2017

The unique glycoproteins of Cryptosporidium parvum and Toxoplasma gondii

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

THE UNIQUE GLYCOPROTEINS OF *CRYPTOSPORIDIUM PARVUM* AND *TOXOPLASMA GONDII*

by

JOHN ROBERT HASERICK

B.S., University of Connecticut, 2005
M.S., University of Connecticut, 2010

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader _____

Catherine E. Costello, Ph.D.
William Fairfield Warren Distinguished Professor

Second Reader _____

John Samuelson, M.D., Ph.D.
Professor of Molecular and Cell Biology

DEDICATION

I would like to dedicate this work to my late grandfather John Roger Haserick, M.D., for his inspirational work in medical science. To my late mother JoAnn Clark Haserick for her loving support throughout my childhood and always nurturing my mind and advocating education. To my son Joseph Ronen Haserick, you are my motivation.

ACKNOWLEDGMENTS

I would like to thank my advisors Prof. Catherine E. Costello and Prof. John Samuelson for having me as a student and guiding my path. Thank you for both for your mentorship and guidance. You have formed me into a scholar and scientist.

Thank you to my committee members, Prof. Vickery Trinkaus-Randall, Prof. Joe Zaia, Prof. Miklos Sahin-Toth, and Prof. Emeritus Phillips Robbins. Thank you to Prof. Barbara Schreiber for ensuring that things have moved along smoothly.

To Prof. Mark McComb, thank you for introducing me to CE-MS, answering questions I have had regarding LC-MS/MS, orbitrap instrumentation, and proteomics database searches. Furthermore, I thank you for ensuring that the servers were functional and that software and related tools were available to use. Without these, it would have been very challenging to perform proteomic analyses.

Prof. Cheng Lin, thank you for taking time to help setup the EED-FTICR-MS/MS experiments and for answering questions I had surrounding fragmentation mechanisms, interpretations, and FTICR instrumentation theory and function. I wish I had more time to delve into fundamentals of instrumentation, I find it very interesting.

Christian Heckendorf, thank you for ensuring all the office and laboratory computers are functional. Thank you for answering computer and scripting questions, as well as for providing useful scripts and software to the lab.

Thank you to those that have taught me some of the instrumentation: Nancy Leymarie, for introducing me to the nanoAcquity UPLC systems, hand-packing columns, and for teaching me how to use and maintain the ThermoFisher Scientific LTQ-Orbitrap

properly. Thank you to Deborah R. Leon, for spending many hours teaching me the nuances of the LTQ-Orbitrap and the NanoAcquity UPLC systems. Also, Liang Han, for teaching me the use and maintenance of the Bruker AmaZon-ETD ion trap instrument. I also would like to thank Roger Théberge, for training me on the Bruker UltrafleXtreme MALDI TOF/TOF MS.

To all my colleagues, past and present: Carolina Agop Nersesian, Yousuf Aqueel, Giulia Bandini, Marianna Budnik, Gary Guy Bushkin, Kevin B. Chandler, Aparajita Chatterjee, Mengdi Fan, Rebecca S. Glaskin, Liang Han, Yuhuan Ji, Yan Jiang, Amanuel Kehasse, Kshitij Khatri, Nancy Leymarie, Deborah R. Leon, Yanyan Lu, Yang Mao, Le Meng, Edwin Motari, Hyejung Park, Yi Pu, Rekha Raghunathan, Chun Shao, Xiaofeng Shi, Yang Tang, Roger Théberge, Lilla Turiák, Stephen A Whelan, Bo Yan, Xiaobin Xu, Ying Zhou. You have all helped.

Thank you to Aparajita Chatterjee for spending time teaching me the culturing techniques of parasites. Also, thank you for helping answer questions I had regarding cloning and protein expression.

I especially thank Giulia Bandini and Deborah R. Leon. Giulia, you have taught me a lot about glycobiology and molecular biology. I have enjoyed our numerous conversations about science. You are a great collaborator, always willing to work together. Deborah, your help and guidance when I was first in the lab was critical to my success. You taught me how to *de novo* sequence glycopeptides and how to utilize the instrumentation to obtain the best results. Your have helped guide me and educate me, thank you.

To James Weinberg, you have been a wonderful friend and mentor. Your sharing of your knowledge and experiences, helped teach me aspects of synthetic organic chemistry that can only be obtained from decades of experience. Thank you. To my wife Briana, thank you for supporting us during this entire process, it would not have been possible without you.

THE UNIQUE GLYCOPROTEINS OF *CRYPTOSPORIDIUM PARVUM* AND *TOXOPLASMA GONDII*

JOHN ROBERT HASERICK

Boston University School of Medicine, 2017

Major Professor: Catherine E. Costello Ph.D., William Fairfield Warren Distinguished Professor

Cryptosporidium parvum and *Toxoplasma gondii* are obligate intracellular parasites transmitted by ingestion of resilient walled structures called oocysts. Infection is self-limiting in adults with normal immune systems. However, severe disease can occur in immunocompromised individuals, or those without cellular immunity. *Cryptosporidium* is a leading cause of infant mortality in developing countries, due to diarrhea. There are no human vaccines and no broad effective drug treatments. Several vaccine candidates have been described: the glycoproteins Gp900, Gp40, and Gp15 and the protein Cp23, the immuno-dominant-antigen. Details about modifications to these proteins have not previously been reported. Using mass spectrometry, we identified 16 *Cryptosporidium N*-glycosylated proteins, including Gp900 and a possible oocyst wall protein. The observed *N*-glycan structures exhibited only two compositions: $HexNAc_2Hex_5$ and $HexNAc_2Hex_6$; these glycoforms had a single extended arm. The simplicity of *Cryptosporidium N*-glycans contrasts with the complexity of host *N*-glycans. Four heavily *O*-glycosylated proteins included Gp900, Gp40, Gp15, and a novel mucin-like protein, Gp20. Single *O*-HexNAc residues modified Ser/Thr in low density

regions of Gp15 and Gp900, while attachment of *O*-HexNAc residues on tandem Ser/Thr repeats of Gp20 and Gp40 approached saturation. Identification of *N*-acetylgalactosamine (GalNAc) as the HexNAc released from proteins suggests that most *Cryptosporidium O*-glycans resemble the immunogenic Tn antigen (*O*-GalNAc). The immunodominant antigen Cp23, while not glycosylated, was discovered to be *N*-myristoylated and *S*-palmitoylated on the first and second residues, respectively. This is the first identification in *Cryptosporidium* of these modifications. Information about the *N*-glycans, *O*-glycans, and lipid modifications may be useful for design of better serodiagnostic reagents and more effective vaccines. To date, there are no vaccines against *Toxoplasma* infection, and the only available pharmaceutical therapies are expensive. In the second study, a novel *O*-fucose modification was discovered on nuclear pore-associated proteins including nucleoporins. This observation has profound implications on how the organism may regulate trafficking in/out of the nucleus by employing a system parallel to that described for *O*- linked *N*-acetylglucosamine in other organisms. In summary, the new details regarding the vaccine candidates of *Cryptosporidium* and the discovery of the novel *O*-fucose modifications in *T. gondii* provide information that could prove useful for development of effective drugs and vaccines.

TABLE OF CONTENTS

DEDICATION iv

ACKNOWLEDGMENTS v

TABLE OF CONTENTS x

LIST OF TABLES xx

LIST OF FIGURES xxi

LIST OF ABBREVIATIONS xxv

Chapter 1. Introduction. 1

Parasitology 1

Cryptosporidium parvum 1

Infection with Cryptosporidium 1

Cryptosporidiosis: The Disease. 3

Treatment of Cryptosporidiosis 3

The Glycoproteins and Vaccine Candidates of *C. parvum* 5

N-Glycosylated Proteins in *C. parvum* 5

O-Glycosylated Proteins in *C. parvum* 6

Non-Glycosylated Immunogens 9

Toxoplasma gondii 9

Transmission of *T. gondii* and Toxoplasmosis 9

N-Glycosylated Proteins of *T. gondii* 11

Introduction to Protein Glycosylation 12
N-Glycosylation 12
Biosynthesis of the Lipid-Linked Oligosaccharide 12
The Oligosaccharide Transferase 15
N-Glycan Trimming and Modification. 18
Functions of *N*-Glycosylation 19
N-Glycosylation in *C. parvum* and *T. gondii* 19
O-Linked Glycosylation 20
Polypeptide *N*-Acetylgalactosamine Transferases 21
O-*N*-Acetyl-Galactosamine Transferases of *C. parvum* and *T. gondii* 23
Mucins and their *O*-Glycans 24
Mucins in *C. parvum* 26
O-Linked *N*-Acetylglucosamine 27
O-GlcNAc Modification of Nuclear Pore Proteins 28
Nuclear Transportation Mechanisms Pertaining to FG-NUPS 30
O-GlcNAcylation in *C. parvum* and *T. gondii* 31
The Similarities of *T. gondii* and *C. parvum* OGT to the Plant OGT 32
O-Fucosylation 34
Mass Spectrometry 36
Fundamentals 36
Sample Introduction Into a Mass Spectrometer 40
Ionization Techniques 40

Matrix-Assisted Laser Desorption/Ionization (MALDI) 40
Electrospray Ionization (ESI) 43
Nanoflow Electrospray Ionization (nanoESI) 45
Electron-Impact Ionization (EI) 46
Mass Analyzers 48
Quadrupole Mass Analyzers 48
Time-of-Flight Mass Analyzers 50
Fourier Transform Mass Spectrometry (FT-MS) 55
Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS) 56
Orbitrap 57
MS/MS: Fragmentation 59
Collision-Induced Dissociation (CID) 60
Higher-Energy C-trap Dissociation (HCD) 61
Electron Reaction Based Fragmentation Methods (ExD) 62
Nomenclature for Naming Ions Observed in MS/MS Spectra 64
Peptide Fragmentation Nomenclature 64
Glycan Fragmentation and Nomenclature 67
Gas Chromatography-Mass Spectrometry (GC-MS) 70
Gas Chromatography 70
GC Columns 71
Gas Chromatography-Mass Spectrometry (GC-MS) 72
Derivatization 73

Chapter 2. Asparagine-Linked Glycans of *Cryptosporidium parvum*. 75
Abstract 75
Introduction 76
Materials and Methods 78
Parasites and Reagents 78
Protein Extraction 78
Trypsin Digestions 79
Release and Processing of *N*-Glycans 80
MALDI-TOF MS 81
Electron Excitation Dissociation (EED) Fourier Transform-Ion Cyclotron Resonance (FT-ICR) MS/MS 81
LC-MS/MS 82
Manual Interpretation of Glycopeptide MS/MS Spectra 83
LC-MS/MS Proteomics Database Search and Analysis 83
Scaffold Analysis 84
Analysis of *N*-Glycosylation Sites 85
Bioinformatics 85
Analysis of Released *N*-Glycans 86
Results 87
N-Glycans of *C. parvum* are Much Simpler than those of the Host and Most Other Parasites. 87
Confident Assignment of *Cryptosporidium parvum N*-Glycosylated Peptides. 90

Heavily Glycosylated Proteins Include an Immunodominant Protein (Gp900) and a Putative Oocyst Wall Protein (POWP1). 92
Nearly 90% of the Occupied *N*-Glycan Sites Contain Thr Rather than Ser. 93
Discussion 93
Tables 96
Figures 98
Chapter 3. The *N*-Glycosylated Proteins of *T. gondii*. 105
Introduction 105
N-Glycosylation in *T. gondii* 106
N-Glycosylated Proteins of *T. gondii* 107
Materials and Methods 107
Parasite Cell Culture and Protein Extraction 107
Enrichment of *N*-Glycosylated Proteins 108
Protein Precipitation and Trypsin Digestion 109
In-Solution Digestions 109
LC-MS/MS Analyses of In-Solution Digests 110
Data Analysis 111
Manual Interpretation of Glycopeptide Spectra Obtained from LC-MS/MS Experiments 111
Automated Database Searches of all Spectra Obtained from LC-MS/MS Experiments 112
Bioinformatics 112

Results 113
Discussion 116
Conclusions 118
Tables 121
Figures 123
Chapter 4. The *O*-Glycosylated Proteins of *C. parvum.* 128
Abstract 129
Introduction 130
Experimental Procedures 133
Reagents and Parasites 133
Protein Extraction and Trypsin Digestion 133
Mass Spectrometry 134
Manual Interpretation of Mass Spectra 135
Database Searches for Glycopeptides 136
Re-Annotation of Automated Database Searches 137
Other Bioinformatics Methods 138
O-Linked Glycan Release and Characterization 139
Monosaccharide Composition Determination 140
Results 141
The Vast Majority of Peptides with *O*-HexNAc Derive from Gp40, Gp15, and Gp900, which are Vaccine Candidates 141
Dense Arrays of *O*-GalNAc are Present on the Ser-rich Domain of Gp40 142

Isolated *O*-GalNAc Residues Decorate a Glycopeptide of Gp15 143
Dense Arrays of *O*-GalNAc are Present on Thr-rich Glycopeptides of Gp20 144
A Glycopeptide of Gp900 with Consecutive Thr is Lightly Modified by *O*-GalNAc, while Numerous Gp900 Glycopeptides Contain a Single *O*-HexNAc Residue 145
At the N-terminus of Cp23 Myristoyl Modifies Gly1, while Palmitoyl Modifies Cys2. 146
Discussion 147
Acknowledgements: 150
Conflict of Interest: 150
Author Contributions: 150
Tables 151
Figures 152
Chapter 5. The *O*-Fucosylated Proteins of *T. gondii.* 162
Abstract 162
Significance Statement 163
Introduction 163
Materials and Methods: 165
Lectin Pull-Downs. 165
Mass Spectrometry 166
In-Solution Digestion of AAL-Enriched Proteins 166
LC-MS/MS Analyses of In-Solution Digests 167
Electron Transfer Dissociation (ETD) Experiments 168

Monosaccharide Analysis: 168
Chemical Release of Monosaccharides by Reductive β-elimination..... 168
Per-Acetylation of the Released Monosaccharides..... 169
Gas Chromatography Mass Spectrometry (GC-MS) of the Alditol Acetates..... 169
Mass Spectrometry Data Analysis: Protein Database Searches 170
Mass Spectrometry Data Analysis: Comparison of All Data 171
Mass Spectrometry Data Analysis: Manual Verification of Database Search Results 171
Bioinformatics and Statistical Analyses. 172
Results..... 172
AAL Binds to *O*-Fucose in *T. gondii*..... 172
AAL Recognizes Proteins Involved in Gene Regulation. 174
Discussion..... 175
Acknowledgments..... 179
Tables..... 181
Figures..... 182
Chapter 6. Significant Findings and Future Work. 187
APPENDIX 1: The Amino Acid Seqence of Gp900. 193
APPENDIX 2: Ion Assignments, Peptides, Sequons, and Bioinformatics Data for the *N*-Glycosylated Peptides. 194

APPENDIX 2A Fragment Ion Assignments for FT-ICR EED MS/MS Spectra of Deutero-Reduced and Permethylated *N*-Glycans. 194

APPENDIX 2A.1 Assigned Ions for $HexNAc_2Hex_5$. 194

APPENDIX 2A.2 Assigned Ions for $HexNAc_2Hex_6$. 198

APPENDIX 2A.3 MALDI-TOF-MS Assigned Ions: *C. parvum* Released *N*-Glycans Deuteroreduced and Permethylated $[M+Na]^+$ 202

APPENDIX 2B: Bioinformatics and Related Data for the *N*-Glycosylated Proteins. 203

APPENDIX 2B.1 The Amino Acid Sequences Used to Generate the Weblogos. . 203

APPENDIX 2B.2 The *N*-Glycosylation Sequons for the Identified Proteins. 206

APPENDIX 2B.3 Percent NxT *versus* NxS Occupancies of *N*-Glycoyslation Sequons: Occupied and Unoccupied Sequons. 212

APPENDIX 2B.4 Table of Protein Names Cross-Referenced to Various Database Accession Numbers; Conserved Domains Assigned by InterPro. 213

APPENDIX 2C: The Predicted Lipid Linked *N*-Glycan Precursors of *C. parvum* and *T. gondii* 214

APPENDIX 3: The Complete Table of Observed *N*-Glycosylated Peptides and their Variants used to Generate Table 3.1. 215

APPENDIX 4: Ion Assignments, Peptides, Sequons, and Bioinformatics Data for the *O*-Glycosylated Peptides. 216

APPENDIX 4A: Complete Tables of Peptides and Ions for the Spectra Assigned to *O*-Glycosylated Peptides. 216

APPENDIX 4A.1 All HexNAc Modified Peptides used to Generate Table 4.1. ... 216

APPENDIX 4A.2 Spectra Re-Annotated using Glycresoft. ... 218

APPENDIX 4B: Ion Assignments for the Mass Spectra Annotated in Figs. 4.2 to 4.7, and Fig.4.9. ... 225

APPENDIX 4B.1: Ion Assignments for Fig 4.2. ... 225

APPENDIX 4B.2: Ion Assignments for Fig 4.3. ... 226

APPENDIX 4B.3: Ion Assignments for Fig 4.4. ... 228

APPENDIX 4B.4: Ion Assignments for Fig 4.5. ... 230

APPENDIX 4B.5: Ion Assignments for Fig 4.6 ... 231

APPENDIX 4B.6: Ion Assignments for Fig 4.7 ... 232

APPENDIX 4B.7: Ion Assignments for Fig 4.9 ... 234

APPENDIX 5: Complete Tables of the *O*-Fucosylated Peptides and Proteins ... 235

APPENDIX 5A: *O*-Fucosylated Peptides Identified. ... 235

APPENDIX 5B: Serine Rich Proteins or Proteins with Observed *O*-Fucosylated Peptides. ... 238

BIBLIOGRAPHY ... 241

CURRICULUM VITAE ... 290

LIST OF TABLES

Table 2.1. Summary of Identified *N*-Glycosylated Peptides 96

Table 2.2. Glycosyltransferase Enzymes Predicted from the Genomes of *C. parvum*, and a Related Organism, *T. gondii*. 97

Table 3.1 Manually Verified Proteomics Search Results of *N*-Glycosylated Peptides from *T. gondii.* 121

Table 3.2 Occupied *N*-Glycosylation Sequons 122

Table 3.3 Totals of the Different *N*-Glycoforms 122

Table 4.1. Overview of Tryptic Glycopeptides Identified by Mass Spectrometry, Examples of which are Shown in Figs. 4.2 to 4.5, 4.7, 4.9, and 4.10. 151

Table 4.2. Lipid Modifications of the N-terminus of Cp23. 151

Table 5.1: Proteins Containing Ser-Rich Domains Identified by AAL Pull-Down and Grouped by Function (Putative or Annotated) 181

LIST OF FIGURES

Fig. 1 *Cryptosporidium* Life Cycle 2

Fig. 1.2 *Toxoplasma gondii* Infection Routes 11

Fig. 1.3 The Canonical Eukaryotic Lipid-Linked-Oligosaccharide Biosynthetic Pathway 15

Fig. 1.4 Overview of the Different Types of *O*-Glycans 21

Fig. 1.5 The 20 Human GalNAc Transferases Displayed as a Phylogenetic and Genomic Analysis to Illustrate the Classes of Transferases. 23

Fig. 1.6 *O*-Glycan Structures 25

Fig. 1.7 *O*-GlcNAc Modification of Nuclear Pore Proteins 29

Fig. 1.8 Block Diagram of a Mass Spectrometer 37

Fig. 1.9: Mass Resolution and Resolving Power. 39

Fig. 1.10 Schematic of a Typical Electrospray Ionization (ESI) Setup. 44

Fig. 1.10 An Electron Impact Ionization Source 47

Fig. 1.11 Quadrupole Schematic 49

Fig. 1.12 A Schematic Overview of a 3D Quadrupole Ion Trap Mass Spectrometer. 50

Fig. 1.13 MALDI-TOF MS Schematic 53

Fig. 1.14 QqTOF Instrument Schematic. 55

Fig. 1.15 Cross-Cut View of an Orbitrap with an Ion in Motion 58

Fig. 1.16 The QE Series of Mass Spectrometers. 59

Fig. 1.17 Peptide Backbone Fragmentation Nomenclature 65

Fig. 1.18 A Hypothetical Peptide Naming all Theoretical Peptide Fragmentations 65

Fig. 1.19 The Domon and Costello Nomenclature for Glycosidic Bond Fragmentation . 67

Fig. 2.1. MALDI-TOF MS: Released, Deutero-Reduced Permethylated *N*-glycans [M + $Na]^{1+}$ 98

Fig. 2.2. Topology of the Most Abundant Glycoform $Hex_6HexNAc_2$ Determined by EED FT-ICR MS/MS. 14-eV EED FT-ICR MS/MS: $Hex_6HexNAc_2$ $[M + Na]^{1+}$ *m/z* 1800.9192..... 99

Fig. 2.3. Glycosidic Linkage Determination of the Most Abundant Glycoform $Hex_6HexNAc_2$ Determined by EED FT-ICR MS/MS. 14-eV EED FT-ICR MS/MS: $Hex_6HexNAc_2$ $[M+Na]^{1+}$ *m/z* 1800.9192..... 100

Fig. 2.4. *De Novo* Identification of Glycopeptides from a Whole *C. parvum* Oocyst Lysate..... 101

Fig. 2.5. Occupied *N*-Glycosylation Sites of POWP1 (cgd2_490)..... 102

Fig. 2.6. Total and Occupied *N*-Glycosylation Sequons in the Observed *N*-Glycosylated Proteins. 102

Fig. 2.7. Topology of the Second Most Abundant Glycoform $Hex_5HexNAc_2$ Determined by EED FT-ICR MS/MS. 14-eV EED FT-ICR MS/MS: $Hex_5HexNAc_2$ $[M + Na]^{1+}$ *m/z* 1596.8199. 103

Fig. 2.8. Glycosidic Linkage Determination of the Second Most Abundant Glycoform $Hex_5HexNAc_2$ Determined by EED FT-ICR MS/MS. 14-eV EED FT-ICR MS/MS: $Hex_5HexNAc_2$ $[M + Na]^{1+}$ *m/z* 1596.8199..... 104

Fig. 3.1 Cartoon Schematics of the *N*-Glycosylated Proteins, Showing Predicted Features, Total Observed Peptides, and Occupied *N*-Glycosylation Sequons. 123

Fig. 3.2 WebLogo of Occupied *N*-Glycosylation Sequons..... 124

Fig. 3.4A Comparison of MS/MS Fragmentation of the Ion *m/z* 1328.53 $[M+2H]^{2+}$ on an LTQ-Orbi (HCD, CID), QE, and QE+..... 125

Fig. 3.4B HCD MS/MS Fragmentation of the Ion *m/z* 1328.53 $[M+2H]^{2+}$ on a QE+: Revealing the Linear Glycan Sequence $HexNAc_2Hex_8$..... 126

Fig. 3.4C HCD MS/MS Fragmentation of the Ion *m/z* 1328.53 $[M+2H]^{2+}$ on a QE+: Revealing the Peptide Sequence NYTSEALR. 127

Fig. 4.1 Schematics of *Cryptosporidium* Glycoproteins Characterized by Mass Spectrometry. 152

Fig. 4.2 HCD MS/MS Spectrum (@ 30 V) of a Tryptic Glycopeptide of Gp40 Shows Saturation of an Array of Ser with *O*-HexNAc Residues. 153

Fig. 4.3. HCD MS/MS Spectrum (@ 30 V) of a Tryptic Glycopeptide of Gp15 Shows that Three of Four *O*-Glycan Sites Contain the HexNAc Modification. 154

Fig. 4.4 HCD MS/MS Spectrum (@ 30V) of a Tryptic Glycopeptide of Gp20 Shows Near Saturation of Consecutive Thr Residues with the *O*-HexNAc Modification. 155

Fig. 4.5 HCD MS/MS Spectrum (@ 30V) of a Tryptic Glycopeptide of Gp900 Shows a Single *O*-HexNAc Modification on a Thr Residue..... 156

Fig. 4.6. HCD MS/MS Spectrum (@30 V) of an N-terminal Peptide of Cp23 (Minus Met-1) Shows Gly-1 is Modified with Myristate and Cys-2 is Modified with Palmitate. 157

Fig. 4.7. HCD MS/MS Spectrum (@ 45V) of a Tryptic Glycopeptide of Gp40 Gives Complete Sequence of the Peptide. 158

Fig. 4.8. GC-MS Analysis Shows the *O*-Glycan Released by β–Elimination from *C. parvum* Glycoproteins is GalNAc. 159

Fig. 4.9. HCD MS/MS Spectrum (@ 30V) of a Tryptic Glycopeptide of Gp900 Shows Partial Glycosylation of a Thr-rich Repeat. 160

Fig. 4.10. A Representative GlycReSoft Re-Annotated Spectrum, One of 345 Tandem Mass Spectra Deposited into the PRIDE Repository. 161

Fig. 5.1: AAL Bound Proteins are Modified by *O*-Fucose on Ser/Thr Residues. 182

Fig. 5.3: AAL Binds Proteins Modified by One or More *O*-Fuc on Ser/Thr Residues, as Verified by Neutral Loss of dHex, MS/MS Fragmentation and GC-MS Monosaccharide Analysis. 184

Fig. 5.4:Proteins and Peptides with Serine Rich Sequences. 185

Fig. 5.5: ITMS ETD MS/MS *m/z* 943.4736 $[M+2H]^{2+}$, (390)ATGGATGLLGSSSLFGDTK(408) + 1 dHex 186

LIST OF ABBREVIATIONS

3D.....three dimensional

AA..... amino acids

AcOH ..... acetic acid

AIDS .....acquired immunodeficiency syndrome

ALC.....average local confidence

Alg.....asparagine-linked glycosylation

BU ..... Boston University

CAD .....collisionally activated decomposition

CHO .....Chinese hamster ovary cells

CID.....collision-induced dissociation

CLMPT1 .....Cleft lip and palate transmembrane protein 1

ConA .....Concanavalin A lectin (*Canavalia ensiformis* lectin)

Cxn.....calnexin

DBA ..... *Dolichos biflorus* agglutinin

DC .....direct current

DH..... dehydratase

Dol-PP.....dolichol pyrophosphate

DTT.....dithiothreitol

ECD.....electron capture dissociation

EDD .....electron detachment dissociation

EDTA.....ethylenediaminetetraacetic acid

EED ..... electron excitation dissociation

EGF ..... epidermal growth factor

EI ..... electron impact

ER ..... endoplasmic reticulum

ERAD ..... endoplasmic reticulum-associated degradation

ESI ..... electrospray ionization

ETD ..... electron transfer dissociation

eV ..... electron volts

FA ..... formic acid

FDR ..... false discovery rate

FG-NUPS ..... phenylalanine glycine repeat containing nucleoporins

Fuc ..... fucose

FWHM ..... full width half maximum

FT-ICR ..... Fourier transform ion cyclotron resonance

FT-MS ..... Fourier transform mass spectrometry

GA ..... giberellin

Gal ..... galactose

GalNAc ..... *N*-acetylgalactosamine

Gap50 ..... glideosome-associated protein 50

GC ..... gas chromatography

GDP-Man ..... guanosine diphosphate mannose

Glc ..... glucose

GlcNAc ..... *N*-acetylglucosamine

Gls1 ..... α-glucosidase I

Gls2 ..... α-glucosidase II

GS ..... glucan synthase

GSL ..... *Griffonia simplicifolia* lectin

GTP ..... guanosine triphosphate

GTPases ..... guanosine triphosphatases

HCD ..... higher energy collisional dissociation

HEAT ..... Huntington elongation factor three

HexNAc ..... *N*-acetylhexosamine

HFF ..... human foreskin fibroblasts

HPLC ..... high-pressure liquid chromatography

IAA ..... iodoacetamide

ID_{50} ..... infectious dose to cause disease in 50% of individuals

IRMPD ..... infrared multiphoton dissociation

Jacalin ..... *Artocarpus integrifolia* lectin

kDa ..... kilodaltons

LC ..... liquid chromatography

M ..... molecular ion

mAb ..... monoclonal antibody

MALDI ..... matrix-assisted laser desorption/ionization

Man ..... mannose

MeOH ..... methanol

MS .....mass spectrometry

MS/MS .....tandem mass spectrometry

Nano-ESI.....nanoflow electrospray ionization

N-glycans .....asparagine-linked glycans

NCI.....negative chemical ionization

Nd:YAG..... neodymium-doped yttrium aluminium garnet

NETD .....negative electron transfer dissociation

nL ..... nanoliter

NLS .....nuclear localization signal

NUPS .....nucleoporins

O-glycans .....serine or threonine-linked glycans

OGT ..... *O*-*N*-acetylglucosamine transferase

OST .....oligosaccharyltransferase

OWPS ..... oocyst wall proteins

PNGase F ..... peptide *N*-glycanase-F

POFuT.....protein *O*-fucosyl transferase

POWP1 ..... possible oocyst wall protein 1

pp-Dol .....dolichol pyrophosphate

ppGalNAcT.....polypeptide *N*-acetylgalactosamine transferase

ppm ..... parts-per-million

ppb.....parts-per-billion

PQD.....pulsed Q-dissociation

QIT ..... quadrupole ion trap

Ran ..... Ras-related nuclear protein

rf..... radio frequency

RNA .....ribonucleic acid

RT .....room temperature

SA ..... supplemental activation

SAG..... surface antigen

Ser .....serine

siRNA ..... short interfering RNA

SPY ..... Spindly

SRS .....SAG1-related sequence

T .....Tesla

TEM .....transmission electron microscopy

Thr..... threonine

TIC .....total ion chromatogram

TOF .....time-of-flight

TPR .....tetratricopeptide repeats

TSR .....thrombospondin repeats

UCG ..... unique *Cryptosporidium* glycoprotein

UDP.....uridine diphosphate

UDP-GalNAc..... uridine diphosphate *N*-acetylgalactosamine

UDP-GlcNAc ..... uridine diphosphate *N*-acetylglucosamine

UEA ..... *Ulex europaeus* agglutinin

UPLC ..... Ultra Performance Liquid Chromatography

UPR ..... unfolded protein response

UV ..... ultraviolet

VVA ..... *Vicia villosa* agglutinin

WGA ..... wheat germ agglutinin

x ..... any amino acid

Chapter 1. Introduction.

Parasitology

Cryptosporidium parvum

Cryptosporidium spp. are eukaryotic parasites belonging to the phylum Apicomplexa. *Cryptosporidium* can infect a wide range of warm blooded animal hosts including humans. *Cryptosporidium parvum* and *Cryptosporidium hominis* are typically spread via the fecal-oral route and infection inflicts prolonged watery diarrhea. Both *C. parvum* and *C. hominis* cause cryptosporidiosis in humans, however, transmission of *C. hominis* occurs only human-to-human, unlike *C. parvum* that is transmissible between humans and animals (1).

Infection with Cryptosporidium

Infections are commonplace amongst those that live in close proximity to farm animals, in particular ruminants, which are the major source of sporadic outbreaks (2). Other farm animals that carry *Cryptosporidium* are pigs, poultry, and rabbits; however, it is not known what role these animals play in the transmission to humans (3). Zoonotic transmission occurs frequently when agricultural run-off contaminates drinking water supplies, or when people come into contact with infected animals (4, 5). Human-to-human infection can happen through contaminated drinking water or from contaminated recreational water (6, 7).The infectious stage of *Cryptosporidium* sp. is the oocyst, a small thick-walled shell approximately 5 μm in diameter, which is impervious to most disinfectants (see Fig.1)(8).

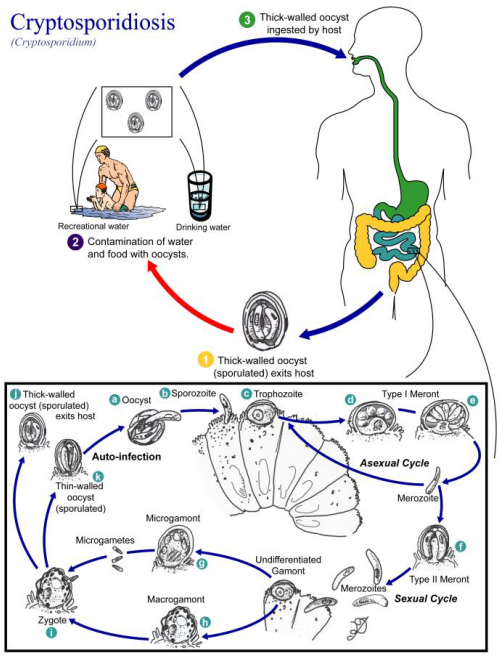


Fig. 1 *Cryptosporidium* Life Cycle

This image is in the public domain and thus free of any copyright restrictions. Content provider: CDC/Alexander J. da Silva, PhD/Melanie Moser. http://www.dpd.cdc.gov/dpdx/HTML/Cryptosporidiosis.htm

Inside the oocyst are four sporozoites (length ~1.3 μm, width 0.5 μm) and, after ingestion, while in the small intestine, the individual parasites excyst from the oocyst, infecting the epithelium (9). The oocysts are extremely infectious and, depending upon the strain, the ID_{50} is approximately 10-100 oocysts, a value determined from experimentally infecting healthy human volunteers, as well as from data extrapolated from widespread infections (10-13). There is reason to believe that transmission may also occur from inhalation, as it has been shown to occur in both immune-competent and immune-incompetent individuals (14). Furthermore, respiratory cryptosporidiosis is well established in poultry, being the primary manifestation in these animals (15). This suggests that inhalation may be another main entry point for infection, in places where dusty environments or water mist from fountains could potentially carry the small

oocysts. In addition, it has been found that *Cryptosporidium* sp. are capable of infecting the conjunctiva of immunodeficient monkeys and the conjunctiva of pigs, indicating that any exposed mucosal epithelia may be an entry point for infection (16, 17).

Cryptosporidiosis: The Disease.

For most healthy adults, cryptosporidiosis is a self-limiting disease. However, for people who do not have a functional immune system (e.g. AIDS patients, those with congenital hypogammaglobulinemia, or organ transplant recipients undergoing immunosuppressive therapy), the disease can be life-threatening due to chronic diarrhea that leads to severe malnutrition and dehydration (18, 19). Furthermore, *Cryptosporidium* is a severe problem for children under two years old in developing countries, for whom it is has been identified as a leading cause of infant mortality due to diarrheal disease (20). Disease is not limited to developing countries, as there are numerous reports of occasional outbreaks in modernized cities when public drinking water filtration units fail. One of the best known cases in the United States took place in 1993 in Milwaukee, WI, where an estimated 400,000 households were affected (6).

Treatment of Cryptosporidiosis

Treatment of cryptosporidiosis is mostly limited to fluid and nutritional replacement until the infection subsides. Nitazoxanide is the only drug treatment available; it has been approved for use against *Cryptosporidium* in persons ≥ one year old (21). Unfortunately, this drug has been found to be totally ineffective for the treatment of cryptosporidiosis in otherwise healthy children, children with HIV, and

adults with HIV (22-24). The only approach which has been found to prevent infection is to treat *Cryptosporidium* infections in otherwise healthy adults, children with HIV, and adults with HIV or AIDS with bovine hyperimmune colostrum (25-29). Most of these studies inoculated cows with *C. parvum* oocyst lysates to induce the hyperimmune response, with the subsequent production of anti-cryptosporidial antibodies concentrated in the colostrum. Two of the proteins which reacted strongly with the bovine colostrum were identified as Gp900 and Cp23 (30, 31).

Attempts have been made to further characterize the immunogens of *Cryptosporidium,* which confer the protective and therapeutic properties of bovine hyperimmune colostrum, but, unfortunately, these have had limited success (30, 32). Two proteins in particular, Cp23 (also referred to as rC7), which has been identified as the immunodominant antigen and the later as an immunogenic glycoprotein, and Gp15/40 have been used as recombinant proteins or DNA-based vaccines to induce production of protective colostrums in cows (30, 32). These strategies for inoculation present to the cow's immune system an antigen that does not represent the native immunogen and therefore could produce sub-optimal results in a real world application. One of the pitfalls from inoculation with recombinant proteins is that any protein modifications present on the native immunogen would not be present on recombinant proteins. This is also true for DNA-based vaccines which utilize the biosynthetic machinery of the cells of the inoculated, which may not modify the proteins in the same manner that the parasite does. Most of the immunogenic proteins identified in previous studies are glycoproteins (15, 33-35). In many instances, it has been shown that the glycosylation is part of the

immunogenic epitope on the native immunogens (36, 37). Despite the knowledge that many of the immunogenic proteins are glycoproteins and that antibodies against these immunogens can provide protection against cryptosporidiosis, the glycosylation remains uncharacterized. The knowledge of which proteins are modified with glycans, where the modification occurs, and what the glycan composition is could be used to help produce effective vaccines. The primary goal of the research presented in this dissertation is to identify and fully characterize the glycoproteins of *C. parvum*.

The Glycoproteins and Vaccine Candidates of *C. parvum*

N-Glycosylated Proteins in *C. parvum*

One of the first studies of the immune response to *C. parvum* immunogens was performed by Luft et al. in 1987; they found that there are approximately 15 glycoproteins that are immunogenic and eight of these appear to be *N*-glycosylated, as determined by lectin blots of SDS-PAGE separated proteins. Interestingly, they also found that, when treated with mixed glycosidases, the antibody recognition was markedly reduced for most of the antigens, and specifically, that PNGase F had a similar reduction on antibody recognition of the same eight proteins, suggesting that the carbohydrate component of the glycoproteins was important to antigenicity (36). A later study confirmed the assertion that the carbohydrates play a critical role in the immune response; it was found that two mAbs had broad recognition of several *C. parvum* proteins, and that this recognition was almost completely abolished when the proteins were treated with sodium periodate (38).

One of the very large relative molecular weight immunodominant proteins, Gp900, was later determined to be both *O*-glycosylated and *N*-glycosylated (15, 35). Gp900 is immunogenic, and antibodies against the protein offer protection against infection (15, 31). This was shown in both in an *in vitro* infection model using monoclonal antibodies that recognize Gp900 and blocked *C. parvum* sporozoite cell invasion, as well as antibodies against Gp900 in bovine hyperimmune colostrum, which provided some protection against infection in calves (15, 31). Interestingly, after treatment with PNGase F, three of the mAbs lost their reactivities with Gp900, suggesting again that *N*-glycosylation plays an important role in the immune response to *C. parvum* immunogens (31).

O*-Glycosylated Proteins in *C. parvum

In *C. parvum*, several additional glycoproteins have been shown to be important to infection and infiltration of the host organism. As already mentioned, Gp900, a large mucin-like protein, contains both *O*-linked and *N*-linked glycans. It is highly immunogenic and is present in both oocysts and on freshly excysted sporozoites. It has been observed to be deposited onto the host surface by infiltrating sporozoites (33). If cells are incubated with recombinant Gp900, specific recombinant domains of Gp900, or mAbs against Gp900, sporozoite infiltration is dramatically reduced, thus demonstrating the importance of this protein during host parasite interactions (33). Two other surface glycoproteins, Gp40 and Gp15, are located on the surface of the sporozoite, and, similar to Gp900, are deposited on the host surface as the sporozoites migrate (39). In a different study, a monoclonal antibody, mAb 4E9, also showed strong protection against infection

in a similar *in vitro* invasion assay (15) Interestingly, mAb 4E9, an IgM, recognizes several proteins, Gp900, Gp40, and another protein, with M_r around 220-kDa, that is only found in the oocysts. When treated with periodate, the reactivies are lost, suggesting that the Ab is recognizing a glycan epitope (15). In addition, when a *C. parvum* lysate is incubated with different concentrations of α-galactosidase, the positive bands on a Western that uses mAb 4E9 as a probe shift to lower apparent molecular weights and eventually disappear for some of the positive proteins (15). These results suggest the presence of terminal 1,-3-*N*-acetyl-galactosamine residues or the presence of Ser/Thr-linked *N*-acetyl-galactosamine in the epitope (15).

One of the other immunodominant antigens, Gp15, is encoded by that same gene as Gp40; however, the translated protein is processed by a furin-like protease to create an N-terminal portion (Gp40) and a C-terminal portion (Gp15). Gp15 has been found to contain a GPI-anchor (40-42). The Gp15 antigen was also found to be one of the immunodominant antigens present in bovine hyperimmune colostrum generated from oocysts, and freshly excysted sporozoites (43). A monoclonal antibody recognizing Gp15, an IgA, referred to as mAb5C3, was shown to reduce the number of shed oocysts in experimentally infected mice (37). This demonstration of passive immunity against Gp15 and its ability to reduce *C. parvum* infection suggests that Gp15 would be a good vaccine candidate (37). The authors also demonstrated that the antibody appears to recognize a glycan epitope on Gp15 (37). When co-incubated with 200 mM *N*-acetyl-glucosamine (GlcNAc) or *N*-acetyl-galactosamine (GalNAc), the binding on a Western blot was decreased; this effect was much more pronounced for GlcNAc than for GalNAc

(37). In addition, when mAb5C3 was used for immunofluorescent microscopy, Gp15 appeared to localize to both the oocyst and sporozoite, where it was present as a coat on the sporozoites inside the oocyst that was quickly shed after excystation (37).

It was postulated that Gp15 was likely modified at three sites near the N-terminus, since Edman sequencing of tryptic peptides yielded an aberrant result, where certain amino acids could not be sequenced properly, and the determined sequence deviated from the predicted sequence (ETSEAAATVDLFAFTLDGGKR) by giving ET?EAAA?VDLFAF?LDGGKR, where the question marks represent the deviations (40). Later, using an array of different synthetic glycopeptides from the N-terminus of Gp15 containing different numbers of GalNAc modified residues on the available Ser/Thr, it was found that sera from people previously infected with *C. parvum* had the strongest response to the peptide ETS∗EAAAT∗VDLFAFT∗LDGGK (where the asterisks indicate modification with GalNAc) (44). Furthermore, there was very little response to the unmodified peptide; for those that contained fewer GalNAc modifications, the response was markedly decreased (44).

An additional protein has been described in the literature as a glycoprotein of possible interest. A C-type lectin, which contains a mucin-like domain, has been shown to be prevalent during infection of cells, with the expression peaking at 48 hours post-infection (45). The presence of glycosylation was predicted, based upon the presence of a mucin-like domain, but the actual modification of the protein was never established (45).

Another class of proteins that has been shown to be important during the transmittable oocyst stage of the parasite is made up of the oocyst wall proteins (OWPS)

(35). This set of proteins appears to be involved in the connection between the sporozoite and the inner oocyst wall. These proteins are thought to be possibly *N*-glycosylated or *O*-glycosylated (35). The inner portion of the oocyst wall appears to be composed largely of these proteins, connected to a rigid bilayer of lipid and an unknown structural component (35). The composition and structure of the *O*- and *N*-glycans of *C. parvum* have only been partially detailed, via lectin stains and glycosidase treatments. Mass spectrometry should provide a means to identify not only the compositions and structures of the glycans, but it also should provide a means to identify the proteins that contain a particular glycoform.

Non-Glycosylated Immunogens

When studying the sera and the immune response in patients known to have had cryptosporidiosis, it was discovered that most of the patients had a very strong antibody response to an approximately 23-kDa protein, as determined by Western blots of *C. parvum* lysates with patient sera (46). Furthermore, it was shown that IgM levels rose during infection; this was followed by a rise in IgG levels that tapered off after 12 months, suggesting that this protein could possibly be used for vaccination (46).

Toxoplasma gondii

Transmission of *T. gondii* and Toxoplasmosis

Toxoplasma gondii, like *C. parvum*, is an apicomplexan parasite that can infect most animals through ingestion of oocysts shed in feces of felines (see Fig. 1.2), which are the definitive host (47). Unlike *C. parvum* oocysts that are passed from the host as

sporulated and infectious, *T. gondii* oocysts are passed unsporulated and must sporulate outside the host in an oxygen-rich environment and later become infectious (48). *T. gondii* oocysts contaminate the soil where cats defecate, posing a threat to gardeners and children that play in this soil. The oocysts can be carried on produce grown in contaminated soil and, if they are not washed properly, persons who consume the fresh produce can acquire the infection (49). Individuals who change the litter boxes of pet cats that have consumed wild animals are also at risk for toxoplasmosis. Another method of infection occurs through ingestion of undercooked or raw meat where *T. gondii* tissue cysts reside (see Fig. 1.2). Potentially, any animal or bird can harbor tissue cysts and, therefore, care must be taken to thoroughly cook any meat prior to eating.

Epidemiological studies estimate that people in the United States, ages 6 – 49, are approximately 9% sero-positive for *T. gondii*; in other parts of the world, the positive rate could be as high as 50% (49, 50). Most people infected with *Toxoplasma* may only have mild flu-like symptoms for a short period of time, with the parasite going into a dormant state, hiding in tissue cysts after the acute stage of infection. However, systemic toxoplasmosis can manifest itself in those who lack cell-mediated immunity, such as AIDS patients, organ transplant recipients undergoing immunosuppressive therapy, or fetuses. If pregnant women becomes infected, or if a latent infection becomes active, *T. gondii* can pass through the placenta to the fetus, and then spontaneous miscarriage or birth defects can arise (47).

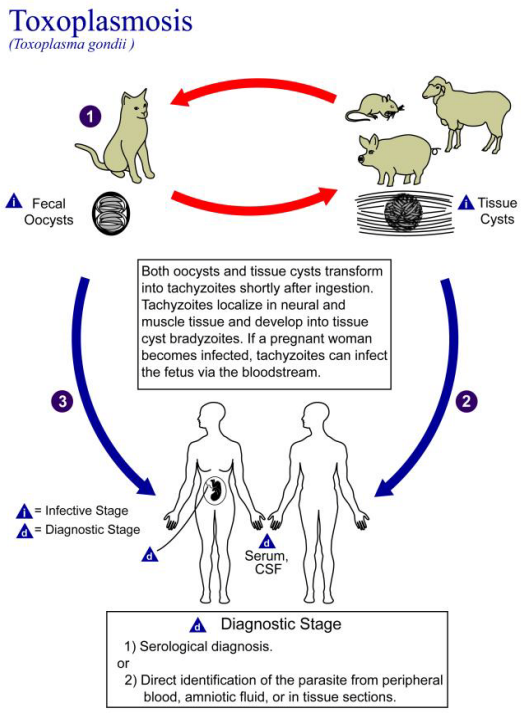


Fig. 1.2 *Toxoplasma gondii* Infection Routes

This image is in the public domain and thus free of any copyright restrictions. Content provider: CDC/Alexander J. da Silva, PhD/Melanie Moser, 2002. http://www.dpd.cdc.gov/dpdx/HTML/Toxoplasmosis.htm

N-Glycosylated Proteins of *T. gondii*

Very few reports have identified any of the *N*-glycosylated proteins in *T. gondii*. An early study of *T. gondii N*-glycans first isolated the Dol-PP-glycans, and used MALDI-TOF MS to show that the lipid-linked *N*-glycan precursor contained mostly $HexNAc_2Hex_9$ and a small amount of $HexNAc_2Hex_8$ (51). Then, global release of *N*-glycans utilizing PNGase F, and examination of the MS profiles, in conjunction with mannosidase treatments, revealed the presence of several glycoforms with the compositions $Hex_{(5-9)}HexNAc_2$ (51). The authors suggested that *T. gondii* could possibly utilize the host cells' *N*-glycan precursor; however, this was never proven. The only direct evidence and characterization of *N*-glycosylation in any apicomplexa was a study

of the protein Gap50 (52, 53). This protein appears to contain several glycoforms consisting of $Hex_{(5-8)}HexNAc_2$, attached to Asn residues on multiple locations within the protein.

Introduction to Protein Glycosylation

This section provides an introduction into the biochemical pathways utilized for protein glycosylation. Described herein are canonical pathways that have been studied in model organisms. The first part of this section describes the *N*-glycosylation pathway. Historically speaking, much of our current understanding of this pathway has been determined through studies involving *Saccharomyces cerevisiae.* The *N*-glycosylation biosynthetic pathway is described within the context of the historical discoveries in S. cerevisiae, as well other significant findings in higher eukaryotes, archaea, and prokaryotes. The second section of this introduction describes the various *O*-glycosylation pathways, with a focus on GalNAc transferases and mucin/mucin-like *O*-GalNAc modifications. In addition, details describing the *O*-GlcNAc modification are covered, as they pertain to nuclear pore proteins and FG-NUPs.

N-Glycosylation

Biosynthesis of the Lipid-Linked Oligosaccharide

Protein *N*-glycosylation is the addition of a polysaccharide to the amide group of asparagine on polypeptides containing the consensus sequence Asn-x-(Ser/Thr), where x is not Pro (54). This sequon is common amongst eukaryotes, prokaryotes, and archaea (55-57). Amongst eukaryotes, a conserved lipid-linked oligosaccharide precursor is transferred, as a group, to secreted proteins within the ER (58). The lipid-linked precursor

is conserved amongst higher eukaryotic organisms, but varies for bacteria and archaea. For the higher organisms, it contains an oligosaccharide with the composition $GlcNAc_2Man_9Glc_3$ (59). Yeast (*Saccharomyces cerevisiae*) has served as a model eukaryotic organism for studying the pathways and mechanisms within the *N*-glycosylation pathway.

N-glycosylation begins with the construction of the lipid-linked *N*-glycan precursor on the cytosolic side of the ER , utilizing the Alg enzymes that sequentially add monosaccharides onto a dolichol-phosphate anchor embedded in the ER membrane (see Fig. 1.3) (60, 61). Initiation of the canonical *N*-glycan biosynthesis pathway requires Alg7, which utilizes the sugar nucleotide UDP-GlcNAc and transfers GlcNAc to dolichol phosphate, creating GlcNAc-PP-Dol (62). The second GlcNAc is then added by the Alg13/Alg14 complex, forming $GlcNAc_2$-PP-Dol (63). The double GlcNAc, often called the chitobiose core, appears to be the minimal *N*-glycan observed in eukaryotic organisms (64).

The next five mannoses are added by the Alg enzymes Alg1, Alg11, and Alg2. The glycosyl transferases all utilize GDP-Man as the activated sugar donor, with the first Man transfer occurring directly to the terminal GlcNAc on the already constructed $GlcNAc_2$-PP-Dol. First, Alg1 adds a Man residue via a β-1,4- glycosidic linkage to the terminal GlcNAc (65). Next, Alg2 adds the second and third Man residues to the previously added Man residue through α-1,3- and α-1,6 glycosidic linkages (60). Then, Alg11 sequentially adds the fourth and fifth Man residues via α-1,2-linkages, starting on the terminal α-1,3- Man (66). The constructed $Man_5GlcNAc_2$-PP-Dol intermediate is then

flipped from the cytosolic face of the ER into the lumen, via a mechanism that is not yet not fully understood, but, at least in humans appears to be dependent upon the protein Rft1 (67, 68). Unlike the cytosolic-facing oligosaccharide transferases, the luminal-facing transferases require Man-P-Dol as a donor, as opposed to GDP-Man. The synthesis of Man-P-Dol is performed by the enzyme Dpm1, which conjugates Man to P-Dol, using GDP-Man as a donor (69). Within the lumen of the ER, four additional Man residues are added to the now luminal-facing $Man_5GlcNAc_2$-PP-Dol, all utilizing the lipid-linked Man-P-Dol donor. First, Alg3 adds a sixth α-1,3-linked Man residue onto the α-1,6-linked Man residue located on the shorter terminal arm (70). Next, Alg12 adds a seventh α-1,6-linked Man residue to the Man residue that was previously modified by Alg3 (71). Finally, Alg9 adds the eighth and ninth Man residues, both α-1,2-linked, to separate terminal mannose residues (the α-1,3 and α-1,6 linked Man) (72).

The final steps of lipid-linked *N*-glycan precursor biosynthesis involve capping with a glucose trisaccharide. Again, this process takes place in the lumen of the ER; the enzymes which perform this function require Glc-P-Dol as the donor. The enzyme Alg5 synthesizes Glc-P-Dol, requiring UDP-Glc as the donor and p-Dol as the substrate (73). The first Glc residue is α-1,3-linked to the terminal Man residue on the first arm of Dol-PP-$GlcNAc_2Man_9$ (74). A second Glc residue is then added to the first, via an α-1,3-linkage by the enzyme Alg8 (75). Finally, Alg10 adds the terminal Glc residue, connecting it to the second Glc residue via an α-1,2-linkage (61). The biosynthesis of the lipid-linked precursor is concluded with the addition of the last Glc residue, creating Dol-PP-$GlcNAc_2Man_9Glc_3$. This oligosaccharide is transferred to the nascent protein in a co-

translational (or, in some cases, post-translational), event. These processes will be outlined in the following section.

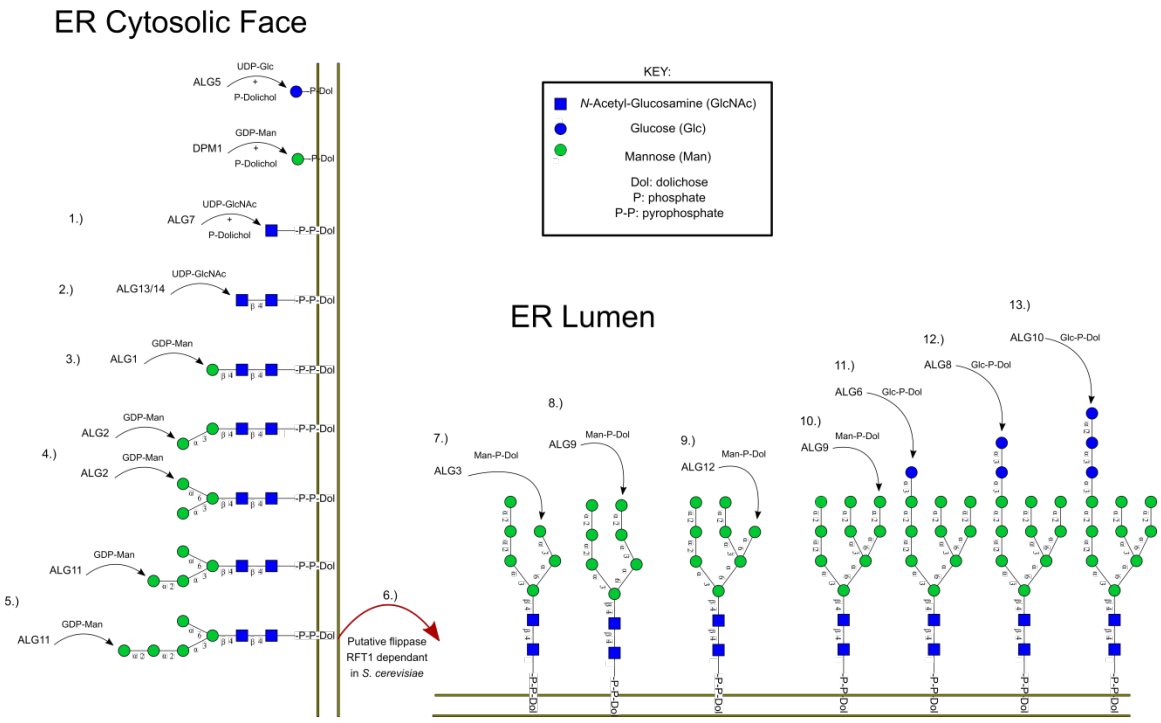


Fig. 1.3 The Canonical Eukaryotic Lipid-Linked-Oligosaccharide Biosynthetic Pathway

The Oligosaccharide Transferase

N-glycosylation has primarily been described as a co-translational event, occurring within the ER during the translation of the protein inside the lumen of the ER (76, 77). There is, however, some evidence suggesting that *N*-glycosylation may occur post-translationally under certain conditions (78, 79). In fact, the catalytic subunits of the mammalian OST, Stt3a and Stt3b, appear to have different activities, with Stt3a having a higher co-translational and Stt3b having higher post-translational OST activity (80).

The OST is multi-subunit complex of proteins, some of which partake in the co-translational or post-translational transfer of the lipid-linked *N*-glycan precursor in

eukaryotes (76). In *S. cerevisiae*, the OST contains eleven subunits: OST1, Stt3, Wbp1, OST3, OST6, Swp1, OST2, OST5, OST4, Wbp1, and Swp1p (81). The roles of all the yeast OST subunits and their higher eukaryotic homologs are still being investigated, and several have already been identified as essential for *N*-glycosylation.

The Stt3 subunit is one of the most evolutionarily conserved and contains the catalytic domain for the transfer of the lipid-linked-oligosaccharide to the Asn of the polypeptide (82). Direct evidence obtained from cross-linking experiments supports the hypothesis that Stt3 contains the catalytic domain, demonstrating that it intimately interacts with the nascent peptide chain (83).

The evolutionary conservation of Stt3 extends into some of the prokaryotes capable of making *N*-glycosylated proteins, such as *C. jejuni.* This organism contains an OST referred to as PglB, a homolog to the yeast *Stt3 (56)*. The function of PglB as an OST was proven by its introduction into *E. coli* , an organism that normally does not make *N*-glycosylated proteins, but was shown to be capable of such activity after PglB was transferred (56). Interestingly, there is a notable similarity to the mechanism found in eukaryotes, in that PglB requires an acetamido group on carbon 2 of the reducing end monosaccharide on the lipid-linked precursor, suggesting that there is an evolutionarily conserved transferase mechanism (84).

The *N*-glycosylation sequon to which the glycan is transferred in *C. jejuni* is similar to that found in eukaryotes. However, the sequon appears to require an acidic residue at the -2 position, extending it to (Asp/Glu)-x-Asn-x-(Ser/Thr), where x is any amino acid except Pro (85).

It has also been found that different Stt3 variants in eukaryotes confer substrate specificity and donor preference. One such instance has been demonstrated in *Trypanosma brucei*, where TbStt3a appears to preferentially transfer $Man_5HexNAc_2$ with a single elongated arm to peptides with *N*-glycosylation sequons close to acidic amino acids, whereas TbStt3b and TbStt3c prefer to transfer $Man_9HexNAc_2$ and are not as substrate-selective (86). In addition, it has been shown that *T. brucei* is capable of utilizing both $Man_5GlcNAc_2$ and $Glc_1Man_5GlcNAc$ as lipid-linked precursor donors when either Alg3 or glucosidase-II has been knocked out of the genome (87, 88).

There are four additional conserved OST subunits which were first characterized in yeast and have homologs in mammalians. These are: Ribophorin I (OST1), Ribophorin II (Swp1), Dad1 (OST2), and OST48 (Wbp1) (89). Ribophorin I and Ribophorin II, have consistently been co-isolated with ribosomal preparations, suggesting there is a close relationship between the OST, translation, and translocation into the ER (90-92). Ribophorin I has been tied to OST activity; when it is removed from an *in vitro* system by immune depletion, there was a marked decrease in OST activity. It has been suggested that Ribophorin I may have a dolichol-binding domain, which may facilitate the transfer of the lipid-linked precursor (93).

The fourth conserved OST subunit is a small hydrophobic protein Dad1 whose function is to keep Ribophorin I/II and OST48 in a stable structure, allowing the OST to function properly (94, 95). The fifth sub-unit, OST48, appears to have a role similar to Dad1. It keeps the OST complex stable. When OST48 was knocked down using siRNA in HeLa cells, hypoglycosylation of proteins was quite apparent (96).

N-Glycan Trimming and Modification.

After the OST transfers the oligosaccharide moiety from the lipid-linked precursor to the polypeptide, the transferred glycan is trimmed of the terminal glucose residues by two exoglucosidases, α-glucosidases I and II (97). Glucosidase I removes the terminal α-1,3- Glc which then allows glucosidase II (Gls2) to trim back the two remaining α-1,2-linked glucose residues (98). Only after all Glc residues are removed will the protein exit the ER; this is, in part, due to the binding of the mono-glucosylated *N*-glycan structure by the ER resident protein calnexin (99). A single terminal glucose can be added back onto the already trimmed high mannose *N*-glycan, sequestering the protein inside the ER through interactions with calnexin, calreticulin, and a thiol oxidoreductase, and allowing the protein to undergo proper folding (100). When the protein cannot be folded properly, or, under conditions of stress, the protein is exported and degraded. This process has been described as the unfolded protein response (UPR) and ER-associated degradation (ERAD) (101, 102).

A properly folded *N*-glycosylated protein can be exported from the ER as a high mannose type *N*-glycosylated protein, or it can be further processed in the Golgi. Typically the first step is trimming back the Man arms with α-mannosidases. Elongation is then initiated with Golgi-resident glycosyltransferases, such as β-GlcNAc transferase, thus beginning the synthesis of complex *N*-glycans, including bi-antennary, multi-antennary, or hybrid structures (103). The branches of these structures are often built up by addition of lactose units (Gal-β-1,4-GlcNAc-β), and can be terminated there, capped with α-sialic acid, α-galactose, or α-fucose (104).

Functions of *N*-Glycosylation

One function of *N*-glycosylation which has already been mentioned is its contribution to control of the folding of secreted proteins. *N*-glycosylation also affects intracellular vesicle trafficking, since the presence of specific glycoforms can determine how a vesicle will be transported between organelles and the cell exterior (105). The roles of *N*-glycosylation extend well beyond intracellular mechanisms. *N*-glycosylation alters the structure and function of immunoglobulins and is involved in the mechanics of cancer metastasis, cell-cell adhesions, and embryonic development (106-109).

N-Glycosylation in *C. parvum* and *T. gondii*

The *N*-glycosylation pathway has not been extensively functionally characterized in either of the organisms described in this dissertation, however, several *in silico* studies have characterized the pathways. It was first predicted that *C. parvum* has a lipid-linked *N*-glycan precursor of Dol-PP-$GlcNAc_2Man_5Glc$, due to the presence of the yeast homologs Alg7, Alg1, Alg2, Alg11, Rft1, Dpm1, Alg5, Alg6, Rft1 and Stt3, and the absence of Alg3, Alg9, Alg12, Alg8, and Alg10 in its sequenced genome (110). The observation that the genome of *T. gondii* contains the same set of predicted enzymes as *C. parvum*, with the exceptions of missing Alg11 and Rft1, and the addition of Alg8 and Alg10, suggested that the lipid-linked *N*-glycan would be Dol-PP-$GlcNAc_2Man_5Glc_3$ (110). In 2009, Cui, et al., showed that there has been evolutionary selective pressure for the *N*-glycosylation consensus sites containing Thr in the sequon as opposed to Ser, and that this selective pressure was driven by the quality control and folding mechanism, which rely upon the co-translational *N*-glycosylation machinery (111). Later, it was

reported that the genomes of *C. parvum* and *T. gondii* do not have quality control and folding machinery, since they lack the required enzymes. *T. gondii* has calnexin (Cxn), although it is noted that the gene appears to be lacking the protein disulfide isomerase binding domain, and the remaining sequence is quite divergent from any characterized calnexin (112). Prior to this dissertation, there have been neither definitions of the *C. parvum N*-glycosylated proteins nor any reports directly showing the structures of their *N*-glycans.

O-Linked Glycosylation

O-glycosylation is a protein modification in which saccharides are linked to the hydroxyl groups or amino acid residues, usually Ser or Thr; in rare cases, it can occur on Tyr or hydroxyproline. The process of *O*-glycosylation typically occurs in the Golgi for mucin and mucin-like glycoproteins. However, there are other classes of *O*-glycosylation where protein modification is initiated in the cytosol, nucleus, or in the ER (113). Fig. 1.4 presents a brief graphical representation of the various types of *O*-glycosylation.

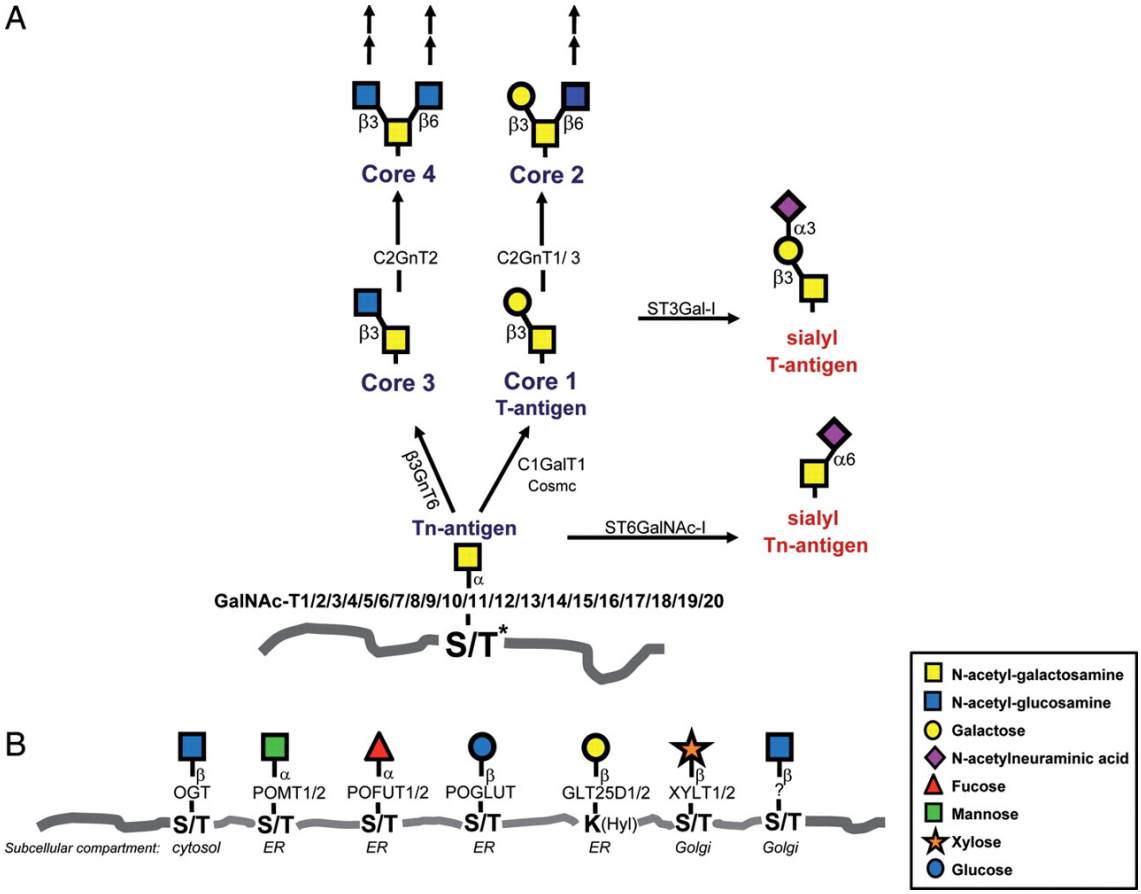


Fig. 1.4 Overview of the Different Types of *O*-Glycans

A. The initiation of the core-type *O*-glycosylation is started by polypeptide GalNAc transferases within the Golgi. If the initial GalNAc is not elongated, it is referred to as the Tn-antigen. The Tn-antigen can, in some organisms, be modified by sialic acid, creating the sialyl Tn-antigen. If a single Gal residue is added, this structure is the T-antigen; the Gal residue may also serve as the basis of the Core-1 structure. B. Other types of *O*-glycosylation are initiated by the transfer of different monosaccharides by their respective glycosyltransferases, which can be located in various intracellular compartments. (Reproduced, with permission, from Bennett, et al., 2012).

Polypeptide *N*-Acetylgalactosamine Transferases

The largest and most evolutionarily conserved group of enzymes is the GalNAc transferases, also referred to as polypeptide GalNAc transferases (ppGalNAcTs). These enzymes modify mucins or mucin-like proteins (114, 115). They utilize sugar nucleotides as sugar donors, most often UDP-GalNAc, and transfer the GalNAc to Ser/Thr in the

Golgi (116-118). The ppGalNAcTs have very similar topologies and functional domain structures. Each contains an N-terminal membrane anchor sequence, a central GalNAc transferase domain and a C-terminal Gal/GalNAc-lectin (ricin) domain (119). To date, 20 different human ppGalNAcT enzymes have been described. These can be subdivided into classes and subclasses, based upon their molecular phylogenetic order (114). The 20 human ppGalNAcTs and their similarities are shown in Fig. 1.5.

Modification of the mucins and mucin-like proteins appears to be a processive procedure, requiring an initial glycosylation event by certain subclasses of GalNAcTs (120). The subsequent binding to the same or additional GalNAc residues on the already *O*-GalNAc modified polypeptide is mediated through the GalNAc-lectin domain. The binding provides a means for anchoring down the ppGalNAcTs so the catalytic domain can modify free Ser/Thr to yield a high density of *O*-glycosylation on the polypeptide (121). Interestingly, the different GalNAcT subtypes appear to have preference for the direction in which they can add GalNAc to an available Ser/Thr, with respect to the initial Ser/Thr-*O*-GalNAc. It was found that T1, T2, T14 had preference for N-terminal addition; T3 and T6 preferred C-terminal addition; T5, T13, T16 could favor either; no preference could be determined for T4, T7-12, T15, T17-20 (122).

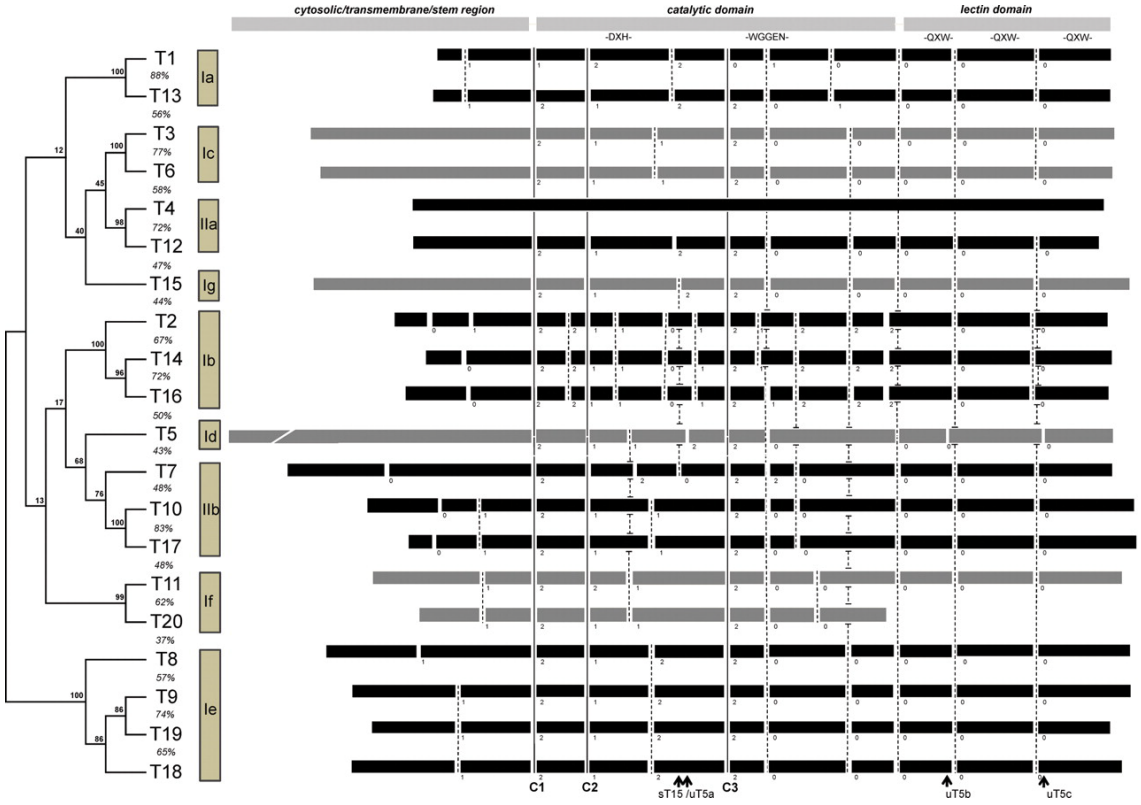


Fig. 1.5 The 20 Human GalNAc Transferases Displayed as a Phylogenetic and Genomic Analysis to Illustrate the Classes of Transferases.

Fig. 1.5 Displayed is a comparison of twenty human ppGalNAcTs. The left side of the figure shows a phylogenetic tree, to help characterize and classify the groups and sub-groups of ppGalNAcT. Common amongst almost all of the enzymes are an N-terminal cytosolic-transmembrane region, a central catalytic domain, and a C-terminal Gal/GalNAc lectin domain. (Reproduced, with permission, from Bennett, 2012)

O-N-Acetyl-Galactosamine Transferases of *C. parvum* and *T. gondii*

In the context of the two parasites described in this dissertation, *C. parvum* and *T. gondii*, there have only been a few studies aimed at identifying the ppGalNAcTs. Most of these studies have identified potential ppGalNAcTs in these organisms at a genomic level; some functional studies have been performed. Functional characterization of all the putative ppGalNAcTs in these organisms have not been fully described.

For *Toxoplasma gondii,* five ppGalNAcTs have been partially functionally characterized; these have been shown to modify mucin-like protein sequences. However, mucin like-proteins *per se* have not been described in the literature (123). A recent study showed that knockouts of ppGalNAc-T2 and T3 in *T. gondii* are responsible for the glycosylation of a tissue cyst protein, as determined by loss of VVL and Jacalin-positive lectin stains, lectins specific for Gal/GalNAc, as observed by fluorescent microscopy and Western blots (124).

There has only been a single study published describing the ppGalNAcTs in *Cryptosporidium.* Four ppGalNAcTs were identified, but only preliminary work has been performed to characterize them (125). The RNA levels for the four ppGalNAcTs (1-4) were monitored at three time points in an infection model. The results showed that T1 and T4 transferases are high at initial infection; after 48 hours only T3 expression is high while T2 is at a low level; only low levels of T2 and T3 can be observed after 72 hours (125). In addition, it was reported that a *C. parvum* lysate was capable of transferring a ^{14}C-labeled GalNAc from radioactive UDP-GalNAc to synthetic peptide substrates (125).

Mucins and their *O*-Glycans

By definition, mucins, and mucin-like proteins always have GalNAc as the first *O*-linked sugar residue; the GalNAc can remain unmodified or additional monosaccharide residues can be attached during passage of the glycoprotein through the Golgi (126). The *O*-linked glycan(s) can be developed into linear or branched structures (see Fig. 1.6 for a review) (127).

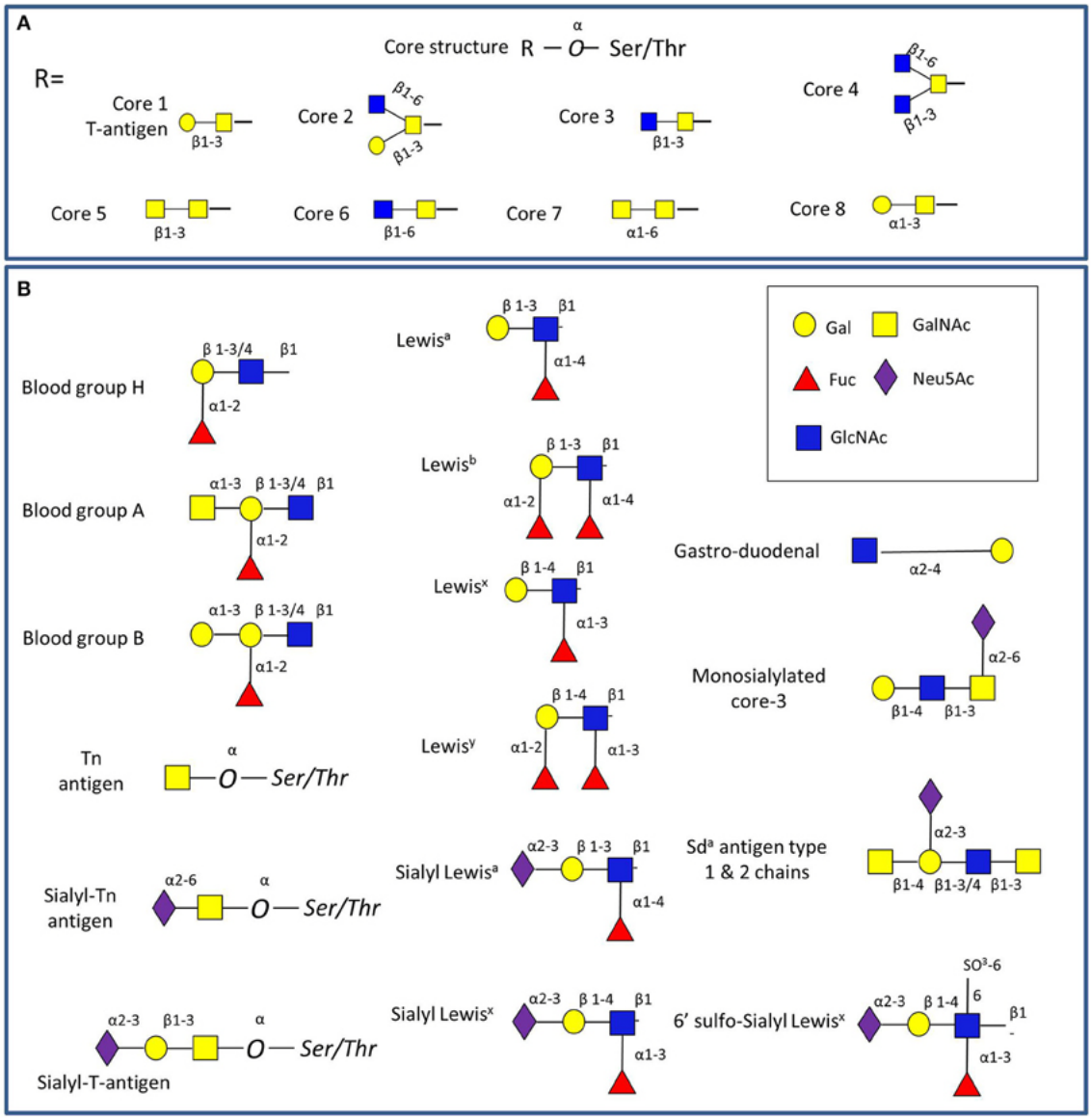


Fig. 1.6 *O*-Glycan Structures

A.) The eight *O*-glycan core structures are shown. B.) Epitopes found on *O*-glycans, and those often encountered in the human mucosa. (Reproduced in accordance with the Creative Commons Attribution License (CC BY), https://creativecommons.org/licenses/by/4.0/, no changes made other than replacement of the figure legend, obtained from Tailford, 2015)

There are eight *O*-glycan core structures (Core 1 - Core 8) and several other important *O*-glycan structures such as the Tn antigen, T antigen, and the Sialy-Tn/T, which are displayed in Fig. 1.6. All of these *O*-glycans are initiated with the addition of a

GalNAc as the first monosaccharide (128). Addition of *O*-Glycans can have varied effects, with the result that the physical properties of mucin proteins can be altered quite dramatically. Mucin-like proteins often can have massive hydrodynamic volumes due to the extent of modification. Dense *O*-glycosylation, occurring at low complexity sequences with disordered secondary structure, results in collective steric interactions, thereby driving the protein to take on a rigid and extended shape and preventing a compact protein conformation (129).

There are two types of mucins, secreted (MUC2/5AC/5B/6/7) and membrane-bound (MUC1/3/4/12/13/15/16/17/20/21) (130-132). The membrane-bound mucins, when abnormally regulated, have been implicated as contributors to the properties of cancerous tumors, such as their increased proliferation and resistance to apoptosis (133). Many of the secreted mucins have numerous Cys residues and are thought to form both inter- and intra-disulfide bonds, thereby creating large soluble and insoluble barriers (131). The mucosal epithelia rely upon both the secreted and membrane-bound mucins to protect against environmental challenges, including pathogens (134, 135).

Mucins in *C. parvum*

C. parvum, the organism which is the main focus of the research described within this dissertation, has its own set of mucin/mucin-like proteins; several of these have been shown to be important for invasion (Gp900 and Gp40) (15, 33, 34). Gp900, a very large mucin-like protein, contains two domains composed of long tandem Thr stretches (see Appendix 1). When sporozoites (the invasive life cycle stage of *C. parvum),* are incubated with a mAb which recognizes the polypeptide between the two Thr rich

domains, it prevents invasion of cells in an *in vitro* model (33). The protein Gp900 has been demonstrated to be very immunogenic in cows. However, this response may not depend on *O*-glycan modifications; much of the observed antigenicity has been attributed to the *N*-glycans, as PNGase F nearly eliminated reactivity for three monoclonal antibodies raised against it (31).

O-Linked *N*-Acetylglucosamine

Another important type of *O*-glycosylation is *O*-GlcNAcylation, the addition of a single GlcNAc residue at Ser or Thr. This modification was first detected on the surface of lymphocytes, where it was discovered to be the first sugar conjugated to the hydroxyl of Ser/Thr in a ^{3}H-Gal oligosaccharide transferase assay (136). This group of researchers explored their novel finding in greater detail, utilizing the same assay to probe subcellular fractions from varied sources, and they found that the highest concentration of *O*-GlcNAc-modified proteins resided in the nuclear fraction, more specifically on the nuclear envelope subfraction (137). The localization was further investigated, resulting in the discovery that the cytoplasmic face of nuclear pores and the lumen of the nuclear pores were the regions modified with *O*-GlcNAc (138-140). Extensive further studies have explored the biochemical pathways and functional roles of the *O*-GlcNAc modification. Utilization of UDP-GlcNAc as a substrate revealed that *O*-GlcNAc transferases (OGT), located both in the cytosol and the nucleus, are responsible for the addition of GlcNAc to protein acceptors (141). Most eukaryotes appear to have the genes encoding for OGT, suggesting that they may serve a vital role in the cell; furthermore, many proteins involved gene regulation and nuclear transport are modified in a dynamic

manner based upon the activity of an *O*-GlcNAcase (OGA), suggesting that *O*-GlcNAcylation may be a signaling mechanism (142). It has been demonstrated that the dynamic nature of *O*-GlcNAc modification is in balance with phosphorylation, indicating that it may serve as a switching mechanism (143).

O-GlcNAc Modification of Nuclear Pore Proteins

When it was first described, the *O*-GlcNAc modification was found to be most pronounced in the nuclear periphery; it was later pinpointed to the nuclear pores (137, 138). These initial observations were explored in greater detail, with the results that the nuclear pore protein NUP-62 was found to be modified with GlcNAc, and an OGT was determined to be responsible for this transfer of the GlcNAc (144). It has been realized that many of the identified nuclear-pore proteins are *O*-GlcNAc modified, with almost all of the FG-repeat containing NUPS found in cytoplasmic, nucleoplasmic regions, and within the lumen of the pore containing *O*-GlcNAc (see Fig. 1.7) (145).

The FG-NUPS are proteins associated with the nuclear pore complex that contain repeating amino acid patterns of (FxFG, GLFG, PSFG, or FG). These proteins are disordered, with the GLFG repeat-containing FG-NUPS being more disordered than the others (146, 147). In *S.* cerevisiae, it has been found that the GLFG repeat-containing FG-NUPS are positioned in the orifice of the nuclear pore, and they are essential for viability, while the FG repeats on the cytoplasmic or nucleoplasmic sides could be deleted without noticeable deleterious effect (148). The disordered regions of the FG-NUPS extend out and into the nuclear pore where they interact with karyopherins, proteins that facilitate

nuclear import/export and karyopherin-cargo complexes. It is thought that the FG regions are the crux of these interactions and confer specificity for transport into and out of the nucleus (149, 150). *O*-GlcNAcylation has not been reported for the organisms investigated during the research conducted for this thesis, but the growing appreciation for its importance mandated that its involvement should be considered. In addition, the aforementioned targets of *O*-GlcNAc modification, (*i.e.,* FG-NUPS and nuclear-periphery localized proteins), were encountered during the research conducted for this thesis, albeit *O*-fucosylation was observed as a modification instead of *O*-GlcNAcylation.

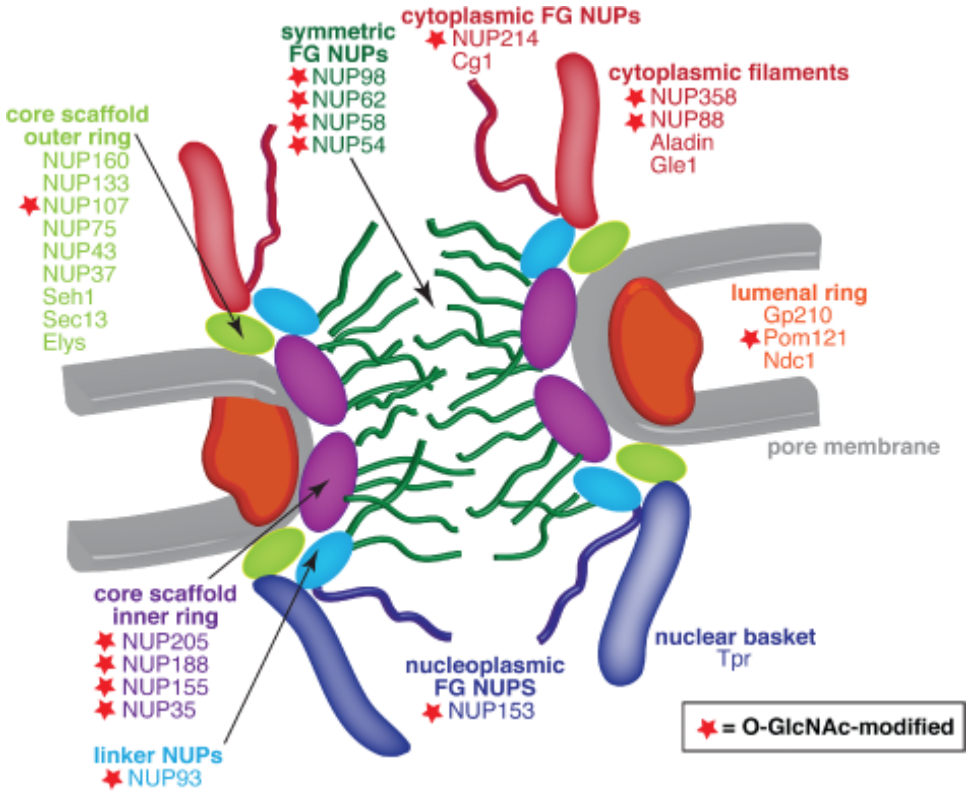


Fig. 1.7 *O*-GlcNAc Modification of Nuclear Pore Proteins

Fig. 1.7 The nuclear pore proteins described in mammals have been shown to be *O*-glycosylated with β-*N*-acetylglucosamine. Modification of these proteins occurs on the nucleoporins, most often those containing FG repeat domains. The *O*-GlcNAc modification favors proteins on the cytoplasmic face or within the lumen of the pore itself. (Reproduced, with permission, from Li and Kohler, 2014).

Nuclear Transportation Mechanisms Pertaining to FG-NUPS

The interactions and relationships between FG-NUPS and nuclear transport were briefly touched upon during the introduction of O-GlcNAc modification. Supplementary details and related topics are introduced here in order to provide readers a proper introduction to some of the novel discoveries discussed in Chapter. 5.

Transportation in and out of a cell nucleus must be a tightly regulated process involving specific interactions with the karyopherins, a broad range of proteins involved in the transportation of cargo into and out of the nucleus. These proteins can loosely be grouped into two classes: 1.) Karyopherin α type (Kapα), which binds to the classical NLS signal (a single- or multi-cluster of basic amino acids); 2.) Karyopherin β type (Kapβ), which bind to specific proteins which lack the canonical NLS (Kapβ proteins can bind Kapα proteins, utilizing them as adapters to transport proteins containing a NLS) (150).

Crucial details which helped to define some of the processes involved in nuclear transport began with the identification of a small GTPase called Ran (Ras-related nuclear protein. The catalyst of this novel finding was made during the course of UV-irradiation photo-labeling experiments using [α-32P]GDP and [α-32P]GTP on nuclear envelope preparations from rat liver nuclei (151) That study identified a 28-kDa GTP/GDP binding protein, which was later found to have GTPase activity; transport of proteins containing a NLS into the nucleus was dependent upon the hydrolysis of GTP (152). The name Ran was coined due to its nuclear localization and its similarity to the conserved series of

GTPases which were named after oncogenes identified in rats, the rat sarcoma genes (Ras) (153).

It was revealed that the switching back and forth between the GTP- and GDP-bound states of Ran helps drive the nuclear import and export mechanism (154). The hydrolysis of Ran-GTP drives importation of Ran along with the proteins which bind to Ran, the Kapβ class of importins, and the cargo bound to Kapβ or its adapters (150).

A defining feature that is shared across the Kapβ and Kapα classes of proteins are some tertiary structural features of repeating domains, referred to as the HEAT motifs and ARM repeats (Armadillo as described in Drosophila). Both these features are helices exposing ideally spaced binding sites on the helices, or in between the repeats (155, 156). The ARM domain repeats are a found on Kapα, where, in between the ARM domains, there are NLS binding pockets that are defined by the WxxxN motif. Using a similar mechanism, Kapβ binds to the FG on FG-NUPS in hydrophobic grooves buried between the HEAT repeats, binding more tightly to the GLFG repeat as compared to other sequences (150, 157).

O*-GlcNAcylation in *C. parvum* and *T. gondii

To date, relatively only a few studies have described *O*-GlcNAc modifications in either *C. parvum* or *T. gondii*. Most reports have been based on indirect observations, *in silico* analysis, or, limited to preliminary functional transferase activity studies. The OGTs in both *C. parvum* and *T. gondii* were first identified from an *in silico* analysis of the genomes searching for homologous proteins to the human OGT (158). It was found that both *T. gondii* and *C. parvum* have divergent topologies when compared to the

human OGT, since they contain only two TPR repeat domains in the N-terminus (substrate recognition domains), whereas the human OGT has 11 (158). Furthermore, the functional domain at the C-terminus is a complete SPY domain (catalytic domain), but the human OGT is split by amino acid insertions (158). In the same study performed by Banerjee et al., the recombinant *C. parvum* OGT was shown to be a functional transferase capable of transferring GlcNAc from tritiated UDP-GlcNAc to the substrate casein kinase (158). To date, this remains the only study that has demonstrated the functionality of an OGT in *C. parvum*. Experiments performed in *T. gondii* have demonstrated transfer of GlcNAc onto the peptide substrate Pro-Tyr-Thr-Val-Val, using radioactive UDP-GlcNAc and a cell free lysate (159).

The Similarities of *T. gondii* and *C. parvum* OGT to the Plant OGT

As was noted by Banerjee et al., 2009, the putative *T. gondii* and *C. parvum* OGT contains a complete SPY domain and is more similar to the plant SPY domain than the human. This observation was also mentioned by Perez-Cervera et al., 2011, where they described the similarities of the putative *T. gondii* OGT and *Arabidopsis* SPY. The authors showed that GlcNAc could be transferred to proteins in a *T. gondii* crude lysate when incubated with tritiated UDP-GlcNAc; however, they did not identify any proteins, nor did they perform any experiments to evaluate the function of their putative OGT.

Due to the significant sequence similarity of the *T. gondii* and *C. parvum* putative OGT to the plant SPY domain, additional background information will be presented here in order to aid in the understanding of results discussed in later chapters of this dissertation. This information is necessary to help frame the discoveries described in

Chapter 5. While this dissertation was being written, a new publication revealed that the plant the SPY domain, also known as SPINDLY, may be a fucosyl transferase (160). The authors demonstrated that a recombinant *Arabidopsis thaliana* SPY was capable of hydrolyzing GDP-fucose in the presence of a substrate. For this, they used a colorimetric assay to measure free phosphate, but they did not present any direct evidence to substantiate their hypothesis that SPY is exclusively a fucosyl transferase. The behavior of SPY as a fucosyl transferase needs to be proven with additional non-ambiguous experiments, especially since SPY has previously been implicated as having GlcNAc transferase activity (161). The SPY protein exhibits an interesting pattern of localization in plant cells. It has sometimes been observed in the cytosol, and at other times, in the nucleus (162). The determinant for localization has not been established; however, since it is known to be involved in plant signaling processes, there is the possibility that the differential localization may be due to a downstream signaling event (162). SPY has been established as a negative regulator of the effects of giberellin (GA), a dicyclic terpenoid plant hormone produced in the plastid from polyprenols, which alters growth and development in the plant (163-165). It is interesting to note that the organisms studied throughout the course of this dissertation are in the phylum Apicomplexa, gaining this name from their secondary plastid called the apicoplast which is evolutionarily divergent from the plant plastid, the chloroplast. It has lost its ability to photosynthesize, but is essential for isoprenoid biosynthesis (166). Of the two organisms studied, *C. parvum* and *T. gondii*, only *T. gondii* appears to have retained the apicoplast (167). What is even more interesting is that *Toxoplasma* and *Plasmodium* are both known to produce some of the

same plant hormones (e.g., giberellin), and these appear to affect the growth and development in these organisms, similar to their effects in plants (168). While there has been no established connection between apicomplexa gibberellins and the SPY domain-containing transferase, or any other glycosyl transferases, the overlap with biosynthetic pathways unique to organisms that contain chloroplast or chloroplast-like plastids raises a question as to this possibility.

O-Fucosylation

O-fucosylation is a modification to proteins that has not been described extensively in the literature. The first description of fucose directly conjugated to an amino acid (as opposed to attachment via a glycosidic linkage) occurred less than thirty years ago; the substrate was human recombinant pro-urokinase, which was expressed in a mouse hybridoma cell line (169). There was suggestion that one of the Thr was modified, however, the exact location could not be verified (169). Human Factor IX was soon discovered to be *O*-fucosylated within the EGF domain, and fucose was determined to be directly conjugated to the hydroxyl of Ser, thus demonstrating that *O*-fucose was a true modification (170). In 1993, Harris et al., proposed the *O*-fucosyl transferase consensus sequence of CxxGG(T/S)C, where x is any amino acid. During the course of studying a migratory locust, *Locusta migratoria,* a neuropeptide (QISCQPGK**T**FKDKCNTCRCG) was found to contain a single Thr *O*-fucose (171). This discovery suggested that *O*-fucose modification occurs not only in mammalian cells but spans across a broad range of animals and could possibly be a universal modification that had been previously overlooked. Soon thereafter, a fucosyl transferase, which utilizes GDP-fucose as a donor,

was isolated from CHO cells using affinity purification with the peptide from the EGF domain containing the previously described *O*-fucosylation consensus sequence (172). Using a library of substrates based upon the proposed fucosyltransferase consensus sequence, these investigators determined that the two Gly residues of the sequence CxxGG(T/S)C were important for proper substrate recognition by the protein *O*-fucosyltransferase (POFuT). They further determined that, if the two Gly were replaced with amino acids that made the peptide inflexible, there would be no fucosylation, suggesting that the two Cys residues need to form a disulfide bridge to properly present the Ser/Thr to the POFuT. The modification of the *Locusta migratoria* neuropeptide also occurred within a stretch of amino acids in between two Cys, suggesting that there may be a common mechanism requiring a tight loop to be presented to a POFuT. They determined that Cys position was indeed crucial for modification with *O*-fucose. In an experiment that utilized the recombinant *Drosophila* POFuT1, Luo, et al. later found that thrombospondin type 1 repeats (TSR 1), $Cx_{(2–3)}(S/T)Cx_2G$, similar to EGF repeats, $Cx_{(4-5)}(S/T)C$, could be modified with *O*-fucose, but POFuT1 was not the enzyme; they noted that the similarities between substrates was evident, but there was another enzyme, not yet identified, which modifies TSR repeats (173). Later, with GDP-fucose utilized as a donor, the human enzyme, POFuT2, was revealed to be the enzyme which modifies TSR type 1 repeats; like POFuT1, this protein has to be folded correctly with disulfide bridges to present the substrate to POFuT2 (174).

There have also been a few reports of nuclear proteins or nucleus-localized proteins in plants. The best example showed modification of proteins localized at the

nuclear periphery in *Daucus carota* protoplasts, as found by a positive stain with Ulex europaeus agglutinin 1 (UEA-1) a fucose-binding lectin (175). Specificity was verified when the binding of the lectin could be abrogated by direct competition using fucose; unfortunately, there have not been any protein identifications resulting from this or any other studies (175).

Mass Spectrometry

Mass spectrometry has become a very useful and informative tool for biomedical research. It has undergone enormous technological advancement since it was first developed into an analytic instrument. Modern mass spectrometric techniques have application in the fields of genomics, proteomics, and glycomics (176-178). This section will briefly touch upon the fundamentals of mass spectrometry and provide a more detailed outline of the techniques and methodologies as they pertain to this dissertation.

Fundamentals

Modern mass spectrometry can trace its roots to the studies of charged particles described in late 19th Century and early 20th century. The physics which dictate how charged particles behave in electric or magnetic fields are the core of how all mass spectrometers function. Studies of cathode ray tubes helped to initiate the development and ideas surrounding mass spectrometry instrumentation (179). This area of research developed rapidly in the 20th century, moving from recordings made with phosphor screens and photographic plates, to oscilloscopes, leading to modern day, high throughput instruments with unparalleled speed and accuracy. The block diagram in Fig. 1.8 shows the general schema of a mass spectrometer.

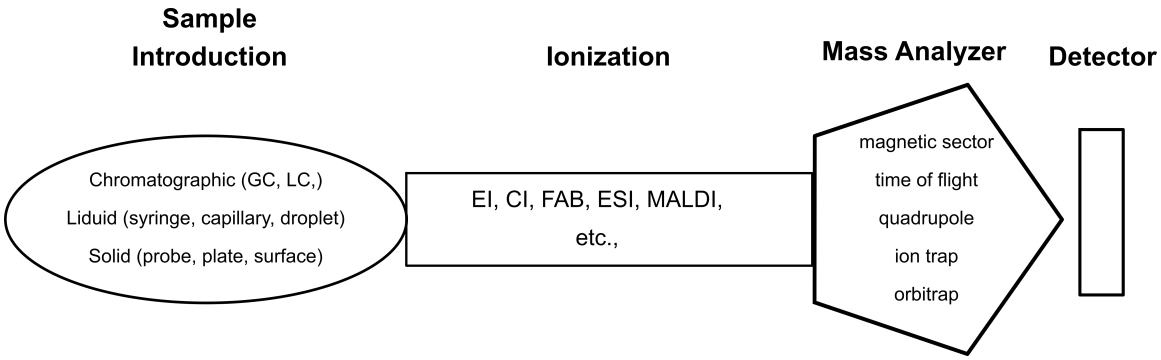


Fig. 1.8 Block Diagram of a Mass Spectrometer

This block diagram shows the essential components of a mass spectrometer. The diagram is simplified for clarity, and does not include MS/MS, or the options for fragmentation ; these will be described later. Hybrid instruments combine mass analyzers of different types.

Understanding the terminology used in mass spectrometry is the first step towards understanding what a mass spectrum is, how to properly interpret it, and what information is needed for good experimental design. A mass spectrum is a measurement of mass-to-charge ratios (*m/z*), the accuracy of which is determined by the characteristics of the mass analyzer(s), but in a given experiment, it also depends on how well an instrument is calibrated with standards, and the reproducibility of the measurements (180). Mass accuracy, or, mass error, is often reported as the parts-per-million (ppm) error from the calculated theoretical value. The deviation can easily be calculated from the experimentally observed *m/z* and the calculated *m/z* (see equation 1.1).

$$\text{mass accuracy (ppm)} = \left| \frac{\left((\text{observed } m/z) - (\text{calculated } m/z) \right)}{(\text{calculated } m/z)} \right| \text{x } 10^6 \qquad \text{Equation 1.1}$$

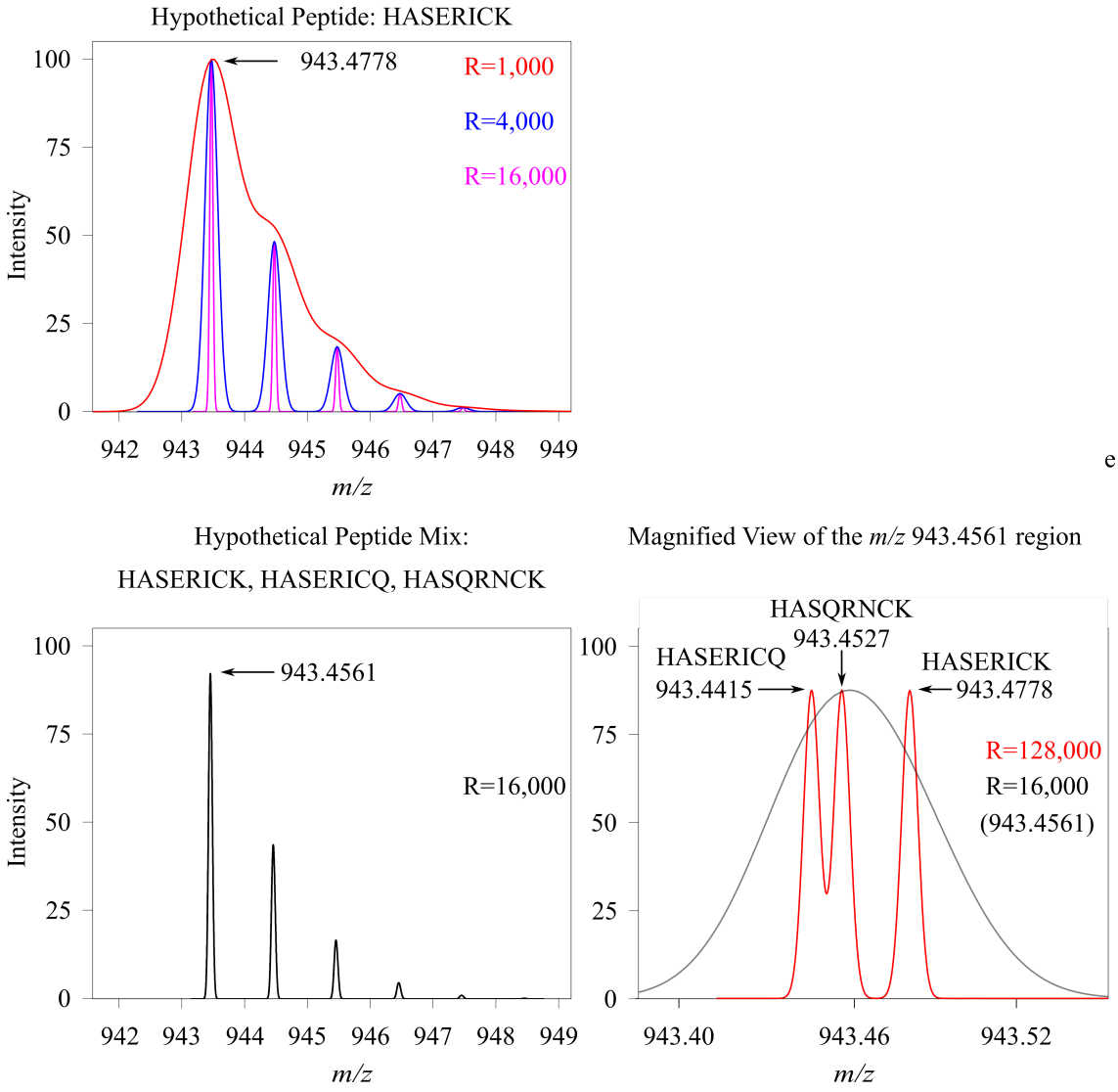
Mass accuracy is instrument-dependent. Instruments that are capable of extremely accurate measurements are indispensable tools for determining chemical formulae (181). FT-ICR MS instruments can routinely make measurements with mass accuracies around 1 ppm, and if both internal and external standards are used, the mass spectrum can be

post-processed to achieve sub-ppm mass errors (182). Recent instrument designs, such as the 21-T FT-ICR mass spectrometers that have recently been constructed, can routinely make measurements with errors in the range of 50 - 200 parts-per-billion (ppb), depending upon the *m/z* scan range and scan time (183, 184).

The resolving power of an instrument, and, resolution, are two different terms used to describe instrument performance and to describe the observations within a spectrum. Resolution, is used to describe an experimental measurement, defining how two peaks of equal height are separated by a valley of a set value, (e.g. 10%, 50%). The resolving power of an instrument, describes an instrument property, as it pertains to a single peak (185). Resolving power, defined as R, is calculated from the full width half maximum (FWHM) of the measured peak (see Equation 1.3).

$$R = \frac{m}{\Delta m}, \text{ at FWHM} \qquad \text{Equation 1.3}$$

Resolving power of an instrument is critical to for accurate identification of analytes. This becomes increasingly more important with an increase of sample complexity and concurrent existence of multiple charge states in a spectrum (186). A simple example is provided in Fig. 1.9 to demonstrate the effects of resolution, and how having an instrument with higher resolving power helps differentiate the components within a mix of hypothetical peptides. It also illustrates a possible pitfall in interpretation, if there is a measurement error due to the presence of unresolved components which could fall within the expected error tolerances (see Fig. 1.9).



e

Fig. 1.9: Mass Resolution and Resolving Power.

The effects of mass resolution on peak shape are shown in the top panel of the figure. A series of peaks recorded with different mass resolutions shown for the for the hypothetical peptide HASERICK, which has a mono-isotopic *m/z 943.4778* for the $[M + H]^+$ ion. Shown are theoretical resolutions of 1,000 (red), 4,000 (blue), and 16,000 (fuschia). The lower resolution of 1,000 is inadequate for most applications, where the isotope distribution can not be observed. The lower two panels show a mixture of three peptides (HASERICK, HASERICQ, and HASQRNCK), all represented as the $[M + H]^+$ ions. The lower left panel shows the sum of all three peptides with a resolution of 16,000. As shown, the apex of the apparent mono-isotopic peak is reported as *m/z* 942.4488, which would be -23.06 ppm error if it was assumed to be the peptide HASERICK. The lower right panel, shows the same peak with R=16,000 (black), and with R = 128,000 (red). The higher resolution of 128,000 allows for the resolution of all three components, revealing that the perceived error of -23.06 ppm was due to the summation of three separate components, which consequently broadens the first peak, with a FWHM of 0.0712, revealing a resolution of 13,241. All theoretical isotopic patterns were calculated using the program mMass version. 5.5.0 (187)

Sample Introduction Into a Mass Spectrometer

Prior to the ionization step, the sample must be carefully prepared to avoid the introduction of salts or other contaminants which may otherwise hinder the analysis, and can even abrogate detection (188, 189). As portrayed in Fig. 1.8, the sample must first be introduced to the ion source. The method of introduction is dependent upon which type of ion source is used and the instrument used for making the measurements. The most commonly used methods for sample introduction are as dried spots on a surface plate (for Matrix-assisted laser desorption/ionization (MALDI)); liquids, either directly infused or eluting from a liquid chromatography system (for electrospray ionization (ESI)/ nanoflow electrospray ionization (nano-ESI)), or in the gas phase from a gas chromatograph (for electron impact (EI)).

Ionization Techniques

Matrix-Assisted Laser Desorption/Ionization (MALDI)

Matrix-assisted laser desorption/ionization is considered to be a “soft” ionization technique, meaning that there is minimal fragmentation of the molecular ion. It is an ionization method which is suitable for a wide range of biological analytes, including carbohydrates, peptides, proteins, nucleic acids, and lipids (190-195). Typically, for measurements to be made in the positive-ion mode, the sample and the matrix are dissolved in a mix of aqueous and organic solvents along with a volatile acid, such as formic acid, to ensure the solution is below the pKa of the matrix, and ~0.5 - 1 μL volumes of each, or a mix of sample and matrix are spotted onto a metal plate and dried. This ionization technique has seen applications for the analysis of macromolecular

complexes and even whole bacterial cell walls or of intact oocysts from *Cryptosporidium* species (196-198). There is limited information to be gained from the analysis of intact cells other than the observation of signature *m/z* values for a particular microorganism acting as a fingerprint for identification (199). Microbial identification using MALDI-TOF MS is now commonly used for the diagnosis of infectious agents and is currently in use in leading medical centers and hospitals worldwide. Initial approaches based on lipid or peptide patterns raised questions of reproducibility and reliability which hindered its implementation (200, 201), but the development and implementation of methods that depend on capsid protein databases has proven extremely successful (202-204). An approach for the analysis of intact organisms has been demonstrated with the analysis of *Bacillus* spp. spores using a MALDI-TOF/TOF mass spectrometer (205). This type of instrument is capable of fragmenting intact proteins, and this provides some amino acid sequence information, a process called top-down protein sequencing. Typically fragmentation is most efficient near the N- and C-termini, and is blocked if there are disulfide bridges, so this method works best if the disulfide bonds are reduced and the free cysteines are alkylated (206, 207).

The idea of utilizing a matrix to assist in the ionization of analytes traces its origin to laser desorption studies of amino acids. During one desorption ionization study of amino acids and dipeptides, the third and fourth harmonics of a neodymium-doped yttrium aluminium garnet (Nd:YAG) laser (208) with a λ of 355nm and 266nm and respectively, were used to irradiate a metal target with the dried amino acids and dipeptides (209). It was found that aromatic amino acids that absorbed the shorter

wavelength could be observed as the $[M+H]^+$, with a relatively low level of irradiance, however, the same aromatic amino acids which did not absorb the 355 nm wavelength, required much higher irradiance, and could be observed as $[M+Na]^+$ or $[M+K]^+$, or as protonated fragments of the amino acid, just at the threshold of ionization for the alkali metals (209). The true breakthrough which Karas et al. made was the observation of an aliphatic amino acid as the $[M+H]^+$ ion when analyzed in a mixture with an ultraviolet-absorbing amino acid. The observation of the non-absorbing amino acid at a level of irradiance required only for the aromatic amino acid, was described as "matrix-assisted laser desorption" (209). Observation of this phenomenon led to the development of an array of MALDI matrices and providing a means for ionizing a wide range of biomolecules, using low levels of UV or infrared (IR) irradiation well below destructive power levels (210). The choice of MALDI matrix, the solvent composition, the presence of salts or detergents, and the sample complexity can all affect the ionization efficiency of the analytes, and therefore care must be taken during sample preparation (211). The matrices which are most commonly used for the analysis of peptides, proteins, or oligosaccharides are 2,5-dihydroxybenzoic acid and sinapic acid (190, 212).

The mechanisms describing the ionization process during MALDI have been widely discussed in the literature, however, there is no unified model to describe the mechanisms for generation of all the ion types observable in a MALDI-MS experiment – likely multiple processes occur in even a single experiment (213-215). The most simplistic model is that a photon is absorbed by the matrix and, while in the excited state, the matrix molecule becomes much more acidic and transfers a proton to an analyte,

together with the charge. However, this model doesn't fully explain why the same matrix can be used in both the positive and negative modes (213). Irradiating the matrix/analyte mixture causes rapid ablation of the surface, rapidly accelerating clusters of matrix and analyte away from the plate surface. There may be thermal component to charge transfer, as well as charge-charge secondary and accumulative reactions which can lead the formation of ions that cannot be explained by the simplistic absorptive model (215). In IR MALDI-MS, localized heating occurs, and large chunks of the sample surface are often ablated (216)

Electrospray Ionization (ESI)

Electrospray ionization has been described as another "soft" ionization technique. It came into use around the same time as MALDI (217). Although early work on ESI was performed in Russia and published in 1984 in a Russian language journal, the work did not become known at the time, largely because of the political situation, an English translation of the earlier report was published in 2008 (218). The sample dissolved in a volatile liquid that flows through a conductive needle terminated at finely tipped cone that is held at a voltage potential in the low-kilovolt range, relative to the inlet of the mass spectrometer (219). The finely-tipped needle sustains a very large charge density and, when the liquid exits the highly charged orifice, it picks up charge, exploding into plume of droplets from Coulombic repulsions. A heated neutral gas aids in the de-solvation process, yielding charged molecules which can then be manipulated through the mass analyzer (217) (See Fig. 1.10).

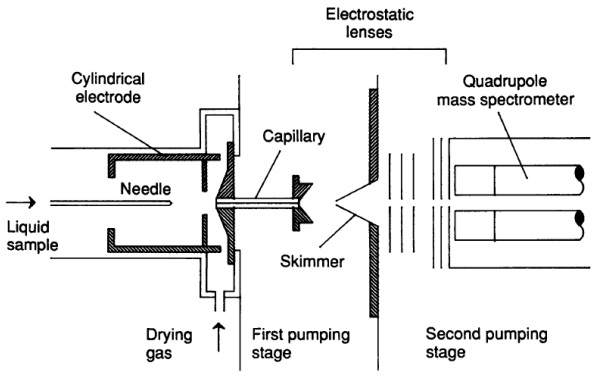


Fig. 1.10 Schematic of a Typical Electrospray Ionization (ESI) Setup.

At the left is the entry point for the liquid sample, with a heated drying gas to aid in desolvation. A high voltage potential is applied between the exit of the needle and the capillary inlet, where ions are transferred into the mass spectrometer. A skimmer is in place to only allow ions to enter into the mass analyzer section, excluding any neutral contaminants. (Reproduced, with permission, from Fenn et al., 1989).

While MALDI-MS has certain advantages of ease of use and in high throughput screening of many samples, it is not as flexible as other techniques, since it cannot be readily coupled to a separation instrument (e.g. HPLC) as is possible with ESI (219). Direct coupling of a separation technique can only be achieved with laboratory built equipment and has never advanced beyond proof of concept demonstrations Another advantage of ESI is that ions are usually observed as multiply charged ions (220). This means that the detection and analysis of larger molecules can be carried out with mass analyzers with limited *m/z* range (221). It should be recognized, however, that chromatographic eluents can be deposited onto MALDI plates (222-224). The subsequent MALDI-MS analysis is much more rapid than the online separation and may be repeated with the same target, since only a tiny fraction of the analyte spot is ablated in a single experiment(225). For proteins and other larger analytes, MALDI (especially IR MALDI) also produces abundant multiply-charged ions (216, 226-228).

Nanoflow Electrospray Ionization (nanoESI)

As the name implies, nano-ESI is electrospray ionization performed using much lower liquid flow rates. The nano-ESI needles have much finer tips (one to two microns) and nano-ESI flow rates range from around 9 nL/min. to 500 nL/min. (229, 230). The decreased flow rate boosts ionization efficiency of the analytes, and therefore sensitivity increases (231). In mixtures, the signals from detergents such as *n*-octyl-β-D-glucoside, and other surface-active species, increase in total ion intensity as flow rates increase, and they can suppress signals from the analytes of interest. However, when the same solution is sprayed at lower (nL/min.) ranges, the signal of interest goes up, demonstrating that the decreased flow rates of nano-ESI can help overcome some of the ionization suppression effects from competing species (232). This realization spawned the idea of using an HPLC attached to a split-flow device fitted with a nanoESI emitter which was coupled to a mass spectrometer (233). This setup allowed for simultaneous semi-preparative scale separation on a liquid chromatography system with the efficient and sensitive detection capabilities of nanoESI (233). Although HPLC separation coupled to a split-flow device can be very informative, it is only feasible large amounts of a sample are available for separation on the standard-flow HPLC. Obtaining such large quantities is not always feasible; certainly this is the case for clinically obtained human samples, environmental samples, and samples from microorganisms which cannot be grown in the lab. Therefore, methods were developed to place chromatography medium inline, just prior to the nanoESI emitter, so that minute quantities of material could be separated with nL/min

flow rates. This strategy provided a means to efficiently separate and analyze peptides generated from nanogram quantities of protein (234).

Commercially available UPLC systems that are now available use smaller diameter separation media (<3μm) to increase the surface area and therefore boost separation efficiency. In addition, these system used capillary scale (<200 μm) inner diameter columns so smaller samples can be separated. Both the smaller diameter resin and smaller tubing increase back pressure and therefore the UPLC systems operate at lower flow rates (<1μL/min), and thus, these systems are perfectly suited for direct interfacing to a nanoESI source (235-237). With the advent of commercially available UPLC systems came the ability to perform high quality shotgun proteomic analyses, because they offer the capability to efficiently and quickly separate peptides generated from minute quantities of proteins (238, 239). Reliable quantification should not be assumed, due to a variety of effects such as ion suppression of the other components which may be present in the sample matrix (240).

Electron-Impact Ionization (EI)

Electron impact ionization (EI), also known as electron ionization, is an ionization technique wherein an electron beam collides with a sample in the gas phase to generate radical ions (241). The source of electrons is provided by a heated filament, either rhenium or tungsten; the emitted electron energy levels are modulated by adjusting the voltage supplied to the filament (see Fig. 1.10). The typical electron beam energy is 70 eV; lower energies may be used to decrease sample fragmentation – but at the cost of substantially reduced sensitivity. Electrons emitted from the filament are collimated into

a beam using permanent magnets so that the beam intersects the flow path of the sample (242). EI is considered to be a hard ionization technique, meaning that most analytes fragment upon ionization, leaving little or no intact molecular ion. Positive-mode ionization occurs when an electron with sufficient energy collides with the molecule, knocking out an electron and generating a cationic radical (see Equation 1.3) (241, 243). This phenomenon is observed when the electron colliding with the molecule has sufficient energy, expressed in electronvolts (eV), to reach the ionization potential of the molecule and the molecule has a large enough cross section for an impact to be probable (244). It is also possible to generate multiply-charged radicals, but this typically occurs with much higher energies or with atoms and gases (245).

$$M + e^- \rightarrow M^{+\bullet} + 2e^- \qquad \text{Equation 1.3}$$

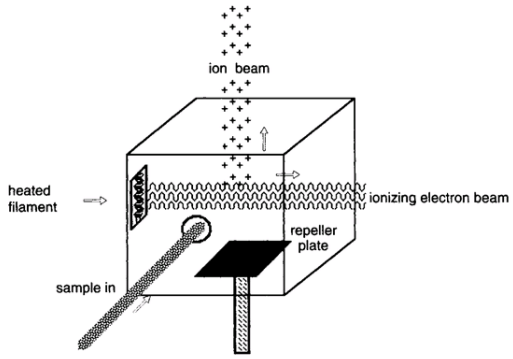


Fig. 1.10 An Electron Impact Ionization Source

The cartoon schematic shows a simplified view of an Electron Impact (EI) source. Electrons are generated from a heated filament. Permanent magnets help to collimate the beam of electrons, so that it interacts efficiently with sample that is introduced in the gas phase. (Reproduced, with permission, Siuzdak, 1996)

Within the scope of studying large or fragile biomolecules, this method of ionization has limited applications, but has great utility for studying such analytes as steroids, lipids and drugs and their metabolites (246-248). One of the limitations of EI is

that the sample needs to be introduced in the gas phase under vacuum conditions of 10^{-6} Torr or less, so this technique is best suited for GC-MS applications of volatile samples (242, 249). In order to analyze large or fragile biomolecules, especially those which are polar, or bear free amine, carboxylic acid, or hydroxyl groups, the samples must first be derivatized to allow the samples to be vaporized without decomposition, and to make them volatile, allowing for gas phase introduction. Although some established derivatization methods are highly efficient, even for small amounts of sample, other derivatization methods can involve lengthy procedures not always suitable for trace amounts of analytes (250, 251).

Mass Analyzers

Quadrupole Mass Analyzers

Quadrupoles can be used in mass spectrometers for both storage of ions and for mass selection. Linear quadrupole mass filter and three dimensional (3D) quadrupole ion traps (QIT) are based on similar principles, but with different designs, properties, and operational parameters. Typically, a 3D ion trap has higher resolution than a linear ion trap; however, a linear ion trap has a higher storage capacity (252). The linear quadrupole, in its simplest form is a set of four parallel rods (see Fig. 1.11) which have both direct current (DC) and radio frequency (rf) voltages applied to opposing pairs of the poles.

In a linear quadrupole, the rf is applied to alternate sets of coupled poles, inducing oscillations of the ions in both the x- and y- planes. The ions oscillate in both the x and y planes, differently for each m/z value. When specific combinations of the rf frequency, rf

voltage, and DC offsets are applied, ions of discrete m/z values have stable trajectories along the z-axis and continue through to the detector (253). A linear quadrupole, becomes a linear QIT when there are end electrodes in place, which can be held with a DC bias, trapping the ions in the central axis rf quadrupole field (254).

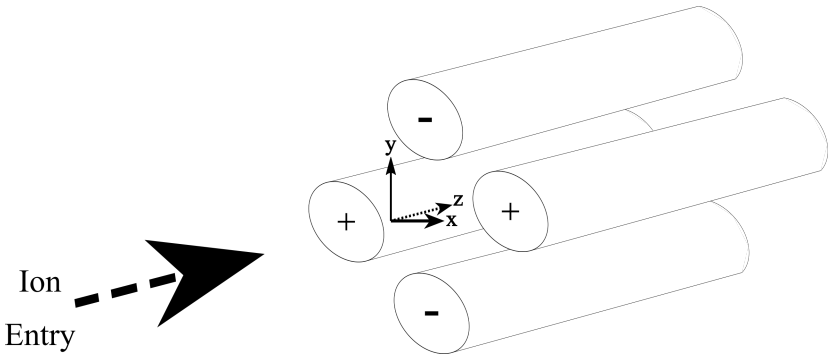


Fig. 1.11 Quadrupole Schematic

This Cartoon Schematic Shows a Linear Quadrupole Setup.

A quadrupole device may be operated in two ways: 1.) ions may be pushed through the quadrupole by simultaneously scanning both rf V and DC V, keeping the ratio of rf V to DC V constant, while the rf frequency remains fixed, or, 2.) scanning the rf frequency, while the rf V and DC V are kept constant (255). Quadrupoles can be stand-alone analyzers in an instrument (as single, double, or triple quadrupoles), or, they are utilized as part of a hybrid instrument for transient trapping of ions, or as part of a mass filter (252). Three dimensional ion traps have a distinct advantage over linear ion traps in that ions of opposite charge can simultaneously be stored in the trap, since 3D ion traps manipulate ions using radio frequency, a property which allows for ion-ion reaction monitoring and the performance of efficient electron transfer dissociation (ETD) reactions (see Fig. 1.12) (256-258). Furthermore, the compact three dimensional design is

more accommodating towards hardware modifications, such as retrofitting them with a He cation source to perform charge transfer dissociation experiments in the ion trap by boring a hole through the end caps (259). Three dimensional ion traps are very efficient at performing multiple sequential MS/MS experiments, or, MSn experiments, due to the ability to selectively eject all ions except those of interest using rf filtering. This is not a unique attribute for 3D traps, as 2D linear traps can perform similar experiments. However, there are extended capabilities and benefits with a 3D trap which have already been mentioned. MSn provides a way to selectively pick apart an analyte of interest, tracking where fragments originate, following multiple fragmentation pathways, a retrograde reconstruction from each MS/MS stage provides a detailed map of the original structure (260).

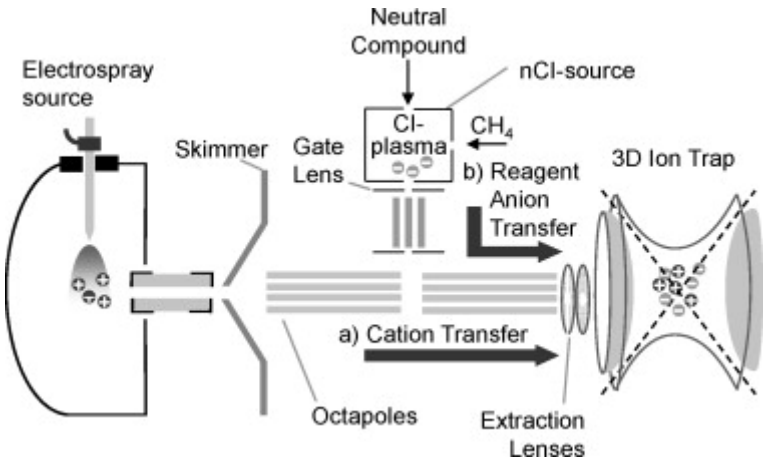


Fig. 1.12 A Schematic Overview of a 3D Quadrupole Ion Trap Mass Spectrometer.

This instrument is capable of trapping both positive and negative ions in the 3D trap, utilizing a separate negative chemical ionization source and a split multipole lens system (Reproduced with permission from Hartmer, et al., 2008).

Time-of-Flight Mass Analyzers

Time-of-flight (TOF) mass analyzers, as the name implies, measure the amount of time an ion traverses a known distance to calculate its *m/z.* Conceptually, this design is

the most straightforward type of mass analyzer. The underlying principles that define how this mass analyzer works can be described with Newtonian physics and a few simple equations; these provide a model which is easy to comprehend. The key to a functional TOF MS is the controlled introduction of ions in discrete packets so that their times of flight can be accurately measured. This is one reason why instruments that produce a continuous stream of ions must be configured to produce ion packets, such as with hybrid quadrupole time-of-flight (Q-TOF) mass analyzers. MALDI is a pulsed ionization mode and therefore can be mated directly with a TOF mass analyzer. However, irregularities in the sample surface may result in the generation of analyte ions with a broad energy spread that cannot be fully compensated even when a reflectron is employed. In these cases, MALDI-MS will produce better results when an orthogonal configuration is used. In a Q-TOF MS, the quadrupole is perpendicular to the flight tube, so ions can be injected as discrete packets into the flight tube, by controlling a stack of high voltage ion lenses that direct the ions orthogonal to their original trajectories and into the flight tube (see Fig. 1.14). A MALDI-TOF instrument controls the ion formation. The ions are generated by precisely timed pulses of a laser. The pulsing of the ion extraction electric field follows the timing of the ion generation. A slight delay in extraction may be imposed to minimize the energy distribution within the ion packet. The ion extractor accelerates the ions at a selected kV potential. This ensures that the amount of force exerted onto the ions is known; it is equal to the electric field (E) times the charge (z) (see equation 1.5). The ions are accelerated through the ion extractor and sent down the flight tube. It is assumed that all the ions will have the same kinetic energy (E_k), which is equal to the

force exerted from the accelerating electric field (see Equation 1.4). Ions with different masses, assuming the charges are the same, will have different velocities as they enter the flight tube. This should be equal to the square root of the reciprocal mass ($m^{-1/2}$), since (z) and (E) would be equal. Therefore, combining equations 1.4 and 1.5, solving for the time of flight (t), and using the flight tube length (L) for the distance traveled for the velocity function, results in equation 1.6. Thus, measuring the time of flight (t), with a known length of flight (L), and a known accelerating electric field (E), the *m/z* can be determined (see Equation 1.7; also refer to Fig. 1.13 for a schematic representation).

$E_k = \frac{1}{2}mv^2 = zE$.....Equation 1.4

$zE = mv^2$.....Equation 1.5

$t = \frac{L\sqrt{m}}{\sqrt{2zE}}$ .....Equation 1.6

$\frac{m}{z} = 2E\left(\frac{t^2}{L^2}\right)$.....Equation 1.7

The variables used in the above equations are: E = accelerating potential, E_k = kinetic energy, m = mass, v^2 = acceleration, or velocity squared, z = elemental charge, t = time of flight, L = length of tube $v^2 = \left(\frac{L}{t}\right)^2$ L is always the length of the drift tube.

An ion with a larger mass will have more inertia than an ion with less mass, and therefore when accelerated in a uniform electric field, the larger ion will have a lower velocity than the ions with less mass when entering in the flight tube.

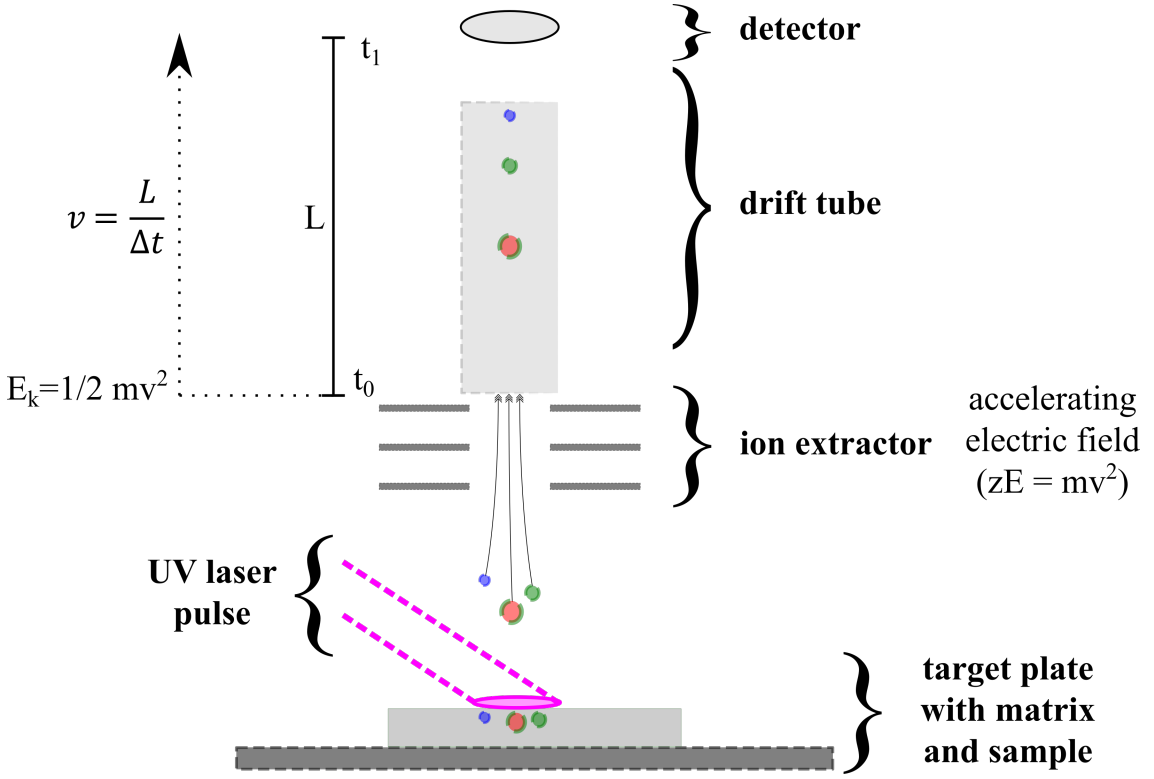


Fig. 1.13 MALDI-TOF MS Schematic

Shown here is a linear MALDI-TOFMS, to help illustrate the equations outlined in the text. The fuschia color represents a pulse of light, which is absorbed by the matrix, partially ablating the surface. Charges are transferred. The ion extractor accelerates the ions with kV of potential. The ions exit into the drift tube with kinetic energy (Ek), which is equal for all ions, the time of flight is measured from t_0 until detected at t_1, for travel of the ions through the length (L) of the drift tube. Using equation 1.7 the *m/z* can be determined, since t is the only variable, E and L are known values.

The equations explain the theoretical ideal. However, there are factors at play that can alter the time of flight. For instance, if the surface of the matrix is uneven, then the total distance to travel for ions of equal *m/z* is not the same, resulting decreased resolution. If the target is too thick, the distance between it and the extraction lens may be different from that used for calibration, and this will introduce a measurement error. Another occurrence which may alter resolution is that the ions are not accelerated through the extractor at exactly the same time from ion plume, resulting in different E_k for ions of the same *m/z*. To combat these issues, several strategies have been developed.

One such strategy involves reflecting ions back at an acute angle before they reach the first detector. This is done by introducing a region with an electrostatic field gradient, into which the more energetic ions penetrate more deeply before being reversed in their flight paths. The effect of the reflectron is a reduction in the energy spread of the sample beam that reaches the second detector (261) (see Fig. 1.14). A reflectron helps in two ways, is the first being that increases the overall length of the flight path – but this could be achieved by simply increasing the length of a linear flight path. The other, more important, effect is that it refocuses ions with the same *m/z,* but slightly different initial kinetic energies. This is accomplished since an ion with the same *m/z*, but with more kinetic energy will penetrate deeper into the reflectron field. When the ions are reflected, the total path length will have increased for the ions with more kinetic energy, re-focusing ions with the same *m/z* but different kinetic energies, putting them closer to the same plane and allowing a tighter ion packet to hit the detector (262). The other method which was developed to improve resolution by correcting for differences in initial kinetic energies that could arise during the ionization process was the use of a delayed ion extraction technique (263). This method involves waiting a precisely timed interval after laser irradiation to apply the ion extraction field, either in a single stage, or multistage extraction process (263-266).

A distinct advantage which TOF instruments have over quadrupole-based instruments is that they have wider *m/z* range, making them quite suitable for intact protein analysis (267). On the other hand, for protein analysis, quadrupole instruments are frequently used observe multiply charged ions, when they are interfaced to ESI or

nano-ESI sources. Hybrid instruments which contain both a quadrupole and a TOF analyzer combine features of both designs, as shown in the double quadrupole TOF MS example in Fig. 1.14. However, the use of a quadrupole before the TOF analyzer restricts the *m/z* transmission range to that of the quadrupole.

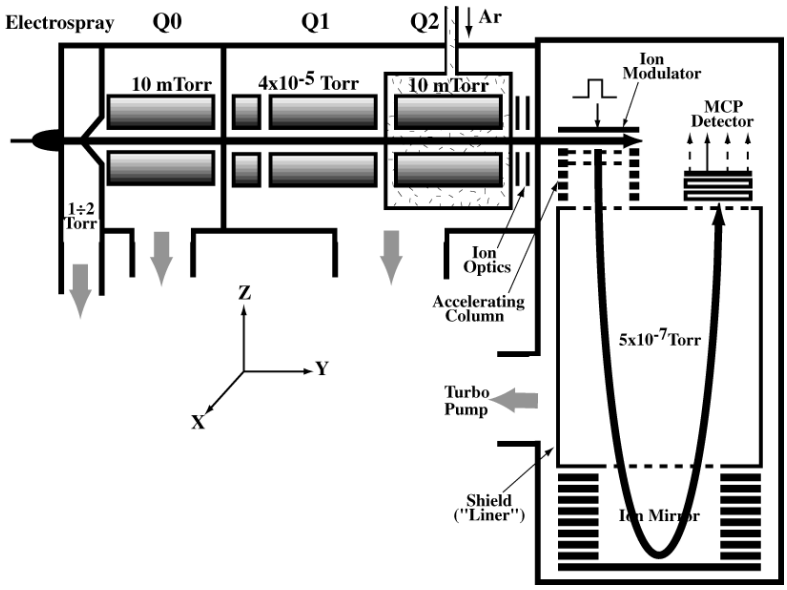


Fig. 1.14 QqTOF Instrument Schematic.

This is a schematic diagram of a quadrupole time-of-flight mass spectrometer. Ions are introduced into the flight tube from the perpendicularly placed quadrupole assembly. A pulsed electric field accelerates ions orthogonal to their original trajectories so their times of flight can be measured. (Reproduced with permission from Chernushevich, et al., 2001)

Fourier Transform Mass Spectrometry (FT-MS)

There are two major classes of FT (Fourier transform) mass spectrometers. The Fourier transform ion cyclotron resonance (FT-ICR), and the Orbitrap instruments. Both types of instruments require the application of a Fourier transform to deconvolute complex time-domain data and convert it into frequency values that can be converted into *m/z* measurements.

A Fourier transform is a mathematical conversion of time-domain data to frequency-domain data, for which the mathematics were described by Joseph Fourier in 1807. He showed that a function could be represented as a trigonometric series. This was later expanded upon and proven by Peter Gustav Lejeune Dirichlet (268). FT has been applied for processing data from many types of analytical chemical instrumentation; Comisarow and Marshall were the first to utilize FT for ion cyclotron mass spectrometry (269).

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS)

FT-ICR MS instruments use a strong magnetic field to trap ions, along with periodic rf pulses to ensure that the ions remain in motion. Ions are detected when they pass between detector plates where a small current is induced and measured. A basic experimental sequence involves first quenching the ICR cell to remove any remaining ions. Analyte ions are then introduced into the ICR cell, and cyclotron motion is induced by an rf chirp to excite ions with their cyclotron frequencies (270). The ions in a cyclotron motion pass over pairs of plates, inducing a current which is measured, and is observed as a composite of numerous frequencies and amplitudes (269, 270). This image, then has a Fourier transform applied to convert the time-domain to frequency-domain, a calibration curve is then applied to this data to obtain a visual spectrum (269, 270). One of the wonderful things about any ion trap mass spectrometer is that ions are detected in a non-destructive manner, and therefore, under ideal circumstances, the ions can be trapped for extremely long times and studied in various ways. The frequency at which ions rotate perpendicular to the magnetic field in an FT-ICR cell is described by equation 1.8, where

a charged particle with a charge (q) and a mass (m), has a cyclotron frequency (f_c) when in a uniform magnetic field (B) (see Equation 1.8) (269). Therefore, if a frequency can be measured, an *m/z* can be determined.

$$f_c = \frac{qB}{2\pi m} \qquad \text{Equation 1.8}$$

The first report of a mass spectrometer utilizing a Fourier transform showed the spectrum of methane as an ion cyclotron resonance (ICR) as an absorption spectrum (269). These investigators published another short communication soon thereafter and showed that frequency-sweep excitation could be performed, thus demonstrating proof of the concept that a wide range of *m/z* could be detected using this type of instrument design (270). The technology has developed quite dramatically since its first description in 1974, as was discussed earlier regarding the recent development of 21-T FT-ICR mass spectrometers which have unparalleled mass accuracies <≤ 1ppm and resolutions R>>1,000,000 (183, 184). Most FT-ICR MS instruments use large, liquid helium cooled, superconducting electromagnets, which have a horizontal bore, which enables the mounting of lasers in line with the ICR the cell, and thus they are easily amenable to unique methods of fragmentation, such as infrared multiphoton dissociation (IRMPD) and ultraviolet photodissociation (UVPD) (271, 272). Infrared irradiation at low power may be used to unfold large molecules in order to make them more susceptible to CID and ExD.

Orbitrap

The other type of Fourier transform mass spectrometer is the basis of a series of instruments that have been broadly adopted in the last decade, the Orbitrap mass

analyzer-based instruments. Their design originated from an old concept that used only an electrostatic field imposed on a fine wire to trap ions, without the need for magnetic or rf fields (273). The unique design by Makarov, utilizes an electrostatic field, but similar to an FT-ICR, it images the ion harmonic frequencies as they oscillate back and forth along the z-axis, and applies a Fourier transform (see Fig. 1.15)(274). The key to the success of this design for the analysis of organic molecules is Makarov's invention of the C-trap for the efficient introduction of ions to the Orbitrap mass analyzer.

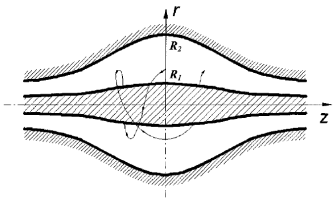


Fig. 1.15 Cross-Cut View of an Orbitrap with an Ion in Motion

(Reproduced, with permission, from Makarov, 2000)

This design allows for high resolution spectra to be obtained quickly on instrumentation which is neither difficult nor expensive to maintain (275). The compact nature of this mass spectrometer allows for bench-top instrumentation, with extremely fast acquisition rates for MS and MS/MS spectra, superb sensitivity, an overall ideal platform for performing in-depth proteomics experiments (see Fig. 1.16) (276).

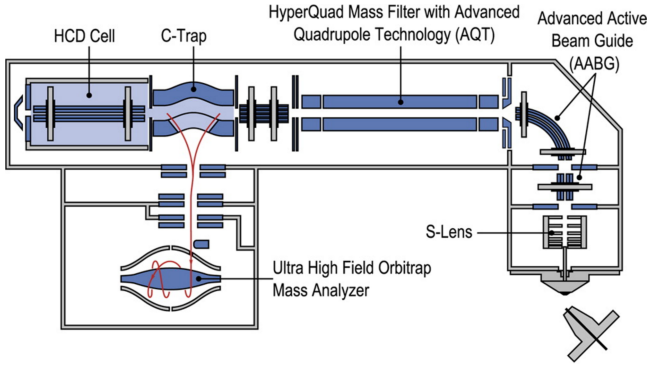


Fig. 1.16 The QE Series of Mass Spectrometers.

QE series of mass spectrometers sold by ThermoFisherScientific. Adapted from reference (276), under the Creative Commons Attribution (CC-BY) License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

MS/MS: Fragmentation

Multiple stages of mass spectrometry can be performed, and this type of analysis is often referred to as tandem MS or MS/MS. The first stage of MS involves the measurement of the parent ion(s), the second stage involves the isolation of a selected precursor ion observed in the MS, which is then fragmented. The resulting product ions are measured in a second stage of mass analysis, to produce the MS/MS spectrum. In ion trap instruments, a fragment ion can be isolated and fragmented, and the resulting spectrum recorded, and then produce ions can be selected for further study and this procedure can continue until no signal remains. This is referred to as MS^n, where *n* are the number of sequential stages of fragmentation and mass analysis.

Fragmentation can occur even with mass spectrometers under many circumstances, whenever the analyte is raised to an energy level that is sufficient to cleave its weakest bonds. Fragmentation can occur during the ionization process, in the ion source (prompt fragmentation), or post-source (metastable post-source fragmentation

or deliberately-induced dissociation) (277, 278). Sometimes this can be utilized to expand the capabilities of an instrument, other times it is an unintentional consequence of an improper operating mode. It may result in the loss of extremely labile functional groups or interesting modifications on molecules (279-281). There is a multitude of different ways to fragment ions; only the most common methods and the methods utilized during the research for this dissertation will be discussed.

Collision-Induced Dissociation (CID)

Collision-induced dissociation is the fragmentation of an ion due to its interaction with a neutral gas (e.g., helium, nitrogen, argon, or, xenon), that involves the transfer of kinetic energy and results in breakage of one or more chemical bonds. CID is also called collisionally activated decomposition (CAD). CID fragmentation has been described as either a high or low energy process; the categorical grouping is based upon how much kinetic energy is involved (usually 3-10 keV for high, and 5-100 eV for low), and how quickly the energy transfer takes place (282-284).

Low energy CID fragmentation is typical of rf-only quadrupole ion traps, where an ion of interest is isolated and excited at its resonant frequency, with just enough amplitude to excite the ion, but not enough to eject the ion. The ion collides with the helium buffer gas, yielding fragment ions (282). Similar low-energy CID fragmentation has been performed with a FT-ICR MS, usually by exciting the ions of interest at a value slightly removed from their resonant frequencies, in the presence of a higher pressure neutral gas in the ICR cell with enough energy to induce cyclotron resonance of the ions but not enough to eject them, in a process called sustained off-resonance collisionally

activated dissociation (SORI-CAD)(285). Triple quadrupole instruments can also perform low energy CID fragmentation with performances similar to those of an ion trap (286), but there are differences in the spectra because the number of collisions is lower in this type of instrument. In a triple quadrupole design, the first quadrupole is used to select the ion of interest, the second quadrupole performs as a collision cell, where it is filled with nitrogen or argon with the quadrupole operating in rf mode only, and finally the third quadrupole is used to analyze the product ions (287).

High energy CID fragmentation occurs when the collision energy is 5-10 keV. This type of CID fragmentation is rarely encountered with modern instrumentation; it was the pathway which was standard for double-focusing magnetic sector instruments, but few of these remain in use today (288). In TOF/TOF instruments, then collide with argon at low keV to produce high energy CID fragmentation ions, but this is not nearly so extensive as that observed with true high-energy fragmentation magnetic sector instruments (289) and more.

Higher-Energy C-trap Dissociation (HCD)

Another type of collision-induced fragmentation is called higher-energy C-trap dissociation (HCD), which is superficially similar to the CID fragmentation described for a triple quadrupole operating with a modified collision cell operating at higher pressures, in place of the original lower pressure open design of quadrupole two (286, 287, 290). HCD was originally performed in the C-trap of an LTQ-Orbitrap instrument, whose designation is based on its shape, where the rf frequency voltage was increased from 1,500 V to 2,500 V., which allowed for the detection of lower *m/z* ions since

fragmentation occurs outside the ion trap (see Fig. 1.16), however, soon a dedicated octapole collision cell was used, and became a commercially available option (290, 291).This type of fragmentation has also been referred to "high energy collision dissociation" using the same acronym (HCD) on the later generation of Orbitrap-based instruments in which the dedicated collision cell is integrated into the design (290, 292). One of the advantages that HCD fragmentation has over CID fragmentation is that the lower *m/z* ions are retained, where in an ion trap the lower 1/3 *m/z* are not trapped after fragmentation (292). There was a technology that was developed to try and overcome the loss of the lower *m/z* range in an ion trap, referred to as pulsed Q-dissociation (PQD) (293). This technology was quickly surpassed by the increased sensitivity, more even ion intensity distribution across the *m/z* range, and ease of use by HCD making PQD is obsolete (294).

Electron Reaction Based Fragmentation Methods (ExD)

Other fragmentation techniques used for MS/MS experiments utilize electrons to induce fragmentation; these include electron capture dissociation (ECD), electron excitation dissociation (EED), electron detachment dissociation (EDD), electron transfer dissociation (ETD), and (295-298). These ExD methods can all be performed in an FTICR MS. In addition, ETD is successfully carried out in quadrupole ion traps that are operated alone or as part of hybrid systems and ECD is sometimes performed in Q-TOF instruments (256, 299)

Electron capture dissociation (ECD) was first reported in an FTICR MS and is still usually performed in these systems but may be adapted to other mass analyzers. A

heated filament or solid electron source introduces electrons into the ICR cell, where electrons with energies in the range ca. 1-3 eV react with multiply positively charged ions, capturing an electron in an exothermic process, releasing additional energy and breaking nearby molecular bonds (295). Electron excitation dissociation (EED) employs an experimental setup very similar to ECD. In EED, even singly protonated ions can be reacted with >10 eV electrons, which subsequently gain charge through a slow radical cation formation, or secondary ionization, then undergo a subsequent fast electron capture process, or intramolecular rearrangement that leads to fragmentation (296). Other electron-induced dissociation processes use electrons generated at various energy levels and may also employed to dissociate negatively-charged analytes.

The last type of ExD fragmentation that will be covered here is electron transfer dissociation (ETD). The fragmentation achieved from ETD is similar to ECD, however, it is easily compatible with a broader range of mass analyzers, the only requirement being that the instrument needs to be fitted with a negative chemical ionization (NCI) source that can introduce an anionic radical to transfer an electron to ions of interest to induce fragmentation. During the development of this technology, a wide range of molecules were tested for their ETD efficiencies as reagent ions, and it was found that SO2, *m*-dinitrobenzene, *o*-dinitrobenzene, *p*-dinitrobenzene, azobenzene, fluoranthene, and perylene appeared to work well for the fragmentation of polycationic peptides (300). Another report compared reaction efficiencies using anthracene and 9,10-diphenylanthracene as NCI reagents to perform ETD experiments with polypeptides (301). Coon et al. found that 9,10-diphenylanthracene formed a more stable anionic

radical with improved electron transfer capacity, while anthracene reacted very quickly with the methane carrier gas, showing again that a polycyclic aromatic hydrocarbon was suitable for ETD reactions. The most practical ETD reagents are the polycyclic aromatic hydrocarbons; azobenzene and the nitrobenzenes and are somewhat unstable and SO2 is a toxic gas. Ultimately, fluoranthene became the reagent of choice for ETD experiments, likely because it sublimes at a lower temperature than the other polycyclic aromatic hydrocarbons and is very efficient at ETD fragmentation of multiply charged peptides (302-304). A limitation of ETD fragmentation, similar to that of ECD fragmentation, is that multiply charged cations are required in order for fragmentation to occur; there is positive relationship between peptide charge state and the production of fragments (302).

Fragmentation utilizing ETD has been successful for peptides, glycopeptides, phosphopeptides, proteins, and for native and permethylated glycans adducted to multi-cationic metals (305, 306). In addition to the studies performed in the positive mode, negative ETD (NETD) fragmentation has been performed on multiply anionic peptides, as well as for polyanionic glycosaminoglycans, the fragmentation mechanism was described as a fragmentation process akin to the EDD mechanism (307, 308).

Nomenclature for Naming Ions Observed in MS/MS Spectra

Peptide Fragmentation Nomenclature

A naming system for peptide fragmentation was proposed by Roepstorff and Fohlman in 1984 (300). Lower case letters indicate the type of ion, with a subscript number indicating the location within the peptide sequence with respect to the N- or C-terminus (see Fig. 1.17). The a_n, b_n, and c_n ions all contain the N-terminus, n = 1 for the

first amino acid and the numbers increasing sequentially towards the C-terminus. Conversely, the x_n, y_n, and z_n ions all contain the C-terminus and are numbered starting with the C-terminal amino acid as n = 1, and increase towards the N-terminus (see Fig. 1.17 for the general scheme, Fig. 1.18 for an example peptide).

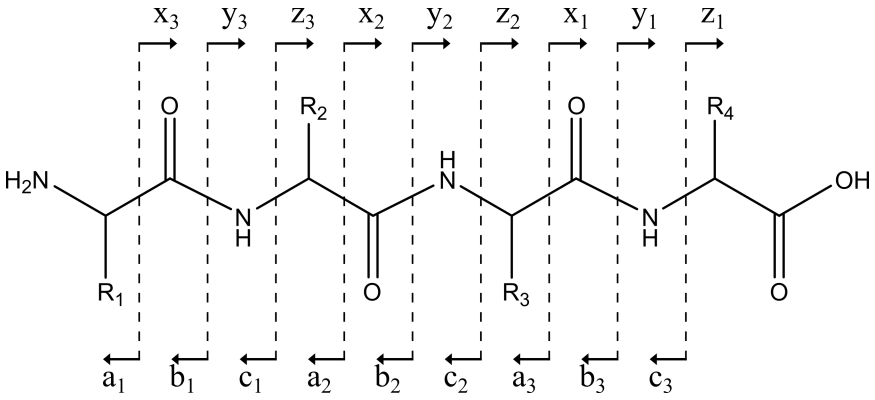


Fig. 1.17 Peptide Backbone Fragmentation Nomenclature

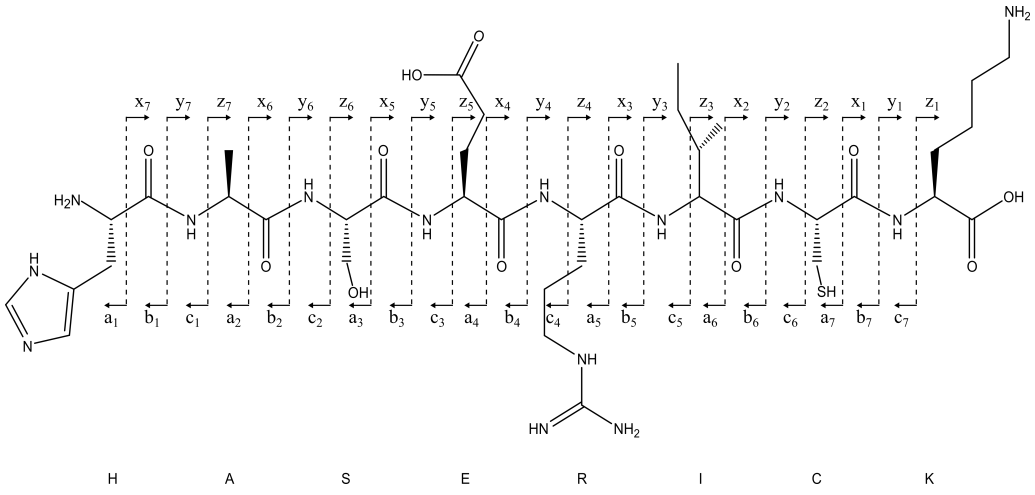


Fig. 1.18 A Hypothetical Peptide Naming all Theoretical Peptide Fragmentations

Two additional types of ions that can be observed, d- and w-ions, are the result of amino acid side chain loss. These result from higher-energy processes and are rarely observed in low-energy CID spectra. Side chain fragmentation can distinguish between

isomeric amino acid residues, e.g., leucine and iso-leucine (284, 301). High-energy CID (keV) fragmentation also produces abundant a -ions, as well as the possibility of (a +1) ions, furthermore, z, x, and y -ions are also observable (302, 303).

Low energy CID fragmentation and HCD fragmentation typically produce b/y-ions as well as a-ions (290, 304, 305). The c- and z-ions are typically found in ECD fragmentation spectrum (295). Similarly, ETD fragmentation yields c/z -ions, although, there have been observations of a -ions and occasionally low intensity b/y –ions, however, these lower intensity ions are usually a CID process from residual collision gas or rf heating (298). Using a gentle CID activation of the charge-reduced ions after electron transfer, referred to as supplemental activation (SA), there is observation of exclusively c/z -ions (306). Supplemental activation has become a standard practice for most ETD experiments, with isolation of the charge-reduced species not required, only a gentle CID activation is performed after the ETD reaction yielding much higher fragmentation efficiencies (306). However, it should be noted that the separation of the charged reduced species from residual precursor and low intensity c/z –ions would yield much cleaner fragmentation pathways, since a mix of odd and even electron c/z -ions can occur (306). The even-electron (c^{+}) and odd-electron ($z^{+\cdot}$) are typical of ETD, but, the co-observation of even-electron (z^{+}) + 1H, and odd electron ($c^{+\cdot}$) – 1H ions can occur from abstraction of hydrogen from an α carbon, yielding the additional c/z-ion types (306).

Glycan Fragmentation and Nomenclature

Fragment ions from glycans or glycoconjugates have their own nomenclature, which was defined by Domon and Costello, 1988 (307)(see Fig. 1.19). It uses upper-case letters, in order to be complementary to the system for peptides, and is thus suitable for designating the fragments from glycopeptides and glycoproteins.

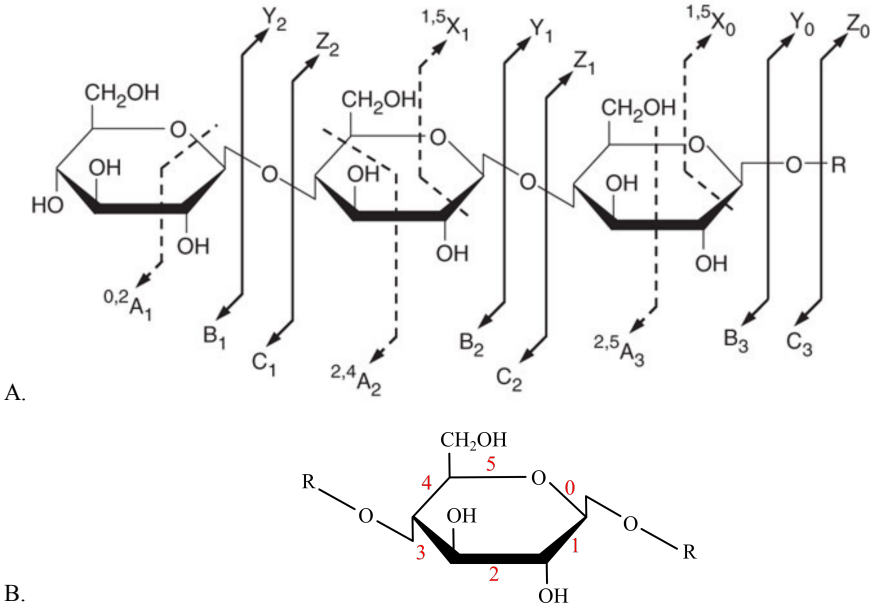


Fig. 1.19 The Domon and Costello Nomenclature for Glycosidic Bond Fragmentation

A. Glycosidic fragments (B_i, C_i) contain the non-reducing end numbering starts with i = 1 and increases sequentially towards the reducing end with each monosaccharide addition. Similarly, the glycosidic fragments (Y_j, Z_j) contain the reducing end sugar and increase sequentially as monosaccharides are added towards the non-reducing end, starting with j = 0, indicating the bond if the glycan were part of a conjugate; the first interglycosidic bond is j = 1. The cross-ring fragments, ${}^{k,l}A_i$ and ${}^{k,l}X_j$, contain superscript numbers "k,l" which indicate the two bonds within the ring that are broken (see Fig. 1.19B for numbering). The ${}^{k,l}A_i$ ions, designate non-reducing-end fragments, with the subscript number "i" indicating the residue within monosaccharide the cross-ring fragments occurred, starting with i = 1 for the terminal non-reducing residue. Reducing end fragments are annotated as ${}^{k,l}X_j$ ions, with the cross ring fragments on the reducing end starting with j = 0. B. The numbers in red show the numbering used to indicate which bonds are broken across the ring If the glycan is branched, each branch is assigned a subscript Greek letter, with a lowercase Greek letter assigned to each branch, starting with the largest branch labeled as α. If the branches have additional branch points, then the Greek subscript is given a prime, double prime, triple prime, etc., for each sequential branch. If an ion is the product of multiple fragments, then the combination is indicated with a forward slash designating the cleavage positions. (Reproduced, with permission, from Domon and Costello, 1988)

This nomenclature can be applied to native and derivatized glycans and glycoconjugates. Glycosidic bonds of the B-, C-, Y- and Z-types are frequently observed in low energy CID fragmentation (308). Low energy CID fragmentation of underivatized glycans, cationized with sodium, [M + Na]+, often generate dominant Y-ion fragments which may be quite useful for determining the topology of released *N*-linked glycans (309); the presence and completeness of this series depends on the composition, sequence and branching of the glycan. Both low energy and high energy CID MS/MS experiments of protonated, [M + H]+, native glycans and glycoconjugates can produce internal monosaccharide losses as well as monosaccharide rearrangements, making the interpretation of such spectra challenging (310, 311). Therefore, when studying native glycans, it is preferable to perform CID MS/MS experiments on metal-cationized species, as rearrangements and internal losses are rarely observed, and thus the spectra are less ambiguous (309).

MALDI-TOF MS analyses of underivatized *N*-glycans often produce the sodiated ion of the composition [M + Na]+, and, when fragmented with high energy CID in a MALDI-TOF/TOF MS, the resulting spectrum usually contains a mixture of glycosidic bond fragments of the (B-, Y-, C-, and Z- types), as well as cross ring fragments of the (A- and X-type); however, the most abundant cross-ring fragments are $^{1,5}X_j$ ions which do not provide interglycosidic linkage information (312).

Permethylation of glycans aids in the interpretation of MS/MS spectrum, since ions resulting from multiple fragmentations can easily be assigned due to the absence of one or more methyl groups where a previous glycosidic bond was present (308, 313). The

analysis of permethylated glycans increases both the sensitivity and fragmentation coverages, as was observed in the spectra of high energy CID fragmentation of sodiated [M + Na]+ ions, where abundant (B-, Y-, C-, Z-, A-, and X-type) ions are observed (314). Furthermore, many X-ions can be of the type that provide interglycosidic linkage information (314).

The electron-based fragmentation techniques also generate informative fragments from the permethylated glycans. For doubly-sodiated permethylated *N*-glycans and linear oligosaccharides, $[M+2Na]^{2+}$, when irradiated with 5-14 eV, in a process referred to as "hot" ECD, monosodiated, disodiated, and protonated fragment ions of the (C-, Z-, B-, Y-, A-, and X-type) are observed (315). In addition to the typical C- and Y-ions of the composition $[M + Na]^{+}$, glycosidic fragments with concurrent loss of two hydrogens can be observed as $[Y + Na - 2H]^{+}$ and $[C + Na - 2H]^{+}$ (315).

Similar fragmentation patterns are observed for permethylated glycans using EED fragmentation, with ion intensities being dependent upon which metal is adducted and how many charges are present. An important advantage of EED is that singly-charged metal cation adducts such as $[M + Na]^{+}$ or $[M + Cs]^{+}$ can be analyzed (316, 317). ETD fragmentation of reduced and permethylated glycans containing a magnesium adduct, $[M + Mg]^{2+}$, produces abundant glycosidic (C-, Z-type ions), and informative cross ring fragmentation ions (A-, and X-type) in a much gentler and controlled manner, where the fragments are typically $[P + Mg]^{\cdot +}$ or $[P – H + Mg]^{+}$, where P = product, however, fragments can also be observed as $[P + H]^{+}$, usually as B-ions (318).

Gas Chromatography-Mass Spectrometry (GC-MS)

Gas Chromatography

Gas chromatography separates compounds in a mixture based primarily upon differences in the boiling point (bp)of the compounds and the interaction of these compounds with the stationary phase (liquid or solid) and, to some extent, the interaction with the mobile phase (flowing carrier gas)(319, 320). The first practical GC setup was described for the separation of volatile fatty acids (320). In this setup, the mixture of volatile fatty acids was dissolved in a stationary phase composed of a high boiling point silicone oil, and the solution was placed onto a glass wool plug on the end of a capillary tube filled with diatomaceous earth in a temperature-adjustable sheath, and a stream of nitrogen flowed through the column to transport any volatile compounds. Separation of the components in the mixture was achieved by increasing the temperature of system to the boiling points of the analytes, where they left the stationary phase (the high boiling point silicone oil, entered into the stream of gas, and were resolved from one another along the length of the column as they travelled to the detector (a titration cell)(320). The necessity for the analytes to be volatile limits the scope of materials that can be analyzed using GC; however, the introduction of a heated injector port and advancement in derivatization procedures allows for lower boiling or non-volatile components to be transferred into the gas phase without decomposition and therefore provides a way to analyze these materials. There has also been a simultaneous advancement in GC detector technology, which greatly expanded the capabilities of GC analysis. One of the first widely utilized GC instruments was developed at The Dow Chemical Co. (321). That

apparatus was capable of separating mixtures with boiling points up to 350 °C and used a thermal conductivity detector (TCD) (321). Other methods of detection were developed for GC, including flame ionization (FI) (322), flame photometric (FP)(323) and electron capture detectors (EC) (324). These detection techniques do offer sensitive detection capabilities, however, some of these are compound-specific, such as flame photometric which relies upon the phosphorescence of phosphorous of sulfur-containing molecules (323), or rely on the electronegativity of atoms (such as halogens and fluorine) with electron capture detectors (324). One of the major drawbacks of these detectors is that they don't necessarily provide additional information on the compositions of the effluents, and therefore mass spectrometric detection techniques were developed. It should be noted that modern GCs can utilize multiple detectors by splitting the flow of gas; such a setup can be advantageous for specialized applications.

GC Columns

The initial descriptions of GC using a liquid stationary phase were limited with the choices of stationary phases and the analytes which could be analyzed. The development of capillary columns containing a thin inner coat of a solid support medium which could be impregnated with any stationary phase, leaving an open bore for increased carrier gas flow rates provided faster, more efficient, and extended separation capabilities (325). Thin film open bore capillary columns are encountered most often in the modern laboratory for general use; however, porous medium-packed columns are still used for specialized applications. Numerous books, reviews, and guides have been published which detail the different stationary phases used for GC and GC-MS columns,

and how these can be applied for specific purposes (326-329). For general GC-MS applications, a low bleed, non-polar DB-5 (5% phenyl-polymethylsiloxane/95% polymethylsiloxane) column is most often employed. This stationary phase was used for all the studies described in this dissertation.

Gas Chromatography-Mass Spectrometry (GC-MS)

The first publication describing the coupling of a gas chromatography instrument to a mass spectrometer was by Holmes and Morrell in 1957 (330). This early system demonstrated the utility of GC-MS in analytical chemistry. Gas chromatography provides a means to rapidly separate mixtures and showed very good sensitivity compared to most analytical methods at the time; coupling a GC to a mass spectrometer provided an additional dimension of data, thus allowing for the confident assignment of compounds, while not completely relying upon retention time matching to authentic standards, something very useful for identifying an unknown substance (330). In this proof-of-concept GC-MS, only a narrow *m/z* window could be scanned within a reasonable time frame, providing limited information per GC-MS analysis (330). This was due to the limitations of the mass spectrometer being used, a Consolidated Electrodynamics Corporation model 21-103B magnetic sector instrument (330). Soon after this first publication, the report of coupling GC to a TOF-MS for detection was published, this TOF-MS could scan a broad *m/z* range very quickly, demonstrating that GC-MS could be a very informative technique when coupled to faster scanning instruments, although the low mass resolution (<400) and requirement for volatility limited the applications to fairly low MW and handling the flow rate was challenging for the pumping system

compounds (331). The development of GC-MS separation devices to enrich the sample delivered to the MS ion source and divert most of the gas away from the sample flow greatly accelerated the utilization of GC-MS for analysis of biological samples (332-334). From these first reports, this area of research expanded and developed quite rapidly, however, the basic principles remains the same. Currently, a typical GC-MS system will use a low-flow capillary column, with helium as the carrier gas; the effluent enters into the ion source of the mass spectrometer under vacuum conditions, where ions are produced with an electron impact ionization source similar in concept to the ion gauge setup of the pre-GC-MS era (335), see Fig. 1.10 for an EI source schematic and description, also refer to the "Electron-Impact Ionization" sub-section of Chapter 1 for a detailed review of the ionization process.

Derivatization

As mentioned previously, gas chromatography requires that the analytes are volatile, due to the reliance upon boiling point as a major determinant of separation. In addition, most biological molecules are polar (fatty acids, sugars, amino acids, etc), and the polar surfaces can interact with the surfaces of the GC system leading to chromatographic peak shape artifacts. Derivatization minimizes the interaction with any active surfaces, and, more importantly derivatization confers volatility. The derivatization product must be chemically and thermally stable. For carbohydrates, there are three common methods of derivatization for GC analysis. These include formation of trimethylsilyl (TMS) esters and ethers of monosaccharides (336), methyl ethers and esters (337, 338), and acetates (339, 340).

Permethylation of monosaccharides without reduction will yield two peaks due to the presence of an anomeric carbon on the reducing end that may be either the α or β epimer. Methanolysis opens the ring and subsequent permethylation (or other derivatization method) yields a single product (338). The acetylation of polyhydroxyl molecules, in particular monosaccharides, is often employed for their analysis by GC (339). The acetyl esters are not as volatile as the methyl ethers; this offers some advantage, in that the acetyl derivatives have higher boiling points, and show better resolution (340). The analysis of monosaccharides as the alditol acetates is well suited for the analysis of neutral sugars but cannot accommodate sialic acids (341). The analysis of alditol acetates has advantages over that of the TMS-esters, because of their increased stability, and their presence as a single chromatographic peak, whereas the TMS-esters give rise to two or more peaks per monosaccharide (336, 342). Formation and analysis of alditol acetates has been used for determining the monosaccharide composition of glycoproteins for a very long time, and has become a standard procedure (341). Derivatization of the non-reduced monosaccharides can also be performed, but this will yield two products for the α /β epimers, as noted above (340). Analysis of the monosaccharides as the alditol acetates is the method used most often for the studies described in this dissertation; however, analysis as the permethylated alditols was employed with good success for GC-MS analysis of the acetamido sugars GalNAc/GlcNAc.

Chapter 2. Asparagine-Linked Glycans of *Cryptosporidium parvum*.

This chapter was first published online on February 8, 2017 in the Molecular and Cell Proteomics Journal. It was published April 1st, 2017 in the special issue, "Proteomics and Infectious Disease: Defining Pathogens Properties and Virulence". The full citation is shown below:

Haserick, J. R., Leon, D. R., Samuelson, J., and Costello, C. E. (2017) Asparagine-Linked Glycans of *Cryptosporidium Parvum* Contain a Single Long Arm, Are Barely Processed in the Endoplasmic Reticulum (Er) or Golgi, and Show a Strong Bias for Sites with Threonine. *Molecular & Cellular Proteomics* 16, S42-S53

John R. Haserick performed all the experiments described within this chapter. Deborah R. Leon played critical roles by assisting with setting up the instrument, and with acquiring and interpreting the data.

Abstract

Cryptosporidium parvum causes severe diarrhea in infants in developing countries and in immunosuppressed persons, including those with AIDS. We are interested in the Asn-linked glycans (*N*-glycans) of *C. parvum*, because (1) the *N*-glycan precursor is predicted to contain five mannose and two glucose residues on a single long arm versus nine mannose and three glucose residues on the three-armed structure common in host *N*-glycans, (2) C. parvum is a rare eukaryote that lacks the machinery for *N*-glycan-dependent quality control of protein folding in the lumen of the Endoplasmic Reticulum (ER), and (3) ER and Golgi mannosidases, as well as glycosyltransferases that build complex *N*-glycans, are absent from the predicted proteome. The *C. parvum N*-glycans

reported here, which were determined using a combination of collision-induced dissociation and electronic excitation dissociation, contain a single, unprocessed mannose arm ± terminal glucose on the trimannosyl chitobiose core. Upon nanoUPLC-MS/MS separation and analysis of the *C. parvum* tryptic peptides, the total ion and extracted oxonium ion chromatograms delineated 32 peptides with occupied *N*-glycan sites; these were derived from 16 glycoproteins. Although the number of potential *N*-glycan sites with Thr (NxT) is only about twice that with Ser (NxS), almost 90% of the occupied *N*-glycan sites contain NxT. The two most abundant *C. parvum* proteins modified with *N*-glycans were an immunodominant antigen on the surface of sporozoites (Gp900) and the possible oocyst wall protein 1 (POWP1). Seven other glycoproteins with *N*-glycans were unique to *C. parvum*; five shared common ancestry with other apicomplexans; two glycoproteins shared common ancestry with many organisms. In summary, *C. parvum N*-glycans are remarkable for the absence of ER and Golgi modification and for the strong bias toward occupancy of *N*-glycan motifs containing Thr.

Introduction

Cryptosporidium parvum and *Cryptosporidium hominis* are coccidian parasites (walled apicomplexans) that infect humans (both) and cows (*C. parvum* only) (49, 343, 344). *C. parvum* and *C. hominis* are leading causes of diarrhea and death in children in the developing world and cause chronic diarrhea in AIDS patients (20, 345-347). While the massive outbreak of *C. parvum* in Milwaukee in 1993 was associated with contamination of municipal water, *Cryptosporidium* in developing countries is likely spread by poor hygiene (6, 348). Although there are mouse vaccine models and

veterinary vaccines for *C. parvum*, there are no human vaccines (349-353). Furthermore, Nitazoxanide, the drug used to treat *C. parvum*, is not effective in immune suppressed persons (354).

We are interested in the *N*-glycans of *Cryptosporidium* for numerous reasons. Like *Toxoplasma gondii*, *Entamoeba histolytica*, and *Trichomonas vaginalis*, there is secondary loss of Alg genes, so that the predicted *N*-glycan precursor of *C. parvum* has a single long arm rather than the three-arm structure common in the host (Table 2.2) (110, 355-359) . In contrast to most other eukaryotes, *C. parvum* has a paucity of predicted mannosidases and glycosyltransferases, which could modify *N*-glycans in the Endoplasmic Reticulum (ER) and Golgi (360) . *C. parvum* sporozoites label with cyanovirin-N, an anti-retroviral lectin that binds to the high mannose *N*-glycans of gp120 in HIV (35, 361) . *C. parvum* is a rare eukaryote that lacks the machinery for *N*-glycan-dependent quality control of protein folding, and there is no positive selection for *N*-glycan sites in secreted proteins of *C. parvum* (111, 112, 362). Antigenic proteins on the surface of *Cryptosporidium* sporozoites (*e.g.*, Gp900 and Gp40/Gp15), termed oocyst wall proteins (COWPs) and possible oocyst wall proteins (POWPs), are glycoproteins with numerous predicted *N*-glycan sites (31, 33-35, 39, 40, 42, 363-369). Finally, Concanavalin A, which binds some *N*-glycans and other mannose-containing structures, recognizes numerous *C. parvum* antigens, while release of *N*-glycans reduces binding of immune sera to parasite proteins on Western blots (36).

Here we used tandem mass spectrometry to identify the proteins that contain *N*-glycans, and determine the structures of *N*-glycans released with PNGase F. We found

that the *N*-glycans of *C. parvum* contain a single long arm, are barely processed in the ER or Golgi, and show an extreme bias for sequons with threonine.

Materials and Methods

Parasites and Reagents

C. parvum oocysts were purchased from Bunch Grass Farm (Deary, ID) and handled under BSL-2 protocols approved by the Boston University Institutional Biosafety Committee. All chemicals and reagents, including proteomics grade trypsin, were obtained from Sigma-Aldrich (St. Louis, MO), unless otherwise stated. All solvents used for LC-MS were Fisher Scientific Optima™ grade (Thermo-Fisher Scientific, Waltham, MA). PNGase F was from New England Biolabs (Ipswich, MA).

Protein Extraction

Two distinct methods were utilized to extract proteins from whole *C. parvum* oocysts. The first method used a combination of mechanical disruption and detergent extraction. Briefly, 10^9 oocysts were concentrated by centrifugation at 1000 x g for 10 min at 4 °C. The oocysts were resuspended in phosphate buffered saline (PBS) with EDTA-free cOmplete™ protease inhibitor (Roche). The oocysts were broken using 0.5-mm glass beads with 4 x 5 min cycles of vigorous bead beating at 4 °C. Samples were placed in an ice bath between cycles to mitigate any heating effect. Proteins were extracted using a buffer containing protease inhibitor (10 mM HEPES, 25 mM KCl, 1 mM CaCl2, 10 mM MgCl2, 2% CHAPS, 6 M guanidine HCl, 50 mM dithiothreitol (DTT), pH 7.4). Insoluble material was removed by centrifugation at 21,130 x g for 5

min at 4 °C in an Eppendorf 5424R microcentrifuge. The supernatant was removed and added to a new microcentrifuge tube; proteins were precipitated by the addition of -20 °C acetone (acetone/sample v/v 8:1) and the tube was allowed to sit undisturbed for ≥18 h at -80 °C. The proteins were concentrated by centrifugation at 21,130 x g for 20 min at 4 °C. The supernatant was discarded, and the pellet was washed 3x with ice-cold acetone. Any remaining solvent was removed in an unheated Speed Vac Plus speed vacuum (Savant, Thermo-Fisher Scientific).

The second chemical method used hot phenol to kill and extract total proteins from 109 *C. parvum* oocysts (370, 371). *C. parvum* oocysts were pelleted by centrifugation, resuspended in 500 µl of distilled water, and added to a conical vial containing 1 ml of phenol, pre-heated to 68 °C in a heating block filled with sand. The vial was sealed, and the contents mixed by inversion every 2 min for 20 min. The vial was removed, placed on ice, and gently centrifuged to facilitate good phase separation. The aqueous layer was removed and discarded. The interphase and phenol layers were carefully separated and saved. The proteins were subsequently precipitated from the phenol and interphase layers by the addition of eight volumes of -20 °C MeOH containing 100 mM NH_4OAc, and allowed to sit undisturbed for ≥18 h at -20 °C. The precipitated proteins were concentrated by centrifugation, and pellets were washed 3x with -20 °C MeOH/0.1 M NH_4OAc prior to lyophilization.

Trypsin Digestions

Three sets of samples were prepared for proteomics experiments. The fraction obtained from the mechanical extraction is referred to as “CHAPS” in the analysis. Two

fractions from the chemical extraction method came from the phenol layer (referred to as "phenol') and the interphase layer (referred to as "interphase"). Precipitated proteins from these three samples were dissolved into 50 mM NH4HCO3, pH 8.0, reduced with 50 mM DTT for 20 min at 60 °C, cooled to RT, and then alkylated with iodoacetamide (IAA) for 20 min at RT, while protected from light. Excess IAA was quenched with DTT, and peptides were generated by digestion with proteomics grade trypsin, overnight at 37 °C (1:20, w/w). The resulting tryptic peptides were dried by speed vacuum and desalted with C18 ZipTip concentrators (EMD Millipore, Danvers, MA), according to the manufacturer's protocol.

Release and Processing of *N*-Glycans

N-glycans were released from total protein isolated from oocysts (100 µg) by overnight treatment at 37 °C with ten units of glycerol-free PNGase F (New England Biololabs), according to the manufacturer's instructions, without the addition of NP-40. The product mixture was lyophilized, and the released *N*-glycans were separated from the proteins by addition of 0.1% trifluoroacetic acid (TFA) in LC-MS grade water. The aqueous phase was passed onto C-18 Sep-Pak cartridges (Waters Corporation, Milford, MA). The cartridges were washed with three bed volumes of 0.1% TFA/water, and the eluents were pooled and lyophilized. The *N*-glycan pool was reduced with 0.5 M $NaBD_4$/2 M NH_3 (aq) overnight at 55 °C. The reaction was quenched by dropwise addition of glacial acetic acid. The products were washed multiple times with 10% acetic acid/MeOH, dried with a gentle stream of nitrogen, and washed again multiple times with 100% MeOH. Permethylation was performed by published methods (372, 373). Briefly, a

slurry of finely ground NaOH in dimethyl sulfoxide was added to the deutero-reduced sample; the suspension was mixed and methyl iodide was added. The solution was gently mixed at RT for one hr. To assure complete derivatization, the process was repeated three times. The product was isolated by extraction with water/chloroform, and the chloroform layer was dried in the SpeedVac.

MALDI-TOF MS

The purified deutero-reduced sample was dissolved in 20 µl of 1:1 MeOH/water, and 0.5 µl of this solution was spotted onto a stainless steel MALDI target with 2,5-dihydroxybenzoic acid as the matrix. The mass spectra were recorded with an ultrafleXtreme MALDI-TOF/TOF MS (Bruker Daltonics, Bremen, Germany) equipped with a smartbeam II Nd-YAG laser (355 nm, 3 nsec, 2 kHz). Each spectrum was acquired by summing the signals recorded after 500 shots from each of 10 locations within the sample spot.

Electron Excitation Dissociation (EED) Fourier Transform-Ion Cyclotron Resonance (FT-ICR) MS/MS

The released, deutero-reduced, and permethylated N-glycans were dried and re-suspended in 10 µl of 50% MeOH, 20 µM sodium acetate. The solution was loaded into a pulled glass capillary tube and directly infused into the ion source of a SolariX 12-T hybrid Qh-FT-ICR mass spectrometer (Bruker Daltonics), using a nano-ESI source. Each $[M + Na]^{1+}$ parent ion was isolated by the quadrupole and accumulated in the collision cell for 8 seconds. The accumulated ions were then transferred into the ICR cell.

Fragmentation by EED was achieved via 14-eV electrons generated from a cathode source heated with 1.5 A current. Electron density and energy were modulated using the following parameters: bias, 14 V; ECD lens, -13.85 V; pulse width 1.0 sec. For each spectrum, 80 transients were averaged.

LC-MS/MS

The dried and desalted peptides were reconstituted in 2% ACN, 0.1% formic acid (FA) and separated on a NanoAcquity Ultra Performance Liquid Chromatography (UPLC) system (Waters), fitted with a nanoAcquity Symmetry C18 trap column (5-µm packing, 180 µm x 20 mm) and a BEH130C18 analytical column (1.7-µm packing, 150 µm x 10 cm). The mobile phase A was 99:1:0.1 (HPLC grade water/ACN/FA), and mobile phase B was 99:1:0.1 ACN/HPLC grade water/FA. Each sample was loaded on the trapping column for 4 min at 4 µl/min flow rate and then separated on the analytical column using a 45 or 90 min 2-40% mobile phase B linear gradient at 0.5 µl/min flow rate. The column was washed between runs and equilibrated for 30 min. The analytical column was coupled to a TriVersa NanoMate ion source (Advion, Ithaca, NY), and the ions were introduced into either an LTQ-Orbitrap-XL-ETD or a QE Plus mass spectrometer (both from Thermo-Fisher Scientific, San Jose, CA), which was operated in the positive-ion mode. MS spectra were obtained by scanning over the range *m/z* 350-2000. MS/MS HCD spectra were acquired by isolating the top 5 (LTQ-Orbitrap) or top 20 (QE+) precursor ions with a 2-*m/z* window and fragmenting the selected precursor ions with 27, 35, or 45 V HCD energy. The MS/MS HCD spectra were scanned from *m/z* 100 to a value that was dependent upon the parent ion.

Manual Interpretation of Glycopeptide MS/MS Spectra

Raw data files from LC-MS/MS experiments were manually interpreted using Qual Browser in the Xcalibur 2.2 software suite (Thermo-Fisher Scientific). HCD MS/MS spectra containing oxonium ions were manually interpreted to determine the peptide sequence and the linear arrangement of the glycan. The $_{y1}$ ion, corresponding to the residue K or R, was used as the starting point for most of the manually interpreted spectra. The resulting peptide tag was then searched using the online NCBI BLASTP algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the predicted *C. parvum* proteome, and the entire nr database (356-358) . When a match was found, we determined the mass difference between the predicted trypsin generated peptide $[M + H]^{1+}$ and that of the precursor, converted to $[M + H]^{1+}$. The glycosidic bond fragment series, typically accounting for the most abundant peaks in the spectra, were sequenced in a similar manner, so far as each series could be followed. Missing residues were accounted for by calculating the difference between the highest member of the assigned series and the total observed molecular weight. Extracted ion chromatograms were generated to aid in the assignment of the numerous glycoconjugates.

LC-MS/MS Proteomics Database Search and Analysis

Once the possible *N*-glycoforms were discovered from the manual interpretation, these values could be utilized to search against the predicted *C. parvum* proteome as possible dynamic modifications. Database searches were performed using the PEAKS software suite version 7.5 (Bioinformatics Solutions Inc., Waterloo, ON, Canada). The following parameters were set for the search: the data refinement step corrected for the

precursor *m/z*, for the PEAKSdenovo search stages, trypsin was specified as the enzyme, 8.0 ppm parent mass error tolerance, 0.05 Da fragment mass error tolerance, with carbamidomethyl cysteine set as a fixed modification, and possible dynamic modifications set to include methionine oxidation, HexNAc at serine/threonine; $Hex_6HexNAc_2$ and $Hex_5HexNAc_2$ on Asn. A maximum of five dynamic modifications was specified. The PEAKSDB search stage was identical to the PEAKSdenovo stage, with the exception that up to three missed trypsin cleavages were allowed, with the possibility of one non-specific cleavage. Searches were performed against the *C. parvum* Iowa-II predicted proteome release-5.0 obtained from the Cryptosporidium Genome Resource (cryptodb.org) which contained 3,803 entries (357, 358). False discovery rate (FDR) estimation was enabled. For the final PEAKSPTM stage, the de novo score average local confidence (ALC) threshold was 15 and the peptide hit threshold (-10 logP) was set to 30. All possible Unimod modifications were considered for this stage. The PEAKSPTM report was exported as a mzidentML with a FDR set to 5%, ALC 50% for de novo only, and proteins with a score of (-10 logP) ≥ 20 containing unique peptides ≥ 2. Each data file was analyzed individually for all samples and replicates.

Scaffold Analysis

The mzidentML files from the PEAKSPTM searches were imported into the computer program Scaffold version 4.6 for further analysis (Proteome Software, Inc., Portland, Oregon). Three “Biosamples” and two “categories” were specified for the samples. The two categories corresponded to the method of protein extraction, either “mechanical” or “chemical”. The sample names correspond to the sub-sample

classification, the “CHAPS” was the mechanically broken 2% CHAPS extraction buffer soluble portion, and the “phenol” and “interphase” samples correspond to the phenol and interphase layers from the chemical extraction procedure. Each sample was analyzed independently, with experiment wide grouping and protein clustering. The probability model utilized was Peptide Prophet with delta mass correction. All spectra that were assigned by the software as possible *N*-glycosylated peptides were manually reviewed for quality and proper assignment to compile the final lists of glycopeptides and proteins that are available in the online supplemental file Excel S3 (http://www.mcponline.org/content/16/4_suppl_1/S42/suppl/DC1).

Analysis of *N*-Glycosylation Sites

For each protein observed to be *N*-glycosylated, the lists of occupied and total potential *N*-linked sites were compared. The “occupied” dataset was created from the list of peptides modified with an *N*-glycan, taking for each a nine-amino-acid window, centered on the modified asparagine. The same window was taken for all tryptic peptides which contained a canonical *N*-glycosylation sequon (NxS/T, N≠P) that could theoretically be generated from the group of observed glycoproteins. The program WebLogo v3.5.0 from the Department of Plant and Microbial Biology, University of California, Berkeley, was used to generate logos (374).

Bioinformatics

Predicted proteins of *C. parvum* with occupied *N*-glycan sites were analyzed for signal peptides and transmembrane helices using SignalP 4.0, TMHMM 2.0, and Phobius

(375-379). Conserved domains were identified, and proteins were compared with those of other apicomplexans, eukaryotes, and bacteria (356-358, 378, 379). Alg enzymes, glucosidases, mannosidases, and OST peptides were predicted from whole genome sequences of *C. parvum* and *T. gondii*, using *S. cerevisiae* as a model (63, 76, 110, 112, 355, 359). Protein cartoon schematics were drawn using the program DOG 1.0 in combination with the software Inkscape 0.91 (380)

Analysis of Released *N*-Glycans

To assist in the interpretation of the MS/MS spectra, we used the software GlycoWorkBench 2.1 (release 146) to generate theoretical fragmentation lists. Additional theoretical *m/z* values were generated using Microsoft Excel. Observed and theoretical peak lists were compared to obtain the best match. Assignments within 1-ppm error were considered to be a likely match. In the event that there were isobaric ion values, annotations were preferentially assigned to the ion that would be generated from a single fragmentation event. A single cross-ring fragment in combination with one or more glycosidic cleavages was considered only if the simple glycosidic bond fragment was also observed within the spectrum. All annotations were assigned only after a thorough manual review of the spectrum using Bruker DataAnalysis software suite version 4.0 SP5 build 283. Manual inspection helped to assign ions that didn't fit the list of theoretical values for expected cleavages.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005503 and 10.6019/PXD005503.

Results

N-Glycans of *C. parvum* are Much Simpler than those of the Host and Most Other Parasites.

The predicted *N*-glycan precursor of *C. parvum,* based upon its Alg enzymes, is $Glc_2Man_5GlcNAc_2$ (Table 2.2) (110, 355, 359, 362). MALDI-TOF MS of *N*-glycans that had been deutero-reduced and permethylated after being released by PNGase F from oocyst glycoproteins showed $Hex_6HexNAc_2$ ($[M + Na]^{1+}$ *m/z* 1800.906) to be the most abundant form, while $Hex_5HexNAc_2$ ($[M + Na]^{1+}$ *m/z* 1596.805) is less abundant (see Fig. 2.1). The *C. parvum N*-glycans are much simpler than those of calf glycoproteins, suggesting oocysts, which are washed with PBS and purified on a CsCl gradient, are clean of host tissues (125). The released *N*-glycans also match those present on glycopeptides which were separated by reversed phase C18 nanoflow chromatography, and identified by manual interpretation of HCD MS/MS spectra (see Fig. 2.4 and Table 2.1).

EED FT-ICR MS/MS was performed on the deutero-reduced and permethylated $Hex_6HexNAc_2$ in order to generate the glycosidic fragments that provide topographic information (Fig. 2.2) and the cross-ring fragments which provide linkage information

The glycan topology is indicated by complete glycosidic bond fragmentation, shown dominated by the non-reducing end fragments (C_n-2H) series; in addition, reducing end glycosidic bond fragments of the (Y_n-2H), and Z_n series are prominent (Fig. 2.2). The single long arm topology is suggested by the sequential $Z_{3\alpha}$ to $Z_{6\alpha}$ ions, and the parallel sequential ($Y_{3\alpha}$-2H) to ($Y_{6\alpha}$-2H) series, where 1-4 hexoses are attached without

branching. If branching were present in this tetrasaccharide moiety, there would be a gap in the linear series, and double glycosidic bond fragments might be observed along the chain. Instead, the only double glycosidic bond fragment series ($Z_{3\alpha}/Z_{3\beta}$ + 2H) through ($Z_{6\alpha}/Z_{3\beta}$ + 2H) correspond to cleavages involving the short arm (Fig. 2.3, Appendix 2A). The ($Z_{3\alpha}/Z_{3\beta}$ + 2H) ion indicates there is no branching from the chitobiose core. The remaining double glycosidic fragments of the ($Z_n/Z_{3\beta}$) series, where Z_n is $Z_{4\alpha}$ to $Z_{6\alpha}$, show loss of the single hexose $Z_{3\beta}$ branch, with no branch points down the long arm (Fig. 2.3, Appendix 2A).

The EED spectrum also allowed assignment of the linkage positions, as shown in Fig. 2.3. The key cross-ring fragment $^{0,4}A_5$ and its paired reducing end fragment ($^{0,4}X_2$-2H), show that the short arm has a single hexose attached via a 1,6-linkage to the central hexose. The observation of $^{3,5}A_5$ and $^{0,3}A_5$ ions support this assignment. The ion pairs ($^{1,3}A_5$ and $^{1,3}X_2$) indicate the longer arm, containing four hexose residues, is attached at the 2 or 3 position to this central hexose. Observation of the $^{0,2}X_2$ ion eliminates the 2 position as the linkage site, thus, narrowing the possibility for the long arm linkage to position 3. Although the $^{1,3}A_5$ and $^{1,3}X_2$ ions are isobaric to the $^{2,4}A_5$ and $^{2,4}X_2$ ions and these might allow assignment of the linkage to the 3 or 4 position, the presence of the $^{3,5}A_5$ ion rules out linkage at the 4 position. In sum, these cross-ring fragments indicate that the long arm containing four hexoses is attached by a 1,3-linkage to the central mannose.

The observation of $^{1,3}A_{4\alpha}$ and $^{1,3}A_{3\alpha}$ ions suggest that second and third hexoses on the long arm are either 2- or 3-linked. $^{0,2}A_{3\alpha}$ or $^{0,2}A_{4\alpha}$ fragments would be expected if

these hexoses were linked at the 3 position; their absence suggests that the links are on the 2 position. The presence of an internal fragment $^{1,3}X_{5\alpha}/B_5$ indicates that the terminal hexose is 1,3-linked and this assignment is confirmed by the observation of the $^{0,2}A_{2\alpha}$-2H ion that rules out the possibility that the terminal hexose is 2-linked (Fig. 2.3).

EED FT-ICR MS/MS was also performed on $Hex_5HexNAc_2$, which is the less abundant *C. parvum N*-glycan (Figs. 2.7 and 2.8, Appendix 2A). The same topology as described for the aforementioned $Hex_6HexNAc_2$ (minus a terminal hexose on the long arm) is indicated by the complete (C_n - 2H), Z, (Y_n - 2H), and ($Z_{n\alpha}/Z_{3\beta}$ + 2H) series where n = 3, 4, or 5 (Fig. 2.7 and Appendix 2A). A series of cross-ring fragments similar to the $Hex_6HexNAc_2$ glycan were observed for this glycoform. The observation of the ion pairs $^{0,4}A_4$ and ($^{0,4}X_2$-2H) indicate that a single hexose is linked at the 6 position; this assignment is further supported by the ions $^{3,5}A_4$ and $^{0,3}A_4$ (Fig. 2.7). The longer trihexose arm is linked at the 3 position, as indicated by the ion pairs $^{1,3}A_4$ and ($^{1,3}X_2$-2H), in conjunction with the $^{0,2}X_2$ ion, ruling out the possibility of a 2 link for the isobaric pairs $^{2,4}A_4$ and ($^{2,4}X_2$-2H). Therefore, it can be concluded that a single hexose is linked 1,6 to the first hexose on the core with the trisaccharide series linked 1,3 to the same residue. These results suggest the less abundant *C. parvum N*-glycan is likely $Man_5GlcNAc_2$ and has a structure identical to $Hex_6HexNAc_2$ without the terminal 1,3 linked hexose on the long arm. While these methods cannot differentiate between isobaric monosaccharides, (e.g., mannose from glucose), these methods can accurately define the topology of the glycan and the linkages connecting each monosaccharide. The structures we have defined are consistent with the *N*-glycan structure which has been proposed on the basis of the

presence or absence of the highly conserved *N*-glycosylation biosynthetic pathway enzymes identified in the *C. parvum* genome (Table 2.2 and Fig. 2.8). The topology of the $Hex_5HexNAc_2$ structure defined here, presumably $Man_5GlcNAc_2$, is different from the $Man_5GlcNAc_2$ glycoform produced when host $Man_9GlcNAc_2$ is processed by ER mannosidase 1 in higher organisms, which has 1,3 and 1,6 dimannosyl branches off the first 1,6-linked Man (112, 355).

Confident Assignment of *Cryptosporidium parvum N*-Glycosylated Peptides.

To rule out the possibility of host cell contamination that could occur since *C. parvum* is an obligate intracellular parasite, the MS/MS spectra of glycopeptides from a whole oocyst lysate were manually interpreted. Peptides generated from a trypsin digestion of proteins isolated from oocysts were separated on a reversed phase-C18 nanoflow column interfaced to a mass spectrometer, as described in the methods section. Extracted oxonium ion chromatograms for *m/z* 204.08 (HexNAc) and *m/z* 366.13 (HexNAc-Hex) were very abundant throughout the MS/MS spectra recorded across the HPLC separation (Fig. 4A). These XIC pointed out which spectra should be manually interpreted to obtain sequence information on both glycan and peptide. A representative HCD spectrum for $[M + 2H]^{2+}$ *m/z* 1092.9426 eluting at 15 min, marked by the carat in Fig. 2.4A, is interpreted in Figs 4B and 4C. Ions containing sequential glycosidic bond fragments dominate the spectrum and provide the linear sequence for $Hex_6HexNAc_2$ (Fig. 2.4B). The glycosidic fragments could be traced down to the aglycon, the peptide with $[M + H]^+$ *m/z* 806.3992. Lower abundance peptide backbone fragments are observable in the magnified view of the spectrum (Fig. 2.4C). The complete y-series is interpretable,

starting from y1 *m/z* 175.1189*,* to the full-length peptide, observed at *m/z* 806.3992, thus revealing the peptide sequence NSTTEVR, and indicating the $Hex_6HexNAc_2$ was linked to Asn (Fig. 2.4C). Checking the peptide sequence against the NCBI nr and *C. parvum* proteome databases, utilizing the blastp algorithm, revealed that the peptide belongs to the *C. parvum* protein POWP1 (Table 2.1). In summary, the linear glycan sequence, conjugation site, and complete peptide sequence, can all be determined from a single spectrum.

This method of manual interpretation was continued systematically for the remaining spectra containing one or more oxonium ion(s). Many of the most abundant peptides mapped to the same protein, POWP1. Five *N*-glycosylation sites were mapped to this protein; all contain either $Hex_5HexNAc_2$ or $Hex_6HexNAc_2$ (Fig. 2.5 and Table 2.1). Many of the MS/MS spectra assigned to glycopeptides contained Y_1 and Y_2 ions that arose via glycosidic cleavages adjacent to the HexNAc residues in the chitobiose core, with charge retention on the peptide fragment, but none of these MS/MS spectra contained (Y_1 + 146) or (Y_1 + 162) ions that would indicate the presence of a deoxyhexose or hexose branch on the inner HexNAc residue. No spectra indicated the presence of glycopeptides that did not originate from *C. parvum*. This result demonstrated that the preparation was clean from contaminating host material, and assured that the released *N*-glycans are of parasite origin, as the results had already suggested (Figs. 2.2, 2.3, 2.7, 2.8). The manual interpretations of the glycopeptide spectra are consistent with the results obtained by analysis of the released glycans and underscore the very limited repertoire *N*-glycans in this organism. This information could then be

applied to perform semi-automated database searches to dig deeper in the spectra, allowing for faster processing of replicate samples.

Heavily Glycosylated Proteins Include an Immunodominant Protein (Gp900) and a Putative Oocyst Wall Protein (POWP1).

Thirty-two tryptic peptides with occupied *N*-glycan sites were identified. These peptides derive from 16 glycoproteins, which include the vaccine candidate Gp900 and probable oocyst wall protein POWP1 (Fig. 2.5) (31, 33-35). Some proteins (*e.g.*, Gp900) but not others (*e.g.*, POWP1) showed a higher relative abundance for $Hex_6HexNAc_2$ versus $Hex_5HexNAc_2$ (Table 2.1), as did the released *N*-glycans (Fig. 2.1) (35). No other *N*-glycoforms were detected.

Like Gp900, two *Cryptosporidium N*-linked glycoproteins that are also unique (UCG1 and UCG2) contain long runs of Thr, which are likely modified by *O*-linked GalNAc (31, 33, 53). Five other unique glycoproteins with occupied *N*-glycan sites (UCG3 to UCG7) remain uncharacterized. Other observed glycoproteins have analogs elsewhere in apicomplexa: the glideosome-associated protein (GAP50), three putative adhesion proteins with a *Limulus* coagulation factor C lectin (LCCL) domain (CCp1, CCp2, and FNPA), and a copper amine oxidase (CAO) are conserved throughout apicomplexa (53, 63, 356, 358, 379, 381, 382). An *O*-GalNAc transferase 4 is present in apicomplexans and mammalian hosts, while GMC oxidoreductase (GMCO) is present in apicomplexans, metazoans, fungi, plants, and bacteria (125) .

Nearly 90% of the Occupied *N*-Glycan Sites Contain Thr Rather than Ser.

N-glycans of *C. parvum* are not used for quality control of protein folding, and there is no positive selection for *N*-glycan sites in its secreted proteins (111, 112, 359). However, we observed a large difference in the rate of occupancy of potential *N*-glycosylation sequons. Despite the 5:3 ratio of Thr (100) and Ser (61) in the second position relative to Asn, the number of occupied *N*-glycan sites overwhelmingly (9:1) favors Thr (35) over Ser (4), as shown in the WebLogo in Fig. 2.6 and the data in Table 1. (374). Notably, only 11 total spectra were assigned to peptides with the asparagine modified in an NxS sequon, compared to the 412 that correspond to *N*-glycosylation on the NxT motif (Table 2.1, Appendix 2B).

Discussion

Alg enzymes, which are required for the synthesis of *N*-glycan precursors, are reliable predictors of the types of *N*-glycans transferred to the nascent peptides, because these glycosyltransferases are constitutively expressed in the ER (110, 355). Two peculiarities present themselves with regards to the Alg enzymes of *C. parvum*. First, the Alg13 peptide of the glycosyltransferase that adds the second GlcNAc to the pyrophosphate-linked precursor can easily be identified in *Cryptosporidium muris* and in all other organisms that make *N*-glycans, but Alg 13 is absent from the predicted proteins of *C. parvum* and *C. hominis* (Table 2.2) (63, 356-358). Second, while the *C. parvum N*-glycan precursor is predicted to be $Glc_2Man_5GlcNAc_2$, $Hex_7HexNAc_2$ was absent from the *N*-glycans released with PNGase F and from tryptic glycopeptides (Figs. 2.1 to 2. 4 and Table 2.1). This result suggests that Alg8, which adds the second glucose to the *N*-

glycan precursor, is not active, or a glucose residue is rapidly removed by glucosidase 2 from $Glc_2Man_5GlcNAc_2$ after it is transferred to the nascent peptide. In contrast, *T. gondii*, which has a predicted *N*-glycan precursor composed of $Glc_3Man_5GlcNAc_2$, has been shown to have glycoproteins containing $Hex_8HexNAc_2$ and $Hex_7HexNAc_2$ (53).

While it is well-known that the oligosaccharyltransferase (OST) that adds *N*-glycans to the nascent peptide prefers *N*-glycan sites with Thr over those with Ser, such a strong bias for Thr as that observed here for occupied *N*-glycan sites of *C. parvum* has not previously been described, to our knowledge (111, 383). The composition of the *C. parvum* OST, which includes the catalytic Stt3 subunit and three non-catalytic subunits (Table 2.2), is similar to that found in other apicomplexans, while the OSTs of some parasites (e.g., Giardia and Trypanosoma) only contain Stt3 (76).

The binding of the anti-retroviral lectin cyanovirin-N to *C. parvum* strongly suggested that the parasite contains a high mannose *N*-glycan (35, 361). Cyanovirin-N also binds to Entamoeba and Trichomonas, each of which builds its *N*-glycans from a precursor composed of $Man_5GlcNAc_2$ (384-386). The *N*-glycan profile of *C. parvum* differs from those of the other parasites in the relative abundance of $GlcMan_5GlcNAc_2$ and the absence of mannosidase products ($Man_4GlcNAc_2$ and $Man_3GlcNAc_2$) and/or hybrid and complex *N*-glycans, which contain LacNAc arms (Trichomonas) or galactose capped with Glc (Entamoeba) (386, 387). Other parasites (Trypanosoma, Leishmania, and Acanthamoeba) and Dictyostelium have *N*-glycan precursors with three mannose arms and make numerous complex *N*-glycans that contain LacNAc, fucose, and xylose (87, 388-390). Finally, the *N*-glycans of the mammalian hosts (mice, humans, cats, etc.)

are much more complex than those of *C. parvum*, which are remarkable for their simplicity (125). Whether the high mannose *N*-glycans of *C. parvum* are involved in antigen masking and/or pathogenesis, as has been shown for high mannose *N*-glycans on gp120 of HIV and on HA of influenza virus, remains to be determined (391-393). It has been established that many of the *C. parvum* proteins which elicit a strong immune response are *N*-linked glycoproteins (31, 36). Attempts have been made to develop vaccines from several of these glycoproteins; however, the critical details such as which amino acids are modified and with what glycan structure(s) were left unanswered (394, 395). The results we have presented here fill in the missing details regarding the *N*-glycosylation of the immunodominant antigen Gp900, and we also expand upon the number of *N*-glycosylated proteins previously described in the literature. Of particular interest is the abundant and densely glycosylated protein POWP1. The function of POWP1 remains to be determined. These details may be crucial in providing a means to finally developing an effective, synthetic glycoprotein or glycopeptide-based vaccine against cryptosporidiosis.

Tables

Name	Ac. Number	Unique *N*-Glycosylated Peptides	# Spectra $Hex_5HexNAc_2$	# Spectra $Hex_6HexNAc_2$
gp900	cgd7_4020	(107)-RMVDPVSLMLFD***N***STGVMYDPNTNSILEGSIAGIR-(141)	2	3
		(989)-SGNLVHPYT***N***QTMSGLSVSYLAAK-(1012)	6	0
		(1224)-LINPTN***N***NTMDSSFAGAYK-(1242)	4	0
		(1388)-DPVTNTQYS***N***TTGNIINPETGK-(1409)	0	1
		(1460)-LPIPGSVAGDEILTEVL***N***ITTDEVTGLPIDLETGLPR-(1496)	20	20
		(1497)-DPVSGLPQLP***N***GTLVDPSNK-(1516)	17	8
		(1524)-SGFI***N***GTSGEQSHEK-(1538)	11	0
POWP1	cgd2_490	(31)-***N***STTEVR-(37)	1	3
		(195)-***N***QTSSSGNNPVNNLLNR-(211)	7	44
		(261)-NNPLY***N***ETSISSDGK-(275)	6	47
		(277)-Y***N***DTASPIK-(285)	11	48
		(286)-TPEIVYY***N***NTSNLR-(299)	3	58
UCG1	cgd5_1210	(328)-APGSSNTM***N***QTTNLNNER-(345)	5	2
		(362)-FALSPL***N***GTEVAPLFSK-(378)	0	10
		(97)-VLLG***N***DSTVK-(106)	1	4
UCG2	cgd6_710	(130)-GTIY***N***ITSVDDLIQNSR-(146)	0	22
UCG3	cgd1_640	(122)-EHVF***N***VTGQVPTLGEVK-(138)	1	1
UCG4	cgd1_660	(290)-EIDDIVPH***N***ETIMK-(303)	17	3
UCG5	cgd2_1290	(37)-ATNQT***N***DSWFNLDLLR-(52)	0	1
UCG6	cgd3_660	(58)-A***N***VSTIFGDLLNSK-(71)	0	1
UCG7	cgd8_4660	(175)-ISFLESGSITET***N***FTMSTYR-(194)	0	1
		(195)-***N***ETGLLTNPK-(204)	0	6
GMCO	cgd2_2510	(209)-VYNLF***N***VSDHGFR-(221)	0	1
		(236)-***N***NTVIETSPVDILTNHLVTK-(255)	0	4
GAP50	cgd2_640	(109)-N***N***VTYDSNNDIFPR-(122)	0	2
		(169)-VSEQAFQNL***N***ATLHYGHK-(186)	0	3
COA	cgd3_3430	(994)-VTLFV***N***K-(1000)-T	0	1
O-GNT4	cgd7_1310	(311)-SL***N***ETQSGVDLEQR-(324)	1	0
CCP1	cgd7_1730	(339)-VSINLSA***N***MTYQLK-(352)	3	0
CCP2	cgd7_300	(44)-SIIDTQDLGES***N***DTKK-(59)	0	2
		(59)-KL***N***ETQILSDAYEANINK-(76)	0	1
FNPA	cgd7_4810	(144)-Y***N***STCGSQQSIVSSR-(158)	0	10

Table 2.1. Summary of Identified *N*-Glycosylated Peptides

The table contains the list of proteins, peptides observed to contain intact *N*-glycans, as well as the number of spectra corresponding to the HexNAc2Hex5 or $HexNAc_2Hex_6$ glycoforms.

Protein	***C. parvum***	***T. gondii***
Alg7	cgd5_2240	TGGT1_244520
Alg13	absent*	TGGT1_268340
Alg14	cgd7_4930	TGGT1_207070
Alg1	cgd7_1810	TGGT1_230590
Alg2	cgd1_230	TGGT1_227790
Alg11	cgd4_2990	TGGT1_246982
DPM1	cgd5_2040	TGGT1_277970
Alg5	cgd5_2590	TGGT1_216540
Alg6	cgd4_3120	TGGT1_262030
Alg8	cgd1_2100	TGGT1_314730
Alg10	absent	TGGT1_321660
Gls 1	absent	TGGT1_242020
Gls2-α	cgd8_1420	TGGT1_253030
ER MNS1	absent**	absent
Golgi MNS2	absent	absent
UGGT	absent	absent
Calnexin	absent	TGGT1_310320
ERGIC53	cgd6_5140	TGGT1_258950
STT3	cgd6_2040	TGGT1_231430
WBP1	cgd2_1650	TGGT1_203970
Ribophorin1	cgd6_5070	TGGT1_202572
DAD1	cgd5_2300	TGGT1_305870

Table 2.2. Glycosyltransferase Enzymes Predicted from the Genomes of *C. parvum*, and a Related Organism, *T. gondii*.

Figures

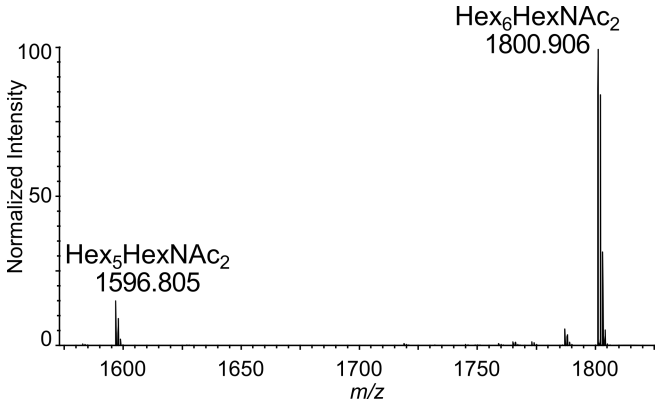


Fig. 2.1. MALDI-TOF MS: Released, Deutero-Reduced Permethylated *N*-glycans [M + Na]$^{1+}$

Total *N*-glycans released from *C. parvum* glycoproteins with PNGase F, reduced with sodium borodeuteride, and permethylated. Only two glycoforms are observed: $Hex_5HexNAc_2$ and $Hex_6HexNAc_2$.

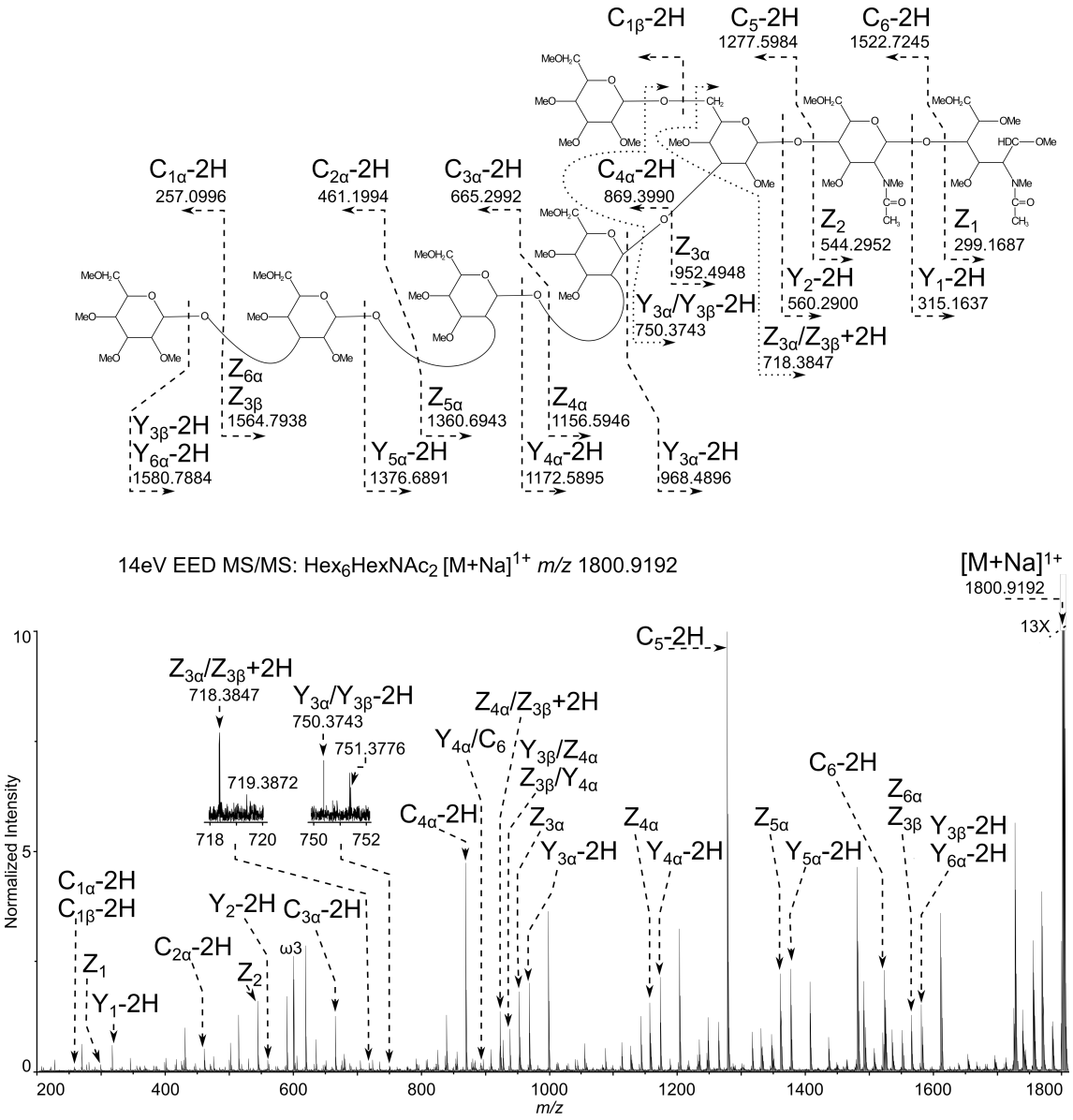


Fig. 2.2. Topology of the Most Abundant Glycoform $Hex_6HexNAc_2$ Determined by EED FT-ICR MS/MS. 14-eV EED FT-ICR MS/MS: $Hex_6HexNAc_2$ $[M + Na]^{1+}$ *m/z* 1800.9192.

Glycosidic fragments provide topological information. The spectrum is labeled only with glycosidic fragments that indicate the topology of the glycoform, revealing a single long arm and an unmodified core. All assignments can be viewed in Appendix 2A.

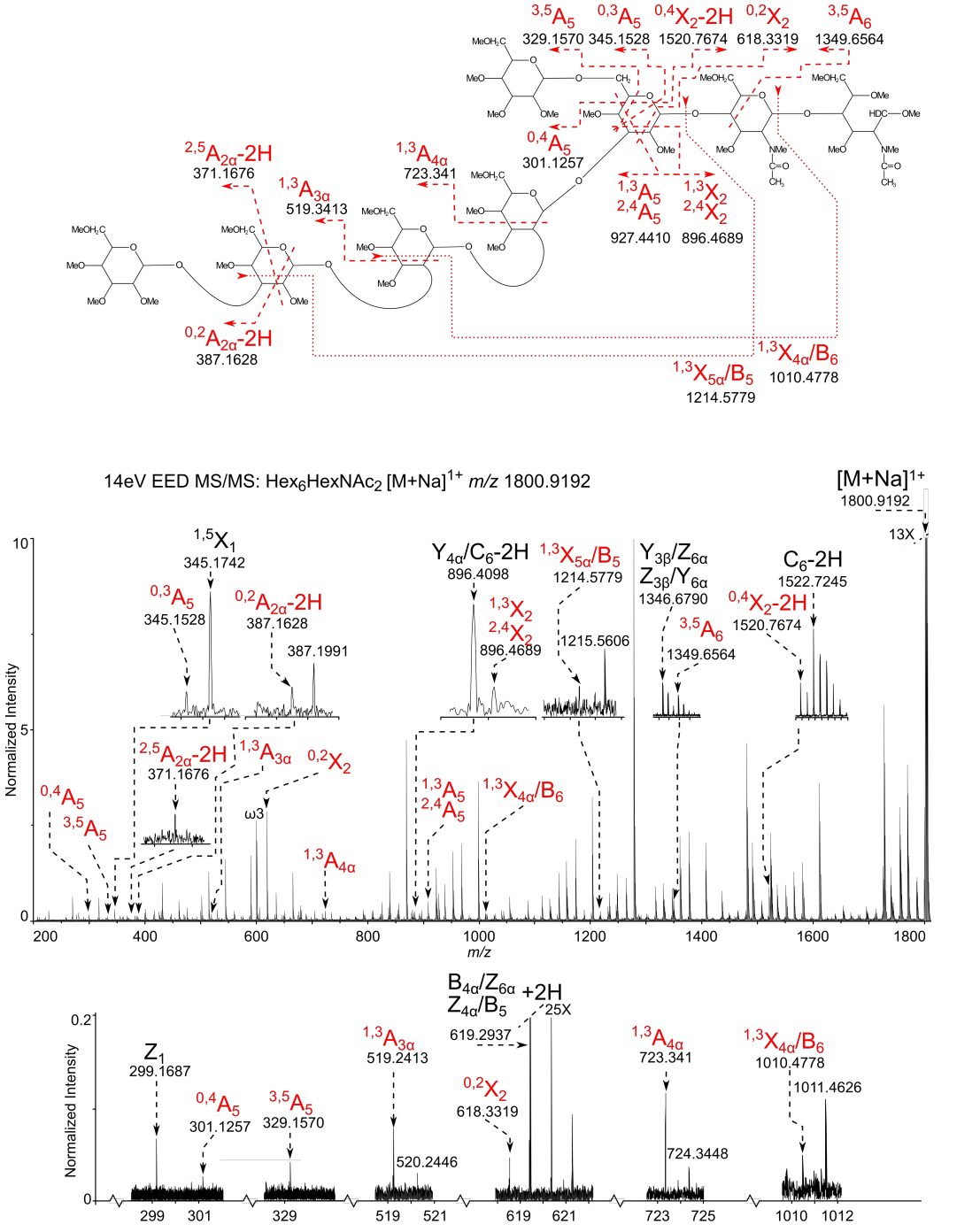


Fig. 2.3. Glycosidic Linkage Determination of the Most Abundant Glycoform $Hex_6HexNAc_2$ Determined by EED FT-ICR MS/MS. 14-eV EED FT-ICR MS/MS: $Hex_6HexNAc_2$ $[M+Na]^{1+}$ *m/z* 1800.9192.

Cross-ring fragments provide linkage information. The spectrum is labeled with only the informative cross ring fragments that provide linkage information. All assignments can be viewed in Appendix 2A.

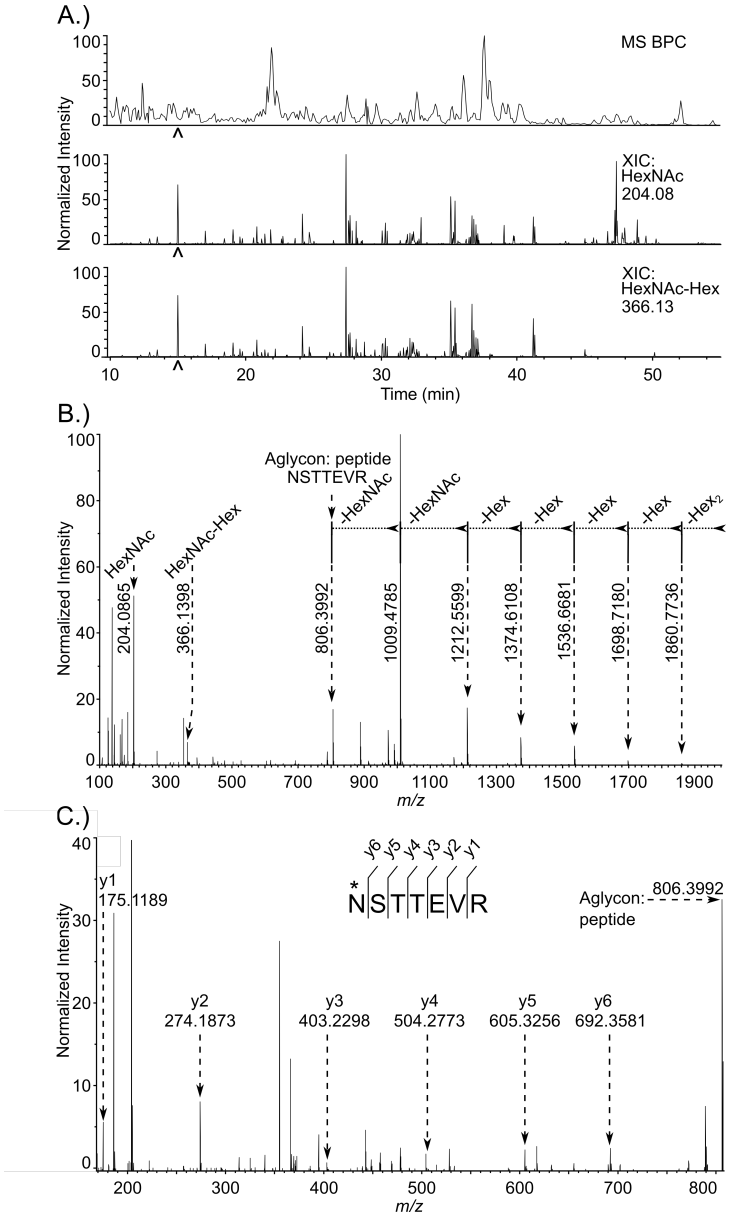


Fig. 2.4. *De Novo* Identification of Glycopeptides from a Whole *C. parvum* Oocyst Lysate.

(A) Base peak and extracted oxonium ion chromatograms from a reversed phase C18 separation. The top trace shows the base peak chromatogram from the MS. The middle and bottom chromatograms are extracted oxonium ion chromatograms (XIC) from the 35-V HCD MS/MS spectra, corresponding to *m/z* 204.08 (HexNAc) and *m/z* 366.13 (HexNAc-Hex), respectively. The carat located at 15 min. indicates the time point for recording of the MS/MS spectrum shown in Fig. 4B and 4C. (B). 35-V HCD MS/MS spectrum of an *N*-glycosylated Peptide: NSTTEVR modified with $Hex_6HexNAc_2$, $[M + 2H]^{2+}$ *m/z* 1092.9426. Prominent glycosidic bond fragmentation is observed, delineating the sequence of the glycan. (C). Peptide sequence of the aglycon. Lower intensity y-ion peptide backbone fragments are observed in the same spectrum. Peptide fragment ion assignments are shown on this magnified view; these extend from y1 (Arg) at *m/z* 175.1189 to the complete aglycon, *m/z* 806.3992, defining the complete peptide sequence as NSTTEVR**.**

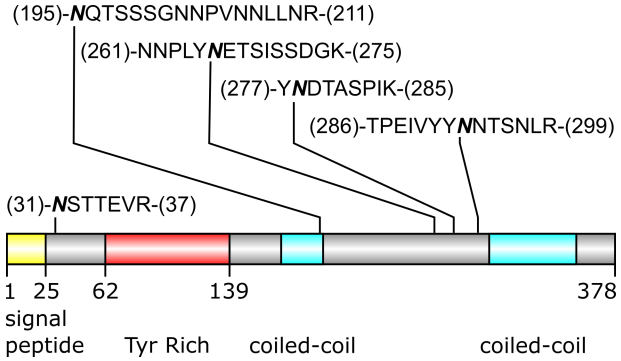


Fig. 2.5. Occupied *N*-Glycosylation Sites of POWP1 (cgd2_490).

The most densely *N*-glycosylated protein is represented as a cartoon schematic. The occupied peptides are shown; the bold and italicized asparagine residues indicate the site of attachment of the glycan.

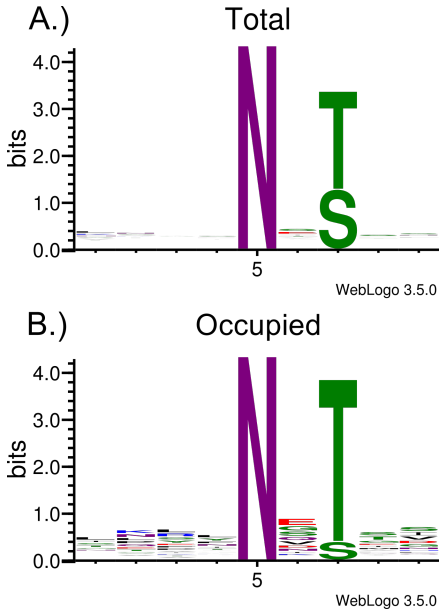


Fig. 2.6. Total and Occupied *N*-Glycosylation Sequons in the Observed *N*-Glycosylated Proteins.

(A). Web-Logos, generated using WebLogo 3.5.0, were compiled for all the peptides containing the canonical *N*-glycosylation sequons from proteins observed to be glycosylated (Total). (B). Peptides observed to be occupied are represented by the (Occupied) logo.

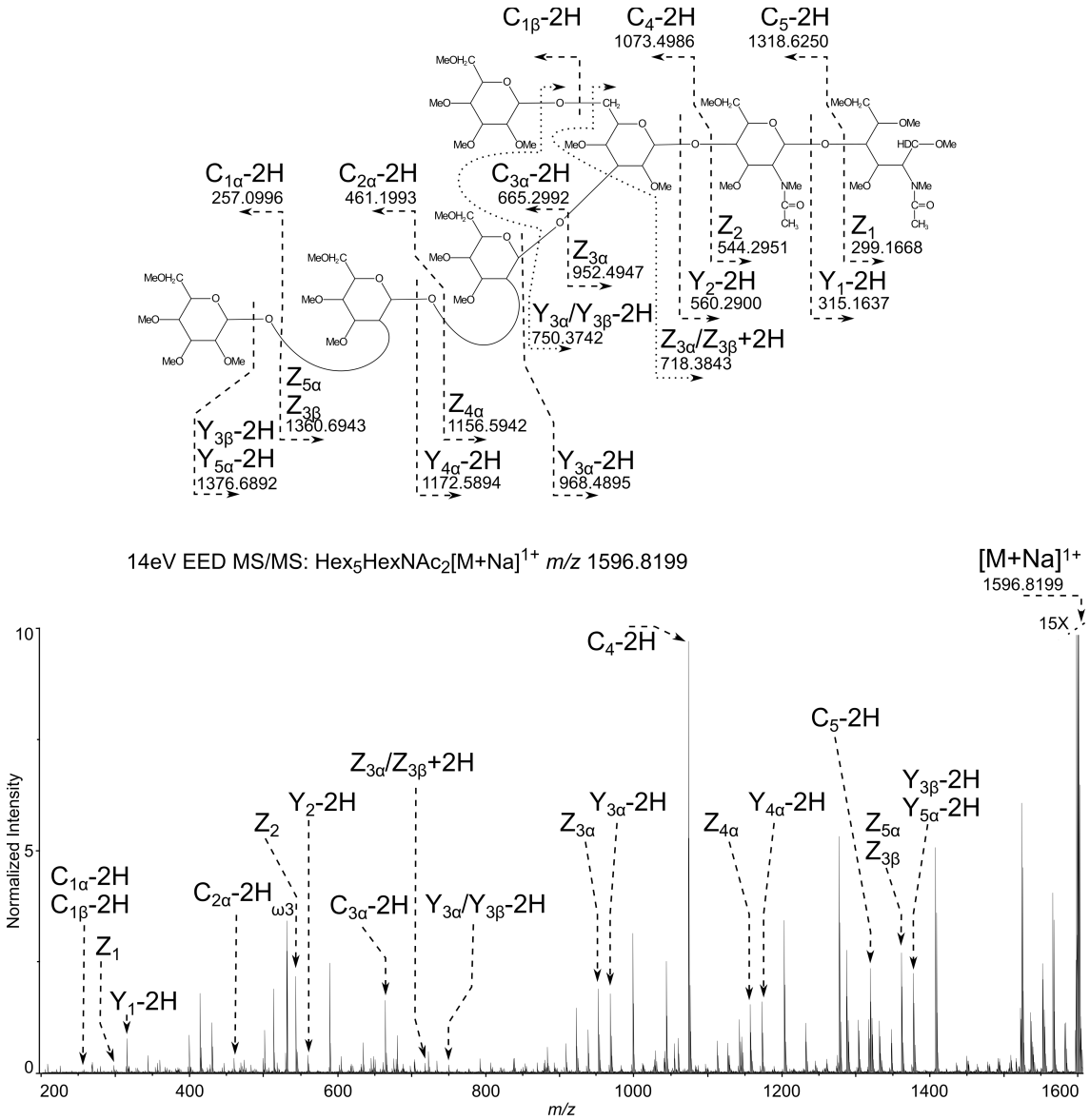


Fig. 2.7. Topology of the Second Most Abundant Glycoform $Hex_5HexNAc_2$ Determined by EED FT-ICR MS/MS. 14-eV EED FT-ICR MS/MS: $Hex_5HexNAc_2$ $[M + Na]^{1+}$ *m/z* 1596.8199.

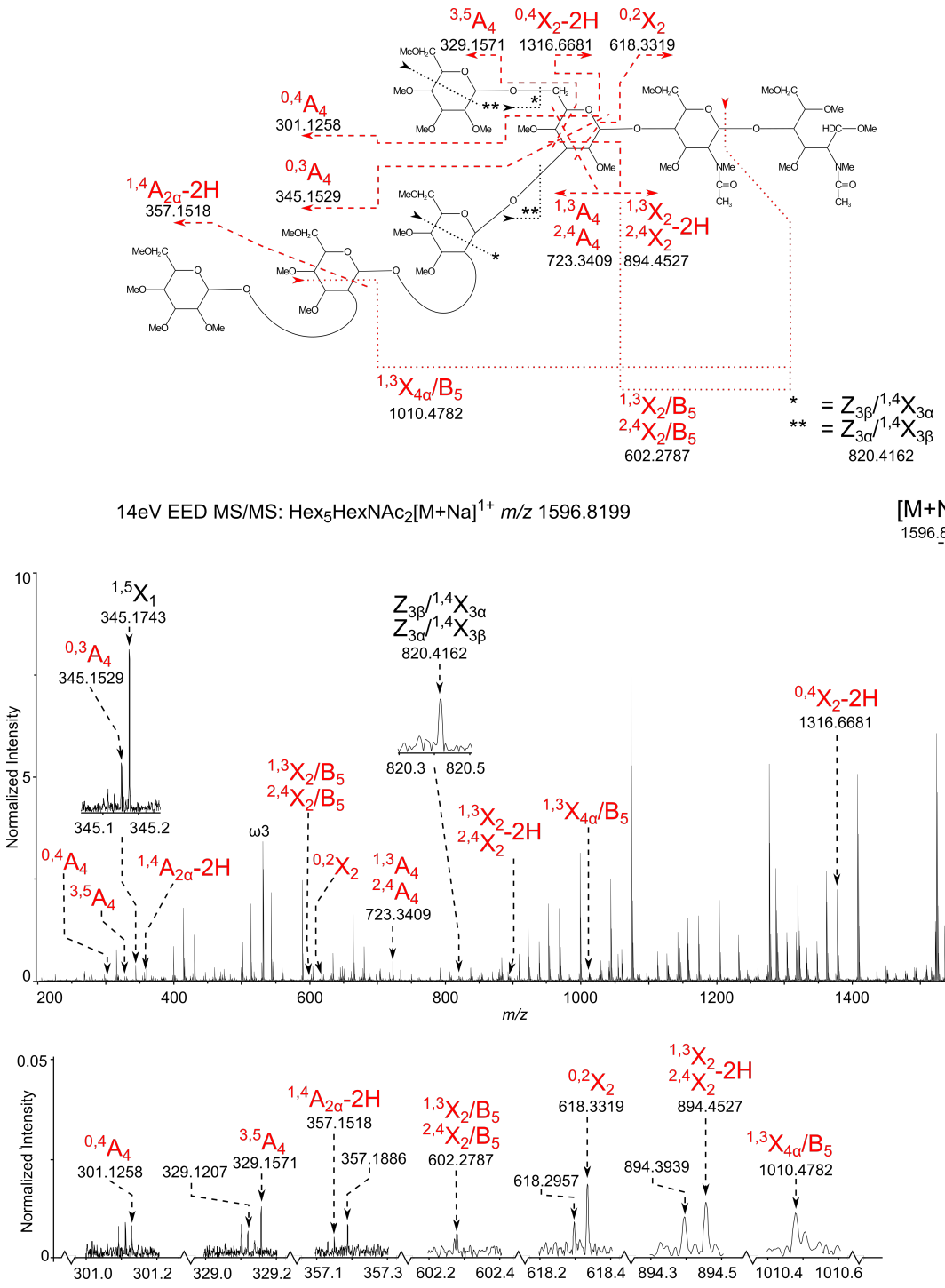


Fig. 2.8. Glycosidic Linkage Determination of the Second Most Abundant Glycoform $Hex_5HexNAc_2$ Determined by EED FT-ICR MS/MS. 14-eV EED FT-ICR MS/MS: $Hex_5HexNAc_2$ $[M + Na]^{1+}$ *m/z* 1596.8199.

Chapter 3. The *N*-Glycosylated Proteins of *T. gondii.*

Introduction

Toxoplasma gondii, as described in the general introduction, is an obligate intracellular parasite capable of infecting most animals and birds. Infection is primarily acquired through the ingestion of oocysts which are secreted in feline feces, the definitive host of *T. gondii* (47, 48, 396). There are several possible sources of infection via oocysts. Pet cats which spend time outdoors and consume animals are at risk of carrying toxoplasmosis, and thus oocysts could potentially contaminate litter boxes in the home, posing a risk to household members (47). It has been demonstrated that oocysts remain infectious for very long periods of time under common environmental conditions (397, 398). In one study, sporulated oocysts remained infectious even after 18 months and two winters buried in soil in Kansas, and oocysts are likely infectious for longer periods of time than could be ascertained (398). Soil contamination poses a risk to gardeners and children who dig contaminated soils and to consumers of improperly washed produce grown in contaminated soil (49). Infection may also occur through the ingestion of undercooked or raw meat of infected animals, where *T. gondii* bradyzoites may reside in tissue cysts (399).

Toxoplasmosis is observed worldwide, with significant percentages of people testing seropositive for previous infection (400). Infection with *Toxoplasma gondii* is often asymptomatic, or only presented with short flu-like symptoms and lymphadenopathy; however, chorioretinitis can occur in immunocompetent individuals (401). Protective immunity keeps infection in check, but latent infections persist in tissue

cysts which can become active systemic infections if the immune system becomes deficient, such as in AIDS or during immunosuppressive therapy (402, 403). Disease in the immunodeficient often leads to severe neurological complications and can be lethal (404-406). Toxoplasmosis in pregnant women puts the fetus as risk, and spontaneous miscarriage, stillbirths, or neonatal death can occur (407).

There are relatively few pharmacological treatments available to treat *T. gondii* infection. These are limited to pyrimethamine co-administered with sulfadiazine, or pyrimethamine and azithromycin (408). There does not appear to be a totally effective treatment for those with AIDS, and infection with *T. gondii* requires long-term continual treatment (409). In pregnant women, spiramycin may be administered to try to prevent congenital transmission, however, this is considered to be an experimental treatment by the Food and Drug Administration, and its efficacy is questionable (410)

N-Glycosylation in *T. gondii*

Although several *in silico* studies have partially described the pathways, the *N*-glycosylation pathway in *T. gondii* remains to be fully characterized. It was first predicted that *T. gondii* has the lipid-linked glycan Dol-PP-$GlcNAc_2Man_5Glc_3$, due to the presence of the yeast homologs Alg7, Alg1, Alg2, Dpm1, Alg5, Alg6, Alg8, Alg10, Stt3, and the absence of Alg3, Alg9, Alg12, in its sequenced genome (110). This prediction was corroborated by the results from Garénaux et al., 2008, where the authors isolated the Dol-PP-glycans from *T. gondii*, finding mostly $HexNAc_2Hex_9$ and a small amount of $HexNAc_2Hex_8$ via MALDI-TOF MS analysis. The $HexNAc_2Hex_9$-containing conjugate is likely host cell contamination, while the $HexNAc_2Hex_8$ originates from the parasite (51).

N-Glycosylated Proteins of *T. gondii*

There have been relatively few reports describing the *N*-glycans or the *N*-glycosylated proteins in *T. gondii*. One early report definitively identified the tachyzoite surface protein Gp23 as being *N*-glycosylated, as *T. gondii* tachyzoites incorporated tritiated monosaccharides into the surface protein gp23, and the radioactivity could be released with PNGase-F treatment (411). Several putative glycoforms with compositions $HexNAc_2Hex_{(5\text{-}9)}$, were identified in another report after PNGase-F treatment of a *T. gondii* lysate; however, there was concern of host cell contribution to the pool of released *N*-glycans (51). The protein glideosome associated protein 50 (Gap50), an integral membrane protein, part of the "glideosomal" complex important for cell invasion, was identified as being *N*-glycosylated by shift on an SDS-PAGE gel after PNGase-F treatment (412). The first study to clearly identify glycopeptides observed, two glycoforms, $HexNAc_2Hex_7$ and HexNAc2Hex8, on a single Gap50 peptide (136)NYTSEALR(143) (52). The same research group followed up with additional experiments, and determined that the two remaining *N*-glycosylation sequons were occupied with similar *N*-glycan compositions (53).

Materials and Methods

Parasite Cell Culture and Protein Extraction

T. gondii type I RH tachyzoites were cultured on a confluent layer of human foreskin fibroblasts (HFF), as previously described (413). The extracellular tachyzoites were separated from any residual host cell material by filtration through a 3-µm Whatman Nucleopore polycarbonate membrane (GE Healthcare Life Sciences, Marlborough, MA),

as previously described (414). Briefly, 1.9x109 extracellular tachyzoites were filtered through 47-mm diameter, 3-µm Nucleopore polycarbonate membranes. The material which passed though the membrane was lightly centrifuged (100 x g, 3 min., 4 °C) to facilitate the removal of residual host debris. To collect the T. gondii tachyzoites, the supernatant was carefully decanted to a new tube and then concentrated by centrifugation (1,450 x g, 15 min., 4 °C). The cell pellet was washed four times with 4 °C PBS, with centrifugation between washes, as described above.

Enrichment of *N*-Glycosylated Proteins

Proteins containing *N*-linked glycans were enriched using methods similar to those described for the enrichment of *O*-fucosylated proteins from *T. gondii* extracellular tachyzoites, with modifications to accommodate the lectin Con-A (415). The concentrated tachyzoites were resuspended in lysis buffer (2% SDS, 0.15 M NaCl, 0.1 mM DTT in 40 mM TrisHCl pH 7.4) to a concentration of ~10^9 cells/mL, heated at 50 °C for 20 min., and were then allowed to cool to RT for 30 min. The lysis solution was then diluted with PD Buffer (0.8% (w/v) *n*-octyl-glucopyranoside, 0.15 M NaCl, 40 mM DTT, 2 mM $MnCl_2$, 2mM $CaCl_2$, 20 mM TrisHCl pH 7.4, EDTA-free complete protease inhibitor tablets (Roche, Basel, Switzerland)), with a volume large enough to dilute the SDS to 0.03%. To the diluted lysate, 12 µg of biotinylated-ConA (Vector Labs, Burlingame, CA) was added to bind the *N*-glycosylated proteins. To ensure maximum binding, the solution was mixed on a rotary mixer for 2 h at 4 °C. Next, 240 µL of Dynabeads® MyOne™ Streptavidin T1 (ThermoFisher Scientific, Waltham, MA) which were pre-washed with PD Buffer, were added and mixed for 30 min at 4 °C on a rotary

mixer. The superparamagnetic beads were collected using a DynaMag™-15 magnetic rack (ThermoFisher Scientific), then washed five times with 4 °C PD Buffer. The *N*-glycosylated proteins were then specifically eluted from the Con-A lectin by incubation with 100 µL of (0.2 M methyl α-D-glucopyranoside/0.2 M methyl α-D-mannopyranoside, in PD Buffer) and lightly shaken at 4 °C for 16 h. The eluted proteins were stored at -80 °C until they were processed for mass spectrometry experiments.

Protein Precipitation and Trypsin Digestion

Two samples were prepared for this study. The first sample originated from the wash (un-bound) fractions from a previous AAL-lectin enrichment experiment, as described previously (415), and were stored at -80 °C before analysis. The second sample was freshly prepared using the methods described above. The proteins were precipitated from solution by the addition of -20 °C MeOH containing 0.1M ammonium acetate for at least 18 h in a -20 °C freezer. The proteins were further processed and isolated as described previously (415).

In-Solution Digestions

The purified and dried protein samples were reconstituted with 50 mM ammonium bicarbonate pH 8.0. The disulfides were reduced with dithiothreitol (DTT) and free sulfhydryls were alkylated with iodoacetamide (IAA). Proteomics grade trypsin was added ~1:20 w/w protein and the solution was incubated for (18 - 20 h) at 37 °C in a mixing heat-block to digest the proteins. The resulting peptides generated from the trypsin digestion were purified using C18 ZipTip® concentrators (EMD Millipore,

Billerica MA) according to the manufacturer's protocol, and as described previously (416).

LC-MS/MS Analyses of In-Solution Digests

The desalted samples were dried in a speed-vacuum and reconstituted in 2% acetonitrile (ACN), 98% LC-MS grade water, and 0.1% formic acid (FA). The peptides were separated using a nano-flow ultra-performance liquid chromatograph (nano-UPLC) (Waters, Milford, MA) using a reversed-phase C18 column. The peptides were separated using a 90-min linear gradient, with a total 120-min run time, the details of which have been described previously (415, 416). The end of the analytical chromatography column was coupled to a NanoMate ion source (Advion, Ithaca, NY), to produce ions in the positive mode. The ions were introduced into a LTQ-Orbitrap-XL-ETD, QE, or a QE Plus mass spectrometer (ThermoFisher Scientific).

The mass spectrometer settings were similar to those described in our previous work (415, 416), with the following exceptions. Experiments performed on the LTQ-Orbitrap-XL, the MS spectra were acquired at 15,000 resolution in the Orbitrap; both CID and HCD MS/MS experiments were performed as technical replicates. For the MS stage, an isolation window of 3 *m/z* was used with a maximum 25 ms injection, fragmentation was performed using 35V CID energy, three microscans, and the product ions were analyzed using the linear ion-trap in the "enhanced" resolution mode. For HCD MS/MS experiments on the LTQ-Orbitrap-XL, an isolation window of 2 *m/z* was used, with a 500 ms maximum injection time, fragments were generated with 35 V HCD, and scanned with two microscans in the Orbitrap analyzer. For experiments on the QE mass

spectrometer, an isolation window of 1.8 *m/z* was used, with a 60 ms maximum injection time with an AGC target of $5x10^5$; fragments were generated with 27 V HCD, and scanned with one microscan. MS/MS experiments on the QE-Plus mass spectrometer were performed using a method devised to maximize simultaneous observation of glycosidic bond fragments and peptide backbone fragments. Here, a two-stepped normalized-collision-energy (NCE), multi-injection method was used. An isolation window of 2.0 *m/z* isolated precursor ions, with a maximum 100 ms maximum injection time per injection (200 ms maximum total per precursor) with an AGC target of $5x10^6$, fragments were generated with 30 V and 35 V HCD NCE for each stage, MS/MS scans were performed with two microscans.

Data Analysis

Manual Interpretation of Glycopeptide Spectra Obtained from LC-MS/MS Experiments

The MS/MS spectra were manually sequenced using the same techniques described for the manual interpretation of *N*-glycosylated spectra from *C. parvum* (416). A brief description is provided here. The raw data from the nano-UPLC-MS/MS analyses, were opened using Qual Browser in the Xcalibur 2.2 software suite (Thermo-Fisher Scientific). Extracted ion chromatograms (XIC) were generated from the MS/MS spectra, corresponding to the *m/z* of oxonium ions, indicative of the fragmentation of glycoconjugates (*m/z* 204.0866 for HexNAc, and, *m/z* 366.1395 for HexNAc-Hex).The spectra which contained one or more oxonium ion(s) were manually sequenced so that

the linear glycan sequence, amino acid localization, and peptide sequence could be obtained from a single spectrum (416).

Automated Database Searches of all Spectra Obtained from LC-MS/MS Experiments

The information gathered from the manual interpretation was then utilized to formulate search parameters for the processing of the UPLC-MS/MS data using the PEAKS software suite version 8.0 (Bioinformatics Solutions Inc., Waterloo, ON, Canada). The searches were performed as described previously (416), with the exceptions of specifying a parent ion mass tolerance of 5.0 ppm, a fragmentation ion mass tolerance of 0.05 Da, and $HexNAc_2Hex_{(5-8)}$ as potential dynamic modifications with a maximum of three variable modifications per peptide. The protein database used for the searches was identical to the one used previously (415). A PEAKSPTM search was performed with the identical parameters, with the consideration of all Unimod modifications. All assigned *N*-glycosylated peptide spectral matches (PSM) were evaluated with a (-logP) value ≥ 13. The verified spectra were tabulated for the two biological samples and their technical replicates, only for the data acquired on the QE+ mass spectrometer.

Bioinformatics

The protein results were visualized using the online web-application Protter (http://wlab.ethz.ch/protter/start/)(417). GPI anchor prediction was performed using the online GPI-prediction tool, (http://mendel.imp.ac.at/gpi/cgi-bin/gpi_pred.cgi) (418). The occupied *N*-glycosylated peptides were used to generate a sequence logo, using Weblogo

ver 3.5 (http://weblogo.threeplusone.com/) (374). Conserved functional domains were predicted using the online tool, InterPro scan (419) (https://www.ebi.ac.uk/interpro/) to search against the InterPro database of conserved protein domains and protein families (420).

Results

Eight proteins were identified as *N*-glycosylated in *T. gondii* tachyzoites, with a total of 44 spectra belonging to *N*-glycosylated peptides (see Table 3.1, Fig. 3.1, and Appendix 3). These proteins are roughly divided into three categories, based upon their defining features (see Fig. 3.1). Group I includes all the proteins that are predicted to have a signal peptide (Gap50, SRS29C, SRS22E, and the hypothetical protein TGGT1_243930). Two of the proteins in group I are predicted to have a GPI-anchor, SRS22E and TGGT1_243930. Group II proteins lack a predicted signal peptide and contain a single transmembrane domain. The two proteins included in group II are a hypothetical protein TGGT1_217680 which does not contain any conserved domains, and the second protein is a CS-domain containing protein (TGGT1_290730). For a description on the CS-domain, please refer to the Prosite database entry PS51203 (http://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS51203) (421). Lastly, group III proteins lack a predicted signal peptide and contain multiple predicted transmembrane regions. The two proteins included in this last group are a hypothetical protein (TGGT1_258870A), and a protein which fits the hidden Markov model for the cleft lip and palate transmembrane protein 1 (CLMPT1) (TGGT1_299110). The InterPro database

match for the CLMPT1 hidden Markov model, matches the PFAM database entry PF05602 (http://pfam.xfam.org/family/PF05602) (422) .

Group I contains the two most abundant proteins identified in this study, Gap50 and SRS29C. Both Gap50 and SRS29C are predicted to have two transmembrane spans which are positioned very close to their N- and C-termini. Furthermore, both of these proteins have two (of four possible) occupied *N*-glycosylation sites identified. However, there is no peptide coverage for the remaining two *N*-glycosylation sites, and therefore their occupancies remain unknown.

Gap50 is the most abundant *N*-glycosylated protein identified in this study based upon the number of assigned peptide spectra. More than half of the 44 spectra corresponding to peptides modified with an *N*-glycan are assigned to Gap50 (24 of 44). The majority of these spectra (18 of 24) are for the peptide (136)NYTSEALR(143). Ten of the 18 spectra contain the modification $HexNAc_2Hex_8$, seven for $HexNAc_2Hex_7$, and only one for $HexNAc_2Hex_6$ (see Table 3.1). The six remaining *N*-glycosylated spectra assigned to Gap50 belong to a different peptide, (76)K.VAANEHISFIASPGSNFLGGVSSLNDTR.W(105). The distribution of glycoforms seems to be slightly different on this peptide: here, the most abundant glycoform appears to be $HexNAc_2Hex_7$ with three spectra observed, followed by $HexNAc_2Hex_6$ with two spectra, and only one spectrum observed for the $HexNAc_2Hex_8$ glycoform (see Table 3.1).

The second most abundant *N*-glycosylated protein is SRS29C, again, based upon the number of observed spectra, corresponding to peptides containing *N*-glycans (see

Table 3.1). Interestingly, this protein has zero spectra which correspond to a peptide containing a $HexNAc_2Hex_8$ modification. Of the two peptides that were identified, nearly all the spectra are for peptides containing the glycan composition $HexNAc_2Hex_6$, two for (197)CSYTENSTLPK(207) and eight for (301)YNCTVPVQLGGEDPSEGSR(319). Only one spectrum was observed for a different glycoform on this protein, $HexNAc_2Hex_7$, which was present on the larger peptide (AA301-AA319). The glycoform distribution is also skewed on this protein, similar to that on Gap50, but in the opposite direction, with a higher abundance of $HexNAc_2Hex_6$, compared to that of $HexNAc_2Hex_8$ (see Table 3.1).

Similar trends are observed for the other proteins identified, where the large hypothetical protein, TGGT1_217680, placed into group II, has two of its nine possible *N*-glycosylation sequons occupied, and in both instances all spectra contain $HexNAc_2Hex_6$. The other proteins, the hypothetical protein (TGGT1_258870A), the CS-domain containing protein, CLMPT1, and SRS22E each have single-spectrum examples of a glycopeptide, far too low for a detailed discussion of distribution; their glycoforms are listed in Table 3.1. However, the single spectrum examples for these proteins are either $HexNAc_2Hex_6$ or $HexNAc_2Hex_8$, which fits the trend that was just discussed

A total of 12 unique *N*-glycosylated peptides containing eight unique occupied sequons have been identified; the peptides are occupied with glycan compositions ranging from $HexNAc_2Hex_{(6-8)}$ (see Tables 3.2, 3.3). The most abundant glycoform based upon the number of observed spectra is $HexNAc_2Hex_6$ with 19 observed spectra, followed by $HexNAc_2Hex_8$ with 14 observations, then $HexNAc_2Hex_7$ with 11 spectra. A

Weblogo for all the observed-to-be-occupied *N*-sequons from these experiments is shown in Fig. 3.2.

The most abundant peptide (136)NYTSEALR(143), from Gap50, is modified with $HexNAc_2Hex_8$ on Asn 136. A comparison of the MS/MS spectra of the same peptide, *m/z* 1328.53 $[M+2H]^{2+}$, as analyzed with the three different instruments (LTQ-Orbitrap CID and HCD fragmentation, as well as the QE, and QE+) which utilized slightly different methods, is shown in Fig 3.4A. Fig 3.4B shows the manual interpretation of the lower spectrum from 3.4A, from the QE+ which used stepped NCE of 30 V HCD and 35 V HCD with multiple injections. The demonstration here is the ability to obtain both the linear glycan's sequence, first with the loss of the mass equal to three hexoses from the parent ion of the monoisotopic $[M+2H]^{2+}$ *m/z* 1328.5298 to the singly charged ion *m/z* 2007.8063. The mass difference of hexose (162.0528 u) can be followed until the mass difference of two sequential HexNAc residues (203.0793 u) is observed, reaching the aglycon peptide $[M+H]^{+}$ *m/z* 953.4677. These data provide the glycan's linear sequence of $HexNAc_2Hex_8$. Here, in the same spectrum, a magnified view is provided to show the detail of sequential losses of masses corresponding to amino acid residues, revealing the sequence NYTSEALR. Thus, in the same spectrum the glycan's linear sequence, its conjugation site, and the amino acid sequence of the peptide are all revealed (Fig 3.4B, Fig 3.4C).

Discussion

The reasons for these observed differences in distributions of glycoforms on the different proteins, and in some cases on different peptides within the same protein

remains unknown. One explanation is that the *T. gondii* glucosidases Glc1 and Glc2 may not work well on certain peptides. The observance of abundant spectra for peptides containing the minimally trimmed glycans $HexNAc_2Hex_8$ and $HexNAc_2Hex_{7,}$ (the presumed product of Glc1) support this as a possibility. For example, the peptide from Gap50 (136)NYTSEALR(143) appears to generate a disproportionate number of spectra for the $HexNAc_2Hex_8$ glycoform (see Table 3.1). One other explanation is that *T. gondii* may transfer different glycans, where the $HexNAc_2Hex_8$ glycoform could originate from the fully assembled Dol-PP-$GlcNAc_2Man_5Glc_3$ and the $HexNAc_2Hex_6$ could originate from the transfer from Dol-PP- $GlcNAc_2Man_5$ with a subsequent glucose transfer by ALG6. The preference for different donors and substrates has been demonstrated in other parasites, such as in *T. brucei*, however, *T. gondii* does not appear to have multiple STT3 variants (86). The other perplexing observation with Gap50, as well as with all the proteins identified, is that the smallest observed *N*-glycoform is $HexNAc_2Hex_6$. This composition is contrary to what would be predicted if Glc1 and Glc2 were to trim off the terminal glucose residues, leaving $GlcNAc_2Man_5$ (observable as $HexNAc_2Hex_5$) (see Fig 3.3 for the predicted un-trimmed *N*-glycan). It could be that the calnexin which *T. gondii* does appear to have, which lacks the protein-protein binding domain, may bind to the glucose on the Asn-linked $GlcNAc_2Man_5Glc$, protecting it from cleavage by Glc2. This is a plausible explanation, since *C. parvum*, which lacks Cxn, is observed to have $HexNAc_2Hex_5$ modified peptides, but no such modification is observed on *T. gondii* proteins (416).

Conclusions

In this report, we were able to identify *N*-glycosylated proteins that are in *T. gondii*. There have been surprisingly few reports in the literature which characterize *N*-glycosylated proteins in this organism, with the exception of Gap50 (52, 53). Our results confirm the previous findings of Fauquenoy, 2008. However, they were able identify additional *N*-glycosylated peptides from Gap50 in their 2011 study, which we were not able to observe, likely due to their use of chymotrypsin which helps cover the gap in peptide coverage that results when only trypsin is used for digestion (see Fig 3.1 for potential trypsin cleavage sites). This report shows our preliminary observations. The optimizations of the Con-A pulldown conditions, as well as the utilization of additional proteases should expand the peptide coverage, as well as expand the numbers of observed *N*-linked proteins. Interestingly, several of the *N*-glycosylated proteins identified in this study such as Gap50 and the SRS proteins, are known virulence factors. It has previously been demonstrated that Gap50 plays a crucial role in host cell invasion, acting as part of the glideosomal apparatus, and that the *N*-glycosylation of Gap50 is key to its functionality (53). In addition, two of the other proteins identified in this study, SRS22E and SRS29C, belong to the SAG1-related sequence (SRS) superfamily of proteins (423).

This family of proteins contains over 120 members, many of them GPI-anchored proteins located on the surface of *T. gondii,* which are thought to play a role invasion of host cells, as well as evasion of the immune system (423). Our findings support the hypothesis that the glycans are transferred to Asn in *T. gondii via* the lipid-linked glycan precursor Dol-PP-$GlcNAc_2Man_5Glc_3$, due to the presence of $HexNAc_2Hex_8$ as one of the

dominant glycoforms. Follow-up studies would be useful to expand the peptide coverage of the proteins described in this report, since many of the peptides which contain an *N*-glycosylation motif were not observed. The optimization of the Con-A pull-down, as well as the use of different proteases would considerably expand the list of detected peptides. Furthermore, it would be interesting to perform analyses on the released *N*-glycans to verify their topologies and their linkages, utilizing techniques described previously, such as EED MS/MS on the released deutero-reduced and permethylated *N*-glycans(416). Furthermore, treatment of released *N*-glycans with mannosidases, in conjunction with MS/MS analyses, could offer additional insight into the glycans structures. An in-depth analysis of the *N*-glycans from *T. gondii* could help answer the remaining questions surrounding the unexpected *N*-glycoform distributions. It would be informative to perform functional assays using recombinant *T. gondii* Glc1/Glc2 to test on synthetic substrates, to determine whether or not *T. gondii* is capable of completely trimming the terminal glucoses from Glc(α1-2)Glc(α1-3)Glc(α1-3)-Man-X, or, if the trimming is arrested at Glc(α1-3)-Man-X. Another informative experiment would be to determine whether or not the *T. gondii* OST is capable of transferring different glycans from different lipid-linked oligosaccharide donors, and, if there are preferences for substrates.

In summary, the work presented here expands the number of proteins identified in *T. gondii* that are confirmed to be *N*-glycosylated. Our preliminary results do suggest there is preference for the modification of the NxT *versus* the NxS motif, similar to what was observed in *C. parvum* (416). Furthermore, our observations confirm previous

observations that there appear to be two distinct populations of *N*-glycans on the glycoproteins of *T. gondii* (52, 53).

Tables

Protein Name	Acc.#	Peptide‡	#Spec.	Glycoform
GAP50	TGGT1_219320	(76)K.VAANEHISFIASPGSNFLGGVSSL**N**DTR.W(105)	2	$HexNAc_2Hex_6$
			3	$HexNAc_2Hex_7$
			1	$HexNAc_2Hex_8$
		(135)R.**N**YTSEALR.T(144)	1	$HexNAc_2Hex_6$
			7	$HexNAc_2Hex_7$
			10	$HexNAc_2Hex_8$
SRS29C	TGGT1_233480	(196)R.CSYTE**N**STLPK.I(208)	2	$HexNAc_2Hex_6$
		(300)K.Y**N**CTVPVQLGGEDPSEGSR.P(320)	8	$HexNAc_2Hex_6$
			1	$HexNAc_2Hex_7$
hyp. protein	TGGT1_217680	(406)R.T**N**STLFESQLR.E(418)	1	$HexNAc_2Hex_6$
		(457)R.GV***N***VTIDR.H(466)	2	$HexNAc_2Hex_6$
hyp. protein	TGGT1_243930	(696)K.MN**N**ETVLYEPDTEIIEK.T(714)	1	$HexNAc_2Hex_8$
		(763)K.TMNSEGVISDGLQSQLPV**N**HTR.L(786)	1	$HexNAc_2Hex_8$
hyp. protein	TGGT1_258870A	(178)R.AHTGDERPF**N**VTTGSSER.R(197)	1	$HexNAc_2Hex_8$
CS-dom. prot.	TGGT1_290730	(333)K.DLDDFHHG**N**YTAR.Y(347)	1	$HexNAc_2Hex_6$
CLMPT1	TGGT1_299110	(141)R.N**N**TTLYVHVR.T(152)	1	$HexNAc_2Hex_6$
SRS22E	TGGT1_359770	(36)K.IETCAPDKPISF***N***VTEAGQSILFK.C(61)	1	$HexNAc_2Hex_6$

Table 3.1 Manually Verified Proteomics Search Results of *N*-Glycosylated Peptides from *T. gondii.*

The results from the two samples, including the technical replicates, all analyzed on the QE+ mass spectrometer, and searched using the PEAKS software suite. The results are collated from the PEAKS-PTM searches, the ppm error is shown for the best scoring spectrum (-10logP).

‡Shown are representative peptides to simplify the table. The complete listing of all the spectra used to generate this table is shown in Appendix 3. The abbreviations used for the protein names are as follows: [(GAP50), acid phosphatase GAP50]; [(clptm1), cleft lip and palate transmembrane protein 1]; [(CS-dom. prot.), Co-chaperone (p23-like) domain containing protein]; [(hyp. Protein), hypothetical protein];[(SRS29C), SAG-related sequence 29C]; [(SRS22E),SAG-related sequence 22E].

Table 3.2 Occupied *N*-Glycosylation Sequons

Sequon	#Unique Pep. Seq.
NVT	3
NYT	2
NST	2
NTT	1
NDT	1
NET	1
NCT	1
NHT	1

Table 3.3 Totals of the Different *N*-Glycoforms

Glycoform	Total Spectra
$HexNAc_2Hex_6$	19
$HexNAc_2Hex_7$	11
$HexNAc_2Hex_8$	14

Figures

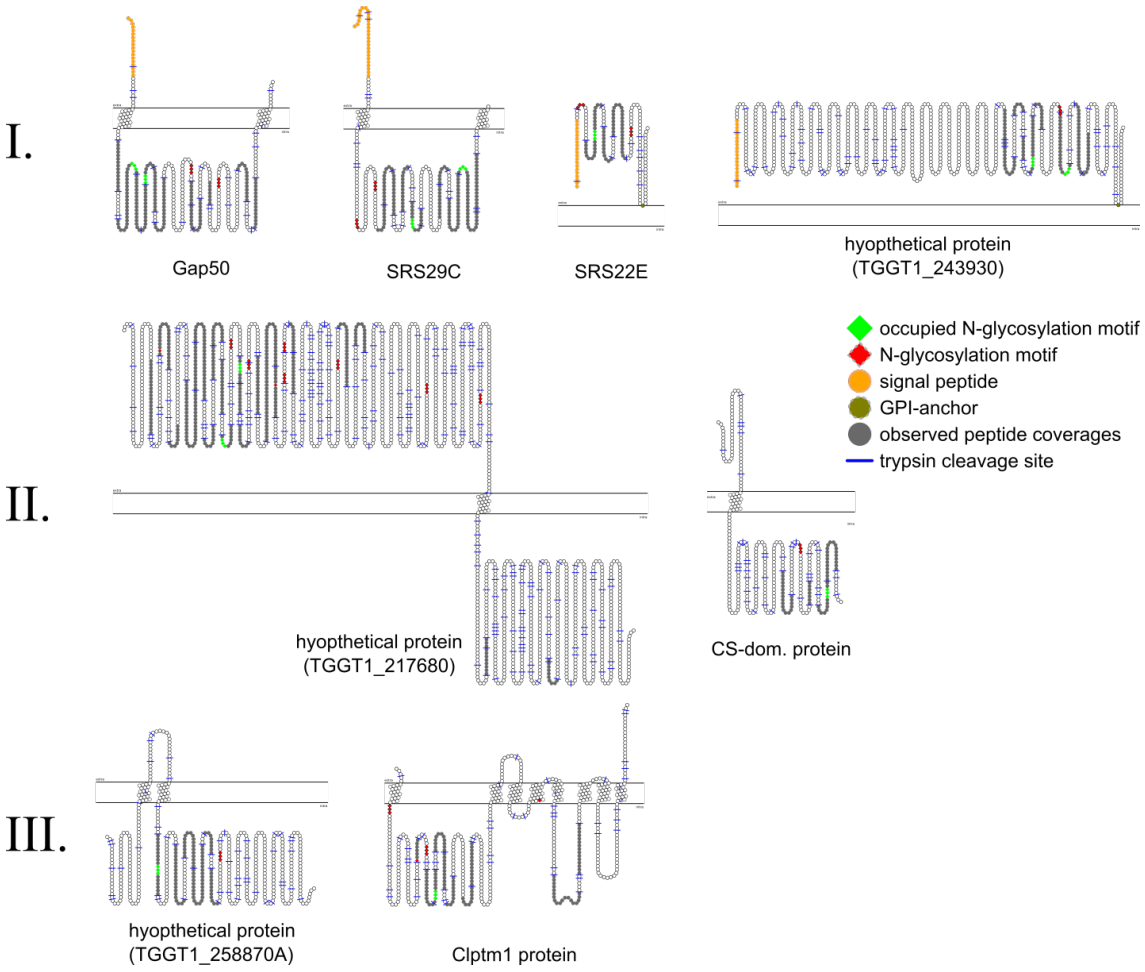


Fig. 3.1 Cartoon Schematics of the *N*-Glycosylated Proteins, Showing Predicted Features, Total Observed Peptides, and Occupied *N*-Glycosylation Sequons.

These proteins are divided into three different groups based upon if there is (I.) a predicted signal peptide, (II.) no signal peptide, but a predicted transmembrane region, and (III.) no predicted signal peptide with the presence of multiple trans-membrane spans. The sequences are marked according to the key, observed peptides are dark grey, signal peptides are orange, a predicted GPI-Anchor is olive color (also indicated with squiggle-line to anchor it to the membrane), observed-to-be-occupied *N*-glycosylated motifs are green diamonds, unobserved canonical *N*-glycosylation motifs are red diamonds (if the peptide was observed, but not observed to be *N*-glycosylated then only the Asn is a red diamond), predicted trypsin cleavage sites are shown as blue lines.

Fig. 3.2 WebLogo of Occupied *N*-Glycosylation Sequons

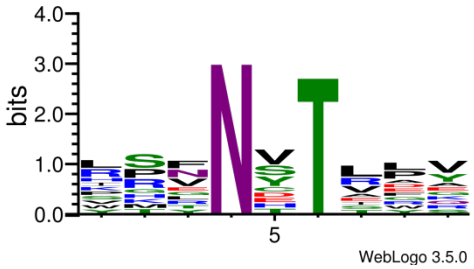
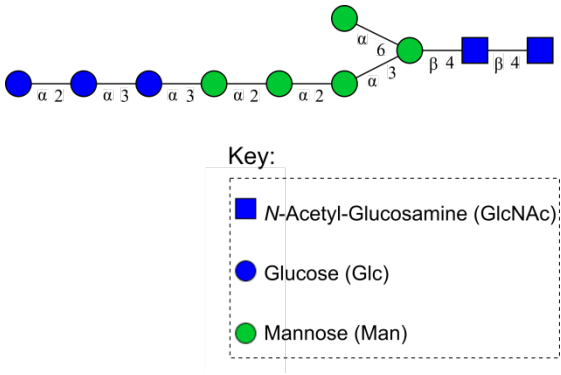


Fig. 3.3 The Predicted Lipid Linked *N*-Glycan of *T. gondii*



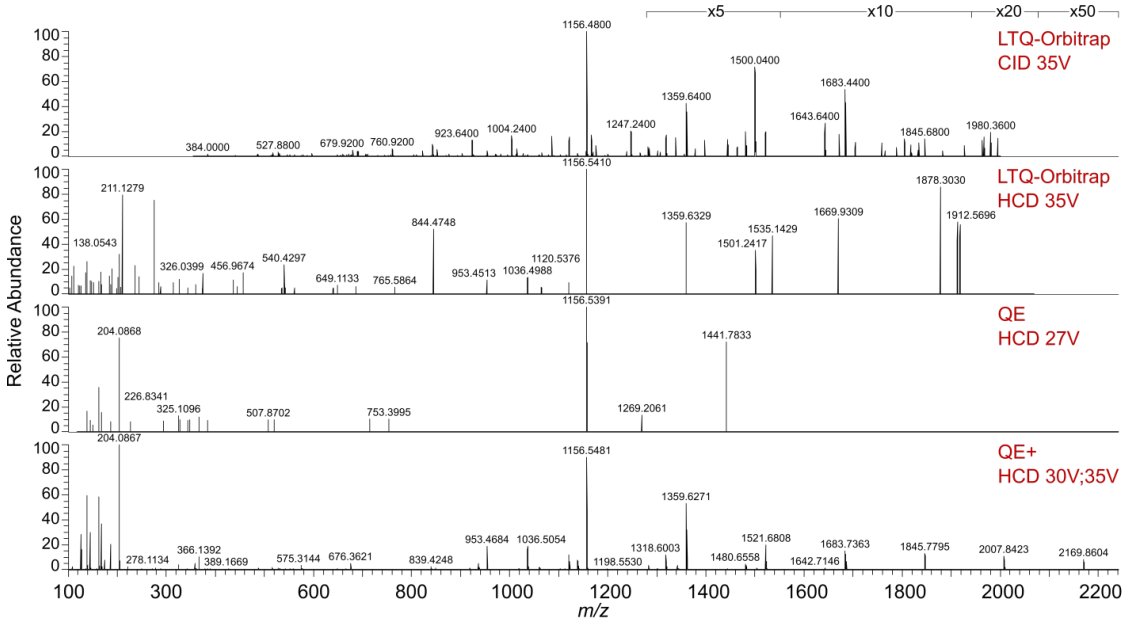


Fig. 3.4A Comparison of MS/MS Fragmentation of the Ion *m/z* 1328.53 $[M+2H]^{2+}$ on an LTQ-Orbi (HCD, CID), QE, and QE+.

This comparison demonstrates the differences in fragmentation of the ion *m/z* 1328.53 $[M+2H]^{2+}$, which corresponds to the glycopeptide NYTSEALR with a $HexNAc_2Hex_8$ glycan modification. The differences in the fragmentation for each spectrum can be attributed in part due to the type of fragmentation (CID or HCD), method of detection (ion trap, Orbitrap), instrument specific differences, and instrument setup (e.g. microscan count, injection time, collision energy, etc.).

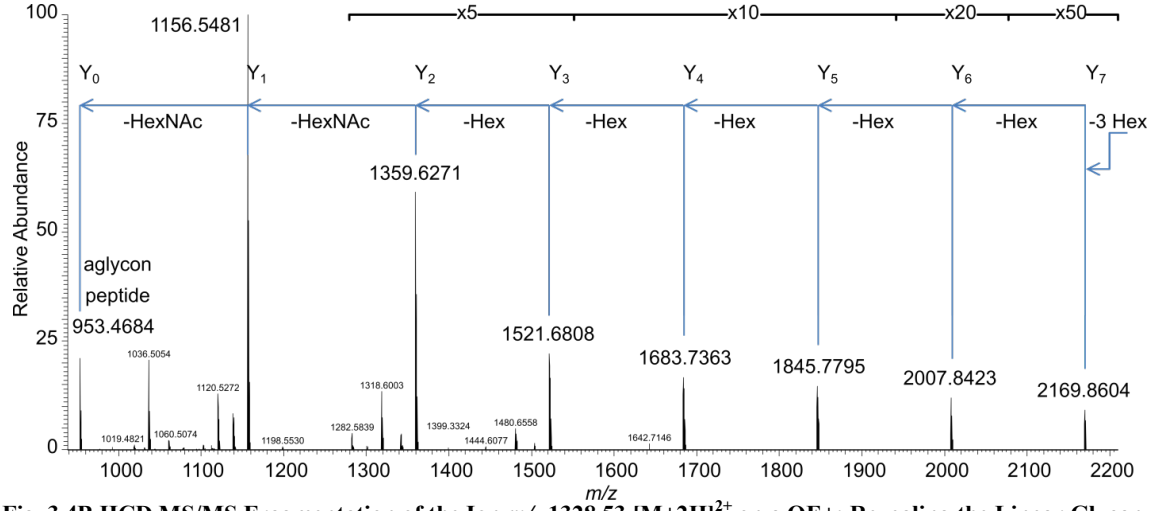


Fig. 3.4B HCD MS/MS Fragmentation of the Ion *m/z* 1328.53 $[M+2H]^{2+}$ on a QE+: Revealing the Linear Glycan Sequence $HexNAc_2Hex_8$.

The ion *m/z* 1328.53 $[M+2H]^{2+}$ was fragmented using HCD on a QE+ mass spectrometer using a method utilizing a multi-injection stepped NCE program as described in the a methods section. This spectrum is the same spectrum shown in the bottom panel of Fig 3.4A, but, a limited range *m/z* (950 - 2,200) is shown. The mass differences calculated from the monoisotopic $[M+H]^+$ parent ion and the largest observed ion (*m/z* 2169.8604), and, between the (*m/z* 2169.8604) ion and sequentially smaller ions down to the aglycon peptide (*m/z* 953.4684), which are equal to the Δmass of one or more sugar(s), reveals a linear glycan sequence of $HexNAc_2Hex_8$.

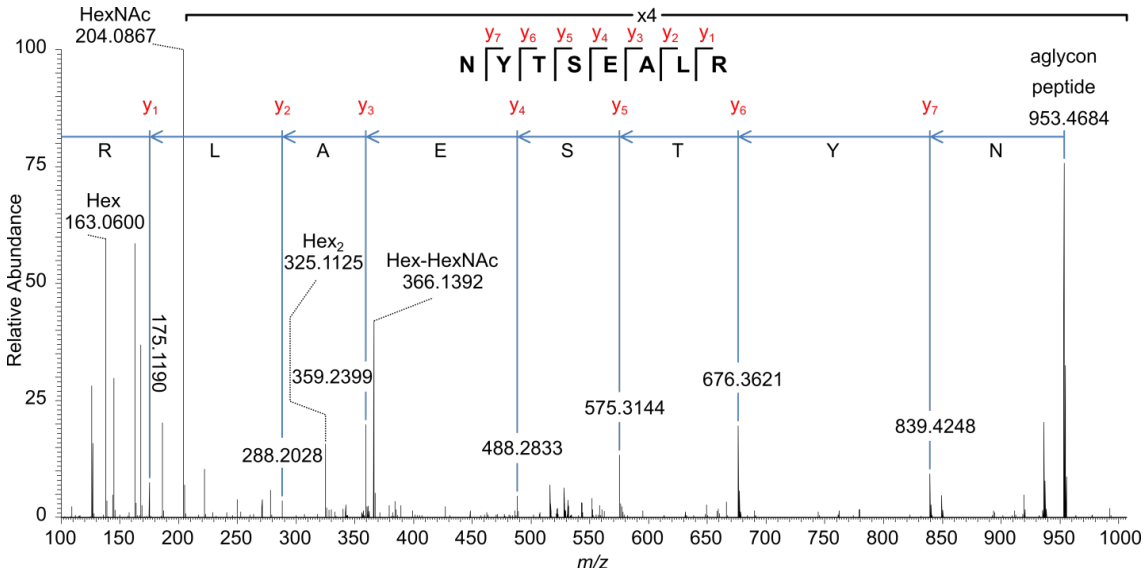


Fig. 3.4C HCD MS/MS Fragmentation of the Ion *m/z* 1328.53 $[M+2H]^{2+}$ on a QE+: Revealing the Peptide Sequence NYTSEALR.

This MS/MS of spectrum the ion *m/z* 1328.53 $[M+2H]^{2+}$, is the same spectrum in the bottom panel of Fig 3.4A, and in Fig 3.4B, but, only the lower *m/z* range from the start of the scan to the aglycon peptide is shown (*m/z* 100-1,000) to reveal the peptide sequence. The complete y-ion series are observed, revealing the peptide sequence NYTSEALR. The spectrum is also labeled to show the abundant oxonium ions (Hex, HexNAc, Hex_2, and Hex-HexNAc)

Chapter 4. The *O*-Glycosylated Proteins of *C. parvum*.

The work contained within this chapter has been submitted to: *The Journal of Biological Chemistry*, an American Society for Biochemistry and Molecular Biology journal. It is presented here in a manuscript format.

The research presented here was exclusively my own with the exception of a single figure which is contained as part of the supplementary data set. This figure was produced by Joshua A. Klein utilizing software which he has been developing as part of his own Ph.D. degree in the Boston University Bioinformatics Program. In addition, Joshua A. Klein helped to re-annotated spectra which are deposited online as part of the entire data-set for publication purposes. That work is not presented here, but, it will be viewable once the manuscript has been accepted and the dataset released to the public. The data has been deposited to the online data repository, the Proteomics Identifications Database (PRIDE), and may be accessed by connecting to http://www.ebi.ac.uk/pride/archive/projects/PXD005989

Cryptosporidium parvum Vaccine Candidates are Heavily Modified with *O*-linked-*N*-acetylgalactosamine or Contain N-terminal *N*-Myristate and *S*-Palmitate.

John R. Haserick, Joshua A. Klein, Catherine E. Costello, and John Samuelson

Abstract

Cryptosporidium parvum (studied here) and *Cryptosporidium hominis* are important causes of diarrhea in infants and immunosuppressed persons. *C. parvum* vaccine candidates, which are on the surface of sporozoites, include glycoproteins with Ser- and Thr-rich domains (Gp15, Gp40, Gp900) and a low complexity, acidic protein (Cp23). Here we used mass spectrometry to determine that *O*-linked GalNAc is present in dense arrays on a glycopeptide with consecutive Ser derived from Gp40 and on glycopeptides with consecutive Thr derived from Gp20, a novel *C. parvum* glycoprotein with a formula weight of ~20 kDa. In contrast, the occupied Ser or Thr residues in glycopeptides from Gp15 and Gp900 are isolated from one another. Gly at the N-terminus of Cp23 is *N*-myristoylated, while Cys, the second amino acid, is *S*-palmitoylated. In summary, *C. parvum O*-GalNAc transferases, which are homologs of host enzymes, densely modify arrays of Ser or Thr, as well as isolated Ser and Thr residues on *C. parvum* vaccine candidates and the N-terminus of an immunodominant antigen has lipid modifications similar to those of host cells and other apicomplexan parasites. The significance of these post-translational modifications is discussed with regards to the function of these proteins and the design of serological tests and vaccines.

Introduction

Cryptosporidium parvum infects humans and cows, while *Cryptosporidium hominis* only infects humans (49, 343, 344). *C. parvum* was first identified as an opportunistic infection and cause of severe diarrhea in AIDS patients (424, 425). In 1993, *C. parvum* contaminated the municipal water supply and a caused a massive outbreak of diarrhea among immunocompetent persons in Milwaukee (6, 345). More recently *C. parvum* has been shown to be the second most important cause (after rotavirus) of diarrhea and death in infants in low resource countries where the parasite is endemic (20, 346, 348). Presently, there are no human vaccines for *C. parvum*, although numerous candidates have been identified (352, 353) and see below. Treatment of *C. parvum* is difficult in populations with the most severe disease: infants and immunosuppressed persons (343, 354).

Oocysts of *C. parvum* have acid-fast walls, which are resistant to environmental insults and to gastrointestinal acids, proteases, and bile (425, 426). Oocysts each contain four infectious sporozoites, which have on their surface Ser- and Thr-rich glycoproteins (*e.g.* Gp900 and Gp40) (31, 33-35, 40, 369). The precursor protein (Gp40/Gp15), which is specific for *C. parvum* and *C. hominis*, is cleaved by a furin-like protease into an N-segment (Gp40) and a C-segment (Gp15) (Fig. 4.1) (42). Subsequently, the N-terminal signal peptide of Gp40 is removed, and a glycosylphosphatidylinositol (GPI) anchor is added to the C-terminus of Gp15 (427). Gp40 contains a domain of 17 consecutive Ser residues followed by Thr-Ser-Thr, while the Ser and Thr residues of Gp15 are dispersed. Gp900, which is much larger than Gp40/Gp15, is a secreted protein present in *C. parvum,*

C. hominis, and *C. muris*; it has four sets of consecutive Thr residues, ranging in length from 33 to 155 residues, as well as dispersed Ser and Thr (31, 33). Gp40 and Gp900, which are shed from the surface of sporozoites during gliding motility, tether sporozoites to the interior surface of the oocyst wall (33, 35, 39), . In contrast, Gp15 is present at the apical end of sporozoites and on the outer surface of the oocyst wall (35).

Polymorphisms in Gp15 and Gp40 have been used to distinguish isolates of *C. parvum*, while recombinant Gp15 has been used to measure the serological response in epidemiological studies (15, 363-367, 428). Vaccination studies have been performed using recombinant *C. parvum* proteins, bacterial vectors (*e.g. Salmonella*), or DNA - encoding *C. parvum* proteins (352, 353, 429). These vaccines either contain no *O*-glycans (bacterially expressed proteins) or may display host *O*-glycans (DNA vaccines). The presence of *O*-glycans (most likely *O*-GalNAc) on *C. parvum* glycoproteins has not previously been detected by mass spectrometry, but it has been suggested by the following observations: The *C. parvum* genome predicts four *O*-GalNAc transferases (*O*-GalNAcTs), and parasite lysates add *O*-GalNAc to synthetic peptides (125). A lectin that recognizes *O*-GalNAc (*Helix pomatia* agglutinin) (HPA) binds to the surface of sporozoites, while binding of a monoclonal antibody (4E9) to Western blots of *C. parvum* proteins is competed by HPA and reduced by treating proteins with an *O*-GalNAcase (15). The *Maclura pomiphera* agglutinin, which binds *O*-GalNAc, dramatically enriches Gp40, Gp900, and other mucin-like glycoproteins of *C. parvum* (35). Lastly, sera from patients infected with *C. parvum* bind to synthetic peptides to which an *O*-linked GalNAc has been added (44).

Cp23, which is also known as the immunodominant antigen, is a small, low complexity, acidic protein present on the surface of sporozoites (30). Monoclonal antibodies to Cp23 partially protect neonatal mice against oral infection with *C. parvum*, while antibodies to Cp23 have more frequently been found in HIV/AIDS patients infected with *C. parvum* but without diarrhea (347, 430). Recombinant Cp23 has been used to demonstrate humoral and cellular immune responses to *C. parvum* in human, cattle, and mouse infections, whereas recombinant Cp23 and DNA-based vaccines have been used to immunize mice and to elicit an innate immune response from mouse and human dendritic cells *in vitro* (349, 350, 366, 431-436). Previous mass spectrometry studies of *C. parvum* sporozoites and oocysts have identified numerous peptides from Gp40, Gp15, Gp900, and Cp23, but none of these studies described post-translational modifications (PTMs), which may include *O*-linked glycans, Asn-linked glycans (*N*-glycans), and fatty acyl chains (35, 44, 125, 427, 437-441). Recently we used mass spectrometry to determine that *C. parvum N*-glycans, which are built on a predicted precursor with a single long mannose arm, appear to be processed by glucosidase-2 but not by ER mannosidases or elongated with Golgi glycosyltransferases (110, 359, 416). The resulting *N*-glycans, which are likely $GlcMan_5GlcNAc_2$ and $Man_5GlcNAc_2$, are remarkable for their simplicity, as compared to the complicated *N*-glycans identified in other protists (389). In this report, we used mass spectrometry to characterize tryptic glycopeptides of lysates of *C. parvum* oocysts and thereby directly determine the number and some of the positions of *O*-GalNAc residues on Gp40, Gp15, Gp900, and a previously uncharacterized glycoprotein with a predicted weight of 20-kDa (named here

Gp20) (Fig. 4.1). Mass spectrometry also detected the addition of myristoyl and palmitoyl groups to the first and second residues at the N-terminus of Cp23 (442, 443).

Experimental Procedures

Reagents and Parasites

Freshly passaged *C. parvum* oocysts were purchased from Bunch Grass Farm (Deary, ID), and handled under BSL-2 protocols with the approval of the Boston University Institutional Biosafety Committee. All reagents and chemicals were purchased from Sigma-Aldrich (St. Louis, MO), unless noted otherwise. Solvents used for LC-MS were Optima™ grade, procured from Fisher Scientific (Thermo-Fisher Scientific, Waltham, MA).

Protein Extraction and Trypsin Digestion

Procedures for extracting proteins from *C. parvum* oocysts and digesting them with trypsin have recently been described in detail (416), and therefore only a brief summary of the methods is presented here. Briefly, 10^9 *C. parvum* oocysts were concentrated by centrifugation, washed 3X with PBS, and re-suspended with PBS containing EDTA-free cOmplete™ protease inhibitor (Roche, Basel, Switzerland). Oocysts walls were disrupted in a bead beater with 0.5-mm glass beads and centrifuged. The PBS supernatant was removed and saved, while the remaining insoluble materials and beads were extracted with a solution composed of 10 mM HEPES, 25 mM KCl, 1 mM $CaCl_2$, 10 mM $MgCl_2$, 2% CHAPS, 6 M guanidine HCl, 50 mM dithiothreitol, 1X protease inhibitor, pH 7.4). This guanidine-DTT supernatant was combined with the PBS

supernatant, and the insoluble material was discarded. The proteins were then precipitated, and the pellet was washed with methanol and vacuum dried. Alternatively, oocyst proteins were extracted with hot phenol, and phenol and interphase layers were kept, while the aqueous layer was discarded. Proteins were precipitated with methanol containing 100 mM NH_4OAc and dried, as described above. The pelleted proteins were re-suspended in 50 mM NH_4HCO_3, pH 8.0, reduced with 50 mM DTT, alkylated with iodoacetamide, and then digested with proteomics grade trypsin (Sigma-Aldrich, St. Louis, MO). Tryptic peptides were dried and desalted using C18 ZipTip concentrators following the manufacturer's protocol (EMD Millipore, Danvers, MA).

Mass Spectrometry

The LC-MS/MS methodologies and the manual interpretation of MS/MS spectra of *C. parvum* glycopeptides containing *O*-glycans were performed using the methods described for *C. parvum N*-glycosylated peptides, as recently described in detail (416). A brief summary of the methods is provided here. Desalted and dried peptides were dissolved in 2% ACN, 0.1% formic acid (FA) and separated using a NanoAcquity Ultra Performance Liquid Chromatography (UPLC) system (Waters. Milford, MA), using a nanoAcquity Symmetry C18 trap column and a BEH130C18 analytical column. Solvent mixtures for the mobile phase gradient were 99:1:0.1 HPLC grade water/ACN/FA and 99:1:0.1 ACN/HPLC grade water/FA. The UPLC was coupled to a TriVersa NanoMate ion source (Advion, Ithaca, NY), operated at 1.5 kV, introducing ions into either an LTQ-Orbitrap-XL-ETD or a QE Plus mass spectrometer (Thermo-Fisher Scientific, San Jose, CA), both operating in the positive-ion mode. MS spectra were recorded over the range

m/z 350-2000. MS/MS HCD spectra were acquired by isolating the top 5 (LTQ-Orbitrap) or top 20 (QE+) precursor ions with a 2-*m/z* window and fragmenting the selected precursor ions with 15 to 45 V HCD energy. The lower energy MS/MS HCD spectra were scanned from *m/z* 100 to an upper *m/z* value which was dependent upon the parent ion m/z. For the 45-V HCD spectra, ions below *m/z* 210 were excluded to avoid trapping the very abundant HexNAc oxonium ion.

Manual Interpretation of Mass Spectra

Data obtained from LC-MS/MS experiments were first examined using Qual Browser in the Xcalibur 2.2 software suite (Thermo-Fisher Scientific). Extracted ion chromatograms were generated from MS/MS spectra for oxonium ions of interest (HexNAc, *m/z* 204.0866; Hex-HexNAc *m/z* 366.1395; $HexNAc_2$, *m/z* 407.1670). Spectra containing one or more of these ion(s) were then manually interpreted (416). Once a sequence was obtained, it was searched against the 3,803 entries within the *C. parvum* Iowa-II predicted proteome and cross-searched within the entire NCBI nr database, using the online NCBI BLASTP algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (356, 358, 444). The software, Glycoworkbench v2.1 release 146, was used to help calculate glycan compositions (445). It was found that the *O*-glycosylated peptides utilized HexNAc almost exclusively. Due to the labile nature of *O*-linked glycans, b and y ions containing one or more HexNAc residues typically had very low abundances. The charge-reduced molecular ion, with the loss of one or more HexNAc residues was often observed. The information obtained from manual interpretations was then used for database searches, allowing for deeper sequencing of the data, and allowing for higher-throughput

processing of samples. A spreadsheet is supplied for all of the manually annotated spectra contained within this manuscript; it shows the peak list, the assigned ions, and their mass errors (see Appendix 4B).

Database Searches for Glycopeptides

Automated database searches were performed on the RP C18 UPLC-MS/MS data using the PEAKS software suite version 8.0 (Bioinformatics Solutions Inc., Waterloo, ON, Canada), against the 3,803 entries within the *C. parvum* Iowa-II predicted proteome release-5.0, available from the *Cryptosporidium* Genome Resource (cryptodb.org). Trypsin was specified as the enzyme, with a 5.0 ppm parent mass error tolerance, a 0.01 Da fragment mass error tolerance, and carbamidomethyl cysteine set as a fixed modification for the PEAKSdenovo step. For the PEAKSDB stage of the search, the following options were specified: trypsin as the enzyme with ≤ two missed cleavages and ≤ one non-specific cleavage, the error tolerances for precursor of 6 ppm and 0.02 Da for fragment ions, carbamidomethyl cysteine as a fixed modification, and the dynamic modifications HexNAc, $HexNAc_2$, $HexNAc_3$, or $HexNAc_4$ with ≤ six/peptide. The peptide match threshold (-10 logP) was set to 15, with estimation of the false discovery rate (FDR) enabled. A 5.6 FDR was calculated using the aforementioned parameters. A multi-round search was performed, using the *de novo* only results from the first PEAKSDB search. The search parameters were identical to the prior PEAKSdenovo and PEAKSDB searches, with the exception that myristate (*N*-term, S,T) and palmitate (C,S,T,K) were specified as dynamic modifications, and HexNAc was removed. A PEAKSPTM search stage was used for the multiple-round search, considering all

Unimod modifications, with ≤ three PTMs per peptide. The results from the PEAKSDB search, and the multiple-round PEAKSPTM search were exported in Excel and collated.

Re-Annotation of Automated Database Searches

The PEAKS DB search algorithm does not consider neutral loss of HexNAc from b/y ions, and therefore does not annotate the MS/MS product ions appropriately, with the exception of peptides modified with only a single HexNAc residue. Therefore, the proteomics results from PEAKS DB were exported in mzIdentML 1.1 format (446) and the manually verified spectra provided to GlycReSoft, (a software package developed in-house for glycopeptide discovery and annotation), to re-annotate the supplied glycopeptide spectra. The code for GlycReSoft, which is currently in active development with periodic updates and improvements, is open source and freely available from the online repository: https://github.com/BostonUniversityCBMS/glycresoft. All peptides listed in the mzIdentML document and all non-redundant theoretical tryptic digest peptides for each included protein were used as templates, upon which a database of theoretical glycopeptides was constructed. Glycosylation was permitted at up to 20 putative sites. All distinct combinations-with-replacement with the putative glycan compositions were generated. For each template peptide, theoretical glycopeptides were produced by assigning glycosylation events for combinations of between 1 and k glycosylation sites, where k is the total number of potential glycosylation sites. The combinatorial complexity was reduced by limiting the number of possibilities to the first 100 combinations, for glycopeptides having an excess of 100 possible placements. Each dataset was deisotoped, charge state deconvolved, and searched independently against the

database described above. Individual datasets which acquired MS/MS scans in the range *m/z* 100-240 were filtered, and only tandem mass spectra for which the average ratio of oxonium ion signal to maximum signal exceeded 5% were considered; otherwise, every scan was considered. In addition to including normal peptide backbone fragments, the search considered spectra containing peaks that indicated either the presence of a HexNAc residue or its loss. The software also searched for the intact peptide backbone with zero or more partial losses of each potential glycan. Glycopeptide-spectrum matches were evaluated based upon joint binomial intensity-backbone coverage criteria included in a novel algorithm that is based in part on a binomial scoring function described previously (447) The lists of ions assigned for each of these spectra are located in Appendix 4A. Fig. 4.10 shows a representative spectrum annotated by GlycReSoft (one of 345 submitted to the ProteomeXchange Consortium (448).

Other Bioinformatics Methods

The furin-like protease site that separates Gp40/Gp15 was predicted by the online tool “ProP 1.0 Server”, made available by the Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark (www.cbs.dtu.dk/services/ProP/) (449). Signal peptides and transmembrane helices were predicted using the online tool Phobius (http://phobius.sbc.su.se/index.html) (375). The GPI-anchor site of Gp15 was predicted using the BIG-PI prediction server (http://mendel.imp.ac.at/gpi/gpi_server.html) (418). Cartoon representations of proteins, mapped with all the peptides across all MS/MS experiments and protein features were generated using the online software tool, Protter v.1.0, (http://wlab.ethz.ch/protter/start/)

(417). The assigned peptides from the PEAKSDB search results were used to map to the proteins of interest. The protein features were mapped using the results from the bioinformatics searches.

O-Linked Glycan Release and Characterization

Ser-linked or Thr-linked glycans were released from the proteins via reductive alkaline β-elimination. Briefly, purified proteins from a total oocyst lysate were first lyophilized in glass conical vials. To the dried protein extract, an aqueous solution of 0.1 M NaOH + 1 M $NaBD_4$ was added. The loosely capped vials were placed into an oven and kept at 45 °C for 18 h. After the incubation period, the borate was removed by extensive washes with 10% acetic acid in methanol and then neat methanol. The released glycans were subsequently separated from the proteins by solid phase extraction columns. To the dried sample, LC-MS grade water containing 0.1% trifluoroacetic acid (TFA) was added; the tube was vigorously vortexed, and the contents were then passed through a C-18 Sep-Pak cartridge (Waters Corporation, Milford, MA). Three bed volumes of 0.1% TFA/water were subsequently passed through the column, and the eluent fractions were pooled and lyophilized. The released *O*-glycans were permethylated using previously described methods (372, 373). A slurry of powdered NaOH in DMSO was added to the dried, released *O*-glycans. An equal volume of methyl iodide was added, and the reaction mixture was agitated gently while protected from the light. The process was repeated three times to ensure complete permethylation. The product was extracted with chloroform/water, and the aqueous layer was removed and discarded. Washes with water were repeated until the pH of the solution was that of the LC-MS grade water being used

for the washes. The chloroform layer was removed, placed into a new clean vial, dried in a speed vacuum and stored in a dissector at -20 °C until it was analyzed.

Monosaccharide Composition Determination

The permethylated sugars were identified using GC-MS with a Bruker Scion SQ interfaced to a 436-GC (Bruker Daltonics, Billerica, MA). Separation was performed using a (30 m x 0.25 mm x 0.25 μm) Restek™ Rxi™- 5ms capillary column (Restek Corporation, Bellefonte, PA), using helium as the carrier gas. Samples were dissolved in hexane with 1 μl introduced via an auto-injector, using a split/split-less injection program, maintaining a constant column flow rate of 1 ml/min. The injector temperature was set to 220°C. The split-less injection sampling was set for 1 min before starting the split flow at 100 ml/min for 1 min. Then a split flow of 50 ml/min was used for the remainder of the program. The initial oven temperature of 60 °C was maintained for 1 min, then ramped at 4 °C/min to 250 °C, with a final ramp to 300 °C at 20 °C/min, and held there for 10 min. Ions generated by an electron impact (EI) ionization source (70 eV) were introduced into the mass analyzer after a 5-min solvent delay. Centroid mass spectra were acquired in the positive mode, scanning the range *m/z* 50-500, taking 500 ms/scan. An internal standard of permethylated *myo*-inositol was added to all samples to verify retention time repeatability. Four spectra were averaged and background subtracted. The retention times and EI spectra of the released and permethylated glycans were compared to those of genuine deutero-reduced, permethylated monosaccharide standards. Data analysis was performed using the software MS Data Review 8.0 (Bruker). Retention times were compared using extracted ion chromatograms (XIC), for the ion signal at *m/z*

101, an ion common to GalNAc and GlcNAc. The EI spectra recorded for the standards and β-elimination products were compared at the same time points.

Results

The Vast Majority of Peptides with *O*-HexNAc Derive from Gp40, Gp15, and Gp900, which are Vaccine Candidates

Peptides obtained from trypsin digestion of total proteins of *C. parvum* oocysts were separated using a UPLC reversed phase C18 column that was online with a mass spectrometer. Peptides were subjected to Higher-energy C-trap Dissociation (HCD), and *O*-glycosylated peptides were recognized by the observation of an oxonium ion signal at *m/z* 204.0866 in the MS/MS spectra, corresponding to the fragmentation of a glycopeptide containing a HexNAc residue. The utilization of larger oxonium ions (corresponding to Hex-HexNAc, HexNAc-HexNAc, and Hex-HexNAc-HexNAc) all derived from *N*-glycans, was described in (416). Since there was no enrichment for glycoproteins in the protein preparations (*e.g.* lectin chromatography), we identified the most abundant glycopeptides without selection bias. These included glycopeptides with *O*-linked glycans, originating from three *C. parvum* vaccine candidates (Gp15, Gp40, and Gp900), as well as Gp20, the immunogenicity of which is unknown (Table 4.1 and Appendix 4A). We also detected the presence of myristate and palmitate on an N-terminal peptide of the immunodominant antigen Cp23. In addition, we identified at least two peptides without *O*-HexNAc from each of 811 other *C. parvum* proteins. Information about these peptides and the glycopeptides described below has been deposited in the ProteomeXchange Consortium.

Dense Arrays of *O*-GalNAc are Present on the Ser-rich Domain of Gp40

Gp40/Gp15 precursor (cgd6_1080) has an N-terminal signal peptide, a furin cleavage site that separates Gp40 (AA-22 to 220) from Gp15 (AA-221 to 324), and a C-terminal site for the addition of a GPI-anchor (Fig. 4.1) (34, 40, 42, 369). A tryptic glycopeptide (AA-43 to 60) of Gp40, which contains 17 consecutive Ser residues followed by Thr-Ser-Thr, was modified with 15 to 20 HexNAc residues (Table 4.1 and Appendix 4A). For example, the monoisotopic mass of the precursor ion *m/z* 1757.2272 $[M + 4H]^{4+}$ corresponds to the value calculated for the peptide DVPVEGSSSSSSSSSSSSSSSSTSTVAPANK with the addition of 20 HexNAc residues (Fig. 4.2 and Appendix 4B). The very abundant HexNAc oxonium ion (*m/z* 204.0866) and a very low abundance peak that fits the value for $HexNAc_2$ (*m/z* 407.1670) are present in the 30-V HCD MS/MS spectrum. The observed dimer could be an artifact generated from the high population of HexNAc monomers. To a very large extent, glycan loss occurs prior to fragmentation of the peptide, with the result that the observed b and y ions contain zero to four HexNAc residues. The only product ion that can be used to assign the HexNAc modification to a specific amino acid is the y_7* ion, indicating the presence of HexNAc on the Thr closest to the C-terminus. In a second experiment, to avoid overpopulating the Orbitrap analyzer with the less informative HexNAc oxonium ion, the start of the selection window was raised from *m/z* 100 to *m/z* 210, and to ensure the generation of more peptide backbone fragments, the HCD energy was increased to 45 V (Fig. 4.7 and Appendix 4B). The 45-V HCD MS/MS spectrum exhibited extensive fragmentation of the aglycon peptide, which resulted in product ions that composed a

nearly complete y -series (y_2 - y_{25}, y_{30}) and included several b ions (b_2 - b_4, b_6 - b_7). Because we saw little evidence for the presence of HexNAc-HexNAc, we assume that each of the 20 potential *O*-glycan sites is occupied with a single HexNAc residue. We were unable to localize site occupancy in glycopeptides with 15-19 HexNAc residues, due to the labile nature of the *O*-glycans. Quite likely, the peptides modified with 15 to 19 HexNAc residues are a mixture of components having different occupancies.

Release of *O*-glycans from *C. parvum* sporulated oocyst proteins by reductive β–elimination, followed by monosaccharide analysis versus sugar standards using GC/MS, showed that the HexNAc residues in the Gp40 glycopeptide and in glycopeptides of the other vaccine candidates are likely GalNAc (Fig. 4.8). In support of this assignment are the previous reports that *C. parvum* has four *O*-GalNAcTs, and patient sera recognize synthetic glycopeptides derived from Gp40 and Gp15 with *O*-GalNAc (44, 125). In summary, the Gp40 spectra presented here show that the *C. parvum O*-GalNAcTs are capable of saturating or nearly saturating consecutive arrays of Ser residues.

Isolated *O*-GalNAc Residues Decorate a Glycopeptide of Gp15

A non-tryptic glycopeptide of Gp15 (AA-221 to 240), which results from cleavage of the Gp40/Gp15 precursor by the furin-like protease, contained one to four HexNAc modifications (Fig. 4.1, Table 4.1, and Appendix 4A) (34, 40, 42, 369). For example, the precursor ion *m/z* 1326.6164 $[M + 2H]^{2+}$ of the most abundant Gp15 glycopeptide has a monoisotopic mass equal to that of the peptide ETSEAAATVDLFAFTLDGGK with the addition of three HexNAc residues (Fig. 4.3 and Appendix 4B). Fragmentation with 30-V HCD yielded a prominent HexNAc

oxonium ion (*m/z* 204.0868) and full series of b and y ions, some of which retained a single HexNAc modification (marked with an asterisk). The product ion series y_6* to y_{12}*, indicates Thr-235 is modified and the series y_{13}* to y_{15}* suggests that either Thr-228 or Thr-235 is modified. The b_3* ion indicates that either Thr-222 or Ser-223 is modified. Thus there is evidence for distribution of the three HexNAc residues over the four available sites in this peptide. In the glycopeptide with four HexNAc modifications, all possible *O*-glycan sites must be occupied. Analyses of the fragmentation patterns of numerous other peptides (Appendix 4A), both tryptic and non-tryptic, suggest that Thr-222 is preferentially modified over Ser-223, while Thr-228 and Thr-235 are nearly always modified.

Dense Arrays of *O*-GalNAc are Present on Thr-rich Glycopeptides of Gp20

Gp20 (cgd7_1280), is a small, acidic, secreted protein with four domains with consecutive Thr residues, two of which are described here (Fig. 4.1). The first Gp20 glycopeptide (87)EGEETDENTDETTTTTTTASPKPK(110) has 10 potential *O*-glycan sites and was found to be decorated with six to eight HexNAc residues (Table 4.1 and Appendix 4A). For example, the peak corresponding to the precursor ion of the most abundant Gp20 glycopeptide has a monoisotopic $[M + 4H]^{4+}$ *m/z* 1001.9305, equal to the value calculated for the peptide modified by seven HexNAc residues (Fig. 4.4 and Appendix 4B). The 30-V HCD MS/MS spectrum includes a HexNAc oxonium ion (*m/z* 204.0868) and numerous b and y-ions retaining zero to two HexNAc residues (marked with asterisks). Because the vast majority of HexNAc residues were lost prior to peptide fragmentation, it was not possible to define the seven occupied sites or to determine

whether the occupancy was heterogeneous. A second Gp20 glycopeptide (135)SSTTTTTTTAPVSSEDNKPEDSEDEK(160) with 12 potential *O*-glycan sites has a monoisotopic mass equal to that of the peptide with the addition of eight HexNAc residues (Table 4.1 and Appendix 4A). Again glycan loss prior to peptide backbone fragmentation made it impossible to localize the occupied *O*-glycans sites. Two other Thr-rich domains of Gp20 are present in a 55-amino acid long tryptic peptide, which was not identified. Regardless, the two Gp20 spectra show that the *C. parvum O*-GalNAcTs are capable of nearly saturating arrays of Thr residues.

A Glycopeptide of Gp900 with Consecutive Thr is Lightly Modified by *O*-GalNAc, while Numerous Gp900 Glycopeptides Contain a Single *O*-HexNAc Residue

Gp900 (cgd7_4020), which has an N-terminal signal peptide and a transmembrane domain near its *C*-terminus, is by far the largest of the *C. parvum* vaccine candidates (1912 amino acids minus the signal peptide) (Fig. 4.1) (31, 33) . One reason for the large size of Gp900 is the presence of a vast array of consecutive Thr residues, which extends from AA-304 to 640. A second Thr-rich region extends from AA-797 to 908. Because of the paucity of tryptic sites in the Thr-rich arrays of Gp900, and the likelihood that the Thr stretches are also heavily *O*-glycosylated, these regions were not observed by mass spectrometry, with one exception (Table 4.1, Appendix 4A, and Fig. 4.8). The precursor ion *m/z* 732.0284 $[M + 3H]^{3+}$ has a monoisotopic mass equal to that calculated for the peptide (609)KPTTTTTTTTTTTTK(623) with the addition of only three HexNAc residues, despite the presence of 12 available sites (Fig. 4.8 and Appendix 4B). The 30-V HCD MS/MS spectrum includes a HexNAc oxonium ion (*m/z* 204.0868)

and numerous b and y ions containing zero to two HexNAc residues (marked with asterisks). Here again, because of the lability of the glycans, it was not possible to precisely define the occupied sites or to determine whether the occupancy was heterogeneous. Many of the most abundant glycopeptides of Gp900 have a single HexNAc modification at an isolated Ser or Thr residue (Table 4.1 and Appendix 4A). For example, the precursor ion *m/z* 895.4646 $[M + 4H]^{4+}$ has a monoisotopic mass corresponding to the value calculated for the peptide (1712)NIVTEAAYGLPVDPK(1726) plus a single HexNAc residue (Fig. 4.5 and Appendix 4B). The b_4*, b_6*, and b_7* ions show that Thr-1715 is modified. The mass spectra of 12 unique peptides from Gp900, each with a single HexNAc modification, together with the spectra from Gp15, suggest that the *C. parvum O*-GalNAcTs are capable of modifying isolated Ser and Thr residues, in addition to stretches of consecutive Ser residues in Gp40 and Thr in Gp20 and Gp900.

At the N-terminus of Cp23 Myristoyl Modifies Gly1, while Palmitoyl Modifies Cys2.

The immunodominant antigen Cp23 (cgd4_3620) contains no signal peptide but has an N-terminal sequence (MGCSSSKPETK) similar to those modified by fatty acyl chains in the host and other apicomplexans (Fig. 4.1) (110, 450-456). Consistent with this resemblance, numerous hydrophobic peptides were identified by mass spectrometry containing the N-terminus of Cp23 minus Met-1, with no modification, substituted by either myristate or palmitate, or both (Table 4.2) (442). For example, the precursor ion $[M + 2H]^{2+}$ *m/z* 736.4573 has a monoisotopic mass equal to that calculated for the peptide GCSSSKPETK with the addition myristate and palmitate (Fig. 4.6 and Appendix 4B).

Fragmentation using 30-V HCD showed the presence of myristate (*m/z* 211.2056) and palmitate (*m/z* 239.2370), as well as the charge reduced $[M + H]^{1+}$ molecular ion with loss of palmitate (*m/z* 1233.6782) or myristate (*m/z* 1261.7001). Complete y-ion series and partial b-ion series allow us to assign myristate to the N-terminal Gly and palmitate to the Cys. We believe the example given is what is present on the native protein. The peptides where palmitate is absent and Cys is carbamidomethylated or palmitate modifies Ser residues, for which numerous other spectra were recorded, are likely artifacts occurring during sample processing (442, 443).

Discussion

Mass spectrometry here directly demonstrated that addition of *O*-GalNAc is a widespread modification of *C. parvum* vaccine candidates (Gp15, Gp40, and Gp900) (352, 353). Previous Evidence for the addition of *O*-GalNAc to these proteins has previously been obtained through the use of synthetic glycopeptides, lectins, patient sera, or a monoclonal anti-carbohydrate antibody to *C. parvum* (15, 44). The remarkable result reported here is that the *O*-GalNAc modifications saturate Ser-rich sequences of Gp40 and they nearly saturate Thr-rich sequences of Gp20, a protein which has not previously been characterized. A second remarkable result is that nearly all of peptides with *O*-glycans derive from just four proteins (Gp40, Gp15, Gp900, and Gp20), even though >800 proteins were identified by mass spectrometry. Limitations of our observations include 1) failure to observe most of the very extended Thr-rich domains of Gp900 and two of the Thr-rich domains of Gp20, 2) inability to assign *O*-glycans sites on many of the peptides due to very facile elimination of the *O*-linked glycan residues during HCD

fragmentation, and 3) limited sampling of glycoproteins with *O*-*G*alNAc. In particular, the GalNAc-binding *Maclura pomifera* agglutinin enriched six mucin-like glycoproteins in addition to Gp40, Gp15, and Gp900 from lysed oocysts (35).

These results show that the four *O*-GalNAcTs of *C. parvum*, each of which has a lectin domain in addition to its glycosyltransferase domain, efficiently continue to glycosylate regions of glycoproteins that are already glycosylated (125, 457, 458). Indeed the four *O*-GalNAcTs of *C. parvum* are able to make the same kind of modifications to arrays of Ser and Thr and to isolated Ser and Thr as the 20 *O*-GalNAcTs of the host. The activity of each *O*-GalNAcT can only be determined by knockouts of the genes encoding these enzymes, a technology that is now available in *C. parvum* grown in mice (459). *O*-glycans of *C. parvum* differ from those of the host in that *O*-GalNAc is not extended by other sugars (458). *C. parvum* then is the equivalent of the "SimpleCell" lines engineered to express truncated *O*-GalNAc (knockout of cosmic gene), which have been used to map occupied *O*-glycan sites (460-462). Similarly, the very short *C. parvum O*-glycans are recognized by anti-Tn antibodies, which bind to *O*-GalNAc (44, 463).

Properties that distinguish *C. parvum* Gp15 and Gp40 include glycosylation (discrete *O*-GalNAc residues versus densely clustered *O*-GalNAc residues) (shown here), localization on sporozoites (apically associated versus diffusely covering surface), localization on oocyst walls (outer surface versus inner surface), and structure (GPI-anchored versus secreted) (15, 31, 33-35, 39, 40, 42, 369, 427). We infer that the densely clustered *O*-GalNAc residues make the Ser-rich regions of Gp40 and Thr-rich regions of Gp20 and Gp900 rigid and extended rather than unstructured (419, 464, 465). These

extended regions of *O*-glycosylation may contribute to the tethering functions of Gp40 and Gp900, which attach sporozoites to the inner layer of the oocyst wall (35). *O*-glycosylation on these glycoproteins that coat the sporozoite surface may also affect host cell invasion and/or the innate and acquired immune responses to infecting parasites (15, 363, 365-367). In contrast, addition of myristate to N-terminal Gly and palmitate to Cys likely directs cp23 from the cytosol to membranes of *C. parvum* and thus is important for its function, as has been extensively studied in *Toxoplasma, Plasmodium*, and the host (441, 450-456). Chemical biology experiments or mutation of sites for addition of myristate and palmitate on cp23 would be useful to test the roles of fatty acyl modifications in *C. parvum*. Serological screens for *C. parvum* use recombinant proteins which are made in bacterial systems that fail to add *O*-GalNAc (Gp40 and Gp15) or fatty acyl chains (Cp23) (347, 363-367, 430, 431). Because the host antibody response includes antibodies to glycopeptides with *O*-glycans (44), these serological screens are likely lacking sensitivity and might be improved by expressing *C. parvum* proteins in SimpleCells that add only *O*-GalNAc to glycoproteins (460-462). Similarly, vaccination with recombinant proteins produced in bacteria provokes an immune response to the unmodified peptides, whereas acquired immunity to *C. parvum* infections includes responses to the *O*-glycans on Gp40 and Gp15 (44) and possibly lipid-modifications of Cp23. Again, the cosmc knockout might be used to produce recombinant Gp40 or Gp15 coated with *O*-GalNAc for vaccination. Production of Cp23 in mammalian cells that add fatty acyl chains might increase the sensitivity of serological screens for this antigen and generate a better vaccine.

Acknowledgements:

Support for this study came from NIH grants R01 AI110638, R01 GM031318, and UL1TR001430 (J.S.) and P41 GM104603, S10 RR025082, and S10 OD010724 (C.E.C.) and from NIH-NHLBI contract HHSN268201000031C (C.E.C.). Thanks to Joseph Zaia for comments and supervision of J.A.K. We also thank Thermo Fisher Scientific for generously loaning the QE-Plus mass spectrometer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005989 and 10.6019/PXD005989 (448). The data may be accessed at the uniform resource locator: http://www.ebi.ac.uk/pride/archive/projects/PXD005989.

Conflict of Interest:

The authors declare that they have no conflicts of interest with the contents of this article.

Author Contributions:

J.R.H, C.E.C., and J.S. conceived and designed the study. J.R.H. performed experiments. J.R.H., J.A.K, and C.E.C. analyzed data. J.R.H, C.E.C., and J.S. wrote the manuscript. C.E.C. and J.S. acquired funding. All authors reviewed the results and approved the final version of the manuscript.

Tables

Protein Name	Representative Glycopeptide†	#HexNAcs
Gp40	(31)DVPVEGSSSSSSSSSSSSSSSSSTSTVAPANK(62)	15,17,18,19,20*
Gp15	(221)ETSEAAATVDLFAFTLDGGK(240)	1,2 ,3*,4
Gp20	(87)EGEETDENTDETTTTTTTASPKPK(110)	6,7*,8
	(135)SSTTTTTTTAPVSSEDNKPEDSEDEK(160)	8
Gp900	(609)KPTTTTTTTTTTTTK(623)	3
	(958)IADTSNLFPVQTHK(971)	1
	(1197)TPTQTDSVTGKPIDPTTGLPFNPPTGH(1223)	1
	(1197)TPTQTDSVTGK(1207)	1
	(1243)YAVSNGIKTDNVYGLPVDEITGLPK(1267)	1
	(1248)NGIKTDNVYGLPVDEITGLPK(1267)	1
	(1373)GKDGLIVPPTNSINK(1387)	1
	(1410)VIPGSLPGSLNYPSFNTPQQTDEITGK(1436)	1
	(1646)TIPGSAASVIHTALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIK(1698)	1
	(1658)ALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIK(1698)	1
	(1710)EKNIVTEAAYGLPVDPK(1726)	1
	(1795)LIDPESGIAIDNSVSGVFATVPGTAAPK(1822)	1
	(1813)ATVPGTAAPK(1822)	1

Table 4.1. Overview of Tryptic Glycopeptides Identified by Mass Spectrometry, Examples of which are Shown in Figs. 4.2 to 4.5, 4.7, 4.9, and 4.10.

The table provides an overview of the tryptic glycopeptides identified using mass spectrometry. Example spectra are shown in Figs. 4.2 to 4.5, 4.7, 4.9, and 4.10. The number of HexNAc residues is listed for each glycopeptide, and the most abundant glycopeptide is marked with an asterisk. Missed and non-tryptic cleavages are omitted, unless these occurrences led to a change in the number of Ser or Thr residues. †The complete list of glycopeptides used to generate this table can be viewed in Appendix 4A.

Protein Name	Mass (MH)	Peptide	#Spectra
Cp23	1290.6972	(2)G(myr)C(cm)SSSKPETK(11)	16*
	1528.9268	(2)G(myr)C(cm)SSS(palm)KPETK(11)	2
	1528.9268	(2)G(myr)C(cm)SS(palm)SKPETK(11)	2
	1528.9268	(2)G(myr)C(cm)SSSK(palm)PETK(11)	2
	1471.9054	(2)G(myr)C(palm)SSSKPETK(11)	2*
	1471.9054	(2)GC(cm)S(palm)SSKPETK(11)	1
	1556.9581	(2)GC(cm)S(palm)S(palm)SKPETK(11)	2

Table 4.2. Lipid Modifications of the N-terminus of Cp23.

Lipid Modifications of the N-terminus of Cp23. Asterisks mark peptides that are likely real. The other peptides could be artifacts from sample preparation (442, 443). (cm = carbamidomethyl, 57.0214 Da, myr = myristate, 210.1983 Da, palm = palmitate, 238.2296 Da)

Figures

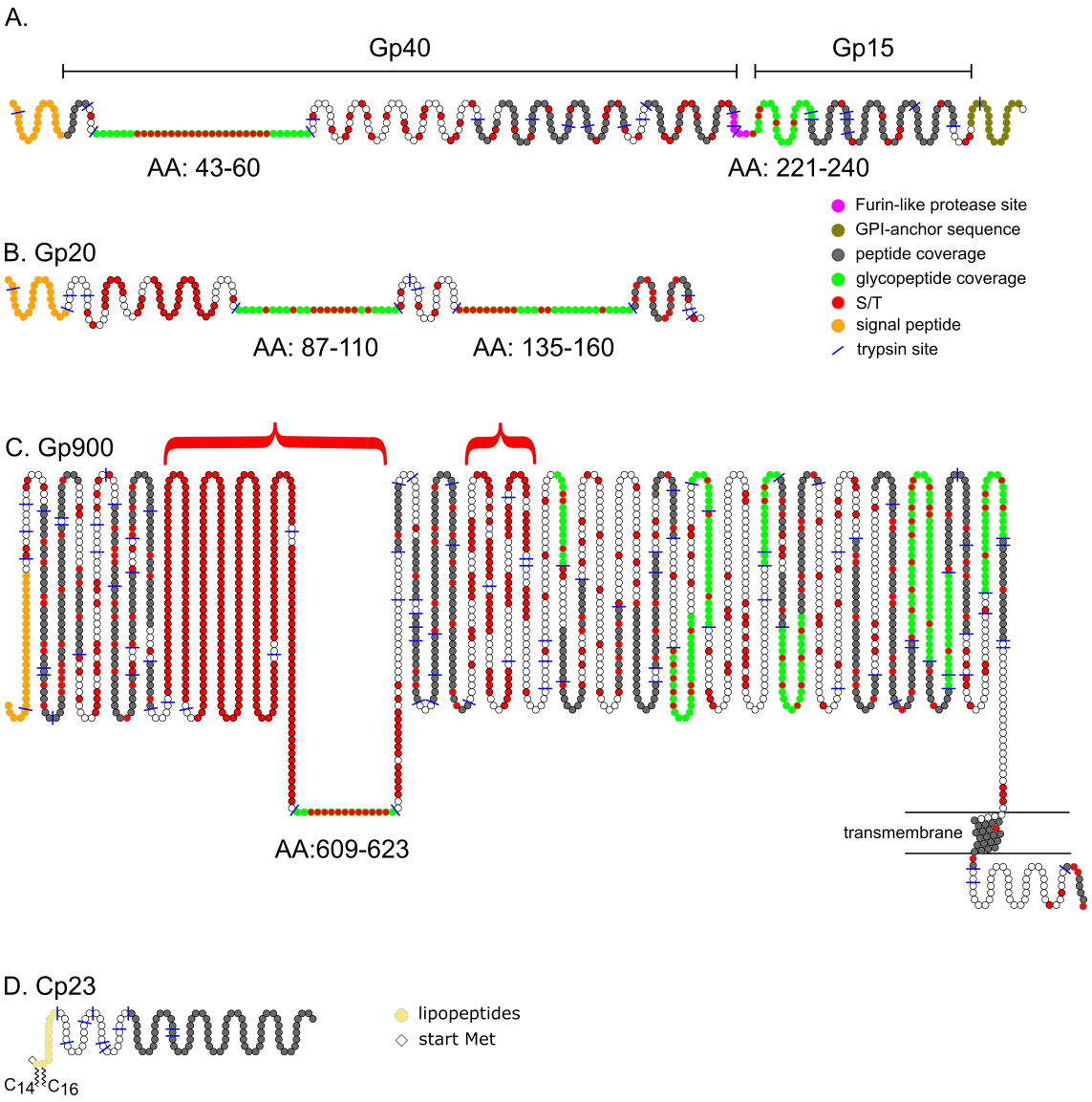


Fig. 4.1 Schematics of *Cryptosporidium* Glycoproteins Characterized by Mass Spectrometry.

(A.) Gp40/Gp15 precursor is cleaved at a furin-like protease site (pink) into Gp40 and Gp15. Mass spectrometry showed Gp40 has a Thr-rich domain (AA-43 to 60) with numerous *O*-linked HexNAc modifications (marked in green, with Ser and Thr residues marked in red); Gp15 contains a single domain (AA-221 to 240) that is glycosylated. Other peptides identified with mass spectrometry are marked in grey. Predicted N-terminal signal peptide is marked in orange, while GPI-anchor signal is marked in olive. (*B.*) A 20-kDa glycoprotein (Gp20) contains two Thr-rich domains (AA-87 to110 and AA-135 to 160), which contain numerous HexNAc modifications. (*C.)* Gp900 contains two very large Thr-rich domains (red brackets), one of which contains a peptide with three HexNAc residues (AA-609 to 623). The transmembrane helix near the C-terminus is encompassed by two horizontal lines, representing a membrane. (*D.*) The N-terminus of Cp23 is modified with *N*-myristate (C_{14}) and *S*-palmitate (C_{16}). The start Met is absent (diamond).

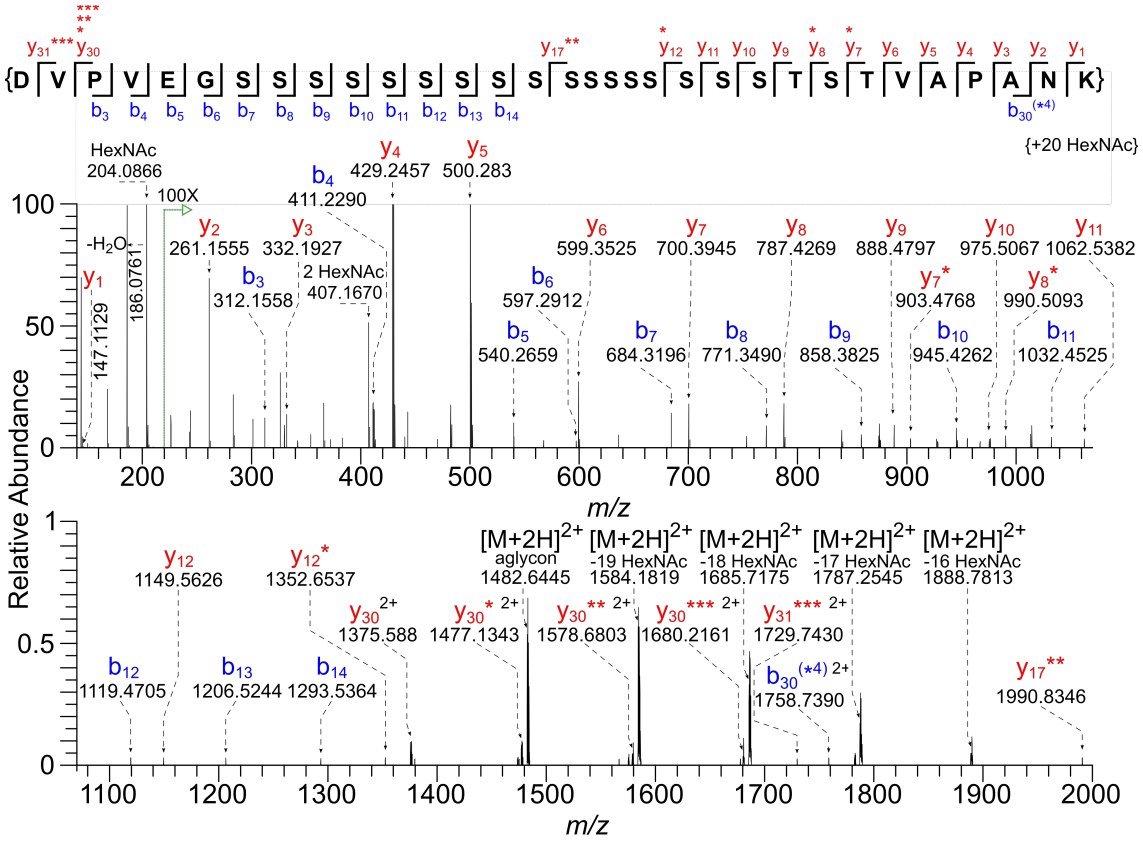


Fig. 4.2 HCD MS/MS Spectrum (@ 30 V) of a Tryptic Glycopeptide of Gp40 Shows Saturation of an Array of Ser with *O*-HexNAc Residues.

The monoisotopic mass of the precursor ion $[M + 4H]^{4+}$ *m/z* 1757.2272 corresponds to the value calculated for the peptide (43)DVPVEGSSSSSSSSSSSSSSSSTSTVAPANK(60) with the addition of 20 HexNAc residues (Δ 0.2 ppm). A very abundant HexNAc oxonium ion (*m/z* 204.0866) and a very low abundance peak (0.5%) corresponding to a HexNAc dimer (*m/z* 407.1670) are present. Asterisks mark the number of HexNAc residues present on b and y ions. Please see Fig. 4.7 for a 45-V HCD MS/MS spectrum of the same Gp40 glycopeptide. The lists of all of the b and y ions assigned for MS/MS spectra shown in this figure and others can be found in Appendix 4B.

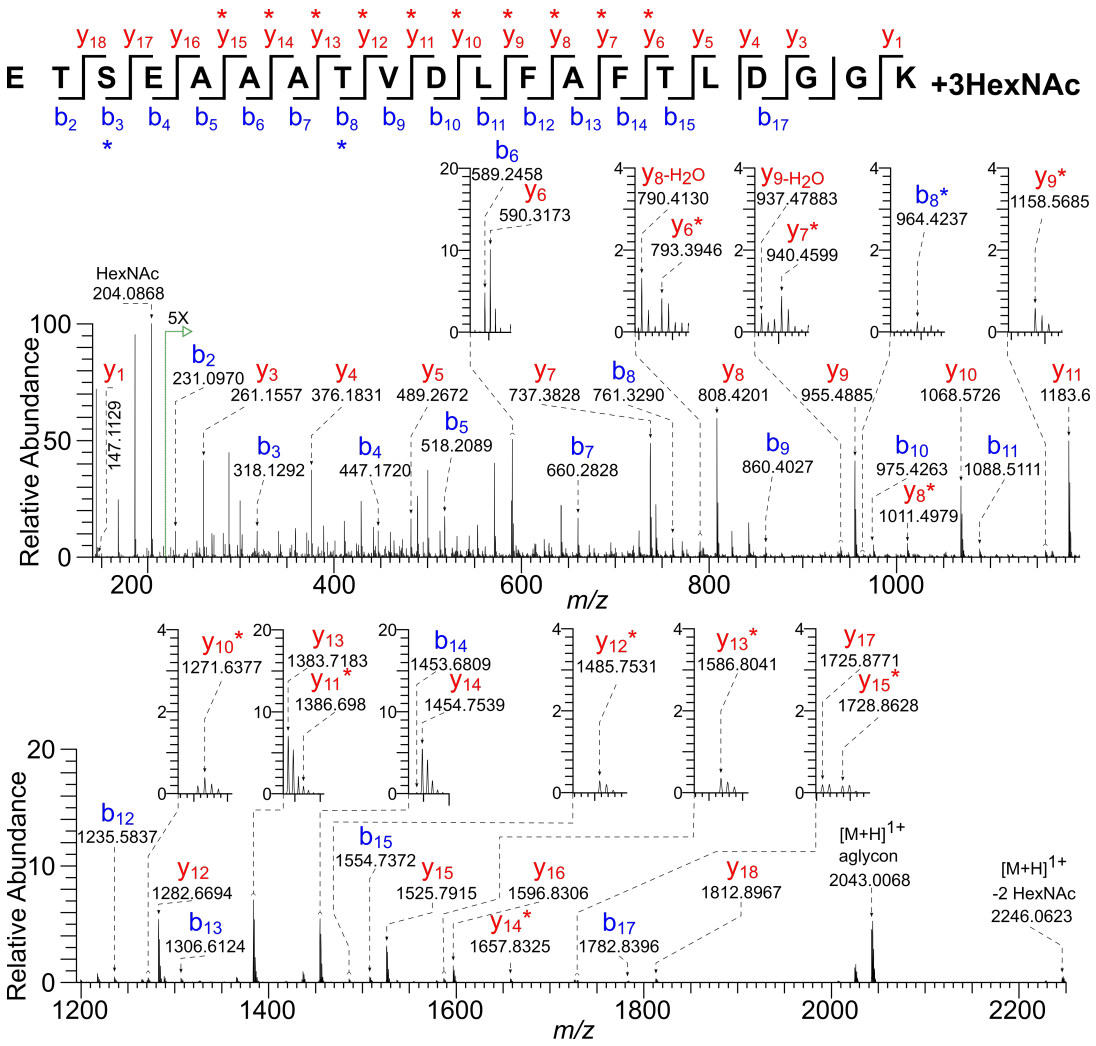


Fig. 4.3. HCD MS/MS Spectrum (@ 30 V) of a Tryptic Glycopeptide of Gp15 Shows that Three of Four *O*-Glycan Sites Contain the HexNAc Modification.

The precursor ion [M + 2H]2+ m/z 1326.6164 of the most abundant Gp15 glycopeptide has a monoisotopic mass corresponding to that calculated for the peptide (221)ETSEAAATVDLFAFTLDGGK(240) with the addition of three HexNAc residues (Δ 0.3 ppm). There is a prominent HexNAc oxonium ion (m/z 204.0868) and full series of b and y ions, some of which contain a single HexNAc residue (*).

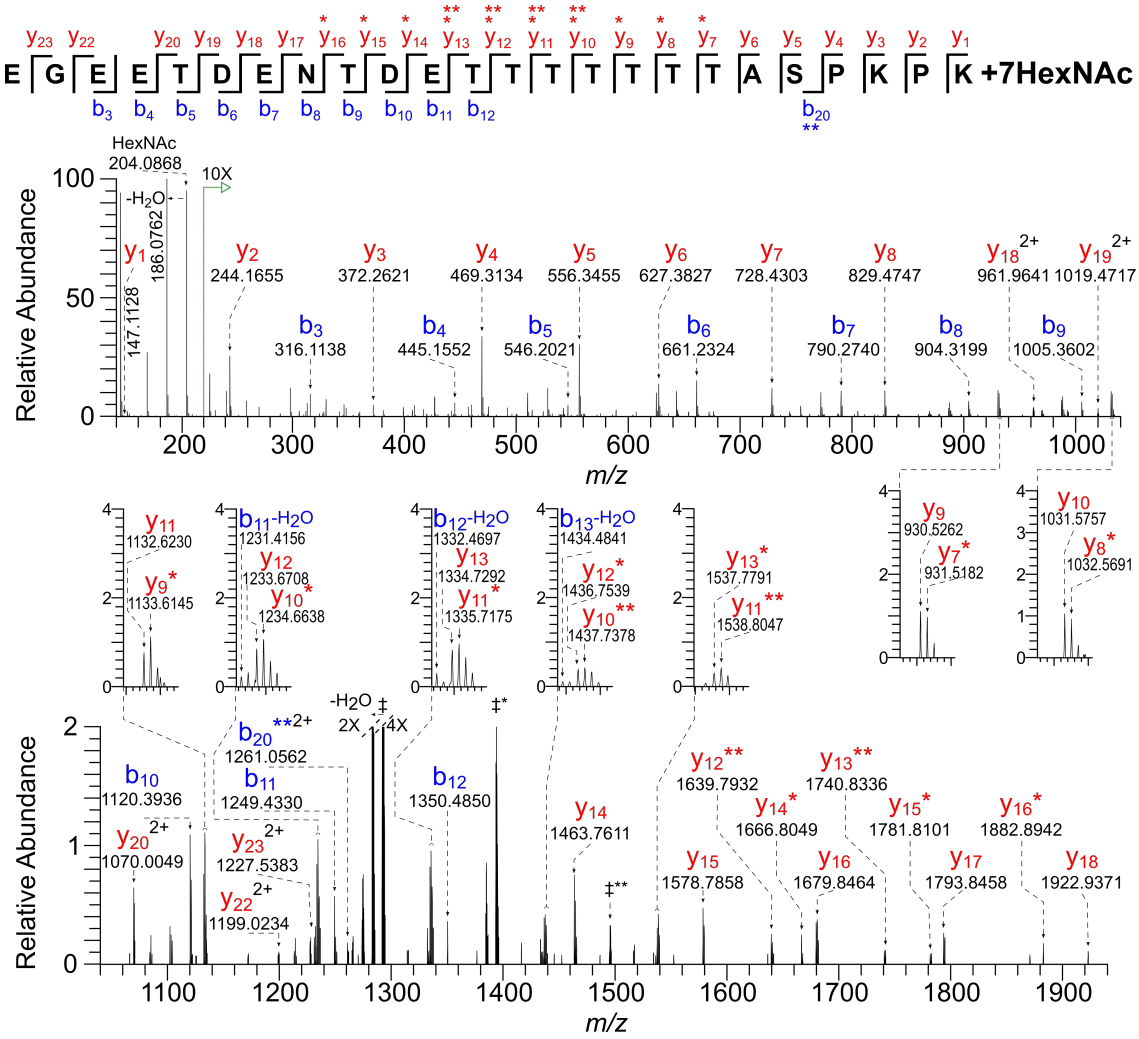


Fig. 4.4 HCD MS/MS Spectrum (@ 30V) of a Tryptic Glycopeptide of Gp20 Shows Near Saturation of Consecutive Thr Residues with the *O*-HexNAc Modification.

The precursor ion $[M + 4H]^{4+}$ *m/z* 1001.9305 has a monoisotopic mass corresponding to that calculated for the peptide (87)EGEETDENTDETTTTTTTASPKPK(110) plus seven HexNAc residues (Δ 0.5 ppm). There is a prominent HexNAc oxonium ion (*m/z* 204.0868) and full series of b and y ions, some of which contain one (*) or two (**) HexNAc residues. All ions are singly charged, except where indicated. In addition, charge-reduced ions, all 2+, corresponding to species that have undergone consecutive losses of HexNAc residues, are observed and are labeled as follows: ‡** = $[M + 2H]^{2+}$ - $HexNAc_5$ (*m/z* 1495.1470), ‡* = $[M + 2H]^{2+}$ - $HexNAc_6$ (*m/z* 1393.6224), ‡ = $[M + 2H]^{2+}$ aglycon peptide (*m/z* 1292.0851) .

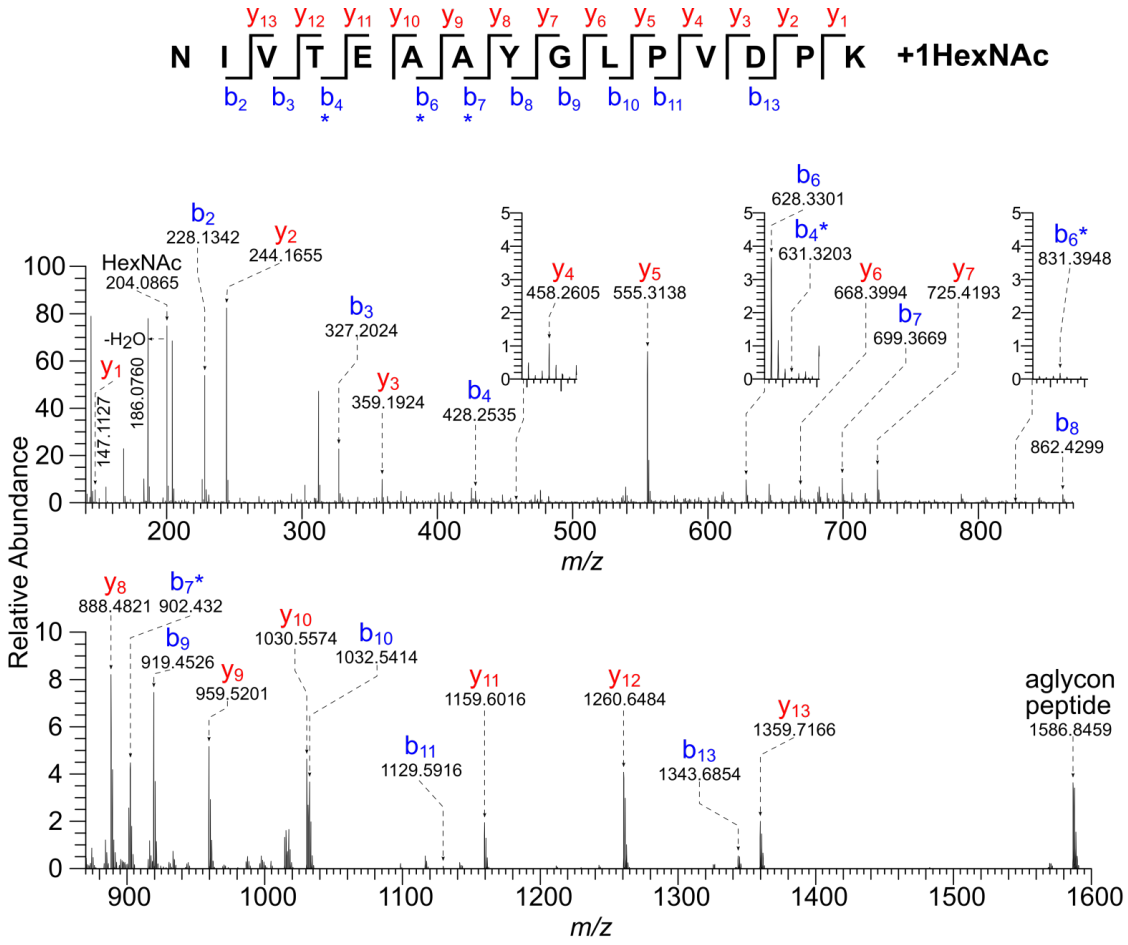


Fig. 4.5 HCD MS/MS Spectrum (@ 30V) of a Tryptic Glycopeptide of Gp900 Shows a Single *O*-HexNAc Modification on a Thr Residue.

The precursor ion $[M + 4H]^{4+}$ *m/z* 895.4646 has a monoisotopic mass corresponding to that calculated for the peptide (1712)NIVTEAAYGLPVDPK(1726) plus a single HexNAc residue (Δ 0.1 ppm). There is a prominent HexNAc oxonium ion (*m/z* 204.0865) and full series of b and y ions, some of which contain a HexNAc residue (*). The b_4*, b_6*, and b_7* ions show that Thr-1715 is modified.

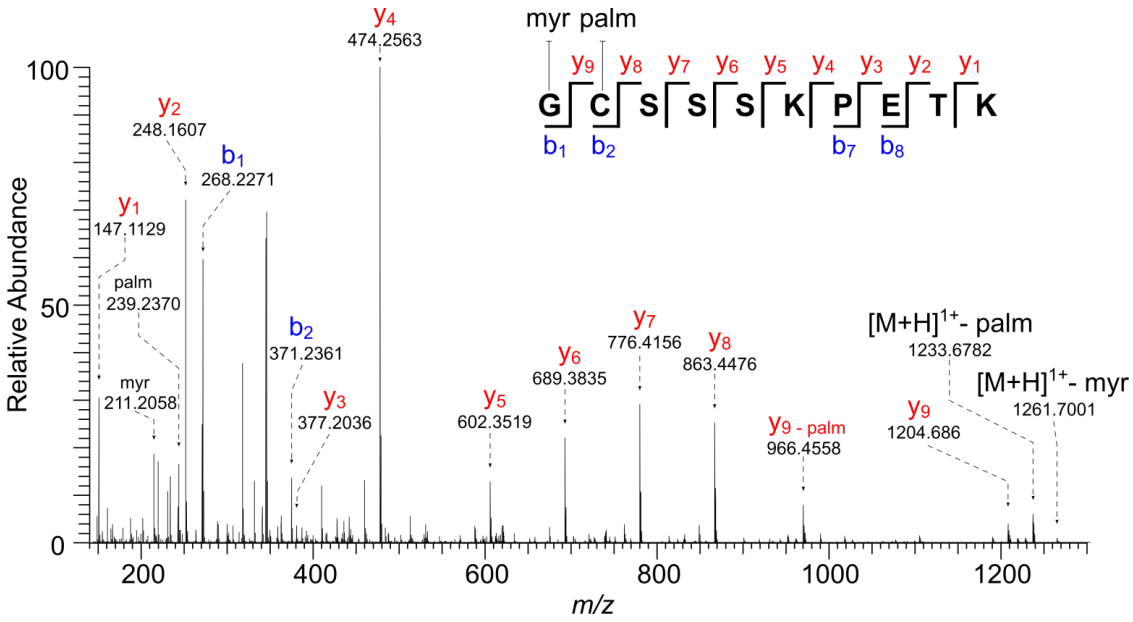


Fig. 4.6. HCD MS/MS Spectrum (@30 V) of an N-terminal Peptide of Cp23 (Minus Met-1) Shows Gly-1 is Modified with Myristate and Cys-2 is Modified with Palmitate.

The precursor ion $[M + 2H]^{2+}$ *m/z* 736.4573 has a monoisotopic mass corresponding to that calculated for the peptide (2)GCSSSKPETK(11) plus myristate and palmitate (Δ 1.1 ppm). Fragment ions could be assigned to myristate (*m/z* 211.2056) and palmitate (*m/z* 239.2370), as well as the charge-reduced $[M + H]^{1+}$ molecular ions that have undergone loss of palmitate (*m/z* 1233.6782) or myristate (*m/z* 1261.7001). All the b/y ions contain the lipid modification, unless otherwise indicated.

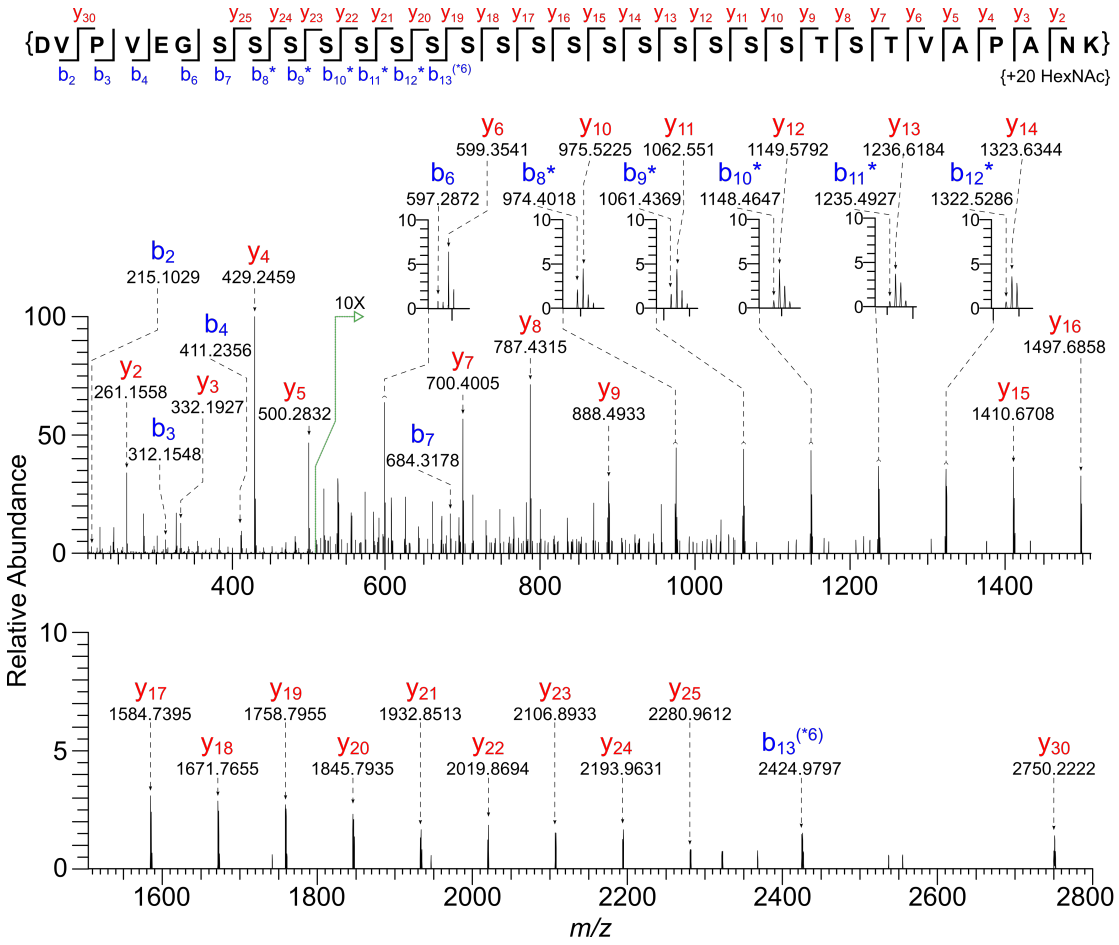


Fig. 4.7. HCD MS/MS Spectrum (@ 45V) of a Tryptic Glycopeptide of Gp40 Gives Complete Sequence of the Peptide.

The precursor ion $[M + 4H]^{4+}$ *m/z* 1757.2272 corresponds to the monoisotopic mass equal to that of the peptide (43)DVPVEGSSSSSSSSSSSSSSSSSTSTVAPANK(60) with the addition of 20 HexNAc residues (Δ 0.6 ppm). The selection window was set to start at *m/z* 210 in order to exclude the very abundant oxonium ion (*m/z* 204.0866). There is extensive fragmentation of the aglycon peptide ((y_2 - y_{25}, y_{30}) and (b_2 - b_4, b_6 - b_7) ions). Fig. 4.2 shows the 30-V HCD MS/MS spectrum of the same peptide.

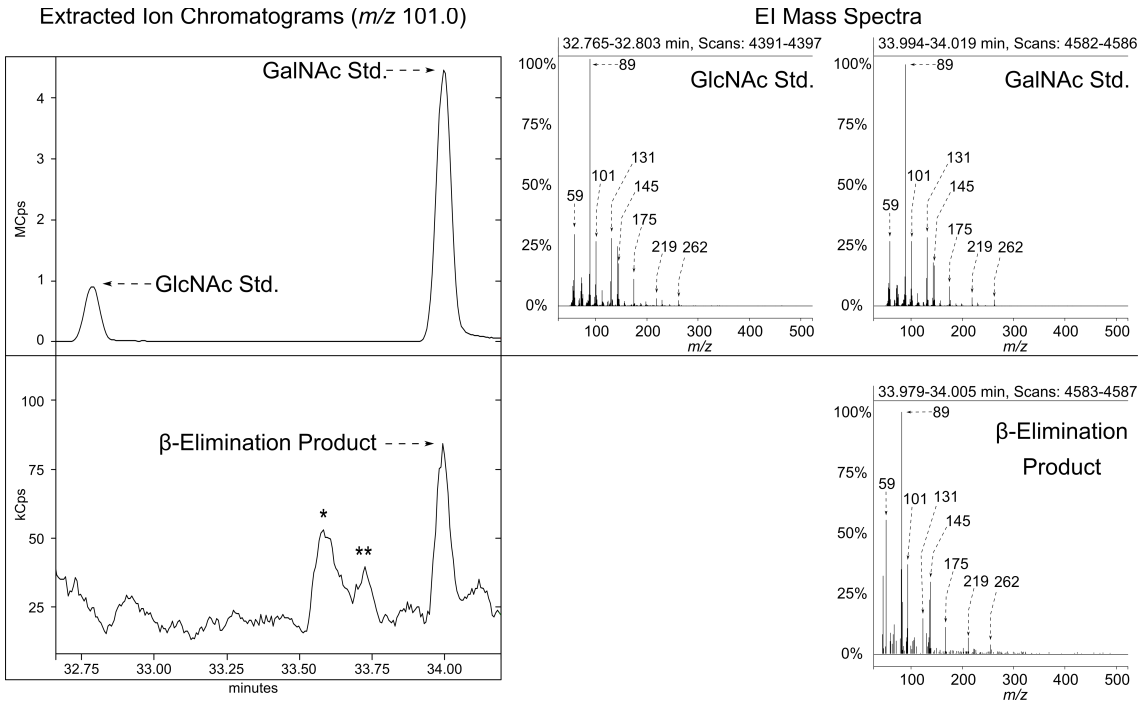


Fig. 4.8. GC-MS Analysis Shows the *O*-Glycan Released by β–Elimination from *C. parvum* Glycoproteins is GalNAc.

GC/MS data obtained for the deuteroreduced and permethylated glycan released by reductive β-elimination of *C. parvum* oocyst glycoproteins are compared with results observed for the GlcNAc and GalNAc standards, which were similarly treated. Left, extracted ion chromatograms of *m/z* 101. Right, electron impact mass spectra of the standards and the product from reductive β-elimination of *C. parvum* oocyst glycoproteins. The sugar released from *C. parvum* is assigned as GalNAc, because the retention time (34 min.) and EI mass spectrum both match that of the standard GalNAc. The EI mass spectra of the components eluting at positions marked with the asterisk (*) and double asterisk (**) symbols do not correspond to sugar derivatives.

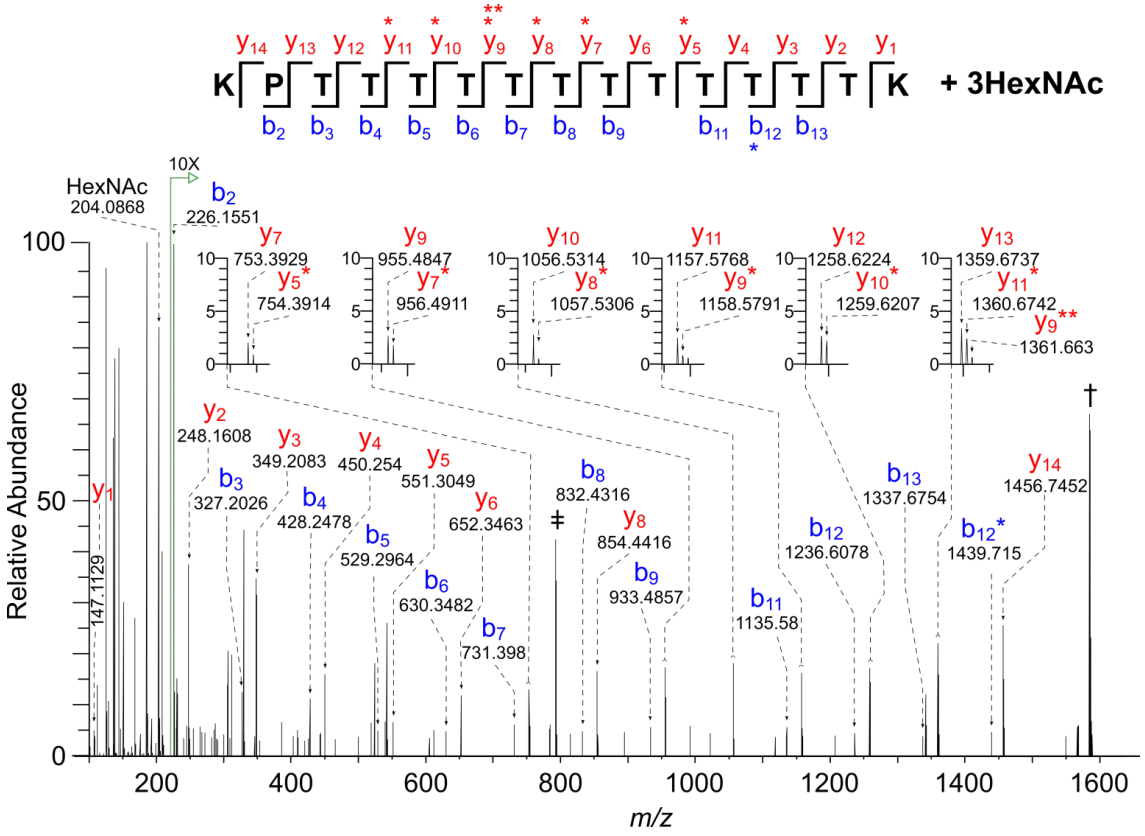


Fig. 4.9. HCD MS/MS Spectrum (@ 30V) of a Tryptic Glycopeptide of Gp900 Shows Partial Glycosylation of a Thr-rich Repeat.

The precursor ion $[M + 3H]^{3+}$ *m/z* 732.0284 has a monoisotopic mass corresponding to the peptide (609)KPTTTTTTTTTTTTK(623) with the addition of three HexNAc residues (Δ -0.1 ppm)(Fig. 4.9). There is a prominent HexNAc oxonium ion (*m/z* 204.0868) and a full series of b and y ions, some of which contain one (*) or two (**) HexNAc residues. The facile loss of HexNAc residues made it impossible to localize occupied sites. Charged-reduced aglycon peptide ions are observed in the spectrum, ǂ = $[M + 2H]^{2+}$ *m/z* 792.9193 and † = $[M + H]^{1+}$ *m/z* 1584.8348.

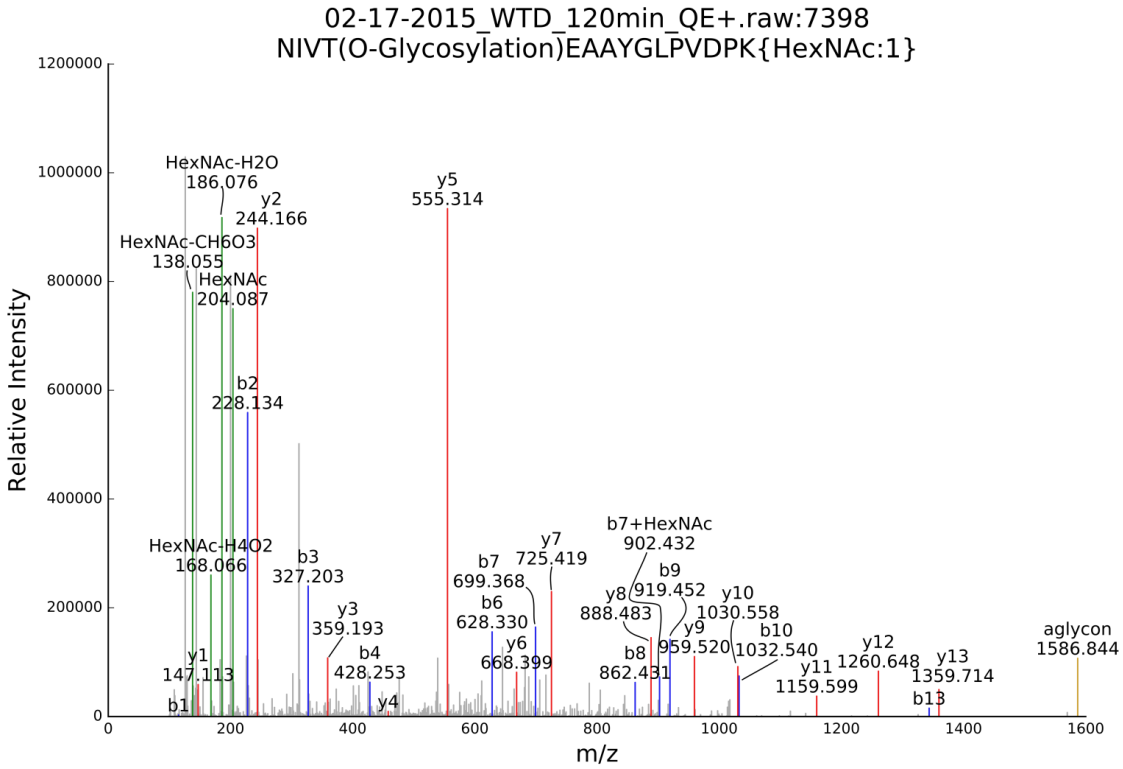


Fig. 4.10. A Representative GlycReSoft Re-Annotated Spectrum, One of 345 Tandem Mass Spectra Deposited into the PRIDE Repository.

This figure presents an example of an MS/MS spectrum initially assigned by PEAKS DB, then manually verified and re-annotated using the in-house software GlycReSoft. GlycReSoft is capable of both discovery and annotation, but only the annotation capabilities were utilized here.

Chapter 5. The *O*-Fucosylated Proteins of *T. gondii.*

The majority of this chapter was published in PNAS, 2016 (415). The full reference is provided below. I present here the mass spectrometry work, which I performed.

Bandini, G., Haserick, J. R., Motari, E., Ouologuem, D. T., Lourido, S., Roos, D. S., Costello, C. E., Robbins, P. W., and Samuelson, J. (2016) *O*-Fucosylated Glycoproteins Form Assemblies in Close Proximity to the Nuclear Pore Complexes of *Toxoplasma gondii*. *Proceedings of the National Academy of Sciences of the United States of America* 113, 11567-11572

Abstract

Toxoplasma gondii is an intracellular parasite that causes disseminated infections in fetuses and immunocompromised individuals. While gene regulation is important for parasite differentiation and pathogenesis, little is known about protein organization in the nucleus. Giuila Bandini, a post-doctoral fellow in the Samuelson lab, showed that the fucose-binding *Aleuria aurantia* lectin (AAL) binds to numerous punctate structures in the nuclei of tachyzoites, bradyzoites, sporozoites, but not to oocysts. AAL also binds to *Hammondia* and *Neospora* nuclei, but not to more distantly related apicomplexans. Analyses of the AAL-enriched fraction with mass spectrometry that I performed indicate that AAL binds *O*-linked fucose added to Ser/Thr residues present in, or adjacent to, Ser-rich domains (SRDs). Sixty-nine Ser-rich proteins were reproducibly enriched with AAL, including nucleoporins, mRNA processing enzymes and cell signaling proteins. Dr. Bandini showed that two endogenous SRDs-containing proteins, a SNF2/SWI2 transcriptional co-activator and a GPN-loop GTPase, and a SRD-YFP fusion localize with AAL to the nuclear membrane. She showed by super-resolution microscopy that the

majority of the AAL signal localizes in proximity to nuclear pore complexes (NPCs). Although host cells modify secreted proteins with *O*-fucose, this is the first *O*-fucosylation pathway described in the nucleocytosol of a eukaryote. Furthermore, these results suggest a novel mechanism by which proteins involved in gene expression accumulate at the nuclear membrane and in close proximity to the NPC.

Significance Statement

We describe here the discovery that assemblies of *O*-fucosylated proteins localize to the nuclear membrane of *Toxoplasma gondii*, particularly in proximity to the nuclear pore complexes (NPCs). *O*-fucose is added to Ser and/or Thr residues in some of the FG domain-containing proteins that characterize the NPC channel, as well as to some Ser-rich sequences in many proteins predicted to have roles in transcription, mRNA processing, and cell signaling. *O*-fucosylation of nucleocytosolic proteins has not been previously described in any eukaryote. In *T. gondii* this appears to be a unique means by which proteins important for gene expression accumulate at the nuclear membrane and in close proximity, likely in association, to the NPC.

Introduction

Toxoplasma gondii is an apicomplexan parasite that causes disseminated infections in humans and these can lead to severe damage in immunocompromised individuals and fetuses (466, 467). There is no human vaccine against *T. gondii* and the price of the drug used to treat toxoplasmosis in the United States (pyrimethamine) has recently increased more than 50-fold (466).

T. gondii has a complex life cycle, and the ability of the parasite to differentiate through its life stages in response to stresses and environmental conditions is fundamental for its pathogenicity and transmission (468). Transcriptome analyses have revealed that a large percentage of mRNAs show life stage-specific expression (469) and/or cell cycle regulation (470). Recent studies have increased our understanding of gene expression in *T. gondii* by identifying the AP2 family of transcription factors (471-473), and by describing post-translational modifications (PTMs) of histones and some of the enzymes responsible for them (474-476). However, little is known about protein organization at the nuclear periphery, a sub-nuclear compartment that plays a critical role in transcriptional regulation in many eukaryotes. In particular, the gene gating model (477)assigns a role to nuclear pore complex (NPC) proteins, not only in protein and mRNA transport, but also in transcriptional regulation and chromatin organization (478-480).

In *T. gondii* chromodomain protein 1 localizes with heterochromatin at the nuclear periphery (481), and centromeres sequester to an apical nuclear region (482). While the nuclear localization signal (NLS) and importin-α system are present, key nuclear import and export molecules are not easily identified (483-485). Furthermore, NPC composition is divergent, so that only phenylalanine-glycine (FG) repeats-containing nucleoporins (FG-Nups) and a putative Nup54 can be predicted by primary sequence homology searches (486).

Here we report the discovery of numerous assemblies of *O*-fucosylated proteins, which associate with the nuclear membrane of *T. gondii*, especially in proximity of the

NPCs. These results improve our understanding the architecture of the nuclear periphery of *T. gondii* and highlight *O*-fucosylation as a PTM involved in assemblies associated with the NPC.

Materials and Methods:

Lectin Pull-Downs.

T. gondii type I RH tachyzoite culture, manipulation, *and in vitro* differentiation to bradyzoites were performed as previously described. For LC-MS/MS analyses, $1.5x10^9$ extracellular tachyzoites were harvested by centrifugation and washed 4x in 1x PBS. Cells were lysed in 2% SDS, 0.15 M NaCl, 0.1 mM DTT in 40 mM TrisHCl pH 7.4, heated 20 min at 50 °C and cooled down to RT. After dilution to 0.03% SDS, 0.8% (w/v) *N*-octyl-glucopyranoside, 0.15 M NaCl, 40 mM DTT, 20 mM TrisHCl pH 7.4 (PD Buffer) plus EDTA-free complete protease inhibitor tablets (Roche), lysate was incubated 2 h rotating at 4 °C with 30 µg biotinylated-AAL, followed by 30 min incubation with Dynabeads® MyOne™ Streptavidin T1. Beads were collected and washed 5x in PD Buffer. Proteins were eluted by incubation with 0.2 M αMeFuc in PD Buffer for 16 h, shaking at 4 °C (E1). E2 was performed by incubating the beads for an additional 2 h in 0.2 M αMeFuc in PD Buffer. For analysis of tagged cell lines, pull-downs were performed on $4.5x10^8$ tachyzoites and amounts of biotinylated-AAL and beads were adjusted accordingly.

Mass Spectrometry.

Proteins in the elution were reduced, alkylated, and digested with proteomics grade trypsin as either in-solution or in-gel digests. The trypsin generated peptides were analyzed by LC-MS/MS using a Chip or UPLC C18 column either on an 6550-QTOF (Agilent Technologies), an LTQ-Orbitrap-XL-ETD, or a Q Exactive Plus Quadrupole hybrid Orbitrap (Thermo Scientific) MS system. The PTM search function within the analysis software PEAKS (Bioinformatic Solutions) (487) was used to produce a list of putative *O*-fucosylated peptides that was manually verified. Scaffold (Proteome Software) (488) was used to compare the five biological repeats and further analyses were performed using RStudio. Monosaccharide composition analysis was performed on a Bruker Scion-SQ GC-MS (Bruker).

In-Solution Digestion of AAL-Enriched Proteins

The following section describes the processing of biological replicates 4 and 5. The proteins in the AAL enrichment eluent, (E1), were precipitated in cold methanol (MeOH) containing 0.1 M ammonium acetate, for at least 18 h at -20 °C. The precipitate was washed with MeOH/ 0.1 M ammonium acetate, any remaining solvent was removed by speed vacuum (Speed Vac® Plus, Savant). The dried samples were re-suspended in 50 mM ammonium bicarbonate pH 8.0, reduced with DTT, alkylated with iodoacetamide, and digested for 18 h at 37°C with proteomics grade trypsin. The resulting digest was dried by speed vacuum and desalted on C18 ZipTip® concentrators (EMD Millipore).

LC-MS/MS Analyses of In-Solution Digests

Samples were reconstituted in 2% ACN, 0.1% FA and separated on an ultra-performance liquid chromatography (UPLC) capillary system (Waters), composed of a nanoAcquity 5 µm Symmetry C18 180 µm x 20 mm trap and a 1.7 µm BEH130C18 150 µm x 10 cm analytical column. The TriVersa NanoMate ion source (Advion) was coupled to either an LTQ-Orbitrap-XL-ETD or QE Plus. Samples were loaded on the trapping column for 4 min at 4 µl/ min, then separated at 0.5 µl/ min using the following conditions: 2-40% B linear gradient (43 min), hold at 40% B for 9 min, wash to 98% B and hold at 98% B (5 min), and re-equilibrate to 2% B (18 min). Alternatively, samples were re-constituted in 1% B and injected on the trapping column at 4 µl/min for 3.75 min. Analytical separation was performed as follows: 0.5 µl/ min flow, 1-60% B linear gradient for 90 min, followed by holding at 60% B for 7 min, ramp to 95% B (3 min), hold for 5 min, and re-equilibration at 1% B for 15 min. On the QE Plus, MS spectra were obtained by scanning over the range *m/z* 350-2000 with 1 microscan and a maximum injection time of 50 ms. MS/MS parameters: top 20 MS/MS isolation window, 1.4 *m/z*, 1 microscan, maximum injection time 45 ms, scan range *m/z* 100-2000. HCD fragmentation was performed at either 27 or 45 eV. For HCD MS and MS/MS, spectra were acquired on the Orbitrap as follows. MS: range *m/z* 300-2000, 1 microscan, max. 500 ms injection time. MS/MS: 2 microscans, max. 500 ms injection time, 2 *m/z* isolation window, and 35 eV HCD fragmentation.

Electron Transfer Dissociation (ETD) Experiments

ETD experiments were performed on an LTQ-Orbitrap-XL-ETD using the same chromatography setup described above (43-min linear gradient). MS scans: 1 microscan, range *m/z* 300-2000. MS/MS scans: isolation width 3 *m/z*, max injection time 25 ms/microscan, fragmentation using ETD with supplemental activation with a 180 ms reaction time, 4 microscans averaged per MS/MS, detection in the linear quadrupole ion trap. The resulting spectra were searched using Mascot against the *T. gondii* GT1 predicted proteome (trypsin as the enzyme with maximum three missed cleavages, 10 ppm precursor and 0.8 Da fragmentation mass tolerances). Carbamidomethyl cysteine was specified as fixed modification and dHex (on Ser/Thr) or methionine oxidation as variable modifications. Figure 5.5 was annotated using the PEAKS software suite using similar parameters as described in this manuscript; however, with the exception that the data was imported as ETD ion-trap data, and the MS/MS error tolerances were loosened to 0.5Da for MS/MS ions.

Monosaccharide Analysis:

Chemical Release of Monosaccharides by Reductive β-elimination

The eluent from the AAL-enrichment (starting material $2.5x10^8$ tachyzoites) was precipitated in cold methanol/ 0.1 M ammonium acetate as described above. Reductive β-elimination was performed by dissolving the protein precipitate in 200 µL 50 mM NaOH containing 1 M $NaBD_4$ and incubated in an oven set to 55 °C for 16 h. Residual $NaBD_4$ was eliminated by adding glacial acetic acid drop wise until effervescence ceased. The

sample was subsequently dried in a speedvac. The boric acid was then removed by extensive drying cycles in 10% acetic acid/ MeOH, followed by 100% MeOH. To isolate the released sugars the dried sample was re-suspended in 2% ACN 0.1% TFA and loaded on a SepPak C18 cartridge (Waters), 4 volumes of H_2O 0.1% TFA was washed through the cartridge and the eluate collected. The sample was lyophilized then per-acetylated to generate alditol acetates.

Per-Acetylation of the Released Monosaccharides

Alditol acetates were generated by adding equal volumes of pyridine and acetic anhydride to the lyophilized sample and heating at 110 °C for 40 min. Once cooled, water was added to quench the remaining acetic anhydride. The alditol acetates were extracted into ethyl acetate and concentrated in a speedvac.

Gas Chromatography Mass Spectrometry (GC-MS) of the Alditol Acetates

The alditol acetates were analyzed using a Bruker Scion-SQ equipped with a 436-GC gas chromatography system utilizing helium as a carrier gas. The concentrated alditol acetates were diluted 1:10 vol./vol. into hexane, and 2 µL injected. The injector temperature was set for 220 °C and maintained a constant column flow rate of 1 mL/ min for the duration of the analysis. The initial split less sample injection was followed by a 100 mL/ min split flow (1min), then 50 mL/ min split flow (59 min). The GC temperature was held at 60 °C for 1 min then raised to 250°C at a rate of 4 °C/ min., then to 300 °C at a rate of 20 °C/ min, holding for 10 min. Separation was achieved on a Restek™ Rxi™-5ms capillary column. Electron impact (EI) ionization was performed with a 70eV source

and spectra acquired in positive mode after a 5 min solvent delay, scanning the *m/z* range 50-500 with a scan time of 500 ms. Spectra were background subtracted and 5 scan averages used to make a composite spectrum. The spectra and retention times were compared to those of genuine deutero-reduced alditol acetate standards analyzed on the same column under identical conditions. To differentiate between rhamnose and fucose, XIC (129 + 171 *m/z*) were compared to that of the genuine standards. The software MS Data Review 8.0 (Bruker) was used for data analysis.

Mass Spectrometry Data Analysis: Protein Database Searches

All MS/MS spectra acquired in the five biological repeats, with the exception of the ETD data, were first analyzed with PEAKS (487). The biological samples (including technical replicates) were analyzed individually. The following parameters were specified for the *de novo* search: parent and fragment mass error tolerances 8 ppm and 0.05 Da, respectively, carbamidomethyl cysteine as a fixed modification, maximum of 8 variable PTMs/ peptide. For the PEAKSDB search, the same settings as above were used with ToxoDB-24_TgondiiGT1_RH as database (combined predicted proteins for GT1 and RH, with redundant protein sequences removed) and False Discovery Rate (FDR) estimation enabled. For the PEAKSPTM search: all parameters were identical to prior sections except a maximum of 5 variable PTMs/ peptide, with all built in Unimod PTMs considered and dHex (on Ser/Thr) or methionine oxidation as variable modifications. Data acquired on the Agilent 6550 was analyzed as above, but monoisotopic parent ion tolerance of 12 ppm and fragmentation tolerance of 0.1 Da were specified.

Mass Spectrometry Data Analysis: Comparison of All Data

The PEAKS PTM results from each biological replicate were imported as unique biological samples into Scaffold 7.5 (Proteome Software) with all technical replicates contained within their respective biological sample group (488). The data were filtered to show 99% protein threshold and a minimum of 10 unique peptides. A 3.6% peptide FDR was determined using the Prophet model. Proteins listed in Appendix 5B and used to build Table 5.1 were either Ser-rich (either ≥10% Ser content or presence of one or more SRD as defined by ≥ 5 tandem Ser residues), with no predicted signal peptide and present in at least four out of five biological samples and/or had at least one observed glycopeptide confirmed by either neutral loss or presence of fragment ions plus dHex in the MS/MS spectrum. The list of putatively fucosylated peptides generated from the PEAKS PTM search (5% FDR maximum threshold for peptide-spectrum matches (PSM)) was then manually verified.

Mass Spectrometry Data Analysis: Manual Verification of Database Search Results

The list of PSM generated from the database searches were all manually verified. Raw data were manually examined using XCalibur Qual Browser (Thermo) or MassHunter Qualitative Analysis B.06.00 (Agilent). Each *O*-fucosylated peptide assignment was manually reviewed in two ways. First, a qualitative assessment of each MS/MS spectrum on the list of peptides containing one or more dHex was performed based on signal intensity, peptide coverage, and parent ion accuracy. The higher quality spectra were manually sequenced with the goal of specifying the locations of the dHex(s) on the peptide. Second, the MS spectrum corresponding to the scan used for the data-

dependant decision tree for the automatic MS/MS of the peptide was reviewed for secondary evidence of dHex. This was accomplished by measuring the intervals from the corresponding to the mass of one or more dHex(s) in the same MS scan (within ±5 mDa) that could be attributed to neutral losses from a single precursor. If peaks having mass differences corresponding to one or more dHex(s) were observed at different retention times, these were considered to be indicators of PTM heterogeneity on a single peptide.

Bioinformatics and Statistical Analyses.

Predicted and known functions for the enriched proteins were assigned based on annotation on ToxoDB (489) and/or the presence of conserved PFAM domains identified by BlastP searches. Statistical analyses were performed by Giulia Bandini, in RStudio with a custom script using packages Biostrings, ggplot2 and plotrix. For the boxplot in Fig. 5.1A, only non-redundant peptide sequences were used and the maximum number of dHex observed was used to assign the peptide sequence to either group. The control set in Fig. 5.1B was obtained from a GO Term search on ToxoDB using the following terms: DNA replication (GO:0006260), RNA processing (GO:0006397), nucleus (GO:0005634), and nucleoplasm (GO:0005654).

Results

AAL Binds to *O*-Fucose in *T. gondii*.

LC-MS/MS of tryptic peptides of AAL-enriched proteins from extracellular tachyzoites identified 69 unique glycopeptides containing 1-6 deoxyhexose(s) (dHexs), each linked to Ser or Thr (Appendix 5A). These glycopeptides correspond to 50 different

peptide sequences, and in many cases different numbers of dHexs were observed on the same peptide. All glycopeptides were manually reviewed and confirmed by either prompt neutral loss of at least one dHex, as indicated by the presence of peaks in the MS that correspond to different glycoforms detected at the same retention time, and/or by peaks in the MS/MS spectra that can be confidently assigned to peptide fragments containing dHex (Appendix 5A). No sugars other than dHex were observed, indicating that AAL is binding to *O*-fucose (*O*-Fuc). In a few cases, the same peptide was observed to be modified with a dHex, and at a later retention time was also observed in its unglycosylated form, suggesting the addition of *O*-Fuc may be a probabilistic event (Appendix 5A). The amino acid modified with *O*-Fuc could not be determined for most glycopeptides, due to the labile nature of the *O*-Fuc and the low complexity amino acid sequence of most of the modified peptides (Fig. 5.1B and Fig. 5.3B). Also, we cannot exclude the possibility of heterogeneity in the glycosylation sites. Combining high-energy collision dissociation (HCD) and electron-transfer dissociation (ETD) MS/MS data, it was often possible to narrow down the modification sites to 2-6 likely Ser/Thr residues (Appendix 5A, Fig. 5.3*B,* and Fig. 5.5). Furthermore, we were able to specify the modification site for one glycopeptide, T610 on TGGT1_203780 (Fig. 5.1*B*), a putative FG-Nup and one of the most abundant proteins identified in the AAL-enriched fraction (Appendix 5B). GC-MS monosaccharide composition analysis of the sugars released by reductive β-elimination from the AAL-enriched fraction identified Fuc as the only dHex present (Fig. 5.3*C* and *D*).

AAL Recognizes Proteins Involved in Gene Regulation.

In most cases, *O*-Fuc is added to low complexity Ser-rich domains (SRDs) or to sequences adjacent to these domains (Fig. 5.1A and Fig. 5.3B) and peptides with long SRDs are likely to have more than one dHex (Fig. 5.4B and Appendix 5A). By SRDs we define sequences that have five or more Ser residues in tandem. More than 70% of the peptides identified in the AAL-enriched fraction came from proteins with one or more SRDs and/or contained more than 10% Ser (Fig. 5.1*A*, Fig. 5.4*A* and Fig. 5.4*B*, and Appendix 5B). Consistent with this observation, a higher number of proteins with Ser 15-mer is found in the apicomplexan that show nuclear AAL binding, compared to those which do not (Fig. 5.4*C*). Comparison of the data from five biological repeats resulted in a set of 69 Ser-rich proteins reproducibly pulled down by AAL and/or for which we observed glycopeptides. For 33 of the 69 AAL-enriched proteins we identified glycopeptides (Table 5.1 and Appendix 5B). It is likely that numerous *O*-fucosylated peptides on long SRDs were not detected because of the paucity of flanking trypsin cleavage sites. In contrast, proteins that were present in AAL-pull downs, but which are likely contaminants because of their high abundances in the cytosol (e.g. ribosomal proteins, cytoskeletal components, chaperones, etc.), do not contain SRDs and have an average of 6% Ser. A control set comprising *T. gondii* proteins associated with nuclear and cytoplasmic Gene Ontology (GO) Terms was also analyzed and shown to have an average Ser content of 9%, compared with the 15% Ser present in the *O*-fucosylated proteins or the SRD set (Fig. 5.4*B*). About 40% of the 69 proteins with one or more SRD contain a canonical NLS, as identified by cNLS mapper (490) (Table 5.1, Fig. 5.4*A*, and

Appendix 5B). PTMs analysis showed that the majority of AAL-enriched proteins exhibit some degree of phosphorylation, but none are modified by ubiquitin. This is in agreement with the data sets in (491) and (492). AAL enriches five out of seven predicted *T. gondii* nucleoporins (four FG-Nups and a Nup54 ortholog) each of which has one or more SRDs (Table 5.1 and Appendix 5B). Also present in the AAL-enriched fraction are proteins predicted to be involved in mRNA processing, protein-protein interactions, ubiquitination, and enzymes that catalyze the addition/removal of phosphate groups from proteins and polyphosphate phosphaditylinositol (493, 494). Transcription regulators and proteins with nucleotide-binding and chromosome-binding domains are also present (Table 5.1 and Appendix 5B). Numerous hypothetical proteins are present and these are often conserved in *T. gondii*, *H. hammondi*, and *N. caninum,* but absent in the apicomplexan that do not bind AAL.

Discussion

Almost all eukaryotes present a glycosylation pathway dedicated to the modification of cytosolic and nuclear proteins in which *N*-acetylglucosamine is transferred to Ser/Thr in disordered domains by *O*-GlcNAc transferase (OGT) (495). Yeast, one of the few organisms lacking a OGT, has been recently shown to use *O*-Man instead of *O*-GlcNAc to modify its nucleocytosolic proteins (496). In contrast, *T. gondii* has three cytosolic glycosylation pathways: an OGT (158), the hydroxylase and glycosyltransferases (GTs) that modify Skp1 with a pentasaccharide (497), and the nucleocytosolic *O*-fucosylation system described here. In the host and presumably the parasite, the donor for the OGT reaction, UDP-GlcNAc, is sensitive to the metabolic state

of the cell, and modification by *O*-GlcNAc affects protein activity (495). Similarly, glycosylation of Skp1 is required for normal growth of tachyzoites in culture, and proline hydroxylation on Skp1 is sensitive to the redox status of *T. gondii* (497, 498). In contrast, this study suggests that addition of *O*-Fuc targets *T. gondii* proteins to assemblies closely associated with the nuclear membrane and that targeting of endogenous and exogenous proteins to the AAL-labeled assembly may occur in the absence of an NLS. SRD-YFP and two endogenous proteins, one containing a predicted NLS, are modified with *O*-Fuc and localize to the nuclear periphery, while addition of an NLS to either Cas9 or YFP targets proteins to the nucleoplasm, nucleolus included. Whether addition of *O*-Fuc affects the activity of *T. gondii* proteins or protein-protein interactions was not determined here.

Although *O*-fucosylation of Ser/Thr residues has been previously described in secreted proteins of eukaryotic cells, this is the first time this modification has been identified on nuclear proteins. AAL staining suggests this pathway is conserved only in *T. gondii*, *H. hammondi* and *N. caninum.* Both human fibroblasts and bovine turbinate cells did not show nuclear staining by AAL. Further studies in different taxonomic groups should be performed, but the limited data so far suggest nuclear *O*-fucosylation may be restricted to these three species. Recognizing the limitations we have encountered thus far in precisely defining the modification site(s), it appears the *T. gondii* unidentified *O*-fucosyltransferase (OFucT) differs from host protein *O*-fucosyltransferases (POFUT1 and POFUT2) in its location and acceptor specificity. First, POFUTs are ER-resident glycosyltransferases (486), while we predict that the *T. gondii* OFucT is either cytosolic

or nuclear. Second, the host enzymes transfer *O*-Fuc to epidermal growth factor-like (POFUT1) or thrombospondin type I (POFUT2) repeats, both of which are characterized by conserved disulfide bonds (499, 500), while disulfides do not form in cytosolic and nuclear proteins.

Our working model of protein *O*-fucosylation in *T. gondii* is shown in Fig. 5.2. AAL-enrichment and the identified glycopeptides suggest that the acceptors of the putative OFucT are for the most part SRDs or proteins containing such domains. However, it seems likely that not all nucleocytosolic proteins with SRDs were identified in the AAL enrichment. In our model, we speculate that an as-yet-unidentified fucose-binding lectin would recognize *O*-fucosylated proteins and participate in their accumulation in assemblies closely associated with the nuclear membrane. This would be similar to the host cell secretory pathway where lectins bind glucosylated and mannosylated *N*-glycans (105). Many hypothetical proteins, with no homology to any known conserved domains and specific to *T. gondii*, have been identified in the AAL pull-downs, and they may be important in forming the assemblies, *i.e.* the hypothesized fucose-binding lectin. The AAL-labeled assembly is also likely to contain proteins that are not *O*-fucosylated and, in our model, these proteins would associate via protein-protein interactions. Protein-protein interactions, disrupted during lysis, reform during lectin enrichment, and we therefore cannot exclude the possibility that some of the proteins isolated in the pull-down that were categorized as contaminants are actually non-fucosylated members of the assemblies. Isolation of the intact assemblies will be required to discriminate between non-fucosylated proteins that are members of the assemblies and

true contaminants. This distinction might be further complicated by the possibility that the protein composition of the assembly is itself heterogeneous. Furthermore, we observed addition of *O*-Fuc as a probabilistic event for ~20% of the identified glycopeptides. According to our model, this would suggest the fucosylated form of the protein might be present in the AAL-labeled assembly, while the non-fucosylated form might diffuse into the nucleus, if an NLS is present, or remain in the cytosol, if an NLS is absent, as suggested by the localization of GPN-3xMYC. Whether this is the result of the OFucT mechanism and/or kinetics, or it is because nuclear *O*-Fuc is, like *O*-GlcNAc, a reversible modification cannot be clearly stated at this point. No putative *O*-fucosidase could be identified so far in *T. gondii*.

Multiple pieces of evidence point to the potential importance of the AAL-labeled assemblies in the nucleus of *T. gondii*. First, proteins in AAL pull-downs include numerous putative nucleoporins, mRNA processing enzymes, transcription regulators, and signaling proteins. Second, disruption of GDP-Fuc biosynthesis, which eliminates binding of AAL to nuclei, appears to severely affect growth, as suggested by our inability to clone the AAL-negative cells. Additional experiments will be performed to investigate if *O*-fucosylation and localization to the assemblies are necessary for the function of individual proteins and growth of *T. gondii.* The absence of AAL binding to oocyst nuclei suggests that these proteins are either absent or not glycosylated in these life stages, since electron microscopy showed that the nuclear membrane of sporulating *T. gondii* remains intact (501). Lastly, five out of seven predicted nucleoporins are present in the AAL enrichments, and tagged versions of Nup68 and Nup67 partially co-localize with AAL-

labeled assemblies. Furthermore, mass spectrometry showed that *O*-Fuc is found on three FG-Nups: Nup68, TGGT1_203780, the second most abundant protein in the pull-down (Appendix 5B), and TGGT1_313430, *T. gondii* ortholog of yeast Nup98/96 (502). In all three instances, the sugar modifies the FG repeats region. FG regions are the disordered sequences that characterize the NPC channel and interact with karyopherins to mediate nuclear transport (484). In higher eukaryotes and yeast, FG regions are highly decorated with *O*-GlcNAc and *O*-Man, respectively, and evidence suggests that this PTM could affect cargo selectivity (495, 496).

O-fucosylation appears then to be a novel mechanism by which *T. gondii* proteins involved in gene expression and mRNA processing (gating hypothesis) are gathered at the nuclear membrane often in close proximity to the NPCs (479), and components of the NPC itself are *O*-fucosylated. The mechanism of the association between the NPC and the *O*-fucosylated protein assemblies at the nuclear periphery (*i.e.* by a fucose-binding lectin or by protein-protein interactions) and whether it might be static or dynamic are the next important questions we shall address.

Acknowledgments

We thank Dr. Aparajita Chatterjee and Dr. Carolina Agop-Nersesian for invaluable assistance with oocyst excystations. Thanks to the Harvard Center for Biological Imaging (HCBI) for infrastructure and support. The ELYRA microscope was acquired through NIH SIG award (S10 RR27990) to HCBI. We would also thank Dr. Eliza Vasile at the Swanson Biotechnology Center Microscopy Core Facility for technical assistance with the OMX. We thank Dr. Christopher West for helpful

discussions. Support for this study came from NIH grants R01 AI110638 (J.S.), R01 GM031318 (P.W.R.), and P41 GM104603 (C.E.C.), NIH-NHLBI contract HHSN268201000031C (C.E.C.) and by a grant from the Mizutani Foundation of Glycoscience (J.S).

Tables

Protein families	NLS [a] +	NLS [a] -	Unique *O*-fuc peptides (proteins[b])
Autophagy	1	0	1 (1)
DNA-binding proteins	3	2	3 (3)
Hypothetical proteins	9	15	29 (13)
Kinases/Phosphatases	2	4	2 (1)
mRNA processing	5	2	2 (2)
NPC/Nuclear Transport	2	4	14 (4)
Nucleotide-binding domain	0	2	
Protein/Protein interaction	5	5	16 (7)
Transcription regulators	2	1	1 (1)
tRNA synthesis	2	0	
Ubiquitination related	1	2	1 (1)
	32	37	69 (33)
Total	69		

Table 5.1: Proteins Containing Ser-Rich Domains Identified by AAL Pull-Down and Grouped by Function (Putative or Annotated).

[a] NLS: nuclear localization signal
[b] Number of proteins from the corresponding family for which glycopeptides were observed

Figures

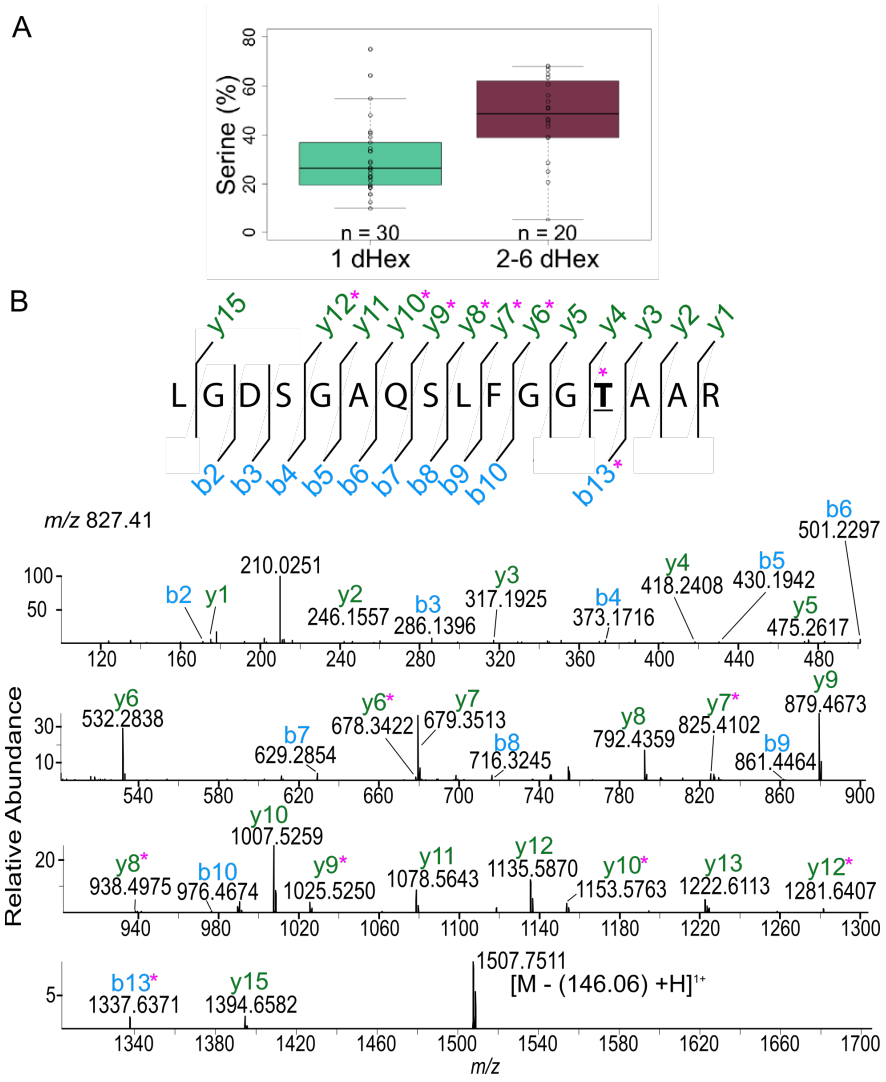


Fig. 5.1: AAL Bound Proteins are Modified by *O*-Fucose on Ser/Thr Residues.

AAL bound proteins are modified by *O*-fucose on Ser/Thr residues. (A.) Peptides with a higher Ser percentage are more likely to be modified with more than one *O*-Fuc (dHex) residue. (B.) HCD MS/MS spectrum of a peptide from a FG-Nup (TGGT1_203780), *m/z* 827.41. Ions b_{13}+dHex and y_6+dHex were observed, indicating Thr610 (*bold underlined*) is the modified residue. Ions plus dHex are marked by a *pink asterisk*.

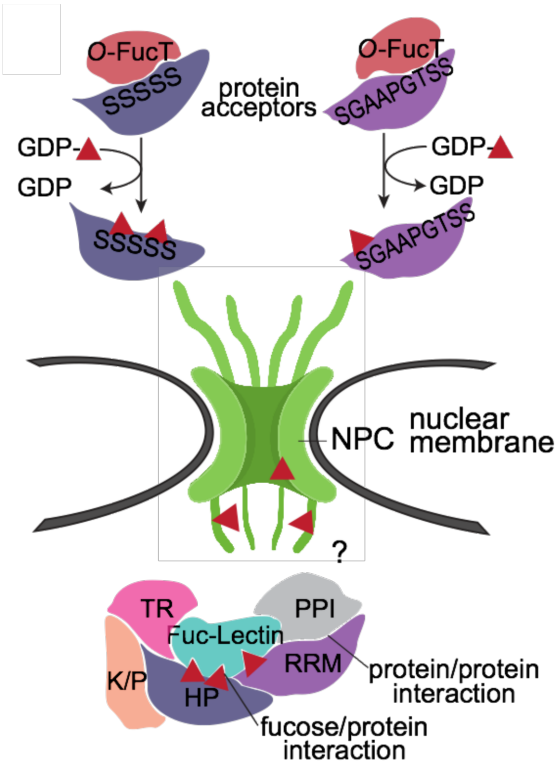


Fig. 5.2: Model for *O*-Fucosylation of Nuclear Proteins in *T. gondii*:

Shown is a model for *O*-fucosylation of nuclear proteins in *T. gondii*: a yet to be identified OFucT transfers fucose (*red triangle*) to Ser/Thr residues on SRD-containing proteins. *O*-fucosylated proteins are then shuttled to the nuclear periphery where they associate in assemblies, via protein-protein interactions and, possibly, an endogenous fucose-binding lectin *(question mark)*. How the assembly of *O*-fucosylated proteins interacts with the NPC remains to be determined (*question mark*). According to the model, non-fucosylated proteins (*asterisks*) will stay in the cytosol or go to nucleus depending on the presence of a NLS. PPI: protein-protein interaction domain; RRM: RNA-recognition motif; K/P: kinase or phosphorylase domain; TR: transcriptional regulator; HP: hypothetical protein.

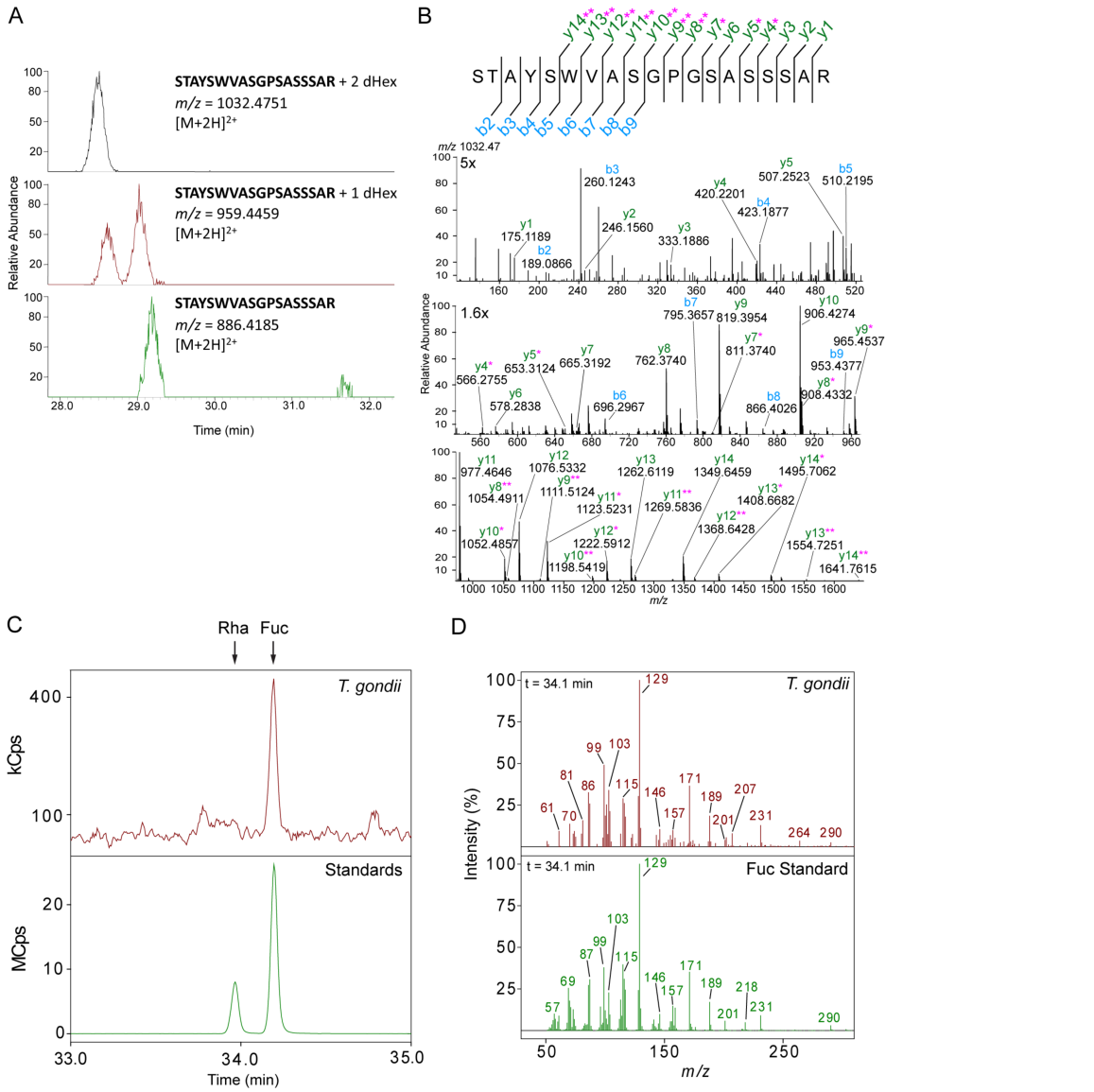
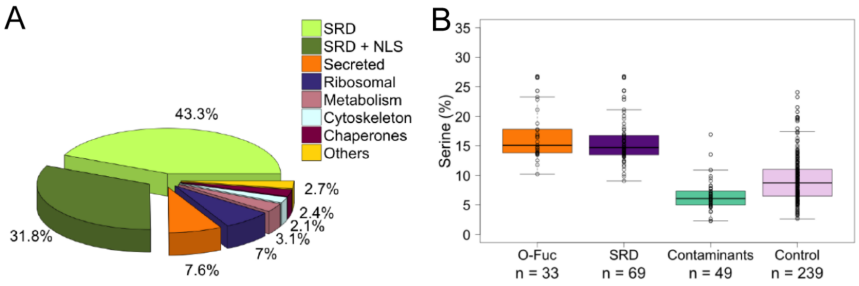


Fig. 5.3: AAL Binds Proteins Modified by One or More *O*-Fuc on Ser/Thr Residues, as Verified by Neutral Loss of dHex, MS/MS Fragmentation and GC-MS Monosaccharide Analysis.

(A.) Extracted ion chromatograms (XIC) obtained from LC-MS spectra of a glycopeptide from TGGT1_285190, show neutral loss of 2 to 1 dHex(s) and in smaller amount from 1 to 0 dHex. (*B.*) MS/MS spectrum for the glycoform with 2 dHexs, *m/z* 1032.47. The y and b series are marked in *green* and *blue*, respectively. Fragment ions plus one or two dHexs are marked with one and two *pink asterisks*, respectively. (*C.*) GC-MS extracted ion chromatograms for the *T. gondii* sample and the standards mixture. The matched retention time compared to the standard, indicate Fuc is the only dHex present after reductive β-elimination release of sugars from AAL-enriched *T. gondii* proteins. XIC were generated from the ions (129 and 171 *m/z)*, these ions are abundant in the EI spectra from dHexs alditol acetates. XICs from 33-35 min are shown and the arrows indicate the rhamnose (Rha) and Fuc retention times. *D*. GC-MS EI spectra of the alditol acetates from *T. gondii* sample and Fuc alditol acetate standard. The retention time (34.1 min) from the XIC shown in *C* was used to generate the spectrum for each. Both the ions pattern and retention time match the Fuc standard and not that of Rha.

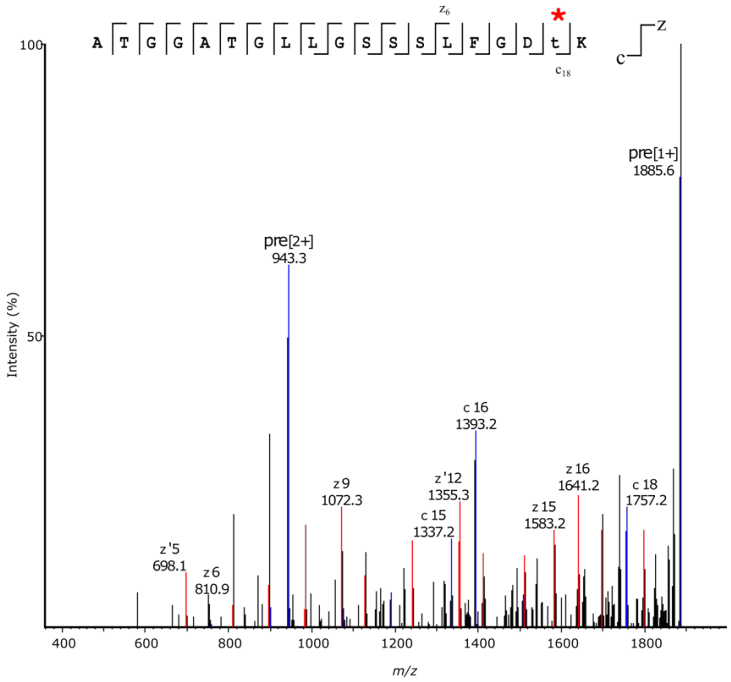


C

		Species					
		T. gondii GT1	*H. hammondi*	*N. caninum*	*S. neurona*	*P. falciparum*	*C. parvum*
Tot. Nb genes		8460	8003	7112	7077	5398	3805
Nb. 15-mer	Ser	161	118	108	14	6	18
	Thr	2	2	3	1	0	20

Fig. 5.4:Proteins and Peptides with Serine Rich Sequences.

(A.) Peptides identified in a representative AAL enrichment. More than 70% of the peptides belong to proteins with SRDs (light and dark green). Of the proteins containing SRDs, 40% are predicted to have a nuclear localization signal (dark green). (B.) Comparison percentages of Ser in proteins for which glycopeptides were identified (*O*-Fuc), Ser-rich proteins with no predicted signal peptide in the AAL enrichment (SRD), other proteins in the enrichment (Contaminants), and a control set of predicted cytosolic and nuclear proteins (Control). (C.) Number of 15-mer of Ser or Thr residues present in various apicomplexan. Higher numbers of Ser 15-mer are present in the species that are bound by AAL.



#	c		Seq	z	z'	#
1	89.1		A			19
2	190.1		T	1798.0	1799.3	18
3	247.1		G	1697.3	1697.4	17
4	304.2		G	1640.2	1641.2	16
5	375.2		A	1583.2	1583.3	15
6	476.2		T	1512.0	1512.4	14
7	533.3		G	1411.1	1412.2	13
8	646.4		L	1354.2	1355.1	12
9	759.4		L	1241.1	1241.4	11
10	816.5		G	1128.0	1128.0	10
11	903.2		S	1071.0	1071.9	9
12	990.0		S	983.7	984.4	8
13	1077.2		S	896.4	897.2	7
14	1190.3		L	809.6	810.9	6
15	1337.2		F	696.3	697.7	5
16	1394.3		G	549.2	550.2	4
17	1509.6		D	492.2	493.2	3
18	1757.2		T(+146.06)	377.2	378.2	2
19			K	130.1	131.1	1

Fig. 5.5: ITMS ETD MS/MS *m/z* 943.4736 $[M+2H]^{2+}$, (390)ATGGATGLLGSSSLFGDTK(408) + 1 dHex

This spectrum was annotated using PEAKS v8.0. Shown is an example ETD MS/MS spectrum, acquired on the LTQ-Orbitrap-XL-ETD. The MS/MS was acquired in the ion trap, with a 180ms. ETD reaction time with supplemental activation (see methods section for details). The complete peptide sequence is observed with near complete c- and z-ion coverage. The peptide sequence (390)ATGGATGLLGSSSLFGDTK(408) belongs to the hypothetical protein (TGGT1_273850) This protein shares homology with FG repeat-containing protein Nup68. The two ions which strongly support the assignment ofThr-407 modified with a dHex (*red asterisk*), are c_{18} and $z_{6.}$ Although, all of the z-ions, and the z +1 ions observed also support this assignment, since they contain the extra mass equal to that of a dHex., the c_{18} ion is the only c-ion observed which contains added mass equal to a dHex. The table in right corner shows all the assignments (ions ≤ 0.5Da of theoretical values.).

Chapter 6. Significant Findings and Future Work.

The work performed and described in this dissertation has concerned two pathogenic parasites*, C. parvum and T. gondii.* Much of the analytical work was descriptive; however, the exploratory work revealed several significant, interesting, and novel discoveries. All of the glycoproteins that have been previously proposed as vaccine candidates in *C. parvum* were characterized through these studies. It has been known that many of the immunodominant antigens are glycosylated, and their importance to the immune response has been established to some degree through indirect methods, but the detailed structures had not been established.

The work presented in Chapters 2 and 4 described the *C. parvum N*- and *O*-glycosylated proteins, as well as the novel finding that the immunodominant antigen (Cp23) is both palmitoylated and myristoylated at its N-terminal domain. These details provide long-sought information missing from all the previous descriptions of the immunogenic proteins in *C. parvum* (e.g. Gp15, Gp40, Gp900, and Cp23). The proteins (Gp15, Gp40, and Gp900) all are known to have epitopes which are carbohydrate-based (15, 31). These epitopes are either *O*-glycans (Gp15/Gp40/Gp900) or *N*-glycans (Gp900). Including the aforementioned proteins which the literature has focused on, there are at least fifteen glycoproteins that are dominant immunogens; eight of these are *N*-glycosylated (36). It is certainly plausible that many of the proteins described here could be among the fifteen immunodominant glycoproteins that have not been identified.

For the *O*-glycosylated proteins of *C. parvum*, the novel findings were made here regarding Gp40 and Gp20. These proteins have long stretches of Ser/Thr within which

the levels of *O*-HexNAcylation almost reach saturation. This certainly appears to be the case with the peptide from Gp40. The immunogenicity of Gp20 is unknown, since this protein has not been described in the literature, but, due to the low complexity and heavy modification with *O*-HexNAc, the sequence of this peptide could not have been easily defined with most techniques used for *C. parvum* in previous studies (such as Edman degradation sequencing). It would be worthwhile to explore the functional role of Gp20 and its localization within the oocyst, and to determine whether or not it is an immunogenic protein. The only other parasites that appear to have abundant numbers of mucin-like proteins with simple Tn-antigen *O*-glycans are some helminths (503).

It is intriguing that other pathogens which encounter the epithelia mucosa for part of their life cycle have abundant sets of mucins or mucin-like proteins, such as in *Trypanosoma cruzi, Leishmania spp., Toxocara canis, Echinococcus multilocularis, Fasciola hepatica,* and *Mesocestoides vogae,* (39, 504-509). The functions of these mucin-like proteins are not entirely known. However, it has been suggested that, as with the mucins of the helminths, they may mimic the host glycoproteins to try to evade the immune system, or, they may be used to actively modulate the host's immune response (510). Furthermore, it has been demonstrated that the parasitic nematode *Toxocara canis* has several secreted mucins, which have long stretches of tandem Thr, are heavily glycosylated, and are thought to help the parasite evade the immune system (506). These mucins resemble those of *C. parvum*, and, by comparison, could help provide insight into additional mechanisms in which the *C. parvum* mucins may play roles.

One of the most abundant glycoproteins in the *C. parvum* oocyst preparations was POWP1. The function of this protein remains to be determined, as there are no homologous proteins in any other organisms, other than *C. hominis*, and distantly in *C. muris*. The novelty to the human-pathogenic cryptosporidium species implies that it may have a novel function in this organism. The density of *N*-glycosylation sites on such a relatively small protein, the presence of a coil-coil domain, and the stretch of Tyr-rich residues, suggest that it could be a structural component of the oocyst wall. Additional studies should be performed to characterize the localization of this protein POWP1 (Cgd2_490) and to ascertain its function. A brief example of a follow-up study would be to generate anti-sera against the recombinant protein for immunofluorescence studies, or to use the sera from previously infected individuals to test for reactivity against a recombinant POWP1.

The discovery that the *C. parvum* immunodominant antigen Cp23 is a lipoprotein, has profound implications in the development of vaccines or therapies. To date, all studies with Cp23 have used the native immunogen as whole/broken oocysts, or as a recombinant protein. Ideally for vaccine development, it would be best to use a well-defined antigen. Isolation of the native immunogen is not always feasible, and a recombinant protein is not likely modified in the same manner as that of the native immunogen. It was previously demonstrated that bovine hyperimmune colostrum from whole oocyst inoculated cows, can be used as an effective treatment to prevent cryptosporidium infection in otherwise healthy human subjects (25). In one study, acquired immunity against cryptosporidium appears to have minimal efficacy in the

prevention of re-infection of humans; however, detectable IgA, IgG and IgM isotypes were detectable prior to secondary challenge with oocysts(25). Whereas diarrheal symptoms did not appear to be reduced, the secretion of oocysts was markedly reduced (25).

Cryptosporidium infections do increase fecal IgA levels, further implicating that mucosal secretion of antibodies may be the most viable method of acquiring immunity (511). It has been demonstrated that there is an inverse relationship between an increase of fecal IgA levels and oocyst shedding in mice, these antibodies specifically recognize (15-, 31-, 38-, 42-, 56-, and 254-kDa) bands on a Western blot, which are likely the same Gp15, Cp23, Gp40, and Gp900 components described elsewhere (512). These studies suggest that a well-characterized antigen could be used for nasal or oral inoculation to prevent infection, or, to help reduce secretion of oocysts to circumvent re-infections.

In *P. falciparum*, an apicomplexa like *C. parvum*, an artificial construct of tripalmitoyl-*S*-glyceryl-cysteinyl-serine conjugated to the N-terminus of the immunodominant epitope of the circumsprozoite protein, showed a marked increase in immune response in BALB/c mice without the need for additional adjuvants (513). A similar strategy was performed with chimpanzees, against a liver stage-specific antigen from *P. falciparum*, which had a palmitate conjugated to the N-terminus (514). Taking these facts into consideration, a similar strategy for immunization against *C. parvum* seems very plausible. The sequences we identified as lipid modified on Cp23 could be used as a template to develop a vaccine utilizing a lipid-peptide-conjugate, or stable-lipid-mimic construct. To further support this concept, inoculation with synthetic di-

palmitoylated peptides is proven to be an effective method of immunization, especially for eliciting a mucosal response (515).

I have concluded this dissertation with a discussion of the findings made in the project spearheaded by Giulia Bandini, with regards the *O*-fucosylated proteins in *T. gondii*. The collaborative efforts we made proved to be very fruitful. The arduous manual interpretation and verification of every single possible *O*-fucosylated peptide at the MS level, looking for neutral losses, and the MS/MS level looking for the exact localization of the very labile *O*-fucose, produced a very confident and high quality result. The data suggests that there are two populations of *O*-fucosylated proteins, FG-NUPS, and proteins containing stretches of tandem Ser. One of the most abundant O-fucosylated proteins is an FG-NUP. This protein and the other FG-NUPs identified, do not have the same low complexity tandem Ser-repeating regions modified with O-Fuc, as seen in the other proteins, but contain discrete regions within the FG-repeating units, modified with *O*-fucose on Thr, or, to a lesser extent, on Ser. The proteins in the other group are as a whole quite abundant. This group contains the patches of Ser-rich repeats, often flanking conserved functional domains. The hydrophilic stretches of serines carry one or more fucose(s) and are interrupted with aliphatic or hydrophobic amino acids. This pattern is reminiscent of components involved in other nuclear transport mechanisms, such as the Kapβ and Kapα classes of nuclear transporters. These proteins have a tertiary structure which features repeating domains, such as the HEAT motifs and ARM repeats. These features are helices exposing ideally spaced binding sites on the helices or in between the repeats (155, 156). I hypothesize, that the Ser-rich repeats modified with *O*-fucose act as

a binding signal, allowing pairing up only to specific FG-NUPs modified with *O*-fucose. In this model, the hydrophobic FG repeats would interact with the repeating hydrophobic tracks located between the Ser stretches, with the fucose residues on the Ser repeats, and along the FG-NUPS, would hydrogen bond, facilitating a lock-and-key mechanism. This would be analogous to the mechanism of, Kapβ, which binds to the FG on FG-NUPS in hydrophobic grooves (150, 157).

Finding abundant *O*-fucosylation directly conjugated to Ser/Thr is of itself an interesting discovery, due to the rarity of such modifications. However, what makes this discovery even more interesting and profound is that the modification is most abundant on the nucleoporins. Furthermore, these modifications are on the FG-NUPs, which are known to be involved in selective transport into the nucleus, but the mechanics have never been understood. I believe that the work and the discoveries made here in *T. gondii*, clearly set up a plausible mechanism for selective nuclear transport. Future research should be performed to test this hypothesis. The mechanism, if understood, not only would impact the field of biology, but could potentially be exploited to develop a therapeutic to treat toxoplasmosis.

APPENDIX 1: The Amino Acid Sequence of Gp900.

>Gp900 | cgd7_4020 | organism=Cryptosporidium_parvum_Iowa_II

MVNIKVSSSAIALVAVIMNPLFSLAFKSSNRLEMRIESSGAVSNEKFVIPSLPSDLDPTTFLLIDSTGKKFS
PYTGKHADASTTSSAYSAPFELDVSGVPIEPNTRRMVDPVSLMLFDNSTGVMYDPNTNSILEGSIAGIRSES
CIVSELNFTSTTGFTTDTSMNWPVSITSGELKDPNKQATISGSRSCGWKQGYSIDSSTGFRVDSITGLPTDP
YSNCPFNPVTGNLVSRSTGKTIPNTYAGVYRSNETKTTEPSANTNFLLVDPKINAPCNSENSFEQGQIFDMG
SKVYIPYTKCVGVKH

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT {Thr Repeat}
TTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT AA 309-637
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 317 Thr
TTTTTTTTTTTTTTTTTTTTTTTTSTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTT 94.91%
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTKKPTTTTTATTTTTT

SETESVIKPDE
WCWLEKNGECEAKGATYVGVIGKDGRIENGMAFTMIPNDDTHVRFRFKVKDVGNTISVRCRKGAGKLEFPDR AA 638-796
SLDFTIPPVAGHNSCSIIVGVSGDGKIHVSPYGSKDVSLISAPIQPSELFNEVYCDTCTAKYGAIHSGYQTS
ADFV

TTTTAKPTTTTTGAPGQPTTTTTGSPSKPTTTTTTKATTTTTTLNPIITTTTQKPTTTTTTKVPGKPP {Thr Repeat} AA 797-908
IATTTTTLKPIVTTTTTKATTTTTTTVPTTTTTTKRDEMTTTTT 64 Thr 57.14%

PLPDIGDIEITPIPIEKMLDKYTRMIYD
YNSGLLLDSNDEPIPGSQAGQIADTSNLFPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSVSY
LAAKNLTVDTDETYGLPIDTLTGYPLDPVSLIPFNPETGELFDPISDEIMNGTIAGIVSGISASESLLSQKS
APIDPATNMVVGEFGGLLNPATGVMIPGSLGPSEQTPFSPEIEDGGIIPPEVAAANADKFKLSIPPSVPESI
PEKDQKIDSISELMYDIESGRLIGQVSKRPIPGSIAGDLNPIMKTPTQTDSVTGKPIDPTTGLPFNPPTGHL
INPTNNNTMDSSFAGAYKYAVSNGIKTDNVYGLPVDEITGLPKDPVSDIPFNSTTGELVDPSTGKPINNSTA
GIVSGKPGLPPIEDENGNLFDPSTNLPIDGNNQLVNPETNSTVSGSTSGTTKPKPGIPVNGGGVVPDEEAKD
QADKGKDGLIVPPTNSINKDPVTNTQYSNTTGNIINPETGKVIPGSLPGSLNYPSFNTPQQTDEITGKPVDT
VTGLPYDPSTGEIIDPATKLPIPGSVAGDEILTEVLNITTDEVTGLPIDLETGLPRDPVSGLPQLPNGTLVD
PSNKKPIPGSHSGFINGTSGEQSHEKDPSTGKPLDPNTGLPFDEDSGSLINPETGDKLQGSHSGTFMPVPGK
PQGENGGIMTPEQILEALNKLPTSNEVNISPRPSSDAVPDRPTNTWWNKISGQTYQVDGKKTIPGSAASVIH
TALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIKGSVPYVSLYVKEKNIVTEAAYGLPVDPKTG
FPIDPISYLPFAKNGELIDPISGKYFSGSIAGFISGKAGSQSKSSDESGNPIDPSTNMPYDPKTGKLIDPES
GIAIDNSVSGVFATVPGTAAPKKGGVIPESVAAEAAKKYFAANVEGEGEGEEVPPPPESSSNIAIQAAGGAS
AAVGLVAAGVGAWYASRNRQEGEDDDDYADGFEAEYEEEEEEEGDEAANETVVTIERDSSFWNES

APPENDIX 2: Ion Assignments, Peptides, Sequons, and Bioinformatics Data for the *N*-Glycosylated Peptides.

APPENDIX 2A Fragment Ion Assignments for FT-ICR EED MS/MS Spectra of Deutero-Reduced and Permethylated *N*-Glycans.

APPENDIX 2A.1 Assigned Ions for $HexNAc_2Hex_5$.

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_5$, Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
1596.8199	1.03E+10	$Hex_5HexNAc_2$-(^{2}H)-reduced (un-fragmented precursor)	-0.17	1596.8202
1522.7831	4.17E+08	$(^{0,4}X_{5\alpha})$,$(^{0,4}X_{4\alpha})$,$(^{0,4}X_{3\alpha})$,$(^{0,4}X_{3\beta})$,$(^{0,4}X_1)$	-0.19	1522.7834
1478.7569	8.23E+06	$(^{0,3}X_{5\alpha})$,$(^{0,3}X_{4\alpha})$,$(^{0,3}X_{3\alpha})$,$(^{0,3}X_{3\beta})$	-0.19	1478.7572
1406.6997	3.50E+08	$(^{1,5}X_{5\alpha})$,$(^{1,5}X_{3\beta})$	0.02	1406.6997
1376.6892	1.54E+08	$(Y_{5\alpha})$-2H,$(Y_{3\beta})$-2H	0.07	1376.6891
1360.6943	1.85E+08	$(Z_{5\alpha})$,$(Z_{3\beta})$	0.08	1360.6942
1320.6404	8.52E+07	(C_5)	-0.17	1320.6406
1318.625	1.62E+08	(C_5)-2H	0.02	1318.6250
1302.6298	8.24E+07	(B_5)	-0.20	1302.6301
1286.6572	1.61E+07	$(^{0,4}X_2/Z_{5\alpha})$,$(^{0,4}X_{3\alpha}/Z_{5\alpha})$,$(^{0,4}X_{3\beta}/Z_{5\alpha})$,$(^{0,4}X_1/Z_{3\beta})$,$(^{0,4}X_{3\alpha}/Z_{3\beta})$, $(^{0,4}X_{4\alpha}/Z_{3\beta})$,$(^{0,4}X_{5\alpha}/Z_{3\beta})$	-0.16	1286.6574
1274.6356	2.87E+07	$(^{1,5}A_5)$	0.36	1274.6351
1246.603	9.51E+06	$(^{0,4}X_{5\alpha}/C_5)$,$(^{0,4}X_{4\alpha}/C_5)$,$(^{0,4}X_{3\alpha}/C_5)$,$(^{0,4}X_1/C_5)$,$(^{0,4}X_{3\beta}/C_5)$	-0.68	1246.6038
1214.5779	8.01E+06	$(^{1,3}X_{5\alpha}/B_5)$,$(^{2,4}X_{5\alpha}/B_5)$,$(^{2,4}X_{4\alpha}/B_5)$,$(^{2,4}X_{3\alpha}/B_5)$,$(^{1,3}X_{3\beta}/B_5)$,$(^{2,4}X_{3\beta}/B_5)$	0.22	1214.5776
1202.5999	2.34E+08	$(^{1,5}X_{4\alpha})$	0.01	1202.5999
1188.5852	1.17E+07	$(Y_{3\beta}/^{1,5}X_{5\alpha})$,$(^{1,5}X_{3\beta}/Y_{5\alpha})$	0.80	1188.5842
1172.5894	1.11E+08	$(Y_{4\alpha})$-2H	0.06	1172.5893
1158.574	3.64E+07	$(Y_{3\beta}/Y_{5\alpha})$-2H	0.28	1158.5737
1145.5564	5.65E+07	$(^{3,5}A_5)$	0.20	1145.5562
1142.5789	8.35E+07	$(Y_{3\beta}/Z_{5\alpha})$,$(Z_{3\beta}/Y_{5\alpha})$	0.12	1142.5788
1140.5629	1.23E+07	$(Y_{3\beta}/Z_{5\alpha})$-2H,$(Z_{3\beta}/Y_{5\alpha})$-2H	-0.19	1140.5631
1130.5196	8.02E+06	$(^{1,5}X_{5\alpha}/C_5)$,$(^{1,5}X_{3\beta}/C_5)$	-0.46	1130.5201
1126.5838	4.68E+07	$(Z_{3\beta}/Z_{5\alpha})$+2H	-0.04	1126.5838
1112.5102	1.41E+07	$(^{1,5}X_{5\alpha}/B_5)$,$(^{1,5}X_{3\beta}/B_5)$	0.58	1112.5096
1100.5089	7.06E+06	$(C_5/Y_{5\alpha})$-2H,$(C_5/Y_{3\beta})$-2H	-0.59	1100.5096
1082.5576	1.29E+07	$(^{0,4}X_{3\beta}/Z_{4\alpha})$,$(^{0,4}X_2/Z_{5\alpha})$,$(^{0,4}X_1/Z_{4\alpha})$	-0.03	1082.5576
1073.4986	6.61E+08	(C_4)-2H	-0.05	1073.4987
1068.5418	8.52E+06	$(^{1,3}X_{4\alpha}/Z_{3\beta})$,$(^{1,3}X_{3\beta}/Z_{4\alpha})$,$(^{2,4}X_{3\beta}/Z_{4\alpha})$	-0.17	1068.5420

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_5$ Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
1054.5266	4.63E+07	$(^{3,5}X_{3\beta}/Z_{4\alpha}),(^{3,5}X_2/Z_{5\alpha})$	0.25	1054.5263
1029.5089	3.31E+07	$(^{1,5}A_4)$	0.07	1029.5088
1024.5152	6.27E+06	$(^{1,4}X_{3\beta}/Z_{4\alpha}),(^{1,4}X_{4\alpha}/Z_{3\beta}),(^{0,4}X_2/B_5)$	-0.56	1024.5158
1010.4782	7.70E+06	$(^{1,3}A_{2\alpha}/B_5)$	0.34	1010.4779
998.5002	2.15E+08	$(^{1,5}X_{3\alpha})$	0.08	998.5001
968.4895	1.23E+08	$(Y_{3\alpha})$-2H	-0.06	968.4896
966.4734	1.34E+07	$(^{1,4}X_{4\alpha}/B_5)$	-0.52	966.4739
955.4349	9.15E+06	$(^{3,5}X_{3\beta}/B_4),(^{3,5}X_{3\alpha}/B_4),(^{3,5}X_{4\alpha}/B_4),(^{3,5}X_{5\alpha}/B_4)$	-0.80	955.4357
954.4743	2.13E+07	$(Y_{4\alpha}/Y_{3\beta})$-2H	0.41	954.4739
952.4947	1.31E+08	$(Z_{3\alpha})$	0.06	952.4946
938.479	6.72E+07	$(Z_{4\alpha}/Y_{3\beta}),(Y_{4\alpha}/Z_{3\beta})$	0.01	938.4790
936.4631	1.23E+07	$(Z_{4\alpha}/Y_{3\beta})$-2H,$(Y_{4\alpha}/Z_{3\beta})$-2H	-0.26	936.4633
926.4203	4.45E+06	$(^{1,5}X_{4\alpha}/C_4)$	-0.05	926.4203
925.4257	4.09E+06	$(^{1,4}X_{5\alpha}/B_4),(^{1,4}X_{3\beta}/B_4),(^{0,3}A_{1\alpha}/Z_{3\beta}),(^{0,3}A_{1\beta}/Z_{5\alpha})$	0.65	925.4251
922.4842	1.01E+08	$(Z_{4\alpha}/Z_{3\beta})$+2H	0.13	922.4841
911.4093	4.33E+06	$(^{2,5}X_{5\alpha}/B_4),(^{2,5}X_{4\alpha}/B_4),(^{2,5}X_{3\alpha}/B_4),(^{2,5}X_{3\beta}/B_4)$	-0.16	911.4094
908.41	1.50E+07	$(^{1,5}X_{4\alpha}/B_5)$	0.24	908.4098
896.4677	4.23E+06	$(^{1,3}X_2),(^{2,4}X_2)$	-0.81	896.4684
896.4106	8.79E+06	$(C_5/Y_{4\alpha})$-2H	0.92	896.4098
894.4527	9.76E+06	$(^{1,3}X_2)$-2H,$(^{2,4}X_2)$-2H	-0.08	894.4528
894.3939	7.17E+06	$(C_5/Y_{4\alpha})$-4H	-0.26	894.3941
885.3943	1.23E+07	$(^{1,5}X_{5\alpha}/C_4),(^{1,5}X_{3\beta}/C_4)$	0.57	885.3938
882.3935	9.48E+06	$(C_5/Y_{5\alpha}/Y_{3\beta})$-2H	-0.71	882.3941
880.4149	2.09E+07	$(Y_{4\alpha}/B_5),(Z_{4\alpha}/C_5)$	0.04	880.4149
878.4579	1.65E+07	$(^{0,4}X_{3\beta}/Z_{3\alpha}),(^{0,4}X_2/Z_{4\alpha}),(^{0,4}X_1/Z_{3\alpha})$	0.04	878.4579
866.4	3.35E+06	$(Y_{3\beta}/Z_{5\alpha}/C_5),(Z_{3\beta}/Y_{5\alpha}/C_5),(Y_{3\beta}/Y_{5\alpha}/B_5)$	0.91	866.3992
864.4429	7.50E+06	$(^{1,3}X_{3\alpha}/Z_{3\beta}),(^{1,3}X_{3\beta}/Z_{3\alpha}),(^{2,4}X_{3\beta}/Z_{3\alpha})$	0.80	864.4422
864.3836	5.13E+06	$(Y_{3\beta}/Z_{5\alpha}/C_5)$-2H,$(Z_{3\beta}/Y_{5\alpha}/C_5)$-2H,$(Y_{3\beta}/Y_{5\alpha}/B_5)$-2H	0.04	864.3836
855.3832	1.15E+07	$(C_4/Y_{5\alpha})$-2H,$(C_4/Y_{3\beta})$-2H	-0.04	855.3832
853.3673	1.53E+07	$(C_4/Y_{5\alpha})$-4H,$(C_4/Y_{3\beta})$-4H	-0.33	853.3676
850.4268	9.97E+06	$(^{3,5}X_2/Z_{4\alpha}),(^{3,5}X_{3\beta}/Z_{3\alpha})$	0.28	850.4266
839.3883	2.36E+07	$(C_4/Z_{5\alpha}),(B_4/Y_{5\alpha}),(C_4/Z_{3\beta}),(B_4/Y_{3\beta})$	-0.02	839.3883
837.3727	2.30E+07	$(C_4/Z_{5\alpha})$-2H,$(B_4/Y_{5\alpha})$-2H,$(C_4/Z_{3\beta})$-2H,$(B_4/Y_{3\beta})$-2H	0.04	837.3727
834.431	3.14E+06	$(Z_{4\alpha}/^{0,3}X_2),(Z_{3\alpha}/^{0,3}X_{3\beta}),$	-0.77	834.4316
820.4162	9.39E+06	$(Z_{3\alpha}/^{1,4}X_{3\beta}),(Z_{3\beta}/^{1,4}X_{3\alpha})$	0.28	820.4160
806.4006	3.48E+06	$(^{2,5}X_{3\beta}/Z_{3\alpha})$	0.31	806.4003
762.3743	7.86E+06	$(^{1,4}X_{3\alpha}/B_5)$	0.22	762.3741
750.3742	1.38E+07	$(Y_{3\alpha}/Y_{3\beta})$-2H	0.09	750.3741

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_5$, Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
734.3794	1.95E+07	$(Z_{3\alpha}/Y_{3\beta}),(Y_{3\alpha}/Z_{3\beta})$	0.25	734.3792
732.3635	5.51E+06	$(Z_{3\alpha}/Y_{3\beta})$-2H,$(Y_{3\alpha}/Z_{3\beta})$-2H	-0.09	732.3636
723.3409	3.46E+07	$(^{1,3}A_{4\alpha}),(^{1,3}A_{4\alpha})$	-0.10	723.3410
718.3843	1.62E+07	$(Z_{3\alpha}/Z_{3\beta})$+2H	0.00	718.3843
704.3098	1.83E+07	$(^{1,5}X_{3\alpha}/B_5)$	-0.29	704.3100
690.2944	8.42E+06	$(C_5/Y_{3\alpha})$-4H	0.06	690.2944
681.2939	9.89E+06	$(^{1,5}X_{4\alpha}/C_4)$	-0.18	681.2940
678.2939	8.38E+06	$(C_5/Y_{4\alpha}/Y_{3\beta})$-2H	-0.67	678.2944
676.3152	2.30E+07	$(C_5/Z_{3\alpha}),(B_5/Y_{3\alpha})$	0.16	676.3151
674.2996	8.29E+06	$(C_5/Z_{3\alpha})$-2H,$(B_5/Y_{3\alpha})$-2H	0.24	674.2994
667.3151	1.40E+07	$(C_{3\alpha})$	0.51	667.3148
665.2992	1.13E+08	$(C_{3\alpha})$-2H	0.14	665.2991
662.2996	1.78E+07	$(C_5/Z_{4\alpha}/Y_{3\beta}),(C_5/Y_{4\alpha}/Z_{3\beta}),(B_5/Y_{4\alpha}/Y_{3\beta})$	0.24	662.2994
660.2833	6.86E+06	$(C_5/Z_{4\alpha}/Y_{3\beta})$-2H,$(C_5/Y_{4\alpha}/Z_{3\beta})$-2H,$(B_5/Y_{4\alpha}/Y_{3\beta})$-2H	-0.74	660.2838
658.2679	3.71E+06	$(C_5/Z_{4\alpha}/Y_{3\beta})$-4H,$(C_5/Y_{4\alpha}/Z_{3\beta})$-4H,$(B_5/Y_{4\alpha}/Y_{3\beta})$-4H	-0.37	658.2681
651.2835	2.21E+07	$(Y_{4\alpha}/C_4)$-2H	0.07	651.2835
649.3042	2.70E+07	$(B_{3\alpha})$	0.01	649.3042
649.2679	1.19E+07	$(Y_{4\alpha}/C_4)$-4H	0.14	649.2678
646.3046	2.38E+07	$(C_5/Z_{3\beta}/Z_{4\alpha})$+2H,$(B_5/Z_{4\alpha}/Y_{3\beta})$+2H,$(B_5/Y_{4\alpha}/Z_{3\beta})$+2H	0.11	646.3045
644.2888	8.16E+06	$(C_5/Z_{3\beta}/Z_{4\alpha}),(B_5/Z_{4\alpha}/Y_{3\beta}),(B_5/Y_{4\alpha}/Z_{3\beta})$	-0.12	644.2889
642.2728	3.36E+06	$(C_5/Z_{3\beta}/Z_{4\alpha})$-2H,$(B_5/Z_{4\alpha}/Y_{3\beta})$-2H,$(B_5/Y_{4\alpha}/Z_{3\beta})$-2H	-0.66	642.2732
637.3036	3.57E+06	$(C_4/Z_{4\alpha})$+2H,$(B_4/Y_{4\alpha})$+2H	-0.93	637.3042
635.2886	4.77E+07	$(C_4/Z_{4\alpha}),(B_4/Y_{4\alpha})$	0.09	635.2885
635.2527	5.60E+06	$(Y_{3\beta}/Y_{5\alpha}/C_4)$	0.86	635.2522
633.273	1.53E+07	$(C_4/Z_{4\alpha})$-2H,$(B_4/Y_{4\alpha})$-2H	0.17	633.2729
621.3096	1.21E+07	$(^{1,5}A_{3\alpha})$	0.52	621.3093
621.2727	9.84E+06	$(C_4/Z_{5\alpha}/Y_{3\beta}),(C_4/Y_{5\alpha}/Z_{3\beta}),(B_4/Y_{5\alpha}/Y_{3\beta})$	-0.31	621.2729
619.2936	1.53E+07	$(B_4/Z_{4\alpha})$+2H	-0.04	619.2936
619.2574	1.03E+07	$(C_4/Z_{5\alpha}/Y_{3\beta})$-2H,$(C_4/Y_{5\alpha}/Z_{3\beta})$-2H,$(B_4/Y_{5\alpha}/Y_{3\beta})$-2H	0.26	619.2572
618.3319	1.23E+07	$(^{0,2}X_2)$	0.04	618.3319
617.2784	3.12E+06	$(B_4/Z_{4\alpha})$	0.68	617.2780
617.2416	4.38E+06	$(C_4/Z_{5\alpha}/Y_{3\beta})$-4H,$(C_4/Y_{5\alpha}/Z_{3\beta})$-4H,$(B_4/Y_{5\alpha}/Y_{3\beta})$-4H	0.01	617.2416
616.316	4.31E+06	$(^{1,4}X_2/Z_{3\beta})$	-0.36	616.3162
607.2937	4.28E+06	$(Y_{4\alpha}/^{1,5}A_4)$	0.12	607.2936
605.278	2.66E+07	$(Z_{3\beta}/B_4/Y_{5\alpha})$+2H, $(Y_{3\beta}/B_4/Z_{5\alpha})$+2H, $(Z_{3\beta}/C_4/Z_{5\alpha})$+2H	0.04	605.2780
603.2622	1.37E+07	$(Z_{3\beta}/B_4/Y_{5\alpha})$, $(Y_{3\beta}/B_4/Z_{5\alpha})$, $(Z_{3\beta}/C_4/Z_{5\alpha})$	-0.21	603.2623
602.2787	4.17E+06	$(^{1,3}X_2/B_5),(^{2,4}X_2/B_5)$	0.64	602.2783
598.2833	4.13E+06	$(Z_{3\beta}/Z_{4\alpha}/^{1,5}A_5)$	-0.16	598.2834

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_5$ Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
590.3006	1.70E+08	$(^{1,5}X_2)$	0.05	590.3006
589.2833	1.02E+07	$(Z_{4\alpha}/^{1,5}A_4)$	0.40	589.2831
587.2674	3.91E+06	$(Z_{3\beta}/B_4/Z_{5\alpha})$+2H	-0.02	587.2674
575.2677	4.58E+06	$(B_4/^{0,4}X_{5\alpha}),(B_4/^{0,4}X_{4\alpha})$	0.50	575.2674
562.3058	1.58E+07	(Y_2)	0.25	562.3057
561.2517	3.23E+06	$(B_{3\alpha}/^{1,3}X_{5\alpha}),(B_{3\alpha}/^{2,4}X_{5\alpha}),(B_{3\alpha}/^{2,4}X_{4\alpha}),(^{1,3}X_{3\alpha}/B_4),(Z_{4\alpha}/^{0,2}A_4)$	-0.11	561.2518
560.29	2.90E+07	(Y_2)-2H	-0.01	560.2900
547.2362	7.49E+06	$(B_{3\alpha}/^{3,5}X_{5\alpha}),(B_{3\alpha}/^{3,5}X_{4\alpha})$	0.16	547.2361
546.3107	3.37E+07	(Z_2)+2H	-0.08	546.3107
545.2568	4.72E+06	$(Z_{3\alpha}/^{2,5}A_5),(Z_{4\alpha}/^{2,5}A_4)$	-0.09	545.2568
544.2951	1.49E+08	(Z_2)	0.01	544.2951
519.2414	1.70E+07	$(^{1,3}A_{3\alpha})$	0.39	519.2412
505.2257	4.98E+06	$(C_4/^{0,2}X_{5\alpha}),(C_4/^{0,2}X_{4\alpha}),(Y_{5\alpha}/^{1,3}A_4),(Y_{5\alpha}/^{2,4}A_4),(Y_{3\alpha}/^{2,4}A_5)$	0.30	505.2255
503.2097	5.11E+06	$(B_{3\alpha}/^{2,5}X_{4\alpha}),(B_{3\alpha}/^{2,5}X_{5\alpha})$	-0.39	503.2099
501.2306	3.00E+06	$(Z_{3\alpha}/^{3,5}A_5)$	-0.07	501.2306
486.1944	3.93E+06	$(^{1,5}X_{3\beta}/Z_{3\alpha}/C_5),(^{1,5}X_{3\alpha}/Z_{3\beta}/C_5),(Y_{3\alpha}/^{1,5}X_{3\beta}/B_5),(Y_{3\beta}/^{1,5}X_{3\alpha}/B_5)$	-0.37	486.1946
477.1944	8.69E+06	$(^{1,5}X_{3\alpha}/C_4),(^{1,5}X_{5\alpha}/C_{3\alpha})$	0.32	477.1942
470.2584	1.68E+07	$(^{0,4}X_1/Z_2)$	0.18	470.2583
463.2152	4.88E+06	$(C_{2\alpha})$	0.47	463.2150
461.1993	2.39E+07	$(C_{2\alpha})$-2H	-0.07	461.1993
458.2587	4.78E+06	$(^{0,3}X_1)$	0.84	458.2583
456.1844	2.89E+06	$(Y_{3\alpha}/Y_{3\beta}/B_5)$-2H,$(Y_{3\alpha}/Z_{3\beta}/C_5)$-2H,$(Z_{3\alpha}/Y_{3\beta}/C_5)$-2H	0.84	456.1840
449.1995	3.90E+06	$(Y_{3\alpha}/C_4),(C_{3\alpha}/Y_{5\alpha})$	0.37	449.1993
447.2198	3.58E+06	$(B_{2\alpha}$+2H)	-0.60	447.2201
447.1837	1.63E+07	$(Y_{3\alpha}/C_4)$-2H,$(C_{3\alpha}/Y_{5\alpha})$-2H	0.04	447.1837
445.2044	6.25E+06	$(B_{2\alpha})$	-0.04	445.2044
445.1684	9.48E+06	$(Y_{3\alpha}/C_4)$-4H,$(C_{3\alpha}/Y_{5\alpha})$-4H	0.82	445.1680
440.1895	3.70E+06	$(Y_{3\alpha}/Z_{3\beta}/B_5),(Y_{3\beta}/Z_{3\alpha}/B_5),(Z_{3\alpha}/Z_{3\beta}/C_5)$	0.90	440.1891
431.1888	7.80E+07	$(Y_{3\alpha}/B_4),(Z_{3\alpha}/C_4),(B_{3\alpha}/Y_{5\alpha}),(C_{3\alpha}/Z_{5\alpha})$	0.07	431.1888
424.1938	3.12E+06	$(Z_{3\alpha}/Z_{3\beta}/B_5)$	-0.91	424.1942
417.1732	4.04E+07	$(Y_{3\beta}/Y_{4\alpha}/B_4),(Z_{3\beta}/Y_{4\alpha}/C_4),(Z_{4\alpha}/Y_{3\beta}/C_4)$	0.20	417.1731
415.1939	1.24E+08	$(Z_{3\alpha}/B_4)$+2H,$(B_{3\alpha}/Z_{5\alpha})$+2H	0.11	415.1939
414.232	4.05E+06	$(^{0,2}X_1)$	-0.24	414.2321
403.1939	9.99E+06	$(Y_{3\alpha}/^{1,5}A_4)$, $(^{1,5}A_{3\alpha}/Y_{5\alpha})$	0.11	403.1939
401.1782	5.98E+07	$(Y_{3\alpha}/Z_{3\beta}/B_4)$+2H,$(Y_{3\beta}/Z_{4\alpha}/B_4)$+2H,$(Z_{3\beta}/Z_{4\alpha}/C_4)$	-0.01	401.1782
399.1626	1.12E+07	$(Y_{3\alpha}/Z_{3\beta}/B_4),(Y_{3\beta}/Z_{4\alpha}/B_4),(Z_{3\beta}/Z_{4\alpha}/C_4)$	0.12	399.1626
385.1833	9.50E+06	$(^{1,5}A_{3\alpha}/Z_{5\alpha})$, $(Z_{3\alpha}/^{1,5}A_4)$	0.03	385.1833
383.1677	5.37E+06	$(Z_{3\beta}/Z_{4\alpha}/B_4)$	0.16	383.1676

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_5$, Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
371.1676	8.88E+06	$(B_{2\alpha}/^{0,4}X_{5\alpha})$	-0.10	371.1676
357.1518	3.48E+06	$(^{1,4}A_{2\alpha})$-2H	-0.53	357.1520
345.1743	2.78E+07	$(^{1,5}X_1)$	0.29	345.1742
341.1572	4.97E+06	$(Z_{3\alpha}/^{2,5}A_5)$	0.29	341.1571
329.1571	8.74E+06	$(^{3,5}A_4)$	0.00	329.1571
317.1794	5.39E+07	(Y_1)	0.20	317.1793
315.1637	1.18E+07	(Y_1)-2H	0.05	315.1637
301.1258	5.54E+06	$(^{0,4}A_{2\alpha})$	0.09	301.1258
299.1688	1.27E+07	(Z_1)	0.10	299.1688
296.1103	5.37E+06	$(^{1,5}X_2/B_5)$	-0.53	296.1105
259.1153	6.62E+06	$(C_{1\alpha}),(C_{1\beta})$	0.35	259.1152
257.0996	4.80E+06	$(C_{1\alpha})$-2H,$(C_{1\beta})$-2H	0.16	257.0996
243.0839	4.83E+06	$(Y_{5\alpha}/C_{2\alpha})$-2H,$(Y_{4\alpha}/C_{3\alpha})$-2H	-0.04	243.0839
227.089	1.25E+07	$(B_{2\alpha}/Y_{5\alpha}),(C_{2\alpha}/Z_{5\alpha}),(B_{3\alpha}/Y_{4\alpha}),(C_{3\alpha}/Z_{4\alpha})$	0.02	227.0890

APPENDIX 2A.2 Assigned Ions for $HexNAc_2Hex_6$.

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_6$, Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
1800.9192	8.23E+09	$Hex_6HexNAc_2$-(^{2}H)-reduced (un-fragmented precursor)	-0.42	1800.9199
1726.8826	3.57E+08	$(^{0,4}X_1),(^{0,4}X_{3\alpha}),(^{0,4}X_{3\beta}),(^{0,4}X_{4\alpha}),(^{0,4}X_{5\alpha}),(^{0,4}X_{6\alpha})$	-0.33	1726.8832
1610.7991	2.26E+08	$(^{1,5}X_{3\beta}),(^{1,5}X_{6\alpha})$	-0.21	1610.7994
1580.7884	9.58E+07	$(Y_{6\alpha})$-2H,$(Y_{3\beta})$-2H	-0.30	1580.7889
1564.7938	8.14E+07	$(Z_{6\alpha}),(Z_{3\beta})$	-0.10	1564.7940
1522.7245	1.45E+08	(C_6)-2H	-0.16	1522.7248
1520.7674	5.63E+07	$(^{0,4}X_2)$-2H	-0.23	1520.7677
1478.7342	1.83E+07	$(^{1,5}X_2)$	-0.49	1478.7349
1450.703	9.54E+06	$(^{0,4}X_1/C_6),(^{0,4}X_{3\alpha}/C_6),(^{0,4}X_{3\beta}/C_6),(^{0,4}X_{4\alpha}/C_6),(^{0,4}X_{5\alpha}/C_6),$ $(^{0,4}X_{6\alpha}/C_6)$	-0.43	1450.7036
1418.6769	6.65E+06	$(^{1,3}X_{6\alpha}/B_6),(^{2,4}X_{6\alpha}/B_6),(^{1,3}X_{3\beta}/B_6),(^{2,4}X_{3\beta}/B_6),(^{2,4}X_{3\alpha}/B_6),$ $(^{2,4}X_{4\alpha}/B_6)$	-0.36	1418.6774
1406.6997	1.30E+08	$(^{1,5}X_{5\alpha})$	0.02	1406.6997
1376.6891	1.46E+08	$(Y_{5\alpha})$-2H	0.00	1376.6891
1374.6732	8.08E+06	$(Z_{6\alpha}/^{1,5}X_{3\beta}),(Z_{3\beta}/^{1,5}X_{6\alpha})$	-0.18	1374.6735
1362.6745	1.42E+07	$(Y_{6\alpha}/Y_{3\beta})$-2H	0.77	1362.6735
1360.6943	1.39E+08	$(Z_{5\alpha})$	0.08	1360.6942
1349.6564	3.42E+07	$(^{3,5}A_6)$	0.34	1349.6559
1346.679	5.45E+07	$(Y_{6\alpha}/Z_{3\beta}),(Z_{6\alpha}/Y_{3\beta})$	0.34	1346.6785
1330.6837	6.22E+07	$(Z_{6\alpha}/Z_{3\beta})$+2H	0.06	1330.6836

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_6$, Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
1304.609	4.98E+06	$(C_6/Y_{3\beta})$-2H,$(C_6/Y_{6\alpha})$-2H	-0.25	1304.6093
1302.5936	4.13E+06	$(C_6/Y_{3\beta})$-2H,$(C_6/Y_{6\alpha})$-4H	-0.06	1302.5937
1277.5984	6.23E+08	(C_5)-2H	-0.02	1277.5984
1272.6418	6.41E+06	$(^{1,3}X_{5\alpha}/Z_{3\beta})$,$(^{2,4}X_{5\alpha}/Z_{3\beta})$,$(^{1,3}X_{3\beta}/Z_{5\alpha})$,$(^{2,4}X_{3\beta}/Z_{5\alpha})$	0.03	1272.6418
1233.6088	4.63E+07	$(^{1,5}A_5)$	0.16	1233.6086
1230.5726	3.39E+06	$(^{1,3}5A/C_6)$-2H	0.04	1230.5725
1214.5779	4.96E+06	$(^{1,3}X_{5\alpha}/B_6)$,$(^{2,4}X_{5\alpha}/B_6)$	0.22	1214.5776
1202.5999	2.03E+08	$(^{1,5}X_{4\alpha})$	0.01	1202.5999
1188.5615	3.77E+06	$(^{1,4}X_{5\alpha}/C_6)$	-0.41	1188.5620
1172.5895	1.36E+08	$(Y_{4\alpha})$-2H	0.15	1172.5893
1170.5739	1.02E+07	$(^{1,5}X_{5\alpha}/Z_{3\beta})$,$(^{1,5}X_{3\beta}/Z_{5\alpha})$	0.19	1170.5737
1158.5741	4.12E+07	$(Y_{5\alpha}/Y_{3\beta})$-2H	0.36	1158.5737
1156.5946	9.85E+07	$(Z_{4\alpha})$	0.16	1156.5944
1142.579	7.97E+07	$(Y_{3\beta}/Z_{5\alpha})$,$(Z_{3\beta}/Y_{5\alpha})$	0.21	1142.5788
1126.5844	3.74E+07	$(Z_{3\beta}/Z_{5\alpha})$+2H	0.49	1126.5838
1100.51	4.80E+06	$(C_6/Y_{5\alpha})$-2H	0.41	1100.5096
1082.558	1.09E+07	$(^{0,4}X_{3\beta}/Z_{4\alpha})$,$(^{0,4}X_2/Z_{5\alpha})$,$(^{0,4}X_1/Z_{4\alpha})$	0.34	1082.5576
1082.4984	6.88E+06	$(B_6/Y_{5\alpha})$-2H,$(C_6/Z_{5\alpha})$-2H	-0.54	1082.4990
1068.5426	6.88E+06	$(^{1,3}X_{4\alpha}/Z_{3\beta})$,$(^{1,3}X_{3\beta}/Z_{4\alpha})$,$(^{2,4}X_{3\beta}/Z_{4\alpha})$	0.58	1068.5420
1057.4681	1.28E+07	$(C_5/Y_{3\beta})$-4H,$(C_5/Y_{6\alpha})$-4H	0.71	1057.4674
1054.5268	4.11E+07	$(^{3,5}X_{3\beta}/Z_{4\alpha})$,$(^{3,5}X_2/Z_{5\alpha})$	0.44	1054.5263
1050.4726	3.41E+06	$(C_6/Z_{3\beta}/Z_{6\alpha})$-2H,$(B_6/Z_{6\alpha}/Y_{3\beta})$-2H,$(B_6/Y_{6\alpha}/Z_{3\beta})$-2H	-0.17	1050.4728
1043.4882	1.76E+07	$(C_5/Z_{3\beta})$,$(C_5/Z_{6\alpha})$,$(B_5/Y_{3\beta})$,$(B_5/Y_{6\alpha})$	0.11	1043.4881
1010.4778	5.29E+06	$(^{1,3}X_{4\alpha}/B_6)$	-0.06	1010.4779
998.5003	2.28E+08	$(^{1,5}X_{3\alpha})$	0.18	998.5001
980.4894	4.43E+06	$(^{1,4}A_6/Z_{5\alpha})$	-0.16	980.4896
968.4896	1.27E+08	$(Y_{3\alpha})$-2H	0.05	968.4896
966.4733	1.05E+07	$(^{1,4}X_{4\alpha}/B_6)$	-0.63	966.4739
954.4742	2.07E+07	$(Y_{4\alpha}/Y_{3\beta})$-2H	0.31	954.4739
952.4948	1.15E+08	$(Z_{3\alpha})$	0.17	952.4946
938.4791	6.12E+07	$(Z_{4\alpha}/Y_{3\beta})$,$(Y_{4\alpha}/Z_{3\beta})$	0.12	938.4790
936.4638	9.12E+06	$(Z_{4\alpha}/Y_{3\beta})$-2H,$(Y_{4\alpha}/Z_{3\beta})$-2H	0.49	936.4633
927.441	4.55E+07	$(^{1,3}A_5)$,$(^{2,4}A_5)$	0.27	927.4407
922.4843	8.57E+07	$(Z_{4\alpha}/Z_{3\beta})$+2H	0.24	922.4841
920.4677	5.39E+06	$(Z_{4\alpha}/Z_{3\beta})$	-0.79	920.4684
908.4101	1.27E+07	$(^{1,5}X_{4\alpha}/B_6)$	0.35	908.4098
896.4689	4.94E+06	$(^{1,3}X_2)$,$(^{2,4}X_2)$	0.53	896.4684
896.4098	1.83E+07	$(C_6/Y_{4\alpha})$-2H	0.02	896.4098

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_6$, Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
894.3942	7.97E+06	$(C_6/Y_{4\alpha})$-4H	0.08	894.3941
885.3937	6.27E+06	$(^{1,5}X_{5\alpha}/C_5),(^{1,5}X_{3\beta}/C_5)$	-0.11	885.3938
882.3943	6.67E+06	$(C_6/Y_{5\alpha}/Y_{3\beta})$-2H	0.19	882.3941
880.4151	1.94E+07	$(Y_{4\alpha}/B_6),(Z_{4\alpha}/C_6)$	0.27	880.4149
878.4579	1.38E+07	$(^{0,4}X_{3\beta}/Z_{3\alpha}),(^{0,4}X_2/Z_{4\alpha}),(^{0,4}X_1/Z_{3\alpha})$	0.04	878.4579
878.3999	4.39E+06	$(B_6/Y_{4\alpha})$-2H,$(C_6/Z_{4\alpha})$-2H	0.78	878.3992
869.399	2.99E+08	$(C_{4\alpha})$-2H	0.14	869.3989
866.399	3.28E+06	$(Y_{3\beta}/Z_{5\alpha}/C_6),(Z_{3\beta}/Y_{5\alpha}/C_6),(Y_{3\beta}/Y_{5\alpha}/B_6)$	-0.25	866.3992
864.442	7.05E+06	$(^{1,3}X_{3\alpha}/Z_{3\beta}),(^{1,3}X_{3\beta}/Z_{3\alpha}),(^{2,4}X_{3\beta}/Z_{3\alpha})$	-0.24	864.4422
864.3842	6.19E+06	$(C_6/Z_{5\alpha}/Y_{3\beta})$-2H,$(C_6/Y_{5\alpha}/Z_{3\beta})$-2H,$(B_6/Y_{5\alpha}/Y_{3\beta})$-2H	0.74	864.3836
853.4042	1.92E+07	$(B_{4\alpha})$	0.27	853.4040
853.3679	1.33E+07	$(C_5/Y_{5\alpha})$-4H,$(C_5/Y_{3\beta})$-4H	0.37	853.3676
850.4261	7.69E+06	$(^{3,5}X_2/Z_{4\alpha}),(^{3,5}X_{3\beta}/Z_{3\alpha})$	-0.54	850.4266
848.3893	3.68E+06	$(C_6/Z_{3\beta}/Z_{5\alpha}),(B_6/Z_{5\alpha}/Y_{3\beta}),(B_6/Y_{5\alpha}/Z_{3\beta})$	0.77	848.3886
839.3884	8.13E+07	$(Y_{5\alpha}/B_5),(Z_{5\alpha}/C_5)$	0.10	839.3883
839.352	4.71E+06	$(Y_{3\beta}/Y_{6\alpha}/C_5)$-4H	0.08	839.3519
838.4038	3.11E+06	$(Y_{3\beta}/Y_{5\alpha}/^{1,5}A_{1\alpha})$	-0.60	838.4043
837.3728	2.34E+07	$(C_5/Z_{5\alpha})$-2H,$(B_5/Y_{5\alpha})$-2H	0.16	837.3727
825.4092	3.21E+07	$(^{1,5}A_{4\alpha})$	0.18	825.4091
825.3733	1.45E+07	$(C_5/Z_{6\alpha}/Y_{3\beta}),(C_5/Y_{6\alpha}/Z_{3\beta}),(B_5/Y_{6\alpha}/Y_{3\beta})$	0.77	825.3727
823.3572	1.15E+07	$(C_5/Z_{6\alpha}/Y_{3\beta})$-2H,$(C_5/Y_{6\alpha}/Z_{3\beta})$-2H,$(B_5/Y_{6\alpha}/Y_{3\beta})$-2H	0.22	823.3570
811.3931	6.22E+06	$(Y_{5\alpha}/^{1,5}A_5)$	-0.37	811.3934
807.3621	1.67E+07	$(B_5/Z_{6\alpha}/Y_{3\beta}),(B_5/Y_{6\alpha}/Z_{3\beta}),(C_5/Z_{3\beta}/Z_{5\alpha})$	0.00	807.3621
793.3822	8.96E+06	$(Z_{5\alpha}/^{1,5}A_5)$	-0.80	793.3828
765.3508	4.14E+06	$(Z_{4\alpha}/^{0,2}A_6),(^{1,3}X_{4\alpha}/B_5),(Z_{5\alpha}/^{0,2}A_5),(^{2,4}X_{6\alpha}/B_{4\alpha}),(^{1,3}X_{6\alpha}/B_{4\alpha}),$ $(^{2,4}X_{4\alpha}/B_{4\alpha})$	-0.96	765.3515
762.3516	3.79E+06	$(^{1,4}X_{3\alpha}/B_6)$	-0.35	762.3519
751.3363	7.21E+06	$(^{3,5}X_{6\alpha}/B_{4\alpha}),(^{3,5}X_{5\alpha}/B_{4\alpha}),(^{3,5}X_{4\alpha}/B_{4\alpha})$	0.55	751.3359
750.3743	9.17E+06	$(Y_{3\alpha}/Y_{3\beta})$-2H	0.22	750.3741
734.3793	1.47E+07	$(Z_{3\alpha}/Y_{3\beta}),(Y_{3\alpha}/Z_{3\beta})$	0.11	734.3792
732.3635	3.86E+06	$(Z_{3\alpha}/Y_{3\beta})$-2H,$(Y_{3\alpha}/Z_{3\beta})$-2H	-0.09	732.3636
723.341	1.77E+07	$(^{1,3}A_{4\alpha}),(Y_{4\alpha}/^{3,5}A_6)$	0.04	723.3410
722.321	5.05E+06	$(^{1,5}X_{3\alpha}/C_6)$	0.59	722.3206
718.3847	1.33E+07	$(Z_{3\alpha}/Z_{3\beta})$+2H	0.55	718.3843
704.31	1.68E+07	$(^{1,5}X_{3\alpha}/B_6)$	-0.01	704.3100
692.3102	3.34E+06	$(C_6/Y_{3\alpha})$-2H	0.28	692.3100
681.2935	5.28E+06	$(^{1,5}X_{4\alpha}/C_5),(^{1,5}X_{6\alpha}/C_{4\alpha})$	-0.77	681.2940
678.2944	6.35E+06	$(C_6/Y_{4\alpha}/Y_{3\beta})$-2H	0.07	678.2944
676.3153	1.68E+07	$(C_6/Z_{3\alpha}),(B_6/Y_{3\alpha})$	0.31	676.3151

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_6$, Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
674.2996	9.89E+06	$(C_6/Z_{3\alpha})$-2H,$(B_6/Y_{3\alpha})$-2H	0.24	674.2994
667.3153	8.78E+06	$(C_{3\alpha})$	0.81	667.3148
665.2992	7.99E+07	$(C_{3\alpha})$-2H	0.14	665.2991
662.2996	5.76E+06	$(C_6/Z_{4\alpha}/Y_{3\beta})$,$(C_6/Y_{4\alpha}/Z_{3\beta})$,$(B_6/Y_{4\alpha}/Y_{3\beta})$	0.24	662.2994
660.2842	7.16E+06	$(C_6/Z_{4\alpha}/Y_{3\beta})$-2H,$(C_6/Y_{4\alpha}/Z_{3\beta})$-2H,$(B_6/Y_{4\alpha}/Y_{3\beta})$-2H	0.62	660.2838
651.2837	1.63E+07	$(Y_{4\alpha}/C_5)$-2H,$(C_{4\alpha}/Y_{6\alpha})$-2H	0.37	651.2835
649.3037	4.69E+06	$(B_{3\alpha})$	-0.76	649.3042
649.268	1.10E+07	$(Y_{4\alpha}/C_5)$-4H,$(C_{4\alpha}/Y_{6\alpha})$-4H	0.30	649.2678
646.3047	1.00E+07	$(C_6/Z_{3\beta}/Z_{4\alpha})$+2H,$(B_6/Z_{4\alpha}/Y_{3\beta})$+2H,$(B_6/Y_{4\alpha}/Z_{3\beta})$+2H	0.27	646.3045
635.2887	4.62E+07	$(C_{4\alpha}/Z_{6\alpha})$,$(C_5/Z_{4\alpha})$,$(B_{4\alpha}/Y_{6\alpha})$,$(B_5/Y_{4\alpha})$	0.25	635.2885
634.3048	5.01E+06	$(Y_{3\beta}/Y_{4\alpha}{}^{1,5}A_6)$	0.43	634.3045
633.2732	1.17E+07	$(C_{4\alpha}/Z_{6\alpha})$-2H,$(C_5/Z_{4\alpha})$-2H,$(B_{4\alpha}/Y_{6\alpha})$-2H,$(B_5/Y_{4\alpha})$-2H	0.49	633.2729
621.3091	5.38E+06	$({}^{1,5}A_{3\alpha})$	-0.29	621.3093
621.2732	1.42E+07	$(C_5/Z_{5\alpha}/Y_{3\beta})$,$(C_5/Y_{5\alpha}/Z_{3\beta})$,$(B_5/Y_{5\alpha}/Y_{3\beta})$	0.50	621.2729
619.2937	1.80E+08	$(B_{4\alpha}/Z_{6\alpha})$+2H,$(B_5/Z_{4\alpha})$+2H	0.12	619.2936
619.2577	1.01E+07	$(C_5/Z_{5\alpha}/Y_{3\beta})$-2H,$(C_5/Y_{5\alpha}/Z_{3\beta})$-2H,$(B_5/Y_{5\alpha}/Y_{3\beta})$-2H	0.74	619.2572
618.3319	7.14E+06	$({}^{0,2}X_2)$	0.04	618.3319
605.2781	2.28E+07	$(C_5/Z_{3\beta}/Z_{5\alpha})$+2H,$(B_5/Z_{5\alpha}/Y_{3\beta})$+2H,$(B_5/Y_{5\alpha}/Z_{3\beta})$+2H	0.20	605.2780
603.2623	1.31E+07	$(C_5/Z_{3\beta}/Z_{5\alpha})$,$(B_5/Z_{5\alpha}/Y_{3\beta})$,$(B_5/Y_{5\alpha}/Z_{3\beta})$	-0.05	603.2623
590.3007	1.08E+08	$({}^{1,5}X_2)$	0.22	590.3006
589.2829	9.74E+06	$(Z_{6\alpha}/{}^{1,5}A_{4\alpha})$,$(Z_{4\alpha}/{}^{1,5}A_5)$	-0.28	589.2831
587.2676	8.59E+06	$(B_5/Z_{5\alpha}/Z_{3\beta})$+2H	0.32	587.2674
575.2676	8.28E+06	$({}^{0,4}X_{6\alpha}/B_{3\alpha})$,$({}^{0,4}X_{5\alpha}/B_{3\alpha})$	0.33	575.2674
562.3056	1.18E+07	(Y_2)	-0.10	562.3057
560.29	1.74E+07	(Y_2)-2H	-0.01	560.2900
547.2364	5.79E+06	$({}^{3,5}X_{6\alpha}/B_{3\alpha})$,$({}^{3,5}X_{5\alpha}/B_{3\alpha})$	0.53	547.2361
546.311	1.08E+07	(Z_2)+2H	0.47	546.3107
545.2571	7.86E+06	$(Z_{3\alpha}/{}^{2,5}A_6)$,$(Z_{4\alpha}/{}^{2,5}A_5)$	0.46	545.2568
544.2952	1.00E+08	(Z_2)	0.20	544.2951
519.2413	1.15E+07	$({}^{1,3}A_{3\alpha})$	0.20	519.2412
517.2257	4.71E+06	$({}^{1,4}X_{6\alpha}/B_{3\alpha})$,$(Z_{3\alpha}/{}^{0,3}A_6)$,$({}^{1,4}X_{3\alpha}/B_5)$,$({}^{1,4}X_{5\alpha}/B_{4\alpha})$	0.29	517.2255
503.2103	3.98E+06	$({}^{2,5}X_{6\alpha}/B_{3\alpha})$,$({}^{2,5}X_{5\alpha}/B_{4\alpha})$	0.80	503.2099
487.2153	3.55E+06	$({}^{1,3}A_{4\alpha}/Z_{6\alpha})$,$({}^{0,2}X_{6\alpha}/B_{3\alpha})$,$({}^{2,4}A_6/Z_{3\alpha})$,$({}^{1,3}A_5/Z_{5\alpha})$,$({}^{2,4}A_5/Z_{5\alpha})$, $({}^{0,2}X_{5\alpha}/B_{4\alpha})$	0.65	487.2150
477.1943	5.54E+06	$(C_{3\alpha}/{}^{1,5}X_{6\alpha})$,$(C_{4\alpha}/{}^{1,5}X_{5\alpha})$,$({}^{1,5}X_{3\alpha}/C_5)$	0.11	477.1942
470.2582	1.17E+07	$({}^{0,4}X_1/Z_2)$	-0.24	470.2583
463.2149	6.72E+06	$(C_{2\alpha})$	-0.18	463.2150
461.1994	3.28E+07	$(C_{2\alpha})$-2H	0.15	461.1993
449.1991	2.98E+06	$(C_{4\alpha}/Y_{5\alpha})$,$(C_5/Y_{3\alpha})$,$(C_{3\alpha}/Y_{6\alpha})$	-0.52	449.1993

Observed *m/z*	Intensity	14 eV EED $HexNAc_2Hex_6$, Assigned Ions	Accuracy (ppm)	Theoretical *m/z*
447.1837	1.27E+07	$(C_{4\alpha}/Y_{5\alpha})$-2H,$(C_5/Y_{3\alpha})$-2H,$(C_{3\alpha}/Y_{6\alpha})$-2H	0.04	447.1837
445.2044	3.82E+06	$(B_{2\alpha})$	-0.04	445.2044
433.2044	7.86E+06	$(C_{4\alpha}/Z_{5\alpha})$+2H,$(C_5/Z_{3\alpha})$+2H,$(C_{3\alpha}/Z_{6\alpha})$+2H,$(B_{4\alpha}/Y_{5\alpha})$+2H, $(B_5/Y_{3\alpha})$+2H,$(B_{3\alpha}/Y_{6\alpha})$+2H	-0.04	433.2044
431.1888	6.27E+07	$(C_{4\alpha}/Z_{5\alpha})$,$(C_5/Z_{3\alpha})$,$(C_{3\alpha}/Z_{6\alpha})$,$(B_{4\alpha}/Y_{5\alpha})$,$(B_5/Y_{3\alpha})$,$(B_{3\alpha}/Y_{6\alpha})$	0.07	431.1888
429.1732	1.77E+07	$(C_{4\alpha}/Z_{5\alpha})$-2H,$(C_5/Z_{3\alpha})$-2H,$(C_{3\alpha}/Z_{6\alpha})$-2H,$(B_{4\alpha}/Y_{5\alpha})$-2H, $(B_5/Y_{3\alpha})$-2H,$(B_{3\alpha}/Y_{6\alpha})$-2H	0.19	429.1731
417.1732	1.81E+07	$(C_5/Z_{4\alpha}/Y_{3\beta})$,$(C_5/Y_{4\alpha}/Z_{3\beta})$,$(B_5/Y_{4\alpha}/Y_{3\beta})$	0.20	417.1731
403.194	6.46E+06	$(Y_{3\alpha}/{}^{1,5}A_5)$,$({}^{1,5}A_{4\alpha}/Y_{5\alpha})$,$({}^{1,5}A_{3\alpha}/Y_{6\alpha})$	0.36	403.1939
401.1782	2.01E+07	$(Y_{4\alpha}/Z_{3\beta}/B_5)$+2H,$(Z_{4\alpha}/Y_{3\beta}/B_5)$+2H,$(Z_{4\alpha}/Z_{3\beta}/C_5)$+2H	-0.01	401.1782
399.1625	1.39E+07	$(Y_{4\alpha}/Z_{3\beta}/B_5)$,$(Z_{4\alpha}/Y_{3\beta}/B_5)$,$(Z_{4\alpha}/Z_{3\beta}/C_5)$	-0.13	399.1626
385.1832	7.96E+06	$({}^{1,5}A_{4\alpha}/Z_{5\alpha})$,$({}^{1,5}A_{3\alpha}/Z_{6\alpha})$,$({}^{1,5}A_5/Z_{3\alpha})$	-0.23	385.1833
371.1676	4.83E+06	$(B_{2\alpha}/{}^{0,4}X_{6\alpha})$	-0.10	371.1676
345.1742	2.01E+07	$({}^{1,5}X_1)$	-0.15	345.1743
341.1571	3.67E+06	$(Z_{3\alpha}/{}^{2,5}A_5)$	0.08	341.1571
329.157	6.15E+06	$({}^{3,5}A_{5\alpha})$	-0.23	329.1571
317.1793	3.83E+07	(Y_1)	-0.11	317.1793
315.1637	6.53E+06	(Y_1)-2H	0.05	315.1637
301.1257	4.13E+06	$({}^{0,2}A_{2\alpha})$,$(Y_{4\alpha}/{}^{1,3}A_5)$,$(Y_{5\alpha}/{}^{1,3}A_{4\alpha})$,$(Y_{6\alpha}/{}^{1,3}A_{3\alpha})$,$(C_{2\alpha}/{}^{0,2}X_{6\alpha})$, $(C_{3\alpha}/{}^{0,2}X_{5\alpha})$,$({}^{0,2}A_{2\alpha})$,$(Y_{4\alpha}/{}^{2,4}A_5)$	-0.25	301.1258
299.1687	1.05E+07	(Z_1)	-0.24	299.1688
296.1103	3.07E+06	$({}^{1,5}X_2/B_6)$	-0.53	296.1105
259.1151	5.34E+06	$(C_{1\beta})$,$(C_{1\alpha})$	-0.42	259.1152
257.0996	3.62E+06	$(C_{1\beta})$-2H,$(C_{1\alpha})$-2H	0.16	257.0996
227.0889	1.74E+07	$(C_{2\alpha}/Z_{6\alpha})$,$(C_{3\alpha}/Z_{5\alpha})$,$(C_{4\alpha}/Z_{4\alpha})$,$(B_{2\alpha}/Y_{6\alpha})$,$(B_{3\alpha}/Y_{5\alpha})$,$(B_{4\alpha}/Y_{4\alpha})$	-0.42	227.0890

APPENDIX 2A.3 MALDI-TOF-MS Assigned Ions: *C. parvum* Released *N*-Glycans

Deuteroreduced and Permethylated $[M+Na]^+$

	$Hex_5HexNAc_2$-(^{2}H)-reduced	$Hex_6HexNAc_2$-(^{2}H)-reduced
Theoretical (*m/z*)	1596.820	1800.920
observed (*m/z*)	1596.805	1800.906
ppm error	-9.50	-7.75

APPENDIX 2B: Bioinformatics and Related Data for the *N*-Glycosylated Proteins.

APPENDIX 2B.1 The Amino Acid Sequences Used to Generate the Weblogos.

The unoccupied and occupied *N*-glycosylation sequons are shown, with 11 amino acids flanking the centered Asn. For the final figure, nine amino acids were used, with three amino acids flanking both sides of the three amino acid sequon, Nx[T/S] x ≠ P.

Unoccupied Sites	N.T	N.S	Occupied Sites	N.T	N.S
AERETNAYSMLNETHHHPKAYFV	1	0	ASRLTDLLERYNSTCGSQQSIVS	1	0
CVWGEVRLVSSNFTTEKSWEVEG	1	0	CELFIVNSKATNQTNDSWFNLDL	1	0
DGNNQLVNPETNSTVSGSTSGTT	1	0	DILLSDYQNAKNNTVIETSPVDI	1	0
DLQNPVDFVNPNCTGRHALSEGC	1	0	DLGESNDTKKLNETQILSDAYEA	1	0
DWIPCCLPGYGNYTVSLHDSYGD	1	0	ERENISIENSINETSIPNEETNL	1	0
EDAIKLYSDKCNFTERKGQFSIS	1	0	ETSISSDGKRYNDTASPIKTPEI	1	0
EEEEEEGDEAANETVVTIERDSS	1	0	GEHKVSINLSANMTYQLKNLRID	1	0
ELFDPISDEIMNGTIAGIVSGIS	1	0	GNVVKSQKQEKNSTTEVRSLGHR	1	0
ENIGLWNVADSNTTSNVSEDEKI	1	0	HKVSEQAFQNLNATLHYGHKHHD	1	0
EQIEGQREIRGNKTQSTGFPFLR	1	0	KLLEENSGASFNKTILKESYDWI	1	0
ERKMDVMIPMENMTPGFINDLQD	1	0	KPIPGSHSGFINGTSGEQSHEKD	1	0
ERYARVRFGSCNETENMVDPLGL	1	0	KTKEIDDIVPHNETIMKDAGNDS	1	0
ERYISSSLILTNGTSQGKFALSP	1	0	LDRAPGSSNTMNQTTNLNNERYI	1	0
FVAASPQAAASNSTSGALPELVL	1	0	LPKDPVSDIPFNSTTGELVDPST	1	0
GVEIVEFATQSNTTVDENDFKFK	1	0	LQNSNESLMQRNQTSSSGNNPVN	1	0
GVPPLTYLSPANETLLPNNVASI	1	0	LVDPSTGKPINNSTAGIVSGKPG	1	0
IERMGFSVDNDNITLLSDVFMDL	1	0	MMRESFREHVFNVTGQVPTLGEV	1	0
IFNKGDALALNNFTNKFGPPIPV	1	0	MVDPVSLMLFDNSTGVMYDPNTN	1	0
IPNTYAGVYRSNETKTTEPSANT	1	0	NGNANLLLKRNNVTYDSNNDIFP	1	0
IRSESCIVSELNFTSTTGFTTDT	1	0	NKDPVTNTQYSNTTGNIINPETG	1	0
KALIRGSCISVNTTRLDSPSVLV	1	0	NPNSEFSLNRINGTWRVYNLFNV	1	0
KETEPDTKNIENKTKSSVKETED	1	0	PKSGNLVHPYTNQTMSGLSVSYL	1	0
KHFMDYVKKELNATDTEMKSLFG	1	0	PPTGHLINPTNNNTMDSSFAGAY	1	0
KQITKVEGNPSNTTLNEMNGVIS	1	0	RDGHCEKGTIYNITSVDDLIQNS	1	0
KTQIHYYNAVTNTTGSSNQDSTI	1	0	RDKNIKNNPLYNETSISSDGKRY	1	0
LAFANPEALEVNKTLIEEIYNRN	1	0	RDPVSGLPQLPNGTLVDPSNKKP	1	0
LDLLREFLMDENVTSSSTDVFEG	1	0	SFLESGSITETNFTMSTYRNETG	1	0
LEYHLYPENPQNNTKPEERGRNM	1	0	SIIDTQDLGESNDTKKLNETQIL	1	0
LPGMPLSTEYTNSTNLDSSKVLL	1	0	SKGPYRVTLFVNKTEAYEYLDSL	1	0
LVDPSTGKPINNSTAGIVSGKPG	1	0	SPIKTPEIVYYNNTSNLRVLAKK	1	0

Unoccupied Sites	N.T	N.S
LVISGQNRGTYNSTIEGTTFLNC	1	0
MDKVCDLVQQWNNTSSFLPRQTD	1	0
MRNPIHTFFGINFTGLLVVGAPI	1	0
NKGNENVNITSNYTMPQAKSLGM	1	0
NKTLIEEIYNRNTTTDGFNRMLM	1	0
NLIIPAQKKLLNTTSPEATKETN	1	0
NNVSNSSQSQYNGTNINSLNSTE	1	0
NQYSDFEYPFFNITAGKLANEFG	1	0
NTTSPEATKETNKTYDNNSDAKN	1	0
NYCSYCCENSANRTTKEQLKCKQ	1	0
PFITTESLQEMNITIVEKLHASS	1	0
PFRVDSNKDVENSTDKEKTDSIY	1	0
PLTYVTMKPYANCTSFLNPLGEE	1	0
RGGSGETLYLLNETSSIPVDIDD	1	0
RIDDYTLVDINNMTVGQMLSLHT	1	0
RLNYLPRVLANNFTWLIPMRDHY	1	0
SDAKNRINQSNNATDKRSEKQTD	1	0
SEDQFSEFLGINGTFPPFVKWPY	1	0
SENIDTSDSDSNSTSDSNSSSNS	1	0
SGLSVSYLAAKNLTVDTDETYGL	1	0
SIDNDFEKINRNVTLELLNGENT	1	0
SKTLPTSFYWPNETEKWAKISFL	1	0
SNRISFEKVLANLTKRVTTKDRV	1	0
SQYNGTNINSLNSTESNPNYIKD	1	0
SSLGFEDITQANKTSISGNSAFV	1	0
TKEETKTQDGKNLTTTTSEFNLD	1	0
TNDRDSGLKIWNLTFLGRRVAFE	1	0
VELENKGNENVNITSNYTMPQAK	1	0
VKLGPIIYQPHNVTSENEEEQER	1	0
VPPYAHIHKAFNLTEPQRQPCGR	1	0
WTDTISKSSIINMTNAYEKIVKG	1	0
WYSEGETFDFYNVTNNYGKDFYN	1	0
----MVKFTLKNTTVIILILACL	1	0
YKSGTPIDRFANATNFVNDGLFN	1	0
YSETVMDSIEGNITANIERMGFS	1	0
ALNKLPTSNEVNISPRPSSDAVP	0	1
ASVNRRLRENTNSSQKASKSSLF	0	1
DKVCDLVQQWNNTSSFLPRQTDF	0	1
DNDNYYQVIVSNTSTTLVKKTKA	0	1

Occupied Sites	N.T	N.S
TETNFTMSTYRNETGLLTNPKEL	1	0
TSQGKFALSPLNGTEVAPLFSKF	1	0
VAGDEILTEVLNITTDEVTGLPI	1	0
WFSAEEVYERNNGTIPITPRVMD	1	0
INGTWRVYNLFNVSDHGFRMSSD	0	1
PLPGPYFRYLGNESESNERKMDV	0	1
SWDEPIDISKANVSTIFGDLLNS	0	1
TNLDSSKVLLGNDSTVKSSSFFN	0	1
FETIHDVVKSLNETQSGVDLEQR	1	0

Total Occupied:	N.T	**N.S**
	35	4

Unoccupied Sites	N.T	N.S
EALKHMPRVPINASKEEIDAAYL	0	1
ELVDPSTGKPINNSTAGIVSGKP	0	1
ESKSSEDESSRNESEIVLSDKKG	0	1
FDYRLVGIDKDNNSFFPAFGIFN	0	1
FESTGEHKVSINLSANMTYQLKN	0	1
FEVFLMDKFDWNSSFTVEGQSFF	0	1
FIVNSKATNQTNDSWFNLDLLRE	0	1
GNDSGKNDSGKNDSSFAFKVSTS	0	1
GTIWCMADNNNNFSRIRDLANKD	0	1
ICLAAMHANTLNRSTGLAQITPI	0	1
IPKATDISLKVNESKPFILFFKQ	0	1
IRSQKKLLDELNKSKSVLRTKAL	0	1
IYSAGLLNTKENSSFNCDFELQG	0	1
KNLKLGELTKTNWSSTSTVATAK	0	1
KSSLFNQIDLNNISFLPKKEFFK	0	1
LAHQCFCVSPENMSKLAFANPEA	0	1
LCEVNTITYSENKSEAWNSELGG	0	1
LEESVELTLKINESGTIWCMADN	0	1
LIGLISCEERENISIENSINETS	0	1
LWNVADSNTTSNVSEDEKIIAMK	0	1
NGNQNNDLQNSNESLMQRNQTSS	0	1
NIDESYVTNSKNASEGILTGIGG	0	1
NMGNSEWRSASNLSAEQQIVYTA	0	1
NTCIHTCRHSANSSNCKLSANSI	0	1
NWATVCLAGKNNNSPHEFPYRGG	0	1
PHNETIMKDAGNDSGKNDSSFAF	0	1
PIKTPEIVYYNNTSNLRVLAKKG	0	1
QDDSIYLLKNENYSPEKTIKQKN	0	1
QLGVIAFSTNSNDSEMMVKSVEI	0	1
RSLKEENQDENNGSKTLPTSFYW	0	1
SDSDSNSTSDSNSSSNSNSSSNS	0	1
SGSVPLGELVENESFLLSGTRNL	0	1
SKLFSNYKTASNESMTLFEENVD	0	1
SNGVVGGVYFDNVSIKAKSCKKF	0	1
STSDSNSSSNSNSSSNSGENGNQ	0	1
TASNEVSQINSNSSQEIAKPNFF	0	1
TGSSNQDSTIDNSSFEGKKDKSK	0	1
TTPANVINNVSNSSQSQYNGTNI	0	1
TTTTTATTASINSSLSNTLMLLS	0	1

Occupied Sites	N.T	N.S

Unoccupied Sites	N.T	N.S
TTTTTEANLFLNASNVGKIASVN	0	1
TTTTTPANVINNVSNSSQSQYNG	0	1
VDVLDLVKKAINESKLNLGLITF	0	1
VKLIKSSGVEKNISSKDISGDIK	0	1
VSHGRVGLSVKNCSETAFDKVSL	0	1
VVTIERDSSFWNES---------	0	1
----MKSLPLINYSTIFIFLYSC	0	1
---------MMNISNNIFKVSIF	0	1
YDNNSDAKNRINQSNNATDKRSE	0	1
YPYAGVPGYPYNYSYSYTRTTSR	0	1
YSSPPSLNNDKNSSQSNE-----	0	1
YYNSCKSACSINDSYAENSTKEY	0	1
NATKETNKTYDNNSDAKNRINQS	0	1
PCEFKKALEGYNDSQYWNIDAND	0	1

Occupied Sites	N.T	N.S

Total Unoccupied	N.T	N.S
	65	57

APPENDIX 2B.2 The *N*-Glycosylation Sequons for the Identified Proteins.

Each protein that was identified to contain *N*-glycans is shown, and labeled according to the Accession number. The start of each sequon begins on a new line with the Asn is registered as the first position.

Acc. #		Nx[TS] Sequon	N.T	N.S
cgd1_660				
	NCTSFLNPLGEEGLCSGDSE	NCT	1	
	NKTKSSVKETEDDKTKEIDD	NKT	1	
	NETIMKDAG	NET	1	
	NDSGK	NDS		1
	NDSSFAFKVSTSLYIPVIFV	NDS		1
		Sub-Total	**3**	**2**
cgd7_4020				
	NSTGVMYDPNTNSILEGSIA	NST	1	
	NFTSTTGFTTDTSMNWPVSI	NFT	1	
	NETKTTEPSANTNFLLVDPK	NET	1	
	NQTMSGLSVSYLAAK	NQT	1	
	NLTVDTDETYGLPIDTLTGY	NLT	1	
	NGTIAGIVSGISASESLLSQ	NGT	1	

Acc. #		Nx[TS] Sequon	N.T	N.S
	NPTN	NPT	N/A	
	NNTMDSSFAGAYKYAVSNGI	NNT	1	
	NSTTGELVDPSTGKPI	NST	1	
	NNSTAGIVSGKPGLPPIEDE	NNS		1
	NSTAGIVSGKPGLPPIEDE	NST	1	
	NSTVSGSTSGTTKPKPGIPV	NST	1	
	NTTGNIINPETGKVIPGSLP	NTT	1	
	NITTDEVTGLPIDLETGLPR	NIT	1	
	NGTLVDPSNKKPIPGSHSGF	NGT	1	
	NGTSGEQSHEKDPSTGKPLD	NGT	1	
	NISPRPSSDAVPDRPTNTWW	NIS		1
	NETVVTIERDSSFW	NET	1	
	NES	NES		1
		Sub-Total	**15**	**3**
cgd3_660				
	NVSTIFGDLLNSKTFSDYNS	NVS		1
	NKTSISGNSAFVSQFSIPFI	NKT	1	
	NITIVEKLHASSENIDTSDS	NIT	1	
	NSTSDS	NST	1	
	NSSSNS	NSS		1
	NSSSNSGENGNQNENSLTET	NSS		1
	NESKLNLGLITFYGFPKSKD	NES		1
	NWSSTSTVATAKLNSVVLAA	NWS		1
	NSSFTVEGQSFFD	NSS		1
	NPTQKYTTDKSIPPTLLAVP	NPT	N/A	
	NDSEMMVKSVEIDLKCSFND	NDS		1
		Sub-Total	**3**	**7**
cgd2_490				
	NTTVIILILACLGNVVKSQK	NTT	1	
	NSTTEVRSLGHRRGGYVGDQ	NST	1	
	NYSYSYTRTTSRPGGVFTRP	NYS		1
	NESESNERKMDVMIPME	NES		1
	NMTPGFINDLQDGILSNGNQ	NMT	1	
	NESLMQR	NES		**1**
	NQTSSSGNNPVNNLLNRESD	NQT	1	
	NETSISSDGKRY	NET	1	
	NDTASPIKTPEIVYY	NDT	1	

Acc. #		Nx[TS] Sequon	N.T	N.S
	NNTSNLRVLAKKGLLERVEE	NNT	1	
	NTSNLRVLAKKGLLERVEE	NTS		1
		Sub-Total	**7**	**4**
cgd8_4660				
	NESMTLFEENVDHLFDYFFY	NES		1
	NTTVDENDFKFKRINRKKIM	NTT	1	
	NGSKTLPTSFYWP	NGS		1
	NETEKWAKISFLESGSITET	NET	1	
	NFTMSTYR	NFT	1	
	NETGLLTNPKELALVFEVLE	NET	1	
		Sub-Total	**4**	**2**
cgd3_3430				
	NPTEIIDGASSCPEEGNRGY	NPT	N/A	
	NYSPEKTIKQKNLLLSDILR	NYS		1
	NSSFNCDFELQGYLGCGFSS	NSS		1
	NFTTEKSWEVEGALGCSGRT	NFT	1	
	NYTVSLHDSYGDGWNSGSYL	NYT	1	
	NESGTIWCMADNNN	NES		1
	NFSRIRDLANKDHIDEQRLV	NFS		1
	NISSKDISGDIKLSIRNLIP	NIS		1
	NLTKRVTTKDRVPKIEITRI	NLT	1	
	NKTEAYEYLDSLDLHKRTPE	NKT	1	
	NETENMVDPLGLNGLYHQWK	NET	1	
	NTTS	NTT	1	
	NVSEDEKIIAMKHYPISEPR	NVS		1
	NPTGPFGRRLEQIIINDEGK	NPT	N/A	
	NVTSENEEEQERKDVDEEDE	NVT	1	
	NLTFLGRRVAFEVAMVDAMA	NLT	1	
	NETHHHPKAYFVVDEKLQTY	NET	1	
		Sub-Total	**9**	**6**
cgd1_640				
	NVTGQVPTLGEVKKMKKHFM	NVT	1	
	NATDTEMKSLFGVETTKKAD	NAT	1	
	NASKEEIDAAYLKAWEMLGK	NAS		1
	NATNFVNDGLFNDNLILDNF	NAT	1	
	NMTNAYEKIVKGEDVYDNHG	NMT	1	

Acc. #		Nx[TS] Sequon	N.T	N.S
	NMTVGQMLSLHTLIFENLYT	NMT	1	
	NKSKSVLRTKALLENVNIVS	NKS		1
	NESEIVLSDKKGSSPSPSNK	NES		1
		Sub-Total	**5**	**3**
cgd7_1730				
	NLSAEQQIVYTAFISATAKA	NLS		1
	NLSA	NLS		1
	NMTYQLKNLRIDWEFPPFTY	NMT	1	
	NPS	NPS		N/A
	NTTLNEMNGVISDFIQIVME	NTT	1	
	NVSIKAKSCKKFFMSPAPRS	NVS		1
	NTSTTLVKKTKAKTTLLKKV	NTS		1
	NCSETAFDKVSLEPFRVDSN	NCS		1
	NSTDKEKTDSIYKSCTLNTH	NST	1	
	NDSYAE	NDS		1
	NSTKEYINQLATCISGTSKA	NST	1	
		Sub-Total	**4**	**6**
cgd7_4810				
	NSTCGSQQSIVSSRLDNIES	NST	1	
	NESKPFILFFKQHPTEASLA	NES		1
	NPSMWPASLTMYIKQQVDEM	NPS		N/A
		Sub-Total	**1**	**1**
cgd7_300				
	NDTKKL	NDT	1	
	NETQILSDAYEANINKDRDN	NET	1	
	NFTGLLVVGAPIVQLTSGIT	NFT	1	
	NSTSGALPELVLCSTNTNLK	NST	1	
	NRSTGLAQITPIEGLESYGA	NRS		1
	NITS	NIT	1	
	NYTMPQAKSLGMGVAVDTFE	NYT	1	
	NLTTTTSEFNLDSEMTCQAD	NLT	1	
	NRTTKEQLKCKQECHKKDVM	NRT	1	
	NFTERKGQFSISMLNDEWTK	NFT	1	
		Sub-Total	**9**	**1**

Acc. #		Nx[TS] Sequon	N.T	N.S
cgd6_710				
	NISNNIFKVSIFYFVLGLIL	NIS		1
	NITSVDDLIQNSRDLMCILE	NIT	N/A	
	NPTGLITVLFFAVVTYDNVL	NPT	0	0
	NNSFFPAFGIFNKGDALALN	NNS		1
	NFTNKFGPPIPVPTTSTSTT	NFT	1	
		sub-Total	**2**	**2**
cgd2_640				
	NVTYDSNNDIFPRFSFPNYF	NVT	1	
	NATLHYGHKHHDFTVVVGNK	NAT	1	
	NSTIEGTTFLNCDSYCSFKV	NST	1	
		Sub-Total	**3**	**0**
cgd2_2510				
	NKTILKESYDWIVNRVSNKG	NKT	1	
	NGTWRVYNLF	NGT	1	
	NVSDHGFRMSSDILLSDYQN	NVS		1
	NNTVIETSPVDILTNHLVTK	NNT	1	
	NKTQSTGFPFLRNANLGGKF	NKT	1	
	NITAGKLANEFGIEDGIKPS	NIT	1	
	NESFLLSGTRNLGIVDASLL	NES		1
		Sub-Total	**5**	**2**
cgd7_180				
	NISIENSI	NIS		1
	NETSIPNEETNLRRLYGKGQ	NET	1	
	NPSPPQMYVPVTKSEPNVYV	NPS		N/A
		Sub-Total	**1**	**1**
cgd2_1290				
	NYSTIFIFLYSCFTVFNYAT	NYS		1
	NQT	NQT	1	
	NDSWFNLDLLREFLMDE	NDS		1
	NVTSSSTDVFEGFKKMIKVS	NVT	1	
	NNTKPEERGRNMFSFGFTGK	NNT	1	
	NSSNCKLSANSINLINQEIP	NSS		1
	NFTWLIPMRDHYQQVTDPSN	NFT	1	

Acc. #		Nx[TS] Sequon	N.T	N.S
		Sub-Total	**4**	**3**
cgd5_1210				
	NASNVGKIASVNRRLRENT	NAS		1
	NSSQKASKSSLFNQIDLN	NSS		1
	NISFLPKKEFFKPYYTTSTT	NIS		1
	NSSLSNTLMLLSQNNNRLQP	NSS		1
	NTTSPEATKET	NTT	1	
	NKTYD	NKT	1	
	NNSDAKNRI	NNS		1
	NQSN	NQS		1
	NATDKRSEKQTDLLQNQHIF	NAT	1	
	NVS	NVS		1
	NSSQSQY	NSS		1
	NGTNINSL	NGT	1	
	NSTESNPNYIKDKSDQLDRA	NST	1	
	NQTTNLNNERYISSSLILT	NQT	1	
	NGTSQGKFALSPL	NGT	1	
	NGTEVAPLFSKFVPNEDTSD	NGT	1	
	NSSQSNE	NSS		1
		Sub-Total	**8**	**9**
cgd5_1220				
	NKSEAWNSELGGNQITPSNI	NKS		1
	NASEGILTGIGGIDLPGMPL	NAS		1
	NSTNLDSSKVLLG	NST	1	
	NDSTVKSSSFFNPMDLSHPF	NDS		1
	NSSQEIAKPNFFLPIDLNNP	NSS		1
		Sub-Total	**1**	**4**
cgd7_4310				
	NVTLELLNGENTNKNDMSET	NVT	1	
	NNTSSFLPRQTDFSNIFKWD	NNT	1	
	NTSSFLPRQTDFSNIFKWD	NTS		1
	NLTEPQRQPCGRCGFGAKCC	NLT	1	
	NGTFPPFVKWPYRGGSGETL	NGT	1	
	NETSSIPVDIDDQEMANMQL	NET	1	
	NVTNNYGKDFYNYRMILRAE	NVT	1	
	NPSSSYLAHQCFCVSPE	NPS		N/A
	NMSKLAFANPEALEV	NMS		1

Acc. #		Nx[TS] Sequon	N.T	N.S
	NKTLIEEIYNR	NKT	1	
	NTTTDGFNRMLMIDGVSTPY	NTT	1	
	NITANIERMGFSVDND	NIT	1	
	NITLLSDVFMDLPEDDSFTQ	NIT	1	
	NETLLPNNVASIVASWAKPL	NET	1	
	NTTRLDSPSVLVTHNNLLDL	NTT	1	
	NNSPHEFPYRGGESEVVYSS	NNS		1
	NGTIPITPRVMDFITVSQAT	NGT	1	
	NPSKLATPLIAKLSSSRNCI	NPS		N/A
	NTTGSSNQDSTID	NTT	1	
	NSSFEGKKDKSKQEFSRECI	NSS		1
	NCTGRHALSEGCPCSRNIFS	NCT	1	
		Sub-Total	**15**	**4**
cgd7_1310				
	NETQSGVDLEQRHSIKERLK	NET	1	
	NDSQYWNIDANDDLLFDHGY	NDS		1
	NPSNRCVTNLNDKAVMKSCP	NPS		N/A
		Sub-Total	**1**	**1**
			N.T	**N.S**
		Grand Total	**100**	**61**

APPENDIX 2B.3 Percent NxT *versus* NxS Occupancies of *N*-Glycoyslation Sequons: Occupied and Unoccupied Sequons.

	Unoccupied	Occupied	Total	% Occupied	% Unoccupied
NxT	65	35	100	35.0%	65.0%
NxS	57	4	61	6.6%	93.4%

% NxT, of occupied	**89.7%**

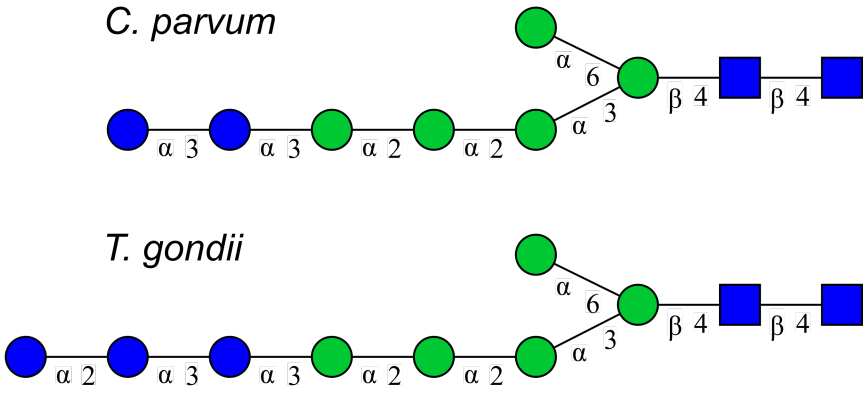
APPENDIX 2B.4 Table of Protein Names Cross-Referenced to Various Database Accession Numbers; Conserved Domains Assigned by InterPro.

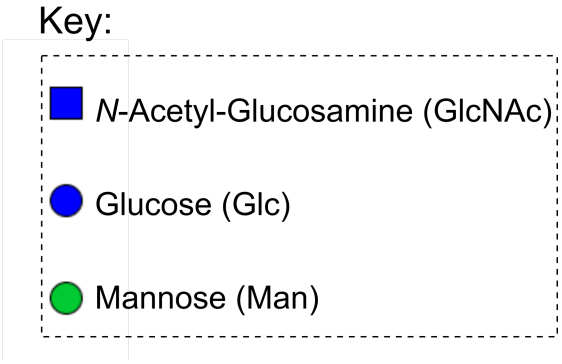
The given protein names are shown cross-referenced to the UniProt accession numbers, PubMed ID, and the gene accession numbers. The conserved domains are shown, as assigned by InterPro, according to the UniProtKB accession numbers for each protein. The amino acid (AA) positions are indicates as per their InterPro entries.

Given name	UniProt Acc. #	PubMed ID	Gene Acc. #	InterPro Domains
UCG3	Q5CT01	15044751	cgd1_640	N/A
UCG4	Q5CSZ9	15044751	cgd1_660	N/A
POWP1	Q5CU33	15044751	cgd2_490	N/A
GAP50	Q5CU19	15044751	cgd2_640	IPR029052(Metallo-dependent phosphatase-like) AA:23-249
UCG5	Q5CTW5	15044751	cgd2_1290	N/A
GMCO	Q5CTL6	15044751	cgd2_2510	IPR023753(FAD/NAD-binding_dom.);IPR000172(Glucose-methanol-choline oxidoreductase, N-terminal)AA:45-197;368-415;IPR007867 (Glucose-methanol-choline oxidoreductase, N-terminal)AA:533-742;
UCG6	Q5CV13	15044751	cgd3_660	
COA	Q5CUC0	15044751	cgd3_3430	IPR013320(Concanavalin A-like lectin/glucanase domain)AA:285-474;IPR000269(Copper amine oxidase)AA:1,244-1,300;1,335-1,750;IPR015798(Copper amine oxidase, C-terminal)AA:1,343-1,740;IPR000998(MAM domain)AA:288-475
UCG1	Q5CRW6	15044751	cgd5_1210	N/A
UCG2	Q5CXK6	15044751	cgd6_710	N/A
CCP2	Q5CZ08	15044751	cgd7_300	IPR008979(Galactose-binding domain-like)AA:138-234;366-503,IPR000772(Ricin B, lectin domain)AA:243-372,IPR000421(Coagulation factor 5/8 C-terminal domain)AA:373-482,IPR014716(Fibrinogen, alpha/beta/gamma chain, C-terminal globular, subdomain 1)AA:597-662,IPR002181(Fibrinogen, alpha/beta/gamma chain, C-terminal globular domain)AA:619-663,IPR004043(LCCL domain)AA:780-883
O-GAT4	Q5CYR4	15044751	cgd7_1310	IPR029044(Nucleotide-diphospho-sugar transferases)AA:52-343 ,IPR001173(Glycosyltransferase 2-like)AA:54-180,IPR000772(Ricin B, lectin domain)AA:361-460
CCP1	Q5CYM9	15044751	cgd7_1730	IPR008979(Galactose-binding domain-like)AA:42-166;302-432,IPR000772(Ricin B, lectin domain)AA:168-289,IPR002181(Fibrinogen, alpha/beta/gamma chain, C-terminal globular domain)AA:541-589 ,IPR004043(LCCL domain)AA:738-834
Gp900	Q5CY21	15044751	cgd7_4020	N/A
FNPA	Q5CXV2	15044751	cgd7_4810	IPR000562(Fibronectin, type II, collagen-binding)AA:39-100,IPR011658(PA14 domain)AA:209-419
UCG7	Q5CVB8	15044751	cgd8_4660	N/A

APPENDIX 2C: The Predicted Lipid Linked *N*-Glycan Precursors of *C. parvum* and *T. gondii*

The cartoons show a visual representation of the predicted complete *N*-glycan which is transferred to the Asn on *N*-glycosylated proteins in each of the organisms. The prediction is based upon the predicted enzymes found in the genomic sequences as shown in Table 2.2.





APPENDIX 3: The Complete Table of Observed *N*-Glycosylated Peptides and their Variants used to Generate Table 3.1.

Protein Name	Accession#	Peptide	Δppm	#Spec	Mods
GAP50	TGGT1_219320	(76)K.VAANEHISFIASPGSNFLGGVSSL**N**DTR.W(105)	0.6	2	-#6-
		(76)K.VAANEHISFIASPGSNFLGGVSSL**N**DTR.W(105)	-0.2	2	-#7-
		(76)K.VAANEHISFIASPGSNFLGGVSSL**N**DTR.W(105)	2.5	1	-#7-, -!-
		(76)K.VAANEHISFIASPGS<u>N</u>FLGGVSSL**N**DTR.W(105)	1.1	1	-%-, -#8-
		(135)R.**N**YTSEALR.T(144)	-2.2	1	-#6-
		(135)R.**N**YTSEALR.T(144)	0.9	5	-#7-
		(135)R.**N**YTSEALR.T(144)	-0.5	2	-#7-, -!-
		(135)R.**N**YTSEALR.T(144)	0.9	9	-#8-
		(135)R.**N**YTSEALR.T(144)	1.4	1	-#8-, -!-
SRS29C	TGGT1_233480	(196)R.<u>C</u>SYTE**N**STLPK.I(208)	0.1	2	-$-, -#6-
		(300)K.Y**N**<u>C</u>TVPVQLGGEDPSEGSR.P(320)	3	2	-#6-, -$-
		(300)K.Y**N**<u>C</u>TVPVQLGGEDPSEGSR.P(320)	2.1	2	-$-, -#6-, -!-
		(300)K.Y**N**<u>C</u>TVPVQLGGEDPSEGSR.P(320)	0.1	1	-#7-, -$-
		(300)K.Y**N**<u>C</u>TVPVQLGGEDPSEGSRPGGGSGGGK.R(329)	2.3	4	-#6-, -$-
hyp. protein	TGGT1_217680	(406)R.T**N**STLFESQLR.E(418)	3.2	1	-#6-
		(457)R.GV**N**VTIDR.H(466)	-1.7	2	-#6-
hyp. protein	TGGT1_243930	(696)K.MN**N**ETVLYEPDTEIIEK.T(714)	1.9	1	-#8-, -!-
		(763)K.TMNSEGVISDGLQSQLPV**N**HTR.L(786)	-0.7	1	-#8-, -!-
hyp. protein	TGGT1_258870A	(178)R.AHTGDERPF**N**VTTGSSER.R(197)	1.9	1	-#8-
CS-dom. prot.	TGGT1_290730	(333)K.DLDDFHHG**N**YTAR.Y(347)	-0.1	1	-#6-, -!-
CLPTM1	TGGT1_299110	(141)R.N**N**TTLYVHVR.T(152)	-0.7	1	-#6-
SRS22E	TGGT1_359770	(36)K.IET<u>C</u>APDKPISF**N**VTEAGQSILFK.C(61)	0.9	1	-$-, -#6-

The results from the two samples, including the technical replicates, all analyzed on the QE+ mass spectrometer, and searched using the PEAKS software suite. The results are collated from the PEAKS-PTM searches, the ppm error is shown for the best scoring spectrum (-10lgP).The symbols used in the table are defined as follows: (N = modified Asn residue; <u>X</u> = other modification, where X is any amino acid; -$- = carbamidomethylation of Cys, 57.02146;-!- = carbamidomethylation artifact, 57.02146; -%- = deamidation, 0.984; -#6- = $HexNAc_2Hex_6$, 1378.47566; -#7- = $HexNAc_2Hex_7$, 1540.52848; -#8- = $HexNAc_2Hex_8$, 1702.5813).

APPENDIX 4: Ion Assignments, Peptides, Sequons, and Bioinformatics Data for the *O*-Glycosylated Peptides.

APPENDIX 4A: Complete Tables of Peptides and Ions for the Spectra Assigned to *O*-Glycosylated Peptides.

APPENDIX 4A.1 All HexNAc Modified Peptides used to Generate Table 4.1.

The representative composite peptide is the shown in Table 4.1. The unique peptides used to make the composite peptide are shown to the right, labeled as "unique peptide" in the table. The number of HexNAc(s) and the number of corresponding spectra are also shown on the right hand side of the table.

Protein	Representative Peptide	Unique Peptide	#HexNAc	#Spec.
Gp40	(31)DVPVEGSSSSSSSSSSSSSSSSSTST VAPANK(62)	(31)DVPVEGSSSSSSSSSSSSSSSSSTST VAPANK(62)	20	2
			19	1
			18	1
			17	1
			15	1
Gp15	(221)ETSEAAATVDLFAFTLDGGK(240)	(219)SEETSEAAATVDLFAFTLDGGK(240)	4	8
			3	31
			2	6
		(219)SEETSEAAATVDLFAFTLDGGKR(241)	4	3
			3	9
			2	2
		(221)ETSEAAATVDLFAFTLDGGK(240)	4	14
			3	36
			2	30
		(221)ETSEAAATVDLFAFTLDGGKR(241)	4	3
			3	8
			2	8
		(222)TSEAAATVDLFAFTLDGGK(240)	4	1
			1	1
		(223)SEAAATVDLFAFTLDGGK(240)	3	2
			2	21
		(223)SEAAATVDLFAFTLDGGKR(241)	2	5
		(225)AAATVDLFAFTLDGGK(240)	2	21
		(225)AAATVDLFAFTLDGGKR(241)	2	5
		(227)ATVDLFAFTLDGGK(240)	2	13
			1	1
		(227)ATVDLFAFTLDGGKR(241)	2	2
		(228)TVDLFAFTLDGGK(240)	2	3
		(232)FAFTLDGGK(240)	1	1
		(233)AFTLDGGK(240)	1	10
Gp20	(87)EGEETDENTDETTTTTTTASPKPK(110)	(87)EGEETDENTDETTTTTTTASPKPK(110)	8	4
			7	7
			6	7

Protein	Representative Peptide	Unique Peptide	#HexNAc	#Spec.
		(87)EGEETDENTDETTTTTTTASPK(108)	7	1
	(135)SSTTTTTTTAPVSSEDNKPEDSEDEK(160)	(135)SSTTTTTTTAPVSSEDNKPEDSEDEK(160)	8	2
Gp900	(609)KPTTTTTTTTTTTTK	(609)KPTTTTTTTTTTTTK(623)	3	1
	(958)IADTSNLFPVQTHK(971)	(958)IADTSNLFPVQTHK(971)	1	2
	(1197)TPTQTDSVTGKPIDPTTGLPFNPPTGH(1223)	(1197)TPTQTDSVTGK(1207)	1	4
	(1197)TPTQTDSVTGK(1207)	(1197)TPTQTDSVTGKPIDPTTGLPFNPPTGH(1223)	1	1
	(1243)YAVSNGIKTDNVYGLPVDEITGLPK(1267)	(1243)YAVSNGIKTDNVYGLPVDEITGLPK(1267)	1	1
	(1247)NGIKTDNVYGLPVDEITGLPK(1267)	(1247)NGIKTDNVYGLPVDEITGLPK(1267)	1	1
		(1248)GIKTDNVYGLPVDEITGLPK(1267)	1	1
		(1251)TDNVYGLPVDEITGLPK(1267)	1	19
	(1373)GKDGLIVPPTNSINK(1387)	(1373)GKDGLIVPPTNS(+203.08)INK(1387)	1	1
	(1410)VIPGSLPGSLNYPSFNTPQQTDEITGK(1436)	(1410)VIPGSLPGS(+203.08)LNYPSFNTPQQTDEITGK(1436)	1	4
	(1646)TIPGSAASVIHTALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIK(1698)	(1646)TIPGSAASVIHTALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIK(1698)	1	4
	(1658)ALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIK(1698)	(1658)ALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIK(1698)	1	2
	(1710)EKNIVTEAAYGLPVDPK(1726)	(1710)EKNIVTEAAYGLPVDPK(1726)	1	2
		(1712)NIVTEAAYGLPVDPK(1726)	1	23
	(1795)LIDPESGIAIDNSVSGVFATVPGTAAPK(1822)	(1795)LIDPESGIAIDNSVSGVFATVPGTAAPK(1822)	1	17
	(1813)ATVPGTAAPK(1822)	(1813)ATVPGTAAPK(1822)	1	4

APPENDIX 4A.2 Spectra Re-Annotated using Glycresoft.

This table indicates the spectra which were re-annotated using Glycresoft. The files are named according to the "Sourcefile.Scan", and have been deposited online into the PRIDE repository as indicated in the text of Chapter 4.

Protein	Peptide	#HexNAc	Source File	Scan#
Gp40/Gp15	K.DVPVEGSSSSSSSSSSSSSSSSSTSTVAPANK.A	20	03-11-2015_WTD_120min_45V_QE+.raw	3987
			02-17-2015_WTD_QE+.raw	3065
		19	03-11-2015_WTD_120min_45V_QE+.raw	4006
		18	03-11-2015_WTD_120min_45V_QE+.raw	4028
		17	03-11-2015_WTD_120min_45V_QE+.raw	4027
		15	03-11-2015_WTD_120min_45V_QE+.raw	4036
	L.SEETSEAAATVDLFAFTLDGGK.R	4	02-17-2015_WTD_QE+.raw	5947
			02-17-2015_WTD_QE+.raw	5947
			02-17-2015_WTD_120min_QE+.raw	9566
			05-08-2015_JRH_CP_PH_QE+.raw	24766
			03-11-2015_WTD_120min_45V_QE+.raw	7985
			06-24-2015_CP_PH_15V&45V_QE+.raw	9827
			02-17-2015_WTD_120min_QE+.raw	9592
			02-17-2015_WTD_120min_QE+.raw	9592
			05-15-2015_wtd_7-2012-A_QE+.raw	7113
			06-24-2015_CP_PH_45V_QE+.raw	10001
	L.SEETSEAAATVDLFAFTLDGGKR.I	4	05-08-2015_JRH_CP_PH_QE+.raw	22981
			05-08-2015_JRH_CP_PH_QE+.raw	23116
			06-24-2015_CP_PH_45V_QE+.raw	9590
	L.SEETSEAAATVDLFAFTLDGGK.R	3	05-08-2015_JRH_CP_PH_QE+.raw	25131
			05-15-2015_cp_inter_QE+.raw	7316
			02-17-2015_WTD_QE+.raw	5996
			02-17-2015_WTD_QE+.raw	5965
			02-17-2015_WTD_QE+.raw	5993
			02-17-2015_WTD_120min_QE+.raw	9697
			07-10-2015_ph_inter.raw	23055
			05-15-2015_wtd_7-2012-A_QE+.raw	7118
			05-08-2015_JRH_CP_PH_QE+.raw	24962
			05-15-2015_wtd_7-2012-A_QE+.raw	7136
			05-08-2015_JRH_CP_PH_QE+.raw	24886
			05-08-2015_JRH_CP_PH_QE+.raw	25139
			02-17-2015_WTD_120min_QE+.raw	9639
			02-17-2015_WTD_120min_QE+.raw	9626
			02-17-2015_WTD_120min_QE+.raw	9659

Protein	Peptide	#HexNAc	Source File	Scan#
			05-08-2015_JRH_CP_PH_QE+.raw	24811
			05-08-2015_JRH_CP_PH_QE+.raw	24749
			07-10-2015_ph_inter.raw	23029
			05-08-2015_JRH_CP_PH_QE+.raw	25012
			03-11-2015_WTD_120min_45V_QE+.raw	8088
			07-10-2015_ph_inter.raw	23180
			06-24-2015_CP_PH_45V_QE+.raw	10037
			06-24-2015_CP_PH_15V&45V_QE+.raw	9862
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9765
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9770
			03-11-2015_WTD_120min_45V_QE+.raw	8017
			06-24-2015_CP_PH_15V&45V_QE+.raw	9875
			06-24-2015_CP_PH_45V_QE+.raw	10054
			03-11-2015_WTD_120min_45V_QE+.raw	8037
			03-11-2015_WTD_120min_45V_QE+.raw	8052
			06-24-2015_CP_PH_15V&45V_QE+.raw	9912
	L.SEETSEAAATVDLFAFTL DGGKR.I	3	05-15-2015_wtd_7-2012-A_QE+.raw	6802
			06-24-2015_CP_PH_45V_QE+.raw	9626
			05-08-2015_JRH_CP_PH_QE+.raw	23272
			06-24-2015_CP_PH_15V&45V_QE+.raw	9454
			07-10-2015_ph_inter.raw	21064
			05-08-2015_JRH_CP_PH_QE+.raw	23143
			05-08-2015_JRH_CP_PH_QE+.raw	23185
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9370
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9370
			05-08-2015_JRH_CP_PH_QE+.raw	23321
	L.SEETSEAAATVDLFAFTL DGGK.R	2	02-17-2015_WTD_QE+.raw	6030
			05-15-2015_wtd_7-2012-A_QE+.raw	7220
			05-08-2015_JRH_CP_PH_QE+.raw	25259
			03-11-2015_WTD_120min_45V_QE+.raw	8100
			02-17-2015_WTD_120min_QE+.raw	9749
			02-17-2015_WTD_120min_QE+.raw	9774
	L.SEETSEAAATVDLFAFTL DGGKR.I	2	05-08-2015_JRH_CP_PH_QE+.raw	23603
			07-10-2015_ph_inter.raw	21498
	E.ETSEAAATVDLFAFTLD GGK.R	4	02-17-2015_WTD_QE+.raw	5960
			05-08-2015_JRH_CP_PH_QE+.raw	24925
			05-08-2015_JRH_CP_PH_QE+.raw	24650
			02-17-2015_WTD_QE+.raw	5960

Protein	Peptide	#HexNAc	Source File	Scan#
			02-17-2015_WTD_QE+.raw	5985
			02-17-2015_WTD_120min_QE+.raw	9614
			02-17-2015_WTD_120min_QE+.raw	9584
			05-08-2015_JRH_CP_PH_QE+.raw	24785
			05-15-2015_wtd_7-2012-A_QE+.raw	7110
			05-08-2015_JRH_CP_PH_QE+.raw	24680
			03-11-2015_WTD_120min_45V_QE+.raw	8010
			06-24-2015_CP_PH_45V_QE+.raw	10000
			06-24-2015_CP_PH_15V&45V_QE+.raw	9816
			03-11-2015_WTD_120min_45V_QE+.raw	8003
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9724
	E.ETSEAAATVDLFAFTLDGGKR.I	4	05-08-2015_JRH_CP_PH_QE+.raw	23130
			05-08-2015_JRH_CP_PH_QE+.raw	22995
			06-24-2015_CP_PH_45V_QE+.raw	9592
	E.TSEAAATVDLFAFTLDGGK.R	4	05-08-2015_JRH_CP_PH_QE+.raw	24958
	E.ETSEAAATVDLFAFTLDGGK.R	3	06-26-2015_CP_PH_15V_QE+.raw	12338
	E.ETSEAAATVDLFAFTLDGGK.R	3	02-17-2015_WTD_QE+.raw	5986
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9758
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9797
			02-17-2015_WTD_QE+.raw	5986
			02-17-2015_WTD_QE+.raw	5975
			02-17-2015_WTD_QE+.raw	6007
			02-17-2015_WTD_120min_QE+.raw	9640
			05-15-2015_wtd_7-2012-A_QE+.raw	7132
			02-17-2015_WTD_120min_QE+.raw	9631
			02-17-2015_WTD_120min_QE+.raw	9669
			02-17-2015_WTD_120min_QE+.raw	9658
			05-08-2015_JRH_CP_PH_QE+.raw	24906
			02-17-2015_WTD_120min_QE+.raw	9690
			05-15-2015_wtd_7-2012-A_QE+.raw	7129
			02-17-2015_WTD_120min_QE+.raw	9720
			05-08-2015_JRH_CP_PH_QE+.raw	24928
			05-08-2015_JRH_CP_PH_QE+.raw	25033
			05-08-2015_JRH_CP_PH_QE+.raw	25160
			05-08-2015_JRH_CP_PH_QE+.raw	24772
			05-15-2015_cp_inter_QE+.raw	7327
			06-24-2015_CP_PH_15V&45V_QE+.raw	9918
			05-15-2015_wtd_7-2012-A_QE+.raw	7161

Protein	Peptide	#HexNAc	Source File	Scan#
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9757
			06-24-2015_CP_PH_45V_QE+.raw	10046
			03-11-2015_WTD_120min_45V_QE+.raw	8036
			03-11-2015_WTD_120min_45V_QE+.raw	8024
			06-24-2015_CP_PH_45V_QE+.raw	10038
			06-24-2015_CP_PH_15V&45V_QE+.raw	9867
			06-24-2015_CP_PH_15V&45V_QE+.raw	9865
			03-11-2015_WTD_120min_45V_QE+.raw	8050
			06-24-2015_CP_PH_45V_QE+.raw	10078
			03-11-2015_WTD_120min_45V_QE+.raw	8096
			03-11-2015_WTD_120min_45V_QE+.raw	8062
			06-24-2015_CP_PH_15V&45V_QE+.raw	9901
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9825
			06-26-2015_CP_PH_15V_QE+.raw	12333
Gp20	K.EGEETDENTDETTTTTTT ASPKPK.S	8	05-15-2015_wtd_7-2012-A_QE+.raw	2922
	K.EGEETDENTDETTTTTTT ASPK.P	7	03-11-2015_WTD_120min_45V_QE+.raw	3385
	K.EGEETDENTDETTTTTTT ASPKPK.S	7	05-15-2015_wtd_7-2012-A_QE+.raw	2962
			05-15-2015_wtd_7-2012-A_QE+.raw	2928
			02-17-2015_WTD_120min_QE+.raw	3246
			05-15-2015_wtd_7-2012-A_QE+.raw	2935
			03-11-2015_WTD_120min_45V_QE+.raw	3380
			03-11-2015_WTD_120min_45V_QE+.raw	3371
			03-11-2015_WTD_120min_45V_QE+.raw	3403
	K.EGEETDENTDETTTTTTT ASPKPK.S	6	02-17-2015_WTD_QE+.raw	2855
			02-17-2015_WTD_120min_QE+.raw	3329
			02-17-2015_WTD_120min_QE+.raw	3332
			05-15-2015_wtd_7-2012-A_QE+.raw	3001
			03-11-2015_WTD_120min_45V_QE+.raw	3471
			03-11-2015_WTD_120min_45V_QE+.raw	3504
			03-11-2015_WTD_120min_45V_QE+.raw	3483
	K.SSTTTTTTTAPVSSEDNK PEDSEDEK.E	8	02-17-2015_WTD_120min_QE+.raw	3513
			03-11-2015_WTD_120min_45V_QE+.raw	3681
Gp900	K.KPTTTTTTTTTTTTK.K	3	02-17-2015_WTD_120min_QE+.raw	3178
	Q.IADTSNLFPVQTHK.S	1	02-17-2015_WTD_120min_QE+.raw	6833
			02-17-2015_WTD_QE+.raw	4633
	K.TPTQTDSVTGK.P	1	02-17-2015_WTD_120min_QE+.raw	3403
			02-17-2015_WTD_QE+.raw	2939

Protein	Peptide	#HexNAc	Source File	Scan#
	K.TPTQTDSVTGKPIDPTTGLPFNPPTGH.L	1	07-10-2015_ph_inter.raw	15875
	K.YAVSNGIKTDNVYGLPVDEITGLPK.D	1	07-10-2015_ph_inter.raw	20433
	S.NGIKTDNVYGLPVDEITGLPK.D	1	05-15-2015_wtd_7-2012-A_QE+.raw	6854
	N.GIKTDNVYGLPVDEITGLPK.D	1	05-15-2015_wtd_7-2012-A_QE+.raw	6580
	K.TDNVYGLPVDEITGLPK.D	1	02-17-2015_WTD_QE+.raw	5725
			05-15-2015_cp_inter_QE+.raw	6931
			02-17-2015_WTD_QE+.raw	5716
			02-17-2015_WTD_120min_QE+.raw	9092
			02-17-2015_WTD_120min_QE+.raw	9113
			05-15-2015_wtd_7-2012-A_QE+.raw	6736
			02-17-2015_WTD_120min_QE+.raw	9124
			05-08-2015_JRH_CP_PH_QE+.raw	22765
			07-10-2015_ph_inter.raw	21361
			02-17-2015_WTD_120min_QE+.raw	9088
			05-08-2015_JRH_CP_PH_QE+.raw	22798
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	9261
			02-17-2015_WTD_QE+.raw	5747
			06-24-2015_CP_PH_45V_QE+.raw	9523
			06-24-2015_CP_PH_15V&45V_QE+.raw	9342
			03-11-2015_WTD_120min_45V_QE+.raw	7583
			03-11-2015_WTD_120min_45V_QE+.raw	7609
			03-11-2015_WTD_120min_45V_QE+.raw	7590
			03-11-2015_WTD_120min_45V_QE+.raw	7620
	K.GKDGLIVPPTNSINK.D	1	05-08-2015_JRH_CP_PH_QE+.raw	26837
	K.VIPGSLPGSLNYPSFNTPQQTDEITGK.P	1	02-17-2015_WTD_QE+.raw	5452
			02-17-2015_WTD_120min_QE+.raw	8622
			02-17-2015_WTD_120min_QE+.raw	8657
			02-17-2015_WTD_120min_QE+.raw	8641
	K.TIPGSAASVIHTALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIK.G	1	05-15-2015_cp_inter_QE+.raw	8634
			07-10-2015_ph_inter.raw	29236
			05-15-2015_cp_inter_QE+.raw	8663
			05-15-2015_wtd_7-2012-A_QE+.raw	8508
	T.ALGTPTQTDPTTGLPSDPSTGLPFIPGFNVLVDPQTGEQIK.G	1	05-15-2015_wtd_7-2012-A_QE+.raw	8401
			05-15-2015_wtd_7-2012-A_QE+.raw	8411
	K.EKNIVTEAAYGLPVDPK.T	1	07-10-2015_ph_inter.raw	13333
			07-10-2015_ph_inter.raw	13307

Protein	Peptide	#HexNAc	Source File	Scan#
	K.NIVTEAAYGLPVDPK.T	1	05-15-2015_cp_inter_QE+.raw	5719
			02-17-2015_WTD_QE+.raw	4891
			06-25-2015_CP_PH_15&45V_HCD_QE+.raw	7647
			02-17-2015_WTD_QE+.raw	4892
			05-15-2015_wtd_7-2012-A_QE+.raw	5552
			05-15-2015_wtd_7-2012-A_QE+.raw	5582
			02-17-2015_WTD_120min_QE+.raw	7409
			02-17-2015_WTD_120min_QE+.raw	7398
			02-17-2015_WTD_120min_QE+.raw	7430
			05-08-2015_JRH_CP_PH_QE+.raw	16458
			02-17-2015_WTD_120min_QE+.raw	7400
			05-08-2015_JRH_CP_PH_QE+.raw	16324
			07-10-2015_ph_inter.raw	15023
			07-10-2015_ph_inter.raw	15152
			05-08-2015_JRH_CP_PH_QE+.raw	16333
			07-10-2015_ph_inter.raw	15144
			06-26-2015_CP_PH_15V_QE+.raw	9181
			03-11-2015_WTD_120min_45V_QE+.raw	6427
			02-17-2015_WTD_120min_QE+.raw	7441
			06-24-2015_CP_PH_45V_QE+.raw	7855
			03-11-2015_WTD_120min_45V_QE+.raw	6456
			06-24-2015_CP_PH_15V&45V_QE+.raw	7648
			03-11-2015_WTD_120min_45V_QE+.raw	6425
	K.LIDPESGIAIDNSVSGVFATVPGTAAPK.K	1	05-15-2015_cp_inter_QE+.raw	7219
			02-17-2015_WTD_QE+.raw	5895
			07-10-2015_ph_inter.raw	22759
			07-10-2015_ph_inter.raw	22613
			07-10-2015_ph_inter.raw	22820
			06-24-2015_CP_PH_15V&45V_QE+.raw	9803
			07-10-2015_ph_inter.raw	22656
			05-15-2015_cp_inter_QE+.raw	7219
			05-15-2015_wtd_7-2012-A_QE+.raw	7030
			02-17-2015_WTD_120min_QE+.raw	9560
			02-17-2015_WTD_120min_QE+.raw	9539
			05-08-2015_JRH_CP_PH_QE+.raw	24449
			05-08-2015_JRH_CP_PH_QE+.raw	24324
			06-24-2015_CP_PH_45V_QE+.raw	9956
			03-11-2015_WTD_120min_45V_QE+.raw	7956
			03-11-2015_WTD_120min_45V_QE+.raw	7926

Protein	Peptide	#HexNAc	Source File	Scan#
			07-10-2015_ph_inter.raw	22686
			07-10-2015_ph_inter.raw	22551
	F.ATVPGTAAPK.K	1	02-17-2015_WTD_120min_QE+.raw	3875
			02-17-2015_WTD_QE+.raw	3136
			02-17-2015_WTD_120min_QE+.raw	3846
			05-15-2015_wtd_7-2012-A_QE+.raw	3361

APPENDIX 4B: Ion Assignments for the Mass Spectra Annotated in Figs. 4.2 to 4.7, and Fig.4.9.

APPENDIX 4B.1: Ion Assignments for Fig 4.2.

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
147.1129	147.1128	0.7	1	y_1
261.1555	261.1557	-0.9	1	y_2
312.1558	312.1554	1.3	1	b_3
332.1927	332.1928	-0.4	1	y_3
411.2290	411.2238	12.6	1	b_4
429.2457	429.2456	0.2	1	y_4
500.2830	500.2827	0.6	1	y_5
540.2659	540.2664	-0.9	1	b_5
597.2912	597.2879	5.6	1	b_6
599.3525	599.3511	2.3	1	y_6
684.3196	684.3199	-0.4	1	b_7
700.3945	700.3988	-6.2	1	y_7
771.3490	771.3519	-3.8	1	b_8
787.4269	787.4308	-5.0	1	y_8
858.3825	858.3840	-1.7	1	b_9
888.4797	888.4785	1.3	1	y_9
903.4768	903.4782	-1.5	1	y_7*
945.4262	945.4160	10.8	1	b_{10}
975.5067	975.5106	-3.9	1	y_{10}
990.5093	990.5102	-0.9	1	y_8*
1032.4525	1032.4480	4.3	1	b_{11}
1062.5382	1062.5426	-4.1	1	y_{11}
1119.4705	1119.4800	-8.5	1	b_{12}
1149.5626	1149.5746	-10.4	1	y_{12}
1206.5244	1206.5121	10.2	1	b_{13}
1293.5364	1293.5441	-5.9	1	b_{14}
1352.6537	1352.6540	-0.2	1	y_{12}*
1375.5880	1375.6078	-14.4	2	y_{30}
1477.1340	1477.1474	-8.9	2	y_{30}*
1482.6445	1482.6554	-7.4	2	aglycon
1578.6803	1578.6871	-4.3	2	y_{30}**
1584.1819	1584.1951	-8.3	2	[M - $HexNAc_{19}$]

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
1680.2161	1680.2268	-6.4	2	y_{30}***
1685.7175	1685.7348	-10.3	2	[M -$HexNAc_{18}$]
1729.7430	1729.7610	-10.4	2	y_{31}***
1758.7390	1758.7400	-0.5	2	b_{30}*4
1787.2545	1787.2745	-11.2	2	[M -$HexNAc_{17}$]
1888.7813	1888.8142	-17.4	2	[M -$HexNAc_{16}$]
1990.8346	1990.8935	-29.6	1	y_{17}**

APPENDIX 4B.2: Ion Assignments for Fig 4.3.

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
102.0554	102.0550	4.4	1	iE
120.0807	120.0808	-0.6	1	iF
129.1023	129.1022	0.5	1	y_1 -H_2O
130.0866	130.0863	2.7	1	y_1 -NH_3
147.1129	147.1128	0.7	1	y_1
187.1073	187.1077	-2.2	1	y_2 -NH_3
213.0870	213.0870	0.1	1	b_2 -H_2O
231.0970	231.0975	-2.4	1	b_2
243.1481	243.1452	12.1	1	y_3 -H_2O
244.1289	244.1292	-1.2	1	y_3 -NH_3
261.1557	261.1557	-0.1	1	y_3
300.1190	300.1190	0.0	1	b_3 -H_2O
318.1292	318.1296	-1.2	1	b_3
358.1722	358.1721	0.3	1	y_4 -H_2O
359.1556	359.1561	-1.5	1	y_4 -NH_3
376.1831	376.1827	1.1	1	y_4
416.1723	416.1664	14.3	1	b_2 -H_2O*
429.1616	429.1616	0.0	1	b_4 -H_2O
447.1720	447.1722	-0.4	1	b_4
471.2566	471.2562	0.9	1	y_5 -H_2O
489.2672	489.2667	0.9	1	y_5
500.1985	500.1987	-0.4	1	b_5 -H_2O
504.1885	504.1824	12.1	1	b_3 -NH_3*
518.2089	518.2093	-0.7	1	b_5
521.2120	521.2089	5.9	1	b_3*
571.2355	571.2358	-0.6	1	b_6 -H_2O

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
572.3088	572.3039	8.6	1	y_6 -H_2O
589.2458	589.2464	-1.0	1	b_6
590.3173	590.3144	4.9	1	y_6
642.2708	642.2729	-3.3	1	b_7 -H_2O
660.2828	660.2835	-1.1	1	b_7
719.3768	719.3723	6.3	1	y_7 -H_2O
737.3828	737.3828	0.0	1	y_7
743.3200	743.3206	-0.8	1	b_8 -H_2O
761.3290	761.3312	-2.9	1	b_8
790.4130	790.4094	4.6	1	y_8 -H_2O
793.3946	793.3938	1.0	1	y_6*
808.4201	808.4199	0.2	1	y_8
842.3888	842.3890	-0.3	1	b_9 -H_2O
860.4027	860.3996	3.6	1	b_9
922.4427	922.4516	-9.7	1	y_7 -H_2O*
937.4783	937.4778	0.5	1	y_9 -H_2O
940.4599	940.4622	-2.4	1	y_7*
955.4885	955.4884	0.1	1	y_9
964.4237	964.4106	13.6	1	b_8*
975.4263	975.4265	-0.3	1	b_{10}
993.4783	993.4888	-10.5	1	y_8 -H_2O*
1011.4979	1011.4993	-1.4	1	y_8 *
1050.5577	1050.5619	-4.0	1	y_{10} -H_2O
1068.5726	1068.5724	0.2	1	y_{10}
1088.5111	1088.5106	0.4	1	b_{11}
1158.5685	1158.5677	0.7	1	y_9*
1165.5850	1165.5888	-3.3	1	y_{11} -H_2O
1166.5791	1166.5728	5.4	1	y_{11} -NH_3
1183.6000	1183.5994	0.5	1	y_{11}
1217.5688	1217.5685	0.3	1	b_{12} -H_2O
1235.5837	1235.5790	3.8	1	b_{12}
1264.6603	1264.6572	2.4	1	y_{12} -H_2O
1265.6534	1265.6412	9.6	1	y_{12} -NH_3
1271.6377	1271.6518	-11.1	1	y_{10}*
1282.6694	1282.6678	1.3	1	y_{12}
1288.6049	1288.6056	-0.5	1	b_{13} -H_2O
1289.6074	1289.5896	13.8	1	b_{13} -NH_3
1306.6124	1306.6161	-2.9	1	b_{13}

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
1365.7001	1365.7049	-3.5	1	y_{13} -H_2O
1383.7183	1383.7155	2.1	1	y_{13}
1386.6980	1386.6787	13.9	1	y_{11}*
1435.6750	1435.6740	0.7	1	b_{14} -H_2O
1436.7368	1436.7420	-3.6	1	y_{14} -H_2O
1437.7405	1437.7260	10.1	1	y_{14} -NH_3
1453.6809	1453.6846	-2.5	1	b_{14}
1454.7539	1454.7526	0.9	1	y_{14}
1485.7531	1485.7472	4.0	1	y_{12}*
1507.7892	1507.7791	6.7	1	y_{15} -H_2O
1508.7775	1508.7631	9.5	1	y_{15} -NH_3
1525.7915	1525.7897	1.2	1	y_{15}
1536.7340	1536.7217	8.0	1	b_{15} -H_2O
1554.7372	1554.7322	3.2	1	b_{15}
1578.8134	1578.8162	-1.8	1	y_{16} -H_2O
1579.8070	1579.8003	4.3	1	y_{16} -NH_3
1586.8041	1586.7948	5.8	1	y_{13}*
1596.8306	1596.8268	2.4	1	y_{16}
1657.8325	1657.8319	0.3	1	y_{14}*
1725.8771	1725.8694	4.5	1	y_{17}
1728.8628	1728.8691	-3.6	1	y_{15}*
1782.8396	1782.8432	-2.0	1	b_{17}
1812.8967	1812.9014	-2.6	1	y_{18}
2025.0002	2024.9811	9.4	1	aglycon -H_2O
2025.9860	2025.9651	10.3	1	aglycon -NH_3
2043.0068	2042.9917	7.4	1	aglycon
2228.0610	2228.0605	0.2	1	[M -$HexNAc_2$] - H_2O
2229.0444	2229.0445	-0.1	1	[M -$HexNAc_2$] -NH_3
2246.0623	2246.0711	-3.9	1	[M -$HexNAc_2$]

APPENDIX 4B.3: Ion Assignments for Fig 4.4.

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
147.1128	147.1128	0.0	1	y_1
187.0800	187.0713	46.3	1	b_2
244.1655	244.1656	-0.3	1	y_2
316.1138	316.1139	-0.4	1	b_3
372.2621	372.2605	4.2	1	y_3

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
445.1552	445.1565	-3.0	1	b_4
469.3134	469.3133	0.2	1	y_4
546.2021	546.2042	-3.8	1	b_5
554.1898	554.2017	-21.4	2	b_8*
556.3455	556.3453	0.3	1	y_5
625.2109	625.2206	-15.5	2	b_{11}
627.3827	627.3824	0.4	1	y_6
661.2324	661.2311	1.9	1	b_6
662.2327	662.2390	-9.5	2	b_{10}*
728.4303	728.4301	0.3	1	y_7
790.2740	790.2737	0.3	1	b_7
820.3894	820.4173	-34.0	2	y_{12}**
829.4747	829.4778	-3.7	1	y_8
830.4783	830.4618	19.9	1	y_6*
897.4425	897.4418	0.8	2	y_{17}
904.3199	904.3167	3.6	1	b_8
930.5262	930.5255	0.8	1	y_9
931.5182	931.5095	9.4	1	y_7*
961.9641	961.9631	1.0	2	y_{18}
1005.3602	1005.3643	-4.1	1	b_9
1019.4717	1019.4766	-4.8	2	y_{19}
1031.5757	1031.5732	2.5	1	y_{10}
1032.5691	1032.5572	11.6	1	y_8*
1070.0049	1070.0004	4.2	2	y_{20}
1120.3936	1120.3913	2.1	1	b_{10}
1132.6230	1132.6208	1.9	1	y_{11}
1133.6145	1133.6048	8.5	1	y_9*
1134.6238	1134.5889	30.8	1	y_7**
1199.0234	1199.0430	-16.3	2	y_{22}
1227.5383	1227.5537	-12.6	2	y_{23}
1233.4164	1233.4747	-47.2	2	b_{15}****
1233.6708	1233.6685	1.9	1	y_{12}
1234.6638	1234.6525	9.1	1	y_{10}*
1235.6639	1235.6365	22.1	1	y_8**
1249.4330	1249.4339	-0.7	1	b_{11}
1261.0562	1261.0014	43.5	2	b_{20}**
1283.5702	1283.5065	49.6	2	b_{18}***
1292.0851	1292.0750	7.8	2	‡, aglycon

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
1334.7292	1334.7162	9.8	1	y_{13}
1335.7175	1335.7002	13.0	1	y_{11}*
1336.7018	1336.6842	13.2	1	y_{9}**
1350.4850	1350.4816	2.6	1	b_{12}
1385.1116	1385.0462	47.2	2	b_{18}****
1385.6013	1385.5382	45.5	2	b_{16}*****
1393.6224	1393.6147	5.5	2	‡*, [M -$HexNAc_6$]
1436.7539	1436.7479	4.2	1	y_{12}*
1437.7378	1437.7319	4.1	1	y_{10}**
1438.7268	1438.7159	7.6	1	y_{8}***
1452.4978	1452.5132	-10.6	1	b_{11}*
1463.7611	1463.7588	1.6	1	y_{14}
1495.1470	1495.1544	-5.0	2	‡**, [M -$HexNAc_5$]
1537.7791	1537.7956	-10.7	1	y_{13}*
1538.8047	1538.7796	16.3	1	y_{11}**
1539.7823	1539.7636	12.2	1	y_{9}***
1552.6135	1552.5769	23.6	1	b_{14}
1578.7858	1578.7857	0.0	1	y_{15}
1639.7932	1639.8273	-20.8	1	y_{12}**
1640.8153	1640.8113	2.5	1	y_{10}***
1666.8049	1666.8382	-19.9	1	y_{14}*
1679.8464	1679.8334	7.7	1	y_{16}
1740.8336	1740.8749	-23.7	1	y_{13}**
1741.8633	1741.8589	2.5	1	y_{11}***
1781.8101	1781.8651	-30.9	1	y_{15}*
1793.8458	1793.8763	-17.0	1	y_{17}
1882.8942	1882.9128	-9.9	1	y_{16}*
1922.9371	1922.9189	9.5	1	y_{18}

APPENDIX 4B.4: Ion Assignments for Fig 4.5.

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
147.1127	147.1128	-0.7	1	y_1
228.1342	228.1343	-0.3	1	b_2
244.1655	244.1656	-0.3	1	y_2
327.2024	327.2027	-0.9	1	b_3
359.1924	359.1925	-0.3	1	y_3

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
428.2535	428.2504	7.3	1	b_4
458.2605	458.2609	-0.9	1	y_4
555.3138	555.3137	0.2	1	y_5
628.3301	628.3301	0.1	1	b_6
631.3203	631.3297	-14.9	1	b_4*
668.3994	668.3978	2.5	1	y_6
699.3669	699.3672	-0.4	1	b_7
725.4193	725.4192	0.1	1	y_7
831.3948	831.4094	-17.6	1	b_6*
862.4299	862.4305	-0.7	1	b_8
888.4821	888.4825	-0.5	1	y_8
902.4320	902.4466	-16.1	1	b_7*
919.4526	919.4520	0.7	1	b_9
959.5201	959.5197	0.5	1	y_9
1030.5574	1030.5568	0.6	1	y_{10}
1032.5414	1032.5360	5.2	1	b_{10}
1129.5916	1129.5888	2.5	1	b_{11}
1159.6016	1159.5994	1.9	1	y_{11}
1260.6484	1260.6470	1.1	1	y_{12}
1343.6854	1343.6842	0.9	1	b_{13}
1359.7166	1359.7155	0.8	1	y_{13}
1586.8459	1586.8424	2.2	1	aglycon

APPENDIX 4B.5: Ion Assignments for Fig 4.6

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
147.1129	147.1128	0.7	1	y_1
161.0380	161.0379	0.5	1	b_2 (-palm, -myr)
248.1607	248.1605	0.9	1	y_2
268.2271	268.2276	-1.8	1	b_1
371.2361	371.2363	-0.5	1	b_2 (-palm)
377.2036	377.2031	1.4	1	y_3
399.2685	399.2676	2.3	1	b_2 (-myr)
474.2563	474.2558	1.0	1	y_4
602.3519	602.3508	1.8	1	y_5
609.4655	609.4660	-0.7	1	b_2
689.3835	689.3828	1.0	1	y_6
776.4156	776.4149	1.0	1	y_7

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
863.4476	863.4469	0.8	1	y_8
927.6149	927.6125	2.6	1	y_6 (+palm)
966.4558	966.4561	-0.3	1	y_9 (-palm)
986.5160	986.5227	-6.8	1	b_8 (-palm)
1014.6516	1014.6445	7.0	1	y_7 (+palm)
1023.4797	1023.4775	2.1	1	[M (-palm, -myr)]
1087.5614	1087.5704	-8.3	1	b_9 (-palm)
1095.7035	1095.7098	-5.7	1	b_7
1101.6763	1101.6766	-0.2	1	y_8 (+palm)
1204.6860	1204.6857	0.2	1	y_9
1224.7467	1224.7524	-4.6	1	b_8
1233.6782	1233.6759	1.9	1	[M (-palm)]
1261.7001	1261.7072	-5.6	1	[M (-myr)]

APPENDIX 4B.6: Ion Assignments for Fig 4.7

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
215.1029	215.1026	1.2	1	b_2
261.1558	261.1557	0.3	1	y_2
312.1548	312.1554	-1.9	1	b_3
332.1927	332.1928	-0.4	1	y_3
411.2356	411.2238	28.7	1	b_4
429.2459	429.2456	0.7	1	y_4
500.2832	500.2827	1.0	1	y_5
540.2582	540.2664	-15.2	1	b_5
597.2872	597.2879	-1.1	1	b_6
599.3541	599.3511	4.9	1	y_6
684.3178	684.3199	-3.1	1	b_7
700.4005	700.3988	2.4	1	y_7
787.4315	787.4308	0.8	1	y_8
888.4933	888.4785	16.6	1	y_9
974.4018	974.4313	-30.3	1	b_8*
975.5225	975.5106	12.2	1	y_{10}
1032.4451	1032.4480	-2.8	1	b_{11}
1061.4369	1061.4633	-24.9	1	b_9*
1062.5510	1062.5426	7.9	1	y_{11}
1148.4647	1148.4954	-26.7	1	b_{10}*

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
1149.5792	1149.5746	4.0	1	y_{12}
1235.4927	1235.5274	-28.1	1	b_{11}*
1236.6184	1236.6066	9.5	1	y_{13}
1322.5286	1322.5594	-23.3	1	b_{12}*
1323.6344	1323.6387	-3.2	1	y_{14}
1410.6708	1410.6707	0.1	1	y_{15}
1497.6858	1497.7166	-20.6	1	y_{9}***
1497.6858	1497.7027	-11.3	1	y_{16}
1584.7395	1584.7487	-5.8	1	y_{10}***
1584.7395	1584.7347	3.0	1	y_{17}
1671.7655	1671.7807	-9.1	1	y_{11}***
1671.7655	1671.7668	-0.8	1	y_{18}
1758.7955	1758.8127	-9.8	1	y_{12}***
1758.7955	1758.7988	-1.9	1	y_{19}
1845.7935	1845.8448	-27.8	1	y_{13}***
1845.7935	1845.8308	-20.2	1	y_{20}
1932.8513	1932.8768	-13.2	1	y_{14}***
1932.8513	1932.8629	-6.0	1	y_{21}
2019.8694	2019.9088	-19.5	1	y_{15}***
2019.8694	2019.8949	-12.6	1	y_{22}
2106.8933	2106.9408	-22.6	1	y_{16}***
2106.8933	2106.9269	-16.0	1	y_{23}
2193.9631	2193.9729	-4.5	1	y_{17}***
2193.9631	2193.9589	1.9	1	y_{24}
2280.9612	2281.0049	-19.2	1	y_{18}***
2280.9612	2280.9910	-13.1	1	y_{25}
2367.9712	2368.0508	-33.6	1	y_{12}******
2367.9712	2368.0369	-27.8	1	y_{19}***
2367.9712	2368.0230	-21.9	1	y_{26}
2424.9797	2425.0445	-26.7	1	y_{27}
2424.9797	2424.9883	-3.5	1	b_{13}******
2424.9797	2424.9744	2.2	1	b_{20}***
2426.0076	2426.0676	-24.7	1	y_{15}*****
2426.0076	2426.0536	-19.0	1	y_{22}**
2750.2222	2750.2082	5.1	1	y_{30}
2964.2405	2964.3036	-21.3	1	aglycon

APPENDIX 4B.7: Ion Assignments for Fig 4.9

Meas. *m/z*	Calc. *m/z*	Δppm	z	Annotation
147.1129	147.1128	0.7	1	y_1
226.1551	226.1550	0.4	1	b_2
248.1608	248.1605	1.3	1	y_2
327.2026	327.2027	-0.2	1	b_3
349.2083	349.2082	0.4	1	y_3
428.2478	428.2504	-6.0	1	b_4
450.2540	450.2558	-4.1	1	y_4
529.2964	529.2980	-3.1	1	b_5
551.3049	551.3035	2.5	1	y_5
630.3482	630.3457	3.9	1	b_6
652.3463	652.3512	-7.5	1	y_6
731.3980	731.3934	6.3	1	b_7
753.3929	753.3989	-7.9	1	y_7
754.3914	754.3829	11.3	1	y_5*
792.9193	792.9200	-0.9	2	M
832.4316	832.4411	-11.4	1	b_8
854.4416	854.4466	-5.8	1	y_8
933.4857	933.4888	-3.3	1	b_9
955.4847	955.4942	-10.0	1	y_9
956.4911	956.4782	13.4	1	y_7*
1056.5314	1056.5419	-9.9	1	y_{10}
1057.5306	1057.5259	4.4	1	y_8*
1135.5800	1135.5841	-3.6	1	b_{11}
1157.5768	1157.5896	-11.0	1	y_{11}
1158.5791	1158.5736	4.7	1	y_9*
1236.6078	1236.6318	-19.4	1	b_{12}
1258.6224	1258.6373	-11.8	1	y_{12}
1259.6207	1259.6213	-0.5	1	y_{10}*
1337.6754	1337.6795	-3.0	1	b_{13}
1359.6737	1359.6849	-8.3	1	y_{13}
1360.6742	1360.6690	3.8	1	y_{11}
1361.6630	1361.6530	7.4	1	y_9**
1439.7150	1439.7112	2.7	1	b_{12}*
1456.7452	1456.7377	5.1	1	y_{14}
1584.8348	1584.8327	1.3	1	aglycon

APPENDIX 5: Complete Tables of the *O*-Fucosylated Peptides and Proteins

APPENDIX 5A: *O*-Fucosylated Peptides Identified.

[a]Between square brackets are given the AA number immediately preceding and following the first and last AA in the given peptide sequence. [b]Observed neutral loss from MS spectra, are marked with a Y. In the cases where the neutral loss, *m/z* is consistent with more than one dHex, the number of dHex for which neutral loss was observed is indicated after the dash.

Gene ID	AA Sequence[a]	NL-#[b]	# dHex	Glycosylation site
TGGT1_203150	[348]SSPASNPLSSPSSPSGPGSSSASSPVHTPAAR[381]	Y	1	
TGGT1_203780	[597]LGDSGAQSLFGG**T**AAR[614]		1	T610
	[628]EIAREPPAPFSFPAVGGGAGGTAPLFSSGK[659]		1	
	[832]GASLFGGATFSAVSQPPSTNSK[855]		1	
	[1028]SSSSTGILGTGTQSQTVSSSAPPSLFVFGGGGPAK[1064]	Y	1	on SSSST
	[1036]GTGTQSQTVSSSAPPSLFVFGGGGPAK[1064]	Y	1	
	[1145]TSGTCLFVFGSTVGASQTACASSGGSVK[1174]	Y-2	2	
	[1145]TSGTCLFVFGSTVGASQTACASSGGSVKR[1175]	Y	1	
	[1145]TSGTCLFVFGSTVGASQTACASSGGSVKR[1175]	Y-1	3	
	[1187]SVFGGTTSQGASTTGGLFGVS[1209]		1	
	[1259]GA**S**PFG**T**QSSTPVFGGGTTATGSSSSLSSVFGASK[1295]		1	T1266 or S1262
	[1259]GASPFGTQSSTPVFGGGTTATGSSSSLSSVFGASK[1295]		2	
TGGT1_206450	[532]RSTPVSGASTPQAR[547]	Y	1	
TGGT1_211150	[1854]KRPFYQQAPSSSSSPSSSSSSASSSSSASSSSSASAFRPSR[1896]	Y-1	3	
	[1854]KRPFYQQAPSSSSSPSSSSSSASSSSSASSSSSASAFRPSR[1896]	Y-1	2	
TGGT1_211700	[17]SLPGTSSSLFGR[30]	Y	1	
TGGT1_216030A	[208]IRRPAPHPPSSSSFSSSSSPSSSSFSSSSSPSSSSFSSSSR[250]	Y-1	3	
	[208]IRRPAPHPPSSSSFSSSSSPSSSSFSSSSSPSSSSFSSSSR[250]	Y-1	4	
	[208]IRRPAPHPPSSSSFSSSSSPSSSSFSSSSSPSSSSFSSSSR[250]	Y-1	2	
	[210]RPAPHPPSSSSFSSSSSPSSSSFSSSSSPSSSSFSSSSR[250]	Y-1	3	
	[210]RPAPHPPSSSSFSSSSSPSSSSFSSSSSPSSSSFSSSSR[250]	Y-1	2	
TGGT1_216030A	[914A/24B]LSAVPSLSPPPPSDAASSSSASSSSASSTSSSISSSSCSASWGFLPPPEEAHR[968A/78B]	Y-1	3	
TGGT1_216030A	[914A/24B]LSAVPSLSPPPPSDAASSSSASSSSASSTSSSISSSSCSASWGFLPPPEEAHR[968A/78B]	Y-1	4	
TGGT1_218070	[1685]ARACSSPPSASSQPSVASPSSSSSTSSSSAS	Y-1	2	

Gene ID	AA Sequence[a]	NL-#[b]	# dHex	Glycosylation site
	SSSSPQLISR[1707]			
	[1687]ACSSPPSASSQPSVASPSSSSSTSSSSASSSSSPQLISR[1707]	Y-1	2	
	[1687]ACSSPPSASSQPSVASPSSSSSTSSSSASSSSSPQLISR[1707]	Y-2	3	
TGGT1_223880	[430]EAVPSTGATQPSFAPSSSVSVLLPSHNEEIK[462]	Y	1	
TGGT1_225890	[154]ASVAEPLSTTSSSSSSSAASAPASSSTR[183]	Y-1	2	
TGGT1_226080A	[4]TVASIHPLSASSASR[20]	Y	1	
TGGT1_226950	[759]KSEEEARKTGWVVPTTHLK[779]	Y-1	2	
	[64]NASSSSGCSSSPAAPTGSSLSPSSK[90]	Y	1	
TGGT1_230890	[288]GRPCSVSSSPSTAASSR[306]	Y	1	on one of S
TGGT1_230940	[1083]LLGLGHQAATAGGSQPSSFSSYSHSQQR[1112]	Y	1	on one of S
	[1083]LLGLGHQAATAGGSQPSSFSSYSHSQQR[1112]	Y-2	2	
	[1083]LLGLGHQAATAGGSQPSSFSSYSHSQQREAPVGGQSSSAR[1125]	Y	1	
TGGT1_233010	[418]SGSSSSSYTGSSSASHPTGASAQAPSHFGSSHYASSHQMHR[460]	Y	1	on SSSSS
	[588]DRPAASGSSAGVSHAYYASQFFAPGSATHSPQLSGSSR[647]	Y	1	
TGGT1_234900	[329]APSSPAALLSAALSR[345]	Y	1	
TGGT1_235550	[224]GSSSPSSPSSLSSSASSSASSSSSASASALSSAR[259]	Y-2	2	
TGGT1_244600	[258]SLSSSLSCASSSSR[273]	Y	1	
TGGT1_254490	[1465]VSPPVAAVSWSTSGSAPVSDSTGSASSER[1495]	Y	1	
TGGT1_260240	[450]QVLGASGAAPGTSSATHLR[490]	Y	1	
TGGT1_273850	[389]ATGGATGLLGSSSLFGDTK[409]		1	
TGGT1_277050	[459]LSFSPPVQADQAMCLPDGNSCGSPTSTLHPRPHASSSPAACTQTST[506]	Y	1	
TGGT1_285190	[27]PLGAPASSPANSQSSAAASGSPLR[52]	Y	1	
	[431]GQTSNMSGQSERRPSPSPSANETAGSFSSSSASSSFSPSHR[473]	Y-1	2	
TGGT1_285190	[443]RPSPSPSANE**T**AGSFSSSSASSSFSPSHR[473]	Y	1	likely T454
	[443]RPSPSPSANETAGSFSSSSASSSFSPSHR[473]	Y-2	2	
	[443]RPSPSPSANETAGSFSSSSASSSFSPSHR[473]	Y-2	3	
	[472]SSSSSASSSSSSFPSSSSSDSVPPR[498]	Y	1	
	[472]SSSSSASSSSSSFPSSSSSDSVPPR[498]	Y-1	2	
	[497]STAYSWVASGPSASSSAR[516]	Y	1	S512 or S513
	[497]STAYSWVASGPSASSSAR[516]	Y	2	
	[515]SSSSPVSAADSGGLPFSQSTASNSQSEPFQAFSFSATSQPSSSGPR[562]		1	
TGGT1_291980	[1920]NGTESALLSSLVR[1934]	Y	1	
TGGT1_292235	[1318]NLWTSSAALPSSPSSASSSSSSSPPSSSSAR[1350]		1	
TGGT1_297520	[1139]HYTPTASALLPLAVSSSQTDASGANASLSASAK[1173]		1	

Gene ID	AA Sequence[a]	NL-#[b]	# dHex	Glycosylation site
TGGT1_300180	[258]IPSQSSSQSSSQSSSQFSSQSSSQFSSSSTASK[292]	Y-1	2	
TGGT1_301410	[649]SQSAPSAASSTAASSFFSGLVVKDDKQR[678]	Y	1	
TGGT1_313430	[278]GANSSSSLFSGAGTASTGTSK[300]		1	
TGGT1_313580	[216]AATSSQLPPVAVASESHLSSSSSLSSSSLSSSSLSSSSSVLPR[260]	Y-1	3	
	[216]AATSSQLPPVAVASESHLSSSSSLSSSSLSSSSLSSSSSVLPR[260]	Y-1	4	
	[216]AATSSQLPPVAVASESHLSSSSSLSSSSLSSSSLSSSSSVLPR[260]	Y-2	5	
	[229]SESHLSSSSSLSSSSLSSSSLSSSSSVLPR[260]	Y-1	3	
	[268]SQDPPGRLPVPGASASASPSVSVSPPPERGENEKTEGARPQSGSGGVER[318]	Y	1	
TGGT1_318260	[20]ETVRPMGEPPPSDSSNSSVGDASKPLSSPSSSSASSSASLSSSSSSSLSSSSSSSSVSPVDPR[84]	Y-1	4	
TGGT1_321450	[3228]NLPSSSASSSSSASSSTASSSSSSSASSSSVASGESTETGVSSAKPAGSPLSVPLGVPGGR[3290]	Y-2	6	
TGGT1_321540	[256]SSCLSSSSSSSLSSSR[273]	Y	1	
	[814]LPSSLSSSSSSSLASSSSSSLSASR[840]	Y-1	2	

APPENDIX 5B: Serine Rich Proteins or Proteins with Observed *O*-Fucosylated Peptides.

Ser-rich proteins with no predicted signal peptide, for which glycopeptides and/or at least 10 unique peptides were identified. For the proteins with no observed glycopeptides, only the ones present in 4 out of 5 biological repeats were included. [b]The name given on ToxoDB is in plain text. When a putative function could be assigned by homology searches, an updated name is proposed in italics. [c] Y=yes; N=no. [d]Based on annotation on ToxoDB and homology searches. [e]Proteins for which we identified dHex-containing glycopeptides are marked with Y. The peptides can be found in APPENDIX 5A. [f]Given as average of In gel and in solution replicates.

Gene ID	Full name[b]	NLS[c]	Glycopetides observed[e]	Total spectra[f]	Total unique peptides[f]
TGGT1_230940	hypothetical protein	N	Y	404	94
TGGT1_203780	hypothetical protein FG repeat-containing protein	N	Y	269	86
TGGT1_285190	zinc finger, C3HC4 type (RING finger) domain-containing protein	N	Y	164	28
TGGT1_229750	hypothetical protein	Y	Y	139	43
TGGT1_216030A	hypothetical protein	N	Y	133	3
TGGT1_234230	hypothetical protein	Y		117	49
TGGT1_316650	hypothetical protein	N		113	47
TGGT1_313580	cytochrome b5 family heme/steroid binding domain-containing protein	N	Y	112	31
TGGT1_240220	hypothetical protein SMC-domain protein	Y		106	40
TGGT1_273850	hypothetical protein FG repeat-containing protein Nup68	N	Y	96	28
TGGT1_292200	RNA recognition motif (RRM) domain protein	Y		96	40
TGGT1_270770	PWI domain-containing protein RNA-binding motif (RRM) domain protein	Y		93	33
TGGT1_304650	histidine acid phosphatase superfamily protein	N		93	56
TGGT1_294630	hypothetical protein PHD domain-containing protein	N		92	30
TGGT1_234900	PHD zinc-finger domain-containing protein	Y	Y	89	47
TGGT1_291620	hypothetical protein	Y		87	41
TGGT1_254490	Sel1 repeat-containing protein TPR repeat domain-containing protein	N	Y	81	31
TGGT1_321540	hypothetical protein	Y	Y	80	34
TGGT1_233010	putative Ser-Thr kinase	N	Y	76	23
TGGT1_297520	proteophosphoglycan PPG1 SMC-domain protein	N	Y	76	39
TGGT1_257580	hypothetical protein	N		73	23
TGGT1_321450	Myb family DNA-binding domain-containing protein	Y	Y	73	32
TGGT1_223880	zinc finger, C3HC4 type (RING finger) domain-containing protein	Y	Y	64	30
TGGT1_318260	transcription initiation factor TFIID subunit TAF5	Y	Y	59	24
TGGT1_218070	hypothetical protein	Y	Y	58	16
TGGT1_300180	hypothetical protein	N	Y	57	20
TGGT1_272720	methyltransferase domain-containing protein	Y		51	26

Gene ID	Full name[b]	NLS[c]	Glycopeptides observed[e]	Total spectra[f]	Total unique peptides[f]
TGGT1_226950	hypothetical protein	N	y	50	24
TGGT1_293320	hypothetical protein	Y		47	25
TGGT1_258980	hypothetical protein RCD1 superfamily protein	N		45	9
TGGT1_310310	WD domain-containing protein	N		45	16
TGGT1_291980	HECT domain (ubiquitin-transferase)-containing protein	Y	y	41	24
TGGT1_248500	hypothetical protein NUP54-like	N		40	12
TGGT1_212820	ubiquitin family protein	N		37	20
TGGT1_207370	hypothetical protein SMC-domain protein	N		36	21
TGGT1_239410	hypothetical protein putative CCR4-associated factor NOT4	Y		36	26
TGGT1_226900	hypothetical protein	N		35	16
TGGT1_315720	Smg-4/UPF3 family protein RRM domain protein	N		34	13
TGGT1_225890	hypothetical protein putative PWWP domain-containing protein	Y	Y	33	15
TGGT1_287170	hypothetical protein	N		33	10
TGGT1_206450	putative autophagy-related cysteine peptidase atg4	Y	Y	32	17
TGGT1_210830	putative RIO1 kinase	N		32	20
TGGT1_296010	phosphatidylinositol 3- and 4-kinase	Y		32	19
TGGT1_213790	hypothetical protein Sec14p-like domain containing protein	Y		30	16
TGGT1_292235	RING Zn-finger domain-containing protein	N	Y	30	18
TGGT1_289820	TBC domain-containing protein	N		26	15
TGGT1_313430	hypothetical protein FG repeat-containing protein	Y	Y	26	13
TGGT1_211700	putative SAN3/GANP family protein	Y	Y	25	15
TGGT1_230890	PHD zinc-finger domain-containing protein	Y	Y	23	14
TGGT1_239400	hypothetical protein	N		23	14
TGGT1_285720	ATP binding domain protein GPN-loop GTPase superfamily	N		23	10
TGGT1_206540	hypothetical protein	Y		22	12
TGGT1_236240	Tyrosine kinase-like (TKL) protein	N		22	11
TGGT1_238400	endonuclease/exonuclease/phosphatase domain-containing protein inositol-5' phosphatase	Y		21	11
TGGT1_235550	PHD zinc-finger domain-containing protein	Y	Y	20	12
TGGT1_211150	hypothetical protein	Y	Y	19	14
TGGT1_202740	putative tRNA pseudouridine synthase	Y		17	11
TGGT1_260240	CCR4-associated factor family protein CAF1 family ribonuclease (polyA specific)	N	Y	17	7
TGGT1_280780	dihydrouridine synthase (dus) protein	Y		16	9
TGGT1_294730	hypothetical protein	N		16	9
TGGT1_244600	hypothetical protein	N	Y	15	4
TGGT1_267710	CPSF A subunit protein	Y		13	10
TGGT1_310610	hypothetical protein FG repeat-containing protein	N		13	8

Gene ID	Full name[b]	NLS[c]	Glycopetides observed[e]	Total spectra[f]	Total unique peptides[f]
TGGT1_226080	putative polyA polymerase	Y	Y	11	5
TGGT1_277050	hypothetical protein	Y	Y	10	7
TGGT1_203150	hypothetical protein	N	Y	10	7
TGGT1_301410	hypothetical protein	N	Y	10	8
TGGT1_218830	hypothetical protein	N		8	6
TGGT1_289310	cullin family protein	N		6	6

BIBLIOGRAPHY

1. Morgan-Ryan, U. M., Fall, A., Ward, L. A., Hijjawi, N., Sulaiman, I., Fayer, R., Thompson, R. C., Olson, M., Lal, A., and Xiao, L. (2002) Cryptosporidium Hominis N. Sp. (Apicomplexa: Cryptosporidiidae) from Homo Sapiens. *Journal of Eukaryotic Microbiology* 49, 433-440

2. Feltus, D. C., Giddings, C. W., Schneck, B. L., Monson, T., Warshauer, D., and McEvoy, J. M. (2006) Evidence Supporting Zoonotic Transmission of Cryptosporidium Spp. In Wisconsin. *Journal of Clinical Microbiology* 44, 4303-4308

3. de Graaf, D. C., Vanopdenbosch, E., Ortega-Mora, L. M., Abbassi, H., and Peeters, J. E. (1999) A Review of the Importance of Cryptosporidiosis in Farm Animals. *International Journal for Parasitology* 29, 1269-1287

4. Graczyk, T. K., Fayer, R., and Cranfield, M. R. (1997) Zoonotic Transmission of Cryptosporidium Parvum: Implications for Water-Borne Cryptosporidiosis. *Parasitology Today* 13, 348-351

5. Kiang, K., Scheftel, J., Leano, F., Taylor, C., Belle-Isle, P., Cebelinski, E., Danila, R., and Smith, K. (2006) Recurrent Outbreaks of Cryptosporidiosis Associated with Calves among Students at an Educational Farm Programme, Minnesota, 2003. *Epidemiology and Infection* 134, 878-886

6. MacKenzie, W. R., Schell, W. L., Blair, K. A., Addiss, D. G., Peterson, D. E., Hoxie, N. J., Kazmierczak, J. J., and Davis, J. P. (1995) Massive Outbreak of Waterborne Cryptosporidium Infection in Milwaukee, Wisconsin: Recurrence of Illness and Risk of Secondary Transmission. *Clinical Infectious Diseases* 21, 57-62

7. Mcanulty, J. M., Fleming, D. W., and Gonzalez, A. H. (1994) A Community-Wide Outbreak of Cryptosporidiosis Associated with Swimming at a Wave Pool. *JAMA: The Journal of the American Medical Association* 272, 1597-1600

8. Weir, S. C., Pokorny, N. J., Carreno, R. A., Trevors, J. T., and Lee, H. (2002) Efficacy of Common Laboratory Disinfectants on the Infectivity of Cryptosporidium Parvum Oocysts in Cell Culture. *Applied and Environmental Microbiology* 68, 2576-2579

9. Smith, H. V., Nichols, R. A., and Grimason, A. M. (2005) Cryptosporidium Excystation and Invasion: Getting to the Guts of the Matter. *Trends in Parasitology* 21, 133-142

10. DuPont, H. L., Chappell, C. L., Sterling, C. R., Okhuysen, P. C., Rose, J. B., and Jakubowski, W. (1995) The Infectivity of Cryptosporidium Parvum in Healthy Volunteers. *New England Journal of Medicine* 332, 855-859

11. Chappell, C. L., Okhuysen, P. C., Sterling, C. R., and DuPont, H. L. (1996) Cryptosporidium Parvum: Intensity of Infection and Oocyst Excretion Patterns in Healthy Volunteers. *Journal of Infectious Diseases* 173, 232-236

12. Okhuysen, P. C., Chappell, C. L., Crabb, J. H., Sterling, C. R., and DuPont, H. L. (1999) Virulence of Three Distinct Cryptospovidium Parvum Isolates for Healthy Adults. *Journal of Infectious Diseases* 180, 1275-1281

13. Messner, M. J., Chappell, C. L., and Okhuysen, P. C. (2001) Risk Assessment for Cryptosporidium: A Hierarchical Bayesian Analysis of Human Dose Response Data. *Water Research* 35, 3934-3940

14. Sponseller, J. K., Griffiths, J. K., and Tzipori, S. (2014) The Evolution of Respiratory Cryptosporidiosis: Evidence for Transmission by Inhalation. *Clinical Microbiology Reviews* 27, 575-586

15. Cevallos, A. M., Bhat, N., Verdon, R., Hamer, D. H., Stein, B., Tzipori, S., Pereira, M. E., Keusch, G. T., and Ward, H. D. (2000) Mediation of Cryptosporidium Parvum Infection in Vitro by Mucin-Like Glycoproteins Defined by a Neutralizing Monoclonal Antibody. *Infection and Immunity* 68, 5167-5175

16. Baskin, G. B. (1996) Cryptosporidiosis of the Conjunctiva in Siv-Infected Rhesus Monkeys. *Journal of Parasitology* 82, 630-632

17. Heine, J., Moon, H., Woodmansee, D., and Pohlenz, J. (1984) Experimental Tracheal and Conjunctival Infections with Crytosporidium Sp. In Pigs. *Veterinary Parasitology* 17, 17-25

18. Current, W. L., Reese, N. C., Ernst, J. V., Bailey, W. S., Heyman, M. B., and Weinstein, W. M. (1983) Human Cryptosporidiosis in Immunocompetent and Immunodeficient Persons. Studies of an Outbreak and Experimental Transmission. *New England Journal of Medicine* 308, 1252-1257

19. Hunter, P. R., and Nichols, G. (2002) Epidemiology and Clinical Features of Cryptosporidium Infection in Immunocompromised Patients. *Clinical Microbiology Reviews* 15, 145-154

20. Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., Wu, Y., Sow, S. O., Sur, D., Breiman, R. F., Faruque, A. S., Zaidi, A. K., Saha, D., Alonso, P. L., Tamboura, B., Sanogo, D., Onwuchekwa, U., Manna, B., Ramamurthy, T., Kanungo, S., Ochieng, J. B., Omore, R., Oundo, J. O., Hossain, A., Das, S. K., Ahmed, S., Qureshi, S., Quadri, F., Adegbola, R. A., Antonio, M., Hossain, M. J., Akinsola, A., Mandomando, I., Nhampossa, T., Acacio, S., Biswas, K., O'Reilly, C. E., Mintz, E. D., Berkeley, L. Y., Muhsen, K., Sommerfelt, H., Robins-Browne, R. M., and Levine, M. M. (2013) Burden and Aetiology of Diarrhoeal Disease in Infants and

Young Children in Developing Countries (the Global Enteric Multicenter Study, Gems): A Prospective, Case-Control Study. *Lancet* 382, 209-222

21. Fox, L. M., and Saravolatz, L. D. (2005) Nitazoxanide: A New Thiazolide Antiparasitic Agent. *Clinical Infectious Diseases* 40, 1173-1180

22. Amadi, B., Mwiya, M., Musuku, J., Watuka, A., Sianongo, S., Ayoub, A., and Kelly, P. (2002) Effect of Nitazoxanide on Morbidity and Mortality in Zambian Children with Cryptosporidiosis: A Randomised Controlled Trial. *Lancet* 360, 1375-1380

23. Amadi, B., Mwiya, M., Sianongo, S., Payne, L., Watuka, A., Katubulushi, M., and Kelly, P. (2009) High Dose Prolonged Treatment with Nitazoxanide Is Not Effective for Cryptosporidiosis in Hiv Positive Zambian Children: A Randomised Controlled Trial. *BMC Infectious Diseases* 9, 195

24. Abubakar, I., Aliyu, S. H., Arumugam, C., Usman, N. K., and Hunter, P. R. (2007) Treatment of Cryptosporidiosis in Immunocompromised Individuals: Systematic Review and Meta-Analysis. *British Journal of Clinical Pharmacology* 63, 387-393

25. Okhuysen, P. C., Chappell, C. L., Crabb, J., Valdez, L. M., Douglass, E. T., and DuPont, H. L. (1998) Prophylactic Effect of Bovine Anti-Cryptosporidium Hyperimmune Colostrum Immunoglobulin in Healthy Volunteers Challenged with Cryptosporidium Parvum. *Clinical Infectious Diseases* 26, 1324-1329

26. Tzipori, S., Roberton, D., and Chapman, C. (1986) Remission of Diarrhoea Due to Cryptosporidiosis in an Immunodeficient Child Treated with Hyperimmune Bovine Colostrum. *British Medical Journal (Clinical Research Ed.)* 293, 1276-1277

27. Ungar, B. L., Ward, D. J., Fayer, R., and Quinn, C. A. (1990) Cessation of Cryptosporidium-Associated Diarrhea in an Acquired Immunodeficiency Syndrome Patient after Treatment with Hyperimmune Bovine Colostrum. *Gastroenterology* 98, 486-489

28. Nord, J., Ma, P., DiJohn, D., Tzipori, S., and Tacket, C. O. (1990) Treatment with Bovine Hyperimmune Colostrum of Cryptosporidial Diarrhea in AIDS Patients. *AIDS* 4, 581-584

29. Greenberg, P. D., and Cello, J. P. (1996) Treatment of Severe Diarrhea Caused by Cryptosporidium Parvum with Oral Bovine Immunoglobulin Concentrate in Patients with Aids. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 13, 348-354

30. Perryman, L. E., Jasmer, D. P., Riggs, M. W., Bohnet, S. G., McGuire, T. C., and Arrowood, M. J. (1996) A Cloned Gene of Cryptosporidium Parvum Encodes Neutralization-Sensitive Epitopes. *Molecular and Biochemical Parasitology* 80, 137-147

31. Petersen, C., Gut, J., Doyle, P., Crabb, J., Nelson, R., and Leech, J. (1992) Characterization of a> 900,000-M (R) Cryptosporidium Parvum Sporozoite Glycoprotein Recognized by Protective Hyperimmune Bovine Colostral Immunoglobulin. *Infection and Immunity* 60, 5132-5138

32. Jenkins, M. C., O'Brien, C., Trout, J., Guidry, A., and Fayer, R. (1999) Hyperimmune Bovine Colostrum Specific for Recombinant Cryptosporidium Parvum Antigen Confers Partial Protection against Cryptosporidiosis in Immunosuppressed Adult Mice. *Vaccine* 17, 2453-2460

33. Barnes, D. A., Bonnin, A., Huang, J. X., Gousset, L., Wu, J., Gut, J., Doyle, P., Dubremetz, J. F., Ward, H., and Petersen, C. (1998) A Novel Multi-Domain Mucin-Like Glycoprotein of Cryptosporidium Parvum Mediates Invasion. *Molecular and Biochemical Parasitology* 96, 93-110

34. Cevallos, A. M., Zhang, X., Waldor, M. K., Jaison, S., Zhou, X., Tzipori, S., Neutra, M. R., and Ward, H. D. (2000) Molecular Cloning and Expression of a Gene Encoding Cryptosporidium Parvum Glycoproteins Gp40 and Gp15. *Infection and Immunity* 68, 4108-4116

35. Chatterjee, A., Banerjee, S., Steffen, M., O'Connor, R. M., Ward, H. D., Robbins, P. W., and Samuelson, J. (2010) Evidence for Mucin-Like Glycoproteins That Tether Sporozoites of Cryptosporidium Parvum to the Inner Surface of the Oocyst Wall. *Eukaryotic Cell* 9, 84-96

36. Luft, B. J., Payne, D., Woodmansee, D., and Kim, C. W. (1987) Characterization of the Cryptosporidium Antigens from Sporulated Oocysts of Cryptosporidium Parvum. *Infection and Immunity* 55, 2436-2441

37. Tilley, M., Upton, S. J., Fayer, R., Barta, J. R., Chrisp, C. E., Freed, P. S., Blagburn, B. L., Anderson, B. C., and Barnard, S. M. (1991) Identification of a 15-Kilodalton Surface Glycoprotein on Sporozoites of Cryptosporidium Parvum. *Infection and Immunity* 59, 1002-1007

38. Bonnin, A., Dubremetz, J. F., and Camerlynck, P. (1991) Characterization of Microneme Antigens of Cryptosporidium Parvum (Protozoa, Apicomplexa). *Infection and Immunity* 59, 1703-1708

39. O'Connor, R. M., Wanyiri, J. W., Cevallos, A. M., Priest, J. W., and Ward, H. D. (2007) Cryptosporidium Parvum Glycoprotein Gp40 Localizes to the Sporozoite Surface by Association with Gp15. *Molecular and Biochemical Parasitology* 156, 80-83

40. Priest, J. W., Kwon, J. P., Arrowood, M. J., and Lammie, P. J. (2000) Cloning of the Immunodominant 17-Kda Antigen from Cryptosporidium Parvum. *Molecular and Biochemical Parasitology* 106, 261-271

41. Priest, J. W., Xie, L. T., Arrowood, M. J., and Lammie, P. J. (2001) The Immunodominant 17-Kda Antigen from Cryptosporidium Parvum Is Glycosylphosphatidylinositol-Anchored. *Molecular and Biochemical Parasitology* 113, 117-126

42. Wanyiri, J. W., O'Connor, R., Allison, G., Kim, K., Kane, A., Qiu, J., Plaut, A. G., and Ward, H. D. (2007) Proteolytic Processing of the Cryptosporidium Glycoprotein Gp40/15 by Human Furin and by a Parasite-Derived Furin-Like Protease Activity. *Infection and Immunity* 75, 184-192

43. Tilley, M., Fayer, R., Guidry, A., Upton, S. J., and Blagburn, B. L. (1990) Cryptosporidium Parvum (Apicomplexa: Cryptosporidiidae) Oocyst and Sporozoite Antigens Recognized by Bovine Colostral Antibodies. *Infection and Immunity* 58, 2966-2971

44. Heimburg-Molinaro, J., Priest, J. W., Live, D., Boons, G. J., Song, X., Cummings, R. D., and Mead, J. R. (2013) Microarray Analysis of the Human Antibody Response to Synthetic Cryptosporidium Glycopeptides. *International Journal for Parasitology* 43, 901-907

45. Bhalchandra, S., Ludington, J., Coppens, I., and Ward, H. D. (2013) Identification and Characterization of Cryptosporidium Parvum Clec, a Novel C-Type Lectin Domain-Containing Mucin-Like Glycoprotein. *Infection and Immunity* 81, 3356-3365

46. Ungar, B. L., and Nash, T. E. (1986) Quantification of Specific Antibody Response to Cryptosporidium Antigens by Laser Densitometry. *Infection and Immunity* 53, 124-128

47. Cenci-Goga, B. T., Rossitto, P. V., Sechi, P., McCrindle, C. M., and Cullor, J. S. (2011) Toxoplasma in Animals, Food, and Humans: An Old Parasite of New Concern. *Foodborne Pathogens and Disease* 8, 751-762

48. Dubey, J. P., Lindsay, D. S., and Speer, C. A. (1998) Structures of Toxoplasma Gondii Tachyzoites, Bradyzoites, and Sporozoites and Biology and Development of Tissue Cysts. *Clinical Microbiology Reviews* 11, 267-299

49. Esch, K. J., and Petersen, C. A. (2013) Transmission and Epidemiology of Zoonotic Protozoal Diseases of Companion Animals. *Clinical Microbiology Reviews* 26, 58-85

50. Robert-Gangneux, F., and Darde, M. L. (2012) Epidemiology of and Diagnostic Strategies for Toxoplasmosis. *Clinical Microbiology Reviews* 25, 264-296

51. Garenaux, E., Shams-Eldin, H., Chirat, F., Bieker, U., Schmidt, J., Michalski, J. C., Cacan, R., Guerardel, Y., and Schwarz, R. T. (2008) The Dual Origin of Toxoplasma Gondii N-Glycans. *Biochemistry* 47, 12270-12276

52. Fauquenoy, S., Morelle, W., Hovasse, A., Bednarczyk, A., Slomianny, C., Schaeffer, C., Van Dorsselaer, A., and Tomavo, S. (2008) Proteomics and Glycomics Analyses of N-Glycosylated Structures Involved in Toxoplasma Gondii--Host Cell Interactions. *Molecular & Cellular Proteomics* 7, 891-910

53. Fauquenoy, S., Hovasse, A., Sloves, P.-J., Morelle, W., Alayi, T. D., Slomianny, C., Werkmeister, E., Schaeffer, C., Van Dorsselaer, A., and Tomavo, S. (2011) Unusual N-Glycan Structures Required for Trafficking Toxoplasma Gondii Gap50 to the Inner Membrane Complex Regulate Host Cell Entry through Parasite Motility. *Molecular & Cellular Proteomics* 10, M111. 008953

54. Marshall, R. D. (1972) Glycoproteins. *Annual Review of Biochemistry* 41, 673-702

55. Gavel, Y., and von Heijne, G. (1990) Sequence Differences between Glycosylated and Non-Glycosylated Asn-X-Thr/Ser Acceptor Sites: Implications for Protein Engineering. *Protein Engineering* 3, 433-442

56. Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., and Aebi, M. (2002) N-Linked Glycosylation in Campylobacter Jejuni and Its Functional Transfer into E. Coli. *Science* 298, 1790-1793

57. Jarrell, K. F., Ding, Y., Meyer, B. H., Albers, S. V., Kaminski, L., and Eichler, J. (2014) N-Linked Glycosylation in Archaea: A Structural, Functional, and Genetic Analysis. *Microbiology and Molecular Biology Reviews* 78, 304-341

58. Robbins, P. W., Hubbard, S. C., Turco, S. J., and Wirth, D. F. (1977) Proposal for a Common Oligosaccharide Intermediate in the Synthesis of Membrane Glycoproteins. *Cell* 12, 893-900

59. Burda, P., and Aebi, M. (1998) The Alg10 Locus of Saccharomyces Cerevisiae Encodes the A-1, 2 Glucosyltransferase of the Endoplasmic Reticulum: The Terminal Glucose of the Lipid-Linked Oligosaccharide Is Required for Efficient N-Linked Glycosylation. *Glycobiology* 8, 455-462

60. Huffaker, T. C., and Robbins, P. W. (1983) Yeast Mutants Deficient in Protein Glycosylation. *Proceedings of the National Academy of Sciences of the United States of America* 80, 7466-7470

61. Burda, P., and Aebi, M. (1999) The Dolichol Pathway of N-Linked Glycosylation. *Biochimica et Biophysica Acta* 1426, 239-257

62. Kukuruzinska, M. A., and Robbins, P. W. (1987) Protein Glycosylation in Yeast: Transcript Heterogeneity of the Alg7 Gene. *Proceedings of the National Academy of Sciences of the United States of America* 84, 2145-2149

63. Bickel, T., Lehle, L., Schwarz, M., Aebi, M., and Jakob, C. A. (2005) Biosynthesis of Lipid-Linked Oligosaccharides in Saccharomyces Cerevisiae Alg13p and Alg14p Form a Complex Required for the Formation of Glcnac2-Pp-Dolichol. *Journal of Biological Chemistry* 280, 34500-34506

64. Ratner, D. M., Cui, J., Steffen, M., Moore, L. L., Robbins, P. W., and Samuelson, J. (2008) Changes in the N-Glycome, Glycoproteins with Asn-Linked Glycans, of Giardia Lamblia with Differentiation from Trophozoites to Cysts. *Eukaryotic Cell* 7, 1930-1940

65. Couto, J. R., Huffaker, T. C., and Robbins, P. W. (1984) Cloning and Expression in Escherichia Coli of a Yeast Mannosyltransferase from the Asparagine-Linked Glycosylation Pathway. *Journal of Biological Chemistry* 259, 378-382

66. Cipollo, J. F., Trimble, R. B., Chi, J. H., Yan, Q., and Dean, N. (2001) The Yeast Alg11 Gene Specifies Addition of the Terminal A1, 2-Man to the Man5glcnac2-Pp-Dolicholn-Glycosylation Intermediate Formed on the Cytosolic Side of the Endoplasmic Reticulum. *Journal of Biological Chemistry* 276, 21828-21840

67. Abeijon, C., and Hirschberg, C. B. (1992) Topography of Glycosylation Reactions in the Endoplasmic Reticulum. *Trends in Biochemical Sciences* 17, 32-36

68. Helenius, J., Ng, D. T., Marolda, C. L., Walter, P., Valvano, M. A., and Aebi, M. (2002) Translocation of Lipid-Linked Oligosaccharides across the Er Membrane Requires Rft1 Protein. *Nature* 415, 447-450

69. Orlean, P., Albright, C., and Robbins, P. W. (1988) Cloning and Sequencing of the Yeast Gene for Dolichol Phosphate Mannose Synthase, an Essential Protein. *Journal of Biological Chemistry* 263, 17499-17507

70. Aebi, M., Gassenhuber, J., Domdey, H., and te Heesen, S. (1996) Cloning and Characterization of the Alg3 Gene of Saccharomyces Cerevisiae. *Glycobiology* 6, 439-444

71. Burda, P., Jakob, C. A., Beinhauer, J., Hegemann, J. H., and Aebi, M. (1999) Ordered Assembly of the Asymmetrically Branched Lipid-Linked Oligosaccharide in the Endoplasmic Reticulum Is Ensured by the Substrate Specificity of the Individual Glycosyltransferases. *Glycobiology* 9, 617-625

72. Burda, P., te Heesen, S., Brachat, A., Wach, A., Dusterhoft, A., and Aebi, M. (1996) Stepwise Assembly of the Lipid-Linked Oligosaccharide in the Endoplasmic Reticulum of Saccharomyces Cerevisiae: Identification of the Alg9 Gene Encoding a Putative Mannosyl Transferase. *Proceedings of the National Academy of Sciences of the United States of America* 93, 7160-7165

73. Heesen, S., Lehle, L., Weissmann, A., and Aebi, M. (1994) Isolation of the Alg5 Locus Encoding the Udp-Glucose:Dolichyl-Phosphate Glucosyltransferase from Saccharomyces Cerevisiae. *European Journal of Biochemistry* 224, 71-79

74. Reiss, G., te Heesen, S., Zimmerman, J., Robbins, P. W., and Aebi, M. (1996) Isolation of the Alg6 Locus of Saccharomyces Cerevisiae Required for Glucosylation in the N-Linked Glycosylation Pathway. *Glycobiology* 6, 493-498

75. Runge, K. W., and Robbins, P. W. (1986) A New Yeast Mutation in the Glucosylation Steps of the Asparagine-Linked Glycosylation Pathway. Formation of a Novel Asparagine-Linked Oligosaccharide Containing Two Glucose Residues. *Journal of Biological Chemistry* 261, 15582-15590

76. Kelleher, D. J., and Gilmore, R. (2006) An Evolving View of the Eukaryotic Oligosaccharyltransferase. *Glycobiology* 16, 47R-62R

77. Yan, A., and Lennarz, W. J. (2005) Unraveling the Mechanism of Protein N-Glycosylation. *Journal of Biological Chemistry* 280, 3121-3124

78. Kolhekar, A. S., Quon, A. S., Berard, C. A., Mains, R. E., and Eipper, B. A. (1998) Post-Translational N-Glycosylation of a Truncated Form of a Peptide Processing Enzyme. *Journal of Biological Chemistry* 273, 23012-23018

79. Bolt, G., Kristensen, C., and Steenstrup, T. D. (2005) Posttranslational N-Glycosylation Takes Place During the Normal Processing of Human Coagulation Factor VII. *Glycobiology* 15, 541-547

80. Ruiz-Canada, C., Kelleher, D. J., and Gilmore, R. (2009) Cotranslational and Posttranslational N-Glycosylation of Polypeptides by Distinct Mammalian Ost Isoforms. *Cell* 136, 272-283

81. Knauer, R., and Lehle, L. (1999) The Oligosaccharyltransferase Complex from Yeast. *Biochimica et Biophysica Acta* 1426, 259-273

82. Yan, Q., and Lennarz, W. J. (2002) Studies on the Function of Oligosaccharyl Transferase Subunits. Stt3p Is Directly Involved in the Glycosylation Process. *Journal of Biological Chemistry* 277, 47692-47700

83. Nilsson, I., Kelleher, D. J., Miao, Y., Shao, Y., Kreibich, G., Gilmore, R., von Heijne, G., and Johnson, A. E. (2003) Photocross-Linking of Nascent Chains to the Stt3 Subunit of the Oligosaccharyltransferase Complex. *Journal of Cell Biology* 161, 715-725

84. Wacker, M., Feldman, M. F., Callewaert, N., Kowarik, M., Clarke, B. R., Pohl, N. L., Hernandez, M., Vines, E. D., Valvano, M. A., Whitfield, C., and Aebi, M. (2006) Substrate Specificity of Bacterial Oligosaccharyltransferase Suggests a Common Transfer Mechanism for the Bacterial and Eukaryotic Systems. *Proceedings of the National Academy of Sciences of the United States of America* 103, 7088-7093

85. Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M., and Aebi, M. (2006) Definition of the Bacterial N-Glycosylation Site Consensus Sequence. *EMBO Journal* 25, 1957-1966

86. Izquierdo, L., Schulz, B. L., Rodrigues, J. A., Guther, M. L., Procter, J. B., Barton, G. J., Aebi, M., and Ferguson, M. A. (2009) Distinct Donor and Acceptor Specificities of Trypanosoma Brucei Oligosaccharyltransferases. *EMBO Journal* 28, 2650-2661

87. Manthri, S., Guther, M. L., Izquierdo, L., Acosta-Serrano, A., and Ferguson, M. A. (2008) Deletion of the Tbalg3 Gene Demonstrates Site-Specific N-Glycosylation and N-Glycan Processing in Trypanosoma Brucei. *Glycobiology* 18, 367-383

88. Jones, D. C., Mehlert, A., Guther, M. L., and Ferguson, M. A. (2005) Deletion of the Glucosidase Ii Gene in Trypanosoma Brucei Reveals Novel N-Glycosylation Mechanisms in the Biosynthesis of Variant Surface Glycoprotein. *Journal of Biological Chemistry* 280, 35929-35942

89. Silberstein, S., and Gilmore, R. (1996) Biochemistry, Molecular Biology, and Genetics of the Oligosaccharyltransferase. *FASEB Journal* 10, 849-858

90. Marcantonio, E. E., Amar-Costesec, A., and Kreibich, G. (1984) Segregation of the Polypeptide Translocation Apparatus to Regions of the Endoplasmic Reticulum Containing Ribophorins and Ribosomes. Ii. Rat Liver Microsomal Subfractions Contain Equimolar Amounts of Ribophorins and Ribosomes. *Journal of Cell Biology* 99, 2254-2259

91. Yu, Y. H., Sabatini, D. D., and Kreibich, G. (1990) Antiribophorin Antibodies Inhibit the Targeting to the Er Membrane of Ribosomes Containing Nascent Secretory Polypeptides. *Journal of Cell Biology* 111, 1335-1342

92. Johnson, A. E., and van Waes, M. A. (1999) The Translocon: A Dynamic Gateway at the Er Membrane. *Annual Review of Cell and Developmental Biology* 15, 799-842

93. Kelleher, D. J., Kreibich, G., and Gilmore, R. (1992) Oligosaccharyltransferase Activity Is Associated with a Protein Complex Composed of Ribophorins I and II and a 48 Kd Protein. *Cell* 69, 55-65

94. Makishima, T., Nakashima, T., NagataKuno, K., Fukushima, K., Iida, H., Sakaguchi, M., Ikehara, Y., Komiyama, S., and Nishimoto, T. (1997) The Highly Conserved Dad1 Protein Involved in Apoptosis Is Required for N-Linked Glycosylation. *Genes to Cells* 2, 129-141

95. Sanjay, A., Fu, J., and Kreibich, G. (1998) Dad1 Is Required for the Function and the Structural Integrity of the Oligosaccharyltransferase Complex. *Journal of Biological Chemistry* 273, 26094-26099

96. Roboti, P., and High, S. (2012) The Oligosaccharyltransferase Subunits Ost48, Dad1 and Kcp2 Function as Ubiquitous and Selective Modulators of Mammalian N-Glycosylation. *Journal of Cell Science* 125, 3474-3484

97. Michael, J. M., and Kornfeld, S. (1980) Partial Purification and Characterization of the Glucosidases Involved in the Processing of Asparagine-Linked Oligosaccharides. *Archives of Biochemistry and Biophysics* 199, 249-258

98. Trombetta, E. S., Simons, J. F., and Helenius, A. (1996) Endoplasmic Reticulum Glucosidase Ii Is Composed of a Catalytic Subunit, Conserved from Yeast to Mammals, and a Tightly Bound Noncatalytic Hdel-Containing Subunit. *Journal of Biological Chemistry* 271, 27509-27516

99. Hebert, D. N., Foellmer, B., and Helenius, A. (1995) Glucose Trimming and Reglucosylation Determine Glycoprotein Association with Calnexin in the Endoplasmic Reticulum. *Cell* 81, 425-433

100. Ellgaard, L., Molinari, M., and Helenius, A. (1999) Setting the Standards: Quality Control in the Secretory Pathway. *Science* 286, 1882-1888

101. Schroder, M., and Kaufman, R. J. (2005) The Mammalian Unfolded Protein Response. *Annual Review of Biochemistry* 74, 739-789

102. Ruddock, L. W., and Molinari, M. (2006) N-Glycan Processing in Er Quality Control. *Journal of Cell Science* 119, 4373-4380

103. Schachter, H. (2000) The Joys of Hexnac. The Synthesis and Function of N-Ando-Glycan Branches. *Glycoconjugate Journal* 17, 465-483

104. Kornfeld, R., and Kornfeld, S. (1985) Assembly of Asparagine-Linked Oligosaccharides. *Annual Review of Biochemistry* 54, 631-664

105. Helenius, A., and Aebi, M. (2004) Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annual Review of Biochemistry* 73, 1019-1049

106. Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M., and Dwek, R. A. (2007) The Impact of Glycosylation on the Biological Function and Structure of Human Immunoglobulins. *Annual Review of Immunology* 25, 21-50

107. Zhao, Y. Y., Takahashi, M., Gu, J. G., Miyoshi, E., Matsumoto, A., Kitazume, S., and Taniguchi, N. (2008) Functional Roles of N-Glycans in Cell Signaling and Cell Adhesion in Cancer. *Cancer Science* 99, 1304-1310

108. Pinho, S. S., Seruca, R., Gartner, F., Yamaguchi, Y., Gu, J., Taniguchi, N., and Reis, C. A. (2011) Modulation of E-Cadherin Function and Dysfunction by N-Glycosylation. *Cellular and Molecular Life Sciences* 68, 1011-1020

109. Metzler, M., Gertz, A., Sarkar, M., Schachter, H., Schrader, J. W., and Marth, J. D. (1994) Complex Asparagine-Linked Oligosaccharides Are Required for Morphogenic Events During Post-Implantation Development. *EMBO Journal* 13, 2056-2065

110. Samuelson, J., Banerjee, S., Magnelli, P., Cui, J., Kelleher, D. J., Gilmore, R., and Robbins, P. W. (2005) The Diversity of Dolichol-Linked Precursors to Asn-Linked Glycans Likely Results from Secondary Loss of Sets of Glycosyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* 102, 1548-1553

111. Cui, J., Smith, T., Robbins, P. W., and Samuelson, J. (2009) Darwinian Selection for Sites of Asn-Linked Glycosylation in Phylogenetically Disparate Eukaryotes and Viruses. *Proceedings of the National Academy of Sciences of the United States of America* 106, 13421-13426

112. Banerjee, S., Vishwanath, P., Cui, J., Kelleher, D. J., Gilmore, R., Robbins, P. W., and Samuelson, J. (2007) The Evolution of N-Glycan-Dependent Endoplasmic Reticulum Quality Control Factors for Glycoprotein Folding and Degradation. *Proceedings of the National Academy of Sciences of the United States of America* 104, 11676-11681

113. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2012) Control of Mucin-Type O-Glycosylation: A Classification of the Polypeptide Galnac-Transferase Gene Family. *Glycobiology* 22, 736-756

114. Schwientek, T., Bennett, E. P., Flores, C., Thacker, J., Hollmann, M., Reis, C. A., Behrens, J., Mandel, U., Keck, B., and Schäfer, M. A. (2002) Functional Conservation of Subfamilies of Putative Udp-N-Acetylgalactosamine: Polypeptide N-Acetylgalactosaminyltransferases Indrosophila, Caenorhabditis Elegans, and Mammals One Subfamily Composed of L (2) 35aa Is Essential Indrosophila. *Journal of Biological Chemistry* 277, 22623-22638

115. Julenius, K., Molgaard, A., Gupta, R., and Brunak, S. (2005) Prediction, Conservation Analysis, and Structural Characterization of Mammalian Mucin-Type O-Glycosylation Sites. *Glycobiology* 15, 153-164

116. McGuire, E. J., and Roseman, S. (1967) Enzymatic Synthesis of the Protein-Hexosamine Linkage in Sheep Submaxillary Mucin. *Journal of Biological Chemistry* 242, 3745-3747

117. Hirschberg, C. B., and Snider, M. D. (1987) Topography of Glycosylation in the Rough Endoplasmic Reticulum and Golgi Apparatus. *Annual Review of Biochemistry* 56, 63-87

118. Rottger, S., White, J., Wandall, H. H., Olivo, J. C., Stark, A., Bennett, E. P., Whitehouse, C., Berger, E. G., Clausen, H., and Nilsson, T. (1998) Localization of Three Human Polypeptide Galnac-Transferases in Hela Cells Suggests Initiation of O-Linked Glycosylation Throughout the Golgi Apparatus. *Journal of Cell Science* 111 (Pt 1), 45-60

119. Ten Hagen, K. G., Fritz, T. A., and Tabak, L. A. (2003) All in the Family: The Udp-Galnac: Polypeptide N-Acetylgalactosaminyltransferases. *Glycobiology* 13, 1R-16R

120. Clausen, H., and Bennett, E. P. (1996) A Family of Udp-Galnac: Polypeptide N-Acetylgalactosaminyl-Transferases Control the Initiation of Mucin-Type O-Linked Glycosylation. *Glycobiology* 6, 635-646

121. Wandall, H. H., Irazoqui, F., Tarp, M. A., Bennett, E. P., Mandel, U., Takeuchi, H., Kato, K., Irimura, T., Suryanarayanan, G., Hollingsworth, M. A., and Clausen, H. (2007) The Lectin Domains of Polypeptide Galnac-Transferases Exhibit Carbohydrate-Binding Specificity for Galnac: Lectin Binding to Galnac-Glycopeptide Substrates Is Required for High Density Galnac-O-Glycosylation. *Glycobiology* 17, 374-387

122. Gerken, T. A., Revoredo, L., Thome, J. J., Tabak, L. A., Vester-Christensen, M. B., Clausen, H., Gahlay, G. K., Jarvis, D. L., Johnson, R. W., and Moniz, H. A. (2013) The Lectin Domain of the Polypeptide Galnac Transferase Family of Glycosyltransferases (Ppgalnac Ts) Acts as a Switch Directing Glycopeptide Substrate Glycosylation in an N-or C-Terminal Direction, Further Controlling Mucin Type O-Glycosylation. *Journal of Biological Chemistry* 288, 19900-19914

123. Stwora-Wojczyk, M. M., Kissinger, J. C., Spitalnik, S. L., and Wojczyk, B. S. (2004) O-Glycosylation in Toxoplasma Gondii: Identification and Analysis of a Family of Udp-Galnac:Polypeptide N-Acetylgalactosaminyltransferases. *International Journal for Parasitology* 34, 309-322

124. Tomita, T., Sugi, T., Yakubu, R., Tu, V., Ma, Y., and Weiss, L. M. (2017) Making Home Sweet and Sturdy: Toxoplasma Gondii Ppgalnac-Ts Glycosylate in Hierarchical Order and Confer Cyst Wall Rigidity. *MBio* 8, e02048-02016

125. Bhat, N., Wojczyk, B. S., DeCicco, M., Castrodad, C., Spitalnik, S. L., and Ward, H. D. (2013) Identification of a Family of Four Udp-Polypeptide N-Acetylgalactosaminyl Transferases in Cryptosporidium Species. *Molecular and Biochemical Parasitology* 191, 24-27

126. Strous, G. J. (1979) Initial Glycosylation of Proteins with Acetylgalactosaminylserine Linkages. *Proceedings of the National Academy of Sciences of the United States of America* 76, 2694-2698

127. Brockhausen, I., Schachter, H., and Stanley, P. (2009) O-Galnac Glycans. In: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E., eds. *Essentials of Glycobiology*, 2nd Ed., Cold Spring Harbor Laboratory Press

The Consortium of Glycobiology Editors, La Jolla, California., Cold Spring Harbor (NY)

128. Tailford, L. E., Crost, E. H., Kavanaugh, D., and Juge, N. (2015) Mucin Glycan Foraging in the Human Gut Microbiome. *Frontiers in Genetics* 6, 81

129. Jentoft, N. (1990) Why Are Proteins O-Glycosylated? *Trends in Biochemical Sciences* 15, 291-294

130. Hilkens, J., Ligtenberg, M. J. L., Vos, H. L., and Litvinov, S. V. (1992) Cell Membrane-Associated Mucins and Their Adhesion-Modulating Property. *Trends in Biochemical Sciences* 17, 359-363

131. Perez-Vilar, J., and Hill, R. L. (1999) The Structure and Assembly of Secreted Mucins. *Journal of Biological Chemistry* 274, 31751-31754

132. Jonckheere, N., Skrypek, N., Frenois, F., and Van Seuningen, I. (2013) Membrane-Bound Mucin Modular Domains: From Structure to Function. *Biochimie* 95, 1077-1086

133. Bafna, S., Kaur, S., and Batra, S. K. (2010) Membrane-Bound Mucins: The Mechanistic Basis for Alterations in the Growth and Survival of Cancer Cells. *Oncogene* 29, 2893-2904

134. Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V., and McGuckin, M. A. (2008) Mucins in the Mucosal Barrier to Infection. *Mucosal Immunology* 1, 183-197

135. McAuley, J. L., Linden, S. K., Png, C. W., King, R. M., Pennington, H. L., Gendler, S. J., Florin, T. H., Hill, G. R., Korolik, V., and McGuckin, M. A. (2007) Muc1

Cell Surface Mucin Is a Critical Element of the Mucosal Barrier to Infection. *Journal of Clinical Investigation* 117, 2313-2324

136. Torres, C. R., and Hart, G. W. (1984) Topography and Polypeptide Distribution of Terminal N-Acetylglucosamine Residues on the Surfaces of Intact Lymphocytes. Evidence for O-Linked Glcnac. *Journal of Biological Chemistry* 259, 3308-3317

137. Holt, G. D., and Hart, G. W. (1986) The Subcellular Distribution of Terminal N-Acetylglucosamine Moieties. Localization of a Novel Protein-Saccharide Linkage, O-Linked Glcnac. *Journal of Biological Chemistry* 261, 8049-8057

138. Hanover, J. A., Cohen, C. K., Willingham, M. C., and Park, M. K. (1987) O-Linked N-Acetylglucosamine Is Attached to Proteins of the Nuclear-Pore - Evidence for Cytoplasmic and Nucleoplasmic Glycoproteins. *Journal of Biological Chemistry* 262, 9887-9894

139. Park, M. K., Donofrio, M., Willingham, M. C., and Hanover, J. A. (1987) A Monoclonal-Antibody against a Family of Nuclear-Pore Proteins (Nucleoporins) - O-Linked N-Acetylglucosamine Is Part of the Immunodeterminant. *Proceedings of the National Academy of Sciences of the United States of America* 84, 6462-6466

140. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger, R. S., Gerace, L., and Hart, G. W. (1987) Nuclear Pore Complex Glycoproteins Contain Cytoplasmically Disposed O-Linked N-Acetylglucosamine. *Journal of Cell Biology* 104, 1157-1164

141. Kreppel, L. K., Blomberg, M. A., and Hart, G. W. (1997) Dynamic Glycosylation of Nuclear and Cytosolic Proteins. Cloning and Characterization of a Unique O-Glcnac Transferase with Multiple Tetratricopeptide Repeats. *Journal of Biological Chemistry* 272, 9308-9315

142. Hart, G. W., Housley, M. P., and Slawson, C. (2007) Cycling of O-Linked Beta-N-Acetylglucosamine on Nucleocytoplasmic Proteins. *Nature* 446, 1017-1022

143. Wells, L., Vosseller, K., and Hart, G. W. (2001) Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-Glcnac. *Science* 291, 2376-2378

144. Lubas, W. A., Frank, D. W., Krause, M., and Hanover, J. A. (1997) O-Linked Glcnac Transferase Is a Conserved Nucleocytoplasmic Protein Containing Tetratricopeptide Repeats. *Journal of Biological Chemistry* 272, 9316-9324

145. Li, B., and Kohler, J. J. (2014) Glycosylation of the Nuclear Pore. *Traffic* 15, 347-361

146. Rout, M. P., and Wente, S. R. (1994) Pores for Thought: Nuclear Pore Complex Proteins. *Trends in Cell Biology* 4, 357-365

147. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L., and Rexach, M. (2003) Disorder in the Nuclear Pore Complex: The Fg Repeat Regions of Nucleoporins Are Natively Unfolded. *Proceedings of the National Academy of Sciences of the United States of America* 100, 2450-2455

148. Strawn, L. A., Shen, T., Shulga, N., Goldfarb, D. S., and Wente, S. R. (2004) Minimal Nuclear Pore Complexes Define Fg Repeat Domains Essential for Transport. *Nature Cell Biology* 6, 197-206

149. Wente, S. R. (2000) Gatekeepers of the Nucleus. *Science* 288, 1374-1377

150. Chook, Y. M., and Blobel, G. (2001) Karyopherins and Nuclear Import. *Current Opinion in Structural Biology* 11, 703-715

151. Seydel, U., and Gerace, L. (1991) A 28,000-Da Gdp/Gtp-Binding Protein Specific to the Nuclear Envelope. *Journal of Biological Chemistry* 266, 7602-7608

152. Melchior, F., Paschal, B., Evans, J., and Gerace, L. (1993) Inhibition of Nuclear-Protein Import by Nonhydrolyzable Analogs of Gtp and Identification of the Small Gtpase Ran/Tc4 as an Essential Transport Factor. *Journal of Cell Biology* 123, 1649-1659

153. Wennerberg, K., Rossman, K. L., and Der, C. J. (2005) The Ras Superfamily at a Glance. *Journal of Cell Science* 118, 843-846

154. Pemberton, L. F., Blobel, G., and Rosenblum, J. S. (1998) Transport Routes through the Nuclear Pore Complex. *Current Opinion in Cell Biology* 10, 392-399

155. Vetter, I. R., Arndt, A., Kutay, U., Gorlich, D., and Wittinghofer, A. (1999) Structural View of the Ran-Importin Beta Interaction at 2.3 Angstrom Resolution. *Cell* 97, 635-646

156. Andrade, M. A., and Bork, P. (1995) Heat Repeats in the Huntington's Disease Protein. *Nature Genetics* 11, 115-116

157. Bayliss, R., Littlewood, T., and Stewart, M. (2000) Structural Basis for the Interaction between Fxfg Nucleoporin Repeats and Importin-Beta in Nuclear Trafficking. *Cell* 102, 99-108

158. Banerjee, S., Robbins, P. W., and Samuelson, J. (2009) Molecular Characterization of Nucleocytosolic O-Glcnac Transferases of Giardia Lamblia and Cryptosporidium Parvum. *Glycobiology* 19, 331-336

159. Dieckmann-Schuppert, A., Bause, E., and Schwarz, R. T. (1994) Glycosylation Reactions in Plasmodium Falciparum, Toxoplasma Gondii and Trypanosoma Brucei

Brucei Probed by the Use of Synthetic Peptides. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1199, 37-44

160. Zentella, R., Sui, N., Barnhill, B., Hsieh, W. P., Hu, J., Shabanowitz, J., Boyce, M., Olszewski, N. E., Zhou, P., Hunt, D. F., and Sun, T. P. (2017) The Arabidopsis O-Fucosyltransferase Spindly Activates Nuclear Growth Repressor Della. *Nature Chemical Biology* 13, 479-485

161. Hartweck, L. M., Scott, C. L., and Olszewski, N. E. (2002) Two O-Linked N-Acetylglucosamine Transferase Genes of Arabidopsis Thaliana L. Heynh. Have Overlapping Functions Necessary for Gamete and Seed Development. *Genetics* 161, 1279-1291

162. Swain, S. M., Tseng, T. S., Thornton, T. M., Gopalraj, M., and Olszewski, N. E. (2002) Spindly Is a Nuclear-Localized Repressor of Gibberellin Signal Transduction Expressed Throughout the Plant. *Plant Physiology* 129, 605-615

163. Jacobsen, S. E., Binkowski, K. A., and Olszewski, N. E. (1996) Spindly, a Tetratricopeptide Repeat Protein Involved in Gibberellin Signal Transduction in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* 93, 9292-9296

164. Hedden, P., and Phillips, A. L. (2000) Gibberellin Metabolism: New Insights Revealed by the Genes. *Trends in Plant Science* 5, 523-530

165. Hedden, P., and Thomas, S. G. (2012) Gibberellin Biosynthesis and Its Regulation. *Biochemical Journal* 444, 11-25

166. Sato, S. (2011) The Apicomplexan Plastid and Its Evolution. *Cellular and Molecular Life Sciences* 68, 1285-1296

167. Kohler, S. (2005) Multi-Membrane-Bound Structures of Apicomplexa: I. The Architecture of the Toxoplasma Gondii Apicoplast. *Parasitology Research* 96, 258-272

168. Andrabi, S. B., Tahara, M., Matsubara, R., Toyama, T., Aonuma, H., Sakakibara, H., Suematsu, M., Tanabe, K., Nozaki, T., and Nagamune, K. (2017) Plant Hormone Cytokinins Control Cell Cycle Progression and Plastid Replication in Apicomplexan Parasites. *Parasitology International*

169. Kentzer, E. J., Buko, A., Menon, G., and Sarin, V. K. (1990) Carbohydrate Composition and Presence of a Fucose-Protein Linkage in Recombinant Human Pro-Urokinase. *Biochemical and Biophysical Research Communications* 171, 401-406

170. Harris, R. J., van Halbeek, H., Glushka, J., Basa, L. J., Ling, V. T., Smith, K. J., and Spellman, M. W. (1993) Identification and Structural Analysis of the Tetrasaccharide

Neuac Alpha(2-->6)Gal Beta(1-->4)Glcnac Beta(1-->3)Fuc Alpha 1-->O-Linked to Serine 61 of Human Factor Ix. *Biochemistry* 32, 6539-6547

171. Nakakura, N., Hietter, H., Van Dorsselaer, A., and Luu, B. (1992) Isolation and Structural Determination of Three Peptides from the Insect Locusta Migratoria. Identification of a Deoxyhexose-Linked Peptide. *European Journal of Biochemistry* 204, 147-153

172. Wang, Y., and Spellman, M. W. (1998) Purification and Characterization of a Gdp-Fucose:Polypeptide Fucosyltransferase from Chinese Hamster Ovary Cells. *Journal of Biological Chemistry* 273, 8112-8118

173. Luo, Y., Nita-Lazar, A., and Haltiwanger, R. S. (2006) Two Distinct Pathways for O-Fucosylation of Epidermal Growth Factor-Like or Thrombospondin Type 1 Repeats. *Journal of Biological Chemistry* 281, 9385-9392

174. Chen, C. I., Keusch, J. J., Klein, D., Hess, D., Hofsteenge, J., and Gut, H. (2012) Structure of Human Pofut2: Insights into Thrombospondin Type 1 Repeat Fold and O-Fucosylation. *EMBO Journal* 31, 3183-3197

175. Gruber, P. J., Sweeney, K. A., and Frederick, S. E. (1988) The Detection of Fucose Residues in Plant Nuclear Envelopes. *Planta* 174, 298-304

176. Gstaiger, M., and Aebersold, R. (2009) Applying Mass Spectrometry-Based Proteomics to Genetics, Genomics and Network Biology. *Nature Reviews: Genetics* 10, 617-627

177. Aebersold, R., and Mann, M. (2003) Mass Spectrometry-Based Proteomics. *Nature* 422, 198-207

178. Zaia, J. (2008) Mass Spectrometry and the Emerging Field of Glycomics. *Chemistry and Biology* 15, 881-892

179. Thomson, J. J. (1913) Bakerian Lecture: Rays of Positive Electricity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 89, 1-20

180. Brenton, A. G., and Godfrey, A. R. (2010) Accurate Mass Measurement: Terminology and Treatment of Data. *Journal of the American Society for Mass Spectrometry* 21, 1821-1835

181. Bristow, A. W. (2006) Accurate Mass Measurement for the Determination of Elemental Formula--a Tutorial. *Mass Spectrometry Reviews* 25, 99-111

182. Qi, Y., Barrow, M. P., Li, H., Meier, J. E., Van Orden, S. L., Thompson, C. J., and O'Connor, P. B. (2012) Absorption-Mode: The Next Generation of Fourier Transform Mass Spectra. *Analytical Chemistry* 84, 2923-2929

183. Hendrickson, C. L., Quinn, J. P., Kaiser, N. K., Smith, D. F., Blakney, G. T., Chen, T., Marshall, A. G., Weisbrod, C. R., and Beu, S. C. (2015) 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *Journal of the American Society for Mass Spectrometry* 26, 1626-1632

184. Shaw, J. B., Lin, T. Y., Leach, F. E., 3rd, Tolmachev, A. V., Tolic, N., Robinson, E. W., Koppenaal, D. W., and Pasa-Tolic, L. (2016) 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox. *Journal of the American Society for Mass Spectrometry* 27, 1929-1936

185. Marshall, A. G., and Hendrickson, C. L. (2008) High-Resolution Mass Spectrometers. *Annual Review of Analytical Chemistry* 1, 579-599

186. Mann, M., and Kelleher, N. L. (2008) Precision Proteomics: The Case for High Resolution and High Mass Accuracy. *Proceedings of the National Academy of Sciences of the United States of America* 105, 18132-18138

187. Strohalm, M., Kavan, D., Novak, P., Volny, M., and Havlicek, V. (2010) Mmass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Analytical Chemistry* 82, 4648-4651

188. Knochenmuss, R., Dubois, F., Dale, M. J., and Zenobi, R. (1996) The Matrix Suppression Effect and Ionization Mechanisms in Matrix-Assisted Laser Desorption/Ionization. *Rapid Communications in Mass Spectrometry* 10, 871-877

189. Muller, C., Schafer, P., Stortzel, M., Vogt, S., and Weinmann, W. (2002) Ion Suppression Effects in Liquid Chromatography-Electrospray-Ionisation Transport-Region Collision Induced Dissociation Mass Spectrometry with Different Serum Extraction Methods for Systematic Toxicological Analysis with Mass Spectra Libraries. *Journal of Chromatography. B: Analytical Technologies in the Biomedical and Life Sciences* 773, 47-52

190. Karas, M., and Hillenkamp, F. (1988) Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Analytical Chemistry* 60, 2299-2301

191. Hillenkamp, F., and Karas, M. (1990) [12] Mass Spectrometry of Peptides and Proteins by Matrix-Assisted Ultraviolet Laser Desorption/Ionization. *Methods in Enzymology* 193, 280-295

192. Harvey, D. J. (2003) Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Carbohydrates and Glycoconjugates. *International Journal of Mass Spectrometry* 226, 1-35

193. Kussmann, M., Nordhoff, E., RahbekNielsen, H., Haebel, S., RosselLarsen, M., Jakobsen, L., Gobom, J., Mirgorodskaya, E., KrollKristensen, A., Palm, L., and Roepstorff, P. (1997) Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Sample Preparation Techniques Designed for Various Peptide and Protein Analytes. *Journal of Mass Spectrometry* 32, 593-601

194. Fitzgerald, M. C., and Smith, L. M. (1995) Mass Spectrometry of Nucleic Acids: The Promise of Matrix-Assisted Laser Desorption-Ionization (Maldi) Mass Spectrometry. *Annual Review of Biophysics and Biomolecular Structure* 24, 117-140

195. Schiller, J., Arnhold, J., Benard, S., Muller, M., Reichl, S., and Arnold, K. (1999) Lipid Analysis by Matrix-Assisted Laser Desorption and Ionization Mass Spectrometry: A Methodological Approach. *Analytical Biochemistry* 267, 46-56

196. Ryzhov, V., and Fenselau, C. (2001) Characterization of the Protein Subset Desorbed by Maldi from Whole Bacterial Cells. *Analytical Chemistry* 73, 746-750

197. Magnuson, M. L., Owens, J. H., and Kelty, C. A. (2000) Characterization of Cryptosporidium Parvum by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. *Applied and Environmental Microbiology* 66, 4720-4724

198. Glassmeyer, S. T., Ware, M. W., Schaefer, F. W., 3rd, Shoemaker, J. A., and Kryak, D. D. (2007) An Improved Method for the Analysis of Cryptosporidium Parvum Oocysts by Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry. *Journal of Eukaryotic Microbiology* 54, 479-481

199. Fenselau, C., and Demirev, P. A. (2001) Characterization of Intact Microorganisms by Maldi Mass Spectrometry. *Mass Spectrometry Reviews* 20, 157-171

200. Bizzini, A., and Greub, G. (2010) Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry, a Revolution in Clinical Microbial Identification. *Clinical Microbiology and Infection* 16, 1614-1619

201. Giebel, R., Worden, C., Rust, S. M., Kleinheinz, G. T., Robbins, M., and Sandrin, T. R. (2010) Microbial Fingerprinting Using Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (Maldi-Tof Ms): Applications and Challenges. *Advances in Applied Microbiology,* 71, 149-184

202. Colquhoun, D. R., Schwab, K. J., Cole, R. N., and Halden, R. U. (2006) Detection of Norovirus Capsid Protein in Authentic Standards and in Stool Extracts by Matrix-

Assisted Laser Desorption Ionization and Nanospray Mass Spectrometry. *Applied and Environmental Microbiology* 72, 2749-2755

203. Calderaro, A., Arcangeletti, M. C., Rodighiero, I., Buttrini, M., Gorrini, C., Motta, F., Germini, D., Medici, M. C., Chezzi, C., and De Conto, F. (2014) Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (Maldi-Tof) Mass Spectrometry Applied to Virus Identification. *Scientific Reports* 4, 6803

204. Hartmann, E. M., Colquhoun, D. R., Schwab, K. J., and Halden, R. U. (2015) Absolute Quantification of Norovirus Capsid Protein in Food, Water, and Soil Using Synthetic Peptides with Electrospray and Maldi Mass Spectrometry. *Journal of Hazardous Materials* 286, 525-532

205. Demirev, P. A., Feldman, A. B., Kowalski, P., and Lin, J. S. (2005) Top-Down Proteomics for Rapid Identification of Intact Microorganisms. *Analytical Chemistry* 77, 7455-7461

206. Zhang, J., Ogorzalek Loo, R. R., and Loo, J. A. (2015) Increasing Fragmentation of Disulfide-Bonded Proteins for Top-Down Mass Spectrometry by Supercharging. *International Journal of Mass Spectrometry* 377, 546-556

207. Theberge, R., Dikler, S., Heckendorf, C., Chui, D. H., Costello, C. E., and McComb, M. E. (2015) Maldi-Isd Mass Spectrometry Analysis of Hemoglobin Variants: A Top-Down Approach to the Characterization of Hemoglobinopathies. *Journal of the American Society for Mass Spectrometry* 26, 1299-1310

208. Geusic, J., Marcos, H., and Van Uitert, L. (1964) Laser Oscillations in Nd-Doped Yttrium Aluminum, Yttrium Gallium and Gadolinium Garnets. *Applied Physics Letters* 4, 182-184

209. Karas, M., Bachmann, D., and Hillenkamp, F. (1985) Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules. *Analytical Chemistry* 57, 2935-2939

210. Knochenmuss, R. (2006) Ion Formation Mechanisms in Uv-Maldi. *Analyst* 131, 966-986

211. Cohen, S. L., and Chait, B. T. (1996) Influence of Matrix Solution Conditions on the Maldi-Ms Analysis of Peptides and Proteins. *Analytical Chemistry* 68, 31-37

212. Strupat, K., Karas, M., and Hillenkamp, F. (1991) 2,5-Dihydroxybenzoic Acid - a New Matrix for Laser Desorption Ionization Mass-Spectrometry. *International Journal of Mass Spectrometry and Ion Processes* 111, 89-102

213. Zenobi, R., and Knochenmuss, R. (1998) Ion Formation in Maldi Mass Spectrometry. *Mass Spectrometry Reviews* 17, 337-366

214. Karas, M., Gluckmann, M., and Schafer, J. (2000) Ionization in Matrix-Assisted Laser Desorption/Ionization: Singly Charged Molecular Ions Are the Lucky Survivors. *Journal of Mass Spectrometry* 35, 1-12

215. Karas, M., and Kruger, R. (2003) Ion Formation in Maldi: The Cluster Ionization Mechanism. *Chemical Reviews* 103, 427-440

216. Menzel, C., Berkenkamp, S., and Hillenkamp, F. (1999) Infrared Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry with a Transversely Excited Atmospheric Pressure Carbon Dioxide Laser at 10.6 Mu M Wavelength with Static and Delayed Ion Extraction. *Rapid Communications in Mass Spectrometry* 13, 26-32

217. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* 246, 64-71

218. Alexandrov, M. L., Gall, L. N., Krasnov, N. V., Nikolaev, V. I., Pavlenko, V. A., and Shkurov, V. A. (2008) Extraction of Ions from Solutions under Atmospheric Pressure as a Method for Mass Spectrometric Analysis of Bioorganic Compounds. *Rapid Communications in Mass Spectrometry* 22, 267-270

219. Smith, R. D., Loo, J. A., Loo, R. R. O., Busman, M., and Udseth, H. R. (1991) Principles and Practice of Electrospray Ionization? Mass Spectrometry for Large Polypeptides and Proteins. *Mass Spectrometry Reviews* 10, 359-452

220. Gaskell, S. J. (1997) Electrospray: Principles and Practice. *Journal of Mass Spectrometry* 32, 677-688

221. Winger, B. E., Light-Wahl, K. J., Ogorzalek Loo, R. R., Udseth, H. R., and Smith, R. D. (1993) Observation and Implications of High Mass-to-Charge Ratio Ions from Electrospray Ionization Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* 4, 536-545

222. Li, L., Wang, A. P. L., and Coulson, L. D. (1993) Continuous-Flow Matrix-Assisted Laser Desorption Ionization Mass-Spectrometry. *Analytical Chemistry* 65, 493-495

223. Little, D. P., Cornish, T. J., ODonnell, M. J., Braun, A., Cotter, R. J., and Koster, H. (1997) Maldi on a Chip: Analysis of Arrays of Low Femtomole to Subfemtomole Quantities of Synthetic Oligonucleotides and DNA Diagnostic Products Dispensed by a Piezoelectric Pipet. *Analytical Chemistry* 69, 4540-4546

224. Hsieh, S., Dreisewerd, K., van der Schors, R. C., Jimenez, C. R., Stahl-Zeng, J., Hillenkamp, F., Jorgenson, J. W., Geraerts, W. P., and Li, K. W. (1998) Separation and Identification of Peptides in Single Neurons by Microcolumn Liquid Chromatography-Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry and Postsource Decay Analysis. *Analytical Chemistry* 70, 1847-1852

225. Beavis, R. C., and Chait, B. T. (1990) Rapid, Sensitive Analysis of Protein Mixtures by Mass Spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 87, 6873-6877

226. Berkenkamp, S., Menzel, C., Karas, M., and Hillenkamp, F. (1997) Performance of Infrared Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry with Lasers Emitting in the 3 Mu M Wavelength Range. *Rapid Communications in Mass Spectrometry* 11, 1399-1406

227. Cramer, R., Hillenkamp, F., and Haglund, R. F. (1996) Infrared Matrix-Assisted Laser Desorption and Ionization by Using a Tunable Mid-Infrared Free-Electron Laser. *Journal of the American Society for Mass Spectrometry* 7, 1187-1193

228. Nordhoff, E., Ingendoh, A., Cramer, R., Overberg, A., Stahl, B., Karas, M., Hillenkamp, F., and Crain, P. F. (1992) Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Nucleic Acids with Wavelengths in the Ultraviolet and Infrared. *Rapid Communications in Mass Spectrometry* 6, 771-776

229. Wilm, M., and Mann, M. (1996) Analytical Properties of the Nanoelectrospray Ion Source. *Analytical Chemistry* 68, 1-8

230. Juraschek, R., Dulcks, T., and Karas, M. (1999) Nanoelectrospray - More Than Just a Minimized-Flow Electrospray Ionization Source. *Journal of the American Society for Mass Spectrometry* 10, 300-308

231. Tang, K., Page, J. S., and Smith, R. D. (2004) Charge Competition and the Linear Dynamic Range of Detection in Electrospray Ionization Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* 15, 1416-1423

232. Schmidt, A., Karas, M., and Dülcks, T. (2003) Effect of Different Solution Flow Rates on Analyte Ion Signals in Nano-Esi Ms, Or: When Does Esi Turn into Nano-Esi? *Journal of the American Society for Mass Spectrometry* 14, 492-500

233. Gangl, E. T., Annan, M., Spooner, N., and Vouros, P. (2001) Reduction of Signal Suppression Effects in Esi-Ms Using a Nanosplitting Device. *Analytical Chemistry* 73, 5635-5644

234. Natsume, T., Yamauchi, Y., Nakayama, H., Shinkawa, T., Yanagida, M., Takahashi, N., and Isobe, T. (2002) A Direct Nanoflow Liquid Chromatography-Tandem

Mass Spectrometry System for Interaction Proteomics. *Analytical Chemistry* 74, 4725-4733

235. Shen, Y. F., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., and Smith, R. D. (2002) High-Efficiency Nanoscale Liquid Chromatography Coupled on-Line with Mass Spectrometry Using Nanoelectrospray Ionization for Proteomics. *Analytical Chemistry* 74, 4235-4249

236. Churchwell, M. I., Twaddle, N. C., Meeker, L. R., and Doerge, D. R. (2005) Improving Lc-Ms Sensitivity through Increases in Chromatographic Performance: Comparisons of Uplc-Es/Ms/Ms to Hplc-Es/Ms/Ms. *Journal of Chromatography. B: Analytical Technologies in the Biomedical and Life Sciences* 825, 134-143

237. Novakova, L., Matysova, L., and Solich, P. (2006) Advantages of Application of Uplc in Pharmaceutical Analysis. *Talanta* 68, 908-918

238. Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and Zubarev, R. A. (2013) Rapid and Deep Human Proteome Analysis by Single-Dimension Shotgun Proteomics. *Molecular & Cellular Proteomics* 12, 3330-3338

239. Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., and Coon, J. J. (2014) The One Hour Yeast Proteome. *Molecular & Cellular Proteomics* 13, 339-347

240. Van Eeckhaut, A., Lanckmans, K., Sarre, S., Smolders, I., and Michotte, Y. (2009) Validation of Bioanalytical Lc-Ms/Ms Assays: Evaluation of Matrix Effects. *Journal of Chromatography. B: Analytical Technologies in the Biomedical and Life Sciences* 877, 2198-2207

241. Hagstrum, H. D. (1953) Instrumentation and Experimental Procedure for Studies of Electron Ejection by Ions and Ionization by Electron Impact. *Review of Scientific Instruments* 24, 1122-1142

242. Busch, K. L. (2006) Mass Spectrometry Forum - Electron Ionization Sources: The Basics. *Spectroscopy* 21, 14-18

243. Smyth, H. (1931) Products and Processes of Ionization by Low Speed Electrons. *Reviews of Modern Physics* 3, 347

244. Hwang, W., Kim, Y. K., and Rudd, M. E. (1996) New Model for Electron-Impact Ionization Cross Sections of Molecules. *The Journal of Chemical Physics* 104, 2956-2966

245. Dorman, F., and Morrison, J. (1961) Double and Triple Ionization in Molecules Induced by Electron Impact. *The Journal of Chemical Physics* 35, 575-581

246. Evershed, R. P., Prescott, M. C., Spooner, N., and Goad, L. J. (1989) Negative Ion Ammonia Chemical Ionization and Electron Impact Ionization Mass Spectrometric Analysis of Steryl Fatty Acyl Esters. *Steroids* 53, 285-309

247. Dodds, E. D., McCoy, M. R., Rea, L. D., and Kennish, J. M. (2005) Gas Chromatographic Quantification of Fatty Acid Methyl Esters: Flame Ionization Detection Vs. Electron Impact Mass Spectrometry. *Lipids* 40, 419-428

248. Garteiz, D. A., and Walle, T. (1972) Electron-Impact Fragmentation Studies of Beta-Blocking Drugs and Their Metabolites by Gc-Mass Spectroscopy. *Journal of Pharmaceutical Sciences* 61, 1728-&

249. Vestal, M. L. (2001) Methods of Ion Generation. *Chemical Reviews* 101, 361-375

250. Schummer, C., Delhomme, O., Appenzeller, B. M., Wennig, R., and Millet, M. (2009) Comparison of Mtbstfa and Bstfa in Derivatization Reactions of Polar Compounds Prior to GC/Ms Analysis. *Talanta* 77, 1473-1482

251. Ruiz-Matute, A. I., Hernandez-Hernandez, O., Rodriguez-Sanchez, S., Sanz, M. L., and Martinez-Castro, I. (2011) Derivatization of Carbohydrates for GC and GC-Ms Analyses. *Journal of Chromatography. B: Analytical Technologies in the Biomedical and Life Sciences* 879, 1226-1240

252. Schwartz, J. C., Senko, M. W., and Syka, J. E. (2002) A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer. *Journal of the American Society for Mass Spectrometry* 13, 659-669

253. Paul, W., and Steinwedel, H. (1953) Notizen: Ein Neues Massenspektrometer Ohne Magnetfeld. *Zeitschrift für Naturforschung A* 8, 448-450

254. Prestage, J. D., Dick, G. J., and Maleki, L. (1989) New Ion Trap for Frequency Standard Applications. *Journal of Applied Physics* 66, 1013-1017

255. March, R. E., and Todd, J. F. (2005) *Quadrupole Ion Trap Mass Spectrometry*, John Wiley & Sons

256. Pitteri, S. J., Chrisman, P. A., Hogan, J. M., and McLuckey, S. A. (2005) Electron Transfer Ion/Ion Reactions in a Three-Dimensional Quadrupole Ion Trap: Reactions of Doubly and Triply Protonated Peptides with So2*. *Analytical Chemistry* 77, 1831-1839

257. Richard, A. (2006) The 3d Quadrupole Ion Trap Mass Spectrometer as a Complete Chemical Laboratory for Fundamental Gas-Phase Studies of Metal Mediated Chemistry. *Chemical Communications*, 1469-1481

258. Hartmer, R., Kaplan, D. A., Gebhardt, C. R., Ledertheil, T., and Brekenfeld, A. (2008) Multiple Ion/Ion Reactions in the 3d Ion Trap: Selective Reagent Anion Production for Etd and Ptr from a Single Compound. *International Journal of Mass Spectrometry* 276, 82-90

259. Li, P., and Jackson, G. P. (2017) Charge Transfer Dissociation (Ctd) Mass Spectrometry of Peptide Cations: Study of Charge State Effects and Side-Chain Losses. *Journal of the American Society for Mass Spectrometry*, 1-11

260. Wong, P. S., and Graham Cooks, R. (1997) Ion Trap Mass Spectrometry. *Current Separations* 16, 85-92

261. Mamyrin, B., Karataev, V., Shmikk, D., and Zagulin, V. (1973) The Mass-Reflectron, a New Nonmagnetic Time-of-Flight Mass Spectrometer with High Resolution. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki* 64, 82-89

262. Doroshenko, V. M., and Cotter, R. J. (1999) Ideal Velocity Focusing in a Reflectron Time-of-Flight Mass Spectrometer. *Journal of the American Society for Mass Spectrometry* 10, 992-999

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

264. Edmondson, R. D., and Russell, D. H. (1996) Evaluation of Matrix-Assisted Laser Desorption Ionization-Time-of-Flight Mass Measurement Accuracy by Using Delayed Extraction. *Journal of the American Society for Mass Spectrometry* 7, 995-1001

265. Barbacci, D. C., Edmondson, R. D., and Russell, D. H. (1997) Evaluation of the Variables That Affect Resolution in Delayed Extraction Maldi-Tof. *International Journal of Mass Spectrometry* 165, 221-235

266. Takach, E. J., Hines, W. M., Patterson, D. H., Juhasz, P., Falick, A. M., Vestal, M. L., and Martin, S. A. (1997) Accurate Mass Measurements Using Maldi-Tof with Delayed Extraction. *Journal of Protein Chemistry* 16, 363-369

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

268. Dirichlet, P. G. L. (2008) Sur La Convergence Des S\'Eries Trigonom\'Etriques Qui Servent\a Repr\'Esenter Une Fonction Arbitraire Entre Des Limites Donn\'Ees. *arXiv preprint arXiv:0806.1294*

269. Comisarow, M. B., and Marshall, A. G. (1974) Fourier-Transform Ion-Cyclotron Resonance Spectroscopy. *Chemical Physics Letters* 25, 282-283

270. Comisarow, M. B., and Marshall, A. G. (1974) Frequency-Sweep Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chemical Physics Letters* 26, 489-490

271. Flora, J. W., and Muddiman, D. C. (2004) Determination of the Relative Energies of Activation for the Dissociation of Aromatic Versus Aliphatic Phosphopeptides by Esi-Fticr-Ms and Irmpd. *Journal of the American Society for Mass Spectrometry* 15, 121-127

272. Shaw, J. B., Robinson, E. W., and Pasa-Tolic, L. (2016) Vacuum Ultraviolet Photodissociation and Fourier Transform-Ion Cyclotron Resonance (Ft-Icr) Mass Spectrometry: Revisited. *Analytical Chemistry* 88, 3019-3023

273. Kingdon, K. (1923) A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review* 21, 408

274. Makarov, A. (2000) Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry* 72, 1156-1162

275. Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005) The Orbitrap: A New Mass Spectrometer. *Journal of Mass Spectrometry* 40, 430-443

276. Kelstrup, C. D., Jersie-Christensen, R. R., Batth, T. S., Arrey, T. N., Kuehn, A., Kellmann, M., and Olsen, J. V. (2014) Rapid and Deep Proteomes by Faster Sequencing on a Benchtop Quadrupole Ultra-High-Field Orbitrap Mass Spectrometer. *Journal of Proteome Research* 13, 6187-6195

277. Gabelica, V., and De Pauw, E. (2005) Internal Energy and Fragmentation of Ions Produced in Electrospray Sources. *Mass Spectrometry Reviews* 24, 566-587

278. Kaufmann, R., Chaurand, P., Kirsch, D., and Spengler, B. (1996) Post-Source Decay and Delayed Extraction in Matrix-Assisted Laser Desorption/Ionization-Reflectron Time-of-Flight Mass Spectrometry. Are There Trade-Offs? *Rapid Communications in Mass Spectrometry* 10, 1199-1208

279. Loo, J. A., Udseth, H. R., Smith, R. D., and Futrell, J. (1988) Collisional Effects on the Charge Distribution of Ions from Large Molecules, Formed by Electrospray-Ionization Mass Spectrometry. *Rapid Communications in Mass Spectrometry* 2, 207-210

280. Zhai, H., Han, X., Breuker, K., and McLafferty, F. W. (2005) Consecutive Ion Activation for Top Down Mass Spectrometry: Improved Protein Sequencing by Nozzle-Skimmer Dissociation. *Analytical Chemistry* 77, 5777-5784

281. Sachon, E., Clodic, G., Blasco, T., Jacquot, Y., and Bolbach, G. (2009) In-Source Fragmentation of Very Labile Peptides in Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Analytical Chemistry* 81, 8986-8992

282. Yost, R. A., Enke, C. G., Mcgilvery, D. C., Smith, D., and Morrison, J. D. (1979) High-Efficiency Collision-Induced Dissociation in an Rf-Only Quadrupole. *International Journal of Mass Spectrometry and Ion Processes* 30, 127-136

283. McLuckey, S. A., and Goeringer, D. E. (1997) Slow Heating Methods in Tandem Mass Spectrometry. *Journal of Mass Spectrometry* 32, 461-474

284. Wells, J. M., and McLuckey, S. A. (2005) Collision-Induced Dissociation (Cid) of Peptides and Proteins. *Methods in Enzymology* 402, 148-185

285. Gauthier, J. W., Trautman, T. R., and Jacobson, D. B. (1991) Sustained Off-Resonance Irradiation for Collision-Activated Dissociation Involving Fourier-Transform Mass-Spectrometry - Collision-Activated Dissociation Technique That Emulates Infrared Multiphoton Dissociation. *Analytica Chimica Acta* 246, 211-225

286. Thomson, B. A., Douglas, D. J., Corr, J. J., Hager, J. W., and Jolliffe, C. L. (1995) Improved Collisionally Activated Dissociation Efficiency and Mass Resolution on a Triple Quadrupole Mass-Spectrometer System. *Analytical Chemistry* 67, 1696-1704

287. Yost, R. A., and Enke, C. G. (1979) Triple Quadrupole Mass Spectrometry for Direct Mixture Analysis and Structure Elucidation. *Analytical Chemistry* 51, 1251-1264

288. Sato, K., Asada, T., Ishihara, M., Kunihiro, F., Kammei, Y., Kubota, E., Costello, C. E., Martin, S. A., Scoble, H. A., and Biemann, K. (1987) High-Performance Tandem Mass Spectrometry: Calibration and Performance of Linked Scans of a Four-Sector Instrument. *Analytical Chemistry* 59, 1652-1659

289. Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L., and Burlingame, A. L. (2000) The Characteristics of Peptide Collision-Induced Dissociation Using a High-Performance Maldi-Tof/Tof Tandem Mass Spectrometer. *Analytical Chemistry* 72, 552-558

290. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nature Methods* 4, 709-712

291. Makarov, A., Denisov, E., Lange, O., and Horning, S. (2006) Dynamic Range of Mass Accuracy in Ltq Orbitrap Hybrid Mass Spectrometer. *Journal of the American Society for Mass Spectrometry* 17, 977-982

292. Boja, E. S., Phillips, D., French, S. A., Harris, R. A., and Balaban, R. S. (2009) Quantitative Mitochondrial Phosphoproteomics Using Itraq on an Ltq-Orbitrap with High Energy Collision Dissociation. *Journal of Proteome Research* 8, 4665-4675

293. Bantscheff, M., Boesche, M., Eberhard, D., Matthieson, T., Sweetman, G., and Kuster, B. (2008) Robust and Sensitive Itraq Quantification on an Ltq Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics* 7, 1702-1713

294. Przybylski, C., and Bonnet, V. (2013) Discrimination of Cyclic and Linear Oligosaccharides by Tandem Mass Spectrometry Using Collision-Induced Dissociation (Cid), Pulsed-Q-Dissociation (Pqd) and the Higher-Energy C-Trap Dissociation Modes. *Rapid Communications in Mass Spectrometry* 27, 75-87

295. Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. (1998) Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society* 120, 3265-3266

296. Nielsen, M. L., Budnik, B. A., Haselmann, K. F., Olsen, J. V., and Zubarev, R. A. (2000) Intramolecular Hydrogen Atom Transfer in Hydrogen-Deficient Polypeptide Radical Cations. *Chemical Physics Letters* 330, 558-562

297. Budnik, B. A., Haselmann, K. F., and Zubarev, R. A. (2001) Electron Detachment Dissociation of Peptide Di-Anions: An Electron-Hole Recombination Phenomenon. *Chemical Physics Letters* 342, 299-302

298. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 101, 9528-9533

299. Voinov, V. G., Deinzer, M. L., Beckman, J. S., and Barofsky, D. F. (2011) Electron Capture, Collision-Induced, and Electron Capture-Collision Induced Dissociation in Q-Tof. *Journal of the American Society for Mass Spectrometry* 22, 607-611

300. Roepstorff, P., and Fohlman, J. (1984) Letter to the Editors. *Biological Mass Spectrometry* 11, 601-601

301. Johnson, R. S., Martin, S. A., and Biemann, K. (1988) Collision-Induced Fragmentation of (M+H)+Ions of Peptides - Side-Chain Specific Sequence Ions. *International Journal of Mass Spectrometry and Ion Processes* 86, 137-154

302. Biemann, K. (1990) [25] Sequencing of Peptides by Tandem Mass Spectrometry and High-Energy Collision-Induced Dissociation. *Methods in Enzymology* 193, 455-479

303. Biemann, K., and Papayannopoulos, I. A. (1994) Amino Acid Sequencing of Proteins. *Accounts of Chemical Research* 27, 370-378

304. Hunt, D. F., Yates, J. R., 3rd, Shabanowitz, J., Winston, S., and Hauer, C. R. (1986) Protein Sequencing by Tandem Mass Spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 83, 6233-6237

305. Klassen, J. S., and Kebarle, P. (1997) Collision-Induced Dissociation Threshold Energies of Protonated Glycine, Glycinamide, and Some Related Small Peptides and Peptide Amino Amides. *Journal of the American Chemical Society* 119, 6552-6563

306. Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., Syka, J. E., and Coon, J. J. (2007) Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. *Analytical Chemistry* 79, 477-485

307. Domon, B., and Costello, C. E. (1988) A Systematic Nomenclature for Carbohydrate Fragmentations in Fab-Ms Ms Spectra of Glycoconjugates. *Glycoconjugate Journal* 5, 397-409

308. Reinhold, V. N., Reinhold, B. B., and Costello, C. E. (1995) Carbohydrate Molecular Weight Profiling, Sequence, Linkage, and Branching Data: Es-Ms and Cid. *Analytical Chemistry* 67, 1772-1784

309. Brüll, L., Kovácik, V., Thomas-Oates, J., Heerma, W., and Haverkamp, J. (1998) Sodium-Cationized Oligosaccharides Do Not Appear to Undergo 'Internal Residue Loss' Rearrangement Processes on Tandem Mass Spectrometry. *Rapid Communications in Mass Spectrometry* 12, 1520-1532

310. Kovacik, V., Hirsch, J., Kovac, P., Heerma, W., Thomasoates, J., and Haverkamp, J. (1995) Oligosaccharide Characterization Using Collision-Induced Dissociation Fast-Atom-Bombardment Mass-Spectrometry - Evidence for Internal Monosaccharide Residue Loss. *Journal of Mass Spectrometry* 30, 949-958

311. Ma, Y. L., Vedernikova, I., Van den Heuvel, H., and Claeys, M. (2000) Internal Glucose Residue Loss in Protonated O-Diglycosyl Flavonoids Upon Low-Energy Collision-Induced Dissociation. *Journal of the American Society for Mass Spectrometry* 11, 136-144

312. Harvey, D. J., Bateman, R. H., and Green, M. R. (1997) High-Energy Collision-Induced Fragmentation of Complex Oligossaccharides Ionized by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Journal of Mass Spectrometry* 32, 167-187

313. Viseux, N., de Hoffmann, E., and Domon, B. (1998) Structural Assignment of Permethylated Oligosaccharide Subunits Using Sequential Tandem Mass Spectrometry. *Analytical Chemistry* 70, 4951-4959

314. Lemoine, J., Fournet, B., Despeyroux, D., Jennings, K. R., Rosenberg, R., and de Hoffmann, E. (1993) Collision-Induced Dissociation of Alkali Metal Cationized and Permethylated Oligosaccharides: Influence of the Collision Energy and of the Collision Gas for the Assignment of Linkage Position. *Journal of the American Society for Mass Spectrometry* 4, 197-203

315. Zhao, C., Xie, B., Chan, S. Y., Costello, C. E., and O'Connor, P. B. (2008) Collisionally Activated Dissociation and Electron Capture Dissociation Provide Complementary Structural Information for Branched Permethylated Oligosaccharides. *Journal of the American Society for Mass Spectrometry* 19, 138-150

316. Yu, X., Huang, Y., Lin, C., and Costello, C. E. (2012) Energy-Dependent Electron Activated Dissociation of Metal-Adducted Permethylated Oligosaccharides. *Analytical Chemistry* 84, 7487-7494

317. Yu, X., Jiang, Y., Chen, Y., Huang, Y., Costello, C. E., and Lin, C. (2013) Detailed Glycan Structural Characterization by Electronic Excitation Dissociation. *Analytical Chemistry* 85, 10017-10021

318. Han, L., and Costello, C. E. (2011) Electron Transfer Dissociation of Milk Oligosaccharides. *Journal of the American Society for Mass Spectrometry* 22, 997-1013

319. James, A., and Martin, A. (1952) Gas-Liquid Partition Chromatography. A Technique for the Analysis of Volatile Materials. *Analyst* 77, 915-932

320. James, A. T., and Martin, A. J. (1952) Gas-Liquid Partition Chromatography; the Separation and Micro-Estimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid. *Biochemical Journal* 50, 679-690

321. Gohlke, G. S. (1957) Instrument Design for Gas-Liquid Partition Chromatography. *Analytical Chemistry* 29, 1723-1724

322. McWilliam, I. G., and Dewar, R. A. (1958) Flame Ionization Detector for Gas Chromatography. *Nature* 181, 760-760

323. Brody, S. S., and Chaney, J. E. (1966) Flame Photometric Detector the Application of a Specific Detector for Phosphorus and for Sulfur Compounds—Sensitive to Subnanogram Quantities. Oxford University Press

324. Lovelock, J. E., and Lipsky, S. R. (1960) Electron Affinity Spectroscopy—a New Method for the Identification of Functional Groups in Chemical Compounds Separated by Gas Chromatography. *Journal of the American Chemical Society* 82, 431-433

325. Halász, I., and Horváth, C. (1963) Open Tube Columns with Impregnated Thin Layer Support for Gas Chromatography. *Analytical Chemistry* 35, 499-505

326. Barry, E. F., and Grob, R. L. (2007) *Columns for Gas Chromatography: Performance and Selection*, John Wiley & Sons

327. David, F., Sandra, P., and Vickers, A. K. (2005) Column Selection for the Analysis of Fatty Acid Methyl Esters. *Research Institute For Chromatography, Agilent Technology, USA*

328. Barry, E. F. (2004) Columns: Packed and Capillary; Column Selection in Gas Chromatography. *Modern Practice of Gas Chromatography* 4, 65-191

329. Agilent (2012) Agilent J&W GC Column Selection Guide. *Speed your selection with this one-stop resource*, pp. 1-268, Agilent Technologies, Inc., https://www.agilent.com/cs/library/catalogs/Public/5990-9867EN_GC_CSG.pdf

330. Holmes, J., and Morrell, F. (1957) Oscillographic Mass Spectrometric Monitoring of Gas Chromatography. *Applied Spectroscopy* 11, 86-87

331. Gohlke, R. S. (1959) Time-of-Flight Mass Spectrometry and Gas-Liquid Partition Chromatography. *Analytical Chemistry* 31, 535-541

332. Biemann, K., and Vetter, W. (1960) Separation of Peptide Derivatives by Gas Chromatography Combined with the Mass Spectrometric Determination of the Amino Acid Sequence. *Biochemical and Biophysical Research Communications* 3, 578-584

333. Ryhage, R. (1964) Use of a Mass Spectrometer as a Detector and Analyzer for Effluent Emerging from High Temperature Gas Liquid Chromatography Columns. *Analytical Chemistry* 36, 759-764

334. Watson, J. T., and Biemann, K. (1964) High-Resolution Mass Spectra of Compounds Emerging from a Gas Chromatograph. *Analytical Chemistry* 36, 1135-1137

335. Ryce, S., and Bryce, W. (1957) An Ionization Gauge Detector for Gas Chromatography. *Canadian Journal of Chemistry* 35, 1293-1297

336. Sweeley, C. C., Bentley, R., Makita, M., and Wells, W. W. (1963) Gas-Liquid Chromatography of Trimethylsilyl Derivatives of Sugars and Related Substances. *Journal of the American Chemical Society* 85, 2497-2507

337. Kircher, H. W. (1960) Gas-Liquid Partition Chromatography of Methylated Sugars. *Analytical Chemistry* 32, 1103-1106

338. Wallenfels, K., Bechtler, G., Kuhn, R., Trischmann, H., and Egge, H. (1963) Permethylation of Oligomeric and Polymeric Carbohydrates and Quantitative Analysis of the Cleavage Products. *Angewandte Chemie International Edition in English* 2, 515-523

339. Gunner, S. W., Jones, J. N., and Perry, M. (1961) The Gas–Liquid Partition Chromatography of Carbohydrate Derivatives: Part I the Separation of Glycitol and Glycose Acetates. *Canadian Journal of Chemistry* 39, 1892-1899

340. VandenHeuvel, W., and Horning, E. (1961) Gas Chromatographic Separations of Sugars and Related Compounds as Acetyl Derivatives. *Biochemical and Biophysical Research Communications* 4, 399-403

341. Kim, J. H., Shome, B., Liao, T. H., and Pierce, J. G. (1967) Analysis of Neutral Sugars by Gas-Liquid Chromatography of Alditol Acetates: Application to Thyrotropic Hormone and Other Glycoproteins. *Analytical Biochemistry* 20, 258-274

342. Sweeley, C. C., Wells, W. W., and Bentley, R. (1966) [7] Gas Chromatography of Carbohydrates. *Methods in Enzymology* 8, 95-108

343. Checkley, W., White, A. C., Jr., Jaganath, D., Arrowood, M. J., Chalmers, R. M., Chen, X. M., Fayer, R., Griffiths, J. K., Guerrant, R. L., Hedstrom, L., Huston, C. D., Kotloff, K. L., Kang, G., Mead, J. R., Miller, M., Petri, W. A., Jr., Priest, J. W., Roos, D. S., Striepen, B., Thompson, R. C., Ward, H. D., Van Voorhis, W. A., Xiao, L., Zhu, G., and Houpt, E. R. (2015) A Review of the Global Burden, Novel Diagnostics, Therapeutics, and Vaccine Targets for Cryptosporidium. *Lancet Infectious Diseases* 15, 85-94

344. Fayer, R., and Xiao, L. (2007) *Cryptosporidium and Cryptosporidiosis*, CRC press

345. Baldursson, S., and Karanis, P. (2011) Waterborne Transmission of Protozoan Parasites: Review of Worldwide Outbreaks - an Update 2004-2010. *Water Research* 45, 6603-6614

346. Platts-Mills, J. A., Babji, S., Bodhidatta, L., Gratz, J., Haque, R., Havt, A., McCormick, B. J., McGrath, M., Olortegui, M. P., Samie, A., Shakoor, S., Mondal, D., Lima, I. F., Hariraju, D., Rayamajhi, B. B., Qureshi, S., Kabir, F., Yori, P. P., Mufamadi, B., Amour, C., Carreon, J. D., Richard, S. A., Lang, D., Bessong, P., Mduma, E., Ahmed, T., Lima, A. A., Mason, C. J., Zaidi, A. K., Bhutta, Z. A., Kosek, M., Guerrant, R. L., Gottlieb, M., Miller, M., Kang, G., Houpt, E. R., and Investigators, M.-E. N. (2015) Pathogen-Specific Burdens of Community Diarrhoea in Developing Countries: A Multisite Birth Cohort Study (Mal-Ed). *The Lancet. Global Health* 3, e564-575

347. Wanyiri, J. W., Kanyi, H., Maina, S., Wang, D. E., Steen, A., Ngugi, P., Kamau, T., Waithera, T., O'Connor, R., Gachuhi, K., Wamae, C. N., Mwamburi, M., and Ward, H. D. (2014) Cryptosporidiosis in Hiv/Aids Patients in Kenya: Clinical Features, Epidemiology, Molecular Characterization and Antibody Responses. *American Journal of Tropical Medicine and Hygiene* 91, 319-328

348. Sarkar, R., Kattula, D., Francis, M. R., Ajjampur, S. S., Prabakaran, A. D., Jayavelu, N., Muliyil, J., Balraj, V., Naumova, E. N., Ward, H. D., and Kang, G. (2014) Risk Factors for Cryptosporidiosis among Children in a Semi Urban Slum in Southern India: A Nested Case-Control Study. *American Journal of Tropical Medicine and Hygiene* 91, 1128-1137

349. Benitez, A. J., McNair, N., and Mead, J. R. (2009) Oral Immunization with Attenuated Salmonella Enterica Serovar Typhimurium Encoding Cryptosporidium Parvum Cp23 and Cp40 Antigens Induces a Specific Immune Response in Mice. *Clinical and Vaccine Immunology* 16, 1272-1278

350. Ehigiator, H. N., Romagnoli, P., Priest, J. W., Secor, W. E., and Mead, J. R. (2007) Induction of Murine Immune Responses by DNA Encoding a 23-Kda Antigen of Cryptosporidium Parvum. *Parasitology Research* 101, 943-950

351. Jenkins, M., Higgins, J., Kniel, K., Trout, J., and Fayer, R. (2004) Protection of Calves against Cryptosporiosis by Oral Inoculation with Gamma-Irradiated Cryptosporidium Parvum Oocysts. *Journal of Parasitology* 90, 1178-1180

352. Ludington, J. G., and Ward, H. D. (2015) Systemic and Mucosal Immune Responses to Cryptosporidium-Vaccine Development. *Curr Trop Med Rep* 2, 171-180

353. Mead, J. R. (2014) Prospects for Immunotherapy and Vaccines against Cryptosporidium. *Human Vaccines & Immunotherapeutics* 10, 1505-1513

354. Cabada, M. M., and White, A. C., Jr. (2010) Treatment of Cryptosporidiosis: Do We Know What We Think We Know? *Current Opinion in Infectious Diseases* 23, 494-499

355. Aebi, M. (2013) N-Linked Protein Glycosylation in the Er. *Biochimica et Biophysica Acta* 1833, 2430-2437

356. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped Blast and Psi-Blast: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* 25, 3389-3402

357. Aurrecoechea, C., Barreto, A., Brestelli, J., Brunk, B. P., Cade, S., Doherty, R., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hu, S., Iodice, J., Kissinger, J. C., Kraemer, E. T., Li, W., Pinney, D. F., Pitts, B., Roos, D. S.,

Srinivasamoorthy, G., Stoeckert, C. J., Jr., Wang, H., and Warrenfeltz, S. (2013) Eupathdb: The Eukaryotic Pathogen Database. *Nucleic Acids Research* 41, D684-691

358. Heiges, M., Wang, H., Robinson, E., Aurrecoechea, C., Gao, X., Kaluskar, N., Rhodes, P., Wang, S., He, C. Z., Su, Y., Miller, J., Kraemer, E., and Kissinger, J. C. (2006) Cryptodb: A Cryptosporidium Bioinformatics Resource Update. *Nucleic Acids Research* 34, D419-422

359. Bushkin, G. G., Ratner, D. M., Cui, J., Banerjee, S., Duraisingh, M. T., Jennings, C. V., Dvorin, J. D., Gubbels, M. J., Robertson, S. D., Steffen, M., O'Keefe, B. R., Robbins, P. W., and Samuelson, J. (2010) Suggestive Evidence for Darwinian Selection against Asparagine-Linked Glycans of Plasmodium Falciparum and Toxoplasma Gondii. *Eukaryotic Cell* 9, 228-241

360. Satoh, T., Yamaguchi, T., and Kato, K. (2015) Emerging Structural Insights into Glycoprotein Quality Control Coupled with N-Glycan Processing in the Endoplasmic Reticulum. *Molecules* 20, 2475-2491

361. Adams, E. W., Ratner, D. M., Bokesch, H. R., McMahon, J. B., O'Keefe, B. R., and Seeberger, P. H. (2004) Oligosaccharide and Glycoprotein Microarrays as Tools in Hiv Glycobiology; Glycan-Dependent Gp120/Protein Interactions. *Chemistry and Biology* 11, 875-881

362. Samuelson, J., and Robbins, P. W. (2015) Effects of N-Glycan Precursor Length Diversity on Quality Control of Protein Folding and on Protein Glycosylation. *Seminars in Cell & Developmental Biology*, pp. 121-128, Elsevier

363. Ajjampur, S. S., Sarkar, R., Allison, G., Banda, K., Kane, A., Muliyil, J., Naumova, E., Ward, H., and Kang, G. (2011) Serum Igg Response to Cryptosporidium Immunodominant Antigen Gp15 and Polymorphic Antigen Gp40 in Children with Cryptosporidiosis in South India. *Clinical and Vaccine Immunology* 18, 633-639

364. Allison, G. M., Rogers, K. A., Borad, A., Ahmed, S., Karim, M. M., Kane, A. V., Hibberd, P. L., Naumova, E. N., Calderwood, S. B., Ryan, E. T., Khan, W. A., and Ward, H. D. (2011) Antibody Responses to the Immunodominant Cryptosporidium Gp15 Antigen and Gp15 Polymorphisms in a Case-Control Study of Cryptosporidiosis in Children in Bangladesh. *American Journal of Tropical Medicine and Hygiene* 85, 97-104

365. Lazarus, R. P., Ajjampur, S. S., Sarkar, R., Geetha, J. C., Prabakaran, A. D., Velusamy, V., Naumova, E. N., Ward, H. D., and Kang, G. (2015) Serum Anti-Cryptosporidial Gp15 Antibodies in Mothers and Children Less Than 2 Years of Age in India. *American Journal of Tropical Medicine and Hygiene* 93, 931-938

366. McDonald, A. C., Mac Kenzie, W. R., Addiss, D. G., Gradus, M. S., Linke, G., Zembrowski, E., Hurd, M. R., Arrowood, M. J., Lammie, P. J., and Priest, J. W. (2001)

Cryptosporidium Parvum-Specific Antibody Responses among Children Residing in Milwaukee During the 1993 Waterborne Outbreak. *Journal of Infectious Diseases* 183, 1373-1379

367. Sarkar, R., Ajjampur, S. S., Muliyil, J., Ward, H., Naumova, E. N., and Kang, G. (2012) Serum Igg Responses and Seroconversion Patterns to Cryptosporidium Gp15 among Children in a Birth Cohort in South India. *Clinical and Vaccine Immunology* 19, 849-854

368. Spano, F., Puri, C., Ranucci, L., Putignani, L., and Crisanti, A. (1997) Cloning of the Entire Cowp Gene of Cryptosporidium Parvum and Ultrastructural Localization of the Protein During Sexual Parasite Development. *Parasitology* 114 (Pt 5), 427-437

369. Strong, W. B., Gut, J., and Nelson, R. G. (2000) Cloning and Sequence Analysis of a Highly Polymorphic Cryptosporidium Parvum Gene Encoding a 60-Kilodalton Glycoprotein and Characterization of Its 15- and 45-Kilodalton Zoite Surface Antigen Products. *Infection and Immunity* 68, 4117-4134

370. Howe, C., Lloyd, K. O., and Lee, L. T. (1972) [20] Isolation of Glycoproteins from Red Cell Membranes Using Phenol. *Methods in Enzymology* 28, 236-245

371. Hurkman, W. J., and Tanaka, C. K. (1986) Solubilization of Plant Membrane Proteins for Analysis by Two-Dimensional Gel Electrophoresis. *Plant Physiology* 81, 802-806

372. Ciucanu, I., and Kerek, F. (1984) A Simple and Rapid Method for the Permethylation of Carbohydrates. *Carbohydrate Research* 131, 209-217

373. Ciucanu, I., and Costello, C. E. (2003) Elimination of Oxidative Degradation During the Per-O-Methylation of Carbohydrates. *Journal of the American Chemical Society* 125, 16213-16219

374. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) Weblogo: A Sequence Logo Generator. *Genome Research* 14, 1188-1190

375. Kall, L., Krogh, A., and Sonnhammer, E. L. (2007) Advantages of Combined Transmembrane Topology and Signal Peptide Prediction--the Phobius Web Server. *Nucleic Acids Research* 35, W429-432

376. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *Journal of Molecular Biology* 305, 567-580

377. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) Signalp 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nature Methods* 8, 785-786

378. Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., and Bryant, S. H. (2015) Cdd: Ncbi's Conserved Domain Database. *Nucleic Acids Research* 43, D222-226

379. Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S. Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D. A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P. D., and Finn, R. D. (2015) The Interpro Protein Families Database: The Classification Resource after 15 Years. *Nucleic Acids Research* 43, D213-221

380. Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., and Yao, X. (2009) Dog 1.0: Illustrator of Protein Domain Structures. *Cell Research* 19, 271-273

381. Baum, J., Richard, D., Healer, J., Rug, M., Krnajski, Z., Gilberger, T. W., Green, J. L., Holder, A. A., and Cowman, A. F. (2006) A Conserved Molecular Motor Drives Cell Invasion and Gliding Motility across Malaria Life Cycle Stages and Other Apicomplexan Parasites. *Journal of Biological Chemistry* 281, 5197-5208

382. Pradel, G., Hayton, K., Aravind, L., Iyer, L. M., Abrahamsen, M. S., Bonawitz, A., Mejia, C., and Templeton, T. J. (2004) A Multidomain Adhesion Protein Family Expressed in Plasmodium Falciparum Is Essential for Transmission to the Mosquito. *Journal of Experimental Medicine* 199, 1533-1544

383. Kasturi, L., Eshleman, J. R., Wunner, W. H., and Shakineshleman, S. H. (1995) The Hydroxy Amino-Acid in an Asn-X-Ser/Thr Sequon Can Influence N-Linked Core Glycosylation Efficiency and the Level of Expression of a Cell-Surface Glycoprotein. *Journal of Biological Chemistry* 270, 14756-14761

384. Carpentieri, A., Ratner, D. M., Ghosh, S. K., Banerjee, S., Bushkin, G. G., Cui, J., Lubrano, M., Steffen, M., Costello, C. E., O'Keefe, B., Robbins, P. W., and Samuelson, J. (2010) The Antiretroviral Lectin Cyanovirin-N Targets Well-Known and Novel Targets on the Surface of Entamoeba Histolytica Trophozoites. *Eukaryotic Cell* 9, 1661-1668

385. Chatterjee, A., Ratner, D. M., Ryan, C. M., Johnson, P. J., O'Keefe, B. R., Secor, W. E., Anderson, D. J., Robbins, P. W., and Samuelson, J. (2015) Anti-Retroviral Lectins Have Modest Effects on Adherence of Trichomonas Vaginalis to Epithelial Cells in Vitro

and on Recovery of Tritrichomonas Foetus in a Mouse Vaginal Model. *PLoS One* 10, e0135340

386. Magnelli, P., Cipollo, J. F., Ratner, D. M., Cui, J., Kelleher, D., Gilmore, R., Costello, C. E., Robbins, P. W., and Samuelson, J. (2008) Unique Asn-Linked Oligosaccharides of the Human Pathogen Entamoeba Histolytica. *Journal of Biological Chemistry* 283, 18355-18364

387. Paschinger, K., Hykollari, A., Razzazi-Fazeli, E., Greenwell, P., Leitsch, D., Walochnik, J., and Wilson, I. B. (2012) The N-Glycans of Trichomonas Vaginalis Contain Variable Core and Antennal Modifications. *Glycobiology* 22, 300-313

388. Hykollari, A., Balog, C. I., Rendic, D., Braulke, T., Wilson, I. B., and Paschinger, K. (2013) Mass Spectrometric Analysis of Neutral and Anionic N-Glycans from a Dictyostelium Discoideum Model for Human Congenital Disorder of Glycosylation Cdg Il. *Journal of Proteome Research* 12, 1173-1187

389. Schiller, B., Hykollari, A., Yan, S., Paschinger, K., and Wilson, I. B. (2012) Complicated N-Linked Glycans in Simple Organisms. *Biological Chemistry* 393, 661-673

390. Schiller, B., Makrypidi, G., Razzazi-Fazeli, E., Paschinger, K., Walochnik, J., and Wilson, I. B. (2012) Exploring the Unique N-Glycome of the Opportunistic Human Pathogen Acanthamoeba. *Journal of Biological Chemistry* 287, 43191-43204

391. Kobayashi, Y., and Suzuki, Y. (2012) Evidence for N-Glycan Shielding of Antigenic Sites During Evolution of Human Influenza a Virus Hemagglutinin. *Journal of Virology* 86, 3446-3451

392. Pentiah, K., Lees, W. D., Moss, D. S., and Shepherd, A. J. (2015) N-Linked Glycans on Influenza a H3n2 Hemagglutinin Constrain Binding of Host Antibodies, but Shielding Is Limited. *Glycobiology* 25, 124-132

393. van Montfort, T., Eggink, D., Boot, M., Tuen, M., Hioe, C. E., Berkhout, B., and Sanders, R. W. (2011) Hiv-1 N-Glycan Composition Governs a Balance between Dendritic Cell-Mediated Viral Transmission and Antigen Presentation. *The Journal of Immunology* 187, 4676-4685

394. Petersen, C. B., CA) (2000) Gp900 Glycoprotein and Fragments for Treatment and Detection/Diagnosis of Cryptosporidium. The Regents of The University of California (Oakland, CA), United States

395. Ward, H., Hamer, D., Keusch, G., and Cevallos, A. M. (2003) Gp40 and Uses Thereof. U.S. Patent US6657045.

396. Miller, N. L., Frenkel, J. K., and Dubey, J. P. (1972) Oral Infections with Toxoplasma Cysts and Oocysts in Felines, Other Mammals, and in Birds. *Journal of Parasitology* 58, 928-937

397. Yilmaz, S. M., and Hopkins, S. H. (1972) Effects of Different Conditions on Duration of Infectivity of Toxoplasma Gondii Oocysts. *Journal of Parasitology* 58, 938-939

398. Frenkel, J. K., Ruiz, A., and Chinchilla, M. (1975) Soil Survival of Toxoplasma Oocysts in Kansas and Costa Rica. *American Journal of Tropical Medicine and Hygiene* 24, 439-443

399. Jacobs, L., Remington, J. S., and Melton, M. L. (1960) A Survey of Meat Samples from Swine, Cattle, and Sheep for the Presence of Encysted Toxoplasma. *Journal of Parasitology* 46, 23-28

400. Tenter, A. M., Heckeroth, A. R., and Weiss, L. M. (2000) Toxoplasma Gondii: From Animals to Humans. *International Journal for Parasitology* 30, 1217-1258

401. Montoya, J. G., and Liesenfeld, O. (2004) Toxoplasmosis. *Lancet* 363, 1965-1976

402. Grant, I. H., Gold, J. W., Rosenblum, M., Niedzwiecki, D., and Armstrong, D. (1990) Toxoplasma Gondii Serology in Hiv-Infected Patients: The Development of Central Nervous System Toxoplasmosis in AIDS. *AIDS* 4, 519-521

403. Shepp, D. H., Hackman, R. C., Conley, F. K., Anderson, J. B., and Meyers, J. D. (1985) Toxoplasma Gondii Reactivation Identified by Detection of Parasitemia in Tissue Culture. *Annals of Internal Medicine* 103, 218-221

404. Cohen, S. N. (1970) Toxoplasmosis in Patients Receiving Immunosuppressive Therapy. *The Journal of the American Medical Association* 211, 657-660

405. Luft, B. J., Brooks, R. G., Conley, F. K., McCabe, R. E., and Remington, J. S. (1984) Toxoplasmic Encephalitis in Patients with Acquired Immune Deficiency Syndrome. *The Journal of the American Medical Association* 252, 913-917

406. Luft, B. J., Hafner, R., Korzun, A. H., Leport, C., Antoniskis, D., Bosler, E. M., Bourland, D. D., 3rd, Uttamchandani, R., Fuhrer, J., Jacobson, J., and et al. (1993) Toxoplasmic Encephalitis in Patients with the Acquired Immunodeficiency Syndrome. Members of the Actg 077p/Anrs 009 Study Team. *New England Journal of Medicine* 329, 995-1000

407. Desmonts, G., and Couvreur, J. (1974) Congenital Toxoplasmosis. A Prospective Study of 378 Pregnancies. *New England Journal of Medicine* 290, 1110-1116

408. Bosch-Driessen, L. H., Verbraak, F. D., Suttorp-Schulten, M. S., van Ruyven, R. L., Klok, A. M., Hoyng, C. B., and Rothova, A. (2002) A Prospective, Randomized Trial of Pyrimethamine and Azithromycin Vs Pyrimethamine and Sulfadiazine for the Treatment of Ocular Toxoplasmosis. *American Journal of Ophthalmology* 134, 34-40

409. Leport, C., Raffi, F., Matheron, S., Katlama, C., Regnier, B., Saimot, A. G., Marche, C., Vedrenne, C., and Vilde, J. L. (1988) Treatment of Central Nervous System Toxoplasmosis with Pyrimethamine/Sulfadiazine Combination in 35 Patients with the Acquired Immunodeficiency Syndrome. Efficacy of Long-Term Continuous Therapy. *American Journal of Medicine* 84, 94-100

410. group, S. s., Thiebaut, R., Leproust, S., Chene, G., and Gilbert, R. (2007) Effectiveness of Prenatal Treatment for Congenital Toxoplasmosis: A Meta-Analysis of Individual Patients' Data. *Lancet* 369, 115-122

411. Odenthal-Schnittler, M., Tomavo, S., Becker, D., Dubremetz, J. F., and Schwarz, R. T. (1993) Evidence for N-Linked Glycosylation in Toxoplasma Gondii. *Biochemical Journal* 291 (Pt 3), 713-721

412. Gaskins, E., Gilk, S., DeVore, N., Mann, T., Ward, G., and Beckers, C. (2004) Identification of the Membrane Receptor of a Class Xiv Myosin in Toxoplasma Gondii. *Journal of Cell Biology* 165, 383-393

413. Roos, D. S., Donald, R. G. K., Morrissette, N. S., and Moulton, A. L. C. (1995) Chapter 3: Molecular Tools for Genetic Dissection of the Protozoan Parasite Toxoplasma Gondii. *Methods in Cell Biology*, pp. 27-63, Elsevier

414. Morrissette, N. S., Bedian, V., Webster, P., and Roos, D. S. (1994) Characterization of Extreme Apical Antigens from Toxoplasma Gondii. *Experimental Parasitology* 79, 445-459

415. Bandini, G., Haserick, J. R., Motari, E., Ouologuem, D. T., Lourido, S., Roos, D. S., Costello, C. E., Robbins, P. W., and Samuelson, J. (2016) O-Fucosylated Glycoproteins Form Assemblies in Close Proximity to the Nuclear Pore Complexes of Toxoplasma Gondii. *Proceedings of the National Academy of Sciences of the United States of America* 113, 11567-11572

416. Haserick, J. R., Leon, D. R., Samuelson, J., and Costello, C. E. (2017) Asparagine-Linked Glycans of Cryptosporidium Parvum Contain a Single Long Arm, Are Barely Processed in the Endoplasmic Reticulum (Er) or Golgi, and Show a Strong Bias for Sites with Threonine. *Molecular & Cellular Proteomics* 16, S42-S53

417. Omasits, U., Ahrens, C. H., Muller, S., and Wollscheid, B. (2014) Protter: Interactive Protein Feature Visualization and Integration with Experimental Proteomic Data. *Bioinformatics* 30, 884-886

418. Eisenhaber, B., Bork, P., and Eisenhaber, F. (1998) Sequence Properties of Gpi-Anchored Proteins near the Omega-Site: Constraints for the Polypeptide Binding Site of the Putative Transamidase. *Protein Engineering* 11, 1155-1161

419. Jones, D. T., and Cozzetto, D. (2015) Disopred3: Precise Disordered Region Predictions with Annotated Protein-Binding Activity. *Bioinformatics* 31, 857-863

420. Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H. Y., Dosztanyi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G. L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D. A., Necci, M., Nuka, G., Orengo, C. A., Park, Y., Pesseat, S., Piovesan, D., Potter, S. C., Rawlings, N. D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P. D., Tosatto, S. C., Wu, C. H., Xenarios, I., Yeh, L. S., Young, S. Y., and Mitchell, A. L. (2017) Interpro in 2017-Beyond Protein Family and Domain Annotations. *Nucleic Acids Research* 45, D190-D199

421. Sigrist, C. J., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2012) New and Continuing Developments at Prosite. *Nucleic Acids Research*, gks1067

422. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam Protein Families Database: Towards a More Sustainable Future. *Nucleic Acids Research* 44, D279-285

423. Jung, C., Lee, C. Y., and Grigg, M. E. (2004) The Srs Superfamily of Toxoplasma Surface Proteins. *International Journal for Parasitology* 34, 285-296

424. Tzipori, S., and Widmer, G. (2008) A Hundred-Year Retrospective on Cryptosporidiosis. *Trends in Parasitology* 24, 184-189

425. Garcia, L. S., Bruckner, D. A., Brewer, T. C., and Shimizu, R. Y. (1983) Techniques for the Recovery and Identification of Cryptosporidium Oocysts from Stool Specimens. *Journal of Clinical Microbiology* 18, 185-190

426. Bushkin, G. G., Motari, E., Carpentieri, A., Dubey, J. P., Costello, C. E., Robbins, P. W., and Samuelson, J. (2013) Evidence for a Structural Role for Acid-Fast Lipids in Oocyst Walls of Cryptosporidium, Toxoplasma, and Eimeria. *MBio* 4, e00387-00313

427. Priest, J. W., Mehlert, A., Moss, D. M., Arrowood, M. J., and Ferguson, M. A. (2006) Characterization of the Glycosylphosphatidylinositol Anchor of the Immunodominant Cryptosporidium Parvum 17-Kda Antigen. *Molecular and Biochemical Parasitology* 149, 108-112

428. Hira, K. G., Mackay, M. R., Hempstead, A. D., Ahmed, S., Karim, M. M., O'Connor, R. M., Hibberd, P. L., Calderwood, S. B., Ryan, E. T., Khan, W. A., and Ward, H. D. (2011) Genetic Diversity of Cryptosporidium Spp. From Bangladeshi Children. *Journal of Clinical Microbiology* 49, 2307-2310

429. Roche, J. K., Rojo, A. L., Costa, L. B., Smeltz, R., Manque, P., Woehlbier, U., Bartelt, L., Galen, J., Buck, G., and Guerrant, R. L. (2013) Intranasal Vaccination in Mice with an Attenuated Salmonella Enterica Serovar 908htr a Expressing Cp15 of Cryptosporidium: Impact of Malnutrition with Preservation of Cytokine Secretion. *Vaccine* 31, 912-918

430. Arrowood, M. J., Mead, J. R., Mahrt, J. L., and Sterling, C. R. (1989) Effects of Immune Colostrum and Orally Administered Antisporozoite Monoclonal Antibodies on the Outcome of Cryptosporidium Parvum Infections in Neonatal Mice. *Infection and Immunity* 57, 2283-2288

431. Smith, L. M., Priest, J. W., Lammie, P. J., and Mead, J. R. (2001) Human T and B Cell Immunoreactivity to a Recombinant 23-Kda Cryptosporidium Parvum Antigen. *Journal of Parasitology* 87, 704-707

432. Bedi, B., and Mead, J. R. (2012) Cryptosporidium Parvum Antigens Induce Mouse and Human Dendritic Cells to Generate Th1-Enhancing Cytokines. *Parasite Immunology* 34, 473-485

433. Borad, A. J., Allison, G. M., Wang, D., Ahmed, S., Karim, M. M., Kane, A. V., Moy, J., Hibberd, P. L., Ajjampur, S. S., Kang, G., Calderwood, S. B., Ryan, E. T., Naumova, E., Khan, W. A., and Ward, H. D. (2012) Systemic Antibody Responses to the Immunodominant P23 Antigen and P23 Polymorphisms in Children with Cryptosporidiosis in Bangladesh. *American Journal of Tropical Medicine and Hygiene* 86, 214-222

434. Du, X. L., Xu, J. M., Hou, M., Yu, R. B., Ge, J. J., Zhu, H. S., and Wu, H. W. (2009) Simultaneous Detection of Serum Immunoglobulin G Antibodies to Cryptosporidium Parvum by Multiplex Microbead Immunoassay Using 3 Recognized Specific Recombinant C. Parvum Antigens. *Diagnostic Microbiology and Infectious Disease* 65, 271-278

435. Liu, K., Zai, D., Zhang, D., Wei, Q., Han, G., Gao, H., and Huang, B. (2010) Divalent Cp15-23 Vaccine Enhances Immune Responses and Protection against Cryptosporidium Parvum Infection. *Parasite Immunology* 32, 335-344

436. Shayan, P., Ebrahimzadeh, E., Mokhber-Dezfouli, M. R., and Rahbari, S. (2008) Recombinant Cryptosporidium Parvum P23 as a Target for the Detection of Cryptosporidium-Specific Antibody in Calf Sera. *Parasitology Research* 103, 1207-1211

437. Snelling, W. J., Lin, Q., Moore, J. E., Millar, B. C., Tosini, F., Pozio, E., Dooley, J. S., and Lowery, C. J. (2007) Proteomics Analysis and Protein Expression During Sporozoite Excystation of Cryptosporidium Parvum (Coccidia, Apicomplexa). *Molecular & Cellular Proteomics* 6, 346-355

438. Sanderson, S. J., Xia, D., Prieto, H., Yates, J., Heiges, M., Kissinger, J. C., Bromley, E., Lal, K., Sinden, R. E., Tomley, F., and Wastling, J. M. (2008) Determining the Protein Repertoire of Cryptosporidium Parvum Sporozoites. *Proteomics* 8, 1398-1414

439. Madrid-Aliste, C. J., Dybas, J. M., Angeletti, R. H., Weiss, L. M., Kim, K., Simon, I., and Fiser, A. (2009) Epic-Db: A Proteomics Database for Studying Apicomplexan Organisms. *BMC Genomics* 10, 38

440. Singh, P., Mirdha, B. R., Srinivasan, A., Rukmangadachar, L. A., Singh, S., Sharma, P., Hariprasad, G., Gururao, H., and Luthra, K. (2015) Identification of Invasion Proteins of Cryptosporidium Parvum. *World Journal of Microbiology & Biotechnology* 31, 1923-1934

441. Resh, M. D. (2016) Fatty Acylation of Proteins: The Long and the Short of It. *Progress in Lipid Research* 63, 120-131

442. Ji, Y., Leymarie, N., Haeussler, D. J., Bachschmid, M. M., Costello, C. E., and Lin, C. (2013) Direct Detection of S-Palmitoylation by Mass Spectrometry. *Analytical Chemistry* 85, 11952-11959

443. Ji, Y., Bachschmid, M. M., Costello, C. E., and Lin, C. (2016) S- to N-Palmitoyl Transfer During Proteomic Sample Preparation. *Journal of the American Society for Mass Spectrometry* 27, 677-685

444. Aurrecoechea, C., Barreto, A., Basenko, E. Y., Brestelli, J., Brunk, B. P., Cade, S., Crouch, K., Doherty, R., Falke, D., Fischer, S., Gajria, B., Harb, O. S., Heiges, M., Hertz-Fowler, C., Hu, S., Iodice, J., Kissinger, J. C., Lawrence, C., Li, W., Pinney, D. F., Pulman, J. A., Roos, D. S., Shanmugasundram, A., Silva-Franco, F., Steinbiss, S., Stoeckert, C. J., Jr., Spruill, D., Wang, H., Warrenfeltz, S., and Zheng, J. (2017) Eupathdb: The Eukaryotic Pathogen Genomics Database Resource. *Nucleic Acids Research* 45, D581-D591

445. Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., and Haslam, S. M. (2008) Glycoworkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. *Journal of Proteome Research* 7, 1650-1659

446. Jones, A. R., Eisenacher, M., Mayer, G., Kohlbacher, O., Siepen, J., Hubbard, S. J., Selley, J. N., Searle, B. C., Shofstahl, J., Seymour, S. L., Julian, R., Binz, P. A., Deutsch, E. W., Hermjakob, H., Reisinger, F., Griss, J., Vizcaino, J. A., Chambers, M.,

Pizarro, A., and Creasy, D. (2012) The Mzidentml Data Standard for Mass Spectrometry-Based Proteomics Results. *Molecular & Cellular Proteomics* 11, M111 014381

447. Risk, B. A., Edwards, N. J., and Giddings, M. C. (2013) A Peptide-Spectrum Scoring System Based on Ion Alignment, Intensity, and Pair Probabilities. *Journal of Proteome Research* 12, 4240-4247

448. Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolome, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) Proteomexchange Provides Globally Coordinated Proteomics Data Submission and Dissemination. *Nature Biotechnology* 32, 223-226

449. Duckert, P., Brunak, S., and Blom, N. (2004) Prediction of Proprotein Convertase Cleavage Sites. *Protein Engineering, Design & Selection* 17, 107-112

450. Child, M. A., Hall, C. I., Beck, J. R., Ofori, L. O., Albrow, V. E., Garland, M., Bowyer, P. W., Bradley, P. J., Powers, J. C., Boothroyd, J. C., Weerapana, E., and Bogyo, M. (2013) Small-Molecule Inhibition of a Depalmitoylase Enhances Toxoplasma Host-Cell Invasion. *Nature Chemical Biology* 9, 651-656

451. Beck, J. R., Fung, C., Straub, K. W., Coppens, I., Vashisht, A. A., Wohlschlegel, J. A., and Bradley, P. J. (2013) A Toxoplasma Palmitoyl Acyl Transferase and the Palmitoylated Armadillo Repeat Protein Tgaro Govern Apical Rhoptry Tethering and Reveal a Critical Role for the Rhoptries in Host Cell Invasion but Not Egress. *PLoS Pathogens* 9, e1003162

452. Frenal, K., Kemp, L. E., and Soldati-Favre, D. (2014) Emerging Roles for Protein S-Palmitoylation in Toxoplasma Biology. *International Journal for Parasitology* 44, 121-131

453. Foe, I. T., Child, M. A., Majmudar, J. D., Krishnamurthy, S., van der Linden, W. A., Ward, G. E., Martin, B. R., and Bogyo, M. (2015) Global Analysis of Palmitoylated Proteins in Toxoplasma Gondii. *Cell Host & Microbe* 18, 501-511

454. Paul, P., Chowdhury, A., Das Talukdar, A., and Choudhury, M. D. (2015) Homology Modeling and Molecular Dynamics Simulation of N-Myristoyltransferase from Plasmodium Falciparum: An Insight into Novel Antimalarial Drug Design. *Journal of Molecular Modeling* 21, 37

455. Wetzel, J., Herrmann, S., Swapna, L. S., Prusty, D., John Peter, A. T., Kono, M., Saini, S., Nellimarla, S., Wong, T. W., Wilcke, L., Ramsay, O., Cabrera, A., Biller, L., Heincke, D., Mossman, K., Spielmann, T., Ungermann, C., Parkinson, J., and Gilberger,

T. W. (2015) The Role of Palmitoylation for Protein Recruitment to the Inner Membrane Complex of the Malaria Parasite. *Journal of Biological Chemistry* 290, 1712-1728

456. Wright, M. H., Clough, B., Rackham, M. D., Rangachari, K., Brannigan, J. A., Grainger, M., Moss, D. K., Bottrill, A. R., Heal, W. P., Broncel, M., Serwa, R. A., Brady, D., Mann, D. J., Leatherbarrow, R. J., Tewari, R., Wilkinson, A. J., Holder, A. A., and Tate, E. W. (2014) Validation of N-Myristoyltransferase as an Antimalarial Drug Target Using an Integrated Chemical Biology Approach. *Nature Chemistry* 6, 112-121

457. Gill, D. J., Clausen, H., and Bard, F. (2011) Location, Location, Location: New Insights into O-Galnac Protein Glycosylation. *Trends in Cell Biology* 21, 149-158

458. Hurtado-Guerrero, R. (2016) Recent Structural and Mechanistic Insights into Protein O-Galnac Glycosylation. *Biochemical Society Transactions* 44, 61-67

459. Vinayak, S., Pawlowic, M. C., Sateriale, A., Brooks, C. F., Studstill, C. J., Bar-Peled, Y., Cipriano, M. J., and Striepen, B. (2015) Genetic Modification of the Diarrhoeal Pathogen Cryptosporidium Parvum. *Nature* 523, 477-480

460. Wang, Y., Ju, T., Ding, X., Xia, B., Wang, W., Xia, L., He, M., and Cummings, R. D. (2010) Cosmc Is an Essential Chaperone for Correct Protein O-Glycosylation. *Proceedings of the National Academy of Sciences U.S.A.* 107, 9228-9233

461. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., and Clausen, H. (2013) Precision Mapping of the Human O-Galnac Glycoproteome through Simplecell Technology. *EMBO Journal* 32, 1478-1488

462. Yang, Z., Halim, A., Narimatsu, Y., Jitendra Joshi, H., Steentoft, C., Schjoldager, K. T., Alder Schulz, M., Sealover, N. R., Kayser, K. J., Paul Bennett, E., Levery, S. B., Vakhrushev, S. Y., and Clausen, H. (2014) The Galnac-Type O-Glycoproteome of Cho Cells Characterized by the Simplecell Strategy. *Molecular & Cellular Proteomics* 13, 3224-3235

463. Ju, T., Otto, V. I., and Cummings, R. D. (2011) The Tn Antigen-Structural Simplicity and Biological Complexity. *Angewandte Chemie. International Ed. In English* 50, 1770-1791

464. Hanisch, F. G. (2001) O-Glycosylation of the Mucin Type. *Biological Chemistry* 382, 143-149

465. Strous, G. J., and Dekker, J. (1992) Mucin-Type Glycoproteins. *Critical Reviews in Biochemistry and Molecular Biology* 27, 57-92

466. Torrey, E. F., and Yolken, R. H. (2013) Toxoplasma Oocysts as a Public Health Problem. *Trends in Parasitology* 29, 380-384

467. Torgerson, P. R., and Mastroiacovo, P. (2013) The Global Burden of Congenital Toxoplasmosis: A Systematic Review. *Bulletin of the World Health Organization* 91, 501-508

468. Lorenzi, H., Khan, A., Behnke, M. S., Namasivayam, S., Swapna, L. S., Hadjithomas, M., Karamycheva, S., Pinney, D., Brunk, B. P., Ajioka, J. W., Ajzenberg, D., Boothroyd, J. C., Boyle, J. P., Darde, M. L., Diaz-Miranda, M. A., Dubey, J. P., Fritz, H. M., Gennari, S. M., Gregory, B. D., Kim, K., Saeij, J. P., Su, C., White, M. W., Zhu, X. Q., Howe, D. K., Rosenthal, B. M., Grigg, M. E., Parkinson, J., Liu, L., Kissinger, J. C., Roos, D. S., and Sibley, L. D. (2016) Local Admixture of Amplified and Diversified Secreted Pathogenesis Determinants Shapes Mosaic Toxoplasma Gondii Genomes. *Nature Communications* 7, 10147

469. Behnke, M. S., Wootton, J. C., Lehmann, M. M., Radke, J. B., Lucas, O., Nawas, J., Sibley, L. D., and White, M. W. (2010) Coordinated Progression through Two Subtranscriptomes Underlies the Tachyzoite Cycle of Toxoplasma Gondii. *PLoS One* 5, e12354

470. Radke, J. R., Behnke, M. S., Mackey, A. J., Radke, J. B., Roos, D. S., and White, M. W. (2005) The Transcriptome of Toxoplasma Gondii. *BMC Biology* 3, 26

471. Balaji, S., Babu, M. M., Iyer, L. M., and Aravind, L. (2005) Discovery of the Principal Specific Transcription Factors of Apicomplexa and Their Implication for the Evolution of the Ap2-Integrase DNA Binding Domains. *Nucleic Acids Research* 33, 3994-4006

472. Walker, R., Gissot, M., Croken, M. M., Huot, L., Hot, D., Kim, K., and Tomavo, S. (2012) The Toxoplasmanuclear Factor Tgap2xi-4 Controls Bradyzoite Gene Expression and Cyst Formation. *Molecular Microbiology* 87, 641-655

473. Walker, R., Gissot, M., Huot, L., Alayi, T. D., Hot, D., Marot, G., Schaeffer-Reiss, C., Van Dorsselaer, A., Kim, K., and Tomavo, S. (2013) Toxoplasma Transcription Factor Tgap2xi-5 Regulates the Expression of Genes Involved in Parasite Virulence and Host Invasion. *Journal of Biological Chemistry* 288, 31127-31138

474. Bhatti, M. M., Livingston, M., Mullapudi, N., and Sullivan, W. J., Jr. (2006) Pair of Unusual Gcn5 Histone Acetyltransferases and Ada2 Homologues in the Protozoan Parasite Toxoplasma Gondii. *Eukaryotic Cell* 5, 62-76

475. Nardelli, S. C., Che, F. Y., Silmon de Monerri, N. C., Xiao, H., Nieves, E., Madrid-Aliste, C., Angel, S. O., Sullivan, W. J., Jr., Angeletti, R. H., Kim, K., and Weiss,

L. M. (2013) The Histone Code of Toxoplasma Gondii Comprises Conserved and Unique Posttranslational Modifications. *MBio* 4, e00922-00913

476. Saksouk, N., Bhatti, M. M., Kieffer, S., Smith, A. T., Musset, K., Garin, J., Sullivan, W. J., Jr., Cesbron-Delauw, M. F., and Hakimi, M. A. (2005) Histone-Modifying Complexes Regulate Gene Expression Pertinent to the Differentiation of the Protozoan Parasite Toxoplasma Gondii. *Molecular and Cellular Biology* 25, 10301-10314

477. Blobel, G. (1985) Gene Gating: A Hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 82, 8527-8529

478. Burns, L. T., and Wente, S. R. (2014) From Hypothesis to Mechanism: Uncovering Nuclear Pore Complex Links to Gene Expression. *Molecular and Cellular Biology* 34, 2114-2120

479. Gordon, M. R., Pope, B. D., Sima, J., and Gilbert, D. M. (2015) Many Paths Lead Chromatin to the Nuclear Periphery. *Bioessays* 37, 862-866

480. Holden, J. M., Koreny, L., Obado, S., Ratushny, A. V., Chen, W. M., Chiang, J. H., Kelly, S., Chait, B. T., Aitchison, J. D., Rout, M. P., and Field, M. C. (2014) Nuclear Pore Complex Evolution: A Trypanosome Mlp Analogue Functions in Chromosomal Segregation but Lacks Transcriptional Barrier Activity. *Molecular Biology of the Cell* 25, 1421-1436

481. Gissot, M., Walker, R., Delhaye, S., Huot, L., Hot, D., and Tomavo, S. (2012) Toxoplasma Gondii Chromodomain Protein 1 Binds to Heterochromatin and Colocalises with Centromeres and Telomeres at the Nuclear Periphery. *PloS One* 7, e32671

482. Brooks, C. F., Francia, M. E., Gissot, M., Croken, M. M., Kim, K., and Striepen, B. (2011) Toxoplasma Gondii Sequesters Centromeres to a Specific Nuclear Region Throughout the Cell Cycle. *Proc. Natl. Acad. Sci. U.S.A.*, pp. 3767-3772

483. Bhatti, M. M., and Sullivan, W. J., Jr. (2005) Histone Acetylase Gcn5 Enters the Nucleus Via Importin-Alpha in Protozoan Parasite Toxoplasma Gondii. *Journal of Biological Chemistry* 280, 5902-5908

484. Dixon, S. E., Bhatti, M. M., Uversky, V. N., Dunker, A. K., and Sullivan, W. J., Jr. (2011) Regions of Intrinsic Disorder Help Identify a Novel Nuclear Localization Signal in Toxoplasma Gondii Histone Acetyltransferase Tggcn5-B. *Molecular and Biochemical Parasitology* 175, 192-195

485. Frankel, M. B., and Knoll, L. J. (2009) The Ins and Outs of Nuclear Trafficking: Unusual Aspects in Apicomplexan Parasites. *DNA and Cell Biology* 28, 277-284

486. Field, M. C., Koreny, L., and Rout, M. P. (2014) Enriching the Pore: Splendid Complexity from Humble Origins. *Traffic* 15, 141-156

487. Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., and Ma, B. (2012) Peaks Db: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Molecular & Cellular Proteomics* 11, M111 010587

488. Searle, B. C. (2010) Scaffold: A Bioinformatic Tool for Validating Ms/Ms-Based Proteomic Studies. *Proteomics* 10, 1265-1269

489. Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J. C., Mackey, A. J., Pinney, D. F., Roos, D. S., Stoeckert, C. J., Jr., Wang, H., and Brunk, B. P. (2008) Toxodb: An Integrated Toxoplasma Gondii Database Resource. *Nucleic Acids Research* 36, D553-556

490. Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009) Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin Alpha. *Journal of Biological Chemistry* 284, 478-485

491. de Monerri, N. C. S., Yakubu, R. R., Chen, A. L., Bradley, P. J., Nieves, E., Weiss, L. M., and Kim, K. (2015) The Ubiquitin Proteome of Toxoplasma Gondii Reveals Roles for Protein Ubiquitination in Cell-Cycle Transitions. *Cell Host & Microbe* 18, 621-633

492. Treeck, M., Sanders, J. L., Elias, J. E., and Boothroyd, J. C. (2011) The Phosphoproteomes of Plasmodium Falciparum and Toxoplasma Gondii Reveal Unusual Adaptations within and Beyond the Parasites&Apos; Boundaries. *Cell Host & Microbe* 10, 410-419

493. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nature Genetics* 25, 25-29

494. Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., and Bryant, S. H. (2013) Cdd: Conserved Domains and Protein Three-Dimensional Structure. *Nucleic Acids Research* 41, D348-352

495. Vanagas, L., Dalmasso, M. C., Dubremetz, J. F., Portiansky, E. L., Olins, D. E., and Angel, S. O. (2013) Epichromatin Is Conserved in Toxoplasma Gondii and Labels the Exterior Parasite Chromatin Throughout the Cell Cycle. *Parasitology* 140, 1104-1110

496. Bond, M. R., and Hanover, J. A. (2015) A Little Sugar Goes a Long Way: The Cell Biology of O-Glcnac. *Journal of Cell Biology* 208, 869-880

497. Ma, B., Simala-Grant, J. L., and Taylor, D. E. (2006) Fucosylation in Prokaryotes and Eukaryotes. *Glycobiology* 16, 158R-184R

498. Shen, B., Brown, K. M., Lee, T. D., and Sibley, L. D. (2014) Efficient Gene Disruption in Diverse Strains of Toxoplasma Gondii Using Crispr/Cas9. *Mbio* 5, e01114-01114

499. Al-Shareffi, E., Chaubard, J. L., Leonhard-Melief, C., Wang, S. K., Wong, C. H., and Haltiwanger, R. S. (2013) 6-Alkynyl Fucose Is a Bioorthogonal Analog for O-Fucosylation of Epidermal Growth Factor-Like Repeats and Thrombospondin Type-1 Repeats by Protein O-Fucosyltransferases 1 and 2. *Glycobiology* 23, 188-198

500. Valero-Gonzalez, J., Leonhard-Melief, C., Lira-Navarrete, E., Jimenez-Oses, G., Hernandez-Ruiz, C., Pallares, M. C., Yruela, I., Vasudevan, D., Lostao, A., Corzana, F., Takeuchi, H., Haltiwanger, R. S., and Hurtado-Guerrero, R. (2016) A Proactive Role of Water Molecules in Acceptor Recognition by Protein O-Fucosyltransferase 2. *Nature Chemical Biology* 12, 240-246

501. Ferguson, D. J., Birch-Andersen, A., Siim, J. C., and Hutchison, W. M. (1979) Ultrastructural Studies on the Sporulation of Oocysts of Toxoplasma Gondii. I. Development of the Zygote and Formation of the Sporoblasts. *Acta Pathologica et Microbiologica Scandinavica. Section B: Microbiology* 87B, 171-181

502. Iwamoto, M., Asakawa, H., Hiraoka, Y., and Haraguchi, T. (2010) Nucleoporin Nup98: A Gatekeeper in the Eukaryotic Kingdoms. *Genes to Cells* 15, 661-669

503. Casaravilla, C., Freire, T., Malgor, R., Medeiros, A., Osinaga, E., and Carmona, C. (2003) Mucin-Type O-Glycosylation in Helminth Parasites from Major Taxonomic Groups: Evidence for Widespread Distribution of the Tn Antigen (Galnac-Ser/Thr) and Identification of Udp-Galnac:Polypeptide N-Acetylgalactosaminyltransferase Activity. *Journal of Parasitology* 89, 709-714

504. Buscaglia, C. A., Campo, V. A., Frasch, A. C., and Di Noia, J. M. (2006) Trypanosoma Cruzi Surface Mucins: Host-Dependent Coat Diversity. *Nature Reviews: Microbiology* 4, 229-236

505. Ilg, T., Stierhof, Y. D., Craik, D., Simpson, R., Handman, E., and Bacic, A. (1996) Purification and Structural Characterization of a Filamentous, Mucin-Like

Proteophosphoglycan Secreted by Leishmania Parasites. *Journal of Biological Chemistry* 271, 21583-21596

506. Corfield, A. P. (2015) Mucins: A Biologically Relevant Glycan Barrier in Mucosal Protection. *Biochimica et Biophysica Acta* 1850, 236-252

507. Hulsmeier, A. J., Gehrig, P. M., Geyer, R., Sack, R., Gottstein, B., Deplazes, P., and Kohler, P. (2002) A Major Echinococcus Multilocularis Antigen Is a Mucin-Type Glycoprotein. *Journal of Biological Chemistry* 277, 5742-5748

508. Cancela, M., Santos, G. B., Carmona, C., Ferreira, H. B., Tort, J. F., and Zaha, A. (2015) Fasciola Hepatica Mucin-Encoding Gene: Expression, Variability and Its Potential Relevance in Host-Parasite Relationship. *Parasitology* 142, 1673-1681

509. Medeiros, A., Chiribao, M. L., Ubillos, L., Festari, M. F., Saldana, J., Robello, C., Dominguez, L., Calvete, J. J., and Osinaga, E. (2008) Mucin-Type O-Glycosylation in Mesocestoides Vogae (Syn. Corti). *International Journal for Parasitology* 38, 265-276

510. van Die, I., and Cummings, R. D. (2010) Glycan Gimmickry by Parasitic Helminths: A Strategy for Modulating the Host Immune Response? *Glycobiology* 20, 2-12

511. Dann, S. M., Okhuysen, P. C., Salameh, B. M., DuPont, H. L., and Chappell, C. L. (2000) Fecal Antibodies to Cryptosporidium Parvum in Healthy Volunteers. *Infection and Immunity* 68, 5068-5074

512. Tarazona, R., Blewett, D. A., and Carmona, M. D. (1998) Cryptosporidium Parvum Infestion in Experimentally Infected Mice: Infection Dynamics and Effect of Immunosuppression. *Folia Parasitologica* 45, 101-107

513. Wiesmuller, K. H., Jung, G., Gillessen, D., Loffl, C., Bessler, W. G., and Boltz, T. (1991) The Antibody Response in Balb/C Mice to the Plasmodium Falciparum Circumsporozoite Repetitive Epitope Covalently Coupled to Synthetic Lipopeptide Adjuvant. *Immunology* 72, 109-113

514. BenMohamed, L., GrasMasse, H., Tartar, A., Daubersies, P., Brahimi, K., Bossus, M., Thomas, A., and Druilhe, P. (1997) Lipopeptide Immunization without Adjuvant Induces Potent and Long-Lasting B, T Helper, and Cytotoxic T Lymphocyte Responses against a Malaria Liver Stage Antigen in Mice and Chimpanzees. *European Journal of Immunology* 27, 1242-1253

515. BenMohamed, L., Krishnan, R., Auge, C., Primus, J. F., and Diamond, D. J. (2002) Intranasal Administration of a Synthetic Lipopeptide without Adjuvant Induces Systemic Immune Responses. *Immunology* 106, 113-121

CURRICULUM VITAE