

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Characterization of protein glycosylation by mass spectrometry

Permalink

<https://escholarship.org/uc/item/8md0r76f>

Author

Settineri, Christine A.

Publication Date

1995

Peer reviewed|Thesis/dissertation

Characterization of Protein Glycosylation by Mass Spectrometry

by Christine A. Settineri

UCSF LIBRARY

This thesis is partial fulfillment of the requirements for the degree

Doctor of Philosophy of Pharmaceutical Chemistry

University of California at San Francisco

December 1994

Acknowledgements

First and foremost, I would like to acknowledge my advisor, Professor Alma L. Burlingame, for providing me with the opportunity to be a part of the Mass Spectrometry Facility, and for enabling me to work on scientifically challenging problems while providing me with the support to successfully complete the work presented herein. I thank him for giving me the independence to complete this work in my own way, and for providing me with the guidance to always uphold to the highest standards in data interpretation and presentation. I also acknowledge my collaborator Dr. Frank R. Masiarz of Chiron Corporation, for providing me with many of the protein samples which are described in this thesis and for being a mentor to me in the field of protein chemistry and protein analysis. His enthusiasm for mass spectrometry as a tool for protein chemists before the days of electrospray and MALDI was always inspiring. I acknowledge my other collaborators, Professor Lewis T. Williams, Dr. Dah-Shun (Roxanne) Duan and Professor Christopher J. Fielding of the University of California, San Francisco, for also providing me with their precious protein samples.

I would also like to acknowledge numerous coworkers for providing support as friends as well as advice and ideas as peers throughout the last five years. I thank Dr. Katalin Medzihradzky for being a constant "sounding board" for my ideas and for always being there to help interpret CID spectra. I thank Dr. Peter Lipniunas for his advice relating to carbohydrate chemistry and analysis and for being there to listen when I was most frustrated. I thank Drs. Patrick Schindler and Andreas Ding for their help with my early days of running the Bio-Q and for hours of stimulating conversation at 4 o'clock coffee. I thank Professor R. Reid Townsend for donating his fetuin standards and glycosidases and for advice on glycoprotein analysis. Last but not least I thank David Maltby, Fred Walls and Professor Arnold Falick for their endless technical help with the mass spectrometers, for keeping the instruments running in the facility and for being technical mentors.

Finally, I gratefully acknowledge my fellow graduate students and close friends, Renée Williard, Theresa Gamble and Sherrie Hans for their friendship and support throughout the last five years, and for helping me to realize the responsibility that I have and the changes I can help institute as a woman scientist. Graduate school would not have been nearly as tolerable or fulfilling without their friendship. Finally, I thank my husband, Dr. David Lloyd, for his understanding, his support for me as a scientist, and for rescuing me from the Midwest and bringing me to California!

UCSF LIBRARY

Characterization of Protein Glycosylation by Mass Spectrometry

by
Christine A. Settineri

Abstract

The structural characterization of oligosaccharides on glycoproteins is essential to the understanding of their function. Protein-linked oligosaccharides are generally located at the molecular surface and are therefore optimally situated to modulate biochemical properties such as bioactivity, folding and immunogenicity, and act as determinants in molecular recognition events such as targeting of enzymes or the uptake of glycoproteins by receptors. Conventional strategies involving chemical degradation and nuclear magnetic resonance spectroscopy require nanomole to micromole quantities of isolated glycans, and chemical degradations may alter labile structural features in unforeseen ways. However, newer techniques such as electrospray ionization mass spectrometry can be used to analyze native glycoconjugates intact which have not been subjected to chemical degradations. Enzymatic and mass spectrometric methods have been developed to characterize glycoprotein oligosaccharides while still attached to the protein as glycopeptides, allowing the glycosylation site-specific characterization of picomole to femtomole quantities of glycopeptides. Using these methods, the sites and structures of the *O*-linked oligosaccharides of recombinant human platelet-derived growth factor and recombinant human p75 nerve growth factor receptor extracellular domain have been characterized. In addition, the sites and structures of the *N*-linked oligosaccharides of recombinant human platelet-derived growth factor receptor and recombinant human p75 nerve growth factor receptor extracellular domains and human lecithin:cholesterol acyl transferase have

UCSF LIBRARY

been determined. All analyses (with the exception of those for the platelet-derived growth factor which were done using liquid secondary ion mass spectrometry) were performed using electrospray ionization mass spectrometry, with pmole to fmole quantities of glycopeptides. The sensitivity of these techniques is further demonstrated by the additional site-specific characterization and sequencing of the *N*-linked oligosaccharides from human apolipoprotein D, a protein which co-purified with the human lecithin:cholesterol acyl transferase protein, present at a level approximately one fifth of the lecithin:cholesterol acyl transferase. Knowledge of the site-specific oligosaccharide structures was used to help understand the various known effects of the glycosylation on the specificity and biological activity of these proteins.


Alina

List of Abbreviations

| | |
|----------|---|
| ABEE | ethyl- <i>p</i> -amino benzoate |
| ABOE | octyl- <i>p</i> -amino benzoate |
| ACN | acetonitrile |
| Apo D | apolipoprotein D |
| AUFS | absorbance units full scale |
| BPI | base peak intensity |
| CHO | chinese hamster ovary |
| CI | chemical ionization |
| CID | collision induced dissociation |
| DCI | direct chemical ionization |
| DLI | direct liquid introduction |
| DMSO | dimethyl sulfoxide |
| EI | electron impact ionization |
| ESIMS | electrospray ionization mass spectrometry |
| FAB | fast atom bombardment |
| FAB-MS | fast atom bombardment mass spectrometry |
| Fuc | fructose |
| FD | field desorption |
| FID | flame ionization detection |
| Gal | galactose |
| GalNAc | N-acetylgalactosamine |
| GC | gas chromatography |
| GLC | gas liquid chromatography |
| Glc | glucose |
| GlcNAc | N-acetylglucosamine |
| Hex | hexose |
| HexNAc | N-acetylhexosamine |
| HPAE-PAD | high performance anion exchange chromatography with pulsed amperometric detection |
| HPLC | high performance liquid chromatography |
| hr | hour(s) |
| LC | liquid chromatography |
| LCAT | lecithin:cholesterol acyltransferase |
| LD | laser desorption |
| LSIMS | liquid secondary ion mass spectrometry |
| MALDI | matrix-assisted laser desorption ionization |
| Man | mannose |
| MeOH | methanol |
| min | minutes |
| MS/MS | tandem mass spectrometry |
| nm | nanometers |
| nmole | nanomole(s) |
| NeuAc | N-acetylneuraminic acid |
| NGF | nerve growth factor |
| NGFR | nerve growth factor receptor |
| NMR | nuclear magnetic resonance spectroscopy |
| PAGE | polyacrylamide gel electrophoresis |
| PD | plasma desorption |
| PDGF | platelet-derived growth factor |
| PDGFR | platelet-derived growth factor receptor |

| | |
|----------|--|
| pmole | picomole(s) |
| PNGase F | peptide-N4-(N-acetyl- β -glucosaminyl) asparagine amidase F |
| RCM | reduced and S-carboxymethylated |
| SIM | selected ion monitoring |
| sNGFR | recombinant nerve growth factor receptor extracellular domain |
| sPDGFR | recombinant platelet-derived growth factor receptor extracellular domain |
| SDS | sodium dodecyl sulphate |
| TIC | total ion current |
| TFA | trifluoroacetic acid |
| TOF | time of flight |
| UV | ultraviolet |

Table of Contents

| | |
|---|------|
| 1. Acknowledgements | iii |
| 2. Abstract | iv |
| 3. List of Abbreviations | vi |
| 4. Table of Contents | viii |
| 5. List of Tables | xi |
| 6. List of Figures | xiii |
| 7. List of Schemes | xvi |
| Chapter 1. Introduction | 1 |
| 1.1 Biological Relevance of Carbohydrates on Glycoproteins | 1 |
| 1.2 Historical Review of the Analysis of Carbohydrates by Mass Spectrometry | 4 |
| 1.3 Ionization Methods Used in Mass Spectrometry | 7 |
| 1.3.1 Electron Ionization | 7 |
| 1.3.2 Field Desorption | 9 |
| 1.3.3 Chemical Ionization | 10 |
| 1.3.4 Liquid Secondary Ion Mass Spectrometry and Fast Atom Bombardment | 11 |
| 1.3.5 Electrospray Ionization | 12 |
| 1.3.6 Matrix-Assisted Laser Desorption Ionization | 14 |
| 1.4 Instrumentation | 15 |
| 1.4.1 Sector | 15 |
| 1.4.2 Quadrupole | 17 |
| 1.4.3 Time of Flight | 19 |
| 1.4.4 Collision-Induced Dissociation | 20 |
| 1.5 Sample Inlet Systems | 21 |
| 1.5.1 Gas-Liquid Chromatography | 22 |
| 1.5.2 Static Probe | 22 |
| 1.5.3 Liquid Chromatography | 23 |
| 1.6 Conclusions | 24 |
| Chapter 2. Methods Development for the Analysis Protein Glycosylation | 26 |
| 2.1 Determination of Monosaccharide Composition | 29 |
| 2.1.1 TMS Ethers and Alditol Acetates | 29 |
| 2.1.2 HPAE-PAD of Hydrolyzed Glycans | 30 |
| 2.2 Asparagine-linked Glycosylation | 32 |
| 2.2.1 Determination of Structural Class | 32 |
| 2.2.1.1 Chemical Release | 33 |
| 2.2.1.2 Enzymatic Release | 33 |
| 2.2.1.3 Peracetylation | 37 |
| 2.2.1.4 Permethylation | 37 |
| 2.2.1.5 Reductive Amination | 41 |
| 2.2.2 Identification at Specific Attachment Sites | 44 |
| 2.2.2.1 Peptide and Glycopeptide Mapping | 45 |
| 2.2.2.2 Glycopeptide Identification | 48 |
| 2.2.3 Determination of the Sequence, Branching and Linkages | 51 |
| 2.2.3.1 Collision Induced Dissociation of Oligosaccharides | 52 |
| 2.2.3.2 Collision Induced Dissociation of Glycopeptides | 53 |
| 2.2.3.3 Exoglycosidase Sequencing of Glycopeptides | 56 |

| | |
|--|------------|
| 2.2.3.4 Methylation Linkage Analysis | 59 |
| 2.2.3.5 Reductive Cleavage Linkage Analysis | 62 |
| 2.2.3.6 Periodate Oxidation Linkage Analysis | 63 |
| 2.3 Serine/Threonine-linked Glycosylation | 64 |
| 2.3.1 Chemical Release | 65 |
| 2.3.2 Derivatization | 65 |
| 2.3.3 Identification at Specific Attachment Sites | 66 |
| 2.3.3.1 Glycopeptide Isolation by Jacalin Agarose | 67 |
| 2.3.3.2 Glycopeptide Identification | 68 |
| 2.3.4 Determination of Sequence, Branching and Linkages | 69 |
| 2.3.4.1 High Energy Collision Induced Dissociation of Glycopeptides | 70 |
| Chapter 3. Characterization of the <i>O</i>-linked Oligosaccharides of Recombinant Human Platelet-derived Growth Factor Expressed in Yeast | 72 |
| 3.1 Abstract | 72 |
| 3.2 Introduction | 73 |
| 3.3 Materials and Methods | 74 |
| 3.4 Results | 78 |
| 3.5 Discussion | 91 |
| Chapter 4. Characterization of the <i>O</i>-linked Oligosaccharides from Recombinant Human Nerve Growth Factor Receptor Extracellular Domain Expressed in Chinese Hamster Ovary Cells | 95 |
| 4.1 Abstract | 95 |
| 4.2 Introduction | 96 |
| 4.3 Materials and Methods | 99 |
| 4.4 Results | 102 |
| 4.5 Discussion | 111 |
| Chapter 5. Characterization and Sequencing of the <i>N</i>-linked Oligosaccharides from Recombinant Human Nerve Growth Factor Receptor Extracellular Domain | 114 |
| 5.1 Abstract | 114 |
| 5.2 Introduction | 115 |
| 5.3 Materials and Methods | 116 |
| 5.4 Results | 118 |
| 5.5 Discussion | 137 |
| Chapter 6. Characterization of the <i>N</i>-linked Oligosaccharides from Recombinant Human Platelet-derived Growth Factor β-Receptor Extracellular Domain Expressed in Chinese Hamster Ovary Cells | 139 |
| 6.1 Abstract | 139 |
| 6.2 Introduction | 140 |
| 6.3 Materials and Methods | 142 |
| 6.4 Results | 143 |
| 6.5 Discussion | 153 |

| | |
|--|------------|
| Chapter 7. Characterization of the <i>N</i>- and <i>O</i>-linked Oligosaccharides from Human Lecithin:Cholesterol Acyltransferase | 155 |
| 7.1 Abstract | 155 |
| 7.2 Introduction | 157 |
| 7.3 Materials and Methods | 158 |
| 7.4 Results | 160 |
| 7.5 Discussion | 177 |
| Chapter 8. Conclusions | 181 |
| References | 185 |

List of Tables

| | | |
|--|---|-----|
| Chapter 2. Methods Development for the Analysis of Protein Glycosylation | | |
| I. | Carbohydrate fragmentation nomenclature as proposed by Domon and Costello (Domon, 1988) and Dell and coworkers (Dell, 1987 #71) | 36 |
| Chapter 3. Characterization of the <i>O</i>-linked Oligosaccharides from Recombinant Human Platelet-derived Growth Factor Expressed in Yeast | | |
| I. | Edman sequence analysis results of peptide T ₁ -41 (glycosylated) and peptide T ₁ -45b (unglycosylated) | 84 |
| II. | Edman sequence analysis results of peptide T ₁ -60 | 85 |
| III. | CID fragmentation of rhPDGF-B-2 tryptic peptides T ₂ -56a and T ₂ -33 | 87 |
| IV. | Glycopeptides identified in rhPDGF-B | 90 |
| Chapter 4. Characterization of the <i>O</i>-linked Oligosaccharides from Recombinant Human Nerve Growth Factor Receptor Extracellular Domain Expressed in Chinese Hamster Ovary Cells | | |
| I. | Monosaccharide composition analysis of the carbohydrates from p75 nerve growth factor receptor. | 102 |
| Chapter 5 Characterization and sequencing of the <i>N</i>-linked Oligosaccharides from Recombinant Human Nerve Growth Factor Receptor Extracellular Domain Using HPLC/Electrospray Mass Spectrometry and Exoglycosidase digestion | | |
| I. | HPLC/Electrospray mass spectrometry data obtained from sequential glycosidase digestions of sNGFR <i>N</i> -linked glycopeptides | 128 |
| II. | Exoglycosidase reaction mixtures used for sequencing sNGFR <i>N</i> -linked oligosaccharides | 132 |
| Chapter 6. Characterization of the <i>N</i>-linked Oligosaccharides from Recombinant Human Platelet-derived Growth Factor β-Receptor Extracellular Domain Expressed in Chinese Hamster Ovary Cells | | |
| I. | Monosaccharide composition analysis results for sPDGFR | 145 |
| II. | Summary of HPLC/ESIMS data obtained on the glycopeptides identified from recombinant PDGF-R glu-C, glu-C + neuraminidase and glu-C + PNGase F digests | 149 |
| Chapter 7. Characterization of the <i>N</i>- and <i>O</i>-linked Oligosaccharides from Human Lecithin:Cholesterol Acyltransferase | | |
| I. | HPLC/Electrospray mass spectrometry data obtained from sequential glycosidase digestion of LCAT <i>N</i> -linked glycopeptides | 166 |

| | | |
|------|---|-----|
| II. | HPLC/Electrospray mass spectrometry data obtained from glycosidase digestion of LCAT <i>O</i> -linked glycopeptides | 166 |
| III. | HPLC/ESIMS data obtained from sequential glycosidase digestion of Apo D <i>N</i> -linked glycopeptides | 171 |

List of Figures

Chapter 2. Methods Development for the Analysis of Protein Glycosylation

| | | |
|------|---|----|
| 1. | Nomenclature for mass spectrometric fragmentation of oligosaccharides as proposed by Domon and Costello | 36 |
| 2A. | Positive ion FAB mass spectra of the reduced and peracetylated human milk oligosaccharide, lacto-N-decaose | 39 |
| 2B. | Positive ion FAB mass spectra of the reduced and permethylated human milk oligosaccharide, lacto-N-decaose | 40 |
| 3. | Negative ion LSIMS spectrum of Man ₆ GlcNAc-ABOE | 43 |
| 4. | Positive ion electrospray mass spectrum of a mixture of three glycopeptides containing amino acid residues (54-85) from bovine fetuin | 47 |
| 5A. | HPLC chromatograms of the tryptic digest of bovine fetuin using water/0.1% TFA as the A mobile phase and acetonitrile/0.08% TFA as the B mobile phase. | 48 |
| 5B. | HPLC chromatograms of the tryptic digest of bovine fetuin using water/0.1% formic acid as the A mobile phase and ethanol/propanol (5:2)/0.08% formic acid as the B mobile phase. | 48 |
| 6. | UV Absorbance, TIC and SIM (<i>m/z</i> 204 and <i>m/z</i> 292) chromatograms from an HPLC/ESIMS analysis of a bovine fetuin tryptic digest | 50 |
| 7. | Negative ion CID spectrum of the <i>p</i> -aminobenzoic acid octyl ester (ABOE) derivative Man ₉ GlcNAc from yeast mannan | 53 |
| 8. | Positive ion ESIMS/MS spectrum of the doubly charged precursor ion at <i>m/z</i> 1249.5 for a complex <i>N</i> -linked glycopeptide from ovomucoid | 54 |
| 9. | MALDI-TOF positive ion reflector mode mass spectrum of rhM-CSF glycopeptide NVFN ¹²² ETK | 56 |
| 10A. | Electrospray mass spectrum of the bovine fetuin tryptic glycopeptides containing Asn ⁸¹ | 58 |
| 10B. | Electrospray mass spectrum of the glycopeptide fraction from spectrum a) treated with neuraminidase from <i>Arthrobacter ureafaciens</i> . | 58 |
| 11. | Positive ion LSIMS/MS (CID) spectrum of an <i>O</i> -linked glycopeptide isolated from a glu-C digest of recombinant human platelet-derived growth factor (rhPDGF) expressed in yeast, containing residues (77-92). | 71 |

Chapter 3. Characterization of the *O*-linked Oligosaccharides from Recombinant Human Platelet-derived Growth Factor Expressed in Yeast

| | | |
|----|--|----|
| 1. | HPLC chromatogram of tryptic digest (T ₁) of recombinant human PDGF-B | 77 |
| 2. | HPLC chromatogram of tryptic digest (T ₂) of recombinant human PDGF-B | 78 |
| 3. | HPLC chromatogram of glu-C (V) of recombinant human PDGF-B | 79 |
| 4. | Tryptic (T ₁) peptide map of recombinant human PDGF-B | 80 |
| 5. | CID spectrum of tryptic peptide T ₁ -41 (KATVTLEDHLACK + 2 Man) from recombinant human PDGF-B | 82 |
| 6. | CID spectrum of glu-C peptide V-42 (IVRKKPIFKKATVTLE + 2 Man) from recombinant human PDGF-B | 83 |
| 7. | CID spectrum of tryptic peptide T ₁ -59 (SLGSLTIAEPAM(O)IAECK) from recombinant human PDGF-B | 86 |
| 8. | Covalent modifications identified for rhPDGF-B expressed in yeast. | 88 |

9. Laser desorption mass spectrum of recombinant human PDGF-B homodimer 89

Chapter 4. Characterization of the *O*-linked Oligosaccharides from Recombinant Human Nerve Growth Factor Receptor Extracellular Domain Expressed in Chinese Hamster Ovary Cells

1. Amino acid sequence of sNGFR, indicating the different domains and glycosylation sites 104
- 2A. Results of HPLC/ESI/CID/MS experiments for two digests of sNGFR HPLC/ESI/CID/MS experiment for the glu-C + neuraminidase digest with SIM at $m/z=204$ 106
- 2B. Results of HPLC/ESI/CID/MS experiments for two digests of sNGFR. HPLC/ESI/CID/MS experiment for the glu-C + neuraminidase + *O*-glycosidase digest with SIM at $m/z=204$ 106
- 3A. Electrospray mass spectra of a major asialo *O*-linked glycopeptide from sNGFR. ESI mass spectrum of the peak at 22 min (starred in Figure 2A). 107
- 3B. ESI mass spectrum of the peak eluting at 60.5 min in the sNGFR + glu-C + neuraminidase + *O*-glycosidase digest. 107
4. Electrospray mass spectra of the peptides resulting from the tryptic digest of the asialo *O*-linked glycopeptide from sNGFR 108
5. Results of HPLC/ESI/CID/MS with SIM experiment of the *O*-linked glycopeptide mixture isolated by jacalin-agarose 110

Chapter 5 Characterization and Sequencing of the *N*-linked Oligosaccharides from Recombinant Human Nerve Growth Factor Receptor Extracellular Domain

1. MALDI-TOF analysis of 100 pmole sNGFR-RCM using sinapinic acid as the matrix. 119
2. SDS-PAGE results for digestion of sNGFR-RCM with N-glycanase and neuraminidase + *O*-glycosidase 120
3. Peptide map of sNGFR-RCM based on the molecular weights obtained from HPLC/ESIMS analysis of the endoproteinase glu-C digest. 121
4. Reverse phase HPLC of ABOE-derivatized *N*-linked oligosaccharides released from sNGFR-RCM 122
5. Negative ion LSIMS of peak 3 from Figure 4 123
6. Negative ion LSIMS of peak 2 from Figure 4 124
7. Results of the HPLC/ESI/CID/MS experiment on the glu-C digest of sNGFR with SIM at $m/z=204$ and $m/z=657$. 126
8. HPLC/ESI mass spectrum of sNGFR glu-C *N*-linked glycopeptide mixture identified by HPLC/ESI/CID/MS with SIM 127
9. HPAE-PAD analysis of sNGFR *N*-linked oligosaccharides released by N-glycanase. 129
- 10A. Exoglycosidase digestion analysis of sNGFR glu-C *N*-linked glycopeptide mixture. HPLC/ESI mass spectrum of the neuraminidase-digested glycopeptide mixture shown in Figure 3. 135
- 10B. HPLC/ESI mass spectrum of the β -galactosidase-digested glycopeptide mixture shown in Figure 5A. 135
- 10C. HPLC/ESI mass spectrum of the β -galactosidase + β -N-acetylhexosaminidase-digested glycopeptide mixture shown in Figure 5B. 135

11. Sequence, branching and linkage data obtained on sNGFR oligosaccharides by exoglycosidase sequencing and HPLC/ESIMS. 137

Chapter 6. Characterization of the *N*-linked Oligosaccharides from Recombinant Human Platelet-derived Growth Factor β -Receptor Extracellular Domain Expressed in Chinese Hamster Ovary Cells

1. Peptide mass map of sPDGFR peptides identified by LSIMS and direct injection ESIMS 144
2. HPLC/ESIMS results of sPDGFR glu-C digest. UV absorbance, base peak intensity and SIM at $m/z=204$ chromatograms 147
3. SIM chromatograms at $m/z=147, 292, 366$ and 657 obtained from HPLC/ESI/CID/MS analysis of the sPDGFR glu-C digest 148
- 4A. HPLC/ESI mass spectrum for the group of glycopeptides representing residues (400-421) from the sPDGFR glu-C digest. 150
- 4B. HPLC/ESI mass spectrum for the group of glycopeptides representing residues (400-421) from the sPDGFR glu-C + neuraminidase digest 150
5. HPLC/ESIMS results of sPDGFR glu-C + neuraminidase digest. UV absorbance, base peak intensity and SIM at $m/z=204$ chromatograms 152

Chapter 7. Characterization of the *N*- and *O*-linked Oligosaccharides from Human Lecithin:Cholesterol Acyltransferase

1. Primary sequence of human LCAT 161
2. HPLC/ESIMS results of LCAT tryptic digest. UV absorbance, base peak intensity and SIM at $m/z=204$ chromatograms 162
3. SIM chromatograms resulting from HPLC/ESI/CID/MS analysis of LCAT with selected ion monitoring at $m/z=204, 292, 366$ and 657 164
4. Electrospray mass spectrum of the LCAT tryptic peptide mixture containing residues 16-39 167
- 5A. Electrospray mass spectrum of LCAT tryptic glycopeptides containing residues 257-276. 169
- 5B. Electrospray mass spectrum of the glycopeptide mixture from (a) after digestion with neuraminidase. 169
- 5C. Electrospray mass spectrum of the glycopeptide mixture from b) after digestion with β -galactosidase. 169
- 6A. Electrospray mass spectrum of LCAT C-terminal tryptic glycopeptides containing residues 400-416. 173
- 6B. Electrospray mass spectrum of the glycopeptide mixture in a) after sequential digestion with neuraminidase followed by Endo- α -N-acetylgalactosaminidase. 173

List of Schemes**Chapter 2. Methods Development for the Analysis of Protein Glycosylation**

1. Representative alditol acetate structures illustrating the primary fragments produced by EI and/or CI 31
2. Illustration of the sites of cleavage for Endo H and PNGase F in an Asn-linked oligosaccharide 33
3. Procedure for oligosaccharide derivatization by reductive amination with the chromophoric derivative *p*-aminobenzoic acid octyl ester 42
4. Illustration of fragmentation observed in EI and/or CI analysis of partially methylated alditol acetates 61
5. Illustration of branch specific ions which are formed by cleavage of the 1→6 linkage with charge retention at C-6 of the 3,6-disubstituted Man residue 64

UCSF LIBRARY

CHAPTER 1. INTRODUCTION

1.1 Biological Relevance of Carbohydrates on Glycoproteins

The oligosaccharides of glycoconjugates contain a remarkable structural complexity and diversity that is unmatched by other biomolecules such as amino acids and nucleic acids. Though there are now many examples of biological roles for specific structures on specific proteins, a single common theory has not been developed to explain this diversity. These examples include a purely structural role, an aid in conformation and stability of proteins, the provision or masking of target structures for antibodies, toxins and microorganisms, control of protein half-life, modulation of protein functions and the provision of ligands for specific binding events in the control of protein targeting or cell-cell interactions. However, in general, the only common features of the varied functions of oligosaccharides are that they either: 1) modulate of biochemical properties of proteins such as bioactivity, folding and immunogenicity [reviewed in (Cumming, 1991; Varki, 1993)], or 2) are determinants in molecular recognition events such as targeting of certain enzymes to lysosomes (Dahms *et al.*, 1989), the uptake of asialoglycoproteins by carbohydrate-specific receptors on hepatocytes and macrophages (Ashwell and Harford, 1982) and the recognition and adhesion of leukocytes by the selectin family of receptor proteins (Phillips *et al.*, 1990; Foxall *et al.*, 1992; Yuen *et al.*, 1992). This recognition and modulation of oligosaccharides provides most of the functional diversity required for the development and differentiation of complex organisms and interactions with their environment.

Protein-linked oligosaccharides are generally located at the molecular surface of proteins, and are therefore perfectly situated for modulating bioactivities and/or serving as recognition determinants. In addition, oligosaccharides are hydrodynamically very large, and therefore can have a tremendous influence on the overall size and shape of a glycoprotein (Montreuil, 1984; Mariuzza *et al.*, 1987; Jentoft, 1990). For example, a

tetraantennary glycan can cover 20-25 nm² of the protein surface, compared to 6-8 nm² for the antigen surface area in an antigen-antibody complex (Montreuil, 1984). Based on light scattering and NMR studies of mucins (Gerken *et al.*, 1989; Shogren *et al.*, 1989), a region of a protein which contains several *O*-linked oligosaccharides should resemble a semi-flexible rod with a length of approximately 2.5 Å per amino acid residue, resulting in a long extended structure, compared to a tightly folded globular protein. For example, an *O*-glycosylated peptide containing 28 amino acid residues would be 70 Å long (Jentoft, 1990). In contrast, the dimensions of hemoglobin, a protein containing 600 amino acids and a molecular weight of 64,000 Da, are only 70 Å x 55 Å x 55 Å (Jentoft, 1990).

There are now hundreds of examples of different effects that the absence or alteration of oligosaccharides have on glycoproteins. For example, in some instances, the "coating" of multiple oligosaccharides on a glycoprotein serves to protect the polypeptide chain from recognition by proteases (Bernard *et al.*, 1983; Gordon *et al.*, 1991) or antibodies (Lace *et al.*, 1990). In addition, there are many examples where oligosaccharides on glycoproteins are required for the initiation of correct polypeptide folding in the rough endoplasmic reticulum (ER), and in the subsequent maintenance of protein conformation and solubility (Oh-eda *et al.*, 1990; Riederer and Hinnen, 1991; Powell and Pain, 1992; Tift *et al.*, 1992). These proteins that are incorrectly glycosylated either fail to fold properly or fail to exit the ER, resulting in degradation. Glycosylation can also modulate the interaction of peptide and protein ligands with their receptors (Cebon *et al.*, 1990; Delorme *et al.*, 1992). For example, some cell surface growth factor receptors appear to acquire their binding functions in a glycosylation-dependent manner, thereby delaying acquisition of function until the receptor is close to the cell surface (Slieker *et al.*, 1986; Olson and Lane, 1989). This might limit unwanted early interactions with a growth factor synthesized in the same cell.

However, there are also many examples where the prevention or alteration of protein glycosylation has little or no apparent effect on the synthesis, folding or targeting; and/or

where the removal of oligosaccharides from a mature protein does not significantly alter its function nor sensitivity to proteolysis or immune recognition. (Kelker *et al.*, 1983; Breitfeld *et al.*, 1984; Hofmann *et al.*, 1992). Unfortunately, even complete knowledge of the structure, biosynthesis and expression of a particular structure does not necessarily indicate its specific function. Nevertheless, to completely understand the biological role of carbohydrates on glycoproteins, the array of structures present at each site on a glycoprotein must first be characterized so that appropriate experiments can be designed to assess their function and determine any structure-function relationships.

The number of detailed site-specific structures of oligosaccharides on glycoproteins continues to expand rapidly due to the significant improvements in enzymatic, chromatographic and mass spectrometric methods of oligosaccharide and glycopeptide analysis over the past five years (discussed in Chapters 1 and 2). The site-specific characterization of glycoprotein oligosaccharides using mass spectrometry combined with enzymatic digestion and liquid chromatography is the core of the work presented in this thesis. Several glycoproteins have been characterized using the methods described in Chapter 2: the *O*-linked oligosaccharides of recombinant human platelet-derived growth factor expressed in yeast (Chapter 3); the *O*-linked oligosaccharides of recombinant human nerve growth factor receptor extracellular domain expressed in chinese hamster ovary cells (Chapter 4); the *N*-linked oligosaccharides of recombinant human nerve growth factor receptor extracellular domain expressed in chinese hamster ovary cells (Chapter 5); the *N*-linked oligosaccharides of recombinant human platelet-derived growth factor receptor extracellular domain expressed in chinese hamster ovary cells (Chapter 6); and the *N*- and *O*-linked oligosaccharides of human plasma lecithin:cholesterol acyltransferase (Chapter 7).

1.2 Historical Review of the Analysis of Carbohydrates by Mass Spectrometry

The use of mass spectrometry for the structural characterization of carbohydrates began in the 1960's with the establishment of fragmentation behavior of peracetylated sugars (Biemann *et al.*, 1963) as well as of methylated sugars (Kochetkov and Chizov, 1965) and methyl glycosides (DeJongh and Biemann, 1963) using electron impact ionization mass spectrometry (EI-MS). During the same period, the jet separator provided the technical basis of a practical interface of a gas-liquid chromatograph (GLC) with a mass spectrometer by Ryhage (Ryhage, 1964). With the report of Hakamori's procedure for complete permethylation of sugars in the same year (Hakamori, 1964), oligosaccharide linkage analysis was developed by Lindberg and colleagues using GLC and EI-MS (Björndal *et al.*, 1967). Glycoconjugates had to be derivatized, e.g. methylated, in order to increase their volatility so that they could be ionized without being pyrolyzed by the high temperatures (200-300°C) required for volatilization. This technique remains a standard for carbohydrate sequencing and linkage analysis with only limited improvements in the methylation chemistry and the GC column technology, as well as increased sensitivity using chemical ionization mass spectrometry (CI-MS). In the years following, trimethylsilyl and acetyl derivatives of the labile hydrogens on sugars were also analyzed by EI-MS and GLC/MS, particularly oligosaccharides and glycolipids. In addition, developments in thermostable bonded stationary phases for GLC columns has extended the analysis of permethylated oligosaccharides by GC/MS to structures containing eight or nine sugar residues (Hansson and Karlsson, 1990). Electron ionization spectra of large permethylated glycosphingolipids were first reported in 1973 (Karlsson, 1973). By 1974, the major types of fragmentation processes occurring in oligosaccharides had been described and were reviewed by Lonngren and Svennson (Lonngren and Svennson, 1974) and again later by Radford and DeJongh (Radford and De Jongh, 1980). Sweeley and

UCSF LIBRARY

coworkers also used stable isotope labeling with GLC/MS to study the turnover rates of glycoconjugates *in vivo* (Sweeley *et al.*, 1966; Vance *et al.*, 1975).

The next major advance in carbohydrate analysis by mass spectrometry came in the late 1970's and early 1980's with the use of soft ionization techniques including field desorption mass spectrometry (FD-MS) (Linscheid *et al.*, 1981) and fast-atom bombardment and liquid secondary ion mass spectrometry (FAB-MS and LSIMS, which will be used interchangeably throughout this chapter) (Barber *et al.*, 1981; Aberth *et al.*, 1982). This technique often yielded molecular ions in addition to fragmentation which could be interpreted to obtain sequence information on glycoconjugates significantly larger than those which could be analyzed by EI and CI-MS. Peter-Katalinic and Egge demonstrated the ability to sequence 1-10 nanomole of native and derivatized glycosphingolipids comprising more than 25 sugar residues (Egge and Peter-Katalinic, 1987), and Reinhold and Carr, as well as Dell and coworkers analyzed permethylated and peracetylated oligosaccharides and glycopeptides up to 5,000-15,000 Da on similar quantities of material (Reinhold and Carr, 1983; Dell and Thomas-Oates, 1989). Sweeley and coworkers demonstrated the ability to analyze and sequence underivatized gangliosides by FAB-MS (Ohashi *et al.*, 1991). New developments in specific reducing-terminal derivatization strategies for oligosaccharides allowed for better separation of mixtures of picomole quantities of samples as well as directed fragmentation by LSIMS (Wang *et al.*, 1984; Webb *et al.*, 1988; Poulter *et al.*, 1991). After array detectors became commercially available in 1987 (Cottrell and Evans, 1987), the addition of these detectors to sector instruments increased the molecular ion detectability by another one to two orders of magnitude, yielding sensitivity in the picomole range (Poulter *et al.*, 1989; Gillece-Castro and Burlingame, 1990). Carr and coworkers first reported that additional fragmentation and therefore sequence information on oligosaccharides could be obtained by tandem mass spectrometry (MS/MS) using FAB coupled with collision-induced dissociation (CID) (Carr *et al.*, 1985), and later, Dell (Dell, 1987) as well as Domon and Costello (Domon and

Costello, 1988) proposed systematic nomenclatures for fragmentation of glycoconjugates identified by FAB-MS and FAB-MS/MS. Both are still used today, so both will be presented in this chapter.

The development of the ionization methods, electrospray ionization mass spectrometry (ESIMS) (Fenn *et al.*, 1989; Smith *et al.*, 1990) and matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOFMS) (Karas *et al.*, 1985; Hillenkamp *et al.*, 1991), has given mass spectrometrists the ability to measure the intact masses of picomole (and sometimes femtomole) quantities of glycoproteins, glycopeptides and other glycoconjugates. In addition, the ease of interfacing ESI mass spectrometers with liquid chromatographic systems enabled the first truly effective HPLC/MS interface, allowing rapid separation and analysis of mixtures of peptides and glycopeptides. Chemical molecular weights of several intact glycoproteins have now been determined using ESIMS (Smith *et al.*, 1990; Green and Oliver, 1991; Wada *et al.*, 1992; Wada *et al.*, 1992) and MALDI-TOFMS (Hillenkamp *et al.*, 1991; Mock *et al.*, 1991; Chait and Kent, 1992; Juhasz *et al.*, 1993). In addition, there are a number of groups who have successfully analyzed picomole and femtomole quantities of glycopeptides using ESIMS and ESIMS/MS (usually coupled to an HPLC) (Hemling *et al.*, 1990; Ling *et al.*, 1991; Carr *et al.*, 1993; Medzihradzky *et al.*, 1994) as well as MALDI-TOFMS (Huberty *et al.*, 1993; Sutton *et al.*, 1994). However, there are limited reports of the analysis of oligosaccharides (Mock *et al.*, 1991; Duffin *et al.*, 1992) or other glycoconjugates such as gangliosides (Juhasz *et al.*, 1993) by either of these techniques. The facile interfacing of microbore or capillary HPLC systems to ESI mass spectrometers (HPLC/ESIMS) has made this technique routine for the analysis of low picomole quantities of glycoproteins which have been previously digested with proteolytic enzymes (Ling *et al.*, 1991; Carr *et al.*, 1993; Huddleston *et al.*, 1993; Medzihradzky *et al.*, 1994; Settineri and Burlingame, 1994).

It is easy to see, from the continuous developments in ionization methods, derivatization strategies and effective coupling of HPLC and MS, why mass spectrometry has become an indispensable ingredient in the overall strategy for structural characterization of glycoconjugates. In the last 20 years, the tractable molecular size has increased dramatically while the amount of sample routinely required for analysis by various mass spectrometric methods has decreased by several orders of magnitude.

1.3 Ionization Methods Used in Mass Spectrometry

All mass spectrometers contain three basic components: the ion source, the mass analyzer and the detector. Ions are produced in the ion source by one of the many ionization methods which will be described in this section, the mass analyzer (described in the next section) separates the ions according to their mass-to-charge ratio (m/z), and then the ions strike the detector, are converted to electrons and produce an ion current proportional to the number of incident ions. A plot of the relative abundances of each ion versus its m/z value is a mass spectrum.

1.3.1 Electron Ionization

In the ion source of electron ionization (EI) mass spectrometers, a stream of electrons is emitted from a resistively heated filament, accelerated through a potential (usually 70 V) and introduced into the ionization region of the mass spectrometer (i.e. the ion source) where a small fraction of the electrons interact with gaseous sample molecules. The sample to be ionized must be in the gas phase at a pressure of less than 10^{-4} torr in the ion source, because higher pressures will lead to undesired ion/molecule reactions between the primary

ions and neutral molecules present. Both positive and negative ions are formed in this high energy deposition process, but since the probability of electron capture is more than 100 times less than the probability of electron removal, a preponderance of positive ions are formed. These positive ions (molecular ions) are in the form of radical cations ($M^{+\cdot}$). These radical-cations have excess internal energy due to their change in electronic state upon becoming radical-cations, and therefore unimolecular dissociation reactions, such as bond cleavages and rearrangements occur and molecular ions are sometimes not seen.

Because electron ionization is limited to compounds which are thermally stable in the gas phase, most biologically important molecules, including glycoconjugates, which are polar, labile and relatively non-volatile, must be derivatized chemically in order to be vaporized and then ionized by this technique without thermal decomposition. Therefore, derivatives such as trimethylsilyl, methyl and acetyl were developed and used extensively in the 1960s and 1970s (Biemann *et al.*, 1963; Chizhov and Kochetkov, 1972). At the same time, gas-liquid chromatographs were interfaced to EI mass spectrometers and GLC/MS quickly became the ideal technique for analysis of small glycoconjugates by EI-MS. Larger samples (larger than hexasaccharides) were introduced using a probe which is resistively heated at the tip when inserted into the mass spectrometer, because they were difficult to chromatograph by GLC due to thermal decomposition and liquid-phase bleeding problems at the high temperatures required for these larger molecules. However, recent developments in GLC thermostable bonded stationary phases allows analysis of relatively large permethylated oligosaccharides comprising up to 11 sugar residues (Karlsson *et al.*, 1987; Hansson *et al.*, 1989). Methylation linkage analysis, using fragmentation rules determined for EI-MS by Kochetkov (Kochetkov and Chizov, 1965), DeJongh (DeJongh and Biemann, 1963) and Biemann (Biemann *et al.*, 1963), was developed during this period using GLC/MS by Hellerqvist and colleagues (Hellerqvist *et al.*, 1968; Hellerqvist and Sweetman, 1990) using Hakamori's permethylation procedure (Hakamori, 1964). Permethylated glycolipids, glycosphingolipids, and oligosaccharides were all sequenced

from their EI spectra during the 1970's and early 1980's (Karlsson, 1973; Karlsson, 1978; Egge *et al.*, 1982; Jardine *et al.*, 1984). Since 100 nanomoles of material was often required for analysis by EI, especially for larger glycoconjugates, EI analysis gave way to the generally more sensitive softer ionization techniques in the 1980's.

1.3.2 Field Desorption

Field desorption mass spectrometry (FD-MS) was the first commercially available method for the ionization of nonvolatile molecules in the condensed state (Beckey, 1970). In FD-MS, a sample is applied directly to a small emitter wire in solution. This emitter wire is then placed in a strong electric field gradient, causing spontaneous ionization (believed to be by electron tunneling) to the emitter. The resulting ions have relatively little internal energy by comparison with molecules ionized by EI. Therefore, this ionization method has worked well for the study of labile species whose molecular ions decompose rapidly when excess internal energy is present (as in EI) (Beckey, 1977) and samples do not require prior chemical modification to produce FD spectra. In contrast to other methods of ionization, the field-desorbed ions are rapidly accelerated away from the site of ionization, due to the very strong repulsive force exerted on the positive ions by the high electric field gradient used for ionization. Field desorption mass spectra are characterized by the formation of even-electron ions (MH^+ or MNa^+) from nonvolatile samples, with minimum fragmentation. For example, FD ionization of a 3-O-methylmannose polysaccharide with the structure $Man_1MeMan_{13}-OCH_3$ was detected as the MNa^+ ion at m/z 2506 (Linscheid *et al.*, 1981). Unfortunately, the inherent experimental difficulties associated with the use of the thin wires in high electric fields have limited the routine use of FD-MS to just a few laboratories, and the development of other desorption ionization methods such as fast atom

bombardment mass spectrometry (see section 1.3.4) replaced the need to use FD-MS in many cases.

1.3.3 Chemical Ionization

Chemical ionization mass spectrometry (CI-MS) was another of the first so-called "soft ionization" techniques developed which had certain advantages for the analysis of glycoconjugates and other polar, nonvolatile compounds. Soft ionization refers to the fact that considerably less energy is transferred to the sample being ionized by CI than with EI (depending on the gas used in CI). During CI, the gaseous sample molecules are protonated by more acidic reagent gas ions such as NH_4^+ or C_2H_5^+ , forming even-electron positive and negative pseudo-molecular ions. This technique deposits less internal energy into the molecules, and therefore can result in less fragmentation and more abundant molecular ions than in EI. The use of GLC/CI-MS for the analysis of partially methylated alditol acetates of sugars added tenfold to the sensitivity over using EI-MS (Laine, 1981). Although there is less fragmentation in CI, the pathways of fragmentation that do occur are related to those observed in EI spectra. As is the case in EI, CI requires volatilization of the sample prior to ionization, so chemical derivatization is still required for analysis.

Direct chemical ionization (DCI) mass spectrometry involves ionization of compounds from the solid state. During DCI, the solid sample is placed directly within the source containing the reagent gas ions on a surface which may or may not be heated. A number of complex carbohydrates, gangliosides and glycosphingolipids were characterized using DCI in the early 1980s (Ariga *et al.*, 1980; Ariga *et al.*, 1982; Reinhold and Carr, 1983; Carr and Reinhold, 1984).

1.3.4 Liquid Secondary Ion Mass Spectrometry and Fast Atom Bombardment

Liquid matrix sputtering methods, developed in the early 1980s are also soft ionization techniques. The sample to be analyzed is first dissolved in a viscous liquid matrix such as glycerol or thioglycerol, loaded on a stainless steel probe tip and bombarded with a beam of Cs^+ ions [LSIMS (Aberth *et al.*, 1982)] or Xe^+ atoms [FAB (Barber *et al.*, 1981)] having kinetic energies of 8-30 keV. The translational kinetic energy of the beam is transferred to the surface molecules which are then sputtered out of the liquid into the high vacuum of the ion source. Ionization takes place in an acid-base like process in the top surface layer of the matrix droplet by a desorption process, so relative surface activities of analytes in mixtures can cause differential ionization, or suppression (Naylor *et al.*, 1986; Ligon, 1990). An alternative to analyzing the sample as a droplet on the surface of a probe (static-FAB or static-LSIMS), which can help to decrease the suppression effects due to differential surface activities, is continuous-flow FAB (Caprioli and Moore, 1990). This technique uses a sample introduction probe which provides a continuous flow of liquid (1-20 $\mu\text{l}/\text{min}$) onto the probe tip where the atom or ion bombardment takes place. This allows the use of much smaller percentages (2-5% rather than 25-50%) of the viscous carriers like glycerol so that the matrix chemical or background noise is dramatically decreased, yielding overall greater signal to chemical noise and a lower limit of analyte ion detection. This method also allows for the use of volatile solvents such as water, methanol and acetonitrile, and more readily permits coupling to HPLC. Although there are significant advantages to using continuous-flow FAB, it is still not widely used because it is technically challenging and successful application depends heavily on probe and probe tip design as well as interface setup.

Since FAB and LSIMS are soft ionization techniques, mass spectra often contain prominent molecular ions (molecular ions with a proton removed or attached), and few or

no fragments are seen. In the positive ion mode, addition of a proton to the molecular ion forms $(M+H)^+$ ions, while in negative ion mode, loss of a hydrogen from the molecular ion forms the $(M-H)^-$ ion. The important advantage to this technique is that one is able to analyze relatively large, often underivatized oligosaccharides using low nanomole quantities of sample. In addition, unlike in EI, molecular ions can often be obtained on these larger molecules. However, sensitivity is very dependent on the chemical nature of the sample being analyzed, the matrix used as well as the instrumentation, and for greatest sensitivity, relatively hydrophilic oligosaccharides are still derivatized to enhance hydrophobicity (surface activity) for analysis by FAB or LSIMS.

1.3.5 Electrospray Ionization

Electrospray ionization (ESI) is a relatively new ionization method and it has only been used routinely for the analysis of biological molecules since 1988. ESI involves the electrostatic nebulization of a solution of charged analyte ions by a large electrostatic field gradient of about 3 kV/cm. As the solution passes through a dry bath gas near atmospheric pressure, highly charged droplets are formed which shrink as the neutral solvent molecules evaporate. As the solvent evaporates, it is thought that the charge repulsion within a droplet becomes greater than the cohesive forces holding the droplet together, and a "Coulombic explosion" results, ejecting desolvated ions into the gas phase (Thomson and Iribarne, 1979). If the solution is acidic, proton or other cation attachment occurs in the positive ion mode, forming multiply charged positive ions (Whitehouse *et al.*, 1985). Proton abstraction occurs in the negative ion mode, forming multiply charged negative ions (Yamashita and Fenn, 1984). The multiple charging effect of analyte molecules permits the analysis of very large molecules using conventional mass analyzers because the mass-to-charge ratio (m/z) remains below 3000 for most biological molecules, such as

glycoproteins, even up to mass 100,000 to 150,000 Da. In addition, fragmentation of the sample usually does not occur during electrospray ionization, depending on the sample and the instrumental conditions. Therefore, molecular weights for intact proteins and some glycoproteins can be determined with high accuracy, including those for various glycoforms. The most commonly used non-mass spectrometric option for determining molecular weights of large biomolecules, polyacrylamide gel electrophoresis (PAGE) (Laemmli, 1970), provides only an estimate of molecular weight by comparison. In addition, large oligosaccharides, for which no molecular ions can be identified using FAB or LSIMS or which require nanomole quantities of sample for analysis, can often be identified using electrospray even when underivatized using only picomole quantities of sample (Duffin *et al.*, 1992; Li *et al.*, 1993).

Besides the increased mass range and sensitivity of this ionization method, electrospray is an ideal method for interfacing to a liquid chromatograph because it is a flow technique operating at atmospheric pressure (Whitehouse *et al.*, 1985). This allows direct separation and mass analysis of low picomole to femtomole quantities of peptides and glycopeptides using microbore or capillary reversed phase HPLC coupled directly to an ESIMS (HPLC/ESIMS). As described in the protein glycosylation section, use of HPLC/ESIMS for analysis of glycoprotein digests greatly decreases sample handling and analysis time. Because of the ease of use, wide mass range and high sensitivity of HPLC/ESIMS, this technique has rapidly become the method of choice for the analysis of protein and glycoprotein digest mixtures.

1.3.6 Matrix-Assisted Laser Desorption Ionization

Matrix-assisted laser desorption ionization (MALDI), another recently developed ionization method for the analysis of biomolecules, involves pulsing a laser of specified wavelength at

a probe containing the analyte co-crystallized with a large excess of a matrix which absorbs at the laser wavelength used. Two commonly used wavelengths are 10.6 μm (CO_2 laser, far IR range) and 337 nm (N_2 laser, far UV). Using a laser in the IR range, the matrix absorbs the energy from the laser and becomes vibrationally excited, causing the matrix molecules to be "ablated" into the gas phase. Using a laser in the UV range, the matrix absorbs the energy from the laser and becomes electronically excited, causing the matrix molecules to "explode" into the gas phase (Karas *et al.*, 1985; Karas *et al.*, 1987). Since the analyte is co-crystallized with the matrix, as the matrix enters into the gas phase, it carries the analyte along with it. The key to success with MALDI is the choice of matrix, and the way in which the analyte and matrix are co-crystallized. Since this is a new technique, new matrices (Currie and Yates III, 1993; Fitzgerald *et al.*, 1993; Juhasz *et al.*, 1993; Pielek *et al.*, 1993) and crystallization methods (Weinberger, 1993; Xiang and Beavis, 1994) are still being introduced, improving the analysis of specific compound classes such as glycopeptides, glycoproteins and other glycoconjugates (Juhasz *et al.*, 1993), as well as oligonucleotides (Currie and Yates III, 1993; Fitzgerald *et al.*, 1993; Pielek *et al.*, 1993).

The advantages of this ionization method are the simplicity of the instrumentation as well as the high mass range and high sensitivity and which can be achieved. The analysis of glycopeptides has been performed at the 500 femtomole level (Huberty *et al.*, 1993), and underivatized oligosaccharides have been analyzed at low picomole levels (Mock *et al.*, 1991), but problems have been reported with cleavage of sialic acid residues during the analysis of these glycoconjugates by MALDI (Huberty *et al.*, 1993). In addition, because of problems with shot to shot reproducibility during MALDI and the large kinetic energy spread of the ions produced, the resolution (typically 100-400) and accuracy of mass determination (0.1%) is significantly inferior to electrospray or FAB/LSIMS instruments, and internal standards are required to achieve 0.01% accuracy mass measurements. Because of these limitations, and because this is a new technique, there are a limited

number of reports of the analysis of glycoconjugates by MALDI-MS. The coming years should show significant improvements in the analysis of glycoconjugates by MALDI, for example, in choice of matrices and matrix additives, leading to a greater use of this type of instrumentation by carbohydrate chemists and mass spectrometrists.

1.4 Instrumentation

Mass analyzers separate ions using electric fields and magnetic fields, as well as their time of flight. As explained below, different mass analyzers can have very different specifications for parameters such as mass or mass-to-charge (m/z) range, resolution ($M/\Delta M$) or resolving power (the ability to separate ions close in m/z value), accuracy of mass measurement, ion transmission, types of detectors possible and ease of use in combination with ancillary equipment such as chromatographs. No one mass analyzer is best for all applications and the choice of analyzer used often depends on the ionization method used as well as the type of information which is required to solve a particular problem. For a review on mass analyzers, see Jennings and Dolnikowski, 1990 and Falick, 1994.

1.4.1 Sector

Most modern sector instruments contain both an electric sector and a magnetic sector. An electric sector consists of two parallel cylindrical plates across which an electric field is applied. The electric sector deflects ion beams as they pass through it according to their kinetic energy. The magnetic sector consists of a magnetic field (usually wedge-shaped) through which a previously accelerated ion beam is passed. Ions pass through the magnet,

where they follow a circular path perpendicular to the direction of the magnetic field. A simple sector has the property of direction focusing; that is, ions of the same mass accelerated from the source at slightly different angles will be refocused on the collector slit after they pass through the magnetic field. A magnetic sector alone has a limited resolution (up to 5000-10,000) because it actually analyzes momentum, and ions of the same mass having slightly different kinetic energies will be refocused in a peak having a distribution of slightly different mass values at the collector.

The basic magnetic sector equation helps to illustrate why a magnetic sector instrument analyzes momentum, and how it is used to determine m/z values. The force exerted on an ion by the perpendicular magnetic field in a magnetic sector is equal to $zevB$, where z is the charge on the ion, e is the charge, v is the ion velocity and B is the magnetic field strength. Since an ion of mass m travels in a circular path of radius R within the magnetic field, $zevB = mv^2/R$ or $zeB = mv/R$. Since $mv = \text{momentum}$, this equation shows that a magnetic sector actually separates ions according to their momentum. By combining this with the kinetic energy ($KE = mv^2/2$) of the ion, $mv^2/2 = zeV$ (where V is the accelerating voltage), the magnetic-sector equation is found to be $m/z = B^2 R^2 e / 2V$. Compensation for this distribution of ion energies may be achieved by appropriate tandem arrangement of an electric sector with a magnetic sector. The electric sector will adjust the trajectories of the ions of the same mass which have slightly different kinetic energies so that the energy-focusing image and the direction-focusing image coincide at the collector.

By arranging an electric and a magnetic sector in tandem using appropriate ion optical design parameters, a "double focusing" mass spectrometer is formed, because one then achieves energy focusing from the electric sector and mass dispersion from the magnetic sector. This focusing effectively increases the resolving power of just a sector mass analyzer alone. Adjustment of the source exit slits and collector slit adjusts the resolution of the instrument. The slits are made smaller for increased resolution, however since this lets fewer ions through, sensitivity is decreased. As mentioned above in the case of a

typical magnetic sector alone, the maximum resolution is approximately 5,000-10,000, but when combined with the appropriate electric sector in a double focusing instrument, resolutions of 100,000 can be achieved for a mass range up to m/z 3000 with an accuracy of mass measurement on the order of 1 ppm. To obtain a higher mass range, a lower accelerating voltage can be used, however, this causes a decrease in source ion extraction efficiency and therefore sensitivity. To increase sensitivity, the source and collector slits can be opened further, however, resolution is then sacrificed. For example, a resolution of 2000 is routinely used on sector instruments for a mass range up to m/z 10,000.

The major advantages of a double focusing sector mass analyzer are high mass range (compared to a quadrupole analyzer), high resolution and accurate mass determination, as well as fast scanning speed and dynamic range. In addition, array detectors, which allow detection of all ions simultaneously over a specific percentage of the mass range, resulting in increased sensitivity, can be used with sector instruments but not with quadrupole instruments. These advantages have made this analyzer, coupled with LSIMS or FAB sources, a very common mass spectrometer used by carbohydrate chemists today. The disadvantages are the cost and size of the instruments (compared to a quadrupole analyzer), and the need to operate the ion source at 3-10 kV above ground potential, which can complicate interfacing with other equipment such as chromatographs.

1.4.2 Quadrupole

A quadrupole mass analyzer contains four parallel rods arranged in a square bundle. The arrangement of these rods is optimal if hyperbolic in cross section. An electric field is created by electrical potentials applied to the rods. If opposite pairs of rods are connected electrically and proper direct current and radio frequency voltages are applied to these rods, ions of a particular m/z value will pass through the quadrupole in stable trajectories. That

is, at a certain dc and rf potential, heavy ions are more influenced by the dc potential than the rf field, causing them to be focused onto the center axis of the instrument and passing on to the detector. Lighter ions, however, are more influenced by the rf potential and will not be focused onto the center axis of the instrument, causing them to collide with the rods and not be transmitted to the detector. In this manner, the electric fields are used to filter or separate ions according to their m/z values. If the rf and dc voltages are scanned such that their ratio remains constant, mass scanning is accomplished and ions of different m/z values will, in turn, follow stable paths, and a mass spectrum which is linear in mass will be obtained.

The quadrupole mass analyzer is probably the most common mass analyzer used today. Its widespread use is based on several features including small size, source operation close to ground potential, rapid scanning ability and no requirement for slits as in sector instruments, yielding high ion transmission at relatively low mass ($\leq m/z$ 1000). However, compared to a sector mass analyzer, its m/z range, accuracy of mass measurement and resolving power are generally not as good. A quadrupole mass analyzer is usually limited to unit mass resolution, which means that $M/\Delta M$ increases with increasing mass. This is because higher mass ions have lower velocities and therefore spend a longer time in the analyzer, experiencing more cycles of the rf field. The mass range is usually limited to m/z 3,000-5,000. Quadrupole mass filters are used often for GC/MS since the gas chromatograph and ion source can be easily interfaced because the source is operated close to ground potential. In addition, quadrupoles may be operated at higher pressures (up to 10^{-4} torr) than sector instruments, which allows them to be easily interfaced with a variety of ion sources. Because there are no slits, transmission of ions is high (esp. $\leq m/z$ 1000) and sensitivity is often comparable to sector instruments below mass 1000 Da. Above m/z 1000, however, sensitivity decreases due to decreased ion transmission.

1.4.3 Time of Flight

The time-of-flight (TOF) mass analyzer is the simplest of the three major types of mass analyzers. Its operation is based on the fact that as ions (when accelerated through a given potential) of the same kinetic energy but different masses are pulsed out of the ion source, the heavier ions must have lower velocities and thus take longer than the lighter ions to travel through the flight tube and reach the detector. A mass spectrum is obtained by measuring the times taken for different mass ions to travel the same distance, and there is theoretically no upper limit to the mass range. The TOF analyzer is used with sources which can produce pulses of ions in the same plane, all having nearly the same kinetic energy. Therefore, TOF mass analyzers are well suited for pulsed laser or plasma desorption ion sources. TOF analyzers also require the use of detector circuitry with nanosecond time resolution. Enhanced resolution and mass measurement accuracy can be achieved by adding a series of reflecting electrostatic lenses at the end of the flight tube, termed a reflector or a reflectron. Using a reflectron, ions of higher energy arrive first and penetrate more deeply into the lens taking a longer path length, allowing those of lower energy to catch up, so that ions of the same mass reach the detector together, and ions of similar velocities but different mass are separated.

As ionization methods have improved, the upper usable mass range of TOF analyzers has risen. The detection of ions of m/z greater than 200,000 has been reported (Hillenkamp *et al.*, 1991), and still appears to be increasing. The accuracy of mass measurement achievable (0.1%-0.05%) is where this analyzer often suffers compared to the others. At m/z 50,000, it is capable of measuring m/z to ± 10 -100, depending on the methods of calibration and ionization. Resolution for analyzers without reflectors is typically below 400, while with reflectors resolution has been increased to 5,000 in some cases. The major advantages of TOF analyzers are very high mass range, high sensitivity, fast repetition rate and simplicity. Typically, the sample is consumed only during the

ionization pulse, so essentially all the ions produced can be collected, yielding very high sensitivity.

1.4.4 Collision-Induced Dissociation (CID)

Collision-induced dissociation (CID) is most often performed by connecting two mass analyzers in tandem (tandem mass spectrometry, MS/MS) with a collision cell in between. Using the first mass analyzer, ions of a particular mass values are selected as the analyte is being ionized. When LSIMS or FAB ionization is used, the possibility of selecting ions having one or a very small range of m/z value with the first mass analyzer eliminates a great deal of the chemical background noise due to the matrix. The selection of ions with the first mass analyzer (MS-1) also acts as a separation or purification step, since ions of all other masses are rejected. Those selected "precursor" ions then pass into the collision cell where they collide with a neutral gas such as helium. These collisions increase the internal energy of the ions (collisional activation), causing unimolecular decomposition (fragmentation) of the precursor ions. The resulting fragment ions then pass into the second mass analyzer and on to the detector, creating a mass spectrum of fragment ions resulting from precursor ions of the particular mass selected with the first mass analyzer (Hayes and Gross, 1990).

Ions can be collisionally activated through high energy (keV) collisions or low energy (eV) collisions, depending on the mass analyzers employed (McLafferty, 1983). Most high energy CID is performed on a tandem four-sector instrument consisting of two double focusing mass spectrometers connected in tandem with a collision cell in between. This type of instrument allows high resolution precursor ion selection in MS-1 with at least unit mass resolution in the second mass analyzer (MS-2). Most low-energy CID is performed on a triple quadrupole instrument where the first and third quadrupoles are analogous to

MS-1 and MS-2 of a four-sector instrument, and the second quadrupole functions as the collision cell. This type of instrument allows much less than unit resolution in the MS-1 and unit resolution in MS-2. Low energy CID can also be performed on a "hybrid" instrument which contains a double focusing sector mass analyzer as MS-1, followed by a quadrupole collision cell and a quadrupole as MS-2. As one might expect, high energy CID produces more types of fragment ions than low energy CID. The advantage of obtaining additional types of fragment ions is that more detailed structurally significant information may be obtained, such as those from ring cleavage processes in oligosaccharides (described in Chapter 2). However, the data obtained is more complex and sometimes more difficult to interpret. A comparative study of high energy and low energy CID (on a hybrid instrument) for the analysis of peptides was published in 1991 (Bean *et al.*, 1991; Tang *et al.*, 1993) which concluded that more complete sequence information and greater sensitivity could be obtained using high energy CID on peptides greater than m/z 1000. In addition, the high energy spectra obtained by the investigators were found to be much more reproducible than the low energy spectra and were less sensitive to the composition of the peptide being analyzed (Bean *et al.*, 1991; Tang *et al.*, 1993). The instrumentation required for high energy CID (four-sector) compared to low energy CID (triple quadrupole or hybrid) has the same advantages and disadvantages as two sector instruments compared to single quadrupole mass analyzers (see above).

1.5 Sample Inlet Systems

There are several different methods of introducing a sample into a mass spectrometer, depending on the type of sample to be analyzed and the ionization method used.

1.5.1 Gas-Liquid Chromatography (GLC)

GLC is commonly used for the analysis of small volatile compounds by EI-MS and CI-MS. The sample is heated in the capillary column of the GLC and elutes from the column in the gas phase directly into the mass spectrometer. This method is very common for the analysis of derivatized monosaccharides and small oligosaccharides as well as other small organic molecules. A special application of GLC/MS of oligosaccharides up to hexasaccharides as permethylated N-trifluoroacetates was developed in the early 1980s (Nilsson and Zopf, 1983), and later, new thermostable bonded GLC stationary phases were introduced permitting oligosaccharides containing up to ten or eleven sugar units to be chromatographed and analyzed by GLC/MS (Karlsson *et al.*, 1987; Hansson *et al.*, 1989). Structures containing more than ten or eleven sugar residues are difficult to introduce via a GLC column because thermal decomposition and bleeding of the column liquid-phase are a problem at the temperatures needed to chromatograph the oligosaccharides. Therefore, static probes are used for EI and CI of higher molecular weight samples.

1.5.2 Static Probe

There are a number of static probes of different designs used for direct insertion of samples into a mass spectrometer. For EI of larger derivatized oligosaccharides, samples are inserted into a small quartz tube at the end of the probe tip. When the probe is inserted into the mass spectrometer, the tip is electrically heated and the sample is distilled off into the gas-phase. Because of the reduced pressure inside the mass spectrometer, the temperature needed to evaporate the large derivatized samples is low enough to avoid thermal decomposition. Static probe is also the most common inlet used with FAB and LSIMS

ionization. The sample is placed on a sample stage as a liquid droplet, inserted into the instrument and the Cs^+ ion or Xe^+ atom beam hits the droplet, causing the sample to be ionized. Static probes are also used for MALDI instruments. In this case, the sample is inserted into the instrument in a solid, crystalline form and the laser pulses hit the sample, causing desorption/ionization.

1.5.3 Liquid chromatography (LC)

The earliest approaches to interfacing liquid chromatography to mass spectrometers (LCMS) involved direct liquid introduction (DLI) (Games, 1984) as well as moving belt transfer interfaces (McFadden *et al.*, 1976; Santikarn *et al.*, 1987) which were useful for many applications (Tomer and Parker, 1989), but they never achieved widespread acceptance due to problems with interfacing and memory effects. Another liquid interface method, continuous-flow FAB (CF-FAB) (Caprioli *et al.*, 1986) was developed to diminish two of the major difficulties with static probe FAB or LSIMS--i.e., the high chemical background noise resulting from using the high matrix concentrations required, and the formation of surface distribution effects in the liquid droplet being ionized. This method involved the constant flow pumping of a mixture of a small amount (1-20 ml/min) of matrix with aqueous solvent. The sample was then introduced into that continuous flow of liquid and brought to the sample stage on the end of a flow probe for ionization. This was then extended to using a microbore HPLC system with the continuous-flow probe to perform direct high performance liquid chromatography/mass spectrometry (HPLC/MS) (Caprioli *et al.*, 1986; Caprioli *et al.*, 1987). Nevertheless, HPLC/MS was not widely and routinely used until it was interfaced to electrospray ionization mass spectrometers (Whitehouse *et al.*, 1985). Since electrospray ionization (ESI) involves the ionization of samples in a flowing liquid near atmospheric pressure, this method is ideally suited to

interfacing with liquid chromatographs. Operation of an ESIMS involves a constant stream of liquid flowing through the probe into the ion source, so the addition of an HPLC to the system is simple and straightforward. The ruggedness and stability of an HPLC/ESIMS system is greater than for a HPLC/CF-FABMS system. Because the mass range and sensitivity are greater as well, HPLC/ESIMS has all but replaced the use of HPLC/CF-FAB for the separation and analysis of mixtures such as protein digests (Hemling *et al.*, 1990; Ling *et al.*, 1991; Carr *et al.*, 1993). However, for a few specific applications such as HPLC combined with high energy CID, CF-FAB and HPLC/CF-FAB are still used today (Hall *et al.*, 1993; Burlingame, 1994).

1.6 Conclusions

Detailed structural analysis of carbohydrates on glycoproteins is a challenging task, and no single analytical tool could detect or characterize all oligosaccharide structures which contain such component variability and complexity. Nevertheless, most detailed studies require the use of a combination of chemical, enzymatic, chromatographic, mass spectrometric and NMR methods. The majority of the strategies used today involve the use of mass spectrometric methods including GC-MS (composition and linkage analysis), FAB-MS and/or FAB-MS/MS of derivatized glycoconjugates. These methods provide information on very small quantities of sample not readily determined from NMR and they provide a large amount of structural information as well. Together, these methods 1) enable one to determine molecular size, 2) reveal heterogeneity directly, 3) provide considerable structural and branching information, and 4) permit structural analysis of all individual components of mixtures (using MS/MS) which have similar chemical properties and thus, are difficult to separate chromatographically.

However, there are now several examples where the value of newer techniques such as ESIMS and MALDI-TOFMS have been demonstrated because of their ability to analyze native glycoconjugates intact which have not been subjected to chemical or enzymatic procedures which may alter labile structural features in previously unforeseen ways. Because of their high mass measurement capabilities, sensitivity and their ability to analyze underivatized samples, these newer methods provide an important compliment to the more established mass spectrometric and NMR techniques used for glycoconjugate analysis. ESIMS has replaced FAB-MS as the method of choice for certain analyses such as peptide and glycopeptide mapping. Using HPLC/ESIMS combined with CID and selected ion monitoring of carbohydrate-specific fragment ions (described in Chapter 2), glycopeptides can be rapidly identified in proteolytic digests of glycoproteins. Once identified, other analytical methods such as glycosidase digestion can then be used to further characterize the oligosaccharides on the glycopeptides.

CHAPTER 2. METHODS DEVELOPMENT FOR THE ANALYSIS OF PROTEIN GLYCOSYLATION

The complete analysis of a glycoprotein involves identification of the entire array of oligosaccharide structures attached to the protein as well as the quantitative assignment of each structure to its site(s) of attachment on the polypeptide chain. Structural analysis of glycoproteins is a formidable challenge to mass spectrometry due to the heterogeneity of oligosaccharide chains which are usually present on a glycoprotein (Carr *et al.*, 1989; Spellman, 1990; Gillece-Castro *et al.*, 1991; Harris *et al.*, 1993; Ohmori *et al.*, 1993). Different oligosaccharide structures are often attached to multiple different sites in a glycoprotein, and heterogeneity may exist not only among these different sites but also within the group of structures which occurs at each single site of glycosylation. Classically molecules called glycoproteins are serum and cell membrane proteins which contain oligosaccharide substituents linked to asparagine (Asn, termed *N*-linked) in the Asn-Xxx-Ser/Thr sequon (where Xxx is any amino acid except proline) (Bause and Hettkamp, 1979), serine or threonine (Ser/Thr, termed *O*-linked) where no consensus sequence for attachment is known. More recently, carboxy-terminal glycosyl-phosphatidylinositol (GPI) lipids have been discovered which anchor certain glycoproteins in the membrane (termed C-terminal glycosyl-phosphatidylinositol (GPI) lipid anchors). Because there is no consensus sequence for *O*-linked glycosylation, the identification of carbohydrate structures occurring at specific attachment sites of *O*-linked sugars presents an even greater analytical challenge. Because of the difficulties in the solubilization of membrane bound glycoproteins without the use of detergents (which inhibit ionization of molecules by all mass spectrometric methods, with the possible exception of MALDI-TOFMS in some cases) the strategies presented in this chapter are those used to characterize soluble glycoproteins.

The analysis of oligosaccharides attached to glycoproteins by mass spectrometry has seen a tremendous revolution in the past few years with the advent of techniques such as FAB and LSIMS (Barber *et al.*, 1981; Aberth *et al.*, 1982), ESIMS (Fenn *et al.*, 1989; Smith *et al.*, 1990) and MALDI-TOFMS (Karas and Hillenkamp, 1988; Hillenkamp *et al.*, 1991). The increased sensitivity and vastly extended mass range of these latter two techniques compared to FAB and LSIMS has permitted the analysis of glycopeptides, peptidoglycans and large oligosaccharides which was previously difficult or even not possible, and at significantly lower sample levels, as well as even certain intact glycoproteins. Direct coupling of microbore or capillary HPLC systems to electrospray ionization mass spectrometers (HPLC/ESIMS) has enabled the routine analysis of proteolytic digests of glycoproteins at the low picomole level (Hemling *et al.*, 1990; Ling *et al.*, 1991; Carr *et al.*, 1993; Medzihradzky *et al.*, 1994). Using HPLC/ESIMS, covalently linked carbohydrate-peptide components (glycopeptides and peptido-glycans) can be detected and analyzed in the same mixture, allowing one to directly determine the attachment sites of the carbohydrates to the peptides as well as address structural heterogeneity at individual glycosylation sites in a single mass spectral analysis.

Developments in methods of CID by HPLC/ESIMS (Conboy and Henion, 1992; Carr *et al.*, 1993) and analysis of metastable ions formed during MALDI-TOFMS (Huberty *et al.*, 1993) have facilitated the location and preparative fractionation of glycopeptides at low picomole levels as well as increasing the sensitivity by which oligosaccharides (Duffin *et al.*, 1992) and glycopeptides can be sequenced (Huberty *et al.*, 1993). Using HPLC/ESIMS, fragmentation can be induced as the analyte eluant enters the mass spectrometer either before the first quadrupole (LC/ESI/CID/MS) or within the second quadrupole (LC/ESI/MS/MS) of a triple quadrupole mass analyzer. Monitoring for fragment ions specific for carbohydrate moieties (selected ion monitoring, SIM) during either of these CID methods allows one to readily identify glycopeptide-containing fractions in even the most complex mixtures of peptides and glycopeptides (Carr *et al.*, 1993;

Medzihradzky *et al.*, 1994). Using MALDI-TOF instruments equipped with reflectron type mass analyzers, the ability to detect and analyze fragments to sequence glycopeptides at the femtomole sample level was recently reported (Huberty *et al.*, 1993). These newer methods, combined with more classical methods such as permethylation and peracetylation of released oligosaccharides for composition, sequencing and linkage analysis have helped to revolutionize the field of glycoprotein analysis by mass spectrometry.

The first approach to obtaining a global look at a soluble glycoprotein involves an estimate of the molecular weight and extent of heterogeneity of the protein if possible. This can be done by ESI-MS and/or MALDI-TOFMS, however, more glycoprotein molecular weights have been obtained by MALDI-TOFMS (Hillenkamp *et al.*, 1991; Mock *et al.*, 1991; Chait and Kent, 1992; Juhasz *et al.*, 1993) than with ESI-MS (Green and Oliver, 1991). Since most glycoproteins are very heterogeneous, ESI-MS is often not useful, because the lack of resolution of the resulting multiply charged peaks make it difficult or impossible to determine the charge state of the individual peaks, and therefore, a measure of the molecular weight of the protein. MALDI-TOFMS of a heterogeneous glycoprotein yields a broad singly charged peak (as shown for yeast expressed platelet-derived growth factor in Chapter 3), however, and if internal standards are used, a reasonably accurate average center of mass can be determined. If the protein sequence is known, the difference between the intact mass and the molecular weight of the amino acids that make up the protein provides an estimate of the size distribution of the glycans, and yields information on the total carbohydrate content of the glycoprotein. By treating the intact glycoprotein with exo- and/or endoglycosidases such as neuraminidase and endoglycosidase-H (Endo H), this heterogeneity is reduced and then reanalysis of the product(s) may yield a more accurate measure of molecular weight as well as additional information on the carbohydrates based on the change in average molecular weight of the intact molecule (Chait and Kent, 1992).

Structural analysis of oligosaccharides attached to a glycoprotein involves several levels of detail which must be addressed with different mass spectrometric methods:

- 1) *The identity and proportion of monosaccharide residues present on the protein.*
- 2) *The structural class(es) of oligosaccharides present on the protein.*
- 3) *The nature and structural heterogeneity of oligosaccharide structure(s) attached at specific sites on the protein.*
- 4) *The sequence, branching and linkages of each oligosaccharide structure.*

These will be addressed in the following sections, and at the appropriate level of detail, the different methods for the analysis of *N*-linked structures and *O*-linked structures will be discussed separately.

2.1 Determination of Monosaccharide Composition

2.1.1 TMS Ethers and Alditol Acetates

The first step involved in the analysis of oligosaccharides is to identify and quantify the monosaccharide composition present. The classical mass spectrometric methods of composition analysis involve GC-FID or GC-MS analysis of trimethylsilyl (TMS) ethers of methyl glycoside derivatives of hydrolyzed oligosaccharides (Sweeley *et al.*, 1963) or of alditol acetates of hydrolyzed oligosaccharides (Sawardeker *et al.*, 1965; Hellerqvist and Sweetman, 1990). Monosaccharides may be identified by their retention times on the GC with flame ionization detection (FID) compared to internal or external standards (as well as their mass spectra if GC-MS is used). The monosaccharide alditol acetates or TMS ethers are also quantitated by the FID response compared to an internal standard such as perseitol (L-glycero-D-manno-heptitol). The tentative identification of particular peaks as

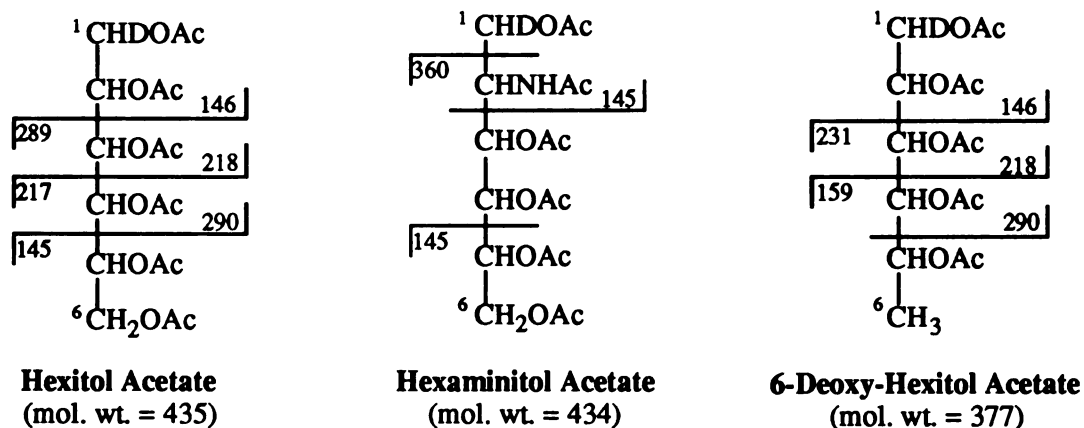
representing specific sugars (based on GC retention time) is confirmed or rejected by GC-MS analysis using a similar column and conditions to that used for GC analysis.

Sample preparation for monosaccharide analysis typically involves hydrolysis of the glycan in 4 M TFA for 4 hrs at 100°C (or 0.8 M HCl in anhydrous methanol at 80-100°C for 24 hrs) into monosaccharides which are reduced to alditols using NaBD₄. The alditols are then dissolved in anhydrous pyridine and treated with hexamethyldisilane and trimethylchlorosilane to form the TMS ethers or acetylated with pyridine and acetic anhydride (1:2, v/v) at 100°C for 30 minutes to form the alditol acetates. Scheme 1 illustrates the resulting alditol acetate structures for hexoses (Hex) (e.g. mannose, Man and galactose, Gal), hexosamines (HexNH₂) (e.g. glucosamine, GlcNH₂ and galactosamine, GalNH₂) and deoxyhexoses (dHex) (e.g. fucose, Fuc), and the primary fragments which result from EI and/or CI analysis. NaBD₄ is used so that the deuteration of C-1 allows one to distinguish fragments containing C-1 from those containing C-6. As illustrated in Scheme 1, the charge is retained on either carbon at the cleavage site, when all the oxygens are acetylated. For monosaccharides containing methoxyl groups, the charge is preferentially retained on the methoxylated carbon(s). This will be discussed further for GC-MS of partially methylated alditol acetates in section 2.2.3 (determination of sequence, branching and linkages) which is used to determine linkages between the individual monosaccharides in an oligosaccharide.

2.1.2 HPAE-PAD of Hydrolyzed Underivatized Glycans

Alternatively, a newer method of monosaccharide composition analysis which does not involve mass spectrometry can be used. This method involves direct analysis of the (underivatized) sugars, hydrolyzed in 2 M TFA (for neutral sugars) or 4-6 M HCl (for amino sugars) at 100°C for 4-6 hrs, and 0.1 N HCl at 80°C for 1 hr (for sialic acids), using

high pH anion exchange chromatography (Dionex CarboPac™ column) with pulsed amperometric detection (HPAE-PAD) (Barr *et al.*, 1991; Townsend and Hardy, 1991).



Scheme 1. Representative alditol acetate structures illustrating the primary fragments produced by EI and/or CI.

Monosaccharides are quantitated by comparing the elution times and peak areas of known quantities of standard monosaccharides with those from the samples.

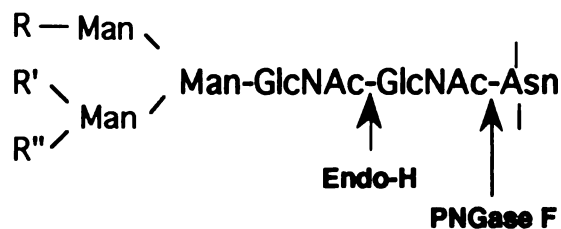
Once the residue composition of the glycosylation is known, a great deal can be inferred concerning putative structural class(es) of oligosaccharides present on the glycoprotein, and then further levels of detail in the analysis can be addressed. For example, detection of fucose and sialic acid indicate the presence of complex *N*-linked structures, while detection of *N*-acetylgalactosamine indicates possible *O*-linked structures. The mass spectrometric analysis of these different classes of structures will be discussed in turn.

2.2 Asparagine-linked Glycosylation

2.2.1 Determination of Structural Class

The first step required in the analysis of *N*-linked oligosaccharides is to determine if they are of the high-mannose, hybrid or complex structural class. As discussed above, the composition analysis should indicate globally what types of structures are present on the entire protein. For more complete analysis, treatment of glycoproteins with endoglycosidases specific for structures of a certain structural class can be performed (Maley *et al.*, 1989). For example, endo- β -*N*-acetylglucosamine H (Endo-H), hydrolyzes the chitobiose core of high mannose and hybrid structures (Trimble and Maley, 1984), while peptide:*N*-glycosidase F (PNGase F, actually a mixture of three enzymes, PNGase F_{1,2,3}) cleaves all classes (high-mannose, hybrid and complex) of *N*-linked structures (Tarentino *et al.*, 1985; Plummer and Tarentino, 1991). As shown in Scheme 2, PNGase F cleaves the entire oligosaccharide from the protein, converting the Asn to which it is attached to an Asp (Tarentino *et al.*, 1985), leaving the rest of the protein or peptide intact; while Endo-H cleaves the reducing terminal GlcNAc of the chitobiose core attached to the Asn, leaving the protein or peptide with a single GlcNAc residue attached to the glycosylated Asn (Trimble and Maley, 1984). Either of these enzymes can be used as the initial step in strategies for the characterization of the oligosaccharides as well as in the identification of their attachment sites.

In order to isolate and analyze free oligosaccharides from a glycoprotein, they must be either chemically or enzymatically released.



Scheme 2. Illustration of the sites of cleavage for Endo H and PNGase F in an Asn-linked oligosaccharide.

2.2.1.1 Chemical Release of N-linked Oligosaccharides

Chemical release of *N*-linked oligosaccharides is most often performed by hydrazinolysis which cleaves the GlcNAc-Asn linkage (Takasaki *et al.*, 1982), leaving a glycosylamine residue as the reducing terminus (all *N*-acetyl groups are removed and the protein is destroyed). The reaction is performed with anhydrous hydrazine at 100°C for 8-12 hr. This reaction can be tricky and if not performed under strict anhydrous conditions, side-products become a severe problem. However improvements and even automation of this reaction have recently been reported (Patel *et al.*, 1993). The oligosaccharides can be re-*N*-acetylated with acetic anhydride and then the resulting acetohydrazone is converted back to the alcohol (the unreduced glycan) by treating the sample with a catalytic amount of Cu²⁺ in acid (Patel *et al.*, 1993).

2.2.1.2 Enzymatic Release of N-linked Oligosaccharides

For enzymatic release of *N*-linked oligosaccharides the intact glycoprotein can be treated with an endoglycosidase such as Endo H (50 mM NH₄HCO₃, pH 8.0, 0.075 mU/ml, 16 hr at 37°C) or PNGase F (also termed *N*-glycanase) (50 mM NH₄HCO₃, pH 8.0, 0.1

mU/ml, 16 hr at 37°C). The oligosaccharides can be separated from the protein by reversed phase HPLC (Settineri *et al.*, 1992) or by a C18 Sep-Pak® cartridge (Poulter *et al.*, 1989) using acetonitrile and water solvents, where the oligosaccharide fraction elutes in the (aqueous) void volume. However, it is recommended that the glycoprotein be digested with a protease first, so that a smaller quantity of endoglycosidase is required and most importantly the reaction is more likely to go to completion. Before proteolytic digestion, any disulfide bonds in the protein are first reduced with a 50 X molar excess of dithiothreitol (DTT) in 100 mM NH₄HCO₃, pH 8.5, 6 M guanidine HCl, 3 mM EDTA, for 1 hr at 50-60°C, followed by alkylation of free thiols for 30 minutes at room temperature with iodoacetic acid (IOAc) in the same buffer. The carboxymethylated protein, typically 200 picomole up to 25 nanomole, is then dialyzed exhaustively with a microdialyzer apparatus into 50 mM NH₄HCO₃, pH 8.0 before proceeding to digestion with a protease such as trypsin or endoproteinase glu-C. Typical protein concentrations for protease digestions are 0.1 to 1.0 mg/ml, so the sample may be concentrated after dialysis if necessary, since the ammonium bicarbonate is a volatile buffer. Trypsin is typically added at a concentration of 1:50 (2%, w/w) enzyme:substrate at time zero and three hrs (for a total of 4%), with a total digestion time of six hrs at 37°C. Endoproteinase glu-C digestions usually require two aliquots of 4% enzyme:substrate over a digestion time of 12 to 20 hrs.

Depending on the amount of sample, a portion of the resulting mixture of peptides and glycopeptides can be analyzed directly by HPLC/ESIMS (*vide infra*) and another portion can be treated with PNGase F or Endo H and vacuum dried. The released oligosaccharides are then derivatized (the types of derivatives, the information obtainable from these derivatives and relative merits of each are discussed below) for EI-MS or LSIMS analysis. Derivatized samples tend to fragment in very predictable ways, which is often not the case for native oligosaccharides. If LSIMS of a derivatized oligosaccharide yields an intact molecular ion (typically MH⁺ or MNa⁺, or both), overall composition can often be defined

(depending on what is known about the source of the sample). In addition any fragment ions obtained can yield sequence information.

Table I illustrates the types of fragment ions that are observed in derivatized as well as underivatized oligosaccharides. Unfortunately, the carbohydrate community has not agreed on the use of a single systematic nomenclature, so Table I illustrates the nomenclature for the two most often described nomenclatures. For example, the common cleavage on the nonreducing side of glycosidic bonds to give oxonium ions are termed A-type cleavage ions by some researchers (Dell, 1987; Peter-Katalinic and Egge, 1987), while it is termed a B_i-ion by others (Domon and Costello, 1988), in a manner analogous to the ions first described for peptide fragmentation (Roepstorff and Fohlman, 1984; Biemann, 1988) (see Table I). The subscript, *i*, represents the number of the glycosidic bond cleaved, relative to the nonreducing terminus. Y- and C-ions, also referred to as β-cleavage ions, involve hydrogen transfer (or sometimes methyl or acetyl transfer in derivatized samples) to the glycosidic oxygen, as shown in Table I.

Finally, X- and A-ions, also referred to as ring cleavage ions, are believed to be formed by various types of fragmentation involving specific hydrogen transfers and result in cleavage of the sugar rings. Figure 1 illustrates the details of the nomenclature proposed by Domon and Costello (Domon and Costello, 1988), which will be used throughout this thesis. Fragments formed where the charge is retained on the nonreducing terminus of the molecule are termed A, B and C, while fragments with charge retention on the reducing terminus are termed X, Y and Z, as shown in Figure 1. The subscripts indicate the position of the cleavage relative to the termini and the superscripts indicate cleavages across a ring for A and X-type fragments.

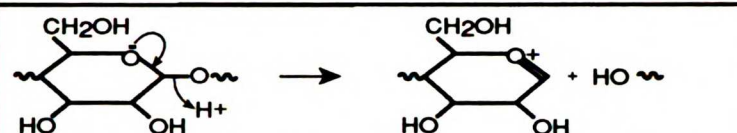
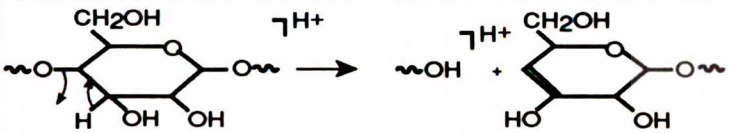
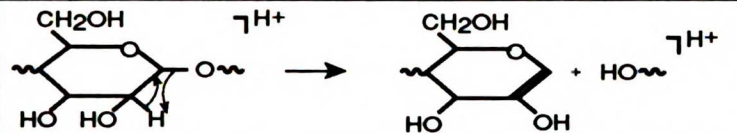
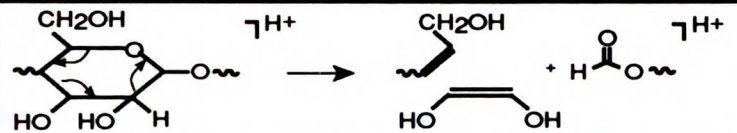
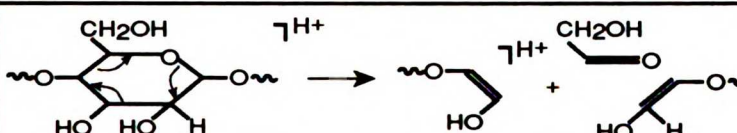
| Proposed fragmentation pathway | Domon & Costello | Dell |
|---|------------------|--------------------------------|
|  | B-ion | A-type oxonium cleavage |
|  | C-ion | β -cleavage (H transfer) |
|  | Y-ion | β -cleavage (H transfer) |
|  | 1,5 X-ion | Ring cleavage |
|  | 2,4 A-ion | Ring cleavage |

Table I. Carbohydrate fragmentation nomenclature as proposed by Domon and Costello (Domon and Costello, 1988) and Dell and coworkers (Dell, 1987)

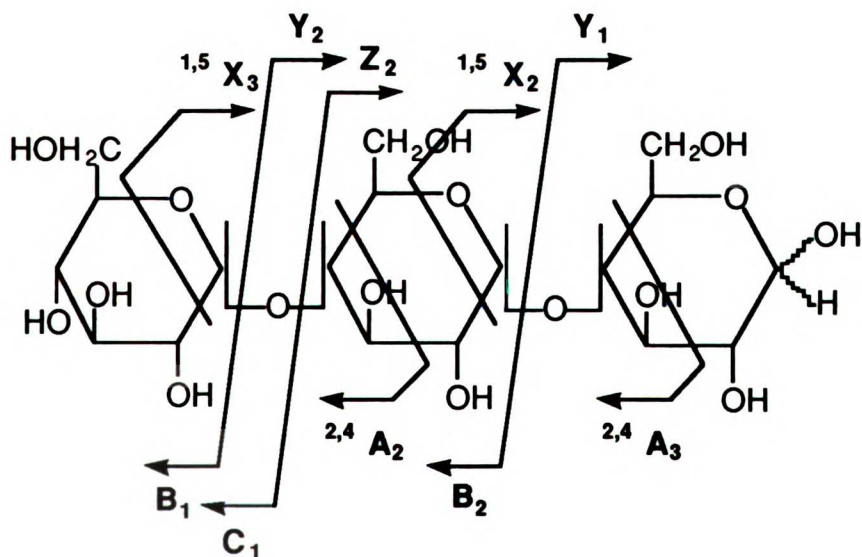


Figure 1. Nomenclature for mass spectrometric fragmentation of oligosaccharides as proposed by Domon and Costello (Domon and Costello, 1988).

2.2.1.3 Peracetylation

One commonly used derivatization procedure, which permits purification of the sample and results in concomitant increased surface activity and therefore, enhanced sensitivity, is peracetylation (Dell and Kobata, 1986). The details of this procedure are as follows: TFAA/acetic acid is added to the sugar/protein/salt mixture of oligosaccharides after PNGase F or Endo H digestion so as neutralize any residual base and the mixture is allowed to react for 10 minutes at room temperature. The reagents are removed with nitrogen and then the acetylated oligosaccharides are dissolved in chloroform and washed several times with water to remove salts and other water-soluble impurities. The acetylated oligosaccharide mixture is analyzed by positive ion FAB or LSIMS, and if the mass of the peracetylated mixture is very high, such mixtures can be subsequently permethylated, thereby reducing the mass (Dell and Thomas-Oates, 1989).

2.2.1.4 Permethylation

Permethylation is a more chemically harsh derivatization method but often yields better quality spectra due to a larger increase in sensitivity upon derivatization than in the case of peracetylation. Permethylated is most often carried out with a procedure which employs NaOH as the base [the Kerek method, a modification of the original Hakamori permethylation procedure (Hakamori, 1964)] (Ciucanu and Kerek, 1984), which is carried out as follows: the sample dissolved in water is first reduced with borodeuteride to prevent β -elimination at the acidic reducing end of the oligosaccharide when exposed to base (alkaline peeling) (Ballou, 1954). Several pellets of NaOH are placed in a dry mortar and approximately 3 ml of dry dimethyl sulfoxide (DMSO) is added using a Pasteur pipette,

and a slurry is formed with a pestle (a modified Kerek method using aqueous NaOH has been recently developed as well (Anumula, 1992)). About 1 ml of the slurry is added to the dry sample. One ml of methyl iodide is added and the mixture shaken for about 10 minutes. The reaction is then quenched by the careful addition of 1 ml of water which is added in 4 aliquots with shaking between additions. When the vigorous reaction subsides, 2 ml of chloroform is added, and the mixture shaken and allowed to settle. The water layer is removed with a pipette and the chloroform layer is washed with water until the water being removed is clear (a Sep pak can also be used as an alternative to extraction). The chloroform layer is dried under nitrogen and the sample lyophilized.

Typically permethylated and peracetylated samples are dissolved in methanol or chloroform, using monothioglycerol or *meta*-nitrobenzylalcohol (*m*-NBA) as the matrix for analysis by positive ion LSIMS. The LSIMS spectra of permethylated and peracetylated samples often yield sodiated molecular ions, MNa^+ , rather than protonated molecular ions, MH^+ , especially if *meta*-nitrobenzylalcohol (*m*-NBA) is used rather than glycerol or monothioglycerol as the LSIMS matrix (Dell *et al.*, 1994). LSIMS analysis of permethylated oligosaccharides is particularly useful for the characterization of complex sugar structures. Permethylated complex structures, like other lactosaminoglycans, give relatively intense fragment ions corresponding to B-ions at every N-acetylhexosamine (HexNAc) (e.g. N-acetylglucosamine, GlcNAc or N-acetylgalactosamine, GalNAc) residue (Fukuda *et al.*, 1984). However, additional fragmentation occurs in permethylated LSIMS spectra which is not observed in peracetylated spectra. For example, preferential elimination of a substituent in the 3-position is often observed in spectra of permethylated oligosaccharides, and permits differentiation between 1→3 and 1→4 linkages (Smith *et al.*, 1975; Grönberg *et al.*, 1989). Permethylated high mannose *N*-linked oligosaccharides, however, yields only minimal sequence information, because of the selective cleavage at the only GlcNAc residues which are in the chitobiose core (Fukuda *et al.*, 1984). Therefore, in the case of high mannose oligosaccharides, peracetylation is preferred over

permethylation to obtain sequence information. The differences between the typical FAB mass spectra of a peracetylated (Figure 2A) and a permethylated (Figure 2B) oligosaccharide are illustrated for lacto-N-decaose from human milk. Both spectra yield

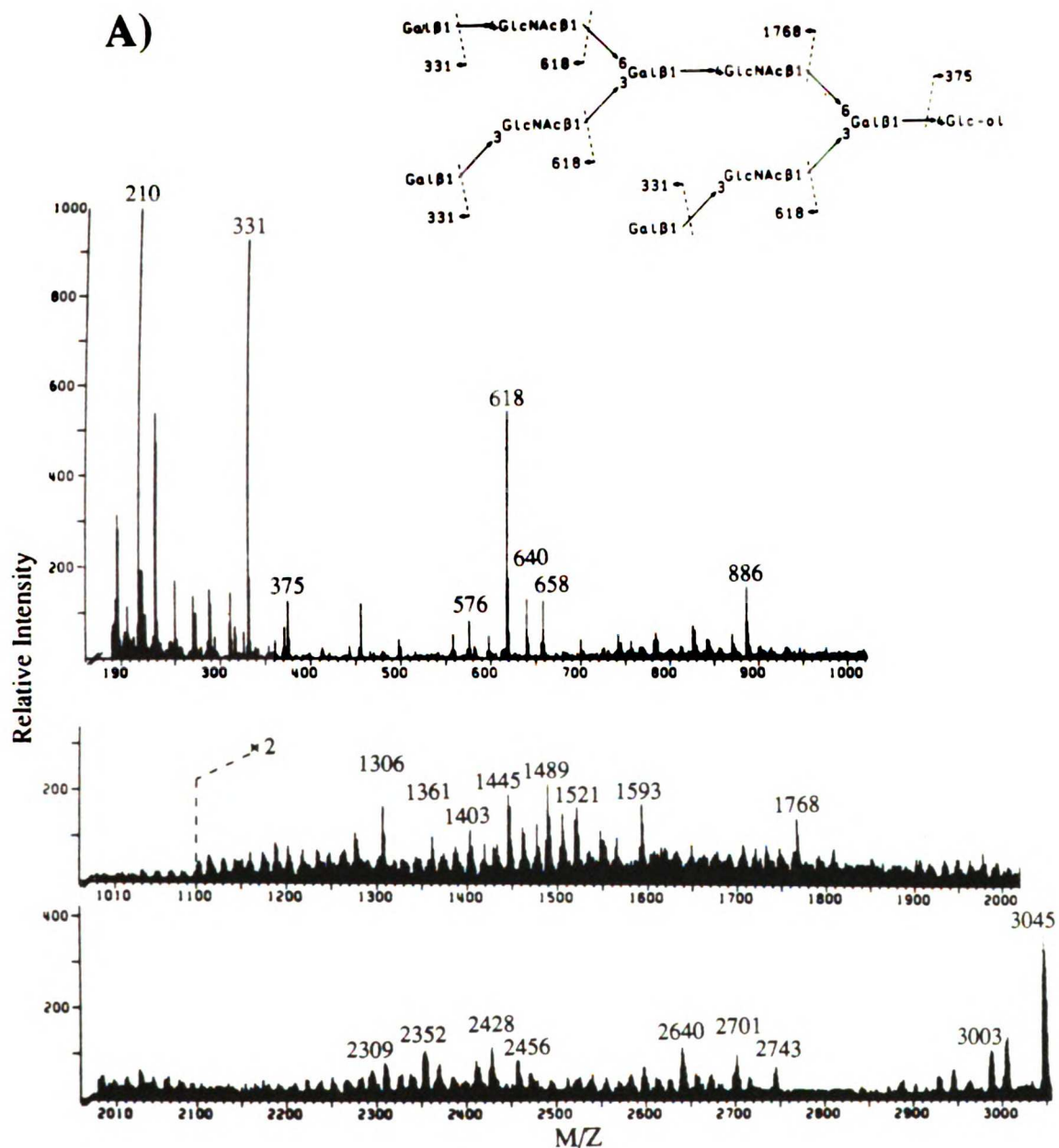


Figure 2. Positive ion FAB mass spectra of the human milk oligosaccharide, lacto-N-decaose. (A) Reduced and peracetylated spectrum run in thioglycerol matrix. $(M+Na)^+ = m/z$ 3045. Adapted with reprint permission from (Egge and Peter-Katalinic, 1987).

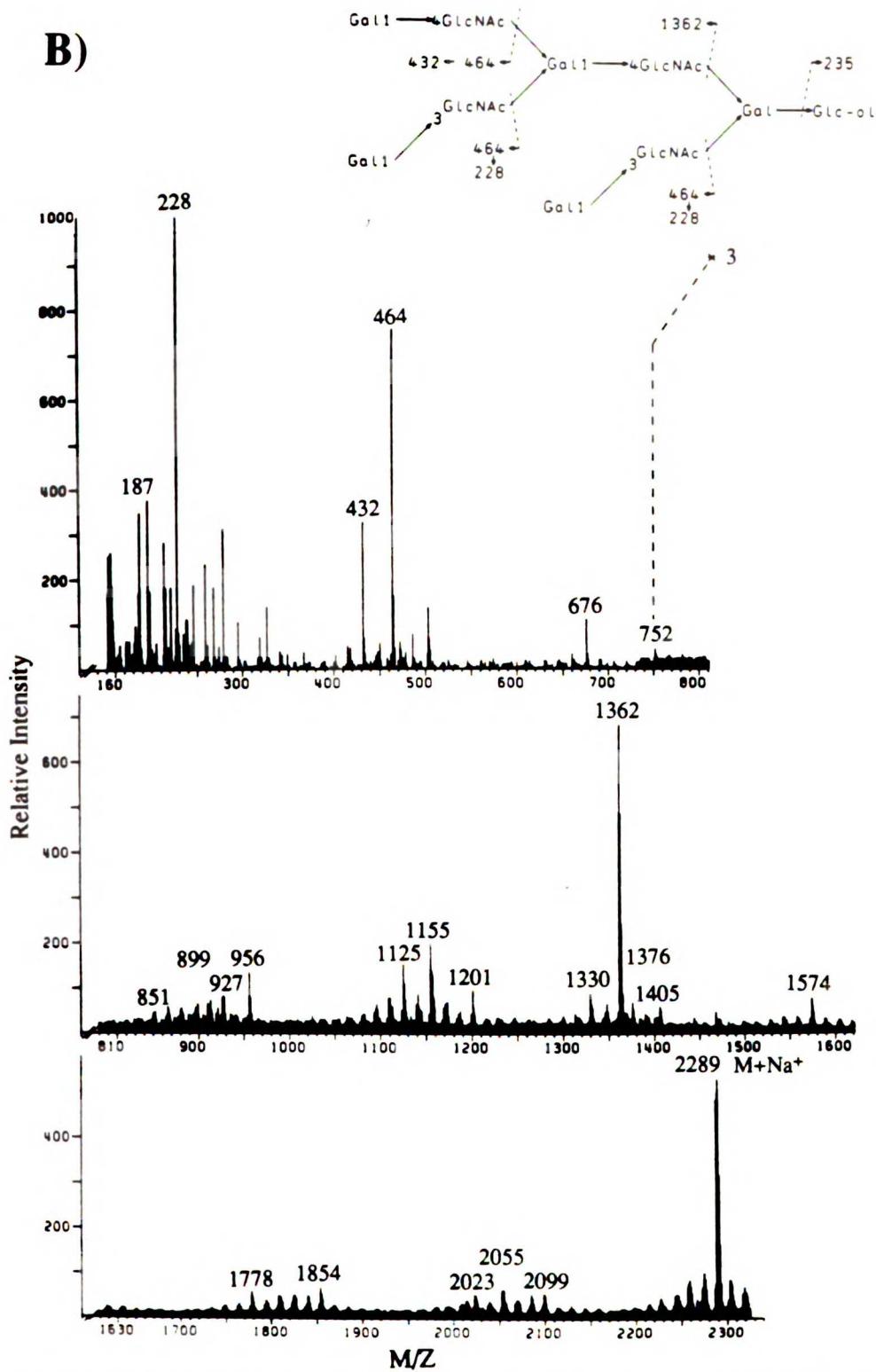


Figure 2 (continued). (B) Reduced and permethylated spectrum, run in thioglycerol matrix with the probe tip precoated with sodium acetate. ($M+Na^+$) = m/z 2289. Adapted with reprint permission from (Egge and Peter-Katalinic, 1987).

sodiated molecular ions at $(M+23)^+$. The peracetylated spectrum is characterized by loss of ketene (-42) from the sequence ions (eg. m/z 289 and 331; m/z 576 and 618). The peracetylated spectrum A) shows the presence of acetylated Gal-GlcNAc branches (m/z 618) and terminal hexose residues (m/z 331). The absence of a tetrasaccharide ion at m/z 1193 and the abundant ion at m/z 1768 indicate the presence of a branched hexasaccharide structure as shown on the 6-linked arm. However, no information can be obtained from this spectrum concerning 1-3 or 1-4 linkages of the terminal galactose residues. The presence of the other intense ions in this spectrum in the region of m/z 1300-1600 and m/z 2300-2750 indicates that this sample was probably not a pure sample containing a single oligosaccharide species. This is a common problem in the structural elucidation of glycoconjugates.

The permethylated spectrum B) contains the analogous permethylated fragment ions as well as additional linkage-specific ions. The Gal-GlcNAc disaccharide units in this structure give rise to an ion at m/z 464. This ion is also accompanied by two fragment ions at m/z 432 and 228. The ions at m/z 228 result from the characteristic preferential elimination of substituents in the 3-position of HexNAc residues, indicating a Gal1→3GlcNAc linkage, while the ion at m/z 432 indicates a Gal1→4GlcNAc linkage.

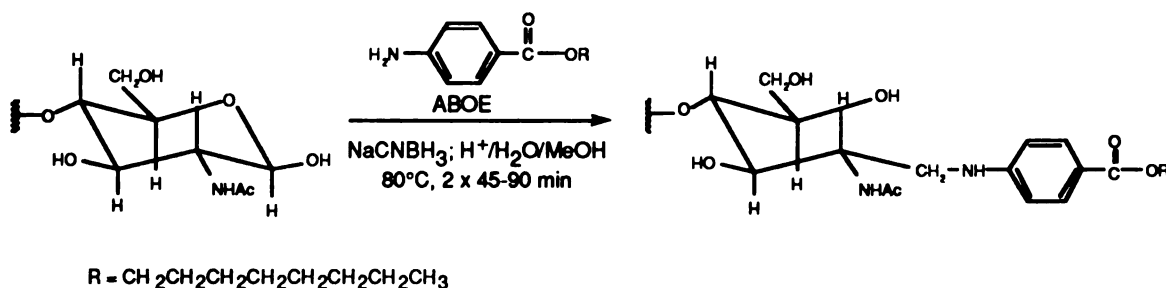
Nevertheless, permethylation and peracetylation chemistries both suffer from the problem of incomplete derivatization (depending on the exact reaction conditions and the sample), which results in a heterogeneous sample, thereby dividing the mass spectrometric signal among several components rather than one, decreasing absolute sensitivity.

2.2.1.5 Reductive Amination

Other commonly used derivatives of oligosaccharides released by PNGase F or Endo H are those formed by reductive amination with hydrophobic chromophores (Wang *et al.*, 1984;

Poulter and Burlingame, 1990; Poulter *et al.*, 1991). This method involves coupling a hydrophobic chromophore to the free reducing terminus of an oligosaccharide via formation of a Schiff base and subsequent reduction to the secondary amine, as shown in Scheme 3 for derivatization with *p*-aminobenzoic acid octyl ester. The procedure for oligosaccharide derivatization with the derivative *p*-aminobenzoic acid octyl ester (ABOE) (Webb *et al.*, 1988; Kaur *et al.*, 1994) or ethyl ester (ABEE), is as follows: Approximately 1-20 mg oligosaccharide is dissolved in 40 ml water in a silylated glass Reacti-vial. In another glass vial, the ABOE or ABEE (0.1 mmol), NaCNBH₄ (35mg), glacial acetic acid (41 ml), and methanol (350 ml) is mixed separately to form the reagent mixture. Forty microliters of the reagent mixture is added to the oligosaccharide solution, and the volume of the total mixture is made up to 200 ml with methanol. The vial is sealed, vortexed, and heated at 80°C for 1 hr. After cooling, 1 ml water and 1 ml of chloroform are added and the chloroform layer extracted two more times with water. The aqueous phases are combined and lyophilized for HPLC analysis.

The chromophoric portion of the derivative is UV absorbing, and permits detection of such derivatized oligosaccharides using conventional HPLC detectors. It has the advantage over permethylation or peracetylation of forming a single product, which adds a small, finite mass to the oligosaccharides when derivatized, yielding better sensitivity.



Scheme 3. Procedure for oligosaccharide derivatization by reductive amination with the chromophoric derivative *p*-aminobenzoic acid octyl ester.

This type of derivatization also allows reversed phase HPLC separation of the derivatized oligosaccharides prior to LSIMS analysis. This is a significant advantage in LSIMS analysis because competition among components in a mixture for the surface layer sputtered using this technique often causes "suppression" of less surface active components, preventing or substantially decreasing their ionization/ejection (Naylor *et al.*, 1986; Poulter *et al.*, 1991). By separating the components first, this problem may be partially circumvented. Another interesting advantage to this derivative is that positive ion analysis of the derivatized oligosaccharides yields primarily intact MH^+ ions, while negative ion analysis yields fragment ions, giving sequence and branching information (Hernandez *et al.*, 1989). Figure 3 illustrates the type of information obtained in a negative ion LSIMS spectrum for an ABOE-derivatized oligosaccharide. This spectrum shows major fragment ions (Y-ions)

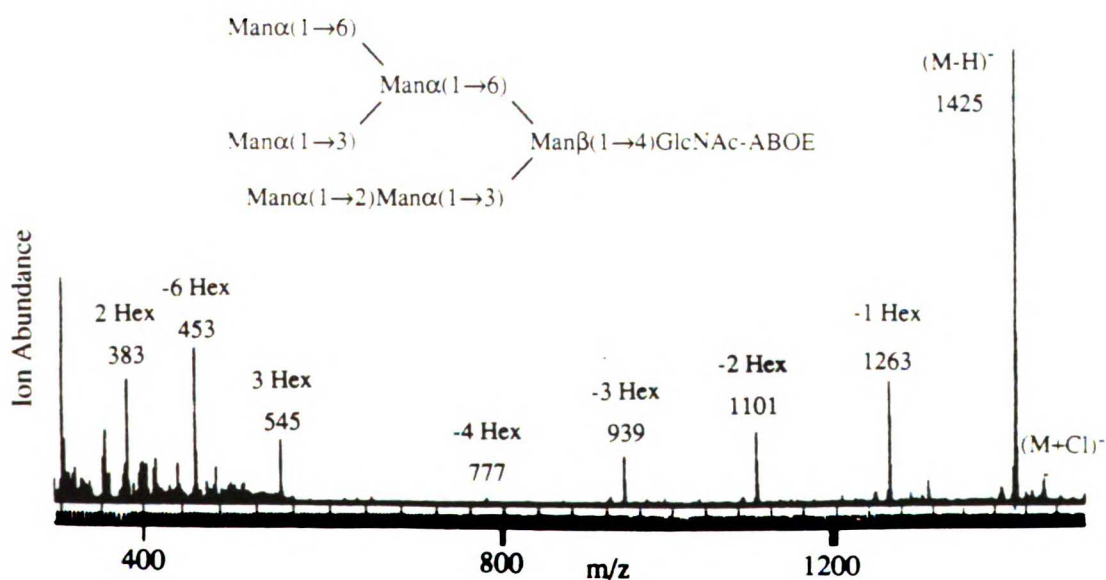


Figure 3. Negative ion LSIMS spectrum of $Man_6GlcNAc-ABOE$, $(M-H)^- = m/z$ 1425. The major Y-type fragment ions show losses of 1, 2, 3 and 6 hexose residues support the structure shown above the spectrum, where no single bond cleavage could result in a loss of 4 or 5 hexose residues. Reprinted with permission from (Hernandez *et al.*, 1989).

corresponding to losses of 1, 2, 3 and 6 hexose residues which support the structure shown above the spectrum, where no single bond cleavage could result in a loss of 4 or 5 hexose residues. The B-ions at m/z 383 and 545 (2 and 3 hexose residues, respectively), also confirm this structure. Other examples of these derivatives are described in Chapter 5. In addition, these derivatized oligosaccharides are substrates for glycosidases, and could therefore be used in a sequencing strategy involving sequential exoglycosidase digestion followed by MALDI-TOFMS or HPLC/ESIMS analysis. These points will be discussed more in section 2.2.3. Several different chromophoric derivatives have been coupled to oligosaccharides, including 2-aminopyridine (Hase *et al.*, 1984), aniline (Wang *et al.*, 1984), and alkyl esters of *p*-aminobenzoic acid (Poulter *et al.*, 1991).

2.2.2 Identification at Specific Attachment Sites

As an alternative to the removal of oligosaccharides from the protein or peptide prior to analysis, more recently developed mass spectrometric techniques have enabled the direct, reproducible analysis of underivatized oligosaccharides while still attached to peptides (i.e. as glycopeptides) for the first time. While glycopeptide mixtures have been successfully analyzed directly by ESIMS as well as MALDI-MS, the ease of interfacing reversed phase HPLC to ESI mass spectrometers (HPLC/ESIMS) is especially well-suited for glycopeptide analysis because HPLC/ESIMS permits separation of glycopeptides from other peptides and glycopeptides in a proteolytic digest followed by direct mass analysis of the separated components in the same experiment. This has the advantages which accrue from less sample handling and greatly reduced analysis time. MALDI also suffers from the problem of metastable cleavage of sialic acid residues during the desorption of sialylated glycopeptides, which could result in a misrepresentation of the actual glycopeptide glycoform species present in a mixture. In addition, by leaving the oligosaccharides

attached to the peptides during the analysis, information concerning both the sites of oligosaccharide attachment and the structural heterogeneity is obtained. Numerous examples of glycopeptide analysis by HPLC/ESIMS are discussed in Chapters 4 to 7.

2.2.2.1 Peptide and Glycopeptide Mapping

As described above, after reduction and alkylation, the glycoprotein (50 pmole up to several nmole) is proteolytically cleaved by a protease such as trypsin. The resulting digest mixture may be analyzed directly by capillary or microbore HPLC/ESIMS, to generate a "peptide mass map" of the entire glycoprotein. While only a 25-50 pmole aliquot is required for such analysis, injection of 100-500 pmole is recommended (if one has that much sample) in order to assure detection of minor glycoforms which may be present. In addition, when a peptide coelutes with a glycopeptide, the intensity of the peptide ion signals are usually 5-50 times more intense than the glycopeptide ion signals, so it is often easy to overlook them, due to the limited usable dynamic range of the ESI quadrupole mass spectrometer (about 100:1). If the protein sequence is known, the deconvoluted masses (molecular weights calculated from the observed multiply charged ions) (Mann *et al.*, 1989) can be attributed to the expected tryptic peptides by comparing the observed molecular weights with the calculated value for each peptide resulting from the specific proteolytic digest of the glycoprotein. Those masses which do not match predicted peptide masses may be glycopeptides. The digest can then be treated with PNGase F or Endo H and analyzed again using the same HPLC/ESIMS conditions, to look for changes in the chromatographic as well as mass spectrometric data. These methods have been used successfully with HPLC (off-line) followed by LSIMS for many years (Carr and Roberts, 1986). However, the sensitivity of these LSIMS instruments above m/z 3000-5000 requires either larger amounts of sample or subdigestion of very large glycopeptides. As a

result, the use of HPLC coupled directly to an ESI-MS has rapidly become the method of choice for the analysis of glycoprotein digests.

By comparing the HPLC and MS data from the different digests it is possible to identify glycopeptides by looking for differences from one HPLC/ESIMS run to another (Spellman, 1990; Carr *et al.*, 1991; Settineri and Burlingame, 1994). For example, as described earlier, PNGase F cleaves all classes of oligosaccharides attached to asparagine and converts the asparagine to an aspartic acid, causing a mass increase of 1 Da) (Tarentino *et al.*, 1985; Carr and Roberts, 1986). Therefore, masses that appear only after PNGase F treatment correspond to deglycosylated peptides containing an Asp instead of an Asn. Masses that appear only before PNGase F treatment are glycopeptides. In addition, the glycosylation heterogeneity which is usually present at each site can help to identify glycopeptides in the data. The spectra will often show an obvious pattern of peaks with masses differing by (the residue mass of sialic acid, 291/number of charges on the glycopeptide) for differences in sialic acid (N-acetylneuraminic acid, NeuAc), (365/number of charges) for differences in hexosyl-N-acetylhexosamines (Hex-HexNAc) or (656/number of charges) for differences sialyl hexosyl-N-acetylhexosamines (NeuAc-Hex-HexNAc), as shown in the electrospray spectrum of a mixture of glycopeptides containing amino acid residues 54-85 from bovine fetuin in Figure 4. This spectrum contains three glycopeptide glycoforms which differ in sialic acid and sialyl hexosyl-N-acetylhexosamine content. Therefore, the three ions at m/z 1196, 1327.1 and 1385.3 which contain five charges are separated by 131.1 and 58.2 Da, corresponding to differences of NeuAc-Hex-HexNAc (656/5) and NeuAc (291/5), respectively.

The use of HPLC combined with ESIMS also permits the introduction of "alternative" HPLC mobile phases to the typically used acetonitrile/water/TFA (which are still compatible with electrospray ionization) in order enhance chromatographic resolution of glycopeptide glycoforms. For example, Medzihradzky and coworkers have recently reported the use of an ethanol/propanol/water/formic acid HPLC mobile phase system

combined with ESIMS which resulted in increased resolution of the sialylated glycopeptide glycoforms from bovine fetuin (Medzihradzsky *et al.*, 1994). As shown in Figure 5, where all of the glycopeptide glycoforms for each of the three *N*-linked glycosylation sites eluted in one fraction using the acetonitrile/water/TFA mobile phases, using the ethanol/propanol/water/formic acid system, the sialylated glycoforms of the carbohydrates attached to Asn¹³⁸ and Asn⁸¹ were either completely or partially separated. This improved chromatographic separation of the glycopeptides during the HPLC/ESIMS experiment resulted in the identification of a new biantennary structure at Asn⁸¹ (Medzihradzsky *et al.*, 1994).

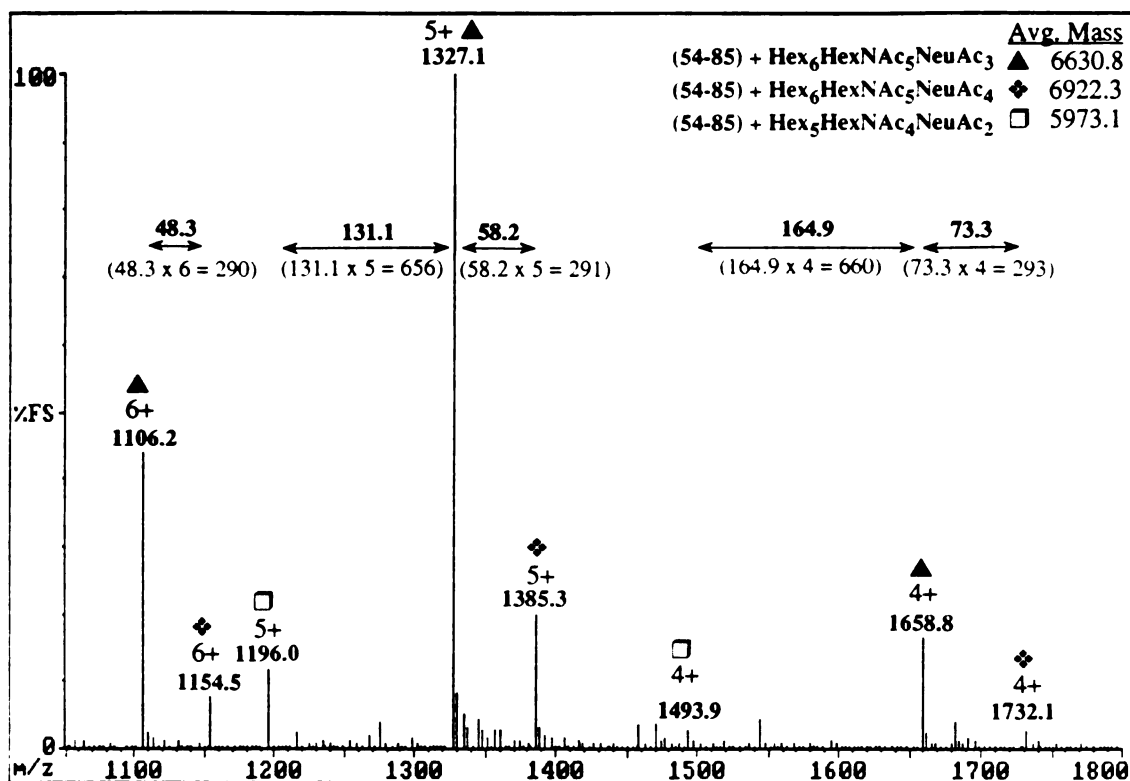


Figure 4. Positive ion electrospray mass spectrum of a mixture of three glycopeptides (two sialylated triantennary complex and one sialylated biantennary complex) all containing amino acid residues (54-85) from bovine fetuin. The arrows illustrate peaks with masses differing by (291/number of charges on the glycopeptide) for differences in *N*-acetylneuraminic acid (NeuAc) as well as (656/number of charges) for differences in *N*-acetylneuraminyl-hexosyl-*N*-acetylhexosamines (NeuAc-Hex-HexNAc).

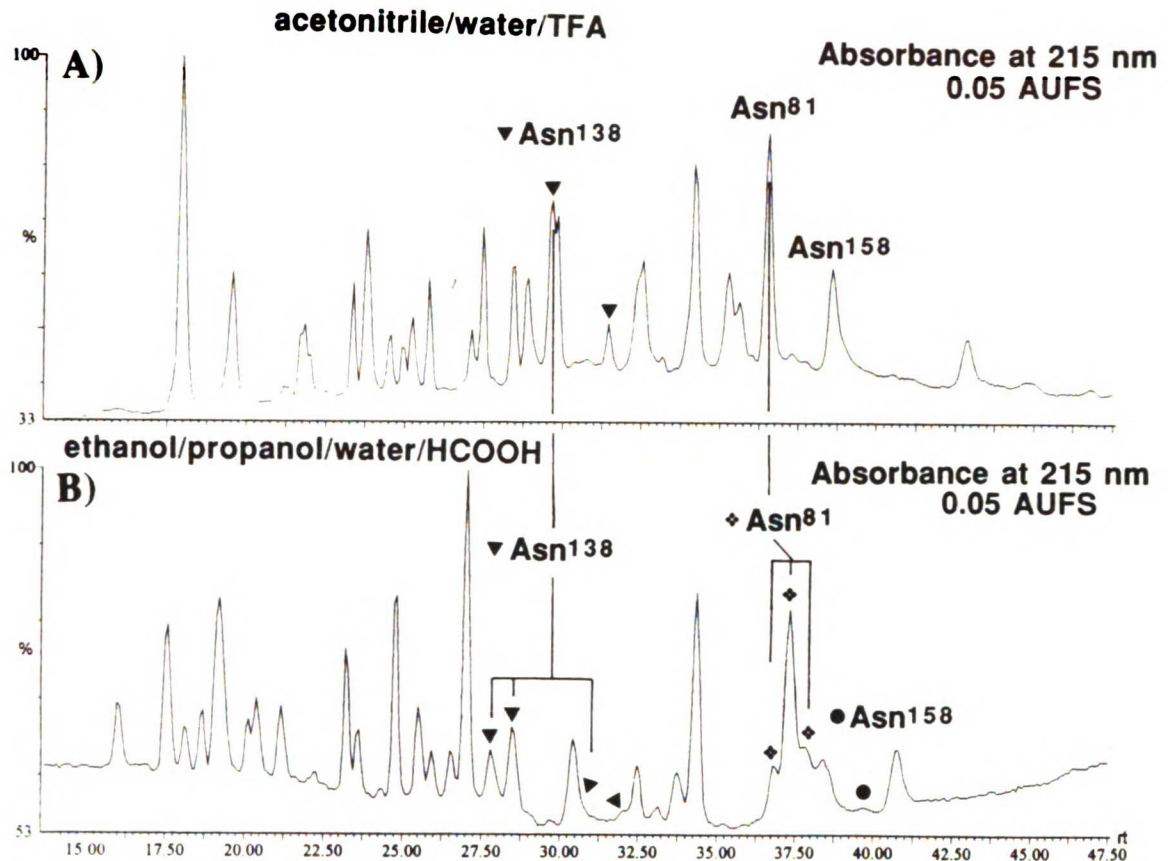


Figure 5. HPLC chromatograms of the tryptic digest of bovine fetuin comparing the two different mobile phase systems. Panel A) illustrates the UV absorbance chromatogram for the separation of a tryptic digest of bovine fetuin using water/0.1% TFA as the A mobile phase and acetonitrile/0.08% TFA as the B mobile phase. Panel B) illustrates the UV absorbance chromatogram for the separation of the fetuin tryptic digest using water/0.1% formic acid as the A mobile phase and ethanol/propanol (5:2)/0.08% formic acid as the B mobile phase. Peaks corresponding to glycopeptides are labeled. Reprinted with permission of Elsevier Science, Inc. from (Medzihradzsky *et al.*, 1994).

2.2.2.2 Glycopeptide Identification using collision induced dissociation and selected ion monitoring

Another complimentary method which has recently been developed for identifying glycopeptides in mixtures also takes advantage of collisional activation during HPLC/ESI-MS analysis. Fragmentation is induced as the HPLC eluant enters the mass spectrometer, by CID. The dissociation of the molecular ions can be performed either in

the electrospray source, by increasing the potential difference between the sampling orifice and the skimmer (HPLC/ESI/CID/MS) or by achieving collisions within the second quadrupole (Q2) of a triple quadrupole mass spectrometer (HPLC/ESI/MS/MS) (Carr, 1993 #19). The first method is the more sensitive method while the latter is more specific, since only precursor ions which decompose to yield the fragments of interest are detected. Diagnostic sugar oxonium-ion fragments (B-ions) are formed from collisionally excited glycopeptides such as m/z 204 for HexNAc, m/z 292 for NeuAc and m/z 366 for Hex-HexNAc. By monitoring these ions which are specific for carbohydrate fragments (selected ion monitoring, SIM) (Conboy and Henion, 1992; Carr *et al.*, 1993), it is possible to readily identify glycopeptide-containing fractions in even the most complex mixtures of peptides and glycopeptides. Several groups have published their own versions of this technique (Huddleston, 1993 #41; Carr, 1993 #19; Duffin, 1992; Medzihradszky, 1994 #51), and have demonstrated its sensitivity and selectivity to detect glycopeptides during liquid chromatographic separation of glycoprotein digests. An example of this type of experiment is shown in Figure 6, where m/z 204 and m/z 292 were selectively monitored during an HPLC/ESIMS analysis of a bovine fetuin tryptic digest. This figure clearly illustrates which fractions in the UV and total ion current (TIC) traces contain glycopeptides, as indicated by the darkened and striped peaks. As with older SIM techniques, this method is extremely sensitive and selective, and many fragment ions can be monitored in a single experiment, yielding considerable information of importance about the types of carbohydrate structures present.

The unique masses of potential glycopeptides identified from a protease digest and a protease followed by a PNGase F digest are compared and combined with the SIM data to determine the mass of the oligosaccharides attached to the glycopeptides. By combining the masses thus determined with further information from the enzyme specificities, one can routinely deduce the composition and structural class of the oligosaccharides from this type

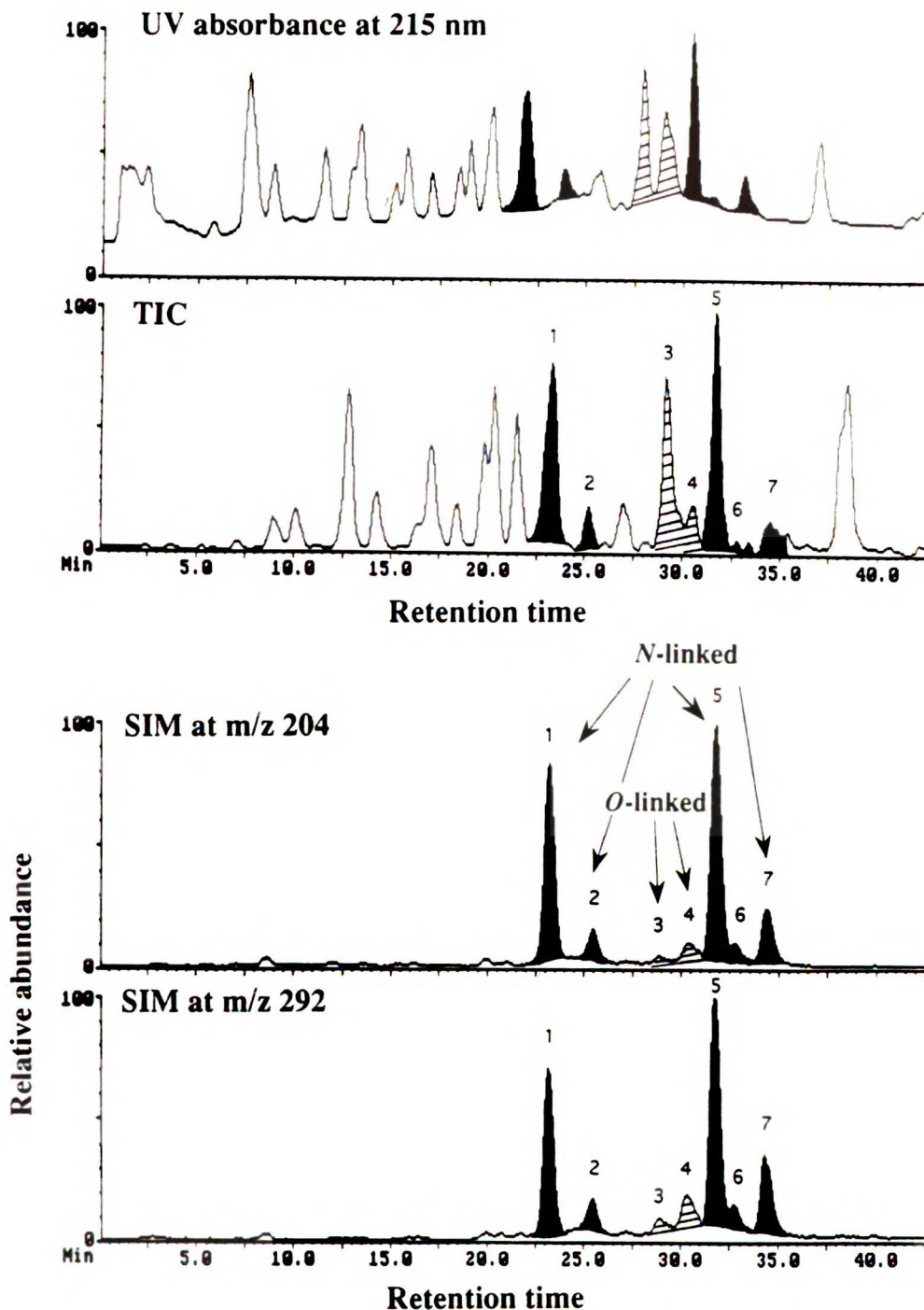


Figure 6. UV Absorbance, TIC and SIM (m/z 204 and m/z 292) chromatograms from an HPLC/ESIMS analysis of a bovine fetuin tryptic digest, using SIM to identify fractions containing glycopeptides. The darkened peaks indicate fractions identified to contain *N*-linked glycopeptides, and the striped peaks indicate fractions identified to contain *O*-linked glycopeptides. (HPLC solvents A=0.1% TFA, B=0.08% TFA in ACN. Elution program was 2-40% A in 40 min. Range=0.05 AUFS).

of mass spectrometric data. This is very straightforward when the sequence of the protein is known, and is slightly more difficult if the sequence is not known, since one cannot correlate the masses observed with the expected specific peptide sequences. If the protein sequence is not known, some additional method must be used to determine the sequence of the peptide part of the glycopeptides by collecting the identified glycopeptide-containing fractions and treating them with PNGase F. The deglycosylated peptides can then be sequenced by MS/MS (Hunt *et al.*, 1986; Biemann and Scoble, 1987) or Edman sequencing.

2.2.3 Determination of the Sequence, Branching and Linkages

Most detailed sequence and branching information obtained on *N*-linked oligosaccharides by mass spectrometry is performed on derivatized oligosaccharides. Originally, derivatization was performed to increase thermal stability in order to permit vaporization and analysis by electron impact and chemical ionization (Radford and De Jongh, 1980; Reinhold and Carr, 1983). More recently, LSIMS has been used extensively to analyze derivatized as well as some underivatized oligosaccharides, particularly negatively-charged oligosaccharides (Egge and Peter-Katalinic, 1987), with considerable success at the picomole to nanomole level. Derivatization, which enhances hydrophobicity, increases the surface activity and thus the absolute sensitivity of analysis by LSIMS, and it provides a center for protonation or deprotonation. In addition, derivatization helps to remove salts and other impurities which are often present in oligosaccharide preparations. As discussed earlier, derivatization may be achieved by alkylation (Hakamori, 1964) or esterification (Dell, 1990) of hydroxyl and acetamido groups or reductive amination of the reducing terminus (Wang *et al.*, 1984; Poulter *et al.*, 1991).

2.2.3.1 Collision Induced Dissociation of Oligosaccharides

While the mass spectra of derivatized *N*-linked oligosaccharides can provide monosaccharide composition from the intact molecular ions as well as some sequence and branching information through the fragment ions observed, in order to generate additional fragmentation such as ring cleavages and eliminate chemical noise and fragments from salt adducts, collision-induced dissociation (CID) or tandem mass spectrometry (MS/MS) is required. This is a process whereby a small amount of the kinetic energy of a mass-selected ion is transformed into vibronic energy upon collision with a gas such as helium for the purpose of inducing unimolecular decompositions, i.e. fragmentation (see section 2.2.4). This method often yields complimentary nonreducing terminal fragments (termed A, B and C) not present in LSIMS spectra, along with more reducing terminal fragments (termed X, Y and Z) both of which are illustrated in Table I and Figure 1 (Domon and Costello, 1988) (see section 2.2.1). In addition MS/MS has the advantage of reducing or eliminating the chemical noise arising from the sputtering of the matrix solvent which is present in LSIMS spectra. In the negative ion mode, CID of the alkyl *p*-aminobenzoate derivatives have been shown to confirm the fragmentation achieved in negative ion LSIMS (Hernandez *et al.*, 1989), eliminate or minimize suspected double cleavage ions and distinguish high mannose branched isomers (Gillece-Castro and Burlingame, 1990; Kaur *et al.*, 1990). This is illustrated in the negative ion CID spectrum of the ABOE derivative of Man₉GlcNAc from yeast mannan shown in Figure 7. The Y₄, Y₃ and Y₂-ions (at *m/z* 1749, 1587 and 1263) correspond to losses of Man₁, Man₂ and Man₄, respectively. These fragments result from single bond cleavages of this oligosaccharide. The weaker ions at *m/z* 1425 and 1101 correspond to cleavage of Man₃ and Man₅, respectively. The fact that these two ions are much weaker than the other ions indicates that they result from multiple bond cleavages. This information is used to distinguish different branched isomers that have the same mass. The A-ions at *m/z* 221, 383 and 707 correspond to similar

nonreducing terminal fragments containing Man₁, Man₂ and Man₄, respectively, and confirm the data obtained from the Y-ions.

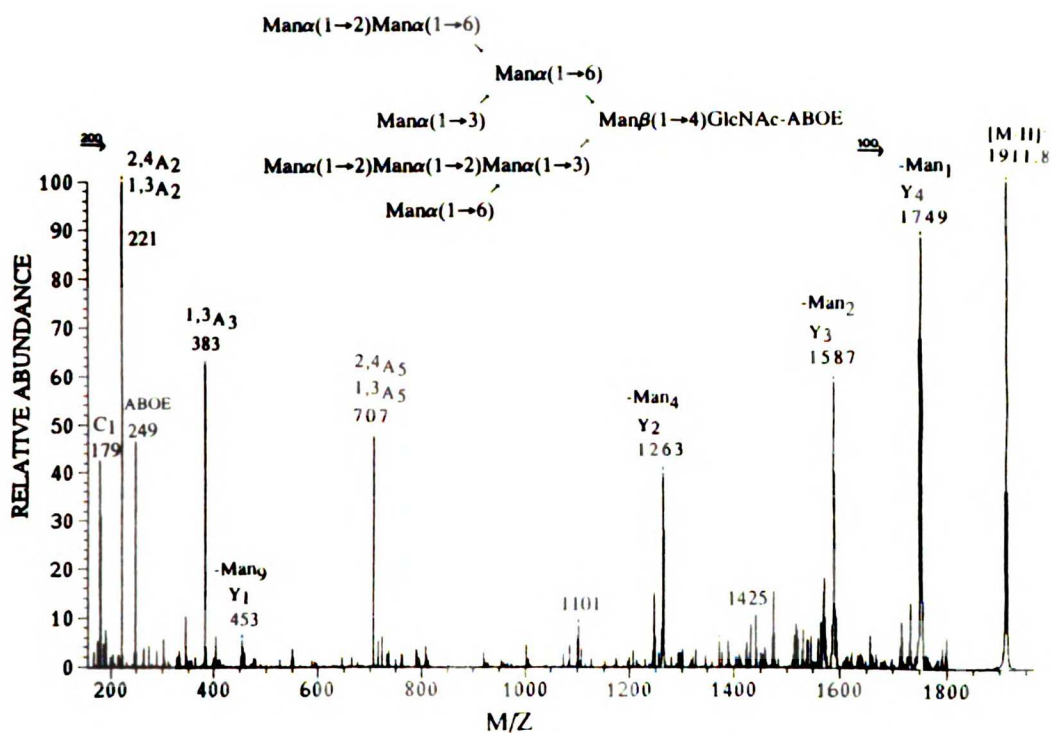


Figure 7. Negative ion CID spectrum of the of *p*-aminobenzoic acid octyl ester (ABOE) derivative Man₉GlcNAc from yeast mannan. Reprinted from (Kaur *et al.*, 1990).

2.2.3.2 Collision Induced Dissociation of Glycopeptides

In the past two years, CID has also been performed by electrospray ionization mass spectrometry (ESIMS) on the double charged species of underivatized *N*-linked

oligosaccharides (Duffin *et al.*, 1992) as well as *N*-linked glycopeptides (Conboy and Henion, 1992) by continuous infusion of purified oligosaccharides and glycopeptides, respectively. With the oligosaccharides, negative ion MS/MS gives sequence information for the negatively charged sialylated oligosaccharides, while positive ion MS/MS yields sequence information for neutral molecules. Both reducing and nonreducing terminal ions are present in these spectra as well. Henion and coworkers have obtained CID spectra of asialo *N*-linked glycopeptide standards in the positive ion mode which yielded reducing terminal fragment ions corresponding to the sequential losses of each monosaccharide residue of the oligosaccharide portion of the glycopeptide, yielding a series of Y-ions, as well as two non-reducing terminal B-ions, as shown in Figure 8.

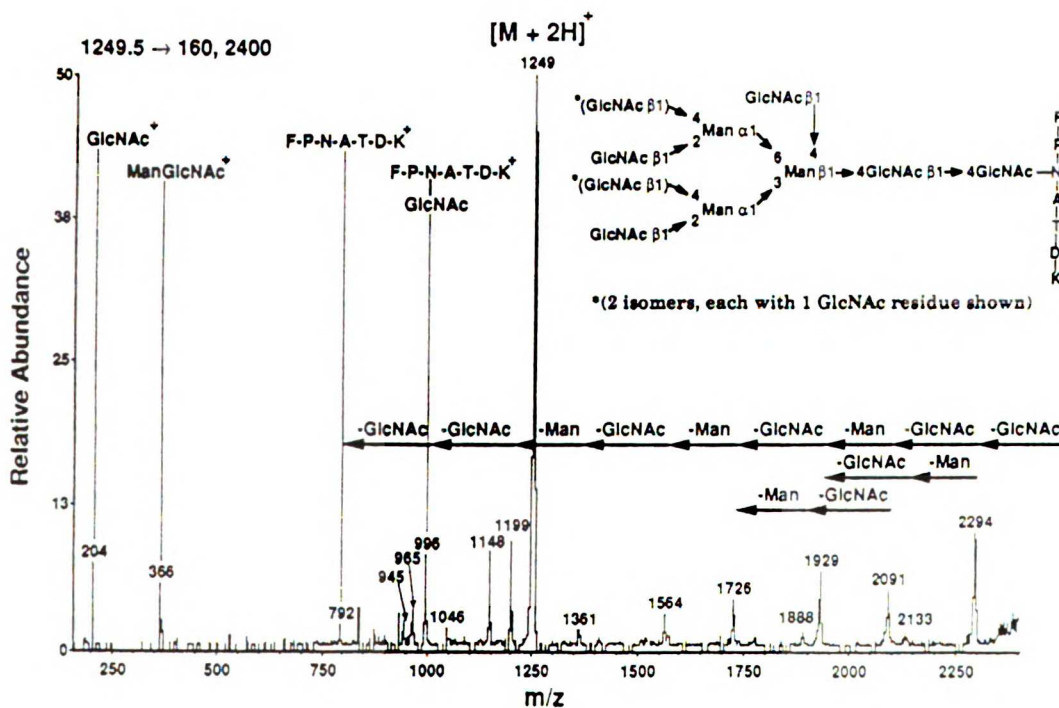


Figure 8. Positive ion ESIMS/MS spectrum of the doubly charged precursor ion at m/z 1249.5 for a complex *N*-linked glycopeptide from ovomucoid, with the structure shown in the inset. Reprinted with permission of Elsevier Science, Inc. from (Conboy and Henion, 1992).

Within the last year, the potential for using MALDI-TOF for sequencing the oligosaccharide components of glycopeptides was demonstrated by obtaining low-energy metastable spectra from stepping the potential of a reflectron attached to the instrument (Huberty *et al.*, 1993). The reflectron is a reflecting electrostatic lens at the end of the flight tube, and the potential is stepped by applying different voltages to these lenses (see section 2.2.6) (Huberty *et al.*, 1993). Using this method, sequence data were obtained on femtomole quantities of sialylated biantennary and triantennary complex *N*-linked glycopeptides, however problems were encountered with metastable cleavage of the sialic acid residues from the sialylated glycopeptides (Huberty *et al.*, 1993). Figure 9 shows an example of the type of data obtained with this method using positive ion MALDI-TOF in the reflector mode to obtain sequence information on an *N*-linked glycopeptide from a Lys-C digest of recombinant human macrophage colony stimulating factor (rhM-CSF) containing residues (119-125) and a disialo biantennary complex sugar structure. As Figure 9 illustrates, sequential glycosidic bond cleavages of each monosaccharide residue with charge retention on the reducing terminal portion of the oligosaccharide (or peptide), a series of Y-ions, were obtained. The inset of the structure with the predicted Y-ion masses shows that as is typical with MALDI-TOF spectra in this mass range, the mass accuracy of the ions obtained is only within 1-3 Da. Another drawback of this method is that the reflectron mass spectra of sialylated glycopeptides exhibit a base peak in the positive ion mode corresponding to the loss of all neuraminic acid residues, and the molecular masses observed for these metastable ions must be corrected on the basis of energy loss due to fragmentation (Huberty *et al.*, 1993). This is the only report to date for sequencing glycopeptides with MALDI-TOF, so it is unclear how widespread this method will become, considering the low mass accuracy compared to CID with ESI-MS and the complications with sialylated glycopeptides.

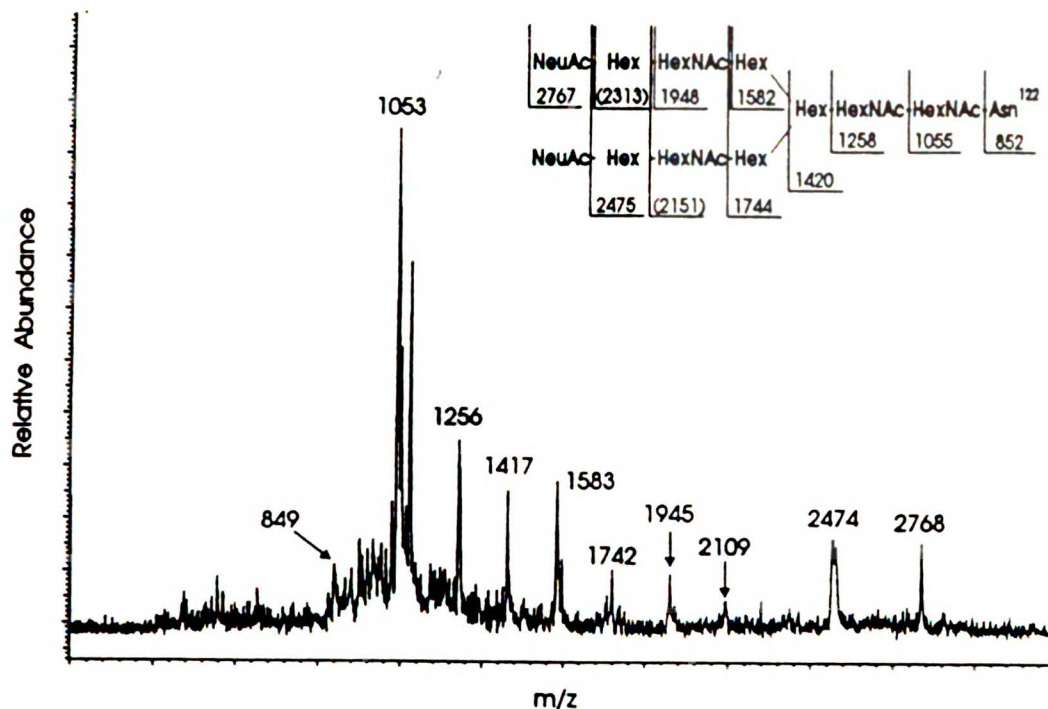
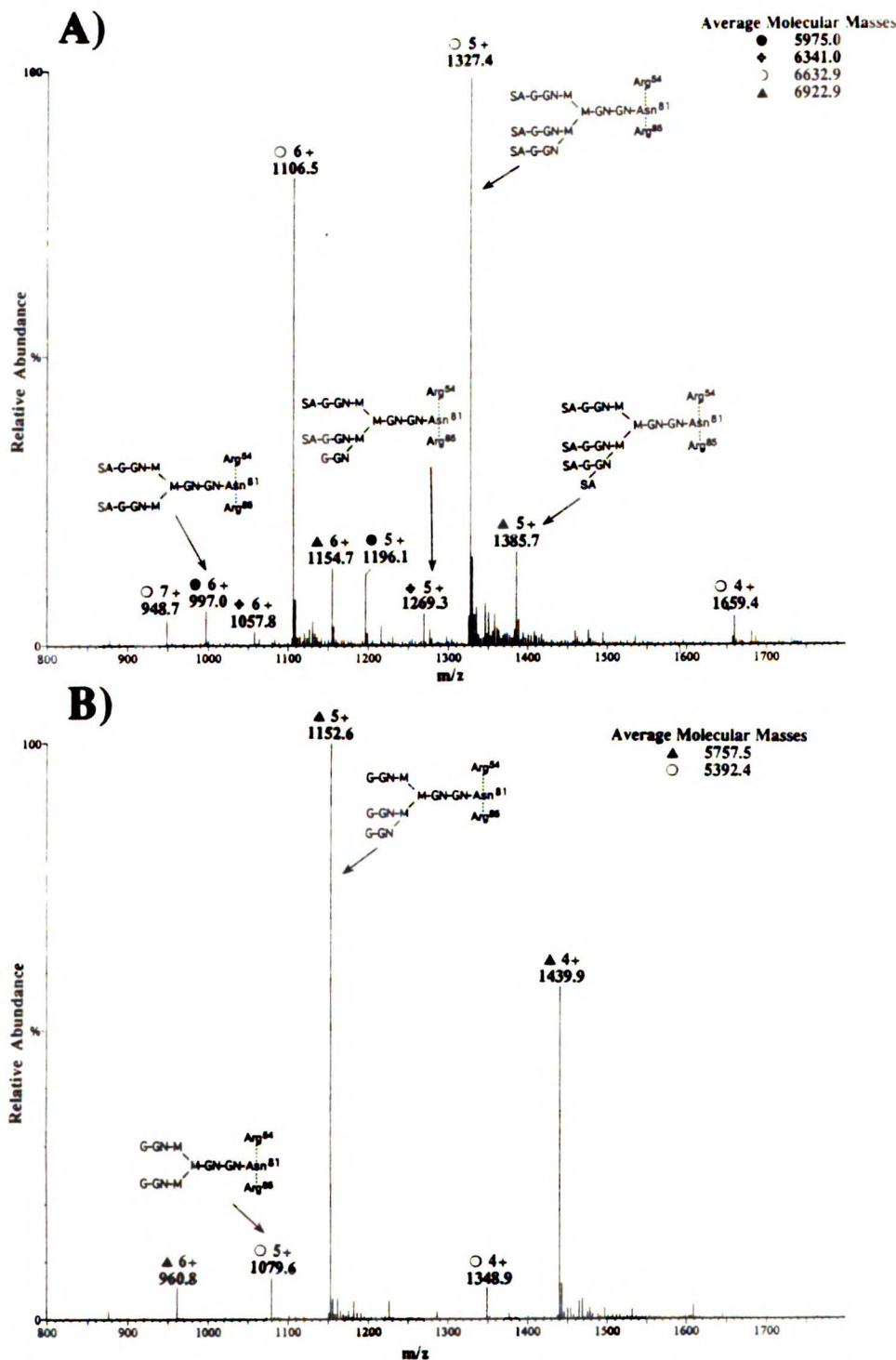


Figure 9. MALDI-TOF positive ion reflector mode mass spectrum obtained from less than 1 picomole of an rhM-CSF glycopeptide NVFN¹²²ETK which contains a disialobiantennary complex sugar structure. The fragment ions are assigned in the inserted structure. Reprinted with permission of the American Chemical Society from (Huberty *et al.*, 1993).

2.2.3.3 Exoglycosidase Sequencing of Glycopeptides

Sequence information as well as some branching and linkage information can be obtained on *N*-linked oligosaccharides while still attached to peptides by HPLC/ESIMS (Settineri and Burlingame, 1993; Settineri and Burlingame, 1994) or MALDI-TOF (Sutton *et al.*, 1994) analysis of glycopeptides before and after treatment with specific exoglycosidases

such as neuraminidase, β -galactosidase, and N-acetylhexosaminidase (Kobata, 1979). By determining the change in mass of a glycopeptide after treatment with one or more glycosidases, and combining this information with the specificity of the enzymes used, oligosaccharide sequence information (as well as some branching and linkage information) can be obtained (Settineri and Burlingame, 1993; Sutton *et al.*, 1994). For example, Figure 10 illustrates a one step exoglycosidase digestion of a mixture of glycopeptide glycoforms from residues (54-85) of bovine fetuin. Spectrum A) is the ESIMS spectrum obtained from a single fraction which eluted during an HPLC/ESIMS analysis of 50 pmole of a fetuin tryptic digest. The molecular masses obtained from deconvolution of the molecular ions in this spectrum correspond to glycopeptides containing disialo biantennary and di-, tri- and tetrasialo triantennary oligosaccharides attached to the peptide containing residues 54-85 from fetuin. In order to confirm the presence of these structures, several sequential digestions with different specific exoglycosidases were performed, and the result of each digestion step was purified and analyzed by HPLC/ESIMS. Spectrum B) shows the result of the first digestion with a neuraminidase from *Arthrobacter ureafaciens*, starting with approximately 800 pmole of isolated glycopeptide. The resulting molecular masses at 5757.5 Da and 5392.4 Da confirmed the presence of both biantennary and triantennary complex structures attached to Asn⁸¹ of the peptide sequence 54-85, due to the complete removal of sialic acid residues from the different glycopeptide glycoforms by the neuraminidase. Further digestions with β -galactosidase (*Streptococcus pneumoniae*), β -N-acetylhexosaminidase (Chicken liver) and then α -mannosidase (Jack bean) using the glycopeptide fraction collected from each previous digestion and HPLC/ESIMS purification/analysis confirmed the original assumptions about the type of oligosaccharide structures present on this particular peptide. In addition, by using these specific enzymes sequentially, the sequences of these oligosaccharides were established to be those shown in Figure 10, because each enzyme can only remove nonreducing terminal monosaccharide units at the end of each branch in the structure.



UCSF LIBRARY

Figure 10. A) Electrospray mass spectrum of the bovine fetuin tryptic glycopeptides containing Asn⁸¹ and disialo biantennary and di-, tri- and tetrasialo triantennary complex oligosaccharides. B) Electrospray mass spectrum of the glycopeptide fraction from spectrum A) treated with neuraminidase from *Arthrobacter ureafaciens*. (GN=N-acetylglucosamine; M=mannose; G=galactose; SA=sialic acid). Reprinted with permission of Elsevier Science, Inc. from (Medzihradszky *et al.*, 1994).

can only remove nonreducing terminal monosaccharide units at the end of each branch in the structure.

More of these purified exoglycosidases, some of which are linkage as well as monosaccharide-specific, are becoming available all the time, making this a more useful technique for determining sequence, branching and some linkages of oligosaccharides. Nevertheless, obtaining complete linkage information for an *N*-linked oligosaccharide by mass spectrometry requires some type of chemical manipulation of at least several nanomoles of carbohydrate, whether one performs the most commonly used methylation analysis (Björndal *et al.*, 1967; Hellerqvist *et al.*, 1968; Hellerqvist, 1990), the reductive cleavage method (Rolf and Gray, 1982), or periodate oxidation (Angel *et al.*, 1987; Angel and Nilsson, 1990).

2.2.3.4 Methylation Linkage Analysis

Methylation analysis involves methylation of all free hydroxyl groups in an oligosaccharide, followed by hydrolysis of the glycosidic bonds, reduction with NaBD₄, and acetylation of the hydrolyzed linkage points (Hellerqvist, 1990). Therefore, the hydroxyls that are acetylated are those that were either involved in the linkages of the monosaccharides in the oligosaccharide or linked to the oxygen within the sugar ring (C-5 in six carbon sugars) before hydrolysis. The resulting partially methylated alditol acetates are then separated by GC and analyzed by EI-MS and/or CI-MS, and identification of residues and linkages is based on relative retention times of standards combined with molecular weight and fragmentation information. As with alditol acetates, reduction with NaBD₄ is performed in order to distinguish fragments containing C-1 from those containing C-6. Fragmentation of partially methylated alditol acetates in electron ionization is well understood (Kochetkov and Chizov, 1965; Björndal *et al.*, 1967), yielding readily

UCSF LIBRARY

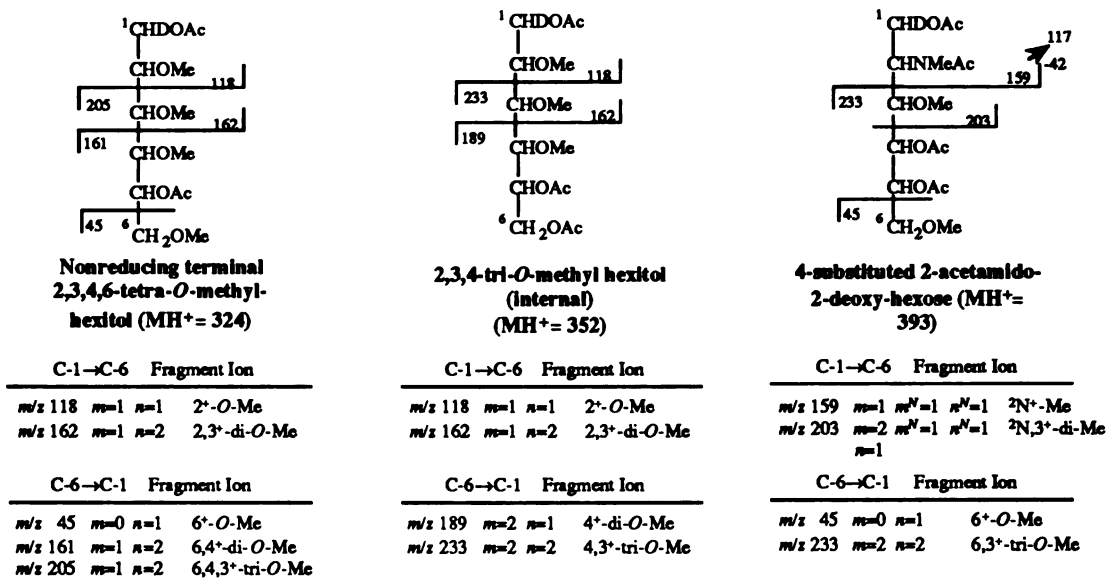
interpretable mass spectra containing specific diagnostic ions for identification of the linkages of alditol acetates. For example, in EI, charge is preferentially retained on the methoxylated carbons, and fragment ions of decreasing abundance are formed by cleavage of the alditol chain between:

1. an (N-acetyl-N-methyl) aminated carbon and a methoxylated carbon with charge retention predominantly on the aminylated carbon, which then loses ketene;
2. two methoxylated carbons, with charge retention on either carbon;
3. a methoxylated and an acetylated carbon with charge retention exclusively on the methoxylated carbon.

The mass of a given fragment ion formed from C-1→C-6 equals $(n \times 72) + (m \times 44) + 2$ and from C-6→C-1 equals $(n \times 72) + (m \times 44) + 1$, where n is the number of acetylated carbons and m is the number of methylated carbons in the fragment ion. Scheme 4 illustrates the fragmentation of partially methylated alditol acetates of a nonreducing terminal hexose, a 1,6-linked internal hexose and a 1,4-linked 2-acetamido-2-deoxy-hexosyl residue. Also included for each of these is a table showing the determination of the structures of the fragment ions using the mass and the equation above. For the amino-deoxy alditol acetates, m^N and n^M refer to N-acetyl and N-methyl, respectively. 2N refers to a 2-amino-2-deoxy function and "+" refers to the site of charge in the alditol chain.

In order to draw firm conclusions about branches and nonreducing terminal residues, the stoichiometry should be internally consistent. For example, a reduced carbohydrate found to contain x nonreducing residues should contain one reducing end and x branch points. Each branch point is represented by one less *O*-methyl group in a residue than found when the same residue is present as an internal chain derivative. The procedure for preparing the sample for methylation analysis by permethylating the sample and then forming the partially methylated alditol acetates is described as follows:

Partially methylated alditol acetates are prepared by dissolving 0.1-100 mg of a permethylated sample in a glass vial in 1 ml 4 M TFA and the sealed mixture is heated at



Scheme 4. Illustration of fragmentation observed in EI and/or CI analysis of partially methylated alditol acetates of a nonreducing terminal hexose, a 1,6-linked internal hexose and a 1,4-linked 2-acetamido-2-deoxy-hexosyl residue, and corresponding tables showing the determination of the structures of the fragment ions using the mass and the equation above.

Acetylation is accomplished by adding 0.5 ml freshly distilled pyridine and 1 ml acetic anhydride to the sample, and the flask stoppered and heated at 100°C for 30 min. Then the pyridine and acetic anhydride are co-distilled with 5 ml toluene to dryness at 40°C, adding more toluene if necessary. The sample is then extracted into CHCl₃ or MeOH or purified through an ion exchange resin and dried down.

2.2.3.5 Reductive Cleavage Linkage Analysis

Linkage analysis using the reductive cleavage method (Rolf and Gray, 1982; Gray, 1990) is a newer method which is based on methylation analysis, but is different with regard to the types of fragments analyzed. This method was developed because methylation analysis

does not distinguish between 4-linked aldopyranosides and 5-linked aldofuranosides. The products of this method are partially methylated anhydroalditols, and are characterized by GC/MS of their acetates using standards, as with methylation analysis. Fragmentation patterns for these derivatives have not been established, but EI spectra do yield diagnostic fragment ions which can be used to identify the residues and linkages. The procedure for reductive cleavage is described as follows: Reductive cleavage is performed with triethylsilane as the reducing agent and with either trimethylsilyltrifluoromethane sulfonate (TMSOTf) or a mixture of trimethylsilylmethane sulfonate (TMSOMs) and boron trifluoride etherate ($\text{BF}_3 \cdot \text{Et}_2\text{O}$) as the catalyst. Briefly, a 5mg sample of methylated oligosaccharide is added to a silanized vial, kept under vacuum for 2 hr, and 0.25 ml of CHCl_2 , 5 Eq Et_3SiH , 5 Eq TMSOMs, and 1 Eq $\text{BF}_3 \cdot \text{Et}_2\text{O}$ are sequentially added. The teflon-capped vial is stirred for 24 hrs at room temperature, 1 ml methanol is added, stirred for 30 more min, and the mixture is then deionized through a 0.5 x 5 cm Bio-Rad AG501-X8 column. The methanol and dichloromethane are removed under vacuum, and the product acetylated by treating with 5 Eq each of acetic anhydride and 1-methylimidazole in 0.2 ml dichloromethane for 30 min. The reaction is terminated by addition of 0.5 ml saturated, aqueous NaHCO_3 , and the CHCl_2 layer washed twice with 1 ml portions of water.

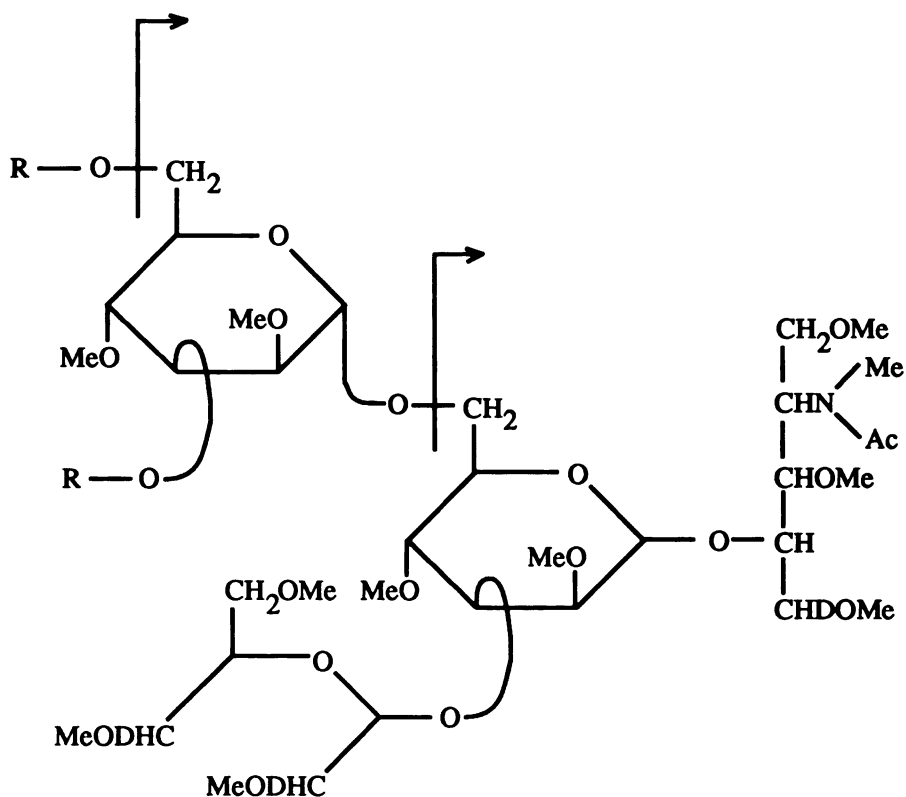
2.2.3.6 Periodate Oxidation Linkage Analysis

Linkage analysis using periodate oxidation of glycoconjugates followed by permethylation and LSIMS (Angel *et al.*, 1987; Angel and Nilsson, 1990) is the third derivatization method developed for linkage analysis by mass spectrometry. Periodate oxidation is a well-known carbohydrate reaction where sugar residues with vicinal hydroxyl groups are cleaved by periodate, resulting in the formation of a dialdehyde (Goldstein *et al.*, 1965).

UCSF LIBRARY

Only vicinal glycols are cleaved by this reaction and the presence of these are dependent on the position of substitution and linkages between monosaccharide residues. After oxidation, reduction and permethylation are carried out and the product analyzed directly by LSIMS or ESIMS, the linkage positions are determined from the diagnostic sequence ions in the spectra. The procedure for periodate oxidation is as follows: Typically 0.1-2.0 mg of oligosaccharides is mixed with 5 ml of 0.1 M acetate buffer, pH 5.5, containing 8 mM sodium periodate, and the reaction takes place in the dark at 4°C for 48 hr. The pH is then adjusted to 7.0 with 0.1 M NaOH for the reduction step, and 25 mg NaBD₄ is added and the sample is kept at 4°C overnight. The pH is then adjusted to 4.5 with acetic acid, and after concentration to dryness, boric acid is removed by repeated evaporations with about 3 ml methanol. The sample is then desalted on a (20 x 1 cm) BioGel P-2 column and eluted with water. The permethylation reaction is then carried out as described above.

In analysis by LSIMS, the sequence ions are formed from the nonreducing terminus by ionization (protonation) of the ring oxygen followed by cleavage of the glycosidic bonds, and specific fragments indicate the structures present. *N*-linked oligosaccharide structures containing internal HexNAc residues give especially intense sequence ions during analysis by LSIMS. For high mannose structures, two and four-substitution can be distinguished, but ions localizing the branches of the 3,6-disubstituted Man residues are often of low abundance in LSIMS. However, these ions are more prevalent in EI, so the same sample (up to Man₉) can be analyzed as the periodate oxidized, permethylated derivative by EI-MS to see these branch specific ions which are formed by cleavage of the 1→6 linkage with charge retention at C-6 of the 3,6-disubstituted Man residue (Angel *et al.*, 1991), as shown in Scheme 5.



Scheme 5. Illustration of branch specific ions which are formed by cleavage of the 1→6 linkage with charge retention at C-6 of the 3,6-disubstituted Man residue. D=deuterium, Me=methyl and Ac=acetyl.

2.3 Serine/Threonine-linked glycosylation

The characterization of *O*-linked oligosaccharides from glycoproteins is similar in many ways to *N*-linked characterization, with some key differences. Firstly, there is no structural class distinction for *O*-linked carbohydrates as with the high mannose, hybrid and complex *N*-linked structures. Secondly, while there is a tendency for *O*-linked glycosylation to occur in regions containing prolines, serines and threonines (Wilson *et al.*, 1991; Pisano *et al.*, 1993), there is no consensus sequence to alert one to the potential presence of *O*-linked structures as is the case with *N*-linked structures. Thirdly, there is no

universal enzyme to remove all *O*-linked oligosaccharides from glycoproteins as with PNGase F for *N*-linked glycans.

2.3.1 Chemical Release

Therefore, to isolate free *O*-linked oligosaccharides from a glycoprotein, the protein is usually treated with base in the presence of a reducing agent, which converts the reducing sugar to an alditol which is stable to β -elimination (Carlson, 1966). Typical reaction conditions contain 200 ml 1 M NaBH₄ in 0.5 M KOH for 12-16 hr at 45°C (Marshall degradation) (Carlson, 1966). It has recently been demonstrated that hydrazinolysis, under the proper conditions, will remove intact *O*-linked oligosaccharides from glycoproteins as well (Patel *et al.*, 1993).

2.3.2 Derivatization

Once the oligosaccharides are removed from the protein, they are commonly permethylated or peracetylated in a manner analogous to the *N*-linked structures described above and analyzed by LSIMS or GC-MS. However, for determination of site specific information, the analytical challenge is significantly greater. Not only is it not possible to predict which serine or threonine residues may be glycosylated, but often *O*-linked oligosaccharides are found clustered at multiple sites in the same region of a glycoprotein, where multiple serine and threonine residues are present which could be glycosylated. In addition, the glycosylation at each site is usually heterogeneous, making analysis even more complicated. However, as described in Chapters 3, 4 and 7 of this thesis, progress has been made with the analysis of intact *O*-linked glycopeptides by mass spectrometry.

2.3.3 Identification at Specific Attachment Sites

As described earlier for the case of *N*-linked glycopeptides, the detection of glycopeptide-containing fractions in glycoprotein digests by HPLC/ESIMS has the advantages of minimal sample handling and analysis time, because no derivatization step is required and an entire glycoprotein proteolytic digest can be analyzed in about an hour using an HPLC connected directly to an ESI mass spectrometer. As with *N*-linked analysis, the reduced and alkylated glycoprotein (50 pmole up to several nanomoles) is digested with a protease such as trypsin or endoproteinase glu-C. An aliquot of the digest, typically 100-300 pmole, is then analyzed by HPLC/ESIMS and a peptide map is generated. Thus, any molecular weights which do not match predicted proteolytically-generated peptides are possible glycopeptides. In order to help identify glycopeptides in the HPLC/ESIMS data, this information is combined with HPLC/ESI/CID/MS of the same digests with selected ion monitoring (SIM) of carbohydrate-specific ions such as *m/z* 204, *m/z* 366 and *m/z* 292 as described previously for *N*-linked glycopeptide analysis.

In order to simplify analysis (sample amount permitting), any *N*-linked glycopeptides present in the digest mixture can first be converted to Asp-peptides with PNGase F (0.1 mU/ml, 16 hr at 37°C) in the protease digestion buffer prior to analysis, and the cleaved *N*-linked oligosaccharides will elute in the void volume during HPLC/ESIMS analysis. This could result in better dynamic range for visualizing the *O*-linked glycopeptides in the SIM analysis, because SIM data for *N*-linked glycopeptides is sometimes much more intense, resulting in missed identification of *O*-linked glycopeptides as described in Chapter 4 (Settineri and Burlingame, 1994).

2.3.3.1 Glycopeptide Isolation by Jacalin Agarose

In order to facilitate characterization of the intact sialylated glycopeptides, which are sometimes difficult to identify by ESIMS (due to decreased ionization in positive ion from the negatively charged sialic acid residues) the *O*-linked glycopeptides can be separated from the rest of the digest using a procedure to isolate *O*-linked glycopeptides by affinity chromatography with a lectin from jackfruit (*Artocarpus integrifolia*), jacalin, bound to agarose beads (Hortin and Trimpe, 1990). Jacalin is a protein that is known to have affinity for the core disaccharide, 1- β -galactospyranosyl-3-(α -2-acetamido-2-deoxygalactopyranoside) (Gal β 1 \rightarrow 3 GalNAc), in *O*-linked oligosaccharides of glycoproteins (Sastry *et al.*, 1986; Young *et al.*, 1989), and has also been shown to bind sialylated Gal β 1 \rightarrow 3 GalNAc structures (Hortin, 1990). As described in Chapter 4, a modified procedure was developed so that the procedure could be performed on a very small scale in a 1.5 ml polypropylene microcentrifuge tube, with minimal sample handling and no desalting step, so that the *O*-linked glycopeptide mixture could be injected directly onto the HPLC/ESIMS system. The sialylated *O*-linked glycopeptides are isolated by adding the entire proteolytic digest to a slurry of jacalin-agarose. The peptides and *N*-linked glycopeptides are removed from the mixture after the *O*-linked glycopeptides bind to the jacalin-agarose, and then the *O*-linked glycopeptides are displaced from the jacalin using methyl- α -D-galactopyranoside. After isolation by jacalin agarose affinity chromatography, the *O*-linked glycopeptide mixture can then be analyzed by HPLC/ESI/CID/MS with SIM as described in the next section.

2.3.3.2 Glycopeptide Identification

Since *O*-linked oligosaccharides are often relatively small structures containing 1-6 monosaccharide units, SIM of larger carbohydrate fragments can be used to help identify the size of the structures expected. For example, in certain cases, the full scan electrospray mass data shows no detectable signal where peptides or glycopeptides have clearly eluted, based on the presence of a UV response from the HPLC. However, SIM experiments in which B-type oxonium ions of m/z 657 (to identify HexNAc-Hex-NeuAc structures) and m/z 948 (to identify HexNAc(NeuAc)-Hex-NeuAc structures) were monitored indicated the presence of glycopeptides in those regions where no signal was detected in the original HPLC/ESIMS data (see Chapter 4). The most likely explanation for these results is that the multiply-charged mass(es) of the intact glycopeptide(s) are out of the usable m/z range of 2500 of the quadrupole mass analyzer used. Therefore, in this case, the SIM data was necessary to detect and confirm the presence of mono- and disialylated *O*-linked glycopeptides.

If sialylated structures are suspected (e.g., from carbohydrate composition analysis as discussed earlier), another aliquot of the protease-treated digest can be treated with a neuraminidase such as the enzyme from *Arthrobacter ureafaciens* (0.05 mU/ml, 8-16 hr at 37°C), and reanalyzed by HPLC/ESIMS to see if there are any changes in the masses or retention times of components in the digest. Similarly, an aliquot of the neuraminidase-treated digest can then be incubated with Endo- α -N-acetylgalactosaminidase (*O*-glycosidase, *Diplococcus pneumoniae*, 0.05 mU/ml, 16 hr at 37°C), which specifically cleaves unsubstituted Gal β 1 \rightarrow 3GalNAc structures commonly found on mucins and other glycoproteins (Umemoto *et al.*, 1977), and the digest reanalyzed by HPLC/ESIMS and LC/ESI/CID/MS. If a change in the masses or retention times of components in the digest occurs after treatment with *O*-glycosidase, this yields very specific information about the size and sequence of the *O*-linked carbohydrates attached to the peptides. By combining

the change in mass after O-glycosidase digestion, the number of oligosaccharides on the glycopeptide can be determined. If there are more Ser and Thr residues in the peptide than the number of oligosaccharide attachment sites, the glycopeptide can be digested with an appropriate protease and reanalyzed by HPLC/ESIMS to determine the sites of glycosylation. If this is unsuccessful or not possible (due to the lack of sites of proteolytic cleavage separating possible glycosylation sites), the O-linked glycopeptide can be sequenced by CID (Medzihradzky *et al.*, 1994) or Edman degradation (Schindler *et al.*, 1994) to determine the sites of glycosylation (see section 2.2.2).

2.3.4 Determination of Sequence, Branching and Linkages

All of the strategies described in the sections above concerning structural characterization of N-linked oligosaccharides apply to sequencing and linkage analysis of O-linked oligosaccharides. These include methylation analysis (Hellerqvist, 1990) and other linkage analysis methods described above for determining branching and linkages, sequential enzyme digestion followed by analysis by MALDI-TOF (Sutton *et al.*, 1994) or HPLC/ESIMS (Settineri and Burlingame, 1993; Settineri and Burlingame, 1994), and fragmentation of oligosaccharides using MALDI-TOF instruments equipped with reflectrons for sequencing (Huberty *et al.*, 1993). However, there is an additional powerful mass spectrometric method for sequencing O-linked oligosaccharides which takes advantage of the generally small size of these structures.

2.3.4.1 High Energy Collision Induced Dissociation of Glycopeptides

This method involves ionization by LSIMS followed by high energy collision induced dissociation (CID, MS/MS) of *O*-linked glycopeptides without derivatization of the oligosaccharides (Gillece-Castro and Burlingame, 1990; Medzihradzky *et al.*, 1990; Settineri *et al.*, 1990; Medzihradzky *et al.*, 1994). As described in Chapter 3, *O*-linked glycopeptides containing mono- and disaccharides can be analyzed by MS/MS at the high pmole to low nmole level, yielding abundant fragment ions due to the loss of carbohydrate residues by glycosidic bond rearrangements (Y-ions, see Figure 1 and Table I) (Gillece-Castro and Burlingame, 1990). In addition to carbohydrate fragments, these spectra contain peptide fragments, yielding the peptide sequence, and sometimes the exact site of glycosylation on the peptide. Figure 11 shows a CID spectrum of an *O*-linked glycopeptide (residues 77-92) from an endoproteinase glu-C digest of recombinant human platelet-derived growth factor (rhPDGF), expressed in yeast. This glycopeptide contains one Man residue, but has two possible sites of glycosylation. This spectrum contains a fragment ion at m/z 1871 (labeled as -Man) corresponding to a Y-ion resulting from cleavage of the Man from the glycopeptide. In addition, the difference in mass between the peptide ions a_{11} and a_{12} , listed in the ion table, is (1545-1282 = 263 Da). This mass difference corresponds to a mannosylated threonine residue (101+162 = 263), and indicates that the twelfth residue in the sequence from the amino-terminus, Thr⁸⁸, is glycosylated, and nearby Thr⁹⁰ is not glycosylated.

| | | | | | | | | | | | | | | | |
|---|---|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|
| | Ile-Val-Arg-Lys-Lys-Pro-Ile-Phe-Lys-Lys-Ala-Thr-Val-Thr-Leu-Glu | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | Man |
| a | 86 | 185 | 341 | 469 | 597 | 694 | 807 | 955 | 1083 | 1211 | 1282 | 1545 | 1644 | 1745 | 1858 |
| b | | 213 | 369 | 497 | 625 | | 835 | 983 | 1111 | 1239 | 1310 | 1573 | 1672 | 1773 | 1886 |
| c | | | 386 | 514 | | | 852 | 1000 | 1128 | 1256 | 1327 | 1590 | 1689 | 1790 | 1903 |
| d | | | | 412 | 540 | | 779 | 879 | 1026 | 1154 | | 1367 | 1630 | 1729 | 1816 |

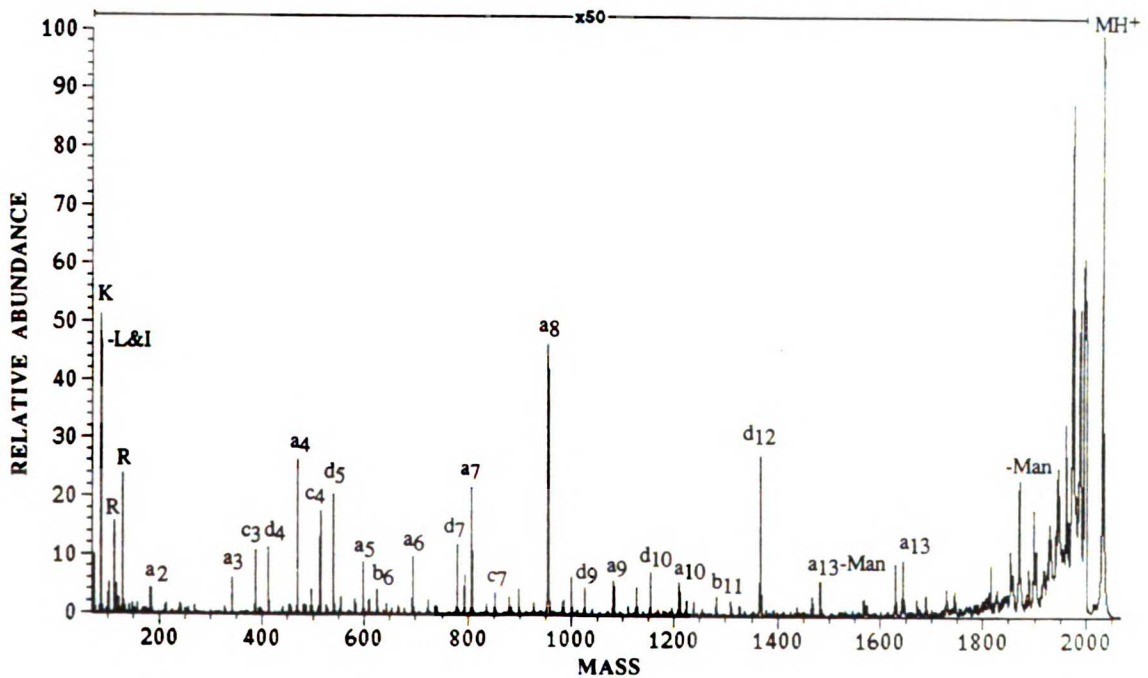
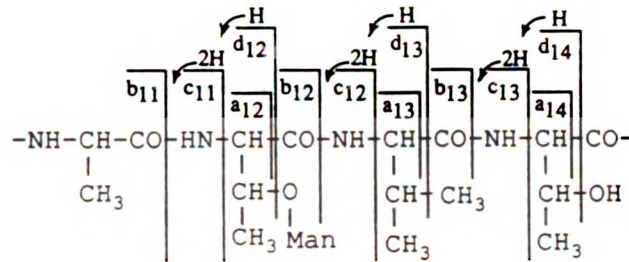


Figure 11. Positive ion LSIMS/MS (CID) spectrum of an *O*-linked glycopeptide isolated from an endoproteinase glu-C digest of recombinant human platelet-derived growth factor (rhPDGF) expressed in yeast, containing residues (77-92), $MH^+=2033.2$. The N-terminal peptide fragment ions belonging to the same ion series are listed in rows below the sequence. The insert represents the N-terminal sequence ions for a portion of the above sequence (-Ala-(Man)Thr-Val-Thr-), illustrating the origin of the a,b, c and d ions and demonstrating the ions which indicate glycosylation of Thr-88 rather than Thr-90. Reprinted by permission of John Wiley & Sons, Ltd. from (Medzihradszky *et al.*, 1990).

CHAPTER 3. CHARACTERIZATION OF THE *O*-LINKED OLIGOSACCHARIDES FROM RECOMBINANT HUMAN PLATELET-DERIVED GROWTH FACTOR EXPRESSED IN YEAST

3.1 Abstract

High energy tandem mass spectrometry was employed to structurally characterize intact *O*-linked glycopeptides and establish the complexity and extent of glycosylation for recombinant human platelet-derived growth factor B chain (rhPDGF-B) expressed in yeast. In addition, liquid secondary ion mass spectrometric (LSIMS) and Edman degradation methods have been employed to verify the protein sequence. LSIMS of HPLC-fractionated proteolytic digests confirmed the complete amino acid sequence predicted by the human PDGF-B gene structure. Potential glycopeptides (as indicated by a mass shift of 162 or 324 Da from the mass of a predicted cleavage product) were sequenced using tandem mass spectrometry and Edman degradation analysis. UV matrix laser desorption MS of recombinant glycoprotein. In addition to the presence of unmodified peptides, corresponding peptides bearing monomannosyl moieties were found on serine-26 and threonines-20, 63, 88, 90 and 101. Further, dimannosyl moieties were found on threonine-6 and 63. These data reveal the presence of *O*-linked glycosylation at sites which do not fortify the concept of a consensus sequence involving proline residues, but which strengthen the concept of secondary and tertiary structure requirements.

UCSF LIBRARY

3.2 Introduction

Human platelet-derived growth factor (PDGF) is a major mitogen in serum which promotes the proliferation of fibroblasts and smooth muscle cells *in vitro*. (Ross *et al.*, 1986). *In vivo*, PDGF is stored in platelet alpha-granules and is released during the platelet release reaction (Scher *et al.*, 1979), which has led to the hypothesis that this protein has a physiological role in wound healing. In addition, the PDGF B-chain precursor is almost identical to p28^{sis}, the transforming protein of simian sarcoma virus (SSV) (Waterfield *et al.*, 1983), and has therefore generated interest in PDGF as having a critical role in neoplasia. PDGF is a cationic glycoprotein dimer of molecular weight range 27,000 to 31,000 Da. The size heterogeneity is thought to result from different extents of glycosylation, processing and the presence of two homologous disulfide-bridged polypeptide chains, A and B. A major percentage of PDGF from human platelets is an A-B heterodimer, but A-A and B-B homodimers exist and have been shown to be biologically active as well (Hart *et al.*, 1990). The literature on the native PDGF derived from mammalian sources indicates the presence of carbohydrate residues (Johnsson *et al.*, 1984).

In this study, the B-chain of PDGF, which is a polypeptide of 109 amino acids (Johnsson *et al.*, 1984), was produced via the alpha-factor secretion pathway in yeast (Brake *et al.*, 1984). The human PDGF-B sequence contains no consensus sites for *N*-linked glycosylation (Asn-Xxx-Ser/Thr), so only *O*-glycosylation could occur in this protein. Therefore, tryptic and endoproteinase glu-C peptide mapping were performed to determine the structure of rhPDGF-B derived from expression in yeast cells. This chapter describes the use of mass spectrometry to characterize the sites and extent of glycosylation on rhPDGF-B.

UCSF LIBRARY

3.3 Materials and Methods

Expression and Purification. Recombinant human PDGF-B dimer (rhPDGF-BB), generously donated by Frank Masiarz and collaborators at Chiron Corporation, was produced via the alpha-factor secretion pathway in *Saccharomyces cerevisiae* (Brake *et al.*, 1984) and was purified from yeast culture medium after centrifugation to remove cells and cell debris. The supernatant liquid was subjected to ion-exchange chromatography on DEAE-Sepharose and S-Sepharose in the first two stages of the purification. After fraction concentration, the material was subjected to hydrophobic interaction chromatography on octyl-Sepharose as a final step. Concentration and buffer exchange of the final product were performed using diafiltration before lyophilization to produce a dry, white powder. Progress of the purification was monitored using reversed phase high performance liquid chromatography (HPLC).

Reduction and Alkylation. Purified rhPDGF-BB was reduced and carboxymethylated at protein concentrations of 1-4 milligrams per milliliter in a buffer containing 0.2 M N-ethylmorpholine acetate (NEMOAc), 6 M guanidine-HCl and 3 mM ethylenediamine-tetraacetic acid (EDTA) (pH 8.5). 2-Mercaptoethanol was used to reduce the disulfide bonds for 60-90 minutes at 60 degrees Centigrade. Carboxymethylation by sodium iodoacetate was performed in the dark for 1 hour at room temperature (Crestfield *et al.*, 1963). The samples were dialyzed into 50 mM NEMOAc (pH 8.5) using a Bethesda Research Labs (Bethesda, MD) microdialysis apparatus with Spectra-Por membrane with a molecular weight cut-off of 3,000 Da. Amino acid analyses were performed on aliquots of the final solutions to obtain yield and characterize the final product.

Proteolytic Digestions. Trypsin digestions were performed using TPCk-trypsin (Serva, Heidelberg, West Germany) at a weight ratio of 5% in 75 mM ammonium bicarbonate buffer (pH 7.8) containing 0.1 mM calcium chloride. Digestions were performed at 37°C for 16 hr. Endoproteinase glu-C (*Staphylococcus aureus* V8)

(Boehringer Mannheim, Indianapolis, IN) digestions were performed in 50 mM ammonium bicarbonate (pH 7.8) at a weight ratio of 4% for 18 hr at 37°C. A fresh aliquot of endoproteinase glu-C (4%) was added to the reaction mixture and the incubation was continued for 8 hr (total 26 hr). Extreme conditions were necessary to obtain complete cleavage as assessed by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of samples at various time points during digestion.

Fractionation of Peptides and Glycopeptides. Reversed phase chromatography was performed on a Waters (Milford, MA) HPLC system using Vydac (The Separations Group, Hesperia, CA) C₁₈ reversed phase columns of 4.6 mm internal diameter and 250 mm length. Gradients were generated using the following solvents: eluant A) 0.1% trifluoroacetic acid (TFA) (w/v) in water and eluant B) 80% acetonitrile (v/v) in solvent A. Enzymatic digests were titrated to pH 2 with 25% (w/v) TFA before injection. Resolution of the tryptic fragments, as shown by the HPLC chromatogram in Figure 1 (18 nmol of PDGF-B monomer), was performed using a gradient from 100% A to 75% B solvent over 1 hr, after an initial isocratic equilibration of 10 min post-injection. For the HPLC chromatogram of the second tryptic digest shown in Figure 2 (30 nmol of monomer), a shallower gradient from 100% A to 50% B solvent over 1 hr was used with the same initial conditions. For the HPLC chromatogram of the endoproteinase glu-C digest shown in Figure 3 (20 nmol of monomer), a gradient from 100% A to 40% B solvent gave the best resolution, with the same initial conditions. Peak detection was achieved at a wavelength of 230 nm.

Mass Spectrometry. LSIMS analysis of tryptic peptides was carried out using a Kratos MS50S double focusing mass spectrometer (Kratos Inc., Manchester, UK) equipped with a high field magnet and a cesium ion source (Aberth *et al.*, 1982) and gun at 10 keV (Falick *et al.*, 1986). Concentrated HPLC fractions (1-2 µl) were added to 1 µl of liquid matrix (1:1 thioglycerol:glycerol with 1% TFA) for analysis. MS/MS experiments were carried out using a Kratos Concept IHH four sector EBEB tandem mass

spectrometer (Walls *et al.*, 1990), equipped with a cesium ion source and a multichannel array detector that allows the simultaneous detection of ions over a 4% mass window (Cottrell and Evans, 1987; Walls *et al.*, 1990). Tandem spectra were acquired while operating at a resolution of approximately 2000. The collision cell was floated at 2 keV and filled with helium so that the parent ion intensity was decreased to one third of its initial value (Walls *et al.*, 1990). Again, 1-2 μ l of sample was used with the same matrix as above or with meta-nitrobenzyl alcohol. CID spectra were interpreted without the aid of computer analysis.

Edman Degradation. Edman degradations were performed using an Applied Biosystems 470A protein sequencer (Applied Biosystems Inc., Foster City, CA) with an on-line 120A phenylthiohydantoin (PTH) amino acid analyzer. Standard programs and reagents were used during these analyses (Hunkapiller *et al.*, 1983). Amino acid analyses were performed using the phenylisothiocyanate derivatization method (Bidlingmeyer *et al.*, 1984) after acid hydrolysis in 6 N HCl at 110°C for 24 hr.

Carbohydrate Composition Analysis. Intact PDGF-BB was hydrolyzed for carbohydrate analysis in 2 M TFA for 4 hr at 100°C. Carbohydrate analysis was performed using a Dionex (Sunnyvale, CA) AI-450 Model II high performance anion exchange (HPAE) chromatography system with pulsed amperometric detection (PAD) (Hardy *et al.*, 1988). Eluant A was 200 mM NaOH and fluent B was 100% water. The program used was 8% A, isocratic for 23 minutes. Identity of the type of hexose was determined by co-injection of a mannose standard with the PDGF-BB hydrolysis sample.

3.4 Results

The first step in the process of confirming proper expression of a given recombinant protein is to verify the amino acid sequence (Gibson and Biemann, 1984). Edman

degradation analysis of the intact rhPDGF-BB revealed the presence of two different sequences (data not shown). The predominant species (65%) possessed the expected N-terminus beginning with serine. The minor species (35%) began with Thr³³ and appears to

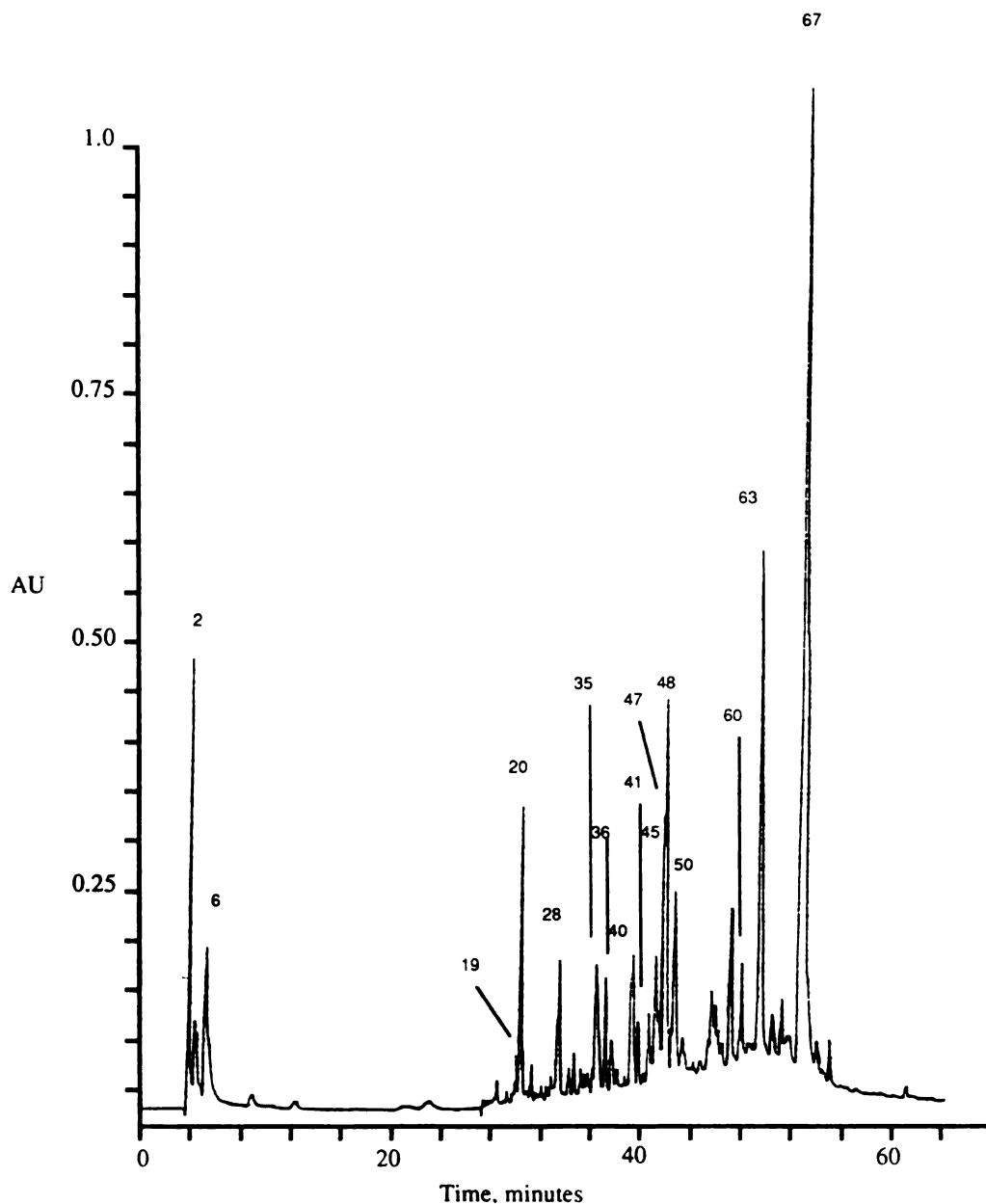


Figure 1. HPLC chromatogram of tryptic digest (T_1) of 18 nmole of rhPDGF-B monomer. Solvent A=0.1% TFA and B=80% ACN in solvent A. The HPLC gradient was 100% A for 10 min, then 100% A to 75% B in 60 min. Wavelength=230 nm.

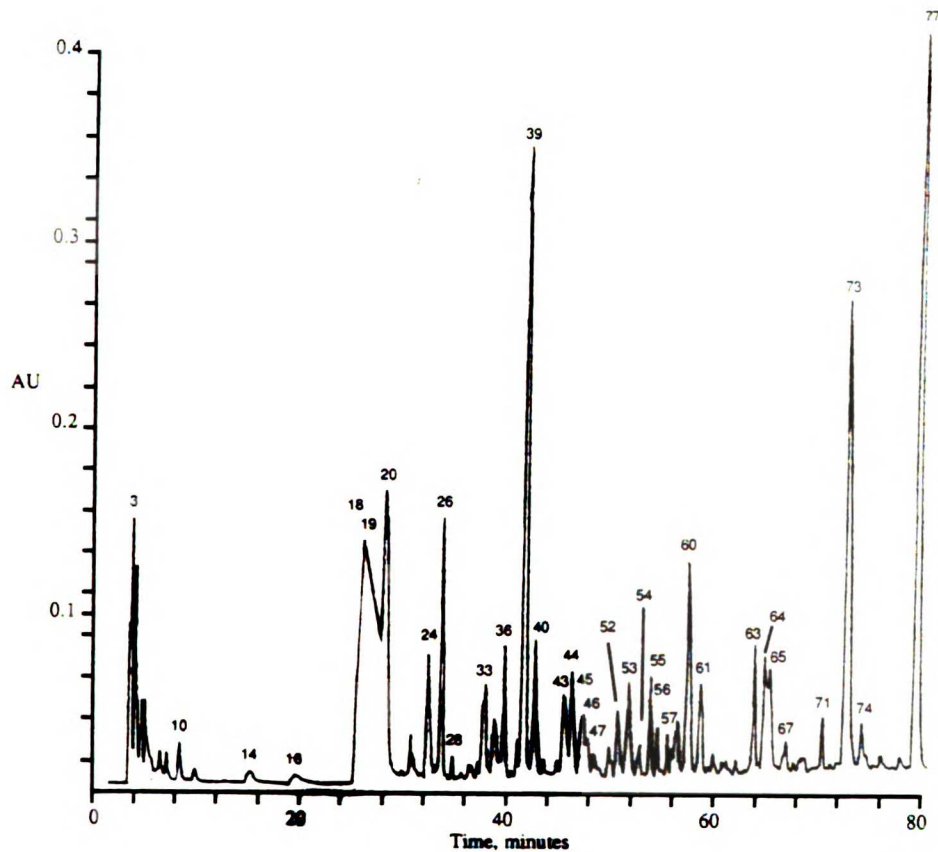


Figure 2. HPLC chromatogram of tryptic digest (T_2) of 30 nmole of rhPDGF-B monomer. Solvent A=0.1% TFA and B=80% ACN in solvent A. The HPLC gradient was 100% A for 10 min, then 100% A to 50% B in 60 min. Wavelength=230 nm.

be the product of a trypsin-like activity that has been observed in all preparations of rhPDGF-BB purified from yeast. This observation was supported by the identification of two peptides found in the endoproteinase glu-C (*Staphylococcus aureus* V8) digest (V-31, residues 25-32, ISRRLIDR, and V-67, residues 33-45, TNANFLVWPPCVE), which begin (V-67) and end (V-31) at a tryptic cleavage site, arginine-32.

To verify the correct sequence, rhPDGF-BB was reduced and carboxymethylated before digestion with trypsin or endoproteinase glu-C (*Staphylococcus aureus* V8) and the

UCSF LIBRARY

peptides were separated by reversed phase HPLC. The HPLC chromatograms for the two tryptic digests (denoted T₁ and T₂), and the endoproteinase glu-C digest (denoted V) are shown in Figures 1, 2 and 3, respectively. The second tryptic digest (T₂) was performed on a rhPDGF-BB preparation (denoted rhPDGF-BB-2) which migrated at a slightly higher molecular weight on SDS-PAGE than the two other preparations, indicating that it might contain more

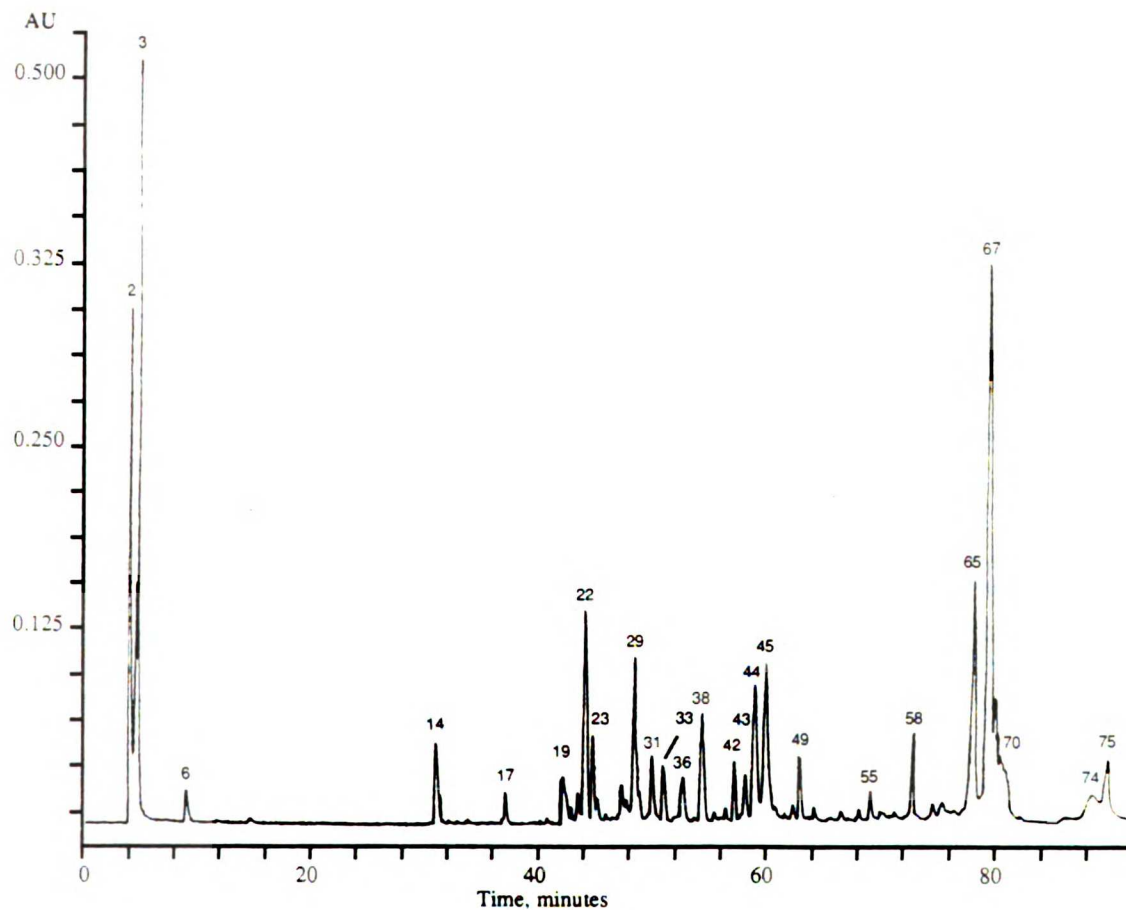


Figure 3. HPLC chromatogram of endoproteinase glu-C (V) digest of 20 nmole of rhPDGF-B monomer. Solvent A=0.1% TFA and B=80% ACN in solvent A. The HPLC gradient was 100% A for 10 min, then 100% A to 40% B in 60 min. Wavelength=230 nm.

UCSF LIBRARY

extensive glycosylation than those used for the T₁ and V digestions. Each HPLC fraction was analyzed by LSIMS and many were also analyzed by amino acid analysis (first tryptic (T₁) and endoproteinase glu-C digests).

Mass spectrometric analysis of the first tryptic digest (T₁) established the presence of all expected tryptic peptides and confirmed the complete sequence predicted by the hPDGF-B gene structure, as shown in Figure 4 (when more than one peptide was identified in the same fraction, the numbers are denoted with a, b, c, etc...). Several overlapping peptides were also identified which resulted from partial tryptic cleavage of the protein in the regions where lysines and arginines are adjacent in sequence. Two fractions, T₁-35 and T₁-40, both contained peptides with masses corresponding to two possible isobaric peptides (T₁-35, KKPIFK or KPIFKK, MH⁺ = *m/z* 760.5 and T₁-40, KIEIVR or IEIVRK, MH⁺ = *m/z* 757.5). Amino acid analyses were used to perform a preliminary identification of 90%

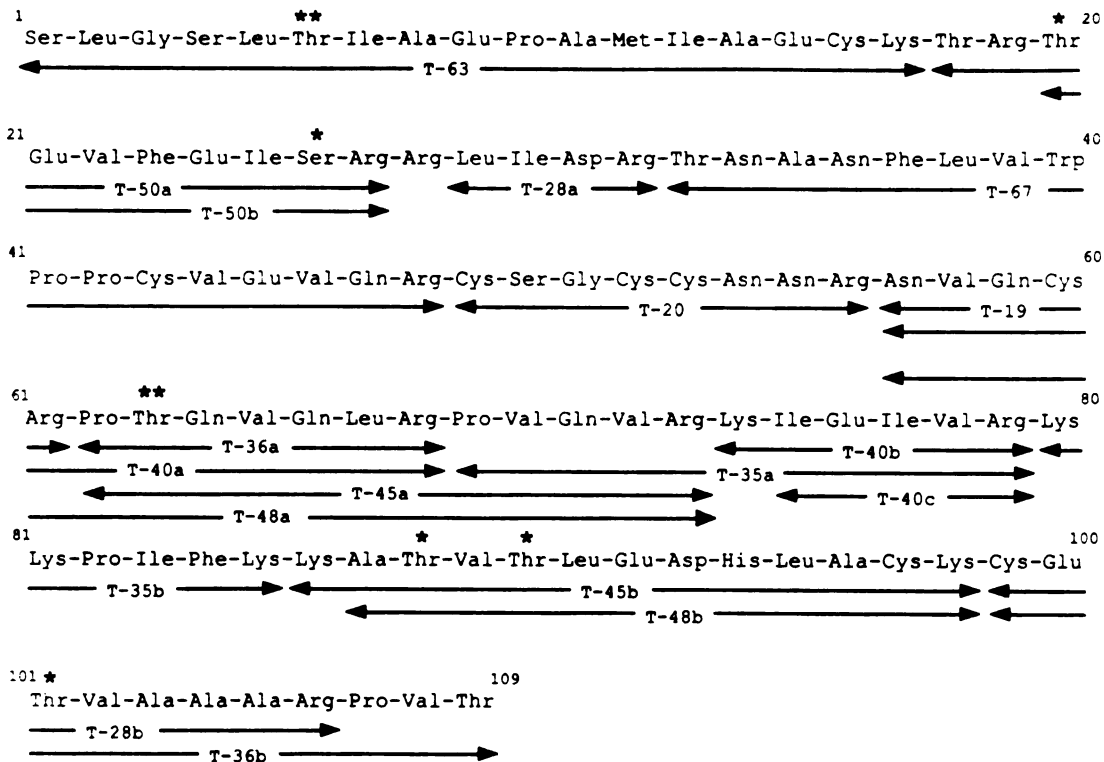


Figure 4. Tryptic (T₁) peptide map of rhPDGF-B. The peptides are numbered according to the fraction numbers shown in Figure 1. Asterisks indicate glycosylation sites.

of the peaks found in the chromatograms. However, the isobaric peptides possessed the same amino acid compositions as well as molecular weight. Therefore, MS/MS was performed on each of these peptides and their sequences were determined to be KKPIFK and KIEIVR, respectively (data not shown).

In addition, several fractions from the tryptic and endoproteinase glu-C digests contained molecular ions corresponding to peptides with masses 162 or 324 Da greater than the masses of predicted peptides, indicating that they may be modified with one or two hexose residues, respectively. UV matrix laser desorption mass spectrometry (Karas *et al.*, 1987; Tanaka *et al.*, 1988) of intact rhPDGF-BB yielded a molecular weight distribution higher than the predicted mass for the amino acid sequence, indicating the presence of extensive post-translational modification such as glycosylation. Since LSIMS and UV laser desorption MS revealed the putative presence of multiple *O*-glycosylation on serine and threonine residues, tandem mass spectrometry was used to establish the nature of glycosylation at each attachment site. Hydrolysis of intact rhPDGF-BB, followed by monosaccharide analysis using HPAE-PAD (Hardy *et al.*, 1988), indicated that only mannose was present in this glycoprotein (data not shown), as expected from expression of this protein in yeast (Brake *et al.*, 1983). Using a combination of LSIMS, high energy CID analysis and Edman sequence analysis, seven different sites of glycosylation were identified in the PDGF sequence. Monomannosyl modifications were found on Thr²⁰, Ser²⁶, Thr⁶³, Thr⁸⁸, Thr⁹⁰ and Thr¹⁰¹. Dimannosyl modifications were found on Thr⁶ and Thr⁶³.

Glycosylation was observed on several peptides containing two putative *O*-linked sites at Thr⁸⁸ and Thr⁹⁰, including T₁-41, V-42, V-43 and T₂-57c. The high energy CID spectrum of the peptide T₁-41 (shown in Figure 5), with the sequence KAT(88)VT(90)LEDHLACK + 2 Man (MH⁺ = *m/z* 1810.8), revealed that both Thr were glycosylated with one Man residue. As is usual for the fragmentation of a peptide with a basic C-terminal residue, this spectrum is dominated by C-terminal sequence ions. The

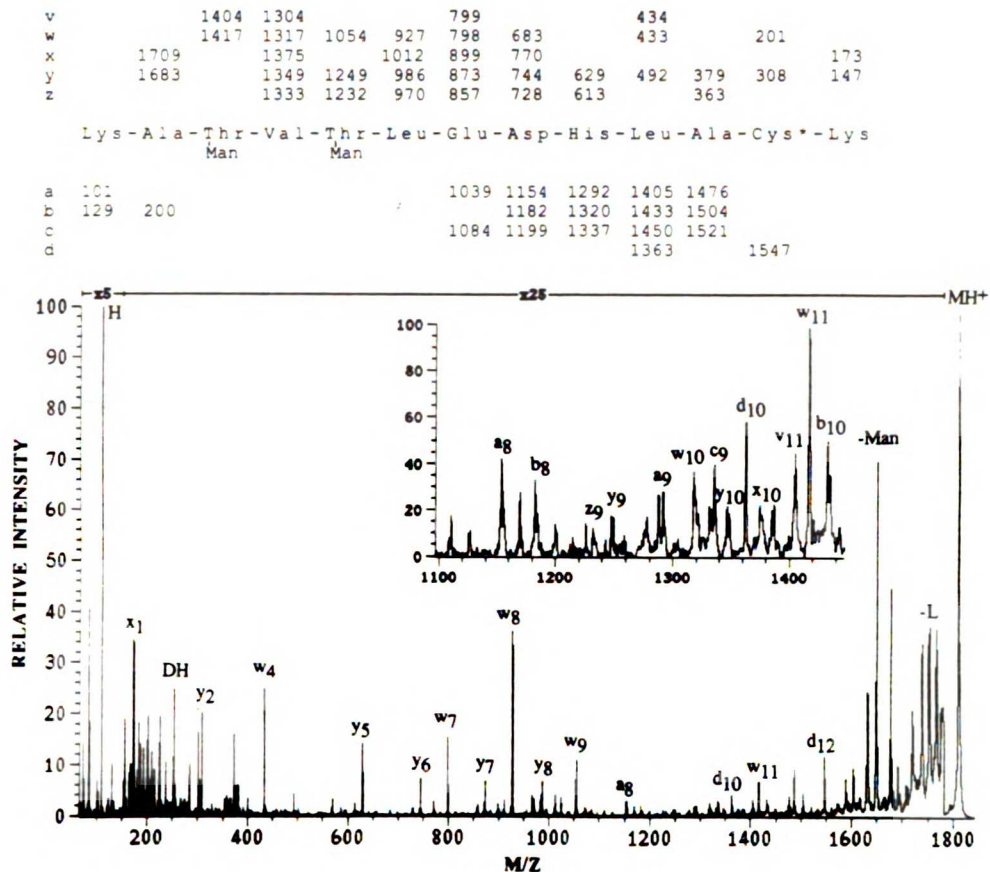


Figure 5. CID spectrum of tryptic peptide T₁-41 (KATVTLEDHLACK + 2 Man, MH⁺ = 1810.8) from rhPDGF-B. All ions identified are included in the ion table above the spectrum where the *m/z* values for ions belonging to the same ion series are listed by rows. For clarity, only the most abundant ions are labeled in the spectrum.

diagnostic ions clearly show that both Thr are modified with one Man residue, by virtue of the shift of y_9 , z_9 , x_{10} , y_{10} , z_{10} , v_{10} , w_{10} , v_{11} and w_{11} by only 162 Da and the shift of x_{12} and y_{12} by 324 Da. Glycosylation of Thr⁸⁸ and Thr⁹⁰ was confirmed by CID analysis of another peptide, V-42, with the sequence IVRKKPIFKKAT(88)VT(90)LE + 2 Man (MH⁺ = *m/z* 2195.8), which covers the same region of protein sequence as T₁-41. In contrast to the CID spectrum of peptide T₁-41, the CID spectrum of V-42 is completely dominated by N-terminal sequence ions, due to the presence of several basic residues close

to the amino terminus of this peptide. CID analysis of V-42, shown in Figure 6, indicated one mannose residue on each Thr, just as in the peptide T₁-41. Therefore, two different ion series in two different peptides of overlapping sequence indicated the same modification of both Thr. In addition, Edman sequence analysis of the glycosylated (T₁-41) and nonglycosylated (T₁-45b) forms of the peptide KAT(88)VT(90)LEDHLACK, shown in Table I, suggested that both Thr were modified by the significant signal losses at both threonine positions. Since the nonglycosylated form of this peptide is also present in these digests, the protein is not completely mannosylated on Thr⁸⁸ and Thr⁹⁰.

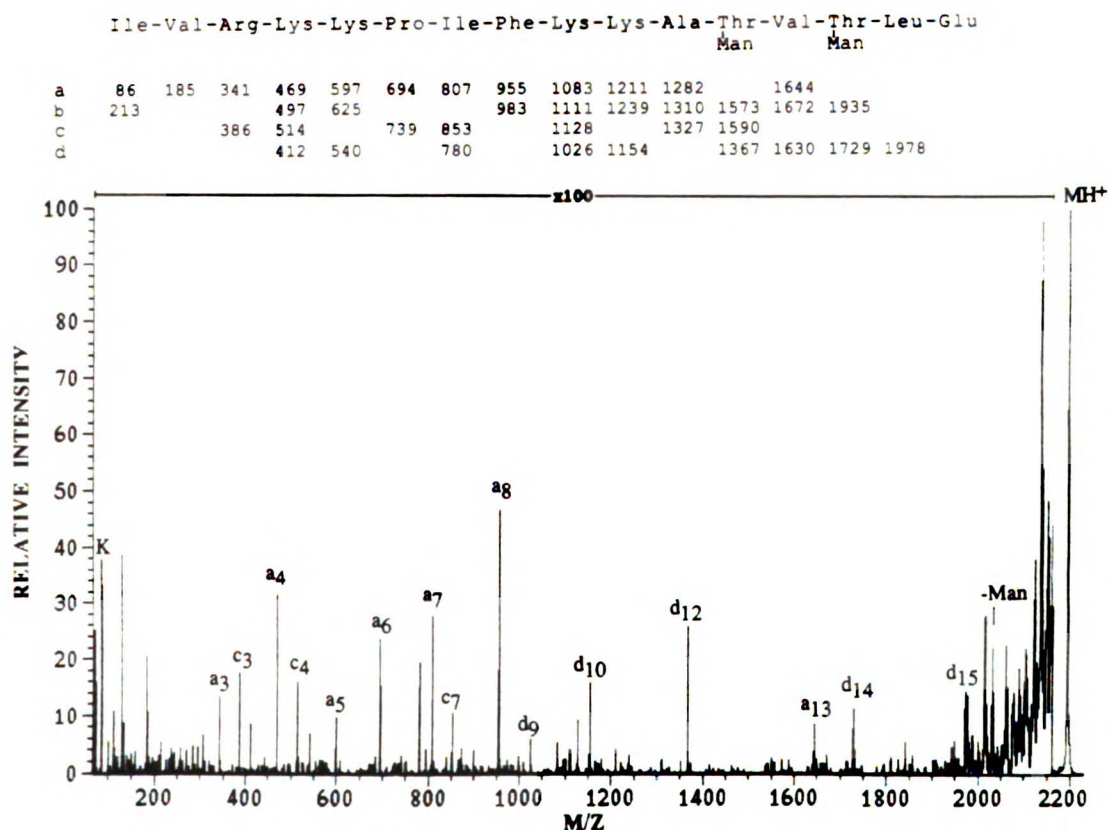


Figure 6. CID spectrum of endoproteinase glu-C peptide V-42 (IVRKKPIFKKATVTLE + 2 Man, MH⁺ = 2195.8) from rhPDGF-B. All ions identified are included in the ion table above the spectrum where the m/z values for ions belonging to the same ion series are listed by rows. For clarity, only the most abundant ions are labeled in the spectrum.

The peptide V-43, with the sequence IVRKKPIFKKAT(88)VT(90)LE + 1 Man (MH⁺ = *m/z* 2034.0), also contains two possible glycosylation sites, Thr⁸⁸ and Thr⁹⁰. Interpretation of the CID spectrum of this monoglycosylated peptide established mannosylation exclusively on Thr⁸⁸. A peptide with the same sequence as V-43 and V-42, but in the nonglycosylated form, was also identified in V-44 (MH⁺ = *m/z* 1871.6), indicating further that the extent of glycosylation is not complete.

| <u>Cycle</u> | <u>Residue</u> | <u>Signal (pmole)</u> | <u>Residue</u> | <u>Signal (pmole)</u> |
|--------------|-------------------------|-----------------------|--------------------------|-----------------------|
| | <u>T₁-41</u> | | <u>T₁-45b</u> | |
| 1 | K | 137 | K | 141 |
| 2 | A | 204 | A | 260 |
| 3 | T | 3 | T | 57 |
| 4 | V | 164 | V | 239 |
| 5 | T | 1 | T | 33 |
| 6 | L | 194 | L | 188 |
| 7 | E | 121 | E | 121 |
| 8 | D | 85 | D | 100 |
| 9 | H | 21 | H | -- |
| 10 | L | 125 | L | 87 |
| 11 | A | 96 | A | 71 |
| 12 | C* | -- | C* | -- |
| 13 | K | 20 | K | 21 |

Table I - Edman sequence analysis results of peptide T₁-41 (glycosylated), MH⁺= 1810.8 and peptide T₁-45b (unglycosylated) MH⁺ = 1486.4. C*=carboxymethyl-cysteine.

Thr⁶ was identified as a site of glycosylation with two Man residues in peptides T₁-60 and T₂-71a. Peptide T₁-60, with the sequence SLGSLTIAEPAMIAECK + 2 Man (MH⁺ = *m/z* 2115.8), contains three possible glycosylation sites: Ser-1, Ser-4 and Thr-6. CID analysis of this peptide showed C-terminal fragment ions which indicated dimannosylation on Thr⁶ (data not shown). Typical C-terminal fragmentation was observed up to the Thr⁶ residue. However, after Thr⁶, several unusual fragment ions were found which prevented

unambiguous sequencing through the remainder of the peptide. In addition, Edman sequence analysis of fraction T₁-60 showed normal cycles at serine residues 1 and 4 and a blank cycle at position 6, confirming that Thr⁶ must be glycosylated with a dimannosyl moiety. The Edman degradation results of peptide T₁-60 are shown in Table II. Peptide T₂-71a (SLGLSTIAEPAM(O)IAECK + 2 Man, MH⁺ = *m/z* 2131.6) contained the same sequence as T₁-60, but its mass was 16 Da higher than the predicted mass, indicating that it may be oxidized. CID analysis of the oxidized, nonglycosylated form of this peptide (T₁-59) confirmed that Met¹² is oxidized to the sulfoxide (Figure 7).

| <u>Cycle</u> | <u>Residue</u> | <u>Signal (pmole)</u> |
|--------------|----------------|-----------------------|
| 1 | S/S' | 33 |
| 2 | L | -- |
| 3 | G | 36 |
| 4 | S/S' | 19 |
| 5 | L | 23 |
| 6 | blank | -- |
| 7 | I | 20 |
| 8 | A | 19 |
| 9 | E | 9 |
| 10 | P | 12 |
| 11 | A | 15 |
| 12 | M | 10 |
| 13 | I | 10 |
| 14 | A | 12 |
| 15 | E | 12 |
| 16 | C* | -- |
| 17 | K | -- |

Table II - Edman sequence analysis results of peptide T₁-60. C*=carboxymethyl-cysteine.

Thr⁶³ was identified with one and two mannose residues attached in peptides T₁-45c, T₂-60, T₂-52 and T₂-53. Thr⁶³ was the only possible site of glycosylation in all of these peptides. MS/MS of T₂-53 with the sequence NVQCRPTQVQLR + 1 Man (MH⁺ = *m/z* 1661.8) showed characteristic fragmentation of a peptide with two non-adjacent arginines.

Both N-terminal and C-terminal ions (a_8 , a_9 , d_8 , d_9 , d_{10} , y_{7-2} and y_8) were found with a 162 Da mass shift (data not shown). Peptides T_1 -45c, T_2 -60 and T_2 -52 are partial peptides containing the same sequence as T_2 -53 (shown in Table IV) with only one possible site of glycosylation. All four of these sequences were also identified as unglycosylated peptides, as shown in Figure 4, indicating that glycosylation is not complete at Thr⁶³.

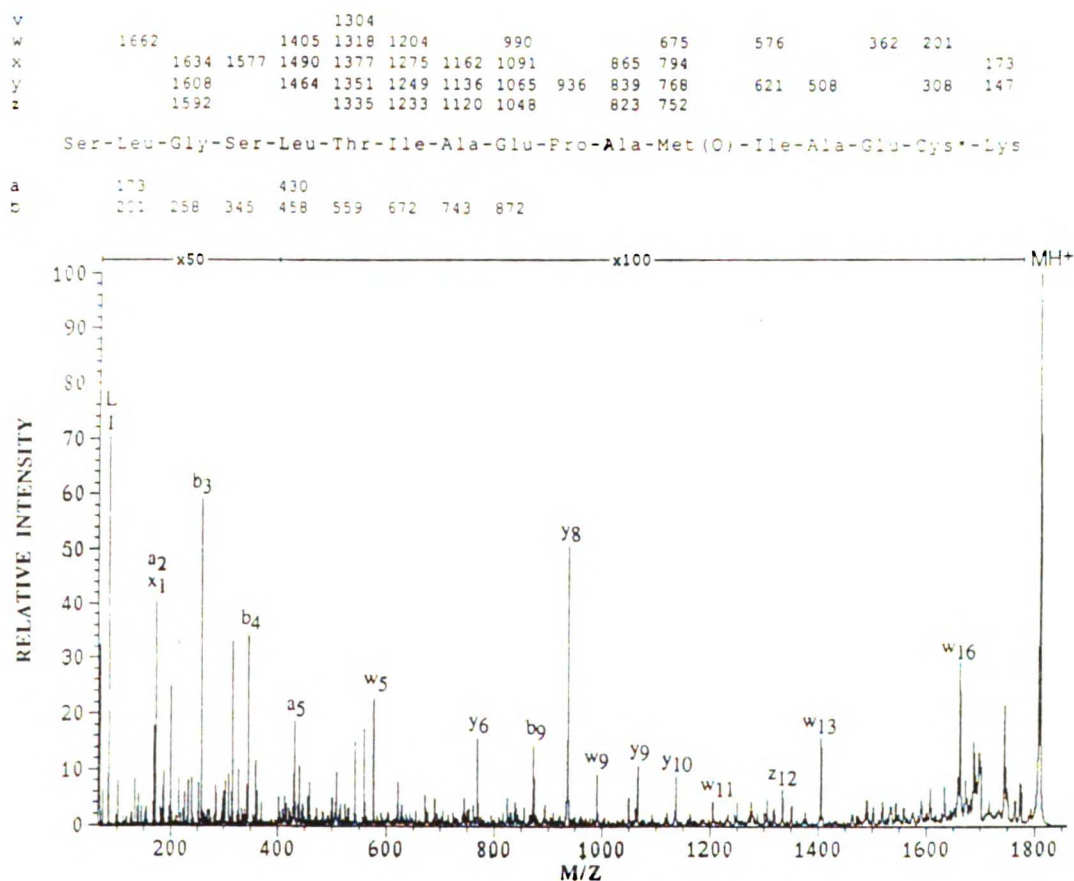


Figure 7. CID spectrum of tryptic peptide T_1 -59 (SLGSLTIAEPAM(O)IAECK, $MH^+ = 1808.0$) from recombinant human PDGF-B. All ions identified are included in the ion table above the spectrum where the m/z values for ions belonging to the same ion series are listed by rows. For clarity, only the most abundant ions are labeled in the spectrum.

The two glycosylation sites at Thr²⁰ and Ser²⁶ were both identified in peptide T₂-56a, with the sequence TEVFEISR + 2 Man (MH⁺ = *m/z* 1304.6), which was modified with two mannose residues. The CID spectrum of this peptide revealed that both sites are glycosylated with one mannose residue. The ions identified are listed in Table III. Thr¹⁰¹

| <u>T₂-56a</u> | MH⁺ = <i>m/z</i> 1304.6 | | | | | |
|--|---|-----|-----|-----|-----|---------|
| v | | 868 | | 592 | 479 | |
| w | | 881 | | 591 | 492 | 229 |
| x | 1067 | 839 | 692 | 563 | 201 | |
| y | | 912 | 813 | 666 | 537 | 175 |
| z | | | | 650 | 521 | 159 |
| T h r - G l u - V a l - P h e - G l u - I l e - S e r - A r g M a n M a n | | | | | | |
| | | | | | | |
| <u>T₂-33</u> | MH⁺ = <i>m/z</i> 1040.4 | | | | | |
| v | | 542 | 443 | | | |
| w | 804 | 555 | 456 | | | |
| x | | | 513 | 414 | 343 | 272 201 |
| y | 879 | 750 | 487 | | 317 | 246 175 |
| z | | 734 | 471 | | 301 | 230 159 |
| C y s * - G l u - T h r - V a l - A l a - A l a - A l a - A r g M a n | | | | | | |
| a | 134 | 263 | | | | |
| b | | 291 | 554 | | | |

Table III. CID fragmentation of rhPDGF-B-2 tryptic peptides T₂-56a and T₂-33. Cys*=carboxymethyl-cysteine. The *m/z* values for ions belonging to the same ion series are listed by rows.

was identified with one Man residue attached in peptide T₂-33 (CETVAAAR + 1 Man, MH⁺ = *m/z* 1040.4), which contained only one possible glycosylation site. CID analysis of this peptide, shown in Table III, confirmed that this peptide was glycosylated on Thr¹⁰¹

with one Man residue. Again, these sequences were found in peptides with no modification, indicating that the extent of glycosylation on Thr²⁰, Ser²⁶ and Thr¹⁰¹ is not complete. A list of all glycopeptides identified in rhPDGF-B is shown in Table IV. Figure 8 shows the rhPDGF-B sequence with all covalent modifications identified in this study.

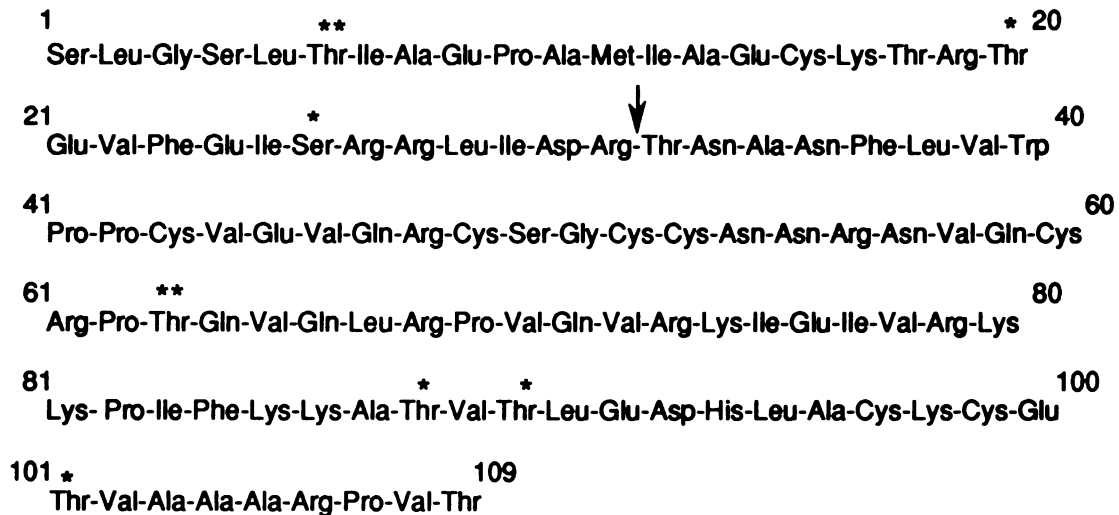


Figure 8. Covalent modifications identified for rhPDGF-B. Asterisks indicate the sites of mannosylation and the maximum number of Man residues observed at the site (as well as none). For example, ** indicates the observation of 0, 1 or 2 Man residues at Thr⁶³. Man resides at position 88 when there is only one Man in the 88-90 region of the molecule. The # indicates the presence of Met sulfoxide at position 12. The arrow between Arg³² and Thr³³ indicates the trypsin-like cleavage found in all preparations of rhPDGF-BB expressed in yeast. All modifications are variable.

Laser desorption mass spectrometry (Karas *et al.*, 1987; Tanaka *et al.*, 1988) of intact rhPDGF-B homodimer was performed to assess the overall degree of glycosylation (Figure 9). The predicted average mass of the dimer with no glycosylation is 24,589 Da. The laser desorption spectrum showed a very broad peak, indicating a large degree of heterogeneity in extent of glycosylation of this protein. The most abundant average mass observed was 25,298 Da. Since the accuracy of the measurement is $\pm 1\%$, this difference corresponds to a mass range of approximately 3 - 6 mannose residues per mole of rhPDGF-B dimer. However, the width of the peak indicates a potential modification range of 0 - 12 (or more)

mannose residues per mole of protein dimer. These data confirm that there is considerable overall heterogeneity in the amount of glycosylation that occurs during expression of this protein.

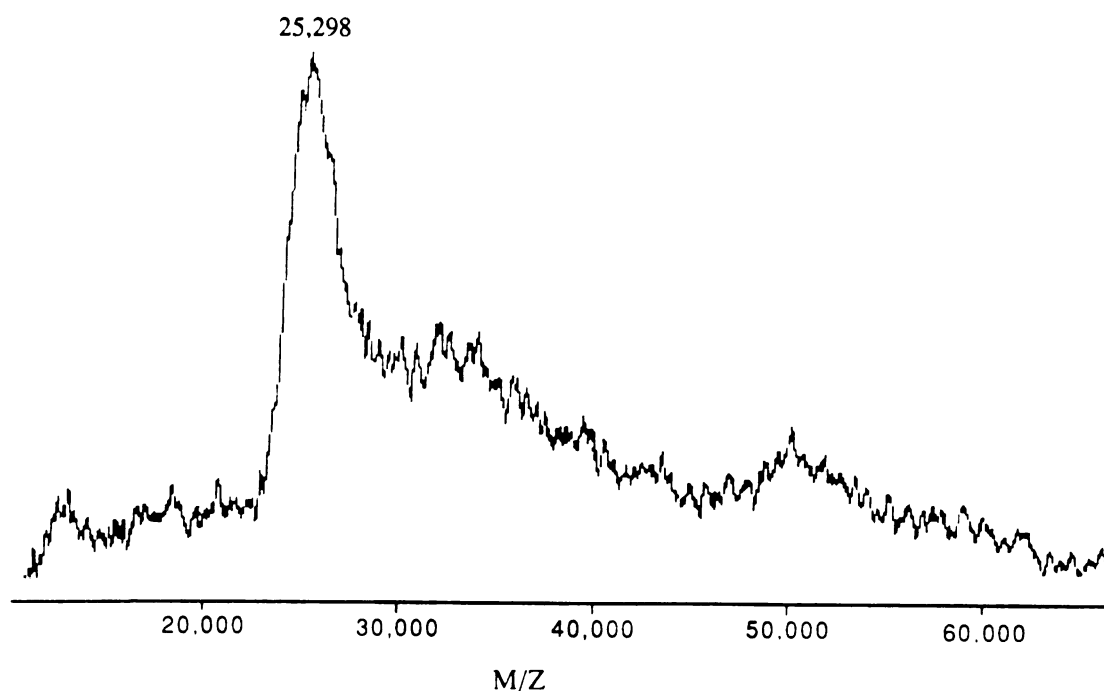


Figure 9. Laser desorption mass spectrum of intact rhPDGF-BB. The mass accuracy is $\pm 1\%$. The average mass indicated is a peak top measurement that was calibrated with chicken egg white lysozyme clusters.

| <u>Fraction</u> | <u>AA Sequence</u> | <u>Sequence Position</u> | <u>Mass (MH⁺)</u> |
|---------------------|----------------------------|--------------------------|------------------------------|
| T ₁ -41 | * * KATVTLEDHLACK | 86 - 98 | 1810.8 |
| T ₂ -57c | * * KKPIFKKATVTLEDHL | 80 - 95 | 2192.8 |
| V-42 | * * IVRKKPIFKKATVTLE | 77 - 92 | 2195.8 |
| V-43 | * IVRKKPIFKKATVTLE | 77 - 92 | 2033.2 |
| T ₁ -60 | ** SLGSLTIAEPAMIAECK | 1 - 17 | 2115.8 |
| T ₂ -71a | ** SLGSLTIAEPAM(O)IEACK | 1 - 17 | 2131.6 |
| T ₁ -45c | ** NVQCRPTQVQLRPVQVR | 57 - 73 | 2403.2 |
| T ₂ -60 | * NVQCRPTQVQLRPVQVR | 57 - 73 | 2241.0 |
| T ₂ -52 | ** NVQCRPTQVQLR | 57 - 68 | 1824.0 |
| T ₂ -53 | * NVQCRPTQVQLR | 57 - 68 | 1661.8 |
| T ₂ -56a | * * TEVFEISR | 20 - 27 | 1304.4 |
| T ₂ -33 | * CETVAAAR | 99 - 106 | 1040.4 |

Table IV- Glycopeptides identified in rhPDGF-B (* = 1 mannose, ** = 2 mannose residues attached to the glycosylation sites which are in bold)

3.5 Discussion

Characterization of *O*-linked glycosylation in glycoproteins requires determination of the carbohydrate structures, the glycopeptide sequence and the sites of carbohydrate attachment. Strategies for the elucidation of *O*-linked structures have included release of oligosaccharides by β -elimination with base in the presence of a reducing agent (Zinn *et al.*, 1977). This method provides information on the type and linkage of carbohydrates, but cannot be used to identify a site of glycosylation on a peptide or protein. This approach was used to determine the structure of *O*-linked carbohydrates in bovine fetuin (Spiro and Bhoyroo, 1974; Nilsson *et al.*, 1979; Edge and Spiro, 1987) but attempts to sequence the *O*-linked peptides produced ambiguous results (Begbie, 1974). Glycosylated amino acids have also been identified following β -elimination of the carbohydrate (Tanaka and Pigman, 1965; Harbon *et al.*, 1968; Downs *et al.*, 1973). The resulting dehydro-peptide can be reduced with borohydride and palladium chloride to form alanine and α -aminobutyric acid (for serine and threonine, respectively) (Tanaka and Pigman, 1965), or sodium sulfite can be added to convert serine to cysteic acid and threonine to α -amino β -sulfonyl butyric acid (Downs *et al.*, 1973; Begbie, 1974). However, this approach has had limited success because the reduced peptide cannot be readily recovered from the palladium catalyst (Tanaka and Pigman, 1965) and base treatment often produces random peptide bond cleavage (Harbon *et al.*, 1968).

Edman degradation of *O*-linked glycopeptides has had limited utility, (Baenziger and Kornfeld, 1974; Begbie, 1974; Tomita *et al.*, 1978) due to poor extraction efficiency, non-identification of glycosylated amino acids and the requirement for a pure glycopeptide. This method has been used to identify sites of modification as described above; however, without mass spectrometric analysis, additional experiments, such as monitoring release of radioactivity during sequence analysis from a previously radiolabeled glycopeptide, are required to unambiguously show that the modified residue contains a particular

carbohydrate residue (Conradt *et al.*, 1985). The often difficult isolation of pure (sequence homogeneous) glycopeptides complicates the identification of glycosylation sites by the classical Edman approach because it is very difficult to determine which amino acid corresponds with which peptide component when sequencing a mixture, unless the peptides are present in very different concentrations. As this study has shown, this is one of the major advantages of CID analysis of peptides and modified peptides over Edman sequence analysis.

Mass spectrometric techniques, such as FAB-MS, are only sufficient for the identification of glycosylation sites if the primary sequence of the peptide is known and there is only one potential glycosylation site in the peptide. Mass spectrometric techniques are commonly used for molecular weight determination of intact *O*-linked glycopeptides, however, chemical sequencing is required to establish the site of glycosylation when more than one serine or threonine is present in the glycopeptides (Sasaki *et al.*, 1988). The use of high energy CID has the major advantage of allowing sequencing of this type of glycopeptide.

Nevertheless, the identification of linkage sites from the CID spectra of *O*-linked glycopeptides bearing larger oligosaccharides may be complicated by the fact that cleavage of the carbohydrate residues from the amino acids often occurs during the high energy collision-induced dissociation (Medzihradszky *et al.*, 1990) yielding ions that do not show the expected mass shift in the peptide sequence ions. This side-chain fragmentation complicates the identification of glycosylation sites, especially for those glycopeptides which contain more than one site of glycosylation. Therefore, one must be cautious in the interpretation of tandem spectra.

The use of recombinant DNA technology in the biotechnology industry has made it possible to obtain large amounts of proteins for a variety of purposes. The complete characterization of these proteins after purification has become a major part of the development of these products. It is ironic that knowledge of the exact structure of many

recombinant proteins will precede information on the natural proteins because of the major differences in amounts readily available. The success of methods used in the study of any given recombinant protein will make it possible to assess the structure of the natural counterpart with an economy of material utilization.

Confirmation of the correct sequence and characterization of modifications that occur on rhPDGF-B as a result of yeast expression are critical to its development as a therapeutic agent. Natural PDGF-B is comprised of 109 amino acid residues of primary sequence and recent experiments involving pure PDGF-B dimer confirm the presence of a cleavage site after Arg³² (Hart *et al.*, 1990). The data of the present study indicate the presence of a portion of the rhPDGF-B molecules as monomers cleaved at the same position (after Arg³²). The presence of a disulfide bond between cysteine-16 and another cysteine residue within the structure would prevent dissociation of the fragment. There is no evidence for cleavage after Leu⁵, as reported for the natural material purified by monoclonal antibody affinity methods.²² However, there is evidence for oxidation of Met¹², which may be the result of aeration during fermentation or oxidation during purification. Methionine oxidation has been seen in both human epidermal growth factor (EGF) (George-Nascimento *et al.*, 1988) and insulin-like growth factor-I (IGF-I) expressed in yeast (Raschdorf *et al.*, 1988).

The results of these experiments demonstrate that the extent of glycosylation of rhPDGF-B is variable at different sites on the protein. The high degree of *O*-glycosylation seen in rhPDGF-B expressed in yeast provides the opportunity to assess the specificity of the machinery of carbohydrate addition. Human PDGF-B contains a total of four serine and nine threonine residues spread throughout the molecule. The majority of the glycosylation occurs on threonine residues (6 of 7 sites). Unlike the situation seen in IGF-I, where the glycosylation site, Thr²⁹, contained an N-terminal proline at position 28 (Gellerfors *et al.*, 1989), there does not appear to be any correlation between the glycosylation of a serine or threonine residue and the N-terminal proximity of a proline

residue. Thr⁶³ is the only site which meets such a criterion in human PDGF-B. It is interesting to note that the mannosylation of Thr⁶ occurs only as a dimer and not as a monomer, while mannosylation of Thr⁶³ occurs as both a monomer and dimer. In addition, when a single mannosylation occurs in the region of PDGF-B containing Thr⁸⁸ and Thr⁹⁰, we have identified mannose only on Thr⁸⁸. The availability of a site is most likely controlled by the secondary or tertiary structure and not a consensus sequence such as that established for N-glycosylation (Bause and Lehle, 1979), so further studies concerning this selective glycosylation could provide some information about the structure and topography of the folded protein.

It is clear that a complete characterization of natural and recombinant proteins is necessary to determine the “natural state” of a protein and the “authenticity” of the recombinant material. The complementary use of amino acid analysis, Edman degradation and the many new modes of mass spectrometry useful for biopolymers will be a necessary prerequisite to providing a complete catalog of this information.

**CHAPTER 4. CHARACTERIZATION OF THE *O*-LINKED OLIGOSACCHARIDES
FROM RECOMBINANT HUMAN NERVE GROWTH FACTOR
RECEPTOR EXTRACELLULAR DOMAIN EXPRESSED IN CHINESE
HAMSTER OVARY CELLS**

4.1 Abstract

Mass spectrometric analysis of the extracellular domain of the human low-affinity p75 nerve growth factor receptor expressed in Chinese hamster ovary cells has identified seven sites of Ser/Thr-linked glycosylation at Ser^{198, 210, 211} and Thr^{199, 205, 206, 216}. Structural characterization of the Ser/Thr-linked oligosaccharides by high performance liquid chromatography/electrospray ionization mass spectrometry coupled with selected ion monitoring and glycosidase digestion, indicates that the structures are galactose β 1 \rightarrow 3N-acetylgalactosamine structures containing 0, 1 and 2 sialic acid residues. These *O*-linked structures fall in a highly conserved portion of the full length receptor near the transmembrane domain, but not within the region which contains multiple proline residues, where *O*-linked glycosylation would be predicted (Wilson *et al.*, 1991). The presence of multiple Ser/Thr-linked glycosylations in a highly conserved region of this protein may have functional significance for ligand binding.

4.2 Introduction

Nerve growth factor (NGF) is a neurotrophic factor that promotes the survival and function of sympathetic and neural crest-derived sensory neurons subsequent to specific interactions with receptors present on the plasma membrane (Banarjee *et al.*, 1973; Frazier *et al.*, 1974). NGF interacts with two different transmembrane receptor proteins, a 75 kDa glycoprotein receptor, p75 NGFR (Johnson *et al.*, 1986; Radeke *et al.*, 1987), and a tyrosine kinase receptor, p140^{trk} (Kaplan *et al.*, 1991; Klein *et al.*, 1991), which are coexpressed in many NGF-responsive neurons (Holtzman *et al.*, 1992; Schechterson and Bothwell, 1992). NGFR is structurally homologous to a family of cysteine-rich receptors that regulate cell death by apoptosis (Smith *et al.*, 1990; Knox and Gordon, 1993; Ogasawara *et al.*, 1993; Tartaglia *et al.*, 1993). Other members of this family include the p55 and p75 tumor necrosis factor receptors (TNFR-I and TNFR-II, respectively) (Smith *et al.*, 1990; Banner *et al.*, 1993; Pennica *et al.*, 1993), the Fas antigen/Apo-I (Itoh *et al.*, 1991), the B cell antigen CD40 (Stamenkovic *et al.*, 1989), the rat T cell antigen OX40 (Mallett *et al.*, 1990), the murine T cell antigen 4-1BB (Kwon and Weissman, 1989) and the lymphocyte-specific receptor CD27 (Loenen *et al.*, 1992). Each of the proteins in this family contain four cysteine-rich units consisting of three disulfide bonds, as shown in Figure 1 for NGFR. Experiments involving mutations and deletions in each of the cysteine-rich units of NGFR (Yan and Chao, 1991) indicate that mutations that affect residues in the third and fourth cysteine-rich units (amino acids 80-160, see Figure 1) have a negative effect upon NGF binding, implying that the binding domain for NGF encompasses this region. However, the crystal structure of the soluble extracellular domain of TNFR-I with TNF- β , indicates the ligand binding domain encompasses the second and third cysteine repeats (Banner *et al.*, 1993). NGFR also contains a domain which connects the ligand-binding part of the structure to the transmembrane region of the receptor. This "stalk" region (amino acids 161-222) contains a high percentage of proline,

serine and threonine residues, indicative of Ser/Thr-linked (*O*-linked) glycosylation (Jentoft, 1990; Wilson *et al.*, 1991). Similar stalk regions have been identified in the 75 kDa TNFR-II (Hohmann *et al.*, 1990), the CD-30 antigen (Dürkop *et al.*, 1992) and the CD-27 antigen (Gravestain *et al.*, 1993).

The neurotrophic stimulation by NGF is mediated by high-affinity binding sites ($K_D=10^{-11}$ M; Sutter, 1979). Another class of binding sites to NGF has a lower affinity ($K_D=10^{-9}$ M) mediated by p140^{trk} (Hempstead *et al.*, 1991; Kaplan *et al.*, 1991) as well as p75 NGFR (Chao *et al.*, 1986; Radeke *et al.*, 1987). Many of NGFR's functions in neurons occur in conjunction with p140^{trk} (Chao, 1992), which regulates neuronal differentiation. In this context, NGFR regulates the affinity of ligand binding to the tyrosine kinase receptor complex (Pleasure *et al.*, 1990; Hempstead *et al.*, 1991; Battleman *et al.*, 1993) and also influences the ligand specificity of *trk* receptors (Rodriguez-Tebar *et al.*, 1992; Benedetti *et al.*, 1993; Davies *et al.*, 1993). In addition, recent studies have shown that the affinity of NGFR for NGF shifts from low ($K_D=10^{-9}$ M, fast dissociating) to high ($K_D=10^{-11}$ M, slow dissociating) upon association with catalytic subunits of p140^{trk} (Hempstead *et al.*, 1992). This shift in the equilibrium binding constant can also be induced by the binding of wheat germ agglutinin (WGA), a carbohydrate-binding protein (or lectin), to NGFR in the presence of NGF (Vale and Shooter, 1982; Buxser *et al.*, 1983; Grob and Bothwell, 1983). The effects of WGA are blocked by the sugar N-acetyl-D-glucosamine (GlcNAc), a specific substrate for WGA, however, Vale and coworkers demonstrated that sialic acid (and not GlcNAc) was the principal target of WGA binding (Vale *et al.*, 1985). In addition, Chapman and coworkers found that WGA binds to sialic acids on the *O*-linked glycans of sNGFR (Chapman *et al.*, 1994). This indicates that binding to *O*-linked carbohydrates on NGFR may be involved in the high affinity binding of NGF to NGF receptors.

In order to study the structure of the NGF binding region in NGFR, and address any possible function of the *O*-linked carbohydrates on NGFR, a recombinant soluble form

(sNGFR) was prepared (Chapman *et al.*, 1994) that contains the amino-terminal 222 amino acids ending just before the transmembrane region of the native full length receptor, and retains its ability to bind NGF ($K_d=10^{-9}$ M) (Chapman *et al.*, 1994). This protein contains one asparagine residue at position 32 in the consensus sequence (Asn-Xxx-Ser/Thr, where Xxx = any amino acid except Pro) required for attachment of Asn-linked (*N*-linked) oligosaccharides, (Kornfeld and Kornfeld, 1976) as well as the Ser, Thr and Pro-rich carboxy-terminal stalk domain which was believed to contain *O*-linked glycosylation. This domain is divided into a variable portion (residues 164-197, 31% identity between avian and mammalian sequences) and a conserved membrane-associated portion (residues 198-222, 76% identity between avian and mammalian sequences). The goal of this study was to determine the exact sites of *O*-linked glycosylation, because if the glycosylation was found to be in the conserved region of the stalk rather than in the variable region, this would imply a functional significance for the carbohydrates, such as for ligand binding.

Microbore HPLC/ESIMS and HPLC/ESI/CID/MS combined with glycosidase digestion was used to determine the location and structures of the *O*-linked oligosaccharides on sNGFR. HPLC/ESI/CID/MS, coupled with SIM of carbohydrate-specific fragment ions was the key analytical tool used to identify and characterize the *O*-linked structures on this glycoprotein. These *O*-linked structures fall in the highly conserved portion of the stalk region of the receptor which falls just before the transmembrane domain, but not within the region which contains multiple proline residues, where *O*-linked glycosylation would be predicted (Wilson *et al.*, 1991). Possible functional implications of the location and presence of this glycosylation include a basic structural function for the stalk region itself or something more complex such as an influence on the folding and orientation of the nearby ligand binding domain and/or NGFR's association with p140^{rk}.

4.3 Materials and Methods

Expression and Purification. The extracellular domain of human NGFR, graciously donated by Frank Masiarz and collaborators at Chiron Corporation, was expressed in Chinese hamster ovary cell line DG44 (Urlaub *et al.*, 1983) as previously described (Chapman *et al.*, 1994). Transfected colonies were screened for expression of the NGFR glycoprotein containing the extracellular domain (sNGFR) by biosynthetic labeling and immunoprecipitation of conditioned media (Chapman *et al.*, 1994). The resulting product bound NGF with a K_D of approximately 10^{-9} M (Kaufman *et al.*, 1993).

Reduction and S-Carboxymethylation. Reduction of purified sNGFR was performed at protein concentrations of 1-4 mg/ml in 200 mM NEMOAc, pH 8.5, or 100 mM ammonium bicarbonate, pH 8.5, containing 6 M guanidine-HCl and 3 mM EDTA with a 50 fold molar excess of dithiothreitol. The protein was reduced with dithiothreitol for 1-1.5 hr at 60°C, followed immediately by alkylation with iodoacetic acid sodium salt for 1 hour at 25°C in the dark.

Protease and Glycosidase Digestions. Endoproteinase glu-C (*Staphylococcus aureus* V8, Boehringer Mannheim, Indianapolis, IN) digestions were performed on reduced and S-carboxymethylated sNGFR in 50 mM ammonium bicarbonate, pH 7.8, at a weight ratio of 8% (2 x 4%) for 18-28 hr at 37°C. Digestion with neuraminidase from *Arthrobacter ureafaciens* (Oxford Glycosystems, Oxford, UK) was performed in 50 mM sodium acetate, pH 5.0, at 37°C for 16-20 hr. Neuraminidase was added to the digestion mixture to obtain a final enzyme concentration of 0.8-1.0 U/ml. Endo- α -N-acetylgalactosaminidase (O-glycosidase, *Diplococcus pneumoniae*, Boehringer Mannheim, Indianapolis, IN) digestions were performed in 50 mM sodium acetate (pH 5.0) at 37°C for 18-24 hr. 1 mUnit was added to the digestion mixture at 0 and 8 hr.

Carbohydrate Composition Analysis. Carbohydrate analyses were performed on acid hydrolyzed, sNGFR-RCM. One nmole of protein was hydrolyzed in each of three

different solutions: 2 M TFA (5 hr), 6 N HCl (3 hr), and 1 N HCl (1 hour) for quantitation of amino sugars, neutral sugars and acidic sugars, respectively. Analyses were performed using a Dionex Bio-LC high performance anion exchange chromatography system consisting of a gradient pump, a pulsed amperometric detector (PAD) with a gold working electrode, and an advanced computer interface with AI450 software for data collection. Samples were injected either manually or using a Spectra Physics SP8780 autosampler onto a CarboPac PA-1 column (4.6 x 250 mm). Eluant 1 was 200 mM NaOH, eluant 2 was 1 M NaOAc and eluant 3 was H₂O. Quantitation was determined by comparing the results to injection of known quantities of a standard monosaccharide mixture.

Isolation of *O*-linked Glycopeptides Using Jacalin Agarose. A slurry of approximately 300 μ l of jacalin agarose beads (jacalin coupled to 4% agarose at a concentration of 4 mg/ml packed volume) in 10 mM HEPES, pH 7.5, 0.15 M NaCl, 0.1 mM Ca²⁺, 20 mM galactose, 20 mM lactose and 0.08 % sodium azide (Vector Laboratories, Burlingame, CA) was washed 5 times in a 1.5 ml polypropylene microcentrifuge tube with 800 μ l 100 mM NH₄HCO₃, pH 7.8 to exchange the buffers and other components. One to two nmole of the endoproteinase glu-C digested, reduced and S-carboxymethylated sNGFR was added and the mixture vortexed periodically for one hour to allow the *O*-linked glycopeptides to bind to the jacalin. After centrifugation, the supernatant was removed and the beads washed three times with 100 mM NH₄HCO₃, pH 7.8 to remove all unbound peptides and glycopeptides. Then 800 μ l 100 mM NH₄HCO₃, pH 7.8 containing 20 mM methyl- α -D-galactopyranoside (Sigma) was added to the jacalin agarose mixture and the tube vortexed periodically for one hour to displace the bound glycopeptides. The supernatant was then removed and the beads washed twice with the buffer containing the methyl- α -D-galactopyranoside and the supernatant removed. The supernatant containing the *O*-linked glycopeptides was then vacuum centrifuged before HPLC/ESIMS analysis.

Peptide and Glycopeptide HPLC. Microbore HPLC/Electrospray ionization mass spectrometric (HPLC/ESIMS) analyses were performed using a Carlo Erba Phoenix 20 gradient HPLC system equipped with a Rheodyne 8125 injector and an Aquapore OD-300 microbore reversed phase column (C₁₈, 100 x 1mm, 7 μm, Applied Biosystems, Foster City, CA). Solvent A was 0.1% TFA/H₂O, solvent B was 0.08% TFA/70% acetonitrile, and the flow rate used was 50 μl/min. Using an Isco μLC-500 syringe pump, a 1:1 mixture of 2-propanol:2-methoxyethanol was mixed with the column effluent post-column (after UV-detection) at a flow rate of 25-40 μl/min. This combined flow of 75-90 ml/min was split at a ratio of approximately 20:1 before entering the mass spectrometer, so that a final flow of approximately 3.5-4.5 μl/min entered directly into the electrospray ionization source. The elution of all compounds was monitored at a wavelength of 215 nm using an Applied Biosystems UV detector model 204 equipped with a U-shaped capillary flow cell (LC Packings, San Francisco, CA).

Mass Spectrometry. ESI mass spectra were acquired on-line, after microbore HPLC elution, on a VG Bio-Q mass spectrometer equipped with an atmospheric pressure electrostatic spray ion source and a quadrupole mass analyzer (maximum m/z = 3000). The flow rate at the probe tip was 3.5-4.5 μl/min. The scan time in non-continuum mode from m/z = 350 to m/z = 2200 was 5 seconds. Typical operating voltages were as follows: probe tip = 4 kV; counter electrode = 1.5 kV; sample cone (nozzle) voltage = 45 V; SIM sample cone voltage: 160-200 V. The nebulizing gas flow rate was typically 150 mL/min, and the drying gas was typically 4 L/min. The source temperature was typically 60-65°C. All data was acquired with Fisons Lab-Base software.

4.4 Results

Carbohydrate Composition Analysis—In order to determine the composition and relative quantities of carbohydrates on sNGFR, the intact, reduced and S-carboxymethylated protein (quantitated by amino acid analysis, data not shown) was hydrolyzed using different conditions specific for quantitation of neutral sugars, amino sugars and sialic acid (Hardy *et al.*, 1988). Analyses were performed using HPAE-PAD (Hardy and Townsend, 1988) and the results are shown in Table I. The large molar values for the monosaccharides galactosamine (from N-acetylgalactosamine, GalNAc), galactose (Gal) and sialic acid (NeuAc) suggested the presence of multiple sialylated *O*-linked carbohydrates.

| | <u>moles of monosaccharide</u> <u>moles hydrolyzed sNGF-R</u> |
|----------------------|--|
| Sialic Acid | 16.5 ± 0.1 |
| Fucose | 1.3 ± 0.1 |
| Galactosamine | 8.7 ± 1.1 |
| Glucosamine | 4.7 ± 0.7 |
| Galactose | 17.4 ± 0.2 |
| Mannose | 3.0 ± 0.1 |

Table I. Monnosaccharide composition analysis of the carbohydrates from p75 nerve growth factor receptor.

Glycopeptide Identification—In order to determine the structures and attachment sites of the *O*-linked oligosaccharides attached to this recombinant glycoprotein, the *O*-linked glycopeptides had to first be identified. sNGFR was reduced and carboxymethylated before digestion with endoproteinase glu-C and the resulting mixture of peptides and glycopeptides was analyzed by microbore reversed phase HPLC/ESIMS. All but residues 195-222 were accounted for as peptides or *N*-linked glycopeptides, based on the molecular weights obtained by HPLC/ESIMS after the endoproteinase glu-C digest. The unaccounted for residues represent the C-terminal portion of the recombinant protein sequence encompasses all of the highly conserved Ser, Thr-rich region (residues 201-222) of sNGFR, as well as a small portion of the Ser, Thr and Pro-rich variable domain (residues 163-200), as shown in Figure 1. Residues 195-222 therefore was predicted to be a glycopeptide which was not identified by HPLC/ESIMS of the endoproteinase glu-C digest because it was not well ionized by the ESI mass spectrometer due to its size and heterogeneity.

Because no *O*-linked glycopeptides were identified in the first analysis, the endoproteinase glu-C digest was treated with neuraminidase from *Arthrobacter ureafaciens*, which hydrolyzes terminal N- or O-acyl-neuraminic acid residues which are $\alpha 2 \rightarrow 3$, $\alpha 2 \rightarrow 6$ or $\alpha 2 \rightarrow 8$ linked to oligosaccharides, glycoproteins or glycopeptides (Uchida *et al.*, 1979; Twining, 1984), in order to decrease the size and the heterogeneity of the glycopeptides. The *O*-linked glycopeptides on sNGFR were then specifically identified using HPLC/ESIMS combined with CID and SIM of carbohydrate-specific fragment ions (Duffin *et al.*, 1992; Carr *et al.*, 1993). These fragment ions were generated by increasing the potential difference between the sampling orifice and the skimmer of the electrospray ion source during the HPLC/ESIMS experiment, thereby causing CID of the molecular ions in the electrospray ion source (HPLC/ESI/CID/MS). By monitoring for the masses of specific diagnostic fragment ions which are generated only from glycopeptides, the

glycopeptides could be identified. In order to perform this experiment, two separate HPLC/ESIMS analyses were required.

The first analysis was performed in the conventional operating mode in which a mass chromatogram representing the total ion abundance, resulting from scanning the entire mass range (m/z 350-2200), and a UV absorbance chromatogram were obtained from the HPLC/ESIMS analysis of the endoproteinase glu-C + neuraminidase digest. A second analysis was then performed using the HPLC/ESI/CID/MS operating conditions with SIM of carbohydrate-specific ions in order to selectively detect where the glycopeptides elute in the HPLC chromatogram of the digest. Once the glycopeptide-containing peaks were identified, the electrospray mass spectra of the glycopeptides from the first analysis could be identified by direct comparison of the two HPLC chromatograms. Based on the data

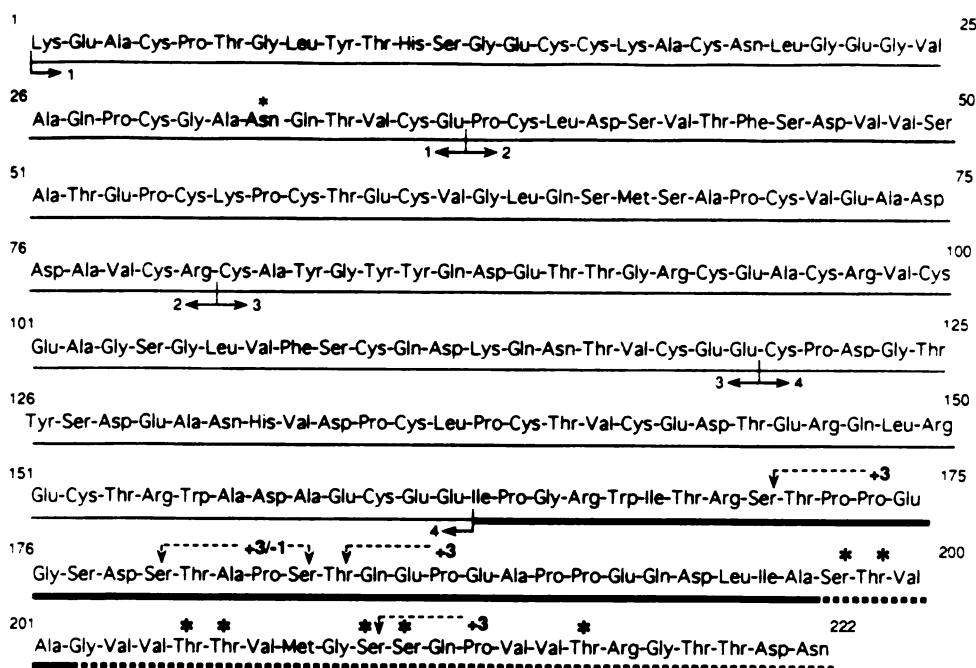


Figure 1. Amino acid sequence of sNGFR, indicating the different domains and glycosylation sites. The thin solid line indicates the section containing the four homologous cysteine repeats, and the boundaries of each repeat are marked with horizontal arrows. The thick solid line indicates the variable domain of the sequence and the thick dotted line indicates the highly conserved membrane-associated domain. The asterisks indicate identified sites of *N*- and *O*-linked glycosylation and the dotted arrows next to the -1 and +3 above the proline residues indicate where *O*-linked glycosylation sites would be predicted based on a general rule for *O*-linked glycosylation proposed by Wilson.

from the carbohydrate composition analysis, the following ions were monitored: $m/z=204$ (GlcNAc^+ or GalNAc^+), $m/z=292$ (NeuAc^+), $m/z=366$ ($[\text{Gal-GlcNAc}]^+$ or $[\text{Gal-GalNAc}]^+$) and $m/z=657$ ($[\text{NeuAc-Gal-GlcNAc}]^+$ or $[\text{NeuAc-Gal-GalNAc}]^+$). Figure 2A illustrates the results of the HPLC/ESI/CID/MS experiment for the endoproteinase glu-C + neuraminidase digest with SIM at $m/z=204$. As illustrated in the lower panel, several peaks are present in the SIM chromatogram. The largest peak is the result of fragmentation of the *N*-linked glycopeptide (residues 24-41) which eluted at approximately 23.5 min. The mass spectrum corresponding to the smaller peak eluting at approximately 22 min (starred in Figure 2A) is shown in Figure 3A. The molecular mass of this single component at 5320.0 Da corresponds to residues (195-222) containing seven hexosyl-*N*-acetylhexosamine residues. Several glycopeptides were also identified in the broad group of peaks at 30 to 35 min which contain larger peptide components encompassing the same C-terminal region of sNGFR and four to seven hexosyl-*N*-acetylhexosamine residues.

In order to further characterize these glycopeptides, the endoproteinase glu-C + neuraminidase digest was then treated with Endo- α -*N*-acetylgalactosaminidase (*O*-glycosidase) from *Diplococcus pneumoniae* which hydrolyzes the disaccharide Gal β 1 \rightarrow 3GalNAc from serine or threonine residues either alone or as part of glycopeptides or glycoproteins (Endo and Kobata, 1976; Umemoto *et al.*, 1977). As shown in Figure 2B, the SIM at $m/z=204$ chromatogram shows only a single large peak corresponding to the *N*-linked glycopeptide, indicating that the *O*-linked glycopeptides are no longer present, because the *O*-glycosidase cleaved the sugars from the glycopeptides. In addition, the endoproteinase glu-C + neuraminidase + *O*-glycosidase digest contained new peptides in the HPLC/ESIMS data corresponding to the unmodified peptides which were previously glycopeptides. The major new peptide eluted in the peak at approximately 60.5 min, as shown in the UV chromatogram in figure 2B. Figure 3B shows the ESI mass spectrum of this peak, which contains the deglycosylated peptide from figure 3A, residues (195-222) at 2762.0 Da. Therefore, this peptide contained the seven *O*-linked structures, Gal β 1 \rightarrow 3

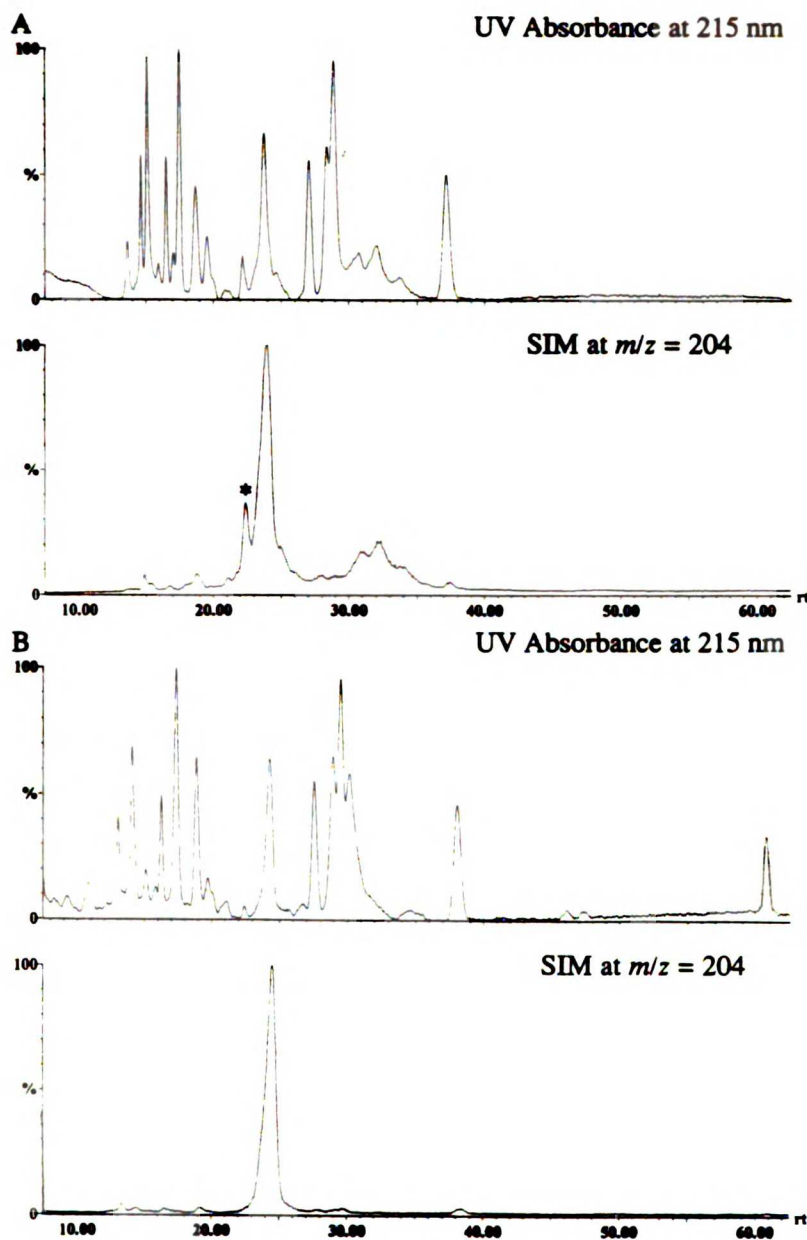


Figure 2. Results of HPLC/ESI/CID/MS experiments for two digests of sNGFR using SIM at $m/z=204$. (A) HPLC/ESI/CID/MS experiment for the endoproteinase glu-C + neuraminidase digest with SIM at $m/z=204$. The single large peak at approximately 23.5 min is the result of fragmentation of the *N*-linked glycopeptide mixture containing amino acid residues (24-41). The starred peak at 22 min represents the *O*-linked glycopeptide for which the mass spectrum is shown in Figure 3A. (B) Results of the HPLC/ESI/CID/MS experiment for the endoproteinase glu-C + neuraminidase + *O*-glycosidase digest with SIM at $m/z=204$. Range=0.05 AUFS. Solvent A=0.1% TFA, solvent B=0.08% TFA in ACN. Gradient program was 0% B for 5 min, 0-20% B in 20 min, 20-40% A in 45 min.

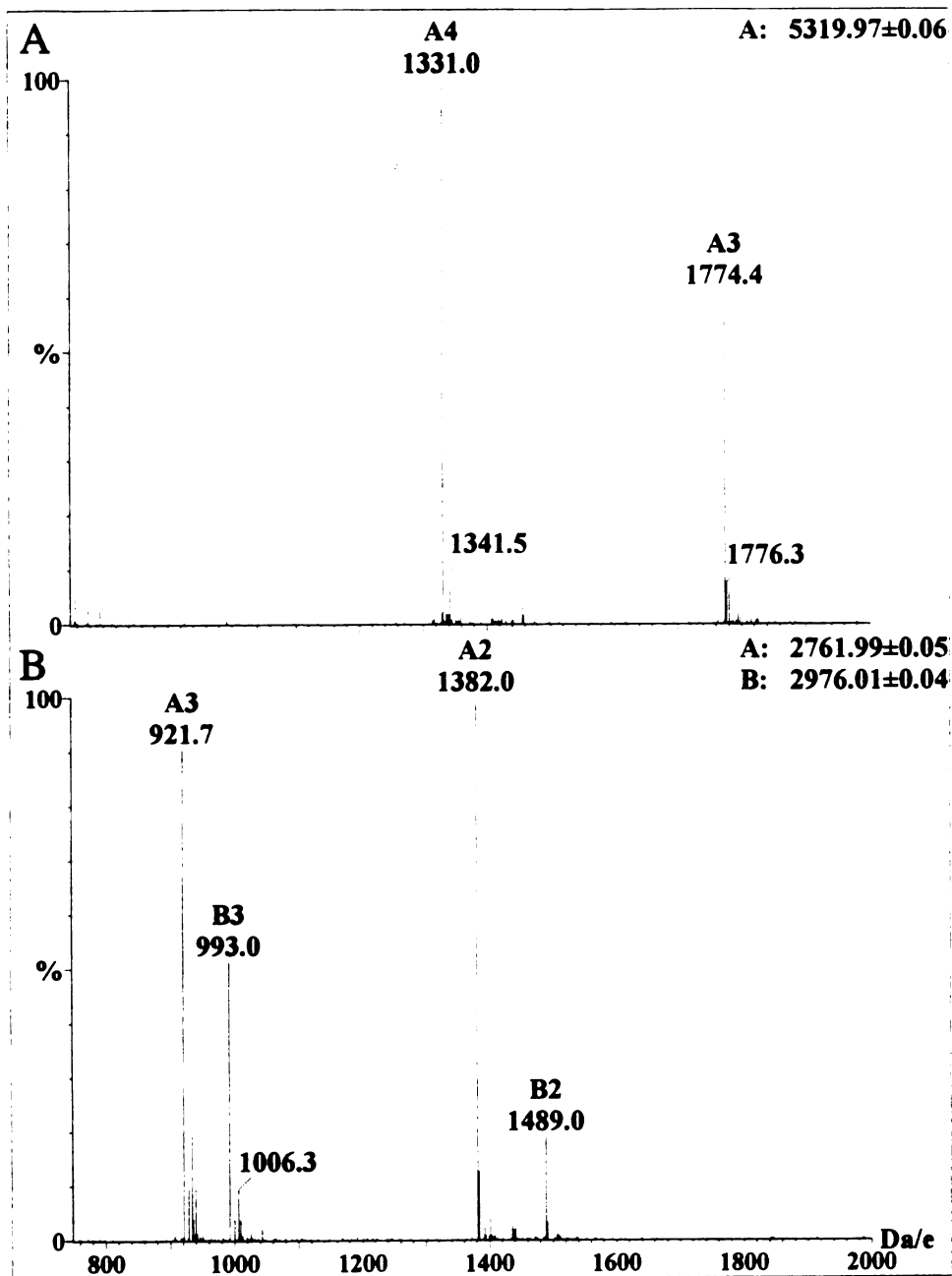


Figure 3. Electrospray mass spectra of a major asialo *O*-linked glycopeptide from sNGFR. (A) ESI mass spectrum of the peak at 22 min (starred in Figure 2A). The molecular mass of this single component at 5320.0 Da corresponds to residues (195-222) containing seven hexosyl-*N*-acetylhexosamine residues. (B) ESI mass spectrum of the peak eluting at 60.5 min in the sNGFR + glu-C + neuraminidase + *O*-glycosidase digest in Figure 2B. The molecular mass of component A (2762.0 Da) corresponds to the deglycosylated peptide from Figure 3A, residues (195-222). Component B is another peptide form sNGFR (152-176) which coeluted with (195-222).

GalNAc, as the glycopeptide, because they were completely removed by *O*-glycosidase, which cleaves only this specific structure.

However, this peptide contains three serine and six threonine residues (see Figure 1), only seven of which are glycosylated. In order to determine which of the nine were glycosylated, the glycopeptide shown in Figure 3A was digested with trypsin, to remove residues (218-222) which contains Thr²¹⁹ and Thr²²⁰, and the digest analyzed by HPLC/ESIMS. As shown in Figure 4, the molecular masses of the two resulting peptides from analysis of the tryptic digest were 4830.4 and 507.5, indicating that residues (218-222) was not glycosylated and that all seven serine and threonine residues in the peptide (195-217) were glycosylated. Therefore, Ser^{198, 210, 211} and Thr^{199, 205, 206, 216} are all glycosylated with Gal β 1 \rightarrow 3 GalNAc.

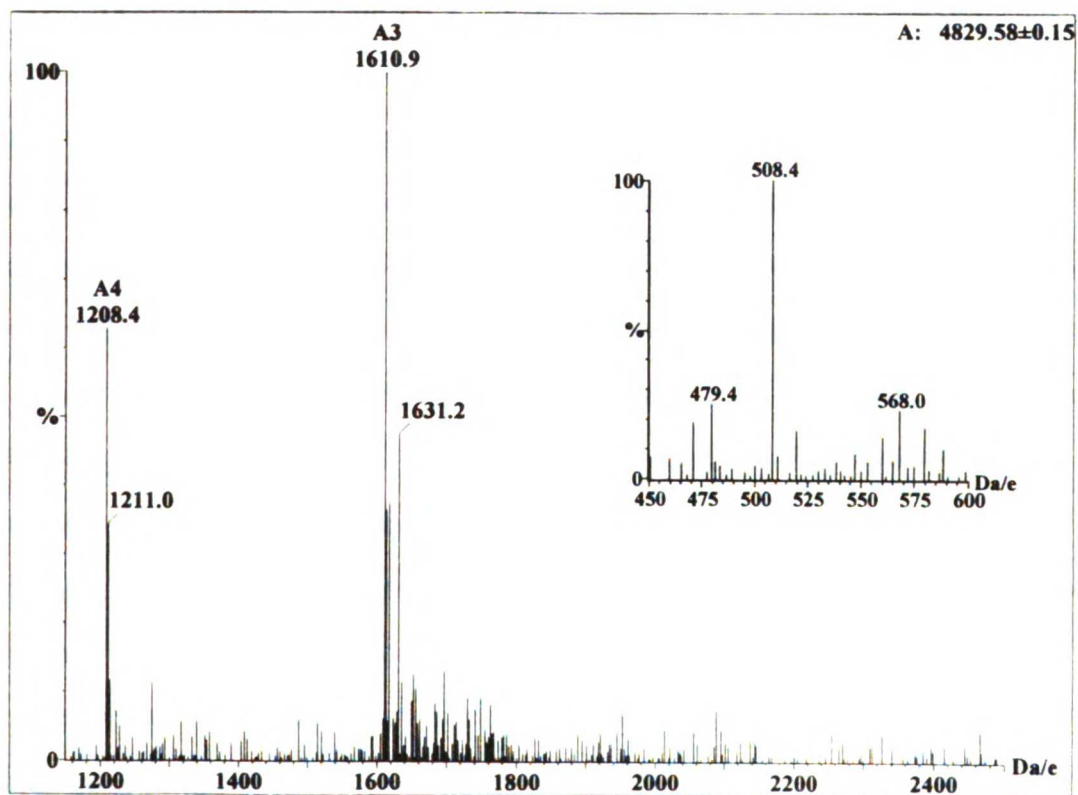


Figure 4. Electrospray mass spectra of the peptides resulting from the tryptic digest of the asialo *O*-linked glycopeptide from sNGFR. ESI mass spectrum of residues (195-217), molecular mass = 4830.4 Da. The insert shows the ESI mass spectrum of residues (218-222), molecular mass = 507.5 Da.

Characterization of Sialylation of O-linked Glycopeptides—The *O*-linked glycopeptides required separation from the rest of the digest for characterization of their sialylation. Therefore, a previously published procedure to isolate *O*-linked glycopeptides using affinity chromatography with a lectin from jackfruit (*Artocarpus integrifolia*), jacalin, bound to agarose beads (Hortin, 1990) was used. Jacalin is a protein that is known to have affinity for the core disaccharide, 1- β -galactospyranosyl-3-(α -2-acetamido-2-deoxygalactopyranoside) (Gal β 1 \rightarrow 3 GalNAc), in *O*-linked oligosaccharides of glycoproteins (Sastry *et al.*, 1986; Young *et al.*, 1989), and has also been shown to bind sialylated Gal β 1 \rightarrow 3 GalNAc structures (Hortin and Trimpe, 1990). Therefore, this method was used to purify the *O*-linked glycopeptides away from the other peptides and glycopeptides in the mixture, in order to facilitate characterization of the intact sialylated glycopeptides.

The procedure was modified to be performed on a smaller scale in a 1.5 ml polypropylene microcentrifuge tube, with minimal sample handling and no desalting step, so that the *O*-linked glycopeptide mixture could be injected directly onto the HPLC/ESIMS system. A glycoprotein with known *O*-linked carbohydrate structures, bovine fetuin (Spiro and Bhojroo, 1974; Edge and Spiro, 1987), was used as a standard to confirm the success of the modified procedure (data not shown). As described in the experimental section, the sialylated *O*-linked glycopeptides were isolated from the endoproteinase glu-C digest of sNGFR by adding the digest to a slurry of jacalin-agarose. The peptides and *N*-linked glycopeptides were removed from the mixture after the *O*-linked glycopeptides were bound to the jacalin-agarose, and then the *O*-linked glycopeptides were displaced from the jacalin using methyl- α -D-galactopyranoside.

After isolation by jacalin agarose affinity chromatography, the *O*-linked glycopeptide mixture was analyzed by HPLC/ESI/CID/MS with SIM of masses corresponding to mono- and disialylated Gal-GalNAc residues ($m/z=657$ [NeuAc-Gal-GalNAc]⁺ and $m/z=948$ [NeuAc-Gal-(NeuAc)GalNAc]⁺). The results, shown in Figure 5, clearly shows the

presence of all of these structures. Figure 5 illustrates the UV absorbance (HPLC) and the SIM data for $m/z=366$, 657 and 948, indicating that the majority of the *O*-linked glycopeptides contain disialyl Gal β 1 \rightarrow 3 GalNAc structures.

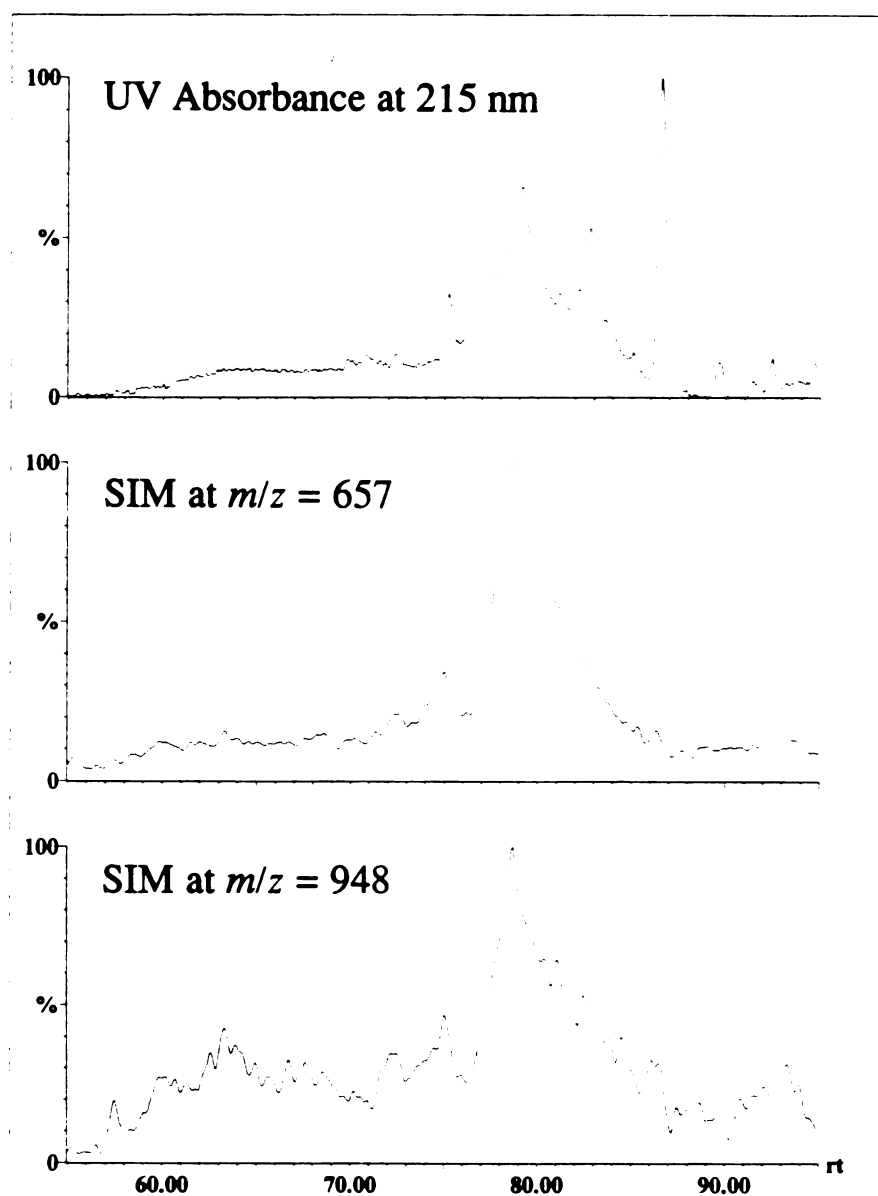


Figure 5. Results of HPLC/ESI/CID/MS with SIM experiment of the *O*-linked glycopeptide mixture isolated by jacalin-agarose. The top panel illustrates the HPLC chromatogram (UV absorbance at 215 nm) and the lower two panels illustrate the total ion current resulting from SIM at $m/z=657$ and 948, respectively, indicating that the majority of the *O*-linked glycopeptides contain disialyl Gal β 1 \rightarrow 3 GalNAc structures. The large peak in the HPLC chromatogram eluting at 87 min is a peptide which purified with the *O*-linked glycopeptides.

4.5 Discussion

While there is no consensus sequence for *O*-linked glycosylation, and therefore, no concrete method for predicting sites of *O*-linked glycosylation, a statistical study of the amino acid sequences around all known *O*-linked glycosylation sites by Wilson and coworkers in 1991 (Wilson *et al.*, 1991) indicated that the most prominent feature in the vicinity of these sites is the increased frequency of proline residues. Furthermore, Wilson and coworkers found that in areas of multiple glycosylation sites, prolines are found most frequently in positions -1 and +3 amino acid residues relative to the glycosylation sites. Based on this theory, the potential sites of *O*-glycosylation on sNGFR, as indicated in Figure 1, would fall primarily in the variable segment of the stalk region of the receptor. Our results using HPLC/ESIMS and HPLC/ESI/CID/MS with SIM of carbohydrate-specific fragment ions indicate that sNGFR contains sialylated *O*-linked glycosylation at Ser^{198, 210, 211} and Thr^{199, 205, 206, 216}, in the highly conserved portion of the stalk, and not where the *O*-linked glycosylation would be predicted. The fact that the glycosylation is located in the conserved portion of the stalk region of sNGFR indicates that the glycosylation has some functional significance which has been preserved through evolution. This could be a basic structural function for the stalk region itself or involve something more complex such as an influence on the folding and orientation of the nearby ligand binding domain and/or NGFR's association with p140^{trk}, the tyrosine kinase receptor which binds NGF, as discussed below.

The effect of multiple glycosylation between residues 198 and 216 may be simply to stabilize and extend the stalk structure of the receptor, since this area is fully available for reaction with lectins (Chapman *et al.*, 1994). It has been proposed by Jentoft that clusters of *O*-linked sugars raise the ligand-binding domain of single transmembrane receptors above the glycocalyx so that they are able to bind to their ligands (Jentoft, 1990). Jentoft has also suggested that clusters of *O*-linked sugars protect the stalk of transmembrane

receptors from proteolytic attack, however, several groups have shown that the extracellular portion of NGFR is released *in vivo* by proteolysis (Zupan *et al.*, 1989; DiStefano *et al.*, 1993), so clearly this is not the function of the glycans. Alternatively, the function of the *O*-linked oligosaccharides in the stalk region may be to influence the folding and orientation of the fourth cysteine repeat domain, which is involved in ligand binding. This is supported by previous findings which stated that deletions in the stalk region of NGFR eliminated cross-linking to NGF and significantly reduced binding to the monoclonal antibody ME20.4, which is known to bind to the fourth cysteine repeat domain of NGFR and blocks NGF binding to the receptor (Yan and Chao, 1991; Vissavajhala *et al.*, 1992).

As of yet, the true structure of the high-affinity NGF receptor has not been clearly defined in biochemical terms. Although there is evidence that the two receptors interact, attempts to directly cross-link the two receptors or co-immunoprecipitate them have been unsuccessful (Meakin *et al.*, 1992). Nevertheless, recent evidence supports the interaction of both receptors in the high affinity binding to NGF (Mahadeo *et al.*, 1994). Our findings that multiple sites of sialylated *O*-linked glycosylation is located in the conserved portion of the stalk region of sNGFR could also be involved with NGFR's association with p140^{trk}. This is supported by the fact that interaction of WGA with NGFR was shown to shift the equilibrium binding constant from low to high (Vale and Shooter, 1982; Grob and Bothwell, 1983) just as interaction of catalytic subunits of the *trk* tyrosine kinase family in the presence of NGFR increase the receptor's affinity for NGF (Hempstead *et al.*, 1992; Mahadeo *et al.*, 1994). In addition, WGA has been shown to interact with sialic acid residues on NGFR because neuraminidase partially blocked the affinity shift by WGA, and a succinylated derivative of WGA which was unable to bind sialic acid was inactive (Vale *et al.*, 1985). Since the seven *O*-glycans are sialylated, and there is only one site of *N*-linked glycosylation (Settineri *et al.*, 1992) it is very likely that WGA is binding to the *O*-linked carbohydrates in the stalk region. The fact that WGA and p140^{trk} have a similar effect on

UJJI LIDIMINI

binding of NGF to NGFR and that WGA is interacting with the *O*-linked carbohydrates, is further suggestion that p140^{trk} does interact with the low affinity NGFR in order for NGF to confer its neurotrophic stimulation. The interaction of p140^{trk} and p75 NGFR is supported by recent findings by Mahadeo and coworkers suggesting that the biological effects of NGF are derived from a novel kinetic binding site that requires the expression of both receptors (Mahadeo *et al.*, 1994). This binding site was proposed to be either (a) a multimeric complex of p75 NGFR and p140^{trk} proteins, or (b) an altered conformation of p140^{trk}, formed in the presence of p75 NGFR. Interestingly, very similar conclusions have been reached with the α , β and γ subunits of the IL-2 receptor (Grant *et al.*, 1992).

In general, the role(s) that protein glycosylation plays involve either modulation of biochemical properties such as bioactivity, folding and immunogenicity or determinants in molecular recognition events such as targeting and uptake of proteins. In addition, the location of these very large glycans must greatly influence the overall structure of the folded glycoproteins. While it is not known what the biological impact of the *O*-linked glycosylation on NGFR is, it is the fact that the glycosylation falls in a highly conserved portion of the receptor, and not in the area one would predict based on sites of *O*-glycosylation identified in many other proteins, is an indication that the glycosylation may be necessary for the structural and/or functional integrity of the receptor.

CHAPTER 5 CHARACTERIZATION AND SEQUENCING OF THE *N*-LINKED OLIGOSACCHARIDES FROM RECOMBINANT HUMAN NERVE GROWTH FACTOR RECEPTOR EXTRACELLULAR DOMAIN

5.1 Abstract

The extracellular domain of the human low-affinity nerve growth factor receptor expressed in Chinese hamster ovary cells (sNGFR) contains 222-amino acids and one site of *N*-linked glycosylation at Asn³². Due to microheterogeneity of *N*-linked oligosaccharide structures on glycoproteins, the characterization and sequencing of all glycoforms present at a given glycosylation site remains a formidable challenge. In order to obtain detailed sequence and branching information on all the *N*-linked oligosaccharides at Asn³² of sNGFR and determine the total number of glycoforms present, a novel method involving sequential exoglycosidase digestion combined with HPLC/electrospray ionization mass spectrometry after each digestion step was developed. Complete sequence and branching information was obtained on all ten glycoforms identified in the isolated glycopeptide mixture starting with only 800 pmole of protein. The results indicate three major oligosaccharide components which are complex fucosylated biantennary structures containing 0, 1 and 2 sialic acid residues and seven minor components which are complex fucosylated triantennary and tetraantennary structures containing 1-2 sialic acid residues and 0-3 additional galactose-*N*-acetylglucosamine units. Only the three major structures were previously identified by liquid secondary ion mass spectrometry of the derivatized released oligosaccharides (Settineri *et al.*, 1992).

5.2 Introduction

In order to study the structure of the NGF binding region in p75 NGFR, a recombinant soluble form of this protein (sNGFR) was prepared (Chapman *et al.*, 1994). This receptor contains the amino-terminal 222 amino acids ending just before the transmembrane region of the native full length receptor, and contains one asparagine residue at position 32 in the consensus sequence (Asn-Xxx-Ser/Thr, where Xxx = any amino acid except Pro) required for attachment of Asn-linked (*N*-linked) oligosaccharides (Kornfeld and Kornfeld, 1976). sNGFR also contains a Ser, Thr and Pro-rich carboxy-terminal domain, indicative of the presence of *O*-linked glycosylation (Jentoft, 1990; Wilson *et al.*, 1991), which is glycosylated with structures as described in Chapter 4.

In order to determine if and how the *N*-linked oligosaccharides were involved in NGF binding, the derivatized oligosaccharide structures were first investigated by LSIMS, before ESIMS was routinely available, and preliminary structures of the three major biantennary complex carbohydrate components were published in 1992 (Settineri *et al.*, 1992). Although there was evidence of the presence of other minor components, the limited quantities of sample available, and the relatively low yield of oligosaccharides obtained after enzymatic release and derivatization, prevented the characterization of these minor components at the time.

When microbore reversed phase HPLC/ESIMS became available in our laboratory, I was then able to characterize the oligosaccharides as glycopeptides in a proteolytic digest of sNGFR, eliminating the requirement for enzymatic release and derivatization of the carbohydrates. The *N*-linked glycopeptide glycoform mixture was identified in an endoproteinase glu-C digest of reduced and S-carboxymethylated sNGFR using HPLC/ESIMS coupled with collision induced dissociation (CID) and selected ion monitoring (SIM) of carbohydrate-specific fragment ions.

To obtain further detailed structural information on the oligosaccharide portion of the identified *N*-linked glycopeptides, a novel method was developed involving sequential digestion of the isolated glycopeptide mixture with exoglycosidases (Kobata, 1979) to remove the nonreducing terminal monosaccharides one at a time, followed by HPLC/ESIMS after each digestion step. By splitting the HPLC eluant 1:20 before entering the mass spectrometer, only 5% of the sample injected actually entered the mass spectrometer, and the remaining 95% of the HPLC purified sample was then recollected for digestion with the next exoglycosidase. Starting with 800 pmole of reduced and S-carboxymethylated sNGFR, after digestion with endoproteinase glu-C and isolation of the *N*-linked glycopeptide mixture, six sequential digestions were performed, enabling the identification and sequencing of ten different *N*-linked oligosaccharides attached to Asn³². The previous characterization of the released and derivatized oligosaccharides required 5 nmole of reduced and S-carboxymethylated protein, and yielded more limited structural information on only three oligosaccharide components (Settineri *et al.*, 1992).

5.3 Materials and Methods

Expression and Purification. The extracellular domain of human NGFR, which was kindly donated by Chiron Corporation, was expressed in Chinese hamster ovary cells as described in Chapter 4, page 99.

Reduction and S-Carboxymethylation. Reduction of purified sNGFR was performed as described in Chapter 4, page 99.

Carbohydrate Composition Analysis. Carbohydrate analyses were performed on acid hydrolyzed, reduced and S-carboxymethylated sNGFR (sNGFR-RCM) as described in Chapter 4, page 100.

Protease Digestion. Endoproteinase glu-C (*Staphylococcus aureus* V8, Boehringer Mannheim, Indianapolis, IN) digestions were performed in 50 mM ammonium bicarbonate, pH 7.8, at a weight ratio of 8% (2 x 4%) for 18-28 hr at 37°C.

Glycosidase Digestions. PNGase F (*Flavobacterium meningosepticum*, glycerol-free N-glycanase, Genzyme, Boston, MA) digestions of 5-10 nmole aliquots of intact reduced and S-carboxymethylated sNGFR were performed at 37°C in 100 mM ammonium bicarbonate, pH 8.0, adding 1 unit of enzyme/day for 5 days under argon. PNGase F (*Flavobacterium meningosepticum*, glycerol-free N-glycanase, Genzyme, Boston, MA) digestion of the isolated bovine fetuin and sNGFR N-linked glycopeptides were performed in water at 37°C for 18-24 hr adding aliquots of 0.25 units of enzyme at 0 and 8 hr (N-linked glycopeptides isolated from a trypsin digest of bovine fetuin were kindly donated by R. Reid Townsend). Digestions with the neuraminidases, β -galactosidase, β -N-acetylhexosaminidase and α -fucosidase (Oxford Glycosystems, Oxford, UK) were performed in 30 mM sodium acetate, pH 5.0, at 37°C for 3-7 hr. Neuraminidase from *Arthrobacter ureafaciens* was added to the digestion mixture to obtain a final enzyme concentration of 0.8-1.0 U/ml. Neuraminidase from Newcastle Disease Virus was added at a final concentration of 0.02-0.05 U/ml. β -galactosidase from *Streptococcus pneumoniae* was added at a concentration of 0.03-0.07 U/ml, β -N-acetylhexosaminidase from Chicken liver was added at 1.0-2.0 U/ml and α -fucosidase was added at 0.2 U/ml. Digestions with α - and β -mannosidases (Oxford Glycosystems, Oxford, UK) were performed in 30 mM sodium acetate, pH 5.0, 25 mM zinc chloride, at 37°C for 5 hr. α -mannosidase from Jack bean was added at final enzyme concentration of 3.5 U/ml, while β -mannosidase from *Helix pomatia* was added at 2.0-4.0 U/ml. Details of each digestion are summarized in Table I.

ABOE Derivatization. *n*-Octyl *p*-aminobenzoate (ABOE) derivatization (Poulter *et al.*, 1991) was performed using a reagent stock solution made up of 500 μ l MeOH, 5 μ l glacial acetic acid, 2 mg *n*-octyl *p*-aminobenzoate and 1 mg NaBH₃CN (final pH 4.5) at a

concentration of 30 μl per nmol of oligosaccharide. The reaction was carried out in an oven at 85°C for 2 hr. After cooling, the mixture was dissolved in 20 μl MeOH and excess reagent extracted twice with 150 μl H₂O/CHCl₃ (1:1, v/v).

***N*-linked Oligosaccharide Profiling.** Separation of sNGFR oligosaccharides was performed on the glycans released from 800 pmole of the isolated *N*-linked glycopeptide by PNGase F. Analyses were performed using a Dionex Bio-LC HPAE-PAD chromatography system described under carbohydrate composition analysis (Hardy and Townsend, 1988). Eluant 1 was H₂O, eluant 2 was 200 mM NaOH, eluant 3 was 100 mM NaOH and eluant 4 was 0.5 M NaOAc, pH 5.5. Identification of oligosaccharides was determined by comparing the results to injection of a standard fetuin oligosaccharide mixture (Dionex).

Peptide and glycopeptide HPLC. Microbore HPLC/ESIMS analyses were performed as described in Chapter 4, page 101.

Mass spectrometry. LSIMS analysis of the derivatized oligosaccharides was carried out using a Kratos MS50S double focusing mass spectrometer as described in Chapter 3, page 74, except that spectra were acquired in the negative ion mode using 1:1 glycerol:thioglycerol as the matrix (no TFA). ESI spectra were acquired as described in Chapter 4, page 100. MALDI-TOF analysis was performed on a Kratos Kompact MALDI 3 in the linear mode using a laser power of 81. Approximately 100 pmole of dried HPLC-purified sNGFR-RCM was dissolved in 1 μl of water and added to 1 μl of sinapinic acid matrix on the sample slide.

5.4 Results

While the molecular weight of the 222-amino acid sequence of sNGFR is approximately 25 kDa, MALDI-TOF analysis of sNGFR-RCM, shown in Figure 1, yielded a broad peak

centered at 31.9 kDa, indicating the presence of some post-translational modification such as glycosylation.

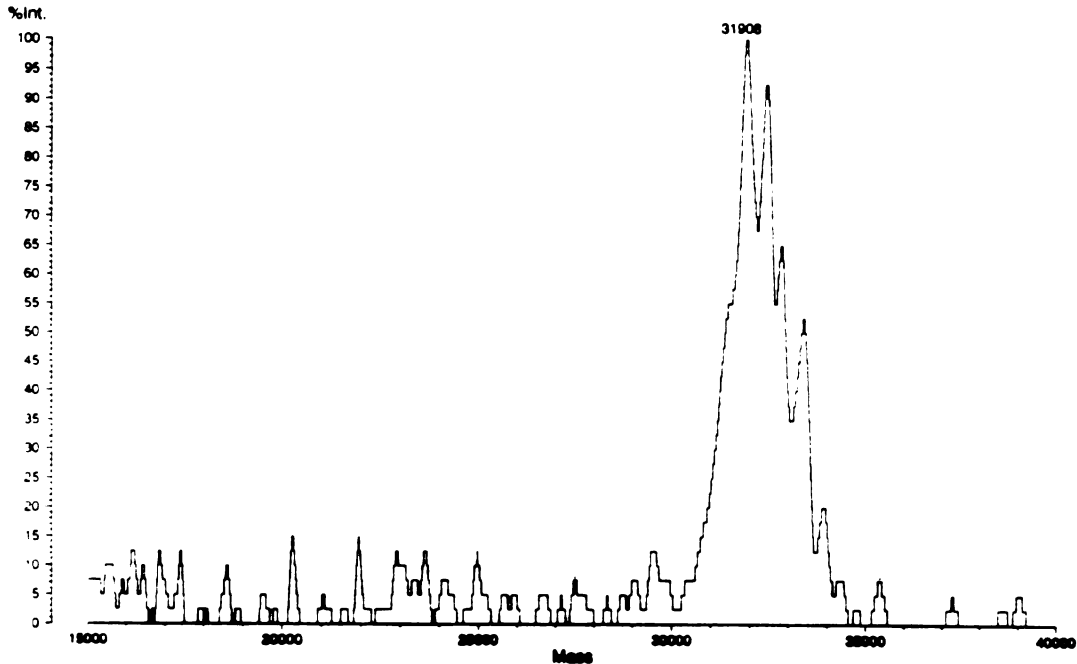


Figure 1. MALDI-TOF analysis of 100 pmole sNGFR-RCM using sinapinic acid as the matrix. Using bovine serum albumin (65,605 Da) for external calibration, corrected average mass is 31.3 kDa. The shoulder peaks are assumed to be more glycosylated species than the peak top mass at 31.3 kDa.

In addition, SDS-PAGE analysis of sNGFR-RCM yielded a wide band of apparent molecular weight of approximately 50-55 kDa, as shown in lane 2 of Figure 2. Treatment of sNGFR-RCM with N-glycanase (lane 3), and neuraminidase + O-glycanase (lane 4) changed the migration of sNGFR-RCM on analytical SDS-PAGE relative to the migration of untreated sNGFR-RCM, as illustrated in Figure 2, indicating the presence of *N*- and *O*-linked glycosylation.

Carbohydrate Composition Analysis—In order to first determine the composition and relative quantities of carbohydrates on sNGFR, intact, reduced and S-carboxymethylated protein was hydrolyzed using different conditions specific for quantitation of neutral sugars, amino sugars and sialic acid (Hardy *et al.*, 1988). Analyses were performed using

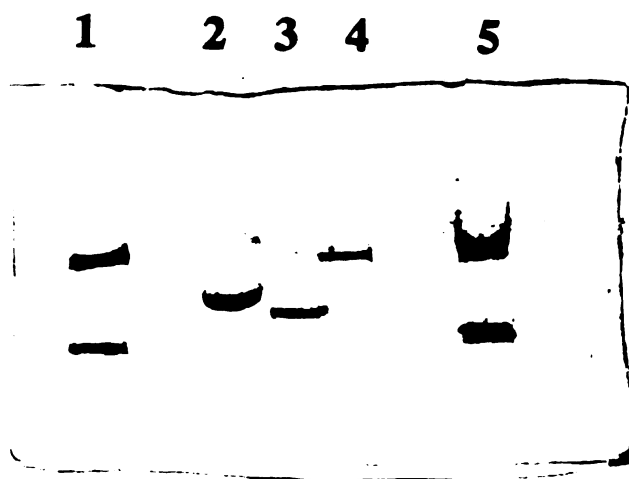


Figure 2. SDS-PAGE results for digestion of sNGFR-RCM (lane 2) with N-glycanase (lane 3) and neuraminidase + O-glycosidase (lane 4). Lanes 1 and 5 are molecular weight standards containing bovine serum albumin (66 kDa, top band) and ovalbumin (45 kDa, lower band).

HPAE-PAD (Hardy and Townsend, 1988) which yielded the results listed in Chapter 4, Table I, page 101. These results suggested the presence of a single site of *N*-linked glycosylation containing fucosylated and sialylated complex type structures as well as many possible sites of sialylated *O*-linked oligosaccharides as described in Chapter 4.

Mass spectrometric peptide mapping—To confirm the predicted amino acid sequence of sNGFR and further characterize the *N*-linked oligosaccharides attached to this recombinant glycoprotein, sNGFR was reduced and carboxymethylated before digestion with

endoproteinase glu-C, and the digest analyzed by ESIMS to generate a peptide map (based on the molecular weights obtained) as shown in Figure 3. The digest was separated by microbore reversed phase HPLC/ESIMS. All but residues (24-41) and (195-222), were identified as unmodified peptides by mass spectrometry of the proteolytic digests, indicated by the thin lines in Figure 3. Residues (24-41), indicated by the thick line in Figure 3, encompasses Asn³², the consensus site for *N*-linked glycosylation (Bause and Hettkamp, 1979). Residues (195-222), also indicated by a thick line, is the C-terminal portion of the protein sequence which is rich in serine and threonine residues, indicating that this part of the protein contains *O*-linked glycosylation as described in Chapter 4.

N-linked oligosaccharide isolation—In order to isolate the *N*-linked oligosaccharides, 5 nmole of reduced and carboxymethylated rhNGF-R were treated with N-glycanase. After

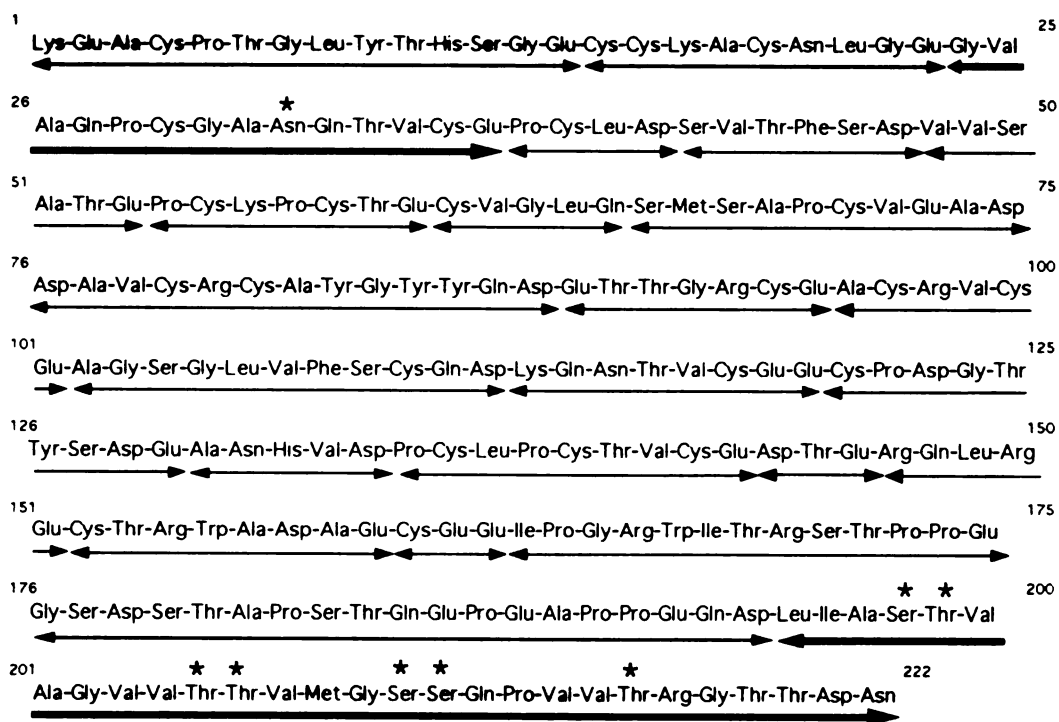


Figure 3. Peptide map of reduced and carboxymethylated sNGFR based on the molecular weights obtained from HPLC/ESIMS analysis of the endoproteinase glu-C digest. Peptides identified as predicted endoproteinase glu-C unmodified peptides are underlined with the thin lines. Peptides identified as endoproteinase glu-C glycopeptides are underlined with the thick lines. Residues (195-222) of the protein which was shown to contain *O*-linked glycosylation, the analysis of which is described in Chapter 4.

N-glycanase treatment, the free oligosaccharides were separated from the intact protein by analytical HPLC on a C4 column. The deglycosylated protein was recovered and the oligosaccharides which eluted in the void volume were derivatized with the hydrophobic chromophore, *n*-octyl *p*-aminobenzoate to form the ABOE derivative (Poulter *et al.*, 1991).

Reversed phase HPLC of the resulting derivatized oligosaccharides indicated three major species labeled 1 through 3 in Figure 4. Since only one chromophore is present per molecule of oligosaccharide, the relative peak intensities of the derivatized oligosaccharides on reversed phase HPLC indicates the relative quantities of each oligosaccharide species present on the glycoprotein.

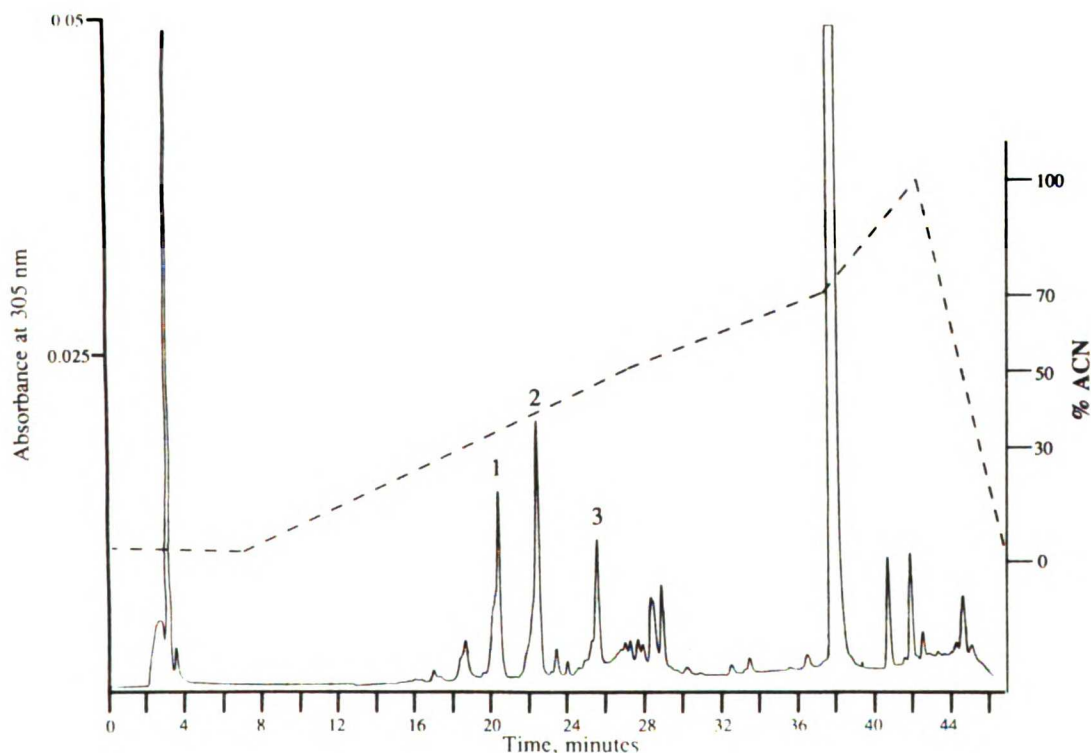


Figure 4. Reversed phase HPLC of ABOE-derivatized *N*-linked oligosaccharides released from sNGF-R RCM. Derivatized oligosaccharides were eluted on an analytical C18 column with water and acetonitrile (no TFA). Peaks eluting after fraction 3 are due to residual ABOE reaction components.

Negative ion LSIMS of peak 3 in Figure 4, shown in Figure 5, yielded an $M-H^- = 2019.4$ and fragment ions corresponding to successive losses of Hex, HexNAc and Hex. Using the molecular weights of the known sugars present from the carbohydrate composition analysis, this molecular ion could only correspond to a monosaccharide composition of Hex₅HexNAc₄fucose₁ABOE₁. Negative ion LSIMS of the largest peak, fraction 2, shown in Figure 6, yielded an $M-H^- = 2310.0$ and fragment ions at $m/z = 2019.4$, 1857.4, 1654.2 and 1491.8, which correspond to successive losses of sialic acid (NeuAc), Hex, HexNAc and Hex. Another weaker series of ions was also identified at $m/z = 2147.8$, 1944.4 and 1783.0, showing successive losses of Hex, HexNAc and Hex, indicating the presence of two branches of different structures.

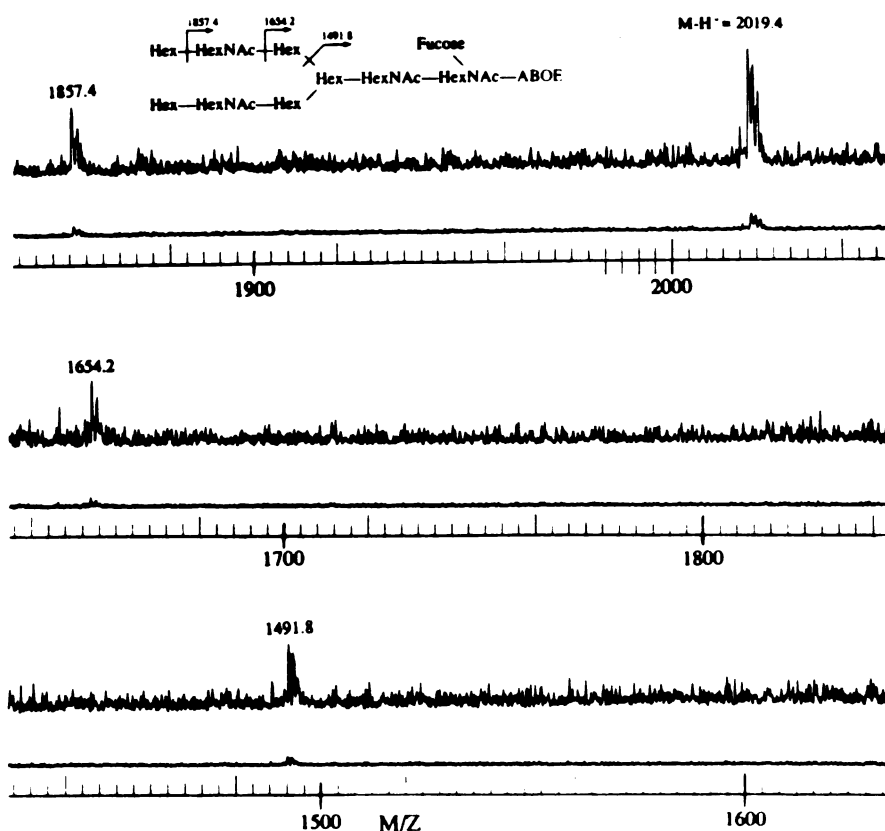


Figure 5. Negative ion LSIMS of peak 3 from figure 4. This fraction yielded an $MH^- = 2019.4$ and fragment ions corresponding to successive losses of Hex, HexNAc and Hex. This molecular ion could only correspond to a monosaccharide composition of Hex₅HexNAc₄fucose₁ABOE₁.

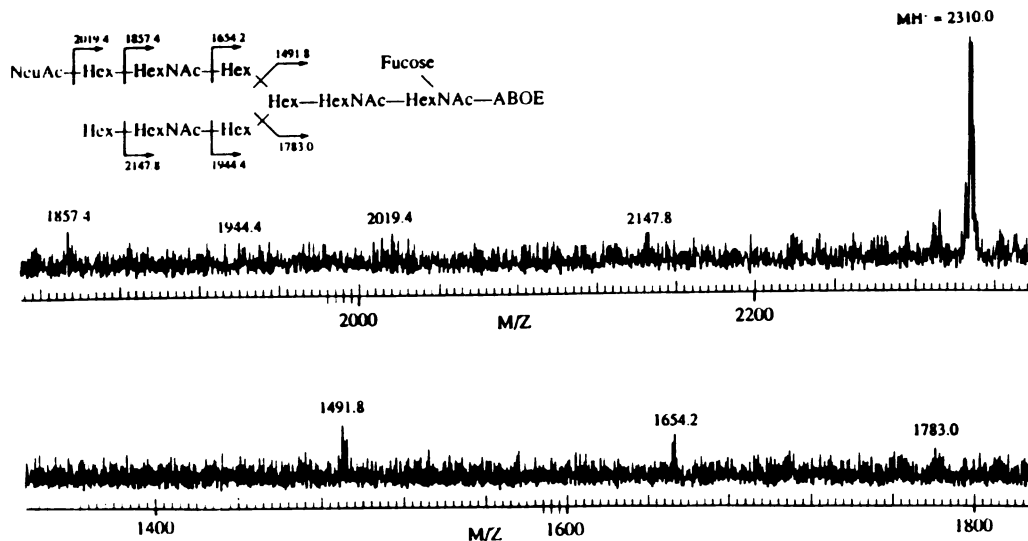


Figure 6. Negative ion LSIMS of peak 2 from figure 4. This fraction yielded an $M-H^- = 2310.0$ and fragment ions at $m/z = 2019.4$, 1857.4 , 1654.2 and 1491.8 , which correspond to successive losses of sialic acid (NeuAc), Hex, HexNAc and Hex. The molecular weight represents a monosaccharide composition of NeuAc₁Hex₅HexNAc₄fucose₁ABOE₁.

Negative ion LSIMS of peak 1 from Figure 4 yielded a weak $M-H^- = 2601.4$ and fragment ions at $m/z = 2309.2$ and 1946.0 , corresponding to successive losses of NeuAc and Hex-HexNAc. Figure 4 contains two smaller peaks eluting before peak 1, which could be oligosaccharide species with additional sialic acid residues. However, LSIMS analysis of these fractions was unsuccessful. From these data, the major *N*-linked structures of rhNGF-R were determined to be fucosylated biantennary complex oligosaccharides. Based on the UV absorbance of the three fractions containing the ABOE derivatized oligosaccharides, approximately 45% are monosialyl, 35% are disialyl and 20% are asialyl.

Glycopeptide Identification—The *N*-linked glycopeptides were specifically identified in the endoproteinase glu-C digest using HPLC/ESIMS combined with CID (HPLC/ESI/CID/MS) and selected ion monitoring (SIM) of carbohydrate-specific fragment ions (Duffin *et al.*, 1992; Carr *et al.*, 1993). Based on the data from the carbohydrate composition analysis, the following ions were monitored: $m/z=204$ (GlcNAc⁺ or GalNAc⁺), $m/z=292$ (NeuAc⁺), $m/z=366$ ([Gal-GlcNAc]⁺ or [Gal-GalNAc]⁺) and $m/z=657$ ([NeuAc-Gal-GlcNAc]⁺ or [NeuAc-Gal-GalNAc]⁺). The peptide containing amino acid residues (24-41) in the endoproteinase glu-C digest was identified using HPLC/ESI/CID/MS, and only as a glycopeptide mixture. Since the nonglycosylated peptide was not found, this suggested that Asn³² was completely glycosylated with *N*-linked oligosaccharides.

One aliquot of the endoproteinase glu-C digest (400 pmole) was analyzed by microbore reversed phase HPLC/ESIMS and another 400 pmole aliquot was analyzed by HPLC/ESI/CID/MS with selected ion monitoring (SIM) in order to selectively detect the glycopeptides in the digest. Once the glycopeptide-containing peaks were identified, the electrospray mass spectra of the glycopeptides from the first analysis could be identified by direct comparison the two UV chromatograms. Figure 7 illustrates the results of the HPLC/ESI/CID/MS experiment for the endoproteinase glu-C digest with SIM at $m/z=204$ and $m/z=657$. The single large peak in the lower two panels is the result of fragmentation of the *N*-linked glycopeptide residues (24-41) which eluted at approximately 22 min. as shown in the corresponding peak in the top panel (UV absorbance trace). The ESI mass spectrum of this glycopeptide mixture is shown in Figure 8, which indicates the presence of at least six glycopeptide glycoforms. The molecular mass measurements of the *N*-linked glycopeptides identified in this spectrum, combined with the carbohydrate composition data and the known sequence of the peptide portion of the glycopeptide (²⁴GVAQPC*GANQTVCEPC*LD⁴¹, 1980.2 Da, where C*=S-carboxymethylcysteine), permitted the determination of the compositions of the oligosaccharides and, therefore, to

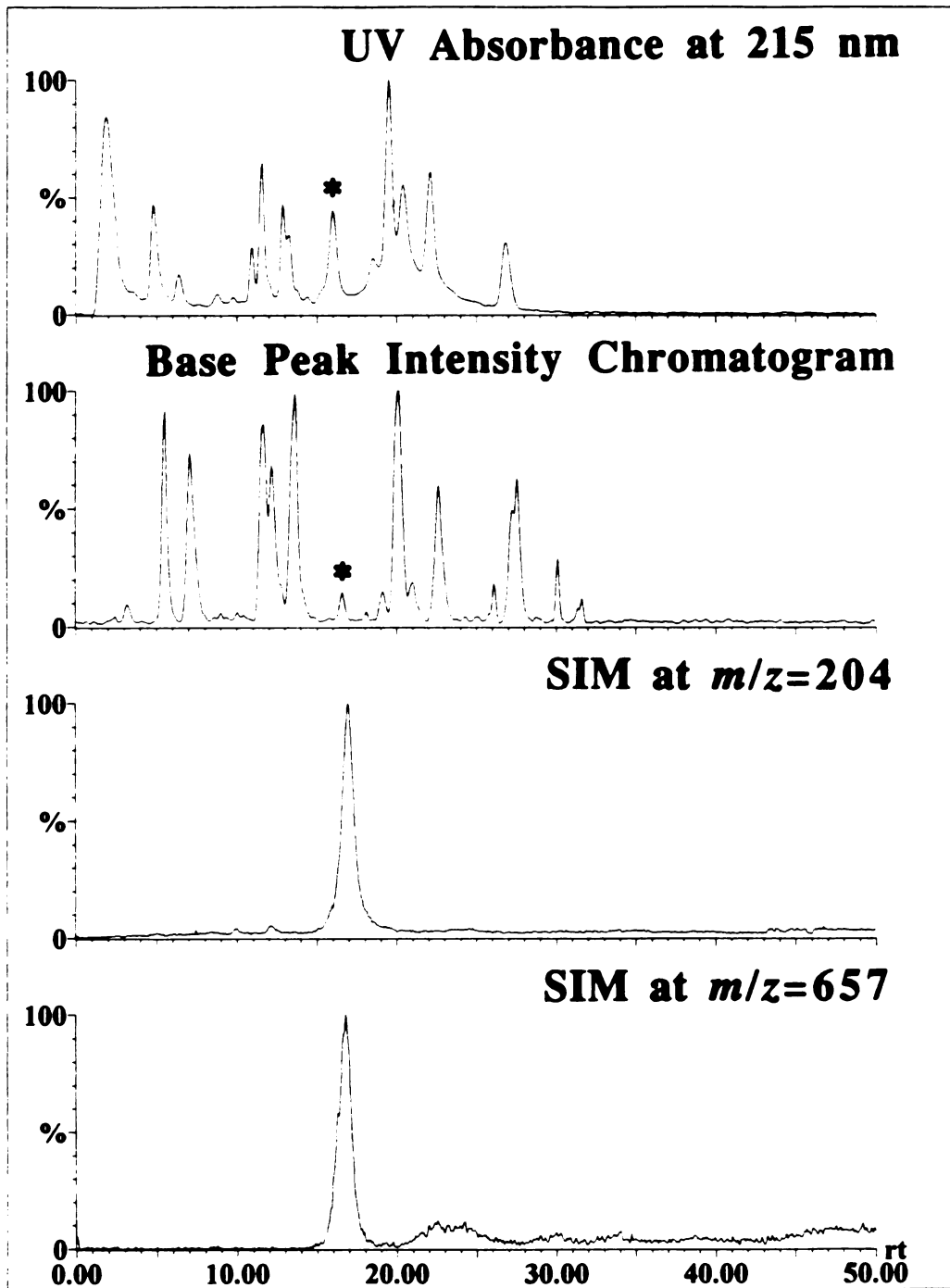


Figure 7. Results of the HPLC/ESI/CID/MS experiment on the endoproteinase glu-C digest of sNGFR with SIM at $m/z=204$ and $m/z=657$. The single large peak in the lower two panels is the result of fragmentation of the *N*-linked glycopeptide residues (24-41) which eluted at approximately 16 min as shown in the corresponding (starred) peaks in the top two panels (UV absorbance and base peak intensity traces). Range=0.05 AUFS. Solvent A=0.1% TFA, solvent B=0.08% TFA in ACN. Gradient program was 0% B for 5 min, 0-20% B in 20 min, 20-40% A in 45 min.

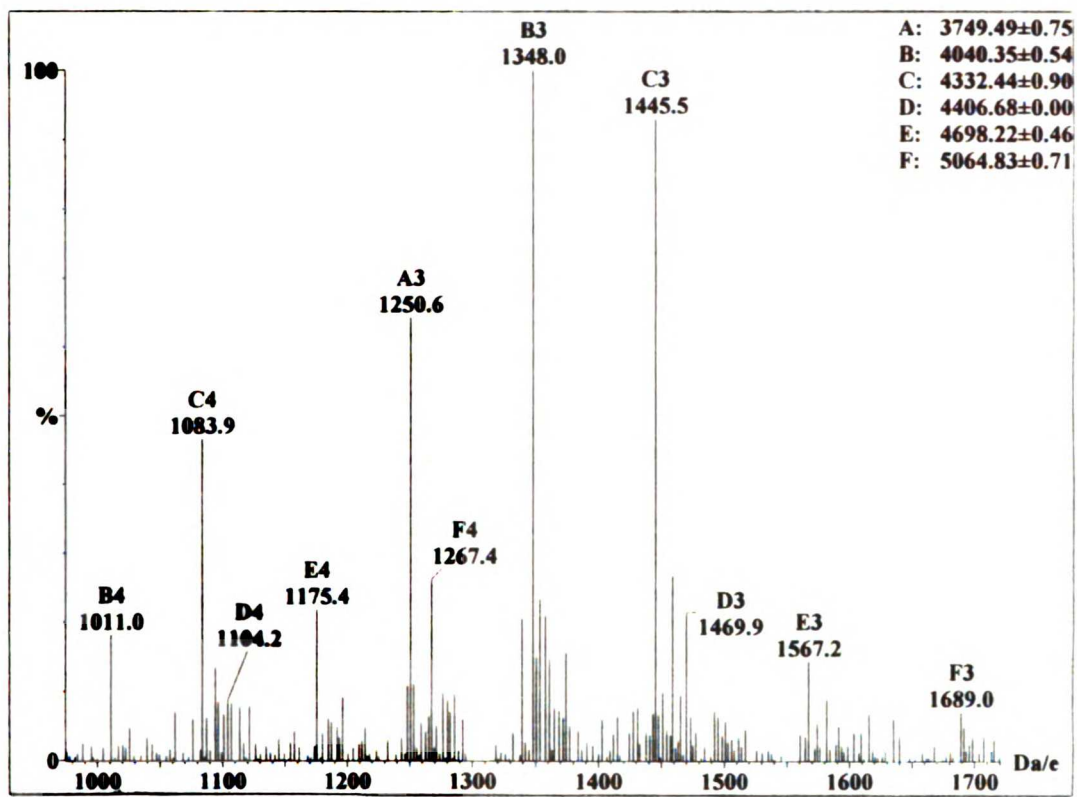


Figure 8. HPLC/ESI mass spectrum of sNGFR endoproteinase glu-C *N*-linked glycopeptide mixture identified by HPLC/ESI/CID/MS with SIM, indicating the presence of at least six glycopeptide glycoforms. Proposed structures for the masses obtained are listed in Table I.

predict the class and size of their structures. By subtracting the mass of the peptide (24-41) from the molecular masses of the intact glycopeptides, the composition of the oligosaccharides attached to this peptide represent fucosylated biantennary, triantennary and tetraantennary (or triantennary with an additional lactosamine (Gal-GlcNAc) structures containing 0 to 3 sialic acid residues, as listed in Table I.

Oligosaccharide profiling by HPAE-PAD—Approximately 800 pmole of the *N*-linked glycopeptide mixture was isolated by microbore reversed phase HPLC and then treated with *N*-glycanase, in order to obtain the mixture of *N*-linked oligosaccharide(s) attached to Asn³² for oligosaccharide mapping by HPAE-PAD. The oligosaccharides with the most negative charges (i.e. sialic acid residues) are retained longest on the anion exchange

ESIMS Molecular Masses (Da) and Proposed Oligosaccharide Structures

| Intact Glycopeptides | Neuraminidase Digest | β -Galactosidase Digest | β -Galactosidase Digest + β -Hexosaminidase Digest | Proposed Structure |
|---------------------------------|--|--|--|--|
| 3749.2 G-GN G-GN | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | Fucosylated Biantennary |
| 4040.4 SA SA-G-GN SA-G-GN | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | Fucosylated Triantennary |
| 4332.3 SA SA-G-GN SA-G-GN | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | Fucosylated Tetraantennary |
| 4405.0 SA SA-G-GN SA-G-GN | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | Fucosylated Triantennary (Lac) ₁ |
| 4698.2 SA SA-G-GN SA-G-GN | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | Fucosylated Tetraantennary (Lac) ₁ |
| 5064.8 SA SA-G-GN SA-G-GN | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | G-GN-GN ₂ F G-GN-GN ₂ F | Fucosylated Tetraantennary (Lac) ₂ |
| | | | | Fucosylated Triantennary (Lac) ₃ |

Table 1. HPLC/Electrospray mass spectrometry data obtained from sequential glycosidase digestion of sNGFR N-linked glycopeptides. SA=sialic acid, G=galactose, GN=N-acetylglucosamine, F=fucose, Lac=lactosamine (galactosyl-N-acetylglucosamine), γ = trimannosyl core common to all N-linked oligosaccharides. The brackets and arrows point to structures resulting from digestion with the next enzyme.

column used for this method (Hardy and Townsend, 1988). A standard mixture of fetuin oligosaccharides which are known to contain disialo, trisialo and tetrasialo components was run before and after the sNGFR oligosaccharides to use as an elution standard. Figure 9 shows an overlay plot of the results of the fetuin standard mixture and the sNGFR oligosaccharide mixture. Three major groups of sNGFR oligosaccharides elute at times

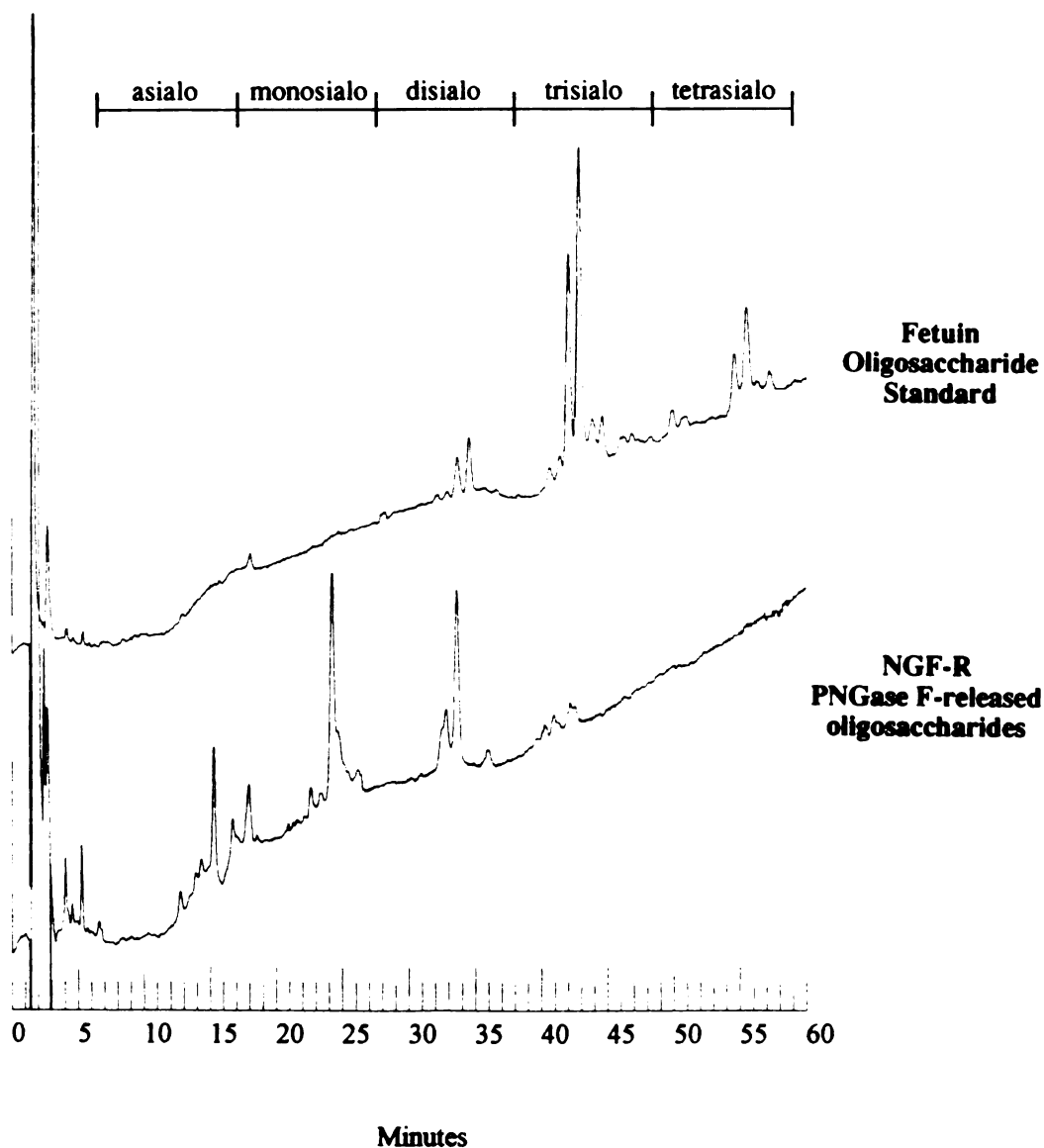


Figure 9. HPAE-PAD analysis of sNGFR *N*-linked oligosaccharides released by N-glycanase. Overlay plot of the results of the bovine fetuin standard oligosaccharide mixture and the sNGFR oligosaccharide mixture. The sNGFR oligosaccharides elute at times which correspond primarily to asialo, monosialo and disialo structures, and a small amount of trisialo structures.

correspond primarily to asialo, monosialo and disialo structures, with a minor amount of possible trisialo structures. This agrees with both the previous ABOE-derivatized oligosaccharide data (Settineri *et al.*, 1992) shown in Figure 4 and the HPLC/ESIMS data of the glycopeptide mixture, shown in Figure 8. Interestingly, the relative distribution of the forms of the biantennary structures are the same by all three methods of analysis. The two independent chromatographic methods (HPAE-PAD of the free oligosaccharides and reversed phase HPLC of the ABOE derivatives) both indicate that the monosialo species is the major species, followed by the disialo and then the asialo species. The HPLC/ESIMS spectrum shown in Figure 8 has a similar distribution for the three triple charged ions corresponding to the monosialo, disialo and asialo biantennary glycopeptides. Similar results were also obtained in the HPLC/ESIMS data for a chymotryptic glycopeptide mixture containing residues (10-45) from sNGFR (Medzihradzky *et al.*, 1993). This indicates that for glycopeptides with the same peptide portion and differing but similar oligosaccharide components, the relative ion abundances (in a limited m/z range) in the electrospray spectra indicate the relative quantities of the species present.

Sequencing of N-linked oligosaccharides using exoglycosidase digestion—In order to confirm the predicted oligosaccharide structures and obtain detailed sequence and branching information on all the *N*-linked oligosaccharides and determine the total number of glycoforms present in the mixture, approximately 800 pmole of the *N*-linked glycopeptide mixture, which was isolated during the HPLC/ESIMS and HPLC/ESI/CID/MS experiments described above, was digested with specific exoglycosidases (one at a time), to sequentially remove the predicted nonreducing terminal monosaccharides, and the resulting mixture analyzed by HPLC/ESIMS after each digestion step. By combining the specificities of the exoglycosidases with the change in the molecular masses before and after each digestion, sequence information was obtained for all the identified oligosaccharides as well as identify additional structures, and their structures determined.

To confirm the published specificity and determine appropriate digestion conditions of

the enzymes to be used, two glycopeptides containing oligosaccharides of known sequence isolated from a tryptic digest bovine fetuin (kindly donated by R. Reid Townsend) were first sequentially digested and analyzed by HPLC/ESIMS. When used at the lowest concentrations suggested by the manufacturer, the specificities of the enzymes were determined to be as published, however using higher concentrations of the enzymes sometimes resulted in a loss of linkage specificity of the enzymes and in extreme cases, caused the enzyme to act as an endoglycosidase. For example digestion of the fetuin glycopeptides with β -N-acetylhexosaminidase from *Streptococcus pneumoniae* at the higher concentration recommended by the seller for nonlinkage-specific cleavage of terminal GlcNAc residues caused complete cleavage of the oligosaccharide from the glycopeptide, leaving only a single GlcNAc attached to the peptide. In addition, the β -galactosidase from *Streptococcus pneumoniae* was found to contain some protease activity, because the larger fetuin glycopeptide, which contained more than 30 amino acid residues, was hydrolyzed at the amino-terminus of the peptide, resulting in the removal of several amino acids from the peptide portion of the glycopeptide. Additional digestion of this peptide with more β -galactosidase resulted in additional hydrolysis of the peptide. It was determined that most of the digestions were complete in a few hr or less, rather than the typical 12-24 hr used for these digestions, most likely due to the very small volumes used (10-25 μ l). It was also determined that the same buffer, 30 mM NaOAc, pH 5.0, worked well for all the enzyme digestions, however each enzyme required a unique final concentration in the reaction mixture, based on enzyme activity. A summary of the reaction mixtures used is given in Table II.

| Enzymes (in sequencing order) | Enzyme Source | Linkage Specificity | Reference | Concentration (U/ml) | Digestion Buffer | Digest Time (hours) |
|----------------------------------|-------------------------------------|------------------------------------|-----------|-------------------------|-----------------------------|------------------------|
| 1. Neuraminidase | Newcastle Disease Virus | NeuNAc } α 2-3,8R NeuGAc | 1,2 | 0.02-0.05 | Buffer A | 16-22 |
| 2. β -Galactosidase | <i>Streptococcus pneumoniae</i> | Gal β 1-(4GlcNAc 4GalNAc) | 3,4 | 0.03-0.07 | Buffer A | 16-22 |
| 3. β -GlcNAcase | Chicken liver | GlcNAc β 1-2,3,4R | 5 | 1.0 | Buffer A | 16-22 |
| 4. α -Mannosidase | Jack bean | Man α 1-2,3,6Man | 6,7 | 3.5 | A + 25 mM ZnCl ₂ | 16-22 |
| 5. β -Mannosidase | <i>Helix pomatia</i> | Man β 1-4GlcNAc | 8 | 2.0-4.0 | A + 25 mM ZnCl ₂ | 15-22 |
| 6. β -GlcNAcase | Chicken liver | GlcNAc β 1-2,3,4R | 5 | 2.0 | Buffer A | 12 |

Buffer A - 30 mM Sodium Acetate, pH 5.0

1. Corfield et al., 1982
2. Paulson et al., 1982
3. Kobata, 1979
4. Distler and Jourdan, 1973
5. Chelibonova-Lorer et al., 1985
6. Li and Li, 1973
7. Tai et al., 1975
8. McCleary, 1983

Table II. Exoglycosidase reaction mixtures used for sequencing sNGFR N-linked oligosaccharides

After confirming the correct enzyme specificities and appropriate digestion conditions, the sNGFR glycopeptide mixture was first digested with Newcastle Disease Virus neuraminidase, which specifically removes any terminal α 2 \rightarrow 3 linked sialic acid residues linked to Gal or GlcNAc from the glycopeptides (Corfield *et al.*, 1982; Paulson *et al.*, 1982). The digest was then analyzed by HPLC/ESIMS and the change in the molecular masses used to determine the number of α 2 \rightarrow 3 linked sialic acid residues removed from each glycopeptide component. Since the HPLC eluant was split 1:20, 95% of the HPLC-purified sample was recollected for digestion with the next exoglycosidase. Figure 10A illustrates the ESIMS spectrum of the neuraminidase-digested glycopeptide mixture. This spectrum contains only four components, one of which corresponds to a structure not

identified in Figure 8, indicating that a great deal of the heterogeneity among the glycopeptide glycoforms was in the number of sialic acid residues attached at the nonreducing terminus. The highest molecular mass structure in Figure 10A was not identified as a sialylated component in Figure 8 because large, sialylated oligosaccharides attached to peptides are often not detected by ESI instruments with quadrupole mass analyzers due to the fact that protonation based solely on the peptide portion occurs, and as the mass of a glycopeptide increases due to the size of the oligosaccharide, the number of protons added to the molecule does not increase. Therefore, the mass-to-charge ratio increases and often falls in the mass range where the sensitivity of a quadrupole mass analyzer is compromised ($>m/z$ 2000). The molecular masses of the four components identified in Figure 10A correspond to fucosylated bi-, tri-, tetra- and pentaantennary oligosaccharides, based on the molecular masses listed in Table I. However, pentaantennary structures are not known to exist, so further enzyme digestions were required to determine if the largest structure was a tetraantennary oligosaccharide with one extra lactosamine unit on one of the branches or possibly a triantennary oligosaccharide with two extra lactosamine units on one or two branches, etc...

The neuraminidase-digested glycopeptide mixture was then digested with β -galactosidase from *Streptococcus pneumoniae*, which removes terminal galactose residues linked $\beta 1 \rightarrow 4$ to GlcNAc or GalNAc (Distler and Jourdan, 1973; Kobata, 1979). After analysis of the β -galactosidase digest by HPLC/ESIMS, the change in the masses from before the digestion indicated whether the oligosaccharides attached to the glycopeptides were biantennary, triantennary or tetraantennary. For example, a triantennary structure will lose three galactose residues (a mass difference of 486 Da) upon β -galactosidase digestion, while a tetraantennary oligosaccharide will lose four galactose residues (a mass difference of 648 Da). In addition, a triantennary oligosaccharide which contains an additional lactosamine unit on one of its branches (the same mass as a tetraantennary structure) will only lose three galactose residues. Therefore, this method was used to distinguish more

branched structures from those with fewer branches which are extended with additional lactosamine units. Seven components were identified in the HPLC/ESIMS spectrum shown in Figure 10B. Component E ($m_{w,avg}=3789.6$) was determined by further β -galactosidase digestion to be an incomplete cleavage product where only 2 Gal residues were removed from a triantennary structure. Based on the remaining molecular masses obtained, however, six different oligosaccharide components were identified, based on the calculated oligosaccharide compositions, as listed in Table I. These include bi-, tri- and tetraantennary structures as well as tri- and tetraantennary structures with additional lactosamine units. Additional digestion with β -galactosidase did not change the spectrum except for the elimination of component E as described above. The compositions of the other components in Table I corresponding to oligosaccharides which still contain galactose residues, result from the presence of lactosamine structures on the branches.

To further confirm the data from the previous digestion, the β -galactosidase-digested glycopeptides were then treated with β -N-acetylhexosaminidase from chicken liver, which removes terminal GlcNAc residues linked $\beta 1\rightarrow 2, 3$ or 4 to any sugar residue. By using this specific enzyme which only removes GlcNAc and not GalNAc residues, the β -galactose residues cleaved by the β -galactosidase must have been linked to GlcNAc and, therefore, the linkage in the structures is Gal $\beta 1\rightarrow 4$ GlcNAc. Four components were identified in the HPLC/ESIMS spectrum shown in Figure 10C and listed in Table I. As shown in Table I, digestion of the bi-, tri- and tetraantennary components of average molecular mass of 3424.6, 3628.3 and 3829.4 Da, respectively, in Figure 10B resulted in a single component with the same mass of 3017.9 Da, which is the trimannosylchitobiose core attached to residues (24-41). The other three components correspond to tri- and tetraantennary oligosaccharides containing one, two and three additional lactosamine units, as shown in Table I. The molecular mass of 4113.9 Da corresponds to a structure which was not identified in the previous β -galactosidase digest. A similar structure was proposed from the molecular mass of 4112.5 Da after the neuraminidase digest, but since this mass was not

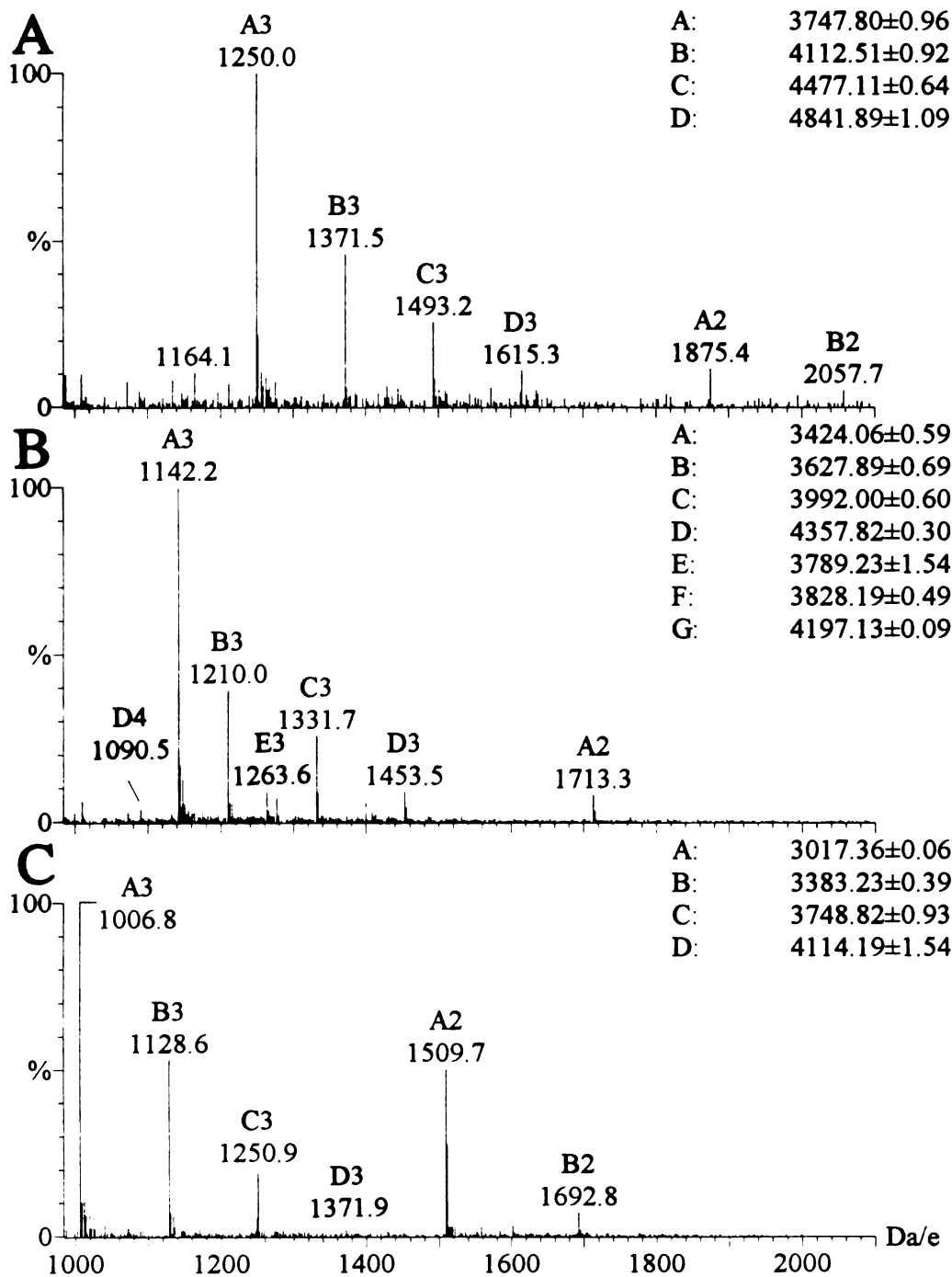
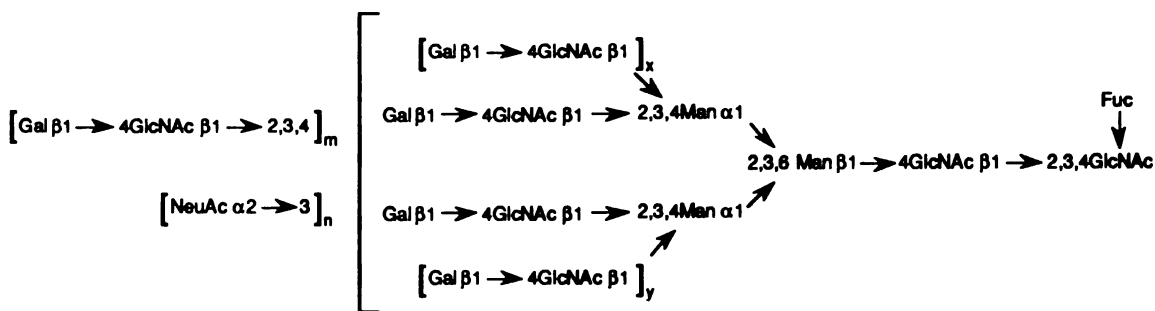


Figure 10. Exoglycosidase digestion analysis of sNGFR endoproteinase glu-C N-linked glycopeptide mixture. (A) HPLC/ESI mass spectrum of the neuraminidase-digested glycopeptide mixture shown in Figure 7. (B) HPLC/ESI mass spectrum of the β -galactosidase-digested glycopeptide mixture shown in Figure 9A. (C) HPLC/ESI mass spectrum of the β -galactosidase + β -N-acetylhexosaminidase-digested glycopeptide mixture shown in Figure 9B. Proposed structures for the masses obtained are listed in Table I.

identified in the β -galactosidase digest, it must have resulted from a previously unidentified structure in the mixture.

The structures proposed that contain additional lactosamine structures were distinguished from incomplete digestion products (or those with linkages not cleaved by the enzymes used) by redigestion of the mixture digested with β -N-acetylhexosaminidase with β -galactosidase and then β -N-acetylhexosaminidase again followed by analysis by HPLC/ESIMS. The data indicated that all the Gal and GlcNAc residues could be removed after further digestion (data not shown), confirming the presence of polylactosamine structures. As listed in the summary of glycosidase digestion information in Table II, three more digestions were performed in order to determine the linkage of the mannose residues in the trimannosyl core using α -mannosidase from Jack bean (Li and Li, 1973; Tai *et al.*, 1975) and β -mannosidase from *Helix pomatia* (McCleary, 1983), and determine where the fucose was attached to the oligosaccharides using the β -N-acetylhexosaminidase from chicken liver. After digestion with α -mannosidase, β -mannosidase, and finally β -N-acetylhexosaminidase, a molecular mass of 2327.7 Da was obtained (data not shown) which corresponds to the peptide (24-41) with a single GlcNAc and a Fuc residue attached, indicating that the fucose residue is linked to the GlcNAc which is attached to the asparagine residue in the peptide. Using the linkage and residue specificities of the five different exoglycosidases and the molecular masses obtained from HPLC/ESIMS data after each digestion step, the sNGFR N-linked oligosaccharides determined to be attached to Asn³² are shown in Figure 11.



where $m = 0-3$, $n = 0-2$ and $x + y = 0-2$

Figure 11. Sequence, branching and linkage data obtained on sNGFR oligosaccharides by exoglycosidase sequencing and HPLC/ESIMS.

5.5 Discussion

HPLC/ESIMS was used to first identify the *N*-linked glycopeptide mixture from sNGFR using HPLC/ESI/CID/MS with selected ion monitoring for carbohydrate-specific fragment ions. Then, HPLC/ESIMS in combination with digestion with highly specific exoglycosidases was employed to determine detailed sequence, branching and linkage information on the *N*-linked oligosaccharide mixture. The spectrum in Figure 8 combined with the HPEA-PAD data in Figure 9 and the ABOE-derivatized oligosaccharide data in Figure 4 indicate that the mono-, di- and trisialo biantennary structures comprise approximately 90% of the oligosaccharides attached to sNGFR. However, only the method described here could identify and characterize the seven minor oligosaccharide components comprising the other 10% of the structures. Since the analysis was started with less than 800 pmole of total glycopeptide, about 720 pmole are biantennary glycopeptides and the remaining seven (or probably more) structures identified could make up 80 pmole or less to start with, leaving 10-12 pmole maximum for any of these other components. Since only 5% of the total sample injected for each HPLC/ESIMS experiment actually entered the mass spectrometer, this means that the maximum amount of each of the

minor components analyzed for the first digestion was 500-600 fmole. After several digestion steps the total amount of glycopeptide was less than one fifth of the starting amount (based on UV absorbance), meaning that the components were measured at less than 100 fmole. This clearly shows the very high sensitivity of this method for sequencing minor glycopeptide components in a complex mixture.

In addition, by gradually simplifying the mixture by removing the heterogeneity inherent in the ends of the branches of these oligosaccharides, new components were identified which were not seen in the more complex mixture due to the limited dynamic range of the ESIMS method. No other analytical method which involves analysis of native or derivatized oligosaccharides can obtain this detailed information on such small quantities of oligosaccharides. Sutton and coworkers have published a similar method using mixtures of enzymes followed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOFMS) rather than HPLC/ESIMS which in theory has equal or better sensitivity (Sutton *et al.*, 1994). However, the authors' experience with the MALDI-TOF method has been that some glycopeptide mixtures work well, while others do not ionize well at all, yielding only some of the components of a mixture in the spectrum. Analysis of standard fetuin glycopeptides by both mass spectrometric methods yielded the identification of more glycopeptide components by the HPLC/ESIMS (Settineri 1993, unpublished results). In addition, it is difficult to identify and/or quantify sialylated glycopeptide components using MALDI-TOF, because the terminal sialic acid residues are almost always lost during the analysis (Huberty *et al.*, 1993; Sutton *et al.*, 1994). While, it may be particular sample handling or techniques or use of appropriate matrix conditions which affected the MALDI-TOF results, I feel that the HPLC/ESIMS method is a simpler, more widely applicable method for sequencing oligosaccharides on glycopeptides using exoglycosidases and mass spectrometry.

CHAPTER 6. CHARACTERIZATION OF THE *N*-LINKED OLIGOSACCHARIDES FROM RECOMBINANT HUMAN PLATELET-DERIVED GROWTH FACTOR β -RECEPTOR EXTRACELLULAR DOMAIN EXPRESSED IN CHINESE HAMSTER OVARY CELLS

6.1 Abstract

The extracellular domain of recombinant murine platelet-derived growth factor β -receptor (sPDGFR) expressed in Chinese hamster ovary cells is composed of 499 amino acids and contains eleven consensus sequences for *N*-linked glycosylation. Enzymatic deglycosylation experiments and carbohydrate composition analysis indicate a large number *N*-linked oligosaccharides attached to this protein. Endoproteinase glu-C peptide mapping using LSIMS, HPLC/ESIMS and MS/MS yielded confirmation of 85% of the theoretical protein sequence. HPLC/ESIMS of PDGF-R proteolytic digests, before and after enzymatic deglycosylation of the glycopeptides, combined with HPLC/ESI/CID/MS with SIM of diagnostic carbohydrate ions was performed to elucidate the structures and attachment sites of the *N*-linked carbohydrates on this protein. Five different groups of glycopeptides containing sialylated complex *N*-linked oligosaccharides were identified, indicating that only five of the eleven consensus sites are glycosylated.

6.2 Introduction

The platelet-derived growth factor β -receptor (PDGFR) is a 180-kDa transmembrane glycoprotein that binds a disulfide-linked dimer of PDGF-B-chains (PDGF-BB) (Ross *et al.*, 1986; Hannink and Donoghue, 1989). Upon binding of PDGF-BB to its receptor, a number of intracellular events occur which ultimately lead to DNA synthesis and cell division (Hannink and Donoghue, 1989; Heldin *et al.*, 1989; Morrison *et al.*, 1989; Morrison *et al.*, 1990). While the normal physiological role of PDGF (acting through the PDGFR) involves tissue repair and embryogenesis, PDGF has been implicated in many pathologic processes where overproduction of PDGF has been linked to atherosclerosis, myeloproliferative disease, and carcinomas (Ross *et al.*, 1986). Development of PDGF antagonists should facilitate studies concerning the role of PDGF in these diseases and may lead to new therapeutic approaches.

PDGFR is a member of a family of related receptors which includes the homologous α -type PDGF receptor (Matsui *et al.*, 1989), the macrophage-colony stimulating factor receptor (Coussens *et al.*, 1986) and the *c-kit* proto-oncogene product (Yarden *et al.*, 1987). Each of these receptors contains a large extracellular region (EXR), a single hydrophobic membrane-spanning region, and a cytoplasmic region with intrinsic tyrosine kinase activity (Yarden *et al.*, 1986; Williams, 1989). The predicted structure of PDGFR EXR comprises five β -strand-rich immunoglobulin-like domains (Williams, 1989) which, except for the fourth domain, contain regularly spaced cysteine residues. In addition, the EXR contains 11 potential sites for *N*-linked glycosylation. Lectin binding and enzymatic studies of intact PDGFR show the presence of both *N*- and *O*-linked glycosylation (Daniel *et al.*, 1987). While there are many implications for the role of glycosylation in relationship to the above structural features, before detailed studies can be performed, the exact structures and sites of attachment of the oligosaccharide chains must first be determined.

In order to study the structure of PDGFR independent of the transmembrane and cytoplasmic regions, a truncated murine PDGF β -receptor which contained only the EXR was expressed in CHO cells (Duan *et al.*, 1991). This soluble PDGFR protein (sPDGFR) fully retains the high affinity binding of the intact PDGFR with PDGF-BB (K_d of sPDGFR = 0.4 nM) and functions as a specific antagonist of PDGF-BB actions (Duan *et al.*, 1991). The recombinant protein contains 499 amino acids (see Figure 1), corresponding to a theoretical molecular mass of 56 kDa, but it migrates as a broad band centering at 110 kDa on SDS-PAGE and Western blots (Duan *et al.*, 1991). However, treatment of EXR with glycosidases that remove *N*-linked and *O*-linked carbohydrates reduced the protein to a single sharp band which migrated at approximately 60-kDa on SDS gels (Duan *et al.*, 1991). These data suggest that the carbohydrate content of sPDGFR may represent up to 50% of the total mass.

As an aid in the understanding of the mechanism of binding and subsequent action of PDGF-BB in terms of receptor function, structural identification of the oligosaccharides on PDGFR or sPDGFR is highly desirable. In addition, if sPDGFR is to be used as a therapeutic, the biological effects of glycosylation must be assessed in terms of consistency, safety and efficacy, due to the effects that carbohydrate chains have on clearance and antigenicity. Knowledge of the attachment sites of the oligosaccharides, would allow mutation studies to be performed on the sPDGFR to test the role of specific glycosylation sites in cellular processing, binding of ligand and subsequent receptor action. In addition, because the carbohydrates on glycoproteins often interfere with their crystallization, the ability to perform site-specific mutation of the glycosylation sites would aid crystallization of this soluble receptor so that its three dimensional structure could eventually be solved. Therefore, the major aims of this project was to determine the structure of the *N*-linked carbohydrates from sPDGFR and their specific sites of attachment using mass spectrometry.

6.3 Materials and Methods

Cloning, Expression, and Purification. The extracellular domain of recombinant platelet-derived growth factor β -receptor was cloned, expressed in CHO cells and purified as described previously (Duan *et al.*, 1991).

Reduction and S-Carboxymethylation. Reduction of sPDGFR was performed at protein concentrations of 1-4 mg/ml in 200 mM NEMOAc, pH 8.5, containing 6 M guanidine-HCl and 3 mM EDTA with a 50 fold molar excess of dithiothreitol. The protein was reduced with dithiothreitol for 1.5 hr at 60°C, followed immediately by alkylation with sodium iodoacetic acid for 1 hr at room temperature in the dark.

Enzymatic and Glycolytic Digestions. Endoproteinase glu-C (*Staphylococcus aureus* V8, Boehringer Mannheim, Indianapolis, IN) digestions were performed in 50 mM ammonium bicarbonate (pH 7.8), at a weight ratio of 8% (2 x 4%) for 18-28 hr at 37°C. PNGase F (glycerol-free, Genzyme, Boston, MA) digestions were performed at 37°C in 100 mM ammonium bicarbonate (pH 8.0), using 2.5-3.5 units of enzyme over two days. Neuraminidase (*Arthrobacter ureafaciens*, Boehringer Mannheim, Indianapolis, IN) digestions were performed in 50 mM sodium acetate (pH 5.0) at 37°C for 18-24 hr. Aliquots of 10-20 mUnits (1-2 μ l) were added at 0 and 8 hr.

Peptide and glycopeptide HPLC. HPLC-ESIMS analyses were performed using a Carlo Erba Phoenix 20 gradient HPLC system and an Aquapore OD-300 microbore column (C₁₈, 100 x 1mm, 7 μ m, Applied Biosystems, Foster City, CA). Solvent A was 0.1% TFA/H₂O, solvent B was 0.08% TFA/ACN, and the flow rate used was 50 μ l/min. Using an Isco μ LC-500 syringe pump, a 1:1 mixture of isopropanol:2-methoxyethanol was mixed with the column effluent post-column (after UV detection) at a flow rate of 35 μ l/min. This combined flow of 85 μ l/min was split at a ratio of approximately 20:1 before entering the mass spectrometer, so that a final flow of approximately 3.5-4.5 μ l/min. entered directly into the electrospray ionization source. The elution of all compounds was

monitored at a wavelength of 215 nm using an Applied Biosystems UV detector model 204 equipped with a U-shaped capillary flow cell (LC Packings, San Francisco, CA).

Mass Spectrometry. LSIMS analysis of peptides was carried out using a Kratos MS50S double focusing mass spectrometer (Kratos Inc., Manchester, UK) equipped with a high field magnet and a cesium ion source (Aberth *et al.*, 1982) and gun at 10 keV (Falick *et al.*, 1986). Concentrated HPLC fractions (1-2 μ l) were added to 1 μ l of liquid matrix (1:1 thioglycerol:glycerol with 1% TFA) for analysis. Electrospray ionization (ESI) spectra were acquired on-line, after microbore HPLC elution, on a VG Bio-Q mass spectrometer (Fisons Instruments, Manchester, U.K.) equipped with an atmospheric pressure electrostatic spray ion source and a quadrupole mass analyzer (maximum m/z = 3000). The scan time in non-continuum mode from m/z = 350 to m/z = 2200 was 5 seconds. Typical operating voltages are as follows: Probe tip = 4 kV; Counter electrode = 1.5 kV; Sample cone (nozzle) voltage = 45 V; SIM sample cone voltage: 150 V-200V. The nebulizing gas flow rate was typically 150 mL/min., and the drying gas was typically 4 L/min. The source temperature was typically 60°C.

6.3 Results

In order to first determine if sPDGFR was expressed with the correct sequence and length, 0.5-1.0 nmole sPDGFR was reduced and S-carboxymethylated. After microdialysis to remove excess reagents, the protein was digested with either trypsin or endoproteinase glu-C, and the resultant peptides were separated by microbore HPLC and manually collected for analysis by LSIMS (or direct injection ESIMS the for high molecular weight, late-eluting fractions). The resulting peptide mass map is shown in Figure 1. As indicated by

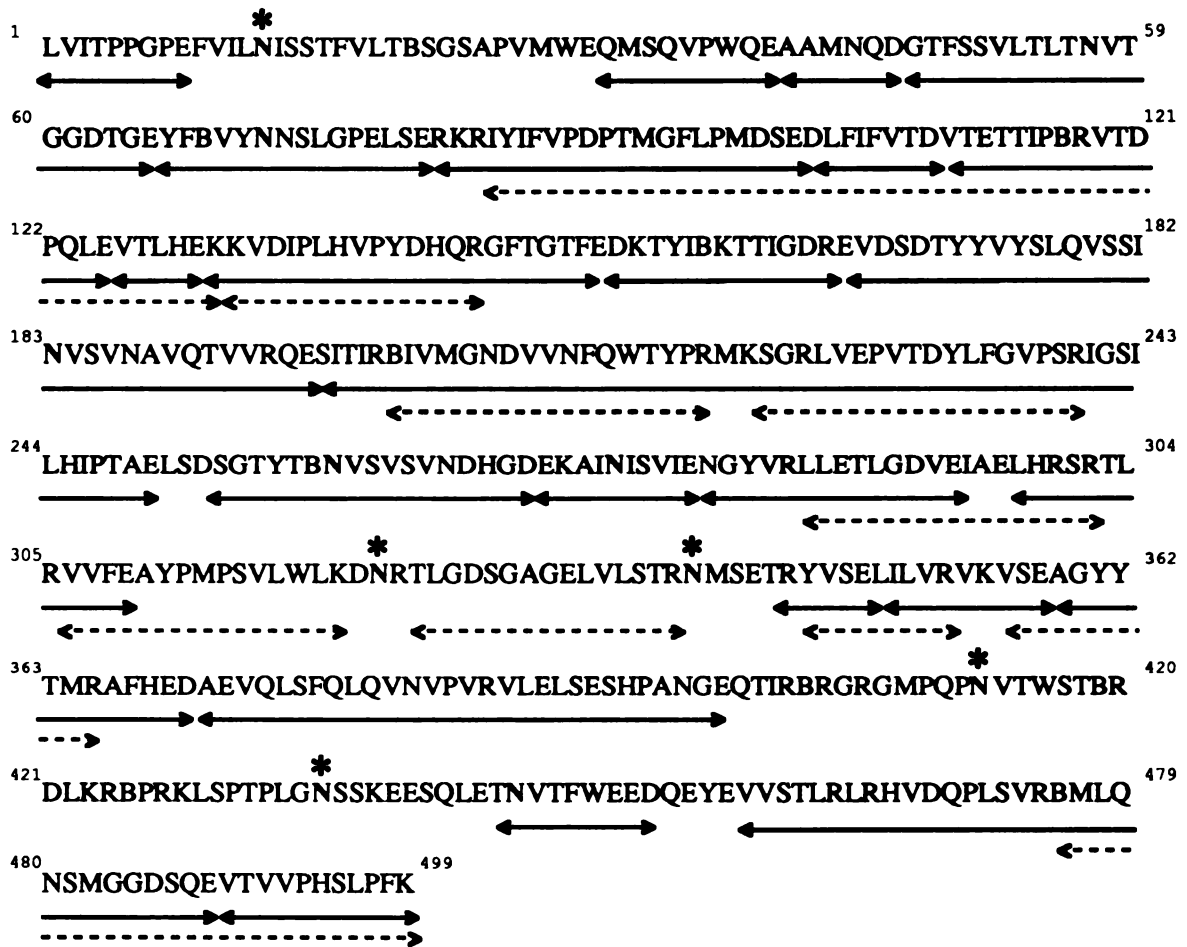


Figure 1. Peptide mass map of sPDGFR peptides identified by LSIMS and direct injection ESIMS. The solid arrows indicate peptides from the endoproteinase glu-C digest, and the dashed arrows indicate peptides identified from the trypsin digest. The starred and bold letters denote sites of Asn-linked glycosylation.

the underlining solid (endoproteinase glu-C digest) and dashed (trypsin digest) arrows, unmodified peptides were found which covered only six of the eleven potential *N*-linked glycosylation sites. This indicated that the other five potential *N*-linked sites may be glycosylated. Therefore, in order to assess the amount of carbohydrate attached to this protein, carbohydrate composition analysis was performed by hydrolysis of 1.0 nmole of the intact sPDGFR protein followed by HPAE-PAD. The results, shown in Table I, supported the peptide mapping results of the presence of five different sites glycosylated on this protein, and the presence of sialic acid, fucose, galactose, and glucosamine residues,

indicated that the structures attached to sPDGFR were of the complex type. For example, every *N*-linked oligosaccharide would contain at least the three mannose residues present in the core of all *N*-linked structures (Kornfeld and Kornfeld, 1976). Since the composition analysis showed the presence of approximately 13 Man residues per mole of sPDGFR, this would represent 4 or 5 ($13/3 = 4.33$) complex type oligosaccharide structures per mole of protein (since complex structures only contain three Man residues per oligosaccharide). In addition, if the average structure on the protein was a complex triantennary structure, which contains 5 GlcNAc residues, then the 24.7 moles of glucosamine from the composition analysis would represent 5 oligosaccharide structures per mole of sPDGFR.

| | <u>moles monosaccharide</u> <u>moles hydrolyzed sPDGFR</u> |
|----------------------|---|
| Sialic acid | 18.4 ± 0.2 |
| Fucose | 7.7 ± 0.2 |
| Galactosamine | 3.4 ± 0.5 |
| Glucosamine | 24.7 ± 0.4 |
| Galactose | 16.4 ± 0.3 |
| Mannose | 12.9 ± 0.1 |

Table I. Monosaccharide composition analysis results for sPDGFR

In order to determine if the "unmapped" sites were glycosylated, and what structures were attached to these sites, another 750 pmole aliquot of this protein was digested with endoproteinase glu-C. 100 pmole of this digest was then analyzed by HPLC/ESIMS, and immediately following, another 75 pmole aliquot of the endoproteinase glu-C only digest

was then analyzed by HPLC/ESIMS with SIM, under identical chromatographic conditions, to identify the glycopeptide-containing fractions. Based on the carbohydrate composition data, the following ions were monitored: $m/z=147$ for [Fuc]⁺, 204 for [GlcNAc]⁺ or [GalNAc]⁺, 292 for [NeuAc]⁺, 366 for [Gal-GlcNAc]⁺ or [Gal-GalNAc]⁺ and 657 for [NeuAc-Gal-GlcNAc]⁺ or [NeuAc-Gal-GalNAc]⁺. Figure 2 illustrates the HPLC/ESIMS data for the SPDGFR endoproteinase glu-C digest run under normal operating conditions where the entire mass range ($m/z=350-2200$) was scanned (base peak intensity chromatogram) and under SIM conditions where only selected masses were scanned (SIM at $m/z=204$). Figure 3 illustrates the SIM chromatograms for the other four masses which were scanned ($m/z=147$, 292, 366 and 657), indicating once again the presence of complex oligosaccharide structures containing fucose and sialic acid. The molecular mass data combined with the SIM data resulted in identification of five different groups of glycopeptide glycoforms. The starred peaks in Figure 2 indicate those that were identified to contain glycopeptides. After deconvolution of the molecular ions, the masses obtained for these glycopeptides matched to different complex carbohydrate structures which could be linked to one of the five "unmapped" peptides containing *N*-linked consensus sites at Asn¹⁴, Asn³²², Asn³³⁹, Asn⁴¹³ and Asn⁴³⁶. The HPLC/ESIMS glycopeptide data and the structures proposed at each *N*-linked site based on molecular weight is summarized in Table II. Figure 4A shows the HPLC/ESI mass spectrum of the group of glycopeptides containing residues (400-421). As shown in Table II, the molecular masses represent fucosylated triantennary oligosaccharides containing 0, 1 and 2 sialic acid residues.

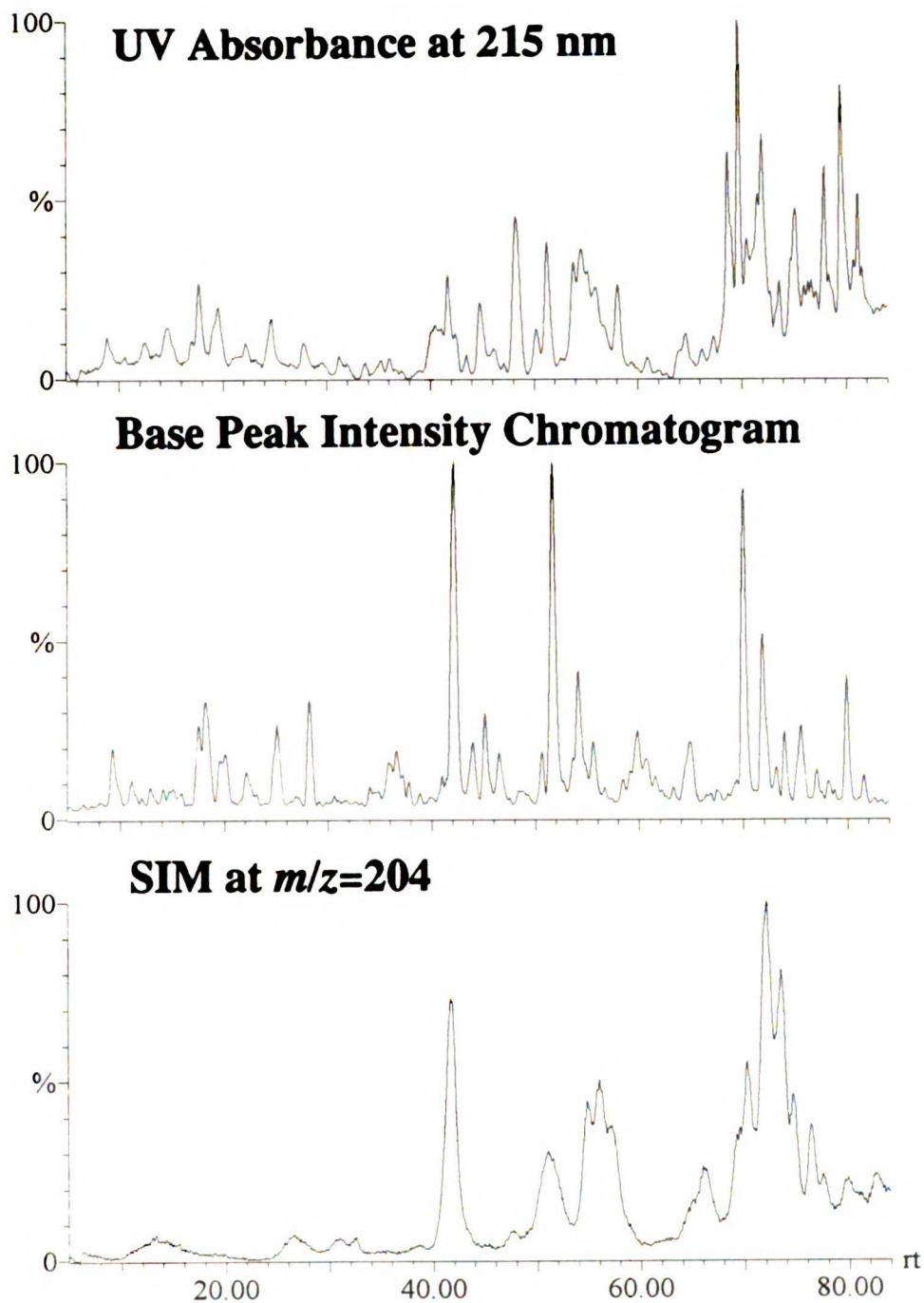


Figure 2. HPLC/ESIMS results of sPDGFR endoproteinase glu-C digest. UV absorbance, base peak intensity and SIM at $m/z=204$ chromatograms. Range=0.04 AUFS. Solvent A=0.1% TFA, solvent B=0.08% TFA in ACN. Gradient program was 0-50% B in 100 min.

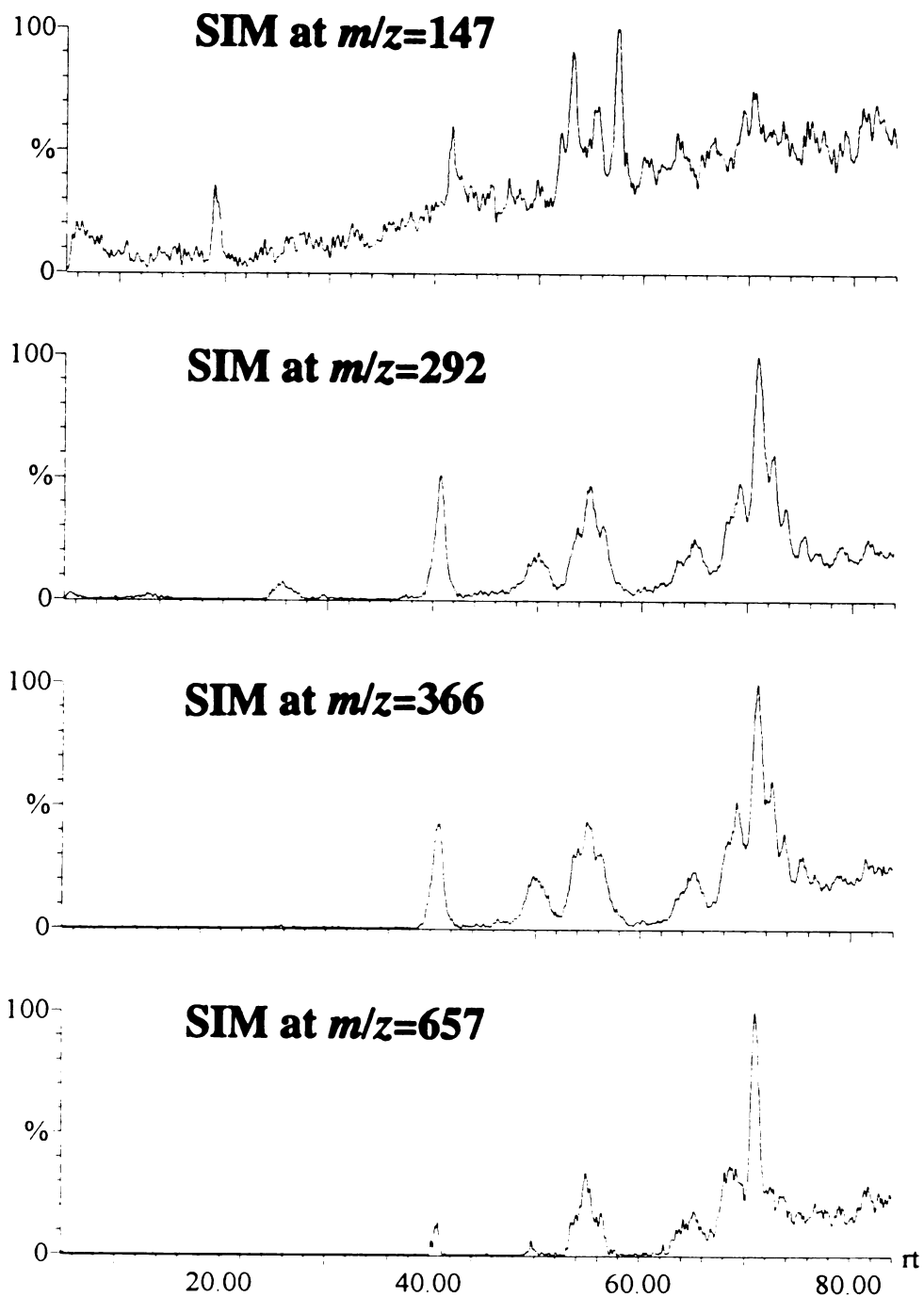


Figure 3. SIM chromatograms at $m/z=147$, 292, 366 and 657 obtained from HPLC/ESI/CID/MS analysis of the sPDGFR endoproteinase glu-C digest.

| Intact Glycopeptide Mass _{avg.} | Glycopeptide Mass _{avg.} & Proposed Structure after Neuraminidase Digest | Peptide Mass _{avg.} after N-glycanase Digest (Asn → Asp) |
|--|--|---|
| y=1 4725.1 Da y=2 5017.6 Da | <p>y=0 4432.7 Da x=1 4800.9 Da</p> <p>Asn-14</p> <p>NeuAc_y-Gal-GlcNAc-Man NeuAc_y-Gal-GlcNAc-Man [NeuAc_y-Gal-GlcNAc]_x</p> <p>Man-GlcNAc-GlcNAc Fucose</p> <p>10 FVLNISSIFVLTC*SGSAPVMWE 31</p> <p>2447.8 Da</p> | |
| x=0 4829.7 Da x=1 5486.1 Da x=2 6142.6 Da | <p>x=0 4247.9 Da x=1 4614.0 Da x=2 4979.6 Da</p> <p>Asn-322</p> <p>[NeuAc]-Gal-GlcNAc]_x NeuAc-Gal-GlcNAc-Man NeuAc-Gal-GlcNAc-Man [NeuAc]-Gal-GlcNAc]_x</p> <p>Man-GlcNAc-GlcNAc Fucose</p> <p>310 AYPMPVSLWLKDNRTLGDGSGAGE 322</p> <p>2478.8 Da</p> | |
| y=1 3210.5 Da y=2 3502.0 Da x=1 4156.4 Da | <p>y=0 2919.6 Da x=1 3285.9 Da x=2 3650.1 Da</p> <p>Asn-339</p> <p>NeuAc_y-Gal-GlcNAc-Man NeuAc_y-Gal-GlcNAc-Man [NeuAc_y-Gal-GlcNAc]_x</p> <p>Man-GlcNAc-GlcNAc Fucose</p> <p>333 LVLSTRNMSE 342</p> <p>1150.3 Da</p> | |
| y=0 4448.7 Da y=1 4738.7 Da y=2 5030.0 Da x=1 5395.0 Da | <p>y=0 4449.4 Da x=1 4815.8 Da x=2 5179.6 Da</p> <p>Asn-413</p> <p>NeuAc_y-Gal-GlcNAc-Man NeuAc_y-Gal-GlcNAc-Man [NeuAc_y-Gal-GlcNAc]_x</p> <p>Man-GlcNAc-GlcNAc Fucose</p> <p>400 QTIRC*RGRGMPQPNVTWSTC*RD 421</p> <p>not found</p> | |
| x=0 4651.2 Da x=1 5307.6 Da x=2 5964.5 Da | <p>x=0 4068.2 Da x=1 4434.4 Da x=2 4799.5 Da</p> <p>Asn-436</p> <p>[NeuAc]-Gal-GlcNAc]_x NeuAc-Gal-GlcNAc-Man NeuAc-Gal-GlcNAc-Man [NeuAc]-Gal-GlcNAc]_x</p> <p>Man-GlcNAc-GlcNAc Fucose</p> <p>422 LKRC*PRKLSPTPLGNSSKEE 441</p> <p>2299.7 Da</p> | |

Table II. Summary of HPLC/ESIMS data obtained on the glycopeptides identified from recombinant PDGF-R endoproteinase glu-C, endoproteinase glu-C + neuraminidase and endoproteinase glu-C + PNGase F digests.

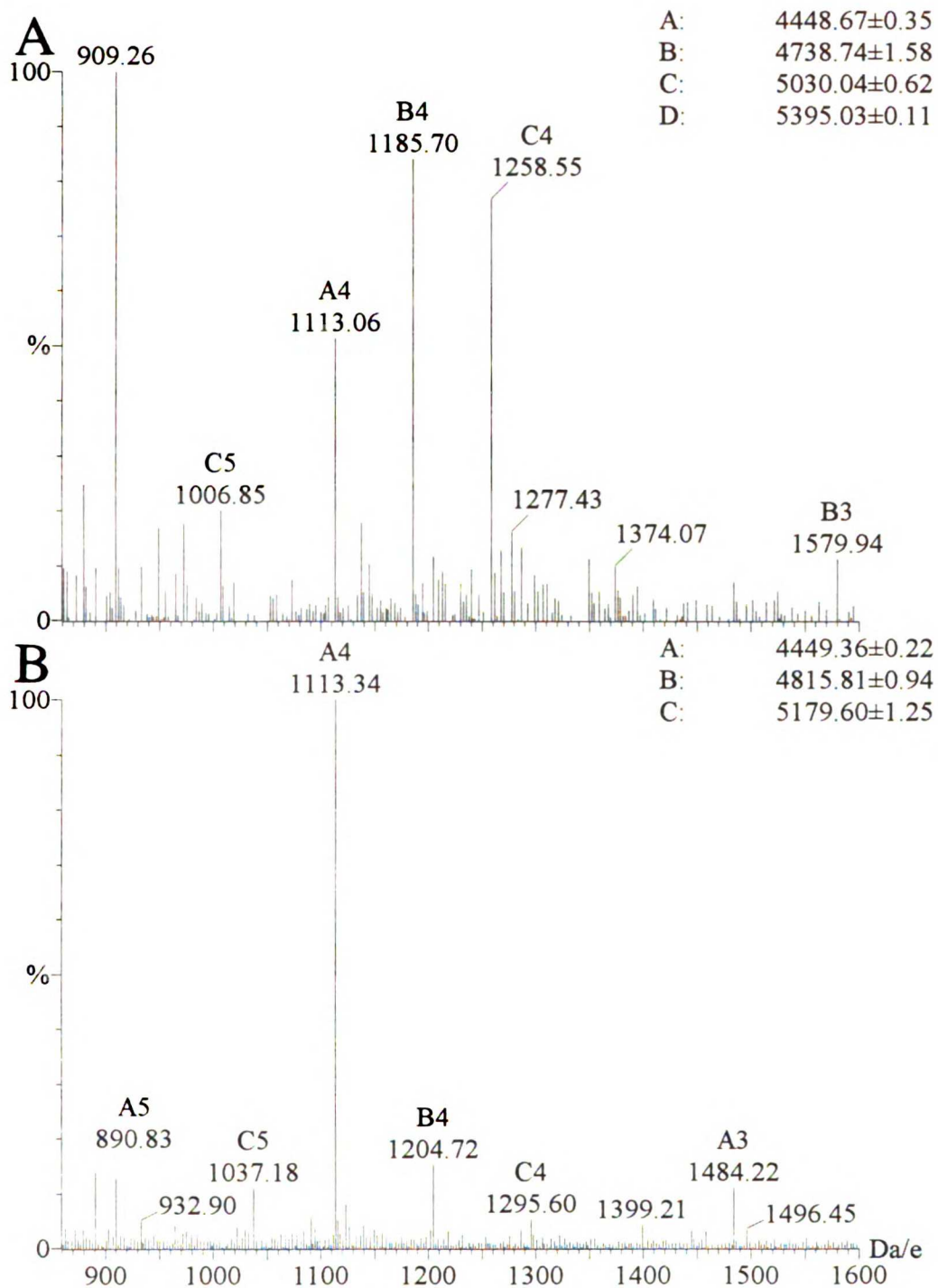


Figure 4. HPLC/ESI mass spectra for the group of glycopeptides representing residues (400-421). A) HPLC/ESI results for glycopeptide (400-421) from the sPDGFR glu-C digest. B) HPLC/ESI results for glycopeptide (400-421) from the sPDGFR glu-C + neuraminidase digest.

However, as showed by the starred peaks, there are many other fractions which should contain glycopeptides, based on the SIM chromatogram. Since all five of the "unmapped" peptides were identified as glycopeptides from this data, it was not clear what other carbohydrate-containing components were present in the digest. Therefore two additional digestions were performed on aliquots of the endoproteinase glu-C digest in order to establish that the proposed glycopeptides identified in the endoproteinase glu-C digest were indeed glycopeptides, and to attempt to identify any additional glycopeptides. The first glycosidase digestion was with neuraminidase from *Arthrobacter ureafaciens*, which hydrolyzes terminal N- or O-acyl-neuraminic acid (sialic acid) residues which are $\alpha 2 \rightarrow 3$, $\alpha 2 \rightarrow 6$ or $\alpha 2 \rightarrow 8$ linked to oligosaccharides (Uchida *et al.*, 1979; Twining, 1984). 175 pmole of the sPDGFR endoproteinase glu-C digest was treated with the *Arthrobacter ureafaciens* neuraminidase. Figure 5 illustrates the HPLC/ESIMS data for the sPDGFR endoproteinase glu-C + neuraminidase digest run under normal operating conditions where the entire mass range ($m/z=350-2200$) was scanned (base peak intensity chromatogram) and under SIM conditions (SIM at $m/z=204$). It looks quite similar to Figure 2, but clearly, the neuraminidase changed the elution time of the glycopeptides. HPLC/ESIMS (100 pmole) and HPLC/ESI/CID/MS with SIM (75 pmole) analysis of the endoproteinase glu-C + neuraminidase digest confirmed that the masses identified to be sialylated glycopeptide structures in the endoproteinase glu-C digest were correct, due to the appearance of masses which were multiples of 291 Da lower in the neuraminidase-treated digest, corresponding to hydrolytic cleavage of sialic acid residues. In addition, larger asialo structures were identified for some of the glycopeptides which were too weak (most likely due to molecular weight) to be seen in the endoproteinase glu-C digest when sialylated (e.g. glycopeptides (333-342) and (400-421)). This data is summarized in Table II. Figure 4B shows the HPLC/ESI mass spectrum of the group of glycopeptides containing residues (400-421) which has been digested with neuraminidase. As shown in

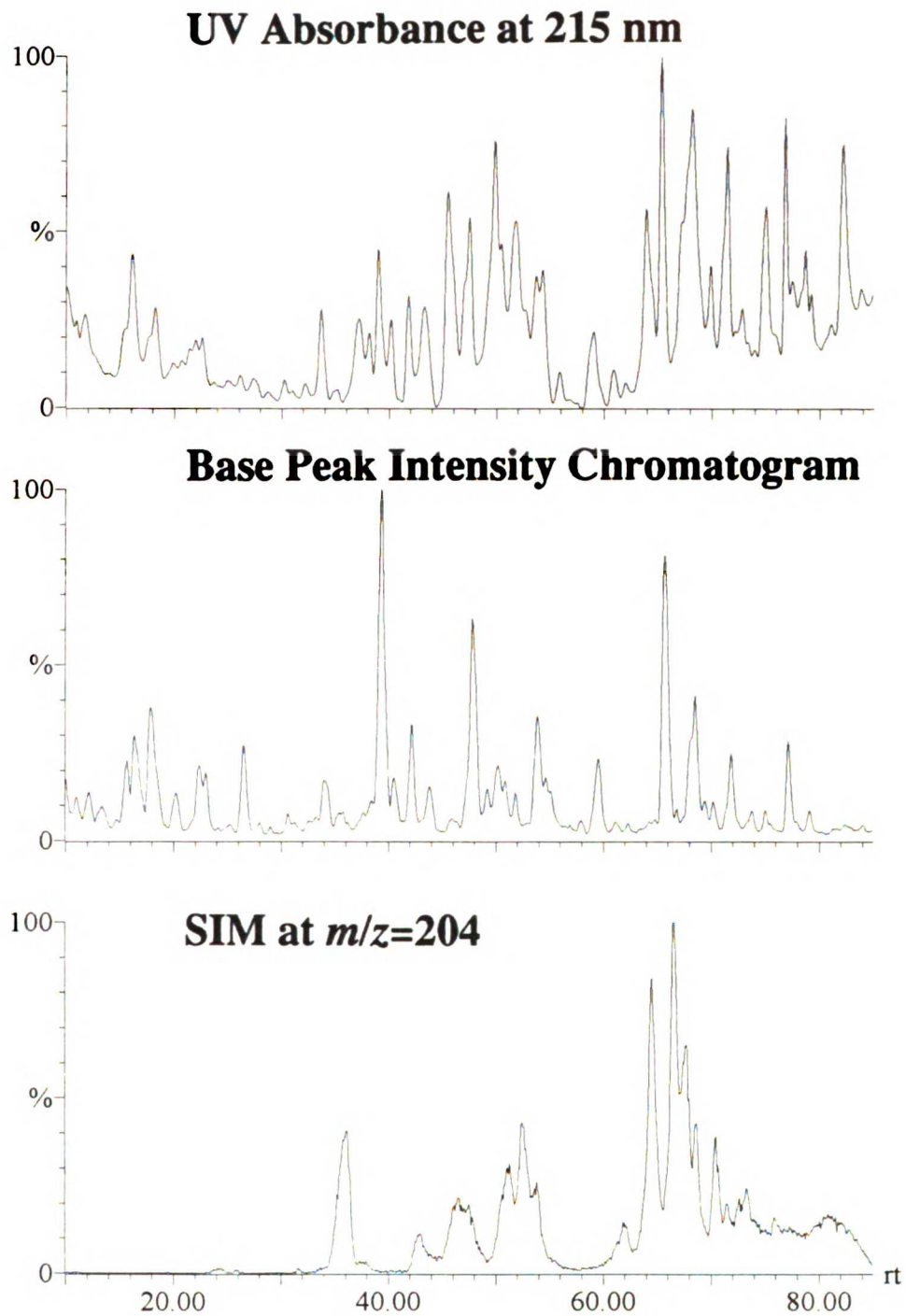


Figure 5. HPLC/ESIMS results of sPDGFR endoproteinase glu-C + neuraminidase digest. UV absorbance, base peak intensity and SIM at $m/z=204$ chromatograms. Range=0.04 AUFS. Solvent A=0.1% TFA, solvent B=0.08% TFA in ACN. Gradient program was 0-50% B in 100 min.

Table II, the molecular masses represent fucosylated triantennary and tetraantennary oligosaccharides which were identified in the endoproteinase glu-C digest.

The second glycosidase digestion of the endoproteinase glu-C digest (375 pmole) was performed using PNGase F, which cleaves all classes of intact *N*-linked oligosaccharides from a protein or peptide, converting the Asn to which it is attached to an Asp (Tarentino *et al.*, 1985), while leaving the rest of the protein or peptide intact (Plummer and Tarentino, 1991). HPLC/ESIMS analysis of the endoproteinase glu-C + PNGase F digest established that the five proposed glycopeptides from the endoproteinase glu-C and endoproteinase glu-C + neuraminidase digests, because the groups of glycopeptides were no longer present. In addition, four of the five original glycopeptides were found as deglycosylated peptides which were one dalton higher in mass than the predicted mass for the non-glycosylated peptides. This data is also summarized in Table II. It is not clear why the deglycosylated peptide containing residues (400-421) was not found. The intact glycopeptide was not present in the endoproteinase glu-C + PNGase F digest, so clearly the peptide should have been present. It is possible that this large slightly hydrophobic peptide did not ionize well or was lost in the sample tube or on the HPLC column.

Unfortunately, no additional glycopeptide components were identified in either of the glycosidase digests. Due to sample limitations, no additional experiments were performed to determine the source of the other carbohydrate-containing components in the digest.

6.5 Discussion

Since no additional experiments could be performed to try to determine the source of the unidentified carbohydrate-containing components, one can only speculate on their source. It is likely that the components came from a contaminating protein(s) which was not purified away from sPDGFR. It is not surprising that a lower level contaminant than

sPDGFR in the digests would not be identified in the full scan HPLC/ESIMS data, because most of the sPDGFR glycopeptide spectra were quite weak, because only 5 pmole of sample (5% of 100 pmole) was actually entering the ESI mass spectrometer. The HPLC/ESI/CID/MS method is so selective and sensitive that glycopeptides are easily identified even for extremely small quantities of sample. In addition, since peptides often yield molecular ions that are 50-100 times more intense than glycopeptides, if there is not enough sample, coeluting peptides will mask the glycopeptides due to the dynamic range limitations of the electrospray mass spectrometer.

In addition, the sPDGFR sample very likely contained other contaminants which prevented optimal ionization at the sample levels analyzed. The original "purified" glycoprotein sample required careful acetone precipitation in order for complete endoproteinase glu-C or trypsin digestion as well as for reasonable electrospray ionization of the peptides and glycopeptides. This problem was not alleviated by extensive dialysis, so clearly the contaminant(s) is something which is either insoluble in the dialysis buffer or strongly associated with the protein. Even after acetone precipitation and extensive dialysis, proteolytic digests of this protein did not ionize as well as digests of other glycoproteins such as sNGFR, described in Chapters 4 and 5. This could have been due to residual amounts of detergents used in the original isolation and purification of the protein.

Chapter 7 describes an example where a contaminating glycoprotein was copurified with the glycoprotein of interest, but since 3 to 4 times more sample was analyzed at a time, and the analysis was done two years later when the HPLC/ESIMS method and the instrument's sensitivity were better optimized, the glycopeptide spectra were much more intense and easier identified when coeluting with peptides. Even though this protein contained an additional protein in the preparation, since it was pure of any other contaminants such as detergents, it was a much easier problem to solve. Clearly, preparation of sample prior to digestion and mass spectrometry can have dramatic effects on the outcome of the analyses performed.

CHAPTER 7. CHARACTERIZATION OF THE *N*-AND *O*-LINKED
OLIGOSACCHARIDES FROM HUMAN PLASMA
LECTITHIN:CHOLESTEROL ACYLTRANSFERASE

7.1 Abstract

Microbore reversed phase HPLC coupled directly with ESIMS (HPLC/ESIMS) and Edman degradation were used to characterize the tryptic peptides and glycopeptides of human lecithin:cholesterol acyltransferase, including four *N*-linked glycopeptides and one *O*-linked glycopeptide. A mass spectrometric technique, involving observation of diagnostic sugar fragment ions produced during the electrospray mass spectrometry experiment, was employed to differentiate the glycopeptides from the peptides in the digest. To obtain structural information on the oligosaccharide portion of the glycopeptides, sequential exo- and/or endoglycosidase digestions combined with HPLC/ESIMS were performed on the isolated glycopeptides. All four potential *N*-linked glycosylation sites (Asn²⁰, Asn⁸⁴, Asn²⁷² and Asn³⁸⁴) were determined to contain primarily sialylated triantennary or biantennary complex carbohydrate structures. Two unexpected *O*-linked glycosylation sites were also identified at Thr⁴⁰⁷ and Ser⁴⁰⁹ in the carboxy-terminal peptide, each of which contain sialylated Galβ1→3GalNAc structures. Site-directed mutagenesis previously indicated that removal of the glycosylation site at Asn²⁷² converted this protein to a phospholipase (O. L. Francone *et al.* (1993) *Biochim. Biophys. Acta* **1166**, 301-304). The results presented herein do not support such a discrimination among the different *N*-glycosylation sites based solely on a knowledge of the carbohydrate structures at each specific site. Three other *N*-linked glycopeptides were also detected and determined to be from another protein, apolipoprotein D, which contains two potential *N*-linked glycosylation sites at Asn⁴⁵ and Asn⁷⁸. Exoglycosidase digestion of these glycopeptides

7.2 Introduction

Lecithin:cholesterol acyltransferase (LCAT) catalyses the transfer of an acyl group from the 2-position of phosphatidylcholine to the 3-position hydroxyl group of cholesterol in blood plasma (Aron *et al.*, 1978). In the absence of cholesterol, long-chain alcohols or water itself can act as acceptors for the acyl group, generating fatty esters or free acids, respectively (Subbaiah *et al.*, 1980). The primary structure of human LCAT has been determined by amino acid (Yang *et al.*, 1987) and cDNA (McLean *et al.*, 1986) sequencing, and the sequence is shown in Figure 1. These studies of LCAT employing Edman degradation data suggested that all four potential *N*-linked glycosylation sites, as indicated by the consensus sequon (Asn-Xxx-Ser/Thr) where Xxx is any amino acid except proline (Bause and Hettkamp, 1979), at Asn²⁰, Asn⁸⁴, Asn²⁷² and Asn³⁸⁴ (shown in Figure 1) were glycosylated (Yang *et al.*, 1987). The carbohydrate composition was previously determined to contain 13% hexose, 6.2% hexosamine and 5.4% sialic acid residues, respectively (Chong *et al.*, 1983). In addition, further characterization of the nature of the glycosylation has been reported involving the use of glycosidases and specific inhibitors of glycoprotein assembly with cultured cells secreting LCAT activity. These studies concluded that LCAT contained high mannose and complex type *N*-linked oligosaccharides (Collet and Fielding, 1991). While the function of the LCAT carbohydrates is not well understood, it has been shown that their presence is necessary for full activity of LCAT, but not for its synthesis or secretion (Collet and Fielding, 1991). It has been established by site-directed mutagenesis that removal of the glycosylation site at Asn²⁷² converts LCAT into a phospholipase, generating fatty acids rather than cholesteryl esters (Francone *et al.*, 1993). Therefore, this study was undertaken to determine the structural nature and heterogeneity of the glycosylation at the four known *N*-linked sites on this protein.

In the present study a combination of trypsin digestion, HPLC/ESIMS, selected ion monitoring, Edman degradation and exo- and endoglycosidase digestions was used to obtain a global picture of the glycosylation on LCAT, including characterization of the structures and their heterogeneity at each glycosylation site. Using these methods, two previously unidentified *O*-linked glycosylation sites were identified and characterized. In addition, the power and sensitivity of these methods are further emphasized by the detection and characterization of three *N*-linked glycopeptides from a low level contaminant in the LCAT preparation, apolipoprotein D (Apo D).

7.3 Materials and Methods

Isolation and Purification of Human LCAT. LCAT, which was graciously provided by Professor Christopher J. Fielding and colleagues at UCSF, was purified from normal human plasma as described earlier (Collet and Fielding, 1991).

Reduction and Alkylation of LCAT. LCAT in 0.1 mM NH_4HCO_3 , 6 M guanidine-HCL, pH 8.0 was reduced by a 50-fold excess of DTT at 50°C for 1.5 hr under argon. A 20-fold excess of iodoacetic acid was used for alkylation in the dark at room temperature for 1.5 hr. A $\text{CHCl}_3/\text{CH}_3\text{OH}$ (1:1, v/v) extraction was then performed to eliminate lipids from the sample followed by dialysis overnight against 50 mM NH_4HCO_3 , pH 8.0.

Tryptic Digestion of LCAT. Reduced and alkylated LCAT was digested with sequencing grade trypsin (Boehringer Mannheim) at an enzyme:substrate ratio of 1:50, w/w, for 8 hr at 37°C.

Fractionation of Peptides and Glycopeptides. Microbore reversed phase HPLC separation was performed using a Carlo Erba Phoenix 20 dual syringe pump system. Separations were accomplished on an Aquapore 300 C-18 column, 1.0mm i.d. x

250mm (Applied Biosystems). Eluant A was 0.1% TFA in water, and eluant B was 0.08% TFA in ACN. Water and acetonitrile were purchased from Fisher, and TFA was obtained from Pierce. Column effluent at 50 μ L/min flowed directly into a variable wavelength UV detector (Linear, model 204) equipped with a high sensitivity U-Z View capillary flow cell (LC Packings). Positioned directly after the detector was an Isco μ LC-500 microflow syringe pump which allowed for post-column addition of a 2-methoxyethanol/2-propanol (Aldrich) (1:1, v/v) mixture to the column effluent at flowrates of 35-40 μ L/min. The effluent was split approximately 1:19 after the mixing tee so that only 4-5 μ L/min went into the mass spectrometer, while the remainder was manually collected. The gradient program used to separate the LCAT tryptic digest was a linear gradient of 2-70% B over 120 min followed by 70-98% B over 10 min. The gradient programs used to analyze the glycopeptides after exo- and endoglycosidase digestions were either 2-40% B over 30 min or 10-50% B over 40 min.

Electrospray Ionization Mass Spectrometry. The microbore HPLC system was interfaced to a VG BIO-Q (Fisons/VG) triple quadrupole mass spectrometer, equipped with an atmospheric pressure electrostatic spray ion source. The mass spectrometer was scanned in non-continuum mode from $m/z = 330-2000$ in 5 seconds. For the SIM experiment the voltage of the sampling cone was adjusted from 60 to 180 volts.

Exo- and Endoglycosidase Digestions. Fractions containing *N*-linked glycopeptides (200-500 pmole) were solubilized in 50mM NaOAc pH 5.0 and 1.0 mU/ μ l neuraminidase (*Arthrobacter ureafaciens*, Boehringer Mannheim) was added. The digestion was performed at 37°C for 5-7 hr. After microbore HPLC/ESIMS analysis, the fractions were collected and vacuum dried. These fractions were then solubilized in 50mM NaOAc pH 5.0 and digested with 0.55 mU/ μ l β -galactosidase (*Streptococcus pneumoniae*, Oxford Glycosystems) at 37°C for 5-7 hr, and the samples reanalyzed HPLC/ESIMS. The fractions of interest were collected and vacuum dried, followed by digestion with 1.0 mU/ μ l *N*-acetylhexosaminidase (Chicken liver, Oxford Glycosystems) at 37°C in 50mM

NaOAc pH 5.0 for another 5-7 hr, and reanalyzed. The dried *O*-linked glycopeptide fraction (50 pmole) was solubilized in 0.1M Na₂HPO₄, pH 7.2 and digested with the neuraminidase (1.0 mU/μl) for 17 hr, followed by 0.5 mU *O*-Glycanase (*Diplococcus pneumoniae* Endo- α -*N*-acetyl-galactosaminidase, Boehringer Mannheim) for an additional 24 hr at 37°C.

Edman Degradation. Microsequencing was performed by Frank Masiarz of Chiron Corporation and Zhonghua Yu and Corey Schwartz of UCSF, on a Applied Biosystems 470A gas phase protein sequencer, with an Applied Biosystems 120A phenylthiohydantoin analyzer (Hunkapiller *et al.*, 1983).

7.4 Results

Localization of glycopeptides by HPLC/ESIMS with selected ion monitoring. Since the goal of this study was not only to characterize the carbohydrates from LCAT but also to assess the heterogeneity at each site, it was necessary to first isolate the corresponding glycopeptides and then further analyze the carbohydrate structures attached to each peptide. Firstly, approximately one nmole of reduced, carboxymethylated LCAT was digested with trypsin. One 300 pmole aliquot of the tryptic digest was analyzed by microbore reversed phase HPLC/ESIMS and another 300 pmole aliquot was analyzed by HPLC/ESI/CID/MS with SIM. The HPLC/ESIMS analysis of the tryptic digest was performed in the conventional operating mode in which a mass chromatogram representing the total ion abundance, resulting from scanning the entire mass range (m/z 330-2000), and a UV absorbance chromatogram were obtained. Only 5% of the material injected flowed into the mass spectrometer, while 95% was manually recollected after HPLC separation for further analysis. Since the primary sequence of LCAT was known, the sequences of the 21 tryptic peptides could be confirmed by identification of the molecular masses of the predicted

tryptic peptides, as indicated by the solid arrows in Figure 1. As illustrated in Figure 2, for each UV peak exists a corresponding MS (base peak intensity) peak, however their relative intensities are not necessarily the same.

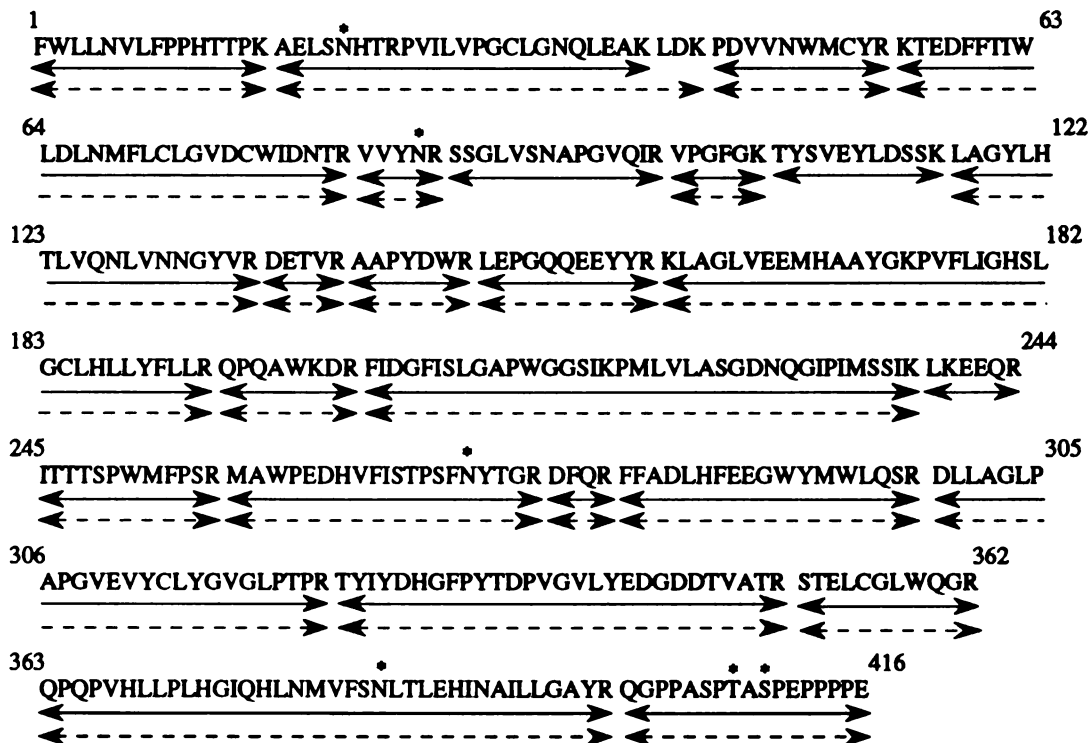


Figure 1. Primary sequence of human LCAT. Glycosylation sites identified at Asn²⁰, Asn⁸⁴, Asn²⁷² and Asn³⁸⁴ (*N*-linked) as well as Thr⁴⁰⁷ and Ser⁴⁰⁹ (*O*-linked) are designated by asterisks. Underlining with the solid arrows indicates peptides and glycopeptides identified by HPLC/ESIMS, while the dashed arrows indicate those identified by Edman sequencing in our laboratory.

Using a second 300 pmole aliquot of the tryptic digest the HPLC/ESI/CID/MS with SIM experiment was then performed in order to identify which of the UV and MS peaks contain glycopeptides. Again, 95% of the injected material was manually collected into the vials containing the fractions which had the same retention time in the first experiment. The remaining 5% was directed into the ESI mass spectrometer which was scanned in the SIM

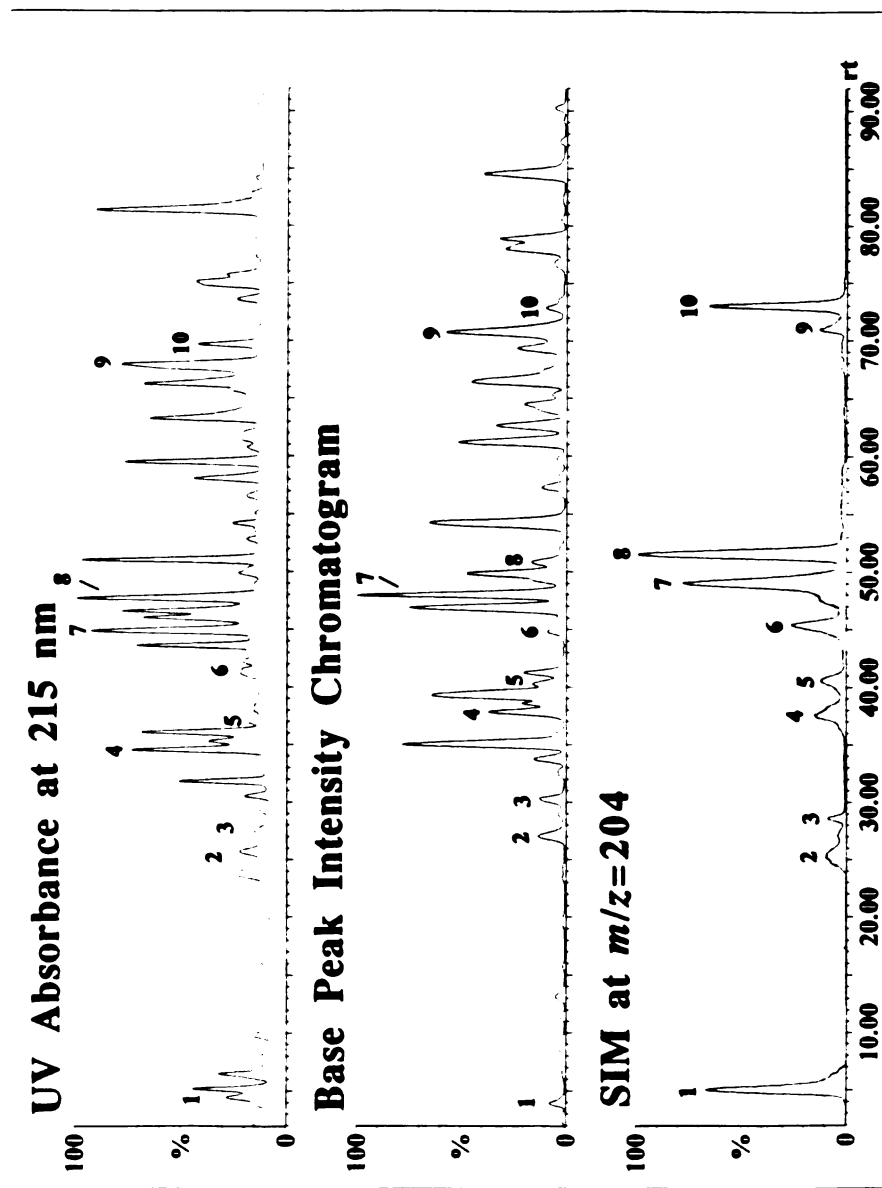


Figure 2. UV absorbance, base peak intensity, and SIM (m/z 204) chromatograms obtained from HPLC/ESI/MS analysis of LCAT tryptic digest. The peaks labeled 1 through 10 were identified as glycopeptides. Range=0.05 AUFS. Solvent A=0.1% TFA, solvent B=0.08% TFA in ACN. Gradient program was 2-70% B in 120 min followed by 70-98% B in 10 min.

mode. The following selected mass values (B-ions) and the indicated carbohydrate moieties were monitored: m/z 204 [HexNAc]⁺, m/z 292 [NeuAc]⁺, m/z 366 [Hex-HexNAc]⁺ and m/z 657 [NeuAc-Hex-HexNAc]⁺. The chromatogram representing SIM at m/z 204 is illustrated in Figure 2. All four SIM chromatograms (shown in Figure 3) could be virtually superimposed, indicating that the results obtained are definitely due to the presence of sialylated complex oligosaccharides (Kornfeld and Kornfeld, 1976) and not fortuitously formed peptide fragments. Ten different glycopeptide-containing fractions were identified, as shown in Figure 2, where the corresponding UV, MS and SIM peaks are labeled 1 through 10. Once the glycopeptide-containing peaks were identified, the electrospray mass spectra of the glycopeptides from the first analysis could be identified by direct comparison of the UV chromatograms from the two experiments and the retention times of these SIM peaks, and the molecular masses of the intact glycopeptides determined.

Because of the many ways that monosaccharides in an oligosaccharide can be linked together, there is a very large number of possible carbohydrate structures which could represent a given composition. However, the type of *N*-linked structures typical of mammalian proteins (represented in three classes termed high mannose, hybrid and complex (Kornfeld, 1976), contain a defined number of sequence and branching combinations, and the molecular weight of the carbohydrate portion of a glycopeptide enables one to predict the oligosaccharide composition as well as the structural class of the structure present on that glycopeptide. By subtracting the calculated molecular weight of the amino acids from the measured mass of the glycopeptides it was possible to deduce the likely composition as well as the class and general structure (e.g. complex, trisialo triantennary of the carbohydrate moieties). Based on the molecular weights and the SIM data, the peaks labeled 1 and 7-10 in Figure 2 contained the four predicted LCAT tryptic *N*-linked glycopeptides (where peaks 9 and 10 contained the same glycopeptide except the peptide in fraction 10 contained an *N*-terminal pyroglutamic acid instead of a Gln based on

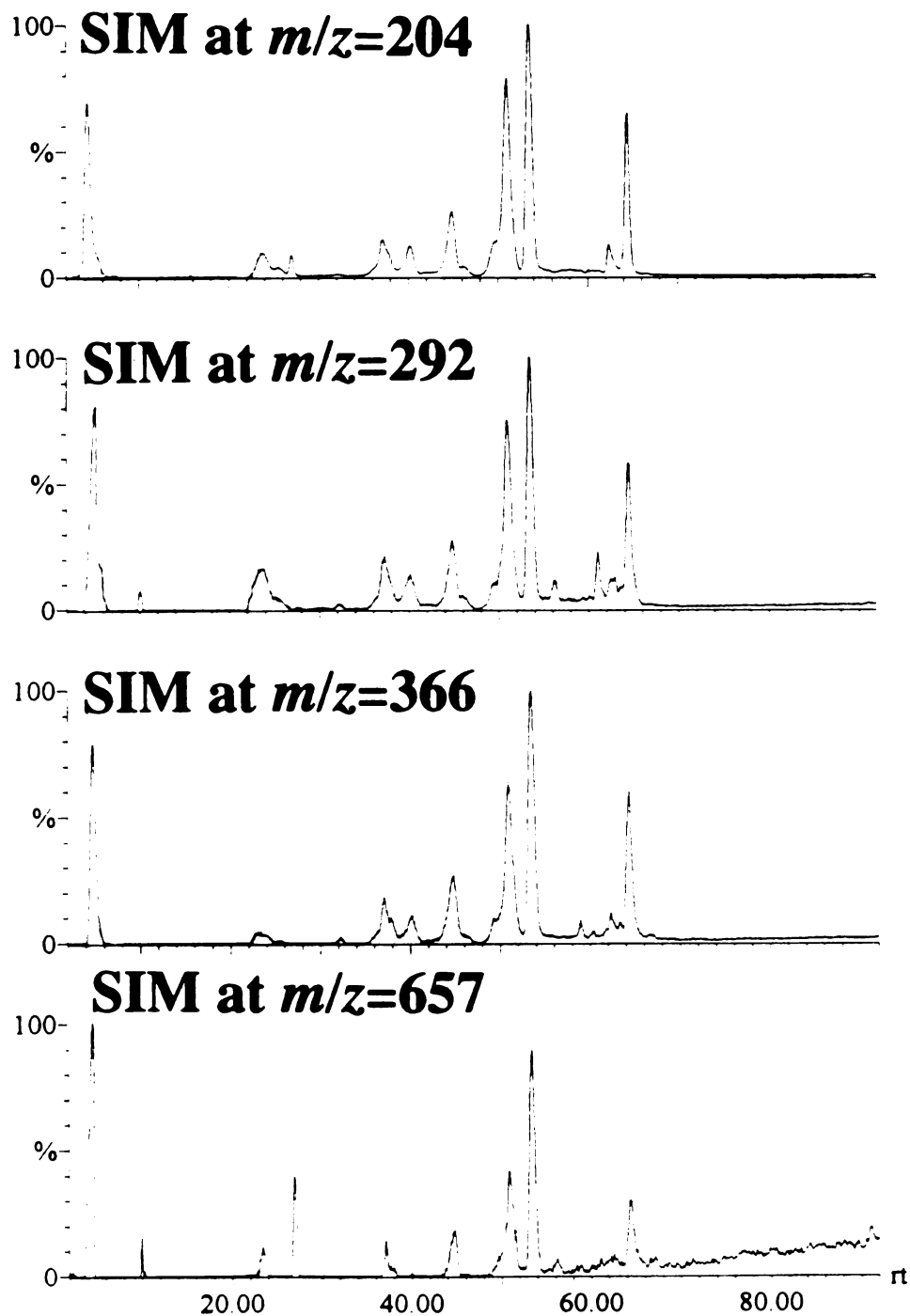


Figure 3. Chromatograms obtained from selected ion monitoring at m/z 204 [HexNAc⁺], m/z 292 [NeuAc⁺], m/z 366 [Hex-HexNAc⁺] and m/z 657 [NeuAc-Hex-HexNAc⁺] from a 300 pmole injection of LCAT.

a molecular weight difference of 17 Da and the presence of an N-terminal Gln in the peptide), and the proposed oligosaccharide composition for each glycoform is described in Table I. Unexpectedly, the peaks labeled 2 and 3 in Figure 2 contained molecular weights which corresponded to the C-terminal tryptic peptide (residues 400-416 with attached *O*-linked oligosaccharides (since this peptide contains no Asn but one threonine and two serine residues), and the proposed oligosaccharide composition for each glycoform is described in Table II. Again, the glycopeptide molecular weight from fraction 3 was 17 Da less than another glycopeptide in fraction 2, and was assumed to be a pyroglutamic acid-containing glycopeptide, since this peptide also contained an N-terminal Gln. Based on the identified molecular weights, fractions 4, 5 and 6 in Figure 2 did not appear to be LCAT *N*-linked or *O*-linked glycopeptides, and further characterization was required to determine their origin.

Characterization of glycopeptides by Edman degradation. Since several unexpected glycopeptide components were identified in the tryptic digest, the ten fractions which were identified as containing glycopeptides from the SIM data were also analyzed by automated Edman degradation. The peptide and glycopeptide sequences determined by our Edman analyses are indicated in Figure 1 by the dotted arrows. By confirming the peptide sequence of the glycopeptides, the Edman data, combined with molecular weight and SIM data obtained from the corresponding mass spectra, established the identification of all of the expected LCAT *N*-linked glycopeptides and confirmed the presence of the unexpected *O*-linked glycopeptides, as described in Tables I and II, respectively. For example, Edman degradation confirmed that the peak labeled 1 in Figure 2 corresponds to the peptide containing residues (⁸¹VVYNR⁸⁵), which contains a consensus sequon for *N*-linked glycosylation at Asn⁸⁴. The carbohydrate composition of the major component (based on signal intensity in the ESI mass spectrum, which was deduced from the difference between the expected (MH⁺=650.8 Da) and the measured (MH⁺=3511.7 Da) molecular mass of this peptide, represents a triantennary complex structure, as listed in Table I. The electrospray

| Fr. # | Glycosylation Site & Sequence Range | Intact Glycopeptide Mw (Da) | Carbohydrate Composition | Mw after Neuraminidase Digestion (Da) | Mw after β -Galactosidase Digestion | Mw after Hexosaminidase Digestion | Confirmed Carbohydrate Structure |
|--------|-------------------------------------|-----------------------------|--|---------------------------------------|---|-----------------------------------|----------------------------------|
| Fr. 1 | Asn 84 (81-85) | 3221.8 | Hex ₆ HexNAc ₅ NeuAc ₂ | 2639.0 | 2152.0 | 1543.6 | Disialo Triantennary |
| | | 3611.7 | Hex ₆ HexNAc ₅ NeuAc ₃ | 2639.0 | 2152.0 | 1543.6 | Trisialo Triantennary |
| | | 3656.8 | Hex ₆ HexNAc ₅ NeuAc ₃ Fuc ₁ | 2784.7 | 2297.6 | 1891.6 | Fuc. Trisialo Triantennary |
| Fr. 7 | Asn 20 (16-39) | 4822.5 | Hex ₅ HexNAc ₄ NeuAc ₂ | 4240.2 | 3914.9 | 3510.8 | Disialo Biantennary |
| | | 5186.4 | Hex ₆ HexNAc ₅ NeuAc ₂ | 4605.0 | 4118.5 | 3510.8 | Disialo Triantennary |
| | | 5479.2 | Hex ₆ HexNAc ₅ NeuAc ₃ | 4685.0 | 4118.5 | 3618.8 | Trisialo Triantennary |
| | | 5625.6 | Hex ₆ HexNAc ₅ NeuAc ₃ Fuc ₁ | 4753.0 | 4426.3 | 4021.0 | Fuc. Trisialo Triantennary |
| | | 5842.4 | Hex ₇ HexNAc ₆ NeuAc ₃ | 4969.9 | 4318.5 | 3510.8 | Trisialo Tetrantennary |
| Fr. 8 | Asn 272 (257-276) | 4268.7 | Hex ₅ HexNAc ₄ NeuAc ₁ | 3979.8 | 3653.8 | 3247.5 | Monosialo Biantennary |
| | | 4560.9 | Hex ₅ HexNAc ₄ NeuAc ₂ | 3979.8 | 3653.8 | 3247.5 | Disialo Biantennary |
| | | 4926.0 | Hex ₆ HexNAc ₅ NeuAc ₂ | 4345.5 | 3856.6 | 3247.5 | Disialo Triantennary |
| | | 5217.3 | Hex ₆ HexNAc ₅ NeuAc ₃ | 4345.5 | 3856.6 | 3247.5 | Trisialo Triantennary |
| | | 5363.0 | Hex ₆ HexNAc ₅ NeuAc ₃ Fuc ₁ | 4492.1 | 4164.8 | 3757.2 | Fuc. Trisialo Triantennary |
| Fr. 9 | Asn 384 (363-399) | 6111.8 | Hex ₅ HexNAc ₄ NeuAc ₁ | 5822.0 | 2715.9 (379-390) | 2310.5 (379-390) | Monosialo Biantennary |
| | | 6404.3 | Hex ₅ HexNAc ₄ NeuAc ₂ | 5822.0 | 2715.9 (379-390) | 2310.5 (379-390) | Disialo Biantennary |
| Fr. 10 | pE(363-399) | 6095.5 | Hex ₅ HexNAc ₄ NeuAc ₁ | 5805.0 | 2715.9 (379-390) | 2310.5 (379-390) | Monosialo Biantennary |
| | | 6386.7 | Hex ₅ HexNAc ₄ NeuAc ₂ | 5805.0 | 2715.9 (379-390) | 2310.5 (379-390) | Disialo Biantennary |

Table I. HPLC/Electrospray mass spectrometry data obtained from sequential glycosidase digestion of LCAT *N*-linked glycopeptides. Bold type indicates the major species identified based on the relative abundance of the molecular ions in the mass spectra.

| Fr. # | Glycosylation Site and Glycopeptide Sequence Range | Intact Glycopeptide Mw (Da) | Carbohydrate Composition | Mw after Neuraminidase + O-Glycanase Digestion | Confirmed Carbohydrate Structures |
|-------|--|-----------------------------|---|--|--|
| Fr. 3 | Thr ⁴⁰⁷ and Ser ⁴⁰⁹ (400-416) | 2968.5 | Hex ₂ HexNAc ₂ NeuAc ₂ | 1656.9 | 2[NeuAc-Gal β 1 \rightarrow 3GalNAc] |
| | | 3259.6 | Hex ₂ HexNAc ₂ NeuAc ₃ | 1656.9 | 1[NeuAc-Gal β 1 \rightarrow 3GalNAc] + 1[NeuAc ₂ -Gal β 1 \rightarrow 3GalNAc] |
| Fr. 2 | | 3551.6 | Hex ₂ HexNAc ₂ NeuAc ₄ | 1656.9 | 2[NeuAc ₂ -Gal β 1 \rightarrow 3GalNAc] |
| | | 3624.9 | Hex ₃ HexNAc ₃ NeuAc ₃ | | |
| | | 3914.9 | Hex ₃ HexNAc ₃ NeuAc ₄ | | |
| | | 4208.4 | Hex ₃ HexNAc ₃ NeuAc ₅ | | |
| Fr. 2 | pE(400-416) | 2951.3 | Hex ₂ HexNAc ₂ NeuAc ₂ | 1638.8 | 2[NeuAc-Gal β 1 \rightarrow 3GalNAc] |

Table II. HPLC/Electrospray mass spectrometry data obtained from glycosidase digestion of LCAT *O*-linked glycopeptides. Bold type indicates the major species identified based on the relative abundance of the molecular ions in the mass spectra.

data also indicated the presence of two minor glycopeptide components, one containing an additional fucose residue and another containing only two sialic acid residues (see Table I). (As discussed in Chapter 5, when comparing glycopeptides with the same peptide sequence and different oligosaccharide components, the relative ion intensities in the ESIMS data correlate closely with the relative quantities of each glycoform). Similarly, Edman degradation confirmed that peak 7 in Figure 2 represents the peptide containing LCAT residues ($^{16}\text{AELSNHTRPVILVPGCLGNQLEAK}^{39}$), which contains a consensus sequon for *N*-linked glycosylation at Asn²⁰. The mass spectrum for this glycopeptide is shown in Figure 4. The molecular weights obtained indicate a trisialo triantennary structure as the major species. In addition, small amounts of biantennary, tetraantennary and fucosylated triantennary structures are present in this spectrum (listed in Table I).

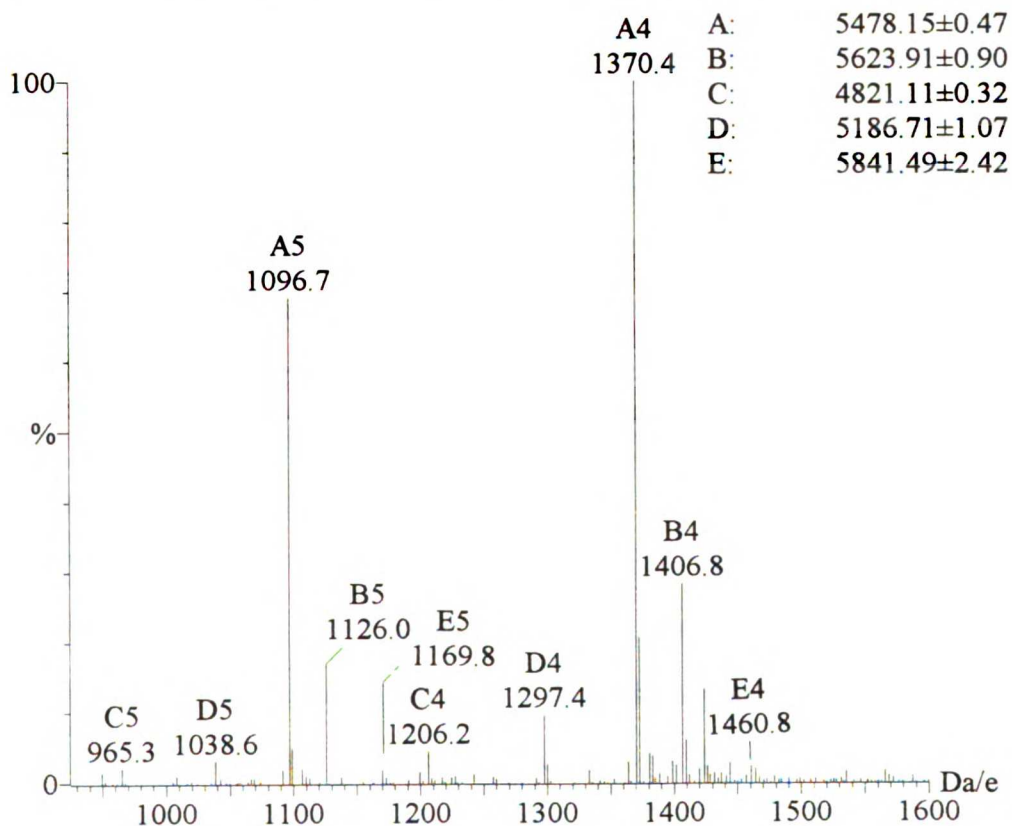


Figure 4. Electrospray mass spectrum of the LCAT tryptic peptide mixture containing residues 16-39. The molecular masses of the four glycopeptide components represent this peptide with complex oligosaccharides attached at Asn²⁰.

Edman degradation of peak 8 in Figure 2 yielded a sequence corresponding to residues (257MAWPEDHVFISTPSFNVTGR²⁷⁶), which contains a consensus sequon for *N*-linked glycosylation at Asn²⁷². The mass spectrum obtained for glycopeptide 257-276 contains five glycopeptide glycoforms as shown in Figure 5A. The molecular weight of the major component in the spectrum corresponds to a trisialo triantennary structure, as listed in Table I, with small amounts of sialylated biantennary and fucosylated trisialo triantennary as well. Similarly, Edman degradation established that the peak labeled 9 in Figure 2 represented the peptide (363QPQPVHLLPLHGIQHLLNMVFSNLTLEHINAILLGAYR³⁹⁹), which contains a consensus sequon for *N*-linked glycosylation at Asn³⁸⁴. The molecular masses obtained for this fraction indicated the presence of glycopeptides containing residues (363-399), with a disialo biantennary carbohydrate structure as the major species, and some monosialo biantennary as well, as shown in Table I. No Edman degradation data was obtained from peak 10, however. The ESIMS data (shown in Table I) indicated that this fraction very likely contained the same glycopeptides as in peak 9, but with a pyroglutamic acid (pGlu) as the amino terminal (*N*-terminal) amino acid rather than the expected Gln, based on the mass difference of 17 Da. The fact that no Edman data was obtained further supports this premise, because the presence of a pGlu at the *N*-terminus of this peptide would block the first cycle of Edman degradation, therefore, preventing a sequence from being obtained.

Edman degradation also established that peak 2 in Figure 2 represented the C-terminal peptide of LCAT (400QGPPASPTASPEPPPPE⁴¹⁶), but this peptide does not contain a consensus sequon for *N*-linked glycosylation and was not previously found to be glycosylated. However, as shown in Figure 1, this peptide contains two serine residues at positions 405 and 409 and one threonine residue at position 407 which could be glycosylated with *O*-linked oligosaccharides. The sequence obtained from Edman degradation of peak 2 yielded two blank cycles which correspond to Thr⁴⁰⁷ and Ser⁴⁰⁹, indicating the likely presence of *O*-linked glycosylation at these sites. Several molecular

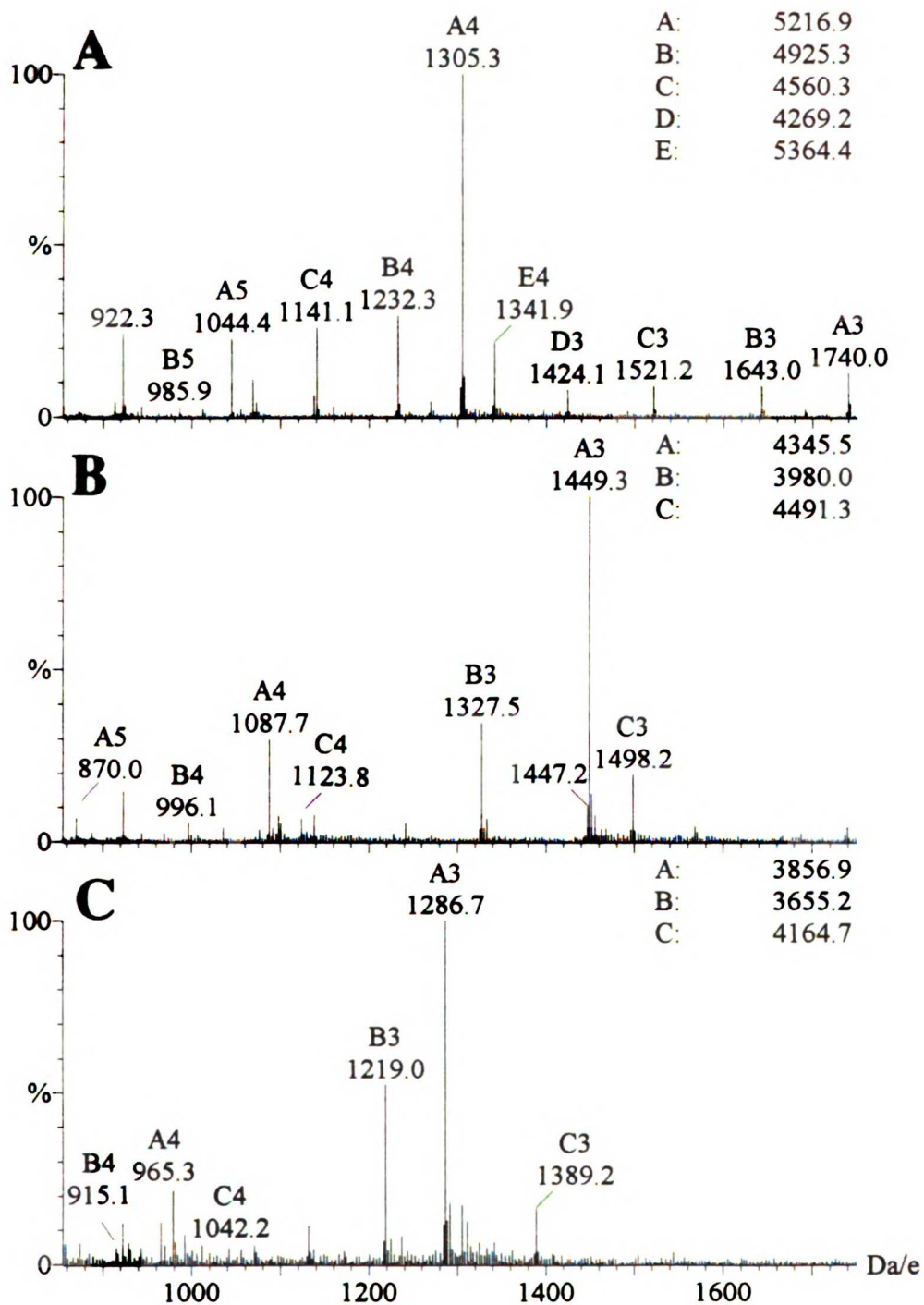


Figure 5. A) Electrospray mass spectrum of LCAT tryptic glycopeptides containing residues 257-276. The molecular masses of the five glycopeptide components represent this peptide with complex oligosaccharides attached at Asn²⁷², respectively. B) Electrospray mass spectrum of the glycopeptide mixture from (A) after digestion with neuraminidase. C) Electrospray mass spectrum of the glycopeptide mixture from B) after digestion with β -galactosidase.

weights were identified in the HPLC/ESIMS data, representing several different carbohydrate compositions as listed in Table II. The major *O*-linked species contained a composition corresponding to two hexoses, two *N*-acetylhexosamines and two sialic acids. Just as with peak 10, peak 3 in Figure 2 did not yield a sequence by Edman degradation because, as the mass spectrometric data indicated, this fraction contained a molecular mass which was 17 Da less than the major glycopeptide species identified in peak 2. This was assumed to be the same major glycopeptide component as in peak 2 (400–416), but with a pGlu as the amino terminal (*N*-terminal) amino acid rather than the expected Gln, and the absence of an Edman sequence further supports this.

Edman degradation analysis of peaks 4, 5 and 6 confirmed that these fractions did not contain LCAT glycopeptides. From a search of the protein data base, the sequences obtained by Edman were identified as peptides from a second glycoprotein, apolipoprotein D (Apo D), which was apparently copurified with the LCAT glycoprotein. This protein, which contains two potential sites of *N*-linked glycosylation at Asn⁴⁵ and Asn⁷⁸, has been previously shown to associate with LCAT (Albers *et al.*, 1976; Utermann *et al.*, 1980), however, no data has previously been reported on its oligosaccharide structures. In addition, exoglycosidase sequencing combined with HPLC/ESIMS (described below) also established that these fractions did indeed contain *N*-linked glycopeptides from Apo D with the following peptide sequences; ⁴¹CIQANYSLMENGK⁵³, ⁴¹CIQANYSLMENGKIK⁵⁵ and ⁶³ADGTVNQIEGEATPVNLTTEPAKKLEVK⁸⁸. Exoglycosidase sequencing of peaks 4, 5 and 6 from Figure 2 also indicated that Asn⁴⁵ contains primarily trisialo triantennary oligosaccharides and Asn⁷⁸ contains primarily fucosylated disialo biantennary oligosaccharides and the data is summarized in Table III.

| Fr. #. | Glycosylation Site & Sequence Range | Intact Glycopeptide Mw (Da) | Proposed Carbohydrate Composition | Mw after Neuraminidase Digestion (Da) | Mw after β -Galactosidase Digestion | Confirmed Carbohydrate Structure |
|--------|-------------------------------------|-----------------------------|--|---------------------------------------|---|---|
| 4 | Asn ⁴⁵ (41-53) | 3734.9 | Hex ₅ HexNAc ₄ NeuAc ₂ | 3167.9 | 2843.9 | Disialo Biantennary |
| | | 4100.3 | Hex ₆ HexNAc ₅ NeuAc ₂ | 3533.1 | 3047.1 | Disialo Triantennary |
| | | 4390.0 | Hex ₆ HexNAc ₅ NeuAc ₃ | 3533.1 | 3047.1 | Trisialo Triantennary |
| | | 4682.3 | Hex ₆ HexNAc ₅ NeuAc ₄ | 3533.1 | 3047.1 | Tetrasialo Triantennary |
| | | 4756.9 | Hex ₇ HexNAc ₆ NeuAc ₃ | 3899.2 | 3413.1 | Trisialo Triantennary Lac ₁ |
| | | 5044.4 | Hex ₇ HexNAc ₆ NeuAc ₄ | 3899.2 | 3413.1 | Tetrasialo Triantennary Lac ₂ |
| | | | | | | |
| 5 | Asn ⁴⁵ (41-55) | 3978.6 | Hex ₅ HexNAc ₄ NeuAc ₂ | 3409.8 | 3086.4 | Disialo Biantennary |
| | | 4268.6 | Hex ₅ HexNAc ₄ NeuAc ₃ | 3409.8 | 3086.4 | Disialo Triantennary |
| | | 4632.5 | Hex ₆ HexNAc ₅ NeuAc ₃ | 3774.9 | 3287.9 | Trisialo Triantennary |
| | | 4924.8 | Hex ₆ HexNAc ₅ NeuAc ₄ | 3774.9 | 3287.9 | Tetrasialo Triantennary |
| | | | | 4140.4 | 3655.7 | Triantennary Lac ₁ |
| | | | | 4509.5 | 4025.1 | Triantennary Lac ₂ |
| 6 | Asn ⁷⁸ (63-88) | 4783.3 | Hex ₅ HexNAc ₄ NeuAc ₁ Fuc ₁ | 4491.9 | 4170.1 | Fuc. Monosialo Biantennary |
| | | 4927.8 | Hex ₅ HexNAc ₄ NeuAc ₂ | 4346.9 | 4022.9 | Disialo Biantennary |
| | | 5075.0 | Hex ₅ HexNAc ₄ NeuAc ₂ Fuc ₁ | 4491.9 | 4170.1 | Fuc. Disialo Biantennary |
| | | 5443.2 | Hex ₆ HexNAc ₅ NeuAc ₂ Fuc ₁ | 4858.2 | 4372.2 | Fuc. Disialo Triantennary |
| | | 5731.0 | Hex ₆ HexNAc ₅ NeuAc ₃ Fuc ₁ | 4858.2 | 4372.2 | Fuc. Trisialo Triantennary |
| | | 6099.5 | Hex ₇ HexNAc ₆ NeuAc ₃ Fuc ₁ | 5222.6 | 4736.6 | Fuc. Trisialo Triantennary Lac ₁ |
| | | | | 5589.6 | 4936.3 | Fuc. Tetraantennary Lac ₁ |
| | | | | | | |

Table III. HPLC/ESIMS data obtained from sequential glycosidase digestion of Apo D N-linked glycopeptides. Bold type indicates the major species identified based on the relative abundance of the molecular ions in the mass spectra. Lac = Gal-GlcNAc, indicating an additional lactosamine group extended on one of the branches.

Characterization of O-linked carbohydrate structures. The carbohydrate composition of the major O-linked glycopeptide (two hexoses, two N-acetylhexosamines and two sialic acids) deduced from the Edman and molecular weight data, was further investigated in order to determine the exact structures of the carbohydrates at Thr⁴⁰⁷ and Ser⁴⁰⁹. Based on the previous structures of O-linked glycopeptides identified on human proteins (Conradt *et al.*, 1989; Clogston *et al.*, 1993; Nimitz *et al.*, 1993), combined with the proposed composition described above, one would predict that both sites would contain sialyl hexosyl N-acetylhexosamine moieties. In order to first determine if each site contained a single hexosyl N-acetylhexosamine disaccharide, the O-linked glycopeptide shown in Figure 6A was digested with neuraminidase (*Arthrobacter ureafaciens*) to

remove terminal sialic acid residues from the glycopeptide (Uchida *et al.*, 1979), followed by endo- α -N-acetylgalactosaminidase (*Diplococcus pneumoniae*, O-Glycanase). The latter enzyme specifically cleaves (Gal β 1 \rightarrow 3GalNAc-Ser/Thr) structures between Ser/Thr and GalNAc residues (Umemoto *et al.*, 1977), resulting in the removal of these very specific O-linked glycans from glycopeptides or glycoproteins. The resulting digest was analyzed by HPLC/ESIMS yielding the spectrum shown in Figure 6B. The molecular weight of the major glycopeptide component shifted from 2951.3 Da in Figure 6A to 1638.8 Da in Figure 6B. This data indicates that the O-linked glycosylation was completely removed from the glycopeptide, leading to the conclusion that Thr⁴⁰⁷ and Ser⁴⁰⁹ were each glycosylated with a Gal β 1 \rightarrow 3GalNAc structure. Based on the molecular weight data summarized in Table II, the major O-linked glycopeptide species very likely contains primarily one sialic acid residue attached to each Gal β 1 \rightarrow 3GalNAc structure, and the other glycoforms such as the component with a molecular weight of 3551.6 and a composition of Hex₂HexNAc₂NeuAc₄, must represent two Gal β 1 \rightarrow 3GalNAc structures each with two sialic acid residues. In addition, the presence of one very small peak at 25-30 min in the SIM at *m/z* 366 combined with the presence of two larger peaks at 25-30 min in the SIM at *m/z* 657 chromatogram shown in Figure 3, led us to conclude that the major glycoform contains one sialic acid residue on each Gal β 1 \rightarrow 3GalNAc structure, rather than two sialic acid residues on one and zero on another. We were unable to determine whether the minor O-linked glycopeptides containing compositions of Hex₃HexNAc₃NeuAc₃₋₅ were glycosylated at all three potential sites or only at Thr⁴⁰⁷ and Ser⁴⁰⁹ (with larger structures), due to limited quantities of sample.

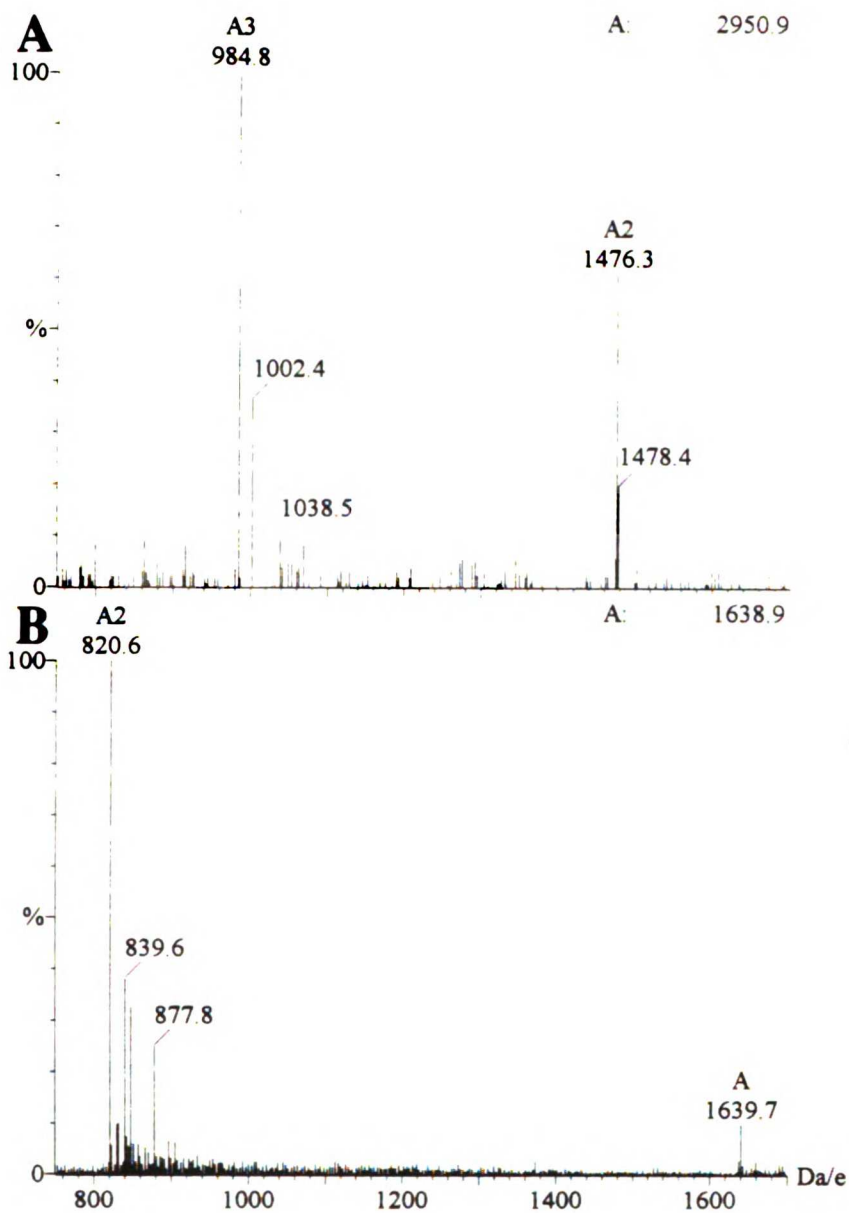


Figure 6. (A) Electrospray mass spectrum of LCAT C-terminal tryptic glycopeptide containing residues 400-416. The molecular mass of the glycopeptide represents this peptide where the N-terminal Glu has been cyclized to pGlu, containing an oligosaccharide composition of HexNAc₂Hex₂NeuAc₂. (B) Electrospray mass spectrum of the glycopeptide in A) after sequential digestion with neuraminidase followed by endo- α -N-acetylgalactosaminidase.

Characterization of N-linked carbohydrate structures. In order to establish the structures possible from knowledge of the carbohydrate compositions listed in Table I, in separate experiments, each of the four isolated glycopeptides was treated sequentially with specific exoglycosidases, to remove the exposed nonreducing terminal monosaccharides, with analysis by HPLC/ESIMS after each step. Each exoglycosidase used was specific for one type of monosaccharide residue such as GlcNAc or Gal. The mass difference after each single exoglycosidase digestion and the monosaccharide specificity of each exoglycosidase were used to determine the sequence of the nonreducing terminal portion of the structures. For example, a change in mass of a glycopeptide after digestion with the neuraminidase, indicates first that sialic acid residues were present at the nonreducing terminus of the oligosaccharide, and the value of the mass difference indicates the number of sialic acid residues removed from the oligosaccharide. By using the enzymes sequentially, one can determine the sequence and branching of the oligosaccharides attached to the glycopeptides, and thereby establish the structures of the oligosaccharides at each site. To confirm the published specificity of the enzymes to be used, we first sequentially digested glycopeptides of known sequence from bovine fetuin (Green and Baenziger, 1988; Cumming *et al.*, 1989) and analyzed them by HPLC/ESIMS (data not shown). After establishing the specificities of the enzymes, each of the eight different glycopeptide mixtures (fractions 1 and 4-10) was first digested with *Arthrobacter ureafaciens* neuraminidase and analyzed by HPLC/ESIMS. The change in the molecular masses was then used to determine the number of sialic acid residues removed from each glycopeptide. As with the earlier HPLC/ESIMS experiments, only 5% of the sample injected actually entered the mass spectrometer, and 95% of the sample was recollected for digestion with the next exoglycosidase, β -galactosidase from *Streptococcus pneumoniae*, which removes terminal galactose residues linked $\beta 1 \rightarrow 4$ to GlcNAc or GalNAc (Distler and Jourdian, 1973; Kobata, 1979). After analysis by HPLC/ESIMS, the change in mass from before the β -galactosidase digestion indicated whether the oligosaccharides were

biantennary, triantennary or tetraantennary. For example, a biantennary structure will lose two galactose residues (a mass difference of 324 Da) upon β -galactosidase digestion, while a triantennary oligosaccharide will lose three galactose residues (a mass difference of 486 Da). Finally, the glycopeptides were digested with β -N-acetylhexosaminidase from chicken liver, which removes GlcNAc residues (Kobata, 1979). This data is summarized in Table I.

The results of the HPLC/ESIMS data before and after digestion of the *N*-linked glycopeptide 257-276 with neuraminidase, shown in Figure 5A and 5B, respectively, provided clear evidence that the major species present contained a trisialylated glycan, while the minor species contained mono- and disialylated glycans. The mass difference of 875 Da (5217.3 Da minus 4342.3 Da) for the major species corresponds to the mass of three sialic acid residues removed from this glycopeptide. After digestion with β -galactosidase the HPLC/ESIMS data, shown in Figure 5C, yielded a mass difference corresponding to three Gal residues (486 Da) for the major glycopeptide, and a mass difference corresponding to two or three Gal residues (324 or 486 Da) for the minor components. After digestion with β -N-acetylhexosaminidase, the HPLC/ESIMS data yielded a mass difference corresponding to three GlcNAc residues (609 Da) for the major glycopeptide, and a mass difference corresponding to two GlcNAc residues (406 Da) for the minor components, consistent with the original compositions suggested in Table I. This data indicates that the original compositions from Table I are consistent with the presence of triantennary complex carbohydrate structures containing two and three sialic acid residues as well as zero and one fucose residues, and biantennary complex structures containing one and two sialic acid residues attached to Asn²⁷². However, only two Gal and two GlcNAc residues could be removed from the fucosylated triantennary structure. This was also found to be the case for the same glycan attached to Asn²⁰, while for the same glycan attached to Asn⁸⁴, the β -N-acetylhexosaminidase could only remove one GlcNAc. It is possible that the fucose on these glycans is attached to a Gal or GlcNAc on one of the three

branches of these structures, because β -galactosidase is known not to cleave Gal residues from branches which are substituted with fucose on the terminal Gal or an attached GlcNAc (Kobata, 1979).

Similar exoglycosidase experiments performed on the other three glycopeptides established the proposed structures listed in Table I. Glycopeptide (81-85) contains primarily a trisialo triantennary structure attached to Asn⁸⁴, with minor fucosylated trisialo triantennary and disialo triantennary species as well. A similar group of structures (trisialo triantennary (major), trisialo tetraantennary, fucosylated trisialo triantennary, disialo triantennary and disialo biantennary) were identified on glycopeptide (16-39). The fourth and largest *N*-linked glycopeptide (363-399) was the only one of the four which was found to contain only biantennary structures. The major component was a disialo biantennary oligosaccharide, and the only other structure identified was a monisialo biantennary oligosaccharide. As shown in Table I, we also experienced proteolytic cleavage of this peptide from digestion with the β -galactosidase. Other experiments with this enzyme, discussed in Chapter 5, have indicated that the β -galactosidase preparation contains some contaminating protease(s) which hydrolyze glycopeptides containing more than 25 amino acids. In addition, the exoglycosidase experiments enabled the determination that the LCAT tryptic digest also contained *N*-linked glycopeptide components from a second glycoprotein, apolipoprotein D (Apo D), which was copurified with the LCAT glycoprotein at a significantly lower quantity, clearly indicating the power of this technique for identifying and sequencing oligosaccharides on glycopeptides. These structures were established to be primarily trisialo triantennary oligosaccharides at Asn⁴⁵ and fucosylated disialo biantennary oligosaccharides at Asn⁷⁸.

Finally, the calculated average molecular weight of LCAT (59,187) based on the combined major carbohydrate structures (12,105) and the amino acid sequence (47,082) is in good agreement with that previously estimated by sedimentation equilibrium data to be approximately 59,000 (Chung *et al.*, 1979) to 60,000 (Chong *et al.*, 1983).

7.5 Discussion

In this study, HPLC/ESIMS, Edman degradation and glycosidase digestion were used to characterize the *N*-glycosylation sites of human LCAT, to assess their heterogeneity at each site and to partially elucidate the structures of the oligosaccharides. This powerful methodology also revealed the presence of unanticipated *O*-glycosylation on the C-terminal tryptic peptide, as well as the presence of another glycoprotein, Apo D, at a significantly lower sample quantity. The key factor in this work was the use of a mass spectrometric-based strategy (HPLC/ESI/CID/MS) for selective identification of glycopeptides in the complex mixture of peptides and glycopeptides by monitoring for diagnostic carbohydrate fragment ions generated by CID.

To obtain detailed structural information on the LCAT oligosaccharides a method described in detail in Chapter 5, which involves specific degradation of the *N*-linked oligosaccharides by sequentially removing the nonreducing terminal monosaccharides from each branch of the structures using exoglycosidases, with analysis by HPLC/ESIMS after each digestion step. Using the specificity of the exoglycosidases, combined with the change in mass of the glycopeptides before and after each digestion, sequence and branching information was obtained on the oligosaccharide portions of the *N*-linked glycopeptides. Of course, the amount of detailed structural information obtained is dependent on the specificity of the glycosidases used. As described in Chapter 5, this method of sequencing glycopeptides using exoglycosidases followed by ESIMS is much more sensitive than the classical strategy of removing oligosaccharides from the glycopeptides or glycoprotein, followed by derivatization such as permethylation or peracetylation and analysis by fast atom bombardment mass spectrometry (FAB-MS) (Carr, 1989 #16).

A very specific endoglycosidase was used to characterize the two *O*-linked oligosaccharides identified on the proline-rich C-terminal peptide of LCAT, each of which

contained sialylated Gal β 1 \rightarrow 3GalNAc structures. Their existence was not suspected before this work, because they are only responsible for a small mass increment compared to that for the *N*-linked carbohydrates, and previous Edman sequencing of this protein did not identify these sites (Yang *et al.*, 1987). The two sites are localized in the proline-rich (8 prolines out of 17 residues in the tryptic peptide) C-terminus of LCAT. While there is no consensus sequence for *O*-linked glycosylation, and therefore, no concrete method for predicting sites of *O*-linked glycosylation, a statistical study of the amino acid sequences around all known *O*-linked glycosylation sites (Wilson *et al.*, 1991) indicated that the most prominent feature in the vicinity of these sites is the increased frequency of proline residues, as described in Chapter 3. Wilson and coworkers also found that in areas of multiple glycosylation sites, prolines are found most frequently in positions -1 as well +3 amino acid residues relative to the glycosylation sites. Interestingly, these findings fit well with the sites of *O*-glycosylation found on LCAT (see Figure 1). Edman degradation showed that only two of the three potential glycosylation sites (Thr⁴⁰⁷ and Ser⁴⁰⁹, but not Ser⁴⁰⁵) in the C-terminal tryptic peptide were found to bear carbohydrates. No proline residues are located in position -1 or +3 relative to Ser⁴⁰⁵, whereas Ser⁴⁰⁹ has a proline located in position +3 and Thr⁴⁰⁷ contains prolines at position -1 as well as +3 to it. The possible biological significance of an *O*-glycosylated C-terminus for LCAT is unclear.

This study established that three of the four *N*-linked glycosylation sites (Asn²⁰, Asn⁸⁴ and Asn²⁷²) were glycosylated with the same group of major (trisialo triantennary) and minor (fucosylated trisialo triantennary, disialo triantennary and disialo biantennary) carbohydrate structures (Asn²⁰ contained a very small percentage of trisialo tetraantennary as well). The fourth site, Asn³⁸⁴, contained only biantennary (mono- and disialo) structures. A previous study involving glycosidases and specific inhibitors of glycoprotein assembly with plasma LCAT and cultured cells secreting LCAT activity concluded that LCAT contained both high mannose and complex type *N*-linked oligosaccharides (Collet and Fielding, 1991). Treatment of plasma LCAT with Endoglycosidase F (which

hydrolyzes the chitobiose core of high mannose and biantennary complex oligosaccharides (Elder,1982) or Endoglycosidase H (which hydrolyzes the chitobiose core of high mannose and some hybrid oligosaccharides (Tarentino, 1974) yielded two species on an SDS-polyacrylamide gel. Therefore, the presence of high mannose as well as complex oligosaccharides was concluded. However, no evidence of high mannose oligosaccharides was found in this work. The Endo F would have removed the biantennary complex structures which we identified, but the Endo H should not have cleaved any complex oligosaccharides. It is very likely that the Endo H preparation contained some contaminating Endo F activity, and therefore cleaved the biantennary structures as well. In fact it was published subsequent to this study that the source for Endo H also contained three different Endo F isozymes, only one of which contained similar activity to Endo H, while the others cleaved all types of *N*-linked oligosaccharides (Trimble and Tarentino, 1991). This clearly shows that one must be cautious when using enzymes isolated from natural sources and low resolution analytical techniques such as SDS-polyacrylamide gel electrophoresis, and again demonstrates the power of the mass spectrometric techniques used for establishing structures.

The exact function of the LCAT oligosaccharides is not known, however, site-directed mutagenesis has indicated that removal of the glycosylation site at Asn²⁷² converts LCAT to a phospholipase, generating fatty acids rather than cholesteryl esters (Francone *et al.*, 1993). The results described herein do not support such a discrimination among the different *N*-glycosylation sites based solely on a knowledge of the carbohydrate structures present at each specific site. Therefore, the enzyme substrate specificity may not depend on the presence of a specific carbohydrate structure at Asn²⁷², but may depend on the existence of any hydrophilic moiety at this special site on the protein and formation of the appropriate substrate binding pocket. An earlier study from some of the authors (Collet and Fielding, 1991) supports this hypothesis, because the data indicated that inhibitors of the processing of *N*-linked oligosaccharides during glycoprotein assembly in cultured cells

secreting LCAT activity had no effect on the catalytic activity of LCAT. This also supports the hypothesis formulated by Francone and coworkers, which states that the oligosaccharides at position 272 play an indirect role, possibly by aligning the sterol 3-hydroxyl group as the lecithin acyl group is transferred from Ser¹⁸¹ in LCAT (Francone *et al.*, 1993). It would be interesting to verify what length or size is necessary for the oligosaccharides at position 272 to maintain full enzyme activity and the appropriate specificity. This could be achieved by trimming down this carbohydrate moiety by sequentially digesting LCAT with specific exoglycosidases, and recording the enzymatic activity after each step.

CHAPTER 8. CONCLUSIONS

While complete knowledge of the structure, biosynthesis and expression of a particular oligosaccharide does not necessarily indicate its specific function, nevertheless, to completely understand the biological role of carbohydrates on glycoproteins, the heterogeneous group of structures present at each site on a glycoprotein must first be characterized so that appropriate experiments can be designed to assess their function and determine any structure-function relationships. As discussed in Chapters 1 and 2, the number of detailed site-specific structures of oligosaccharides on glycoproteins continues to expand rapidly due to the significant improvements in enzymatic, chromatographic and mass spectrometric methods of oligosaccharide and glycopeptide analysis over the past five years. While improvements continue to be made, detailed structural analysis of oligosaccharides on glycoproteins remains a challenging task, and no single analytical tool can be used to detect and/or characterize all oligosaccharide structures which contain such component variability and complexity. Most detailed analyses require the use of a combination of chemical, enzymatic, chromatographic, mass spectrometric and NMR methods.

The majority of the strategies used today involve the use of several mass spectrometric methods including GC-MS (composition and linkage analysis), FAB-MS (LSIMS) and/or FAB-MS/MS of derivatized oligosaccharides which have been released from glycoproteins or glycopeptides. As described in this thesis, ESIMS and MALDI-TOFMS are now used routinely to analyze native glycopeptides and glycoproteins which have not been subjected to chemical or enzymatic procedures which may alter labile structural features in previously unforeseen ways. Because of their ability to analyze underivatized samples, coupled with high mass measurement capability and sensitivity, ESIMS and MALDI-TOFMS provide an important complement to the more established mass spectrometric and NMR techniques used for glycoprotein analysis. As demonstrated in Chapters 4 through 7, HPLC/ESIMS

has replaced FAB-MS and LSIMS as the method of choice for peptide and glycopeptide mapping. HPLC/ESIMS easily requires 10-1000 times less material than FAB-MS, depending on the sample and its molecular weight. Using HPLC/ESIMS combined with collision-induced dissociation and selected ion monitoring of carbohydrate-specific fragment ions, glycopeptides can now be rapidly identified in proteolytic digests of glycoproteins, as described in Chapters 4 through 7. Once identified, exo- and endoglycosidase digestions are then used to determine detailed sequence and branching information on both *N*- and *O*-linked oligosaccharides, including distinguishing between polylactosamine and more branched *N*-linked structures, as demonstrated in Chapter 4. Only by the analysis of glycopeptides can one determine oligosaccharide glycoform heterogeneity at a specific site on a glycoprotein. As described in Chapter 3, certain *O*-linked glycopeptides can also be analyzed by high energy MS/MS to determine the structures and sometimes the attachment sites of their *O*-linked oligosaccharides. Together, the methods described in this thesis 1) enable one to determine molecular size, 2) reveal heterogeneity directly and 3) provide considerable sequence and branching information

The site-specific characterization of glycoprotein oligosaccharides using mass spectrometry combined with enzymatic digestion and liquid chromatography is the core of the work presented in this thesis. Several glycoproteins have been characterized using the methods described in Chapters 1 and 2. Six sites of *O*-linked glycosylation were identified and characterized on recombinant human platelet-derived growth factor expressed in yeast (Chapter 3). High energy CID using four sector tandem mass spectrometry was used to determine the exact glycosylation sites for four of the six sites, because the glycopeptides obtained by trypsin and endoproteinase glu-C digestion contained more than one possible site of glycosylation. Monomannosyl structures were identified at five of the six sites, while dimannosyl structures were identified at two of the sites. Because LSIMS was used to identify and characterize the peptides and glycopeptides from this protein, high picomole and low nanomole quantities of sample was required for all analyses.

The *O*-linked oligosaccharides of recombinant human nerve growth factor receptor extracellular domain expressed in chinese hamster ovary cells (Chapter 4) were identified and characterized by HPLC/ESIMS and HPLC/CID/ESI/MS of an endoproteinase glu-C digest. While the intact sialylated *O*-linked glycopeptide was not successfully ionized, by separating the *O*-linked glycopeptides from the rest of the endoproteinase glu-C digest by lectin affinity chromatography and monitoring for the large sialylated fragments during HPLC/CID/ESI/MS, it was determined that the glycopeptide contains mono- and disialyl Gal β 1—3GalNAc structures. Furthermore, the seven sites of glycosylation fall in a highly conserved portion of the receptor, rather than the expected region which contains multiple proline residues. The ten unique *N*-linked oligosaccharide glycoforms at Asn³² of sNGFR (Chapter 5) were sequenced in detail using exoglycosidase digestion combined with HPLC/ESIMS. The oligosaccharides were determined to be primarily fucosylated, sialylated biantennary and triantennary structures. Triantennary oligosaccharides containing one and two additional lactosamine groups were also identified and distinguished from tetraantennary structures by specific changes in molecular weight after exoglycosidase digestion. Reanalysis after trimming of the branches of the oligosaccharides revealed the presence of larger structures not identified in the original endoproteinase glu-C digest due to their high *m/z* values as well as the dynamic range limitations of the electrospray ionization technique. However, the detailed analyses of sNGFR required only femtomole to low picomole quantities of sample, demonstrating the high sensitivity technique. Original analysis of the released derivatized oligosaccharides identified only three unique components and required nmole quantities of sample due to the losses from the derivatization procedure and the lower sensitivity of the LSIMS technique relative to ESIMS.

Five out of eleven possible sites of *N*-linked glycosylation on recombinant human platelet-derived growth factor β -receptor extracellular domain expressed in chinese hamster ovary cells (Chapter 6) were identified by HPLC/ESIMS and HPLC/ESI/CID/MS with

SIM of an endoproteinase glu-C digest. The predicted structures of the five groups of *N*-linked oligosaccharides were further established by glycosidase digestion with neuraminidase and PNGase F followed by HPLC/ESIMS analysis. The *N*-linked oligosaccharides characterized were fucosylated and/or sialylated bi-, tri- and tetraantennary complex structures. The sensitivity and selectivity of the SIM technique clearly showed the presence of other carbohydrate-containing components in the protein digest, however, no additional structural components were identified due to limited sample as well as possible impurities in the sample, causing poor ionization.

The *N*- and *O*-linked oligosaccharides of human plasma lecithin:cholesterol acyltransferase (Chapter 7) were also identified by HPLC/ESIMS and HPLC/ESI/CID/MS with SIM. The four groups of *N*-linked oligosaccharides were sequenced using exoglycosidase digestion followed by HPLC/ESIMS and determined to be sialylated biantennary and triantennary complex structures with a small percentage of fucosylated structures as well. The unexpected *O*-linked oligosaccharides which were identified were determined to be sialylated Gal β 1—3GalNAc structures. While site-directed mutagenesis had indicated that removal of the glycosylation site at Asn²⁷² changed the enzyme specificity of LCAT, the results described herein do not support such a discrimination among the different *N*-glycosylation sites based solely on a knowledge of the carbohydrate structures present at each specific site.

Interestingly, while the *O*-linked oligosaccharides on human LCAT fall at sites which agree with the (-1/+3) location tendency of *O*-linked sugars relative to proline proposed by Wilson (Wilson *et al.*, 1991), only one of the six *O*-linked structures from recombinant PDGF, and none of the *O*-linked structures from recombinant sNGFR fall at sites which support this "rule". Only but continued characterization and study of the structures and functions of oligosaccharides on glycoproteins, including comparison of structures on natural and recombinant homologs, will the understanding of these complex molecules be understood.

REFERENCES

Aberth, W., K. M. Straub and A. L. Burlingame 1982. "Secondary ion mass spectrometry with cesium ion primary beam and liquid target matrix for analysis of bioorganic compounds." Anal. Chem. **54**: 2029-2034.

Albers, J. J., V. G. Cabana and Y. Dee Barden Stahl 1976. "Purification and characterization of human plasma lecithin:cholesterol acyltransferase." Biochemistry **15**: 1084-1087.

Angel, A.-S., F. Lindh and B. Nilsson 1987. "Determination of binding positions in oligosaccharides and glycosphingolipids by fast-atom-bombardment mass spectrometry." Carbohydr. Res. **168**: 15-31.

Angel, A.-S., P. Lipniunas, K. Erlansson and B. Nilsson 1991. "A procedure for the analysis by mass spectrometry of the structure of oligosaccharides from high-mannose glycoproteins." Carbohydr. Res. **221**: 17-35.

Angel, A.-S. and B. Nilsson 1990. Linkage positions in glycoconjugates by periodate oxidation and fast atom bombardment mass spectrometry. In: J. A. McCloskey, ed. Methods Enzymology **193**. San Diego: Academic Press. pp. 587-607.

Ariga, T., T. Murata, M. Oshima, M. Maezawa and T. Miyatake 1980. "Characterization of glycosphingolipids by direct chemical ionization mass spectrometry." J. Lipid Res. **21**: 879-887.

Ariga, T., R. K. Yu, M. Suzuki, S. Ando and T. Miyatake 1982. "Characterization of GM1 ganglioside by direct chemical ionization mass spectrometry." J. Lipid Res. 23: 437-442.

Aron, L., S. Jones and C. J. Fielding 1978. "Human plasma lecithin-cholesterol acyltransferase. Characterization of cofactor-dependent phospholipase activity." J. Biol. Chem. 253: 7220-7226.

Ashwell, G. and J. Harford 1982. "Carbohydrate-specific receptors of the liver." Annu. Rev. Biochem. 51: 531-554.

Baenziger, J. and S. Kornfeld 1974. "Structure of the carbohydrate units of IgA1 immunoglobulin. I. Composition, glycopeptide isolation, and structure of the asparagine-linked oligosaccharide units." J. Biol. Chem 249: 7260-7269.

Ballou, C. E. 1954. "Alkali-sensitive glycosides." Adv. Carbohydr. Chem. 9: 91-100.

Banarjee, S. P., S. H. Synder, P. Cuatrecasas and L. A. Greene 1973. "Binding of nerve growth factor receptor in sympathetic ganglia." Proc. Natl. Acad. Sci., USA 70: 2519-2523.

Banner, D. W., A. D'Arcy, W. Janes, R. Gentz, H. J. Schoenfeld, C. Broger, H. Loetscher and W. Lesslauer 1993. "Crystal structure of the soluble human 55 kd TNF receptor-human TNF beta complex: implications for TNF receptor activation." Cell 73: 431-445.

Barber, M., R. S. Bordoli, R. D. Sedgwick and A. N. Tyler 1981. "Fast atom bombardment of solids (F.A.B.): a new ion source for mass spectrometry." J. Chem. Soc. Chem. Commun.: 325.

Barr, J. R., K. R. Anumula, M. B. Vettese, P. B. Taylor and S. A. Carr 1991. "Structural classification of carbohydrates in glycoproteins by mass spectrometry and high-performance anion-exchange chromatography." Anal. Biochem. 192: 181-192.

Battleman, D. S., A. I. Geller and M. V. Chao 1993. "HSV-1 vector-mediated gene transfer of the human nerve growth factor receptor p75hNGFR defines high-affinity NGF binding." J. Neurosci. 13: 941-951.

Bause, E. and H. Hettkamp 1979. "Primary structural requirements for *N*-glycosylation of peptides in rat liver." FEBS Lett. 108: 341-344.

Bause, E. and L. Lehle 1979. "Enzymatic *N*-glycosylation and *O*-glycosylation of synthetic peptide acceptors by dolichol-linked sugar derivatives in yeast." Eur. J. Biochem. 101: 531-540.

Bean, M. F., S. A. Carr, G. C. Thorne, M. H. Reilly and S. J. Gaskell 1991. "Tandem mass spectrometry of peptides using hybrid and four-sector instruments: a comparative study." Anal. Chem. 63: 1473-1481.

Beckey, H. D. 1970. In: ed. Recent Developments in mass Spectroscopy. Baltimore: University Park Press. pp. 1154-1165.

Beckey, H. D. 1977. Principles of field ionization and field desorption mass spectrometry. Oxford: Pergamon Press.

Begbie, R. 1974. "Studies on fetuin from foetal bovine serum. The composition and amino acid sequences of glycopeptides from fetuin." Biochim. Biophys. Acta 371: 549-576.

Benedetti, M., A. Levi and M. V. Chao 1993. "Differential expression of nerve growth factor receptors leads to altered binding affinity and neurotrophin responsiveness." Proc. Natl. Acad. Sci. USA 90: 7859-7863.

Bernard, B. A., S. A. Newton and K. Olden 1983. "Effect of size and location of the oligosaccharide chain on protease degradation of bovine pancreatic ribonuclease." J. Biol. Chem. 259: 12198-12202.

Bidlingmeyer, B. A., S. A. Cohen and T. L. Tarvin 1984. "Rapid analysis of amino acids using pre-column derivatization." J. Chromatogr. 336: 93-104.

Biemann, K. 1988. "Contributions of mass spectrometry to peptide and protein structure" Biomed. Environ. Mass Spectrom. 16: 99-111.

Biemann, K., D. C. DeJongh and H. K. Schnoes 1963. "Application of mass spectrometry to structure problems. XIII. Acetates of pentoses and hexoses." J. Am. Chem. Soc. 85: 1763-1771.

Biemann, K. and H. A. Scoble 1987. "Characterization by tandem mass spectrometry of structural modifications in proteins." Science 237: 992-998.

Björndal, H., B. Lindberg and S. Svensson 1967. "Mass spectrometry of partially methylated alditol acetates." Carbohydr. Res. 5: 433-440.

Brake, A. J., D. J. Julius and J. Thorner 1983. "A functional prepro-alpha-factor gene in *Saccharomyces* yeasts can contain three, four, or five repeats of the mature pheromone sequence." Mol. Cell. Biol. 3: 1440-1450.

Brake, A. J., J. P. Merryweather, D. G. Coit, U. A. Heberlein, F. R. Masiarz, G. T. Mullenbach, M. S. Urdea, P. Valenzuela and P. J. Barr 1984. "Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*." Proc. Natl. Acad. Sci. USA 81: 4642-4646.

Breitfeld, P. P., D. Rup and A. L. Schwartz 1984. "Influence of the *N*-linked oligosaccharides on the biosynthesis, intracellular routing, and function of the human asialoglycoprotein receptor." J. Biol. Chem. 259: 10414-10421.

Burlingame, A. L. 1994. Simultaneous Detection in Sector Mass Spectrometry. In: T. Matsuo, Y. Seyama, R. M. Caprioli, M. L. Gross, eds. Biological Mass Spectrometry- Present and Future. Chichester: John Wiley & Sons. pp. 147-164.

Buxser, S. E., D. J. Kelleher, L. Watson, P. Puma and G. L. Johnson 1983. "Change in state of nerve growth factor receptor." J. Biol. Chem. 258: 3741-3749.

Caprioli, R. M., B. DaGue, T. Fan and W. T. Moore 1987. "Microbore HPLC/mass spectrometry for the analysis of peptide mixtures using a continuous flow interface." Biochem. Biophys. Res. Commun. 146: 291-295.

Caprioli, R. M., T. Fan and J. S. Cottrell 1986. "Continuous-flow sample probe for fast atom bombardment mass spectrometry." Anal. Chem. **58**: 2949-2952.

Caprioli, R. M. and W. T. Moore 1990. Continuous-Flow Fast Atom Bombardment Mass Spectrometry. In: J. A. McCloskey, ed. Methods Enzymology **193**. San Diego: Academic Press. pp. 214-237.

Carr, S. A., M. E. Hemling, M. F. Bean and G. D. Roberts 1991. "Integration of mass spectrometry in analytical biotechnology." Anal. Chem. **63**: 2802-2824.

Carr, S. A., M. E. Hemling, W. G. Folena, R. W. Sweet, K. Anumula, J. R. Barr, M. J. Huddleston and P. Taylor 1989. "Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry." J. Biol. Chem. **264**: 21286-21295.

Carr, S. A., M. J. Huddleston and M. F. Bean 1993. "Selective identification and differentiation of *N*- and *O*-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry." Protein Sci. **2**: 183-196.

Carr, S. A. and V. N. Reinhold 1984. "Structural characterization of glycosphingolipids by direct chemical ionization mass spectrometry." Biomed. Mass Spectrom. **11**: 633-642.

Carr, S. A., V. N. Reinhold, B. N. Green and J. R. Hass 1985. "Enhancement of structural information in FAB ionized carbohydrate samples by neutral gas collision." Biomed. Mass Spectrom. **12**: 288-295.

Carr, S. A. and G. D. Roberts 1986. "Carbohydrate Mapping by Mass Spectrometry: A Novel Method for Identifying Attachment Sites of Asn-Linked Sugars in Glycoproteins." Anal. Biochem. **157**: 396-406.

Cebon, J., N. Nicola, M. Ward, I. Gardner, P. Dempsey, J. Layton, U. Dührsen, A. W. Burgess, E. Nice and G. Morstyn 1990. "Granulocyte-macrophage colony stimulating factor from human lymphocytes. The effect of glycosylation on receptor binding and biological activity." J. Biol. Chem. **265**: 4483-4491.

Chait, B. T. and S. B. H. Kent 1992. "Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins." Science **257**: 1885-1894.

Chao, M. V. 1992. "Neurotrophin receptors: a window into neuronal differentiation." Neuron **9**: 583-593.

Chao, M. V., M. A. Bothwell, A. H. Ross, H. Koprowski, A. A. Lanahan, C. R. Buck and A. Sehgal 1986. "Gene transfer and molecular cloning of the human NGF receptor." Science **232**: 518-521.

Chapman, B. S., S. E. Kaufman, M. R. Eckart and G. R. Lapointe 1994. "O-linked oligosaccharide and NGF-binding affinity of recombinant 75 kDa neurotrophin receptor." J. Biol. Chem., *submitted*.

Chizhov, O. S. and N. K. Kochetkov 1972. Methods Carbohydr. Chem. **6**: 540-590.

Chong, K. S., M. Jahani, S. Hara and A. G. Lacko 1983. "Characterization of lecithin-cholesterol acyltransferase from human plasma. 3. Chemical properties of the enzyme." Can. J. Biochem. and Cell Biol. 61: 875-881.

Chung, J., D. A. Abano, G. M. Fless and A. M. Scanu 1979. "Isolation, properties, and mechanism of in vitro action of lecithin:cholesterol acyltransferase from human plasma." J. Biol. Chem. 254: 7456-7464.

Ciucanu, I. and F. Kerek 1984. "A simple and rapid method for permethylation of carbohydrates." Carbohydr. Res. 131: 209-217.

Clogston, C. L., S. Hu, T. C. Boon and H. S. Lu 1993. "Glycosidase digestion, electrophoresis and chromatographic analysis of recombinant human granulocyte colony-stimulating factor glycoforms produced in Chinese hamster ovary cells." J. Chromatogr. 637: 55-62.

Collet, X. and C. J. Fielding 1991. "Effects of inhibitors of *N*-linked oligosaccharide processing on the secretion, stability, and activity of lecithin:cholesterol acyltransferase." Biochemistry 30: 3228-3234.

Conboy, J. J. and J. D. Henion 1992. "The determination of glycopeptides by liquid chromatography/mass spectrometry with collision-induced dissociation." J. Am. Soc. Mass Spectrom. 3: 804-814.

Conradt, H. S., R. Geyer, J. Hoppe, L. Grotjahn, H. Plessing and H. Mohr 1985. "Structures of the major carbohydrates of natural human interleukin-2." Eur. J. Biochem. 153: 255-261.

Conradt, H. S., M. Nimtz, K. E. Dittmar, W. Lindenmaier, J. Hoppe and H. Hauser 1989. "Expression of human interleukin-2 in recombinant baby hamster kidney, Ltk-, and Chinese hamster ovary cells. Structure of *O*-linked carbohydrate chains and their location within the polypeptide." J. Biol. Chem. 264: 17368-17373.

Corfield, A. P., M. Wember, R. Schauer and R. Rott 1982. "The specificity of viral sialidases. The use of oligosaccharide substrates to probe enzymic characteristics and strain-specific differences." Eur. J. Biochem. 124: 521-525.

Cottrell, J. S. and S. Evans 1987. "Characteristics of a multichannel electrooptical detection system and its application to the analysis of large molecules by fast atom bombardment mass spectrometry." Anal. Chem. 59: 1990-1995.

Coussens, L., C. Van Beveren, D. Smith, E. Chen, R. L. Mitchell, C. M. Isacke, I. M. Verma and A. Ullrich 1986. "Structural alteration of viral homologue of receptor proto-oncogene fms at the carboxyl terminus." Nature 320: 277-280.

Crestfield, A. M., S. Moore and W. H. Stein 1963. "The preparation and enzymatic hydrolysis of reduced and s-carboxymethylated proteins." J. Biol. Chem. 238: 622-627.

Cumming, D. A. 1991. "Glycosylation of recombinant protein therapeutics: control and functional implications." Glycobiology 1: 115-130.

Cumming, D. A., C. G. Hellerqvist, M. Harris-Brandts, S. W. Michnick, J. P. Carver and B. Bendiak 1989. "Structures of asparagine-linked oligosaccharides of the

glycoprotein fetuin having sialic acid linked to N-acetylglucosamine." Biochemistry 28: 6500-6512.

Currie, G. J. and J. R. Yates III 1993. "Analysis of oligodeoxynucleotides by negative-ion matrix-assisted laser desorption mass spectrometry." J. Am. Soc. Mass Spectrom. 4: 955-963.

Dahms, N. M., P. Lobel and S. Kornfeld 1989. "Mannose 6-phosphate receptors and lysosomal enzyme targeting." J. Biol. Chem. 264: 12115-12118.

Daniel, T. O., D. F. Milfay, J. Escobedo and L. T. Williams 1987. "Biosynthetic and glycosylation studies of cell surface platelet-derived growth factor receptors." J. Biol. Chem. 262: 9778-9784.

Davies, A. M., K. F. Lee and R. Jaenisch 1993. "p75-deficient trigeminal sensory neurons have an altered response to NGF but not to other neurotrophins." Neuron 11: 565-574.

DeJongh, D. C. and K. Biemann 1963. "Application of mass spectrometry to structure problems. XIV. Acetates of partially methylated pentoses and hexoses." J. Am. Chem. Soc. 85: 2289-2294.

Dell, A. 1987. "F.A.B.-mass spectrometry of carbohydrates." Adv. Carbohydr. Chem. Biochem. 45: 19-72.

Dell, A. 1990. Preparation and desorption mass spectrometry of permethyl and peracetyl derivatives of oligosaccharides. In: ed. Mass Spectrometry. San Diego: Academic Press. pp. 647-660.

Dell, A., K.-H. Khoo, M. Panico, R. A. McDowell, A. T. Etienne, A. J. Reason and H. Morris 1994. FAB-MS and ES-MS of glycoproteins. In: ed. Glycobiology: A Practical Approach. Oxford: IRL Press. pp. 187-222.

Dell, A. and A. Kobata 1986. "Novel FAB-MS procedures for glycoprotein analysis." Biochem. Biophys. Res. Commun. 135: 1126-1132.

Dell, A. and J. E. Thomas-Oates 1989. Fast Atom Bombardment-Mass Spectrometry (FAB-MS): Sample Preparation and Analytical Strategies. In: ed. Analysis of Carbohydrates by GLC and MS. Boca Raton, FL: CRC Press. pp. 217-235.

Delorme, E., T. Lorenzini, J. Griffin, F. Martin, F. Jacobsen, T. Boone and S. Elliott 1992. "Role of glycosylation on the secretion and biological activity of erythropoietin." Biochemistry 31: 9871-9876.

DiStefano, P. S., D. M. Chelsea, C. M. Schick and J. F. Mckelvy 1993. "Involvement of a metalloprotease in low-affinity nerve growth factor receptor truncation - inhibition of truncation *in vitro* and *in vivo*." J. Neurosci. 13: 2405-2414.

Distler, J. J. and G. W. Jourdian 1973. "The purification and properties of beta-galactosidase from bovine testes." J. Biol. Chem. 248: 6772-6780.

Domon, B. and C. E. Costello 1988. "Carbohydrate nomenclature." Glycoconjugate J. 5: 397-409.

Downs, F., A. Herp, J. Moschera and W. Pigman 1973. "Beta-elimination and reduction reactions and some applications of dimethylsulfoxide on submaxillary glycoproteins." Biochim. Biophys. Acta 328: 182-192.

Duan, D. S., M. J. Pazin, L. J. Fretto and L. T. Williams 1991. "A functional soluble extracellular region of the platelet-derived growth factor (PDGF) beta-receptor antagonizes PDGF-stimulated responses." J. Biol. Chem. 266: 413-418.

Duffin, K. L., J. K. Welply, E. Huang and J. D. Henion 1992. "Characterization of N-linked Oligosaccharides by Electrospray and Tandem Mass Spectrometry." Anal. Chem. 64: 1440-1448.

Dürkop, H., U. Latza, M. Hummel, F. Eitelbach, B. Seed and H. Stein 1992. "Molecular cloning and expression of a new member of the nerve growth factor receptor family that is characteristic for Hodgkin's disease." Cell 68: 421-427.

Edge, A. S. B. and R. G. Spiro 1987. "Presence of an O-glycosidically linked hexasaccharide in fetuin." J. Biol. Chem. 262: 16135-16141.

Egge, H., J. C. Michalski and G. Strecker 1982. "Heterogeneity of urinary oligosaccharides from mannosidosis: mass spectrometric analysis of permethylated Man₉, Man₈, and Man₇ derivatives." Arch. Biochem. Biophys. 213: 318-326.

Egge, H. and J. Peter-Katalinic 1987. "Fast atom bombardment mass spectrometry for structural elucidation of glycoconjugates." Mass Spectrom. Rev. 6: 331-393.

Endo, Y. and A. Kobata 1976. "Partial purification and characterization of an endo- α -N-acetylgalactosaminidase from the culture of medium of *Diplococcus pneumoniae*." J. Biochem. (Tokyo) 80: 1-8.

Falick, A. M. 1994. "Mass spectrometric techniques." In: Applications of Mass Spectrometry to Organic Stereochemistry. New York: VCH Publishers. pp. 17-29.

Falick, A. M., G. H. Wang and F. C. Walls 1986. "Ion source for liquid matrix secondary ionization mass spectrometry." Anal. Chem 58: 1308-1311.

Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse 1989. "Electrospray ionization for mass spectrometry of large biomolecules." Science 246: 64-71.

Fitzgerald, M. C., G. R. Parr and L. M. Smith 1993. "Basic matrices for the matrix-assisted laser desorption/ionization mass spectrometry of proteins and oligonucleotides." Anal. Chem. 65: 3204-3211.

Foxall, C., S. R. Watson, D. Dowbenko, C. Fennie, L. A. Lasky, M. Kiso, A. Hasegawa, D. Asa and B. K. Brandkley 1992. "The three members of the selectin receptor family recognize a common carbohydrate epitope sialyl Lewis^x oligosaccharide." J. Cell Biol. 117: 895-902.

Francone, O. L., L. Evangelista and C. J. Fielding 1993. "Lecithin-cholesterol acyltransferase: effects of mutagenesis at N-linked oligosaccharide attachment sites on acyl acceptor specificity." Biochim. Biophys. Acta 1166: 301-304.

Frazier, W. A., L. F. Boyd and R. A. Bradshaw 1974. "Properties of the specific binding of ^{125}I -nerve growth factor to responsive peripheral neurons." J. Biol. Chem. 249: 5513-5519.

Fukuda, M., A. Dell and M. N. Fukuda 1984. "Structure of foetal lactosaminoglycan, the carbohydrate moiety of band 3 isolated from human umbilical cord erythrocytes." J. Biol. Chem. 259: 4782-4791.

Games, D. F. 1984. In Morris H. R., ed. : Soft ionization biological mass spectrometry. London: Heyden. pp. 54-66.

Gellerfors, P., K. Axelsson, A. Helander, S. Jahahsson, L. Kenne, S. Lindqvist, B. Pavlu, A. Skottner and L. Fryklund 1989. "Isolation and characterization of a glycosylated form of human insulin-like growth factor I produced in *Saccharomyces cerevisiae*." J. Biol. Chem. 264: 11444-11449.

George-Nascimento, C., A. Gyenes, S. M. Halloran, J. Merryweather, P. Valenzuela, K. S. Steimer, F. R. Masiarz and A. Randolph 1988. "Characterization of recombinant human epidermal growth factor produced in yeast." Biochemistry 27: 797-802.

Gerken, T. A., K. J. Butenhof and R. L. Shogren 1989. "Effects of glycosylation on the conformation and dynamics of *O*-linked glycoproteins: carbon-13 NMR studies of ovine submaxillary mucin." Biochemistry 28: 5536-5543.

Gibson, B. W. and K. Biemann 1984. "Strategy for the mass spectrometric verification and correction of the primary structures of proteins deduced from their DNA sequences." Proc. Natl. Acad. Sci. USA 81: 1956-1960.

Gillece-Castro, B. L. and A. L. Burlingame 1990. Oligosaccharide characterization with high-energy collision-induced dissociation mass spectrometry. In: ed. Mass Spectrometry. San Diego: Academic Press. pp. 689-712.

Gillece-Castro, B. L., A. Prakobphol, A. L. Burlingame, H. Leffler and S. J. Fisher 1991. "Structure and bacterial receptor activity of a human salivary proline-rich glycoprotein." J. Biol. Chem. 266: 17358-17368.

Goldstein, I. J., G. W. Hay, B. A. Lewis and F. Smith 1965. Methods Carbohydr. Chem. 5: 361-363.

Gordon, M. M., C. Hu, H. Chokshi, J. E. Hewitt and D. H. Alpers 1991. "Glycosylation is not required for ligand or receptor binding by expressed rat intrinsic factor." Am. J. Physiol. Gastrointest. Liver Physiol. 260: G736-G742.

Grant, A. J., E. Roessler, G. Ju, M. Tsudo, K. Sugamura and T. A. Waldmann 1992. "The interleukin 2 receptor (IL-2R): the IL-2R α subunit alters the function of the IL-2R β subunit to enhance IL-2 binding and signaling by mechanisms that do not require binding of IL-2 to IL-2R α subunit." Proc. Natl. Acad. Sci. USA 89: 2165-2169.

Gravestien, L. A., B. Blom, L. A. Nolten, E. de Vries, G. van der Horst, F. Ossendorp, J. Borst and W. A. Loenen 1993. "Cloning and expression of murine CD27: comparison with 4-1BB, another lymphocyte-specific member of the nerve growth factor receptor family." Eur. J. Immunol. 23: 943-950.

Gray, G. R. 1990. Linkage analysis using reductive cleavage method. In: ed. Mass Spectrometry. San Diego: Academic Press. pp. 573-587.

Green, B. N. and R. W. A. Oliver 1991. "The study of intact proteins and glycoproteins by electrospray m. s." Biochem. Soc. Trans. 19: 929-935.

Green, E. D. and J. U. Baenziger 1988. "Asparagine-linked oligosaccharides on lutropin, follitropin, and thyrotropin. I. Structural elucidation of the sulfated and sialylated oligosaccharides on bovine, ovine, and human pituitary glycoprotein hormones." J. Biol. Chem. 263: 25-35.

Grob, P. M. and M. A. Bothwell 1983. "Modification of nerve growth factor receptor properties by wheat germ agglutinin." J. Biol. Chem. 258: 14136-14143.

Grönberg, G., P. Lipniunas, T. Lundgren, K. Erlansson, F. Lindh and B. Nilsson 1989. "Isolation of monosialylated oligosaccharides from human milk and structural analysis of three new compounds." Carbohydr. Res. 191: 261-278

Hakamori, J. 1964. "A rapid permethylation of glycolipid and polysaccharide catalyzed by methylsulfinyl carbanion in dimethylsulfoxide." J. Biochem. (Tokyo) 55: 205-208.

Hall, S. C., D. M. Smith, F. R. Masiarz, V. W. Soo, H. M. Tran, L. B. Epstein and A. L. Burlingame 1993. "Mass spectrometric and Edman sequencing of lipocortin I isolated by two-dimensional SDS/PAGE of human melanoma lysates." Proc. Natl. Acad. Sci. USA 90: 1927-1931.

Hannink, M. and D. J. Donoghue 1989. "Structure and function of platelet-derived growth factor (PDGF) and related proteins." Biochim. et Biophys. Acta 989: 1-10.

Hansson, G. C. and H. Karlsson 1990. High-mass gas chromatography-mass spectrometry of permethylated oligosaccharides. In: ed. Mass Spectrometry. San Diego: Academic Press. pp. 733-738.

Hansson, G. C., Y.-T. Li and H. Karlsson 1989. "Characterization of glycosphingolipid mixtures with up to ten sugars by gas chromatography and gas chromatography-mass spectrometry as permethylated oligosaccharides and ceramides released by ceramide glycanase." Biochemistry 28: 6672-6678.

Harbon, S., G. Herman and S. Clauser 1968. "Quantitative evaluation of *O*-glycosidic linkages between sugars and amino acids in ovine submaxillary gland mucoprotein." Eur. J. Biochem. 4: 265-272.

Hardy, M. R. and R. R. Townsend 1988. "Separation of positional isomers of oligosaccharides and glycopeptides by high-performance anion-exchange chromatography with pulsed amperometric detection." Proc. Natl. Acad. Sci. USA 85: 3289-3293.

Hardy, M. R., R. R. Townsend and Y. C. Lee 1988. "Monosaccharide analysis of glycoconjugates by anion exchange chromatography with pulsed amperometric detection." Anal. Biochem. 170: 54-62.

Harris, R. J., H. Van Halbeek, J. Glushka, L. Basa, V. T. Ling, K. J. Smith and M. W. Spellman 1993. "Identification and structural analysis of the tetrasaccharide

NeuAc α (2 \rightarrow 6)Gal β (1 \rightarrow 4)GlcNAc β (1 \rightarrow 3)Fuc α 1-O-linked to serine 61 of human factor IX." Biochemistry 32: 6539-6547.

Hart, C. E., M. Bailey, D. A. Curtis, S. Osborn, E. Raines, R. Ross and J. W. Forstrom 1990. "Purification of PDGF-AB and PDGF-BB from human platelet extracts and identification of all three PDGF dimers in human platelets." Biochemistry 29: 166-172.

Hase, S., I. Ibuki and T. Ikenaka 1984. "Reexamination O-pyridylamination used for fluorescence labeling of oligosaccharides and its applications to glycoproteins." J. Biochemistry 95: 197-203.

Hayes, R. N. and M. L. Gross 1990. Collision-induced dissociation. In: ed. Mass Spectrometry. San Diego: Academic Press. pp. 237-263.

Heldin, C. H., A. Emlund, C. Rorsman and L. Ronnstrand 1989. "Dimerization of B-type platelet-derived growth factor receptors occurs after ligand binding and is closely associated with receptor kinase activation." J. Biol. Chem. 264: 8905-8912.

Hellerqvist, C. 1990. "Linkage analysis using Lindberg method". In: J. A. McCloskey, ed. Methods in Enzymology, 193: San Diego: Academic Press. pp. 554-573.

Hellerqvist, C. G., B. Lindberg, S. Svensson, T. Holme and A. A. Lindberg 1968. Carbohydr. Res. 8: 43-47.

Hellerqvist, C. G. and B. J. Sweetman 1990. Mass spectrometry of carbohydrates. In: ed. Methods Biochem. Anal. New York: Wiley. pp. 91-143.

Hemling, M. E., G. D. Roberts, W. Johnson, S. A. Carr and T. R. Covey 1990. "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." Biomed. Environ. Mass Spectrom. 19: 677-691.

Hempstead, B. L., D. Martin-Zanca, D. R. Kaplan, L. F. Parada and M. V. Chao 1991. "High-affinity NGF binding requires coexpression of the *trk* proto-oncogene and the low-affinity NGF receptor." Nature 350: 678-683.

Hempstead, B. L., S. J. Rabin, L. Kaplan, S. Reid, L. F. Parada and D. R. Kaplan 1992. "Overexpression of the *trk* tyrosine kinase rapidly accelerates nerve growth factor-induced differentiation." Neuron 9: 883-896.

Hernandez, L. M., L. Ballou, E. Alvarado, B. L. Gillece-Castro, A. L. Burlingame and C. E. Ballou 1989. "A new *saccharomyces cerevisiae mnn* mutant *N*-linked oligosaccharide structure." J. Biol. Chem. 264: 11849-11856.

Hillenkamp, F., M. Karas, R. C. Beavis and B. T. Chait 1991. "Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers." Anal. Chem. 63: 1193A-1203A.

Hofmann, R., A. Joseph, M. M. Bhargava, E. M. Rosen and I. Goldberg 1992. "Scatter factor is a glycoprotein but glycosylation is not required for its activity." Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. 1120: 343-350.

Hohmann, H. P., M. Brockhaus, P. A. Baeuerle, R. Remy, R. Kolbeck and A. P. van Loon 1990. "Expression of the types A and B tumor necrosis factor (TNF) receptors is

independently regulated, and both receptors mediate activation of the transcription factor NF-kappa B. TNF alpha is not needed for induction of a biological effect via TNF receptors." J. Biol. Chem. **265**: 22409-22417.

Holtzman, D. M., Y. ' . Li L. F., S. Kinsman, C.-K. Chen, J. S. Valetta, J. Zhou, J. B. Long and W. C. Mobley 1992. "p140trk mRNA marks NGF-responsive forebrain neurons: evidence that *trk* gene expression is induced by NGF." Neuron **9**: 465-478.

Hortin, G. L. 1990. "Isolation of glycopeptides containing O-linked oligosaccharides by lectin affinity chromatography on jacalin-agarose." Anal. Biochem. **191**: 262-267.

Hortin, G. L. and B. L. Trimpe 1990. "Lectin affinity chromatography of proteins bearing O-linked oligosaccharides: application of jacalin-agarose." Anal. Biochem. **188**: 271-277.

Huberty, M. C., J. E. Vath, W. Yu and S. A. Martin 1993. "Site-specific carbohydrate identification in recombinant proteins using MALD-TOF MS." Anal. Chem. **65**: 2791-2800.

Huddleston, M. J., M. F. Bean and S. A. Carr 1993. "Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests." Anal Chem **65**: 877-884.

Hunkapiller, M. W., R. M. Hewick, W. J. Dreyer and L. E. Hood 1983. "High-sensitivity sequencing with a gas-phase sequenator." Methods Enzymol. **91**: 399-413.

Hunt, D. F., J. R. Y. III, J. Shabanowitz, S. Winston and C. R. Hauer 1986. "Protein sequencing by tandem mass spectrometry." Proc. Natl. Acad. Sci. USA **83**: 6233-6237.

Ii, T., Y. Ohashi, Y. Matsuzaki, T. Ogawa and Y. Nagai 1993. "Electrospray mass spectrometry of pentacosaccharides of blood group I-activity and related compounds." Org. Mass Spectrom. 28: 1340-1344.

Itoh, N., S. Yonehara, A. Ishii, M. Yonehara, S.-I. Mizushima, M. Sameshima, A. Hase, Y. Seto and S. Nagata 1991. "The polypeptide encoded by the cDNA for human cell surface antigen Fas can mediate apoptosis." Cell 66: 233-243.

Jardine, I., F. Matsuura and C. C. Sweeley 1984. "Electron ionization mass spectra of reduced and permethylated urinary oligosaccharides from patients with mannosidosis." Biomed. Mass Spectrom. 11: 562-568.

Jennings, K. R. and G. G. Dolnikowski 1990. Mass analyzers. In: ed. Mass Spectrometry. San Diego: Academic Press. pp. 37-61.

Jentoft, N. 1990. "Why are proteins O-glycosylated?" Trends Biochem. Sci. 15: 291-294.

Johnson, D., A. Lanahan, C. R. Buck, A. Sehgal, C. Morgan, E. Mercer, M. Bothwell and M. Chao 1986. "Expression and structure of the human NGF receptor." Cell 47: 545-554.

Johnsson, A., C. H. Heldin, A. Wasteson, B. Westermark, T. F. Deuel, J. S. Huang, P. H. Seeburg, A. Gray, A. Ullrich, G. Scrace, P. Sroobant and M. D. Waterfield 1984. "The *c-sis* gene encodes a precursor of the B chain of platelet-derived growth factor." EMBO J. 3: 921-928.

Juhasz, P., C. Costello and K. Biemann 1993. "Matrix-assisted laser desorption ionization mass spectrometry with 2-(4-hydroxyphenylazo)benzoic acid matrix." J. Am. Soc. Mass Spec. **4**: 399-409.

Kaplan, D. R., B. L. Hempstead, D. Martin-Zanca, M. V. Chao and L. F. Parada 1991. "The *trk* proto-oncogene product - a signal transducing receptor for nerve growth factor." Science **252**: 554-558.

Karas, M., D. Bachmann, U. Bahr and F. Hillenkamp 1987. Int. J. Mass Spectrom. Ion Proc. **78**: 53-57.

Karas, M., D. Bachmann and F. Hillenkamp 1985. "Influence of wavelength in high-irradiance UV laser desorption mass spectrometry of organic molecules." Anal. Chem. **57**: 2935-2939.

Karas, M. and F. Hillenkamp 1988. "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." Anal. Chem. **60**: 2299-2301.

Karlsson, H., I. Carlstedt and G. C. Hansson 1987. "Rapid characterization of mucin oligosaccharides from rat small intestine with gas chromatography-mass spectrometry." FEBS Lett. **226**: 23-27.

Karlsson, K.-A. 1973. "Carbohydrate composition and sequence analysis of cell surface components by mass spectrometry. Characterization of the major monosialoganglioside of brain." FEBS Lett. **32**: 317-320.

Karlsson, K.-A. 1978. "Mass-spectrometric sequence studies of lipid-linked oligosaccharides: blood-group fucolipids, gangliosides and related cell-surface receptors." Prog. Chem. of Fats and Other Lipids 16: 207-230.

Kaufman, S. E., S. Brown and G. B. Stauber 1993. "Characterization of ligand binding to immobilized biotinylated extracellular domains of three growth factor receptors." Anal. Biochem. 211: 261-266.

Kaur, S., W. Liang, R. R. Townsend, R. Trimble and A. L. Burlingame (1990). Proceedings of the 38th ASMS Conference on Mass Spectrometry and Allied Topics. 38th ASMS Conference on Mass Spectrometry and Allied Topics, Tucson, Arizona, pp. 1361-1362.

Kaur, S., R. R. Townsend, W. Liang, R. B. Trimble and A. L. Burlingame 1994. "Structural investigation of isomerically homogeneous *N*-linked oligosaccharides from glycoproteins by liquid secondary ion and tandem mass spectrometry." Anal. Biochem., *in prep.*

Kelker, H. C., Y. K. Yip, P. Anderson and J. Vilcek 1983. "Effects of glycosidase treatment on the physiochemical properties and biological activity of human interferon-gamma." J. Biol. Chem. 258: 8010-8013.

Klein, R., S. Q. Jing, V. Nanduri, E. O'Rourke and M. Barbacid 1991. "The *trk* proto-oncogene encodes a receptor for nerve growth factor." Cell 65: 189-197.

Knox, K. A. and J. Gordon 1993. "Protein tyrosine phosphorylation is mandatory for CD40-mediated rescue of germinal center B cells from apoptosis." Eur. J. Immunol. **23**: 2578-2584.

Kobata, A. 1979. "Use of endo- and exoglycosidases for structural studies of glycoconjugates." Anal. Biochem. **100**: 1-14.

Kochetkov, N. K. and O. S. Chizov 1965. "Mass spectrometry of methylated methylglycosides." Tetrahedron **21**: 2029-2047.

Kornfeld, R. and S. Kornfeld 1976. "Comparative aspects of glycoprotein structure." Ann. Rev. Biochem. **45**: 217-237.

Kwon, B. S. and S. M. Weissman 1989. "cDNA sequences of two inducible T-cell genes." Proc. Natl. Acad. Sci., USA **86**: 1963-1967.

Lace, D., A. H. Olavesen and P. Gacesa 1990. "The effects of deglycosylation on the properties of native and biotinylated bovine testicular hyaluronidase." Carbohydr. Res. **208**: 306-311.

Laemmli, U. K. 1970. "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**: 680-685.

Laine, R. A. 1981. "Enhancement of detection from partially methylated alditol acetates by chemical ionization mass spectrometry." Anal. Biochem. **116**: 383-388.

Li, Y. T. and S. C. Li 1973. " α -mannosidase, β -N-acetylhexosaminidase, and β -galactosidase from jack bean meal." Methods Enzymol. 28: 699-713.

Ligon, W. 1990. Evaluating the composition of liquid surfaces using mass spectrometry. In: A. L. Burlingame, ed. Biological Mass Spectrometry. Amsterdam: Elsevier. pp. 61-73.

Ling, V., A. W. Guzzetta, E. Canova-Davis, J. T. Stults, W. S. Hancock, T. Covey and B. Shushan 1991. "Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry." Anal. Chem. 63: 2909-2915.

Linscheid, M., J. D'Angona, A. L. Burlingame, A. Dell and C. E. Ballou 1981. "Field desorption mass spectrometry of oligosaccharides." Proc. Natl. Acad. Sci., USA 78: 1471-1475.

Loenen, W. A., E. De Vries, L. A. Gravestien, R. Q. Hintzen, A. Van Lier and J. Borst 1992. "The CD27 membrane receptor, a lymphocyte-specific member of the nerve growth factor receptor family, gives rise to a soluble form by protein processing that does not involve receptor endocytosis." Eur. J. Immunol. 22: 447-455.

Lonngren, J. and S. Svennson 1974. "Mass spectrometry in structural analysis of natural carbohydrates." Adv. Carbohydr. Chem. Biochem. 29: 41-106.

Mahadeo, D., L. Kaplan, M. V. Chao and B. L. Hempstead 1994. "High affinity nerve growth factor binding displays a faster rate of association than p140trk binding: Implications for multi-subunit polypeptide receptors." J. Biol. Chem. 269: 6884-6891.

Maley, F., R. B. Trimble, A. L. Tarentino and T. H. Plummer 1989. "Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases." Anal. Biochem. **180**: 195-204.

Mallett, S., S. Fossum and A. N. Barclay 1990. "Characterization of the MRC OX40 antigen of activated CD4 positive T lymphocytes--a molecule related to nerve growth factor receptor." EMBO J. **9**: 1063-1068.

Mann, M., C. K. Meng and J. B. Fenn 1989. "Interpreting mass spectra of multiply charged ions." Anal. Chem. **61**: 1702-1708.

Mariuzza, R. A., S. E. Phillips and R. J. Poljak 1987. "The structural basis of antigen-antibody recognition." Annu. Rev. Biophys. **16**: 139-159.

Matsui, T., M. Heidaran, T. Miki, N. Popescu, W. La Rochelle, M. Kraus, J. Pierce and S. Aaronson 1989. "Isolation of a novel receptor cDNA establishes the existence of two PDGF receptor genes." Science **243**: 800-804.

McCleary, B. V. 1983. " β -D-Mannosidase from *Helix pomatia*." Carbohydr. Res. **111**: 297-310.

McFadden, W. H., H. L. Schwartz and A. Evans 1976. "Direct analysis of liquid chromatographic effluents." J. Chromatogr. **122**: 389-392.

McLafferty, F. W. (1983). Tandem mass spectrometry. New York, Wiley.

McLean, J., C. Fielding, D. Drayna, H. Dieplinger, B. Baer, W. Kohr, W. Henzel and R. Lawn 1986. "Cloning and expression of human lecithin-cholesterol acyltransferase cDNA." Proc. Natl. Acad. Sci. USA 83: 2335-2339.

Meakin, S. O., U. Suter, C. C. Drinkwater, A. A. Welcher and E. M. Shooter 1992. "The rat *trk* protooncogene product exhibits properties characteristic of the slow nerve growth factor receptor." Proc. Natl. Acad. Sci. USA 89: 2374-2378.

Medzihradzky, K. F., B. L. Gillece-Castro, C. A. Settineri, R. R. Townsend, F. R. Masiarz and A. L. Burlingame 1990. "Structure determination of *O*-linked glycopeptides by tandem mass spectrometry." Biomed. Environ. Mass Spectrom. 19: 777-781.

Medzihradzky, K. F., D. A. Maltby, S. C. Hall, C. A. Settineri and A. L. Burlingame 1994. "Characterization of protein *N*-glycosylation by reversed-phase microbore liquid chromatography/electrospray mass spectrometry, complementary mobile phases and sequential exoglycosidase digestion." J. Am. Soc. Mass Spectrom. 5: 350-358.

Medzihradzky, K. F., C. A. Settineri, F. R. Masiarz, D. A. Maltby and A. L. Burlingame 1993. Post-translational modifications of recombinant proteins determined by LC/electrospray mass spectrometry and high performance tandem mass spectrometry. In: ed. Techniques in Protein Chemistry IV. San Diego, CA: Academic Press. pp. 117-126.

Medzihradzky, K. F., R. R. Townsend and A. L. Burlingame 1994. "Analysis of *O*-glycosylation by tandem mass spectrometry." Biol. Mass Spectrom., *in prep.*

Mock, K. K., M. Davey, M. P. Stevenson and J. S. Cottrell 1991. "The integration of mass spectrometry into the biochemistry laboratory." Biochem. Soc. Trans. 19: 948-953.

Montreuil, J. 1984. "Spatial conformation of glycans and glycoproteins." Biol. of Cell **51**: 115-131.

Morrison, D. K., D. R. Kaplan, J. A. Escobedo, U. R. Rapp, T. M. Roberts and L. T. Williams 1989. "Direct activation of the serine/threonine kinase activity of Raf-1 through tyrosine phosphorylation by the PDGF beta-receptor." Cell **58**: 649-657.

Morrison, D. K., D. R. Kaplan, S. G. Rhee and L. T. Williams 1990. "Platelet-derived growth factor (PDGF)-dependent association of phospholipase C-gamma with the PDGF receptor signaling complex." Mol. Cell. Biol. **10**: 2359-2366.

Naylor, S., A. F. Findeis, B. W. Gibson and D. H. Williams 1986. "An approach toward the complete fast atom bombardment analysis of enzymic digests of peptides and proteins." J. Am. Chem. Soc. **108**: 6359-6363.

Nilsson, B., N. E. Norden and S. Svensson 1979. "Structural studies on the carbohydrate portion of fetuin." J. Biol. Chem. **254**: 4545-4553.

Nilsson, B. and D. Zopf 1983. "Oligosaccharides released from glycolipids by trifluoroacetolysis can be analyzed by gas chromatography-mass spectrometry." Arch. Biochem. Biophys. **222**: 628-648.

Nimtze, M., W. Martin, V. Wray, K. D. Kloppel, J. Augustin and H. S. Conradt 1993. "Structures of sialylated oligosaccharides of human erythropoietin expressed in recombinant BHK-21 cells." Eur. J. Biochem. **213**: 39-56.

Ogasawara, J., R. Watanabe-Fukunaga, M. Adachi, A. Matsuzawa, T. Kasugai, Y. Kitamura, N. Itoh, T. Suda and S. Nagata 1993. "Lethal effect of the anti-Fas antibody in mice." Nature 364: 806-809.

Oh-eda, M., M. Hasegawa, K. Hattori, H. Kuboniwa, T. Kojima, T. Orita, K. Tomonou, T. Yamazaki and N. Ochi 1990. "O-linked sugar chain of human granulocyte colony-stimulating factor protects it against polymerization and denaturation allowing it to retain its biological activity." J. Biol. Chem. 265: 11432-11435.

Ohashi, Y., D. A. Gage and C. C. Sweeley 1991. In: F. Hommes, ed. Techniques in Diagnostic Human Biochemical Genetics. A Laboratory Manual. New York: John Wiley and sons. pp. 239-244.

Ohmori, K., A. Takada, I. Ohwaki, N. Takahashi, Y. Furukawa, M. Maeda, M. Kiso, A. Hasegawa, M. Kannagi and R. Kannagi 1993. "A distinct type of sialyl Lewis X antigen defined by a novel monoclonal antibody is selectively expressed on helper memory T cells." Blood 82: 2797-2805.

Olson, T. S. and M. C. Lane 1989. "A common mechanism for posttranslational activation of plasma membrane receptors." EASEB J. 3: 1618-1624.

Patel, T., J. Bruce, A. Merry, C. Bigge, M. Wormald, A. Jaques and R. Parekh 1993. "Use of hydrazine to release in intact and unreduced form both *N*- and *O*-linked oligosaccharides from glycoproteins." Biochemistry 32: 679-693.

Paulson, J. C., J. Weinstein, L. Dorland, H. van Halbeek and J. F. Vliedenthart 1982. "Newcastle disease virus contains a linkage-specific glycoprotein sialidase. Application to

the localization of sialic acid residues in N-linked oligosaccharides of alpha 1-acid glycoprotein." J. Biol. Chem. **257**: 12734-12738.

Pennica, D., V. T. Lam, R. Weber, W. J. Kohr, L. J. Basa, M. W. Spellman, A. Ashkenazi, S. J. Shire and D. V. Goeddel 1993. "Biochemical characterization of the extracellular domain of the 75-kilodalton tumor necrosis factor receptor." Biochemistry **32**: 3131-3138.

Peter-Katalinic, J. and H. Egge 1987. Mass Spectrom. Rev. **6**: 331-362.

Phillips, M. L., E. Nudelman, F. C. A. Gaeta, M. Perez, A. K. Singhal, S. Hakomori and J. C. Paulson 1990. "ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand, sialyl Le^x." Science **250**: 1130-1132.

Pieles, U., W. Zurcher, M. Schar and H. E. Mozer 1993. "MALDI-TOF MS: A powerful new tool for mass and sequence analysis of natural and modified oligonucleotides." Nucl. Acid Res. **21**: 3191-3196.

Pisano, A., J. W. Redmond, K. L. Williams and A. A. Gooley 1993. "Glycosylation sites identified by solid-phase Edman degradation: O-linked glycosylation motifs on human glycoporphin A." Glycobiology **3**: 429-435.

Pleasure, S. J., U. R. Reddy, G. Venkatakrishnan, A. K. Roy, J. Chen, A. H. Ross, J. Q. Trojanowski, D. E. Pleasure and V. M. Lee 1990. "Introduction of nerve growth factor (NGF) receptors into a medulloblastoma cell line results in expression of high- and low-affinity NGF receptors but not NGF-mediated differentiation." Proc. Natl. Acad. Sci. USA **87**: 8496-8500.

Plummer, T. H. and A. L. Tarentino 1991. "Purification of the oligosaccharide-cleaving enzymes of *Flavobacterium meningosepticum*." Glycobiology 1: 257-263.

Poulter, L. and A. L. Burlingame 1990. Desorption mass spectrometry of oligosaccharides coupled with hydrophobic chromophores. In: ed. Mass Spectrometry. San Diego: Academic Press. pp. 661-689.

Poulter, L., J. P. Earnest, R. M. Stroud and A. L. Burlingame 1989. "Structure, oligosaccharide structures, and posttranslationally modified sites of the nicotinic acetylcholine receptor." Proc. Natl. Acad. Sci. USA 86: 6645-6649.

Poulter, L., R. Karrer and A. L. Burlingame 1991. "n-Alkyl p-aminobenzoates as derivatizing agents in the isolation, separation, and characterization of submicrogram quantities of oligosaccharides by liquid secondary ion mass spectrometry." Anal. Biochem. 195: 1-13.

Powell, L. M. and R. H. Pain 1992. "Effects of glycosylation on the folding and stability of human, recombinant and cleaved α_1 -antitrypsin." J. Mol. Biol. 224: 241-252.

Radeke, M. J., T. P. Misko, C. Hsu, L. A. Herzenberg and E. M. Shooter 1987. "Gene transfer and molecular cloning of the rat nerve growth factor receptor." Nature 325: 593-597.

Radford, T. and D. C. De Jongh 1980. Carbohydrates. In: ed. Biochemical Applications of Mass Spectrometry. New York: Wiley (Interscience). pp. 255-310.

Raschdorf, F., R. Dahinden, W. Maerki, W. J. Richter and J. P. Merryweather 1988. "Location of disulphide bonds in human insulin-like growth factors (IGFs) synthesized by recombinant DNA technology." Biomed. Environ. Mass Spectrom. 16: 3-8.

Reinhold, V. N. and S. A. Carr 1983. Mass Spectrom. Rev. 2: 153-169.

Riederer, M. A. and A. Hinnen 1991. "Removal of *N*-glycosylation sites of the yeast acid phosphatase severely affects protein folding." J. Bacteriol. 173: 3539-3546.

Rodriguez-Tebar, A., G. Dechant, R. Gotz and Y. A. Barde 1992. "Binding of neurotrophin-3 to its neuronal receptors and interactions with nerve growth factor and brain-derived neurotrophic factor." EMBO J. 11: 917-922.

Roepstorff, P. and J. Fohlman 1984. "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed. Environ. Mass Spectrom. 11: 601.

Rolf, D. and G. R. Gray 1982. "Reductive Cleavage of Glycosides." J. Am. Chem. Soc. 104: 3539-3541.

Ross, R., E. W. Raines and D. F. Bowen-Pope 1986. "The biology of platelet-derived growth factor." Cell 46: 155-169.

Ryhage, R. 1964. "Use of a mass spectrometer as a detector and analyzer for effluents emerging from high temperature gas liquid chromatographic columns." Anal. Chem. 36: 759-764.

Santikarn, S., G.-R. Ritter and V. N. Reinhold 1987. "Oligosaccharide structural studies by on-line HPLC-MS using fast atom bombardment ionization." J. Carbohydr. Chem. **6**: 141-154.

Sasaki, H., N. Ochi, A. Dell and M. Fukuda 1988. "Site-specific glycosylation of human recombinant erythropoietin: analysis of glycopeptides or peptides at each glycosylation site by fast atom bombardment mass spectrometry." Biochemistry **27**: 8618-8626.

Sastry, M. V., P. Banarjee, S. R. Patanjali, M. J. Swamy, G. V. Swarnalatha and A. Surolia 1986. "Analysis of saccharide binding to *Artocarpus integrifolia* lectin reveals specific recognition of T-antigen (β -D-Gal(1 \rightarrow 3)D-GalNAc)." J. Biol. Chem. **261**: 11726-11733.

Sawardeker, J. S., J. H. Sloneker and A. Jeanes 1965. "Quantitative determination of monosaccharides as their alditol acetates by gas liquid chromatography." Anal. Chem. **37**: 1602-1604.

Schechterson, L. C. and M. Bothwell 1992. "Novel roles for neurotrophins are suggested by BDNF and NT-3 mRNA expression in developing neurons." Neuron **9**: 449-463.

Scher, C. D., R. C. Shepard, H. N. Antoniades and C. D. Stiles 1979. "Platelet-derived growth factor and the regulation of the mammalian fibroblast cell cycle." Biochim. Biophys. Acta **560**: 217-241.

Schindler, P. A., C. A. Settineri, X. Collet, C. J. Fielding and A. L. Burlingame 1994. "Site-specific determination of the *N*- and *O*-linked carbohydrates on human plasma

lecithin:cholesterol acyl transferase by HPLC/electrospray mass spectrometry and glycosidase digestion.” Protein Science: *submitted*.

Settineri, C. A. and A. L. Burlingame (1993). Sequencing of glycopeptides using sequential digestion followed by HPLC/electrospray mass spectrometry. 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, 94a-94b.

Settineri, C. A. and A. L. Burlingame 1994. Strategies for the characterization of carbohydrates from glycoproteins by mass spectrometry. In: ed. Techniques in Protein Chemistry V. San Diego: Academic Press. pp. 97-104.

Settineri, C. A., B. Chapman, I. Leung, L. S. Cousens and A. L. Burlingame 1992. Characterization of the glycosylation on recombinant human low-affinity nerve growth factor receptor. In: ed. Techniques in Protein Chemistry III. San Diego: Academic Press. pp. 295-302.

Settineri, C. A., K. F. Medzihradzsky, F. R. Masiarz, A. L. Burlingame, C. Chu and C. George-Nascimento 1990. “Characterization of *O*-glycosylation sites in recombinant B-chain of platelet-derived growth factor expressed in yeast using liquid secondary ion mass spectrometry, tandem mass spectrometry and Edman sequence analysis.” Biomed. Environ. Mass Spectrom. **19**: 665-676.

Shogren, R., T. A. Gerken and N. Jentoft 1989. “Role of glycosylation on the conformation and chain dimensions of O-linked glycoproteins: light-scattering studies of ovine submaxillary mucin.” Biochemistry **28**: 5525-5536.

Sliker, L. J., T. M. Martensen and M. D. Lane 1986. "Synthesis of epidermal growth factor receptor in human A431 cells. Glycosylation-dependent acquisition of ligand binding activity occurs post-translationally in the endoplasmic reticulum." J. Biol. Chem. **261**: 15233-15241.

Smith, C. A., T. Davis, D. Anderson, E. Solam, M. P. Beckmann, R. Jerzy, S. K. Dower, D. Cosman and R. G. Goodwin 1990. "A receptor for tumor necrosis factor defines an unusual family of cellular and viral proteins." Science **248**: 1019-1023.

Smith, E. L., J. M. McKibbin, K.-A. Karlsson, I. Pascher and B. E. Samuelsson 1975. "Characterization by Mass Spectrometry of Blood Group A Active Glycolipids from Human and Dog Small Intestine." Biochemistry **14**: 2120-2124.

Smith, R. D., J. A. Loo, C. G. Edmonds, C. J. Barinaga and H. R. Udseth 1990. "New developments in biochemical mass spectrometry: electrospray ionization." Anal. Chem. **62**: 882-899.

Spellman, M. W. 1990. "Carbohydrate characterization of recombinant glycoproteins of pharmaceutical interest." Anal. Chem. **62**: 1714-1722.

Spiro, R. G. and V. D. Bhojroo 1974. "Structure of the *O*-glycosidically linked carbohydrate units of fetuin." J. Biol. Chem. **249**: 5704-5717.

Stamenkovic, I., E. A. Clark and B. Seed 1989. "A B-lymphocyte activation molecule related to the nerve growth factor receptor and induced by cytokines in carcinomas." EMBO J. **8**: 1403-1410.

Subbaiah, P. V., J. J. Albers, C. H. Chen and J. D. Bagdade 1980. "Low density lipoprotein-activated lysolecithin acylation by human plasma lecithin-cholesterol acyltransferase. Identity of lysolecithin acyltransferase and lecithin-cholesterol acyltransferase." J. Biol. Chem. 255: 9275-9280.

Sutton, C., J. O'Neill and J. Cottrell 1994. "Site-specific characterization of glycoprotein carbohydrates by exoglycosidase digestion and laser desorption mass spectrometry." Anal. Biochem. 218: 34-46.

Sweeley, C. C., R. Bentley, M. Makita and W. W. Wells 1963. "Gas-liquid chromatography of trimethyl silyl derivatives of sugars and related substances." J. Am. Chem. Soc. 85: 2497-2507.

Sweeley, C. C., W. H. Elliott, I. Fries and R. Ryhage 1966. "Mass spectrometric determination of unresolved components in gas chromatographic effluents." Anal. Chem. 38: 1549-1553.

Tai, T., K. Yamashita, M. Ogata-Arakawa, N. Koide and T. Muramatsu 1975. "Structural studies of two ovalbumin glycopeptides in relation to the endo- β -N-acetylglucosaminidase specificity." J. Biol. Chem. 250: 8569-8575.

Takasaki, S., T. Mizuochi and A. Kobata 1982. "Hydrazinolysis of asparagine-linked sugar chains to produce free oligosaccharides." Methods Enzymol. 83: 263-268.

Tanaka, K. and W. Pigman 1965. "Improvement in hydrogenation procedure for demonstration of O-threonine glycosidic linkages in bovine submaxillary mucin." J. Biol. Chem. 240: 1487-1488.

Tanaka, K., H. Waki, Y. Ido, S. Akita, Y. Yoshida and T. Yoshida 1988. "Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry." Rapid Commun. Mass Spectrom. **2**: 151-153.

Tang, X. J., P. Thibault and R. K. Boyd 1993. "Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation." Anal. Chem. **65**: 2824-2834.

Tarentino, A. L., C. M. Gómez and T. H. Plummer 1985. "Deglycosylation of asparagine-linked glycans by peptide:*N*-glycosidase F." Biochemistry **24**: 4665-4671.

Tartaglia, L. A., T. M. Ayres, G. H. W. Wong and D. V. Goeddel 1993. "A novel domain within the 55 kd TNF receptor signals cell death." Cell **74**: 845-853.

Thomson, B. A. and J. V. Iribarne 1979. "Field induced ion evaporation from liquid surfaces at atmospheric pressure." J. Chem. Phys. **71**: 4451-4463.

Tiftt, C. J., R. L. Proia and R. D. Camerini-Otero 1992. "Glycosylation of CD4. Tunicamycin inhibits surface expression." J. Biol. Chem. **263**: 9502-9507.

Tomer, K. B. and C. E. Parker 1989. "Biochemical applications of liquid chromatography-mass spectrometry." J. Chromatogr. **492**: 189-221.

Tomita, M., H. Furthmayr and V. T. Marchesi 1978. "Primary structure of human erythrocyte glycoporphin A. Isolation and characterization of peptides and complete amino acid sequence." Biochemistry **17**: 4756-4770.

Townsend, R. R. and M. R. Hardy 1991. "Analysis of glycoprotein oligosaccharides using high-pH anion exchange chromatography." Glycobiology **1**: 139-147.

Trimble, R. B. and F. Maley 1984. "Optimizing hydrolysis of *N*-linked high-mannose oligosaccharides by endo- β -*N*-acetylglucosaminidase H." Anal. Biochem. **141**: 515-522.

Trimble, R. B. and A. L. Tarentino 1991. "Identification of distinct endoglycosidase (endo) activities in *Flavobacterium meningosepticum*: endo F1, endo F2, and endo F3. Endo F1 and endo H hydrolyze only high mannose and hybrid glycans." J. Biol. Chem. **266**: 1646-1651.

Twining, S. S. 1984. "Fluorescein isothiocyanate-labeled casein assay for proteolytic enzymes." Anal. Biochem. **143**: 30-34.

Uchida, Y., Y. Tsukada and T. Sugimori 1979. "Enzymatic properties of neuraminidases from *Arthrobacter ureafaciens*." J. Biochem. **86**: 1573-1585.

Umemoto, J., V. P. Bhavanandan and E. A. Davidson 1977. "Purification and properties of an endo- α -*N*-acetyl-*D*-galactosaminidase from *Diplococcus pneumoniae*." J. Biol. Chem. **252**: 8609-8614.

Urlaub, G., E. Käs, A. M. Carothers and L. A. Chasin 1983. "Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells." Cell **33**: 405-412.

Utermann, G., H. J. Menzel, G. Adler, P. Dieker and W. Weber 1980. "Substitution in vitro of lecithin-cholesterol acyltransferase. Analysis of changes in plasma lipoproteins." Eur. J. Biochem. 107: 225-241.

Vale, R. D., M. Hosang and E. M. Shooter 1985. "Sialic acid residues on NGF receptors on PC12 cells." Dev. Neurosci. 7: 55-64.

Vale, R. D. and E. M. Shooter 1982. "Alteration of binding properties and cytoskeletal attachment of nerve growth factor receptors in PC 12 cells by wheat germ agglutinin." J. Cell Biol. 94: 710-717.

Vance, D. E., W. Krivit and C. C. Sweeley 1975. "Metabolism of neutral glycosphingolipids in plasma of a normal human and a patient with Fabry's disease." J. Biol. Chem. 250: 8119-8125.

Varki, A. 1993. "Biological roles of oligosaccharides." Glycobiology 3: 97-130.

Vissavajhala, P., J. D. Leszyk, J. Lin-Goerke and A. H. Ross 1992. "Structural domains of the extracellular domain of human nerve growth factor receptor detected by partial proteolysis." Arch. Biochem. Biophys. 294: 244-252.

Wada, Y., A. Nishikawa, N. Okamoto, K. Inui, H. Tsukamoto, S. Okada and N. Taniguchi 1992. "Structure of serum transferrin in carbohydrate-deficient glycoprotein syndrome." Biochem. Biophys. Res. Commun. 189: 832-836.

Wada, Y., J. Tamura, B. D. Musselman, D. B. Kassel, T. Sakurai and T. Matsuo 1992. "Electrospray ionization mass spectra of hemoglobin and transferrin by a magnetic sector

mass spectrometer. Comparison with theoretical isotopic distributions.” Rapid Commun. Mass Spectrom. **6**: 9-13.

Walls, F. C., M. A. Baldwin, A. M. Falick, B. W. Gibson, S. Kaur, D. A. Maltby, B. L. Gillece-Castro, K. F. Medzihradzky, S. Evans and A. L. Burlingame 1990. In: ed. Biological Mass Spectrometry. Amsterdam: Elsevier. pp. 197-216.

Wang, W. T., N. C. LeDonne Jr., B. Ackerman and C. C. Sweeley 1984. “Structural characterization of oligosaccharides by high performance liquid chromatography, fast-atom bombardment mass spectrometry and exoglycosidase digestion.” Anal. Biochem. **141**: 366-381.

Waterfield, M. D., G. T. Scrace, N. Whittle, P. Stroobant, A. Johnsson, A. Wasteson, B. Westermark, C. H. Heldin, J. S. Huang and T. F. Deuel 1983. “Platelet-derived growth factor is structurally related to the putative transforming protein p28sis of simian sarcoma virus.” Nature **304**: 35-39.

Webb, J. W., K. Jiang, B. L. Gillece-Castro, A. L. Tarentino, T. H. Plummer, J. C. Byrd, S. J. Fisher and A. L. Burlingame 1988. “Structural characterization of intact, branched oligosaccharides by high performance liquid chromatography and liquid secondary ion mass spectrometry.” Anal. Biochem. **169**: 337-349.

Weinberger, S. R. (1993). An evaluation of crystallization methods for matrix-assisted laser desorption/ionization of proteins. 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, 775a.

Whitehouse, C. M., R. N. Dreyer, M. Yamashita and J. B. Fenn 1985. "Electrospray interface for liquid chromatographs and mass spectrometers." Anal. Chem. **57**: 675-679.

Williams, L. T. 1989. "Signal transduction by the platelet-derived growth factor receptor." Science **243**: 1564-1570.

Wilson, I. B. H., Y. Gavel and G. V. Heijne 1991. "Amino acid distributions around *O*-linked glycosylation sites." Biochem. J. **275**: 529-534.

Xiang, F. and R. C. Beavis 1994. "Alternative matrices for matrix-assisted laser desorption mass spectrometry." Rapid Comm. Mass Spectrom. **8**: 199-205.

Yamashita, M. and J. B. Fenn 1984. "Electrospray ion source. Another variation on the free-jet theme." J. Phys. Chem. **88**: 4451-4459.

Yan, H. and M. V. Chao 1991. "Disruption of cysteine-rich repeats of the p75 nerve growth factor receptor leads to loss of ligand binding." J. Biol. Chem. **266**: 12099-12104.

Yang, C.-Y., D. Manoogian, Q. Pao, F.-S. Lee, R. D. Knapp, A. M. Gotto and H. J. Pownall 1987. "Lecithin:cholesterol acyltransferase. Functional regions and a structural model of the enzyme." J. Biol. Chem. **262**: 3086-3091.

Yarden, Y., J. A. Escobedo, W. J. Kuang, T. L. Yang-Feng, T. O. Daniel, P. M. Tremble, E. Y. Chen, M. E. Ando, R. N. Harkins, U. Francke, V. A. Fried, A. Ullrich and L. T. Williams 1986. "Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors." Nature **323**: 226-232.

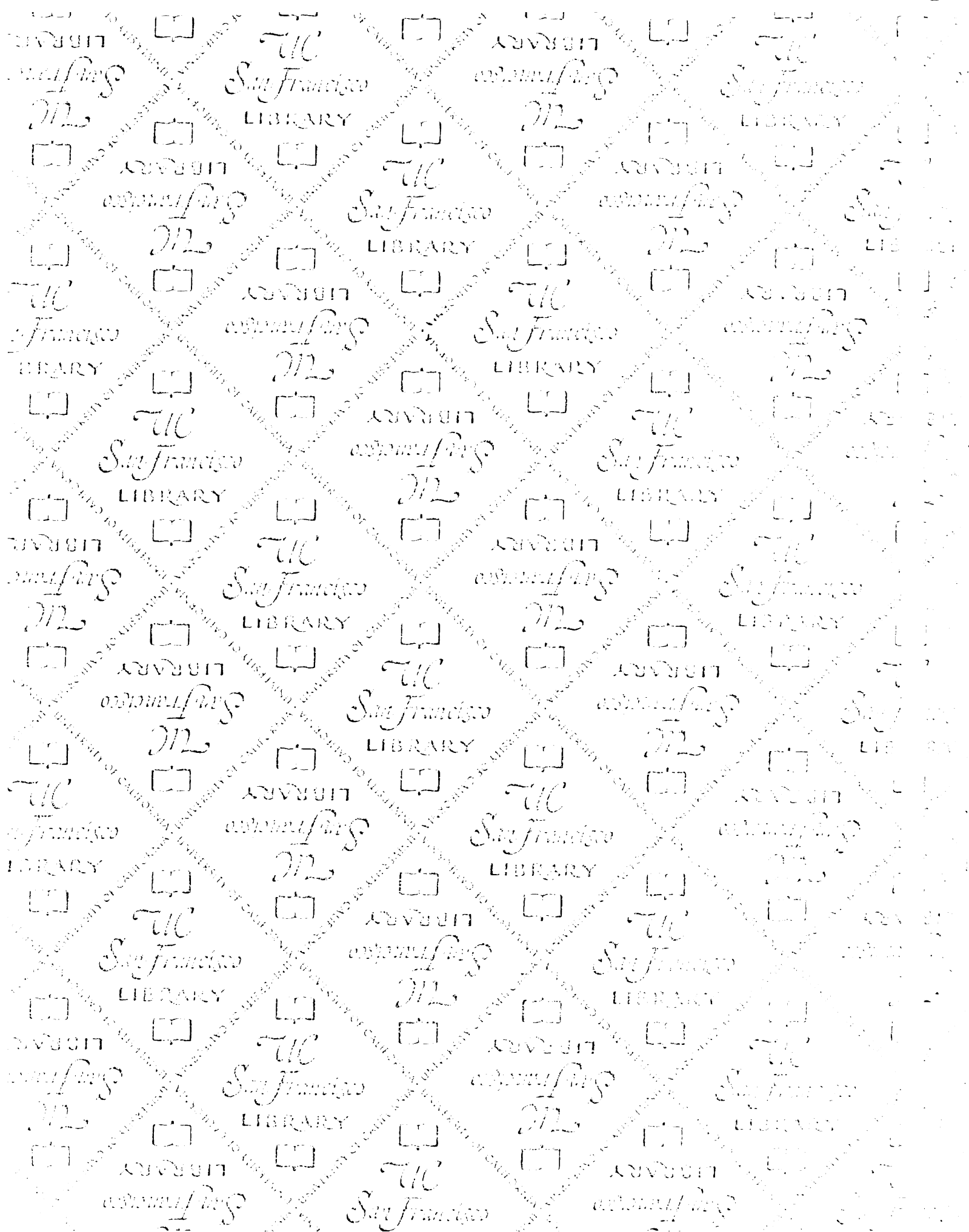
Yarden, Y., W. J. Kuang, T. Yang-Feng, L. Coussens, S. Munemitsu, T. J. Dull, E. Chen, J. Schlessinger, U. Francke and A. Ullrich 1987. "Human proto-oncogene *c-kir*: a new cell surface receptor tyrosine kinase for an unidentified ligand." EMBO J. **6**: 3341-3351.

Young, N. M., R. A. Z. Johnston, A. G. Szabo and D. C. Watson 1989. "Homology of the D-galactose-specific lectins from *Artocarpus integrifolia* and *Maclura pomifera* and the role of an unusual small polypeptide subunit." Arch. Biochem. Biophys. **270**: 596-603.

Yuen, C.-T., A. M. Lawson, W. Chai, M. Larkin, M. S. Stoll, A. C. Stuart, F. X. Sullivan, T. J. Ahern and T. Feizi 1992. "Novel sulfated ligands for the cell adhesion molecule E-selectin revealed by the neoglycolipid technology among O-linked oligosaccharides on an ovarian cystadenoma glycoprotein." Biochemistry **31**: 9126-9131.

Zinn, A. B., J. J. Plantner and D. M. Carlson 1977. "Nature of linkages between protein core and oligosaccharide." In: M.I. Horowitz and W. Pigman, eds. The Glycoconjugates. Vol. I New York: Academic Press. pp. 69-86.

Zupan, A. A., P. A. Osborne, C. E. Smith, N. R. Siegel, R. M. Leimgruber and E. M. J. Johnson 1989. "Identification, purification and characterization of truncated forms of the human nerve growth factor receptor." J. Biol. Chem. **264**: 11714-11720.



For reference

Not to be taken from the room.



