (a) The MALDI spectrum of fraction CT11 after PNGase F deglycosylation. (b) MALDI-PSD spectrum of peptide 84-98 acquired from both precursor ions,  $(M+H)^+ = m/z$  1626.5 and  $(M+Na)^+ = m/z$  1648.9. The sequence and the description placed on top of the spectrum is peptide 84-105 containing sequons 4 and 5 (underlined). Notice all the fragment ions are produced from  $(M+H)^+$  precursor except otherwise marked.

experience that PNGase F can also remove GlcNAc from the GlcNAc-peptide given sufficient enzyme and reaction time. Since fraction CT11 has been found to contain peptide 84-98 bearing no sugar, one GlcNAc, and a total of  $\text{Man}_{3-15}\text{GlcNAc}_2$  (one occupied sequon), either sequons 4 or 5 was partially glycosylated with  $\text{Man}_{3-15}\text{GlcNAc}_2$ . Another possibility is that in peptide 84-98 either sequon may be mono-glycosylated.

By applying equation (II-1) to the relative abundances of the peak at  $m/z$  1525.0 (peptide 152-163) and the peak at  $m/z$  1626.0 (sequons 4 and 5 containing peptide 84-98) before and after PNGase F digestion, one can conclude that this peptide has been approximately 96% glycosylated at either sequon 4 or 5 (see discussion above). As discussed earlier, this peptide seems to bind alkali ions ( $\text{Na}^+$  and  $\text{K}^+$ ) strongly even though the HPLC desalting procedure had been used. It is possible that there is a special alkali ion binding sequence in this peptide and this may be one of the reasons why the glycopeptide signal was so weak. Another explanation for the weak signal is that glycosylation of Asn<sup>99</sup> inhibits cleavage of the peptide bond between Phe<sup>98</sup> and Asn<sup>99</sup>. Therefore, the amount of peptide 84-105 (found in CT10) must be more than peptides 84-97 (found in CT9) and 84-98 (found in CT11). The reasons for the stronger  $(\text{M}+\text{H})^+$  signal (without alkali cations) obtained for glycopeptide 84-105 (see Figure II.17a) may be the large amount of this peptide formed and the fact that it contains a C-terminal arginine.

In an effort to sequence peptide 84-98 by PSD, the  $(\text{M}+\text{Na})^+$  ion was employed as the precursor ion due to the difficulty of obtaining an adequate signal for the  $(\text{M}+\text{H})^+$  ion =  $m/z$  1626.5. During the experiment, special settings were applied in the instrument configuration to improve the



result. By setting the timed ion selector at a mass value between  $(M+Na)^+$  and  $(M+H)^+$ , the window was wide enough to allow both precursor ions to pass into the ion mirror. Under these conditions, a fairly satisfactory PSD spectrum was obtained which identified the peptide sequence as  $^{84}SGSMVVDYDNTSGFF^{98}$  (Figure II.20b). The mass difference between  $b_8$  ( $m/z$  839.0) and  $b_9$  ( $m/z$  954.1) is 115.1 Da indicating that residue 92 is aspartic acid. In addition, the gap between  $b_9$  and  $b_{11}$  ( $m/z$  1169.1) is 215.0 Da which agrees the Asn-Thr in positions 93 and 94. Therefore, sequon 4 (Asn<sup>92</sup>) not sequon 5 (Asn<sup>93</sup>) must have been glycosylated. It should be pointed out that during analysis of the PSD spectrum, most fragment ions were found to arise via decay from the  $(M+H)^+$  precursor ion and only a few fragments arose from  $(M+Na)^+$ . This phenomenon suggest that the  $(M+Na)^+$  ion is much more stable than  $(M+H)^+$  ion (Mallis and Russell, 1986; Russell et al., 1988; Ngoka et. al., 1994).

Figure II.21a shows the MALDI spectrum of fraction T10 of the tryptic digest of external invertase. A well defined glycopattern of 11 peaks was observed in the range of  $m/z$  2682.1 to 4304.2 which collapsed to the peak at  $m/z$  1627.0 after PNGase F treatment (Figure II.21b). This mass matches peptide 332-346 (calc. 1626.8) which contains sequon 10 (Asn<sup>337</sup>-Ile-Ser). The corresponding Endo H peptide,  $(M+H)^+ = 1828.8$ , confirms that this represents the glycosylated peptide 332-346. Therefore, the glycopattern in Figure II.20a can now be assigned to peptide Ala<sup>332</sup>-Arg<sup>346</sup> bearing GlcNAc<sub>2</sub>Man<sub>4-15</sub> at sequon 10. In addition, the unglycosylated peptide 332-346 found in fraction T11 indicates the partial glycosylation at sequon 10. To determine the glycosylation occupancy, a 1:1 mixture of fractions T10 and T11 was digested with Endo H and the relative abundance (averaging of 10 spectra) between the unglycosylated free peptide and the Endo

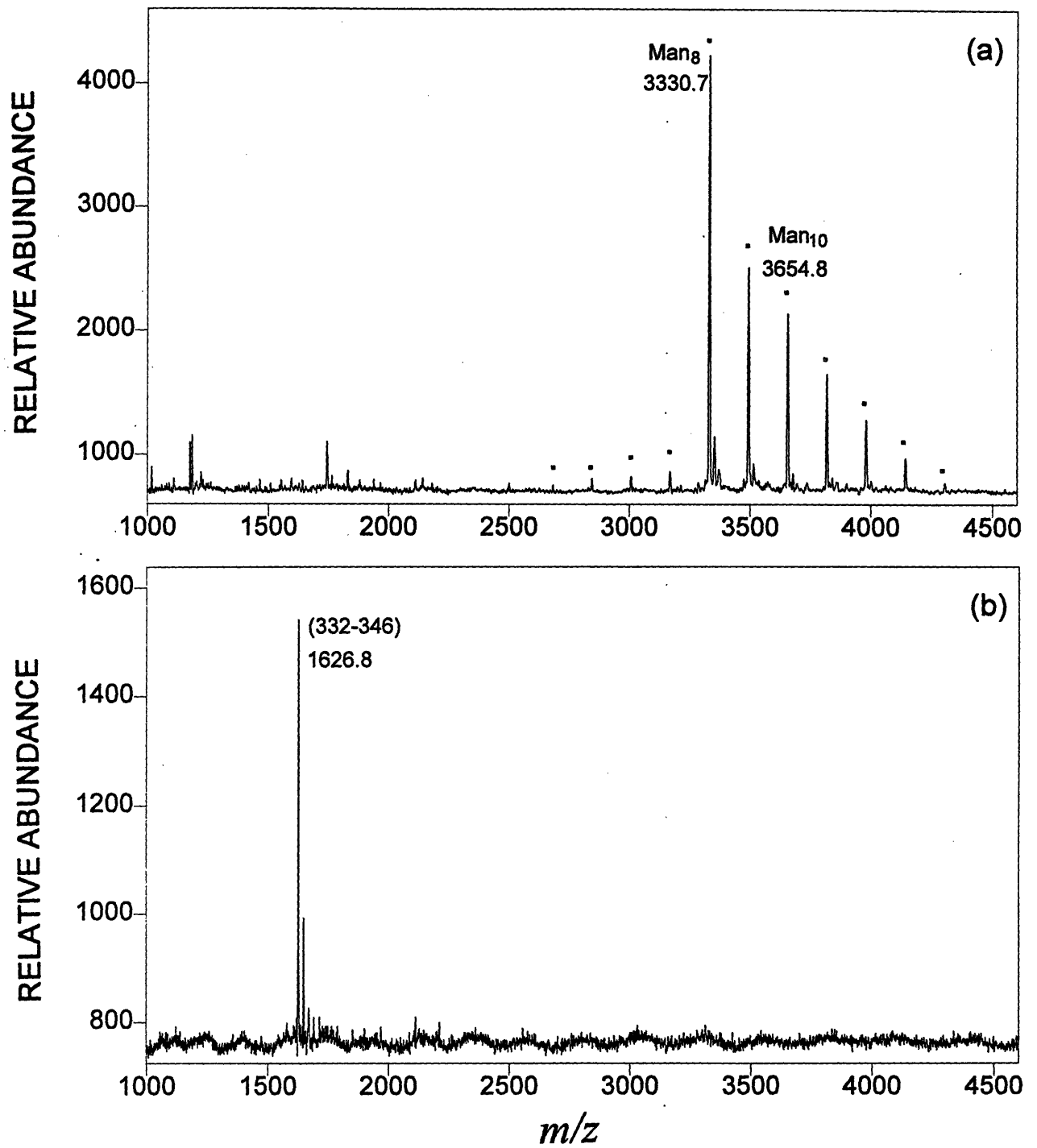


Figure II.21. (a) The MALDI spectrum of fraction T10 of the trypsin digest. (b) The corresponding MALDI spectrum after PNGase F deglycosylation.

H peptide was used to estimate the degree of glycosylation as discussed earlier. Using this method the occupancy at sequon 10 was determined to be 87%.

The MALDI spectrum (Figure II.22a) of fraction T15 exhibits some abundant peaks at  $m/z$  1593.9, 2353.9, 3182.7, and 3433.4 which fit to the tryptic peptides 408-421 (calc. 1593.8), 312-331 (calc. 2353.6), 398-425 (calc. 3182.5), and 189-219 (calc. 3433.8), while others present at  $m/z$  2951.0, and 4286.8 match the tryptic-chymotryptic peptides 115-139 (calc. 2951.1) and 398-434 (calc. 4287.8). The expanded region of  $m/z$  6,000 to 9,100 (Figure II.22a) may present three glycopatterns in fraction T15. One exhibits seven peaks in the mass range of 7,800-8,900, while the other two present in the same mass range in between 6,000 to 7,500 Da differ by 30 Da. Since these two glycopatterns exhibit the identical signal profile, these two glycopeptides may contain the same sequon(s).

After fraction T15 was treated with PNGase F as usual, the spectrum was fairly complicated (Figure II.23a). It turned out (see the later discussion of the Lys-C and Asp-N digests of invertase) that the pair of signals at  $m/z$  4871.3 and 4901.7 correspond to peptide 36-76 (which contains sequon 2 at Asn<sup>45</sup>). According to the published sequence of yeast invertase (Taussig and Carlson, 1983) this peptide should give rise to a  $(M+H)^+$  ion of  $m/z$  4848.3 (with 1N→D after PNGase F digestion) which is about 23 Da less than the lighter mass signal of the pair. As will be shown later, in  $m/z$  4871.3 one amino acid was replaced by another while in  $m/z$  4901.7 two amino acids have been replaced. These signals are spaced 30 Da apart and must correspond to the twin-peak glycopattern described above (see Figure II.22b).

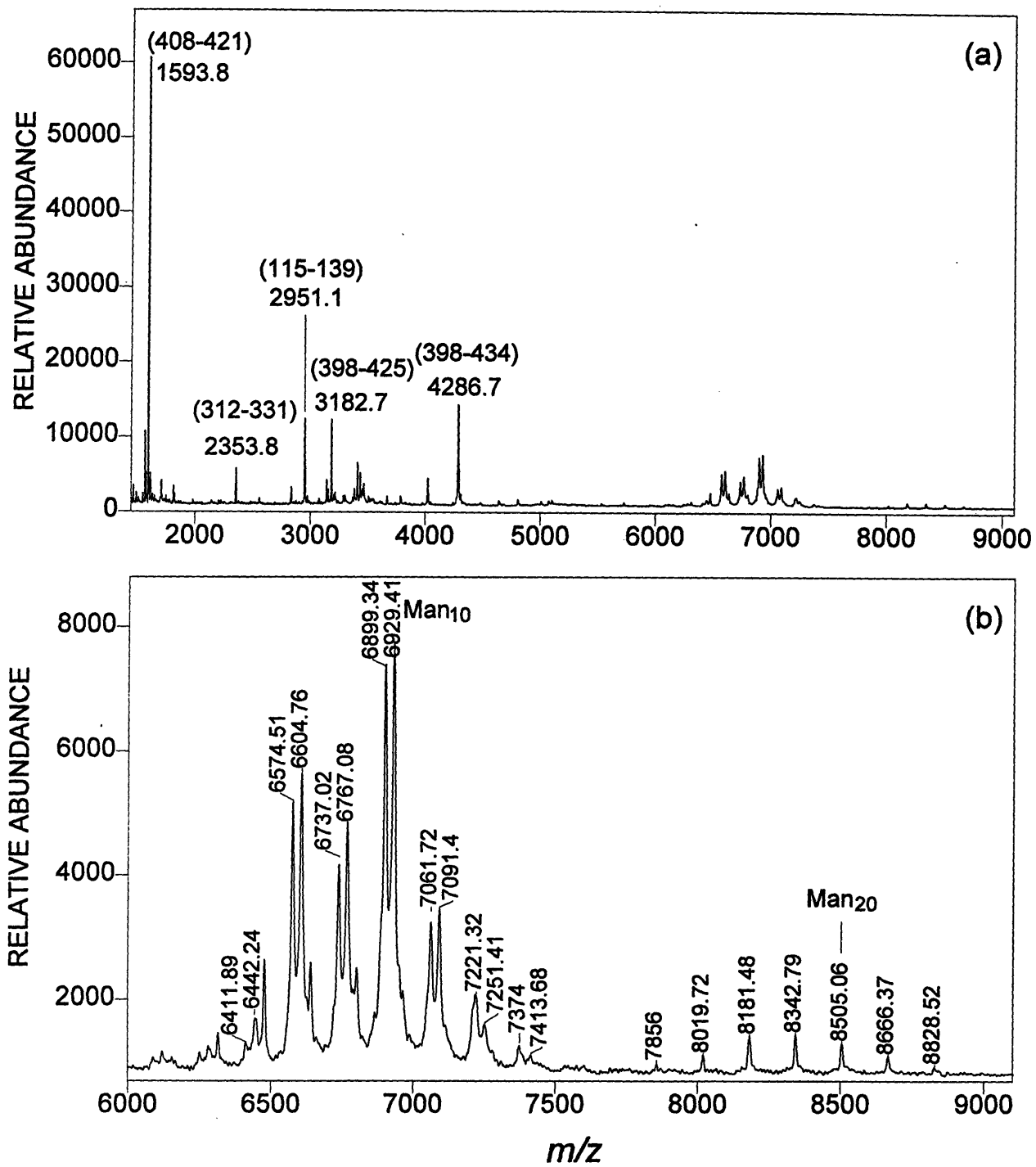


Figure II.22. (a) The MALDI spectra of fraction T15. (b) The region of 6,00-9100 of part (a) expanded (8×) in the Y-axis.

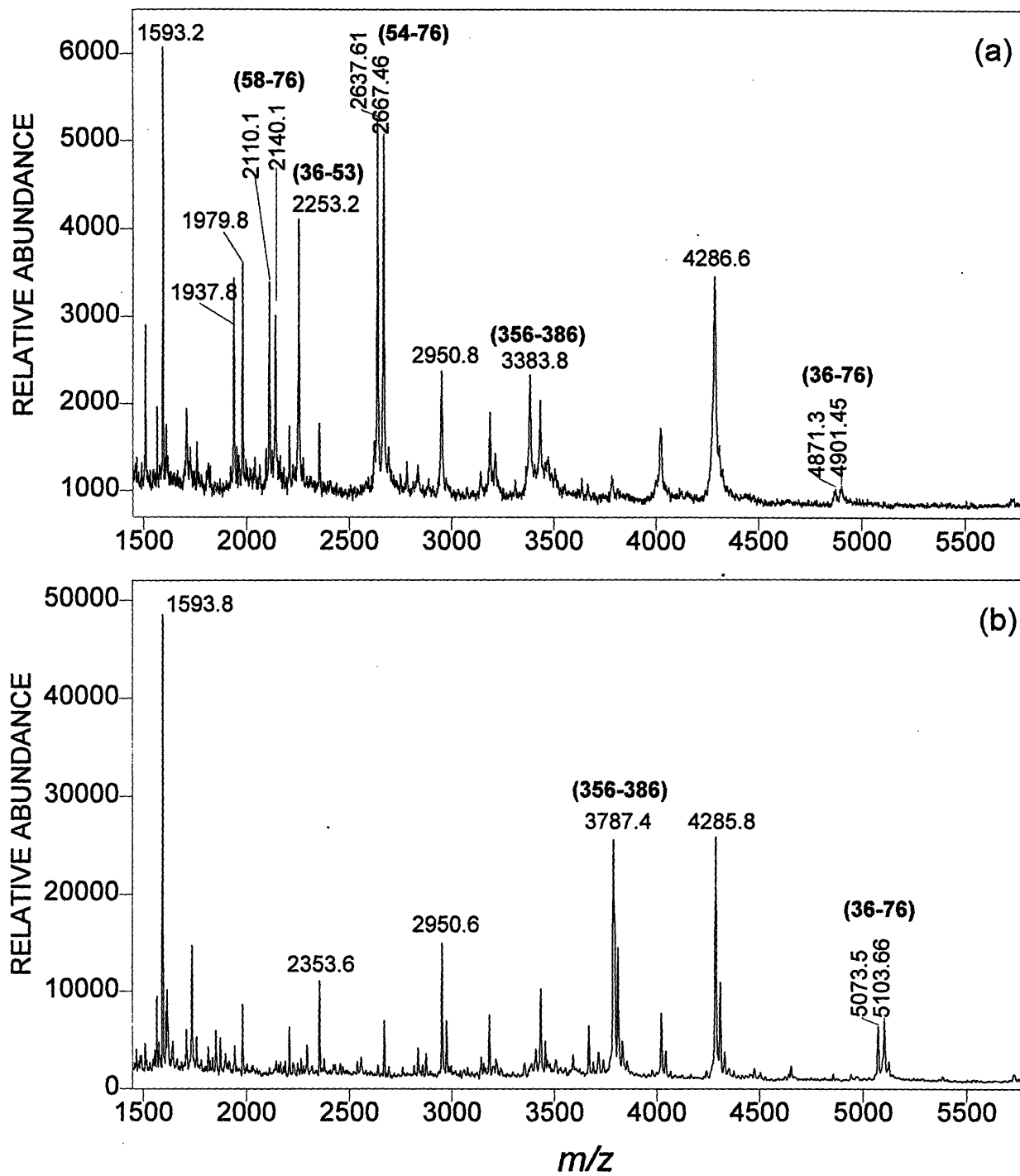


Figure II.23. The MALDI spectra of fraction T15 after deglycosylation with PNGase F (a) and Endo H (b).

During PNGase F deglycosylation this peptide was apparently hydrolyzed to the pair at  $m/z$  2637.6/2667.5, corresponding to peptide 54-76 (with both replaced amino acids), and at  $m/z$  2253.1, which represents the remainder (peptide 36-53). These fragments must have been produced by chymotryptic cleavage between Leu<sup>53</sup> and Phe<sup>54</sup>. A non-specific cleavage has occurred at the carboxyl side of His<sup>57</sup> giving rise to the doublet at  $m/z$  2110.1/2140.1 which was assigned to peptide 58-76. This somewhat surprising result is reproducible and was not observed earlier, so it was thought that PNGase F might have some proteolytic activity to which peptide 36-76 is very sensitive (see discussion of fractions T16-18). This unexpected cleavage did not occur with Endo H where peptide 36-76 was found at  $m/z$  5075.1/5105.3 as expected (Figure II.23b). On the other hand, after PNGase F deglycosylation another new peak appeared at  $m/z$  3383.8 which was assigned to peptide 356-386 (calc. 3383.7 for two occupied sequons) containing sequons 12 and 13 (Asn<sup>365</sup>-Ser-Thr and Asn<sup>379</sup>-Thr-Thr). The signals at  $m/z$  1937.8 and 1979.8 could not be identified.

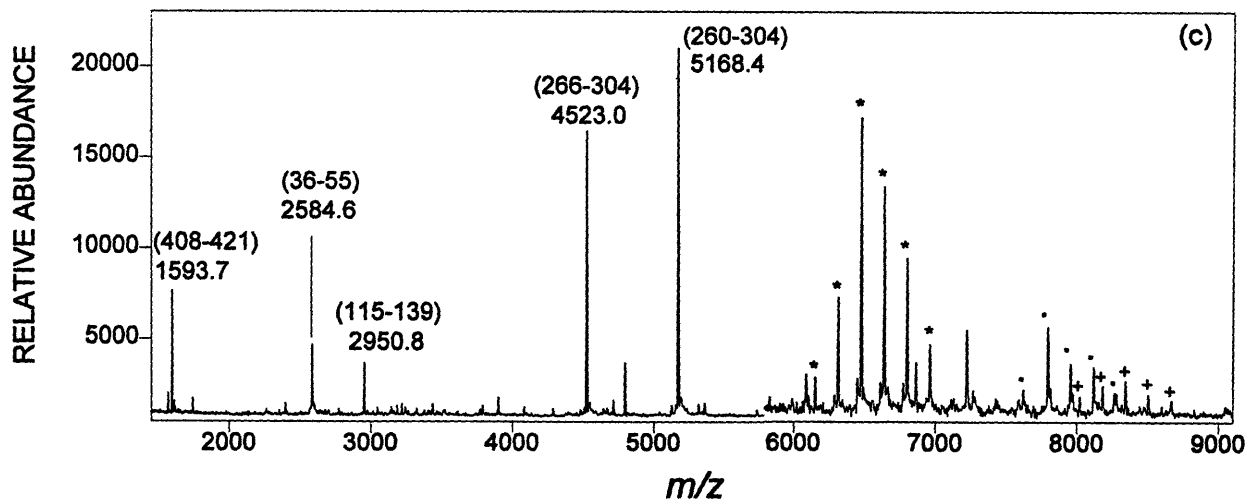
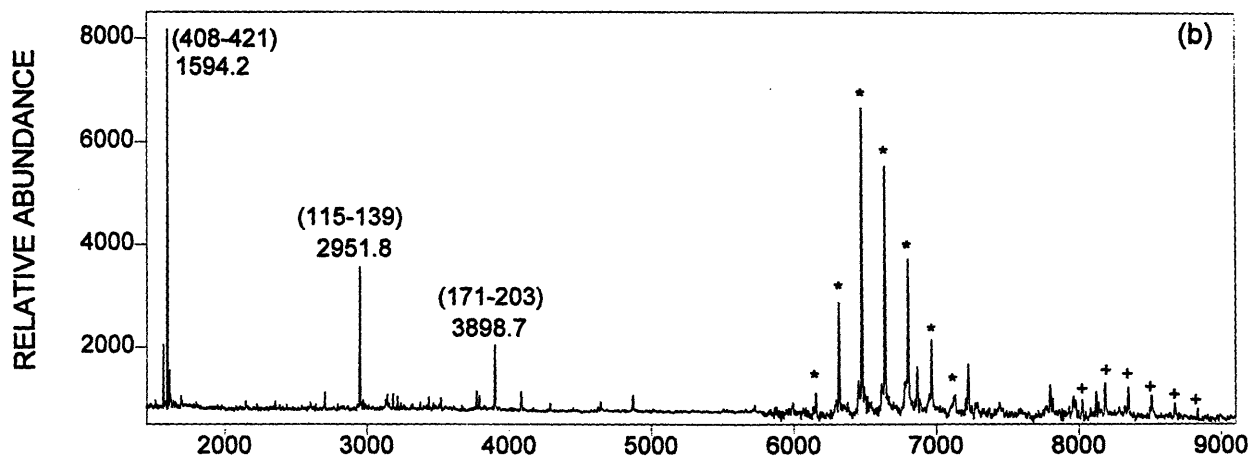
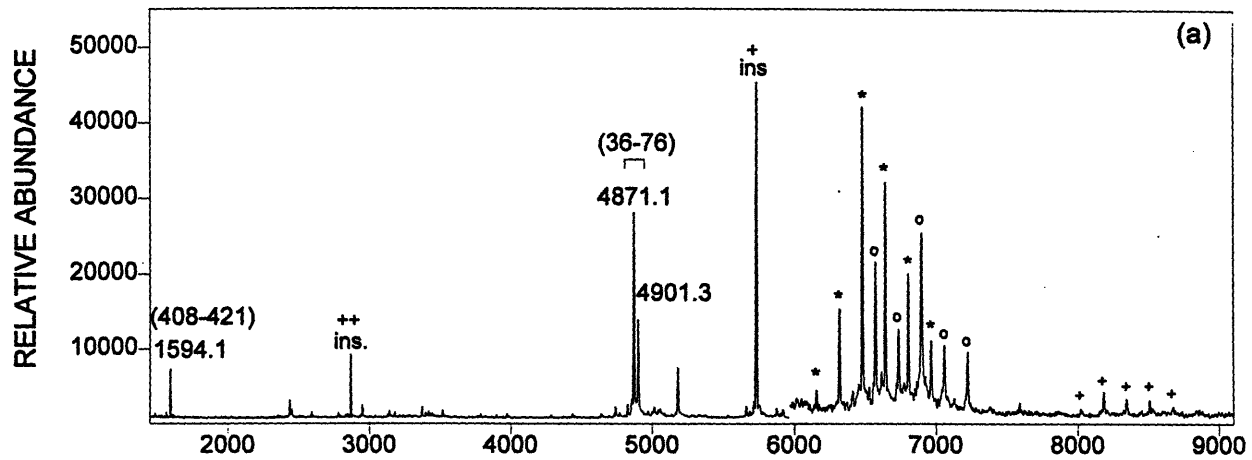
The corresponding Endo H peptide 356-386 was found at  $m/z$  3789.2 which confirmed that sequons 12 and 13 were both glycosylated. However, the glycopattern corresponding to peptide 356-386 with two oligosaccharide chains was apparently too weak to be detected (it was eventually found in the Lys-C digest). In contrast, the glycopattern observed at  $m/z$  7856.4, 8019.8, 8181.5, 8342.8, 8505.5, 8666.8, and 8828.3 could be assigned to peptide 220-259 (calc. 4449.9) with two oligosaccharide chains totalling Man<sub>16-21</sub>GlcNAc<sub>4</sub> (calc. 3407.0-4217.7) at sequons 8 and 9. However, after removal of the sugars with PNGase F and Endo H, the corresponding deglycosylated peptide could not be detected for reasons unknown.

The MALDI mass spectra of fractions T16, T17, and T18 are shown in Figures II.24a, II.24b, and II.24c, respectively. The peptides which do not contain a sequon and the unglycosylated free peptide (36-76) have been indicated and labeled in the spectra. The high mass range (above  $m/z \sim 1,600$ ) of the spectra have been expanded in the Y-axis to show the glycopattern(s) more clearly. All three spectra exhibit the same major glycopatterns which indicates that these glycopeptides eluted slowly over these three fractions. Therefore, equal portions of fractions T16-18 were mixed and deglycosylated with PNGase F and Endo H. The MALDI spectrum of the PNGase F digest (Figure II.25a) indicates that many of the expected peptides may have been proteolytically hydrolyzed, because the abundant higher mass signals at  $m/z$  4871.1 (see Figure II.24a), 4523.0 and 5168.4 (see Figure II.24c) had disappeared and there was no signal that could be correlated with any of the glycopatterns present in Figure II.23a-c. The only peak which may correspond to PNGase F deglycosylation is the one at  $m/z$  3383.6 which matches peptide 356-386 (calc. 3383.6 for two Asn→Asp conversions), but none of the glycopattern in Figure II.24a-c belongs to this glycopeptide.

It should be pointed out that this phenomenon of peptide hydrolysis was only observed during the deglycosylation of fractions T15-18 which are the last four HPLC fractions (see Figure II.2). Since these fractions were collected from an HPLC separation using a linear gradient after the solvent mixture reached 40% of ACN, remaining trypsin and its autolysis products of chymotryptic activity (Keil-Dlouhá et al., 1971) would co-elute. These traces of proteases would be activated during the PNGase F digestion which requires similar conditions (pH 7.8 at 37 °C). This assumption was further supported by the Endo H deglycosylation experiment. Because the conditions for this enzyme (pH 5.8 at 37 °C) is less suitable for these proteases, the corresponding Endo H peptides

Figure II.24 (next page). The MALDI spectra of HPLC fractions T16 (a), T17 (b), and T18 (c). Region above  $\sim m/z$  6,000 expanded ( $5 \times$ ) in the Y-axis. The peaks labeled "ins" in Figure II.24a indicates the singly and doubly charged signal of insulin which was used as the internal calibration standard. In all three spectra, the same glycopeptide was marked with the same label. For example, "★" indicates the glycopattern of peptide 457-497 (which contains sequon 14) with  $\text{Man}_{8-13}\text{GlcNAc}_2$ , the strongest signal ( $m/z$  6477.6) in this glycopattern corresponds to peptide 457-497 with  $\text{Man}_{10}\text{GlcNAc}_2$ ; "°" indicates the glycopattern of peptide 36-76 with  $\text{Man}_{9-12}\text{GlcNAc}_2$ ; "+" indicates the glycopattern of peptide 220-259 (including sequons 8 and 9) with  $\text{Man}_{17-21}\text{GlcNAc}_4$ . The glycopattern labeled "•" could not be identified.





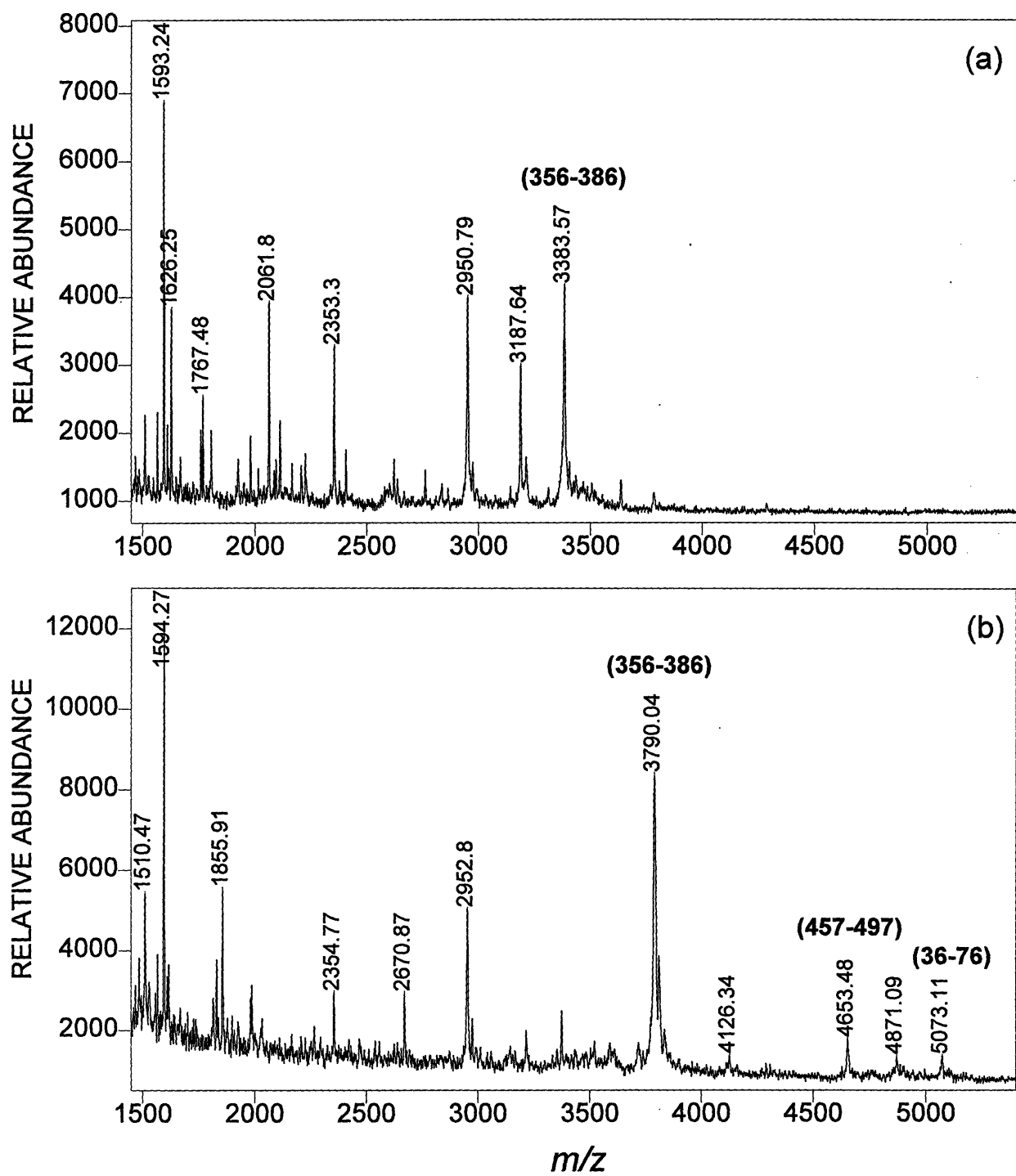


Figure II.25. The MALDI spectra of the mixture of T16, T17, and T18 after deglycosylation with PNGase F (a) and Endo H (b).

were found at  $m/z$  5073.1, 4654.2 and 3790.4 (Figure II.25b) and matched peptide 36-76 (calc. 5073.5), 457-497 (calc. 6478.6 with one GlcNAc) and 356-386 (calc. 3788.1 with two GlcNAc) even though some of the strong signals in Figures II.24a-c are still missing (i.e. have been hydrolyzed). It turned out that the major glycopattern (labeled "★" in Figures II.24a-c) corresponds to the tryptic-chymotryptic peptide 457-497 (which contains sequon 14 at Asn<sup>493</sup>-Met-Thr) with Man<sub>8-14</sub>GlcNAc<sub>2</sub>. The glycopattern with "◦" label (Figure II.24a) belongs to peptide 36-76 (with one modified amino acid, see result and discussion below) with Man<sub>8-12</sub>GlcNAc<sub>2</sub>.

A pair of peaks at 4871.1 and 4901.3 spaced 30 Da apart (Figure II.24a) corresponds to the unglycosylated free peptide 36-76 (which contains sequon 2) with one and two modified amino acid(s), respectively (their corresponding glycopeptides were found in fraction T15 and discussed there). This indicates that sequon 2 is not fully glycosylated. In deglycosylated fraction T15 (Figure II.23a),  $(M+H)^+ = m/z$  4871.1 and  $(M+H)^+ = m/z$  4901.3 were of the same intensity. However, in fraction T16 the latter is 50% less than the former. In addition, the glycopattern (labeled "◦" in Figure II.24a) corresponding to the former was also found in T16. In contrast to the twin-peak glycopatterns shown in Figure II.22b, the signals of this glycopattern indicates only one peptide. It is concluded that the glycosylated peptide  $(M+H)^+ = m/z$  4871.1 was eluted a little bit later than the glycosylated peptide  $(M+H)^+ = m/z$  4901.3, and the former was collected in two fractions (T15 and T16) while the latter was only collected in fraction T15. Because these free peptides were only found in fraction T16, their relative abundance of 2:1 should correspond more closely to the native ratio.

## 2. *Lys-C Digestion of native intact protein:*

In order to confirm the conclusions drawn from the tryptic digest discussed in the previous section and to confirm some of the conclusions thereof, intact external invertase was digested with Lys-C and the digest fractionated by HPLC. Twenty-four fractions were collected (see Figure II.3) and analyzed by MALDI-TOF mass spectrometry. Glycopatterns were found in fractions L6, L7, L8, L12, L13, L14, L16, L18, and L19.

The single signal at  $m/z$  905.8 in the MALDI spectrum of fraction L6 (Figure II.26) corresponds to the Lys-C peptide 507-513 (calc. 906.1) which is the C-terminal peptide of invertase. The glycopattern dominating this spectrum represents the glycosylation distribution at sequon 7 (Asn<sup>247</sup>) in the Lys-C peptide 140-154 with oligosaccharides of Man<sub>5-47</sub>GlcNAc<sub>2</sub> (for details refer to the data from T6, Figure II.5). It should be noted that the signals at the low and the high mass ends of the glycopattern are often weak. Therefore, the spectrum of the Lys-C peptide levels off at Man<sub>47</sub> while that of the tryptic peptide (Figure II.5b) continues to 58 mannoses. For this reason the "glycan lengths" of each sequon shown in Table II.4 (see the conclusion section below) will be the largest values observed. Since peptide Asn<sup>140</sup>-Arg<sup>151</sup> is not produced when using Lys-C, the MALDI spectrum of fraction L6 exhibited only a single glycopattern (Figure II.26) as expected. This observation confirmed the assignments of the two overlapping glycopatterns in the MALDI spectrum of T6 (see Figure II.5b, II.5c) as glycopeptides 140-151 and 140-154. A close comparison of the abundance pattern of the peaks labeled "★" in Figure II.5b and that represented by Figure II.26 (pure glycopeptide 140-154) demonstrates the reproducibility of this bimodal pattern.

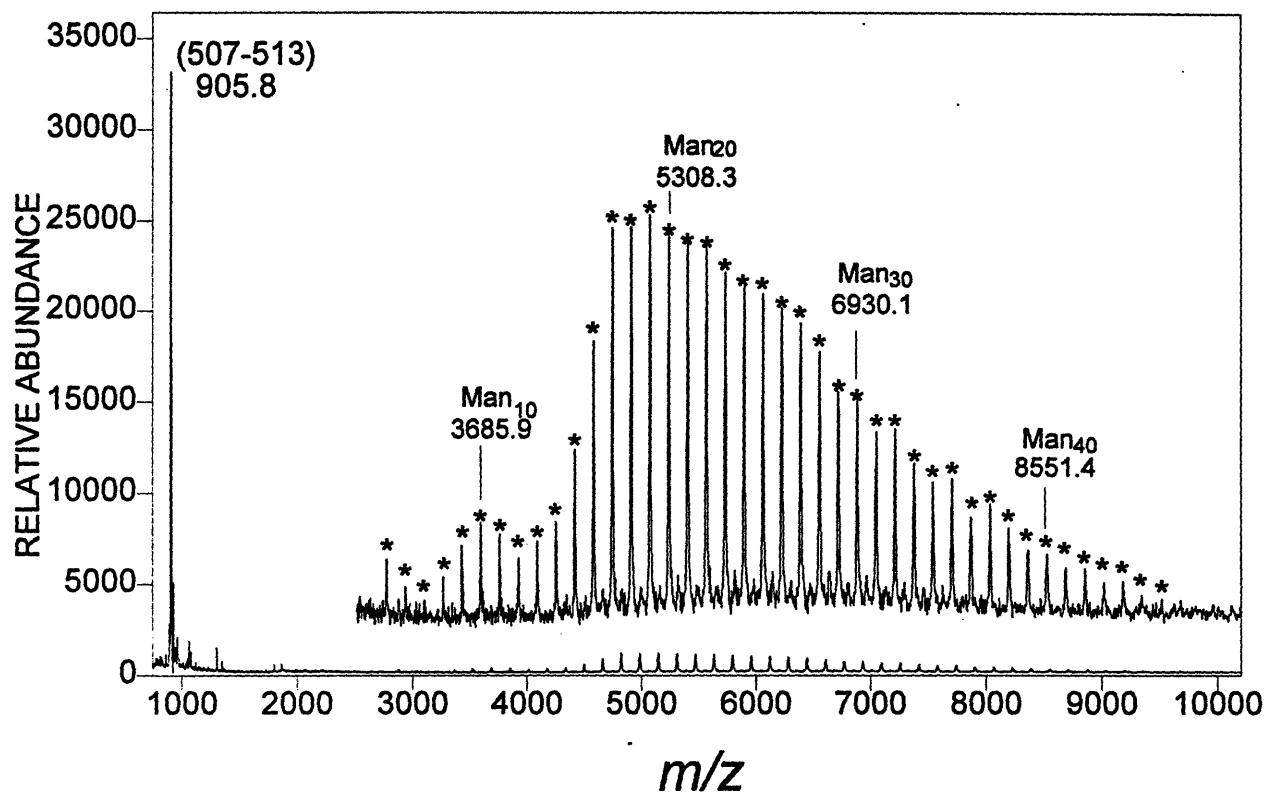


Figure II.26. The MALDI spectrum of the HPLC fraction L6 of the Lys-C digest.

The MALDI spectrum of fraction L7 shown in Figure II.27 exhibits two glycopatterns. The corresponding peptides were identified by digestion with PNGase F and Endo H as peptides 140-154 (containing sequon 7) and 1-18 (containing sequon 1). These results further confirm the assignments made from the tryptic digest (fractions T6 and T7). The series of signals labeled "★" represent peptide 140-154 with the shorter oligosaccharide chains ( $\text{Man}_{5-19}\text{GlcNAc}_2$ ) at sequon 7, while the other (labeled "•") corresponds to the small portion of peptide 1-18 with the longer glycan chains ( $\text{Man}_{15-47}\text{GlcNAc}_2$ ) at sequon 1. As expected, the majority of peptide 1-18 with the  $\text{Man}_{4-42}\text{GlcNAc}_2$  was found in the next fraction (L8, Figure II.28) with a glycosylation distribution profile identical to that observed in fraction T7 of tryptic digest (see Figure II.6b). Again, peptide 1-18 with the oxidized methionine in position 2 was also found both in fractions L7 and L8. The sugar-free peptide 140-154 found in fraction L8 (Figure II.28) indicates sequon 7 has only been partially glycosylated. The degree of glycosylation was estimated (as discussed above) to be 90%.

The MALDI spectrum of fraction L12 shown in Figure II.29a exhibits a "wide range" glycopattern consisting of 66 peaks in the mass range of 5,000 to 14,300 Da. This multiplet collapsed to a single peak at  $m/z$  2604.9 (Figure II.29b) after PNGase F digestion which matches the Lys-C peptide 332-355 (calc. 2605.9) containing sequons 10 ( $\text{Asn}^{337}\text{-Ile-Ser}$ ) and 11 ( $\text{Asn}^{350}\text{-Thr-Thr}$ ). The corresponding signal detected at  $m/z$  3010.0 (Figure II.29c) after Endo H deglycosylation (calc. 3010.3) is about 405 Da higher and indicates that both sequons 10 and 11 are glycosylated. Therefore, the glycopattern presented in Figure II.29a belongs to peptide 332-355 with a total of  $\text{Man}_{13-67}\text{GlcNAc}_4$ , the sum of sequons 10 and 11. It should be recalled that the glycopattern of  $\text{Man}_{4-14}\text{GlcNAc}_2$  at sequon 10 had been deduced from the data of fraction T10 of the tryptic digest

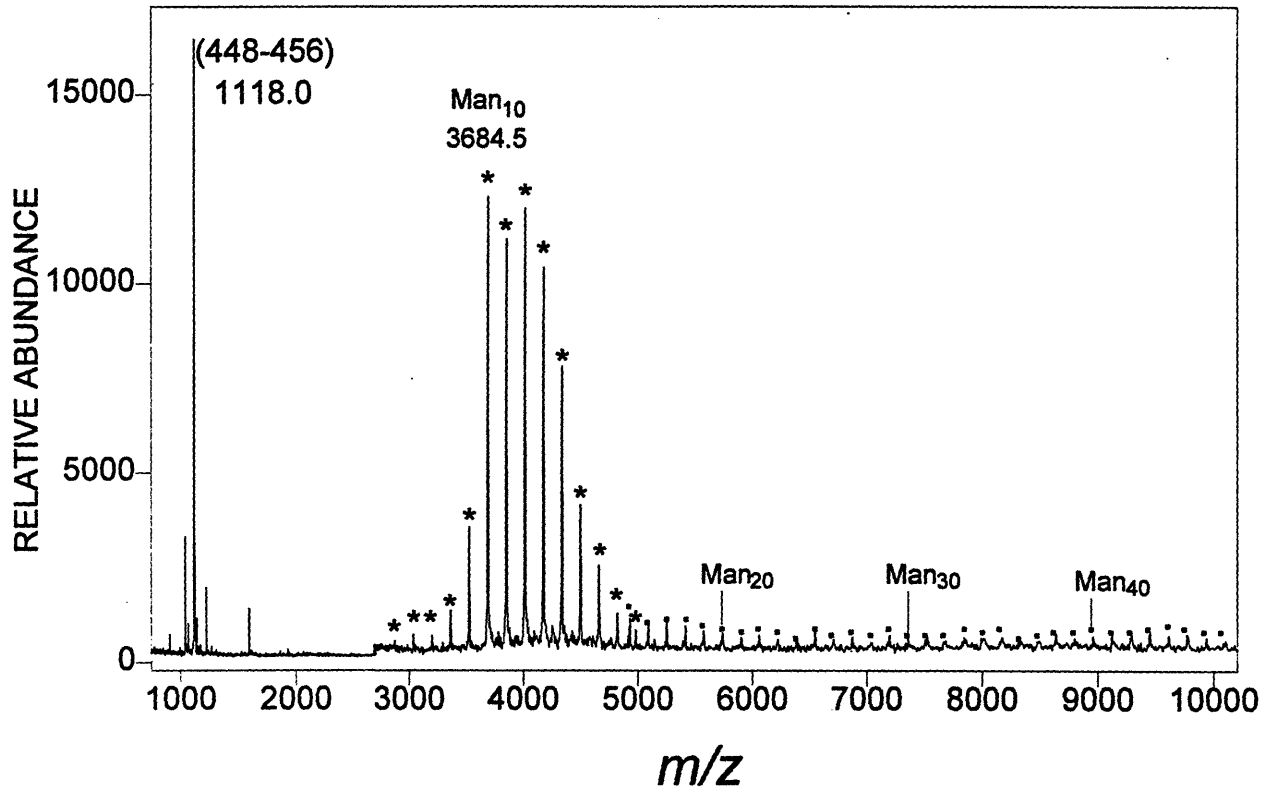


Figure II.27. The MALDI spectrum of the HPLC fraction L7 of the Lys-C digest.

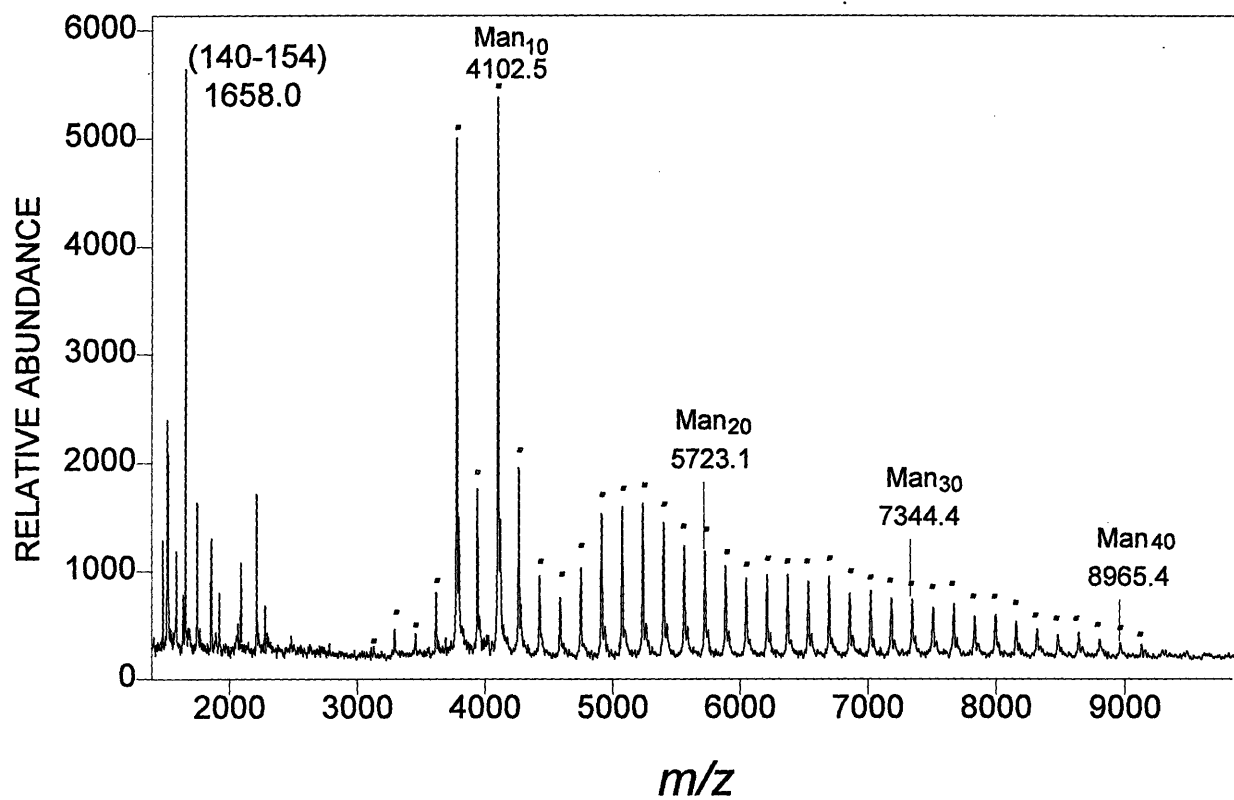


Figure II.28. The MALDI spectrum of the HPLC fraction L8 of the Lys-C digest.



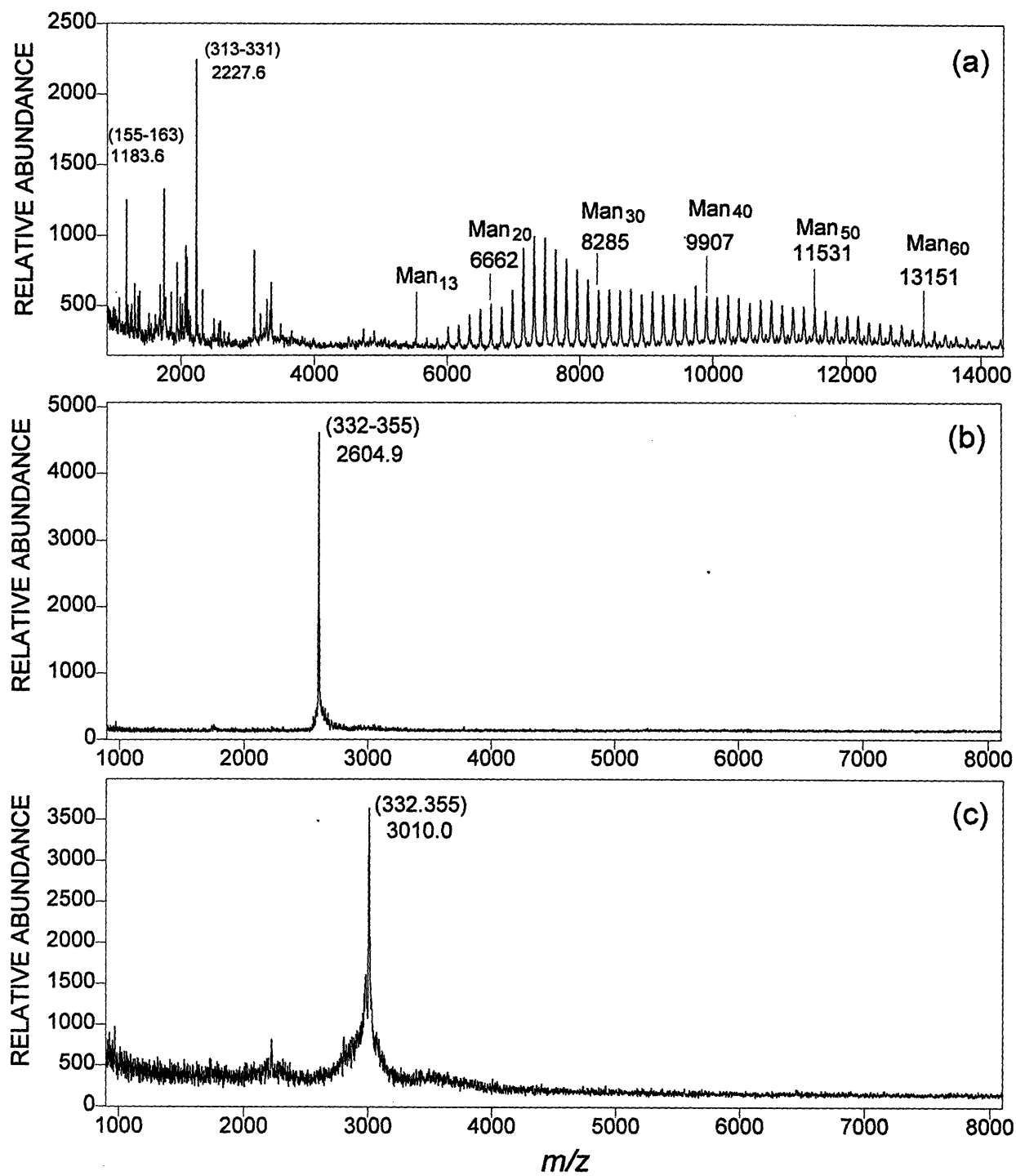


Figure II.29. The MALDI spectra of the HPLC fraction L12 of the Lys-C digest before (a) and after deglycosylated with (b) PNGase F and (c) Endo H.

(see Figure II.21a) but the glycosylation at sequon 11 had not been determined. This can now be done by subtracting the individual glycopattern of sequon 10 (exhibited in Figure II.21a) from the total glycopattern of sequons 10 + 11 as will be discussed later.

Two glycopatterns were found in the mass range of 4,800 to 7,800 Da of the MALDI spectrum of fraction L13 (Figure II.30a). After PNGase F deglycosylation of L13, a single peak at  $m/z$  2604.8 was detected (Figure II.30b). However, two peaks, at  $m/z$  2806.7 and 3009.8, were observed in the MALDI spectrum acquired after Endo H digestion (Figure II.30c). All these data together indicate that this glycopeptide contains amino acids 332-355 (calc. 2604.9 for 1N→D and 2605.9 for 2N→D) bearing one GlcNAc (calc. 2807.1) and two GlcNAc (calc. 3010.3), respectively, at sequons 10 (Asn<sup>337</sup>-Ile-Ser) and/or 11 (Asn<sup>350</sup>-Thr-Thr). These sequons are (at least one), therefore, not fully glycosylated. In Figure II.30a, the masses of the more abundant glycopattern (labeled "★") match to peptide 332-355 with total of Man<sub>13-23</sub>GlcNAc<sub>4</sub> (both sequons are occupied, its longer carbohydrate components has been detected in L12), and the masses of the weaker glycopattern (labeled "•") match to peptide 332-355 with Man<sub>13-29</sub>GlcNAc<sub>2</sub> (either sequon 10 or 11 is occupied).

A single simple glycopattern in the range of  $m/z$  4,000 to 5,000 was observed in the MALDI spectrum of fraction L14 (Figure II.31) which collapsed to the peak at  $m/z$  2605.5 after PNGase F deglycosylation. The corresponding Endo H peptide observed at  $m/z$  2808.0 indicating that the glycopattern shown in Figure II.31 also belongs to peptide 332-355 in which one sequon is occupied by Man<sub>7-11</sub>GlcNAc<sub>2</sub>. The same mono-glycosylated peptide with longer oligosaccharide chains had

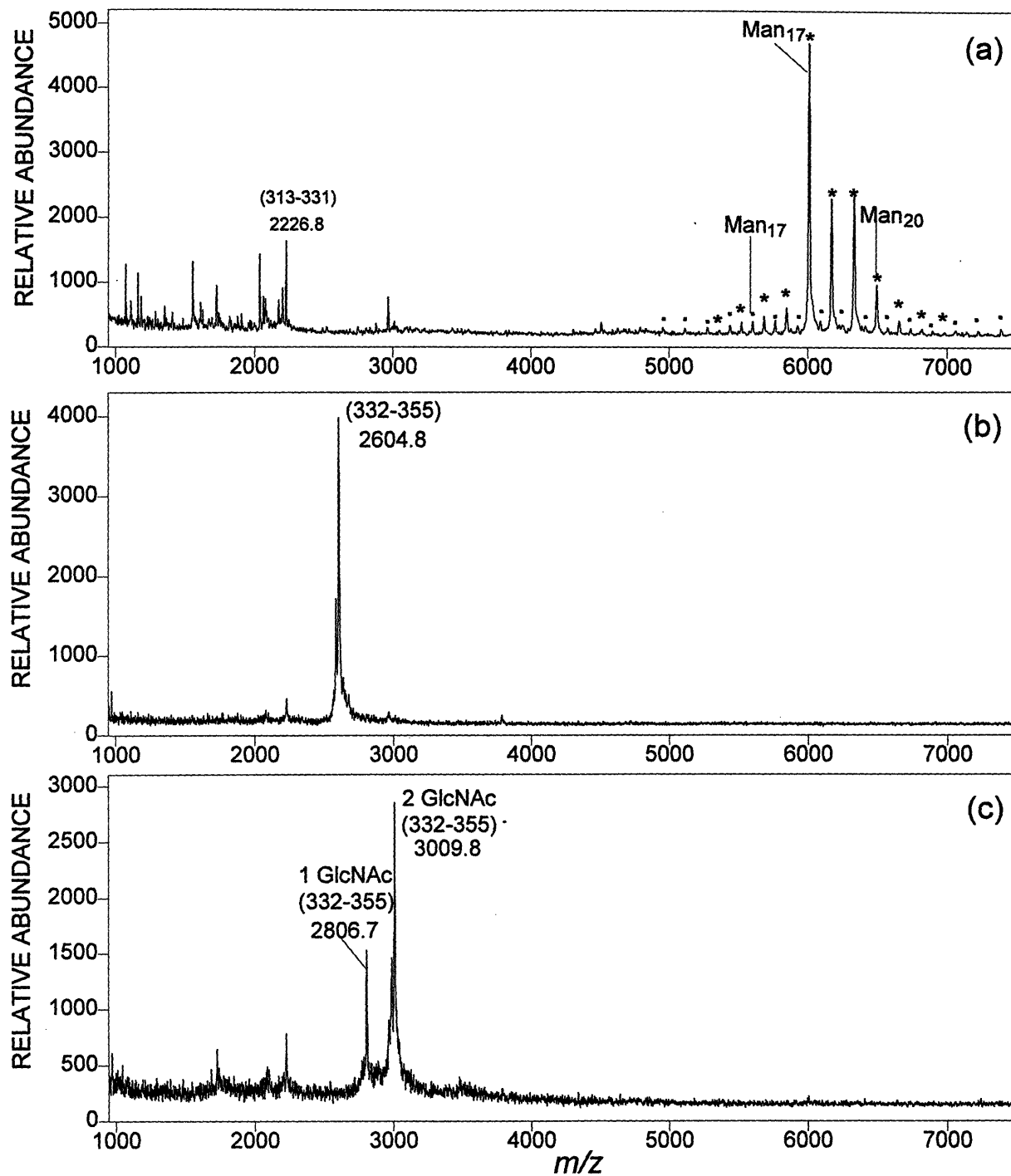


Figure II.30. The MALDI spectra of the HPLC fraction L12 of the Lys-C digest before (a) and after deglycosylated with (b) PNGase F and (c) Endo H.

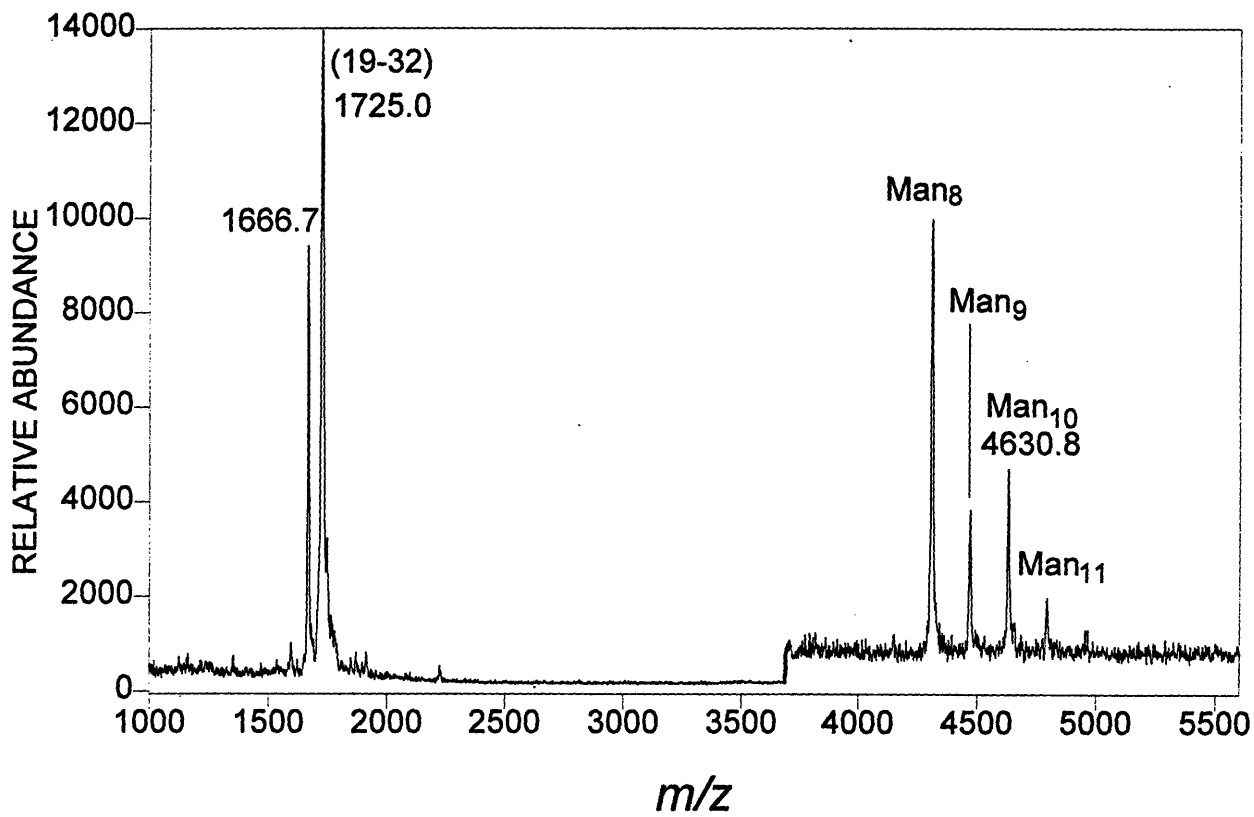
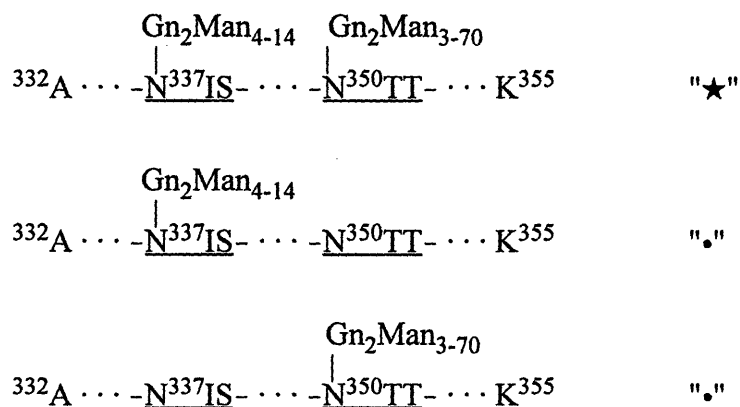


Figure II.31. The MALDI spectrum of the HPLC fraction L14 of the Lys-C digest. The region above  $m/z$  3,700 was expanded ( $3.5 \times$ ) in the Y-axis.

eluted in fraction L13 (see above).

Up to this point, peptide 332-355 (which contains sequons 10 and 11) has been found to elute over fractions 12, 13, and 14 with decreasing size of oligosaccharides. With both sequons occupied this peptide eluted in fraction L12 with a larger carbohydrate ( $\text{Man}_{13-78}\text{GlcNAc}_4$ ) than when eluted in fraction L13 ( $\text{Man}_{13-23}\text{GlcNAc}_4$ ). Similarly, the mono-glycosylated peptide collected in L13 has a total of  $\text{Man}_{13-29}\text{GlcNAc}_2$  while the one collected later in fraction L14 contains only  $\text{Man}_{7-11}\text{GlcNAc}_2$ . In order to obtain the complete glycopattern at sequons 10 and 11 (mono- and bi-occupied), fractions L12, L13, and L14 were combined and the MALDI spectrum of the mixture exhibits two glycopatterns with the almost identical profile (Figure II.32a, the high mass region expanded in Figure II.32b). The masses of the glycopattern signals (labeled "★") match peptide 332-355 with 12 to 78 mannoses and 4 GlcNAc, both sequons occupied, while the masses of those series signals (labeled "•") belong to peptide 332-355 with 3 to 70 mannoses and 2 GlcNAc, either sequons



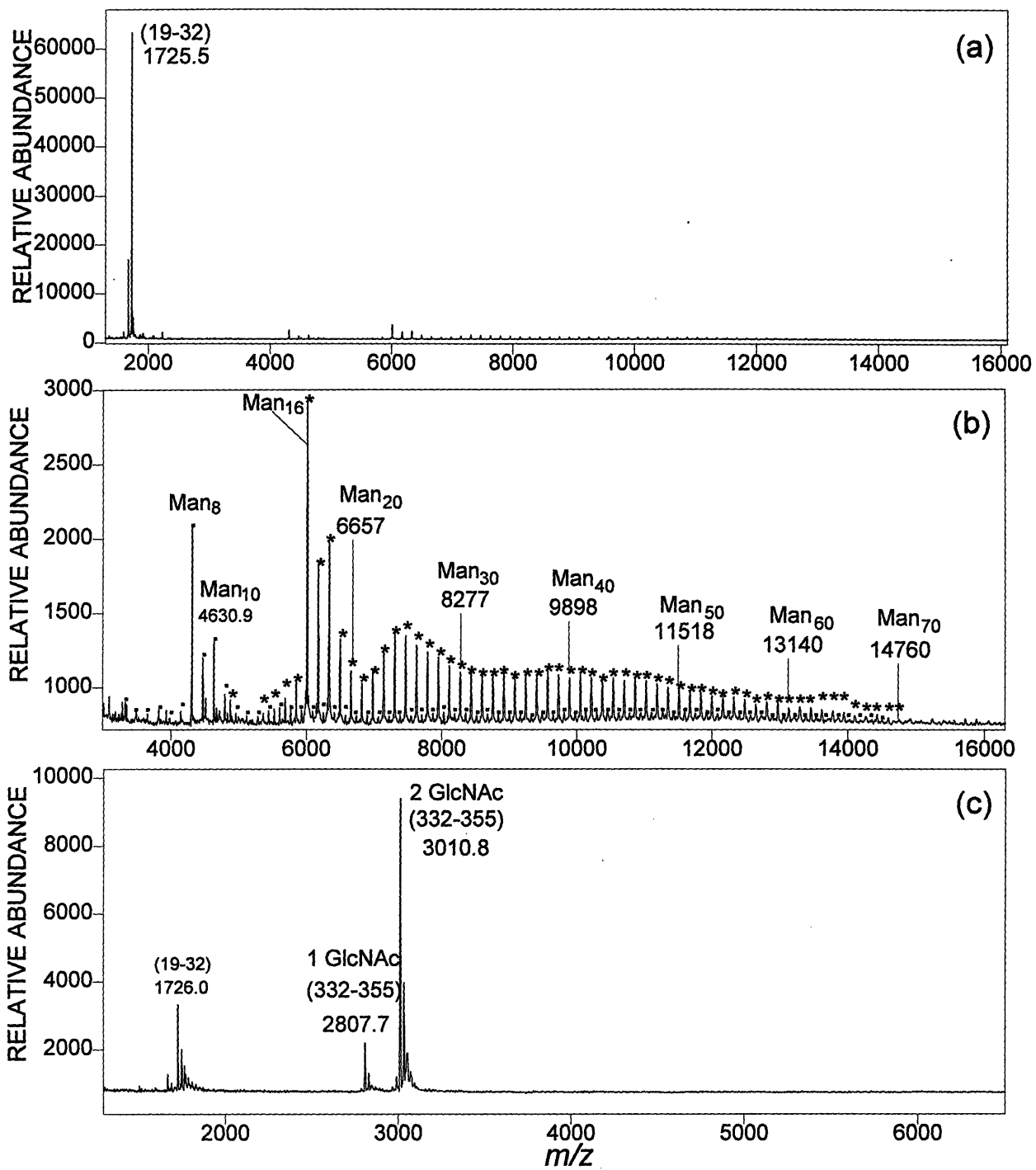


Figure II.32. (a) The MALDI spectra of the mixture of fractions L12, L13, and L14. (b) The expanded region of  $m/z$  3,200-16,200 in (a). (c) After Endo H deglycosylation.

occupied. Sequon 10 has been determined above individually to carry the oligosaccharides with the size ranging  $\text{Man}_{4-14}\text{GlcNAc}_2$ , but the glycopattern in MALDI spectrum (labeled "•" in Figure II.32b) revealed that the mono-glycosylation at sequon 10 or 11 has the range of  $\text{Man}_{3-70}\text{GlcNAc}_2$ . Obviously, peptide 332-355 with  $\text{Man}_3\text{GlcNAc}_2$  and those glycans consisting more than 14 mannoses must not be contributed by sequon 10, therefore, sequon 11 can be inferred to carry the oligosaccharide chains with the size ranging  $\text{Man}_{3-70}\text{GlcNAc}_2$ .

On the other hand, it is suggested that both sequons 10 and 11 could have the similar occupancy based on the two facts: First, the two glycopatterns (Figure II.32b) represent peptide 332-355 with mono-glycosylation (labeled "•") and with bi-glycosylations (labeled "★") are almost identical, therefore, their occupancy must be similar. If the occupancy at one sequon was significantly higher than the other the distribution profiles of these two glycopatterns would not be the same. Second, the ratio of their corresponding Endo H signals at  $m/z$  2807 (for only one occupied sequon) and 3010 (for two occupied sequons) is about 1:4 (averaged 10 spectra, one of these spectra shown in Figure II.32c) which indicates the average total occupancy of 85% for sequons 10 and 11. Since this value is close to the occupancy estimated above for sequon 10 (87%), sequon 11 should have the occupancy of 83%.

In order to obtain further information of sequon 11, the mixture of fractions L12, L13, and L14 was digested with Arg-C which is expected to produce two peptides, 332-346 (with sequon 10) and 347-355 (with sequon 11). However, only the glycopattern corresponding to peptide 332-346 (which contains sequon 10) was found when the digests were analyzed directly by MALDI MS (data

not shown), and the glycopattern observed from this experiment was identical, as expected, to the one observed in the fraction T10 of the trypsin digest (see Figure II.21a). It should be noted that the missing piece of peptide (347-355) was also not detected in the trypsin digest. The reasons for not finding this glycopeptide may be: (1) when the Arg-C digest was analyzed by MALDI, the heterogeneity of peptide 347-355 (glycans at sequon 11) is much higher than peptide 332-346 (glycans at sequon 10), so the signals of glycopeptide 347-355 will be suppressed by the much stronger signals from glycopeptide 332-346; (2) when HPLC fractionation was applied for the trypsin digest, this high hydrophilic peptide, <sup>347</sup>FATNTLLTK<sup>355</sup>, with such a large oligosaccharide portion may elute in the solvent front which was not collected. Based on these assumptions, the Arg-C digest was separated by HPLC and all fractions collected, including solvent front. However, glycopeptide 347-355 was still not detected. If it eluted in the solvent front, the high salt content may have suppressed its ionization. Permethylation of this glycopeptide followed by CHCl<sub>3</sub> extraction might have been necessary to overcome this problem.

Figure II.33a presents the MALDI spectrum of fraction L16. According to the spectrum, there is one or more (signals too weak to recognize) glycopattern(s) in the range of *m/z* 5,500 to 8,200. The corresponding PNGase F (Figure II.33b) and Endo H (Figure II.33c) deglycosylated peptides at *m/z* 3383.9 and 3789.0 revealed a hidden glycopattern of peptide 356-386 (calc. 3383.7 for 2N→D, 3788.1 for 2 GlcNAc) which contains sequons 12 and 13. The signal of the PNGase F peptide at *m/z* 4283.4 corresponds to peptide 226-265 (calc. 4282.6 for one occupied sequon) which is confirmed by the corresponding Endo H peptide at *m/z* 4485.7. The glycopattern (labeled "★") observed in Figure II.33a indicates Man<sub>8-18</sub>GlcNAc<sub>2</sub> at either sequon 8 or sequon 9.



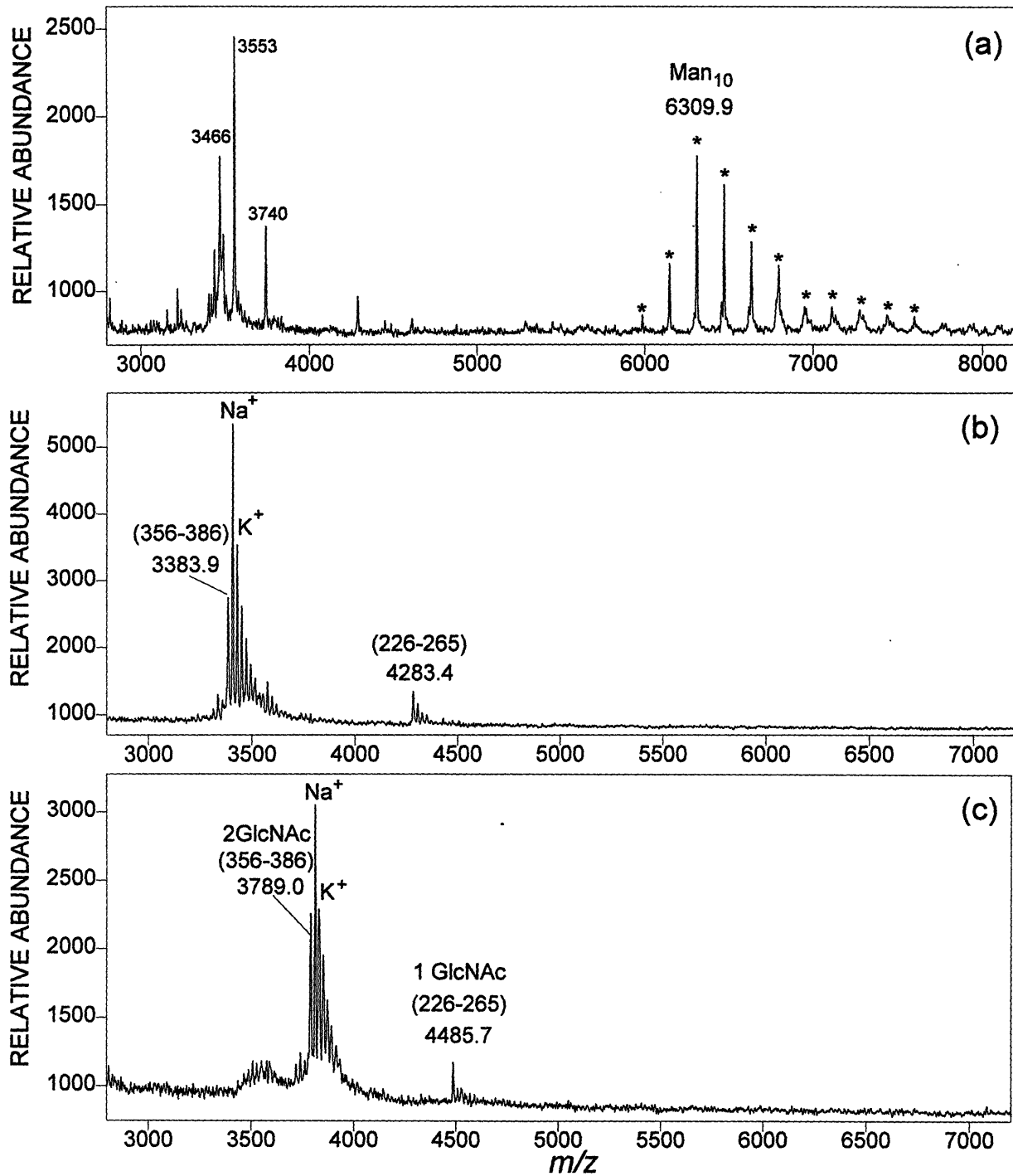


Figure II.33. (a) The MALDI spectra of the HPLC fraction L16 of the Lys-C digest. (b) and (c) Spectra acquired after deglycosylation with PNGase F and Endo H, respectively, without desalting.

In order to isolate sequons 12 and 13 in peptide 356-386, 8 and 9 in peptide 226-265, fraction L16 was further digested with Glu-C and roughly separated by HPLC using an C4 guard column (see experimental section). Fifteen fractions were collected (Figure II.34) and analyzed by MALDI MS which revealed glycopattern(s) in fractions G4, G5, G7, and G11.

Figure II.35a, II.35b, and II.35c show the MALDI spectra of fraction G4, G4 after PNGase F digestion and after Endo H digestion, respectively. The signals at  $m/z$  1440.5, 1485.6, and 1687.3 observed after PNGase F digestion together with the corresponding signals at  $m/z$  1642.8, 1687.9, and 1888.9 after Endo H deglycosylation match peptides 374-386 (which contains sequon 13 at Asn<sup>379</sup>-Thr-Thr, calc. 1439.9), 253-265 (which contains sequon 9 at Asn<sup>256</sup>-Gln-Ser, calc. 1484.3), and 356-371 (which contains sequon 12 at Asn<sup>365</sup>-Ser-Thr, calc. 1686.6), respectively. Therefore, the two glycopatterns, labeled "★" and "•", in Figure II.35a represent peptide 374-386 with Man<sub>4</sub>-<sub>13</sub>GlcNAc<sub>2</sub> at sequon 13 (Asn<sup>379</sup>-Thr-Thr) and peptide 253-265 with Man<sub>4-11</sub>GlcNAc<sub>2</sub> at sequon 9 (Asn<sup>256</sup>-Gln-Ser), respectively. No glycopattern corresponding to peptide 356-371 (sequon 12 at Asn<sup>365</sup>) was observed in Figure II.35a, and one must assume that those signals were suppressed by the others. The signals at  $m/z$  1643.1 and 1890.8 in Figure II.35a fit to the GlcNAc-peptides 374-386 (with sequon 13) and 356-371 (with sequon 12), but their origin or formation is difficult to rationalize. These peptides would need to be sequenced to confirm that the assignments are correct.

The MALDI spectrum of fraction G5 exhibited two glycopatterns in the mass range of  $m/z$  2,600-4,500 (Figure II.36a) which has been enlarged ( $\times 2$ ) in the Y-axis. Figures II.36b and II.36c represent the spectra after PNGase F and Endo H deglycosylation, respectively, of this fraction. The

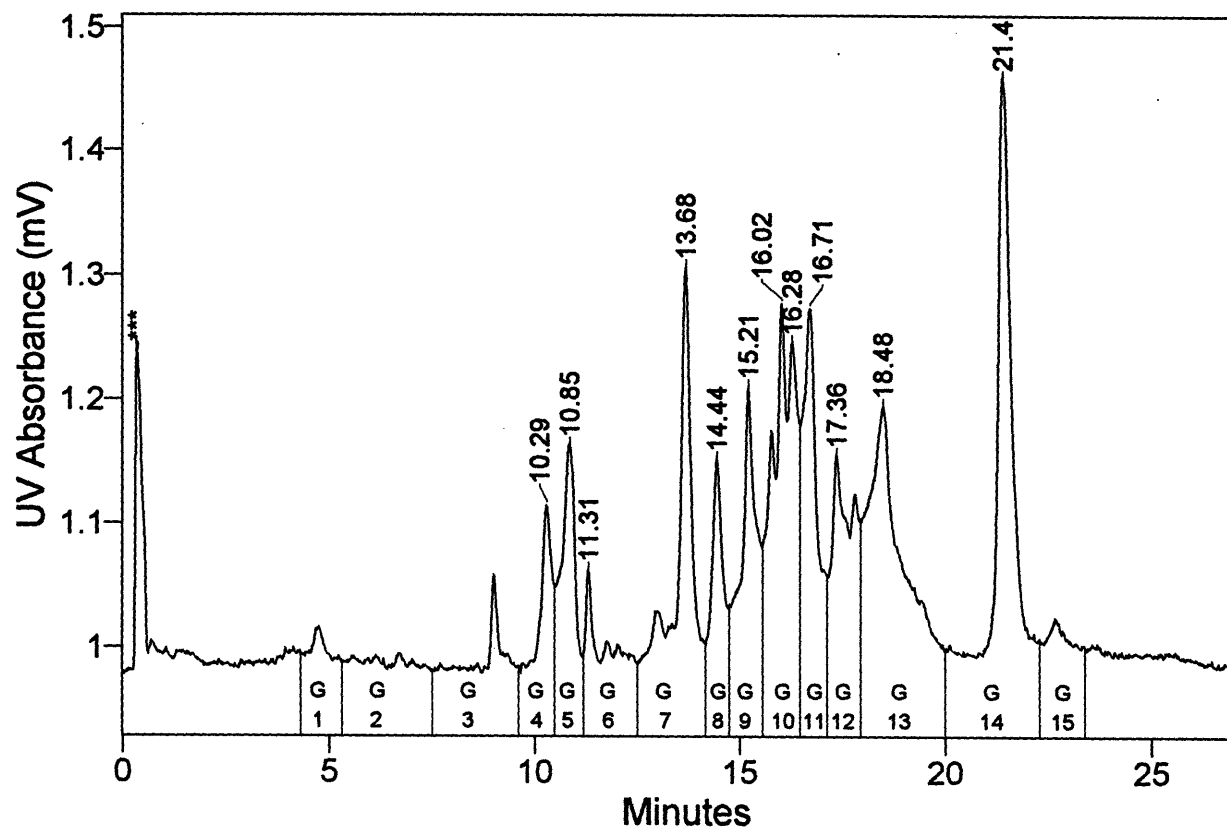


Figure II.34. HPLC trace of the Glu-C subdigest of fraction L16 (from the Lys-C digest). The fraction collected are marked.

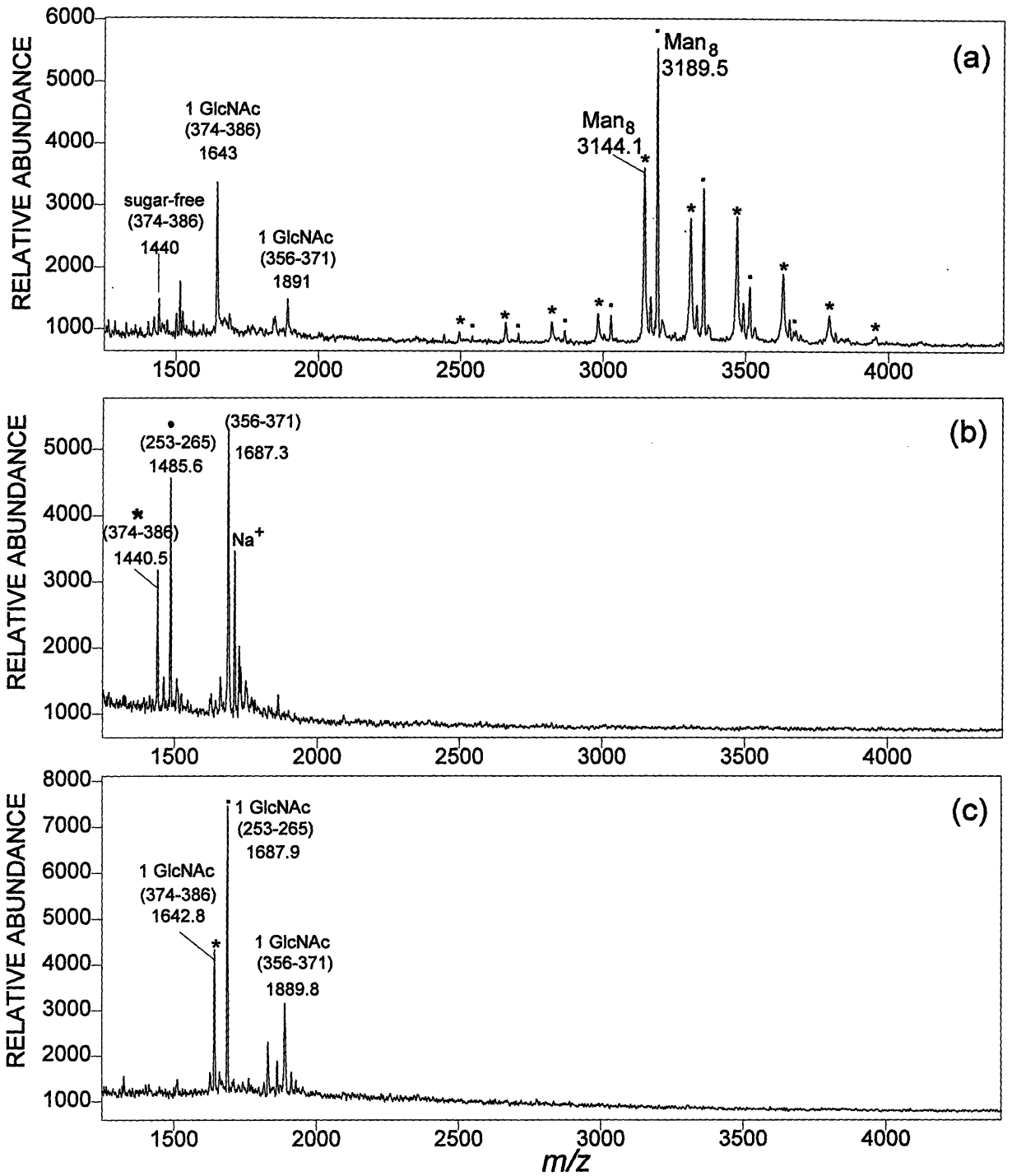


Figure II.35. The MALDI spectra of fraction G4 of the Glu-C subdigest of fraction L16 (from the Lys-C digest) before (a) and after deglycosylated with (b) PNGase F and (c) Endo H.

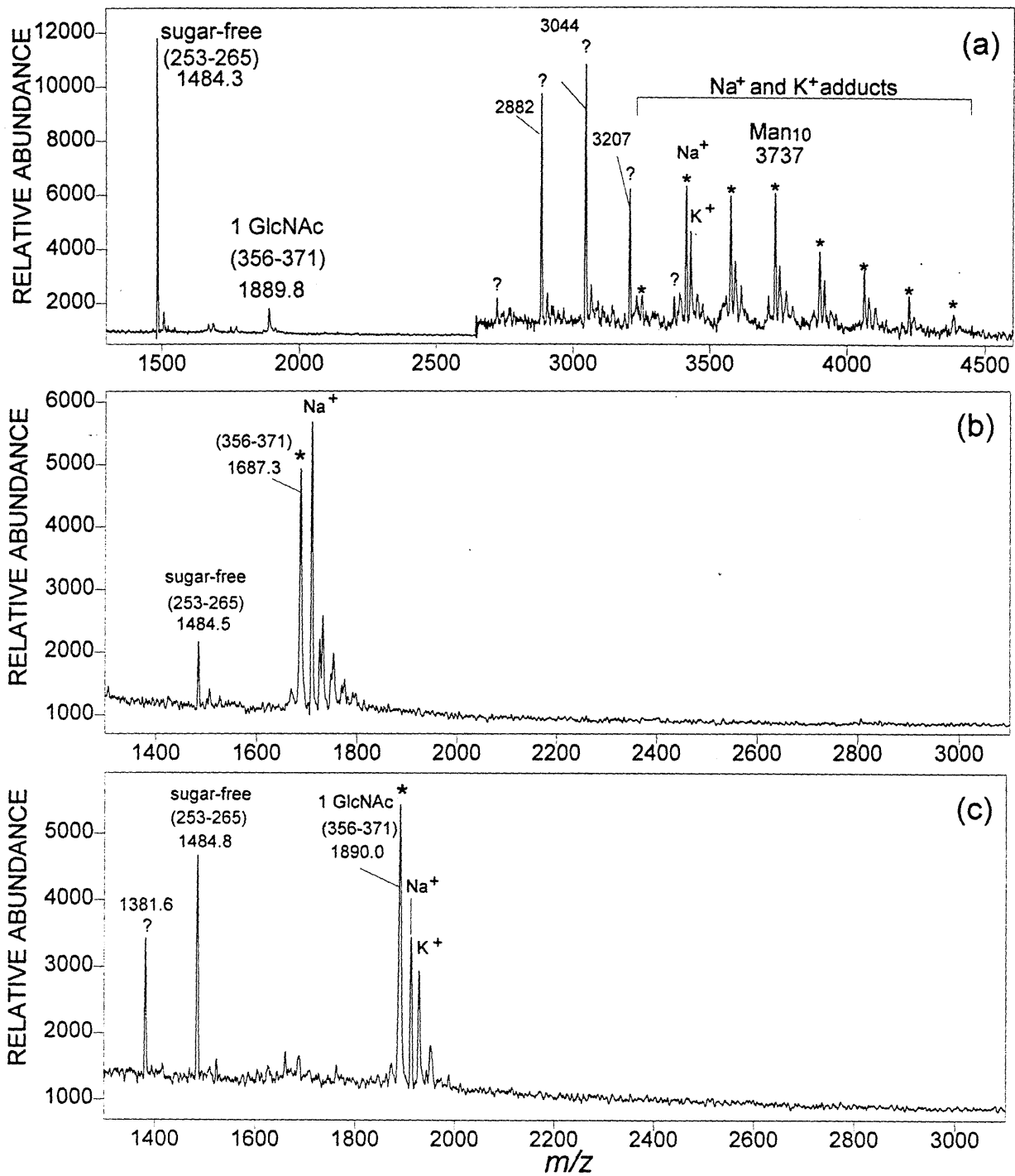


Figure II.36. The MALDI spectra of fraction G5 of the Glu-C subdigest of fraction L16 (from the Lys-C digest) before (a) and after deglycosylation with (b) PNGase F and (c) Endo H.

signals at  $(M+H)^+ = m/z$  1687.3 and  $(M+Na)^+ = m/z$  1709.8 correspond to the PNGase F peptide 356-371 (calc. 1686.7) which contains sequon 12, an assignment confirmed by the corresponding Endo H peptide  $(M+H)^+ = m/z$  1890.0. The glycopattern labeled "★" in Figure II.36a reveals that in this peptide  $Man_{7-14}GlcNAc_2$  is attached at sequon 12 (Asn<sup>365</sup>-Ser-Thr). It should be pointed out that the signals of the glycopattern (labeled "★") represent only the  $Na^+$  and  $K^+$  adducts which indicates a special affinity to the cation. The glycopattern labeled "?" and the corresponding Endo H peptide  $(M+H)^+ = m/z$  1381.6 (the corresponding PNGase F peptide signal was not detected) can not yet be identified. The low mass peak at  $m/z$  1484.2 in Figure II.36a matches unglycosylated peptide 253-265 indicating only partial glycosylation at sequon 9. The other weak signal at  $m/z$  1889.8 has already been identified in fraction G4 as GlcNAc-peptide 356-371.

The MALDI spectrum of fraction G7 exhibits a glycopattern (Figure II.37a) with both  $(M+H)^+$  and  $(M+Na)^+$  ions at  $m/z$  2,700-4,100. These multiplets collapsed to the peak at  $m/z$  1716.3 after PNGase F deglycosylation (Figure II.37b) which corresponds to peptide 372-386 (calc. 1715.9), containing sequon 13. This is confirmed by the peak at  $m/z$  1918.9 in the MALDI spectrum of fraction G7 after Endo H deglycosylation (Figure II.37c). It should be noted that the shorter sequon 13 containing peptide 374-386 has been identified in fraction G4 (Figure II.35a). As expected, the glycopatterns of these two peptides are very similar. In analogy a GlcNAc-peptide 372-386 was also found here ( $m/z$  1919.5).

The glycopattern in the mass range of  $m/z$  4,500 to 5,500 in the MALDI spectrum of fraction G11 (Figure II.38a) was identified as peptide 226-252 bearing  $Man_{8-13}GlcNAc_2$  by the PNGase F

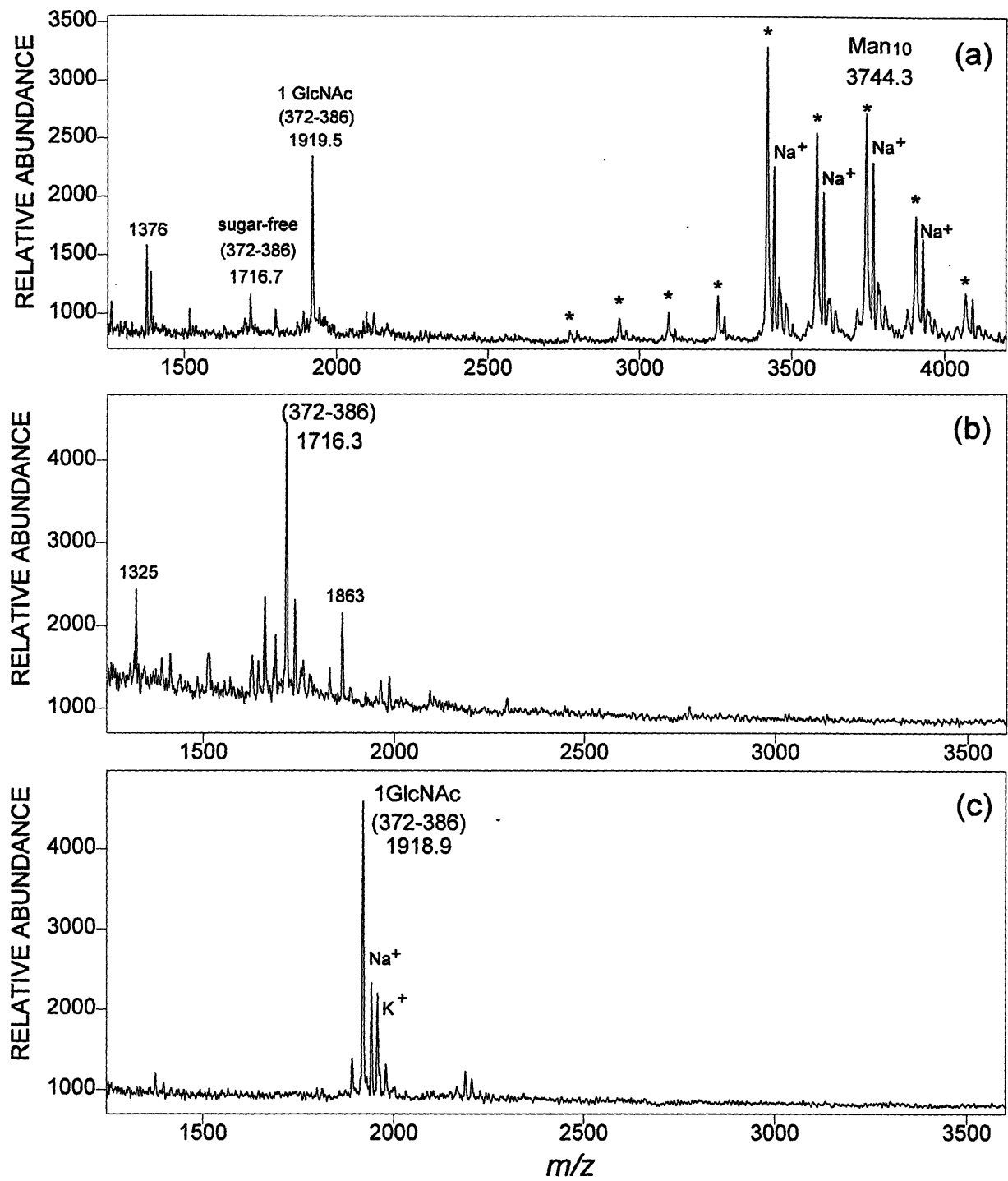


Figure II.37. The MALDI spectra of fraction G7 of the Glu-C subdigest of fraction L16 (from the Lys-C digest) before (a) and after deglycosylated with (b) PNGase F and (c) Endo H.

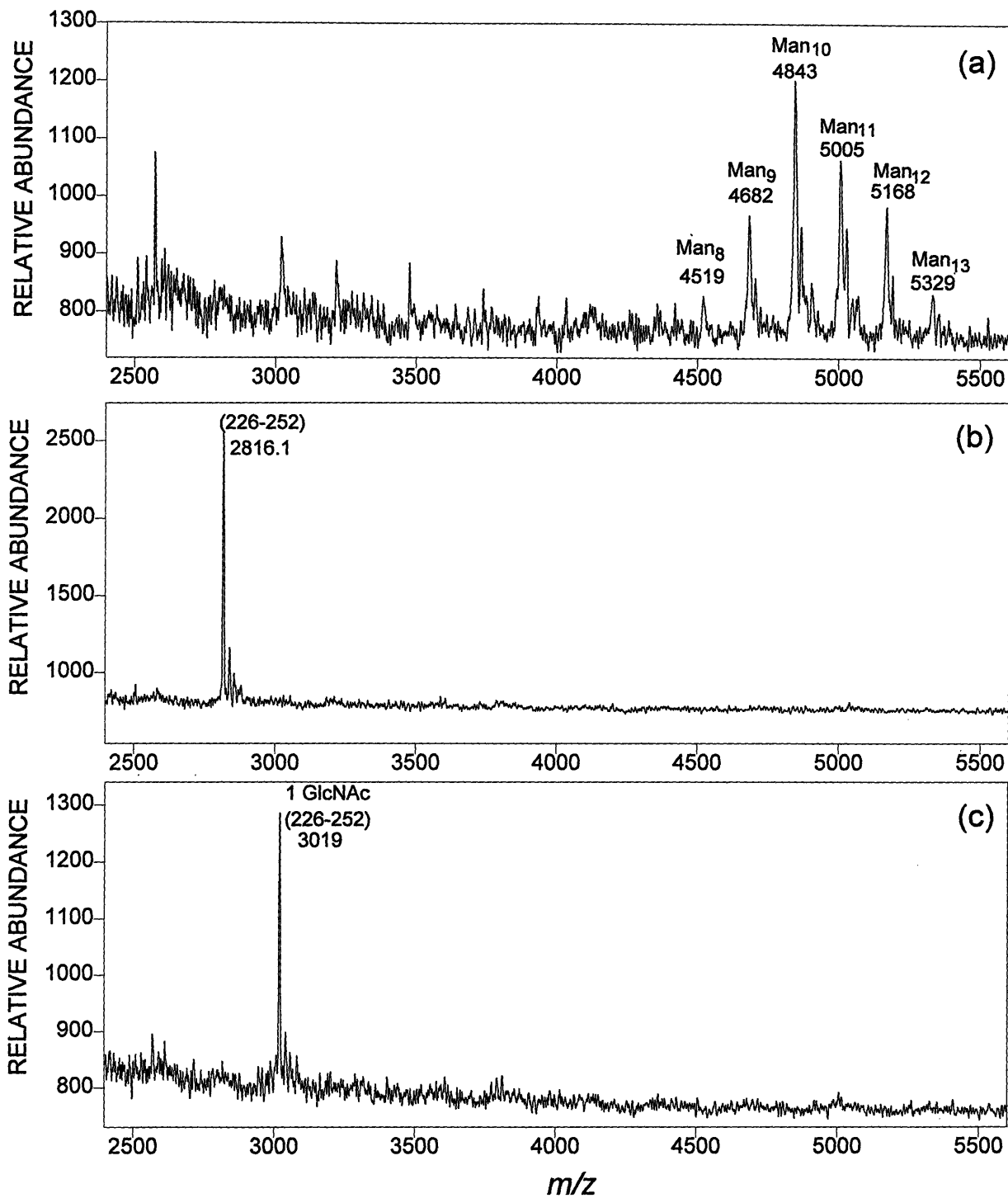


Figure II.38. The MALDI spectra of fraction G11 of the Glu-C subdigest of fraction L16 (from the Lys-C digest) before (a) and after deglycosylated with (b) PNGase F and (c) Endo H.



and Endo H deglycosylation experiments (Figures II.38b and II.38c). The PNGase F deglycosylated peptide  $(M+H)^+ = m/z$  2816.1 and the corresponding Endo H peptide  $(M+H)^+ = m/z$  3019.3 match peptide 226-252 (calc. 2817.5) and GlcNAc-peptide 226-252 (calc. 3020.2).

Up to this point, the individual glycopatterns for sequons 8, 9, 12, and 13 which could not be identified in the tryptic digest have been determined using Lys-C. The glycopattern (labeled "★") observed in Figure II.33a was in fact the one averaged from the individual contribution of sequons 8 and 9. Furthermore, sequons 12 and 13 must be fully glycosylated (because no corresponding carbohydrate-free peptide was detected) although a small amount may have a single GlcNAc attached.

Returning to the Lys-C digest, the MALDI spectrum of fraction L18 is shown in Figure II.39a which indicated three glycopatterns in the mass region of  $m/z$  6,000-10,000. After Endo H deglycosylation, the peaks at  $m/z$  4486.1 and 4689.1 (Figure II.39b) match peptides 226-265 (with sequons 8 and 9) with one GlcNAc (calc. 4484.8) and two GlcNAc (calc. 4688.0), while the signals at  $m/z$  7644.9 and 7848.2 correspond to peptides 77-139 (with sequons 3, 4, 5, 6) with two GlcNAc (calc. 7645.1) and three GlcNAc (calc. 7848.3). The glycopattern (labeled "•") belongs to peptide 226-265 with the oligosaccharides attaching at either sequons 8 or 9 which it has been determined and discussed in fraction L16. The other glycopattern (labeled "#") in the region of  $m/z$  8,300-9,650 corresponds to peptide 226-265 with total of  $\text{Man}_{20-28}\text{GlcNAc}_4$  (both sequons 8 and 9 were occupied). Therefore, sequons 8 and 9 found in peptide 226-265 have been mono-glycosylated at either site or bi-glycosylated at both sites. It is surprising that the peptide (226-265) with more

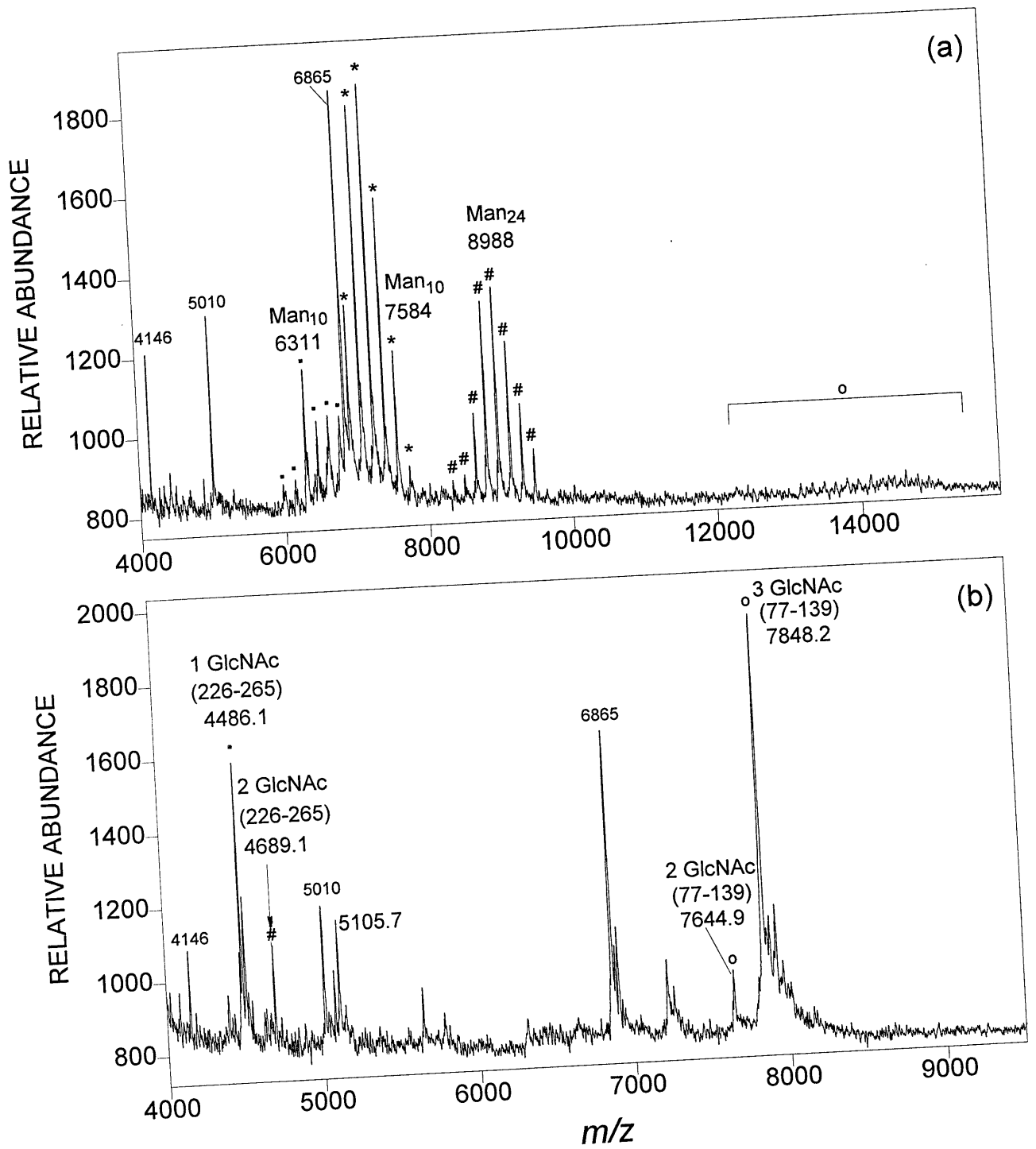


Figure II.39. The MALDI spectra of fraction L18 of the Lys-C digest before (a) and after (b) Endo H deglycosylation.

carbohydrate (bi-glycosylation) eluted in a later fraction than the same peptide with less carbohydrate (mono-glycosylation), which is contrary to the other observations. Based on the signals at  $m/z$  7644.9 (with 2 GlcNAc) and 7848.2 (with 3 GlcNAc) peptide 77-139 (which contains sequons 3, 4, 5, and 6) would be expected to result in glycopatterns at  $m/z$  11,294-14,537 (for  $\text{Man}_{20-40}\text{GlcNAc}_4$ ) and at  $m/z$  11,701-16,565 (for  $\text{Man}_{20-50}\text{GlcNAc}_6$ ). However, the signals in this region (marked "o") were too weak to be recognized. Although no Endo H peptide corresponding to the glycopattern (labeled "★") ranging from  $m/z$  6,938 to 7,745 is found, one can calculate that it is from peptide 457-506 (calc. 5559.2) with  $\text{Man}_{6-11}\text{GlcNAc}_2$  (calc. 1379.2-2189.9) at sequon 14 (Asn<sup>493</sup>-Met-Thr). One must assume that this peptide has somehow been proteolytically hydrolyzed during Endo H deglycosylation.

Figure II.40a shows the MALDI spectrum of fraction L19. The high mass region has been enlarged ( $\times 10$ ) in the Y-axis. The twin-peak glycopattern in the mass region of 6,000-8,000 belongs to glycopeptide 36-76 which was already discussed above in fraction T15 of the trypsin digest. It should be recalled that the lighter mass signal is about 23 Da heavier than the one calculated from the published sequence (Taussig and Carlson, 1983) and the difference between the pair is about 30 Da. To further study this peptide for the origin of the pair of signals, fraction L19 was treated with chymotrypsin followed by HPLC fractionation using a C18 column (Figure II.40b). The pair of signals at  $m/z$  1240 and 1270 (Figure II.41a) was found in fraction 18. Since the former is 23 Da higher than the predicted chymotryptic peptide 56-66 and the components of this pair also differed by 30 Da, the modified amino acid(s) must be in this fragment. The MALDI-PSD spectra of these two ions are shown in Figures II.41b and II.41c. The  $b_3$  to  $b_{10}$  ions of peptide  $(M+H)^+ = m/z$  1269.7

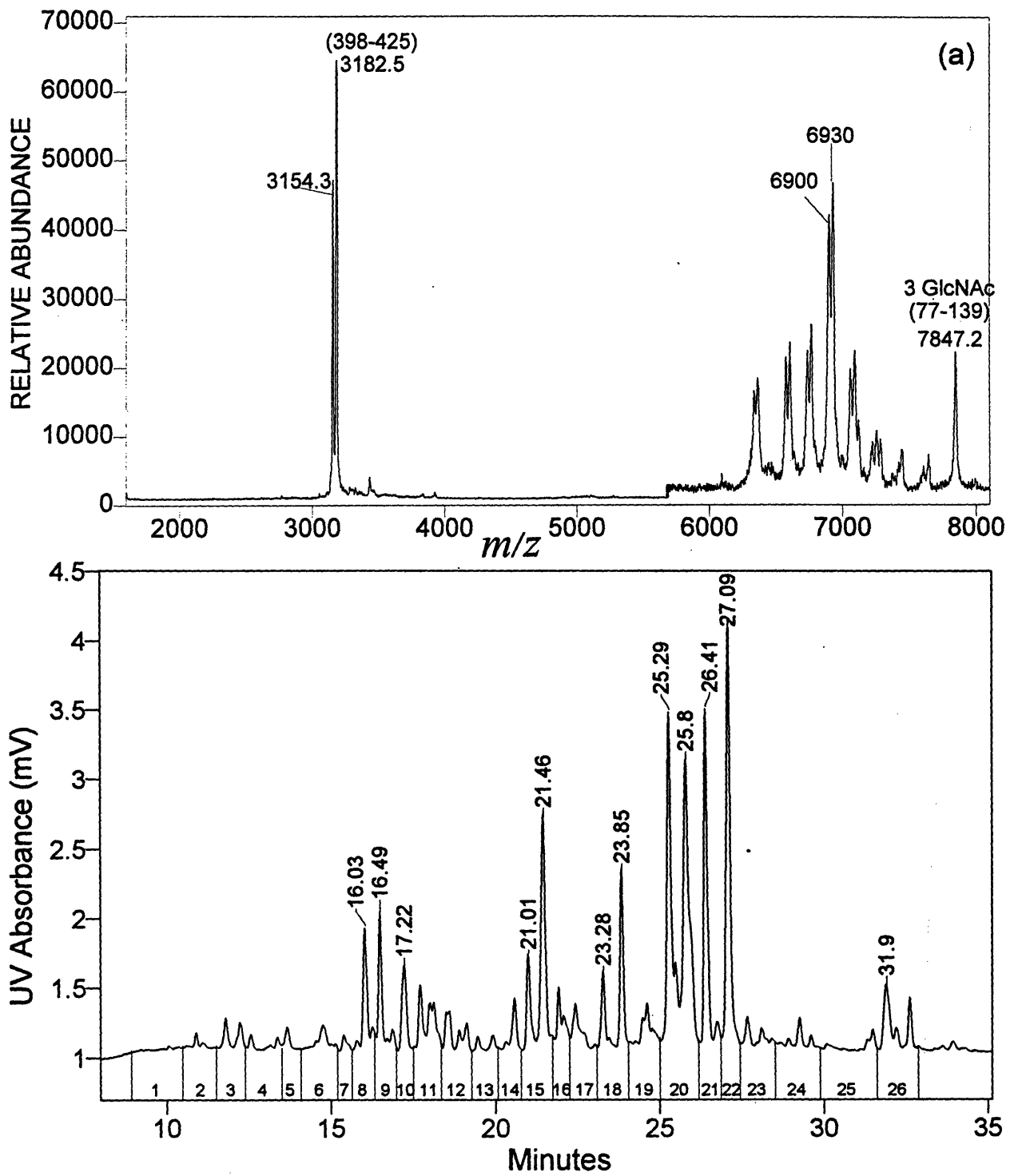


Figure II.40. (a) The MALDI spectra of fraction L19 of the Lys-C digest. (b) HPLC trace of the chymotryptic subdigest of fraction L19 (from the Lys-C digest). The fractions collected are marked.

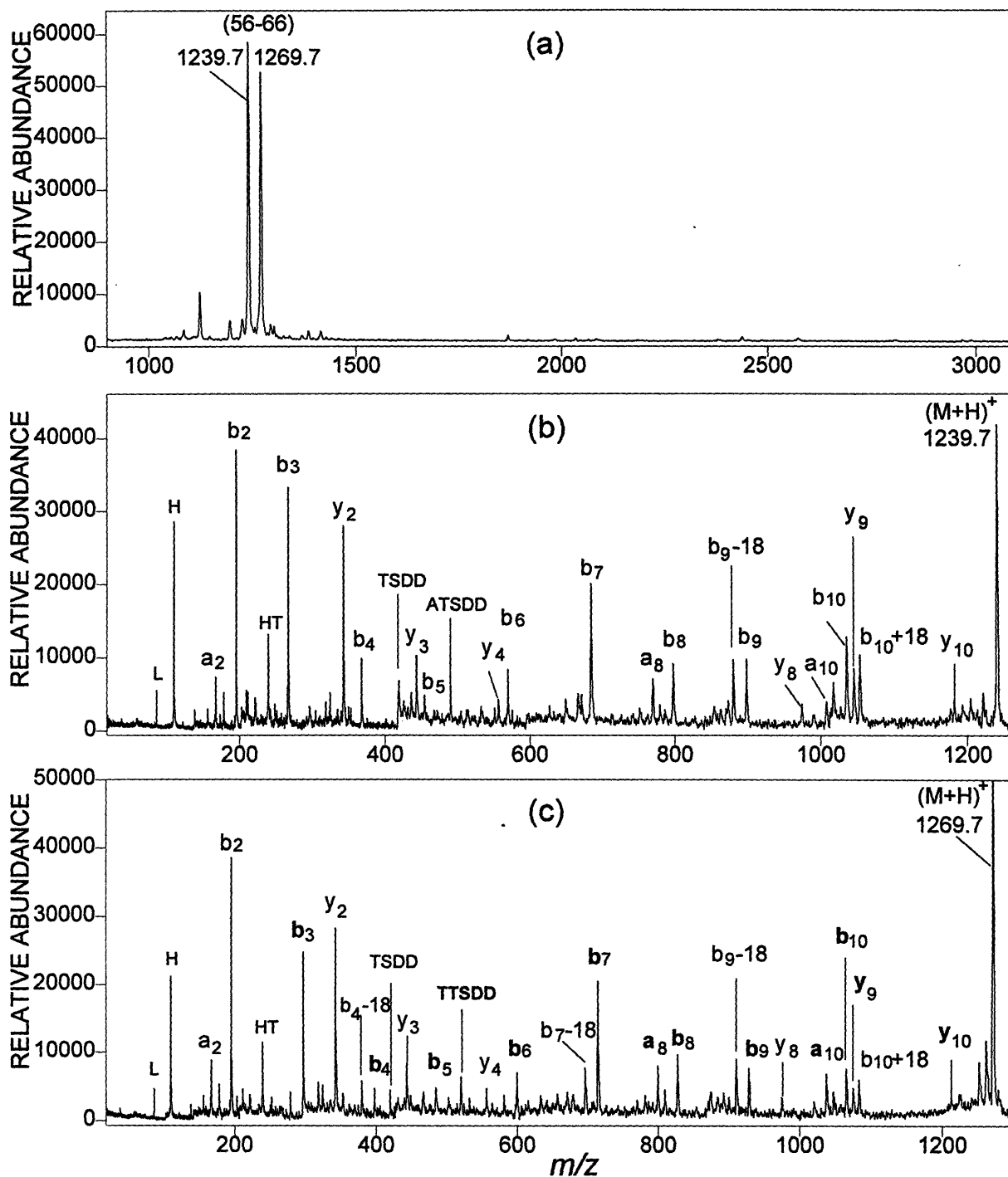


Figure II.41. (a) The MALDI spectrum of fraction 18 of the chymotryptic subdigest of fraction L19 (from the Lys-C digest). (b) and (c) The MALDI-PSD spectra of peptides  $(M+H)^+ = m/z$  1239.7 and 1269.7, respectively. The sequences are identified as **GH****A****T**SDD**L****T****H****W** for the former and **GH****T****T**SDD**L****T****H****W** for the latter. The bold letters in (c) indicate that these ions were used to identify the amino acids replaced.

(Figure II.41c) are all 30 Da heavier than the  $b_3$  to  $b_{10}$  ions of peptide  $(M+H)^+ = m/z$  1239.7 (Figure II.41b) indicated that the third amino acid (Ala<sup>58</sup>) has been partially replaced by Thr. The internal ion corresponding to TTSDD observed in Figure II.41c instead of ATSDD in Figure II.41b and the  $y_9$  and  $y_{10}$  ions all indicate the same amino acid replacement.

On the other hand, the difference between  $b_9$  and  $b_{10}$  is about 137 Da indicating that the 10th amino acid in this peptide (position 65 of the protein) is His instead of the Asn previously reported based on the DNA sequence of the gene coding for invertase (Taussig and Carlson, 1983). The mass difference between His and Asn is 23 Da. Therefore, peptide  $(M+H)^+ = m/z$  1239.7 has the sequence of <sup>56</sup>GHATSDDL**TH**W<sup>66</sup> and peptide  $(M+H)^+ = m/z$  1269.7 has the sequence of <sup>56</sup>GHTSDDL**TH**W<sup>66</sup> (underlining indicates the position of the modification taking place; the modified amino acids are in bold face). Comparing these two sequences with <sup>56</sup>GHATSDDLTNW<sup>66</sup> corresponding to the DNA sequence, one must conclude that two amino acids at position 58 and 65 have been partially and fully modified, respectively. The replacement of Asn<sup>65</sup> by His explains why the lighter peptide in the pair of peaks is 23 Da heavier than the expected one (see discussion above), and it is also found that the peptide 30 Da heavier in the pair of peaks is indeed caused by the additional replacement of Ala<sup>58</sup> by Thr. The modification at position 65 must be complete, while Ala<sup>58</sup> was replaced only partially (approximately 35%) by Thr. Inspection of the published DNA sequence of the yeast *SUC2* gene for invertase shows that the codons used for Ala<sup>58</sup> and Asn<sup>65</sup> are GCT and AAT, respectively (Taussig and Carlson, 1983). Both the replacement of Ala<sup>58</sup> by Thr and Asn<sup>65</sup> by His requires only a single nucleotide change to ACT and CAT, respectively.

In Figure II.40a, there is also a pair of strong signals at  $m/z$  3154.2 and 3182.5. The second fits to peptide 398-425 (calc. 3182.5) but it was not possible to assign the signal at  $m/z$  3154.2 to any reasonable fragment of invertase. In order to confirm the assignment of the former and to identify the latter, another portion of fraction L19 was digested with Arg-C and the digest separated by HPLC (Figure II.42). Two peptides differing by 28 Da ( $m/z$  1565.7 and 1593.8) were found in fractions R6. The PSD spectra of these two ions are shown in Figures II.43a and II.43b. They exhibit a rather similar pattern, but the space between  $b_9$  and  $b_{10}$  is 99 Da in the PSD spectrum of peptide  $(M+H)^+ = m/z$  1593.8 (Figure II.43b) indicating that the 10th amino acid (position 412 in the protein) is Val, as expected from the DNA sequence. In contrast, in the PSD spectrum of peptide  $(M+H)^+ = m/z$  1565.7 (Figure II.43a) the space between  $b_9$  and  $b_{10}$  is 71 Da indicating Ala in position 412. These two PSD spectra, therefore, revealed that both peptides represent the sequence of peptide 408-421 but the amino acid (Val) in position 412 has been partially replaced by Ala. The DNA codon for Val in the yeast invertase is GTC (Taussig and Carlson, 1983). If the second nucleotide T has been partially replaced by C (GTC→GCC) the corresponding amino acid is partially expressed as Ala.

### 3. *Asp-N Digestion of native intact protein:*

In order to resolve the remaining ambiguities, intact external invertase was digested with Asp-N. The digest was fractionated by HPLC and 22 fractions were collected (see Figure II.4). It was found that fractions A4, A5, A7, A8, A9, A11, A12, A14, A15, A18, A19, and A20 contain

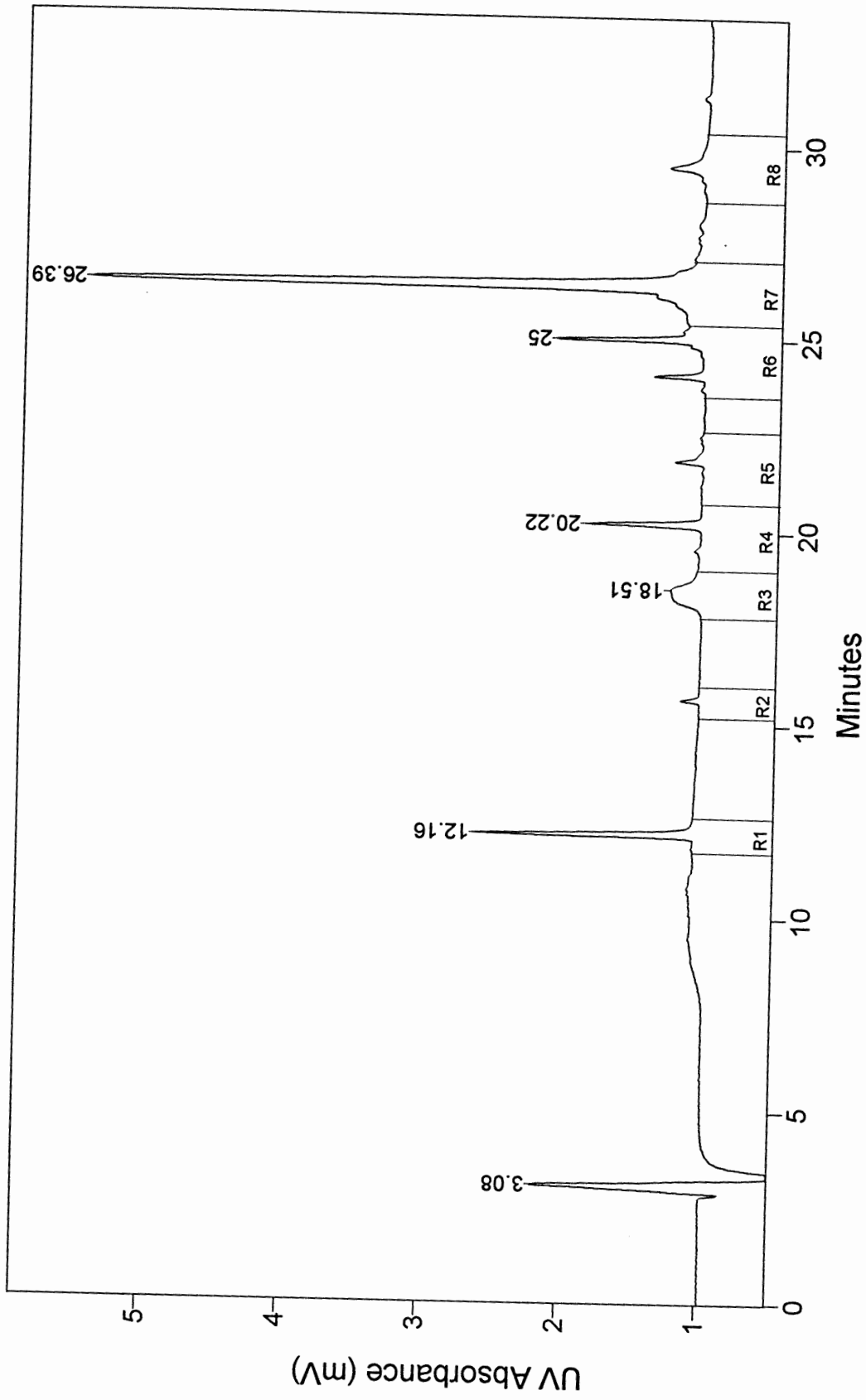


Figure II.42. HPLC fractions of the the Arg-C subdigest of fraction L19 (from the Lys-C digest). Fractions collected are marked.



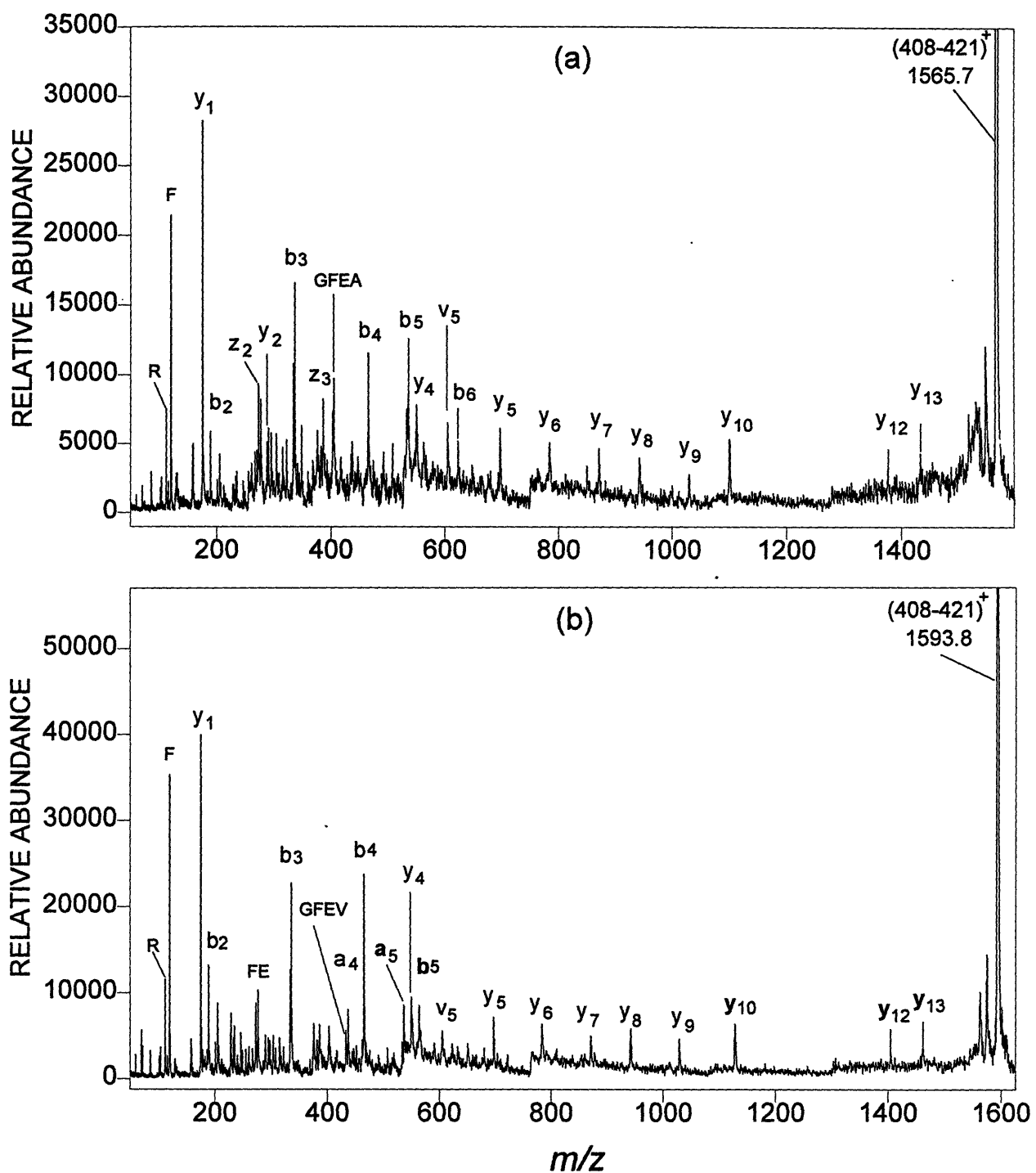


Figure II.43. The MALDI-PSD spectra of peptides (M+H)<sup>+</sup> = m/z 1565.7 (a) and 1593.8 (b). The sequences are identified as MGFE**A**SASSFFLDR for the former and MGFE**V**SASSFFLDR for the latter. The bold letters in (c) indicate that these ions were used to identify the amino acids replaced.

glycopeptide(s) since they presented at least one glycopattern in their MALDI spectra. However, many of the glycopeptides produced were not due to specific cleavage at the N-terminal side of Asp. In fact, the PSD spectra of some of the more abundant small peptides revealed that they were produced by cleaving at the N-terminal side of Glu or C-terminal side of Arg/Lys. Overall, a large fraction (approximate 40%) of the peptides/glycopeptides produced by Asp-N digestion were not the predicted Asp-N specific peptides. To simplify the data presentation, only the results of those fractions that contain only the Asp-N specific glycopeptides will be discussed below.

Figure II.44a and II.44b show the MALDI spectra of fraction A4 before and after PNGase F deglycosylation. The peak at  $m/z$  818.8 in Figure II.44b matches peptide 255-261 (calc. 818.9, containing sequon 9). The  $(M+H)^+$  ions of the eight peaks (labeled "★" in Figure II.44a) correspond to this peptide with  $\text{Man}_{4-11}\text{GlcNAc}_2$  indicating that sequon 9 has only a "narrow range glycosylation". Because sequon 9 was reported (Ziegler et al., 1988) to contain a long chain carbohydrate (more than 50 mannoses), the PSD spectrum of the PNGase F deglycosylated peptide  $(M+H)^+ = m/z$  818.8 was acquired to confirm the sequence, DDQSRVV (Figure II.45). The identical glycopattern of sequon 9 had been found earlier in the Lys-C digest (see Figure II.35a, labeled "•"). Since a signal at  $m/z$  817.8 corresponding to the sugar-free peptide 255-261 was also found in fraction A5, the 1:1 mixture of A4 and A5 was deglycosylated with Endo H. However, the corresponding Endo H peptide was not found as expected due to the low occupancy. It should be noted that the signal of the sugar-free peptide was too strong to find a reference peak in the spectrum of the mixture (selecting reference peak refer to the discussion above). Therefore, in this case another method was tried to estimate the occupancy. The MALDI spectrum of the sugar-free peptide

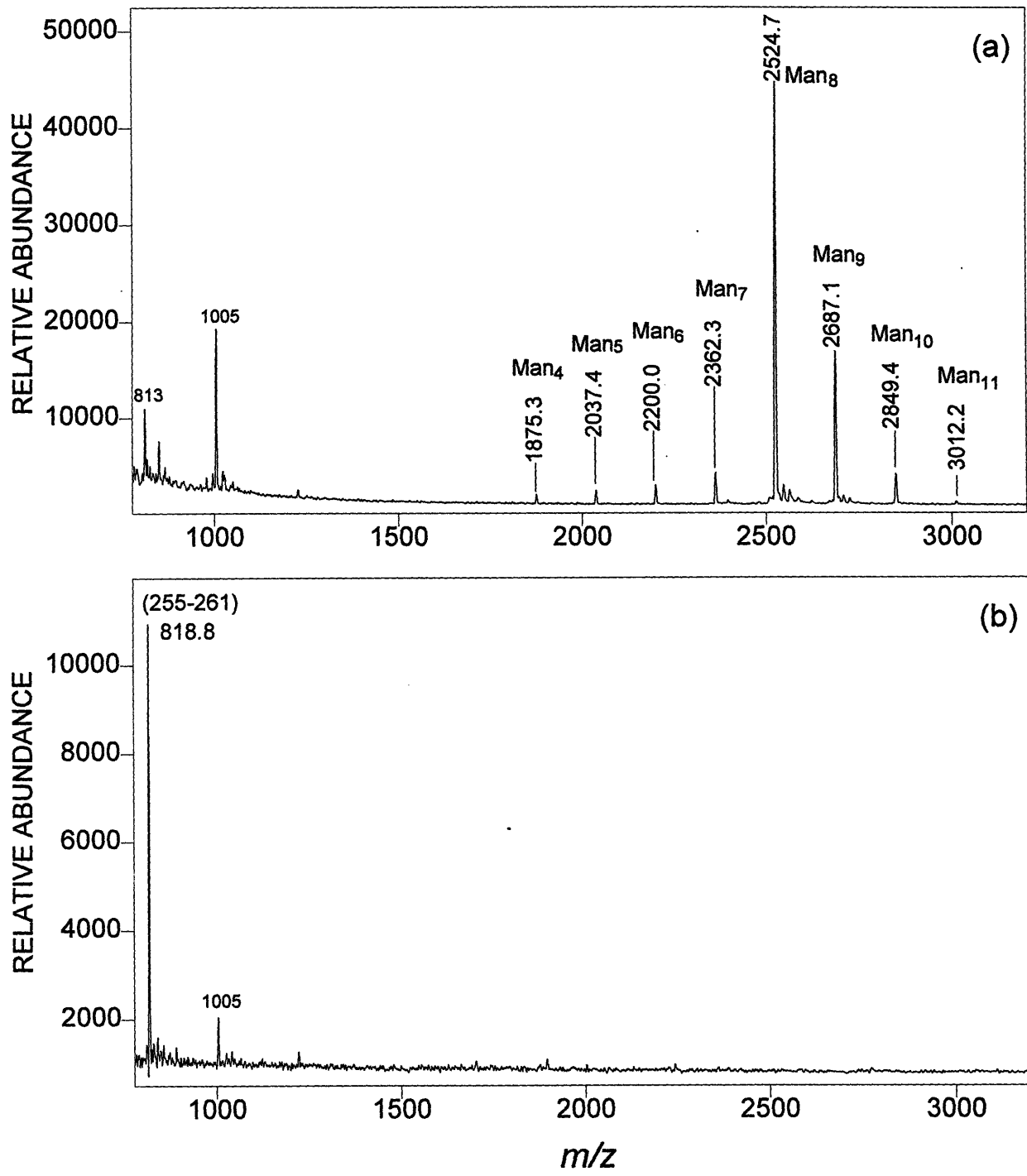


Figure II.44. The MALDI spectra of fraction A4 of the Asp-N digest before (a) and after (b) PNGase F deglycosylation.

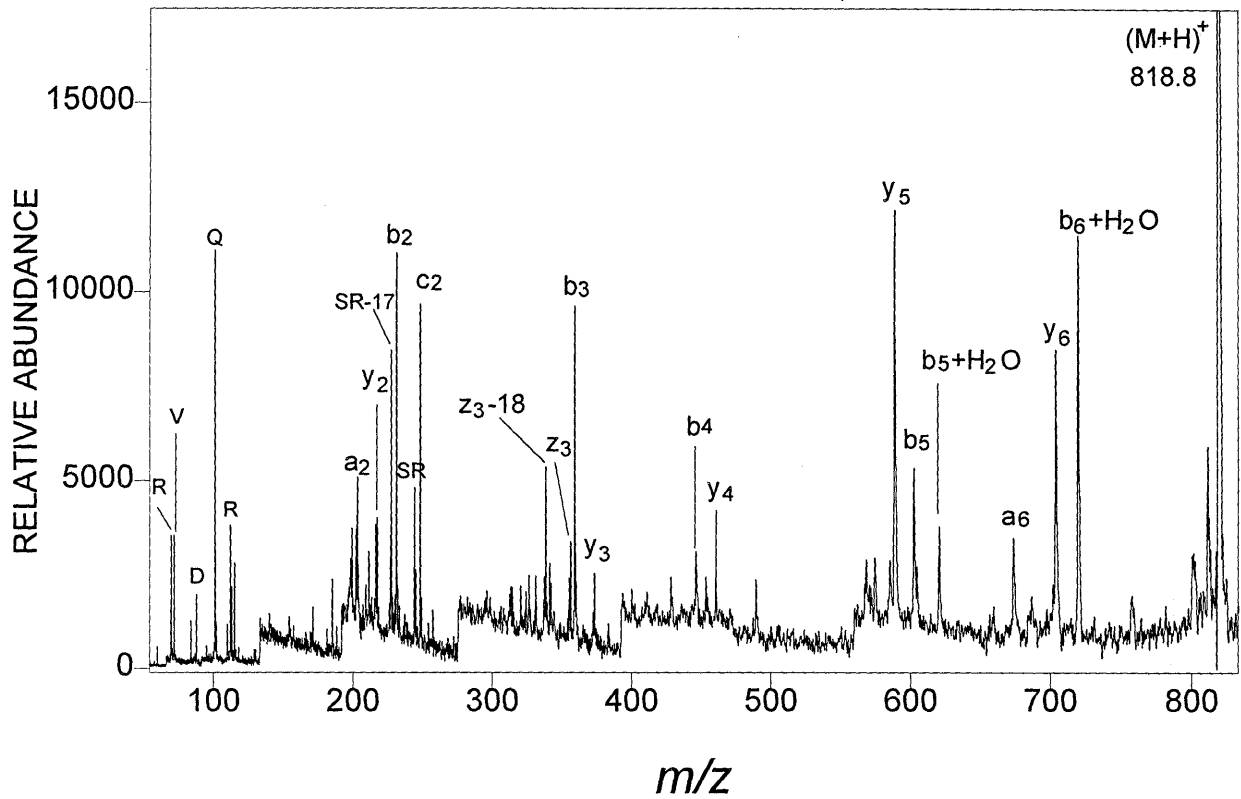


Figure II.45. MALDI-PSD spectrum of the PNGase F peptide  $(M+H)^+ = m/z$  818.8 which confirms the sequence  $^{255}\text{DDQSRV}^{261}$ .

$(M+H)^+ = m/z$  817.2 was acquired in the DE mode to provide a resolution high enough to resolve the isotopic cluster (Figure II.46a). The relative isotopic abundance of the species at  $m/z$  818.2 was measured to be 45% (averaged 10 spectra) of the  $C^{12}$ -species (theoretical  $C^{13}$ ,  $H^2$ ,  $N^{15}$  and  $O^{17}$  abundance for peptide DNQSRVV is 41%). After the mixture of A4 and A5 was digested with PNGase F, the relative abundance of the peak at  $m/z$  818.2 increased to 55% (Figure II.46b, averaged 10 spectra) due to the additional contribution from the  $C^{12}$ -species of the PNGase F deglycosylated peptide at  $m/z$  818.2 (one dalton heavier than the sugar-free peptide), indicating 10% glycosylation.

The MALDI spectrum of fraction A11 exhibits a "wide range glycosylation" with 64 peaks (Figure II.47a). The PNGase F deglycosylated peptide  $(M+H)^+ = m/z$  2610.2 (Figure II.47b) indicates this glycopattern corresponds to the sequon 7 containing peptide 129-151 (calc. 2609.9) with  $Man_{5-68}GlcNAc_2$ . This glycopattern differs from that observed in Figure II.5b (the tryptic digest) and Figure II.26a (the Lys-C digest) due to the different fraction cut off (in both tryptic and Lys-C digest, the glycopeptides with long and short oligosaccharide chains were collected in two or more fractions); thus, it should correspond to the complete profile. In addition, the signal at  $m/z$  2812.0 in Figure II.46a matches peptide 129-151 with one GlcNAc (calc. 2812.0) which indicates some of the oligosaccharide chains have been trimmed by Endo H (discussed earlier). The peak at  $m/z$  2263.6 that appeared after PNGase F deglycosylation corresponds to peptide 68-89 (with sequon 3 at Asn<sup>78</sup>) indicating that this glycopeptide also eluted in fraction A11. In abundances of the corresponding glycopeptides are too low to be observed in the spectrum show in Figure II.47a.

The MALDI spectrum of fraction A12 of the Asp-N digest is shown in Figure II.48a. The

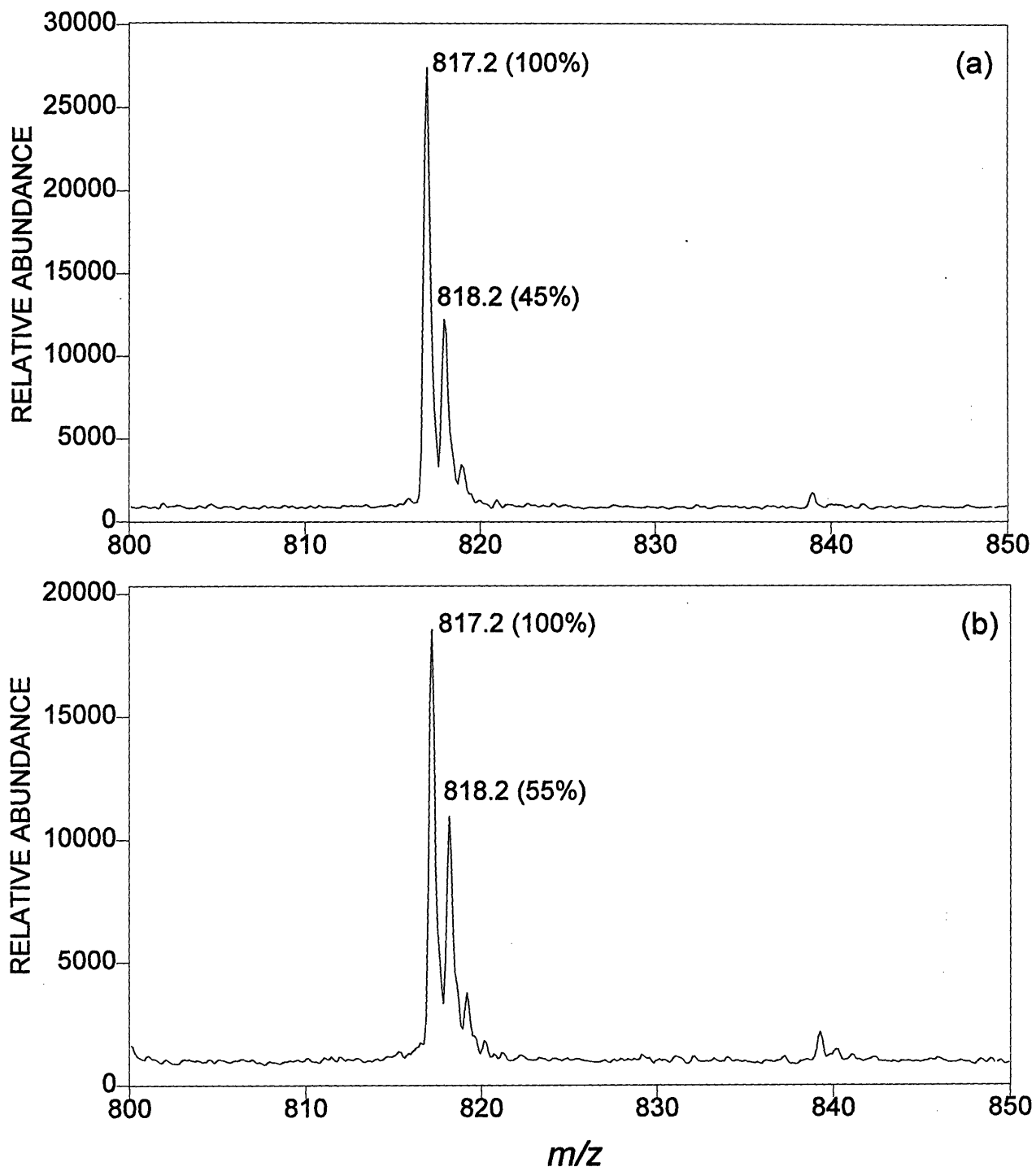


Figure II.46. MALDI-DE spectra of (a) fraction A5 and (b) the 1:1 mixture of A4 and A5 after PNGase F deglycosylation.

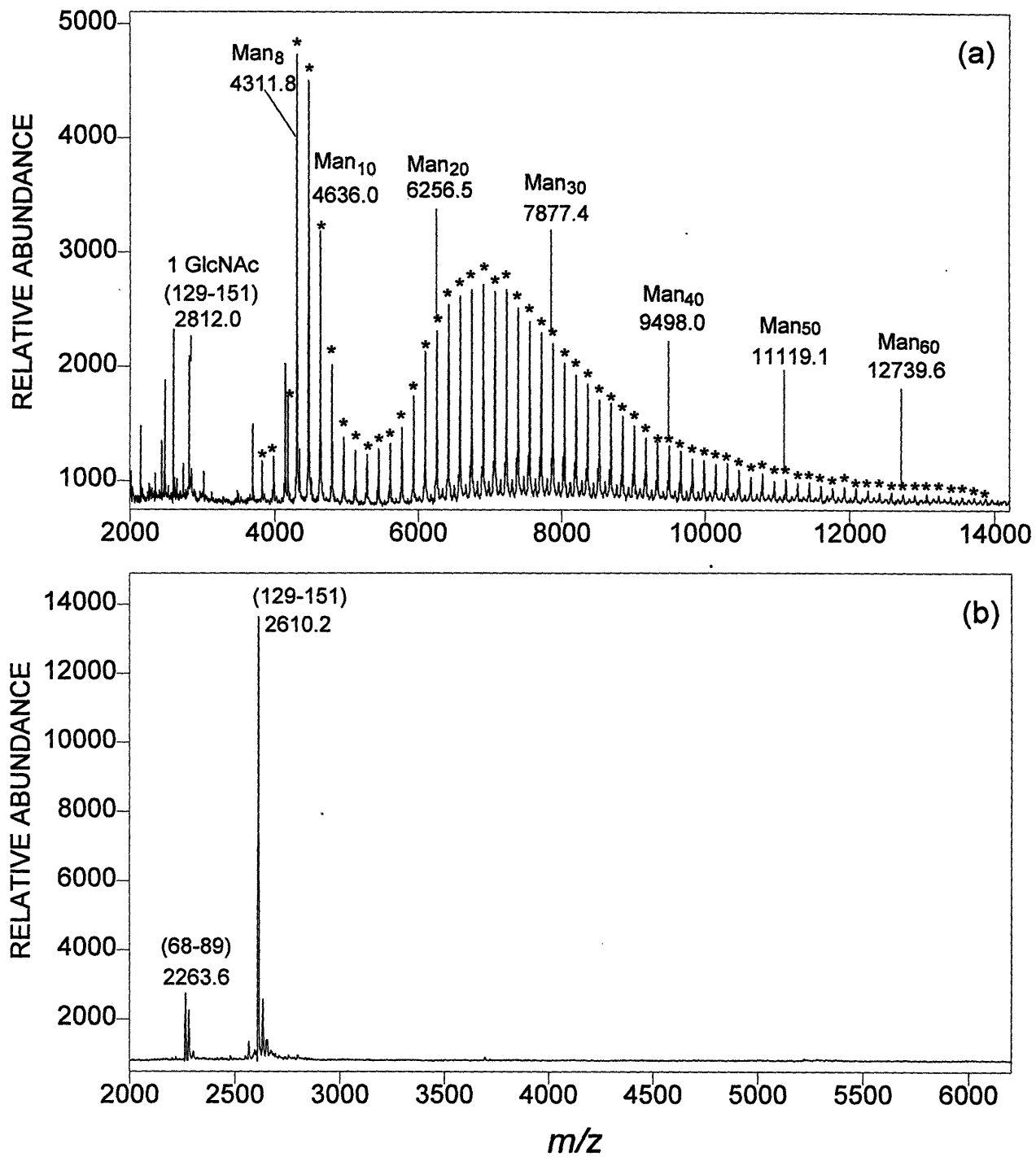


Figure II.47. The MALDI spectra of fraction A11 of the Asp-N digest before (a) and after (b) PNGase F deglycosylation.

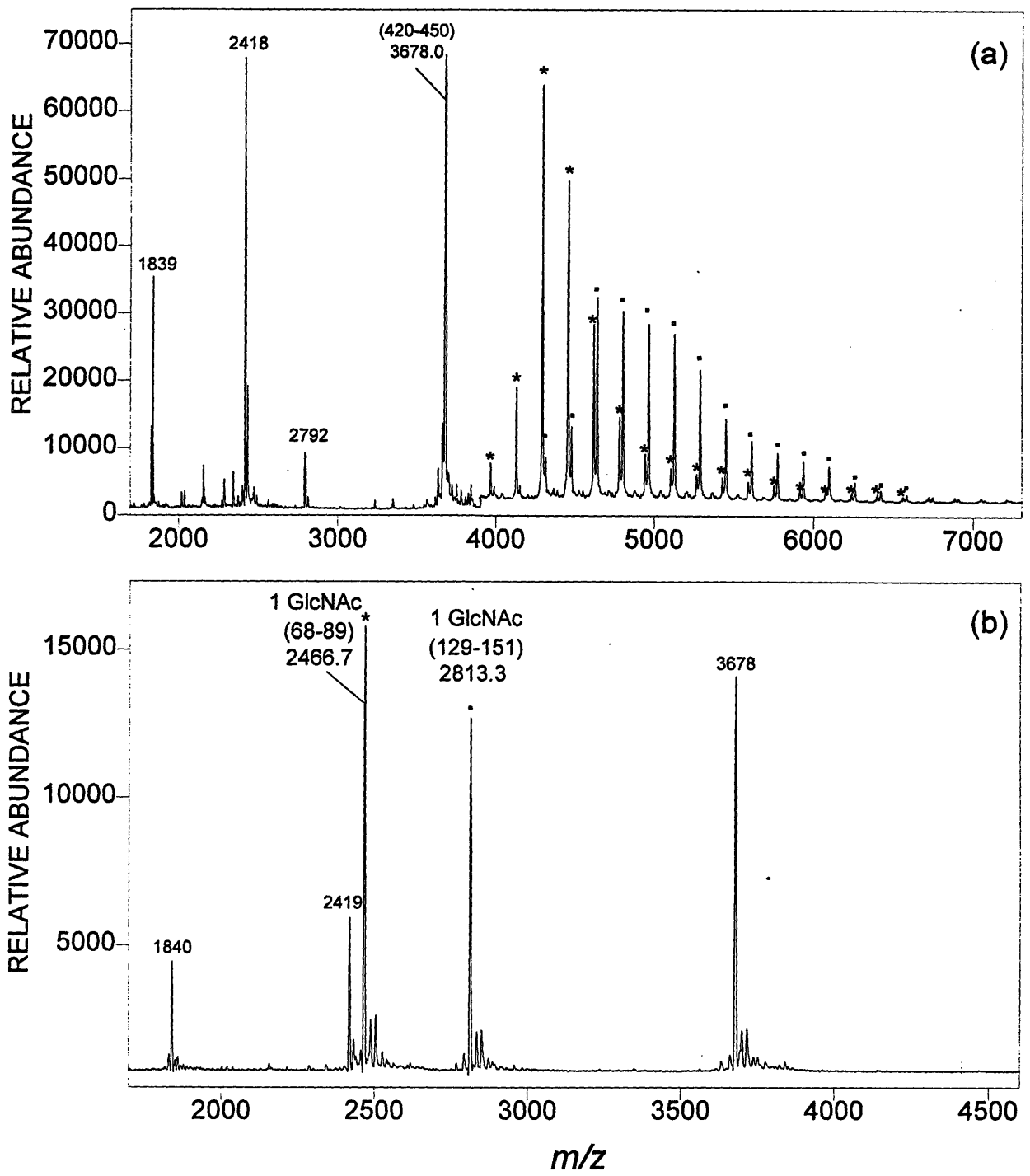


Figure II.48. The MALDI spectra of fraction A12 of the Asp-N digest before (a) and after (b) Endo H deglycosylation.



peak at  $m/z$  3678.0 matches the Asp-N peptide 420-450 (calc. 3677.1). The peaks at  $m/z$  1839.4, 2418.2, and 2792 do not match any reasonable Asp-N fragment of invertase and must, therefore, be produced by non-specific cleavage. The region of  $m/z$  3,800-7,300 has been expanded in the Y-axis to show clearly the glycopatterns (labeled "★" and "•"). After Endo H deglycosylation the glycopattern (labeled "•") collapsed to a peak at  $m/z$  2813.3 which corresponds to the Asp-N peptide 129-151 with one GlcNAc (calc. 2812.0). The glycopattern labeled "•" in Figure II.48a indicates that this peptide is glycosylated with  $\text{Man}_{6-24}\text{GlcNAc}_2$ . The same peptide with the longer glycan chains,  $\text{Man}_{5-68}\text{GlcNAc}_2$ , has been found earlier in fraction A11 (see Figure II.47a). Another Endo H deglycosylated peptide  $(M+H)^+ = m/z$  2466.7 matches the Asp-N peptide 68-89 (calc. 2464.8) containing sequon 3 (Asn<sup>78</sup>-Asp-Ser), the glycopattern of which (labeled "★" in Figure II.48a) is  $\text{Man}_{7-26}\text{GlcNAc}_2$ . This is in agreement with the results of the trypsin digestion experiment. It should be noted that Asp-N could not cleave the peptide bond between Asn<sup>78</sup> and Asp<sup>79</sup> when the former is glycosylated, but does cleave when Asn<sup>78</sup> is free.

The MALDI spectrum of fraction A18 is shown in Figure II.49a, which exhibits the 30 Da twin-peak glycopattern (discussed above with the trypsin digest) and another weaker glycopattern in between (labeled "•"). Removing the oligosaccharides with Endo H the MALDI spectrum (Figure II.49b) exhibits a pair of signals at  $m/z$  3558.1/3588.1, where the former matches peptide 33-60 (calc. 3556.9) and the latter corresponds to peptide 33-60 with the Ala<sup>58</sup> replaced by Thr (see results and discussion of the Lys-C digest). Both contain sequon 2 at Asn<sup>45</sup>. The twin-peak glycopattern revealed peptide 33-60 bearing  $\text{Man}_{5-12}\text{GlcNAc}_2$  at sequon 2. It should be pointed out again that the predicted cleavage site at the N-terminal side of Asp<sup>46</sup> was not cleaved when Asn<sup>45</sup> is

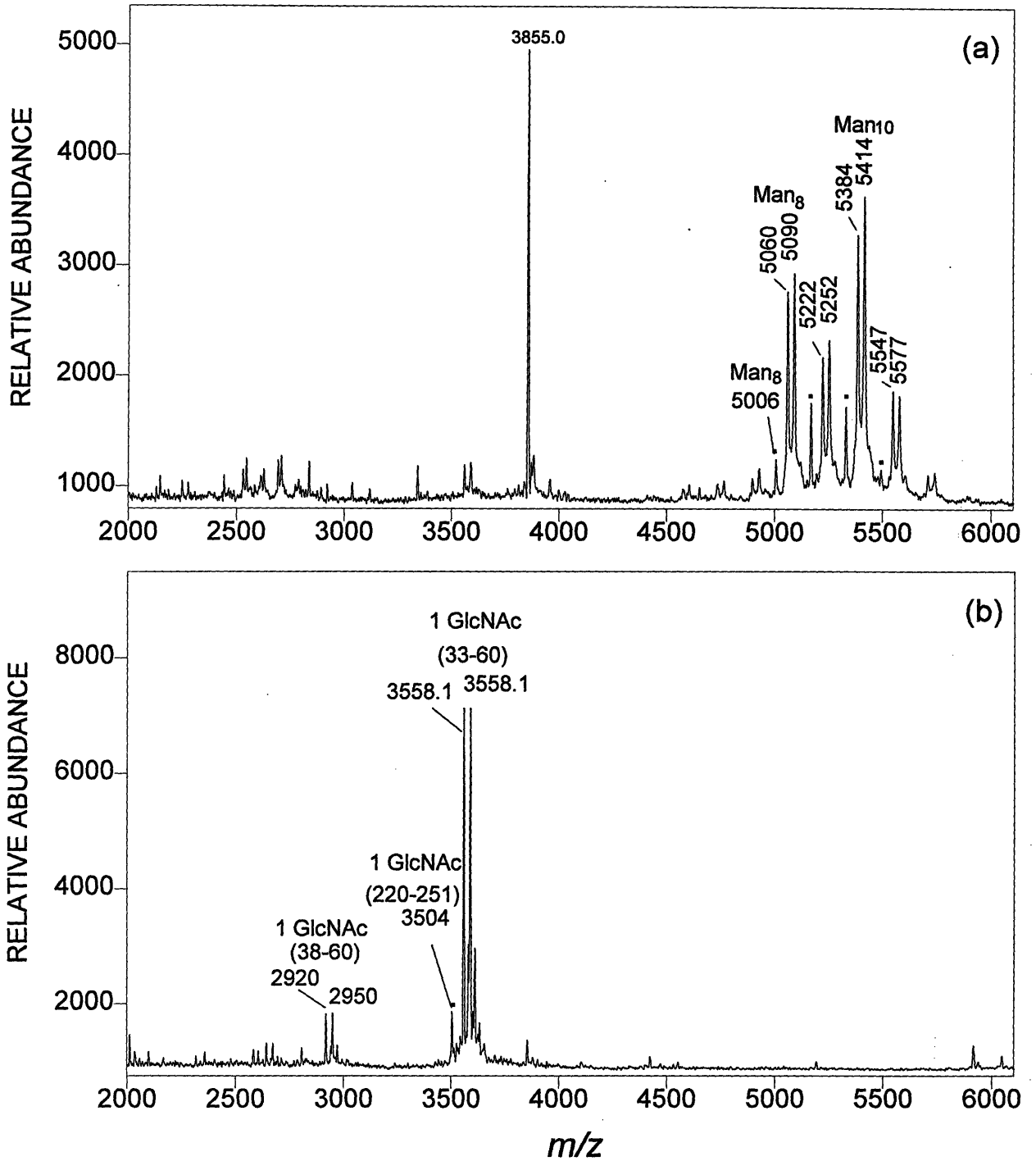


Figure II.49. The MALDI spectra of fraction A18 of the Asp-N digest before (a) and after (b) Endo H deglycosylation.

glycosylated. If it is not, it can be completely cleaved and the corresponding peptides 33-45 and 46-60 were detected in fractions A15 and A17. There is another pair of signals at  $m/z$  2920/2950 which is also spaced 30 Da apart. This turned out to be due to the proteolysis product (peptide 38-60, calc. 2919.2). This fragment was formed by cleaving the peptide bond between His<sup>37</sup> and Leu<sup>38</sup> from peptide 33-60 during Endo H deglycosylation. The other Endo H peptide found at  $m/z$  3504.3 corresponds to peptide 220-251 which contains sequon 8. This peptide was formed by the cleavage at the carboxyl side of Lys<sup>219</sup> (a tryptic site) and the N-terminal side of Glu<sup>252</sup> (a minor specificity of Asp-N). Its corresponding glycopattern (labeled "•") is observed in Figure II.49a and matches peptide 220-251 with Man<sub>8-11</sub>GlcNAc<sub>2</sub>.

### **C. Analysis of an Endo H contaminated sample:**

All structural studies reported in section A and B were carried out on a purified batch of external invertase. Initially, an earlier batch of external invertase was received and stored dry at below -20 °C. Using the same strategy described above, the oligosaccharide ranges observed at sequons 1, 2, 3, 7, 8, 9, 10, 11, 13, and 14 were similar to the results discussed above, except that the corresponding peptides were also detected containing much higher degree of only one GlcNAc. Sequons 4, 6, and 12 had only GlcNAc attached. The results suggested that all the glycosylation sites had been fully or partially deglycosylated by Endo H. If the sample was the native intact

protein (as we originally thought), there must have been Endo H activity in the cells sufficient to trim the oligosaccharides before or after the oligosaccharide elongation.

While the average molecular weight was first measured as 75 kDa, reanalysis of the same sample three years later indicated an average mass 68 kDa along with a different peak shape (Figure II.50a and II.50b). This transformation provided the initial indication that Endo H activity was present in the sample, but we did not know at first why and how the sample carried Endo H. One possibility was that Endo H was co-purified with the invertase. If there was an Endo H inhibitor activity in the cell and it did not become co-purified, that might explain the fact that the GlcNAc-peptide was never detected in the fresh purified sample. Since this would have represented a new type of activity, we then spent some time to confirm the observation until we were informed that this sample had been accidentally contaminated with Endo H before we received it.

Surprisingly, although the contaminated sample was stored under dry condition in the freezer at  $-20^{\circ}\text{C}$ , it was slowly being digested by Endo H resulting in the removal of almost 10 kDa of glycan over two years. On the other hand, the molecular weight of the new sample batch of intact external invertase guaranteed to be free of Endo H contamination was determined to be 97 kDa (see Figure II.1). In contrast to the molecular weight of the first sample (MW 75 kDa, see Figure II.50a) determined a few years ago, there was a difference of about 20 kDa. Therefore, most of the carbohydrate residues in the first sample of external invertase were gradually removed by contaminating Endo H, even under the extreme conditions (dry sample at  $-20^{\circ}\text{C}$ , no buffer, no denaturing reagent).

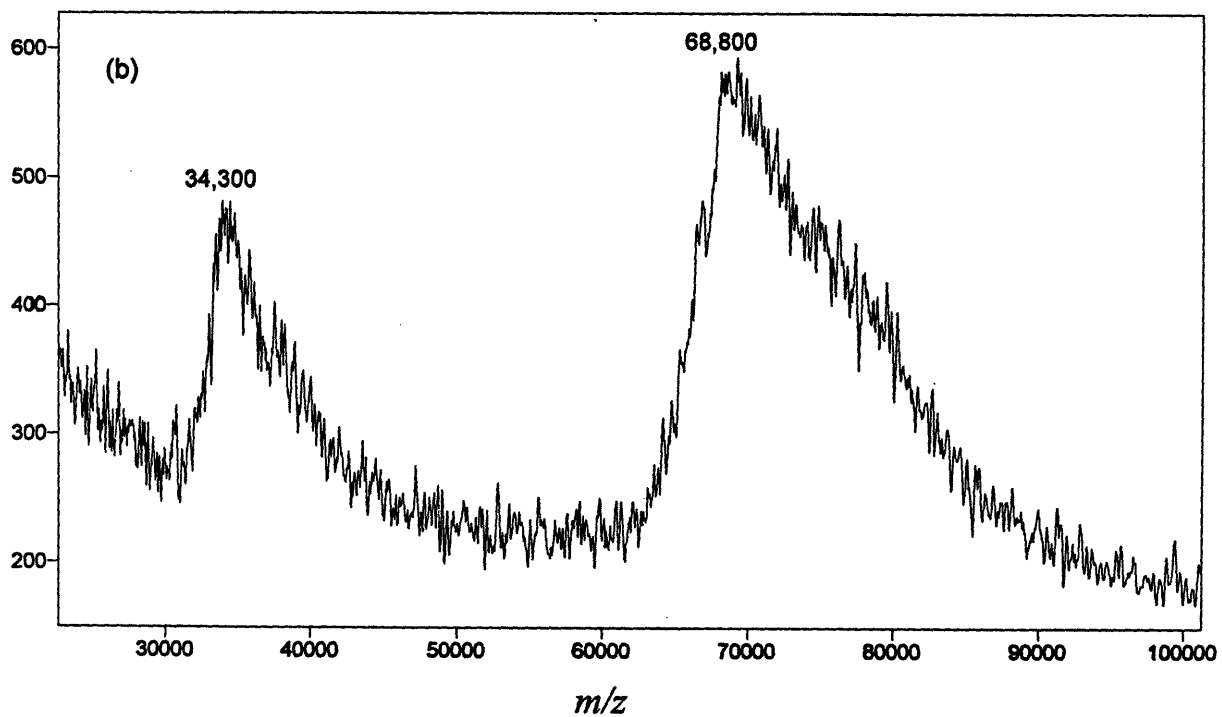
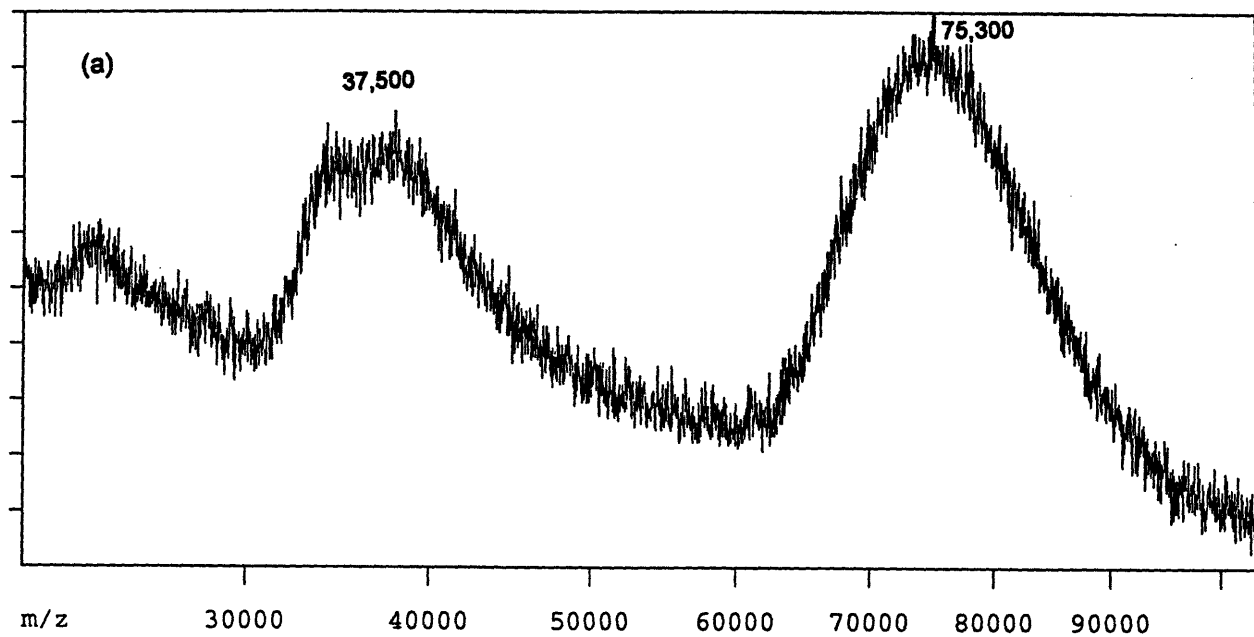


Figure II.50. (a) The MALDI spectrum of the old batch sample of external invertase. (b) The same sample detected three years later.

## D. Conclusion

To summarize the results described above, the glycopeptides and the peptides containing sequon(s) identified from the trypsin, Lys-C, and Asp-N digest are listed in Table II.1, II.2, and II.3, respectively. The extent of glycosylation determined for each sequon is listed in Table II.4. Our results confirmed the previous report that all the oligosaccharides found in yeast external invertase correspond to high mannose structures. It was found that four of the 14 sequons (1, 3, 7, and 11 at positions Asn<sup>4</sup>, Asn<sup>78</sup>, Asn<sup>146</sup> and Asn<sup>350</sup>, respectively) have a "wide range glycosylation" (containing oligosaccharides from short to long chains). The other nine sequons (2, 4, 6, 8, 9, 10, 12, 13, and 14 at positions Asn<sup>45</sup>, Asn<sup>92</sup>, Asn<sup>99</sup>, Asn<sup>165</sup>, Asn<sup>247</sup>, Asn<sup>256</sup>, Asn<sup>337</sup>, Asn<sup>379</sup> and Asn<sup>493</sup>) have a "narrow range glycosylation" (containing only short oligosaccharide chains). In agreement with earlier studies (Reddy et al., 1988) the remaining sequon 5 (Asn<sup>93</sup>) was found not to be glycosylated. The minimum and maximum masses calculated for the observed shortest and longest oligosaccharide chain of each sequon are also listed in Table II.4. The total minimum and maximum value plus the calculated average molecular weight of internal invertase provided a theoretical mass range of external invertase from 74,342 to 120,712 Da. This result generally agrees with the mass range (77 to 122 k Da) of the intact external invertase measured by MALDI-MS (Figure II.1). This confirms that the glycosylation pattern at each individual sequon found in this work corresponds well to the carbohydrate population of the native glycoprotein. The sum of the glycosylation frequency of all of the sequons appeared to be about 10.27 (Table II.4), which is a rough estimate (see the discussion in Chapter I) but still compatible with the reported value of 9-10 *N*-linked oligosaccharide chains per subunit (Trimble and Maley, 1977).

Table II.1. Observed sequons in the HPLC fractions of the tryptic digest.

Fraction	T6	T7	T8	T9	T10	T11	T15	T16	T17	T18
peptides containing sequon(s)	7 <sub>l</sub> <sup>a</sup>	7 <sub>s</sub> <sup>b</sup> 1 <sub>l+s</sub>	1 <sub>u</sub> <sup>c</sup> 8+9 3+4+5+6	8+9 <sup>d</sup> 3+4+5+6	10	10 <sub>u</sub>	2 <sup>e</sup>	2 <sub>u</sub> <sup>f</sup> , 2 <sub>u</sub> 14 12+13	14 12+13	14 12+13

Chymotryptic  
subdigest

Fraction	CT3	CT4	CT6	CT9	CT10	CT11
Sequon(s)	3 <sub>l</sub>	3 <sub>s</sub> , 6	6	4+5 <sup>g</sup>	4+5+6	4+5 <sup>h</sup>

<sup>a</sup> Subscript l represents long oligosaccharide chains.

<sup>b</sup> Subscript s represents short oligosaccharide chains.

<sup>c</sup> Subscript u represents unglycosylated sequon.

<sup>d</sup> Two or more sequons in a peptide which are fully or partially glycosylated.

<sup>e</sup> In peptides 36-76 with one and two replaced amino acid(s).

<sup>f</sup> In peptide 36-76 with one replaced amino acid.

<sup>g</sup> In peptide 84-97.

<sup>h</sup> In peptide 84-98.

Table II.2. Observed sequons in the HPLC fractions of the Lys-C digest.

Fraction	L6	L7	L8	L12	L13	L14	L16	L18	L19
peptides	7 <sub>l</sub> <sup>a</sup>	7 <sub>s</sub> <sup>b</sup>	7 <sub>u</sub> <sup>c</sup>	10+11 <sup>d</sup>	10+11 <sup>e</sup>	10+11 <sup>f</sup>	8+9 <sup>f</sup>	8+9 <sup>d</sup>	8+9 <sup>d</sup>
containing		1 <sub>l</sub>	1 <sub>l+s</sub>				12+13	14	2
sequon(s)			1 <sub>u</sub>					3+4+5+6	

Glu-C subdigest						
Fraction	G4	G5	G6	G7	G9	G11
	9	9 <sub>u</sub>	9 <sub>u</sub>	13 <sup>g</sup>	8 <sub>u</sub> <sup>h</sup>	8 <sup>i</sup>
Sequon(s)	12	12				
	13 <sup>j</sup>					

<sup>a</sup> Subscript l represents long oligosaccharide chains.

<sup>b</sup> Subscript s represents short oligosaccharide chains.

<sup>c</sup> Subscript u represents unglycosylated sequon.

<sup>d</sup> Both sequons are glycosylated.

<sup>e</sup> Sequons are fully and partially glycosylated.

<sup>f</sup> Only one of the two sequons is glycosylated.

<sup>g</sup> In peptide 372-386.

<sup>h</sup> In peptide 231-252.

<sup>i</sup> In peptide 226-252.

<sup>j</sup> In peptide 374-386.



Table II.3. Observed sequons in the HPLC fractions of the Asp-N digest.

Fraction	A4	A9	A7	A8	A9	A11	A12	A14	A15	A18	A19	A20
peptides containing sequon(s)	9	9 <sub>u</sub> <sup>a</sup>	3 <sub>u</sub> <sup>b</sup>	7 <sub>l</sub> <sup>c</sup>	7 <sub>s</sub> <sup>d</sup>	7 <sub>l+s</sub> <sup>e</sup>	3 <sup>f</sup> 7 <sub>s</sub> <sup>e</sup>	7 <sub>l</sub> <sup>g</sup>	7 <sub>s</sub> <sup>g</sup> 2 <sub>u</sub> <sup>i</sup>	2 <sup>h</sup>	7 <sub>u</sub> <sup>g</sup> 8	12+13

<sup>a</sup> Subscript u represents unglycosylated sequon.

<sup>b</sup> In peptide 68-78 with unglycosylated sequon 3 at Asn<sup>78</sup>.

<sup>c</sup> In peptide 136-151. Subscript l represents long oligosaccharide chains.

<sup>d</sup> In peptide 136-151. Subscript s represents short oligosaccharide chains.

<sup>e</sup> In the expected Asp-N peptide 129-151.

<sup>f</sup> In peptide 68-89 with carbohydrates at sequon 3 (Asn<sup>78</sup>).

<sup>g</sup> In peptide 145-182 and 144-182.

<sup>h</sup> In peptide 33-60 when sequon 2 (Asn<sup>45</sup>) is glycosylated.

<sup>i</sup> In peptide 33-45 when sequon 2 (Asn<sup>45</sup>) is unglycosylated.

Table II.4. Summary of results from MALDI MS compared to published data<sup>a</sup>

Sequon number	Observed glycan length	Calculated glycan masses Min. Max.	Estimated glycosylation freq.	Published glycan length	Published glycosylation freq.
1 Asn <sup>4</sup>	Man 4-52Gn <sub>2</sub>	1055 8838	0.98 <sup>b</sup>	L	1.0
2 Asn <sup>45</sup>	Man 5-13Gn <sub>2</sub>	1217 2514	0.79	S	0.6
3 Asn <sup>78</sup>	Man 4-36Gn <sub>2</sub>	1055 6243	0.90	S	0.8
4 Asn <sup>92</sup>	Man 4-14Gn <sub>2</sub>	1055 2676	1.00 <sup>c</sup>	S	0.9
5 Asn <sup>93</sup>	-	-	-	-	-
6 Asn <sup>99</sup>	Man 6-22Gn <sub>2</sub>	1379 3973	1.00	S	0.9
7 Asn <sup>146</sup>	Man 5-68Gn <sub>2</sub>	1217 11432	0.90	L	0.5
8 Asn <sup>247</sup>	Man 8-13Gn <sub>2</sub>	1704 2514	0.10	S	0.3
9 Asn <sup>256</sup>	Man 4-11Gn <sub>2</sub>	1055 2190	0.10	L	0.3
10 Asn <sup>337</sup>	Man 4-14Gn <sub>2</sub>	1055 2676	0.87	S	1.0
11 Asn <sup>350</sup>	Man 3-70Gn <sub>2</sub>	893 11756	0.83	L	1.0
12 Asn <sup>365</sup>	Man 7-14Gn <sub>2</sub>	1541 2676	0.90	S	1.0
13 Asn <sup>379</sup>	Man 4-12Gn <sub>2</sub>	1055 2352	0.90	S	1.0
14 Asn <sup>493</sup>	Man 6-11Gn <sub>2</sub>	1379 2190	1.00	S	0.5
Total	Man 72-353Gn <sub>26</sub>	15660 62030	10.27	4-L, 9-S <sup>d</sup>	9.8

Calc. ave. MW of int. invertase:

Calc. mass range from 74.342 Da to 120.712 Da

58.682 Da

- a See Ziegler et al., (1988)
- b The glycosylation frequency listed here represent the percentage of glycosylation at each sequon (e.g. frequency = 0.98 represents 98% occupancy at sequon 1).
- c The frequency 1.00 means the unglycosylated peptide at this particular sequon was not observed, therefore the occupancy is estimated 100%.
- d L = long chain glycan (Man<sub>>50</sub>); S = short chain glycan (Man 8-14).

Most of the predicted peptide fragments were detected in each of three parallel experiments. In general, the trypsin and Lys-C digests gave better results than Asp-N (Tables II.5, II.6, and II.7). A few predicted glycopeptides as well as unglycosylated peptides were missing from each of three enzyme digestions (Lys-C, trypsin, and Asp-N digestion). Fortunately, there was enough overlap in these experiments that the complementary data could be combined to cover 100% of the amino acid sequence, as shown in Figure II.51. For instance, the glycopeptide of sequons 4, 5, and 6 could not be found in the Lys-C digest but it was detected in the trypsin digest, whereas the glycopeptide of sequon 11 was only found in the Lys-C digest.

The proteolytic glycopeptide can not be separated completely by reversed phase HPLC, with the more heavily glycosylated glycoforms eluting first (Hemling et al., 1990). This phenomenon was certainly effects to the determination of the distribution profile of those "wide range glycosylation" peptides due to fact that glycopeptides with the longer chain glycans have higher hydrophilicity and thus shorter retention time, while the reverse is true for the shorter chain glycans. Under these circumstances, therefore, one must combine the fractions containing the same peptide but with different length of sugar chains in order to observe the complete glycopattern. A good example is shown in Figure II.12a, II.12b, and II.12c. The tryptic-chymotryptic peptide 77-83 with the oligosaccharide of  $\text{Man}_{4-36}\text{GlcNAc}_2$  was collected in two fractions, one contains peptide 77-83 with the oligosaccharide chains longer than 13 mannoses (see Figure II.12a) while the other one contains the same peptide with the oligosaccharide chains shorter than 13 mannoses (see Figure II.12b). The complete glycopattern, therefore, can only be obtained by combining equal parts of these two fractions (see Figure II.12c).

Table II.5. Peptides found in tryptic digest which do not contain a sequon.

Amino acid position	Found in fraction	Observed mass	Calculated mass
19 - 32	12	1726.5	1726.9
108-139			
189-219	16-18	7218.3	7219.9 <sup>a</sup>
155-163	9, 10	1184.0	1184.3
164-170	8	821.1	821.0
171-175	1	641.1	640.7
176-185	8	1183.4	1183.3
260-304	18	5168.4	5168.7
266-304	18	4523.0	4522.9
305-311	7	780.0	779.9
312-331	11, 12	2354.3	2353.6
313-331	12	2226.3	2225.5
387-397	16	1339.7	1339.6
398-407	9	1221.1	1221.3
408-421	13-18	1594/1566	1593.8 <sup>b</sup>
431-438	7	1041.3	1041.1
439-447	7	1065.4	1065.2
448-456	7	1119.4	1119.2
507-510	6	549.7	549.7

<sup>a</sup> The mass of the disulfide linked peptide (108-139)-SS-(189-219).

<sup>b</sup> The mass from the original sequence without modification.

Table II.6. Peptides found in Lys-C digest which do not contain a sequon.

Amino acid position	Found in fraction	Observed mass	Calculated mass
19 - 32	14	1725.9	1726.9
155-163	11	1183.3	1184.3
164-170	9	819.3	821.0
171-175	1	639.6	640.7
176-185	9	1182.0	1183.3
189-219	17	3432.9	3433.8
295-304	10	1174.5	1175.3
266-312	23	5415.2	5412.0
313-331	15	2224.6	2225.5
387-397	20	1339.4	1339.6
398-407	7	1221.0	1221.3
398-425	19,20	3181.5/3153.2	3182.5 <sup>a</sup>
431-438	7	1039.9	1041.1
431-447	11	2085.5	2087.3
448-456	7	1118.0	1119.2
507-513	6	905.8	906.1

<sup>a</sup> The mass from the original sequence without modification.

Table II.7. Peptides found in Asp-N digest which do not contain a sequon.

Amino acid position	Found in fraction	Observed mass	Calculated mass
8-22	13	1813.6	1813.1
23-29	14	865.2	864.9
28-40	19	1802.2	1802.0
46-60	17	1676/1706	1675.8 <sup>a</sup>
61-67	10	916.4	915.9
62-67	8	800.8	800.8
100-114	12	1839.4	1838.1
103-107	2	671.6	671.4
103-128	17	3151.5	3150.5
129-135	3	759.4	760.8
152-158	15	955.5	955.1
152-172	16	2543.1	2541.9
173-181	11	1118.1	1118.2
182-195	14	1625.6	1624.8
247-254	9	923.5	923.0
266-276	17	1383.6	1383.5
391-400	18	1208.6	1208.4
391-407	19	2111.7	2111.4
401-410	15	1257.2	1257.4
401-419	16	2227.5	2225.5
408-419	19	1322.3	1322.5
420-438	10	2302.4	2300.6
420-450	12	3678.0	3677.1
451-461	17	1334.8	1334.6
499-513	10	1915.6	1915.2
505-513	7	1150.1	1149.3

<sup>a</sup> The mass from the original sequence without modification.

Figure II.51 (two pages). The amino acid sequence of invertase based on its DNA sequence (Taussig and Carlson, 1983) and the MS data (Reddy et al., 1988). Peptides determined by MALDI MS are underlined; line 1 indicates the tryptic peptides, line 2 indicates the Lys-C peptides, and line 3 indicates the Asp-N peptides. A bold letter indicates that the amino acid has been fully (*italics*) or partially replaced by another amino acid which is placed on top of that letter. Potential glycosylation sites containing the sequence Asn-Xxx-Ser/Thr (Xxx can be any amino acid but Pro) are overlined and numbered.



<sup>1</sup>  
S M T N E T S D R P L V H F T P N K G W M N D P N G L W Y D 30  
1 —————  
2 —————  
3 —————

<sup>2</sup>  
E K D A K W H L Y F Q Y N P N D T V W G T P L F W G H <sup>T</sup> A T S 60  
1 —————  
2 —————  
3 —————

<sup>H</sup> <sup>3</sup>  
D D L T N W E D Q P I A I A P K R N D S G A F S G S M V V D 90  
1 —————  
2 —————  
3 —————

<sup>4</sup> <sup>5</sup> <sup>6</sup>  
Y N N T S G F F N D T I D P R Q R C V A I W T Y N T P E S E 120  
1 —————  
2 —————  
3 —————

<sup>7</sup>  
E Q Y I S Y S L D G G Y T F T E Y Q K N P V L A A N S T Q F 150  
1 —————  
2 —————  
3 —————

R D P K V F W Y E P S Q K W I M T A A K S Q D Y K I E I Y S 180  
1 —————  
2 —————  
3 —————

S D D L K S W K L E S A F A N E G F L G Y Q Y E C P G L I E 210  
1 —————  
2 —————  
3 —————

V P T E Q D P S K S Y W V M F I S I N P G A P A G G S F N Q 240  
1 —————  
2 —————  
3 —————

<sup>8</sup> <sup>9</sup>  
Y F V G S F N G T H F E A F D N Q S R V V D F G K D Y Y A L 270  
1 —————  
2 —————  
3 —————

Q T F F N T D P T Y G S A L G I A W A S N W E Y S A F V P T 300  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

N P W R S S M S L V R K F S L N T E Y Q A N P E T E L I N L 330  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

K A E P L I N <sup>10</sup> I S N A G P W S R F A T <sup>11</sup> N T T L T K A N S Y N 360  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

V D L S <sup>12</sup> N S T G T L E F E L V Y A V <sup>13</sup> N T T Q T I S K S V F P 390  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

D L S L W F K G L E D P E E Y L R M G F E <sup>A</sup> V S A S S F F L D 420  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

R G N S K V K F V K E N P Y F T N R M S V N N Q P F K S E N 450  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

D L S Y Y K V Y G L L D Q N I L E L Y F N D G D V V S T N T 480  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

Y F M T T G N A L G S V <sup>14</sup> N M T T G V D N L F Y I D K F Q V R 510  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

E V K 513  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_

During this research it was found that when an enzyme cleavage site is adjacent to a glycosylated sequon, this cleavage site will be cleaved very slowly or not at all. For instance, Asp-N did not cleave the bond between Asn<sup>45</sup> and Asp<sup>46</sup> when Asn<sup>45</sup> was glycosylated but it was completely cleaved when the sequon was not occupied (see discussed above). Similarly, trypsin could not cleave the site between Arg<sup>77</sup>-Asn<sup>78</sup> because Asn<sup>78</sup> was glycosylated. This experimental evidence is consistent a known carbohydrate function: they contribute an important role in protecting a protein from protease digestion. However, it should be noted that chymotrypsin did partially cleave (about 30%) peptide Arg<sup>77</sup>-Arg<sup>105</sup> at the C-terminal side of Phe<sup>98</sup> even though Asn<sup>99</sup> has been glycosylated. When this cleavage occurs a glycosylated N-terminus is produced and, it was found that in that case the oligosaccharide attached to the terminal Asn is very difficult to be removed with PNGase F but is easily cleaved off with Endo H.

DHB has been reported to be the best matrix for glycopeptides for MALDI (Stahl et al., 1991). However, we compared DHB and  $\alpha$ -CHCA and found that they produce very similar spectra. In fact,  $\alpha$ -CHCA forms a much more homogeneous surface during crystallization on the laser target, so it is easier to obtain a satisfactory spectrum.

As described above, this study produced a much more detailed picture of the glycosylation of external invertase than previous studies. First of all, sequon 9 was found to be glycosylated with 4-11 mannose oligosaccharides which is definitely not a "long chain" glycosylation as reported by Ziegler et al., 1988. On the other hand, sequon 3 carries sugar chains of GlcNAc<sub>2</sub>Man<sub>4-36</sub> that is more likely to be a "wide range glycosylation" rather than a "short chain". Secondly, even though

sequons 1, 7, and 11 were observed to be composed with a few mannose to over 50 mannose rather than the over 50 mannose long chains (Ziegler et al, 1988). In fact, the major components of these "wide range glycosylation" are below 30 mannoses (see Figure II.6b, II.46a, and II.32b). We therefore prefer to use the term "wide range glycosylation" instead of the previous term "long chain" glycosylation (Ziegler et al, 1988) for this type of glycosylation.

The work reported in this thesis demonstrates that MALDI MS is an excellent method to determine clearly the distribution of each single component in the wide range type glycosylation. For this type of glycoprotein (like invertase), other mass spectrometric techniques are not capable of measuring the complete glycosylation distribution. For example, FAB (or LSIMS) does not ionize highly glycosylated peptides well and such complex mixtures give ESI spectra that are impossible to interpret easily and correctly. Third, it was discovered that the shortest chain oligosaccharide in many sequons had as few as 4 or 5 mannose residues, even though the signals for these short chains are weak. As early as 1982, it was stated "*That no oligosaccharides smaller than  $Man_8GlcNAc$  were seen during either short or long term labeling implies that the metabolic turnover of  $Man_8GlcNAc$  involves elongation rather than removal of additional mannose*" (Byrd et al, 1982), and this has become the accepted description of the synthesis process for yeast glycoproteins (Kukuruzinska et al, 1987; Ziegler and Trimble 1991). It has also been concluded that unlike mammalian glycoproteins, yeast glycoproteins have only two distinct sizes of high mannose *N*-linked oligosaccharides: a  $Man_{8-15}$  class and a  $Man_{>50}$  class (Byrd et al, 1982). Based on our results, at least for the yeast invertase, this is not true. This inconsistency may be caused by the detection techniques used to produce their results compared to ours, because of the limited sensitivity of gel

filtration and NMR which did not allow the detection of relatively low abundance of oligosaccharide components with less than eight mannoses. Thus, our results suggest the following hypothesis for yeast glycoprotein biosynthesis based on the well known glycoprotein biosynthesis process (see discussion in Chapter I): After the precursor  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$  is assembled and transferred to nascent protein, it is first trimmed to an intermediate,  $\text{Man}_8\text{GlcNAc}_2$ . This intermediate then faces the fate of further trimming (to the sizes of  $\text{Man}_{4-7}$  or even  $\text{Man}_3$ ) or elongation (to any sizes of  $\text{Man}_{>9}$ ). Assuming that the cell contains both glycosidases (mannosidases) and mannosyltransferases, the elongation processes from  $\text{Man}_8\text{GlcNAc}_2$  should not be only the addition of mannose but also reflect the result of a competition between addition and trimming. Therefore, the final oligosaccharide distribution may be the equilibrium products of these two reactions. Nature may have a requirement to control this equilibrium based on certain need(s). This would explain why sequons 1, 7, and 11 of invertase were glycosylated with a few to over 50 mannose oligosaccharides instead of a class of very large glycan. To confirm this hypothesis, one possible experiment is to quench the mannosidase reaction (by adding a mannosidase inhibitor for instance) after 10 min, since this time would be sufficient to allow the formation of the essential elongation intermediate  $\text{Man}_8\text{GlcNAc}_2$  (Byrd et al, 1982), so that the rest of the process would consist of the addition of mannose only. If the hypothesis is correct, this experiment would result in a distribution of oligosaccharides that is different from the normal one.

If there is no mannosidase activity in the Golgi apparatus, the place where the oligosaccharide elongation takes place, the short oligosaccharide chains containing less than eight mannoses were produced in the endoplasmic reticulum (ER) rather than in the Golgi. Thus, another

possible hypothesis has to be proposed to explain our results where the glycoprotein synthesized in yeast cells could go through the similar pathway as they did in animal cells (see the description of glycoprotein biosynthesis in Chapter I): After the carbohydrate precursor  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  is linked to the protein, it is first trimmed completely to the core  $\text{Man}_3\text{GlcNAc}_2$  oligosaccharide and then elongation takes place in the Golgi. The final distribution of the carbohydrate residues represents an elongation starting from an intermediate  $\text{Man}_3\text{GlcNAc}$  instead of  $\text{Man}_8\text{GlcNAc}_2$ . Again, the distribution of the oligosaccharide polymers would depend upon what nature needs. If this hypothesis is true, the pathway of yeast and mammalian glycoprotein synthesis is similar. The only difference would be that more complex carbohydrate residues are utilized in animal cells while yeast produced only high mannose *N*-linked oligosaccharides.

As discussed in the section on the Lys-C digests and shown in Figure II.52, three amino acids have been found to be fully or partially substituted by other amino acids. Asparagine in position 65 is completely replaced by histidine, while alanine and valine in position 58 and 412 were found to be partially replaced by threonine (~ 35%) and alanine (~ 40%), respectively. This results suggest that there are two amino acid sequences. One with only one amino acid replaced ( $\text{N}^{65} \rightarrow \text{H}^{65}$ ) while the other has all three replaced, or another possibility is that both sequences have two replaced amino acids, ( $\text{A}^{58} \rightarrow \text{T}^{58}$ ,  $\text{N}^{65} \rightarrow \text{H}^{65}$ ) and ( $\text{N}^{65} \rightarrow \text{H}^{65}$ ,  $\text{V}^{412} \rightarrow \text{A}^{412}$ ). Since the amino acid sequence (Figure II.52) is otherwise identical with the one reported for the SUC2 gene (Taussig and Carlson, 1983), and its DNA sequence has been recently verified by Dr. F. Maley, these two invertase sequences must be derived from two different modified DNA strains which may be called gene SUC2a and SUC2b.

## E. Materials and Methods

### *Materials:*

The yeast external invertase which we studied was provided by Dr. Frank Maley who purified it (Reddy et al, 1988) from commercially available material (Boehringer Mannheim, Indianapolis, IN) presumably derived from the SUC2 gene (Taussig and Carlson, 1983). The fully deglycosylated invertase was produced by digesting 1.3 nmol of external invertase in 50  $\mu$ L 3 M urea, 0.1 M sodium phosphate pH 5.7 with 0.03 units of Endo H at 37 °C for 20 hrs, or in 50  $\mu$ L 2-3 M urea, 0.05 M sodium phosphate pH 7.5 with 0.5 units of PNGase F at 37 °C for 20 hrs. Protease Asp-N, chymotrypsin, and Glu-C were purchased from Boehringer Mannheim. Endo Lys-C was obtained from Wako Pure Chemical (Dallas, TX) and, Arg-C and trypsin were purchased from Promega (Madison, WI) and from Boehringer Mannheim. All the proteases used in this study were sequencing grade. PNGase F and Endo H were also obtained from Boehringer Mannheim. Matrix materials (sinapinic acid,  $\alpha$ -cyano-4-hydroxy-cinnamic acid, 2,5-dihydroxybenzoic acid, and HABA) were purchased from Aldrich (Milwaukee, WI).

### *Enzymatic digests:*

Primary digestion. Approximately 10 nmol of invertase were digested with modified trypsin

(0.8% by weight) in 10% ACN, 0.03 M  $\text{NH}_4\text{HCO}_3$  pH 7.8 37 °C for 18 hrs, followed by HPLC fractionation with a linear gradient from 100% A and 0% B to 40% A and 60% B in 60 min (for solvent A and B compositions, refer to the HPLC experimental section). All the fractions were lyophilized in a Speed-Vac vacuum concentrator and redissolved in 200  $\mu\text{L}$  water or 30% ACN in water, then analyzed by MALDI-TOF MS to locate the glycopeptides as well as to confirm the masses of the proteolytic peptide fragments. About one to five percent by volume of a selected HPLC fraction (containing glycopeptides) was treated with PNGase F and/or Endo H and, ten to twenty percent by volume of a selected HPLC fraction (where more than one sequons in a peptide fragment was found) was subdigested with one of the following enzymes: chymotrypsin, Arg-C or Glu-C.

In two other series of parallel experiments, Lys-C and Asp-N were used for the primary digestion of intact external invertase: Approximately 5 nmol of intact protein was cleaved using 1.0% w/w Lys-C in 100  $\mu\text{L}$  3.0 M urea, 0.03 M Tris-HCl (pH 9.0), at rt for 6 hrs. Another 5 nmol of intact external invertase was digested using Asp-N (100:1 substrate/enzyme) in 100  $\mu\text{L}$  2.5 M urea and 0.03 M sodium phosphate (pH 7.8) at 37°C for 3 hrs.

*Subdigestion with chymotrypsin.* Chymotrypsin digestion of fraction T8 + T9 of the trypsin digest was carried out in 100  $\mu\text{L}$  8.0% ACN, 0.01 M Tris-HCl pH 7.6, 0.001 M  $\text{CaCl}_2$ , with 3% w/w chymotrypsin incubating at 37 °C for 18 hrs. The digests were fractionated by HPLC using a Vydac  $\text{C}_{18}$  column (4.6  $\times$  250 mm) with a linear gradient of 100% A and 0% B to 65% A and 35% B in 60 min.



Chymotryptic digestion of fraction L19 of the Lys-C digest was carried out in 42  $\mu$ L 10% ACN, 0.05 M Tris-HCl pH 8.0, with 3% by weight chymotrypsin at 37°C for 3 hrs, followed by HPLC fractionation on a Vydac C<sub>18</sub> column (4.6  $\times$  250 mm) using a linear gradient of 100% A and 0% B to 50% A and 50% B in 45 min.

Subdigestion with endo Arg-C. Selected HPLC fractions were digested in 45  $\mu$ L 0.05 M NH<sub>4</sub>HCO<sub>3</sub> pH 8.0, 0.01 M CaCl<sub>2</sub>, using 0.5  $\mu$ g (about 5% by weight) of Arg-C, incubating at 33 °C for 4.5 hrs. The digests were separated by HPLC on a C<sub>18</sub> column with a linear gradient from 100% A and 0% B to 50% A and 50% B in 30 min.

Subdigestion with endo Glu-C. Endo Glu-C digestion conditions were chosen to cleave at the C-terminal side of glutamic acid only (pH 4) and it was applied to fraction L16 of the Lys-C digest. The lyophilized HPLC fraction was redissolved in 34  $\mu$ L 10% ACN, 0.03 M sodium phosphate buffer (pH 4), with 1  $\mu$ g (3% by weight) Glu-C at rt for 18 hrs. The digests were fractionated by HPLC using a C<sub>4</sub> guard column with a linear gradient from 100% A and 0% B to 50% A and 50% B in 30 min.

Endo H deglycosylation and PNGase F deglycosylation. For the Endo H deglycosylation, 1-5% of each fraction that contained glycopeptide was digested in 50  $\mu$ L 10% ACN, 0.05 M sodium phosphate pH 5.8 using 0.02-0.04 units of Endo H at 37 °C for 18 hrs. In contrast, the same amount of sample was digested with 0.2-0.4 units of PNGase F in 50  $\mu$ L 10% ACN, 0.05 M NH<sub>4</sub>HCO<sub>3</sub> pH 7.8 solution at 37 °C for 18 hrs. After deglycosylation, the digest solution was analyzed directly by

MALDI and/or after desalting using a Vydac C<sub>4</sub> guard column in a HPLC system prior to MALDI MS.

*Reversed Phase High Performance Liquid Chromatography:*

The proteolytic peptides were fractionated on a Vydac C<sub>4</sub> or C<sub>18</sub> column (4.6 × 250 mm) by rp-HPLC using two Waters HPLC pumps (model 510), a gradient programmer (model 680), and a Waters model 490 UV multiwavelength detector monitoring at  $\lambda$  214 nm with a solvent system of A=H<sub>2</sub>O, 0.05% TFA and B=ACN, 0.035% TFA. All the digests were fractionated at a flow rate of 1.0 mL/min with a linear gradient described individually.

HPLC desalting was carried out on a Vydac C<sub>4</sub> guard column operating initially with 100% of solvent A and changing directly to 40% A and 60% B once the solvent front had passed.

*MALDI-TOF MS:*

The average molecular weight of intact external invertase and of the enzymatic digest peptides were determined by MALDI-TOF MS, using a Voyager Elite (PerSeptive Biosystem, Framingham, MA) time-of-flight mass spectrometer. Samples deposited on a stainless steel target were desorbed and ionized by a 337 nm nitrogen laser (Laser Science Inc, Newton, MA) which operated at a 5 Hz repetition rate with 3.5 ns pulse width, using an accelerating voltage of 30 kV or 25 kV. The PerSeptive Voyager Elite can be operated in either the linear mode (flight path 200 cm)

or reflector mode (flight path 300 cm). External calibration was applied to the mass measurement of proteolytic peptides by using a mixture of angiotensin I and bovine insulin (Sigma, St. Louis, MO) which covers the mass range of 500-7000 Da, while bovine serum albumin (BSA,  $M_r$  66430) was used for the mass measurement of intact and deglycosylated invertase. The external calibration method provides a  $\pm 0.1\%$  of mass accuracy in the linear mode and  $\pm 0.05\%$  in the reflector mode. The instrument resolution in mass range of 500-5000 Da was about  $m/\Delta m$  300-600 (FWHM) for the linear mode,  $m/\Delta m$  2000-4000 for the DE mode, and  $m/\Delta m$  1000-2000 for the reflector mode.

Sequencing some of the peptides and deglycosylated peptides was accomplished by post source decay (PSD) measurements. PSD spectra were generated with the Perseptive Voyager Elite time-of-flight instrument, operated in the reflector mode at an accelerating voltage of 20 kV, 50% grid voltage, and 0.1-0.01% stepwise decreased guide wire voltages. The metastable ions which decayed during the free flight time were refocused by decreasing the mirror voltage from 20 to 1.2 kV in 9-11 steps.

Sample preparation for MALDI-TOF MS. Matrix solutions were prepared in acetonitrile/ $H_2O$  50:50 for  $\alpha$ -CHCA (10 g/L) and in acetonitrile/ $H_2O$  15:85 for DHB (10 g/L). Sinapinic acid was used as a saturated solution in 30% acetonitrile. HABA solution was made by first diluting a freshly prepared saturated solution (HABA in 50% ACN) with the same amount of 50% ACN, followed by addition of 20% of MeOH. Sample solutions were mixed with matrix (sinapinic acid (SA) or HABA for the intact molecule and  $\alpha$ CHCA or DHB for the proteolytic peptides) to a final concentration of 1-30 pmol/ $\mu$ L, and 0.5  $\mu$ L (0.5-15 pmol) of this solution was

placed on the target for the MALDI experiment. It has to be noted that a relatively large amount of sample (~ 15 pmol total) is necessary for acquiring the MALDI mass spectra of the entire glycopeptide distribution at each sequon. For instance, a "wide range" type glycopeptide consisting of oligosaccharides from 10 to 50 mannoses would need 40-times more sample than would ordinarily be required for the detection of a single component.

### **III. Identification of a Human Axillary Odor Carrying Protein**

#### **A. Background**

Several studies in humans have suggested that axillary odors and secretions from both males and females are a source of chemical signals containing physiologically active components capable of altering the menstrual cycle. These alterations include the well-accepted menstrual synchrony affect first documented by McClintock in an all female living group (McClintock, 1971) and later replicated by others in co-educational facilities (Graham and McGrew ,1980; Quadagno et al., 1981). In non-human mammals such as rodents, estrus synchrony has been shown to be mediated by airborne chemical signals (McClintock, 1978).

The involvement of axillary components in mediating menstrual synchrony was first suggested by Russell et al. (1980). Studies by Cutler et al. (1986) and Preti et al. (1986) were the first prospectively conducted double-blind studies to attempt menstrual cycle alterations using axillary extracts from male and female donors. The above studies suggest that certain axillary components function as chemical signals involved in the regulation of reproductive function via alteration of the hypothalamic pituitary-gonadal axis; chemical signals with this mode of action are termed primer pheromones (Albone, 1984).

Axillary secretions and odors are derived from an area of the body with exceptional odor-producing capabilities. Several types of skin glands, including apocrine, eccrine, sebaceous, and

apoeccrine glands, contribute moisture and substrate to a large permanent population of cutaneous microflora (Leyden et al., 1981). These consist of lipophilic and large colony diptheroids as well as micrococci. These microorganisms generate a variety of odoriferous compounds that characterize the axillary region. *In vivo* correlations of odor quality and axillary bacterial populations have demonstrated that the aerobic diptheroids are associated with the stronger more distinct axillary odor (Leyden et al., 1981).

A number of investigations of axillary constituents have focused upon the interesting steroidal molecules found there (Rennie et al., 1991; Gower et al., 1993). Volatile odoriferous steroids such as 5  $\alpha$ -androst-16-en-3 $\beta$ -ol (androst-enol) and 5 $\alpha$ -androst-16-en-3-one (androst-enone) as well as non-volatile steroid sulfates were identified and quantitated by radioimmunoassay and gas chromatography/mass spectrometry (Rennie et al., 1991; Gower et al., 1993). The urine/musky-like odors of androst-enone and androst-enol were thought by some investigators to be suggestive of axillary odor (Leyden et al., 1981; Rennie et al., 1991; Gower et al., 1993). However, recent studies (X-N. Zeng et al., 1991; 1992) have presented both organoleptic and analytical evidence that a mixture of C<sub>6</sub>-C<sub>11</sub>, straight-chain, branched and unsaturated acids constitute the characteristic axillary odor. In combined male samples, the *E*-isomer of 3-methyl-2-hexenoic acid (3M2H) is the dominant analytical component of the mixture, while in combined female samples the straight-chain acids are present in greater relative abundance (X-N. Zeng et al., 1996). The *Z*-isomer is also present in both genders, however in different relative abundance: 10:1 (*E*:*Z*) in males (X-N. Zeng et al., 1991); 16:1 (*E*:*Z*) in females (X-N. Zeng et al., 1996).

More than 30 years ago, it was demonstrated that odorless precursors of axillary odor are present in apocrine gland secretions and that the characteristic odor arises from interaction of the odorless apocrine secretion precursors with the axillary microflora (Labows et al., 1982). The water-soluble components of apocrine secretion were found to contain the odorless precursors of the characteristic odor (X-N. Zeng et al., 1992). Furthermore, it was recently shown that 3M2H is carried to the skin surface bound to two proteins, which have been designated as apocrine secretion odor binding proteins: ASOB1 and ASOB2, molecular mass of 45 kDa and 26 kDa, respectively, as determined from SDS-PAGE (Spielman et al., 1995). Spielman et al. (1995) also partially characterized ASOB1 and 2 and produced polyclonal antibodies to each purified protein isolated from male subjects. Antisera raised against ASOB1 demonstrated reactivity to both proteins; however, antisera to ASOB2 reacted only with ASOB2 suggesting relatedness or perhaps a precursor-product relationship. In addition, both ASOB1 and 2 were found to be glycosylated (Spielman et al., 1995).

Although the structures of the physiologically active molecules in the axillae that alter menstrual cycle length and timing are not known, the above results demonstrate that the chemistry of axillary odor production is similar to that found in sources of chemical signals in non-human mammals. In the hamster (Henzel et al., 1988; Singer and Macrides, 1990) and house mouse (Robertson et al., 1993), volatile signals found in vaginal secretions and urine, respectively, are bound to proteins. To better understand the mechanism of axillary odor production and the structural relationship between the ASOB proteins and 3M2H, we used matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) in conjunction with specific

enzymatic cleavages to determine the structure of ASOB2 and the extent of glycosylation of this protein.

## **B. Materials and Methods**

### *Materials:*

Apocrine secretion was collected from paid healthy male donors by Dr. James J. Leyden and Dr. George Preti of the Monell Chemical Senses Center, Philadelphia, PA. Protein (ASOB2) was isolated and purified from sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions by Dr. Andrew I. Spielman, New York University, College of Dentistry, Division of Basic Sciences, New York, NY. Apocrine gland biopsies and *in situ* hybridization were performed by Dr. James J. Leyden and Dr. Benjamin R. Vowels of the Department of Dermatology, School of Medicine, University of Pennsylvania, Philadelphia, PA, respectively. The odor-producing compound, 3-methyl-2-hexenoic acid (3M2H) was identified and quantitated by Dr. George Preti using gas chromatography/mass spectrometry as reported earlier (X-N. Zeng, et al., 1991; X-N. Zeng, et al., 1992). The detailed experimental procedures of the donor selection, apocrine secretion collection, protein purification, assay for 3-methyl-2-hexenoic acid (3M2H) bound to proteins, and *in situ* hybridization have been reported by C. Zeng et al. (1996).



*Reversed-phase high performance liquid chromatography (rp-HPLC):*

Enzymatic digests were separated by rp-HPLC on a Vydac C<sub>4</sub> column (4.6 × 250 mm) at a flow rate of 1 mL/min and using a gradient of 100% solvent A (0.05% TFA in deionized, distilled water) to 60% solvent B (0.035% TFA in ACN) over 60 min using a UV detector (214 nm).

*Enzymatic digests:*

Approximately 200-300 pmol ASOB2 in 100 μL PBS directly eluted from SDS-PAGE was taken to dryness in a Speed-Vac. It was digested using 3% w/w endo Lys-C (Wako Pure Chemical, Dallas, TX) in 60 μL 2.5 M urea, 0.025 M Tris-HCl (pH 9.0) at room temperature (rt) for 5 h, followed by rp-HPLC fractionation. The HPLC fractions containing glycopeptides (as judged from their mass spectra) were deglycosylated with 0.5 units PNGase F (Boehringer Mannheim, Indianapolis, IN) in 45 μL 13% ACN, 0.025 M NH<sub>4</sub>HCO<sub>3</sub> (pH 7.8) at 37 °C for 9 h, further digested with 10% w/w endo Arg-C (Boehringer Mannheim, sequencing grade) in 10 μL 30% ACN, 20 μL buffer (0.1 M Tris-HCl, pH 7.6, 0.01 M CaCl<sub>2</sub>) and 10 μL activation solution (Boehringer Mannheim) at 37 °C for 18 h.

Another HPLC fraction containing the glycopeptide with a complex-type oligosaccharide was treated sequentially with (a) 10 mU neuraminidase in 12 μL 0.02 M NH<sub>4</sub>OAc, pH 5.0, at 37 °C for 5 h; (b) 1 mU β-galactosidase in 11 μL 0.02 M NH<sub>4</sub>OAc, pH 5.0, at 37 °C for 5 h; and (c) 0.6 units PNGase F in 16 μL 0.02 M NH<sub>4</sub>HCO<sub>3</sub>, pH 7.5, incubated at 37 °C overnight. The sugar-free

peptide was again further cleaved to smaller peptides by Arg-C digestion under the conditions described above.

In a separate experiment, another batch of approximately 300-500 pmol of ASOB2, collected from five male donors and dissolved in PBS, was divided into two equal portions; one was hydrolyzed with 5% NaOH, as described above, and analyzed organoleptically and by GC/MS for the presence of 3M2H (X-N. Zeng et al., 1991; Spielman et al., 1995). The other portion was digested with endo Glu-C (about 20% w/w) in 50  $\mu$ L 6% ACN, 0.01 M  $\text{NH}_4\text{OAc}$ , pH 4.0, at rt overnight in a sealed tube. Immediately after opening, the content was first evaluated organoleptically and then extracted with  $3 \times 150 \mu\text{L}$   $\text{CHCl}_3$ . The extract was concentrated and analyzed by GC/MS for 3M2H. The aqueous phase was injected onto an HPLC  $\text{C}_4$ -column and fractions analyzed by MALDI-TOF-MS.

#### *MALDI-TOF MS:*

The molecular weights of ASOB2, deglycosylated ASOB2, and of proteolytic peptides formed by enzymatic digestions were determined by MALDI-TOF MS, using a Voyager Elite (PerSeptive Biosystem, Framingham, MA) or a VT2000 (Vestec Corp., Houston, TX) time-of-flight mass spectrometer.  $\alpha$ -Cyano-4-hydroxycinnamic acid was used as the matrix. The sample on the target was irradiated with 3 nsec pulses of a 337 nm  $\text{N}_2$  laser at a 5 Hz repetition rate. The accelerating voltage was 30 kV (except in the post-source decay mode). Angiotensin I and bovine insulin were used as external calibrants for peptides and trypsinogen for intact and deglycosylated

proteins, resulting in  $\pm 0.1\%$  mass accuracy at a resolution  $m/\Delta m$  of about 300-500.

To obtain the sequence of some of the proteolytic peptides, their post-source decay (PSD) mass spectra (Spengler et al., 1992) were measured with the Voyager Elite TOF spectrometer, operated in the reflector mode at an accelerating voltage of 20 kV. To cover the entire spectrum of fragment ions, the mirror voltage was decreased from 20 to 1.2 kV in 9-11 steps.

#### *Reduction of disulfide bonds:*

About 10 pmol of sample was dissolved in 2  $\mu\text{L}$  0.01 M  $\text{NH}_4\text{HCO}_3$  buffer (pH 7.8) and 1  $\mu\text{L}$  DTT solution (20  $\mu\text{g}/\mu\text{L}$  in water) was added. After reaction for two minutes at rt or 37°C, 0.5  $\mu\text{L}$  of this solution was mixed with 1.5  $\mu\text{L}$  matrix ( $\alpha$ -CHCA) and 0.5  $\mu\text{L}$  was loaded on the target for MALDI-TOF MS.

### **C. Results and Discussion**

Attempts to determine the N-terminal sequence of ASOB2 by the Edman method had failed, indicating that the N-terminus was blocked. Consequently, the structure of this protein was

determined by mass spectrometry. The molecular weights,  $M_r$ , of three independently isolated ASOB2 samples were found to be approximately 22,582 (Figure III.1a), 23,396 and 22,920 by MALDI-TOF-MS. Upon deglycosylation with PNGase F, the  $M_r$  shifted to 19,359 (Figure III.1b), 19,428, and 19,400, respectively. These results suggested that each of the three isolates consisted of the same glycoprotein differing slightly in the size of the attached *N*-linked oligosaccharides. The variations in the molecular weights considerably exceed the 0.1% of the method and are due to the common inhomogeneity of the carbohydrate portion and artifacts of the SDS-PAGE separation (see below).

In order to determine the primary structure of the ASOB2 protein, approximately 250 pmol (based on the intensity of Coomassie blue staining of the SDS gel from which it was eluted) was digested with endo Lys-C. The digest was fractionated on rp-HPLC and 28 fractions were collected (Figure III.2). The more abundant  $(M + H)^+$  ions in the MALDI mass spectra of these fractions were subjected to PSD in an effort to obtain sequence information. Two of the spectra, peptide  $(M+H)^+ = 1671.9$  and peptide  $(M+H)^+ = 1306.6$ , are shown in Figure III.3a and III.3b. Each of these provided a definite partial sequence based on the  $m/z$  values of consecutive fragment ions differing from another by one of the amino acid residue masses (NHCHRCO). For example, the spectrum in Figure III.3a would fit a peptide with a partial sequence ...VQENFDVYPNP..., if one assumes that all the ions of that series are of the N-terminal type  $b_n$  (Biemann, 1990). The N-terminal fragment ( $m/z$  195.1) could correspond to PP, but this peak may also possibly be related to a form of ligand

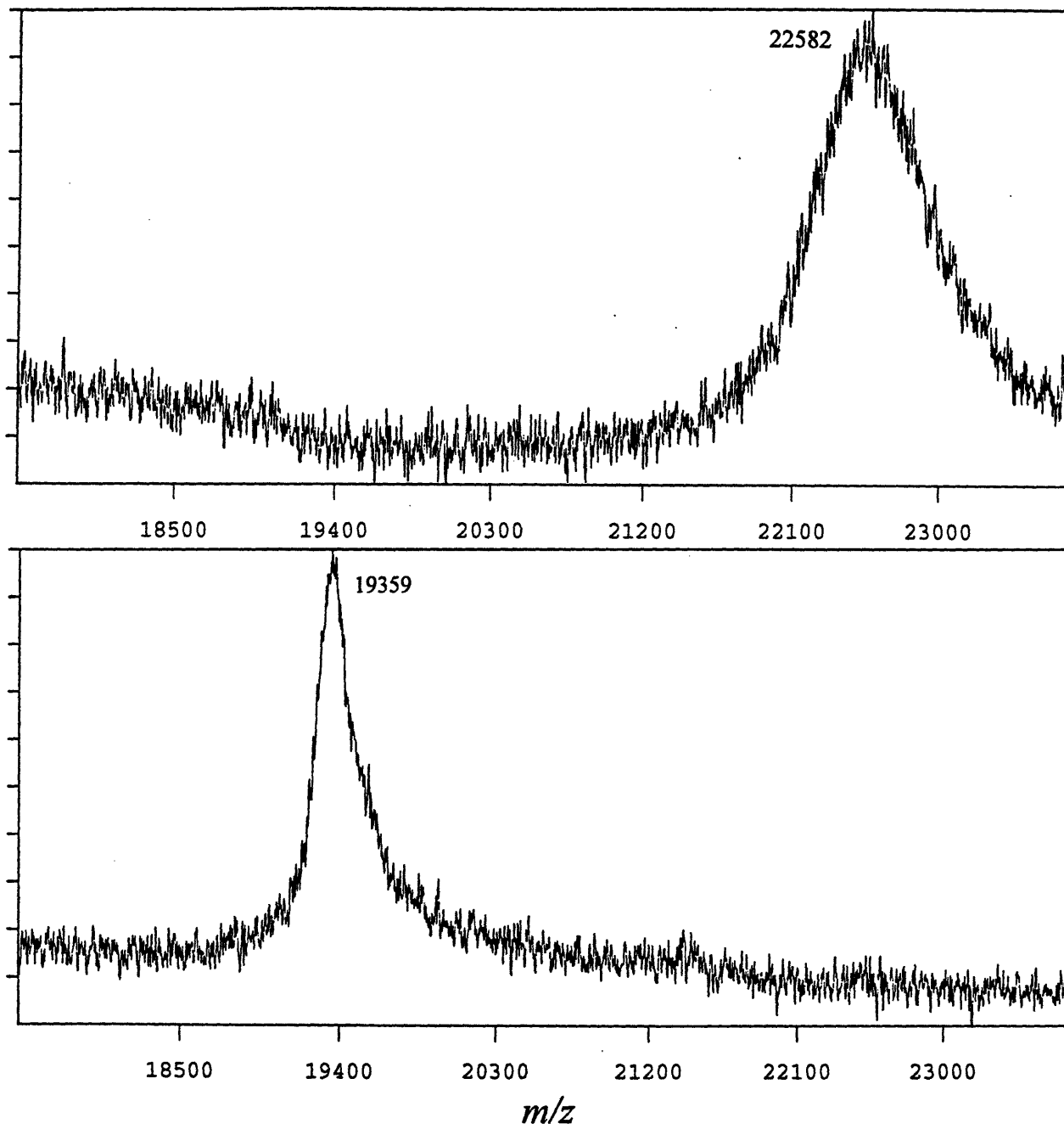


Figure III.1. MALDI-TOF mass spectra of ASOB2. (a) Intact sample isolated by SDS-PAGE. (b) After deglycosylation with PNGase F.

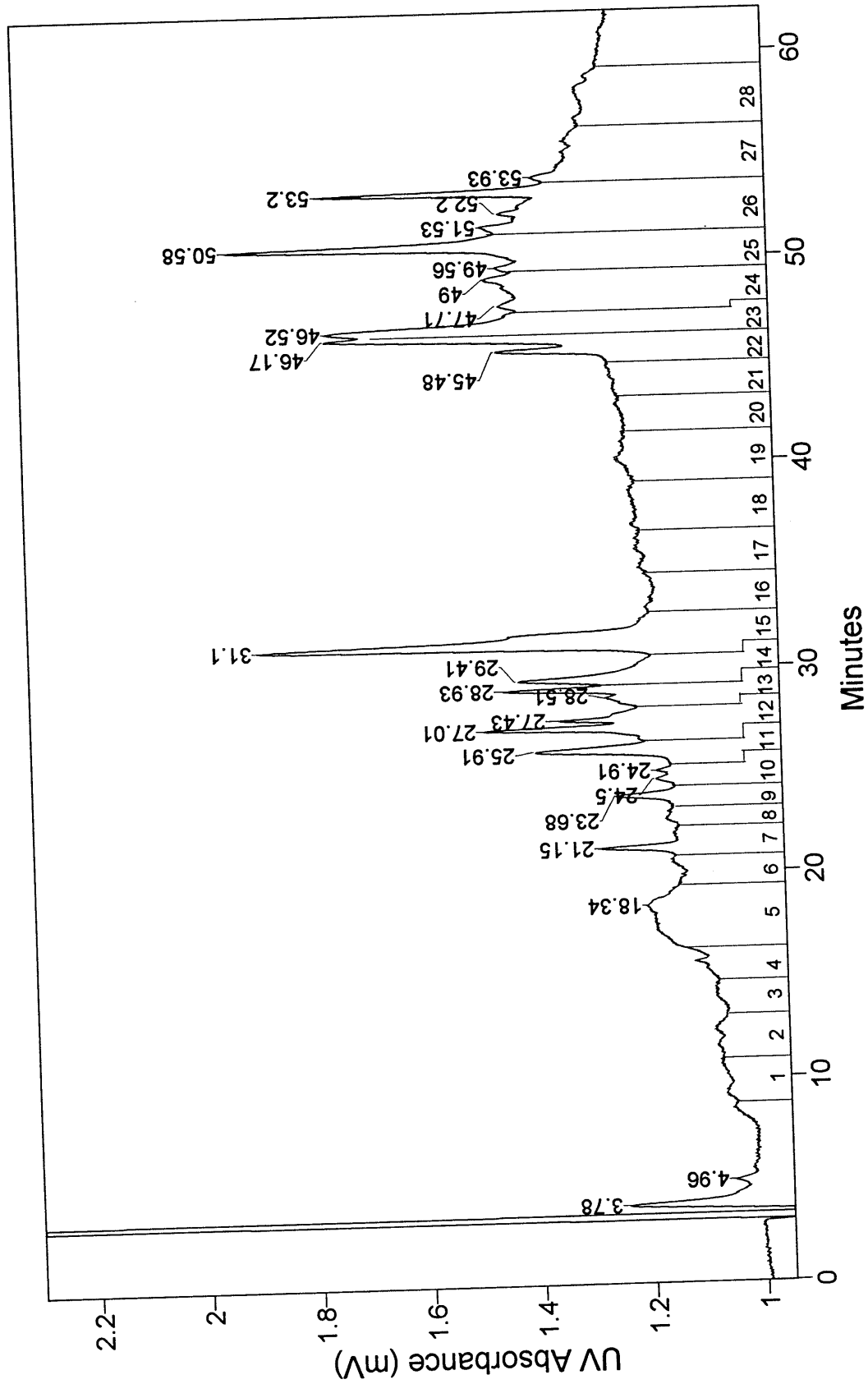


Figure III.2. HPLC trace of the Lys-C digest of ASOB2. The fractions collected are marked.

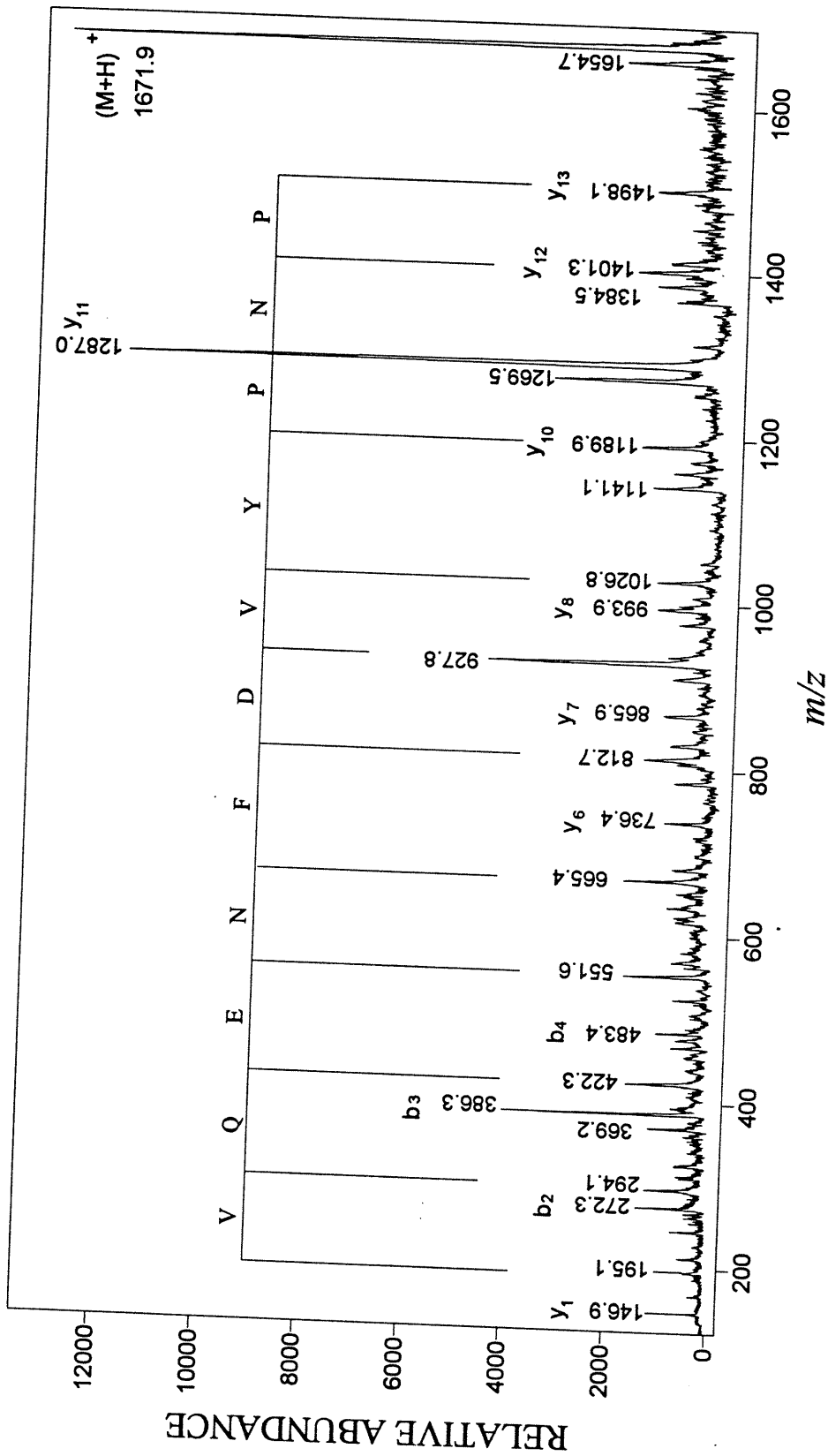


Figure III.3.a. PSD mass spectrum of the (M + H)<sup>+</sup> ion (m/z 1672.9) of C\*PNNPVQENFDVKNK. The notations b<sub>n</sub>, y<sub>n</sub>, etc. refer to fragment ions corresponding to these sequences. The single letter notation in the figure refers to the preliminary partial interpretation discussed in the text and used for database matching.

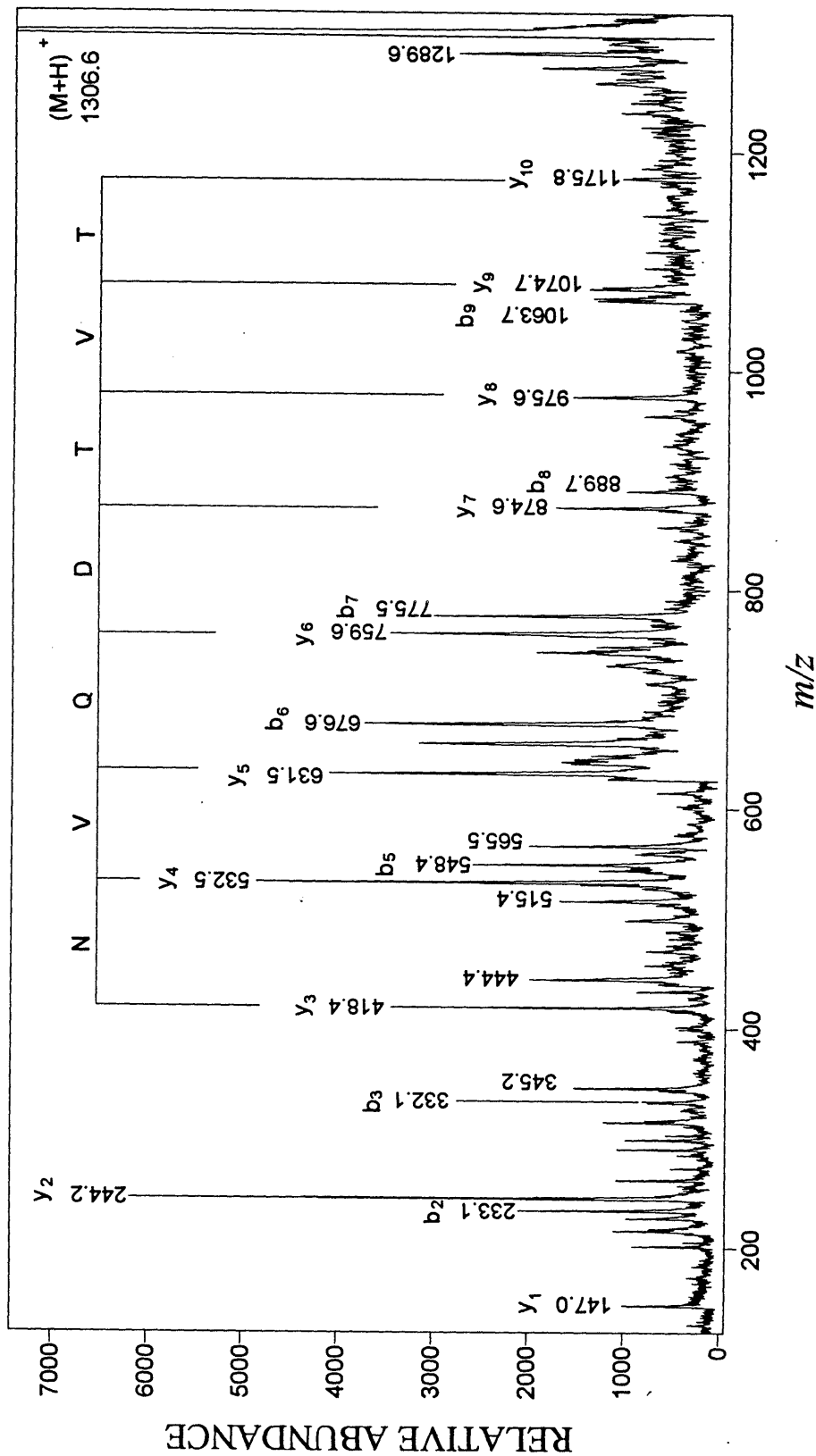
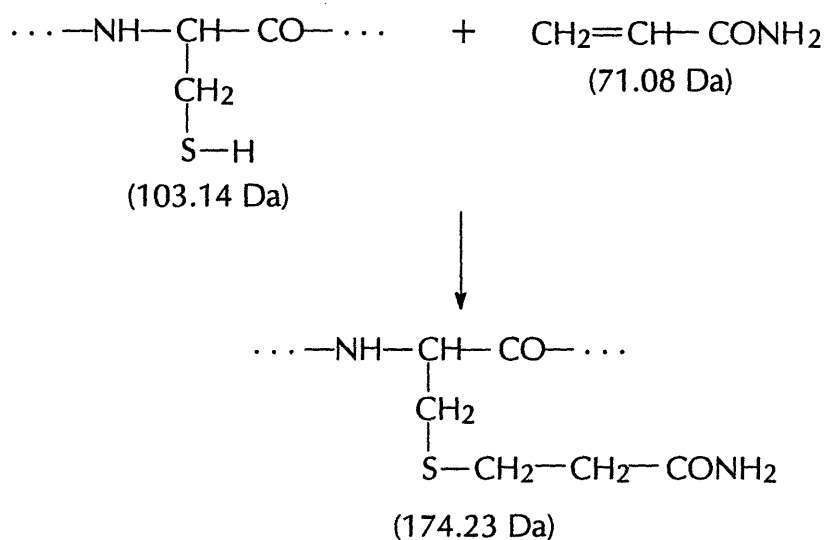


Figure III.3b. PSD spectrum of the  $(M + H)^+$  ion ( $m/z$  1306.6) of MTVTDQVNC\*PK ( $C^* = \text{Cys-CH}_2\text{-CH}_2\text{-CONH}_2$ ). The notations  $b_n$ ,  $y_n$ , etc. refer to fragment ions corresponding to these sequences. The single letter notation in the figure refers to the preliminary partial interpretation discussed in the text and used for database matching.



covalently bound to valine. Similarly, a definite sequence ...NVQDTV... can be deduced from the other spectrum (Figure III.3b). Whether it is to be read from the N- to the C-terminus or the reverse depends on the type of fragments,  $b_n$  or  $y_n$  (Biemann, 1990). Rather than attempting a more complete interpretation, which was hampered by some unassignable mass differences in the low and high mass region (i.e., the N- and/or C-termini of the peptides that later turned out to be due to an unexpected modification of a cysteine), we decided to first check whether the protein is a known one. When independently searching (Collins and Coulson, 1990) the SwissProt protein database (containing 43,470 protein sequences) using the BLITZ server of the European Molecular Biology Laboratory (EMBL) for a match with ...VQENFDVYPNP..., as well as ...NVQDTV... (both forward and reverse), human apolipoprotein D (Hillier and Green, 1991) was found to be a perfect fit for the latter (in reverse) heptapeptide sequence and for the first seven amino acids of the former, respectively. We are grateful to Mr. W. Yu of Genetics Institute (Andover, MA) for carrying out this computer search for us. The mass spectra (Figure III.3) turned out to represent peptides 8-21 and 157-167 of apoD, both of which contained a cysteine that had been unexpectedly S-carbamido-



ethylated by addition of monomeric acrylamide still present in the SDS-PAGE system (Hall et al., 1993). The major ion-series (single letter code) shown in the spectrum (Figure III.3a) and used for database matching turned out, in retrospect, to correspond to a set of "internal" fragment ions (Biemann, 1990) representing the sequence PPVQENFDV and triggered by the preferential primary cleavage at the N-P peptide bond. This result is an impressive example of the power of protein database searching utilizing partial sequence information derived from the mass spectra of peptides, even if they contain an unexpected and unknown modification of one of the amino acids (see Figure III.3). The complete amino acid sequence of mature apoD and the matching proteolytic peptides (underlined) are shown in Figure III.4. Three PSD mass spectra, including that of the N-terminal pyroglutamic acid fragment (<QAFHLGK), are shown in Figures III.5, III.6, and III.7 which provide additional examples of the interpretation of such spectra in terms of peptide sequence.

The PSD spectrum (Figure III.8) of  $(M+H)^+ = m/z$  1601.7, an ion of low abundance in the MALDI spectrum of HPLC fraction 8 was identified as peptide 8-21 with free Cys<sup>8</sup> (calc. 1601.8). Even though the signals in the PSD spectrum were weak, they compared well with those of the spectrum of the same peptide (see Figure III.3a) which is a major component but carbamidoethylated. The first hint of this identification was the mass difference of 71 Da between the  $(M+H)^+$  ions. This finding indicates that the S-carbamidoethylation was not complete. The partial carbamidoethylation of at least two cysteines explains the previously mentioned variations in the measured  $(M + H)^+$  ions of the PNGase F-deglycosylated ASOB2, which were higher than 19,289.1 as calculated for apoD (fully reduced, Asp in positions 45 and 78). It should be noted that another minor component of  $(M + H)^+ = m/z$  1677.1 (detected in the DE mode) found in fraction 10

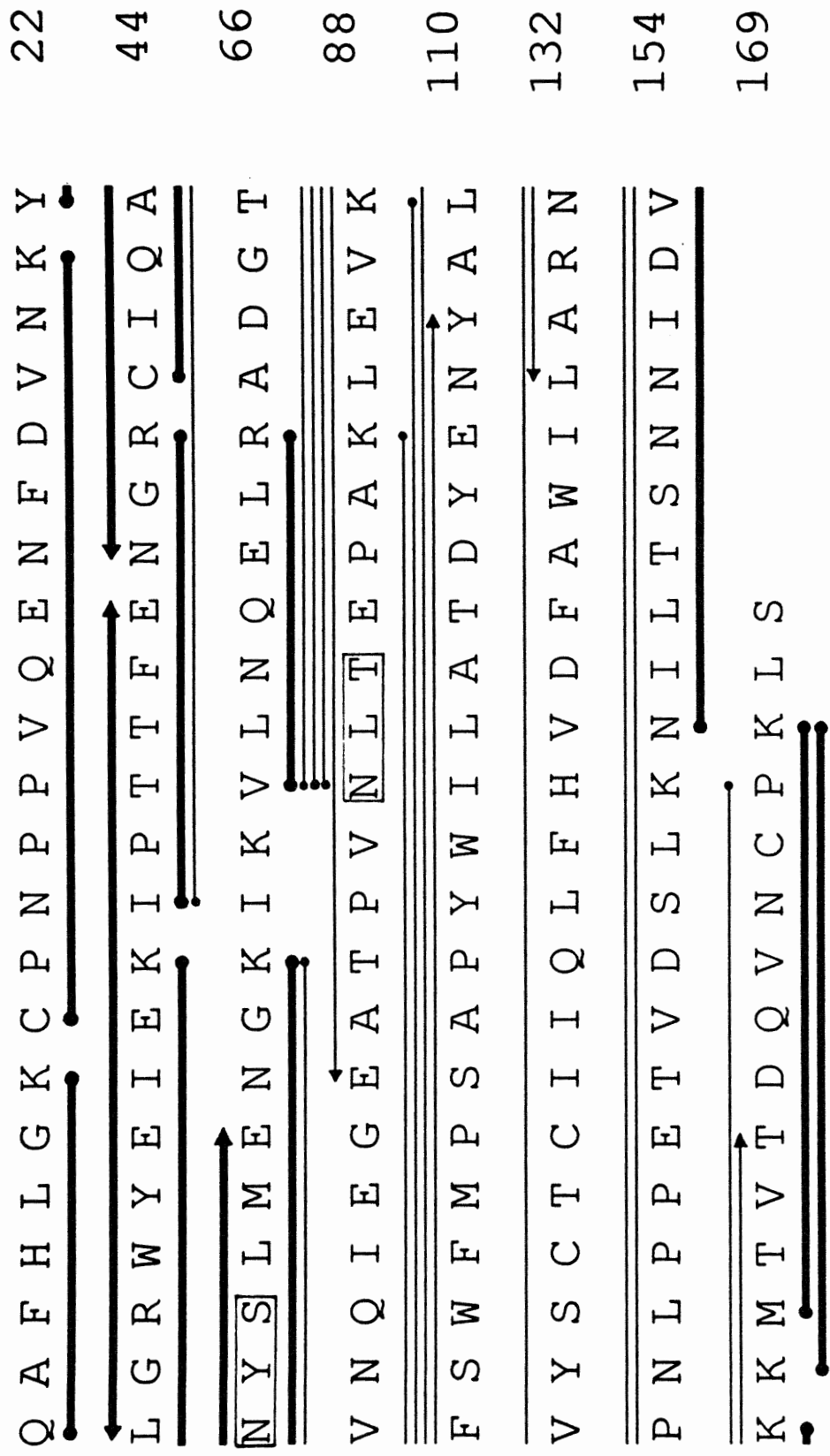


Figure III.4. Amino acid sequence of human apolipoprotein D (Drayna et al., 1986). —•—: peptides observed in endo Lys-C digest; ←—→: peptides observed in endo Glu-C digest. Heavy lines indicate sequences confirmed by MALDI-PSD mass spectra. The two glycosylated sequons (N-X-S/T) are boxed.

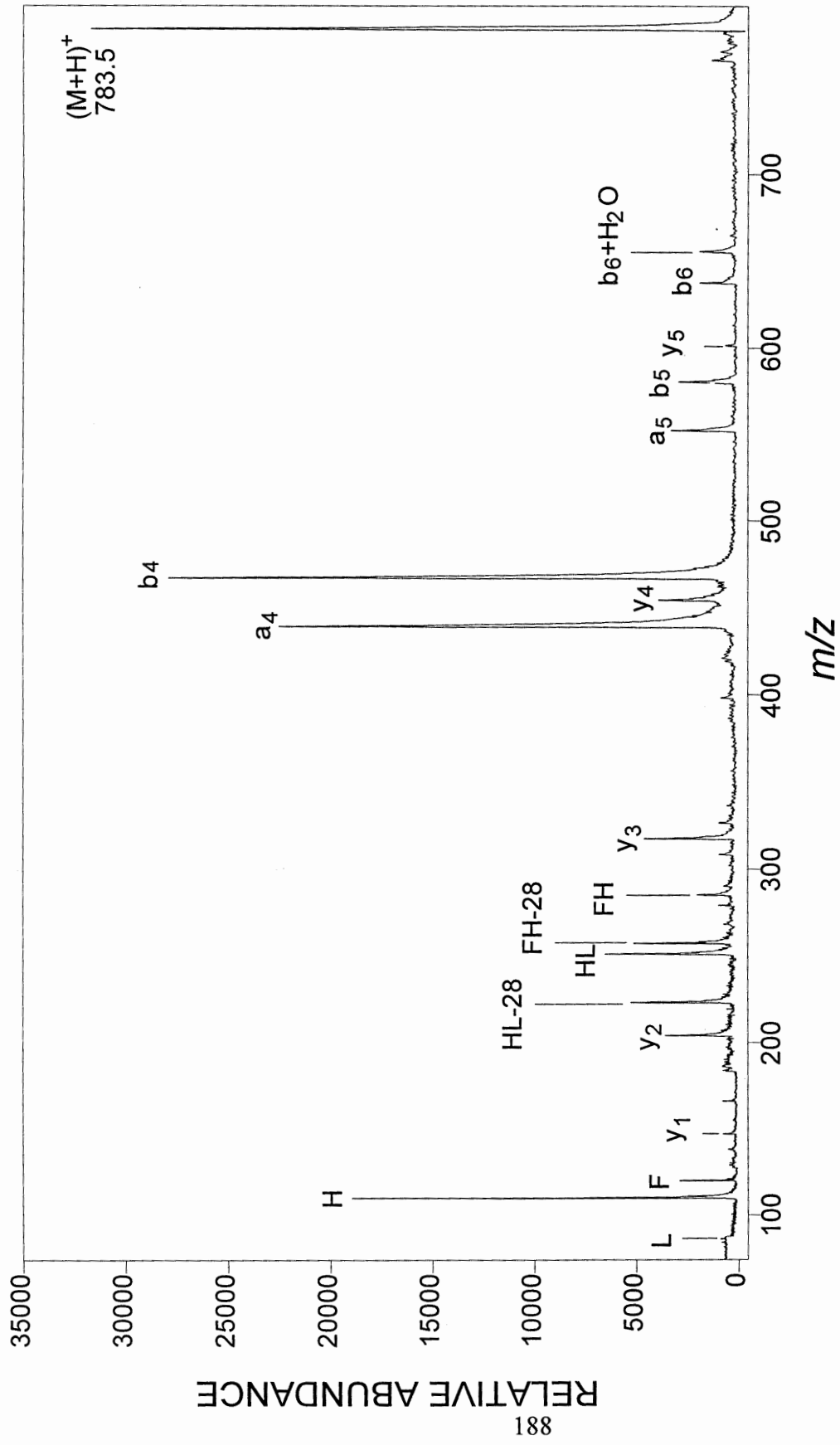


Figure III.5. PSD spectrum of (M+H)<sup>+</sup> = m/z 783.5, the N-terminal peptide of apoD, <QAFHLGK.

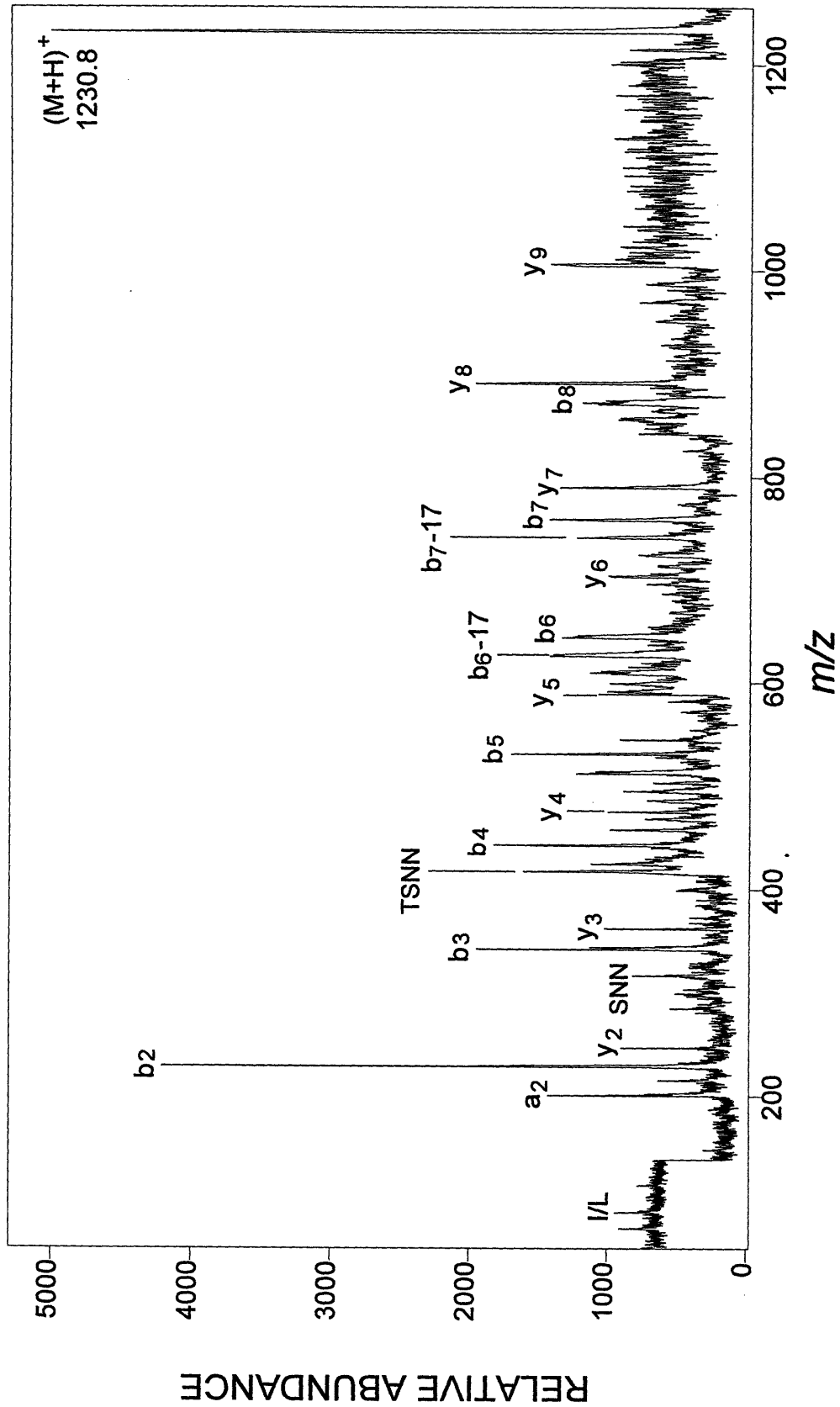


Figure III.6. PSD spectrum of (M+H)<sup>+</sup> = m/z 1230.8, peptide 145-155, NILTSNNIDVK.

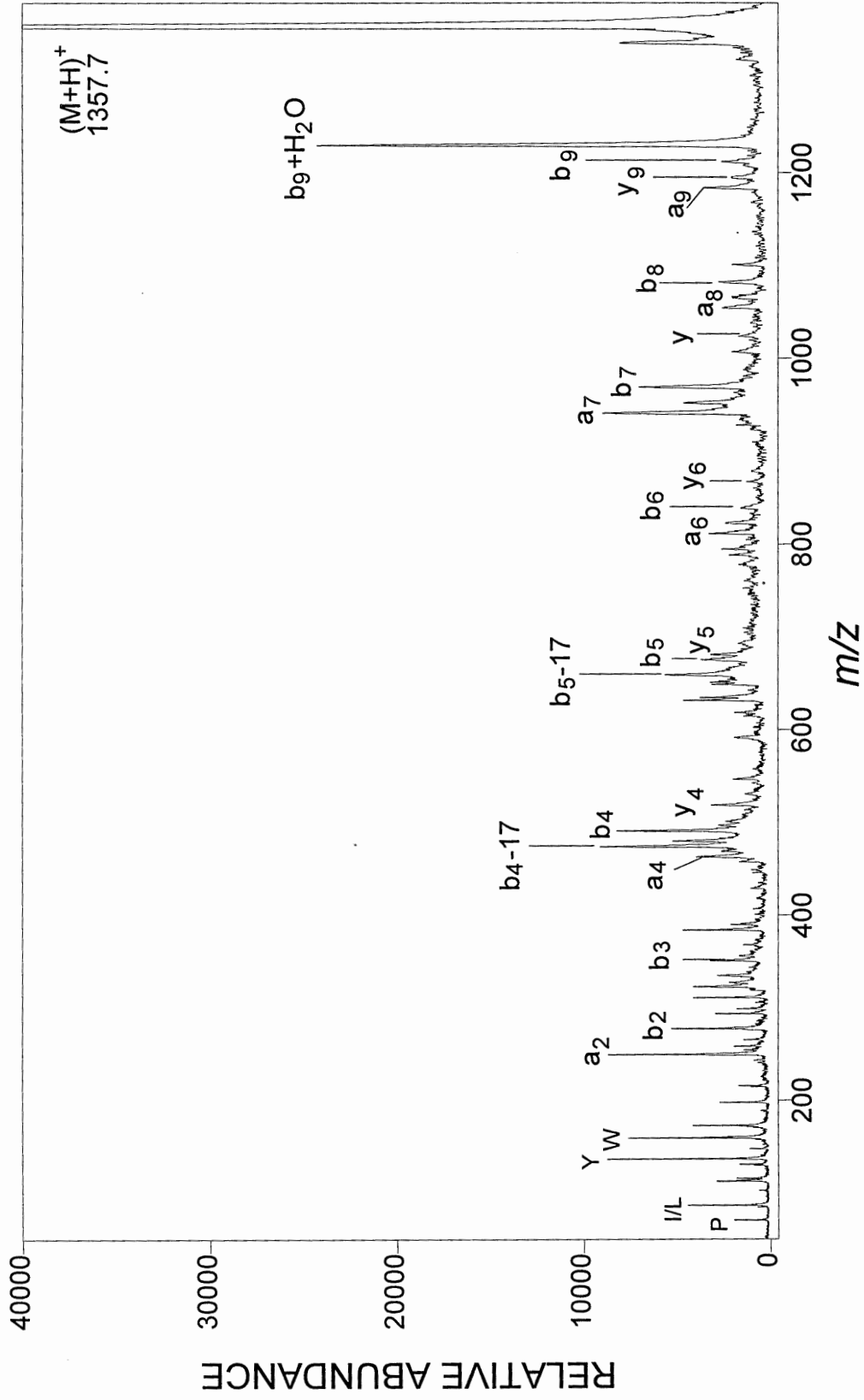


Figure III.7. PSD spectrum of (M+H)<sup>+</sup> = m/z 1357.7, peptide 22-31, YLGRWYEIEK.

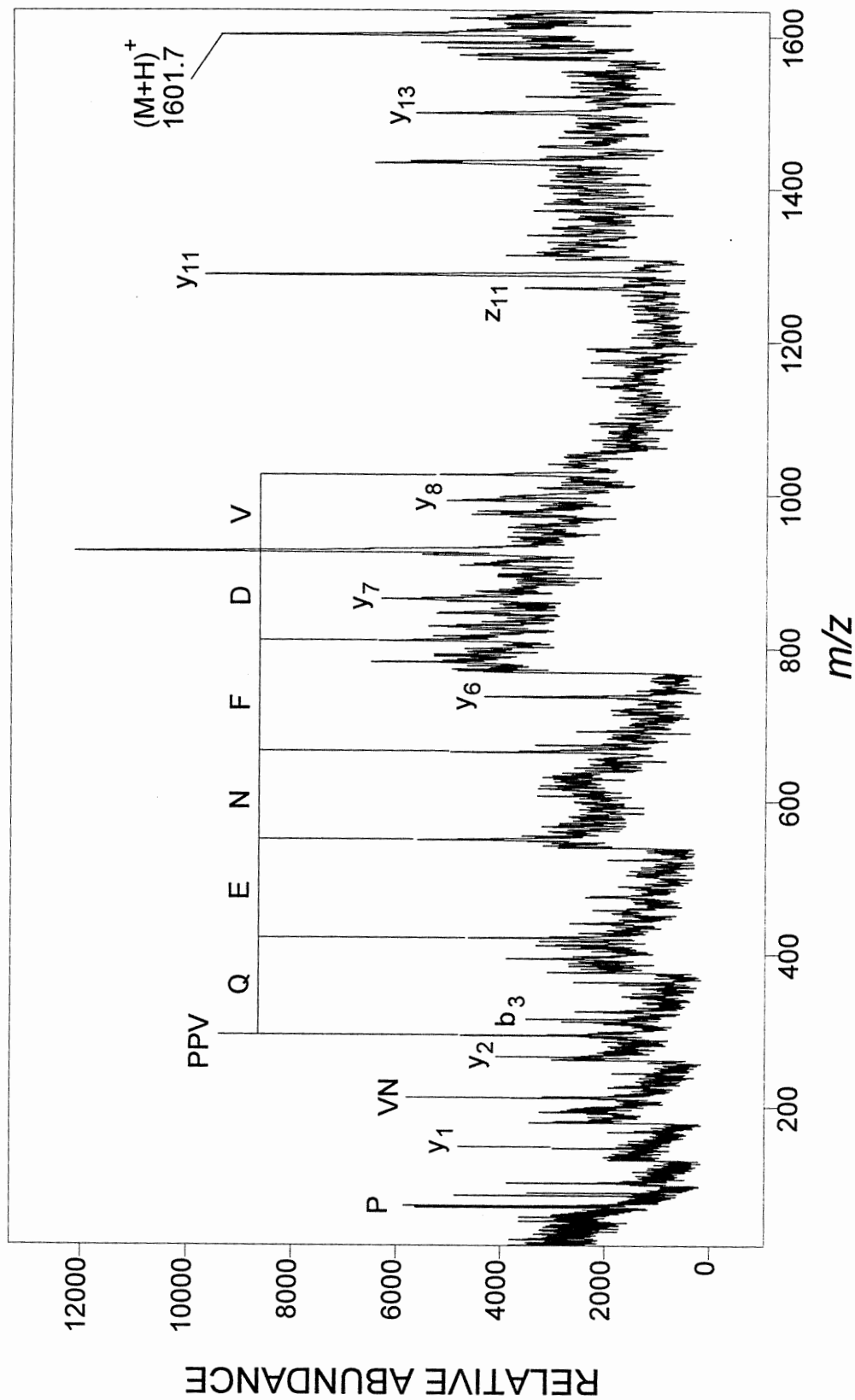
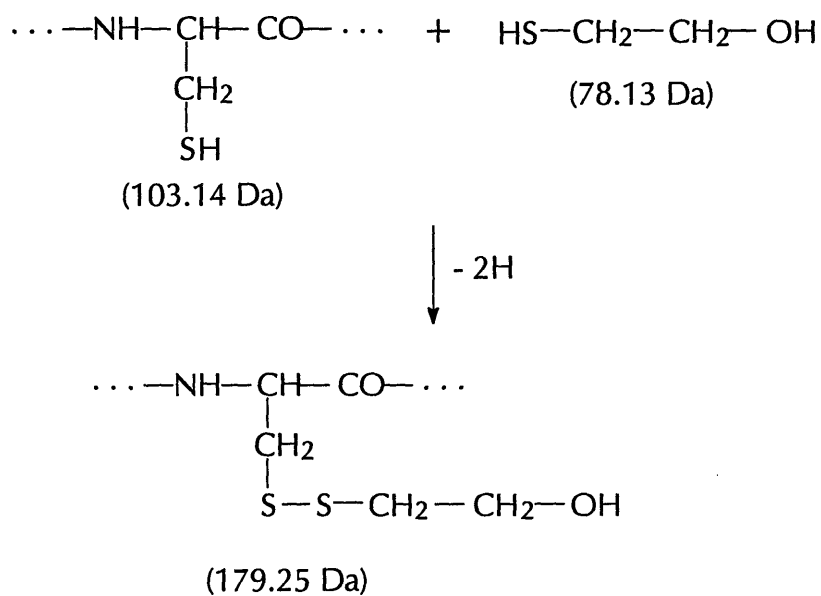


Figure III.8. PSD spectrum of  $(M+H)^+ = m/z$  1601.7, peptide 8-21, CPNPPVQENFDVVK. Notice the space of 103 Da between the  $(M+H)^+$  ion and  $y_{13}$  ion indicates the free cysteine in the N-terminus. Notations refer to Figure III.3.

(see Figure III.2) corresponds to peptide 8-21, where Cys<sup>8</sup> was disulfide linked with mercaptoethanol (calc. 1678.8 for the monoisotopic protonated ion). This was deduced from the N-terminal ions (b<sub>2</sub>, b<sub>3</sub>, and b<sub>4</sub>) and the C-terminal ions (y<sub>11</sub>, y<sub>12</sub>, and y<sub>13</sub>) in the PSD spectrum (Figure III.9), which otherwise was very similar to the major component of fraction 9 (peptide 8-21) shown in Figure III.3a.



The MALDI mass spectrum of fraction 12 (see Figure III.2) exhibited a multiplet of peaks spaced 291 Da apart in the high mass region (Figure III.10a and Table III.1), indicative of a carbohydrate moiety terminating in sialic acids. Successive treatment with neuraminidase, β-galactosidase, and PNGase F produced single peaks at *m/z* 4182, 3858, and 2559, respectively (Figure III.10b, 10c, and 10d). These results indicated that this glycopeptide spans the region of amino acids 32-53 (carbamidoethylcysteine at position 41), (M + H)<sup>+</sup> calc. *m/z* 2559.9 (for the



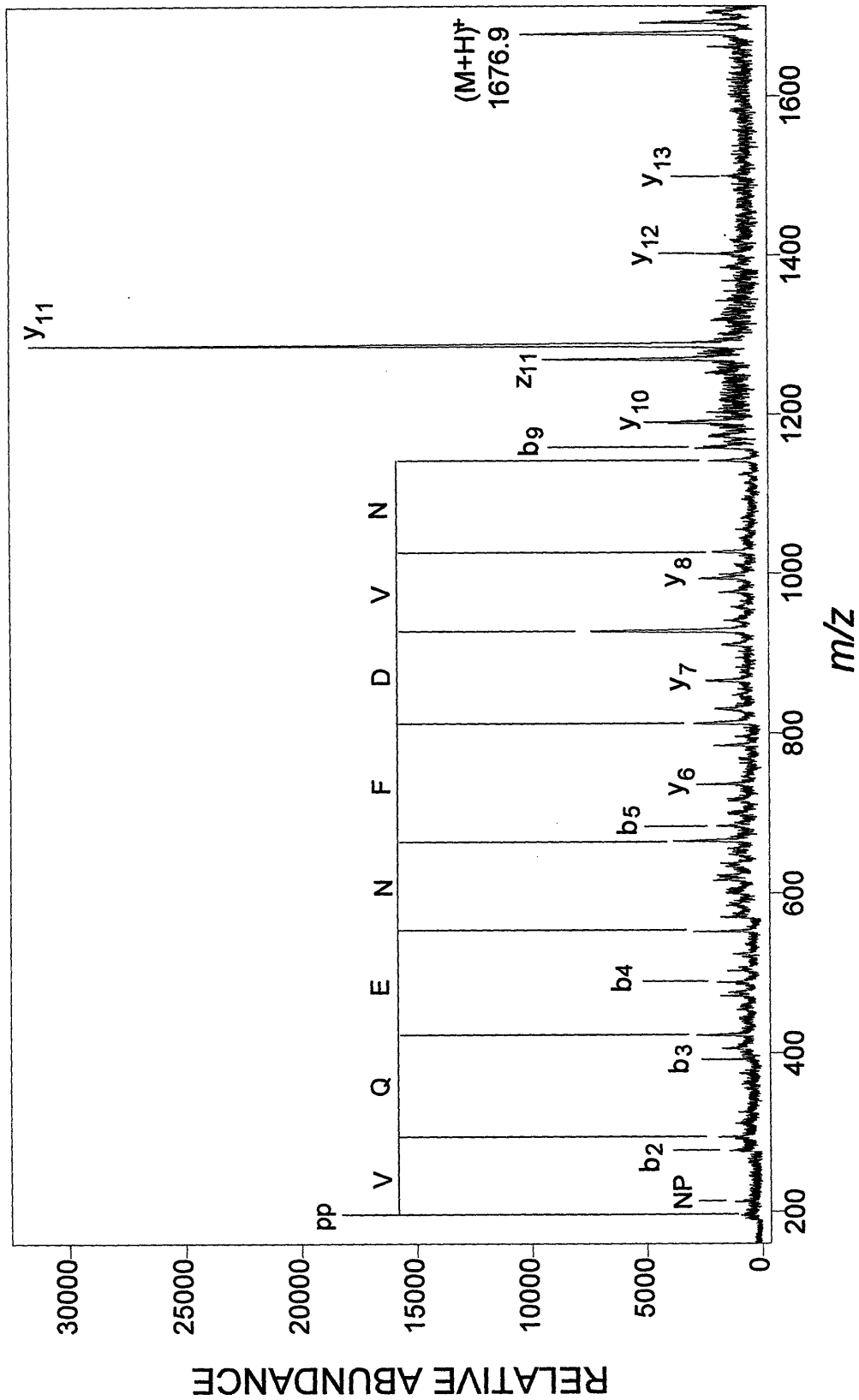
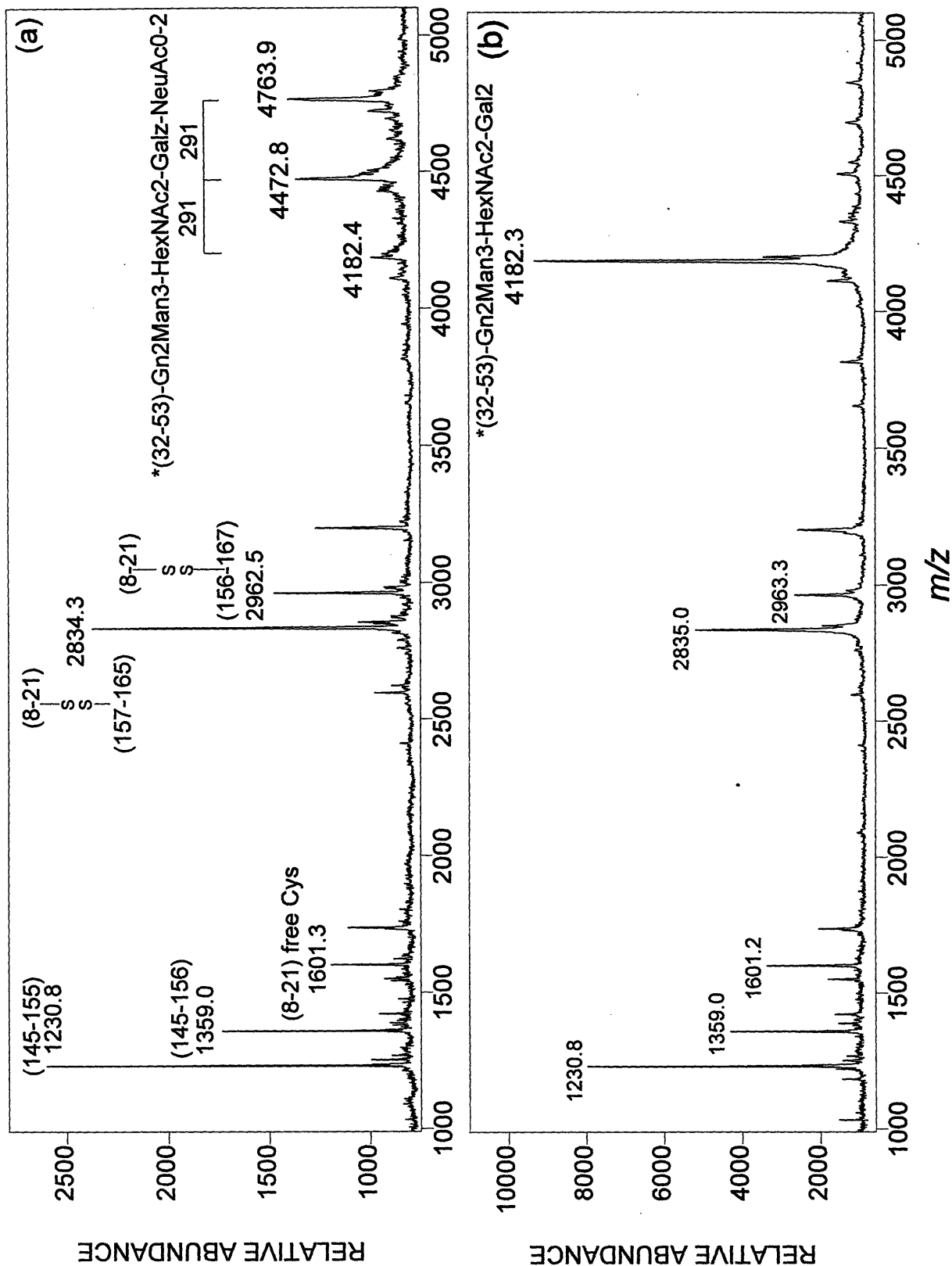
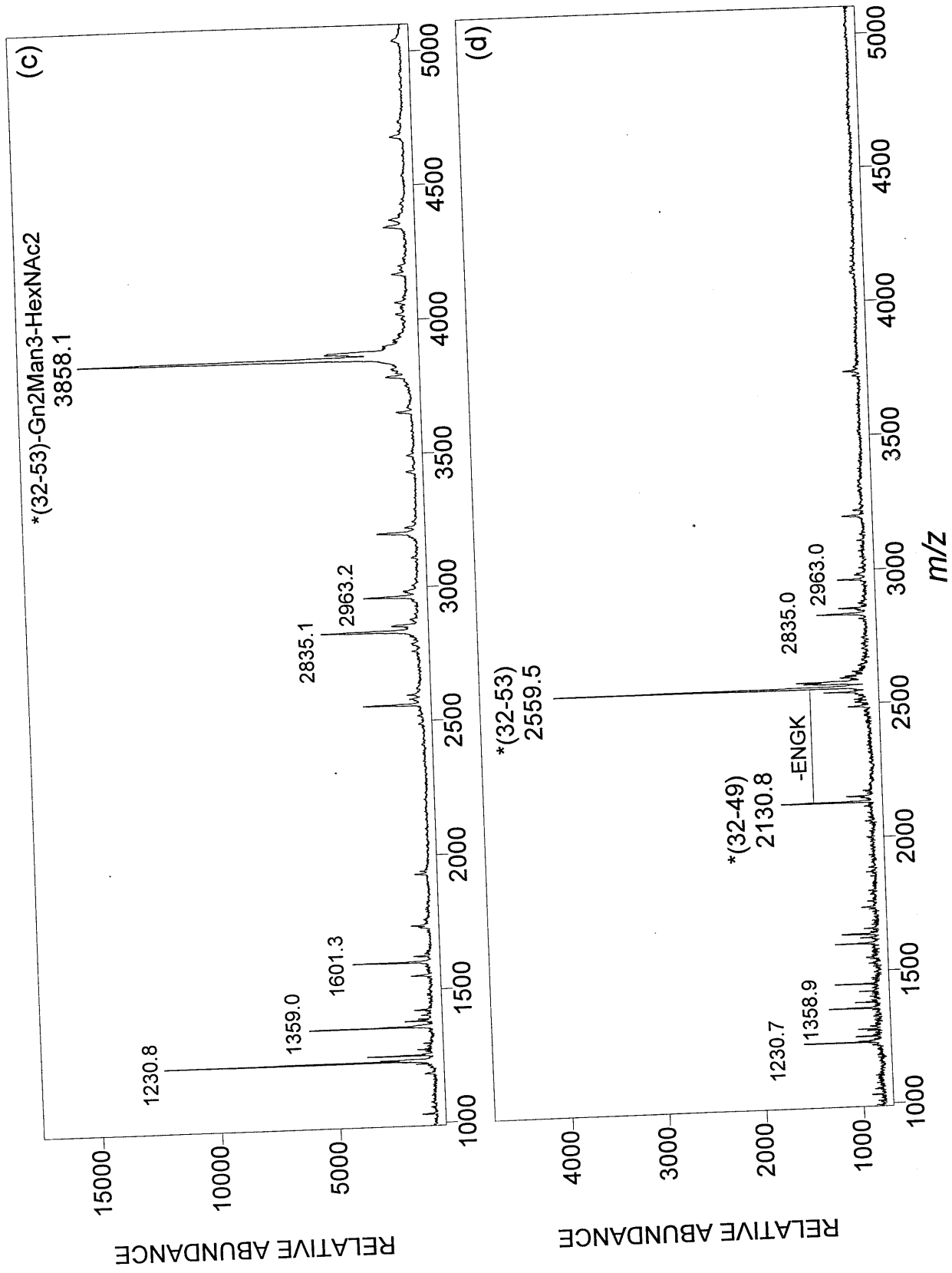


Figure III.9. PSD spectrum of  $(M+H)^+ = m/z$  1676.9, peptide 8-21 with the modified Cys<sup>8</sup>. Notice the space of 179 Da between the  $(M+H)^+$  ion and  $y_{13}$  ion indicates the N-terminal cysteine is disulfide linked with mercaptoethanol.

Figure III.10 (two pages). (a) The MALDI spectrum of HPLC fraction 12. (b), (c), and (d) The MALDI spectra of fraction 12 treated sequentially with neuraminidase,  $\beta$ -galactosidase, and PNGase F, respectively. In Figure III.10a, the peak at  $m/z$  1230.7 has been identified to be peptide 145-155 by PSD (see Figure III.6) and the peaks at  $m/z$  1359.0 and 1601.3 matched to the longer Lys-C peptide chain 145-156 and peptide 8-21 with the free Cys<sup>8</sup>. The peaks in the region of  $m/z$  2,500 to 3,500 are due to a pair of disulfide linked peptides as discussed later.



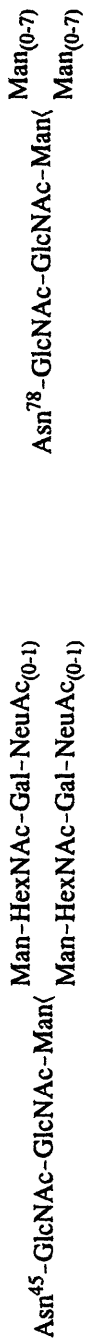


**Table III.1.** Summary of the characterization of the carbohydrate moieties at Asn<sup>45</sup> and Asn<sup>78</sup>

HPLC fraction	$m/z$ of (M+H) <sup>+</sup>	$m/z$ of (M+H) <sup>+</sup> after digestion with 1. neuraminidase 2. $\beta$ -galactosidase 3. PNGase F; Endo H <sup>a</sup>	Glycosylation Site	Amino acid position	proposed carbohydrate composition <sup>b</sup>
12	4182		Asn <sup>45</sup>	32-53	HexNAc <sub>4</sub> Hex <sub>5</sub> NeuAc <sub>0</sub>
	4473	4182			HexNAc <sub>4</sub> Hex <sub>5</sub> NeuAc <sub>1</sub>
	<b>4764</b>	3858			<b>HexNAc<sub>4</sub>Hex<sub>5</sub>NeuAc<sub>2</sub></b>
13	3839		Asn <sup>78</sup>	56-84	HexNAc <sub>2</sub> Hex <sub>2</sub>
	4001				HexNAc <sub>2</sub> Hex <sub>3</sub>
	4163				HexNAc <sub>2</sub> Hex <sub>4</sub>
	<b>4325</b>	no change			<b>HexNAc<sub>2</sub>Hex<sub>5</sub></b>
	4488	3109			HexNAc <sub>2</sub> Hex <sub>6</sub>
	4652	3312			HexNAc <sub>2</sub> Hex <sub>7</sub>
4817		HexNAc <sub>2</sub> Hex <sub>8</sub>			
	4980		HexNAc <sub>2</sub> Hex <sub>9</sub>		

<sup>a</sup> For fraction 13 only.

<sup>b</sup> Structures proposed based on these data and known biogenesis of glycosylation :



deglycosylation product), which must bear a complex oligosaccharide of the composition (HexNAc)<sub>4</sub> (Hex)<sub>5</sub> (NeuAc)<sub>0-2</sub> at the glycosylation site Asn<sup>45</sup>. The identity of the deglycosylated peptide was further confirmed by digestion with Arg-C, producing (M + H)<sup>+</sup> ions at *m/z* 1034.1 (calc. 1034.5 for the monoisotopic species) and *m/z* 1543.0 (calc. 1543.4), respectively, corresponding to peptides 32-40 and 41-53. Their PSD spectra corroborated the expected amino acid sequences. The PSD spectrum of one of them, (M+H)<sup>+</sup> = *m/z* 1034.1 (Figure III.11), was clearly interpreted (from the sequence-defining concatenated set of y<sub>1</sub>-y<sub>8</sub> ions along with other a<sub>n</sub>, b<sub>n</sub> ions and internal acyl and immonium ions) to be IPTTFENGR.

The MALDI mass spectrum of fraction 13 (see Figure III.2) consisted of a pattern of peaks (Figure III.12a and Table 1) spaced about 162 Da apart, typical of glycopeptides differing in the number of hexose residues. Upon deglycosylation with PNGase F, the multiplet collapsed to *m/z* 3109 (Figure III.12b), which best fit the peptide 56-84 (calc. 3109.9) and includes the other glycosylation site, Asn<sup>78</sup>. This identification was further confirmed by digestion of the peptide with Arg-C, producing the expected smaller peptides, 56-62 of (M + H)<sup>+</sup> = *m/z* 872 (calc. 871.7), and 63-84 of (M + H)<sup>+</sup> = *m/z* 2257 (calc. 2257.2). The PSD spectrum of the former revealed the expected sequence, VLNQELR (Figure III.13).

Because there was no Lys-C peptide observed with a mass indicating the covalent attachment of 3M2H, it was important to prove that it was still present in the protein sample being analyzed. Therefore, another batch of ASOB2 (estimated to represent 300-500 pmol from Coomassie blue staining of the SDS-gel from which the protein had been eluted) dissolved in PBS was divided into

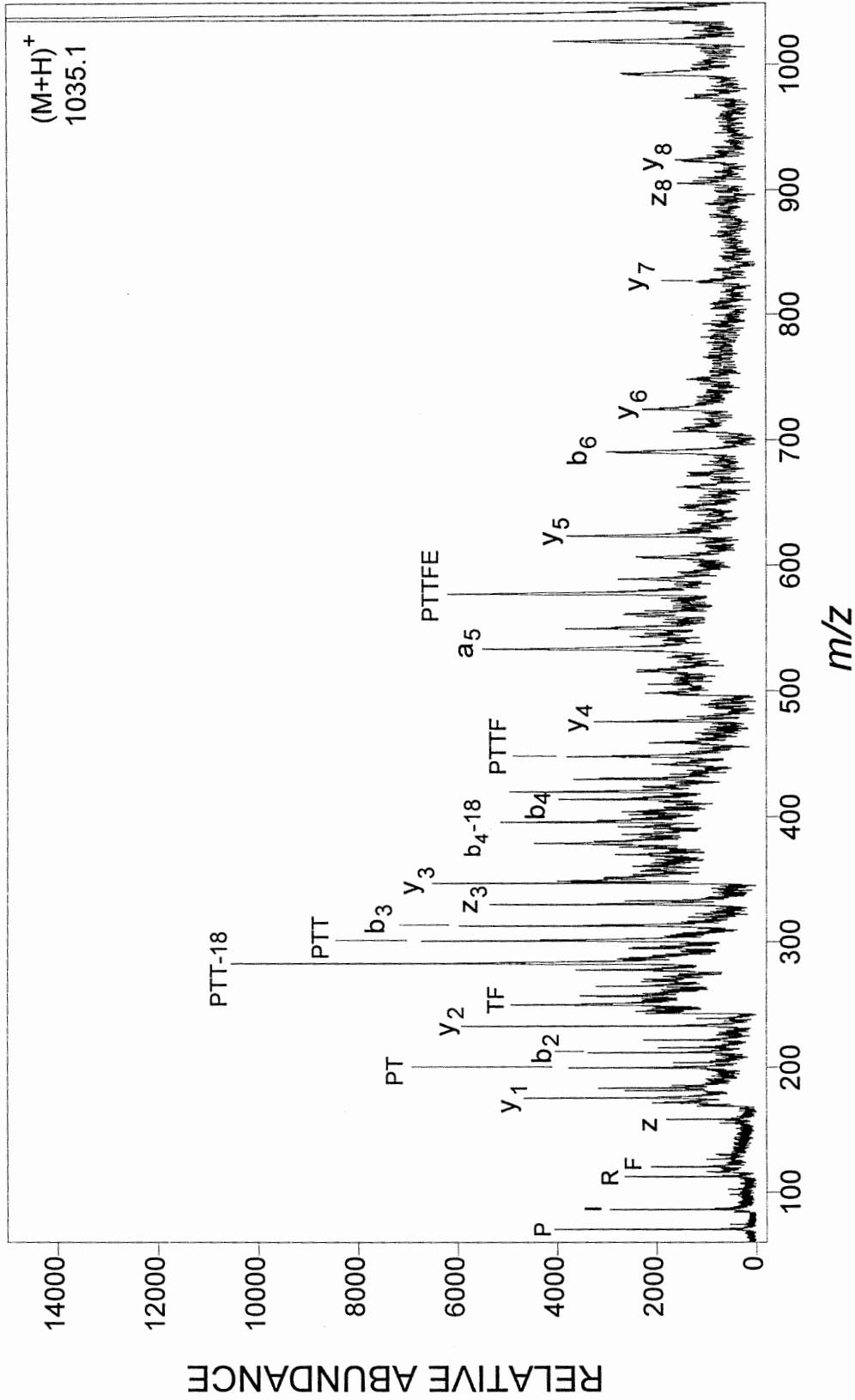


Figure III.11. PSD spectrum of peptide  $(M+H)^+ = m/z$  1035.1, peptide 32-40, IPTTFENGR.

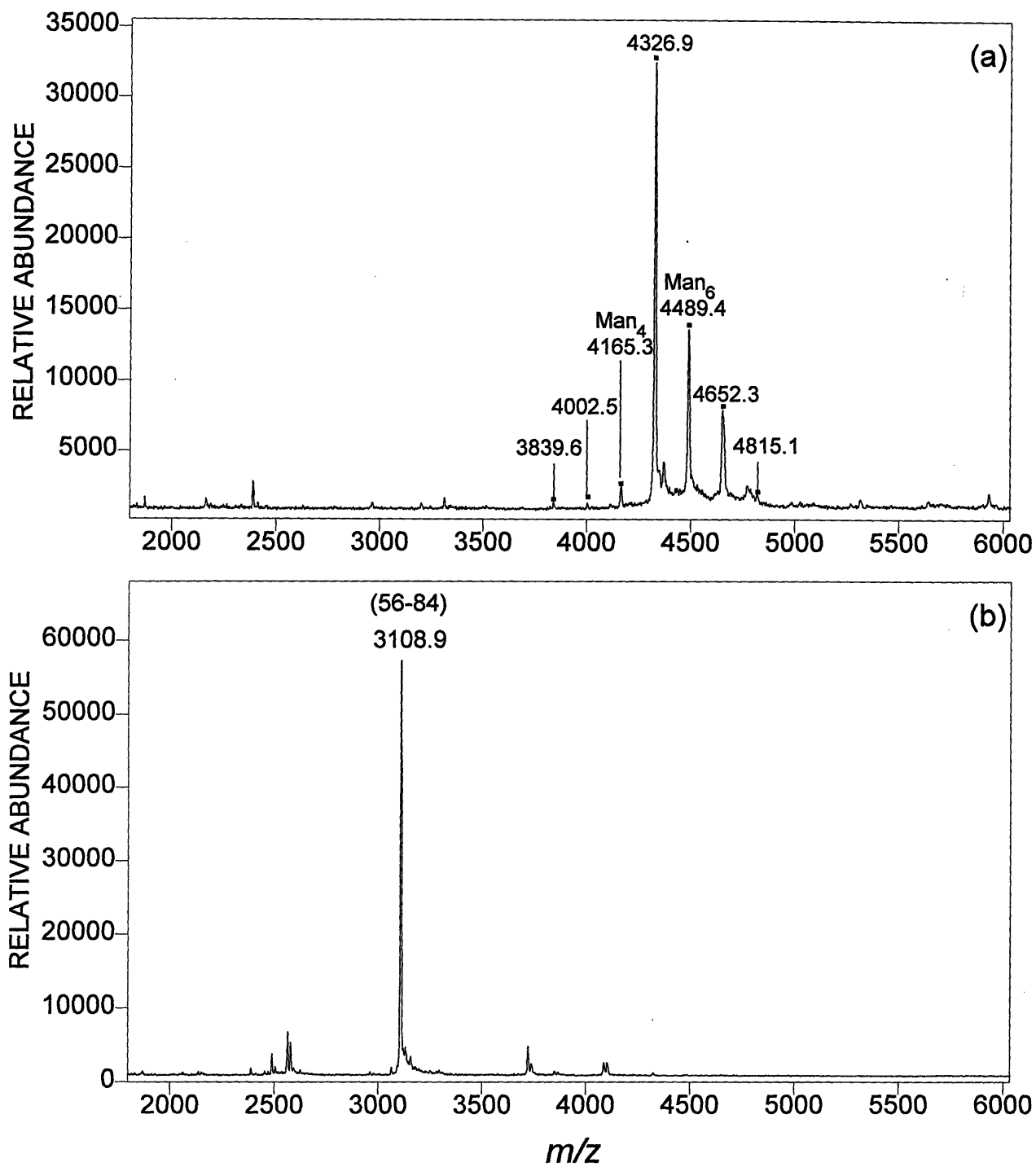


Figure III.12. The MALDI spectra of HPLC fraction 13 before (a) and after (b) deglycosylation with PNGase F.



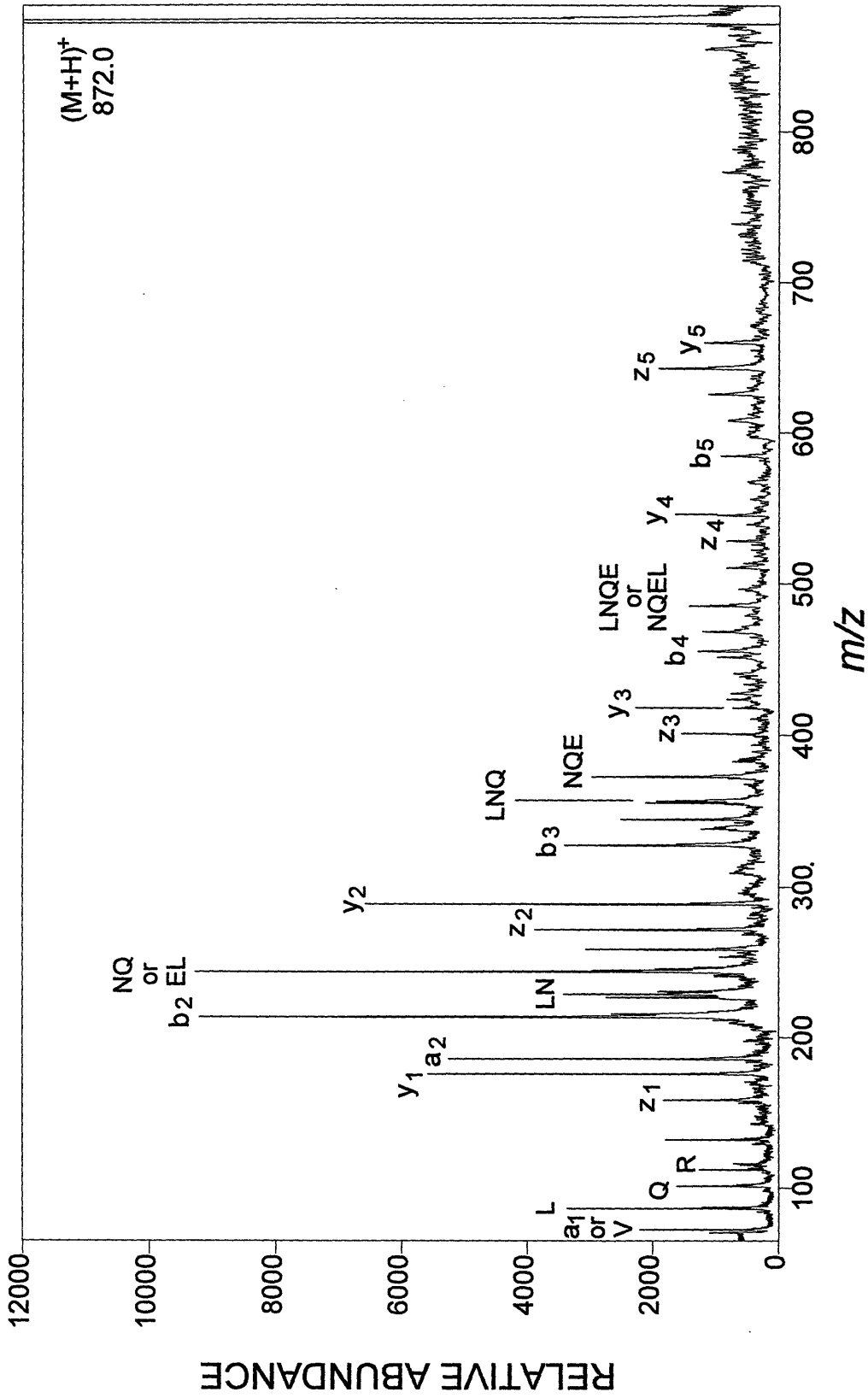


Figure III.13. PSD spectrum of peptide (M+H)<sup>+</sup> = m/z 872.0, peptide 56-62, VLNQLER.

two equal portions. One portion was subjected to base hydrolysis and the hydrolyzate analyzed by GC/MS for the presence of 3M2H, which was indeed detected. To liberate and detect non-covalently bound odor components, the remaining portion of unhydrolyzed ASOB2 was digested with endo Glu-C at pH 4.0, conditions under which ester bonds would remain intact. A  $\text{CHCl}_3$  extraction of the digest was concentrated and analyzed for free 3M2H, both organoleptically and by GC/MS. The latter indicated the presence of approximately 280 pmol 3M2H, with the *E*-isomer predominating over the *Z*-isomer. The  $(M + H)^+$  ions of the peptides and some of their PSD spectra from this digest further confirmed that the carrier protein is apoD (Figure III.4). In addition, the strength of the signals in the HPLC (Figure III.14), the abundance of the  $(M + H)^+$  ions of the intact ASOB2 in this sample, and of those of the peptides present in the digest suggested that the amount of ASOB2 digested in this experiment was approximately 100-150 pmoles, instead of 150-250 pmol as the Coomassie stain would have indicated.

During the study of the Glu-C digest, the Cys<sup>8</sup> in peptide 1-15 was found again to be modified in two different ways. One was the addition of monomeric acrylamide discussed earlier, giving rise to an  $(M+H)^+$  ion of  $m/z$  1720. The other peptide,  $(M+H)^+ = m/z$  1724, found in fraction 4 (Figure III.15a) gave a PSD spectrum (Figure III.15b), exhibiting a gap of 179 Da between the  $b_7$  and  $b_8$  ions, indicating disulfide bond formation with 2-mercaptoethanol. The latter was confirmed by reduction with DTT which resulted in a  $(M+H)^+$  ion of  $m/z$  1648 (calc. 1648.9) corresponding to the reductive loss of mercaptoethanol (Figure III.16a and III.16b).

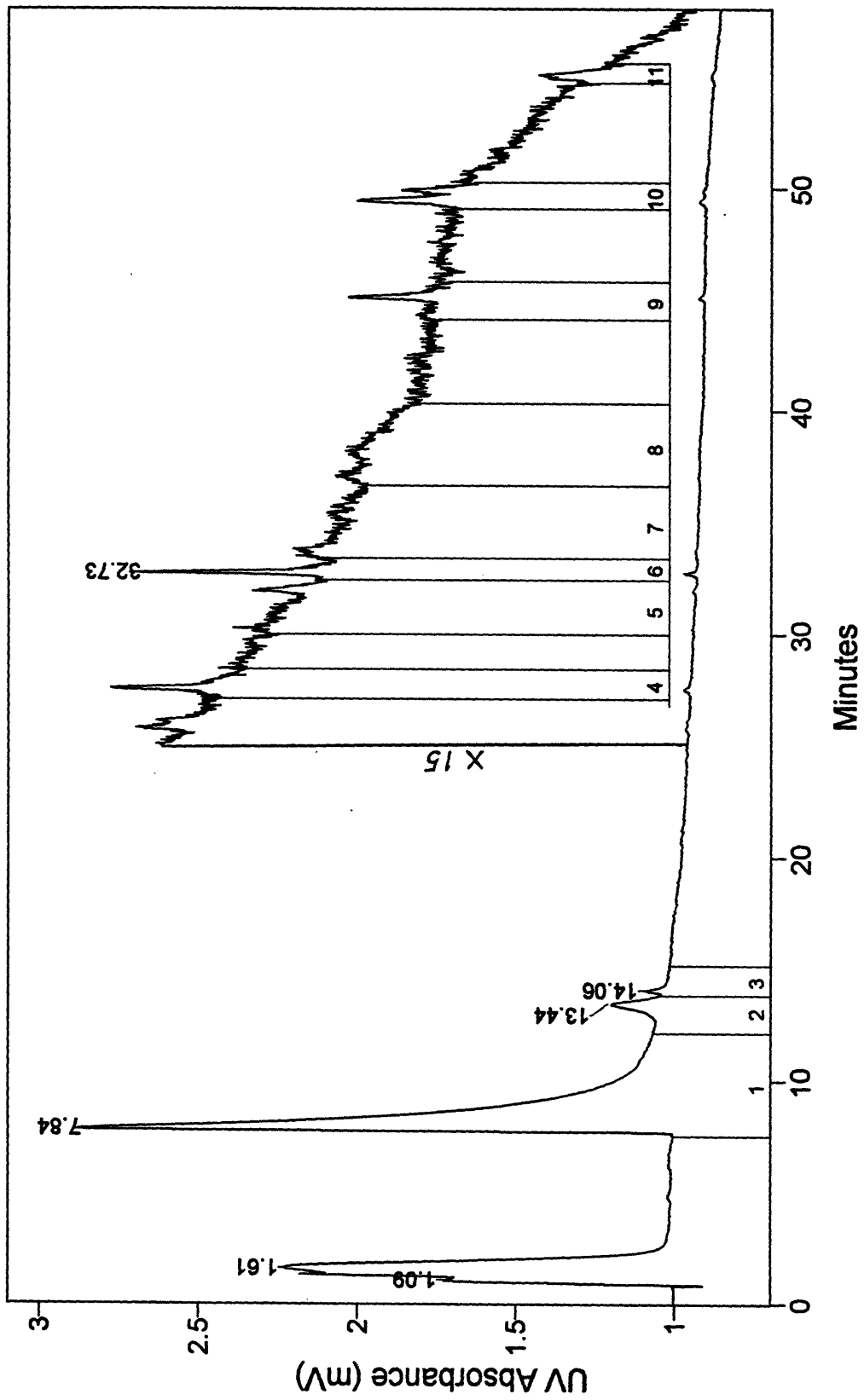


Figure III.14. HPLC trace of Glu-C digest of ASOB2. The fractions collected are marked.

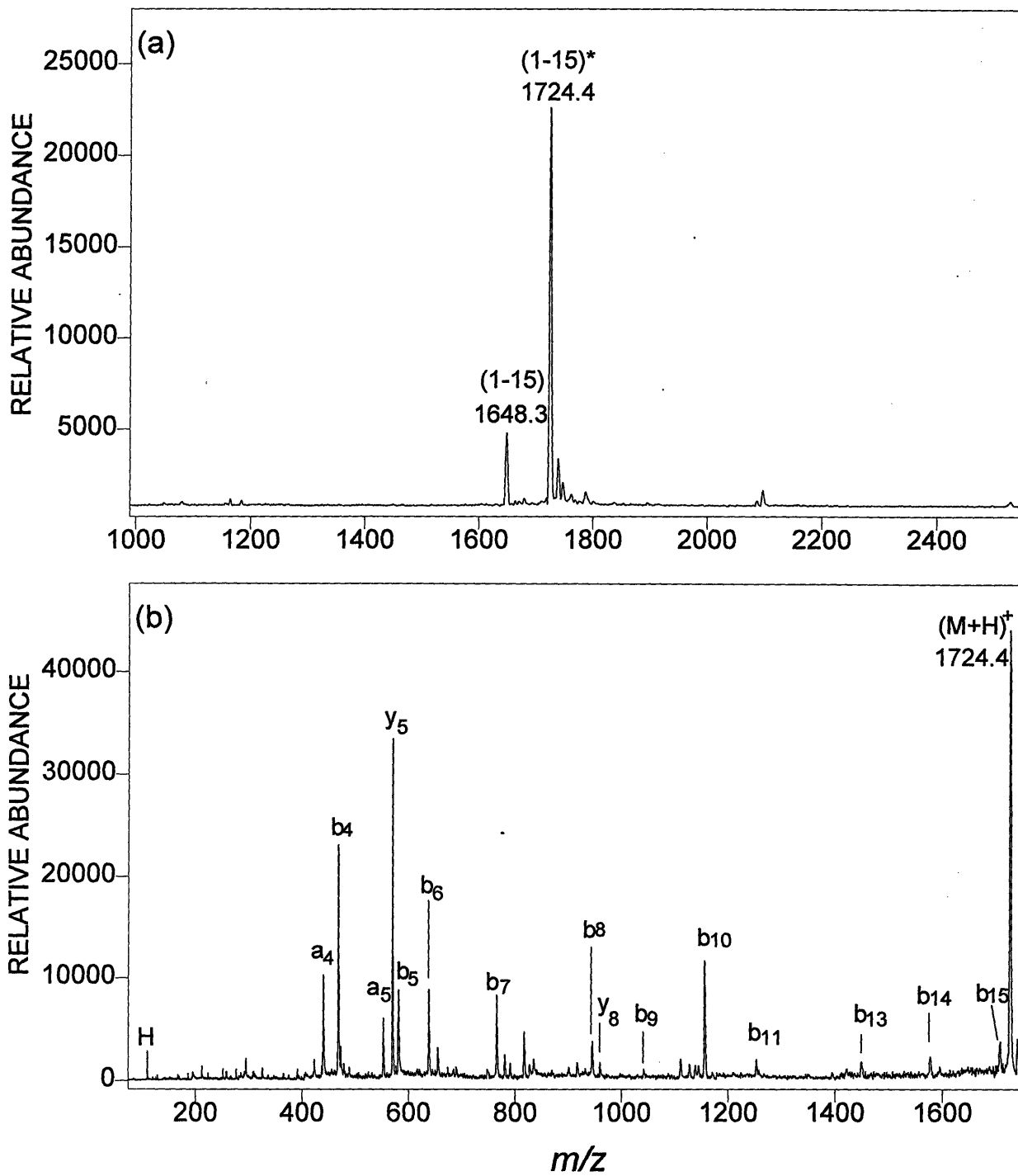


Figure III.15. (a) MALDI spectrum of Glu-C digest fraction 4. (b) PSD spectrum of  $(M+H)^+ = m/z$  1724.

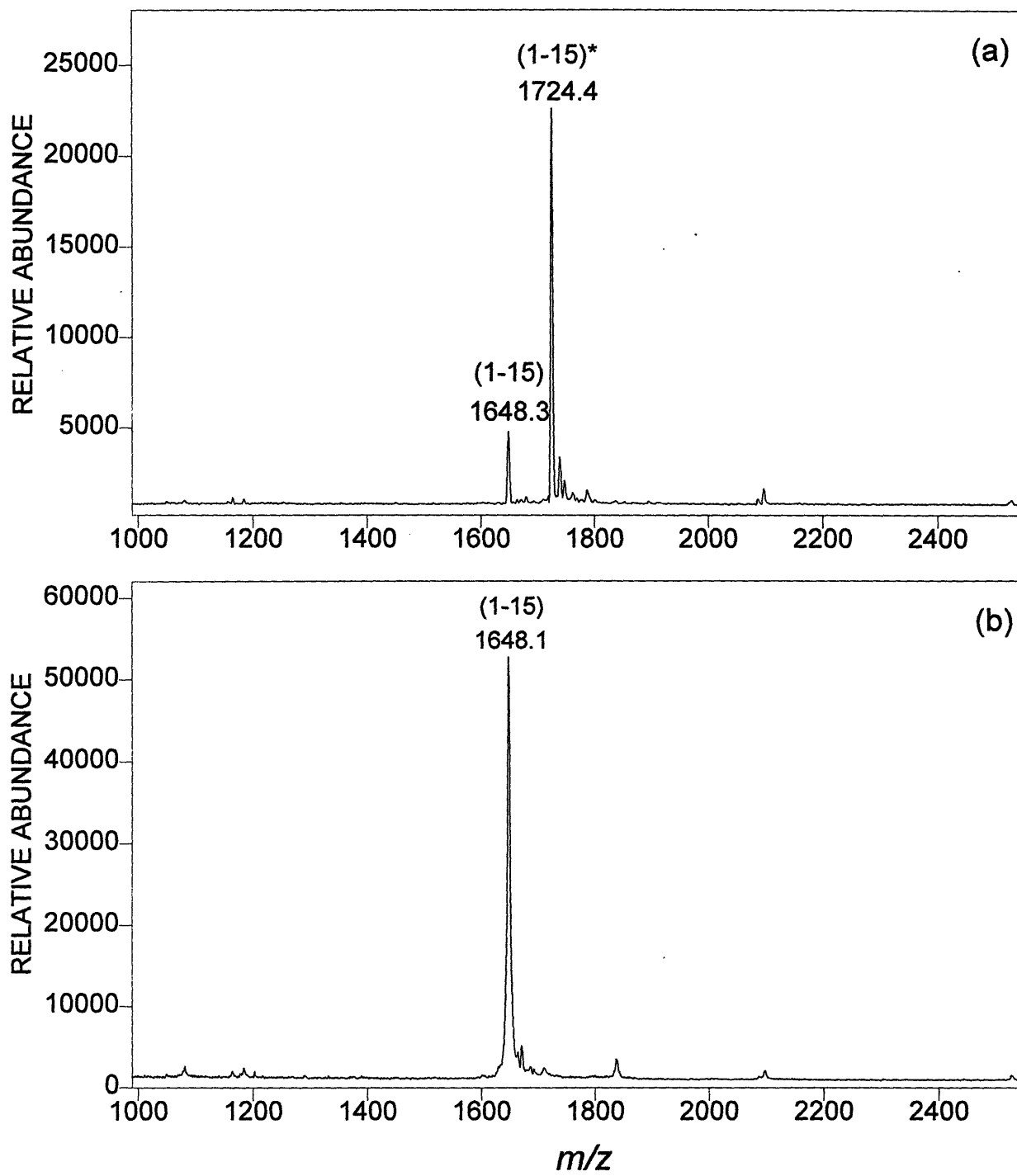


Figure III.16. MALDI spectra of fraction 4 of the Glu-C digest before (a) and after (b) DTT treatment.

It has been known that apolipoprotein D contains five cysteines among its 169 amino acids. The disulfide bonds of human plasma apoD have been identified to be Cys<sup>8</sup>-Cys<sup>114</sup> and Cys<sup>41</sup>-Cys<sup>165</sup> while Cys<sup>116</sup> is intramolecularly linked to Cys<sup>6</sup> of apoA-II (Yang et al., 1994). The predicted (M+H)<sup>+</sup> ion of PNGase F deglycosylated apoD would be 19,283.07 (calculated for both Asn<sup>45</sup> and Asn<sup>78</sup> converted to Asp and a free Cys<sup>116</sup>). It was found at about *m/z* 19,359, the center of a relatively broad peak which indicates that this protein is a mixture of apoD and its various reaction products discussed above, formed during isolation and purification on SDS-PAGE .

However, there must also have been some re-shuffling of the native disulfide bonds because a peptide with disulfide bonds other than those predicted was found. As shown in Figure III.10a, two Lys-C fragments, the strong signals at *m/z* 2834.3 and 2962.5 could be the Lys-C peptide 8-21 disulfide linked to peptide 157-167 and 156-167. In order to confirm these disulfide linkages, fraction 12 was refractionated by HPLC to further isolate the peptides. The isolated peptide of (M+H)<sup>+</sup> = *m/z* 2836 (Figure III.17a) was observed together with two other peaks at *m/z* 1236.1 and 1601.5 which may be fragments of peptide (M+H)<sup>+</sup> = *m/z* 2836 generated by cleavage of the disulfide bond in the mass spectrometer (Crimmins et al., 1995).

To chemically identify the disulfide linkage, the isolated peptide (M+H)<sup>+</sup> = *m/z* 2836 was treated with DTT (see experimental section) to reduce the disulfide bond. After this treatment, the MALDI spectrum (Figure III.17b) shows only the two peaks at *m/z* 1235.6 (calc. 1236.5 for peptide 157-167) and at *m/z* 1601.0 (calc. 1601.8 for peptide 8-21) which confirmed the disulfide linkage between Cys<sup>8</sup> and Cys<sup>165</sup>. One of the other fractions (refractionated fraction 12) contained the

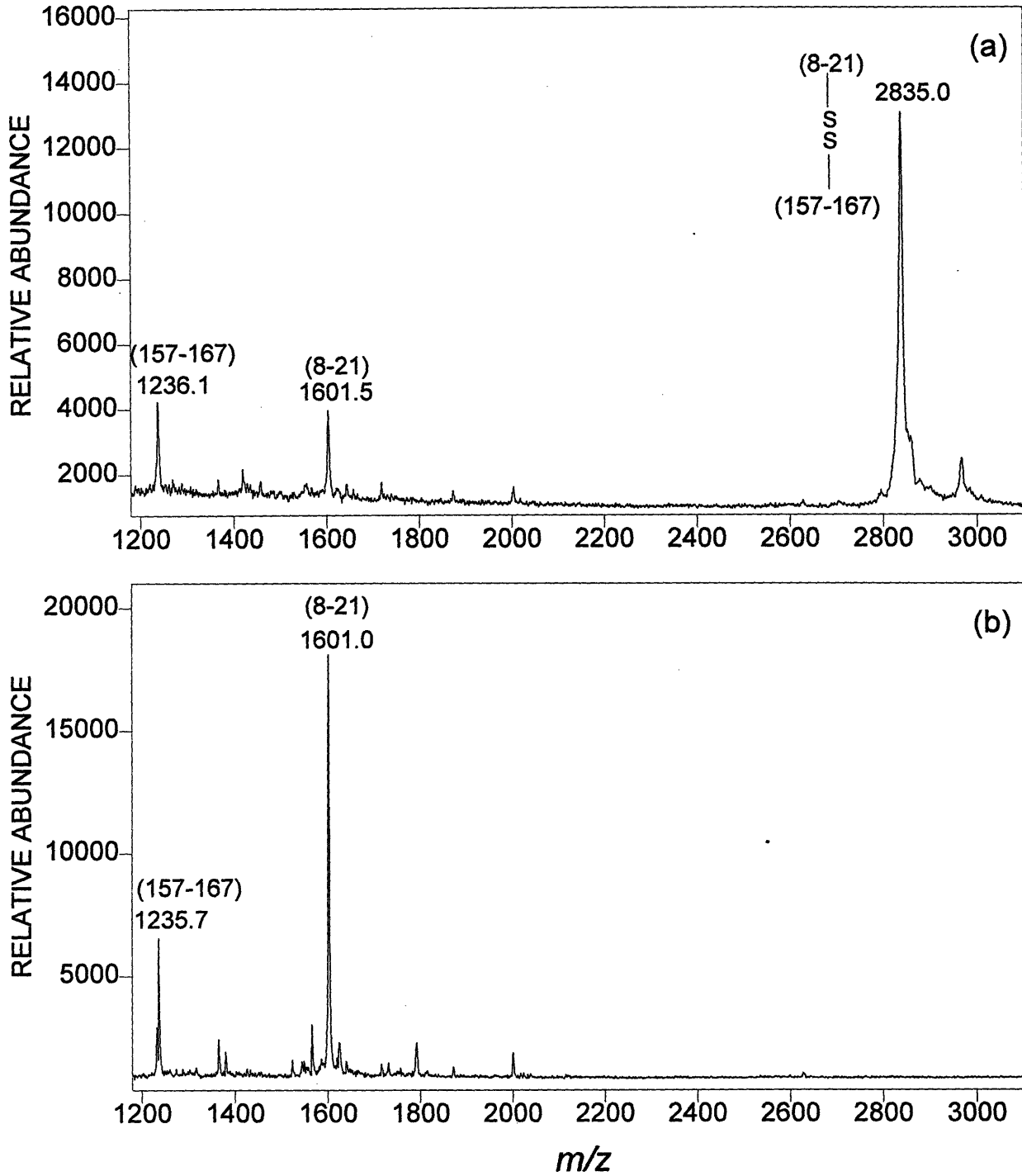


Figure III.17. MALDI spectra of peptide  $(M+H)^+ = m/z$  2835.2 isolated from fraction 12 of the Lys-C digest before (a) and after (b) DTT treatment.

peptide of  $(M+H)^+ = m/z$  2962.5 previously observed (see Figure III.10a) and assumed to be (8-21)-SS-(156-167). This was also confirmed by reduction to the two individual peptides 8-21 and 156-167 (data not shown). Therefore, some disulfide interchange must have taken place (perhaps initiated by the free -SH group of Csy<sup>116</sup>) during the isolation, purification and perhaps even proteolytic cleavage of ASOB2.

In order to determine if expression of apoD occurred in axillary tissue, an anti-sense oligonucleotide probe for apoD was synthesized. It was found that hybridization occurred exclusively in the apocrine gland. The reactivity was predominately in the cytoplasm; however, in some cells, a slight degree of nuclear staining was evident. There was no reactivity in any other cell type in the epidermis or dermis including keratinocytes, fibroblasts, eccrine glands, sebaceous glands, vascular epithelium, and perivascular mononuclear cells.



## D. Conclusion

To date, the role, if any, of apoD in the plasma lipoprotein system is not known. Several hydrophobic ligands for apoD have been suggested by both *in vitro* studies (progesterone type steroids) (Milne et al. 1993; López-Boado et al., 1994) and theoretical considerations (bilirubin) (Peitsch and Boguski, 1990). Our data identify, for the first time, a ligand (3M2H) that apoD carries in apocrine secretion. These results further demonstrate a remarkable similarity between human axillary secretions and odors and non-human mammalian odor sources used in chemical signalling. It is known that chemical signals, which modify reproductive endocrinology and behavior, are carried by proteins of the lipocalin family in the female golden hamster (aphrodisin) (Singer and Macrides, 1990), house mouse (major urinary proteins) (Robertson et al., 1993) and possibly the pig (pheromaxein) (Booth and White, 1988). As discussed above, human axillary extracts contain components that act as primer pheromones because the extracts can alter the length and timing of the female menstrual cycle, presumably by alteration of neuroendocrine pathways.

Both the three-dimensional structure (from x-ray crystallography) and amino acid sequence of other known lipocalins, e.g., retinol-binding protein and bilin-binding protein, suggest that the interactions between a member of the lipocalin protein superfamily and its ligand(s) are non-covalent (Boyles et al., 1990). Our results also show that although 3M2H was associated with the ASOB2 samples analyzed, none of the proteolytic peptides were found to be 3M2H adducts by mass spectrometry, indicating that the ligand is not covalently bound to the protein.

Although first isolated from mixtures of plasma lipoproteins, apoD is not a member of the protein family containing the other apolipoproteins that include apo A-I, A-II, A-IV, C-I, C-II, C-III, B-100 and E. The complete amino acid sequence of apoD was first inferred from its cDNA and then confirmed by partial sequencing using the Edman degradation (Drayna et al., 1986). The mature protein is cleaved from the 20 amino acid leader peptide between glycine and glutamine, which now becomes the N-terminal amino acid and converts to pyroglutamic acid by the elimination of  $\text{NH}_3$  thus blocking the N-terminus. The 169 amino acid sequence of apoD is homologous with members of the  $\alpha_{2\mu}$ -microglobulin or lipocalin superfamily of proteins (Drayna et al., 1986). In contrast to the other apolipoproteins that are synthesized primarily in the liver and intestines, apoD is produced in a variety of organs including the liver, intestines, pancreas, kidney, placenta, adrenals, spleen, and foetal brain (Drayna et al., 1986; Milne et al., 1993). The amount of apoD mRNA varies from organ to organ and is cell-type specific (Drayna et al., 1986; Milne et al., 1993; López-Boado et al., 1994; Boyles et al., 1990); in addition, the mRNA appears to be transcriptionally inducible because its amount increases several-fold in regenerating rat sciatic nerve (Boyles et al., 1990) and breast tumors under the influence of retinoic acid (López-Boado et al., 1994). Because apoD is expressed in a variety of tissues and under different circumstances, it has been suggested that the ligands associated with this protein may differ with tissue and circumstances, e.g., regenerating nerves and breast tumor cells in the presence of retinoic acid (Milne et al., 1993). However, when apoD is expressed in apocrine glands the *E*- and *Z*-isomers of 3M2H are the major ligands. As outlined above, it is difficult to know the amount of protein actually used in our experiments, but the best estimates indicate that there may be two moles of 3M2H for each mole of apoD in apocrine secretions.

Previous studies to localize expression of apoD mRNA in human tissue have not specifically included skin; however, it appears that the majority of cells expressing apoD *in vivo* are fibroblasts (Milne et al., 1993). In five separate axillary tissues examined, no expression of apoD was observed in fibroblasts. In agreement with these results, an *in vitro* study of a human diploid fibroblast line indicated that apoD mRNA occurred in quiescent and senescent fibroblasts but not in actively proliferating cells (Provost et al., 1991).

It should be noted that the glycosylation pattern of the apoD from axillary glands differs significantly from that recently found for apoD from serum (Schindler et al., 1995). There, the carbohydrate moieties are more complex and more extended. Furthermore, not only at the Asn<sup>45</sup> site but also at the Asn<sup>78</sup> site the attached glycan terminates in sialic acids. Clearly, the sites of expression (or, at least, of post-translational glycosylation) differ for plasma apoD and apocrine apoD.

#### IV. References

- Aberth, W., Straub, K.M., Burlingame, A.L. "Secondary Ion Mass Spectrometry with Cesium Ion Primary Beam and Liquid Target Matrix of Bioorganic Compounds" (1982) *Anal. Chem.* **54**, 2029-2034.
- Albone, E. S. (1984) in *Mammalian Semiochemistry: The Investigation of Chemical Signals Between Mammals* (Wiley, New York), pp. 2-5.
- Angel, A.S., Lindh, F., Nilsson, B. "Determination of Binding Positions in Oligosaccharides and Glycosphingolipids by Fast-Atom-Bombardment Mass Spectrometry" (1987) *Carbohydr. Res.* **168**, 15-31.
- Angel, A.S. and Nilsson, B. "Analysis of Glycoprotein Oligosaccharides by Fast Atom Bombardment Mass Spectrometry" (1990) *Biomed. Environ. Mass Spectrom.* **19**, 721-730.
- Barber, M., Bordoli, R.S., Sedgwick, R.D., Tyler, A.N. "Fast Atom Bombardment of Solids (F.A.B.): A New Ion Source for Mass Spectrometry" (1981) *J. Chem. Soc. Chem. Comm.* **7**, 325-327.
- Bartlet-Jones, M., Jeffery, W.A., Hansen, H.F., Pappin, D.J.C. "Peptide Ladder Sequencing by Mass Spectrometry Using a Novel, Volatile Degradation Reagent" (1994) *Rapid Commun. Mass Spectrom.* **8**, 737-742.
- Beavis, R.C. and Chait, B.T. "High-Accuracy Molecular Mass Determination of Proteins Using Matrix-Assisted Laser Desorption Mass Spectrometry" (1990) *Anal. Chem.* **62**, 1836-1840.
- Beavis, R.C. and Chait, B.T. "Cinnamic Acid Derivatives as Matrices for Ultraviolet Laser Desorption of Proteins" (1989) *Rapid Commun. Mass Spectrom.* **3**, 432-435.
- Beavis, R.C., Chaudhary, T., Chait, B.T. " $\alpha$ -Cyano-4-Hydroxycinnamic Acid as a Matrix for Matrix-Assisted Laser Desorption Mass Spectrometry" (1992) *Org. Mass Spectrom.* **27**, 156-158.
- Benninghoven, A. and Sichtermann, W.K. "Detection, Identification and Structural Investigation of Biologically Important Compounds by Secondary Ion Mass Spectrometry" (1978) *Anal. Chem.* **50**, 1180-1184.
- Biemann, K. "Mass Spectrometric Methods for Protein Sequencing" (1986) *Anal. Chem.* **58**, 1289A-1300A.
- Biemann, K. "Contribution of Mass Spectrometry to Peptide and Protein Structure" (1988) *Biomed.*

*Environ. Mass Spectrom.* **16**, 99-111.

Biemann, K. "Sequencing of Peptides by Tandem Mass Spectrometry and High-Energy Collision-Induced Dissociation" (1990a) *Meth. Enzym.* **193**, 455-480.

Biemann, K. "Appendix 5. Nomenclature for Peptide Fragment Ions (Positive Ions)" (1990b) *Meth. Enzym.* **193**, 886-887.

Biemann, K. "Mass Spectrometry of Peptides and Proteins" (1992) *Annu. Rev. Biochem.* **61**, 977-1010.

Billeci, T.M. and Stults, J.T. "Tryptic Mapping of Recombinant Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry" (1993) *Anal. Chem.* **65**, 1709-1716.

Blakley, C.R., Carmody, J.J., Vestal, M.L. "Liquid Chromatograph-Mass Spectrometry for Analysis of Nonvolatile Samples" (1980) *Anal. Chem.* **52**, 1636-1641.

Bock, K. and Pedersen, C. "Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Monosaccharides" (1983) *Adv. Carbohydr. Chem. Biochem.* **41**, 27-66.

Bock, K. and Thorgersen, H. "Nuclear Magnetic Resonance Spectroscopy in the Study of Mono- and Oligosaccharides" (1982) *Ann. Rep. NMR Spectrosc.* **13**, 1-57.

Booth, W. D. and White, C. A. "The Isolation, Purification and Some Properties of Pheromaxein, the Pheromonal Steroid-Binding Protein, in Porcine Submaxillary Glands and Saliva" (1988) *J. Endocrin.* **118**, 47-57.

Boyles, J. K., Notterpek, L. M., Anderson, L. J. "Accumulation of Apolipoproteins in the Regenerating and Remyelinating Mammalian Peripheral Nerve. Identification of Apolipoprotein D, Apolipoprotein A-IV, Apolipoprotein E, and Apolipoprotein A-I" (1990) *J. Biol. Chem.* **265**, 17805-17815.

Bradley, C.V., Williams, D.H., Hanley, M.R. "Peptide Sequencing Using the Combination of Edman Degradation, Carboxypeptidase Digestion and Fast Atom Bombardment Mass Spectrometry" (1982) *Biochem. Biophys. Res. Commun.* **104**, 1223-1230.

Brown, J.A., Segal, H.L., Maley, F., Trimble, R.B., Chu, F. "Effect of Deglycosylation of Yeast Invertase on its Uptake and Digestion in Rat Yolk Sacs" (1979) *J. Biol. Chem.* **254**, 3689-3691.

Brown, R.S. and Lennon, J.J. "Mass Resolution Improvement by Incorporation of Pulsed Ion Extraction in a Matrix-Assisted Laser Desorption/Ionization Linear Time-of-Flight Mass Spectrometer" (1995) *Anal. Chem.* **67**, 1998-2003.

Burlingame, A.L. Baillie, T.A. Derrick, P.J. "Mass Spectrometry" (1986) *Anal. Chem.* **58**, 165R-211R.

Byrd, J.C., Tarentino, A.L., Maley, F., Atkinson, P.H., Trimble, R.B. "Glycoprotein Synthesis in Yeast. Identification of Man<sub>8</sub>GlcNAc<sub>2</sub> as an Essential Intermediate in Oligosaccharide" (1982) *J. Biol. Chem.* **257**, 14657-14666.

Carr, S.A., Barr, J.R., Robert, G.D., Anumula, K.R., Taylor, P.B. "Identification of Attachment Sites and Structural Classes of Asparagine-Linked Carbohydrates in Glycoproteins" (1990) *Meth. Enzym.* **193**, 501-518.

Carr, S.A., Hemling, M.E., Bean, M.F., Roberts, G. D. "Integration of Mass Spectrometry in Analytical Biotechnology" (1991) *Anal. Chem.* **63**, 2802-2824.

Carr, S.A., Huddleston, M.J., Bean, M.F. "Selective Identification and Differentiation of N- and O-linked Oligosaccharides in Glycoproteins by Liquid Chromatography-Mass Spectrometry" (1993) *Protein Sci.* **2**, 183-196.

Carr, S.A., Roberts, G., Annan, R.S., Hemling, M.E., Hoyes, J. "Evaluation of PSD-MALDI and CID-MALDI for Structural Analysis" (1995) *Proc. 43<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics* p. 620.

Chait, B.T., Chaudhary, T., Field, F.H., "Mass spectrometric Characterization of Microscale enzyme Catalyzed Reaction of Surface-Bound Peptide and Proteins" (1986) *Methods in Protein Sequence Analysis* Walsh, K. A. Ed. pp 483-492.

Chait, B.T. and Kent, S.B.H. "Weighing Naked Proteins: Practical, High-Accuracy Mass Measurement of Peptides and Proteins" (1992) *Science*, **257**, 1885-1894.

Chait, B.T., Wang, R., Beavis, R.C., Kent, S.B.H. "Protein Ladder Sequencing" (1993) *Science* **262**, 89-92.

Chiarelli, M.P., Sharkey, A.G., Hercules, D.M. "Excited-State Proton Transfer in Laser Mass Spectrometry" (1993) *Anal. Chem.* **65**, 307-311.

Chu, F.K., Trimble, R.B., Maley, F. "The Effect of Carbohydrate Depletion on the Properties of Yeast External Invertase" (1978) *J. Biol. Chem.* **253**, 8691-8693.

Cohen, R.E., Zhang, W., Ballou, C.E. "Effects of Mannoprotein Mutations on *Saccharomyces cerevisiae* Core Oligosaccharide Structure" (1982) *J. Biol. Chem.* **257**, 5730-5737.

Colby, S.M., King, T.B., Reilly, J.P. "Improving the Resolution of Matrix-assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry by Exploiting the Correlation between Ion

- Position and Velocity" (1994) *Rapid Commun. Mass Spectrom.* **8**, 865-868.
- Collins, J. F. and Coulson, A. F. W. "Significance of Protein Sequence Similarities" (1990) *Meth. Enzymol.* **183**, 474-487.
- Cotter, R.J. (Ed.), "Time-of-Flight Mass Spectrometry" (1994) in *Time-of-Flight Mass Spectrometry, Acs Symposium Series 549*, American Chemical Society, Washington DC, pp 16-48.
- Crimmins, D.L., Saylor, M., Rush, J., Thoma, R.S. "Facile, *in Situ* Matrix-Assisted Laser Desorption Ionization-Mass Spectrometry Analysis and Assignment of Disulfide Pairings in Heteropeptide Molecules" (1995) *Anly. Chem.* **226**, 355-361.
- Cummings, R.D. "Use of Lectins in Analysis of Glycoconjugates" (1994) *Meth. Enzym.* **230**, 66-86.
- Cutler, W. B., Preti, G., Krieger, A. M., Huggins, G. R., Garcia, C. R., Lawley, H. J. "Human Axillary Secretions Influence Women's Menstrual Cycles: The Role of Donor Extract from Men" (1986) *Horm. & Behav.* **20**, 463-473.
- Darnell, J., Lodish, H., Baltimore, D. *Molecular Cell Biology. Sci. Am. Books, distributed by W.H. Freeman and company*, New York, (1990), p 80 and p 662.
- Dell, A., Reason, A. J., Khoo, K-H., Panico, M., McDowell, R.A., Morris, H.R. "Mass Spectrometry of Carbohydrate-Containing Biopolymers" (1994) *Meth. Enzym.* **230**, 108-132.
- Domon, B. and Costello, C.E. "A systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates" (1988) *Glycoconjugate J.* **5**, 397-409.
- Drayna, D., Fielding, C., McLean, J., Baer, B., Castro, G., Chen, E., Comstock, L., Henzel, W., Kohr, W., Rhee, L., Wion, K., Lawn, R. "Cloning and Expression of Human Apolipoprotein D cDNA" (1986) *J. Biol. Chem.* **261**, 16535-16539.
- Edman, P. and Begg, G.E. "A Protein Sequenator" (1967) *Eur. J. Biochem.* **1**, 80-91
- Ehring, H., Karas, M., Hillenkamp, F. "Role of Photoionization and Photochemistry in Ionization Processes of Organic Molecules and Relevance for Matrix-Assisted Laser Desorption Ionization Mass Spectrometry" (1992) *Org. Mass Spectrom.* **27**, 472-480.
- Esmon, B., Novick, P. Schekman, R. "Compartmentalized Assembly of Oligosaccharides on Exported Glycoproteins in Yeast" (1981) *Cell* **25**, 451-460.
- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C. M. "Electrospray Ionization for Mass Spectrometry of Large Biomolecules" (1989) *Science* **246**, 64-71.

- Fitzgerald, M.C., Parr, G.R., Smith, L.M. "Basic Matrices for the Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Proteins and Oligosaccharides" (1993) *Anal. Chem.* **65**, 3204-3211.
- Garrozzo, D., Impallomeni, G., Spina, E., Sturiale, L., Zanetti, F. "Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Polysaccharides" (1995) *Rapid Commun. Mass Spectrom.* **9**, 937-941.
- Gascon, S., Neumann, N. P., Lampen, J.O. "Comparative Study of the Properties of the Purified Internal and External Invertase from Yeast" (1968) *J. Biol. Chem.* **243**, 1573-1577.
- Gascon, S. and Lampen, J.D. "Purification of the Internal Invertase of Yeast" (1968) *J. Biol. Chem.* **243**, 1567-1572.
- Gillece-Castro, B.L. and Burlingame, A.L. "Oligosaccharide Characterization with High-Energy Collision-Induced Dissociation Mass Spectrometry" (1990) *Meth. Enzym.* **193**, 689-712.
- Gimon, M.E., Preston, L.M., Solouki, T., White, A.M., Russell, D.H. "Are Proton Transfer Reactions of Excited States Involved in UV Laser Desorption Ionization" (1992) *Org. Mass Spectrom.* **27**, 827-830.
- Gorin, P.A.J. "Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Polysaccharides" (1981) *Adv. Carbohydr. Chem. Biochem.* **38**, 13-104.
- Gower, D. B. and Ruparelia, B. A. "Olfaction in Humans with Special reference to odorous 16-androstenes: Their Occurrence, Perception and Possible Social, Psychological and Sexual Impact" (1993) *J. Endocrinol.* **137**, 167-187.
- Graham, C. A. and McGrew, W. C. "Menstrual Synchrony in Female Undergraduates Living on a Coeducational Campus" (1980) *Psychoneuroendocrinol.* **5**, 245-252.
- Guilhaus, M. "Principles and Instrumentation in Time-of-Flight Mass Spectrometry" (1995) *J. Mass Spectrom.* **30**, 1519-1532.
- Gusev, A.I., Wilkinson, W.R., Proctor, A., Hercules, D.M. "Improvement of Signal Reproducibility and Matrix/Comatrix Effects in MALDI Analysis" (1995) *Anal. Chem.* **67**, 1034-1041.
- Halbeek, H.V. "<sup>1</sup>H Nuclear Magnetic Resonance Spectroscopy of Carbohydrate Chains of Glycoproteins" (1994) *Meth. Enzym.* **230**, 132-280.
- Hall, S. C., Smith, D. M., Masiarz, F. R., Soo, V. W., Tran, H. M., Epstein, L. B. Burlingame, A. L. "Mass Spectrometric and Edman Sequencing of Lipocortin I Isolated by Two-Dimensional SD/PAGE of Human Melanoma Lysates" (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1927-1931.



- Hardy, M.R. and Townsend, R.R. "High-pH Anion-Exchange Chromatography of Glycoprotein-Derived Carbohydrates" (1994) *Meth. Enzym.* **230**, 208-225.
- Hemling, M.E., Roberts, G.D., Johnson, W., Carr, S.A. Covey, T.R. "Analysis of Proteins and Glycoproteins at the Picomole level by On-line Coupling of Microbore High-performance Liquid Chromatography with Flow Fast Atom Bombardment and Electrospray Mass Spectrometry: A Comparative Evaluation" (1990) *Biomed. Environ. Mass Spectrom.* **19**, 677-691.
- Henzel, W. J., Rodriguez, H., Singer, A. G., Stults, J. T., Macrides, F., Agosta, W. C., Niall, H. "The Primary Structure of Aphrodisin" (1988) *J. Biol. Chem.* **263**, 16682-16687.
- Hillenkamp, F., Karas, M., Beavis, R.C., Chait, B.T. "Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers" (1991) *Anal. Chem.* **63**, 1193A-1203A.
- Hillier, L., and Green, P. "OSP: A Computer Program for Choosing PCR and DNA Sequencing Primers" (1991) *PCR Meth. Appl.* **1**, 124-128.
- Hines, W., Peltier, J., Hsieh, F., Martin, S.A. "Peptide Fragment Mass Spectra: Comparison of PSD and CID" (1995b) *Proc. 43<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics* p. 387.
- Hubbard, S.C., and Ivatt, R.J. "Synthesis and Processing of Asparagine-Linked Oligosaccharides" (1981) *Ann. Rev. Biochem.* **50**, 555-583.
- Huberty, M.C., Vath, J.E., Yu, W., Martin, S.A. "Site-Specific Carbohydrate Identification in Recombinant Protein Using MALDI-TOF MS" (1993) *Anal. Chem.* **65**, 2791-2800.
- Hunt, D.F., Fates III, J.R., Shabanowitz, J., Winston, S., Hauer, C.R. "Protein Sequencing by Tandem Mass Spectrometry" (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6233-6237.
- Jackson, P. "High-Resolution Polyacrylamide Gel Electrophoresis of Fluorophore-Labeled Reducing Saccharides" (1994) *Meth. Enzym.* **230**, 250-265.
- Jacob, G.S. and Scudder, P. "Glycosidases in Structural Analysis" (1994) *Meth. Enzym.* **230**, 280-299.
- Jenning, K.R. and Mason, R.S. "Tandem Mass Spectrometry Utilizing Linked Scanning of Double Focussing Instrument" (1983) in McLafferty, F. M. Ed. *Tandem Mass Spectrometry* John Wiley and Sons, New York, pp 197-222.
- Johnson, R.S. Ph.D. Dissertation, Massachusetts Institute of Technology, 1988.
- Juhasz, P., Costello, C.E., Biemann, K. "Matrix-Assisted Laser Desorption Ionization Mass Spectrometry with 2-(4-Hydroxyphenylazo)benzoic Acid Matrix" (1993) *J. Am. Soc. Mass.*

*Spectrom.* **4**, 399-409.

Karas, M., Bachmann, D., Bahr, U., Hillenkamp, F. "Matrix-Assisted Ultraviolet Laser Desorption of Non-Volatile Compounds" (1987) *Int. J. Mass Spectrom. Ion Proc.* **78**, 53-68.

Karas, M., Bahr, U., Hillenkamp, F. "UV Laser Matrix Desorption/ionization Mass Spectrometry of Proteins in the 100,000 Dalton Range" (1989) *Int. J. Mass Spectrom. Ion Proc.* **92**, 231-242.

Karas, M., Bahr, U., Strupat, K., Hillenkamp, F., Tsarbopoulos, A., Pramanik, B.N. "Matrix Dependence of Metastable Fragmentation of Glycoproteins in MALDI TOF Mass Spectrometry" (1995) *Anal. Chem.* **67**, 675-679.

Karas, M., Bahr, U., Ingendoh, A., Nordhoff, E., Stahl, B., Strupat, K., Hillenkamp, F. "Principles and Applications of Matrix-Assisted UV-Laser Desorption/Ionization Mass Spectrometry" (1990) *Anal. Chim. Acta.* **241**, 175-185.

Karas, M. and Hillenkamp, F. "Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons" (1988) *Anal. Chem.* **60**, 2299-2301.

Karas, M., Nordhoff, E., Stahl, B., Strupat, K., Hillenkamp, F. "Matrix-Mixtures for a Superior Performance of Matrix-Assisted Laser Desorption Ionization Mass Spectrometry" (1992) *Proc. 40<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics*, 368-369.

Kaufmann, R., Kirsch, D., Spengler, B. "Sequencing of Peptides in a Time-of-Flight Mass Spectrometer: Evaluation of Postsource Decay following Matrix-Assisted Laser Desorption Ionization (MALDI)" (1994) *Int. J. Mass Spectrom. Ion Processes* **131**, 355-385.

Kobata, A. "Use of Endo- and Exoglycosidases for Structural Studies of Glycoconjugates" (1979) *Anal. Biochem.* **100**, 1-14.

Kochetkov, N.K. and Chizhov, O.S. "Mass Spectrometry of Carbohydrate Derivatives" (1966) In *Adv. Carbohydr. Chem.* Wolfrom, M.L. Ed.; Academic Press: New York, pp. 39-93.

Kochetkov, N.K. and Chizhov, O.S. "Mass Spectrometry of Carbohydrates" (1972) In *Meth. Carbohydr. Chem.* Whistler, R.L. and BeMiller, J.N. Eds.; Academic Press: New York and London, pp. 540-554.

Köster, C., Castoro, J.A., Wilkins, C.L. "High-Resolution Matrix-Assisted Laser Desorption/Ionization of Biomolecules by Fourier Transform Mass Spectrometry" (1992) *J. Am. Chem. Soc.* **114**, 7572-7574.

Kukuruzinska, M.A., Bergh, M.L.E., Jackson, B.J. "Protein Glycosylation in Yeast" (1987) *Ann. Rev. Biochem.* **56**, 915-944.

- Labows, J. B., McGinley, K. J., Kligman, A. M. "Perspectives on Axillary Odor" (1982) *J. Soc. Cosmet. Chem.* **34**, 193-202.
- Laine, R.A. "Glycoconjugates: Overview and Strategy" (1990) *Meth. Enzym.* **193**, 539-553.
- Lee, Y.C., and Scocca, J.R. "A Common Structural Unit in Asparagine-Oligosaccharides of Several Glycoproteins from Different Sources" (1972) *J. Biol. Chem.* **247**, 5753-5758.
- Lehle, L., Cohen, R.E., Ballou, C.E. "Carbohydrate Structure of Yeast Invertase" (1979) *J. Biol. Chem.* **254**, 12209-12218.
- Leyden, J. J., McGinley, K. J., Hoelzle, K., Labows, J. N., Kligman, A. M. "The Microbiology of the Human Axilla and Its Relationship to Axillary Odor" (1981) *J. Invest. Dermatol.* **77**, 413-416.
- Liao, P-C. and Allison, J. "Ionization Processes in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry: Matrix-Dependent Formation of  $[M+H]^+$  vs  $[M+Na]^+$  Ions of Small Peptides and Some Mechanistic Comments" (1995) *J. Mass Spectrom.* **30**, 408-423.
- Linhardt, R.J. "Capillary Electrophoresis of Oligosaccharides" (1994) *Meth. Enzym.* **230**, 265-280
- Linsley, K.B., Chan, S-Y., Chan, S., Reinhold, B.B., Lisi, P., Reinhold, V.N. "Applications of Electrospray Mass Spectrometry to Erythropoietin *N*- and *O*-linked Glycans" (1994) *Anal. Biochem.* **219**, 207-217.
- Loo, J.A., Udseth, H.R., Smith R.D. "Peptide and Protein Analysis by Electrospray Ionization-Mass Spectrometry and Capillary Electrophoresis-Mass Spectrometry" (1989) *Anal. Biochem.* **179**, 404-412.
- López-Boado, Y. S., Tolivia, J. López-Otin, C. "Apolipoprotein D Gene Induction by Retinoic Acid Is Concomitant with Growth Arrest and Cell Differentiation in Human Breast Cancer Cells" (1994) *J. Biol. Chem.* **269**, 26871-26878.
- Macfarlane, R.D. and Torgerson, D.F. "Californium-252 Plasma Desorption Mass Spectroscopy" (1976) *Science* **191**, 920-925.
- Maley, F., Trimble, R.B., Tarentino, A.L., Plummer, Jr. T.H. "Characterization of Glycoproteins and Their Associated Oligosaccharides through the Use of Endoglycosidases" (1989) *Anal. Biochem.* **180**, 195-204.
- Mallis, L.M. and Russell, D.H. "Fast Atom Bombardment-Tandem Mass Spectrometry Studies of Organo-Alkali Metal Ions of Small peptides" (1986) *Anal. Chem.* **58**, 1076-1080.
- Mann, M., Meng, C.K., Fenn J. B. "Interpreting Mass Spectra of Multiply Charged Ions" (1989)

*Anal. Chem.* **61**, 1702-1708.

McClintock, M. K. "Menstrual Synchrony and Suppression" (1971) *Nature* **229**, 244-245.

McClintock, M.K. "Estrous Synchrony and its Mediation by Airborne Chemical Communication" (1978) *Horm. & Behav.* **10**, 264-276.

Meng, C.K., Mann, M., Fenn, J.B. "Of Protons or Proteins" A Beam's a beam for a' that" (O.S. Burns)" (1988) *Z. Phys. D* **10**, 361-368.

Milne, R. W ., Rassart, E., Marcel, Y. L. "Molecular Biology of Apolipoprotein D" (1993) *Curr. Opin. Lipidol.* **4**, 100-106.

Mumma, R.O. and Vastola, F.J. "Analysis of Organic Salts by Laser Ionization Mass Spectrometry. Sulfonates, Sulfates and Thiosulfates" (1972) *Org. Mass Spectrom.* **6**, 1373-1376.

Nakajima, T., and Ballou, C.E. "Microheterogeneity of the Inner core Region of Yeast Manno-protein" (1975) *Biochem. Biophys. Res. Commun.* **66**, 870-879.

Neumann, N.P. and Lampen J. O. "Purification and Properties of Yeast Invertase" (1967) *Biochemistry* **6**, 468-475.

Ngoka, L.C., Gal, J-F., Lebrilla, C.B. "Effects of Cations and Charge Type on the Metastable Decay Rates of Oligosaccharides" (1994) *Anal. Chem.* **66**, 692-698.

Novick, p., Ferro, S., Schekman, R. "Order of Events in the Yeast Secretory Pathway" (1981) *Cell* **25**, 461-469.

Ohta, M., Hamako, J., Yamamoto, S., Hatta, H., Kim, M., Yamamoto, T., Oka, S., Mizuochi, T., Matsuura, F. "Structures of Asparagine-Linked Oligosaccharides from Hen Egg-Yolk Antibody (IgY). Occurrence of Unusual Glucosylated Oligo-Mannose Type Oligosaccharides in a Mature Glycoprotein" (1991) *Glycoconj. J.* **8**, 400-413.

Orlando, R., Bush, A.C., Fenselau, C. "Structural Analysis of Oligosaccharides by Tandem Mass Spectrometry: Collision Activation of Sodium Adduct ions" (1990) *Biomed. Environ. Mass Spectrom.* **19**, 747-754.

Overberg, A., Karas, M., Bahr, U., Kaufmann, R., Hillenkamp, F. "Matrix-assisted Infrared-laser (2.94  $\mu\text{m}$ ) Desorption/Ionization Mass Spectrometry of Large Biomolecules" (1990) *Rapid Commun. Mass Spectrom.* **4**, 293-296.

Overberg, A., Karas, M., Hillenkamp, F. "Matrix-assisted Laser Desorption of Large Biomolecules with a TEA-CO<sub>2</sub>-Laser" (1991) *Rapid Commun. Mass Spectrom.* **5**, 128-131.

- Papayannopoulos, I.A. Ph.D. Dissertation, Massachusetts Institute of Technology, 1987.
- Papayannopoulos, I.A., and Biemann, K. "A Computer Program (COMPOST) for Predicting Mass Spectrometric Information from Known Amino Acid Sequences" (1991) *J. Am. Soc. Mass Spectrom.* **2**, 174-177.
- Parekh, R.B. "Glycoform Analysis of Glycoproteins" (1994) *Meth. Enzym.* **230**, 340-348.
- Patel T.P., and Parekh, R.B. "Release of Oligosaccharides from Glycoproteins by Hydrazinolysis" (1994) *Meth. Enzym.* **230**, 57-66.
- Peitsch, M. C. and Boguski, M. S. "Is Apolipoprotein D A Mammalian Bilin-Binding Protein" (1990) *New Biol.* **2**, 197-206.
- Plummer, T.H., Jr., Elder, J.H., Alexander, S., Phelan, A.W., Tarentino, A.L. "Demonstration of Peptide:N-Glycosidase F Activity in Endo- $\beta$ -N-acetylglucosaminidase F Preparations" (1984) *J. Biol. Chem.* **259**, 10700-10704.
- Posthumus, M.A., Kistemaker, P.G., Meuzelaar, H.L.C., Ten Noever de Brauw, M.C. "Laser Desorption-Mass Spectrometry of Polar Nonvolatile Bio-Organic Molecules" (1978) *Anal. Chem.* **50**, 985-991.
- Preti, G., Cutler, W. B., Garcia, G. R., Huggins, G. R., Lawley, H. J. "Human Axillary Secretions Influence Women's Menstrual Cycles: The Role of Donor Extract of Females" (1986) *Horm. & Behav.* **20**, 474-482.
- Provost, P. R., Marcel, Y. L., Milne, R. W., Weech, P. K. Rassart, E. "Apolipoprotein D Transcription Occurs Specifically in Nonproliferating Quiescent and Senescent Fibroblast Cultures" (1991) *FEBS Lett.* **290**, 139-141.
- Quadagno, D. M., Shubeita, H. E., Deck, J., Francoeur, D. "Influence of Male Social Contacts Exercise and All-Female Living Conditions on the Manstrual cycle" (1981) *Psychoneuroendocrinol.* **6**, 239-244.
- Reddy, A.V., Johnson, R.S., Biemann, K., William, R.S., Ziegler, F.D., Trimble, R.B., Maley, F. "Characterization of the Glycosylation Sites in Yeast External Invertase; I. N-linked Oligosaccharide content of the Individual Sequons" (1988) *J. Biol. Chem.* **263**, 6978-6985.
- Reddy, A.V., MacColl, R., Maley, F. "Effect of Oligosaccharides and Chloride on the Oligomeric Structures of External, Internal, and Deglycosylated Invertase" (1990) *Biochemistry* **29**, 2482-2487.
- Reinhold, B.B., Chan, S-Y., Chan, S., Reinhold, V.N. "Profiling Glycosphingolipid Structural Detail: Periodate Oxidation, Electrospray, Collision-induced Dissociation and Tandem Mass

- Spectrometry" (1994) *Org. Mass Spectrom.* **29**, 736-746.
- Rennie, P. J., Gower, D. B., Holland, K. T. "In-vitro and in-vivo Studies of Human Axillary Odour and the Cutaneous Microflora" (1991) *Brit. J. Dermatol.* **124**, 596-602.
- Robertson, D. H. L., Benyon, R. L., Evershed, R.P. "Extraction, Characterization and Binding Analysis of Two Pheromonally Active Ligands Associated With the Major Urinary Protein of the House Mouse (*Mus musculus*)" (1993) *J. Chem. Ecol.* **19**, 1405-1416.
- Rouse, J.C., Yu, W., Martin, S.A. "A comparison of the Peptide Fragmentation Obtained from a Reflector Matrix-Assisted Laser Desorption-Ionization Time-of-Flight and a Tandem Four Sector Mass Spectrometer" (1995a) *J. Am. Soc. Mass Spectrom.* **6**, 822-835.
- Rouse, J.C., Vath, J.E., Scoble, H.A. "MALDI and PSD Characterization of Released Carbohydrates from Glycoproteins" (1995b) *Proc. 43<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics* p. 359.
- Russell, D.H., McGlohon, E.s., Mallis, L.M. "Fast-Atom Bombardment Tandem Mass Spectrometry Studies of Organo-Alkali-Metal Ions of Small Peptides. Competitive Interaction of Sodium with Basic Amino Acid Substituents" (1988) *Anal. Chem.* **60**, 1818-1824.
- Russell, M. J., Switz, G.M., Thompson, K. "Olfactory Influences on the Human Menstrual Cycle" (1980) *Pharmac. Biochem. Behav.* **13**, 737-738.
- Sato, K., Asada, R., Ishihara, M., Kunihiro, F., Kammei, Y., Kubota, E., Costello, C.E., Martin, S.A., Scoble, H., Biemann, K. "High-Performance Tandem Mass Spectrometry: Calibration and Performance of Linked Scans of a Four-Sector Instrument" (1987) *Anal. Chem.* **59**, 1652-1659
- Schär, M., Börnsen, K.O., Gassmann, E., "Fast Protein Sequence Determination with Matrix-Assisted Laser Desorption and Ionization Mass Spectrometry" (1991) *Rapid Commun. Mass Spectrom.* **5**, 319-326.
- Schindler, P.A., Settineri, C.A., Collet, X., Fielding, C.J., Burlingame, A.L. "Site-Specific Detection and Structural Characterization of the Glycosylation of Human Plasma Proteins Lecithin: Cholesterol Acyltransferase and Apolipoprotein D Using HPLC/Electrospray Mass Spectrometry and Sequential Glycosidase Digestion" (1995) *Protein Sci.* **4**, 791-803.
- Siegel, M.M., Hollander, I.J., Hamann, P.R., James, J.P., Hinman, L., Smith, B.J., Farnsworth, A.P.H., Phipps, A., King, D.J. "Matrix-Assisted UV-Laser Desorption/Ionization Mass Spectrometry Analysis of Monoclonal Antibodies for the Determination of Carbohydrate, Conjugated Chelator, and Conjugated Drug Content" (1991) *Anal. Chem.* **63**, 2470-2481.
- Singer, A. G. and Macrides, F. "Aphrodisin: Pheromone or Transducer?" (1990) *Chem. Senses* **15**, 199-203.

Smith, R.D., Loo, J.A., Edmonds, C.G, Barinaga, C.J., Udseth, H.R. "New Developments in Biochemical Mass Spectrometry: Electrospray Ionization" (1990) *Anal. Chem.* **62**, 882-899. ESI

Spellman, M.W. "Carbohydrate Characterization of Recombinant Glycoproteins of Pharmaceutical Interest" (1990) *Anal. Chem.* **62**, 1714-1722.

Spengler, B. and Cotter, R.J. "Ultraviolet Laser Desorption/Ionization Mass Spectrometry of Proteins above 100,000 Daltons by Pulsed Ion Extraction Time-of-Flight Analysis" (1990) *Anal. Chem.* **62**, 793-796.

Spengler, B., Hirsch, D., Kaufmann, R. "Peptide Sequencing by Matrix-assisted Laser-Desorption Mass Spectrometry" (1992a) *Rapid Commun. Mass Spectrom.* **6**, 105-108.

Spengler, B., Kirsch, D., Kaufmann, R. "Fundamental Aspects of Postsource Decay in Matrix-Assisted Laser Desorption Mass Spectrometry. 1. Residual Gas Effects" (1992b) *J. Phys. Chem.* **96**, 9678-9684.

Spielman, A.I., Zeng, X-N., Leyden, J.J., Preti, G. "Proteinaceous Precursors of Human Axillary Odor: Isolation of Two Novel Odor-Binding Proteins" (1995) *Experientia* **50**, 40-47.

Spiro, R.G. "Analysis of Sugars Found in Glycoproteins" (1966) *Meth. Enzym.* **8**, 3-53.

Spiro, R.G. "Study of the Carbohydrates of Glycoproteins" (1972) *Meth. Enzym.* **28B**, 3-43.

Stahl, B., Setup, M., Karas, M., Hillenkamp, F. "Analysis of Neutral Oligosaccharides by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry" (1991) *Anal. Chem.* **63**, 1463-1466.

Strupat, K., Karas, M., Hillenkamp, F. "2,5-Dihydroxybenzoic Acid: a New Matrix for Laser Desorption-Ionization Mass Spectrometry" (1991) *Int. J. Mass Spectrom. Ion. Processes.* **111**, 89-102.

Sweeley, C.C. and Nunez, H.A. "Structural Analysis of Glycoconjugates by Mass Spectrometry and Nuclear Magnetic Resonance Spectroscopy" (1985) *Ann. Rev. Biochem.* **54**, 765-801.

Takegawa, K., Miki, S., Iwahara, S. "Effect of Deglycosylation of Polymannose Chains on the Properties of Yeast External Invertase" (1990) *J. Ferment. Bioeng.* **70**, 131-133.

Takegawa, K., Yoshikana, M., Mishima, T., Nakoshi, M., Iwahara, S. "Determination of Glycosylation Sites Using a Protein Sequencer and Deglycosylation of Native Yeast Invertase by Endo- $\beta$ -N-Acetylglucosaminidase" (1991) *Biochem. Intern.* **25**, 585-592.

Tang, X., Ens, W., Standing, K.G., Westmore, J.B. "Daughter Ion Mass Spectra from Cationized Molecules of Small Oligopeptides in a Reflecting Time-of-Flight Mass Spectrometer" (1988) *Anal.*

*Chem.* **60**, 1791-1799.

Tarentino, A.L., Gómez, C.M., Plummer, Jr. T.H. "Deglycosylation of Asparagine-Linked Glycans by Peptide: *N*-glycosidase F" (1985) *Biochemistry* **24**, 4665-4671.

Tarentino, A.L., and Maley, F. "Purification and Properties of an Endo- $\beta$ -*N*-acetylglucosaminidase from *Streptomyces griseus*" (1974) *J. Biol. Chem.* **249**, 811-817.

Tarentino, A. L., Plummer, T. H., Jr., Maley, F. "The Release of Intact Oligosaccharides from Specific Glycoproteins by Endo- $\beta$ -*N*-acetylglucosaminidase H" (1974) *J. Biol. Chem.* **249**, 818-824.

Tarentino, A.L., Plummer, R.B., Plummer, T.H., Jr. "Enzymatic Approaches for Studying the Structure, Synthesis, and Processing of Glycoproteins" (1989) *Meth. cell biol.* **32**, 111-139.

Taussig, R. and Carlson, M. "Nucleotide Sequence of the Yeast *SUC2* Gene for Invertase" (1983) *Nucleic Acid Res.* **11**, 1943-1954.

Trimble, R.B., and Atkinson, P.H. "Structure of Yeast External Invertase  $\text{Man}_{8-14}\text{GlcNAc}$  Processing Intermediates by 500-Megahertz  $^1\text{H}$  NMR Spectroscopy" (1986) *J. Biol. Chem.* **261**, 9815-9824.

Trimble, R.B., and Atkinson, P.H. "Structural Heterogeneity in the  $\text{Man}_{8-13}\text{GlcNAc}$  Oligosaccharides from Log-phase *Saccharomyces* Yeast: a One- and Two-dimensional  $^1\text{H}$  NMR Spectroscopic Study" (1992) *Glycobiology* **2**, 57-75.

Trimble, R.B., and Maley, F. "Subunit Structure of External Invertase from *Saccharomyces cerevisiae*" (1977) *J. Biol. Chem.* **252**, 4409-4412.

Varki, A. "Biological Roles of Oligosaccharides: All of the Theories are Correct" (1993) *Glycobiology* **3**, 97-130.

Vastola, F.J., Mumma, R.O., Pirone, A.J. "Analysis of Organic Salts by Laser Ionization" (1970) *Org. Mass Spectrom.* **3**, 101-104.

Vastola, F.J. and Pirone, A.J. "Ionization of Organic Solids by Laser Irradiation" (1968) *Adv. mass spectrom.* **4**, 107-111.

Vertes, A., Gijbels, R., Levine, R.D. "Homogeneous Bottleneck Model of Matrix-Assisted Ultraviolet Laser Desorption of Large Molecules" (1990) *Rapid Commun. Mass Spectrom.* **4**, 228-233.

Vestal, M.L., Juhasz, P., Martin, S.A. "Delayed Extraction Matrix-assisted Laser Desorption Time-of-Flight Mass Spectrometry" (1995) *Rapid Commun. Mass Spectrom.* **9**, 1044-1050.



Wagner, R.M., Fraser, B.A. "Use of Immobilized Exopeptidases and Volatile Buffers for Analysis of Peptides by Fast Atom Bombardment Mass Spectrometry" (1987) *Biomed. Environ. Mass Spectrom.* **14**, 235-239.

Whittall, R.M. and Li, L. "High-Resolution Matrix-Assisted Laser Desorption/Ionization in a Linear Time-of-Flight Mass Spectrometer" (1995) *Anal. Chem.* **67**, 1950-1954.

Wiley, W.C. and McLaren, I.H. "Time-of-Flight Mass Spectrometry with Improved Resolution" (1955) *Rev. Sci. Instrum.* **26**, 1150-1157.

Yang, C-Y., Gu, Z-W., Blanco-Vaca, F., Gaskell, S.J., Yang, M., Massey, J.B., Gotto, A.M., Pownall, H.J. "Structure of Human Apolipoprotein D: Locations of the Intermolecular and Intramolecular Disulfide Links" (1994) *Biochemistry* **33**, 12451-12455.

Yu, W., Vath, J.E., Huberty, M.C., Martin, S.A. "Identification of the Facile Gas-Phase Cleavage of the Asp-Pro and Asp-Xxx Peptide Bonds in Matrix-Assisted Laser Desorption Time-of-Flight Mass Spectrometry" (1993) *Anal. Chem.* **65**, 3015-3023.

Zeng, C., Spielman, A.I., Vowels, B.R., Leyden, J.J., Biemann, K., Preti, G. "A Human Axillary Odorant is Carried by Apolipoprotein D" (1996) *Proc. Natl. Acad. Sci. U.S.A.* in press.

Zeng, X-N., Leyden, J.J., Brand, J.G., Spielman, A.I., McGinley, K., Preti, G. "An Investigation of Human Apocrine Gland Secretion for Axillary Odor Precursors" (1992) *J. Chem. Ecol.* **18**, 1039-1055.

Zeng, X-N., Leyden, J.J., Lawley, H.J., Sawano, K., Nohara, I., Preti, G. "Analysis of Characteristic Odors from Human Male Axillae" (1991) *J. Chem. Ecol.* **17**, 1469-1492.

Zeng, X-N., Leyden, J.J., Spielman, A.I., Preti, G. "Analysis of Characteristic Human Female Axillary Odors: Qualitative Comparison to Males" (1996) *J. Chem. Ecol.* in press.

Ziegler, F.D., Maley, F., Trimble, R.B. "Characterization of the Glycosylation Sites in Yeast External Invertase. II. Location of the Endo- $\beta$ -N-Acetylglucosaminidase H-Resistant Sequons" (1988) *J. Biol. Chem.* **263**, 6986-6992.

Ziegler, F.D. and Trimble, R.B. "Glycoprotein Biosynthesis in Yeast: Purification and Characterization of the Endoplasmic Reticulum Man<sub>9</sub> Processing  $\alpha$ -Mannosidase" (1991) *Glycobiology* **1**, 605-614.

## V. Appendices

### Appendix I.

#### Codes and Mass Values of Amino Acid Residues and the Corresponding Immonium Ions and Side Chain (R) Loss

<u>Amino Acid</u>	<u>Code</u>		<u>Residue mass</u>		<u>Immonium Ion Mass<sup>i</sup></u>	<u>R-group Loss</u>
	<u>3-letter</u>	<u>1-letter</u>	<u>Monoisotopic<sup>g</sup></u>	<u>Average<sup>h</sup></u>		
Glycine	Gly	G	57.0215	57.05	30.05	- 1
Alanine	Ala	A	71.0371	71.08	44.08	- 15
Serine	Ser	S	87.0320	87.08	60.08	- 31
Proline	Pro	P	97.0528	97.12	70.12	- 41
Valine	Val	V	99.0684	99.13	72.13	- 43
Threonine	Thr	T	101.0477	101.11	74.11	- 45
Cysteine	Cys	C	103.0092	103.14	76.14	- 47
Isoleucine	Ile	I	113.0841	113.16	86.16	- 57
Leucine	Leu	L	113.0841	113.16	86.16	- 57
Asparagine	Asn	N	114.0429	114.10	87.10	- 58
Aspartic acid	Asp	D	115.0270	115.09	88.09	- 59
Glutamine	Gln	Q	128.0586	128.13	101.13	- 72
Lysine	Lys	K	128.0950	128.17	101.17	- 72
Glutamic acid	Glu	E	129.0426	129.12	102.12	- 73
Methionine	Met	M	131.0405	131.20	104.20	- 75
Histidine	His	H	137.0589	137.14	110.14	- 81
Phenylalanine	Phe	F	147.0684	147.18	120.18	- 91
Arginine	Arg	R	156.1011	156.19	129.18	- 100
Tyrosine	Tyr	Y	163.0633	163.17	136.17	- 107
Tryptophan	Trp	W	186.0793	186.21	159.21	- 130

<sup>g</sup>Calculated using the light isotopic mass of <sup>1</sup>H 1.0078(3); <sup>12</sup>C 12.0000(0); <sup>14</sup>N 14.0030(7); <sup>16</sup>O 15.9949(1); and <sup>32</sup>S 31.9720(7). (numbers in parentheses are the next uncertain digit)

<sup>h</sup>Calculated using the average atomic weight value of <sup>1</sup>H 1.00(8); <sup>12</sup>C 12.01(1); <sup>14</sup>N 14.00(7); <sup>16</sup>O 15.99(9); AND <sup>32</sup>S 32.06(7).

<sup>i</sup>Only the average values are shown.

## Appendix II.

### Abbreviations and Masses of the Common Monosaccharide Residues in Glycoproteins

<u>Monosaccharide Residue</u>	<u>Abbreviation</u>	<u>Formular</u>	<u>Monoisotopic</u>	<u>Average</u>
<b>Deoxypentose</b> Deoxyribose	<b>DeoxyPent</b> dRib	$C_5H_{10}O_4$	<b>116.0473</b>	<b>116.12</b>
<b>Pentose</b> Arabinose Ribose Xylose	<b>Pent</b> Ara Rib Xyl	$C_5H_{10}O_5$	<b>132.0423</b>	<b>132.12</b>
<b>Deoxyhexose</b> Fucose	<b>DeoxyHex</b> Fuc	$C_6H_{10}O_5$	<b>146.0579</b>	<b>146.14</b>
<b>Hexose</b> Galactose Glucose Mannose	<b>Hex</b> Gal Glc Man	$C_6H_{12}O_6$	<b>162.0528</b>	<b>162.14</b>
<b>Hexuronic acid</b> Glucuronic acid	<b>HexA</b> GlcA	$C_6H_{10}O_7$	<b>176.0321</b>	<b>176.13</b>
<b>N-acetylhexosamine</b> N-acetylgalactosamine N-acetylglucosamine	<b>HexNAc</b> GalNAc GlcNAc	$C_8H_{15}NO_6$	<b>203.0794</b>	<b>203.19</b>
<b>Sialic acid</b> (N-acetylneuraminic acid)	<b>SA</b> (NeuAc)	$C_{11}H_{19}NO_9$	<b>291.0954</b>	<b>291.26</b>

## VI. Biographical note

Chenhui Zeng was born in Guangzhou (Canton) of the People's Republic of China on August 14, 1956, the only child of Professors Longmei Zeng and Jingyu Su. He was raised and completed high school education in Guangzhou. He subsequently worked in a maintenance workshop at the Guangzhou Sewing Machine Factory. After the "Cultural Revolution", he had the chance to enroll at Zhongshan University and graduated in July 1982 having majored in analytical chemistry. He then took his second job in Instrumentation Analysis and Research Center at Zhongshan University, and there he began his mass spectrometry career. In the spring of 1983, he completed the advanced mass spectrometry training program at Dalian Institute of Technology (Dalian, P. R. China) managed by the Chinese Education Bureau. Shortly after he returned to Zhongshan University he was married to Yiling Su-Zeng and they had their first daughter Sandy (Xinran) Zeng in 1984. In the summer of 1986, Chenhui Zeng obtained an appointment to work in the Mass Spectrometry Facility at Massachusetts Institute of Technology (Cambridge, MA. USA) under Professor Biemann and Dr. Catherine E. Costello. In 1989, he had his second daughter Angela (Weiran) Zeng right after the "Tian-An-Men Square Demonstration" movement in Beijing (P. R. China). The following year, he matriculated in the Department of Chemistry at Massachusetts Institute of Technology as a Doctoral Candidate and has worked there until the present time under the supervision of Professor Biemann.

## VII. Publications and Abstracts

### Publications:

Pearlstein, R. M.; Lock, C. J.; Faggiani, R.; Costello, C. E.; Zeng, C.; Jones, A. G.; and Davison, A. "Synthesis and Characterization of Technetium (V) Complexes with Amine, Alcoholate and Chloride Ligands" *Inorganic Chemistry*, **1988**, *27*, 2409-2413.

Chin, D.; Zeng, C.; Costello, C. E.; and Goldberg, I. H. "Sites in the Diyne-ene Bicyclic Core of Neocarzinostatin Chromophore Responsible for Hydrogen Abstraction from DNA" *Biochemistry*, **1988**, *27*, 8106-8114

Qian, F.; Du, S.; Shu, G.; Li, M.; Zeng, C.; Shu, Y. "Chemical Constituents of Hougujun (*Hericium erinaceus*). I. Isolation and Determination of Oleanolic Acid Saponins" *Zhongcaoyao (Chinese Traditional and Herbal Drugs)*, **1988**, *19*, 290-296.

Plaziak, A. S.; Zeng, C.; Costello, C. E.; Lis, S.; Elbanowski, M. "Mass Spectrometric Behavior of  $M(\text{acac})_3$  Complexes ( $M=\text{Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, and Lu}$ ;  $\text{acac}=\text{acetylacetonate}$  ligand) in Triethanolamine upon Negative Liquid Secondary Ion Mass Spectrometry" *Inorganic Chimica Acta*, **1991**, *184*, 229-233.

Zeng, C.; Spielman, A. I.; Vowels, B. R.; Leyden, J. J.; Biemann, K.; Preti, G. "A Human Axillary Odorant Carried by Apolipoprotein D" *Proc. Natl. Acad. Sci. U.S.A.* **1996**, in press.

Van der Donk, W. A.; Zeng, C.; Biemann, K.; Stubbe, J.; Hanlon, A.; Kyte, J. "Identification of the Active Site of the R1 Subunit of Ribonucleotide Reductase from *E. coli*: Characterization of Substrate Induced Polypeptide Cleavage by C225SR1" *Biochemistry*, **1996**, in press.

Abstracts and presentation:

Costello, C. E.; Domon, B.; Zeng, C.; Du, S.; Qian, F.; and Li, M.; "FABMS and MS/MS Analyses of Glycosidic Triterpenes from Chinese Medicines", *Proc. 35<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics*, 1987, 1.

Costello, C. E.; Zeng, C.; Vries, N. H.; O'Connell, L. A.; Pearlstein, R. M.; Davison, A.; and Jones, A. G.; "FAB MS and FAB MS/MS Investigation of Technetium and Rhenium Complexes", *Proc. 36<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics*, 1988, 143.

Costello, C. E.; Zeng, C.; and Singh, B. N.; "Structure Determination of Biosynthetic Intermediates in the Cell Membranes of Parasitic Protozoans", *Proc. 38<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics*, 1990, 1403.

Costello, C. E.; Juhasz, P.; Köchling, H. J.; Papayannopoulos, I. A.; Zeng, C.; and Biemann, K.; "Utilization of a Matrix-Assisted Laser Desorption TOF MS Modified for Biological Studies at Multiple Wavelengths", *Proc. 40<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics*, 1992, 1909.

Juhasz, P.; Papayannopoulos, I. A.; Zeng, C.; Papov, V.; and Biemann, K.; "The Utility of Matrix-Assisted Laser Desorption for the Direct Analysis of Enzymatic Digests of Proteins", *Proc. 40<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics*, 1992, 1913.

Donk, W.; Zeng, C.; Biemann, K. and Stubbe, J.; "Identification of the Active Site of Ribonucleotide Reductase from *E. coli*."; presented at *14<sup>th</sup> Enzyme Mechanisms Conference, Scottsdale, Arizona, 1995*.

Zeng, C. and Biemann, K.; "Determination of Glycosylation in External Invertase by MALDI-TOF MS", *Proc. 43<sup>th</sup> Am. Soc. Mass Spectrom. Allied Topics*, 1995, 1655.

## VIII. Acknowledgements

I wish to express my gratitude to Professor Klaus Biemann for his support and guidance in the research, described in this thesis. I also appreciate the opportunity I have had to work in his laboratory which has provided a stimulating and challenging research environment. The National Institute of Health is acknowledged for funding a research assistantship (grants RR0317 and GM005472 to Klaus Biemann).

I owe a great debt to Dr. Catherine E. Costello, who was Associate Director of the MIT-NIH Mass Spectrometry Facility through 1994, for her outstanding help and friendly support; and for helping to give me the chance to work and study at MIT. I especially thank Ms. Kathleen F. O'Brien, Prof. Biemann's secretary, for her kind concern for me and for my family; for her warm advice and for correcting my mistakes, for her kind words and for listening. I would also like to thank all the members (past and present) of the Biemann group for their support and friendship which I really care about.

Thanks are also due to my parents for their support and always wise advice, and to my two beautiful and lovely daughters for their love and the happy time when we are together. The last person but certainly not the least I would like to thank is my wife Yiling Su-Zeng, for her love and for taking care of the whole family. Without her I could not have accomplished this work.