THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Structural characterisation of glycosylation of paramyxovirus attachment and fusion proteins by mass spectrometry

Cassandra Lee Pegg

BHlthSc., MPH., MMolBiol.

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2016

School of Chemistry and Molecular Biosciences

Abstract

The family *Paramyxoviridae* (paramyxovirus) contains several significant human and animal pathogens. Represented within this family are human respiratory syncytial virus (hRSV) and Newcastle disease virus (NDV). The former contributes significantly to severe respiratory tract disease in infants, children and immunocompromised individuals. At present, highly efficacious therapeutics or safe and effective vaccines are not available for hRSV. NDV is the causative agent of Newcastle disease, afflicting a wide range of avian species. The desire to study NDV is due not only to the significant economic impact it has on the poultry industry worldwide, but also its potential use as an oncolytic agent and vaccine vector in humans and animals. Additionally, findings on NDV may be translated to closely related viruses that cause disease in humans, such as parainfluenza viruses.

As outlined in Chapter 1 of this thesis, the infectious processes of all members of this family are driven by two major membrane glycoproteins whose ectodomains project from the viral envelope: the fusion (F) and attachment glycoproteins. The F glycoprotein is responsible for viral entry by means of fusion with host cell membranes while the variable attachment glycoproteins, haemagglutinin, haemagglutinin-neuraminidase (HN) and major surface glycoprotein (G), are involved in viral attachment to host cells. Previous studies have shown that altering the glycosylation profile of these proteins can modulate the ability of the virus to infect host cells and stimulate the host immune system. As yet, site-specific glycan heterogeneity of hRSV and NDV surface glycoproteins has not been defined at a chemical level. As described in Chapter 2, reverse-phase liquid chromatography tandem mass spectrometry (MS) strategies were implemented to characterise intact glycopeptides from the attachment and F glycoproteins of hRSV and NDV. Site-occupancy and monosaccharide compositions at a given site were determined using collision-induced dissociation, higher-energy collision dissociation, electron-transfer dissociation and electron-transfer dissociation combined with collision dissociation. As described in Chapter 3, a spectral processing program was developed called OxoExtract to aid identification of intact glycopeptides that were fragmented with higher-energy collision dissociation.

The F and HN proteins of NDV were derived from virions propagated in embryonic eggs. Analyses of HN in Chapter 4 revealed high mannose N-linked glycans and complex or hybrid N-linked glycans that were variably fucosylated, sialylated and sulfated or phosphorylated. In total 63, 58, and 37 glycans were identified at sites N341, N433 and N481, respectively. In addition, a previously undocumented O-linked glycosylation site was identified in the stalk domain of the protein. Observed glycans from NDV F described in Chapter 5 were mainly high mannose, containing variations of five to nine mannose

residues across four sites, N85, N191, N366 and N471. There was also evidence of fucosylated complex or hybrid glycans at site N191, which is the first site-specific description of such glycans on NDV F. Although this work was not completed with virions from naturally infected tissue or cells, the results presented herein on NDV HN and F may help elucidate mechanisms of viral glycoprotein synthesis. The observation of hybrid or complex glycans on F and HN suggests that at least some protein species are transported through the Golgi apparatus. This supports the theory that the two proteins may interact before reaching the cell surface. The stalk region of HN is thought to play an important role in triggering the F protein to induce fusion. Given the position of the observed O-linked site in the stalk domain of HN, it could be postulated that O-linked glycans impact oligomerisation of HN or the interaction between HN and F. These results will form the basis for future studies of the distribution of glycan structures across glycosylation sites and may increase the understanding of the role glycosylation plays in the functionality of HN and F. Furthermore, it will enable comparisons to be made of HN and F from other species of NDV and paramyxoviruses.

Soluble forms of hRSV F and G proteins were recombinantly produced in human embryonic kidney 293 FreeStyle™ cells and analysed with the aforementioned MS strategies. The work in Chapter 6 focussed on a homotrimeric form hRSV F, which has been extensively structurally characterised in pursuit of drug and vaccine design objectives. It was revealed that all five N-linked sites can be occupied with the identification of 20, 19, 7, 24 and 70 different glycans at N-linked sites N27, N70, N116, N126 and N500, respectively. Many of the observed N-linked glycans exhibited fragmentation characteristics consistent with N-acetylhexosamine units, which could potentially represent GalNAcβ1–4GlcNAc or LacdiNAc motifs. These units are not typically observed in mammals and are potentially immunogenic. Moreover, O-linked glycosylation of F is described for the first time. This work also demonstrated extensive O-linked glycosylation of G in Chapter 7. A total of 28 O-linked sites were identified with attached compositions presumed to be Tn, T and mono- and di-sialylated T antigens. Twelve N-linked glycans were also identified at site N135 of G. Although the work was not completed with proteins derived from virions, the techniques applied herein can be used to investigate native forms of G and F. The results also provide the first step in the elucidation of site-specific and compositional differences of glycans from F and G produced in different cell lines.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

<u>Peer-reviewed papers</u>

C. L. Pegg, C. Hoogland and J. J. Gorman (2016). Site-specific glycosylation of the Newcastle disease virus haemagglutinin-neuraminidase. Submitted to the Glycoconjugate Journal (accepted for publication DOI: 10.1007/s10719-016-9750-7).

Conference Abstracts

C. L. Pegg, C. Hoogland, S. M. Johnson, C. C. Gonzalez, M. E. Peeples and J. J. Gorman. Poster presentation: Characterisation of glycosylation of paramyxovirus surface glycoproteins by mass spectrometry. Twenty-third International Symposium on Glycoconjugates, Split, Croatia (September, 2015). *Received a Susan Hamilton Travel Award for Science Education and a University of Queensland Graduate School International Travel Award (GSITA).*

C. L. Pegg, C. Hoogland and J. J. Gorman. Poster presentation: Characterisation of glycosylation of the Newcastle disease virus haemagglutinin-neuraminidase protein. Australian Society for Medical Research Student Conference, Brisbane, Australia (April, 2015).

C. L. Pegg, C. Hoogland and J. J. Gorman. Lightening Talk (Invited) and Poster presentation: Characterisation of glycosylation of the Newcastle disease virus haemagglutinin-neuraminidase surface glycoprotein. Lorne Proteomics Symposium, Lorne, Australia (February, 2015). *Awarded a student poster prize from the conference organisers.*

C. L. Pegg, C. Hoogland, S.M. Johnson, C. C. Gonzalez, M. E. Peeples and J. J. Gorman. Poster presentation: Characterisation of glycosylation of paramyxovirus surface glycoproteins by mass spectrometry. Eleventh International Symposium on Mass Spectrometry in the Health and Life Sciences (August, 2014). *Received a student travel award from the conference organisers and a QIMR Berghofer Travel Award.*

C. L. Pegg, C. Hoogland, S. M. Johnson, C. C. Gonzalez, M. E. Peeples and J. J. Gorman. Poster presentation: Characterisation of glycosylation of paramyxovirus surface glycoproteins by mass spectrometry. Australian Society for Medical Research Student Conference, Brisbane, Australia (April, 2014).

C. L. Pegg, C. Hoogland, S. M. Johnson, C. C. Gonzalez, M. E. Peeples and J. J. Gorman. Poster presentation: Deciphering glycopeptides using higher-energy collisional dissociation and electron-transfer dissociation mass spectrometry. Lorne Proteomics Symposium, Lorne, Australia (February, 2014). *Awarded a student poster prize from the conference organisers.*

From the above manuscript, the spectral processing program was incorporated as Chapter 3. The analysis of glycosylation of the haemagglutinin-neuraminidase protein of Newcastle disease virus was incorporated as Chapter 4.

Contributor	Statement of contribution
Pegg, C. L. (Candidate)	Conceived and designed experiments (50%) Performed experiments (100%) Analysed data (95%) Defining rules for a spectral processing program (95%) Wrote and edited the paper (60%)
Hoogland, C.	Defining rules for a spectral processing program (5%) Translating prescribed rules into a Java application (100%)
Gorman, J. J.	Conceived and designed experiments (50%) Contributed reagents and materials (100%) Analysed data (5%) Wrote and edited the paper (40%)

Contributions by others to the thesis

<u>All research carried out:</u>

Professor Jeffrey Gorman contributed to the conception and design of projects, critical revision of results and drafting and corrections of all written work.

<u>Development of in-house spectral processing program:</u>

Dr Christine Hoogland translated prescribed rules into a Java application.

<u>Training and technical assistance:</u>

Dr Keyur Dave and Dr Marcus Hastie provided training and technical assistance on the mass spectrometers. Dr Rosa Viner provided expert advice on mass spectrometry acquisition strategies. Dr Keyur Dave and Dr Madeline Headlam provided training in techniques for protein analysis. Dr Christine Hoogland and Emma Norris provided advice for bioinformatic searches. Dr Keyur Dave, Emma Norris and Dr Marcus Hastie provided critical revision of manuscripts for publication and the thesis document.

<u>Provision of proteins:</u>

Purified Newcastle disease virion stocks were provided by Professor Jeffrey Gorman.

Recombinant attachment and fusion proteins of human respiratory syncytial virus were provided by Professor Mark Peeples.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Jeffery Gorman for his support, enthusiasm and valuable mentorship. In addition to establishing a meaningful and exciting project he determinedly ensured that I had access to state-of-the-art mass spectrometers and encouraged me to take my own path.

I would also like to thank my associate supervisor Professor Paul Young for his helpful comments and for sharing his insights into the viruses.

To the Research Administration staff at the School of Chemistry and Molecular Biosciences, in particular, Dr Abigail Downie and Jill Sheridan, thank you for your guidance and for helping me to navigate the paperwork and formalities of the degree.

Professor Andrew Boyd and Leanne Cooper, thank you for introducing me to the Eph/ephrin system. It was an incredibly interesting project, which was made all the more enjoyable by working with you both.

Completing this work would have been all the more difficult were it not for the support and encouragement from members, past and present, of the Protein Discovery Centre. I gratefully thank Dr Keyur Dave for his guidance and training on the mass spectrometers, particularly during the initial phase of my studies. I would also like to thank Dr Christine Hoogland, who without hesitation assisted me as I jumped into the complex world of glycoproteomics. I would also like to thank Dr Keyur Dave, Dr Madeleine Headlam, Dr Marcus Hastie and Emma Norris for the technical advice they provided and for the numerous informative discussions about the intricacies of spectra. Emma, thank you for continually volunteering to edit papers and chapters of this thesis. To everyone in the lab, thank you for the shared food and lunchtime conversations and for the advice, both scientific and life.

Lastly, I would like to thank my friends and family. To the girls, thank you for your endless cheer and making it easy to forget the hard times.

Keywords

Paramyxoviridae, Newcastle disease virus, human respiratory syncytial virus, fusion glycoprotein, haemagglutinin-neuraminidase, major surface glycoprotein, glycosylation, mass spectrometry, higher-energy collision dissociation, electron-transfer dissociation

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060101, Analytical Biochemistry, 20%

ANZSRC code: 060199, Biochemistry and Cell Biology not elsewhere classified, 20%

ANZSRC code: 060109, Proteomics and Intermolecular Interactions (excl. Medical Proteomics), 60%

Fields of Research (FoR) Classification

FoR code: 0601, Biochemistry and Cell Biology, 70%

FoR code: 0605, Microbiology, 30%

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A list of abbreviations used in this thesis is provided below. Standard abbreviations for units of measurement and amino acids have not been defined.

Abbreviation	Definition
4HB	Four-helix bundle
β4GalNAc-T	β1–4–N-acetylgalactosaminyltransferase
β4Gal-T	β1–4–galactosyltransferase
aa	Amino acid
AGC	Automatic gain control
CA6	Carbonic anhydrase VI
CCD	Central conserved domain
CDV	Canine distemper virus
CID	Collision-induced dissociation
CT	Cytoplasmic tail
dHex	Deoxyhexose
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
DTT	Dithiothreitol
EIC	Extracted ion chromatogram
ER	Endoplasmic reticulum
ETciD	Electron-transfer dissociation with supplemental collision-induced dissociation
ETD	Electron-transfer dissociation
EThcD	Electron-transfer dissociation with supplemental higher-energy collision dissociation
ETnoD	Electron-transfer no dissociation
FDR	False discovery rate
F	Fusion glycoprotein
F_0	The fusion glycoprotein precursor
F_1	Carboxyl-terminal subunit of the fusion glycoprotein
F_2	Amino-terminal subunit of the fusion glycoprotein
FP	Fusion peptide
Fuc	Fucose
FWHM	Full width at half-maximum
G	Major surface glycoprotein
GAG	Glycosaminoglycans
Gal	Galactose
GalNAc	N-acetylgalactosamine
Glc	Glucose
GlcNAc	N-acetylglucosamine

GlcNAcT-I	N-acetylglucosaminyltransferase-I
H	Hydrogen
HA	Haemagglutinin
HBD	Heparin-binding domain
HCD	Higher-energy collision dissociation
HCD-pd-	Higher-energy collision dissociation-product-dependent-
HCV	Hepatitis C virus
HEK	Human embryonic kidney
Hep-2	Human epithelial type 2
HeV	Hendra virus
Hex	Hexose
HexA	Hexuronic acid
HexNAc	N-acetyl-hexosamine
hGC	Human chorionic gonadotropin
hMPV	Human metapneumovirus
HN	Haemagglutinin-neuraminidase
hPIV	Human parainfluenza virus
HPLC	High performance liquid chromatography
HR	Heptad repeat
hRSV	Human respiratory syncytial virus
HSV	Herpes simplex virus
IAA	Iodoacetamide
KDN	Ketodeoxynonulonic acid
LacdiNAc	(GalNAcβ1–4GlcNAc) extension
LacNAc	(Galβ1–4GlcNAc) extension
LTQ	Linear ion trap
m/z	Mass-to-charge ratio
Man	Mannose
MDBK	Madin-Darby bovine kidney
MeV	Measles virus
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MuV	Mumps virus
MW	Molecular weight
NCE	Normalised collision energy
ND	Newcastle disease
NDV	Newcastle disease virus
NeuAc	N-acetylneuraminic acid
NeuGc	N-glycolyl-neuraminic acid
NK	Natural killer cells
nUHPLC	Nano-ultra-high pressure liquid chromatography
NiV	Nipah virus
pep27	Soluble 27 amino acid peptide lost from hRSV fusion protein
Phos	Phosphate

PNGase F	Peptide-N4-(N-acetyl-ß-glucosaminyl) asparagine amidase
ppGalNAcTs	N-acetylgalactosaminyltransferases
PSM	Peptide-to-spectrum match
PTM	Post-translational modification
sF	Soluble fusion glycoprotein
S/N	Signal-to-noise
SD	Stalk domain
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeV	Sendai virus
sG	Soluble major surface glycoprotein
SLe^X	Sialyl Lewis X
SORL1/LR11	Sorting protein-related receptor
SP	Signal peptide
Sulf	Sulfate
SV5	Simian virus 5
TEAB	Triethylammonium bicarbonate buffer
TM	Transmembrane
V4-QLD	QLD/66 isolate of Newcastle disease virus
V4-VAR	Variant of the QLD/66 isolate of Newcastle disease virus

Chapter 1: Introduction to the structure and function of glycosylation with reference to paramyxovirus surface proteins

1.1   PROTEIN GLYCOSYLATION

Protein glycosylation is a ubiquitous post-translational modification (PTM) whereby oligosaccharide structures known as glycans are covalently linked to amino acid side chains (1). It confers additional layers of complexity to the structural and functional properties of proteins and can have a profound influence on both normal and irregular biological processes (2). In mammals, glycosylation has been shown to play a role in cell signalling, inflammation, immune responses and several disease states as well as protein activity and folding (3-5). The impact of glycosylation has also been highlighted in pharmaceutical and biotech industries where glycans can affect the stability, solubility, bioactivity and pharmacokinetic properties of protein therapeutics (6-8). Characterising the glycosylation patterns of glycoproteins has therefore become an important aspect of biological and biomedical studies.

Glycans are composed of monosaccharides which are the simplest forms of carbohydrates (9). Although other terms such as sugars, carbohydrates and oligosaccharides can be used to describe monosaccharide compositions attached to proteins, the term glycan has been chosen as the preferred term in this thesis. Commonly observed monosaccharides found in mammals, mannose (Man), glucose (Glc), galactose (Gal), fucose (Fuc), N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc) and N-acetylneuraminic acid (NeuAc) are presented in **Figure 1-1**, following the Symbol Nomenclature for Glycans system (10). The isomers Man, Glc and Gal contain a six-carbon backbone as do the isomers GalNAc and GlcNAc, but the latter two contain an acetylated amino group. Fuc is also a six-carbon monosaccharide but lacks a hydroxyl group at one carbon while NeuAc belongs to a family of nine-carbon acidic monosaccharides called sialic acids.

The great diversity observed in glycan structures is due to the non-template driven biosynthetic processing of glycans and the branching or non-linear nature of the glycosidic linkages (9). Glycosidic linkages can be formed in two different configurations between the anomeric carbon of one monosaccharide and a hydroxyl group of another monosaccharide, with the configurations

defined as α and β linkages. The anomeric carbons (C) are at C-1 or C-2, depending on the monosaccharide and can be attached to several different hydroxyl groups on the other monosaccharide. The bond configurations (α and β), anomeric carbon numbers (C-1 or C-2) and positions of the hydroxyl groups (defined by the carbon it is attached to) are used to name glycosidic linkages (for example, β1-4 or α2-3) (9). The final glycan structures are dependent on cell type and the physiological status of the cell (11, 12). The dynamic nature of glycosylation means both native and recombinant glycoproteins can be observed as many different glycoforms. Glycoforms of a single protein can differ in the glycan structures attached at a specific site (defined as microheterogeneity) or in the proportion of sites that are occupied (defined as macroheterogeneity) (13). The complex nature of protein glycosylation and subsequent structural diversity of glycoproteins present significant challenges when defining the glycosylation profile of glycoproteins.

Figure 1-1. Symbol representation of some common mammalian monosaccharides. Monosaccharides that have been designated circles or squares are isomeric and represent hexoses and N-acetylhexosamines, respectively, following the Symbol Nomenclature for Glycans system (10). If stereochemistry cannot be determined the shape is filled white.

The work herein addresses two major types of extracellular glycosylation, namely, N-linked and O-linked. These types of glycosylation are differentiated by mechanisms of biosynthesis and the amino acid side chains to which they are conjugated. These two specific types of glycosylation will be discussed in more detail in the following sections.

1.1.1 N-linked glycosylation

N-linked glycans are predominately conjugated to the side-chain amide nitrogen of a peptide bound Asn in the amino acid consensus Asn-Xaa-Ser/Thr (where Xaa is not proline). N-linked glycans share a common Manα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAcβ1 trimannosylchitobiose core (**Figure 1-2**) which originates from a $Glc_3Man_9GlcNAc_2$ precursor structure (14). In eukaryotes, step-wise production of this precursor structure begins in the cytoplasm while it is attached to a dolichyl pyrophosphate carrier in the membrane of the endoplasmic reticulum (ER) (14-16). During production, it is flipped into the lumen of the ER and once the structure is complete it is transferred *en bloc* to nascent polypeptides. Subsequent processing of the tetradecasaccharide occurs in the ER and Golgi apparatus through the actions of glycosidases and glycosyltransferases (**Figure 1-3**). The degree of glycan processing during glycoprotein biosynthesis places the final N-linked glycan in one of three structural categories (16).

The high mannose (or oligomannose) class contains five to nine mannose residues and arises from trimming of the $Glc_3Man_9GlcNAc_2$ precursor to a $Man_8GlcNAc_2$ oligosaccharide in the ER or subsequent trimming to a $Man_5GlcNAc_2$ structure in the *cis*-Golgi. This $Man_5GlcNAc_2$ (or Man_5) structure (illustrated in **Figure 1-2**) is the substrate for N-acetylglucosaminyltransferase-I (GlcNAcT-I) and a required intermediate for the production of the remaining two categories, hybrid or complex glycans (15). During maturation of the glycans GlcNAcT-I adds a GlcNAc residue to the common core α1–3 mannose. If the peripheral α-mannose residues are not removed from the α1–6 arm of the Man_5 structure, a hybrid glycan results (15, 16). Removal of the peripheral α-mannose residues in the medial-Golgi enables the addition of a GlcNAc to α1–6 arm of the common core and results in the precursor for complex biantennary structures. The antennae can be extended through the addition of Gal and NeuAc or non-monosaccharide substituents such as sulfate (Sulf) or phosphate (Phos) in the *trans*-Golgi (**Figure 1-3**) (15, 16).
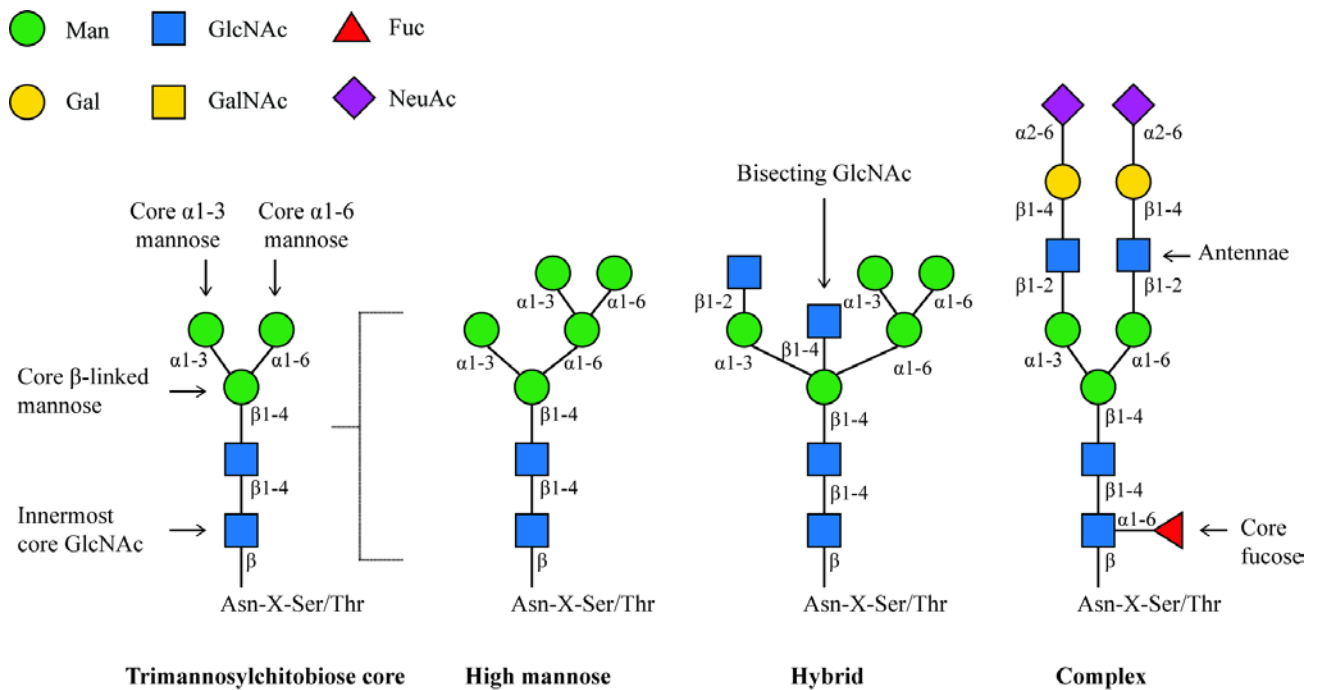
Figure 1-2. Structural categories of N-linked glycans and their common trimannosylchitobiose core. The common trimannosylchitobiose core has been identified with commonly used terms for some of the constituent monosaccharides. Structures from the three main categories of N-linked glycans, high mannose (with a $Man_5GlcNAc_2$ structure illustrated), hybrid and complex have been represented. Bisecting GlcNAc, branching and core fucosylation have also been presented.

During processing N-linked glycans can also be fucosylated, typically in a α1–6 linkage on the innermost core GlcNAc (**Figure 1-2**). Hybrid and complex N-linked glycans may also have a bisecting GlcNAc residue attached to the core β-linked mannose. Glycosylation site accessibility on the folded protein has been found to correlate with more highly processed glycans, glycan branching and core fucosylation (17, 18). Thus, the mature glycoproteins can bear a mixture of high mannose, hybrid or complex glycans. More recently, an unconventional glycan processing pathway has been described, prompted by the growing observation of truncated N-linked structures in mammals (19-23). These glycans stem from trimming of hybrid or complex glycans and the truncated N-glycans containing one to four mannose residues (Man_{1-4}) are described as paucimannose and those without Man residues are described as chitobiose core type glycans (19).
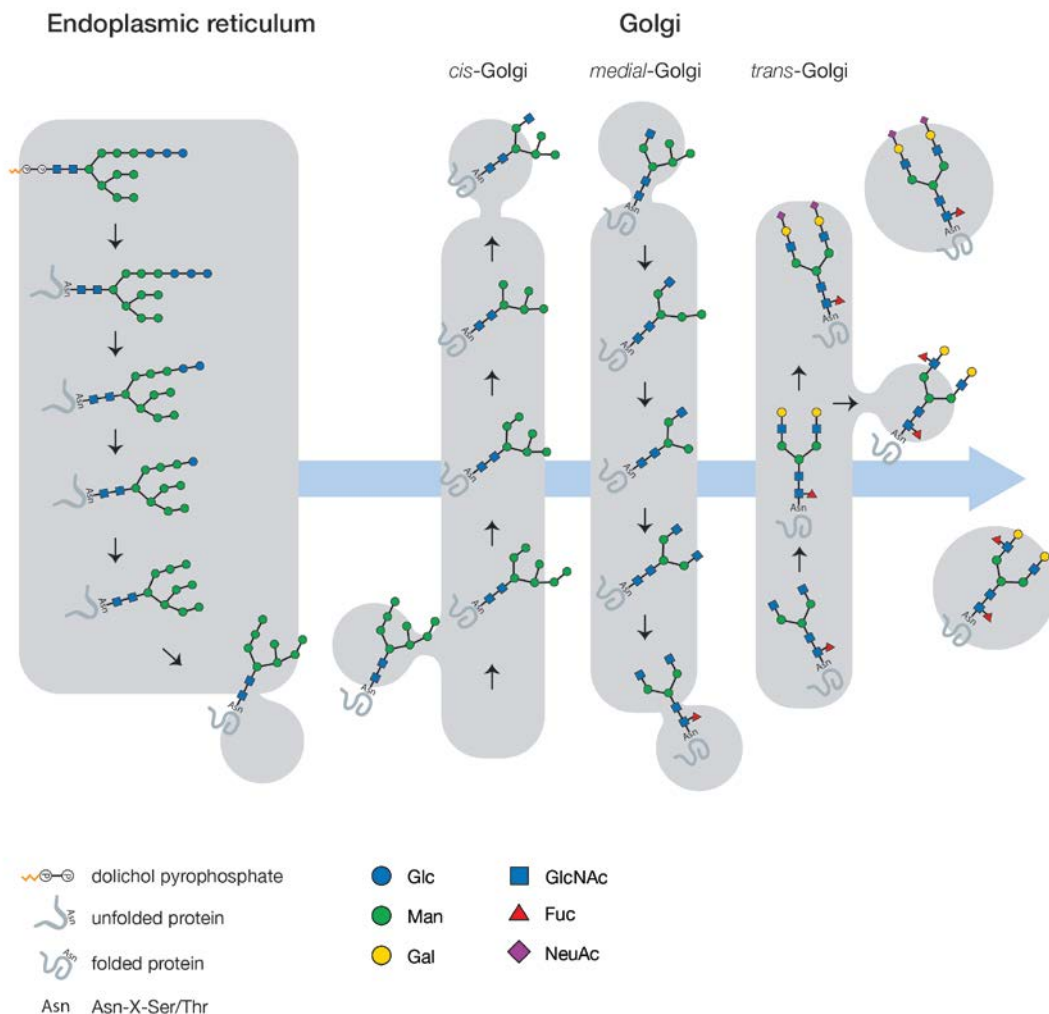
Figure 1-3. Processing of N-linked glycans on glycoproteins. Remodelling of the $Glc_3Man_9GlcNAc_2$ structure takes place after it is transferred to a nascent polypeptide chain in the ER (top left). During trimming in ER the glycoprotein undergoes folding and oligomeric assembly if required, before moving to the Golgi apparatus. Further glycan processing takes place in the *cis-*, medial- and *trans*-Golgi before secretion of the final glycoprotein. Please refer to the text for a description of the pathway. Not all glycan structures and pathways are represented. Figure adapted from (1).

1.1.2 O-linked glycosylation

Rather than *en bloc* transfer of a polysaccharide structure; O-linked glycosylation results from the sequential addition of monosaccharides to oxygen, typically from the side chain hydroxyl group of Ser and Thr residues. Unlike N-linked glycosylation, there is no amino acid consensus sequence to predict the addition of O-linked glycans to Ser and Thr residues (24). Several types of O-linked modifications exist, one highly dynamic form typically occurs in the nuclear and cytoplasmic compartments of a cell through the addition of β-linked GlcNAc (O-GlcNAc). The core GlcNAc is

not normally extended and is associated with intracellular signalling and the regulation of pathways in disease states such as diabetes (25). More recently, this type of glycosylation has also been observed on extracellular proteins (26).

Alternatively, another type of O-linked glycosylation, generally referred to as mucin-like glycosylation, is derived from the ER-Golgi pathway (24). It is initiated by the transfer of GalNAc (O-GalNAc) in an α-linkage to acceptor substrates by a large family of polypeptide-N-acetylgalactosaminyltransferases (ppGalNAcTs) (24). The base GalNAc residue (Tn antigen) is built on to form eight different cores, four of which are illustrated in (**Figure 1-4**). The core-1 structure (T antigen) is the most common core observed in mucin-like glycosylation and can be extended to form long branching structures (24). The addition of NeuAc residues to core-1 forms mono- and di-sialylated structures, described as sialyl-T and disialyl-T antigens, respectively. The addition of GlcNAc to core-1 forms a core-2 structure which can be elongated with Gal on the terminal GlcNAc. The unusual addition of GalNAc to the terminal GlcNAc on a core-2 structure forms the less commonly observed LacdiNAc motif (24). Core-3 and core-4 structures are not commonly observed but are produced by certain mucin-secreting tissues (24). Mucin-like or dense O-GalNAc glycosylation is typically found on regions of glycoproteins that are rich in Ser, Thr and Pro and can serve to maintain the conformation of a protein, act a barrier for mucosa and shield proteins or cellular surfaces from proteolytic digestion (27-30). However, O-GalNAc glycosylation is also found in isolated positions on non-mucinous proteins (31, 32). These isolated O-linked sites are predicted to have a range of biological functions including regulating protein cleavage and potential involvement in several disease states (31, 33, 34).
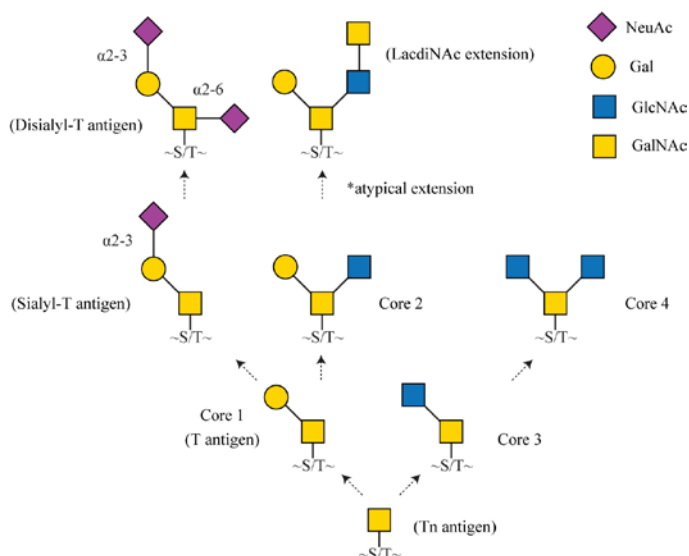
Figure 1-4. Biosynthesis of mucin-like O-linked glycans.

1.1.3 Viral protein glycosylation

The importance of glycosylation in host-pathogen interactions is well established (29, 35, 36). With respect to viral infection, glycosylation of viral surface proteins has been implicated in attachment and entry, induction of immune responses and evasion of host-immune responses (37-40). Furthermore, viral glycoproteins can be used as targets for vaccines and antiviral therapies (41-44). Glycosylation is specific to various organisms, tissues and cell lines (22, 45, 46), as such, production of viral proteins in different expression systems can result in substantially different glycosylation profiles (47-50). Defining glycosylation of viral surface proteins produced *in vivo* or *in vitro* is therefore important for the elucidation of host-virus interactions and for the design of viral therapeutics. This thesis focuses on characterising site-specific N- and O-linked glycosylation of the attachment and fusion (F) proteins from two members of the *Paramyxoviridae* (paramyxovirus) family, human respiratory syncytial virus (hRSV) and Newcastle disease virus (NDV). *Paramyxoviridae* replication occurs in cytoplasm of the host cell and the viruses use existing host biosynthetic pathways, including glycosylation pathways, to produce progeny virions incorporating glycosylated proteins (51). Glycoproteomic studies of viral proteins can be very challenging and adequate quantities of the proteins can be difficult to obtain from natural systems, and in some cases, even from *in vivo* models or from virions propagated in cells lines (49). Therefore, studies often involve viral proteins produced from virions in non-natural systems or from recombinant methods, as is the case in this thesis.

1.2 PARAMYXOVIRUSES

1.2.1 Taxonomy and global burden

The family *Paramyxoviridae*, within the order *Mononegavirales*, is a group of enveloped viruses possessing a negative-sense, single-stranded RNA genome. This family of viruses afflicts a diverse range of hosts and can be divided into two main subfamilies, *Pneumovirinae* and *Paramyxovirinae* (51) (**Figure 1-5**). Within *Pneumovirinae* there are two significant human pathogens, hRSV and human metapneumovirus (hMPV), while *Paramyxovirinae* contains the human pathogens mumps virus (MuV), measles virus (MeV) and human parainfluenza viruses (hPIV 1-4), PIV 5 and the zoonotic viruses Hendra (HeV) and Nipah (NiV). Also classified within *Paramyxovirinae* are Sendai virus (SeV) and NDV, causing disease in mouse and bird species, respectively.

Collectively, these viruses contribute significant disease and economic burdens worldwide. Effective vaccines have been licensed for two human pathogens, MeV and MuV, but outbreaks in vaccinated and unvaccinated individuals continue to add to childhood morbidity and mortality rates worldwide (52-54). An effective human treatment or vaccine is not available for HeV or NiV, which cause severe disease in humans with high case fatality rates (55). A vaccine is available for the amplifying host of HeV but measures remain in place to limit human exposure to HeV and NiV pending treatments for other livestock and humans (56). Importantly, effective vaccines for hRSV, hMPV and hPIV 3, which are substantial contributors to respiratory disease in children and immunocompromised individuals, remain elusive (57-59). Besides the serious risk of disease to humans, the paramyxoviruses HeV, NiV and NDV, also pose an economic threat through the infection of livestock. Outbreaks of these viruses not only result in the death of diseased stock but also necessitate the controlled destruction of healthy animals and reduced movement of stock (60, 61).
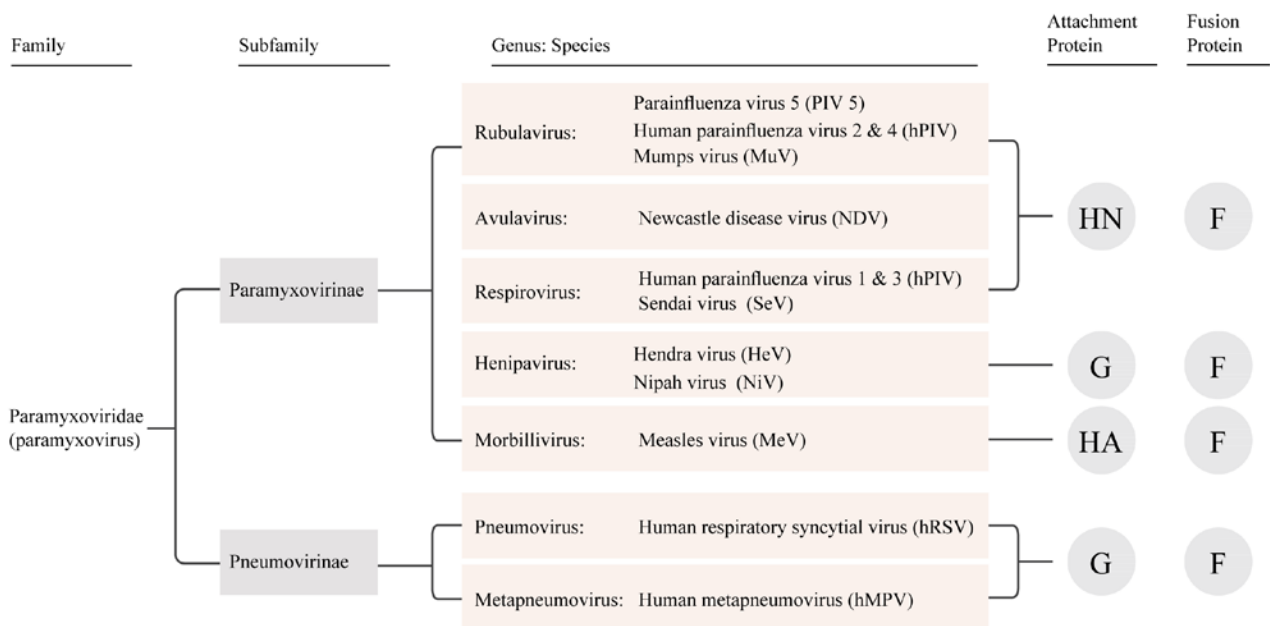
Figure 1-5. Overview of paramyxovirus taxonomy. Not all genera and species have been listed and branching does not represent evolutionary distance. The fusion (F) and attachment proteins, haemagglutinin (HA), haemagglutinin neuraminidase (HN) or major surface glycoprotein (G) of genera of interest have been noted. Adapted from (39).

1.2.2 Paramyxovirus structure

Viruses belonging to the paramyxovirus family have six similar proteins (62) (**Figure 1-6**), two of which are known to be glycosylated. These are the surface glycoproteins that project from the lipid envelope, known as the F and attachment glycoproteins. The F glycoprotein is common to all species discussed above and induces fusion of the host cell membrane and the viral envelope (39, 63). The attachment glycoprotein belonging to each virus can be one of three protein types; haemagglutinin (HA), haemagglutinin neuraminidase (HN) or major surface glycoprotein (G). These proteins bind to host cells via specific carbohydrate residues or host cell proteins and may also facilitate viral entry (37, 64). The matrix protein is found on the inner surface of the lipid bilayer, stabilising the virion structure and directing assembly and budding. Within virions, the RNA genome is encapsidated by the nucleocapsid protein, forming a helical nucleocapsid structure. The phosphoprotein and large polymerase protein, associate with the nucleocapsid and are responsible for polymerase activity (51, 65).
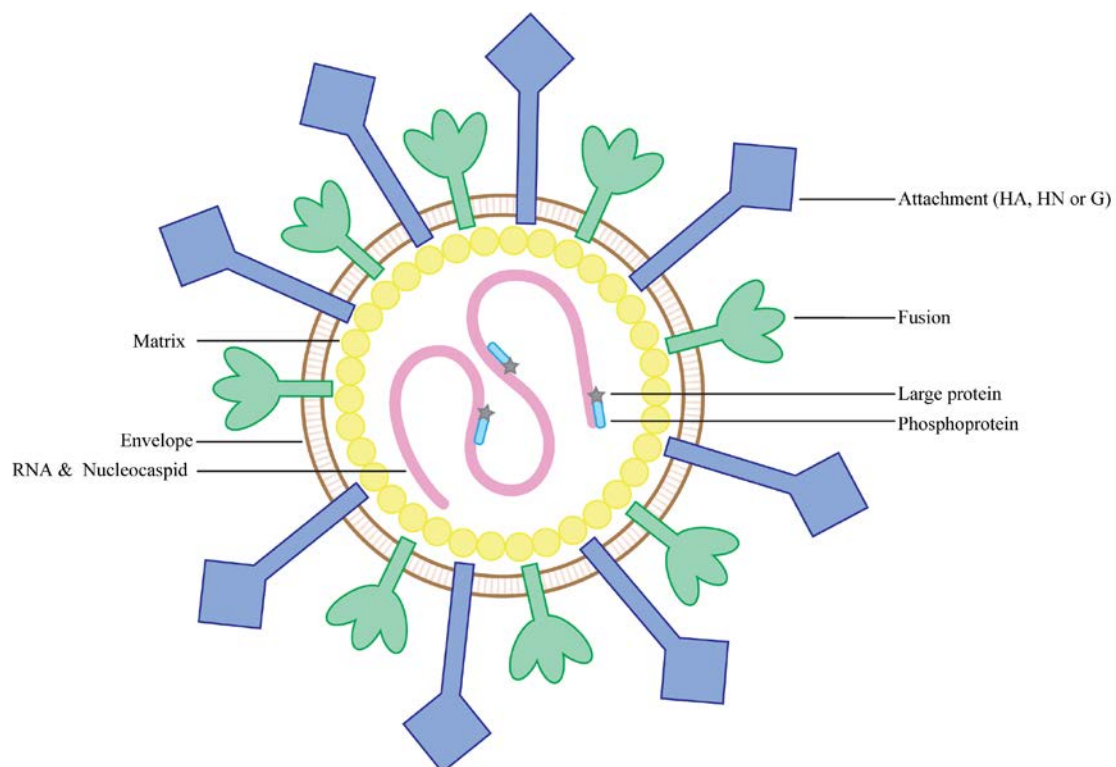
Figure 1-6. Schematic illustrating the structure of a paramyxovirus. The schematic (not to scale) identifies six proteins common to paramyxoviruses. Please refer to the text for a description of each protein. The attachment protein represents haemagglutinin (HA), haemagglutinin neuraminidase (HN) or major surface glycoprotein (G).

1.2.3 Paramyxovirus attachment and fusion surface glycoproteins

The infectious processes of all members of the paramyxovirus family are driven by the attachment and F proteins. The protein ectodomains are displayed on the external aspects of the lipid bilayers of the viruses, as such; they are also the major antigenic components of the viruses and the main target for vaccines and neutralising therapeutics (39). The F protein is a homotrimeric type I integral membrane glycoprotein and comparisons of the crystal structures of F from NDV, hPIV3, PIV5 and hRSV reveal that they exhibit relatively conserved molecular architecture (63). The F protein is produced as a precursor (F_0), which is cleaved into an active form by host proteases, resulting in disulfide-linked polypeptide chains derived from the amino-terminal (F_2) and the carboxyl-terminal (F_1) domains. Cleavage of F_0 into the F_2 and F_1 subunits releases a hydrophobic peptide or "fusion peptide" (FP) that inserts into target membranes after the fusion process has been initiated (66). What triggers the F protein and the exact process behind fusion has not been determined and may differ for each virus, but F undergoes a conformational change from a metastable prefusion state to a stable post-fusion state (67). This fusion process may also produce syncytia if the F protein is expressed on the infected cell's surface (68).

In contrast, the attachment glycoproteins of paramyxoviruses have a diversity of functional attributes. Viruses bearing HA can agglutinate red blood cells while those bearing HN additionally display neuraminidase (sialidase) activity (51). Viruses expressing G exhibit no haemagglutinating or sialidase activities (51). These glycoproteins are oligomeric type II integral membrane proteins and comparisons of the atomic structures of HA, HN and G from *Paramyxovirinae* found the secondary and overall structures were highly conserved (63). Within this subfamily the G and HA proteins bind to cellular receptors while HN recognises and binds to sialylated glycans (69). After receptor binding it is thought the attachment proteins of *Paramyxovirinae* trigger F and induce fusion. Such fusion promotion usually requires expression of F and the stalk domain of the homotypic attachment protein (39, 70). Unlike HA, HN and G from *Paramyxovirinae* the G protein from *Pneumovirinae* is less conserved and there is conflicting evidence as to whether the protein binds a specific receptor *in vivo* and if G is required for infectious processes (71-73).

1.3 NEWCASTLE DISEASE VIRUS (NDV)

1.3.1 Significance

The avian paramyxovirus NDV is a model paramyxovirus and the causative agent of Newcastle disease (ND). The highly pathogenic nature of ND has resulted in significant economic losses to poultry industries worldwide and the disease remains endemic in some countries (74-76). Several vaccines are available to control outbreaks, although the inability to treat wild avian populations means the virus can persist and new strains of NDV are continually emerging worldwide (61, 77).

Although NDV is not a significant human pathogen, the virus is easily propagated in embryonic eggs and previous work specific to NDV has been translated into other enveloped viruses (78). Furthermore, therapeutic agents developed for avian viruses can then be translated to closely related human pathogens; evident in the development of influenza neuraminidase inhibitors (79). A better understanding on the molecular virology of this virus may also be beneficial for treatment of human disease. For instance, NDV has been used as an oncolytic agent and vaccine vector for animal and human use (43, 80-83) and for the production of NDV-HN virus-like particles and peptide vaccines (84-86).

1.3.2 Attachment (HN) and fusion (F) proteins of NDV

The F and HN glycoproteins are presented on the envelope of NDV (**Figure 1-7** and **Figure 1-8**, respectively). The levels of virulence of strains seen in NDV, lentogenic (low virulence), mesogenic (moderate virulence) and velogenic (highly virulent), are associated with the susceptibility of F_0 to cleavage into the F_2 and F_1 disulfide linked chains (78). Increased cleavage correlates with more virulent forms of the virus and the level of cleavage is dictated by the amino acid sequences of the cleavage-activation site (87, 88) and cellular distribution of host cell proteases that recognise or accommodate particular sequences (77, 89). Further differentiation of virulence is based on whether or not HN is expressed with a C-terminal extension (88, 89). In avirulent strains of NDV, such as the V4-VAR strain studied in Chapter 4, removal of this extension is required to produce biologically active HN. Addition of a C-terminal extension to other strains of NDV that do not encode an extension, including highly virulent strains, can reduce pathogenicity (90-92).
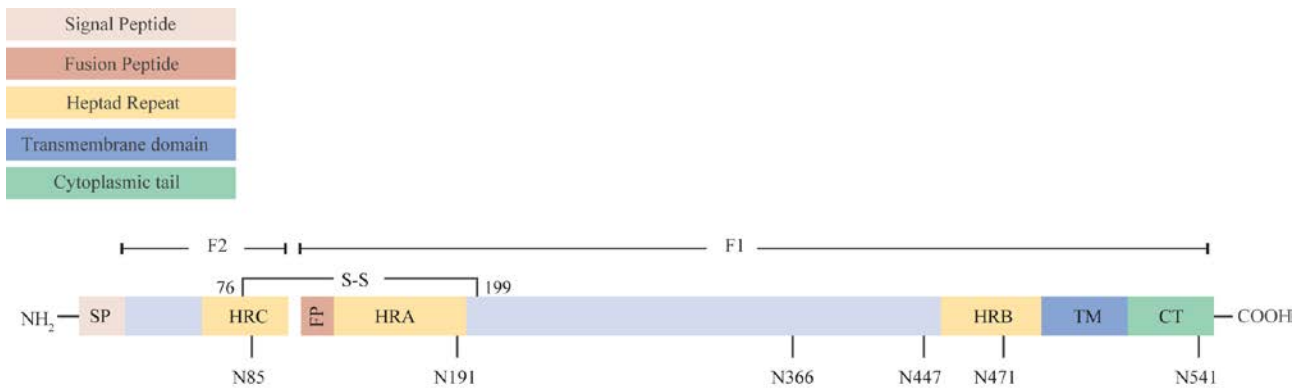
Figure 1-7. Schematic of NDV F (Strain QLD 66/UniProt ID P33615). The schematic (not to scale) identifies the signal peptide (SP), heptad repeats (HR A-C), fusion peptide (FP), transmembrane domain (TM) and cytoplasmic tail (CT). N-linked consensus sites (N-X-S/T) are marked with vertical lines and the amino acid number of the respective Asn residue. Cleavage of F_0 at the amino-terminal end of FP produces F_2 and F_1 chains which are linked by a disulfide bond (represented by a connected line with the amino acid number of the Cys residues).

To undertake its function of receptor binding, HN recognises and binds to gangliosides and sialylated N-linked glycans containing α2–3 and α2–6 sialyl linkages (93). The proximity of the virions to host cells then enables fusion of the viral and host cell membranes. Further to its attachment function, the stalk domain of NDV HN is also thought to modulate the fusion activity of F (70, 94, 95). X-ray crystallography of the stalk domains of NDV HN has revealed a four-helix bundle structure (4HB) with an 11-residue repeat forming the hydrophobic core (96, 97). Site-directed mutagenesis or introduced N-linked sites into the stalk domain of NDV HN both result in significantly reduced or blocked fusion (98-100).
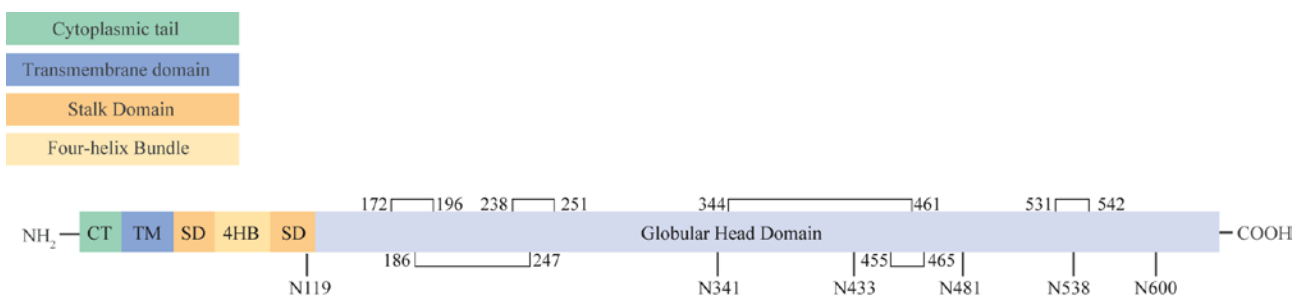
Figure 1-8. Schematic of NDV HN (Strain QLD 66/UniProt ID P13850). The schematic (not to scale) identifies the cytoplasmic tail (CT), transmembrane domain (TM), stalk domain (SD), the four-helix bundle (4HB) found in the SD and the globular head domain. N-linked consensus sites (N-X-S/T) are marked with vertical lines and the amino acid number of the respective Asn residue. Disulfide bonds are represented by connected lines and are derived from (101) with the amino acid number of the Cys residues.

1.3.3 Glycosylation of NDV HN and F proteins

A mutation study has shown that the N-linked glycosylation site N85 found on F_2 of NDV F, is highly important for cleavage and fusion activity (102). Sites N191, N366 and N471 also affect the functional activity of F whereby mutation of N191 was associated with a small decrease in fusion activity and combined mutations of N191 and N471 were found to increase the replication, virulence and immunogenicity of viruses (103). Released glycans from NDV F reveal predominately high mannose glycosylation of F (104), which has been confirmed at site N85 (105).

With respect to NDV HN, mutation studies predict that four sites are occupied, N119, N341, N433 and N481 (106, 107). Abolishing these sites can modulate viral replication, transport of HN, protein folding and subsequent reactivity with antibodies, and attachment and fusion promotion activity (106, 107). Crystal structures of HN have also revealed that sites N341, N433 and N481 are occupied, with up two GlcNAc residues identified at each site (91, 108). These residues likely form part of the trimannosylchitobiose core while further monosaccharides were likely not observed due to the flexibility of glycan chains. It has been shown that N538, which is flanked by two Cys residues, is not glycosylated, presumably due to disulfide bond formation (101). Mutation of cysteine residues either side of N538 resulted in the addition of glycans as determined by gel electrophoresis (109). Site N600 illustrated in **Figure 1-8** is found in a limited number of NDV strains and is situated in the C-terminal extension that undergoes cleavage to activate HN (88, 89). Site N481, and to a lesser extent N341, are required for maturation and proper folding of HN (107), while site N119 is thought to mask epitopes recognised by neutralising antibodies (106, 107).

1.4 HUMAN RESPIRATORY SYNCYTIAL VIRUS (HRSV)

1.4.1 Significance

Human respiratory syncytial virus is considered an important cause of acute lower respiratory infection in infants, children (110-112) and the elderly and immunocompromised individuals (113). Clinical manifestations of severe infection, such as bronchiolitis and pneumonia, can cause considerable morbidity and occasional mortality. It is estimated that in 2005 hRSV was responsible for between 66,000-199,000 deaths worldwide in children younger than five years (111). There is

currently no effective therapeutic or approved vaccine for hRSV and a strong desire exists within the scientific and medical communities to develop safe and effective treatments (58, 59, 114).

Several aspects of hRSV infection have hampered the development and application of vaccines and antivirals, specifically: the highly pathogenic nature of the virus; the ability of the virus to infect at a young age, thus limiting the time-frame to administer vaccines; the ability of the virus to re-infect symptomatically throughout life without significant antigenic diversity; and the potential for infection to cause enhanced respiratory disease (115, 116). With respect to the latter, considerable caution has been applied when developing vaccines and treatments for hRSV since the failure of a formalin-inactivated whole virion vaccine in the 1960s (117). After administration of the inactivated vaccine to infants, subsequent natural exposure to hRSV resulted in significant enhanced lung pathology. Of the recipients who were seronegative for hRSV before vaccination, 80% required hospitalisation after natural exposure. This enhanced pathology was suggestive of a Th2 biased immune response, which is characterised by the production of Th2-type cytokines that promote B cell class switching to IgE and induction of eosinophilia and mast cell proliferation and degranulation. This Th2 bias has since been documented in clinical and animal models of enhanced hRSV disease where there is a predominance of Th2-type cytokines and an excess of lung eosinophils and neutrophils (116).

1.4.2 Attachment (G) and fusion (F) proteins of hRSV

Virions of hRSV present the G and F glycoproteins on their envelopes (**Figure 1-9** and **Figure 1-10**, respectively), which are the main source of epitopes for hRSV antibody responses (118). Human RSV isolates are classified into subtypes A and B based on the antigenicity of the G and F proteins. The G protein does not follow the same molecular architecture as HA, HN and G from *Paramyxovirinae* and also lacks haemagglutinin and sialidase activity (63). The extracellular domain of G is not well conserved within strains of hRSV (119, 120), with the exception of a central conserved domain (CCD) containing a cystine noose (121). The G protein of hRSV has been implicated in immune evasion (122) and attachment to cell surface glycosaminoglycans (GAG) (123, 124) and a chemokine receptor (125). The process of attachment may be reliant on the C-terminus of the G which can be removed proteolytically in certain cell lines (73). It has been shown that G is not essential for hRSV infectivity or replication in cultured cells (72). This dispensability

of G may be due to different viral attachment processes between *Paramyxovirinae* and *Pneumovirinae,* whereby F from hRSV can bind to receptors and facilitate attachment (126).
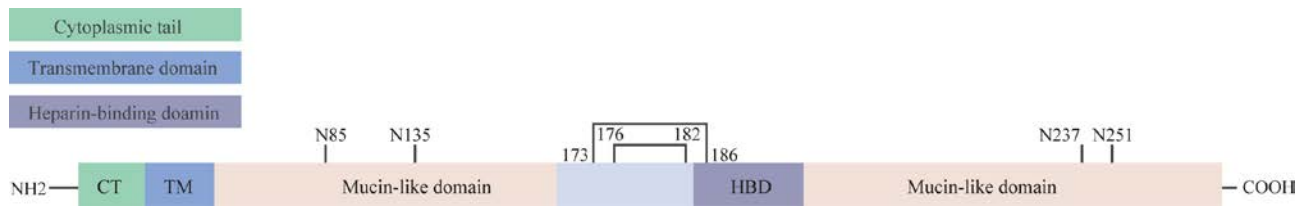
Figure 1-9. Schematic of hRSV G (Strain A2/UniProt ID P03423). The schematic (not to scale) identifies the cytoplasmic tail (CT), transmembrane (TM), mucin-like domains and heparin-binding domain (HBD). N-linked consensus sites (N-X-S/T) are marked with vertical lines and the amino acid number of the respective Asn residue. A cystine noose is found between the two mucin-like domains and is represented by connected lines according to (121) with the amino acid number of the Cys residues.
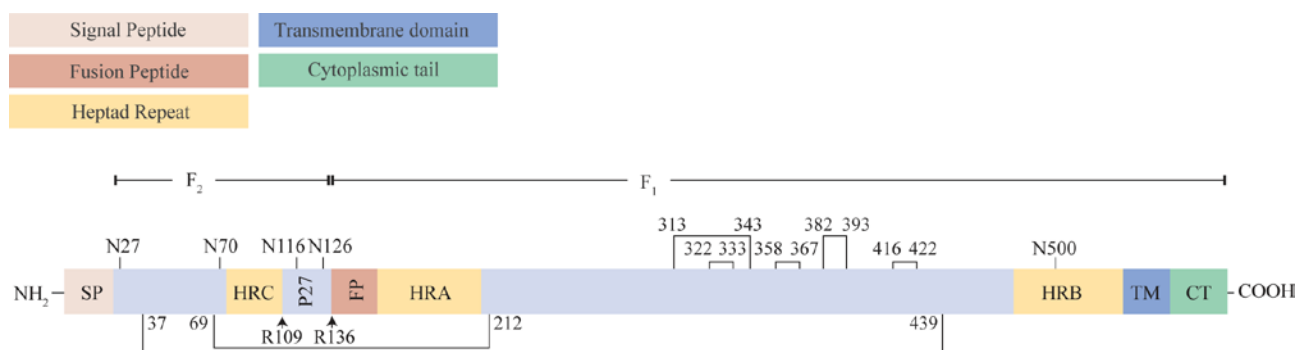
Figure 1-10. Schematic of hRSV F (Strain A2/UniProt ID P03420). The schematic (not to scale) identifies the signal peptide (SP), fusion peptide (FP), heptad repeats (HR A-C) (131), transmembrane domain (TM) and the cytoplasmic tail (CT). N-linked consensus sites (N-X-S/T) are marked with vertical lines and the amino acid number of the respective Asn residue. Disulfide bonds are represented by connected lines according to (132) with the amino acid number of the Cys residues. The furin-like cleavge sites R109 and R136 are denoted by arrows. Cleavage of F_0 at the amino-terminal end of FP produces the F_2 and F_1 chains. Additional cleavage at R109 produces soluble pep27 (P27).

Another potential explanation for the dispensability of hRSV G is the difference in structural rearrangement of the F protein in *Pneumovirinae* compared to *Paramyxovirinae.* Cleavage of F_0 in hRSV by furin-like proteases occurs at two cleavage sites (**Figure 1-10**); the first site is similar in all paramyxoviruses and is found at the amino-terminal end of the fusion peptide with cleavage

occurring C-terminal to Arg at amino acid position 136. An additional second cleavage site is found at the carboxyl-terminal end of the F_2 chain (R109) and cleavage of both sites is a requirement for syncytium formation (127, 128). Cleavage at R136 and R109 also results in the loss of a soluble 27 amino acid peptide, designated pep27 or P27 (128). An interesting study has shown that recombinant SeV, expressing the two hRSV F cleavage sites on the SeV F protein, resulted in a reduced dependency on HN (129). More recently, one group has suggested that hRSV F is cleaved at the amino-terminal site (R109) during transport and expression in host cells, while cleavage at the remaining site (R136) occurs after endocytosis which induces the F protein to trigger (130).

1.4.3 Glycosylation of hRSV G and F proteins

The G protein is highly glycosylated, with over half of the mass of the 90 kDa mature form attributed to glycosylation, a high proportion of which is O-linked (133-135). The high level of O-linked glycosylation is thought to be associated with the mucin-like domains found in the ectodomain of the protein (115), while N-linked glycosylation (approximately 13 kDa of the mature form) may occur at one or more of the four N-linked consensus sites (133, 136). Production of G in different cell lines results in distinct migration patterns that can be attributed to changes in the glycosylation profile (50, 135). Differential O-linked glycosylation has been noted in G after infection of lower respiratory tract primary cells and alveolar epithelial cells, which resulted in changes in reactivity of the protein with antibodies (50). The effect of differential glycosylation on other functions of G can be seen from a study that produced recombinant hRSV G devoid of O-linked glycans. Subsequent vaccination reduced lung pathology and Th2 cytokine production in a mouse hRSV challenge model (137). Furthermore, removal of a section of the second mucin-like domain of hRSV G expressed on a recombinant vaccinia virus reduced the Th2 response in mice (138).

With respect to hRSV F, three N-linked sites are thought to be occupied, N27, N70 and N500, with the latter site being required for syncytium formation (139). Mutation of site N70 to Gln was seen to increase fusion activity while dual mutations N27Q and N70Q decreased fusion activity. A significant decrease in fusion activity (90%) was observed after the mutation N500Q. Furthermore, the importance of maturation and proper formation of N-linked glycans in infection and subsequent syncytia has been reported for hRSV F (135). Prevention of glycan maturation beyond high mannose structures resulted in a one hundred-fold reduction in viral infectivity (135). The overall importance of glycosylation for hRSV infection has also been shown through treatment of hRSV

virions with an exo-sialidase, that removes non-reducing terminal α(2–3,6,8,9) linked NeuAc from glycans and resulted in an increase in infectivity (140). Furthermore, treatment of virions with specific endoglycosidases, which cleave within glycan structures and remove N-linked or core 1 and Tn O-linked glycans, significantly decreased viral infectivity (140).

Taken together, the above studies reveal that glycosylation of the attachment and F proteins of paramyxoviruses can play important roles in host-pathogen interactions. They highlight the importance of characterising site-specific glycan heterogeneity, not only to further elucidate potential functional roles of the glycans, but also to ensure accurate glycosylation profiles when producing these proteins for therapeutic use. The work presented herein utilised mass spectrometry (MS) to characterise the monosaccharide compositions of glycans in a site-specific manner on glycoproteins from paramyxoviruses. An overview of MS as a tool to analyse protein glycosylation will be discussed in the coming sections.

1.5   MASS SPECTROMETRY AS A TOOL TO ANALYSE GLYCOSYLATION

1.5.1   Methods to analyse protein glycosylation

Glycan heterogeneity and the potential substoichiometric presence of glycoforms make characterising glycans a bioanalytical challenge. A range of tools and techniques can be used individually or in combination to characterise protein glycosylation. Some more commonly used techniques include lectin or antibody binding to selectively identify glycan structures or glycosidic linkages, and staining and visualisation of electrophoretically separated proteins to identify glycoproteins or evaluate changes in protein glycosylation (141). Protein glycosylation can be altered by enzymatic trimming or removal of glycans, disruption of cellular glycosylation pathways or site-directed mutagenesis of proteins to abolish glycosylation sites (141). These techniques typically provide a global view of the glycosylation profile of proteins but do not enable comprehensive structural information to be elucidated. To obtain such information, glycans can be chemically or enzymatically released from glycoproteins and analysed by high performance liquid chromatography (HPLC) to deduce structural or compositional information through comparison with known glycan standards. However, many proteins have more than one putative site of glycosylation and site-specificity is lost with glycan release.

Structural characterisation of glycoproteins or glycans may also be achieved using X-ray crystallography or nuclear magnetic resonance spectroscopy. However, the inherent flexibility of glycan chains and potential glycan heterogeneity can lead to disordered electron density maps or incomplete resonance assignments (142). For disordered glycoproteins, small-angle X-ray scattering may be used to provide molecular weight (MW) determination or low resolution characterisation of protein conformation (143). Another tool that is widely used for the structural characterisation of protein glycosylation is MS [reviewed in (19, 144-154)]. Depending on the type of analysis implemented, a range of qualitative, quantitative, structural and site-specific information can be obtained. Typical analyses involve releasing glycans from the glycoproteins, analysing intact glycoproteins or digesting the glycoprotein with proteases before analyses. To adequately describe the use of MS for the analysis of protein glycosylation some of the basic concepts of MS will first be introduced.

1.5.2   Mass spectrometry

As an analytical tool, MS provides an elegant means of characterising the compositional and structural elements of biological macromolecules. The applications of MS are far-reaching but this section will introduce MS concepts that pertain mainly to the analysis of proteins and glycoproteins and the techniques implemented in this work.

The fundamental principle behind MS is the production of charged particles or ions, which are separated and measured according to their mass-to-charge ratio (*m/z*) and abundance. Therefore, the basic components of a mass spectrometer are the ionisation source which ionises the particles; the mass analyser, which separates ions according to *m/z*; and the detector, which measures and detects the ions (155). One technique used to produce gas-phase macromolecular ions is electrospray ionisation (156, 157). During this technique, a high voltage is applied to a sample in a volatile solvent travelling through a narrow capillary creating an aerosol of charged particles. The charged particles are produced within an electric field and are transferred from solution into the gas phase as the droplets undergo rounds of solvent evaporation and Coulomb fission, which is typically aided by high temperatures and a flow of gas in the source (156). Positive or negative ions are formed depending on the electric potential applied at the source (156, 157). In positive ion mode, a molecular ion is typically formed from the addition of hydrogen (H) cations or protons (that is H

which has lost an electron) in a process called protonation. This results in multiply charged species $[M + nH]^{n+}$ which can be separated and detected by the mass spectrometer.

There are several types of mass analysers all with varying degrees of speed, sensitivity, *m/z* range, dynamic range, resolution and mass accuracy (157). The mass resolving power and mass accuracy of an instrument are important parameters as they enable closely related *m/z* values to be separated, including the isotopic profile of a molecular ion, and thus charge states and accurate masses to be determined (158). A mass analyser that is capable of high resolution and mass accuracy is the Orbitrap (159), available on the Orbitrap Fusion™ Tribrid™ mass spectrometer used in this work (160). This instrument also contains quadrupole and linear ion trap (LTQ) mass separation technologies in a configuration that allows many different tandem MS (MS/MS) acquisition strategies to be implemented. These are discussed in the coming sections.

Tandem MS can be implemented in the common "bottom-up" approach in proteomics, where proteins are digested and the resultant peptides are typically separated by reverse-phase-HPLC before analyses by MS. During MS precursor scans are conducted to measure the *m/z* values of these peptides, described as precursor ions, before they are selected for fragmentation. Tandem MS involves increasing the internal energy of multiply charged precursor cations to induce the dissociation of covalent bonds, producing product ions. The *m/z* values of these product ions are also measured. Tandem MS methods include, but are not limited to, collision-induced dissociation (CID) (161), higher-energy collision dissociation (also known as higher-energy C-trap dissociation and abbreviated HCD) (162) and electron-transfer dissociation (ETD) (163). During fragmentation with CID and HCD, dissociation of the peptide bond (C—N) is induced by vibrational activation with neutral gas molecules. This predominantly produces b- and y- series ions (164-166) according to nomenclature described in (167-169) (**Figure 1-11**). The collision energies applied during CID and HCD on the Fusion™ Tribrid™ mass spectrometer are both considered low-energy with collision energies in the low electron-volt (eV) range rather than keV, however, there may be subtle differences in the fragment ions that are produced (165). The CID process is performed in the ion trap by resonant excitation of a selected precursor ion; primary fragment ions that do not have the same *m/z* value as the precursor ion are therefore not activated further. During the HCD process, which is akin to beam-type CID, ions are emitted into a collision cell where primary fragment ions retain kinetic energy and can undergo further fragmentation through multiple collisions (170-172). An advantage of the latter approach is that fragment ions are not subject to cut-off at the low *m/z* range, as observed in CID MS/MS spectra from trapping analysers (170, 172).

In contrast to collision dissociation, ETD is a non-ergodic fragmentation process that occurs after transfer of an electron from a radical anion to a multiply charged peptide cation (163). In the case of the Fusion™ Tribrid™ mass spectrometer, radical anions of fluoranthene (m/z 202) are utilised as the ETD reagent (173). Several mechanisms have been proposed for the fragmentation patterns observed during ETD (163, 174) [reviewed in (175, 176)]. Upon electron capture by a charged precursor cation the original authors, describing the closely related electron-capture dissociation, predicted that a hydrogen atom is transferred to an amide oxygen from a nearby positively charged site (177). An intermediate radical site is formed at the carbonyl carbon which induces cleavage of the N—Cα bond resulting in c- and z-ions (**Figure 1-11**). The c- and z-ions produced in ETD are even-electron (prime or "′") and odd-electron radical species (dot or "·"), respectively. However, abstraction of H from the z· radical fragment can also result in radical c-ions and even-electron z-ions (176). During the ETD process, a multiply protonated precursor ion can accept electrons from radical anions producing charge-reduced precursors that fail to dissociate (ETnoD) (174, 178, 179). The ETD process also promotes neutral losses such as ammonia and water from the charge reduced precursors. Ions that are derived from unreacted precursors, ETnoD and neutral losses are often the most intense ions in MS/MS spectra (180).
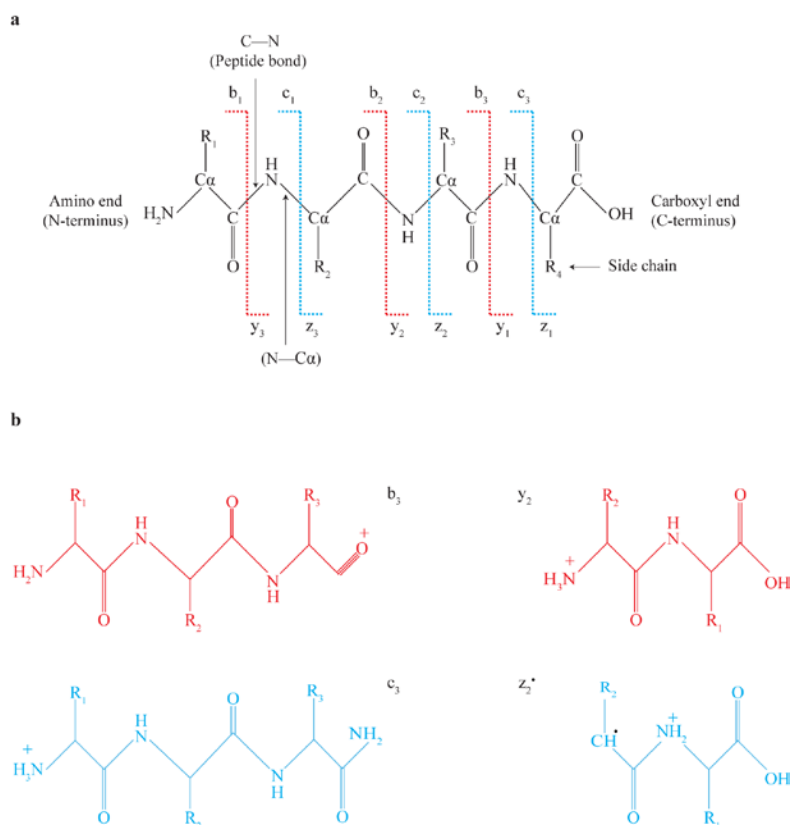
Figure 1-11. Fragmentation of the peptide backbone using collision dissociation and ETD. (a) Representation of peptide fragmentation identifiying fragment ions that are predominantly observed with CID and HCD (b- and y-ions), and ETD (c- and z-ions). (b) Representation of fragment ions, b_3 and y_2 and c_3 and z_2. Image adapted from (181).

1.5.3   Mass spectrometry for protein glycosylation

As discussed, the inherent complexity of glycosylation creates significant challenges when characterising protein glycosylation. Using MS this challenge can be met in several different ways; the glycans can be released and studied, the intact protein can be analysed or the glycoprotein can be digested with proteases before analysis. Each strategy has benefits and disadvantages and the level at which glycosylation is analysed should be determined by the biological or technical question to be answered.

Releasing N-glycans is the most well-documented strategy [reviewed in (182-186)], and typically relies on peptide-N4-(N-acetyl-ß-glucosaminyl) asparagine amidase (PNGase F). This enzyme cleaves N-linked glycans at the innermost GlcNAc residue and converts the previously glycosylated Asn to Asp (187). O-linked glycans are usually released chemically, although these methods are not as widely reported (185, 188). The released glycans can be labelled and analysed using a variety of different MS methods in positive and negative ion mode (189, 190). A major advantage to releasing the glycans is the simple methodology and level of structural information that is acquired (190). With certain approaches the structures can also be quantified with reasonable consistency (185, 191, 192). However, if a protein contains multiple sites of glycosylation, site-specific heterogeneity or occupancy cannot be established by analysing released glycans.

The field of intact (top-down) glycoproteomics is continually expanding with technical advances in mass spectrometers. A particular advantage of intact MS is that it can quantitatively elucidate microheterogeneity and macroheterogeneity of the glycoprotein species (193-196). One drawback of this technique is that it typically requires a high level of protein homogeneity, and a small number of putative sites of glycosylation.

For the purposes of this work, MS was used to characterise site-specific glycan heterogeneity of glycoproteins subsequent to their digestion with proteolytic enzymes (197-201). The resultant glycopeptides (peptide + attached glycan) preserve information about site occupancy, the monosaccharide composition of the glycan and the site of glycosylation, which can be elucidated after MS/MS using CID, HCD and ETD (22, 202-213). At this point it would be appropriate to reiterate that particular vertebrate monosaccharides are isomeric (**Figure 1-1**) (E.g.

GalNAc/GlcNAc or Man/Gal/Glc), thus mass spectral analysis of intact glycopeptides cannot distinguish between monosaccharide isomers in the attached glycans. Nomenclature has been implemented in this work to denote this ambiguity, where deoxyhexosamine (dHex) represents Fuc and N-acetylhexosamine (HexNAc) and hexosamine (Hex) are used to represent GlcNAc/GalNAc and Man/Gal/Glc, respectively.

When analysing glycopeptides HCD and CID predominantly break the monosaccharide linkages of the attached glycan producing diagnostic glycan oxonium ions and glycopeptide fragment ions (peptide with varying degrees of the glycan structure) (**Figure 1-12**) (203, 204, 214, 215). Fragmentation can also result in the production of peptide b- and y-ions but factors such as peptide sequence, precursor *m/z* and the level of dissociation energy can affect sequence coverage (214, 216). When employing collisional energies, a stepped method can be applied where a precursor is fragmented with multiple collision energies (typically low, medium and high) and the fragment ions are measured in one MS/MS spectrum. By including the lower and higher energies it may produce more glycosidic fragments and peptide fragments, respectively. An important facet of HCD fragmentation of N-linked glycopeptides is production of a fragment ion consisting of the peptide with the core GlcNAc attached (203). This is referred to as the Y1 ion according to nomenclature described in (217). This Y1 ion and other fragment ions such as Y2 (peptide+HexNAc$_2$) and Y1+dHex$_1$ can be used to infer the mass of the peptide portion of the fragmented glycopeptide. These fragment ions, in combination with peptide sequence ions, can therefore facilitate the identification of the peptide moiety. When analysing O-linked glycopeptides HCD often results in complete loss of the glycan moiety and an intense ion corresponding to the mass of the peptide (Y0) (209). Again, the Y0 ion can be used to infer the mass of the peptide moiety and support the identification of the peptide through sequence ions. When analysed using high resolution and mass accuracy, the oxonium ions produced from the glycan portion of the glycopeptide can be used to identify the components of the glycan (203, 218, 219). In some cases, the relative abundance ratio of oxonium derived from HexNAc can indicate the presence of certain linkages of monosaccharides (GlcNAc or GalNAc) in a glycan (220-222).

Alternatively, ETD predominately induces fragmentation of the peptide backbone and thus preserves labile post-translational modifications such as glycosylation (**Figure 1-12**) (175, 202, 205). This is an important factor when two or more glycosylation sites are present on a glycopeptide. In some instances glycosidic bond cleavages can occur in ETD fragmentation of N- (207) and O-linked (223) glycopeptides. More recently, the combination of collision activation

(224-226) with ETD, (termed EThcD or ETciD when combined with HCD or CID, respectively) has been used to enhance fragmentation of N-linked (227) and O-linked glycopeptides (221). In addition to increasing the relative amounts of c- and z-ions these combined fragmentation methods also produce ions that are characteristic of CID/HCD fragmentation of glycopeptides (19). This proves beneficial as it provides informative spectra where both peptide and monosaccharide constituents can be elucidated along with the site of glycan attachment.
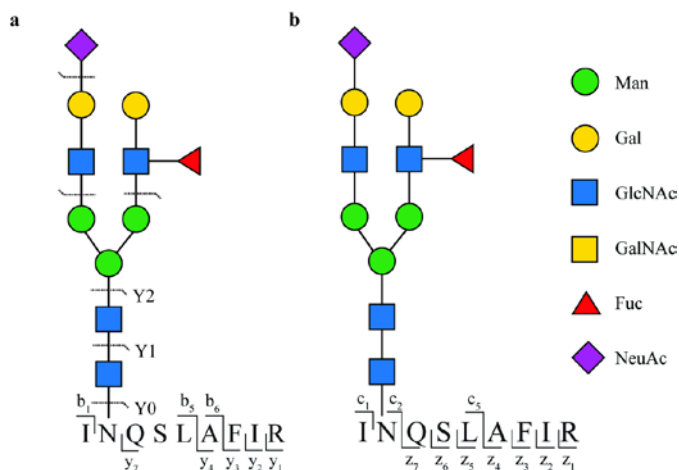
Figure 1-12. Representation of tandem MS of glycopeptides. Theoretical examples of bond cleavages using HCD or CID (a) and ETD (b).

Within this work the emerging acquisition strategies, HCD-product-dependent (HCD-pd)-ETD (with or without supplemental activation) or HCD-pd-CID, were also implemented. These strategies used high mass accuracy glycan oxonium ions detected in the first pass Orbitrap HCD scans to trigger 'on-the-fly' ETD or CID fragmentation of the same precursor ion detected in the Orbitrap or ion trap, respectively (227-231). These strategies allow targeted sampling of glycopeptides by complementary fragmentation techniques in one MS run without a drastic increase in duty cycle, which is the time taken to monitor an analyte. This is important, as a large number glycoforms may elute in a narrow retention time window, and a short duty cycle is required to analyse as many of the eluting glycoforms as possible.

Prior to MS there are many sample preparation methods that may be implemented. One method used to determine site occupancy at the glycopeptide level is achieved by digesting a sample with PNGase F before or after proteolytic digestion. As noted, removal of the glycan with PNGase F results in the conversion of Asn to Asp. The peptides are then subjected to MS/MS where the mass difference (+0.984 Da) can be sequenced to Asn in an N-linked consensus site, thus confirming site occupancy (187). A general measure of site occupancy can be achieved by digesting a sample with PNGase F and comparing the intensities of non-glycosylated peptides containing an N-linked site

with the de-glycosylated versions of the same peptide (where Asn is deamidated) (232, 233). However, spontaneous deamidation can occur and measures need to be taken to minimise such false positive data.

Due to the potential substoichiometric presence of glycopeptides and lower signal strength compared to non-glycosylated peptides during MS (232), enrichment strategies are also often implemented prior to analyses (234). Enrichment is particularly relevant for complex samples and, depending on the approach, can facilitate the isolation of a broad range of glycopeptides (206, 213, 227, 230) or those carrying certain monosaccharides or substituents (235-239). Hydrophilic interaction liquid chromatography is commonly reported in the literature (234) with several different types of resins and methods available (240-242). By enriching for glycopeptides, or glycoproteins prior to proteolytic digestion, it reduces ion suppression by co-eluting non-glycosylated peptides. This increases the relative intensities of glycosylated peptides and the likelihood of their selection and subsequent detection.

1.5.4   Analysis of data derived from tandem MS of intact glycopeptides

Many excellent data analysis tools have been developed to analyse glycopeptides and peptides from complex or single protein samples using different enzyme digestions and MS fragmentation techniques [reviewed in (243-248)]. However, the analysis of spectra derived from fragmentation of intact glycopeptides remains a substantial challenge and still relies on manual validation of assignments made by software (248, 249).

The main aim when analysing glycopeptides is to correctly identify the peptide and glycan portions of the glycopeptides and the site of glycan attachment. Many factors come into play when trying to confirm the components of a glycopeptide, these include where the sample was derived (e.g. the organism, tissue or cells), the complexity of the sample and how the sample was prepared. This information can be used to narrow the search space by limiting the number of glycans or certain monosaccharides or substitutes, specifying protein sequences and including parameters specific to sample preparation. Precursor mass is typically not adequate for confident identification of glycopeptides (250) and MS/MS spectra may provide additional information that enables confident assignment of the peptide moiety and attached monosaccharides or substituents. Chromatographic

retention time can also be used to validate glycopeptide assignments as identical peptides with different glycans attached will elute in a narrow retention time range.

As aptly described by Lee and colleagues (249), glycopeptide assignments may be false (both the peptide and glycan portions are incorrect), or partially false (either the peptide or glycan portion is incorrect). There are several potential pitfalls that may be encountered when identifying glycopeptides, involving isobaric or near isobaric modifications or reactions that lead to misassignment of either the peptide or glycan moiety or both. Partial false positives that stem from incorrect assignment of the glycan portion include the allocation of Fuc_2 in place of $NeuAc_1$ ($\Delta m=1.0204$), which can occur if the second peak (^{13}C) is selected in the isotopic distribution of a precursor ion rather than the monoisotopic peak ($\Delta m=1.0033$). The isobaric monosaccharide pairs Hex_1NeuAc_1 and Fuc_1NeuGc_1 (where NeuGc = N-glycolyl-neuraminic acid) ($\Delta m=0$) may also prove problematic. To circumvent such problems HCD MS/MS spectra can be interrogated for NeuAc fragment ions, the presence of which indicate Fuc_2 or Fuc_1NeuGc_1 are not components of the attached glycan. The former can be further confirmed by manual inspection of the isotopic distribution to ensure correct selection of the monoisotopic peak. The latter can be further confirmed by inspecting HCD MS/MS spectra for the presence of oxonium ion for NeuGc and Fuc (251).

Another isobaric difference that may lead to a false assignment is oxidation of Met (O=15.995 Da), which equates to the difference between Hex and dHex (C6O5H10-C6O4H10=15.995 Da). Care needs to be taken to ensure the peptide modification is correctly identified to prevent the mass difference being transferred to the glycan. Side reactions on peptides or glycans need to be considered (252, 253) as do non-covalent dimer formations involving glycopeptides (254). Other factors that may also need to be taken into consideration when analysing glycopeptides, include the position of putative glycosylation sites in the amino acid sequence and residues surrounding potential proteolytic cleavage sites. If the putative glycosylation site is adjacent to a proteolytic cleavage site the presence of a glycan at that site may hinder cleavage (255). Furthermore, negatively charged residues neighbouring a tryptic cleavage site can reduce the cleavage efficiency of trypsin, particularly at lysine bonds (256). This may result in missed cleavages and thus larger peptides (and potentially glycopeptides) than expected using *in silico* digestions.

1.6   SIGNIFICANCE AND SCOPE OF THIS WORK

A large body of research surrounds the attachment and F proteins of NDV and hRSV including functional, structural and immunological studies. It has been shown that glycosylation of NDV HN and F and hRSV F and G proteins can affect protein structure and folding and modulate the ability of the viruses to infect host cells and stimulate host immune systems. Moreover, mutation studies have identified important glycosylation sites on these proteins, and in some cases, have highlighted distinct roles for these sites. However, site-specific glycan heterogeneity has not been defined and this remains one of the unexplored areas related to NDV and hRSV surface glycoproteins. The work presented herein aims to bridge this gap by providing the first site-specific characterisation of glycosylation of these proteins. The findings may help elucidate mechanisms of viral attachment, replication and immune evasion within paramyxoviruses. Furthermore, it will enable comparisons to be made between native viral proteins and those produced in different expression systems for the purposes of vaccine and therapeutic design objectives.

1.7   RESEARCH AIMS

Given the paucity of data defining the specific glycosylation patterns of NDV HN and F and hRSV F and G the work presented herein was predominately exploratory in nature. During this study it became evident that the glycosylation patterns of each protein were quite varied and required specialised sample preparation, MS techniques and data analysis pipelines. Therefore, the overall research aims in this project were defined as follows:

1.   To assess occupancy and the monosaccharide compositions of glycans at each putative N-linked glycosylation site of NDV HN and F and hRSV F and G.

2.   To define the monosaccharide compositions of O-linked glycans from hRSV G in a site-specific manner.

3.   To identify unanticipated O-linked glycans on the attachment and F proteins of NDV and F protein of hRSV.

Chapter 2: Common materials and methods

This section contains general theories and methods that were applied throughout this work. Any deviations to these methods have been noted in the relevant Chapters.

2.1   GENERAL REAGENTS

Ammonium bicarbonate (NH$_4$HCO$_3$), dithiothreitol (DTT), iodoacetamide (IAA), coomassie brilliant blue (R250), bromophenol blue and triethylammonium bicarbonate buffer (TEAB) were all purchased from Sigma-Aldrich (St. Louis, MO, USA). Tris(hydroxymethyl)aminomethane (Tris) UltraPure™ was purchased from Invitrogen (Carlsbad, California, USA). Water was purified by a Synergy water purification system (EMD Millipore, MA, USA) and an EZ-2 Plus vacuum centrifuge was used (Genevac, Ipswich, England). Lyophilised sequencing grade bovine trypsin and endoproteinase Glu-C from *Staphylococcus aureus* and solubilised PNGase F derived from *Flavobacterium meningosepticum* were purchased from Roche Diagnostics GmbH (Mannheim, Germany). Solubilised sequencing grade Sialidase A™ (cleaves α(2–3,6,8,9) linked NeuAc) from *Arthrobacter ureafaciens* and lyophilised Sialidase S™ (cleaves α2–3 linked NeuAc) from *Streptococcus pneumoniae* were purchased from ProZyme, Inc. (San Leandro, CA, USA). Trypsin was reconstituted in 1 mM HCl and Glu-C and Sialidase S™ in purified water. The proteolytic enzymes were aliquoted into 5 µL quantities and trypsin stored at -80 °C and Glu-C at -20 °C. PNGase F was stored at -20 °C and the sialidases at 4 °C.

2.2   REDUCTION AND ALKYLATION

Protein samples were reduced with a freshly prepared solution of 10 mM DTT in 100 mM of NH$_4$HCO$_3$ buffer with 1-2% (*w/v*) SDS in microcentrifuge tubes. After the addition of DTT, the microcentrifuge tubes were flushed with N$_2$ and reduction was allowed to proceed at 4°C for 18 h. This was followed by further reduction for 2 h at room temperature. This strategy was based on previously published protocols which demonstrated complete unfolding of the proteins and largely penetrant reduction and alkylation of the hydrophobic regions of HN and other proteins (88, 257, 258). Cys residues were alkylated using a freshly prepared solution of 25 mM IAA in H$_2$O. After

the addition of IAA the microcentrifuge tubes were again flushed with N_2 and alkylation was allowed to proceed in the dark for 2 h at room temperature.

2.3 SODIUM DODECYL SULFATE-POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE)

One dimensional SDS-PAGE was carried out using the discontinuous buffer system based on Laemmli (259). Unless stated otherwise, gels were handcast using stacking [4% acrylamide, 0.1% SDS (*w/v*), 0.125 mM Tris-HCl, pH 6.8] and resolving [10% acrylamide, 0.1% SDS (*w/v*), 0.375 mM Tris-HCl, pH 8.8] gels. Polymerisation proceeded at room temperature for a minimum of 2 h before continuing over night at 4°C. Prior to SDS-PAGE, protein samples were mixed with equal volumes of reducing sample buffer [62.5 mM Tris-HCl (pH 6.8), 2% (*w/v*) SDS, 20% (*w/v*) glycerol, 10 mM DTT and 0.01% (*w/v*) bromophenol blue]. SDS-PAGE was performed using a Mini PROTEAN® system (Bio-Rad Laboratories, Hercules, CA, USA) with a Tris-Glycine running buffer. After electrophoresis the gels were briefly washed in water, stained with 0.03 % (*w/v*) CBB R-250 in 50/8.75/41.25 methanol/acetic acid/water (*v/v/v*) for 20 min, destained in the same ratio methanol/acetic acid/water (60 min), destained again in 8.75/91.25 acetic acid/water (60 min) then finally, equilibrated in water (60 min). If required, bands of interest were excised with a clean scalpel blade and peptides or proteins were extracted from the gel by in-gel digestion or electroelution, respectively.

2.4 IN-GEL DIGESTION

Bands excised from SDS-PAGE separated proteins were sliced into small fragments (~2 mm cubes) and destained twice with 200 mM NH_4HCO_3 in 50 % (*v/v*) aqueous CH_3CN at 37°C for 45 min then dried using a vacuum centrifuge. The gel fragments were rehydrated with 40 mM NH_4HCO_3 in 10% (*v/v*) aqueous CH_3CN containing 0.02 µg/µL trypsin for 1 h at room temperature. Additional 40 mM NH_4HCO_3 in 10% (*v/v*) aqueous CH_3CN was added and the samples were incubated at 37°C overnight. The supernatant containing the tryptic peptides was removed and the gel fragments were further extracted three times with 0.1% (*v/v*) aqueous trifluoroacetic acid at 37°C for 45 min, the supernatants from all extracts were pooled with the original digest supernatant. The samples were reduced to ~10 µL in a vacuum centrifuge for MS analysis.

2.5   ELECTROELUTION

Intact proteins were harvested from SDS-PAGE bands by electroelution, in this technique excised gel pieces are placed in a submerged chamber in an electroelution tank and proteins are drawn from the gel using an electric field and concentrated on a dialysis membrane. A buffer of 0.1M NH_4HCO_3 buffer and 0.05% (*w/v*) SDS was used and 3,500 MWCO Snakeskin™ pleated dialysis tubing (Pierce, Rockford, IL, USA). Electroelution was performed at a constant application of 2 W of power for 3 h before harvesting then a further 12.5 h before a second harvest.

2.6   METHANOL PRECIPITATION AND IN-SOLUTION TRYPSIN DIGESTION

In order to concentrate proteins and remove any impurities protein samples were subjected to methanol precipitation before in-solution digestion with trypsin. The precipitation method followed previously published protocols (257, 260). Protein samples were reduced to a volume of less than 100 µL in a vacuum centrifuge. The protein samples were co-precipitated with an aliquot of trypsin (protein to enzyme ratio 100:1) using ten volumes of methanol (-20°C) and incubated at -20°C overnight. After overnight precipitation, the samples were centrifuged for 15 min at 16,000 x g, 4°C and the supernatant aspirated. The pellets were washed twice more with -20°C methanol (90%) then dried and resuspended in 0.1M TEAB. If required, the samples were gently vortex mixed to solubilise the proteins. The samples were briefly centrifuged and incubated at 37°C for 2 h. Finally, another aliquot of trypsin was added (protein to enzyme ratio 33:1) and the digests were incubated at 37°C for a further 6 h. Co-precipitation enhances the efficiency of trypsin digestion as trypsin can work its way out of incompletely solubilised proteins.

2.7   NANO-ULTRA-HIGH PRESSURE LIQUID CHROMATOGRAPHY-TANDEM MS

Mass spectral analyses of samples acidified with aqueous trifluoroacetic acid were performed on an Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific Inc. Bremen, Germany) in positive ion mode. The mass spectrometer was coupled to a nanoACQUITY nano-ultra-high pressure liquid chromatography (nUHPLC) system (Waters Corporation, MA, USA), where

trapping was performed on a Waters C18 2G Symmetry (100 Å, 5 µm, 180 µm x 20 mm) trap column and gradient elution on a Waters C18 BEH (130 Å, 1.7 µm, 75 µm x 200 mm) column in-line with the trap column. Solvent A was 0.1% (*v/v*) aqueous formic acid and solvent B was 100% (*v/v*) CH_3CN containing 0.1% (*v/v*) formic acid. Eluates from the analytical column were continuously introduced into the mass spectrometer via a Nanospray Flex™ (NG) ion Source (Thermo Scientific) fitted with a PicoTip™ emitter (coating P200P, tip 10 ± 1 µm, New Objective, MA, USA). The MS methods implemented were specific to the NDV and hRSV glycoproteins and have been detailed in the relevant Chapters. All RAW files discussed in this thesis have been included in the supplementary files.

2.8   DATA ANALYSIS OF PEPTIDES

Proteome Discoverer (v. 1.4.1.14 or 2.1.0.81, Thermo Scientific) and the search engine Mascot (v. 2.5.1, Matrix Science Ltd., London, UK) were used to search the RAW data files generated from MS analysis of digested NDV and hRSV glycoproteins. The input parameters and databases were specific to each protein and have been detailed in the relevant Chapters.

2.9   DATA ANALYSIS OF GLYCOPEPTIDES

At the commencement of this study several software tools were investigated for their ability to identify glycopeptides and correctly assign the monosaccharide compositions of the glycan from MS/MS spectra using a variety of fragmentation methods (261-264). Byonic (v. 2.0-3) was selected due to its capability in handling multiple fragmentation types (CID, HCD, ETD and EThcD) (261, 265). For searches in Byonic, data from RAW files were converted to either MGF or mzML in Proteome Discoverer using a signal-to-noise (S/N) threshold of zero. An in-house spectral processing program, OxoExtract, was also developed for the analysis of HCD MS/MS spectra of glycopeptides (discussed in Chapter 3). For searches in OxoExtract, data from RAW files were converted to mzML in Proteome Discoverer using an S/N threshold of zero.

All glycopeptides assigned by Byonic or OxoExtract were manually validated. The peptide portion of the glycopeptide was verified using combinations of glycopeptide fragment ions (Y1 or Y0) with peptide b- and y- or c- and z-ions. Glycopeptides were expected to elute within a reasonable

retention time window, taking into consideration neutral and charged glycans and modifications to peptide moieties such as oxidised Met or deamidated Asn. Diagnostic glycan oxonium ions in a spectrum were used to confirm the presence of each monosaccharide or substituent assigned in the glycan. For example, HexNAc, Hex, dHex or NeuAc were confirmed through ions such as HexNAc$_1$ (*m/z* 204.0867), HexNAc$_1$Hex$_1$, (*m/z* 366.1395), HexNAc$_1$Hex$_1$dHex$_1$ (*m/z* 512.1974) or NeuAc$_1$ (*m/z* 292.1027), respectively. The use of diagnostic glycan oxonium ions in this work is discussed in more detail in Chapter 3. If the composition of the glycan could not be unambiguously assigned using oxonium ions (applicable mainly to sulfated and phosphorylated glycans) this has been noted in the relevant chapters.

2.10 ADDITIONAL DATA ANALYSIS TOOLS

For each glycoprotein analysed a degree of manual searching of spectra was required. Xcalibur Qual Browser (v. 3.0.63, Thermo Scientific) was used to search spectra and produce extracted ion chromatograms (EICs) for *m/z* values of interest. Xcalibur Qual Browser was also used to export spectra for manual annotation in Adobe Illustrator (Adobe). The MS-Digest and MS-Product modules of Protein Prospector (http://prospector.ucsf.edu) were used to calculate the theoretical masses of peptides and peptide fragment ions, respectively.

Chapter 3: Development of a spectral processing program for the analysis of glycopeptides using HCD MS/MS

3.1   SUMMARY

Despite the availability of numerous software tools for the identification of glycopeptides, confident assignment of site-specific glycosylation still relies on manual validation of results. Moreover, if a sample contains glycans that are not commonly observed, a specific glycan database may need to be compiled prior to software searches. Accordingly, identification of glycopeptides can prove time-consuming. To this end, a spectral processing program called OxoExtract was developed to automate site-specific assignment of glycans from a single protein. The purpose of the program was two-fold. First, it extracted all MS/MS spectra containing a "query" ion (i.e. a diagnostic glycan oxonium ion) and reported useful data such as precursor information and other oxonium ions or glycopeptide fragment ions observed in the spectrum. Second, it selected and submitted candidate N-linked glycopeptides for searches in GlycoMod to identify the attached N-linked glycan. These results were manually validated and used to create a custom glycan database for searches in Byonic.

3.2   INTRODUCTION

Preliminary investigations of NDV HN glycopeptides (discussed in Chapter 4) revealed many of the monosaccharide compositions at a given site were not present in glycan databases provided by Byonic or in repositories such as UniCarbKB (266). The observed monosaccharide compositions of the glycans were highly fucosylated and variably sulphated or phosphorylated. This prompted the development of a custom glycan database for NDV HN using the web based application GlycoMod (264). A major advantage of GlycoMod is that it searches for all possible monosaccharide compositions within a defined mass tolerance rather than relying on glycan databases. When analysing glycopeptides in GlycoMod the masses of putatively glycosylated peptides are entered, or are computed by GlycoMod if a protein sequence is provided. The user also enters observed precursor masses and GlycoMod predicts all possible monosaccharide compositions from a list of ten monosaccharides and substituents (for each residue there is an upper limit that is imposed for

inclusion in the composition). A large number of matches may result, particularly if all optional monosaccharide residues are considered as potential components of the glycan and if there are several potential sites of glycosylation in the protein sequence. The output from GlycoMod may be limited by refining the search parameters before each search. For example, data from the spectrum of a fragmented glycopeptide can be used to infer information such as the peptide mass and presence of certain monosaccharides or substituents on the attached glycan. By entering this information into GlycoMod it reduces the search space and thus time spent manually reviewing and removing false-positive results. OxoExtract is a program developed during the course of this work to automate the submission of these peptide and compositional constraints to GlycoMod for N-linked glycopeptide searches. OxoExtract selects all HCD MS/MS spectra with a specified "query" ion and searches them for oxonium and glycopeptide fragment ions. OxoExtract then extracts candidate glycopeptides and submits them to GlycoMod, customising the search parameters for each spectrum of a putative glycopeptide. Finally, OxoExtract creates a single output file that contains data for all spectra with a "query" ion and the results of the GlycoMod searches.

3.3   DEVELOPMENT OF OXOEXTRACT

The rules implemented in OxoExtract were based on glycopeptide fragmentation patterns using HCD where the fragment ions were detected in the Orbitrap. Fragmentation of glycopeptides using HCD typically results in the production of the Y1 ion and diagnostic glycan oxonium ions, in particular the HexNAc oxonium ion (203, 204). Similar fragmentation patterns were observed using the Fusion™ Tribrid™ mass spectrometer after manual interpretation of spectra from the analysis of NDV HN (discussed in Chapter 4). Therefore, the main principle behind OxoExtract is to isolate all spectra containing a "query" m/z value (such as 204.0867 for [HexNAc+H]$^+$). OxoExtract then searches those spectra for other oxonium ions and glycopeptide fragment ions including the Y1 ion. Observations of other oxonium ions can be used to confidently assign the presence of monosaccharides or substituents in assigned glycans (218-220, 238, 251, 267, 268). For example, sialylated glycopeptides produce abundant oxonium ions (NeuAc$_1$-H$_2$O and NeuAc$_1$) (218) while glycans with terminating Fuc residues may produce ions corresponding to HexNAc$_1$Hex$_1$dHex$_1$ (268). Such oxonium ions can also be used to prevent misassignment of glycans due to isobaric or near isobaric monosaccharide compositions and peptide modifications (discussed in Chapter 1). Furthermore, the identification of Y1+dHex can indicate core fucosylation (146, 216). The prescribed rules were translated into the OxoExtract Java application by Dr Christine Hoogland.

3.3.1 Searching with OxoExtract

The graphical user interface for OxoExtract is shown in **Figure 3-1** and identifies all input parameters.

Figure 3-1. Graphical user interface of OxoExtract

When using OxoExtract the first step is to upload an mzML file of HCD MS/MS spectra. The user then enters a "query" *m/z* value along with the mass tolerance to be applied (ppm). For all OxoExtract searches completed in this work the "query" ion was [HexNAc+H]$^+$ (204.0867) and a ±10 ppm mass tolerance was applied. As stated, OxoExtract searches all spectra with a "query" oxonium ion for predefined mono- and oligosaccharide oxonium ions. The list of the oxonium ions is stored in an Excel spreadsheet within the data folder of OxoExtract which can be modified by the user. The oxonium ions used in this work were observed in samples, theoretical or were defined in the literature (219, 220, 238, 251, 267-269) and have been represented in **Table 3-1**. For the purposes this work the monoisotopic masses of the oxonium ions were calculated with ChemCalc (http://www.chemcalc.org/) (270) using the chemical formula of the ions.

Spectra with the "query" oxonium are also searched for theoretical N-linked and O-linked glycopeptide fragment ions derived from an *in silico* digest of the protein of interest. Protein sequences searched by OxoExtract are stored in an Excel spreadsheet within the data folder of OxoExtract, which can be modified by the user. Input parameters for the *in silico* digest include the protein sequence, the proteolytic enzyme and the number of allowable missed cleavages (between zero and two). A fixed modification of Cys carbamidomethylation and a variable modification of oxidised Met can be selected for use in the calculation of theoretical fragment ions. The program also requires a maximum charge (between one and five) and *m/z* range (e.g. 0-2000) to be set for the calculation of theoretical fragment ions. For N-linked searches *m/z* values for Y1, Y2, Y1+dHex and peptide b- and y-ions are calculated for all peptides containing the specified N-linked consensus sequence (default is N-X-S/T). There is also an option to add the mass of HexNAc$_1$ to peptide y-series ions for proline. For O-linked searches *m/z* values are calculated for Y0 and Y1 ions and peptide b- and y-ions for all peptides containing Ser or Thr.

When specified, OxoExtract queries GlycoMod with a list of candidate N-linked glycopeptides to predict the compositions of the attached glycans. All precursors where fragmentation resulted in the production of the "query" oxonium ion and an N-linked Y1 ion are considered candidate glycopeptides. The previously entered mass tolerance (ppm) is applied as the mass tolerance for GlycoMod searches. For each candidate glycopeptide search (i.e. each spectrum), OxoExtract submits the precursor mass [M] and the peptide mass [M] (derived from the peptide portion of the theoretical Y1 ion in the spectrum). Pre-programmed rules implemented in OxoExtract define inclusion of certain monosaccharides for the GlycoMod searches. OxoExtract stipulates that for every search HexNAc and Hex must be components of the N-glycan (equivalent to selecting "Yes"

for a monosaccharide in a GlycoMod search). This was implemented as they form the common trimannosylchitobiose core. The residues NeuAc and dHex are considered "Possible" components of the N-linked glycan if the relevant dHex or NeuAc containing-oxonium ions are present in the spectrum of the candidate glycopeptide. The oxonium ions used for inclusion of dHex are listed in italics in **Table 3-1** under "dHex". The presence of *m/z* values for Y1+dHex also triggers the dHex to be considered "Possible". The oxonium ions used for inclusion of NeuAc are listed in italics in **Table 3-1** under "NeuAc". The user can modify the list of oxonium ions that trigger inclusion of dHex or NeuAc in the search. If the relevant dHex or NeuAc containing-oxonium ions are not present in the spectrum of the candidate glycopeptide, OxoExtract stipulates that they should not be considered in the monosaccharide compositions (equivalent to selecting "No" for a monosaccharide in a GlycoMod search). The inclusion of residues NeuGc, pentose, Sulf, Phos, ketodeoxynonulonic acid (KDN) and hexuronic acid (HexA) as potential components of the N-glycans is defined by the user (Options: Yes, Possible, No) (**Figure 3-1**).

Often there are multiple matching compositions in the results of GlycoMod searches. The criteria for ranking the final N-linked composition that is displayed in the OxoExtract output file are as follows: compositions with a common trimannosylchitobiose core are selected over those without a core; if more than one match contains the common core then the composition that is listed in UniCarbKB is selected and; if no matches or more than one match is listed in UniCarbKB then the composition with the lowest Δmass (ppm) value is selected. A link is also provided with the results of the individual GlycoMod search and the user can assess other matching compositions for that search if needed.

Table 3-1. Oxonium ions investigated in OxoExtract searches of spectra

Oxonium ion	$[M+H]^+$	Oxonium ion	$[M+H]^+$
HexNAc: N-acetylhexosamine			
$HexNAc_1$	204.0866	$HexNAc_2-H_2O$	389.1555
$HexNAc_1-C_2H_6O_3$	126.0550	$HexNAc_2$	407.1660
$HexNAc_1-CH_6O_3$	138.0550	$HexNAc_2+H_2O$	425.1766
$HexNAc_1-C_2H_4O_2$	144.0655	$HexNAc_3-H_2O$	592.2348
$HexNAc_1-2H_2O$	168.0655	$HexNAc_3$	610.2454
$HexNAc_1-H_2O$	186.0761	$HexNAc_3+H_2O$	628.2560
$HexNAc_1+H_2O$	222.0972		
Hex: Hexose			
Hex_1	163.0601	Hex_3	487.1657
Hex_1-H_2O	145.0495	Hex_4	649.5707
Hex_2	325.1129	Hex_5	811.2714
HexNAc/Hex			
$HexNAc_1Hex_1-H_2O$	348.1289	$HexNAc_1Hex_3$	690.2451
$HexNAc_1Hex_1$	366.1395	$HexNAc_2Hex_2$	731.2717
$HexNAc_1Hex_1+H_2O$	384.1500	$HexNAc_3Hex_1$	772.2982
$HexNAc_1Hex_2$	528.1923	$HexNAc_2Hex_3$	893.3245
$HexNAc_2Hex_1$	569.2188	$HexNAc_3Hex_2$	934.3510
dHex: Deoxyhexose			
$dHex_1$	*147.0652*	*$HexNAc_1Hex_1dHex_1$*	*512.1974*
$dHex_2$	*293.1231*	*$HexNAc_2dHex_1$*	*553.2239*
Hex_1dHex_1	*309.1180*	$HexNAc_1Hex_1dHex_2$	658.2553
$Hex_1dHex_1+H_2O$	*327.1286*	$HexNAc_1Hex_3dHex_1$	836.3030
$HexNAc_1dHex_1$	*350.1446*	$HexNAc_2Hex_2dHex_1$	877.3296
NeuAc: N-acetylneuraminic acid			
$NeuAc_1$	*292.1027*	$HexNAc_1Hex_1dHex_1NeuAc_1$	803.2928
$NeuAc_1-H_2O$	*274.0921*	$HexNAc_1Hex_2NeuAc_1$	819.2877
$NeuAc_1-2H_2O$	*256.0816*	$HexNAc_2Hex_1NeuAc_1$	860.3143
$HexNAc_1Hex_1NeuAc_1$	*657.2349*	$HexNAc_1Hex_1NeuAc_2$	948.3303
$NeuAc_1-2H_2O$	256.0816	$HexNAc_1Hex_3NeuAc_1$	981.3405
Hex_1NeuAc_1	454.1555	**NeuAc + O-acetylation (Ac)**	
$HexNAc_1NeuAc_1$	495.1821	$NeuAc_1Ac_1$	334.1133
$HexNAc_1dHex_1NeuAc_1$	641.2400	$NeuAc_1Ac_1-H_2O$	316.1027
$HexNAc_2NeuAc_1$	698.2614	$HexNAc_1NeuAc_1Ac_1$	537.1926
Sulf: Sulfate			
$Sulf_1$	80.9641	$HexNAc_2Sulf_1$	487.1228
Hex_1Sulf_1	243.0169	$HexNAc_1Hex_1dHex_1Sulf_1$	592.1542
$HexNAc_1Sulf_1$	284.0435	$HexNAc_1Hex_1NeuAc_1Sulf_1$	737.1917
$HexNAc_1Hex_1Sulf_1$	446.0963		

Phos: Phosphate			
$Phos_1$	80.9736	$HexNAc_1Hex_1Phos_1$	446.1058
Hex_1Phos_1	243.0264	$HexNAc_2Phos_1$	487.1324
$HexNAc_1Phos_1$	284.0530		
Pent: Pentose			
$Pent_1$	133.0495	$HexNAc_1Pent_1$	336.1289
Hex_1Pent_1	295.1024		
HexA: Hexuronic acid			
$HexA_1$	177.0394		
KDN: Ketodeoxynonulonic acid			
KDN_1	251.0761		
NeuGc: N-glycolyl-neuraminic acid			
$NeuGc_1$	308.0976	**NeuGc + O-acetylation**	
$NeuGc_1-H_2O$	290.0870	$NeuGc_1Ac_1-H_2O$	332.0976
$HexNAc_1NeuGc_1$	511.1770	$NeuGc_1Ac_1$	350.1082
		$HexNAc_1NeuGc_1Ac_1$	553.1875

3.3.2 Limitations

It is important to note that OxoExtract does not apply a scoring algorithm. As such the program should not be used without manual validation. Only two peptide modifications were included in the parameters (carbamidomethyl Cys and oxidised Met), thus glycopeptides with other peptide modifications need to be investigated manually. Another limitation is that OxoExtract can only be used for single protein searches, and is therefore not useful for investigations into complex samples. Finally, OxoExtract requires both the "query" oxonium ion and the Y1 ion to be present in a spectrum for it to be considered a candidate N-linked glycopeptide. Production and intensities of the Y1 ion has been shown to be dependent on the collision energy used, the charge state of the precursor ion and the amino acid sequence of the glycopeptide (203). Reliance on the Y1 ion may lead to false-negatives (e.g. the Y1 ion is not produced and the glycopeptide is not submitted for GlycoMod searches) and false-positives (e.g. a fragment ion matches the Y1 but is not a true Y1 ion).

3.4   DISCUSSION

At the commencement of this work most available software relied on pre-defined glycan databases (255, 261, 263, 271-273), in a manner analogous to proteomic software. Software that did not require prior knowledge of glycan databases [reviewed in (244)] were not freely available or did not suit the MS methods being applied herein. Manual identification of glycopeptides from NDV HN (discussed in Chapter 4) revealed monosaccharide compositions that were not present in glycan databases. OxoExtract was therefore developed to automate manual searches of N-linked glycopeptides in GlycoMod while also permitting a global view of the fragmentation characteristics of glycopeptides. The optional feature of searching with GlycoMod reduced the time spent identifying glycopeptide candidates and manually validating glycopeptide assignments. By specifying peptide masses and monosaccharide residues for each search it reduced the number of hits allocated by GlycoMod but did not limit the search to structures not yet defined in glycan databases.

In this work, OxoExtract was often used several times to search a single mzML file. The first search was typically completed without the GlycoMod feature. Putative glycopeptides were investigated using information such as the presence of oxonium ions, retention time and glycopeptide and peptide fragment ions. This information was used to build a "glycosylation oxonium ion profile". This profile was then used to set the parameters for further searches using the GlycoMod feature or to ascertain if glycan databases in Byonic were appropriate for the sample being analysed. For example, the presence or absence of diagnostic oxonium ions for NeuGc (251) was used to include or exclude the residue in searches. This helped prevent misassignment of the isobaric monosaccharide pairs Hex_1NeuAc_1 and Fuc_1NeuGc_1 in subsequent searches using OxoExtract and Byonic (249). As another example, the presence of sulfated oxonium ions was used to identify the potential substituent on glycans from each sample. Overall, OxoExtract allowed flexible interpretation of glycopeptide data, which was used to assign N-linked compositions and compile protein-specific glycan databases. These were then combined with the predefined Byonic glycan databases ensuring a more complete profile of glycosylation was obtained.

Chapter 4: Characterisation of glycosylation of Newcastle disease virus haemagglutinin-neuraminidase (HN) protein

4.1 SUMMARY

Members of the *Avulavirus*, *Respirovirus* and *Rubulavirus* genera of paramyxoviruses utilise HN glycoproteins as their attachment proteins. Previous studies have shown that the N-linked glycans present on these proteins can modulate the ability of the virus to infect host cells and stimulate the host immune system. However, site-specific heterogeneity of these glycans has not been defined. This study concerns characterisation of the monosaccharide compositions attached to HN of the *Avulavirus* NDV which causes ND in a range of avian species. The HN protein was derived from egg propagated virions of V4-VAR, an isolate of the avirulent strain QLD/66. Tandem mass spectrometry strategies including CID, HCD and ETD were implemented to characterise the heterogeneity of glycans. Overall 63, 58, and 37 glycans were identified at Asn residues 341, 433 and 481, respectively. Sites N433 and N481 were observed to contain high mannose glycans with paucimannose glycans also observed at site N481. Sites N341, N433 and N481 contained complex or hybrid glycans with many of the glycans containing variations of Fuc and Sulf or Phos. Sialyation of complex or hybrid N-linked glycans was additionally observed at N341 and N433. In addition, a previously undocumented O-linked glycosylation site was identified on the stalk domain of the HN protein. These findings will form the basis for future quantitative glycomic studies of N-linked glycans from NDV HN and assessment of the functional significance of the O-linked glycan in the stalk domain of this protein.

4.2 INTRODUCTION

Highly pathogenic and infectious strains of NDV can result in significant economic losses to the poultry industry worldwide (74, 75). Although vaccines are available to control outbreaks it is not feasible to effectively vaccinate wild avian populations. Consequently, this provides a niche for the virus to propagate and new strains of NDV are continually emerging (61, 77). Effective therapeutic agents may be a very useful adjunct to vaccines to control NDV, particularly for protection of elite

breeding stocks. In addition to the interest on NDV in relation to avian disease, a better understanding of this virus may be beneficial in the context of comparative paramyxoviruses that cause human disease. The HN protein possesses haemagglutination and sialidase activity and the overall architecture and functions of HN from the *Paramyxovirinae* subfamily are quite well conserved. Thus mechanisms of action identified from NDV HN may translate to other human viruses such as hPIV.

As discussed in Chapter 1, mutation studies have shown that sites N119, N341, N433 and N481 from NDV HN strains are likely to be occupied (106, 107). Furthermore, these studies show that the loss of glycosylation sites can affect protein folding and transport of HN, viral replication and pathogenesis. Despite the functional importance of the glycans on NDV HN, no in-depth report of these glycans has been described. Glycan site-specific heterogeneity of HN from NDV is reported for the first time herein. The avirulent isolate of NDV used in this study, herein termed V4-VAR, was a variant of the avirulent QLD/66 strain of NDV (274). The V4-VAR isolate has been used in previous studies to provide important structural insights into HN and F proteins from paramyxoviruses (101, 275, 276). A total of 63, 58 and 37 different N-linked glycans were identified across HN glycosylation sites N341, N433 and N481, respectively. These included high mannose and complex or hybrid glycans that were variably fucosylated, sialylated and sulfated or phosphorylated. In addition, O-linked glycans were identified on a previously undocumented O-linked site from the stalk domain of the protein.

4.3 METHODS

4.3.1 Provision of samples

The viral preparation V4-VAR was provided by Professor Jeffrey Gorman from the Protein Discovery Centre at QIMR Berghofer. The isolate of NDV used was a variant of the Queensland (QLD)/66 strain of NDV (274) propagated and purified in embryonic chicken eggs as previously described (88, 277). Briefly, the virus was grown in nine-day-old embryonic eggs and virions were isolated by centrifugation using a linear sucrose density gradient.

4.3.2 SDS-PAGE separation of V4-VAR virions and electroelution of HN proteins

From stocks of purified V4-VAR virions, approximately 48 µg was reduced and alkylated as per the methods described in Chapter 2 except that 50 mM of IAA was used. Reduced and alkylated V4-VAR viral proteins were subjected to SDS-PAGE followed by staining and de-staining. Bands of interest that migrated to ~75 kDa were excised and stored overnight at 4°C until electroelution as per the methods described in Chapter 2. Electroelution has previously been shown to be an effective technique for recovery of HN proteins in a form compatible with downstream in-solution sample handling strategies (88, 257).

4.3.3 Enzymatic digestions of V4-VAR HN proteins

After electroelution the protein samples were methanol precipitated overnight as per the methods described in Chapter 2, except that 0.05 µg of trypsin was added before precipitation and a further 0.1 µg of trypsin was added for the final incubation in 0.1 M TEAB. An aliquot of the V4-VAR HN tryptic digest was subject to digestion with 1 µL of 1U/µL recombinant PNGase F at 37°C overnight. Resultant peptides were desalted with a C18 ZipTip (10 µL pipette tip with a 0.6 µL resin bed; Millipore, MA, USA) using the manufacturers' guidelines for MS analysis.

4.3.4 Nano-ultra-high pressure liquid chromatography

Approximately 1/10 of the electroeluted and digested sample was injected for each analysis using a nUHPLC system as described in Chapter 2. Samples were loaded onto the trap column and washed for 5 min at 15 µL/min in 98% solvent A and 2% solvent B. Peptides and glycopeptides were subsequently eluted onto the analytical column at flow rate of 0.3 µL/min whilst ramping through a sequence of linear gradients from 2% to 40% solvent B in 60 min, to 70% B over 15 min, to 95% B in 5 min then holding at 95% B for 5 min. The column was then re-equilibrated with 2% B for 20 min.

4.3.5 Mass spectrometry data acquisition

Data were acquired using a variety of fragmentation strategies (203, 204, 228-231) set out in Appendix A. The tryptic V4-VAR sample was run as three separate chromatographic experiments (HCD, HCD-pd-ETD and HCD-pd-CID) and the trypsin/PNGase F sample as one chromatographic experiment (HCD-pd-ETD). It is important to note that this study was not designed to compare the effectiveness of HCD, CID or ETD fragmentation. Different MS acquisition strategies and data analysis pipelines were implemented to qualitatively characterise site-specific glycan heterogeneity of V4-VAR HN.

4.3.6 Sequence conservation of HN across multiple NDV strains

A multiple sequence alignment was used to assess sequence conservation and the presence of N-linked consensus sites in HN across strains of NDV. All reviewed NDV HN sequences were retrieved from UniProt (fifteen HN sequences retrieved on 12 March 2016) (278). The protein sequences were aligned using Clustal Omega (279) available at www.uniprot.org applying the default parameters (Gonnet transition matrix, gap opening penalty of six bits and gap extension of one bit) (280).

4.3.7 Determining amino acid sequence changes in V4-VAR HN

Proteome Discoverer (v. 1.4.1.14) and the search engine Mascot were used to search spectra from the HCD MS/MS analysis of tryptic peptides from V4-VAR. The protein database used contained the complete proteome of chicken and reviewed protein sequences from NDV (downloaded from www.uniprot.org on 18 April 2016 consisting of 24,094 chicken and 43 NDV protein sequences) and 247 common contaminant sequences (281). The following parameters were used for the Mascot search: digestion with trypsin; maximum of two missed cleavages; 10 ppm precursor mass tolerance; 0.02 Da fragment tolerance; fixed modification of carbamidomethylation of Cys and dynamic modifications of mono-oxidised methionine Met and deamidation of Asn and Gln residues. Confident peptide-to-spectrum matches (PSMs) were assigned using the Proteome Discoverer "Target Decoy PSM Validator" and a PSM false discovery rate (FDR) threshold of 0.05 was applied. At least two unique peptides were required for confident protein identifications. To

confirm the amino acid sequence of HN from V4-VAR, PSMs for NDV HN were manually validated. Peptides not identified in the HN sequence by the Mascot search were investigated by manual *de novo* sequencing in Xcalibur Qual Browser.

4.3.8 Assignment of N-linked monosaccharide compositions and glycopeptides to V4-VAR HN using HCD fragmentation

Spectra from the HCD MS/MS analysis of tryptic peptides from V4-VAR were analysed with OxoExtract. The following parameters were used: digestion with trypsin; maximum of two missed cleavages; fixed modification of carbamidomethylation of Cys and a dynamic modification of mono-oxidised Met. The protein database queried contained the V4-VAR HN sequence. The optional feature of GlycoMod was enabled. In addition to the default monosaccharide parameters the substituents Phos and Sulf were considered as possible components of the N-glycans. In addition to assigning glycopeptides to V4-VAR HN the observed glycans from the OxoExtract search were also used for subsequent searches in Byonic (detailed below).

4.3.9 Assignment of V4-VAR HN glycopeptides using ETD and CID fragmentation

Spectra obtained from ETD and CID fragmentation of trypsin digested V4-VAR HN were converted to MGF format and analysed in Byonic. The majority of the ETD and CID scans were putative glycopeptides due to the product dependent function. Cleavage specificity was set C-terminal to Lys and Arg residues allowing a maximum of two missed cleavages, a precursor mass tolerance of 10 ppm and a fixed modification of Cys carbamidomethylation. The relevant fragmentation type was selected for each search, ETD or CID, with 0.02 Da and 0.6 Da fragment ion tolerances applied, respectively. One glycan attached at an N-linked consensus site and one mono-oxidised Met were allowed per peptide. The protein database queried contained the V4-VAR HN sequence. The N-linked glycan database queried was a combination of the Byonic mammalian glycan database (309_Mammalian no sodium) and any additional glycans (73 in total) assigned to V4-VAR HN from the OxoExtract search. The default protein false FDR and peptide output options were changed to "Show all N-glycopeptides" which is recommended by the manufacturer when analysing simple samples. By selecting this parameter all identified N-linked glycopeptides are shown irrespective of the assigned Byonic PSM score or FDR. A manual cut-off score of 100 was applied for glycopeptides containing N341 and N433, while no cut-off score was applied to

glycopeptides containing N481 as the small size of the peptide backbone and poor peptide fragmentation likely resulted in lower scores. All glycans assigned by Byonic to ETD or CID spectra were confirmed in Xcalibur Qual Browser through relevant oxonium ions from the corresponding HCD MS/MS spectrum that triggered the ETD or CID scan of interest.

For all glycopeptides assigned by ETD and CID, the HCD MS/MS spectrum that triggered ETD or CID scan of interest was investigated for diagnostic oxonium ions which were used to confirm the monosaccharide composition of an assigned N-linked glycan. Glycopeptides identified by ETD were accepted if the mass of the attached glycan could be localised to Asn in the N-linked consensus (N-X-S/T) with complete sequencing information of the intervening Ser/Thr/Tyr residues. When assigning glycopeptides using CID spectra, Byonic did not annotate the majority of fragment ions that were derived from fragmentation of the glycan portion of glycopeptides. Manual validation of CID was completed with the aid of a glycan calculator created manually in Microsoft Excel (Supplementary Table S4-1). To calculate glycopeptide fragment ions the masses of theoretical oligosaccharides and glycans were first calculated using "Carbo Calculator" provided by IonSource (http://www.ionsource.com/Card/carbo/carbstr.htm). These oligosaccharide and glycan masses were added to the mass of peptides containing the N-linked consensus sites and m/z values for different charge states were calculated for each theoretical glycan fragment ion. Glycopeptides identified by CID were accepted if all monosaccharide residues from the allocated glycan were observed at least once in the MS/MS spectrum and all intense peaks (greater than 10% relative abundance) could be accounted for. If there wasn't clear evidence of the peptide portions of the assigned glycopeptides (typically an intense ion representing Y1) the assignment was further confirmed through manual inspection of the corresponding HCD spectrum.

4.4   RESULTS

4.4.1 Prediction of potential glycosylation sites in V4-VAR HN

The V4-VAR viral preparation used in this study was a variant of the QLD/66 strain of NDV (UniProt accession number P13850). As such, the precise amino acid sequence of HN V4-VAR was not known before analyses were undertaken which potentially precluded identification of potential N-linked glycosylation sites for this protein. To predict potential sites of N-linked glycosylation in V4-VAR HN the conservation of potential sites amongst NDV strains was examined by alignment

of HN sequences. Only reviewed HN sequences from UniProt were used in the alignment and this included HN from the QLD/66 strain. **Figure 4-1** highlights potential tryptic peptides containing each N-linked consensus site from the alignment of the HN strains (the full alignment is presented in Appendix B with the names of the strains from each Uniprot Identifier). **Figure 4-1** also illustrates the position of the N-linked sites in HN. The alignment revealed that five N-linked consensus sites were relatively conserved across all strains of NDV. The consensus sites correspond to potential glycosylation at N119, N341, N433, N481 and N538 in the sequence of HN from QLD/66. Some strains of NDV have one additional N-linked consensus site corresponding to N144, N508 or N600. Of the additional potential sites, the sequence of HN from the QLD/66 strain contained the consensus site for N600.

```
                    119                         341                        508              538
P13850 NDVQ   (R)INGAANSSGCGAPIHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIQMAK(S)  (R)LNPVSAVFDSISR(S)  (V)KTNK(T)
P12558 NDVU   (Q)INGAANSSGCGAPIHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPVSAVFDSISR(S)  (V)KTNK(T)
P12555 NDVD   (Q)INGAANSSGCGAPIHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPVSAVFDSISR(S)  (V)KTNK(T)
Q9Q2W5 NDVK   (Q)INGAANNSGWGAPIHDPDFIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPASAVFDSTSR(S)  (V)KTNK(T)
P32884 NDVB   (Q)INGAANNSGWGAPIHDPDFIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPTSAVFDSTSR(S)  (V)KTNK(T)
P35743 NDVL   (Q)INGAANNSGWGAPIHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPASAVFDSTSR(S)  (V)KTNK(T)
P12559 NDVH4  (Q)INGAANNSGWGAPIHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPASAVFDSTSR(S)  (V)KTNK(T)
P12553 NDVTG  (Q)INGAANNSGWGAPIHDPDFIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPASAVFDSTSR(S)  (V)KTNK(T)
Q91UL0 NDVB1  (Q)INGAANNSGWGAPIHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPASAVFDSTSR(S)  (V)KTNK(T)
P12556 NDVI   (Q)INGAANNSGCGAPVHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPVSAVFDNISR(S)  (V)KTNK(T)
P35742 NDVJ   (Q)INGAANNSGCGAPVHDPDYIGGIGK(E)  (K)RYNNTFPDKQDYQIRMAK(S)  (R)LNPVSAVFDNISR(S)  (V)KTSKA
P35741 NDVH3  (Q)INGAANNSGCGAPVHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPVSAVFDNISR(S)  (V)KTNK(T)
P35740 NDVC   (Q)INGAANNSGCGEPVHDPDYIGGIGK(E)  (K)RYNNTCPDEQDYQIRMAK(S)  (R)LNPVSAVFDNISR(S)  (V)KTNKA
P12557 NDVM   (Q)INGAANNSGCGAPVHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNPVSAVFDNVSR(S)  (V)KTNK(T)
P12554 NDVA   (Q)INGAANNSGCGAPVHDPDYIGGIGK(E)  (K)RYNDTCPDEQDYQIRMAK(S)  (R)LNLVSAVFDNISR(S)  (V)KTNK(T)
              :.******.** * *.:****.:*******   * ***:* **:*****.:****   ***  *****. ***   ***.*:

NH2 — CT  TM   SD  [                Globular Head Domain                ] — COOH

                    144                             433           481            600
P13850 NDVQ   (K)ELIVDDASDVTSYYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMIVSNK(T)  (R)NHTLR(G)  (K)NQTEYR(R)
P12558 NDVU   (K)ELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMVSNK(T)   (R)NHTLR(G)  (K)NQTEYR(R)
P12555 NDVD   (K)ELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMVSNK(T)   (R)NHTLR(G)  (K)NQTEYR(R)
Q9Q2W5 NDVK   (K)ELIVDNASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMVSNK(T)   (R)NHTLR(G)  --------
P32884 NDVB   (K)ELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMVSNK(T)   (R)NHTLR(G)  --------
P35743 NDVL   (K)ELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMVSNK(T)   (R)NHTLR(G)  --------
P12559 NDVH4  (K)ELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMVSNK(T)   (R)NHTLR(G)  --------
P12553 NDVTG  (K)ELIVDDASDVTSFYPSAFQEHHNFIPAPTTGSGCIR(I)  (R)GSSYFSPALLYPMVSNK(T)   (R)NHTLR(G)  --------
Q91UL0 NDVB1  (K)ELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMVSNK(T)   (R)NHTLR(G)  --------
P12556 NDVI   (K)ELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMTVNNK(T)  (R)NHTLR(G)  --------
P35742 NDVJ   (K)ELIVDDTSDVTSFYPSAYQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMTVYQQ(T)  (R)NHTLR(G)  --------
P35741 NDVH3  (K)ELIVDDASDVTSFYPSAYQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMTVNNK(T)  (R)NHTLR(G)  --------
P35740 NDVC   (K)ELIVDDISDVTSFYPSAYQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMTVNNK(T)  (R)NHTLR(G)  --------
P12557 NDVM   (K)ELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMTVNNK(T)  (A)NHTLR(G)  --------
P12554 NDVA   (K)ELIVDDTSDVTSFYPSAFQEHLNFIPAPTTGSGCTR(I)  (R)GSSYFSPALLYPMTVNNN(T)  (R)NHTLR(G)  --------
              * *****: *****:****:*** ************* **   * ***********  * ::*   ******
```

* Positions have a single, fully conserved residue.
: Conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix.
. Conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix.

Figure 4-1. Alignment of HN sequences revealing potential tryptic peptides that contain N-linked consensus sites. Multiple sequence alignment of NDV HN was implemented using fifteen annotated NDV HN sequences from the UniProt website. Sequence identifiers are listed on the left and the NDV Queensland/66 strain (UniProt entry P13850) is the first identifier. All N-linked consensus (N-X-S/T) sites have been highlighted in dark grey. A schematic (not to scale) is presented of NDV HN (UniProt ID P13850) identifying the cytoplasmic tail (CT), transmembrane (TM), stalk domain (SD) and the globular head domain.

4.4.2 Isolation of V4-VAR HN

The HN protein was observed to migrate as a diffuse band corresponding to approximately 75 kDa during SDS-PAGE of V4-VAR proteins (**Figure 4-2**). As the calculated mass of HN is 67,656 Da (UniProt entry P13850) the diffuse nature of this band is likely attributable to varying degrees of glycosylation of the N-linked sites. Slices corresponding to the 75 kDa regions were excised and pooled before the intact proteins were extracted using electroelution. The harvested proteins were subsequently subjected to enzymatic digestion followed by MS analysis.
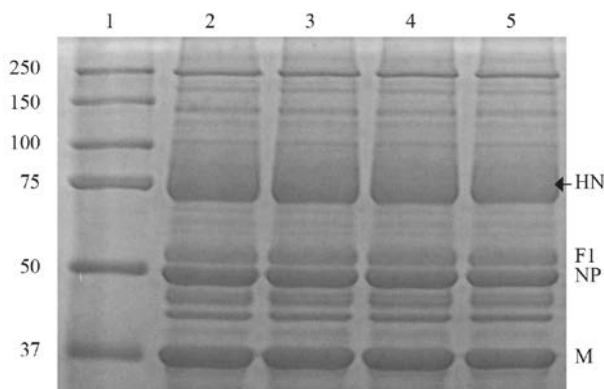
Figure 4-2. SDS-PAGE separation of NDV V4-VAR proteins. Lane 1 contains the MW markers with the protein masses shown in kDa. Lanes 2-5 were each loaded with ~12 µg of reduced and alkylated V4-VAR virion proteins. Bands previously identified as NDV HN, fusion F_1 chain (F1), nucleocapsid (NP) and matrix (M) proteins (88) are indicated. Bands from each lane containing HN (denoted with an arrow) were excised from the gel and subject to electroelution.

4.4.3 Occupancy status of putative N-linked sites on HN from V4-VAR

The RAW file generated from HCD MS/MS of tryptic peptides from V4-VAR was analysed in Xcalibur Qual Browser by producing an EIC within a ±10 ppm window for the theoretical *m/z* of [HexNAc+H]$^+$. Precursors that produced fragmentation ions corresponding to HexNAc were seen to elute from six to 55 min interspersed between other apparently non-glycosylated tryptic peptides (**Figure 4-3a** and **b**). The sequence of HN from QLD/66 was used to calculate theoretical *m/z* values for Y1 ions for tryptic peptides containing each potential N-linked site (N119, N341, N433, N481, N538 and N600). An EIC was produced for these theoretical Y1 ions and inspection of spectra that additionally yielded abundant fragment ions corresponding to HexNAc revealed potentially glycosylated peptides containing sites N481 (^{481}<u>N</u>HT<u>L</u>R^{485}), N341 (^{340}Y<u>NDT</u>cPDEQDYQIQMAK^{356} and ^{339}RY<u>NDT</u>cPDEQDYQIQMAK^{356}) and N433 (^{417}GSSYFSPALLYPMIVS<u>NK</u>^{434}) (**Figure 4-3c, d** and **e**, respectively) where lowercase "c"

represents Cys carbamidomethylation. Glycopeptide Y1 ions for all three tryptic peptides containing sites N341 and N433 were also observed with an additional mass of 16 Da, which was later attributed to oxidised Met in the peptide sequence (**Figure 4-3d** and **e**, respectively). As expected, glycopeptides that contained oxidised peptide species eluted several minutes earlier than their non-oxidised counterparts. This initial manual investigation revealed glycosylated tryptic peptides containing the N-linked sites N341, N433 and N481 as predicted from the parental QLD/66 isolate. Evidence of glycosylation at predicted sites N119, N538 and N600 was not observed.
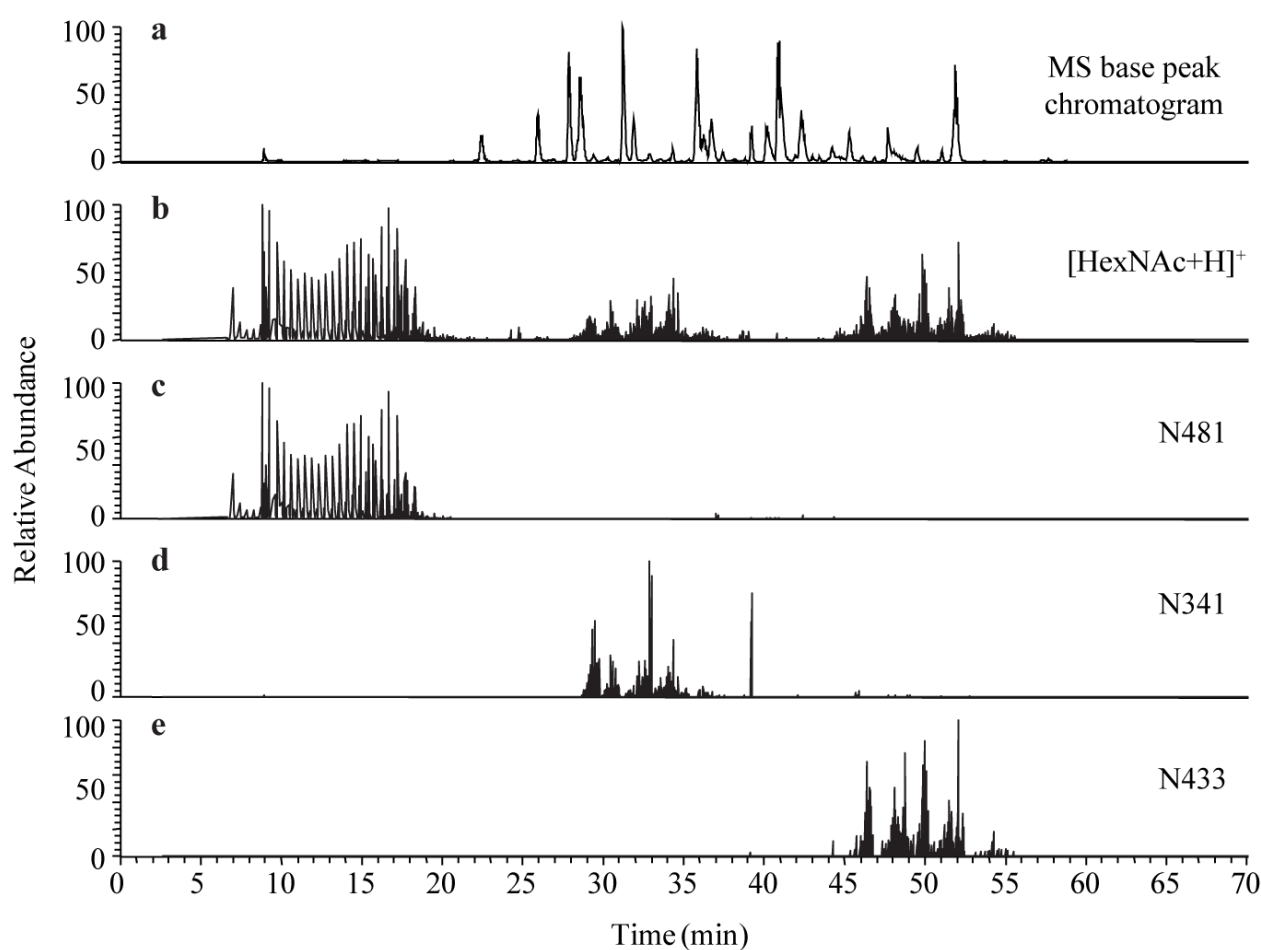
Figure 4-3. Retention time profiles of putative N-linked glycopeptides from HCD MS/MS analysis of NDV V4-VAR HN digested by trypsin. The MS base peak chromatogram (a) depicts the retention time profile of peptides and glycopeptides eluted throughout the LC-MS/MS experiment. The EIC at the MS/MS level of *m/z* 204.0867 (b) indicated the presence of putatively glycosylated peptides. The EICs at the MS/MS for the calculated Y1 ions of glycopeptides containing N-linked sites N481 ([481]NHTLR[485]), N341 ([340]YNDTcPDEQDYQIQMAK[356] and [339]RYNDTcPDEQDYQIQMAK[356] with and without oxidised Met) and N433 ([417]GSSYFSPALLYPMIVSNK[434] with and without oxidised Met) delineate the retention times of the glycopeptides (c, d and e, respectively).

4.4.4 Amino acid sequence changes in V4-VAR HN

The Mascot search of the RAW file generated from HCD MS/MS of tryptic peptides from V4-VAR identified 146 proteins (PSM FDR of 0.05, two distinct peptides per protein) (Supplementary Table S4-2). Mascot-assigned PSMs resulted in 75% sequence coverage of HN (QLD/66 strain, UniProt entry P13850) (**Figure 4-4**). PSMs containing the potential N-linked sites N433 and N538 were detected by the Mascot search while PSMs corresponding to tryptic peptides containing other N-linked sites of the parental QLD\66 isolate (N119, N341 and N600) were undetected. Theoretical *m/z* values for the potential tryptic peptides containing site N481 in a non-glycosylated form were below the *m/z* scan range set in the MS acquisition parameters.

Figure 4-4. Protein sequence coverage of HN based on identified peptides from HCD MS/MS analysis of V4-VAR HN digested with trypsin. Sequence coverage derived from the Mascot search is underlined. Sequence variations from the parental QLD/66 isolate are included in the main sequence with the variant amino acid residues of the QLD/66 sequence listed below the variations. The amino acid sequence number for Asn residues within N-linked consensus sequences (N-X-S/T) from QLD/66 have been noted above the relevant amino acid residue (N119, N341, N433, N481, N538 and N600).

When considering initial investigations of N-linked site occupancy and the results from the Mascot search, four tryptic peptides of the HN sequence were undetected (Q87-R113, I114-K138, I175-R197 and N600-R605). Theoretical b- and y-ions were calculated for tryptic peptides from these undetected sequences using the QLD/66 HN sequence. Manual *de novo* sequencing identified four amino acid sequence variations, R113Q, N115S, N119S and I175M in three of the missing tryptic peptides, Q87-R113, I114-K138 and I175-R197. The amino acid sequence variation R113Q resulted in one large tryptic peptide Q87-K138. The position of the amino acid variations in the HN protein sequence are indicated in **Figure 4-4**. Spectral evidence for the sequence variations from each tryptic peptide Q87-K138 and I175-R197 are presented in Appendix C and D, respectively. The observed mutations have been previously documented in virions derived from the QLD/66 strain (101, 282, 283). Importantly, the N119S mutation abolished one of the six potential N-linked glycosylation sites predicted in V4-VAR HN. No evidence of the tryptic peptide N600-R605, which contains N-linked site N600, was observed. The results from the initial investigations of N-linked site occupancy, Mascot search and manual *de novo* sequencing were combined to construct an amino acid sequence for HN of V4-VAR.

4.4.5 Characterisation of N-linked glycopeptides of V4-VAR HN using HCD

Analysis of the HCD MS/MS file with OxoExtract identified a total of 255 different N-linked glycopeptides (Supplementary Table S4-3). These were distributed across seven different peptide masses and three N-linked sites (N341, N433 and N481). Representative spectra from manually validated glycopeptides containing N341, N433 and N481 can be found in **Figure 4-5a**, **b** and **c**, respectively. Glycopeptide Y1 ions (identified as "peptide+HexNAc$_1$" in annotated spectra in this thesis) were observed for all three glycopeptides and oxonium ions supported the proposed composition of the glycans. In **Figure 4-5a** fragmentation of a glycopeptide containing N341 resulted in near complete peptide sequence coverage. The sequence ions b$_2$ and b$_5$ were identified with HexNAc$_1$ attached, the former of these indicated glycosylation at N341. In **Figure 4-5b** fragmentation of a glycopeptide containing N433 resulted in significant peptide sequence coverage. Sequence ions from the y-series (with the addition of HexNAc$_1$ and HexNAc$_2$) were observed as quite intense ions. As can be seen in **Figure 4-5c** HCD fragmentation of glycopeptides containing N481 did not result in significant peptide fragmentation and the only sequence ions (y$_2$ and y$_3$) in that spectrum were of low relative abundance (the numbers of b- and y-ions observed for every glycopeptide assigned by HCD are presented in Supplementary Table S4-3).

Glycans that contained Sulf or Phos, as determined by precursors mass, were confidently assigned as sulfated if oxonium ions $HexNAc_1Sulf_1$ or $HexNAc_1Hex_1Sulf_1$ were present. The distinction between sulfated and phosphorylated glycopeptides can be made using mass shifts between glycan oxonium ions in a single spectrum (238, 267). For all confidently assigned sulfated compositions the mass shifts between ions matching the *m/z* for $HexNAc_1$ and $HexNAc_1Sulf_1$ or $HexNAc_1Hex_1$ and $HexNAc_1Hex_1Sulf_1$ were observed to range between 79.9545 Da and 79.9582 Da (mass differences listed in Supplementary Table S4-3). An example is presented in **Figure 4-6**, where sulfation was evident by the oxonium ion at *m/z* 446.096 (1+) corresponding to $HexNAc_1Sulf_1Hex_1$ which was calculated as being 79.957 Da higher than the observed $HexNAc_1Hex_1$ at *m/z* 366.139 (1+). This confirmed the presence of a Sulf group (79.9568) as opposed to Phos (79.9663). Sulfated oxonium ions were only present in 23% of spectra of glycopeptides where the monosaccharide compositions contained Sulf or Phos. No phosphate-specific oxonium ions were observed; therefore the remaining "Sulf or Phos" monosaccharide compositions were assigned as ambiguous (Supplementary Table S4-3).

4.4.6 Characterisation of N-linked glycopeptides of V4-VAR HN using CID and ETD

The ETD search identified 13 different glycopeptides, all of which contained site N481 in the peptide sequence [481]NHTLR[485] (manually validated Byonic results in Supplementary Table S4-4). The fragmentation patterns produced by ETD enabled the site of glycan attachment to be localised to N481. An example is presented in **Figure 4-7** revealing near complete peptide sequence coverage. The glycan structure can be assigned to N481 using the c_1 ion and losses of monosaccharides from the precursor ion (Hex_1 and $HexNAc_1$) also help confirm the monosaccharide composition of allocated glycan. The peptide sequence information provided by ETD was advantageous as HCD produced minimal peptide fragmentation of glycopeptides containing N481 (**Figure 4-5c**). As such, ETD confirmed that the observed Y1 ions from HCD spectra of glycopeptides containing N481 were in fact from the predicted N-linked peptide [481]NHTLR[485]. All 13 glycopeptides assigned using ETD contained glycan structures that were neutral. Most precursors were triply charged with precursor (*m/z*) values between 633 and 1172, making them amenable to fragmentation by ETD (284). Due to minimal fragmentation of the peptide backbone ETD did not confirm glycan site localisation for glycopeptides containing sites N341 and N433. However, the amino acid sequence of peptides containing these sites was confirmed by HCD fragmentation and it was assumed N341 and N433 were the sites of attachment.
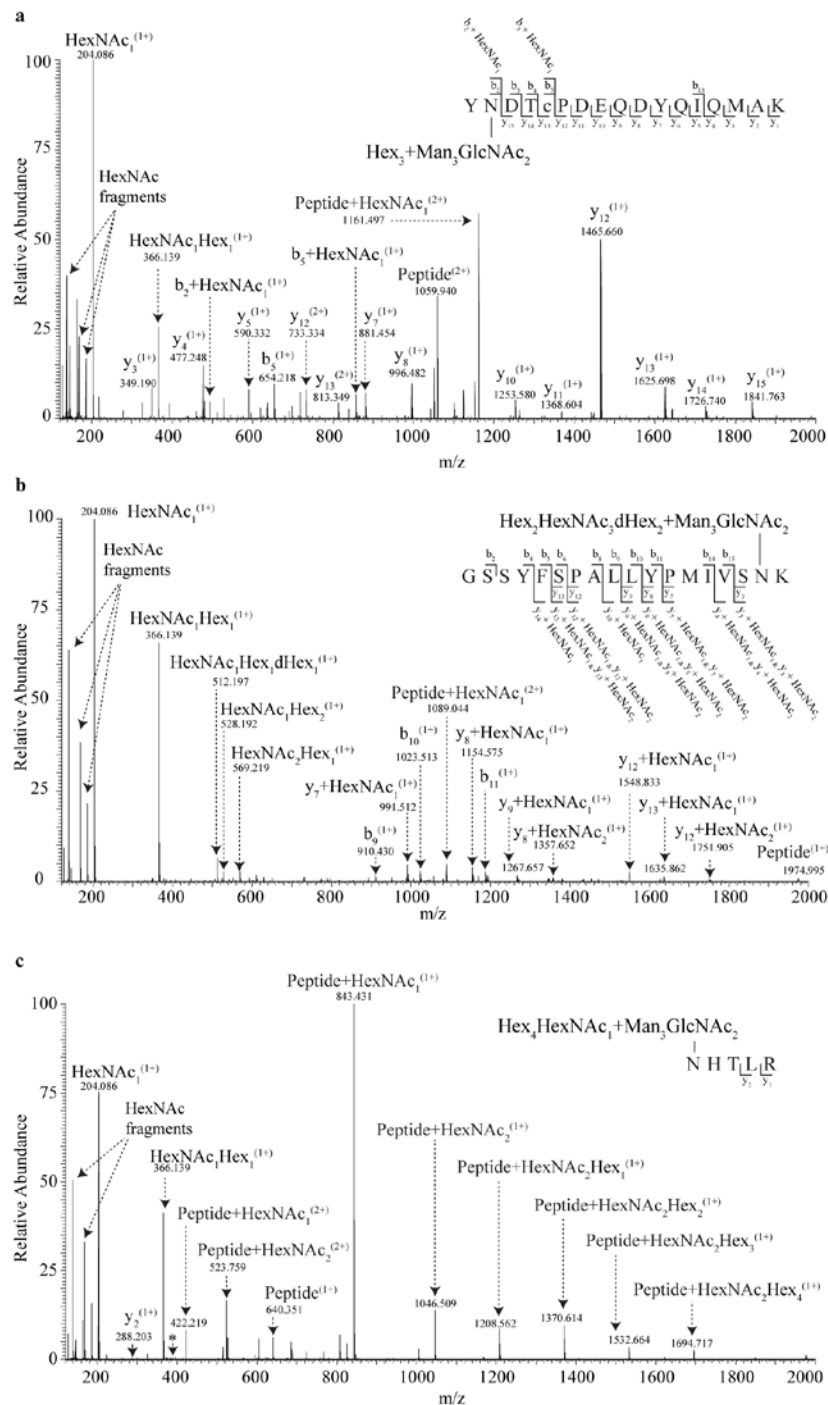
Figure 4-5. HCD fragmentation of N-linked glycopeptides from V4-VAR HN. Each panel has a schematic of the peptide fragmentation pattern observed for the respective glycopeptides. Lowercase "c" in the schematic respresent Cys carbamidomethylation. Not all ions have been labelled in the spectra for ease of interpretation. (a) HCD fragmentation of a precursor at *m/z* 1166.457 (3+) containing N341 (aa 340-356). (b) HCD fragmentation of a precursor at *m/z* 1345.590 (3+) containing N433 (aa 417-434). (c) HCD fragmentation of a precursor at *m/z* 795.325 (3+) containing N481 (aa 481-485) where peptide sequence ion y_3 at *m/z* 389.250 is denoted by "*".
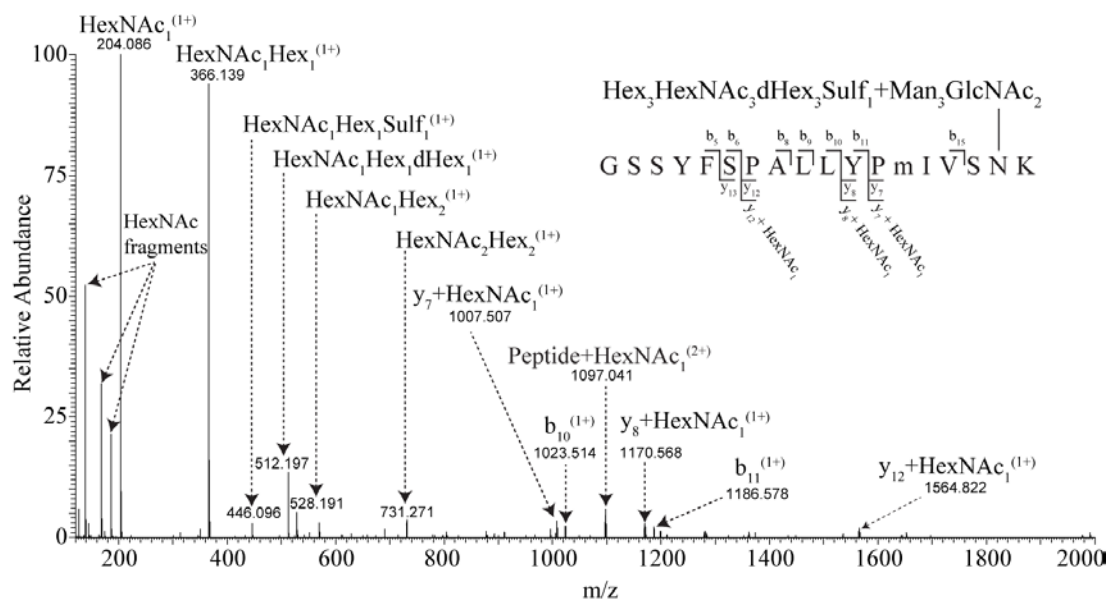
Figure 4-6. HCD fragmentation of a glycopeptide from V4-VAR HN containing site N433 with a mono-sulfated glycan attached. HCD fragmentation of a precursor at *m/z* 1124.714 (4+) containing N433 (aa 417-434). A schematic of the peptide fragmentation pattern is shown but not all ions have been labelled in the spectra for ease of interpretation. Lowercase "m" in the schematic respresents oxidised Met.
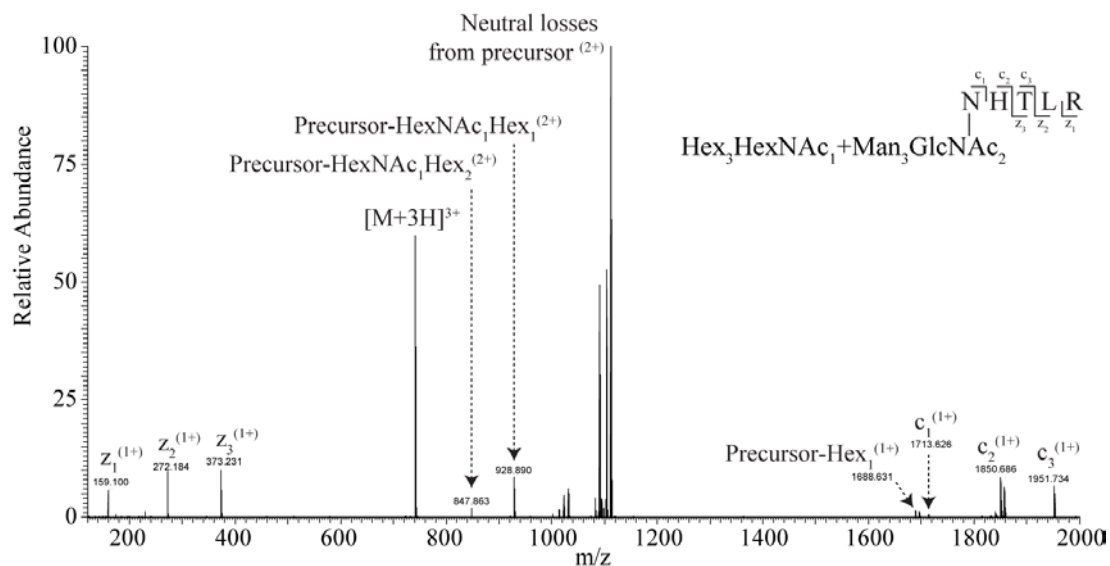
Figure 4-7. ETD sequence coverage of an N-linked glycopeptide containing site N481 from V4-VAR. HCD fragmentation of a precursor at *m/z* 741.308 (3+) containing N481 (aa 481-485). A schematic of the fragmentation pattern is shown with the spectrum labelled accordingly.

The fragmentation patterns produced by CID enabled identification of 125 different glycopeptides containing three N-linked sites (N341, N433 and N481) (Supplementary Table S4-4). Fragmentation of glycopeptides using CID provided sequence information for the attached glycan moieties that was complementary to that observed in HCD. In particular, CID proved useful for the analysis of sulfated or phosphorylated glycopeptides enabling masses of ~80 Da to be identified in the peptide+glycan ions in spectra from both mono- and di-sulfated/phosphorylated glycans (**Figure 4-8** and **Figure 4-9**, respectively). Fragment ions produced by CID were analysed in the ion trap and therefore did not enable the distinction between Sulf and Phos to be made, but the masses of ~80 Da could be used to confirm the glycan contained "Sulf or Phos". In **Figure 4-8** and **Figure 4-9** masses of HexNAc (203), Hex (162) and dHex (146) were observed in the peptide+glycan ions confirming the monosaccharide composition of the attached glycan. Sulfation/phosphorylation was supported by the fragment ions at *m/z* 1790.90 and 1830.90 in **Figure 4-8**. These ions are likely to be doubly charged with a mass difference equating to ~80 Da. In **Figure 4-9** fragment ions at *m/z* 1363 and 1385 reveal losses of dHex and ~80 Da from the precursor, respectively, and masses of ~80 Da can be observed throughout the peptide+glycan ions.
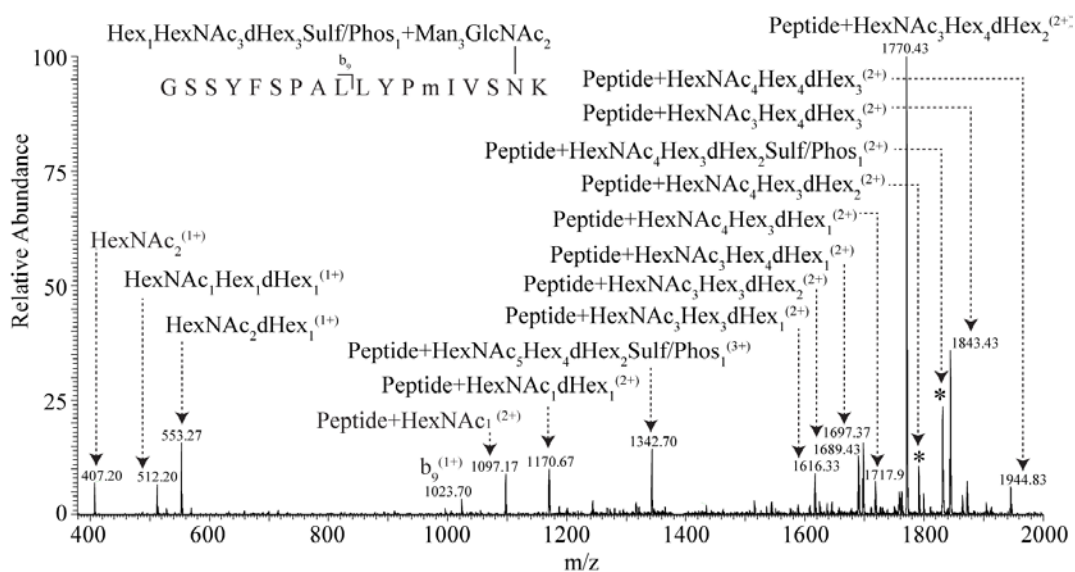
Figure 4-8. CID fragmentation of a glycopeptide from V4-VAR HN containing site N433 with a mono-sulfated/phosphorylated glycan attached. CID fragmentation of a precursor at *m/z* 1391.253 (3+) containing N433 (aa 417-434). A schematic of the fragmentation pattern is shown with the spectrum labelled accordingly. Lowercase "m" in the schematic respresent oxidised Met. Fragment ions at *m/z* 1790.90 and 1830.90 in the spectra are denoted by an "*".
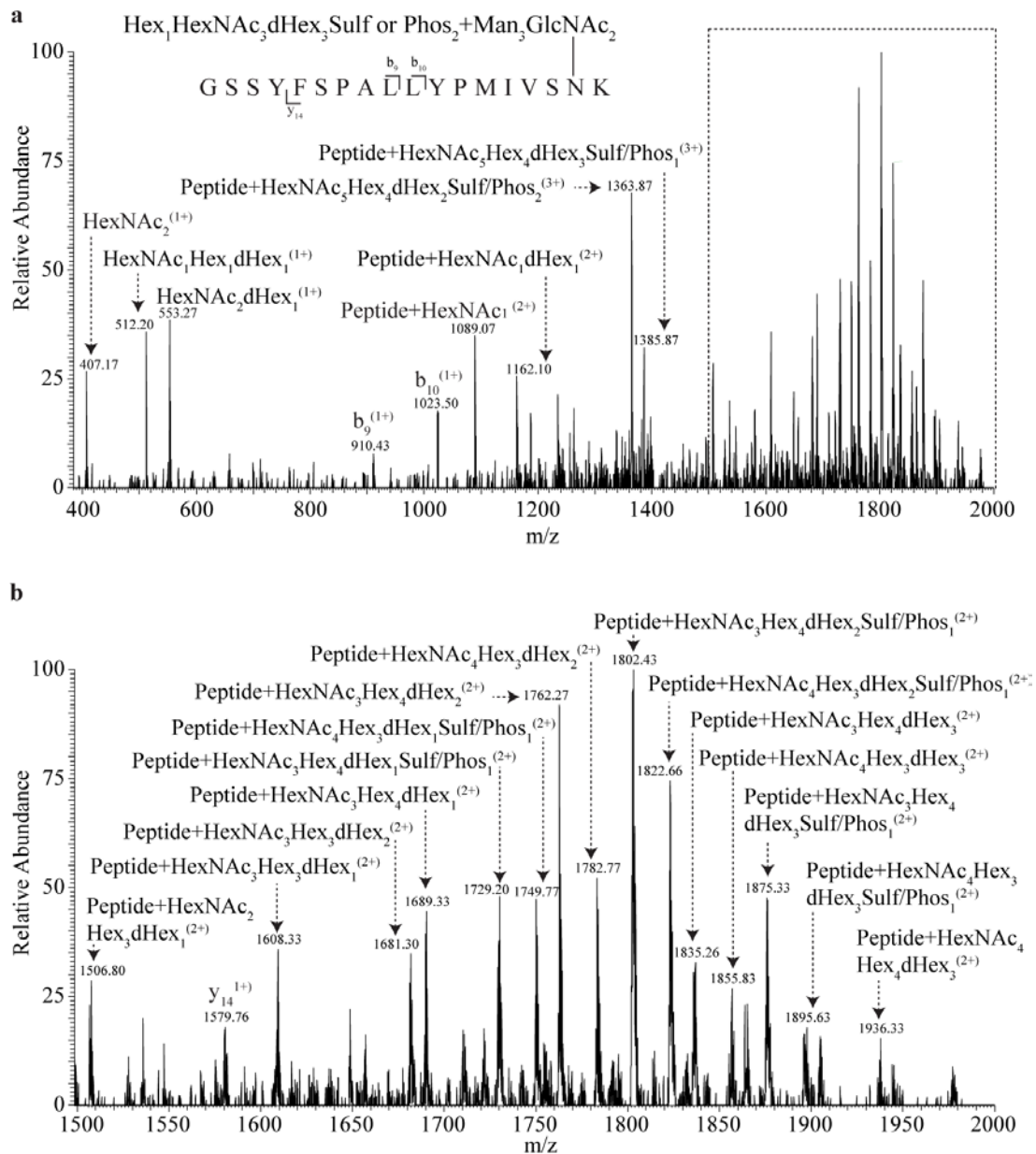
Figure 4-9. CID fragmentation of a glycopeptide from V4-VAR HN containing site N433 with a di-sulfated/phosphorylated glycan attached. CID fragmentation of a precursor at *m/z* 1412.573 (3+) containing N433 (aa 417-434). (a) Reveals the spectrum from *m/z* 400 to 2000 with a schematic of the fragmentation pattern and the spectrum labelled accordingly. (b) Highlights the same spectrum at *m/z* 1500 to 2000.

4.4.7 Diversity of N-linked glycosylation of V4-VAR HN

The glycan heterogeneity documented across V4-VAR HN was quite extensive. Combining the results from the HCD, ETD and CID searches and then considering each glycan only once at each N-linked site, enabled 63, 58 and 37 N-linked glycans to be identified at sites N341, N433 and N481, respectively. These included paucimannose ($HexNAc_2Hex_{3-4}$), high mannose and complex or hybrid structures that were neutral, sialylated and sulfated or phosphorylated (**Figure 4-10**). High levels of fucosylation were also observed with some glycans containing up to five Fuc residues.
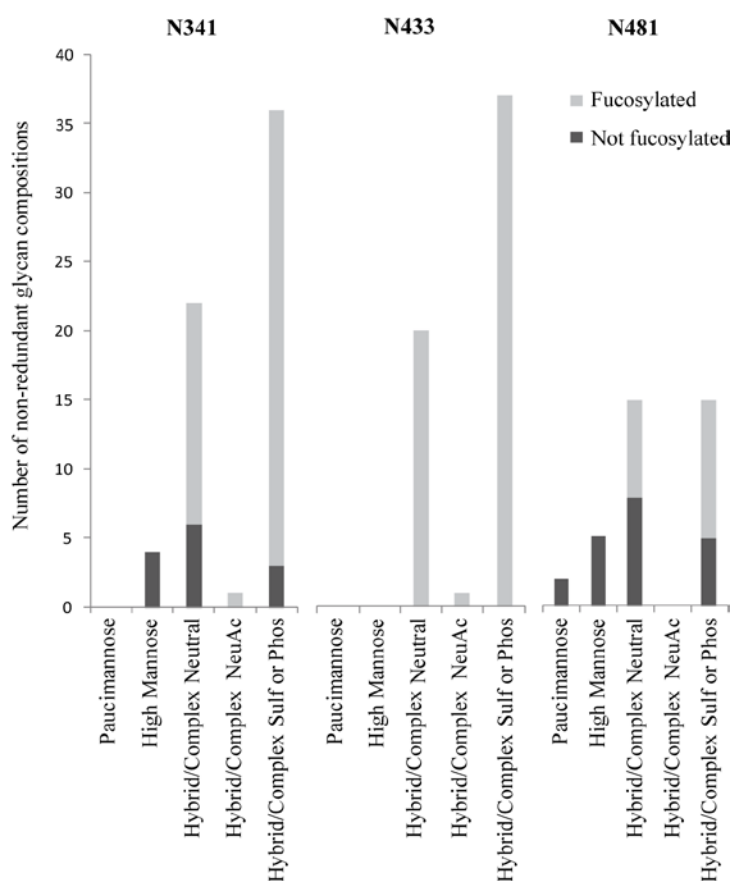
Figure 4-10. Qualitative differences in glycans identified at sites N341, N433 and N481 of V4-VAR HN. Glycans have been grouped into paucimannose, high mannose and hybrid or complex. The latter group has been separated into those with neutral glycans or those containing NeuAc or Sulf/Phos. The light grey and dark grey areas of the histogram represent the number of glycan with and without fucose, respectively.

4.4.8 Assignment of the monosaccharide compositions of O-linked glycans from HN of NDV isolates

During initial investigations of NDV V4-HN evidence of an O-linked glycopeptide was observed (data not included) but with poor fragmentation. To investigate potential O-linked glycosylation of HN from V4-VAR the tryptic digest was subjected to PNGase F digestion followed by MS analysis using an HCD-pd-ETD strategy. Xcalibur Qual Browser was used to analyse the RAW file and produce an EIC for the theoretical m/z of $[HexNAc+H]^+$. Manual inspection of HCD spectra yielding fragment signals corresponding to HexNAc identified an O-linked glycopeptide (^{66}AEEKITSALGSNQVVDR^{83}) with two sialylated glycoforms ($HexNex_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$) (**Figure 4-11a** and **b**, respectively). The glycopeptide Y0 ions (identified as "peptide" in annotated spectra in this thesis) were observed for both glycopeptides and represent the complete loss of the glycan moiety. Oxonium ions for HexNAc, Hex and NeuAc supported the proposed monosaccharide compositions of the glycans. In **Figure 4-11a** near complete peptide sequence coverage was obtained confirming the amino acid sequence. The glycosylated peptide contained three potential sites of O-linked glycosylation T71, S72 and S76. The site of attachment of the O-linked glycans from V4-VAR could not be identified in the pd-ETD spectra, as there was little or no dissociation of the precursor ions. However, a preliminary study using a different NDV isolate propagated from the QLD/66 strain (herein termed V4-QLD) identified an additional glycoform ($HexNAc_1Hex_1$). Fragmentation of this glycoform using HCD (**Figure 4-12a**) revealed near complete peptide sequence coverage and oxonium ions supported the proposed monosaccharide composition of the glycan. Fragmentation of this glycoform using ETD (**Figure 4-12b**) revealed that S76 was not the site of glycosylation through ions c_{10}, c_{11}, z_7 and z_8. The c_6 ion (denoted with a "*" in **Figure 4-12b**) indicated that the site of glycosylation was T71.
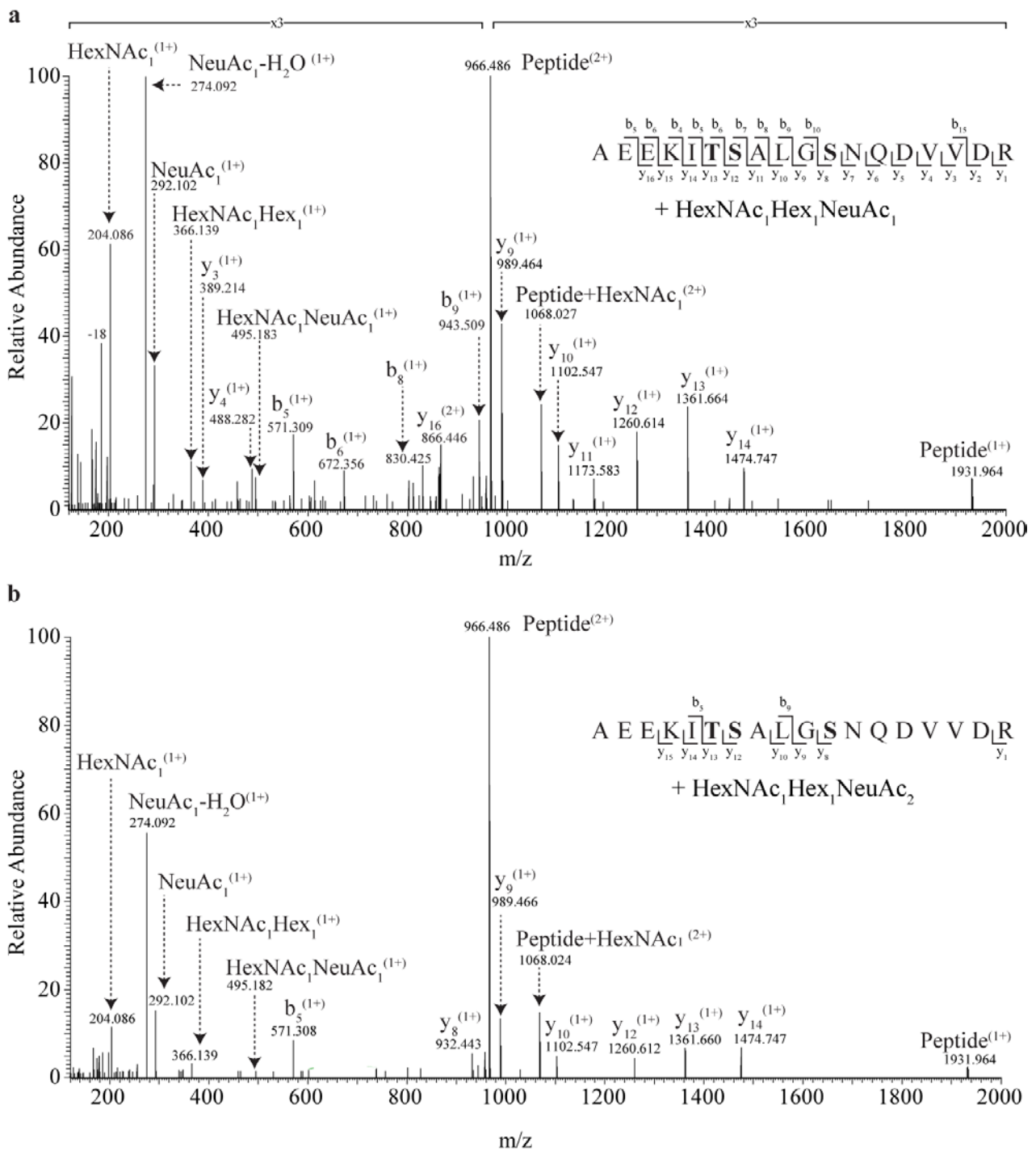
Figure 4-11. HCD fragmentation of two O-linked glycoforms of the same peptide sequence from V4-VAR HN. Each panel has a schematic of the peptide fragmentation pattern observed for the respective glycopeptides (aa 66-83). Potential sites of O-linked glycosylation are denoted in bold. Not all ions have been labelled in the spectra for ease of interpretation. (a) HCD fragmentation of the precursor ion at *m/z* 863.404 (3+). (b) HCD fragmentation of the precursor ion at *m/z* 960.437 (3+).
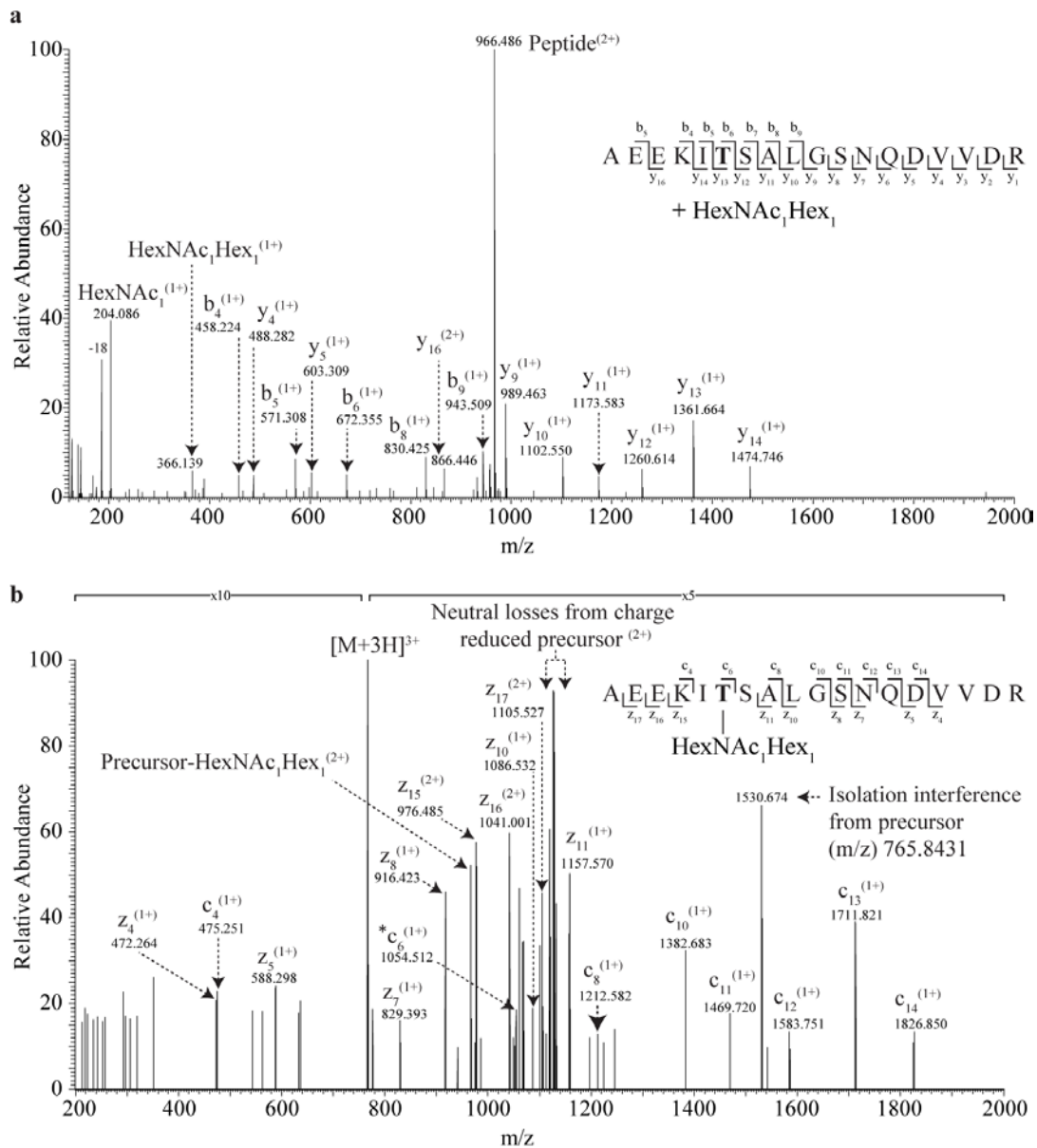
Figure 4-12. HCD and ETD fragmentation of an O-linked glycopeptide from HN of the V4-QLD isolate. Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and the likely site of O-linked glycosylation is denoted in bold. Not all ions have been labelled in the spectra for ease of interpretation. Fragmentation of the same precursor ion at *m/z* 766.372 (3+) (aa 66-83) using (a) HCD fragmentation and (b) ETD fragmentation.

4.5 DISCUSSION

The sequence of HN from the QLD/66 isolate contains six potential sites of N-linked glycosylation, five of which are relatively conserved across strains of NDV (N119, N341, N433, N481 and N538) (**Figure 4-1**). The sixth site in QLD/66 (N600) is found in the carboxy-terminal region of the protein, which is not well conserved and is proteolytically removed during maturation of this virus

(88, 285). Analysis of NDV V4-VAR, a variant of QLD/66, revealed that HN sites N341, N433 and N481 were conserved in this isolate and were used for the addition of N-linked glycans (**Figure 4-13**). Occupancy of these sites is consistent with X-ray crystallography studies of HN from other strains of NDV, where up two GlcNAc residues were identified at each site (91, 108). Other studies also indicated that sites N341, N433 and N481 are occupied involving mutation of individual N-X-S/T consensus sites, which increased the migration of NDV HN compared to wild type during gel electrophoresis (106, 107). These mutation studies also identified site N119 as occupied but analysis of HN from V4-VAR revealed the natural mutation Asn119Ser (**Figure 4-13**). As such, this site no longer includes the consensus motif N-X-S/T required for N-linked glycosylation and implies that it is not biologically essential.
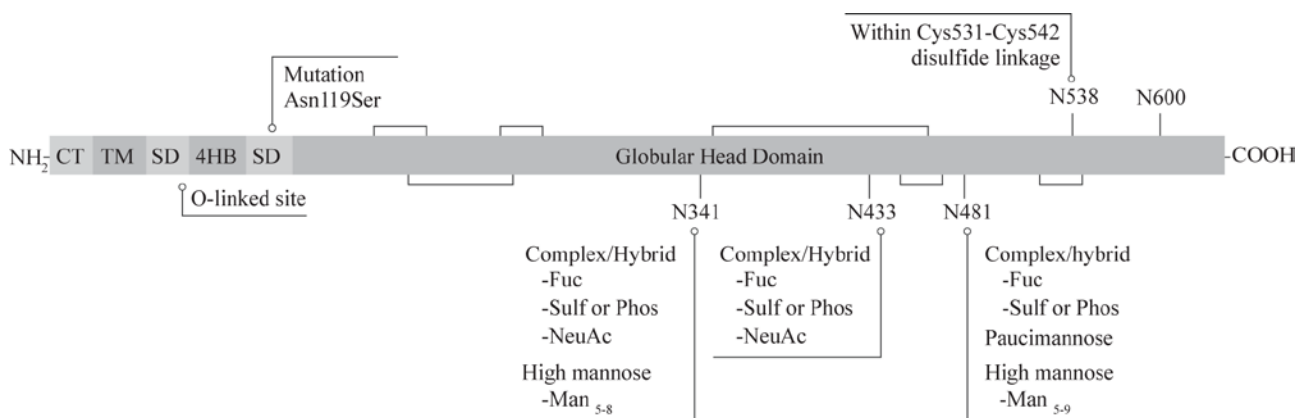
Figure 4-13. Glycosylation profile of V4-VAR HN. Schematic (not to scale) of NDV V4-VAR HN identifying the cytoplasmic tail (CT), transmembrane (TM), stalk domain (SD), four-helix bundle (4HB) and the globular head domain. Disulfide bonds are represented by connected lines and are derived from (101). N-linked consensus sites (N-X-S/T) are marked with a vertical line and the amino acid number of the observed glycosylated residue. Of the sites that were identified as occupied, N341, N433 and N481, the type of N-linked structures and additional monosaccharides contributing to the glycans have been listed. The potential N-linked site N119 in V4-VAR was mutated to serine thus precluding N-linked glycosylation at this site. Sites N538 was observed without glycosylation and occupancy at site N600 could not be determined. The region in the stalk domain where the O-linked site was observed has also been indicated.

Sites N538 and N600 were not identified as occupied in this analysis of HN from V4-VAR. It is noteworthy that complete tryptic digestion (i.e. no missed cleavages) of the HN protein would yield peptides bearing sites N538 and N600 that are small and hydrophilic. Such peptides, or potentially glycosylated versions of these peptides, may not have been retained on the reverse-phase chromatography material used in this study. However, a non-glycosylated peptide containing the potential N-linked site N538 was detected. Observation of this peptide confirmed that the N-linked

consensus site was conserved in V4-VAR HN. This is consistent with previous work using V4-VAR that revealed that N538, which is flanked by two Cys residues (sequence [531]CFKVVKT<u>NKT</u>YC[542]), is not glycosylated due to disulfide bond formation (101). Lack of glycosylation of N538 has also previously been demonstrated through mutation studies in other strains (106). In contrast to site N538, a tryptic peptide containing the potential N-linked site N600 was not observed during the analysis of V4-VAR. Electroeluted V4-VAR HN was also digested with endoproteinase Lys-C (from *Lysobacter enzymogenes*) and analysed by MS (data not included). Digestion with Lys-C could have produced the theoretical peptide [600]N<u>QT</u>EYRRELESYAASWP[616]. Using manual *de novo* and database searching with Mascot and Byonic, peptides containing N600 were still not detected. Unfortunately in the MS analysis of HN digested with trypsin then PNGase F, the *m/z* values (550-1500) were set to increase the likelihood of detecting the O-linked site. Thus the theoretical *m/z* values for the potential tryptic peptides containing site N600 were below the *m/z* scan range set. Site N600 resides within a 45 aa C-terminal extension which undergoes post-translational processing and is likely to be very low in comparative stoichiometric abundance relative to all other peptides (88, 286). Accordingly, substantial cleavage of this extension site may be the reason a peptide or glycopeptide bearing this site was not identified. As such, potential occupancy at site N600 was not established.

The N-linked glycan heterogeneity documented across V4-VAR HN was quite extensive with the identification of paucimannose and high mannose compositions and complex or hybrid structures that were variably fucosylated, sialylated and sulfated or phosphorylated. Paucimannose structures have not been widely reported on vertebrate proteins, however, over the last few years a number of studies have begun to highlight their presence in vertebrate tissue and cells (19-23), including proteins produced in embryonic eggs (287). The remaining glycan types (high mannose and complex or hybrid) and observed monosaccharides and substituents have been previously documented on proteins or glycans produced in embryonic eggs (47, 48, 288, 289). During viral replication viruses can utilise host-cell protein synthesis pathways and glycosylation machinery to produce the components required for progeny virions. Differential glycosylation of viral proteins can be seen across strains, cell culture and viral replication pathways (290, 291). As discussed in the introduction, the types of N-linked structures observed at each N-linked site and variations in monosaccharide compositions of the glycans reflect the degree of glycan processing during biosynthesis. *Paramyxoviridae* replication occurs in cytoplasm of the host cell where HN is transported to the infected cell surface after maturation in the Golgi complex (51). The present results indicate that such transport applies to V4-VAR HN as glycans containing multiple Fuc

residues, NeuAc and Sulf or Phos were detected on HN (1). The identification of high mannose structures ($Man_{5-9}GlcNAc_2$) at sites N481 and N341 indicate that these glycans escaped glycosyl trimming in the ER and Golgi to some extent. Thus, site N481 and to an extent N341 may be partially shielded from glycosylation machinery after protein folding occurs in the ER. Site N481 has previously been shown to contain mainly high mannose or hybrid structures through endoglycosidase H (Endo H) digestion of HN from transfected Cos-7 cells (106). Thus minimal processing of glycans attached at site N481 may be observed on HN from other strains and when using different protein production methods. Site N481 and to a lesser extent N341, are required for maturation and proper folding of HN (107). This suggests that these sites containing the high mannose glycans are buried within the structure during the folding process and remain there to an extent. In contrast, the higher level of fucosylation seen at site N433 (all glycans contained Fuc) and lack of high mannose glycans may indicate that it remains fully exposed to glycosylation machinery during glycoprotein synthesis. Finally, the diversity of glycans seen at each site may also imply different conformers of HN as suggested in 3D crystal structures (108, 292). Such differential glycosylation has been reported for NDV HN, where a form of HN containing intermolecular disulfide bonds was thought to contain high mannose, hybrid and complex glycan structures, while a form without intermolecular disulfide bonds contained primarily high mannose or hybrid structures (293).

The HN protein of NDV facilitates attachment to the host cells via interactions with sialylated cellular structures (69, 93). To undertake its function of receptor binding HN recognises and binds to gangliosides and sialylated N-linked glycans containing NeuAc in α2–3 and α2–6 linkages (93, 294). On the other hand, the protein also exhibits sialidase activity which allows the virus to cleave itself from already infected host cells and from budding progeny virions, thus preventing non-productive interactions with sialylated structures (62). The sialidase activity of HN is specific for NeuAc residues in a α2–3 linkage and to lesser extent residues in a α2–8 linkage (295), while leaving α2–6 linked NeuAc intact (296). Given the receptor binding and sialidase activity of HN it is interesting that sialylated glycans were observed in this study. Two N-linked sites (N341 and N433) contained a sialylated N-linked glycan ($Hex_3HexNAc_3dhex_3NeuAc_1$ + $Man_3GlcNAc_2$). Additionally, two O-linked glycans were observed that contained sialic acid ($Hex_1NAcHex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$). The presence of NeuAc on the surface of NDV virions could induce HN-mediated self-aggregation which would be detrimental for pathogenesis. O-linked glycans do not seem to be a cellular receptor for NDV (294), as such the sialylated O-linked glycan observed on HN may not induce such self-aggregation. However, sialylated N-linked glycans have been

shown to be a receptor for HN. Qualitatively, the number of N-linked glycans observed with sialylation was quite low (one glycan at two N-linked sites). The presence of low levels of NeuAc on NDV virions has been previously described (297) where it was suggested that some NeuAc content was associated with HN. Comparison with a neuraminidase-deficient mutant of NDV suggested NeuAc was also associated with the F protein in the mutant virions (297). These results indicate that HN may remove NeuAc from progeny virions. It has been shown that cells from embryonic chicken eggs are able to produce sialylated N-linked glycans with both α2–3 and α2–6 linked NeuAc (47). Given that host cell glycosylation machinery includes sialyltransferases, NDV virion glycoproteins may contain both α2–3 and α2–6 linked NeuAc on N-glycans. Thus, the low level of NeuAc on V4-VAR may be predominantly α2–6 linked as the sialidase activity of NDV HN is not specific for this linkage.

It has also been shown that cells from embryonic chicken eggs produce N-linked glycans that are sulfated (47). Interestingly, an analysis of released glycans from HN and F derived from Sendai virions propagated in eggs revealed acidic glycans that were not removed after sialidase digestion (298). Sulfated glycans have also been observed on the HA glycoprotein derived from influenza virions propagated in eggs (48, 288, 289). Given the observation of sulfation on other viral proteins produced in embryonic eggs and the presence of sulfated oxonium in the analysis of V4-VAR it is likely that glycans that were assigned ambiguously as "sulfated or phosphorylated" are in fact sulfated. Nevertheless, as sulfation was unable to be confirmed for all glycans caution was applied while representing the data.

The present study revealed that the stalk domain of HN from V4-VAR bore O-linked glycosylation (**Figure 4-11**). A preliminary analysis of NDV virions derived from another isolate of the strain QLD/66 indicated the O-linked glycans are attached at T71 (**Figure 4-12**). The peptide identified as O-linked ([66]AEEKITSALGSNQVVDR[83]) contains an internal Lys at position K69, which should have been cleaved by trypsin. Conversely, HCD MS/MS analysis of tryptic V4-VAR identified a peptide that was cleaved by trypsin ([70]ITSALGSNQVVDR[83]) with no evidence of glycosylation. This indicated that steric hindrance from an O-linked glycan possibly impeded tryptic cleavage at K69. It also revealed that the O-linked site is not always glycosylated.

Sequence alignments of HN reveal that T71 and the neighbouring S72 are highly conserved between NDV strains (299, 300). X-ray crystallography of the stalk domains of NDV (residues

D79-N115) and PIV5 HN has revealed a 4HB with an 11-residue (i, i + 3, i + 4, i + 4) repeat forming the hydrophobic core (96, 97). The identified O-linked site (T71) is positioned N-terminally to the 11-residue repeats of NDV HN. Sequence alignment of HN from NDV, PIV5 and hPIV (serotypes 1-4) based on the 11-residue repeat, reveals a conserved Thr in PIV5 (T62) and hPIV1 (T88) that align with the NDV O-linked site (T71) (96).

The stalk region of HN from the *Paramyxovirinae* subfamily is thought to play an important role in triggering the F protein to induce fusion (94, 95, 301). Introduced N-linked sites or site-directed mutagenesis of residues in the in the stalk domain of NDV HN resulted in significantly reduced or blocked fusion as judged by content-mixing assays (98-100). This inhibition of fusion also occurred when N-linked sites were introduced into the PIV5 stalk domain (97). Interestingly, introduction of an N-linked site at position K69 of NDV HN and S60 of PIV5 HN (two residues N-terminal to the conserved NDV T71 and PIV5 T62) blocked fusion while maintaining the structure and other functionalities of HN (97, 99). Dual mutation of sites N119 and N341, the former of which lies in the stalk domain of HN, have been shown to increase fusogenicity *in vitro* (106, 107). As noted by the authors of these studies, it is likely that the removal of these N-linked glycans increases oligomerisation of HN or allows greater non-covalent bonding of the HN stalk region with F, triggering F and increasing fusogenicity. Increased virulence was not translated *in vivo*, leading the authors to conclude that increased fusion resulted in an enhanced host immune response and faster clearance of the virus. Given the position of the O-linked site in the stalk domain of HN and that the introduction of N-linked sites into the stalk domain decreases fusion, while the removal of N-linked sites increases fusion, it could be postulated that glycosylation of T71 impacts on oligomerisation of HN or the interaction between HN and F and affects fusion kinetics.

Although T71 is highly conserved between NDV strains, genome sequencing of an avirulent Australian strain of NDV (I-2) revealed a sequence variation of T71A (283). Strain NDV I-2 is able to produce haemagglutinin-inhibition antibodies in chickens, elicit cytopathic changes in cultured chicken embryo kidney cells and spread and propagate in in the organs of chickens (302-304). Thus the removal of T71 is not deleterious for antibody production, HN expression and immunogenicity. Further experimental investigations are required to determine the biological significance of the O-linked site. The NDV I-2 strain provides a natural mutant of T71 and it would be of interest to analyse this strain to determine if, in the absence of T71, S72 is glycosylated. It would also be of interest to determine if the conserved T71 is glycosylated in more virulent strains of NDV and T62 or T88 of PIV5 and hPIV1, respectively.

The present report is the first to characterise site-specific glycan heterogeneity of HN from paramyxoviruses. To achieve this NDV HN derived from virions propagated in embryonic eggs was analysed by MS. Overall the characterisation of N-linked glycosylation from NDV HN agreed with previous findings with respect to site occupancy and observation of high mannose glycans attached at site N481. The complex nature of the glycans indicates Golgi processing in accordance with reports. In addition, O-linked glycoforms were identified on a previously undocumented O-linked glycopeptide from the stalk domain of the protein. These findings will provide the basis for further research to increase the understanding of the role glycosylation plays in the functionality of HN and will enable comparisons to be made of HN from other species of NDV and paramyxoviruses.

Chapter 5: Characterisation of glycosylation of Newcastle disease virus fusion (F) protein

5.1 SUMMARY

To enable a more complete insight into the glycosylation of NDV surface glycoproteins, the F protein of the avirulent V4-VAR strain was also analysed by MS. The F protein was derived from the same egg propagated virions used to produce HN in Chapter 4. In stark contrast to the high number of monosaccharide compositions observed at N-linked sites of HN, which included high mannose and hybrid or complex types, the N-linked sites from F contained predominately high mannose glycans. Of the six predicted N-linked sites of NDV F, the high mannose compositions were observed at sites N85, N191, N366 and N471. Four complex or hybrid compositions with variable fucosylation were also observed at site N191. The observation of fucosylated complex or hybrid compositions at N191 confirms predictions that some NDV F species are transported through the late-Golgi during processing. The work herein provides the first comprehensive site-specific description of N-linked glycans from NDV F and provides a comparison for future studies into site-specific glycosylation of F from virulent strains of NDV and from other paramyxoviruses.

5.2 INTRODUCTION

The F proteins of paramyxoviruses are essential for viral infectivity and mediating fusion between viral and host cell membranes, or infected and uninfected cell membranes. The fusogenic property of NDV F relies on cleavage of the F_0 precursor into two disulfide linked subunits (F_1 and F_2) (305) and, at minimum, expression of stalk domain of HN (70). As discussed in the Chapter 1, strains of NDV are categorised by their level of virulence (lentogenic, mesogenic and velogenic), which is defined by the degree of pathology observed during infection. Lentogenic strains cause mild respiratory infection while mesogenic strains produce respiratory and neurological disease with occasional mortality (74). Velogenic strains are further classified into viscerotropic or neurotropic, distinguished by fatal infection with intestinal haemorrhagic lesions and high mortality with neurological signs, respectively (74). The F protein plays an important role in NDV pathology as the level of virulence typically correlates with cleavability of F_0 into the F_1 and F_2 subunits,

whereby increased cleavage results in increased virulence. Cleavage is dictated by the amino acid sequence of the cleavage-activation site (87). Monobasic and polybasic cleavage sites render the F_0 protein susceptible to cleavage by host cell proteases, which can be limited to specific cells of the respiratory and digestive tract, or proteases distributed throughout several systems, respectively (77, 89, 306). Recently it was shown that clinically diseased wild bird populations encoding polybasic residues at the cleavage site were able to transmit ND to vaccinated domesticated chickens (76). However, monobasic and polybasic residues at the cleavage sites of NDV F are not the sole determinant of virulence. Expression of a 45 aa C-terminal extension on HN (90, 91) and regions of the stalk and head domain of HN have been implicated in lowered virulence (307). Furthermore, it has been shown that elimination of an N-linked consensus site in the F_2 subunit of NDV F significantly reduces cleavage of F_0 (102), while eliminating two sites in the F_1 subunit increases virulence (103). These mutation studies highlight the need for site-specific characterisation of NDV F glycosylation, particularly as it has also been revealed that specific glycosylation sites play a role in cell surface expression and fusion (102, 103).

The F_0 protein of NDV V4-VAR contains six putative sites of N-linked glycosylation (275), the first site, N85, resides on F_2 and the remaining five sites are found on F_1, N191, N366, N447, N471 and N541. These six N-linked sites on NDV F are highly conserved amongst strains (308). Site-specific glycosylation of NDV V4-VAR at site N85 has already been described, with the observation of high mannose compositions (105). Analyses of glycans released from the F protein of an avirulent strain of NDV produced in Madin-Darby bovine kidney (MDBK) cells revealed primarily high mannose structures (Man_{5-9}) (104). Interestingly, the work presented herein also identified predominantly high mannose glycans which were assigned to sites N85, N191, N366 and N471. Further processing of attached glycans was observed at site N191, with the identification of complex or hybrid glycans with variable fucosylation. Occupancy at site N447 could not be determined in this work and evidence of glycosylation was not observed at site N541, however, peptides containing N541 were observed without glycosylation indicating this site may not be glycosylated.

5.3 METHODS

5.3.1 Sample preparation

The viral preparation was provided by Professor Jeffrey Gorman from the Protein Discovery Centre at QIMR Berghofer and the terminology used for the preparation, V4-VAR, follows that which was described in Chapter 4. The isolate of NDV used was a variant of the Queensland (QLD)/66 strain of NDV (274) propagated and purified in embryonic chicken eggs as previously described (88, 277). From stocks of purified V4-VAR virions, approximately 72 µg, in 48 µg and 24 µg aliquots, were reduced and alkylated as per the methods described in Chapter 2 except that 50 mM of IAA was used. Targeted analyses were conducted to characterise glycopeptides from the F_1 or F_2 region of the protein and these are described in each section below.

5.3.2 SDS-PAGE separation of V4-VAR virions and electroelution of F_1 proteins

Approximately 48 µg of reduced and alkylated V4-VAR viral proteins were subjected to SDS-PAGE followed by staining and de-staining as per the methods described in Chapter 2. Bands of interest corresponding to F_1 were excised and stored overnight at 4°C until electroelution as per the methods described in Chapter 2.

5.3.3 Enzymatic digestions of V4-VAR F_1 proteins

Following electroelution, approximately half of the harvested V4-VAR F_1 proteins were methanol precipitated overnight without trypsin as per the methods described in Chapter 2. The dried pellet was stored at -80 °C then resuspended in 50 mM NH_4HCO_3 before digestion with 0.1 µg of trypsin at 37°C overnight. An aliquot (~1/4) of the V4-VAR F_1 tryptic digest was further digested with 1 µL of 1U/µL recombinant PNGase F (hereafter referred to as trypsin/PNGase F). An aliquot (~1/4) of the V4-VAR F_1 tryptic digest was also further digested with 0.1 µg of Glu-C (hereafter referred to as trypsin/Glu-C). Both additional digests were incubated at 37°C overnight. Resultant peptides from the trypsin/PNGase F digest were desalted with a C18 ZipTip (10 µL pipette tip with a 0.6 µL resin bed; Millipore, MA, USA) using the manufacturers' guidelines for MS analysis.

5.3.4 SDS-PAGE separation of V4-VAR virions and in-gel digestion of F_1 and F_2 proteins

Approximately 24 μg of reduced and alkylated V4-VAR viral proteins were methanol precipitated overnight without trypsin. These proteins were subjected to SDS-PAGE separation as per the methods described in Chapter 2 except that a gradient gel was used (4-20% precast polyacrylamide gel, Mini-PROTEAN® TGX™, BioRad Laboratories). Bands of interest corresponding to F_1 and F_2 were excised and subjected to in-gel digestion with trypsin as per the methods in Chapter 2.

5.3.5 Nano-ultra-high pressure liquid chromatography

Approximately 1/8 of the original electroeluted sample and half the in-gel digested samples were injected for each analysis using a nUHPLC system as described in Chapter 2. Samples were loaded onto the trap column and washed for 5 min at 15 μL/min in 98% solvent A and 2% solvent B. Peptides and glycopeptides were subsequently eluted onto the analytical column at flow rate of 0.3 μL/min whilst ramping through a sequence of linear gradients from 2% to 40% solvent B in 60 min, to 70% B over 15 min, to 95% B in 5 min then holding at 95% B for 5 min. The column was then re-equilibrated with 2% B for 20 min.

5.3.6 Mass spectrometry data acquisition

Survey scans of peptide and glycopeptide precursors from *m/z* 300 to 2000 were acquired in the Orbitrap at a resolution of 120K (full width at half-maximum, FWHM) at *m/z* 200 using an automatic gain control (AGC) target of 400,000 and maximum injection time of 50 ms. For internal mass calibration the lock mass option was enabled using the polycyclodimethylsiloxane ion at *m/z* 445.1200 (310). The most intense precursors within *m/z* 600-1800 (trypsin and trypsin/Glu-C digests of F_1 and in-gel digestions of F_1 and F_2) or *m/z* 300 to 1800 (trypsin/PNGase F digest of F_1) were selected for fragmentation by HCD. Precursors with intensities over 5,000 counts (charges 2-6) were isolated with a mass selecting quadrupole using an isolation window of *m/z* 2. Precursors were fragmented in the ion routing multipole, using a stepped method of ±5% around a normalised collision energy (NCE) of 25%. Previously selected ions within a ±10 ppm window were dynamically excluded for 15 s. Fragment ions were acquired in the Orbitrap at a resolution of 30K

using an AGC target of 50,000 and maximum injection time of 100 ms. If fragment ions were produced within the top 20 ions corresponding to 204.0867 (HexNAc), 163.0601 (Hex) or 292.1027 (NeuAc) within a ±10 ppm window, the precursor ions were re-isolated and subjected to ETciD with 25% supplemental activation. Fragment ions produced by ETciD were acquired in the Orbitrap at a resolution of 60K using an AGC target of 100,000 and maximum injection time of 250 ms with 2 microscans.

In total five chromatographic runs (one run for each sample) were performed using a HCD-pd-ETciD method. These were conducted with the electroeluted F_1 sample that was digested with trypsin, trypsin/Glu-C and trypsin/PNGase F and the in-gel tryptic digests of F_1 and F_2.

5.3.7 Data processing of non-glycosylated and deglycosylated peptides from V4-VAR F

Proteome Discoverer (v2.1.0.81) and the search engine Mascot were used to search HCD MS/MS spectra from all RAW files. The protein database contained the sequence for V4-VAR F (275). Cleavage specificity was set as semi-tryptic for the samples digested with trypsin. For the trypsin/Glu-C sample cleavage specificity was set as both trypsin and Glu-C with the latter allowing cleavage at Glu and Asp residues. A maximum of two missed cleavages were allowed for each search. Mass tolerances of 10 ppm and 0.02 Da were applied to precursor and fragment ions, respectively. Carbamidomethylation of Cys was set as a fixed modification and dynamic modifications included mono-oxidised Met, deamidation of Asn and Gln residues and conversion of N-terminal Gln to pyroglutamate. The "Fixed Value PSM Validator" node was used and a cut-off score of 30 was applied to all PSMs.

In addition to the Mascot searches, a mzML file of HCD spectra from the trypsin/PNGase F sample was also analysed with Byonic. This was completed to investigate possible amino acid modifications not included in the Mascot search. A wild card search was conducted allowing -40 or +100 mass tolerance (Da) on all amino acid residues. Cleavage specificity was set C-terminal to Lys and Arg residues allowing ragged cleavage at the C-terminus of peptides and a maximum of two missed cleavages. Mass tolerances of 10 ppm and 0.02 Da were applied to precursor and fragment ions, respectively. Carbamidomethylation of Cys was set as a fixed modification and two dynamic modifications were allowed per peptide, including mono-oxidised Met, deamidation of

Asn and Gln residues and conversion of N-terminal Gln to pyroglutamate. The automatic peptide score cut-off was disabled for the wild-card search.

5.3.8 Oxonium ion profile and preliminary investigations of glycopeptides from V4-VAR F

Spectra from the analyses of tryptic peptides from electroeluted NDV F_1 and in-gel digested F_2 were analysed with OxoExtract. The following parameters were used: digestion with trypsin; maximum of two missed cleavages; a fixed modification of carbamidomethylation of Cys and a dynamic modification of mono-oxidised Met. The protein database queried contained the sequence of V4-VAR F_0 (275). The optional feature of GlycoMod was enabled. In addition to the default monosaccharide parameters the substituents Phos and Sulf were considered as possible components of the N-glycans. The observed oxonium ion profile and monosaccharide compositions of the glycans were typical high mannose and hybrid or complex structures. As such, a NDV V4-VAR F custom glycan database was not required.

5.3.9 Assignment of glycopeptides from V4-VAR F

For searches in Byonic HCD and ETciD spectra from RAW files were converted separately into mzML files. Cleavage specificity was set C-terminal to Lys and Arg residues for the trypsin digested samples with additional cleavage at Glu and Asp for the trypsin/Glu-C sample. A maximum of two missed cleavages were allowed. Mass tolerances of 10 ppm and 0.02 Da were applied to precursor and fragment ions, respectively. A fixed modification Carbamidomethylation of Cys was allowed. Two common dynamic modifications were allowed per peptide from mono-oxidised Met, deamidation of Asn and Gln residues and conversion of N-terminal Gln to pyroglutamate. A maximum of two glycans were permitted to be attached to one peptide at N-linked consensus sites. The protein database queried contained the sequence for V4-VAR F_0 (275). The N-linked glycan database queried was the Byonic mammalian database (309_Mammalian no sodium) with all glycans containing NeuGc removed. The precursor off-set option was enabled (narrow), which allowed Byonic to consider that the true monoisotopic mass may not have been assigned to a precursor (i.e. it could consider precursors with a difference of 1 Da from the observed mass of a precursor). The default protein FDR and peptide output options were changed to "Show all N-glycopeptides" which is recommended by the manufacturer when analysing simple samples.

Once a list of glycopeptides had been compiled for site N191 using the results from the electroeluted F_1 sample digested with trypsin, the Xtract feature within Xcalibur Qual Browser was used to deconvolute MS precursor spectra of the eluting glycopeptides. Spectra were summed across the elution period (~1 min) for all glycoforms and deconvoluted from multiply charged signals to protonated monoisotopic masses. Precursors were included if they fell within the m/z range of 800 to 1600, had a minimum S/N threshold of two and a maximum charge state of five.

5.4   RESULTS

5.4.1 Isolation of V4-VAR F_1 and F_2

The F_1 subunit of V4-VAR F_0 was observed to migrate just above the 50 kDa MW marker after SDS-PAGE separation using a 10% resolving gel (**Figure 5-1**) and 4-20% gradient gel (**Figure 5-2**). The calculated mass of V4-VAR F_0 is 58,844 Da, while the masses of F_1 and F_2 are 46,691 and 8,910 Da with the signal peptide removed from F_2 (275). Slices corresponding to the region of interest for F_1 were excised and pooled before the intact proteins or peptides were extracted using electroelution (**Figure 5-1**) or in-gel digestion (**Figure 5-2**), respectively. The F_2 subunit was observed to migrate between the 10 and 15 kDa MW markers (**Figure 5-2**). Slices corresponding to region of interest for F_2 were excised and subject to in-gel digestion.

5.4.2 Identification of non-glycosylated and deglycosylated peptides from V4-VAR F

Mascot-assigned PSMs from the analyses of electroeluted V4-VAR F_1 digested with trypsin, trypsin/PNGase F and trypsin/Glu-C and the tryptic in-gel digests of F_1 and F_2 have been included in Supplementary Tables S5-1 to S5-5, respectively. Sequence coverage of F_1 was comparable between the electroeluted sample and the in-gel digested sample. The trypsin/Glu-C digest did not enable any further sequence coverage of F_1 to be obtained. The overall sequence coverage from the Mascot searches has been presented in **Figure 5-3a** and the positions of the potential six N-linked sites on V4-VAR F_0 are illustrated in **Figure 5-3b**.
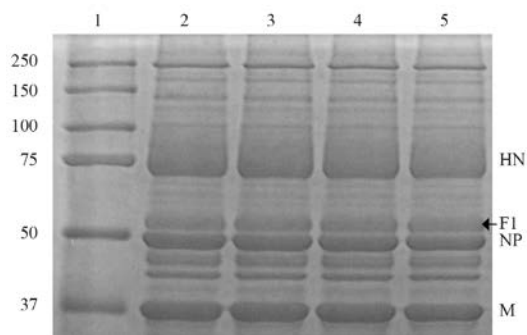
Figure 5-1. SDS-PAGE separation of V4-VAR proteins using a 4% stacking and 10% resolving gel. Lane 1 contains the MW markers with the protein masses shown in kDa. Lanes 2 to 5 were each loaded with ~12 µg of reduced and alkylated V4-VAR virion proteins. Bands previously identified as NDV HN, F_1, (F1) nucleocapsid (NP) and matrix (M) proteins (88) are indicated. Bands from each lane containing F_1 (denoted with an arrow) were excised from the gel and subject to electroelution.
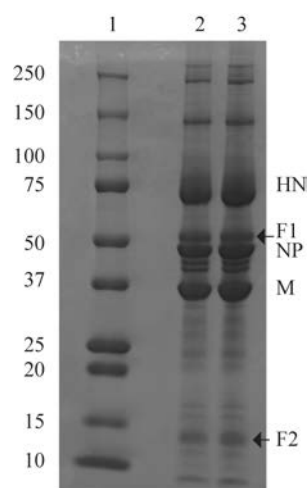
Figure 5-2. SDS-PAGE separation of V4-VAR proteins using a 4-20% gradient gel. Lane 1 contains the MW markers with the protein masses shown in kDa. Lanes 2 and 3 were each loaded with ~12 µg of reduced and alkylated V4-VAR virion proteins. Bands previously identified as NDV HN, F_1, nucleocapsid (NP) and matrix (M) and F_2 proteins (88) are indicated. Bands from each lane containing F_1 and F_2 (denoted with arrows) were excised from the gel and subject to in-gel digestion.

The Mascot search results of electroeluted F_1 digested with trypsin and trypsin/Glu-C and the in-gel tryptic digest of F_1 revealed PSMs containing the N-linked site N541. In these searches PSMs corresponding to peptides containing the remaining four sites in F_1 were not identified. Several of the PSMs containing N541 from the trypsin and trypsin/Glu-C digests of F_1 identified the potential site of N-glycan attachment as being deamidated. The assigned spectra were manually investigated to confirm deamidation of N541 and rule out modification of other residues in the peptide sequence ^{535}TLLWLG<u>NNT</u>LDQMR^{548}.

The Mascot search of the trypsin/PNGase F digest of F_1 identified PSMs containing N191 and N366 (**Figure 5-3a**) where Asn residues within the N-linked consensus sites were deamidated. As observed in the trypsin and trypsin/Glu-C digests of F_1 the Mascot search of the trypsin/PNGase F digest also revealed PSMs where N541 was deamidated. Peptides containing sites N447 and N471, which reside on one theoretical peptide (N47-K480), were not observed in the search of the

trypsin/PNGase F digest of F_1. The Byonic wild card search of the trypsin/PNGase F sample did not enable additional peptide sequence coverage to be obtained.

The Mascot search of the in-gel tryptic digest of F_2 revealed PSMs from the F_2 subunit but did not identify peptides containing N85 (**Figure 5-3a**). Interestingly, the search results also revealed a PSM with the amino acid sequence ^{102}IQESVTTSGGGKQG^{115} that contained a semi-tryptic cleavage event at the C-terminus of the peptide (**Figure 5-3a**). This peptide is positioned N-terminal to the cleavage site that separates the F_2 and F_1 subunits. The observation of this peptide conforms with previous observations where the C-terminus of F_2 from the parental QLD/66 strain was identified as "GKQG" due to carboxypeptidase B-like trimming after cleavage of F_0 (88).

5.4.3 Identification of glycopeptides from V4-VAR F

The results of the Byonic HCD searches of the trypsin and trypsin/Glu-C digestions of F_1 and in-gel digestions of F_1 and F_2 are presented in Supplementary Table S5-6 (all results for F_1) and Supplementary Table S5-7 (results for F_2) with annotated spectra in Supplementary Figures S5-1 to S5-4), respectively. The Byonic searches of ETciD spectra did not yield any results and have not been included in the supplementary data.

The monosaccharide compositions of the glycans observed at each N-linked site of V4-VAR have been represented in (**Figure 5-3b**). Eight glycans were observed at N191 from the F_1 subunit, four high mannose structures (Man$_{5-8}$) and four hybrid or complex glycans including two fucosylated glycans. A representative spectrum of HCD fragmentation of a glycopeptide containing N191 is presented in **Figure 5-4**. The oxonium ion at m/z 512.197, corresponding to HexNAc$_1$Hex$_1$dHex$_1$, confirms the presence of Fuc in the glycan. The relative abundances of the different glycoforms observed at N191 are presented in **Figure 5-5** and reveals that the fucosylated species are of lower relative abundance compared to the high mannose species. It should be noted that the signal intensities of glycopeptides can vary due to signal suppression from co-eluting non-glycosylated peptides and differences in the monosaccharide compositions (232). As such, the summed and deconvoluted spectrum in **Figure 5-5** provides a general indication of the relative abundance of the glycoforms. Three glycans could be assigned to N366 (Man$_{6-8}$) and one glycan could be assigned to N471 (Man$_9$) (**Figure 5-6** and **Figure 5-7**, respectively). Glycopeptides containing N447 and N541

were not detected. In the F_2 subunit four glycans were assigned to N85 (Man_{5-8}) and representative spectrum is presented in **Figure 5-8**.
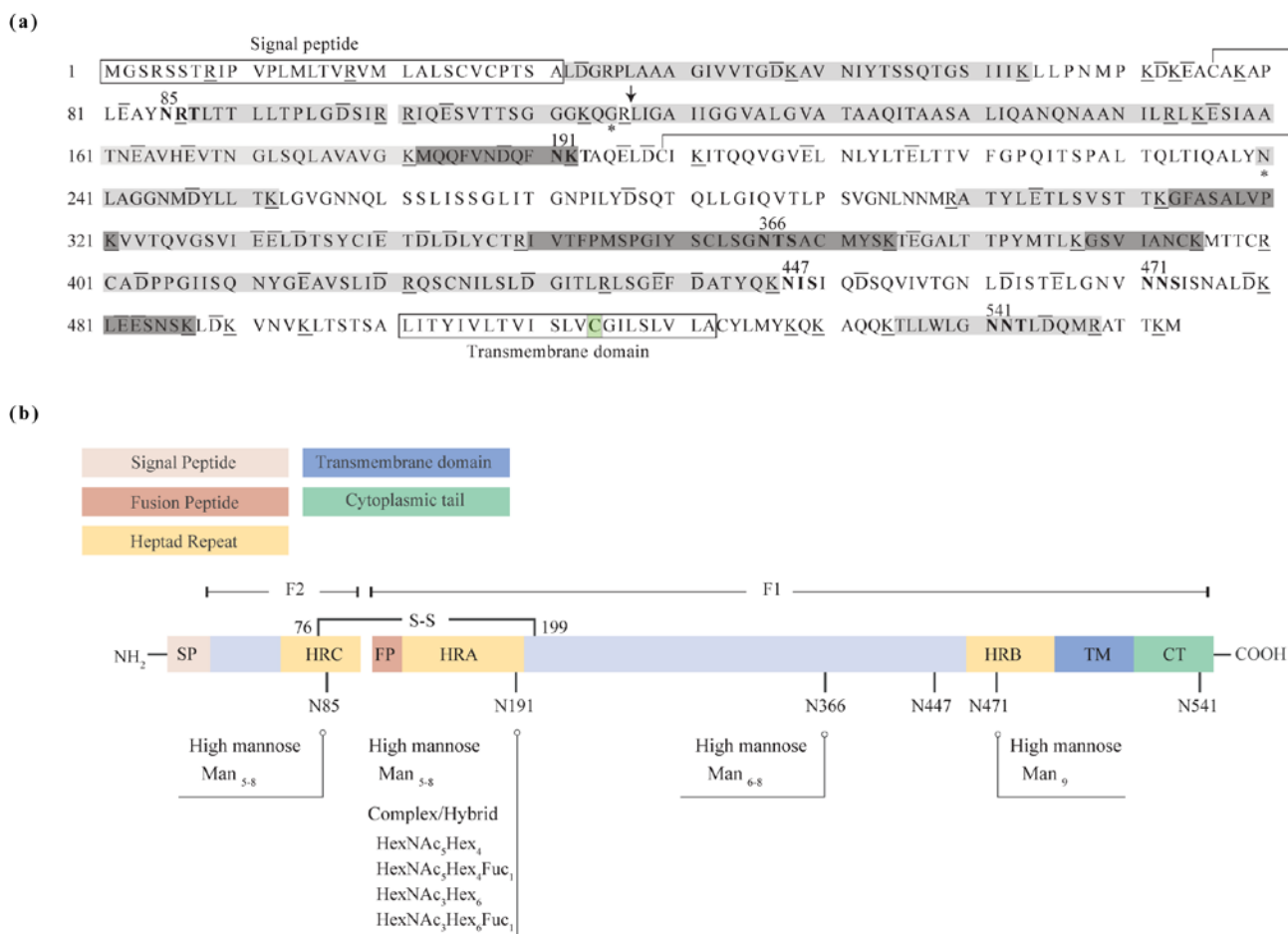
Figure 5-3. Protein sequence coverage and glycans of NDV V4-VAR F_0 based on peptides and glycopeptides identified from HCD MS/MS analyses. (a) Amino acid sequence of NDV V4-VAR F_0 (275), where the F_2/F_1 cleavage site is indicated by an arrow and disulfide bonds are represented by connected lines. The sequences of the signal peptide and transmembrane domain are shown in clear boxes with a predicted site of Cys palmitoylation highlighted in green in the transmembrane domain. The six N-linked consensus sites are in bold font with the amino acid number of the predicted site of glycosylation above each site. Potential cleavage sites for trypsin and Glu-C are indicated by a line above and below the relevant residues, respectively. Sequence coverage derived from the Mascot searches of F_1 and F_2 digested with trypsin are highlighted in light grey. Additional sequence coverage obtained from the trypsin/PNGase F digest of F_1 is highlighted in dark grey. An asterix "*" below a residue indicates semi-tryptic cleavage. (b) A schematic (not to scale) of NDV V4-VAR F_0 presents the signal peptide (SP), heptad repeats (HR A-C), fusion peptide (FP), transmembrane domain (TM) and cytoplasmic tail (CT) (275, 276). Cleavage of F_0 at the amino-terminal end of FP produces F_2 and F_1 chains which are linked by a disulfide bond, represented by a connected line with the amino acid number of the Cys residues [assigned from (275, 276)]. N-linked consensus sites (N-X-S/T) are marked with vertical lines and the amino acid number of the respective Asn residue. If glycans were assigned to an N-linked site, the monosaccharide compositions of the attached glycans have been indicated. High mannose glycoforms have been labelled Man_{5-8} while the remaining glycans have been labelled with the number of each monosaccharide residue.
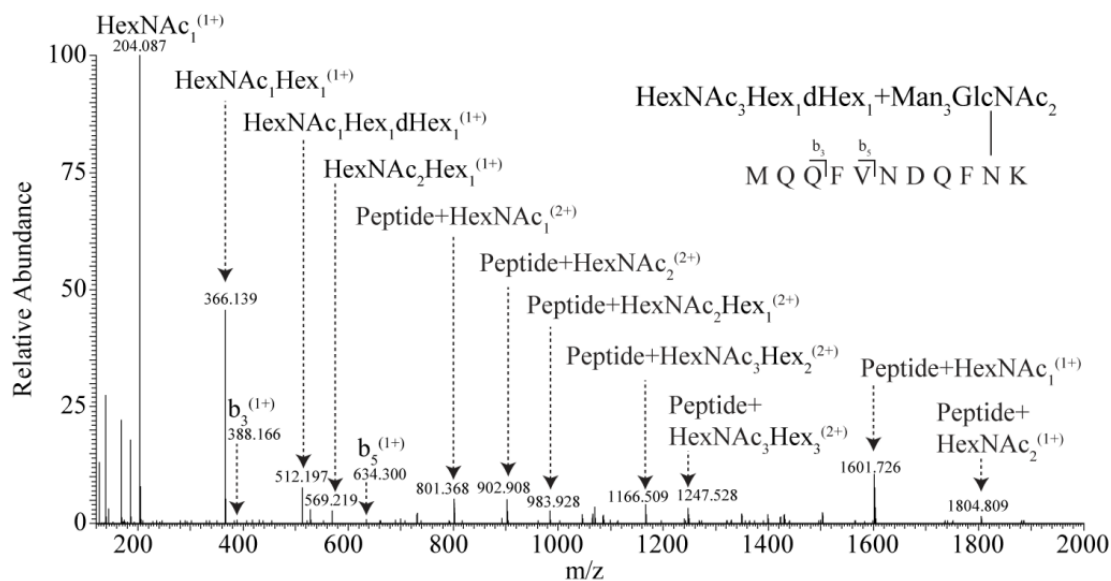
Figure 5-4. HCD spectrum of an N-linked glycopeptide from V4-VAR F₁ containing site N191. HCD fragmentation the precursor ion at *m/z* 1070.110 (3+). The spectrum was obtained from the analysis of electroeluted F₁ digested with trypsin. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (aa 182-192).
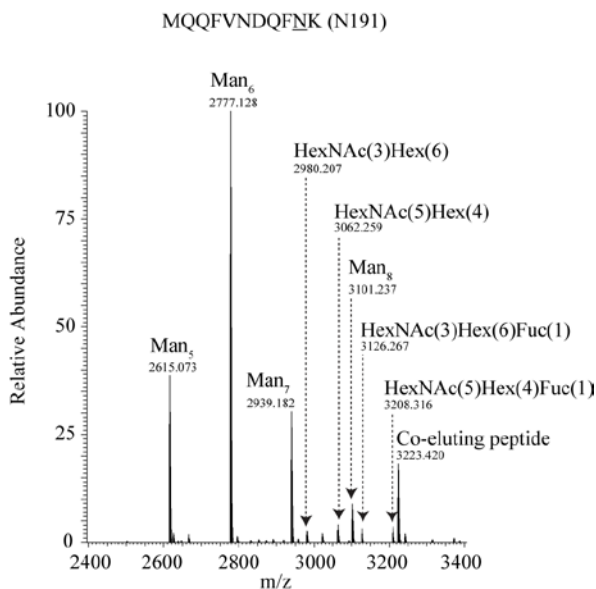
Figure 5-5. Glycoforms of an N-linked glycopeptide from V4-VAR F₁ containing site N191. Spectra for the glycopeptide containing N191 (aa 182-192) was obtained from the analysis of electroeluted F₁ digested with trypsin. Precursor MS spectra were summed and deconvoluted into singly protonated precursor masses. High mannose glycoforms have been labelled Man_{5-8} while the remaining glycans have been labelled with the number of each monosaccharide residue.
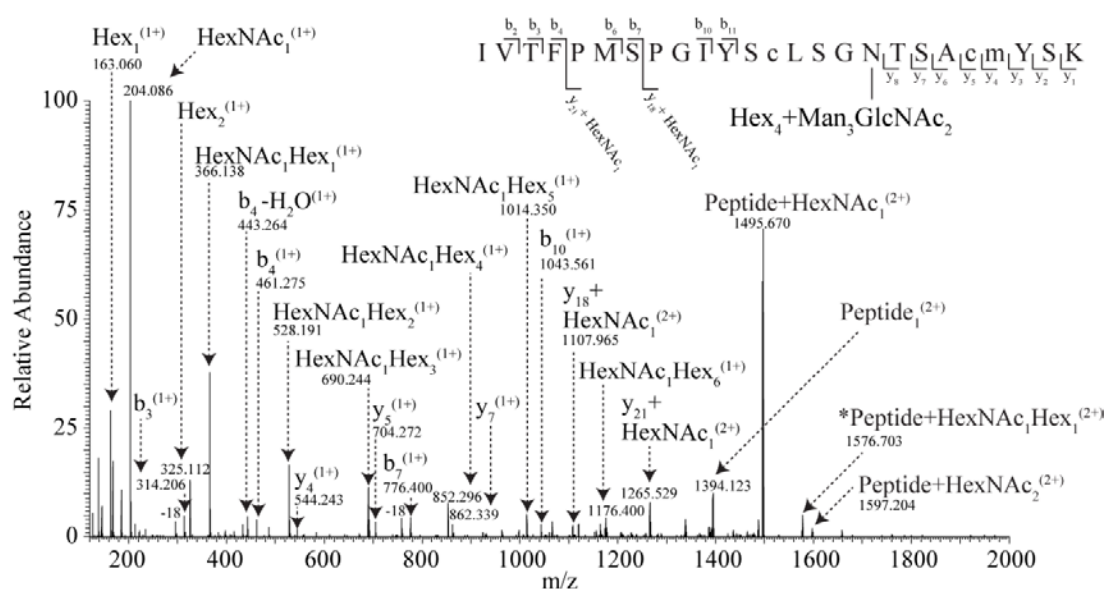
Figure 5-6. HCD spectrum of an N-linked glycopeptide from V4-VAR F$_1$ containing site N366. HCD fragmentation the precursor ion at *m/z* 1443.267 (3+) is presented. The spectrum was obtained after in-gel digestion of F$_1$ with trypsin. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (aa 350-374). Lowercase "m" and lowercase "c" in the peptide sequence represent oxidised Met and carbamidomethylation of Cys, respectively. Not all peptide ions from the schematic have labelled in the spectrum for ease of interpretation.
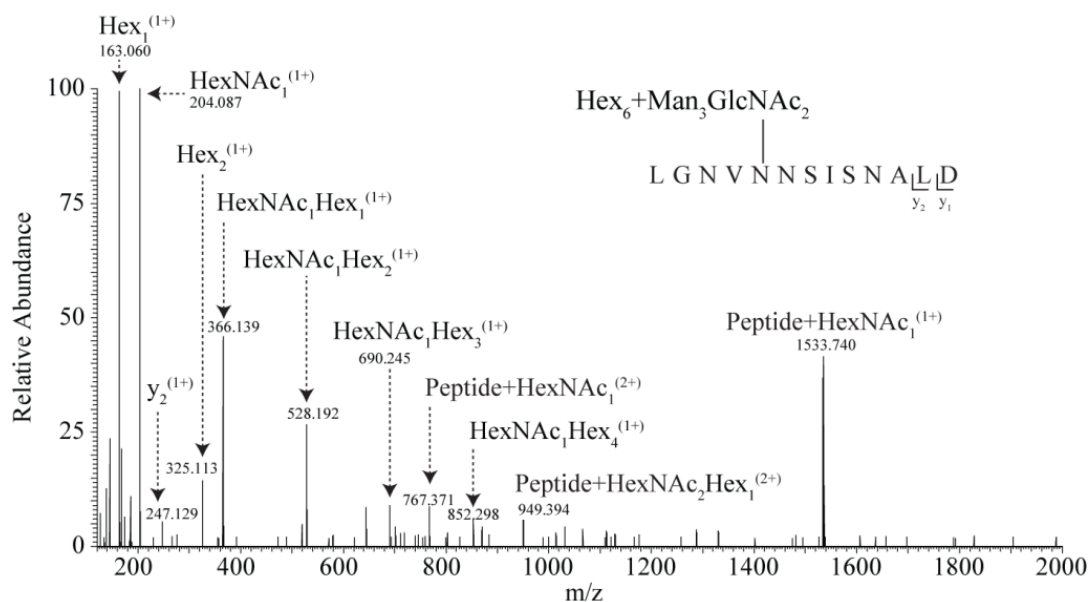
Figure 5-7. HCD spectrum of an N-linked glycopeptide from V4-VAR F$_1$ containing site N471. HCD fragmentation the precursor ion at *m/z* 1065.770 (3+) is presented. The spectrum was obtained from the analysis of electroeluted F$_1$ digested with trypsin then Glu-C. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (aa 467-479). Not all peptide ions from the schematic have labelled in the spectrum for ease of interpretation.
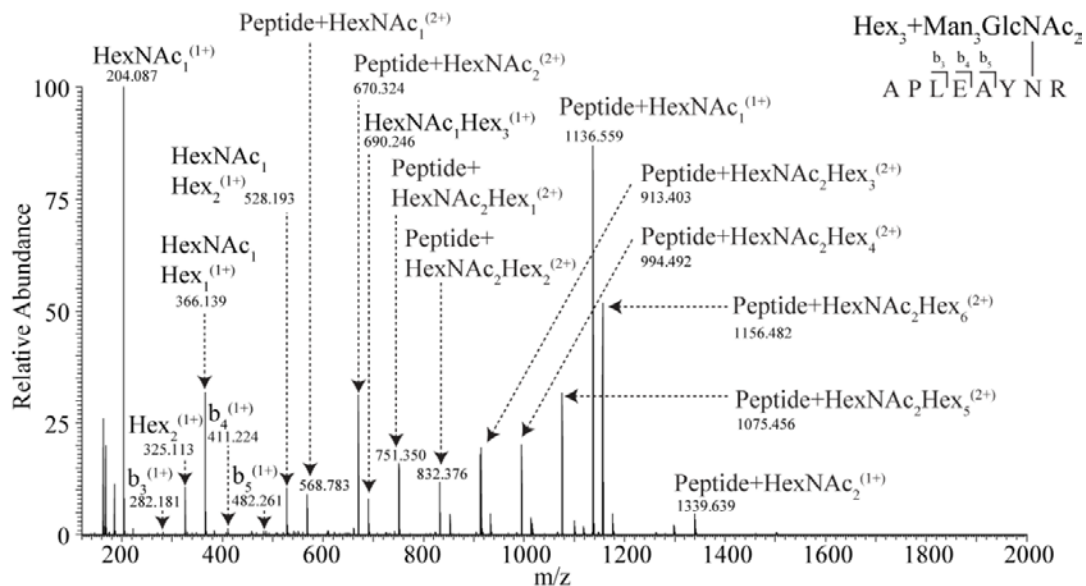
Figure 5-8. HCD spectrum of an N-linked glycopeptide from V4-VAR F$_2$ containing site N85. HCD fragmentation the precursor ion at *m/z* 1156.482 (3+) is presented. The spectrum was obtained after in-gel digestion of F$_2$ with trypsin. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (aa 79-86).

5.5   DISCUSSION

To assess the types of glycans present on F from NDV, virion proteins from the V4-VAR isolate were separated and by SDS-PAGE and F was digested before analysis by MS. Two fragmentation methods, HCD and ETciD, were employed in a product-dependent manner. However, only HCD fragmentation enabled identification of glycopeptides from F of NDV. It has been demonstrated that ETD is not effective for the fragmentation of peptides with precursor charge states less than two or where precursor *m/z* values are greater than 850 (284). The precursor *m/z* values for all glycopeptides identified from V4-VAR F exceeded 850 and those identified from N85 were predominately doubly protonated, thus they were not amenable to fragmentation by ETD. Supplemental collisional activation of non-dissociated peptide precursors has been shown to induce fragmentation of $[M+2H]^{2+}$ ions (225), however ETD fragmentation with 25% supplemental activation did not induce dissociation of glycopeptide ions from NDV F. Using HCD it was revealed that four sites were occupied N85, N191, N366 and N471 with high mannose glycans (**Figure 5-3b**). Hybrid or complex glycans were also observed at N191 with variable fucosylation. This study also demonstrated that site N541 is likely not glycosylated while occupancy at site N447 could not be determined.

Early investigations into the synthesis and processing of NDV F predicted that the protein is modified and transported through the ER-Golgi pathway (89, 311). It was later shown that NDV F and HN likely interact in the ER (312) prior to cleavage of F_0 in the *trans*-Golgi (313). Transport of viral glycoproteins through the ER-Golgi pathway can be monitored by Endo-H digestion (106, 135, 314), an enzyme that removes high mannose and hybrid glycans (315). As the N-linked glycans on the protein mature in the medial- and *trans*-Golgi, they become resistant to Endo-H digestion, which can be monitored by gel electrophoresis. However, as previous studies (104, 105), and the work herein reveal, the glycans present on NDV F are mainly high mannose. As such, monitoring transport of NDV F through the ER-Golgi pathway using Endo H has not proved useful as the protein does not become resistant to the enzyme (313). Studies using ^{3}H labelled Fuc predicted that F_1 from avirulent and virulent strains of NDV produced in MDBK and chicken embryo cells, respectively, contained low levels of Fuc (313, 316). These predictions were based on the expected electrophoretic mobility of F_1 during gel electrophoresis of virion proteins, thus the incorporation of Fuc could not be unequivocally assigned to F. The results presented herein reinforce those above, revealing that some species of the F protein of NDV V4-VAR are transported through the *trans*-Golgi through the presence of fucosylated hybrid or complex structures at site N191.

Interestingly, analyses of glycans released from virion derived F proteins of other paramyxoviruses revealed vastly different glycan profiles to NDV F. Released glycans of F derived from Sendai virus (SeV) propagated in embryonic chicken eggs revealed that 75% of the structures contained charged residues (298). Glycans of F from Simian virus 5 (SV5 now classified as PIV 5) produced in MDBK cells were observed to be complex type glycans (309). It is predicted that the F protein of hRSV produced in Human epithelial type 2 (HEp-2) cells contains predominately complex type glycans (135, 314). It had been postulated that the high mannose glycans observed from NDV F produced in MDBK cells was cell-line specific (104). However, the results herein confirm that NDV F produced in embryonic eggs also presents with mainly high mannose glycans.

As discussed in Chapter 1, the F proteins of paramyxoviruses are all trimeric type I integral membrane proteins, thus the N- and C- terminal domains are conserved in addition to sequences of functional importance such as the fusion peptide and F_2/F_1 cleavage sites. In addition to these features, comparisons of the amino acid sequences of F proteins revealed conservation of Cys residues and heptad repeats A and B (305, 317). Furthermore, the overall architecture of F is conserved between paramyxoviruses (63). Interestingly, there is general conservation of N-linked

sites within a single genus, but not always between genera of paramyxoviruses (317). The functional roles of N-linked sites on F proteins from paramyxoviruses have been investigated through site-directed mutation studies of F from NDV (102, 103), PIV5 (318), SeV (319), MeV (320, 321), Canine distemper virus (CDV) (321), NiV (322), HeV (323) and hRSV (139) (**Figure 5-9**). Most N-linked consensus sites were predicted to be occupied with the exception of a site in the cytoplasmic domain NDV, one from the F_1 subunit of CDV and one from the F_2 subunit of NiV and HeV (**Figure 5-9**).
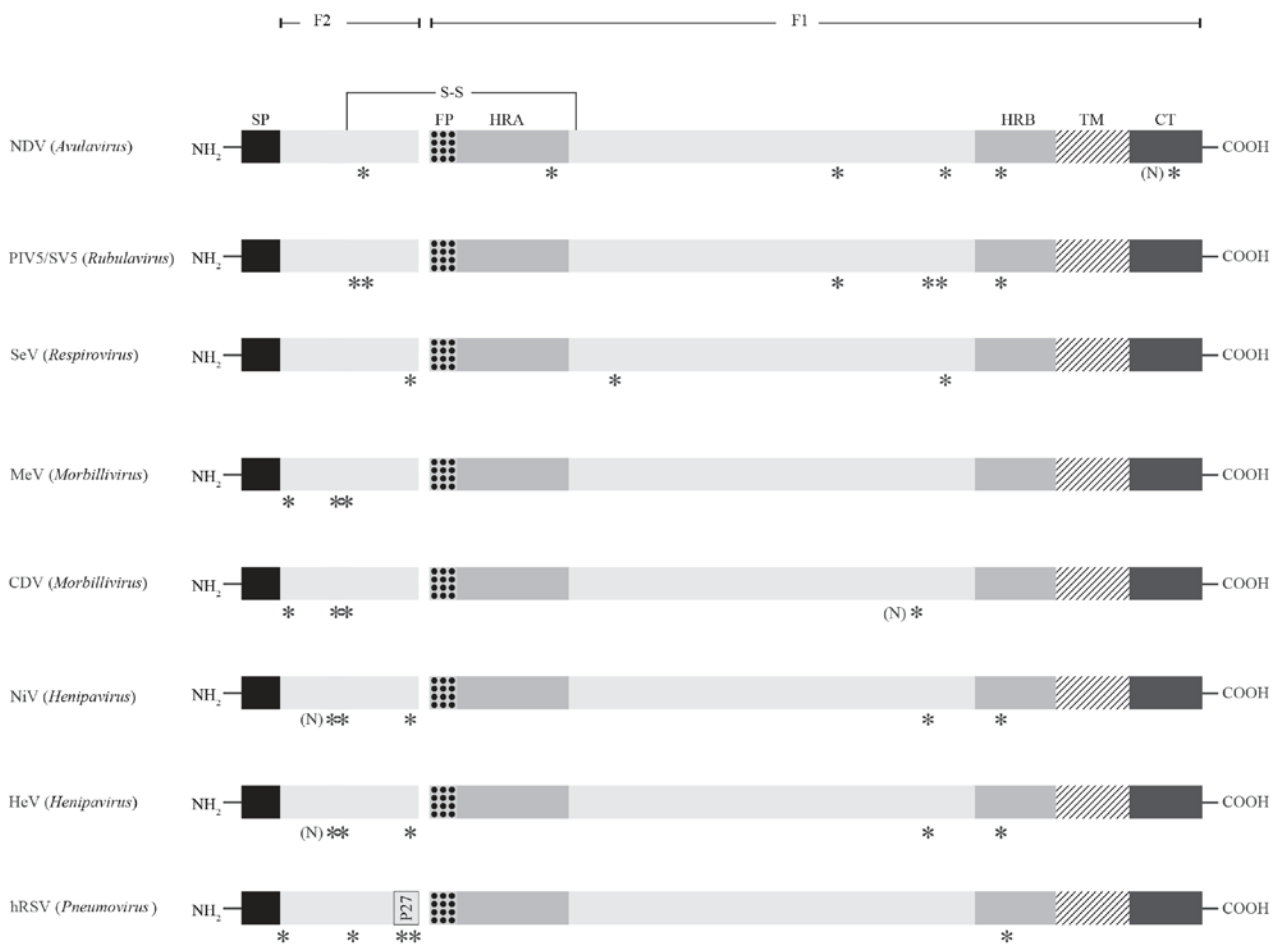
Figure 5-9. Position of N-linked sites from F proteins of paramyxoviruses. Schematic representations (not to scale) of F proteins from different members of the paramyxovirus family. The genus to which each virus belongs has been included in brackets. General conserved regions are represented on the top schematic of NDV, identifying the signal peptide (SP), fusion peptide (FP), heptad repeats (HRA and HRB), transmembrane domain (TM) and cytoplasmic domain. The soluble peptide (P27) that is lost from hRSV F_0 has been represented on the schematic of hRSV (bottom). The position of N-linked sites on each protein are denoted by an "*" and are derived from the studies that have been referenced in the text. If an N-linked site was predicted to be unoccupied an "N" has been placed next to it.

These functional studies revealed the importance of specific sites for some viruses but most highlighted the cumulative effects of eliminating N-linked sites in the F proteins. Furthermore, there were no clear conserved roles for N-linked sites between viruses. For example, elimination of the N-linked consensus site in HRB of NDV and hRSV did not significantly affect cleavage, transport or expression of F (103, 139). Contrastingly, mutation of the N-linked site in HRB of SV5 significantly reduced cleavage and transport of F (318) and moderately reduced expression of NiV and HeV F (322, 323). For HeV and NiV of the *Henipavirus* genus, the N-linked consensus site N-terminal to HRB is required for folding and processing of F (322, 323). Conversely, elimination of sites in similar regions of NDV and SeV did not affect folding and processing of the respective proteins (102, 103, 319). Although the number, location and function of N-linked sites are seemingly not conserved between paramyxoviruses, Messling and Cattaneo predicted roles for N-linked sites based on their position in regions of the tertiary structure of F (321). Using the crystal structure of F from NDV V4-VAR (276) (**Figure 5-10**) and predicted structures of CDV, SV5 and SeV F, the authors assigned folding and transport roles to N-linked sites in the stalk and neck regions, while those sites in the head region were expected to control fusion.

More recently, the crystal structures of F from a virulent strain of NDV (Australia-Victoria/Aus Vic) (324), hPIV3 (325) and hRSV (326) have been solved in a similar conformation to NDV V4-VAR (**Figure 5-10**). As yet, the roles of N-linked sites from F of hPIV3 have not been elucidated. Using the stalk, neck and head theory, it could be predicted that N359 is involved in fusion control, while either N238 or N446 may be involved in folding and transport (**Figure 5-10**). Comparisons of the structures in **Figure 5-10** also reveal that N27 at the N-terminus of hRSV F_2 is positioned in the head region of the F protein. This aligns with the previous predictions by Messling and Cattaneo that N-linked sites from CDV and MeV, which are also positioned at the N-terminus of F_2 (**Figure 5-9**), are situated in the head region of the respective F proteins. However, a functional study of the F protein from hRSV revealed elimination of N500 significantly altered fusion activity while elimination of N27 had little impact on fusion (139), which is at odds with the stalk, neck and head theory.
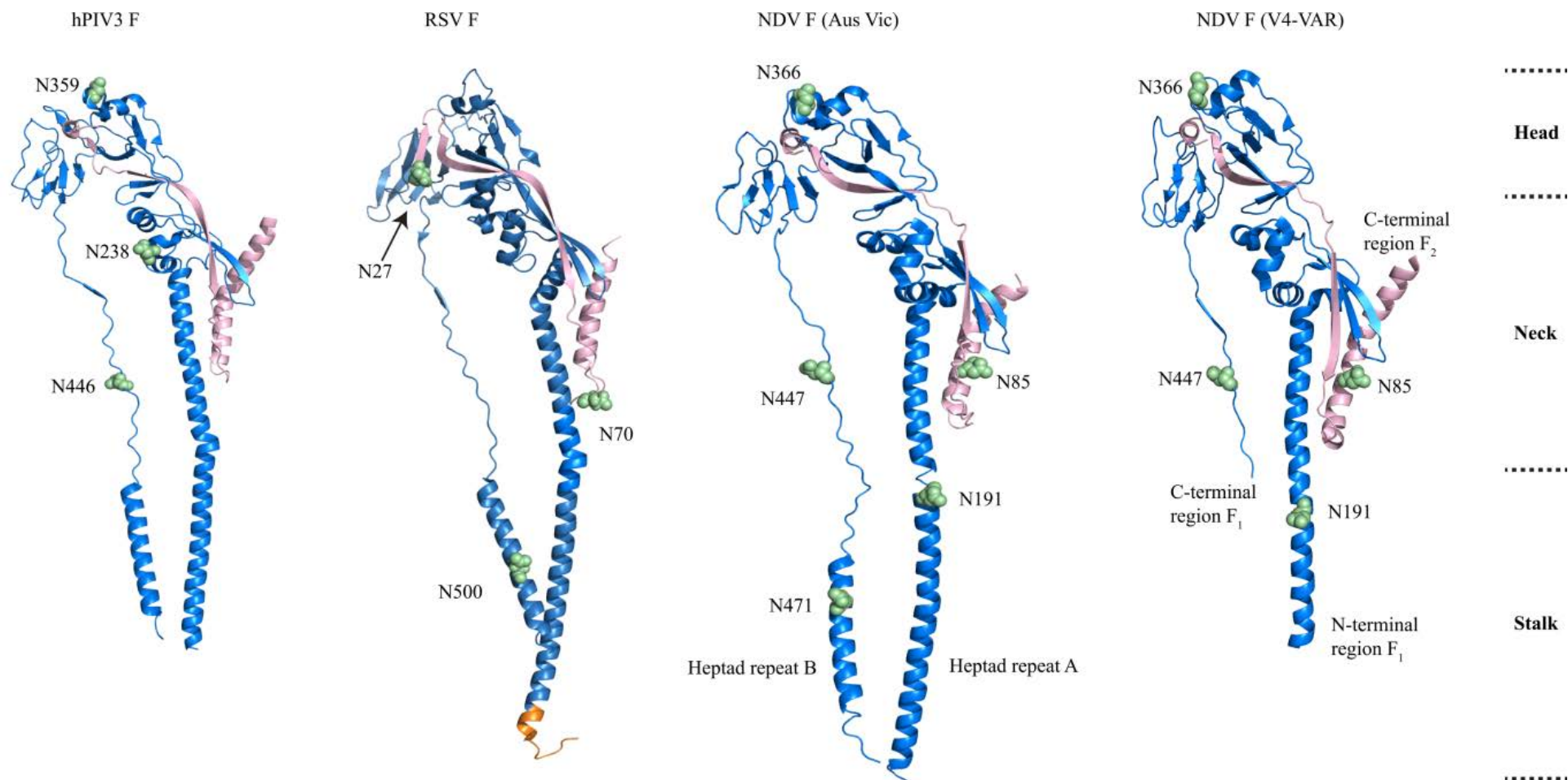
Figure 5-10. Position of N-linked sites on the crystal structures of F proteins from paramyxoviruses. Cartoon representation of F proteins from hPIV3, RSV F, NDV Australia Victoria/32 (Aus Vic) and NDV V4-VAR (RCSB PDB identifiers 1ZTM, 3RRR, 3MAW and 1G5G, respectively). The structures have been represented in their monomeric form to enable easier comparisons. The F_2 and F_1 subunits have been identified in pink and blue, respectively. The fusion peptide has been identified in orange and is only present on hRSV F. N-linked sites are represented by green spheres. General regions of the F proteins are represented on the NDV proteins with the stalk, neck and head domains represented on NDV V4-VAR. All images were created PyMOL (version 1.3).

Over the last several years it has been postulated that the F proteins of paramyxoviruses undergo a transition from a metastable prefusion state to a stable post-fusion state (39, 63, 64, 70, 327). To date, the crystal structures of F in a prefusion conformation have been solved for PIV5 (325), hRSV (328) and HeV (329). These conformational differences in F may account for the different roles observed for N-linked sites that reside in the same stalk, neck and head regions of the previously solved structures. The variable roles observed for N-linked sites in F proteins from paramyxoviruses may also be due to the expression systems used in each mutation study or the type of mutation introduced to eliminate each N-linked site. Furthermore, post-translation and functional differences between the F proteins may account for differential glycosylation or function. For example, the F_0 precursors from HeV, NiV and SeV are not cleaved by furin-like proteases during transport through the *trans*-Golgi, but rather by cathepsin L and trypsin-like proteases (55, 319). Moreover, the crystal structure of F from HeV reveals that the conformation of the fusion peptide adjacent to the cleavage site differs to PIV5, likely due to substrate requirements of the enzymes that cleave each protein (329). There is also a greater reliance on the attachment proteins of *Paramyxovirinae* for fusion promotion than in the *Pneumovirinae* subfamily (37). Even within *Paramyxovirinae* different models of fusion activation are predicted (39, 55). These differences may necessitate divergent roles for N-linked sites on the respective F proteins, highlighting the need for site-specific investigations into occupancy and glycan heterogeneity of F from each virus.

This study represents a step towards defining site-specific glycan heterogeneity of F proteins from paramyxoviruses. Using F derived from V4-VAR virions it was demonstrated that site N85 of the F_2 subunit was occupied. The assignment of high mannose compositions (Man_{5-8}) conforms with previous observations of high mannose glycans at N85 of V4-VAR F (105). The presence of high mannose glycans indicates that accessibility to N85 by glycosidases and glycosyltransferases is hindered during transport of F through the ER and Golgi. Interestingly, peptides containing N85 were only observed in a glycosylated form signifying high occupancy at this site. Elimination of the N-linked consensus site containing N85 has been shown to impair folding of F, significantly reduce cleavage of F_0 and reduce surface expression and fusion activity (102). Site N85 is situated nine residues downstream from a Cys residue that is involved in the disulfide bond linking the F_2 and F_1 subunits. Furthermore, N85 is in close proximity to the F_2/F_1 cleavage site. Thus glycosylation of N85 may act to hold F_0 in a conformation that enables proper disulfide bond formation and allows proteases to access the cleavage site. However, a later study revealed deletion of N85 did not significantly affect cleavage, surface expression or the fusion properties of F (103). The two studies used different expression systems and the latter work did not compare reduced and unreduced

samples, relying on apparent MW to draw conclusions on cleavage. This may account for the differences observed between the two studies.

The results herein also demonstrated that site N191, N366 and N471 of the F_1 subunit are occupied. Peptides containing N191, N366 and N471 were not observed without glycosylation indicating that, like N85, these sites may also be highly occupied. Previous studies have predicted glycosylation at these sites on NDV F, but elimination of each individual consensus site did not significantly alter folding, F_0 cleavage or surface expression of F (102, 103). Mutation of site N191 was associated with decreased fusion, although the effects were not as pronounced as those observed for N85 (102). Interestingly, combined mutations of N471 found in HRB and N191 found in HRA (**Figure 5-10**) increased replication, virulence and immunogenicity of NDV in chickens (103). In paramyxoviruses both HRA and HRB play pivotal roles in the rearrangement of F into an intermediate then post-fusion conformation. They facilitate the insertion of the fusion peptide into host cell membranes and associate to form a stable six-helix bundle in the post fusion conformation of F (67). As discussed, the additional processing of glycans observed at site N191, albeit at low levels, indicates the F protein is transported through the medial- and *trans*-Golgi. The higher degree of processing at site N191 indicates that this region of HRA may remain more accessible to glycosidases and glycosyltransferases than other regions of V4-VAR F containing N-linked sites. Site N471 was the only site identified with a Man_9 structure indicating that HRB may be highly shielded during processing of V4-VAR F.

Unfortunately, occupancy of N447 could not be established in this work as peptides or glycopeptides containing this site were not observed. Site N447 and N471 were both positioned on a large theoretical tryptic peptide (aa 447-480 in **Figure 5-3a)** with a calculated mass of 3,586 Da. During the analyses of V4-VAR F_1 this peptide was not detected. Site N447 is also situated directly C-terminal to a trypsin cleavage site at position K446. It was predicted that potential glycosylation of N447 may hinder trypsin sterically (255). This would produce an even larger theoretical tryptic peptide containing N447 and N471 (aa 436-480 in **Figure 5-3a)** with a calculated mass of 4,825 Da. To avoid potentially large tryptic peptides and to yield separate peptides containing N447 and N471, preliminary work investigated the use of Glu-C to produce the theoretical peptides [442]ATYQKNISIQD[452] and [467]LGNVNNSISNALD[479], respectively. However, digestion of V4-VAR F_1 with Glu-C did not result in detection of these theoretical peptides (data not shown). For this reason V4-VAR F_1 was first digested with trypsin followed by PNGase F or Glu-C to investigate occupancy and heterogeneity at N447 and N471.

Interestingly, Mascot searches of all digests of F_1 identified peptides directly N-terminal to N447. This revealed that trypsin was able to cleave C-terminal to K446. Therefore, it is not clear why the theoretical tryptic peptide containing N447 and N471 (aa 447-480) was not detected after PNGase F treatment. It may be due the large size of the peptide (3,586 Da) or modifications that were not included in the search parameters. To investigate the latter possibility a wild card search was conducted with Byonic, but this did not yield any results. To assess if stepped collision energy was insufficient for the fragmentation of the potentially deglycosylated peptide, manual searching was undertaken in Xcalibur for both precursor and fragment ions with varying levels of deamidation. Again this search did not yield any results. Subsequent digestion of F_1 tryptic peptides with Glu-C also did not enable detection of glycopeptides containing N447. This may be due to the small size of the theoretical peptide [446]KNISIQD[452]. There are conflicting reports on occupancy at N447 using site-directed mutagenesis combined with gel-electrophoresis (13,14). However, the crystal structure of NDV V4-VAR F revealed up to three residues ($GlcNAc_2Man_1$) at N447 in trimers of F, indicting this site can be occupied.

Glycosylation of N541 was not observed in this work, however peptides containing N541 were observed without glycosylation suggesting this site is not occupied. As the site of N-glycan attachment (N541) was deamidated in the trypsin digested sample, PNGase F digestion could not be used to infer occupancy. Site N541 resides within the cytoplasmic domain of V4-VAR F, therefore it is not expected to be glycosylated as transfer of N-linked oligosaccharides occurs in the lumen of the ER (14). Mutation studies have also predicted that N541 is not glycosylated (13,14).

Overall this study provides an interesting insight into glycosylation of NDV F as it reveals that sites N85, N191, N366 and N471 may generally remain inaccessible to glycosyltransferases irrespective of the cell line or tissue used to propagate avirulent NDV virions. This study also demonstrated that species of V4-VAR F are transported through the late Golgi by the presence of fucosylated hybrid or complex glycans at site N191. The high mannose profile of NDV F differs to F derived from virions of other paramyxoviruses. This suggests that despite overall conservation of the tertiary and quaternary structures of F proteins between paramyxoviruses, NDV F may have a different structural arrangement during transport of the protein through the ER and Golgi pathway. Crystal structures of NDV F in a prefusion conformation or uncleaved form may help to shed light on any potential differences in arrangement. It would also be of interest to determine if the high mannose profile observed on V4-VAR is present on more virulent strains of NDV F. As N-linked sites are

highly conserved amongst NDV strains, variations in glycosylation profiles may indicate different conformations of F between virulent and avirulent strains.

Chapter 6: Characterisation of glycosylation of human respiratory syncytial virus fusion (F) protein

6.1 SUMMARY

Human respiratory syncytial virus is recognised as a serious human pathogen that contributes significantly to the burden of respiratory disease worldwide. Despite this burden effective vaccines and treatments for hRSV remain elusive. The F protein of hRSV is a major target for the development of therapies against this pathogen. Substantial structural detail has been described for hRSV F. However, there is a paucity of data describing the glycosylation profile of this protein. Mass spectrometry techniques were applied herein to a soluble recombinant trimeric form of F (sF) produced in human embryonic kidney cells, providing the first site-specific characterisation of glycans from F. Implementation of HCD-pd-EThcD methods enabled the identification of 20, 19, 7, 24 and 70 glycans at N-linked sites N27, N70, N116, N126 and N500, respectively. The use of stepped HCD collision energies and EThcD provided specific advantages for the identification of glycopeptides from sF. Stepped HCD enabled confident identification of glycopeptides containing sites N27 while EThcD was able to distinguish that the isomeric peptide sequence containing N126 was ^{124}KTNVTLSK^{131} rather than ^{125}TNVTLSKK^{132}. Both fragmentation methods facilitated the identification of sulfated glycopeptides. Some of the attached glycans were present in fucosylated, sulfated and/or sialylated forms as evident from the observation of specific oxonium ions. Many of the observed N-linked glycans exhibited fragmentation characteristics consistent with diHexNAc units, which could potentially represent the GalNAcβ1–4GlcNAc or LacdiNAc motif. These antenna are not typically observed in mammals and are potentially immunogenic. Moreover, O-linked glycosylation of the protein is described for the first time, with a glycan attached at T100. These findings form the basis for future functional and comparative studies of the glycan structures and occupancy of N- and O-linked sites from the native F protein. These findings may also have implications for future studies into viral infectivity and the design of therapeutic agents.

6.2 INTRODUCTION

Infection by hRSV is a significant contributor to acute lower respiratory disease in children, the elderly and immunocompromised individuals (110-113). Although hRSV virions present two major glycoproteins on the viral membrane the F protein remains the major target of vaccines and neutralising therapeutics (118). The F protein mediates membrane fusion and is essential for viral replication while G may be dispensable in certain cell cultures (72, 126). Like other members of the paramyxovirus family hRSV F_0 is cleaved into an active form by furin-like host proteases in the *trans*-Golgi (330), resulting in disulfide linked polypeptide chains F_2 and F_1. Despite structural conservation of F throughout the paramyxovirus family (63), hRSV F differs through the presence of a second cleavage site in F_0 which results in the loss of the soluble pep27 (128).

Another post-translational modification hRSV F undergoes following biosynthesis is the addition of N-linked glycans in the ER (140). These N-linked glycans are further processed in the Golgi complex during the maturation of hRSV F proteins (135, 314). As discussed in Chapter 1 of this thesis, N-linked structures may contain elongated antennas. The most common antenna is formed after the addition of Gal to a terminal GlcNAc in a beta-linkage (Galβ1–4GlcNAc) also referred to as LacNAc (15). On a restricted number of mammalian glycoproteins unique beta-linked GalNAc (GalNAcβ1–4GlcNAc) or LacdiNAc motifs have been observed (331) (**Figure 6-1**). The antennas can be decorated with a number of different monosaccharides and substituents including NeuAc, Sulf and Phos. The F protein of hRSV contains five highly conserved N-linked consensus sites; N27, N70, N116, N126 and N500. Sites N27 and N70 reside in F_2, N116 and N126 in the pep27 region of F_2 and N500 in F_1. The Long strain of RSV, a strain of subtype A, contains an additional N-linked site N120, also located in pep27. Despite this conservation, lack of glycosylation at these sites does not affect proteolytic cleavage of F or cell surface presentation (139, 330). Of the five N-linked sites, three are thought to be occupied N27, N70 and N500, with N500 being required for syncytium formation (139). Mutation of site N70 was seen to increase fusion activity while dual mutation of N27 and N70 decreased fusion activity. Furthermore, the importance of maturation and proper formation of N-linked glycans in infection and subsequent syncytia formation has been reported for the hRSV F protein (135).

Elucidation of the monosaccharide compositions of glycans at each N-linked site of hRSV F may provide a basis for experiments to determine the role N-linked glycosylation plays in the infectivity of the virus. Application of MS methods herein to a soluble recombinant trimeric form of RSV F

(sF), which has been extensively structurally characterised in pursuit of drug and vaccine design objectives (332), permitted the first site-specific analysis of glycan heterogeneity for hRSV F. A total of 20, 19, 7, 24 and 70 glycans were identified at N-linked sites N27, N70, N116, N126 and N500, respectively. Many of the observed glycans exhibited characteristics of diHexNAc units, which could potentially be LacdiNAc motifs. These are not typically observed on mammalian glycans and presence of these units on hRSV F or sF may have far-reaching implications for biological studies and the design of vaccines or anti-viral therapies targeting hRSV F. Finally, the observation of O-linked glycosylation of hRSV sF is detailed for the first time. Two O-linked glycans were observed in the F_2 subunit of sF where the site of attachment of $HexNAc_1Hex_1$ could be narrowed down to T100 and the site of attachment of $HexNAc_1Hex_1NeuAc_2$ was localised to S99 or T100.
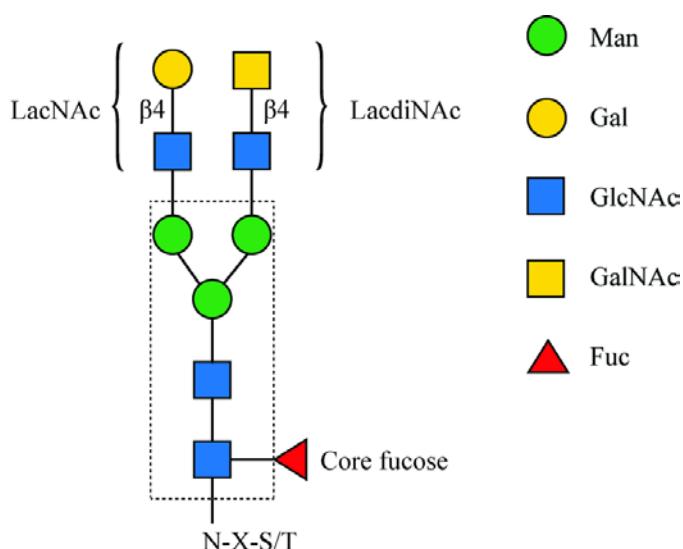
Figure 6-1. Example of a complex N-linked glycan with LacNAc and LacdiNAc antennae. The N-linked trimannosylchitobiose core has been identified in a dotted box. Core fucosylation is represented. The common LacNAc extension and less common LacdiNAc extension have also been illustrated.

6.3   METHODS

6.3.1 Provision of samples

Recombinant sF was kindly provided by Professor Mark. E. Peeples from Nationwide Children's Hospital (Columbus, OH, USA). The gene sequence of sF was based on hRSV strain A2 and contains the following introduced amino acid mutations S155C, S290C, S190F and V207L. The carboxyl-terminal region of the protein containing the transmembrane and cytoplasmic domains have been removed and replaced with a trimerization motif, thrombin site, six His-tag and a StreptagII. The two Ser-to-Cys mutations have been shown to induce a disulfide bond that stabilises

sF in a trimeric prefusion form, the structure and stability of which has been extensively characterised (332). The protein was engineered to maintain an antigenic site (Ø) which is targeted by potent neutralising antibodies. The protein was produced in human embryonic kidney (HEK) 293 FreeStyle™ cells and was and was purified by immobilised metal ion chromatography on a nickel column followed by size exclusion chromatography. It was then concentrated by centrifugation in an Amicon Ultra-30K Filter (Millipore).

6.3.2 Sample preparation and enzymatic digestions of sF

Approximately 50 µg of purified sF was reduced and alkylated before proteins were methanol precipitated with trypsin as per the methods described in Chapter 2. An aliquot (5 µg) of the hRSV sF tryptic digest was also subjected to digestion with 1U of PNGase F in 60 mM NH_4HCO_3 at 37°C overnight. Resultant peptides were desalted with a C18 ZipTip (10 µL pipette tip with a 0.6 µL resin bed; Millipore, MA, USA) using the manufacturers' guidelines for MS analysis.

6.3.3 Nano-ultra-high pressure liquid chromatography

Approximately 200 ng of digested sF was injected for each analysis using a nUHPLC system as described in Chapter 2. Samples were loaded onto the trap column and washed for 3 min at 5 µL/min in 99.5% solvent A and 0.5% solvent B. Peptides and glycopeptides were subsequently eluted onto the analytical column and separated at flow rate of 0.3 µL/min. The analytical column was held at 1% solvent B for 3 min before ramping through a sequence of linear gradients up to 40% solvent B in 60 min, to 70% B over 15 min, to 95% B in 5 min and then holding at 95% B for 5 min. The column was then re-equilibrated with 0.5% B for 20 min.

6.3.4 Mass spectrometry data acquisition

Survey scans of peptide and glycopeptide precursors from *m/z* 300 to 2000 were acquired in the Orbitrap at 120K resolution (FWHM) at *m/z* 200 using an AGC target of 400,000 and maximum injection time of 50 ms. For internal mass calibration the lock mass option was enabled using the polycyclodimethylsiloxane ion at *m/z* 445.1200 (310). The most intense precursors (*m/z* 300 to 1800) with charges 2-6 and intensities over 5,000 counts were selected for fragmentation by HCD.

Precursor ion isolation was performed with a mass selecting quadrupole using an isolation window of *m/z* 2. Precursors were fragmented in the ion routing multipole, using either a NCE of 30% or, in a separate chromatographic run, a stepped method of ±5% around a NCE of 25% (hereafter denoted as 25±5%). Previously selected ions within a ±10 ppm window were dynamically excluded for 15 s. Fragment ions were acquired in the Orbitrap at a resolution of 30K using an AGC target of 50,000 and maximum injection time of 100 ms. If fragment ions were produced within the top 20 ions corresponding to 204.0867 (HexNAc), 163.0601 (Hex) or 292.1027 (NeuAc) within a ±10 ppm window the precursors ions were re-isolated and subjected to EThcD using supplemental activation with a NCE 15%. Fragment ions produced by EThcD were acquired in the Orbitrap at a resolution of 60K using an AGC target of 200,000 and maximum injection time of 250 ms with 2 microscans. The trypsin digest of sF was analysed using both HCD NCE of 30% and HCD NCE of 25±5% protocols. The trypsin digest that was additionally digested with PNGase F was analysed using the HCD NCE of 25±5% protocol

6.3.5 Data processing of non-glycosylated and deglycosylated peptides from sF

Proteome Discoverer (v2.1.0.81) and the search engine Mascot were used to search HCD MS/MS spectra from the RAW files of the analyses of sF. The protein database contained the sequence for sF. Cleavage specificity was set as semi-tryptic with a maximum of two missed cleavages. Mass tolerances of 10 ppm and 0.02 Da were applied to precursor and fragment ions, respectively. Carbamidomethylation of Cys was set as a fixed modification and dynamic modifications included mono-oxidised Met, deamidation of Asn and Gln residues, conversion of N-terminal Gln to pyroglutamate and loss of ammonia from N-terminal carbamidomethyl Cys residues. The "Fixed Value PSM Validator" node of Proteome Discover was used and a cut-off score of 30 was applied to all PSMs.

6.3.6 Assignment of the monosaccharide compositions of N-linked glycans from sF

Spectra from the HCD MS/MS analysis of tryptic peptides from sF were analysed with OxoExtract. The following parameters were used: digestion with trypsin; maximum of two missed cleavages; fixed modification of carbamidomethylation of Cys and a dynamic modification of mono-oxidised Met. The protein database queried contained the sF sequence with the signal peptide removed. The optional feature of GlycoMod was enabled. In addition to the default monosaccharide parameters

the substituents Phos and Sulf were considered as possible components of the N-glycans. For peptide modifications that were not available in OxoExtract (conversion of N-terminal Gln to pyroglutamate and loss of ammonia from N-terminal carbamidomethyl Cys residues) the mass of the peptide was inferred manually from the glycopeptide Y1 ion and the monosaccharide compositions were allocated after manual searches in GlycoMod. The observed compositions from the OxoExtract search were used for subsequent searches in Byonic (detailed below).

6.3.7 Assignment of N-linked and O-linked glycopeptides from sF

Byonic was used to analyse HCD and EThcD mzML files separately with Byonic using a protein database that contained the sequence for sF with the signal peptide removed. Cleavage specificity was set C-terminal to Lys and Arg residues allowing a maximum of two missed cleavages. Mass tolerances of 10 ppm and 0.02 Da were applied to precursor and fragment ions, respectively. A fixed modification Carbamidomethylation of Cys was allowed. Two dynamic modifications were allowed per peptide and these included mono-oxidised Met, deamidation of Asn and Gln residues, conversion of N-terminal Gln to pyroglutamate and loss of ammonia from N-terminal carbamidomethyl Cys residues. One glycan attached at an N-linked consensus site was allowed per peptide. The N-linked glycan database queried was a combination of the Byonic mammalian database (309_Mammalian no sodium) with all glycans containing NeuGc removed and any additional glycans (32 monosaccharide compositions in total) assigned in the OxoExtract search. The default protein false FDR and peptide output options were changed to "Show all N-glycopeptides" as recommended by the manufacturer when analysing simple samples. After manual inspection of the Byonic results a cut-off score of 100 was chosen for glycopeptides fragmented by HCD. For glycopeptides fragmented by EThcD a cut-off score of 100 was applied for those containing sites N27 and N500. No cut-off score was applied for glycopeptides containing N70, N116 and N126 due to low scores assigned by Byonic which was likely due to the small length of the peptide portions of the corresponding glycopeptides. To confirm glycopeptide assignments by EThcD the identification of at least three z- or c-ions was required. For O-linked searches the enzyme specificity was changed to semi-specific with ragged cleavage allowed at the C-terminus of glycopeptides. One O-linked glycan was allowed per peptide attached at Ser or Thr residues and the glycan database queried was the common O-glycan database which contained six common mucin-type glycans ($HexNAc_1$, $HexNAc_2$, $HexNAc_1Hex_1$, $HexNAc_2Hex_1$, $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$).

6.4 RESULTS

6.4.1 Identification of N-linked sites of sF

A schematic of sF is represented in **Figure 6-2** revealing the positions of the five N-linked consensus sites in the protein. Four sites, N27, N70, N116 and N126, are positioned in the F_2 subunit, with the latter two in the pep27 region of F_2. The remaining site, N500, is positioned in the F_1 subunit. The two sites of furin-like cleavage at R136 and R109 are identified in the insert. In the present study trypsin was used to digest sF and predicted theoretical peptides containing sites N27, N70 and N500 are shown in **Figure 6-2**. Theoretical tryptic peptides containing sites N116 (^{114}FMNYTLNNAK^{123}) and N126 (^{125}TNVTLSK^{131}) can be deduced from the amino acid sequence of pep27. Allowing one missed cleavage, or one non-canonical internal trypsin cleavage site, in peptides containing N126 resulted in the isomeric peptides ^{124}KTNVTLSK^{131} and ^{125}TNVTLSKK^{132}.
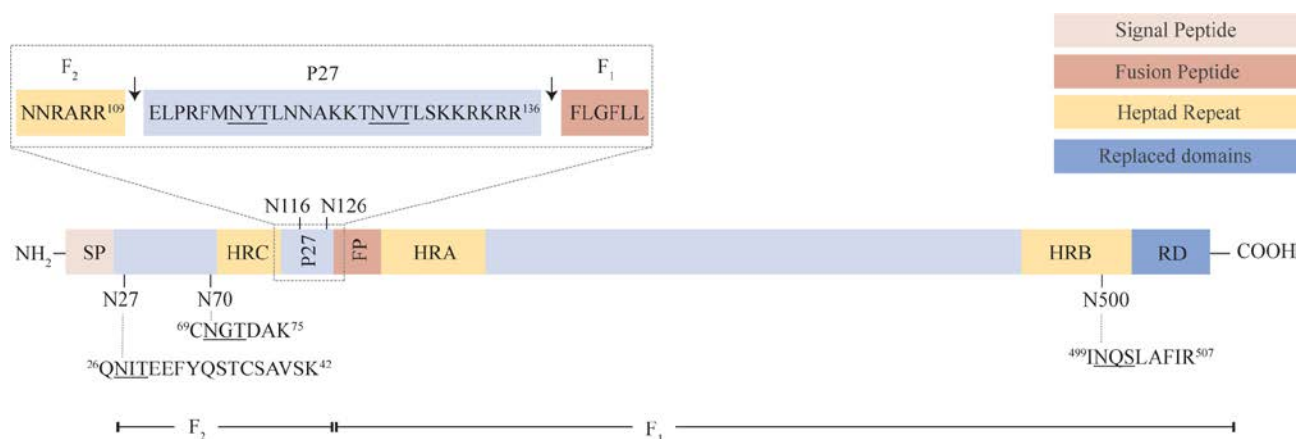
Figure 6-2. Schematic of hRSV sF. The signal peptide (SP), heptad repeats (HRA, HRB and HRC), pep27 (P27), fusion peptide (FP) and the replaced (RD) transmembrane and cytoplasmic domains are identified. The N-linked sites are marked with a vertical line with the amino acid number of the respective Asn residue. Theoretical tryptic peptides containing N-linked sites N27, N70 and N500 are shown with the N-linked consensus sites underlined. An insert reveals the two furin-like cleavage sites (denoted by arrows) that lead to the production of the F_2 and F_1 subunits after cleavage at R136, and pep 27 after cleavage at both R109 and R136. The two N-linked consensus sites within pep27 (N116 and N126) have been underlined.

6.4.2 Identification of non-glycosylated and deglycosylated peptides from sF using HCD fragmentation

Mascot searches were performed on the analyses of trypsin digested sF using 30% and stepped HCD and trypsin/PNGase F digested sF using stepped HCD (Supplementary Tables S6-1 to S6-3, respectively). Greater than 80% sequence coverage was observed in each search (Appendix E). In the samples that were digested with trypsin alone, PSMs containing N27 were not observed while PSMs containing N70, N116, N126 and N500 were observed (Supplementary Tables S6-1 and S6-2).

In the PNGase F digested sample, PSMs were observed that corresponded to sequences containing N27, N116, N126 and N500 where Asn residues in N-linked consensus sites were deamidated (Supplementary Table S6-3). A tryptic peptide containing N70 was not observed in this sample. The peptide [69]cNGTDAK[75], where lowercase "c" represents carbamidomethyl Cys, may be susceptible to NH_3 elimination from the N-terminal S-carbamoylmethylcysteine (333). As such, the variable modification "pyro-carbamidomethyl" was included in the parameters of the Mascot searches. However, peptides containing N70 that had undergone such a reaction were also not identified. The peptide [69]cNGTDAK[75] was hydrophilic and eluted quite early in the trypsin digested sample that was not subjected to a C18 ZipTip clean-up protocol before MS. As the trypsin/PNGase F digest was subjected to a C18 ZipTip clean-up protocol before MS, the potentially deglycosylated versions of this peptide may have been lost during this procedure.

Interestingly, the amino acid sequence [88]NAVTELQLLMQSTPATNNRARR[109], which includes the furin cleavage site (R109), was observed with semi-tryptic cleavages at the C-terminal end of the tryptic peptides (e.g. [88]NAVTELQLLMQSTPATNNRA[107] and [88]NAVTELQLLMQSTPATNN[105]). Trypsin cleaves predominantly C-terminal to Lys and Arg residues (334), therefore the "semi-tryptic" cleavages may originate from trimming by carboxypeptidases (88).

6.4.3 The use of different HCD collision energies to identify glycopeptides from sF

The two HCD NCEs of 30% and 25±5% were investigated for identification of glycopeptides from sF. Performance of two separate chromatographic experiments using 30% NCE or 25±5% NCE provided a general qualitative comparison of the two methods. The manually validated Byonic

search results for glycopeptides fragmented by HCD with a NCE of 30% and 25±5% are presented in Supplementary Table S6-4 and S6-5, respectively. It was revealed that 25±5% NCE permitted more confident assignment of the peptide mass from glycopeptides containing N27 compared to 30% NCE. The N-terminal amino acid sequences of peptides containing N27 were not derived from cleavage by trypsin but rather from removal of the signal peptide. The Mascot search of the trypsin/PNGase F digest identified several different N-terminal sequences and modifications of peptides containing N27 which included [24]SGQNITEEFYQSTcSAVSK[42], [26]QNITEEFYQSTcSAVSK[42] and [26]qNITEEFYQSTcSAVSK[42], where lower case "q" represents conversion of N-terminal Gln to pyroglutamate. Fragmentation of glycopeptides containing N27 with 30% and 25±5% NCE produced peptide sequence ions of low relative abundance predominantly from the C-terminal region of the peptide (**Figure 6-3a** and **Figure 6-3b**, respectively). Both glycopeptide assignments were given high scores of ~554 by Byonic, however, 25±5% NCE also produced glycopeptide Y1 ions with varying degrees of the peptide with monosaccharides attached (**Figure 6-3b**). Production of the Y1 ion was critical for the identification of the peptide portions of glycopeptides containing N27 as the putative site of glycosylation was also at the N-terminal region of the peptides. As peptide sequence ions were not observed from this region of the glycopeptides, the peptide mass could have been incorrectly assigned and the mass difference taken or added to the glycan. Of particular concern was the mass difference between N-terminal Gln and its conversion to pyroglutamate (-17.027 Da). This mass shift could be mistaken for the difference between Hex_1dHex_1 and $NeuAc_1$ (-17.015 Da) in the glycan. Thus, production of glycopeptide Y1 ions using 25±5% enabled confident assignment of glycopeptides containing N27.

Comparison of 30% and 25±5% NCE also revealed that peptide sequence ions were of low relative abundance for glycopeptides containing sites N70, N116, N126 and N500 (Supplementary Figure S6-1a-d, respectively). Finally, 25±5% NCE yielded more informative spectra for sulfated glycopeptides than 30% NCE. This arose through the production of sulfated oxonium ions in 94% and 68% of spectra of sulfated glycopeptides recorded with 25±5% and 30% NCE, respectively. The production of sulfated oxonium ions enabled the distinction between sulfation and phosphorylation of glycopeptides to be made, as illustrated in the following section. Given that the higher dissociation energy did not significantly improve the relative abundance of peptide b- and y-ion and yielded less sulfated oxonium ions a 25±%5 NCE was used for the subsequent analyses of N-linked glycopeptides from hRSV sF.
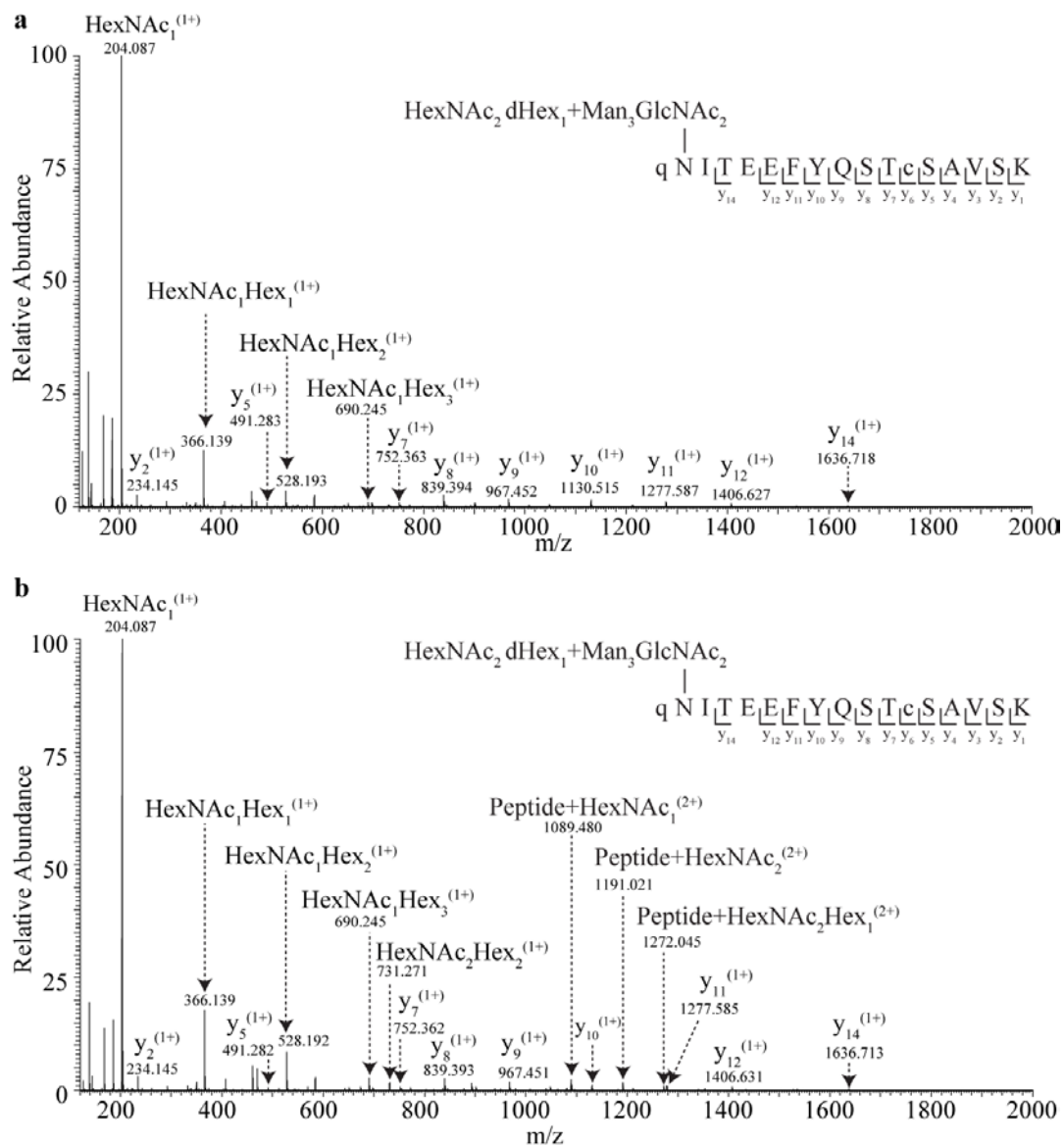
Figure 6-3. Comparison of 30% HCD and stepped HCD for the identification of glycopeptides containing site N27. (A) HCD fragmentation of precursor ion at *m/z* 1140.474 (3+) using a NCE of 30% with a Byonic score of 553.6 for the glycopeptide assignment. (B) HCD fragmentation of precursor ion at *m/z* 1140.474 (3+) using a NCE of 25±5% with a Byonic score of 553.7 for the glycopeptide. Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and the spectra are labelled accordingly. Lowercase "q" and lowercase "c" in the peptide sequence represent pyroglutamate and carbamidomethylation of Cys, respectively. Not all ions have been labelled in the spectra for ease of interpretation.

6.4.4 Stepped HCD and EThcD fragmentation of N-linked glycopeptides from sF

The use of stepped HCD with a NCE of 25±5% and EThcD were investigated for the identification of glycopeptides from sF. The manually validated Byonic search results for glycopeptides fragmented using a stepped HCD and EThcD are presented in Supplementary Table S6-5. In total

156 non-redundant glycopeptides were observed that contained one of the five N-linked sites. Annotated HCD and EThcD spectra for all accepted glycopeptides are presented in Supplementary Figure S6-2 and S6-3, respectively. Of these glycopeptides, 93 were assigned by both HCD and EThcD while 56 were assigned solely by HCD and seven were assigned solely by EThcD (**Figure 6-4a**). When assessing the value of HCD and EThcD for the assignment of hRSV sF glycopeptides containing each N-linked site (**Figure 6-4b**) it was observed that HCD performed substantially better for the assignment of glycopeptides containing sites N27 and N500. At these sites no glycopeptides were identified solely by EThcD, while large proportions were uniquely identified by HCD. Alternatively, EThcD uniquely identified glycopeptides that contained N70, N116 and N126 although a larger proportion of glycopeptides at N70 were uniquely identified by HCD.

Despite EThcD fragmentation resulting in fewer glycopeptide assignments a closer look at the spectra revealed that in many instances EThcD produced considerably more intense peptide fragment ions as evident in the HCD-pd-EThcD spectra presented in **Figure 6-5 to 6-9**, for N-linked sites N27, N70, N116, N126 and N500, respectively. Furthermore, EThcD resulted in greater sequence coverage of the peptide portion of the glycopeptides containing N70, N116, N126 and N500 (**Figures 6-6 to 6-9**). Importantly, EThcD confirmed that the isomeric peptide sequences containing N126, was [124]KTNVTLSK[131] rather than [125]TNVTLSKK[132] (**Figure 6-8**).
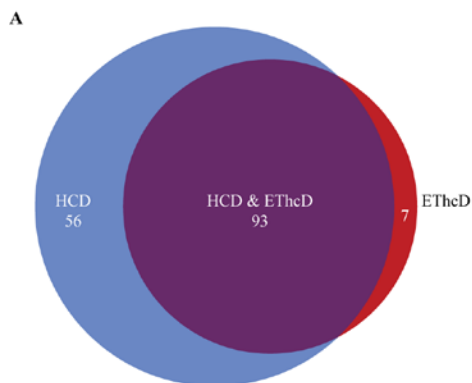
Figure 6-4. Glycopeptides observed from the analysis of trypsin digested hRSV sF using HCD and EThcD fragmentation. (a) Venn diagram illustrating the distribution glycopeptides identified by HCD or EThcD fragmentation. (b) Histogramic representation of glycopeptides identified at each N-linked site using HCD and / or EThcD fragmentation. The peptide amino acid sequences from observed glycopeptides containing each site have been listed.

N27: QNITEEFYQSTCSAVSK

N70: CNGTDAK

N116: FMNYTLNNAK

N126: KTNVTLSK or TNVTLSK

N500 INQSLAFIR
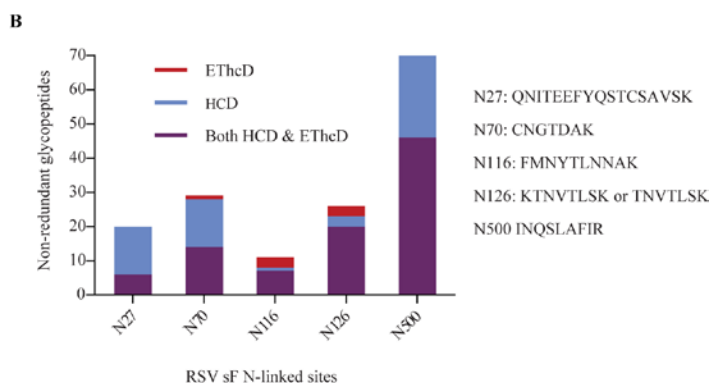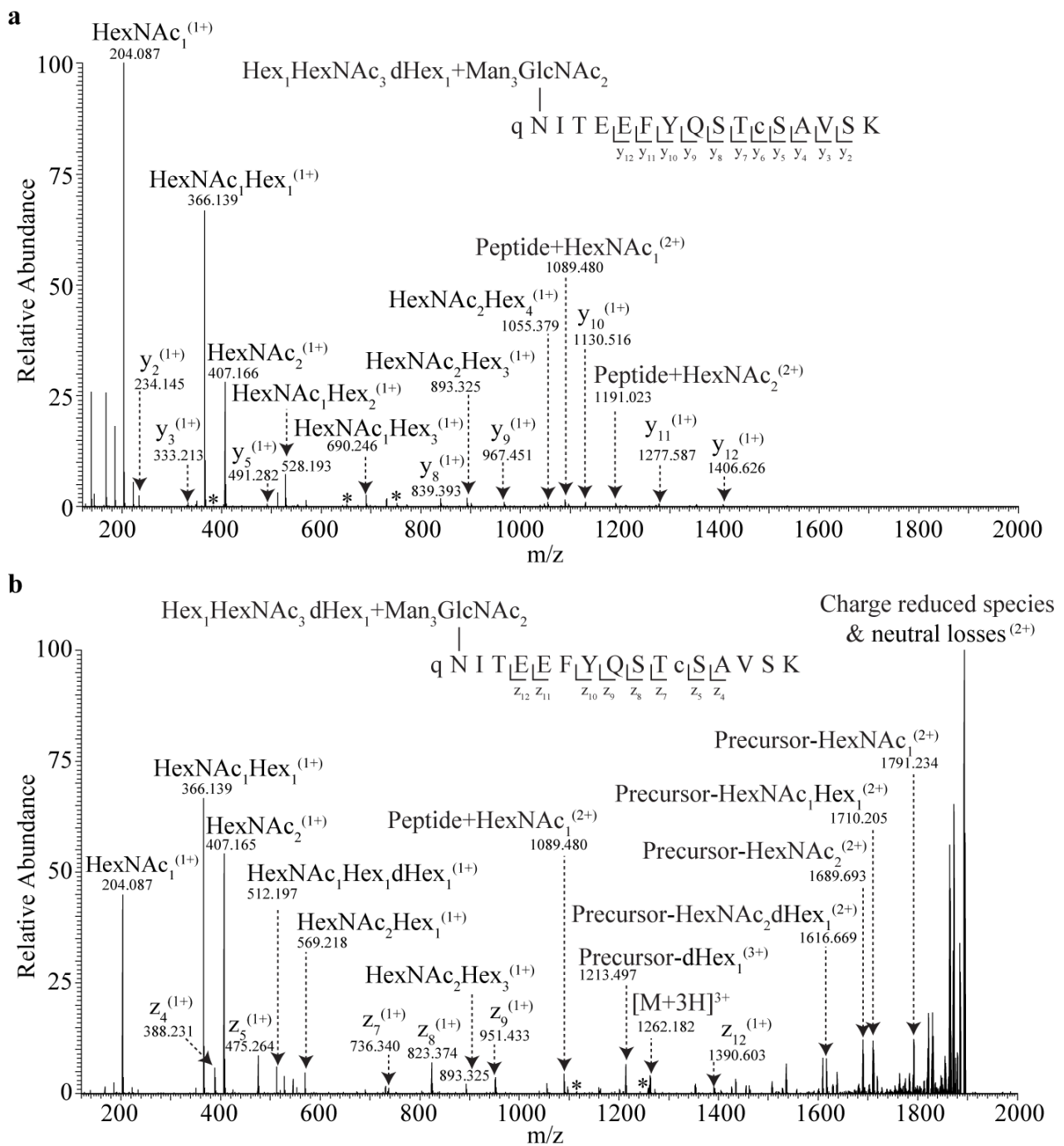
Figure 6-5. HCD and EThcD spectra of an N-linked glycopeptide from hRSV sF containing site N27. (a) HCD fragmentation (NCE of 25±5%) with a Byonic score of 552.6 for the glycopeptide assignment and (b) EThcD fragmentation of the same precursor ion at *m/z* 1262.188 (3+) with a Byonic score of 137.9 for the glycopeptide assignment. Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and peptide sequence ions that have not been labelled in the spectra are denoted with an "*". Lowercase "q" and lowercase "c" in the peptide sequence represent pyroglutamate and carbamidomethylation of Cys, respectively.
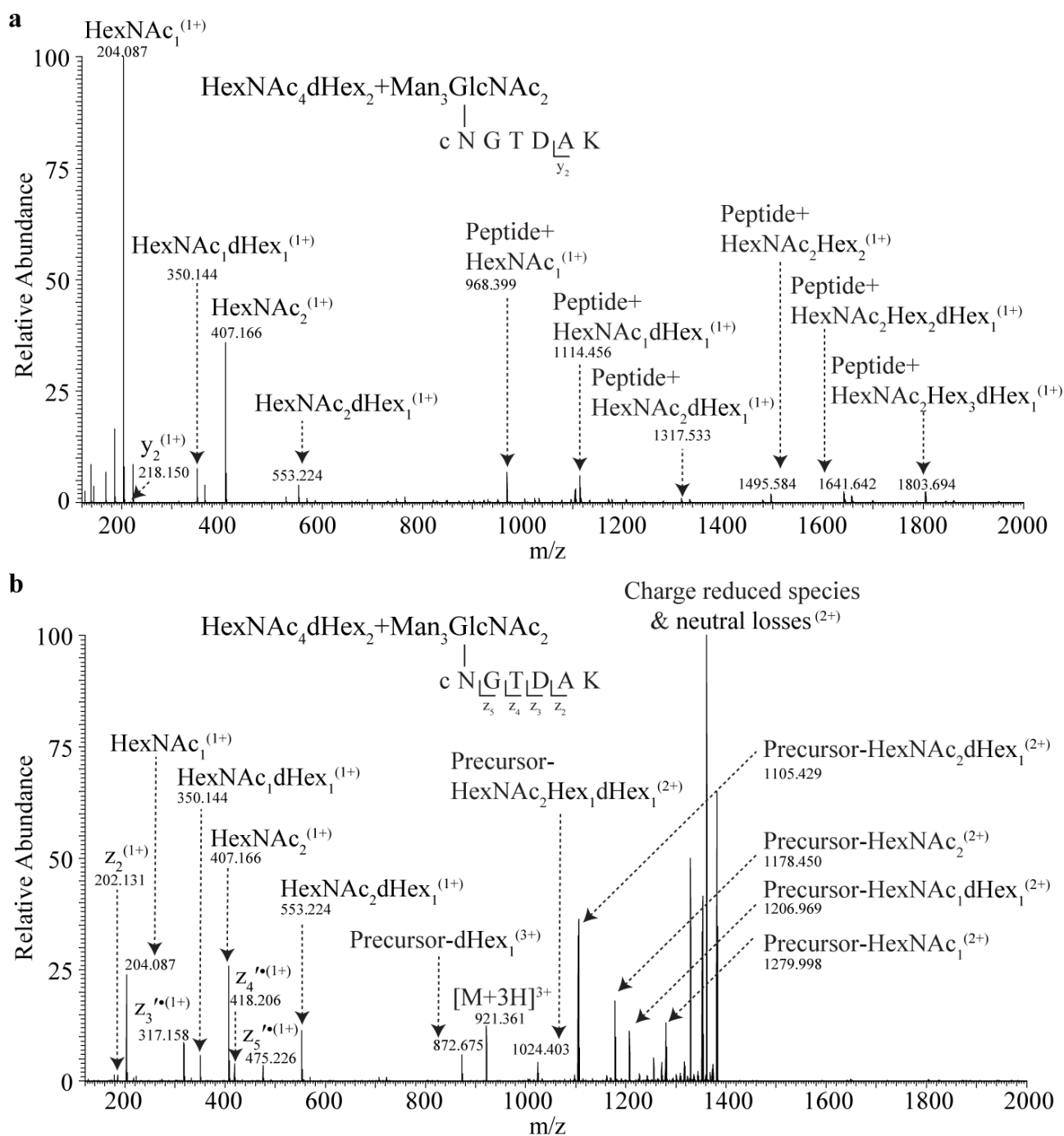
Figure 6-6. HCD and EThcD spectra of an N-linked glycopeptide from hRSV sF containing site N70. (a) HCD fragmentation (NCE of 25±5%) with a Byonic score of 228.7 for the glycopeptide assignment and (b) EThcD fragmentation of the same precursor ion at *m/z* 921.363 (3+) with a Byonic score of 101.8 for the glycopeptide assignment. Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and the spectra are labelled accordingly. Lowercase "c" in the peptide sequence represents carbamidomethylation of Cys. The "'" symbol next to labelled z-ions indicates abstraction of hydrogen (z + H) by the fragment ion. The addition of a dot "·" symbol with the "'" symbol indicates both the radical and z + H ions were observed for the fragment ion as judged by the isotopic distributions described in (176). The remaining z-ions are considered to be typical radical ions as the isotopic distribution of the fragment ions did not enable confirmation of z + H ions.
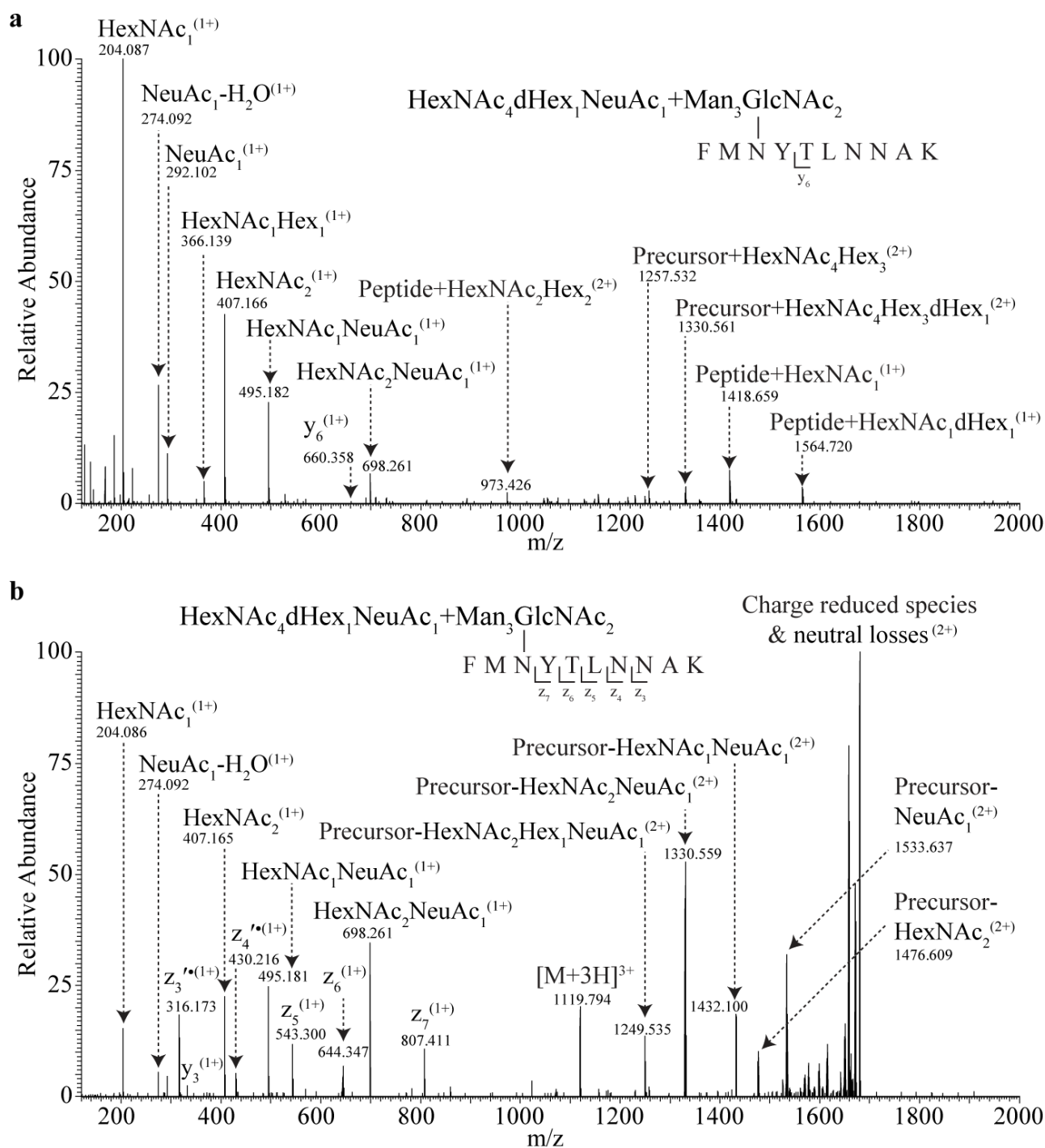
Figure 6-7. HCD and EThcD spectra of an N-linked glycopeptide from hRSV sF containing site N116. (a) HCD fragmentation (NCE of 25±5%) with a Byonic score of 254.7 for the glycopeptide assignment and (b) EThcD fragmentation of the same precursor ion at *m/z* 1119.796 (3+) with a Byonic score of 60.2 for the glycopeptide assignment. Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and the spectra are labelled accordingly. Labelling of radical and even-electron z-ions follows that set out in **Figure 6-6**.

a

K T N V T L S K
|
HexNAc₂+Man₃GlcNAc₂

HexNAc₁$^{(1+)}$ 204.087

Peptide+HexNAc₂Hex₂dHex₁$^{(2+)}$ 883.929

Peptide+HexNAc₂Hex₂$^{(2+)}$ 810.900

Peptide+HexNAc₂Hex₁$^{(2+)}$ 729.873

Peptide+HexNAc₂$^{(2+)}$ 648.849

HexNAc₁Hex₁$^{(1+)}$ 366.139

HexNAc₂$^{(1+)}$ 407.166

Precursor+HexNAc₂Hex₃dHex₁$^{(2+)}$ 964.955

Precursor+HexNAc₃Hex₂dHex₁$^{(2+)}$ 985.971

Precursor+HexNAc₃Hex₃dHex₁$^{(2+)}$ 1066.495

Peptide+ HexNAc₁$^{(1+)}$

Peptide+ HexNAc₁dHex₁$^{(1+)}$ 1093.610

Peptide+ HexNAc₂Hex₁dHex₁$^{(1+)}$

Peptide+ HexNAc₂Hex₂dHex₁$^{(1+)}$ 1239.664 1604.802 1766.849

b

Charge reduced species & neutral losses $^{(2+)}$

Precursor-HexNAc₁$^{(2+)}$ 1066.494

Precursor-HexNAc₁Hex₁$^{(2+)}$ 985.467

Precursor-HexNAc₂$^{(2+)}$ 964.955

c₁ c₂ c₃ c₄
K T N V T L S K
z₅ z₄ z₃ z₂

HexNAc₂+Man₃GlcNAc₂

c₁$^{(1+)}$ 146.129

HexNAc₁$^{(1+)}$ 204.086

HexNAc₂$^{(1+)}$ 407.165

z₂′$^{(1+)}$ 219.134

c₂$^{(1+)}$ 247.176

z₃$^{(1+)}$ 331.209

z₄′•$^{(1+)}$ 433.265

z₅$^{(1+)}$ 531.326

[M+3H]³⁺ 779.027

c₃$^{(1+)}$ 1805.750

c₄$^{(1+)}$ 1904.816

Figure 6-8. HCD and EThcD spectra of an N-linked glycopeptide from hRSV sF containing site N126. (a) HCD fragmentation (NCE of 25±5%) with a Byonic score of 203.6 for the glycopeptide assignment and (b) EThcD fragmentation of the same precursor ion at *m/z* 779.027 (3+) with a Byonic score of 34.6 for the glycopeptide assignment. Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and the spectra are labelled accordingly. Labelling of radical and even-electron z-ions follows that set out in **Figure 6-6**. The radical fragment was not observed for z₂.
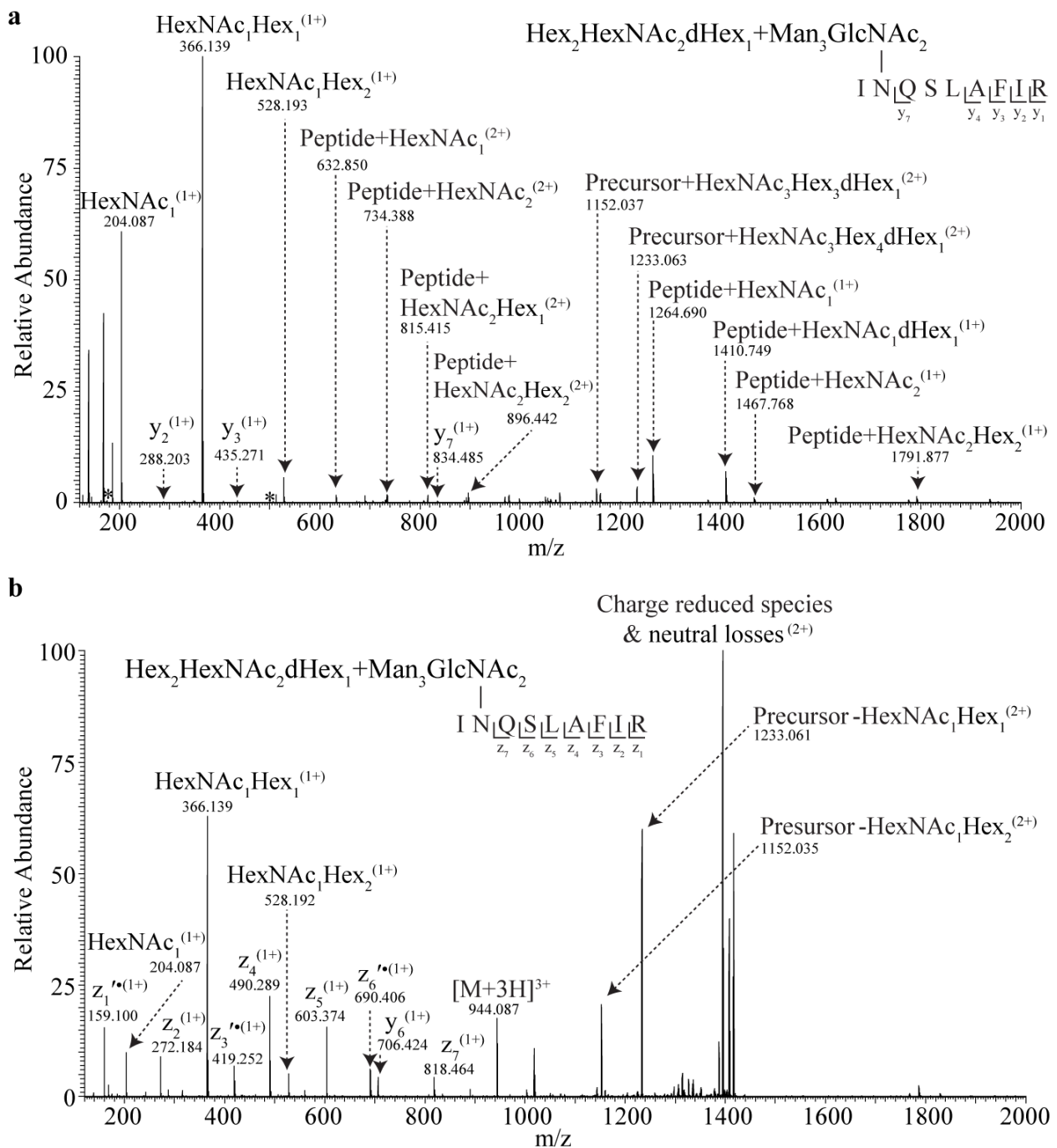
Figure 6-9. HCD and EThcD spectra of an N-linked glycopeptide from hRSV sF containing site N500. (a) HCD fragmentation (NCE of 25±5%) with a Byonic score of 466.9 for the glycopeptide assignment and (b) EThcD fragmentation of the same precursor ion at *m/z* 944.089 (3+) with a Byonic score of 239.4 for the glycopeptide assignment. Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and the spectra are labelled accordingly. Peptide sequence ions that have not been labelled in the spectra are denoted with an "*". Labelling of radical and even-electron z-ions follows that set out in **Figure 6-6**.

Both HCD and EThcD were beneficial for the analysis of sulfated glycopeptides, producing sulfated oxonium ions in 94% of HCD and 100% of EThcD MS/MS spectra (**Figure 6-10**) assigned by precursor mass to sulfated or phosphorylated glycopeptides. In a single spectrum, calculation of the mass difference of 79.954-79.958 Da between sulfated oxonium ions and their non-sulfated equivalents confirmed the presence of Sulf over Phos which have theoretical differences of 79.957 Da and 79.966, respectively (238, 267). This is illustrated in **Figure 6-10a** where the difference between $HexNAc_1$ at *m/z* of 204.087 and $HexNAc_1Sulf_1$ at *m/z* of 284.043 equated to 79.956 Da and in **Figure 6-10b** where the mass shift between $HexNAc_2$ at *m/z* of 407.165 and $HexNAc_2Sulf_1$ at *m/z* of 487.122 (represented by Δ) equated to 79.957 Da.
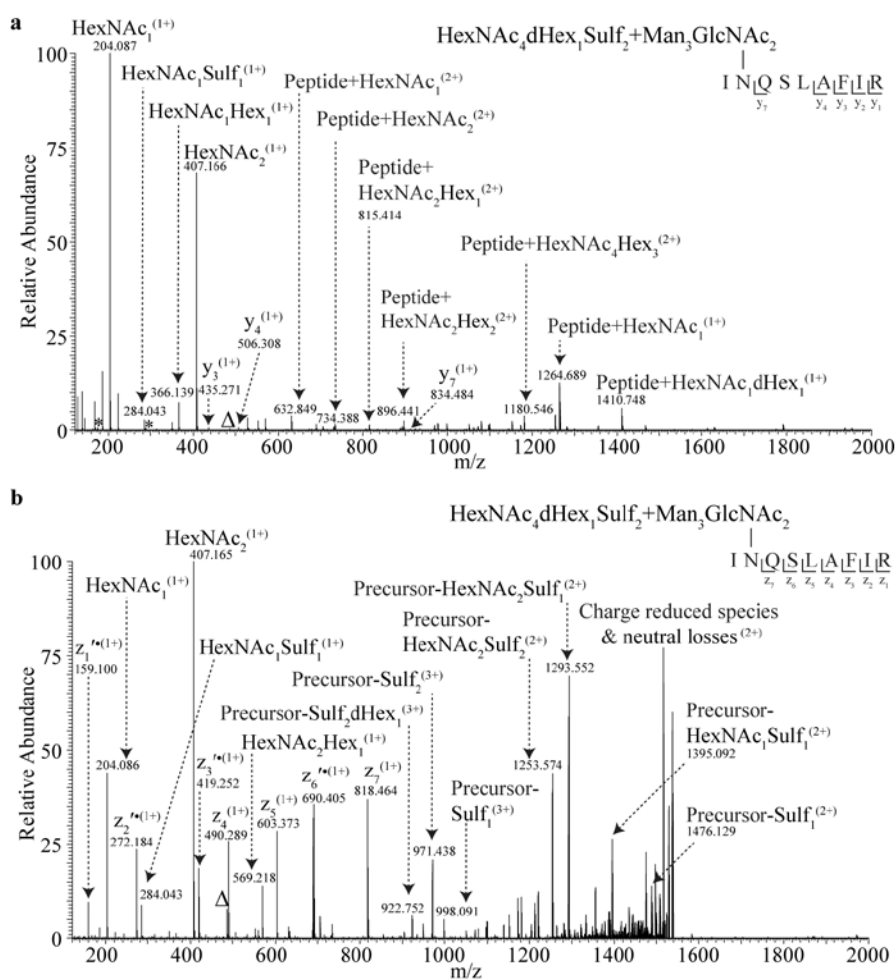
Figure 6-10. HCD and EThcD spectra of a sulfated N-linked glycopeptide from hRSV sF containing site N500. (a) HCD fragmentation with a Byonic score of 456.0 for the glycopeptide assignment and (b) EThcD fragmentation of the same precursor ion at *m/z* 1024.745 (3+) with a Byonic score of 159.1 for the glycopeptide assignment. Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and the spectra are labelled accordingly. Peptide sequence ions that have not been labelled in the spectra are denoted with an "*". In the spectra the ion marked with a "Δ" at *m/z* 487.122 corresponds to the sulfated oxonium ion $HexNAc_2Sulf_1$. Labelling of radical and even-electron z-ions follows that set out in **Figure 6-6**.

6.4.5 Modification of the peptide potion of glycopeptides containing N70

The peptide portion of glycopeptides containing N70 ([69]cNGTDAK[75]) were also observed with a peptide modification that was presumed to be the formation of pyro-carbamidomethyl due to ammonia loss from the N-terminal carbamidomethyl Cys residues (333). This modification is commonly observed on peptides (335) and results in an increased retention time in reversed-phase chromatography systems for peptides (336). Fragmentation of glycopeptides presumed to contain pyro-carbamidomethyl Cys with HCD, produced Y1 ions that were 17.027 Da less than the expected Y1 ion. Furthermore, glycopeptides presumed to contain pyro-carbamidomethyl Cys eluted several minutes later than glycopeptides that contained carbamidomethyl Cys (Appendix F), consistent with previous observations of increased retention time of glycopeptides containing pyro-carbamidomethyl Cys (337). Fragmentation of one of these later eluting glycopeptides with EThcD confirmed that the C-terminal amino acids (GTDAK) were not modified (**Figure 6-11**) while the presence of the Y0 and Y1 ions at *m/z* of 748.2952 and 951.370, respectively, indicated that the loss of ammonia was from the peptide. Given that Asn was already modified with an N-linked glycan the loss of 17.027 Da was likely associated with the N-terminal carbamidomethyl Cys residue.
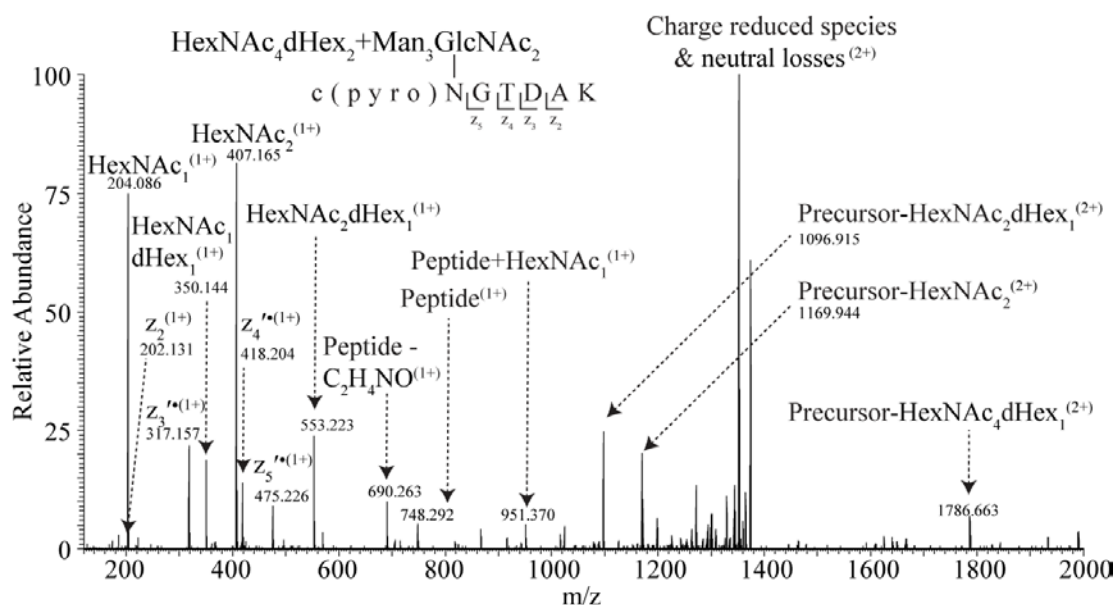
Figure 6-11. EThcD spectrum of an N-linked glycopeptide from hRSV sF containing site N70 with the peptide modification pyro-carbamidomethyl Cys . EThcD fragmentation of a precursor ion at *m/z* 915.6874 (3+). The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide and the spectrum is labelled accordingly. Within the amino acid sequence "c(pyro)" represents pyro-carbamidomethyl. Labelling of radical and even-electron z-ions follows that set out in **Figure 6-6**.

6.4.6 Qualitative distribution of the monosaccharide compositions at N-linked sites of sF

Considering each monosaccharide compositions only once at each N-linked site resulted in 20, 19, 7, 24 and 70 glycans identified at N27, N70, N116, N126 and N500, respectively (Supplementary Table S6-6). All glycans were classified as hybrid or complex except for two compositions, one paucimannose and one high mannose, identified at N500. All assigned hybrid or complex glycans contained at least one dHex residue. Ions corresponding to peptide+HexNAc$_1$dHex$_1$, indicating core fucosylation, were observed in HCD spectra for all hybrid or complex glycopeptides containing N70, N116, N126 and N500 (**Figures 6-6 to 6-9**). The production of peptide+HexNAc$_1$dHex$_1$ ions was less consistent for glycopeptides containing N27 but core fucosylation was apparent in 75% of spectra of assigned glycopeptides.

Inspection of the monosaccharide compositions of glycans observed across hRSV sF indicated that some contained diHexNAc extensions with variable sulfation, fucosylation or sialylation, rather than the more common HexNAc$_1$Hex$_1$ extensions (**Figure 6-1**). This was evident through the observation of abundant oxonium ions for HexNAc$_2$ at *m/z* of 407.166 in HCD and EThcD spectra of sF glycopeptides (**Figure 6-5** to **6-8**). The HEK cell line used in the present study is known to express β1–4–N-acetylgalactosaminyltransferases (β4GalNAc–Ts) that are responsible for the LacdiNAc motif (338). In this study, glycans exhibiting intense HexNAc$_2$ ions also contained a higher number of HexNAc residues than Hex after the N-linked trimannosylchitobiose core, which is consistent with the addition of LacdiNAc. In addition to HexNAc$_2$ ions, fragmentation of sF glycopeptide containing these unusual monosaccharide compositions exhibited the diagnostic oxonium ions HexNAc$_2$dHex$_1$ at *m/z* of 553.224 and HexNAc$_2$NeuAc$_1$ at *m/z* of 698.261 (**Figure 6-6** and **Figure 6-7**), all of which have been identified previously in MS/MS spectra of glycopeptides with LacdiNAc extensions (339, 340). Despite these observations being consistent with LacdiNAc motifs, MS of intact glycopeptides does not enable the linkages of the glycan residues to be determined, thus, these observations can only be represented as evidence of diHexNAc extensions on hRSV sF. The production and intensities of the sulfated oxonium ion HexNAc$_2$Sulf$_1$ were variable in spectra where monosaccharide compositions indicated diHexNAc additions, as illustrated by a low abundant ion at *m/z* 487.122 (represented by Δ) in a HCD spectrum in **Figure 6-10a** and the presence of a more intense ion at *m/z* 487.122 (represented by Δ) in an EThcD spectrum of the same precursor in **Figure 6-10b**. As illustrated in **Figure 6-12** the relative abundances of HexNAc$_2$ ions were low in MS/MS spectra of glycopeptides where the monosaccharide compositions indicated typical LacNAc antenna. Some compositions exhibited the HexNAc$_2$ ions with varying relative abundances such as those glycopeptides where the attached

glycans contained $HexNAc_3dHex_1$, $HexNAc_3dHex_2$ and $Hex_1HexNAc_3dHex_1$ after the trimannosylchitobiose core (**Figure 6-12** denoted by an "*" next to the composition). The diverse fragmentation patterns observed may be indicative of different glycan structures with the same monosaccharide composition. For example, those with low abundant $HexNAc_2$ ions may have bisecting GlcNAc or typical LacNAc antennas while those with relatively intense $HexNAc_2$ ions may contain terminal diHexNAc residues. For N-linked sites N27, N70, N116, N126 and N500 the percentage of assigned spectra where the relative abundance of the $HexNAc_2$ ion was greater than 10% was 65%, 75%, 100%, 86% and 54%, respectively.

6.4.7 Detection of O-linked glycopeptides from sF

Byonic searches of data obtained from trypsin digested sF using HCD NCEs of 30% and 25±5% identified two potential O-linked glycopeptides (Supplementary Table S6-7). The monosaccharide compositions corresponding to $HexNAc_1Hex_1NeuAc_2$ and $HexNA_1Hex_1$ were attached to the peptide, [88]NAVTELQLLMQSTPATNNR[106], which is situated at the C-terminal region of F_2 and N-terminal to the furin-like cleavage site R109. Fragmentation of the sialylated glycoform with HCD produced the glycopeptide Y0 ion which represents the complete loss of the glycan moiety (**Figure 6-13**). Near complete peptide sequence coverage was obtained and the oxonium ions observed in **Figure 6-13a** supported the proposed monosaccharide compositions. Using EThcD the site of attachment of the sialylated glycan was narrowed down to S99 or T100 through the presence of c_8 and z_5 peptide ions (**Figure 6-13b**). Furthermore, the production of an ion at *m/z* 948.329 in **Figure 6-13b**, which corresponds to $HexNAc_1Hex_1NeuAc_2$, confirmed the composition of the O-linked glycan. Using EThcD the site of attachment of the glycan $HexNAc_1Hex_1$ was localised to T100 though the presence of peptide ions z_5, z_7 and c_{12} denoted with "*" in **Figure 6-14**. The O-linked Byonic search allowed ragged cleavage at the C-terminus of peptides to accommodate glycopeptides containing [88]NAVTELQLLMQSTPATNNR[106] that may have undergone carboxyl-trimming after furin-like cleavage, as was observed in the Mascot searches. The Byonic search did not identify O-linked glycopeptides where the peptide portion was formed from a semi-tryptic cleavage.
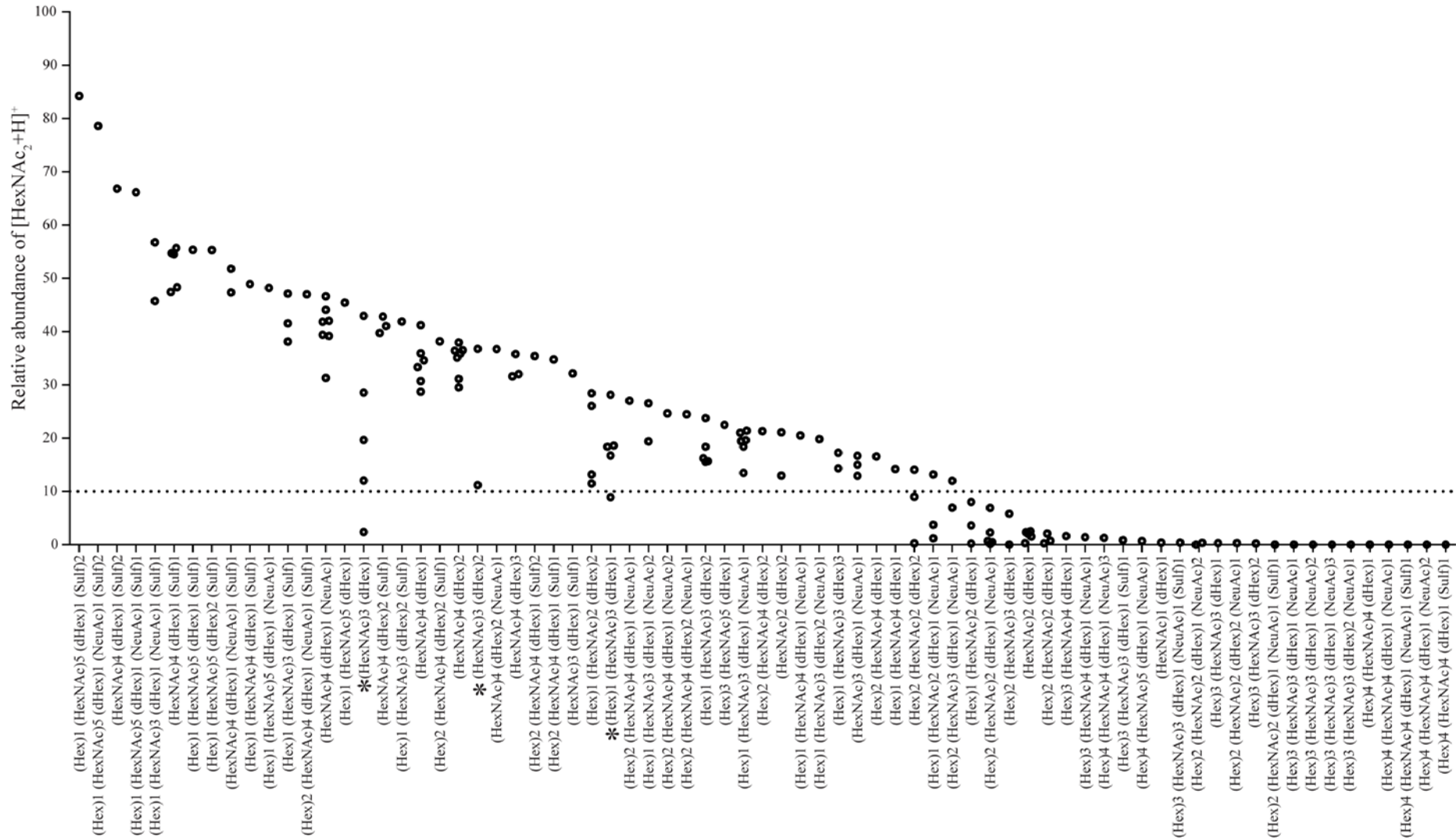
Figure 6-12. Relative abundances of [HexNAc$_2$+H]+ ions identified in MS/MS spectra of all hRSV sF glycopeptides assigned a hybrid or complex glycan by monosaccharide composition. Fragmentation was achieved with 25%±5 NCE. The abundance of the [HexNAc$_2$+H]+ ion is relative to the most intense ion in the MS/MS spectrum. The dotted line represents 10% relative abundance. The non-redundant compositions listed on the x-axis represent monosaccharide residues deduced after the trimannosylchitobiose core for each glycan. Multiple data points for each monosaccharide composition reflect the multiple glycopeptides identified with that composition.
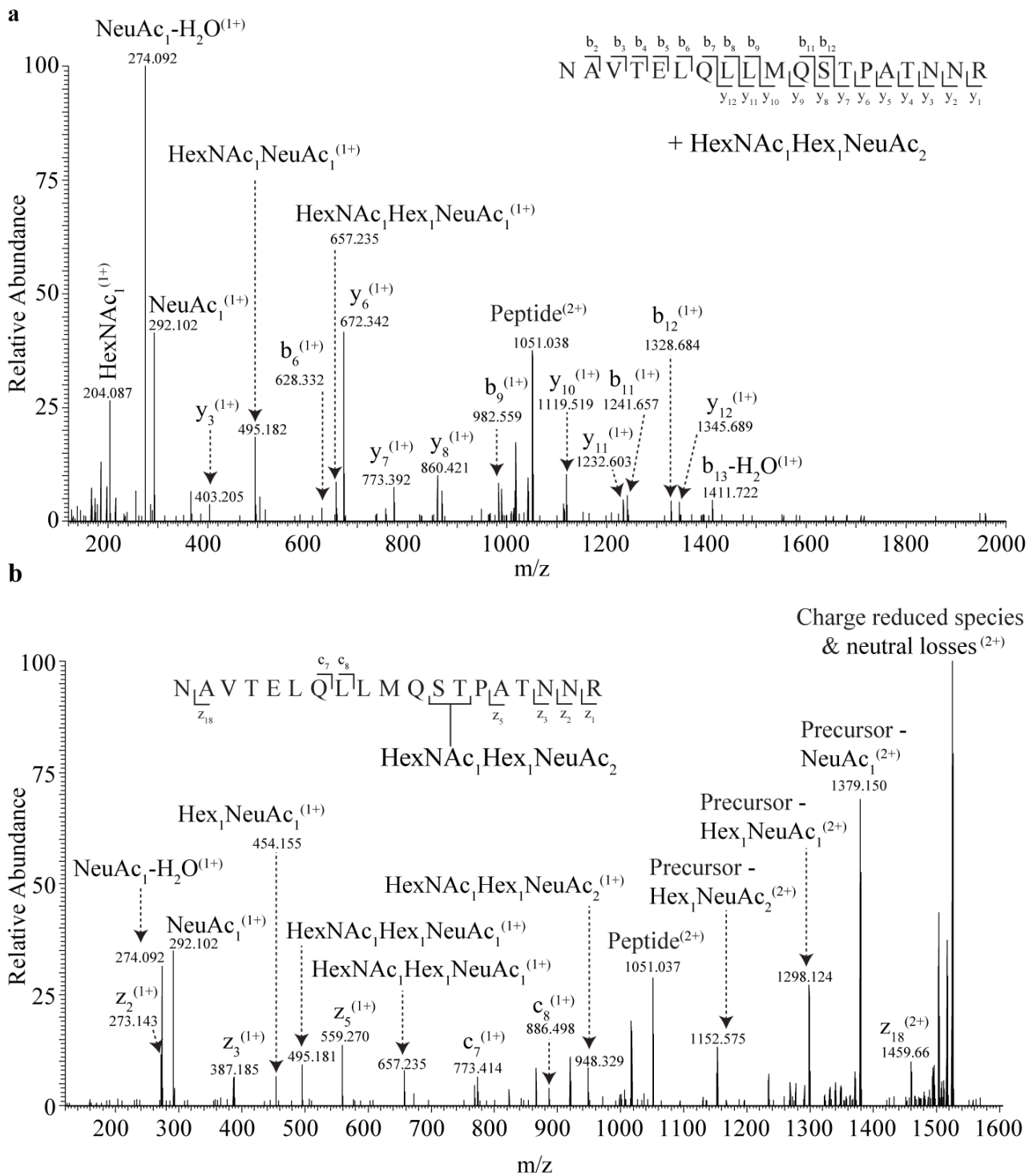
Figure 6-13. HCD and EThcD spectra of a hRSV sF O-linked glycopeptide with HexNAc₁Hex₁NeuAc₂ attached at either S99 or T100. (a) HCD fragmentation using a NCE of 25±5%) and (b) EThcD fragmentation of the same precursor ion at *m/z* 1016.804 (3+). Each panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide. Not all ions have been labelled in the spectra for ease of interpretation.
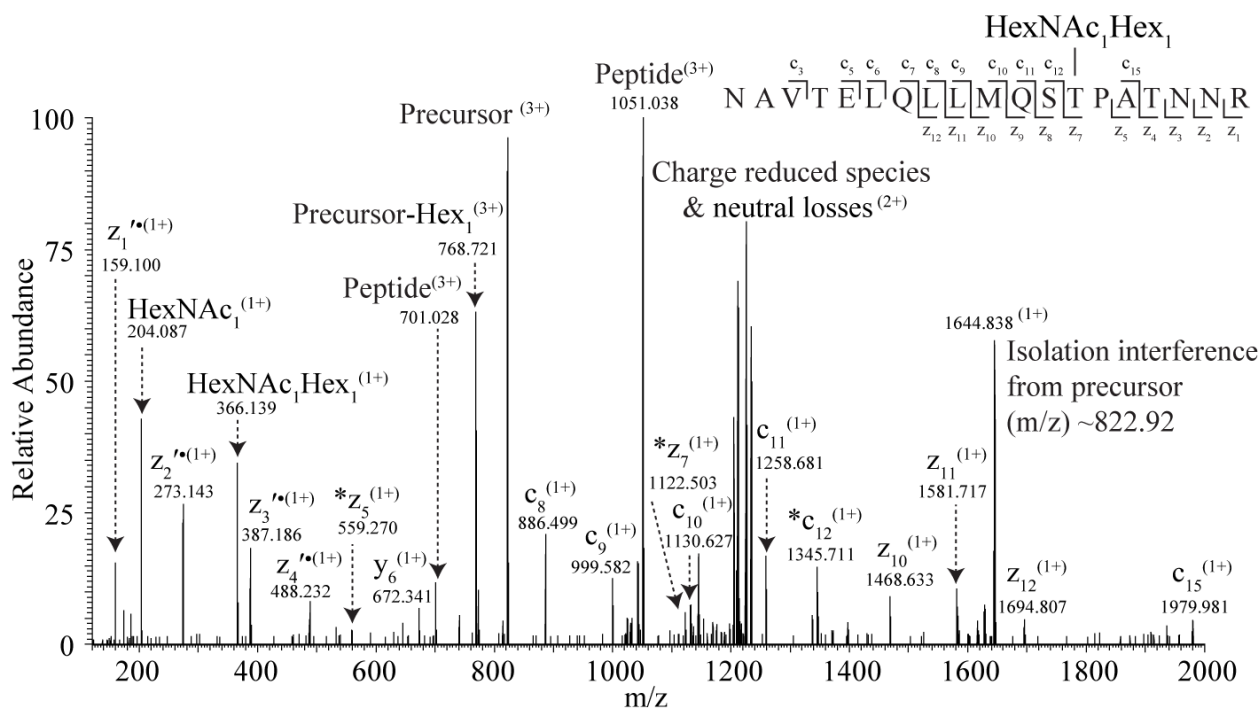
Figure 6-14. EThcD fragmentaion of a hRSV sF O-linked glycopeptide with HexNA$_1$Hex$_1$ attached at T100. EThcD fragmentation of a precursor ion at *m/z* 822.741(3+) is presented. A schematic of the peptide fragmentation pattern is shown. Not all ions have been labelled in the spectra for ease of interpretation. Labelling of radical and even-electron z-ions follows that set out in **Figure 6-6**.

6.5 DISCUSSION

The F protein of hRSV plays an integral role viral infectivity and spread (124, 341) and is a major target for vaccines (58, 114), small molecule antiviral agents and neutralising antibodies (118, 342). The F protein has been characterised structurally (326, 328), however, details regarding the monosaccharide compositions of glycans have not been reported. The methods adopted in this study enabled the first site-specific assessment of glycosylation of hRSV F. The recombinant hRSV F protein analysed in the present study, sF, has been extensively structurally characterized in pursuit of drug and vaccine design objectives (332). This study demonstrated that stepped HCD and EThcD provided data that enabled characterisation of glycopeptides from sF. Using Byonic, stepped HCD enabled the identification of more glycopeptides containing sites N27 and N500 than EThcD. However, EThcD provided valuable peptide fragmentation information for glycopeptides containing all N-linked sites and typically produced more intense peptide ions. Furthermore, only EThcD was able to distinguish that the potential isomeric peptide sequence containing N126 was

[124]KTNVTLSK[131] rather than [125]TNVTLSKK[132]. Importantly these analyses also enabled the identification of N-terminal modifications of glycopeptides containing N27 and N70 produced after removal of the signal peptide and proteolytic cleavage with trypsin, respectively. Such modifications may need to be considered in future analyses of virion derived hRSV F.

Stepped HCD and EThcD also facilitated the identification of sulfated glycopeptides, through the production of sulfated oxonium ions, albeit at lower relative abundances than other oxonium ions. As described in Chapter 4 for NDV HN, this enabled the distinction between Sulf and Phos to be made based on the accuracy of the mass differences between sulfated and non-sulfated oxonium ions (238, 267). Typically Sulf substituents on glycans have been identified in negative ion mode after the glycans have been released from a protein and analysed with or without derivatisation (343). However, site-specificity is lost when releasing the glycan. In positive ion mode, ion-pairing reagents can be used to stabilise labile Sulf substituents on glycopeptides before direct infusion into the mass spectrometer (344). The use of lower dissociation energies in positive ion mode has been used for the characterisation of enriched O-linked glycopeptides containing sulfated glycosaminoglycan chains (238, 267). The present study complements this research by showing that stepped HCD and EThcD can be used to produce sulfated oxonium ions from N-linked glycopeptides in positive ion mode under typical LC-MS conditions.

The observation of non-glycosylated peptides containing N70, N116, N126 and N500 in the trypsin digest of sF suggests these N-linked sites are not always occupied. Conversely, peptides containing N27 were only observed in the trypsin/PNGase F digest, suggesting site N27 is highly occupied. Site directed mutagenesis of hRSV F indicated only sites N27, N70 and N500 were occupied (139) while other work suggests that N116 or N126 in pep27 may be glycosylated (128). The work presented herein reveals that sites N116 and N126 can be occupied. Furthermore, presence of mature glycans at N116 and N126 indicates *trans*-Golgi processing which aligns with predictions in the literature that furin-like cleavage and potential removal of pep27 takes place in the late Golgi (314, 330).

Of significance is the observation of diagnostic diHexNAc oxonium in the analyses of sF, which have been described previously in spectra of glycopeptides with LacdiNAc extensions (339, 340). Glycans with LacdiNAc extensions have been identified in kidney derived cell lines such as the HEK293 cells used in this work (345, 346), however, further studies are required to determine the

linkages of the diHexNAc residues observed on the glycans of hRSV sF. The unique LacdiNAc terminal modification has only been identified on a relatively small number of mammalian proteins, with variable sulfation, α1–3 fucosylation and α2–6 sialylation (339, 345-353). These include native proteins secreted from the stomach (Trefoil factor 2) (339, 345), salivary glands (carbonic anhydrase VI / CA6) (346), placenta (human chorionic gonadotropin / hCG) (348), anterior pituitary (luteinizing hormone, thyroid stimulating hormone and prolactin-like hormones) (347, 349-351), monolayer of the cerebellum (tenascin-R) (352), kidneys and neurons in the brain (sorting protein-related receptor / SORL1/LR11) (353). The functional importance of LacdiNAc extensions is not well established, but the modification has been implicated in regulating serum concentrations of hormones, cell recognition and contraceptive and immunosuppressive activities (331, 345, 354). The two enzymes responsible for LacdiNAc motifs, β4GalNAc-T3 and -T4, are expressed in limited tissues and cell lines (338). The human β4GalNAc-Ts and β1–4–galactosyltransferases (β4Gal-T) responsible for LacdiNAc and LacNAc additions, respectively, are thought to be within the same range of protein expression in HEK293 cells (345) (deduced from the Model Organism Protein Expression database). Despite this, β4GalNAc additions are not highly represented in glycoproteomic and glycomic studies of proteins from HEK293 cell lines, with sulfated β4GalNAc motifs particularly underrepresented (23, 345, 355).

Unlike β4Gal-Ts, the catalytic efficiencies of β4GalNAc-Ts are dependent on the protein, indicating that specific protein recognition determinants drive LacdiNAc addition (356). Native mammalian proteins that contain glycans with LacdiNAc also show a similar glycosylation profile when produced in kidney cell lines (345-347). Several studies suggest that protein sequences proximal to putative N-linked sites act as *cis*-regulatory elements for β4-specific GalNAc-transferases (345, 346, 351, 353). One such study identified the 19 amino acid peptide LRRFIEQKITKRKKEKYMP displaying an alpha-helical structure at carboxyl-terminus of CA6 (346). This 19-residue peptide promoted LacdiNAc extensions on CA6. Furthermore, the addition of the 19-residue peptide to the C-terminus of a protein not previously modified with LacdiNAc directed β4GalNAc-T3 and -T4 activity. Moreover, the recognition determinant PLRSKK situated N-terminal to two N-linked glycosylation sites in the alpha subunit of hCG was shown to promote β4GalNAc-T activity (348). Additional investigations have revealed that specific amino acid residues, particularly the basic residues highlighted in bold within the sequences Q**K**IT**KRKK**E**KY**MP and PL**RSKK**, are important for the induction of β4GalNAc-T3 and -T4 activity (351). In addition, two basic sequences, **K**PL**RRKR** and **K**TVF**KRR**, the former of which is a furin cleavage site, mediate GalNAc additions on SorLA/LR11 (353).

The presence of diHexNAc motifs on paramyxovirus fusion proteins has not been described previously. As discussed in Chapter 5, glycans released from Sendai F proteins derived from virions propagated in chicken eggs revealed high mannose and complex glycans with LacNAc additions (298), while glycans released from NDV F isolated from virions propagated in MDBK cells revealed high mannose structures only (104). Unlike the F protein of other paramyxoviruses and other viral type I fusion proteins hRSV F is cleaved at two sites by furin-like proteases. This is of particular interest, as cleavage of hRSV F at R109 and R136 would result in C-terminal regions of F_2 and pep27 that have clusters of basic residues. In particular, cleavage at R136 occurs at the C-terminus of the highly basic sequence (^{123}**KKT**<u>NVT</u>L**SKKRKRR**^{136}). The propensity of these C-terminal regions, before or after cleavage, to form alpha helices has not been defined in crystallography studies. It has been suggested that cleavage of F_0 into F_1 and F_2 subunits occurs simultaneously with maturation of the glycans and trimerisation of the F protein in the medial-Golgi (314, 357). If the timing of this cleavage is accurate the C-terminal region of F_2 containing pep27 (^{123}**KKT**<u>NVT</u>LS**KKRKRR**^{136}) may indeed act as a cis-regulatory element for β4GalNAc-T in the trans-Golgi and promote GalNAc additions at the N-linked sites N70, N116 and N126. This was particularly evident at N116 where all compositions observed exhibited characteristics of diHexNAc additions. It has been suggested that *cis*-regulatory elements can induce β4GalNAc-T activity on distant N-linked sites and this may be dependent on the quaternary structure of the protein and the proximity of the glycan substrates to the regulatory element (356). This may explain why diHexNAc units were observed at site N27 and N500, as they are not directly adjacent to the two furin-like cleavage sites (104).

Although LacdiNAc motifs have been observed on human glycoproteins the number and location of these glycoproteins is quite limited and expression of such glycans in the lungs may elicit different cellular or immune responses. Interestingly, LacdiNAc motifs have been well established as immunogenic components of helminths and are thought to be responsible for the Th2 bias with eosinophilic granulomas observed in schistosomiasis (358). This type of immune response is also observed in enhanced hRSV disease (116). Antibodies to LacdiNAc and fucosylated LacdiNAc have been identified in sera from humans, monkeys and mice infected with *Schistosoma mansoni*, one of the parasites responsible for schistosomiasis (359, 360). The glycans on these helminths are thought to interact with Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN), a C-type lectin that is expressed on macrophages and dendritic cells (358).

Moreover, monoclonal antibodies to fucosylated LacdiNAc inhibited binding of DC-SIGN to the LacdiNAc antigen expressed from eggs of *Schistosoma mansoni* (361).

Given the immunogenic properties of LacdiNAc glycans it would be of interest to determine if tissues naturally infected by hRSV express β4GalNAc-Ts. Glycomic characterisation of human bronchial and lung tissue has revealed high mannose or complex-type glycans bearing LacNAc extensions (38, 362). Although glycans bearing LacdiNAc were not detected in bronchial and lung tissue, β4GalNAc-T3 is expressed in tracheal tissue, albeit at low levels compared to other tissues such as those derived from ovaries, stomach and brain (338). The transferases β4GalNAc-T3 and -T4 were not expressed or were expressed at very low level in adult lung. However, β4GalNAc-T4 was highly expressed in foetal lung and was expressed in lung squamous and lung adenocarcinoma cell lines at levels equal to or greater than β4GalNAc-T3 and -T4 in HEK293 cells (338). Of interest is the high level of expression of β4GalNAc-T4 in foetal lung tissue compared to adult, as the highest incidence of serious hRSV disease occurs in infants under six months of age (363). Structural predictions show N70 is positioned within a major antigenic site (Ø) in the prefusion form of the protein (118, 332) (**Figure 6-15**). As such, antibodies against the glycans at this site might neutralise the virus effectively. Qualitatively, approximately 75% of the glycopeptides assigned to N70 exhibited fragmentation patterns of diHexNAc units, thus the glycans at this site may modulate immune responses to hRSV. Furthermore, if β4GalNAc-Ts are expressed in the lungs of infants, as potentially indicated by high expression of β4GalNAc-T4 in foetal lung tissue (338), but not of children or adults, this may result in distinct antigenic forms of hRSV virions circulating. Also of interest, is that the first attempt at an inactivated hRSV vaccine resulted in significant enhanced lung pathology in infants upon infection with wild type hRSV. The vaccine was developed after repeated passage of the virus through two kidney cell cultures before intramuscular administration (117). Although purely conjectural, it is possible that production of vaccines or therapeutics or natural infection with hRSV produces these immune stimulating glycan ligands on F and these may contribute to priming Th2 immune responses.
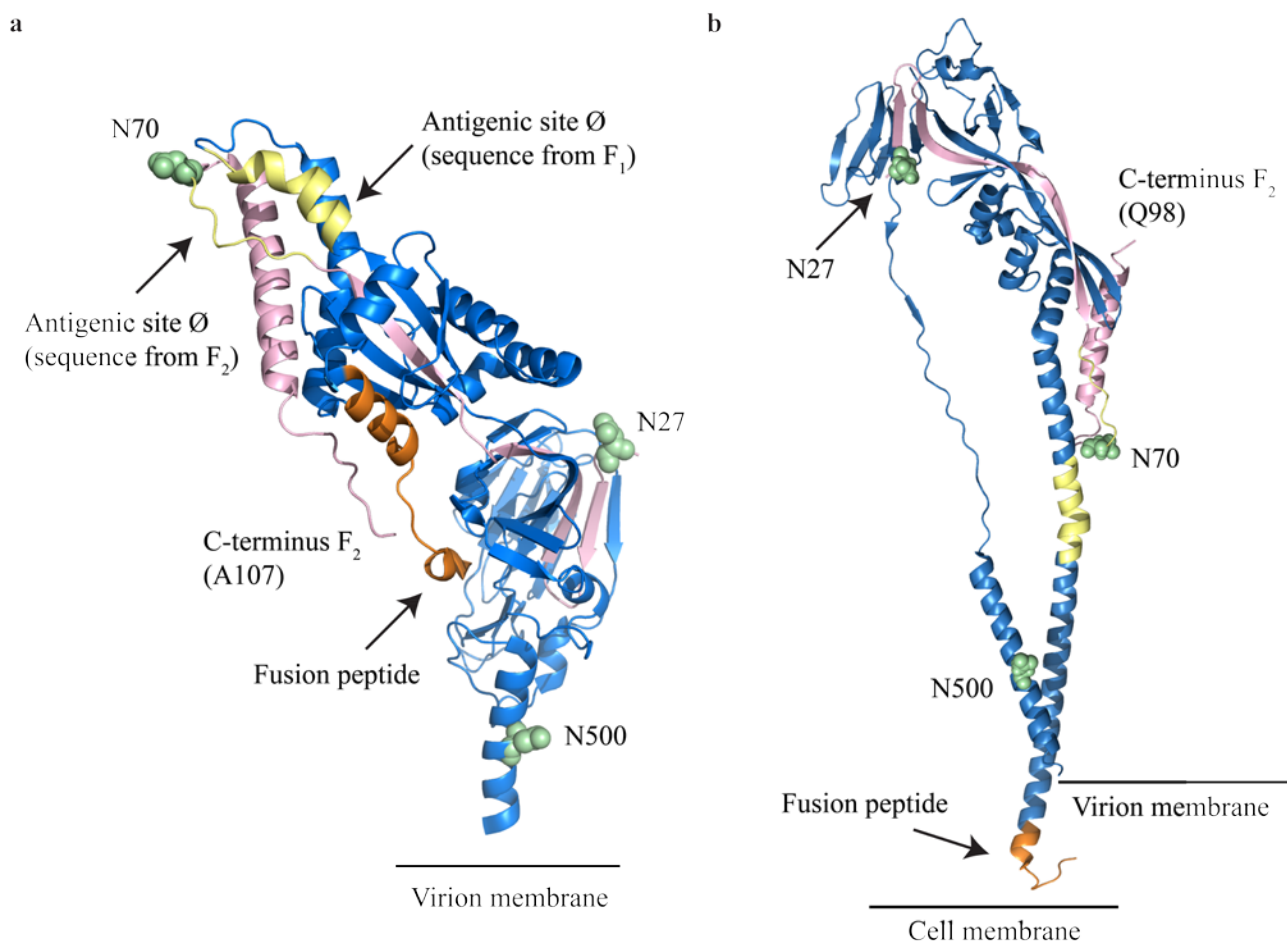
Figure 6-15. Crystal structures of hRSV F in prefusion and post-fusion conformations revealing a major antigenic site (Ø) in the prefusion form. Cartoon representation of monomers of hRSV F where the F_2 and F_1 subunits have been identified in pink and blue, respectively, with the exception of the fusion peptide (orange) and antigenic Ø site (yellow). N-linked sites are represented by green spheres. (a) Prefusion conformation of hRSV F (RCSB PDB identifier 4MMU) revealing F_2 (aa 26-107) and F_1 subunits (aa 137-509). The Ø sites are formed at the apex of a trimer and are composed of two sequences from the monomer, one from F_1 (aa 196-210) and one from F_2 (aa 62-69) (118). Within the structure N70 sits directly adjacent to the Ø site at the apex. (b) Post-fusion conformation of RSV F (RCSB PDB identifier 3RRR) revealing F_2 (aa 26-98) and F_1 subunits (AA147-519). The Ø site is no longer formed after structural rearrangement of F. All images were created PyMOL (version 1.3).

The novel identification of two O-linked glycans on sF was unexpected as hRSV F is thought to be subjected only to N-linked glycosylation (140). The site of attachment, T100, could be confirmed for the $HexNAc_1Hex_1$ O-linked glycan, but could not be distinguished between S99 and T100 for the $HexNAc_1Hex_1NeuAc_2$ O-linked glycan. The proximity of S99 and T100 to the furin cleavage site (R109) is interesting as O-GalNAc glycosylation has been implicated in the regulation of proprotein convertases (33), a family of proteolytic enzymes, including furin, utilised by enveloped viruses in host cells (364). As cleavage of the furin like-sites and O-linked glycosylation are thought

to occur in the later stages of protein processing in the Golgi it is difficult to speculate on a potential role of the O-linked site. Furthermore, the importance of O-linked glycosylation sites observed on recombinantly produced viral proteins is unclear. For example, there is conflicting evidence as to whether O-linked glycans observed on recombinantly produced human immunodeficiency virus gp120 are also present on virally produced gp120 (365, 366). To add to this, it may prove difficult to detect glycoforms on F at O-linked sites in this region of the protein due to carboxyl-trimming. As noted, the potential O-linked sites reside within a tryptic peptide that is directly N-terminal to the R109 furin cleavage site. In the present study, peptides containing the O-linked sites were observed in non-glycosylated forms with multiple C-terminal semi-tryptic cleavages. This is likely due to carboxyl-trimming of F_2 as indicated in the crystal structure of hRSV sF where the C-terminus of F_2 was identified as A107 rather than R109 (332). The use of alternative proteolytic enzymes may be necessary to avoid "diluting" potential signals from glycopeptides containing S99 and T100 due to carboxyl-trimming.

The present study provides as template for future analyses of glycopeptides derived from native hRSV F, highlighting potential pitfalls of using higher dissociation energies for detection of glycopeptides containing N27, and the benefits of using stepped NCEs and EThcD for the identification of glycopeptides containing all five sites. Furthermore, the identification of diHexNAc extensions that may be potential LacdiNAc antennas on N-linked glycans invokes several hypotheses and potential avenues for future studies. A hypothesis has been put forward that the basic residues proximal to the two furin-like cleavage sites of F act as *cis*-regulatory elements and induce the transferases responsible for LacdiNAc additions. These extensions may modulate immune responses upon infection with hRSV or after administration of therapeutics containing hRSV F. The novel observation of an O-linked glycan on hRSV sF also provides a basis for future functional and biological studies. Finally, the results presented herein may be used to compare glycosylation of hRSV F proteins produced recombinantly in different cell lines, virally, or those isolated from virions derived by natural infection.

Chapter 7: Characterisation of glycosylation of human respiratory syncytial virus attachment surface glycoprotein (G)

7.1   SUMMARY

Both N- and O-linked glycosylation of hRSV G is well documented with a considerable portion of the mass of G credited to O-linked glycosylation. Moreover, glycosylation of hRSV G has been implicated in viral infection, immune evasion and modulation of host immune responses. However, site-specific glycosylation of hRSV G and the contribution of such glycans to infectivity remain to be elucidated. Recently, advances in mass spectrometry-based technologies have enabled characterisation of glycans on heavily glycosylated proteins in a site-specific manner. Such technologies have been applied herein to a recombinant soluble form of hRSV G attached to the N-terminus of HA from measles virus (MeV). Both HCD and EThcD methods were employed to address the lack of site-specific detail pertaining to hRSV G. A total of 28 O-linked sites were identified on hRSV G. The O-linked glycans observed were small glycans with the compositions presumed to be Tn, T and mono- and di-sialylated T antigens. PNGase F digestion revealed that N-linked sites N85, N135 and N237 may be occupied, however this could not be confirmed for sites N85 and N237 through the observation of attached monosaccharide compositions. Twelve N-linked compositions were localised to site N135 and represented hybrid or complex-type compositions with core fucosylation. Interestingly, O-linked glycans were also observed on HA from MeV, which has not been previously described as O-glycosylated. The present study also highlights the benefits of EThcD, which enabled up to five O-liked glycans to be assigned to one peptide sequence in a site-specific manner. The combination of Glu-C and trypsin digestions proved useful by increasing the number of glycopeptides detected, particularly form the C-terminal region of the protein. This work is the first to describe site-specific O-linked and N-linked glycosylation of hRSV G. The techniques applied in herein can be used to investigate native forms G while the results provide the first step in the elucidation of site-specific and compositional differences of glycans from hRSV G produced in different cell lines.

7.2   INTRODUCTION

Like the other attachment proteins of paramyxoviruses hRSV G is a type II integral membrane protein containing cytoplasmic and transmembrane domains at the amino-terminal region of the protein. However, the G protein of hRSV does not follow the same molecular architecture as the other attachment glycoproteins from *Paramyxovirinae* (63). Overall G is not well conserved with the exception of the amino-terminal domains and a CCD which is flanked by two mucin like domains (120). The CCD contains four Cys residues which are disulfide-linked to form a cystine noose (121, 367). The two downstream Cys residues form a highly conserved CX3C motif that may bind to the chemokine receptor CX3CR1 in human airway epithelial cells (125, 368, 369). Surface GAG have also been described as a receptor for hRSV G with one site of attachment directly C-terminal to the CCD, described as the heparin-binding domain (HBD) (73, 123, 124, 370). Despite the attachment function of G, the protein is not essential for viral infectivity or replication in cultured cells (72, 370). However, it is required for efficient infection *in-vivo* and for infection of primary well-differentiated human airway epithelial cells (73). A recent comparison of hRSV strains predicted that the two mucin-like domains are intrinsically disordered (371). Within these mucin-like domains it is expected that 25-40 of the potential O-linked sites are utilised (115, 118). The extracellular domain of hRSV G, consisting of the mucin-like domains and CCD, can also be produced as a soluble form and may act as an antibody decoy or to supress host antiviral immune responses (122, 372).

Importantly, glycosylation of hRSV G can modulate viral activity through changes in infectivity, reactivity to antibodies and induction of Th2 responses (50, 137, 138, 140). Studies using endoglycosidases, inhibitors of protein transport or glycosylation and gel electrophoresis have estimated that over half of the MW of the mature form of G can be attributed to glycosylation (133-136, 140, 373). Most of this additional mass is thought to be associated with O-linked glycosylation in the STP-rich mucin-like domains (133, 136, 140, 373). Glycosylation at one or more of the four conserved N-linked sites with complex-type glycans has also been predicted (133, 136, 140, 373). Infection of different cell lines with hRSV has revealed distinct migration patterns of G and differential binding of G to carbohydrate-specific lectins and antibodies, highlighting the importance of elucidating cell-specific differences in glycans and site-occupancy of hRSV G (50, 373).

Despite the numerous studies investigating glycosylation of hRSV G, to date, no study has defined site-specific glycosylation of G. This remains an important aim given recent publications on herpesviruses which revealed conservation of O-linked sites on homologous proteins between viruses and importantly, conservation of sites between viral proteins produced *in-vitro* and from a clinical specimen (374). Clusters of O-linked sites were also seen to occupy regions of functional importance such as sites of proteolytic cleavage, protein-protein interactions, immune evasion and interactions with GAG (374). Furthermore, O-linked sites observed on recombinantly produced HeV G (375) have recently been shown to have functional properties (376). Mutation of individual O-linked sites on both HeV and NiV G proteins altered interactions with F proteins, fusion activity and induced conformational changes in G (376). It should be noted that although the HeV, NiV and hRSV attachment glycoproteins are all termed "G", the HeV and NiV attachment proteins resemble that of other members of the *Paramyxovirinae* subfamily with respect to secondary structure and three-dimensional arrangement (63). Furthermore, the predicted level of O-linked glycosylation is much greater for hRSV G (133, 375, 377). Therefore, the roles defined for the O-linked sites of HeV and NiV G may differ to those observed for hRSV G.

Taken together, the above studies highlight the importance of defining site-specific glycosylation of hRSV G. The type of O-linked glycosylation expected on hRSV, that is mucin-like glycosylation, is initiated by transfer of GalNAc to Ser or Thr residues in the Golgi by a large family ppGalNAcTs (24). The Tn antigen described in (**Figure 1-4**) is built on to form one of eight different cores. The core monosaccharides can also be extended, forming long branching structures that often contain Gal, NeuAc and Fuc residues (378). As described in Chapter 6 for N-linked glycans, GalNAc can also be added to terminal GlcNAc on a core-2 structures to form O-linked LacdiNAc motifs, which have been described on a limited number of proteins (379). Sites that are modified by O-linked glycosylation and the compositions of attached glycans are less random than once thought, with diverse acceptor substrate specificities between ppGalNAcTs (378, 380). Dense O-GalNAc glycosylation serves many functions, including promotion of extended conformations of proteins, shielding proteins or cellular surfaces from proteolysis and modulating immune responses (378, 381).

As mucins contain a high number of O-linked sites, many of which can be clustered, analysing these glycoproteins can be quite challenging. Typically, MS studies employ low-energy electron based fragmentation techniques such as ETD to enable site-specific location of the labile O-linked modifications (374, 382-384). However, multiple glycans attached to one peptide, extended O-

glycans or sialylated structures can result in less efficient fragmentation of the glycopeptides (381, 382, 385). To assess site-specific glycosylation of hRSV G a recombinant soluble form of G (sG) conjugated to the N-terminal region of HA from MeV was analysed. Multiple enzyme digestions were conducted to increase the likelihood of detecting O-linked and N-linked glycopeptides from sG. These included digestion with trypsin and Glu-C to increase amino acid coverage of sG and the identification of glycosylated peptides. Additionally, N-linked glycans were removed with PNGase F to assess occupancy at N-linked sites and to identify O-linked sites on glycopeptides containing N-linked consensus sites. Finally, sialidase treatments were employed to investigate the linkage of sialic acid residues present on sG and to increase the likelihood of assigning N- and O-linked sites after removal of sialic acid residues. Digested sG samples were analysed by MS using HCD and EThcD fragmentation. Overall 28 O-linked sites were unambiguously assigned to Ser or Thr residues in the mucin-like domains of hRSV G. The compositions of O-glycans observed indicated typical mucin-like glycosylation with the structure proposed to be Tn, T, sialyl-T and disialyl-T antigens. Furthermore, N-linked sites N85, N135 and N237 were deemed to be occupied after digestion with PNGase F and occupancy was further confirmed at N135 after the identification of twelve N-linked compositions attached at the site. Three O-linked sites were unambiguously assigned to MeV HA with some O-linked compositions indicative of Core-2 glycans with LacdiNAc additions.

7.3 METHODS

7.3.1 Provision of samples

Recombinant soluble hRSV G, described herein as sG, was kindly provided by Professor Mark. E. Peeples from Nationwide Children's hospital (Columbus, OH, USA). In the construct used to express sG, the amino-terminal cytoplasmic and transmembrane domains of hRSV G were replaced with the amino-terminal coding region of MeV HA (hereafter described as MeV-HA) followed downstream by a furin cleavage site, a six-His tag and a factor Xa site. The final C-terminal sequence of sG contained the extracellular domain of hRSV G (hereafter described as hRSV-G). Construction of sG is described in (386), for the purposes of this work the proteins were expressed in HEK293F cells and purified using wheat germ agglutinin.

7.3.2 Sample preparation and enzymatic digestions of sG

Approximately 150 µg of purified sG was reduced and alkylated before aliquots were methanol precipitated with and without trypsin as per the methods described in Chapter 2. To increase the likelihood of identifying peptides and glycopeptides from sG and in order to investigate the potential linkages of NeuAc on sG aliquots of the trypsin digest were subjected to different enzyme digestions. An aliquot (5 µg) of the trypsin digested sG was further digested with 1U of PNGase F (hereafter referred to as trypsin/PNGase F) in 10 µL of 50 mM NH_4HCO_3. An aliquot (5 µg) of the trypsin digested proteins was also further digested with Glu-C using a final ratio of protein: enzyme, 40:1 (*w/w*) (hereafter referred to as trypsin/Glu-C) in 10 µL of 50 mM NH_4HCO_3. Aliquots (2.5 µg) of the trypsin digest were brought to a volume of ~15 µL with H_2O and heated at 95 ˚C for 5 min to deactivate trypsin before sialidase digestion (32). The aliquots were treated with 5 mU of Sialidase-A or Sialidase-S (hereafter referred to as trypsin/Sialidase-A or -S) in 50 mM sodium phosphate (pH 6.0). The trypsin/PNGase F, trypsin/Glu-C, trypsin/Sialidase-S and trypsin/Sialidase-A digestions were all allowed to proceed for 16 h at 37°C. Resultant peptides and glycopeptides from the trypsin/PNGase F, trypsin/Sialidase-S and trypsin/Sialidase-A digestions were desalted with a C18 ZipTip (10 µL pipette tip with a 0.6 µL resin bed; Millipore, MA, USA) using the manufacturers' guidelines for MS analysis.

7.3.3 SDS-PAGE separation of intact sG treated with sialidase

Methanol precipitated intact sG was also treated with Sialidase-A or Sialidase-S. For each treatment 3.5 µg of sG was digested with 10 mU of sialidase in 50 mM sodium phosphate (pH 6.0). Samples were also prepared for sG (3.5 µg) with no enzyme, Sialidase-A (10 mU) with no protein and Sialidase-S (10 mU) with no protein in the same buffer. All samples were incubated for 16 h at 37°C before separation by SDS-PAGE. Intact sG samples were subjected to SDS-PAGE followed by staining and de-staining as per the methods described in Chapter 2.

7.3.4 Nano-ultra-high pressure liquid chromatography

Approximately 200 ng of digested sG was injected for each analysis using a nUHPLC system as described in Chapter 2. Samples were loaded onto the trap column and washed for 3 min at 5

µL/min in 99% solvent A and 1% solvent B. Peptides and glycopeptides were subsequently eluted onto the analytical column and separated at flow rate of 0.3 µL/min ramping through a sequence of linear gradients from 1% to 2% solvent B in 5 min, to 30% B over 85 min, to 50% B over 30 min, to 95% B in 5 min and then holding at 95% B for 5 min. The column was then re-equilibrated with 1% B for 20 min.

7.3.5 Mass spectrometry data acquisition

Survey scans of precursor ions from m/z 350 to 1800 were acquired in the Orbitrap at 120K resolution (FWHM) at m/z 200 using an AGC target of 400,000 and maximum injection time of 50 ms. For internal mass calibration the lock mass option was enabled using the polycyclodimethylsiloxane ion at m/z 445.1200 (310). The precursor selection priority for fragmentation by HCD-MS/MS was set to highest charge, within a range of 2-8, then highest intensity over 5,000 counts. Precursor ion isolation was performed with a mass selecting quadrupole using an isolation window of m/z 2. Precursors were fragmented in the ion routing multipole, using a NCE of 30%. Previously selected ions within a ±10 ppm window were dynamically excluded for 25 s. Fragment ions were acquired in the Orbitrap at a resolution of 30K using an AGC target of 50,000 and maximum injection time of 60 ms. If fragment ions were produced corresponding to 204.0867 (HexNAc), 138.0550 (HexNAc fragment) or 292.1027 (NeuAc) within a ±10 ppm window the precursors ions were re-isolated and subjected to EThcD using supplemental activation with a NCE of 15%. Fragment ions produced by EThcD were acquired in the Orbitrap at a resolution of 60K using an AGC target of 200,000 and maximum injection time of 250 ms. In total five chromatographic runs were performed (one per sample) using a HCD-pd-EThcD method. These were conducted with the trypsin, trypsin/PNGase F, trypsin/Glu-C, trypsin/Sialidase-S and trypsin/Sialidase-A digests.

7.3.6 Data processing of non-glycosylated and deglycosylated peptides from sG

Proteome Discoverer (v2.1.0.81) and the search engine Mascot were used to search HCD MS/MS spectra from the RAW files of analyses of sG. The protein database contained the sequence for sG. Cleavage specificity was set as semi-tryptic (trypsin and trypsin/PNGase F samples) or trypsin combined with Glu-C (trypsin/Glu-C sample). A maximum of two missed cleavages were allowed. Mass tolerances of 10 ppm and 0.02 Da were applied to precursor and fragment ions, respectively.

Carbamidomethylation of Cys was set as a fixed modification and dynamic modifications included mono-oxidised Met, deamidation of Asn and Gln residues and conversion of N-terminal Gln to pyroglutamate. The "Fixed Value PSM Validator" node was used and a cut-off score of 30 was applied to all PSMs.

7.3.7 Glycan oxonium ion profiles of glycopeptides from sG

To assess the potential monosaccharide compositions of glycans present on sG, the HCD mzML files for the trypsin, trypsin/PNGase, trypsin/Sialidase-A and trypsin/Glu-C digests were analysed with OxoExtract without the N-linked GlycoMod function. The following parameters were used: digestion with trypsin or trypsin/Glu-C; maximum of two missed cleavages; fixed modification of carbamidomethylation of Cys and a dynamic modification of mono-oxidised Met. The protein database queried contained the sG sequence. The oxonium ion profiles were investigated for potential monosaccharides and O-glycans that weren't included in the Byonic O-glycan databases. Sulfate containing oxonium ions were observed and potential compositions were investigated manually in GlycoMod. For searches in GlycoMod the peptide masses were inferred from the theoretical glycopeptide Y0 ion matched in OxoExtract and all monosaccharide residues were considered possible components of a glycan. As a final check, the OxoExtract search results of the trypsin/PNGase F analysis were investigated for O-linked compositions. If more than 100 glycopeptide Y0 ions were observed with the HexNAc ion in spectra from the file, those with more than five peptide b- or y-ions present in a spectrum were investigated. The peptide masses of these putatively O-linked glycopeptides were subtracted from the precursor mass. The mass differences, which were predicted to be the masses of O-linked glycans, were investigated manually in GlycoMod using the "Free Oligosaccharide" option. For this GlycoMod search all monosaccharide residues were considered possible components of a glycan.

7.3.8 Assignment of glycopeptides from sG

Analysing O-linked glycosylation is a computationally intensive task, particularly for sG due to the high content of Ser (total 33) and Thr (total 56) residues, many of which are clustered together in the mucin-like domains of hRSV-G. For this reason Byonic searches were conducted with eight different sets of parameters each targeting different digests of sG, compositions of glycans and number of glycosylation sites allowed per glycopeptide. For the Byonic searches mzML files from

the analyses of the digests with trypsin (abbreviated Tryp in **Table 7-1**), Tryp/PNGase F, Tryp/Glu-C and Tryp/Sialidase-A were used (**Table 7-1**). The HCD and EThcD spectra were searched separately. Mass tolerances of 10 ppm and 0.02 Da were applied to precursor and fragment ions, respectively. Cleavage specificity and the maximum number of missed cleavages allowed for each search are presented in **Table 7-1**. Carbamidomethylation of Cys residues was considered a fixed modification for all searches. The numbers of variable "common" or "rare" modifications allowed per peptide are presented in **Table 7-1**. Variable peptide modifications were set as "common" while variable glycan modifications were set as "rare" or additionally "common" (**Table 7-1**). The protein database queried contained the sG sequence.

For initial Byonic searches, corresponding to numbers 1 and 2 in **Table 7-1**, a large O-linked glycan database was used containing 80 O-linked glycans. This glycan database was a combination of the Byonic database "70 human O-linked" with ten sulfated O-linked glycans identified in the GlycoMod searches. These two initial searches allowed two O-linked glycans per peptide. Once the typical O-linked glycans present on sG had been established, the numbers of O-linked glycans in the databases were reduced and an increased number of O-glycans were allowed per glycopeptide. For searches corresponding to numbers 3 and 4 from **Table 7-1** the O-linked databases contained $HexNAc_1$, $HexNAc_2$, $HexNAc_1Hex_1$, $HexNAc_2Hex_1$, $HexNAc_1Hex_1dHex_1$ (5 O-linked) or additionally $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$ (7 O-linked). These two searches allowed a total of four O-linked glycans per peptide. Search number 5 from **Table 7-1** allowed ragged cleavage at the C-terminus of peptides and was used to investigate O-linked glycopeptides containing the C-terminus of hRSV-G. This search allowed two O-linked glycans per peptide from a small glycan database (7 O-linked). Search numbers 6-8 from **Table 7-1** investigated N-linked glycosylation and the N-glycan database queried was a combination of the Byonic mammalian database (309_Mammalian no sodium) with all glycans containing NeuGc removed. Search number 6 investigated N-linked glycosylation alone while searches 7 and 8 were used to investigate glycopeptides with both N- and O-linked glycosylation.

Table 7-1. Byonic parameters for searches of O-linked and N-linked glycopeptides from hRSV sG

Search	MS files searched	Enzyme specificity	Missed cleavages	Number of modifications allowed	Peptide modifications (Common)	Glycan modification & databases used	Aims of the search: Types of glycosylation &/or positions of the glycans or glycopeptides
1	Tryp/PNGase F	R, K	4	Rare: 2 Common: 2	Deamidated Asn Oxidised Met Pyro-Glu	Rare: 80 O-linked	Up to two O-linked sites per peptide and deamidation of Asn within N-linked consensus sites
2	Tryp/Glu-C	R, K, D, E	2	Rare: 2 Common: 2	Oxidised Met Pyro-Glu	Rare: 80 O-linked (Automatic peptide cut-off disabled)	RSV-G N-terminal peptide (E)[107]ITSQITTILASTTPGVK[123](S) Cysteine noose peptides (K)[150]QNKPPSKPNND[162](F) (E)[167]VFNFVPCSICSNNPTCWAICK[187](R) RSV-G C-terminal peptides (E)[258]LTSQME[263](T) (E)[264]TFHSTSSE[271](G) (E)[272]GNPSPSQVSTTSE[284](Y) (E)[285]YSPSQPSSPPNTPR[297](Q)
3	Tryp/PNGase F (EThcD only)	R, K	4	Rare: 2 Common: 2	Deamidated Asn Oxidised Met Pyro-Glu	Common: 7 O-linked Rare: 7 O-linked	Up to four O-linked glycans per peptide and deamidation of Asn within N-linked consensus sites.
4	Tryp/Sialidase A	R, K	4	Rare: 2 Common: 2	Oxidised Met Pyro-Glu	Common: 5 O-linked (No NeuAc) Rare: 5 O-linked (No NeuAc)	Up to four O-linked glycans per peptide.
5	Tryp/Glu-C	D, E (C-term ragged)	2	Rare: 2 Common: 2	Oxidised Met Pyro-Glu	Rare: 7 O-linked	O-linked glycosylation of RSV-G C-terminal peptide (E)[285]YSPSQPSSPPNTPRQ[298](_)
6	Trypsin	R, K	2	Rare: 1 Common: 2	Oxidised Met Pyro-Glu	Rare: N-linked	N-linked glycans
7	Tryp/Glu-C	R, K, D, E	2	Rare: 2 Common: 2	Oxidised Met Pyro-Glu	Rare: 7 O-linked Common: N-linked	Glycopeptides containing N85, N135, N237 & N251 with N-linked & up to two O-linked glycans attached (K)[85]NTTPTYLTQNPQLGISPSNPSE[106](I) (K)[135]NTTTTQTQPSKPTTK[149](Q) (E)[225]VPTTKPTEEPTINTTK[240](T) (K)[241]TNIITTLLTSNTTGNPE[257](L)
8	Tryp/Sialidase A	R,K	4	Rare: 2 Common: 2	Oxidised Met Pyro-Glu	Rare: 5 O-linked (no NeuAc) Common: N-linked	Glycopeptides with N-linked & up to two O-linked glycans attached

7.4 RESULTS

7.4.1 Linkages of sialic acid on glycans from sG

Purified sG was detected as a major band at ~100 kDa and a minor band at ~50 kDa after separation by SDS-PAGE (**Figure 7-1a**, Lane 2). The mass of sG is calculated to be ~40 kDa, thus the observed apparent mass at ~100 kDa suggests that sG is highly glycosylated or may form dimers. Treatment of sG with Sialidase-S and Sialidase-A decreased the rate of electrophoretic mobility compared to the untreated sample (**Figure 7-1a**, Lanes 3 and 4, respectively). Sialidase-S cleaves α2–3 linked NeuAc while Sialidase-A cleaves α(2–3,6,8,9) linked NeuAc. Inspection of the bands produced after treatment of sG with the sialidases (**Figure 7-1a**, Lanes 3 and 4) revealed that the shift in migration was similar for both treatments. This suggests that the NeuAc present on sG is α2–3 linked as treatment with a α(2–3,6,8,9) specific sialidase did not change the migration profile significantly. However, at the glycopeptide level EICs of the intensities of the NeuAc oxonium ion NeuAc-H_2O, suggested that putatively sialylated glycopeptides were still present after Sialidase-S treatment and that the remaining NeuAc residues were removed after treatment with Sialidase-A treatment (**Figure 7-1**b).
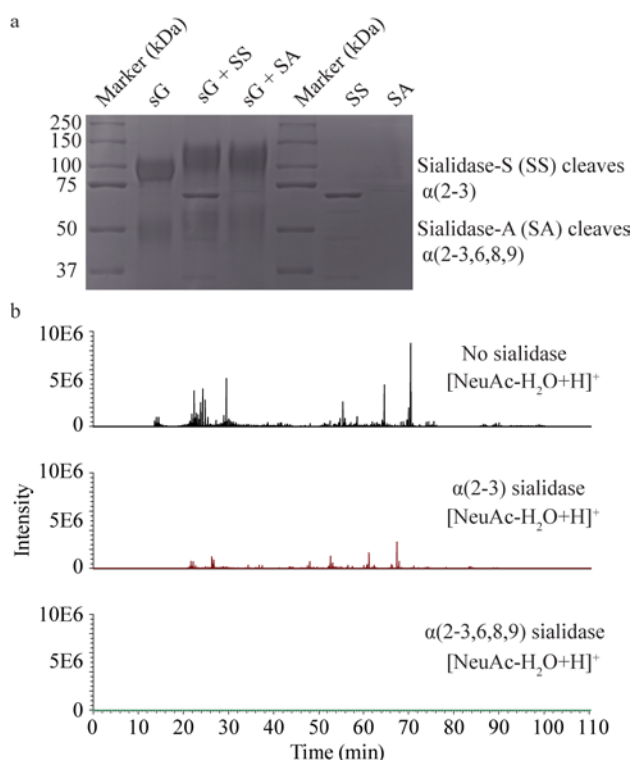
Figure 7-1. Sialidase treatment of intact sG and glycopeptides of sG. (a) SDS-PAGE of sG untreated (Lane 2) and treated with Sialidase-S and Sialidase-A (Lanes 3 and 4, respectively). To identify the contribution of bands by Sialidase-S and Sialidase-A these enzymes were run without sG (Lanes 6 and 7, respectively). (b) EICs of *m/z* 274.0921 after HCD MS/MS analysis of trypsin digested sG untreated (top), treated with Sialidase-S (middle) and Sialidase-A (bottom).

7.4.2 Identification of non-glycosylated and deglycosylated peptides from sG

The sequence of sG (**Figure 7-2**a) contains the first 102 residues from the N-terminal region of MeV HA (UniProt identifier P08362). As described in the methods, this region is followed by a furin cleavage site, six-His tag, a factor Xa site and the extracellular domain of hRSV G. The hRSV-G domain starts with Ala at amino acid position 63 of G from the hRSV A2 strain (UniProt identifier P03423). To enable easy comparisons of the results described herein with those in the literature, the amino acid numbering relating to the hRSV-G portion of sG is described to that of G from hRSV strain A2.

The search engine Mascot was used to assign PSMs from the trypsin, trypsin/PNGase F and trypsin/Glu-C digests of sG (Supplementary Tables S7-1 to S7-3, respectively). Analysis of trypsin digested sG revealed PSMs from the N-terminal region beginning at S48 of MeV-HA with almost complete sequence coverage of the remaining downstream regions of MeV-HA and the His-tag and cleavage sites (**Figure 7-2a**). The Mascot search of the trypsin digest also identified peptides from the hRSV-G region that were predominately from the first mucin-like domain, the CCD and the HBD. The identification of the peptide hRSV-G ^{134}KNTTTQTQPSKPTTK^{149} from the Mascot search of the trypsin/PNGase F digest, resulted in almost complete sequence coverage of the first mucin-like domain. Only one peptide from the second mucin-like domain, hRSV-G ^{213}KDPKPQTTK^{221}, was identified by Mascot searches (**Figure 7-2a**). No additional sequence coverage was obtained from the second mucin-like domain after the analysis of the trypsin/Glu-C digested sample, which was completed with the aim of identifying the peptides ^{258}LTSQME^{263}, ^{264}TFHSTSSE^{271}, ^{272}GNPSPSQVSTTSE^{284} and ^{285}YSPSQPSSPPNTPR^{297}. The lack of peptide coverage from the second mucin-like domain suggested this region was highly glycosylated.

a

Measles virus haemagglutinin (MeV-HA) M1-E102

| Intravirion | Transmembrane |

1　M S P Q R D R I N A　FYK D NPHP K G　S R IVINR E HL　MI D RPYVLLA　VLFVMFL SLI　GLLAIAGIR L　HR AAIYTA E I

71　H K SLSTNL D V　(I)N S IEHQV K D　VL (I) PLF K IIG　DE GK K R K R RG　NSAV D MHHHH　HHGIE G R G

Respiratory syncytial virus attachment G (hRSV-G) A63-Q298

　　　　　　　　　　　　　　　　　　　　　　　　　　AS　ANH K V (I) P (I)(I) A

75　I I Q D A (I) SQI K　NTT PTYLTQN　PQLGISPSNP　S E ITSQITTI　LA S T (I) PGV K S　(I) LQS (I)(I) V K T K　NTT (I)(I) TQTQPS

145　K P (I)(I) K QR Q NK　PP S KPNN D FH　FE VFNFVPCS　ICSNNPTCW A　ICK R IPN K KP　GK K T (I)(I) KP (I) K　KPTL K (I)(I) K K D

215　P K PQ (I)(I) K S K E　VP (I)(I) KP (I) EEP　TIN T T K TNII　TTLLTS NTT G　NP E LTSQME T　FHSTSSE GNP　SPSQVSTTS E

285　Y PSQPSSPPN　TP R Q

b

T76*
S75* S83
S73* T81 T93　　Occupied　　Occupied
　　　　　　　　(PNGase F)　(12 compositions)　　　　　　　　　　　　　　　Occupied
　　　　　　　　　　　　　　　　　　　　　　　C176　C182　　　　　　　　(PNGase F)
　　　　　　　　N85　　　　N135　　　C173 ┌──┐ C186　　　　　N237　　N251

NH$_2$ — [MeV-HA] [ ] Mucin-like domain [HBD] Mucin-like domain — COOH

Legend:
- Measles Virus N-terminus
- His-tag and cleavage sites
- Heparin-binding domain

T70	S117	S124	T136*	S144	S157	T198*	T210	T219	T227	T259*	S267*
T72	T118*	T125	T137	S147		T199	T211	T220	T228	S260*	T268*
T73	T119	S128	T138	T148		T200			T231		
T80		T129				T203			T239		
		T130									

c

MeV-HA (M1-E102)

S73 S75 T76 T81
*HexNAc$_1$Hex$_1$NeuAc$_1$
*HexNAc$_1$Hex$_1$NeuAc$_2$

S73 S75 T76
(s)*HexNAc$_1$Hex$_1$

S73 S75 T76 T81 S83
(H)*HexNAc$_3$Hex$_1$NeuAc$_1$Sulf$_1$
(H)*HexNAc$_2$Hex$_2$NeuAc$_2$Sulf$_1$

T81
HexNAc$_1$Hex$_1$
HexNAc$_3$Hex$_1$
HexNAc$_1$Hex$_1$NeuAc$_1$

T81 S83
*HexNAc$_3$Hex$_1$NeuAc$_1$Sulf$_1$

S83
HexNAc$_1$Hex$_1$NeuAc$_1$

T93
HexNAc$_1$Hex$_1$
HexNAc$_1$Hex$_1$Fuc$_1$
HexNAc$_1$Hex$_1$NeuAc$_2$
(H)HexNAc$_3$Hex$_1$dHex$_1$
(H)HexNAc$_3$Hex$_1$Sulf$_1$
(H)HexNAc$_1$Hex$_1$NeuAc$_1$
(H)HexNAc$_3$Hex$_1$NeuAc$_1$Sulf$_1$

hRSV-G (A63-Q298)

T70
(s)HexNAc$_1$
(s)HexNAc$_1$Hex$_1$

T72
HexNAc$_1$Hex$_1$
HexNAc$_1$Hex$_1$Fuc$_1$
HexNAc$_1$Hex$_1$NeuAc$_1$
HexNAc$_1$Hex$_1$NeuAc$_2$

T72 & T73
*HexNAc$_3$Hex$_1$

T73
(s)HexNAc$_1$Hex$_1$

T80
(s)HexNAc$_1$Hex$_1$

S117
HexNAc$_1$Hex$_1$

T118 & T119
*HexNAc$_1$Hex$_1$NeuAc$_2$

T119
HexNAc$_1$
HexNAc$_1$Hex$_1$

S124
HexNAc$_1$

T125
HexNAc$_1$

S128
HexNAc$_1$
HexNAc$_1$Hex$_1$

T129
HexNAc$_1$
HexNAc$_1$Hex$_1$

T130
HexNAc$_1$
HexNAc$_1$Hex$_1$

T136 & T137
*HexNAc$_1$Hex$_1$

T137
HexNAc$_1$
HexNAc$_1$Hex$_1$

T138
HexNAc$_1$
HexNAc$_1$Hex$_1$

S144
HexNAc$_1$
HexNAc$_1$Hex$_1$
HexNAc$_1$Hex$_1$NeuAc$_1$
HexNAc$_1$Hex$_1$NeuAc$_2$

T147
HexNAc$_1$
HexNAc$_1$Hex$_1$

T148
HexNAc$_1$Hex$_1$

S157
HexNAc$_1$Hex$_1$
(H)HexNAc$_1$Hex$_1$NeuAc$_1$
(H)HexNAc$_1$Hex$_1$NeuAc$_2$

T198 & T199
*HexNAc$_1$Hex$_1$

T199
HexNAc$_1$
(s)HexNAc$_1$Hex$_1$

T200
HexNAc$_1$
HexNAc$_1$Hex$_1$

T203
HexNAc$_1$
HexNAc$_1$Hex$_1$
HexNAc$_1$Hex$_1$Fuc$_1$
HexNAc$_1$Hex$_1$NeuAc$_1$
HexNAc$_1$Hex$_1$NeuAc$_2$

T210
HexNAc$_1$
HexNAc$_1$Hex$_1$

T211
HexNAc$_1$
HexNAc$_1$Hex$_1$

T219
HexNAc$_1$
HexNAc$_1$Hex$_1$

T220
HexNAc$_1$
HexNAc$_1$Hex$_1$

T227
HexNAc$_1$Hex$_1$

T227 & T228
*HexNAc$_1$Hex$_1$NeuAc$_2$

T228
HexNAc$_1$Hex$_1$

T231
HexNAc$_1$
HexNAc$_1$Hex$_1$

T239
HexNAc$_1$
HexNAc$_1$Hex$_1$

T259 & S260
*HexNAc$_1$Hex$_1$
*HexNAc$_2$Hex$_2$
*HexNAc$_2$Hex$_2$NeuAc$_1$

S267 and T268
*HexNAc$_1$Hex$_1$

*Site of attachment was ambiguous
(s)Identified after sialidase digestion
(H)Identified by HCD MS/MS

Figure 7-2. Amino acid sequence coverage and N- and O-linked glycosylation of sG. (a) Amino acid sequence of sG. Potential cleavage sites for trypsin and Glu-C are indicated by a line below and above the relevant residues, respectively. Sequence coverage derived from the Mascot search of sG digested with trypsin is highlighted in light grey. Additional sequence coverage obtained from the trypsin/PNGase F digest is highlighted in dark grey. Residues circled in red were unambiguously assigned as O-linked. Areas of the protein not detected by Mascot or Byonic searches are shown with a clear box. (b) Schematic of sG (not to scale) identifying the MeV-HA region (green), the His-tag and cleavage sites (blue) and the hRSV-G region containing the mucin-like domains (pink) and heparin-binding domain (HBD/purple). Disulfide bonds are represented by connected lines with the amino acid number of the Cys residues which are derived from (121). The N- and O-linked sites are marked with a vertical line and the amino acid number of the expected or observed glycosylated residues. If an O-linked site is marked with an "*" the site of glycosylation was not confidently assigned between two or more residues. The N-linked site N135 was confirmed as occupied after 12 N-linked compositions were sequenced to N135. Sites N85 and N237 were considered to be occupied after PNGase F digestion of tryptic peptides resulted in the identification of peptides or glycopeptides where Asn in the N-linked consensus site was deamidated. (c) Compositions of O-linked glycans observed at each residue.

7.4.3 Glycan oxonium ion profiles of glycopeptides from sG

Sulfated oxonium ions were observed in the all three samples and could be assigned to limited peptides from the MeV-HA region. Ten sulfated O-glycans were added to the Byonic O-glycan database after manual searches using GlycoMod. It was noted that oxonium ions for Fuc ($HexNAc_1dHex_1$, Hex_1dHex_1, $HexNAc_1Hex_1dHex_1$) were mainly present in the trypsin and trypsin/sialidase-A samples but not the trypsin/PNGase F sample. Furthermore, Fuc oxonium ions were predominately observed in spectra that also contained Y1 ions for N-linked glycopeptides, indicating the monosaccharide residue was associated with N-linked glycans. Oxonium ions for Fuc in the trypsin/PNGase F sample could be assigned to limited peptides from the MV-HA and hRSV-G regions of sG suggesting the residue was not highly present on O-linked glycans. The NeuAc oxonium ion NeuAc-H_2O at *m/z* 274.0921 was observed in 20% of MS/MS scans with the HexNAc ion in the trypsin digested sample and 0.1% of scans in the trypsin/sialidase-A sample, indicating NeuAc was removed with the sialidase treatment.

7.4.4 Assignment of O-linked glycopeptides from sG

To assess O-linked glycosylation of sG a number of enzymatic digests and Byonic searches were completed. The fragmentation methods HCD and EThcD were combined in a product dependant manner to provide complementary fragmentation data for the assigned glycopeptides. A summary of the results of all Byonic searches for glycopeptides were combined and the approximate locations of each site of glycosylation on sG have been annotated in **Figure 7-2b** with observed O-linked monosaccharide compositions listed in **Figure 7-2c**. The results of the Byonic and manual searches of the trypsin, trypsin/PNGase F, trypsin/Glu-C and trypsin/Sialidase-A samples are provided in Supplementary Table S7-4 with annotated spectra in Supplementary Figures S7-1 to S7-4, respectively. The results presented in Table S7-4 include information regarding the sample and MS analysis, the Byonic search, the sites of attachment, monosaccharide compositions of the glycans and the outcome of manual validation of each glycopeptide.

Within the MeV-HA portion of sG, O-linked glycans were unambiguously assigned to three sites, T81, S83 and T93 (**Figure 7-2b**). Several of the O-linked compositions assigned to the MeV-HA region were not typical of core-1 (T antigen) glycans which contain the base unit $HexNAc_1Hex_1$. The unusual compositions assigned to the MeV-HA region contained three HexNAc residues, such $HexNAc_3Hex_1$ at MeV-HA T81 and $HexNAc_3Hex_1dHex_1$, $HexNAc_3Hex_1Sulf_1$ and $HexNAc_3Hex_1NeuAc_1Sulf_1$ at MeV-HA T93 (**Figure 7-2c**). Confirmation of sulfation was derived from diagnostic oxonium ions in either the HCD or EThcD scans of the glycopeptides. The peptide ^{90}DVLTPLFK^{97} from the MeV-HA region has only one potential site of O-linked glycosylation, T93, thus fragmentation of this peptide verified the above O-linked compositions could be attached at one site (**Figure 7-3**). Production of the diagnostic oxonium ions $HexNAc_1dHex_1$ at *m/z* 350.145, $HexNAc_1Hex_1$ at *m/z* 366.139 and $HexNAc_2$ at *m/z* 407.166 in **Figure 7-3** further verified the allocated $HexNAc_3Hex_1dHex_1$ O-linked composition.
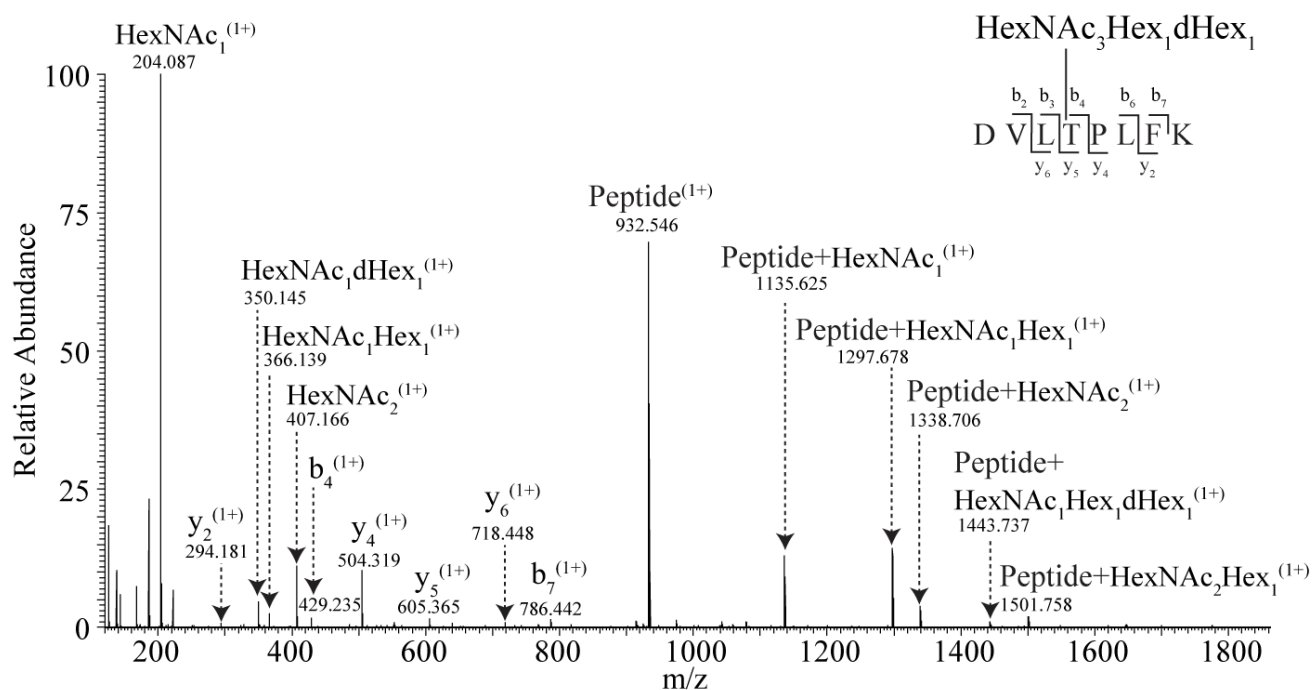
Figure 7-3. HCD fragmentation of an O-linked glycopeptide form the MeV-HA region of sG. HCD fragmentation of a precursor ion at *m/z* 945.452 (2+). The spectrum was obtained after trypsin digestion of sG followed by PNGase F digestion. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (MeV-HA aa 90-97). Not all ions have been labelled on the spectrum for ease of interpretation.

Within the hRSV-G portion of sG O-glycans were unambiguously assigned to 28 sites (**Figure 7-2b**). Although the sites of attachment T259, S260, S267 and T268 at the C-terminus of hRSV-G are labelled ambiguous in **Figure 7-2b**, these residues were observed on two separate peptides [258]LTSQME[263] and [264]TFHSTSSE[271] and thus represent at least two more sites that are occupied. The majority of O-linked compositions observed in the hRSV-G portion were typical of mucin-like glycosylation (**Figure 7-2c**) with the most plausible structures being the Tn antigen and T antigen with additional levels of sialyation. Taken together with the results of the sialidase treatments the observation of compositions $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$ suggest monosialylated core-1 (NeuAcα2–3Galβ1–3GalNAcα1-) and disialylated core-1 (NeuAcα2–3Galβ1–3(NeuAcα2–6)GalNAcα1-) O-linked glycopeptides. Up to five O-linked glycans were assigned to the peptide sequence [124]STLQSTTVK[132] in the first mucin-like domain of hRSV-G (**Figure 7-4**). Fragmentation of the glycopeptide using EThcD enabled $HexNAc_1$ to be sequenced to sites S124, T125, S128, T129 and T130.
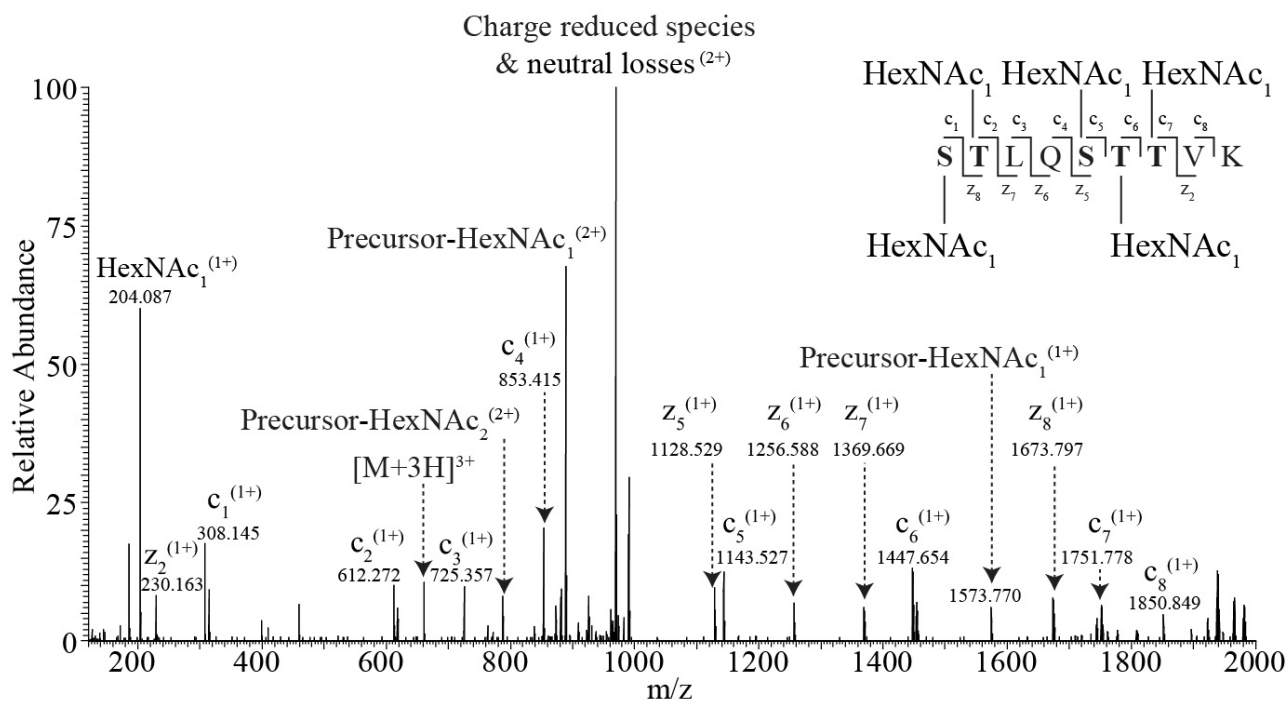
Figure 7-4. EThcD fragmentation of a hRSV-sG glycopeptide with five O-linked glycans attached. HCD fragmentation of a precursor ion at *m/z* 660.6484 (3+) is presented. The spectrum was obtained after trypsin digestion of sG followed by PNGase F digestion. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (hRSV-G aa 124-132) and the spectrum is labelled accordingly.

Up to three glycans could be assigned to single glycopeptides in the second mucin-like domain (Supplementary Table S7-4). Two sites from the hRSV-G region, T72 and T203, were assigned fucosylated O-linked glycans (**Figure 7-2c**). The former site was also assigned an O-linked glycan with three HexNAc residues attached (**Figure 7-5**). In **Figure 7-5**, which reveals the EThcD fragmentation pattern of a glycopeptide containing [69]VTPTTAIIQDATSQIK[84], the production of the oxonium ion $HexNAc_3dHex_1$ at *m/z* 772.299 verified the allocated $HexNAc_3Hex_1$ O-linked composition. The c_3 peptide sequence ion at *m/z* 314.207, which was observed in a radical and prime form, confirmed the glycan was not attached at T70. The z_{12} peptide sequence ion at *m/z* 1171.643 confirmed the O-linked glycan was attached at T72.
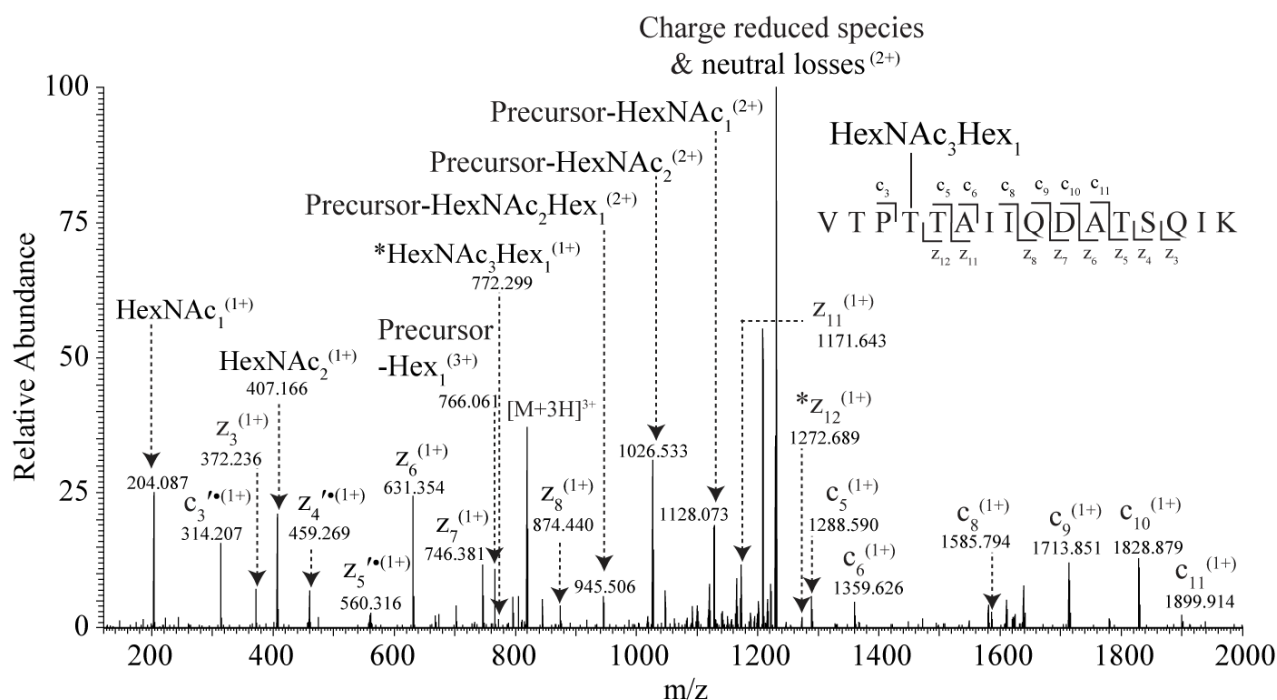
Figure 7-5. EThcD fragmentation of a hRSV-sG glycopeptide with an O-linked composition containing three HexNAc residues localised to T72. EThcD fragmentation of a precursor ion at *m/z* 820.078 (3+) is presented. The spectrum was obtained after trypsin digestion of sG followed by PNGase F digestion. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (hRSV-G aa 69-84) and the spectrum is labelled accordingly. The addition of a dot '·' symbol with the '′' symbol indicates both the radical and z + H ions were observed for the fragment ion as judged by the isotopic distributions described in (176). The remaining z-ions are considered to be typical radical ions as the isotopic distribution of the fragment ions did not enable confirmation of z + H ions.

7.4.5 Occupancy at N-linked sites of sG

The sG protein contains four potential sites of N-linked glycosylation (**Figure 7-2**). The only N-linked site observed without glycosylation was site N85 in the results from Mascot searches of the trypsin and trypsin/Glu-C digests. Sites N85 and N135 were observed in a deamidated form in the Mascot search results of the trypsin/PNGase F digest. Sites N135 and N237 were observed in a deamidated form within O-linked glycopeptides after Byonic searches of the trypsin/PNGase F digest (**Table 7-2 and Figure 7-6**). Byonic searches enabled N-glycans to be localised to N135 but not to N-linked sites N85, N237 and N251 (**Table 7-2**). Byonic searches did not enable the identification of glycopeptides with both N-glycan and O-linked attached.

Table 7-2. Occupancy of N-linked sites from sG

Observation in Mascot or Byonic Searches	N85	N135	N237	N251
Unoccupied	Yes	Not observed	Not observed	Not observed
Occupied: Determined by PNGase F (no attached O-linked)	Yes	Yes	Not observed	Not observed
Occupied: Determined by PNGase F (attached O-linked)	Not observed	Yes	Yes	Not observed
Occupied: Determined by N-glycans (no attached O-linked)	Not observed	Yes	Not observed	Not observed
Occupied: Determined by N-glycans (attached O-linked)	Not observed	No	Not observed	Not observed
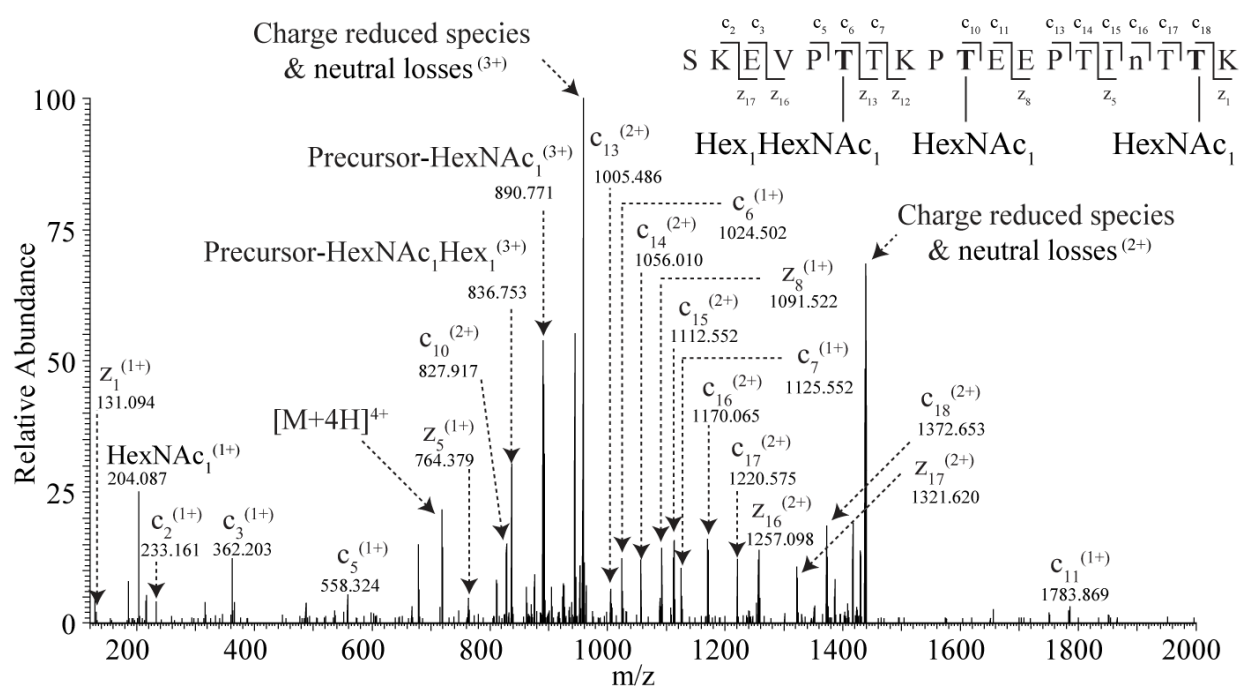
Figure 7-6. EThcD fragmentation of a hRSV-sG O-linked glycopeptide containing N-linked site N237 in a deamidated form. EThcD fragmentation of a precursor ion at *m/z* 719.100 (4+) is presented. The spectrum was obtained after trypsin digestion of sG followed by PNGase F digestion. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (hRSV-G aa 222-240). Not all ions have been labelled on the spectrum for ease of interpretation. Lower case "n" in the schematic represents deamidation of Asn.

7.4.6 Monosaccharide compositions assigned to N-linked sites of sG

As described above N-glycans could only be localised to N135 of sG (**Table 7-2**). The Byonic searches identified twelve N-linked glycans (**Figure 7-7**). The site attachment at N135 in **Figure 7-7** was confirmed through the presence of peptide sequence ions c_2, c_3, z_{14} and z_{15} at *m/z* 247.176, 1805.754, 752.392 and 1531.682, respectively. The HCD fragmentation patterns of the N-linked glycans attached at N135 indicated they all contained core fucosylation. Three of the N-linked compositions attached at N135 contained NeuAc. Interestingly, three of the N-linked compositions observed at site N135 also displayed fragmentation characteristics of diHexNAc similar to those seen in Chapter 6 for hRSV sF.
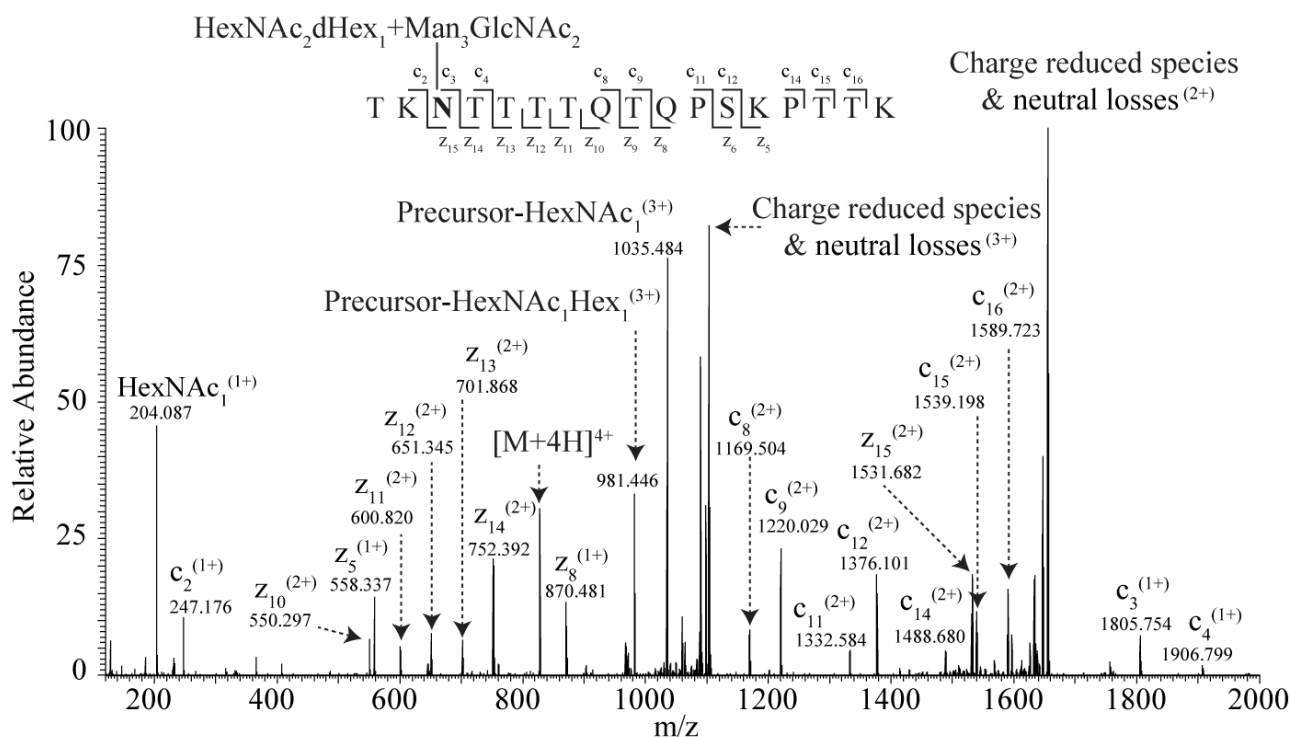
Figure 7-7. EThcD fragmentation of a hRSV-sG glycopeptide containing an N-linked glycan at site N135. EThcD fragmentation of a precursor ion at *m/z* 827.636 (4+) is presented. The spectrum was obtained after trypsin digestion of sG. The panel has a schematic of the peptide fragmentation pattern observed for the glycopeptide (hRSV-G aa 133-149) and the spectrum is labelled accordingly.

7.5   DISCUSSION

Purified sG was detected as a major band at ~100 kDa and a minor band at ~50 kDa after separation by SDS-PAGE. The apparent masses observed are greater than the calculated mass of sG which is ~40 kDa. The electrophoretic mobility of hRSV G produced virally typically corresponds to a mass of ~90 kDa despite a calculated mass of ~32 kDa (136). This mass difference is attributed primarily to O-linked glycosylation (~50%) and a smaller portion to N-linked glycosylation (~15%) (133). The presence of two bands of sG is consistent with previous observations from hRSV G (73, 133, 386, 387). It has been hypothesised that the different species arise from dimer formation, differential glycosylation or significant cleavage of hRSV G (73, 133, 386, 387). Treatment of sG with Sialidase-S and Sialidase-A decreased the migration of both the 100 and 50 kDa bands, an effect that has been observed previously for hRSV G (373). It is probable that the increase in apparent MW was due to a reduction in the intrinsic negative charge of the protein after removal of the anionic NeuAc residues. Before removal, the terminating NeuAc residues on sG may have acted in the same manner as SDS, increasing the migration of the protein towards the anode (388).

Together, the results of the N-linked, O-linked and Mascot searches reveal that within the pool of sG proteins some species may be minimally modified by O-linked glycosylation in the first mucin-like domain. However, the identification of a glycopeptide (^{124}STLQSTTVK^{132}) from the initial mucin-like domain with five occupied O-linked sites reveals that within the pool of sG this region can also be heavily O-linked glycosylated. Glycosylation of the conserved central region containing four Cys residues, specifically peptide hRSV-G ^{167}VFNFVPCSICSNNPTCWAICK^{187}, was not observed. This conforms with observations that the Cys residues form a noose which may prevent ppGalNAcTs from modifying Ser and Thr residues in this region (121). In contrast to the first mucin-like domain of hRSV-G, only one peptide from the second mucin-like domain was observed without glycosylation (^{213}KDPKPQTTK^{221}). This suggests that the C-terminal region of hRSV-G is highly glycosylated.

The Mascot search of tryptic sG identified peptides containing N85 that were present without glycosylation, thus N85 is not always glycosylated. Subsequent treatment of the tryptic peptides with PNGase F produced peptides containing sites N85 and N135 where Asn residues were deamidated. This suggests N85 and N135 are modified by N-linked glycosylation, the latter of which was confirmed after sequencing N-linked glycans to N135 using EThcD. Peptides and O-

linked glycopeptides containing N135 were only observed after PNGase F treatment and all identified peptides contained Asn residues within the N-linked consensus site that were deamidated. This suggests that N135 is substantially occupied or alternatively, that in the absence of N-linked glycosylation a high level of O-linked glycosylation in this region impedes detection and identification of potential glycopeptides by MS. Peptides containing N237 were only observed with O-linked glycosylation and only after PNGase F digestion, again all identified peptides contained Asn residues within the N-linked consensus site that were deamidated. These results conform with those seen in the Mascot searches where minimal sequence coverage was obtained from the second mucin-like domain, revealing that this region of the protein is likely to be highly O-linked glycosylated. The inability to identify O-linked glycopeptides containing N237 without PNGase F digestion suggests that like N135, site N237 may be highly occupied or highly O-linked glycosylated in the absence of N-linked glycosylation. It also suggests deamidation of Asn was not spontaneous and represents occupancy at N237. It was predicted that the trypsin/Glu-C digestion would produce the theoretical peptide [241]TNIITTLLTS<u>NTT</u>GNPE[257] which would enable peptides or glycopeptides containing N251 to be detected. Unfortunately this was not the case, and may be due to a high level of glycosylation in this region. For future studies, combining the proteolytic digestions trypsin/Glu-C with PNGase F may enable identification of deglycosylated versions of peptides or O-linked glycopeptides containing N251.

To date, O-linked glycosylation of MeV HA has not been described. Interestingly, the presence of basic residues such as oligo-His tags are thought to induce GalNAc O-linked glycosylation on sites adjacent to such tags (345). Given the proximity of the glycosylated MeV-HA residues to a basic region of sG containing the furin cleavage site (KKRKRR) and the six-His tag, it is possible O-linked glycosylation of MeV-HA was a result of a peptide recognition motif in sG. Some of the O-linked compositions observed in the MeV-HA region were indicative of core-2 glycans with LacdiNAc motifs (379). The potential extended core-2 glycans from MeV-HA contained three HexNAc residues and one Hex residue and were variably fucosylated, sialylated and sulfated. Interestingly, an envelope protein (E2) of hepatitis C virus (HCV) was produced in HEK293 cells and O-linked compositions with three HexNAc residues were observed with variable sulfation and sialyation (384). The authors surmised that the compositions represented new core-2 configurations. A His-tag was present at the C-terminal region of E2 but the unusual compositions were observed at sites throughout the N-terminal region as well. Thus, it remains to be seen if virally produced MeV HA is O-linked glycosylated or, if singly, or in combination, the furin cleavage site, 6-His tag or mucin-like domains of sG induced O-linked glycosylation of MeV-HA.

The overall importance of O-linked glycosylation for the infectivity of hRSV has been indicated through several studies. Interestingly, removal of O-linked glycans from hRSV virions significantly decreased viral titers, revealing that O-linked glycosylation is required for infection (140). Notably, removal of the O-linked glycans with O-glycosidase did not produce a large shift in the apparent MW of G, suggesting only specific accessible residues were important. To this end, the potential roles of O-linked sites observed in the present work will be discussed. However, it should be emphasised that other O-linked sites may be occupied and may not have been identified in this study for several reasons. These include, but are not limited to, hydrophilic glycopeptides not being retained on the LC-column and glycopeptides with dense areas of glycosylation or large glycans failing to produce adequate signal or fragment ions.

Glycosylation of G may play an important role in priming a pathogenic host immune response as hRSV G produced without O-linked glycosylation or missing regions C-terminal to the CCD reduced Th2 responses in mice (137, 138). Further analyses revealed specific residues [93]KPGKKTTTKPT[203] were responsible for inducing lung eosinophilia but were not required for induction or protective immunity (138). There are four potential sites of O-linked attachment highlighted in bold within these eleven residues [93]KPGKK**TTT**KP**T**[203], three of which were confirmed as occupied in the present study corresponding to T199, T200 and T203. These residues are situated directly C-terminal to the HBD, which has been described as a region of G that binds GAG (73, 123, 124, 370). It has been shown that removal of a cluster of O-linked sites that sit proximal to a GAG-binding domain on the gC attachment protein of herpes simplex virus (HSV) type 1, affects binding of the virus to GAG and reduces the production of new viral particles (389). In view of the above studies, it could be postulated that T199, T200 and T203 of hRSV G serve dual roles, potentially modulating immune responses and also promoting binding to GAG.

Further downstream to the HDB two residues have been implicated as a cleavage site for cathepsin L between residues K209 and T210 (386). In the present study T210 and T211 were also observed to be occupied. A comprehensive analysis of O-linked sites in herpesviruses revealed clusters of O-linked sites in areas susceptible to proteolytic cleavage, prompting the authors to conclude that the O-linked sites served to protect the proteins from proteolytic cleavage. Cleavage of hRSV G, presumed to be between K209 and T210, results in less infectious virions and compromises the attachment function of G (386). Accordingly, the O-linked sites T210 and T211 may play a role in protecting hRSV G from proteolytic cleavage.

In addition to specific sites of O-linked glycosylation, the types of O-linked glycans present on hRSV-G are also of interest. Threonine at position 72 of hRSV-G displayed interesting O-linked glycans. It was one of two sites from hRSV-G where the O-glycans were observed to contain fucosylated glycans and was the only site from hRSV-G that was identified with an O-glycan containing three HexNAc residues. The reason behind this differential glycosylation at T72 is not known, it may reflect altered glycosylation due to its proximity to the basic His-tags and cleavage sites. However, the compositions assigned to T72 and the surrounding residues were typical of the mucin-like glycosylation observed at other sites of hRSV-G. To my knowledge, T72, or this region of G, has not been implicated in any specific mechanistic or biological actions of hRSV G. The remaining site from hRSV-G that contained a fucosylated glycan, T203, has been implicated in an area of G that modulates immune activities, as described above. The differential glycosylation observed at T203 may indicate that this particular site plays an important role in the eleven residues [93]KPGKKTTTKPT[203] thought to induce lung eosinophilia.

The remaining O-linked glycans from the hRSV-G portion of sG were mainly truncated glycans presumed to be Tn and T antigens with a small number of sialyl- and disialyl-T antigens. It should be noted that purification by wheat germ agglutinin may have biased the type of glycans observed in this study. Recombinant HeV G produced in a HEK293 cell line contained O-glycans that were mainly T and sialyl-T antigens (375). Conversely, HCV E2 produced in HEK293 cells contained mainly sialyl-T antigens and sialylated and sulfated core-2 compositions (384). The work presented herein and the studies on HeV and HCV reveal that recombinant viral proteins produced in HEK293 cells display considerably different O-linked glycan profiles. Moreover, the MeV-HA region of sG displayed potential core-2 structures with sulfated substituents revealing that regions of the same protein can be differentially O-glycosylated during production. Truncated glycans, particularly the Tn and T antigens that was observed on hRSV-G have been implicated in a range of aberrant immune responses, including cancer (390). The importance of specific types of O-linked glycosylation in host immune responses has also been defined for viruses. Extended O-linked glycans on HSV induced an innate immune response at the mucosal surface that is independent of the typical antiviral interferon response (391). This mechanism is initiated during the early stages of infection and involves stimulation of the chemokines CXCL9 and CXCL10 and recruitment of neutrophils. Removal of N-linked glycans from HSV virions did not reduce CXCL10 expression, however, removal of O-linked glycans or propagation of virions in a cell line that produces truncated O-glycans decreased CXCL10 expression. Consequently, if the truncated O-glycans

observed on hRSV-G are present on hRSV G *in vivo* it may enable the virus to evade early antiviral activity at the mucosal surface.

The type of glycosylation observed on sG may differ to that of virally produced G in respiratory tissue. Inoculation of human airway epithelium with hRSV produced a form of G that migrated with an apparent mass of 180 kDa (73). This may be due to the different repertoire and expression levels of O-linked transferases in the respiratory tract compared to other tissues or cell lines. For example, bronchial tissue is known to secrete mucins with the less common core-3 and core-4 O-glycans (24). However, G produced virally in tracheal biopsies displayed a similar migration pattern to that of infected HEp-2 cells (~90 kDa) (50). Furthermore, lectin binding studies have shown that viral G produced in HEp-2 cells and colon carcinoma cell lines contains unsubstituted GalNAc and Gal-GalNAc residues (50, 373) with the additional observation of α2–6 linked NeuAc in HEp-2 cells (50). Irrespective of the differences in O-linked glycosylation seen between different cell lines and tissues, the work presented herein reveals that specific residues on hRSV-G can act as substrates for GalNAc transferases and thus may be occupied during viral infection. This provides a foundation for future studies such as those on HeV, where O-linked sites identified on recombinant HeV G were later confirmed to play roles in the activities of HeV and NiV attachment proteins (375, 376).

The present study provides a template for future analyses of glycopeptides derived from hRSV G particularly highlighting the benefits of EThcD and in some instances HCD, combined with multiple enzymatic digestions. This work provides an insight into the type of glycosylation present on hRSV G when produced in a kidney cell line, providing the first step in the elucidation of site-specific and compositional differences of glycans from hRSV G produced in different cell lines. Sites T72 and T203 were observed to contain some monosaccharide compositions that did not correspond to the typical Tn, T and mono- and di-sialylated T antigens and may therefore be of interest in future mutation studies.

Chapter 8: Discussion and conclusions

8.1 SUMMARY

Paramyxoviruses are an important viral family and glycosylation of the surface attachment and F proteins of these viruses has been shown to greatly affect viral replication, infectivity and immunogenicity. Within this family, NDV and hRSV contribute to the significant economical and health burdens of paramyxoviruses worldwide. While the functional, structural and immunological properties of the attachment and F proteins of NDV and hRSV have been explored, there is a paucity of data investigating site-specific glycosylation of these proteins. This body of research therefore seeks to provide the first insight into glycan heterogeneity at N- and O-linked sites of the attachment and F proteins of NDV and hRSV. An overview of the prominent findings from this study are presented below in the context of analysing viral glycoproteins by mass spectrometry and how the described glycosylation patterns of the attachment and F proteins may modulate viral replication and infectivity with potential implications for vaccine or therapeutic design.

8.2 GLYCOSYLATION OF NDV HN AND F

8.2.1 Overview of the results with reference to current literature

In the analysis of glycans from HN, each fragmentation method, HCD, ETD and CID, provided valuable information. However, in the analysis of NDV F, fragmentation with HCD was useful for glycopeptide identification while glycopeptides of F did not readily dissociate using ETciD. As set out in the introduction, HCD can facilitate glycopeptide identification by generating glycopeptide Y1 ions and low mass diagnostic oxonium ions at high mass accuracy when analysed in the Orbitrap. The former can be used to predict the peptide moiety of a glycopeptide, while the latter can be used to identify specific monosaccharide residues and substituents in the attached glycan. Confirming the composition of the attached glycan was particularly important for glycopeptides from HN as they contained a high number of Fuc residues, thus inspection of HCD spectra for diagnostic oxonium ions was instrumental in confirming that Fuc residues were not erroneously assigned in the place of NeuAc. In some instances the oxonium ions also enabled sulfation to be

confirmed on HN glycopeptides. Fragmentation with HCD also provided valuable peptide sequence information for HN glycopeptides containing N341 and N433; however, ETD was required to confidently assign the peptide portion of glycopeptides containing N481. Compared to HCD, fragmentation of HN glycopeptides with CID enabled more peptide+glycan ions to be identified, including masses of ~80 Da for Sulf or Phos substituents. However, manual validation of glycopeptides assigned by Byonic using CID spectra proved time-consuming. This was due to several factors; firstly, trap analysers provide low resolution and mass accuracy data and an inherent drawback is cut-off at the low m/z range. Consequently, each assigned monosaccharide composition was manually validated by interrogating spectra for diagnostic oxonium ions from the HCD spectrum that triggered the CID spectrum of interest. Secondly, Byonic did not annotate the majority of fragment ions from HN glycopeptides assigned by CID due to the usual composition of the glycans. Unassigned fragment ions were therefore manually identified for each glycopeptide assignment. In summary, although CID fragmentation was informative, the benefits were lessened by the high degree of manual interpretation required in the assignment of glycopeptides from HN. Thus for future analyses of HN glycopeptides, HCD and ETD may be preferred fragmentation methods, particularly for high-throughput or larger comparative analyses. For future analyses of NDV F fragmentation with HCD may prove beneficial.

There are also several interesting structural observations that may be taken from this work on NDV HN and F. Analysis of HN revealed some unusual glycans, such as paucimannose or complex or hybrid glycans with high levels of fucosylation that were variably sulfated or phosphorylated. Many of the fucosylated and sulfated or phosphorylated compositions were not present in UniCarbKB at the time the analyses were undertaken. As such, this research contributes novel results of glycans produced in embryonic eggs. (47, 287, 298). The observation of Sulf on NDV HN glycans is likely due to propagation methods. Glycans from human influenza HA produced in embryonic eggs have been observed to contain Sulf (48, 288, 289). Furthermore, paucimannose compositions (287) and Fuc residues have also been observed on glycans from human influenza HA propagated in embryonic eggs (288). The higher degree of fucosylation reported herein on NDV HN compared to influenza HA may be due to differences in protein structure or analytical and data analysis methods.

Differential glycosylation due to protein structure was particularly evident in the present work. The HN and F proteins were both derived from virions of NDV V4-VAR but in contrast to HN, the glycosylation profile of F was limited to high mannose glycans with a small degree of fucosylation at site N191. This high mannose profile has previously been described for NDV F, after release of

the glycans from another avirulent strain produced in MDBK cells (104). The glycosylation profile of NDV F differs to F proteins derived from virions of other paramyxoviruses which revealed predominantly complex-types after release of the glycans (135, 298, 309). As highlighted in Chapter 5, the functional roles of N-linked sites on F proteins also differ between paramyxoviruses despite general conservation of sites between some viruses and conservation of the overall architecture of F proteins in crystal structures. Thus, homology modelling of the F proteins may need to consider occupancy and heterogeneity at glycosylation sites. This highlights the need to investigate site-specific glycosylation of F proteins from the different paramyxoviruses, which has been examined for the first time in this work. The high mannose profile observed for NDV F compared to other paramyxoviruses indicates a different structural arrangement of NDV F during transport of the protein through the ER and Golgi. Hypothetically, such conformational differences may arise from interactions of F and the relevant attachment proteins during protein synthesis and processing. Alternatively, the high mannose profile of NDV F may be restricted to avirulent strains, which remains to be elucidated in future research.

Another important result from the work on NDV was the observation of an O-linked site in the stalk domain of HN. Although site-specificity of the O-linked glycans was not determined on HN from the V4-VAR isolate, T71 was identified as the likely site of attachment in preliminary studies of the second isolate, V4-QLD. This secondary confirmation of O-linked glycosylation revealed that the site is not limited to V4-VAR and indicated it may be present in other strains. In paramyxoviruses, the stalk domain of HN has been shown to modulate the fusion properties of F (63). Structural alignment of the stalk regions reveal T71 is conserved in between NDV HN and PIV5 (T62) and hPIV1 (T88). This finding will form the basis for future assessments of occupancy and the functional significance of the O-linked site in HN from more virulent stains of NDV and other paramyxoviruses such as PIV5 and hPIV1.

8.2.2 Limitations of the work

A limitation of this work is that the virions were not from naturally infected tissue or cells and thus may not accurately reflect glycosylation of the native proteins. Furthermore, only the monosaccharide compositions of the glycans were obtained, which is limitation of the methods and LC-MS/MS strategies applied. Another potential limitation in the analysis of NDV HN was the coverage of peptides or glycopeptides containing N-linked sites N538 and N600. Specifically, the

use of orthogonal proteases and PNGase F could have been implemented to elucidate occupancy at N538 and N600. However, N538 has previously been shown not to be glycosylated by analysis of peptides from a digest of HN from NDV-V4 VAR (101). Furthermore, the carboxy-terminal region of HN from NDV-V4 VAR undergoes substantial trimming (88, 285) and this may be the reason peptides or glycopeptides containing N600 were not detected. Peptides or glycopeptides containing N-linked site N447 from NDV F were also not detected and site occupancy therefore could not be confirmed. There are conflicting reports of occupancy at this site in other strains of NDV (102, 103), but site occupancy has been confirmed in a crystallography study of F from NDV V4-VAR (276).

8.2.3 Future directions

In order to address the unanswered issues in this work regarding coverage of N538 and N600, HN could be digested with alternative enzymes. Specifically, N538 coverage may be increased if an Arg-C digestion is followed with Glu-C. This may produce theoretical peptides terminating in R557 and E547, respectively (^{517}VSSSSTKAAYTTSTCFKVVKT<u>NKT</u>YCLSIAE^{547}ISNTLFGEFR^{557}). For detection of peptides containing N600, PNGase F treatment before or after Lys-C digestion may prove useful producing the theoretical peptide ^{600}<u>n</u>QTEYRRELESYAASWP^{616} or taking into consideration a missed cleavage at K599 ^{588}DDIVSPIFCDAK<u>n</u>QTEYRRELESYAASWP^{616}, where lower case "n" represents deamidated Asn. To assess occupancy at N447 of F the intact protein could be digested with PNGase F to increase the likelihood of proteolytic digestion with Glu-C and potentially produce the peptide ^{442}ATYQK<u>n</u>ISIQD^{452}.

Given the position of the O-linked site in the stalk domain of HN it would be of interest to investigate the presence of the O-linked site on HN from other strains of NDV, particularly in viruses with different levels of pathogenicity. Furthermore, given the conservation of T71 between NDV HN and PIV5 and hPIV1 it would also be of interest to investigate O-linked glycosylation in the latter viruses. If the O-linked site is observed in other strains of NDV or in other viruses, mutation studies could be completed to investigate potential functionality. Lastly, given the disparity of glycosylation profiles between NDV V4-VAR F and those available for F from other paramyxoviruses, it would be of interest to analyse site-specific glycosylation of virion derived F from other viruses and from more virulent strains of NDV.

8.3 GLYCOSYLATION OF HRSV SF AND SG

8.3.1 Overview of the results with reference to current literature

Biochemical analyses of hRSV glycoproteins are generally performed on recombinant forms due to the inherent difficulties in propagating hRSV virions. Importantly, the present results provide a template for future analyses of G and F by mass spectrometry using more limited quantities of these proteins derived from virions. This applies to the use of EThcD in determining site-specific O-linked glycosylation of hRSV G and Glu-C digestions to enable the detection of O-linked sites in the C-terminal region of the protein. With respect to hRSV F, the benefit of stepped HCD and EThcD for identifying the peptide and glycan portions of glycopeptides from F may prove useful in future studies. Furthermore, the identification of N-terminal peptide modifications of glycopeptides containing N27 and N70 of F will enable more tailored search parameters to be applied.

In addition to the technical aspects of characterising glycopeptides derived from hRSV F and G, the present research provides some intriguing structural insights that may have biological implications. Firstly, it confirmed that sites N27, N70 and N500 are occupied (139) and revealed that sites N116 and N126 in pep27 may also be occupied. Furthermore, the work herein confirmed that the N-linked glycans present on hRSV F are predominantly mature glycans (135, 314) as only low levels of high mannose glycans were observed. Importantly, O-linked glycosylation of hRSV F was detected for the first time with confirmation of an O-linked glycan at T100. The O-linked site is positioned N-terminal to the R109 furin-like cleavage site, and could theoretically mediate furin-like cleavage, as observed for O-linked sites of other proteins (33, 364). As cleavage of the furin like-sites and O-linked glycosylation are thought to occur in the later stages of protein processing in the Golgi, it is difficult to speculate on the timing of O-glycan addition versus furin cleavage. In the crystal structure of sF the peptide containing the O-linked site is buried inside the trimeric structure with the fusion peptide and therefore may not be accessible to ppGalNAcTs (332). Furthermore, it has been shown that trimerisation of sF takes place in the Golgi, only after pep27 is removed (357). Thus the O-linked glycans may prevent cleavage and incorporation of sF monomers into the trimer. Alternatively, the O-linked glycans may represent subsequent modifications of monomers that were not cleaved and were thus not incorporated into the trimer.

Also of interest are the potential biological implications of the types of N-linked glycans identified on sF and at site N135 of sG. If the glycans with diHexNAc units are shown to be LacdiNAc motifs there are several avenues of research that could be investigated. Terminal sulfated GalNAc is recognised by liver mannose receptors which mediate the rapid uptake of hormone glycoproteins expressing N-linked sulfated LacdiNAc antennas (392). Given that kidney cell lines are commonly used for the propagation of virions and the production of recombinant viral proteins and vaccines (48, 291, 383, 393), it raises the question of serum clearance of hRSV F and G therapeutics produced in kidney cell lines or other cell lines expressing β4GalNAc-Ts. Another question that could be raised is the potential antigenic or immune stimulating properties of LacdiNAc extensions on sF and sG or hRSV virions. If β4GalNAc-Ts are not expressed in the lungs of infants, children or adults how does the propagation of proteins or virions in cell lines that do express β4GalNAc-Ts, including commonly used Vero, HEK and lung adenocarcinoma cell lines, affect immunological responses *in vitro* and *in vivo*? This question is particularly relevant to the first inactivated hRSV vaccine, where virions were propagated in kidney cell lines and administration of the vaccine resulted in enhanced hRSV disease upon infection with wild type hRSV (117). Conversely, if β4GalNAc-Ts are expressed in the lungs of infants, as potentially indicated by high expression of β4GalNAc-T4 in foetal lung tissue (338), but not of children or adults, this may result in distinct antigenic forms of hRSV virions circulating. In this respect, site N70 may play an important role given its position within a major antigenic site (Ø) in the prefusion form of the protein (332).

As discussed in Chapter 6 LacdiNAc antennas can induce aberrant immune responses such as the Th2 bias observed in schistosomiasis (358), an immune response also observed in enhanced hRSV disease (116). Interestingly, LacdiNAc glycans present on glycodelin, a human endometrial protein, have been implicated in the potent immunosuppressive activities of the protein (354). Modulation of immune activity was thought to be mediated through suppression of human natural killer (NK) cell activity (394). Clinical studies of infants with severe hRSV infection identified low levels of NK cells in the blood and lung tissue (395, 396). Low levels of NK cells have also been shown to regulate hRSV-specific Th2 responses in a mouse model of hRSV infection (397).

Another interesting observation from current research is the overlap in glycan ligands and receptors between hRSV surface glycoproteins and selectins. Selectins (L-, P- and E-) are cell adhesion molecules that promote slow rolling and tethering of leukocytes to the endothelium so they can exit the blood and move to sites of inflammation. They play a role in inflammatory processes by binding to each other or ligands expressed on endothelial cells or leukocytes and inducing signalling

pathways (398, 399). Of particular interest is the role they play in inflammatory lung disease and the accumulation of neutrophils, eosinophils and other leucocytes in the lungs (399, 400). During hRSV-induced bronchiolitis in infants, L-selectin undergoes release or "shedding" from neutrophils, the predominant leukocyte in the airways during hRSV infection (401). More recently it has been shown that L-selectin is required for effector T cell migration to viral infected lungs and induction of protective immunity (402). Interestingly, L-selectin produced in HEK293F cells contained glycans with the same extensions presumed to be on sF in this study, that is sulfated and fucosylated LacdiNAc extensions (403). It has been postulated that the LacdiNAc extensions on L-selectin may be used to mediate self-binding (404). Sialylated, fucosylated and sulfated glycans, in particular those with Sialyl Lewis X (SLe^X) structures, are ligands for selectins (405) (**Figure 8-1**). Fucosylated LacdiNAc has been identified as a potent inhibitor of E-selectin with a greater binding affinity then SLe^X (406). The similar conformation of SLe^X and fucosylated LacdiNAc (**Figure 8-1**) is thought to be the reason E-selectin can bind both ligands. Interestingly, L-selectins have higher affinities for sulfated oligosaccharides (405) and sulfated derivatives of sLe^X have been observed to induce "shedding" of L-selectin from leukocytes (407). Hypothetically, if hRSV virions or airway epithelial cells infected with hRSV express F or G with LacdiNAc extensions, this may influence selectin-mediated processes such as neutrophil migration and promotion of T-cell protective immunity in the lungs.
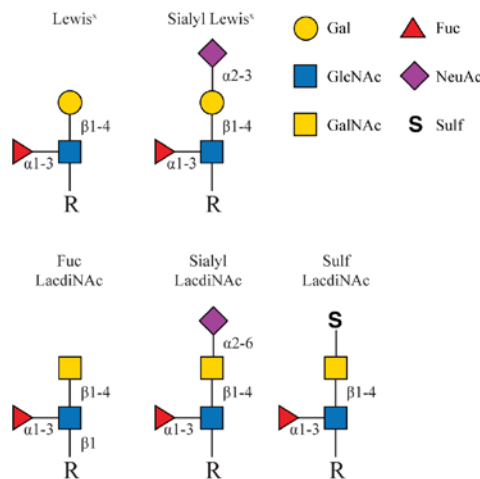
Figure 8-1. Comparison of Lewis^x antigens with LacdiNAc equivalents.

In addition to LacdiNAc extensions, interactions with selectins may be mediated by hRSV G, as the mucin-like glycosylation described on G is also observed on major selectin ligands which mediate trafficking of leukocytes to sites of inflammation (398). It has also been shown that hRSV G can bind directly to L-selectin (408). Furthermore, L-selectin binds nucleolin expressed by myeloid

lineage and haematopoietic progenitor cells (409). Interestingly, nucleolin has also been established as a receptor for the F protein of hRSV (126). Finally, virions of hRSV have been shown to have similar receptor binding characteristics to L-selectin (408). Fucoidan, an inhibitor of L-selectin, blocked hRSV infection of Hep2 cells, as did soluble L-selectin and the pan-selectin antagonist TBC1269 (408). Thus the binding affinities of F and G from hRSV may influence L-selectin-mediated processes.

Taken together, the above research regarding immune modulators, glycan ligands and the results obtained from the analyses of sF and sG invoke several hypotheses and potential avenues for research. A hypothetical paradigm of hRSV infection in the lungs has been presented, where the protein sequences of G and F promote the addition of immunogenic LacdiNAc glycan epitopes by β4GalNAc-Ts. Expression of these transferases in humans may change with age resulting in distinct antigenic forms of hRSV. The presence of these immune stimulating glycans on F or G in vaccines, therapeutics or natural infection with hRSV may also contribute to priming aberrant immune responses.

8.3.2 Limitations of the work

A limitation of this work is that the proteins were not from naturally infected tissue or cells and, as noted in section 8.2.2, only the monosaccharide compositions of the glycans were obtained. Therefore, linkages were not confirmed for N-glycan compositions from sF and sG that exhibited characteristics consistent with LacdiNAc extensions. Furthermore, releasing the N-glycans for quantification would provide important information about the overall level of diHexNAc glycosylation on sF and sG. With respect to sG peptides containing N-linked site N251 were not detected. This meant both occupancy at N251 and occupancy at the seven potential O-linked sites within the peptide (^{241}**TNII**TT**LL**T**S**NTTGNPE^{257}) could not be determined. Finally, two regions of the C-terminus of hRSV-G were not observed (^{272}GNPSPSQVSTTSE^{284} and ^{285}YPSQPSSPPNTPRQ^{298}).

8.3.3 Future directions

For future studies, combining the proteolytic digestions (trypsin and Glu-C) with PNGase F may enable identification of peptides or O-linked glycopeptides containing N251 of G by producing the theoretical peptide ^{241}TNIITTLLTS<u>n</u>TTGNPE^{257} (where lower case "n" represents deamidated Asn). Furthermore, analyses of the C-terminal regions of sG may benefit from the use of Glu-C digestion alone. This would result in significantly fewer "MS amenable" glycopeptides and peptides being produced from the CCD, HBD and regions downstream to these. Analysis of the trypsin/Glu-C digestion revealed glycopeptides and peptides from these regions dominated MS spectra at retention times when the C-terminal Glu-C glycopeptides were eluting (^{258}LTSQME^{263} and ^{264}TFHSTSSE^{271}). By using Glu-C digestion alone it may effectively enrich for the C-terminal glycopeptides and enable detection of the remaining regions that were not observed (^{272}GNPSPSQVSTTSE^{284} and ^{285}YPSQPSSPPNTPRQ^{298}).

Based on observations in this work, future avenues of research may include investigations into the linkages of the diHexNAc units observed on hRSV sF and virions produced in cells that express β4GalNAc-Ts and those that do not. This could be achieved through enzyme treatments and antibody or lectin binding studies to determine linkage. Glycoproteomics or glycomics may also be used to confirm the results. These experiments could be followed by immunological studies where virions produced in these cell lines are applied to animal models to see if there are differences in immune responses. Furthermore, it would be of interest to determine if the two furin sties act as a *cis*-regulatory element for β4 specific GalNAc-Ts. This could be achieved by analysing site-specific glycosylation of hRSV sF proteins where both or one of the furin cleavage consensus sites have been removed (^{106}**RARR**^{109} and/or ^{131}**KKRKRR**^{136}). Additionally, sequences could be removed that contain the furin cleavage consensus site and surrounding basic residues (^{123}**KK**T<u>NVT</u>LS**KKRKRR**^{136}). If removal of these sequences does change the glycosylation profile of hRSV sF mutating individual amino acids within these sequences may highlight important residues for induction of β4-specific GalNAc-transferases. Moreover, if removal of the basic residues from sF changes the glycan profile qualitatively it would be important to quantify the difference at both at the glycopeptide and released glycan level. Finally, to comprehensively assess and quantify the N-linked glycans present on hRSV sG the glycans could be released and analysed. Ultimately, it would be advantageous to analyse hRSV F and G produced from virions expressed in cell lines derived from human airway epithelial cell lines or infected respiratory tissue and compare the results to the work herein.

8.4   CONCLUDING REMARKS

Protein glycosylation is highly heterogeneous and determining macroheterogeneity and microheterogeneity in a pool of proteins remains a challenge. Currently, the analysis of glycoproteomic data is time consuming and requires manual validation of all glycopeptide assignments. As bioinformatic tools evolve to include the rules and observations from manual validations of glycoproteomic results, studies will be applied in a more high-throughput manner with less user input. Furthermore, advances in LC-MS/MS technologies, with increases in speed, sensitivity, m/z range, dynamic range, resolution and mass accuracy will enable studies to be completed on more complex glycoproteomes. However, determining how individual monosaccharide motifs, glycans or types glycans influence protein function within a population of proteins, remains a substantial challenge. Accordingly, although this work provides a technical template for the study of NDV and hRSV surface glycoproteins in a site-specific manner, simply applying these to virions from naturally infected cells or tissues may not fully elucidate the functional role of glycans given the inherent complexity of glycosylation. Nevertheless, this work represents a step towards defining the role of glycosylation from paramyxovirus attachment and fusion proteins, and as tools and techniques develop, it may help decrease the burden of these viruses.

Bibliography

1. Moremen, K. W., Tiemeyer, M., and Nairn, A. V. (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13**, 448-462

2. Hart, G. W., and Copeland, R. J. (2010) Glycomics hits the big time. *Cell* **143**, 672-676

3. Parodi, A. J. (2000) Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem. J.* **348**, 1-13

4. Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867

5. Marth, J. D., and Grewal, P. K. (2008) Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* **8**, 874-887

6. Sinclair, A. M., and Elliott, S. (2005) Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *J. Pharm. Sci.* **94**, 1626-1635

7. Schiestl, M., Stangler, T., Torella, C., Cepeljnik, T., Toll, H., and Grau, R. (2011) Acceptable changes in quality attributes of glycosylated biopharmaceuticals. *Nat. Biotechnol.* **29**, 310-312

8. Higel, F., Seidl, A., Sorgel, F., and Friess, W. (2016) N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm.* **100**, 94-100

9. Bertozzi, C. R., and Rabuka, D. (2009) Structural Basis of Glycan Diversity. in *Essentials of Glycobiology* (Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. eds.), 2 Ed., Cold Spring Harbor Laboratory Press, New York, NY

10. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324

11. Larkin, A., and Imperiali, B. (2011) The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **50**, 4411-4426

12. Goochee, C. F., and Monica, T. (1990) Environmental Effects on Protein Glycosylation. *Nat. Biotechnol.* **8**, 421-427

13. Zacchi, L. F., and Schulz, B. L. (2016) N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconj. J.* **33**, 359-376

14. Aebi, M. (2013) N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta* **1833**, 2430-2437

15. Stanley, P., Schachter, H., and Taniguchi, N. (2009) N-Glycans. in *Essentials of Glycobiology* (Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. eds.), 2 Ed., Cold Spring Harbor Laboratory Press, New York, NY

16. Kornfeld, R., and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631-664

17. Thaysen-Andersen, M., and Packer, N. H. (2012) Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology* **22**, 1440-1452

18. Lee, L. Y., Lin, C. H., Fanayan, S., Packer, N. H., and Thaysen-Andersen, M. (2014) Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants. *Front. Immunol.* **5**, 404

19. Thaysen-Andersen, M., Packer, N. H., and Schulz, B. L. (2016) Maturing glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Mol. Cell. Proteomics* **15**, 1773-1790

20. Dahmen, A.-C., Fergen, M.-T., Laurini, C., Schmitz, B., Loke, I., Thaysen-Andersen, M., and Diestel, S. (2015) Paucimannosidic glycoepitopes are functionally involved in proliferation of neural progenitor cells in the subventricular zone. *Glycobiology* **25**, 869-880

21. Thaysen-Andersen, M., Venkatakrishnan, V., Loke, I., Laurini, C., Diestel, S., Parker, B. L., and Packer, N. H. (2015) Human neutrophils secrete bioactive paucimannosidic proteins from azurophilic granules into pathogen-infected sputum. *J. Biol. Chem.* **290**, 8789-8802

22. Medzihradszky, K. F., Kaasik, K., and Chalkley, R. J. (2015) Tissue-specific glycosylation at the glycopeptide level. *Mol. Cell. Proteomics* **14**, 2103-2110

23. Fujitani, N., Furukawa, J., Araki, K., Fujioka, T., Takegawa, Y., Piao, J., Nishioka, T., Tamura, T., Nikaido, T., Ito, M., Nakamura, Y., and Shinohara, Y. (2013) Total cellular glycomics allows characterizing cells and streamlining the discovery process for cellular biomarkers. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2105-2110

24. Brockhausen, I., Schachter, H., and Stanley, P. (2009) O-GalNAc Glycans. in *Essentials of Glycobiology* (Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. eds.), 2 Ed., Cold Spring Harbor Laboratory Press, New York, NY

25. Hart, G. W., Housley, M. P., and Slawson, C. (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-1022

26. Sakaidani, Y., Ichiyanagi, N., Saito, C., Nomura, T., Ito, M., Nishio, Y., Nadano, D., Matsuda, T., Furukawa, K., and Okajima, T. (2012) O-linked-N-acetylglucosamine modification of mammalian Notch receptors by an atypical O-GlcNAc transferase Eogt1. *Biochem. Biophys. Res. Commun.* **419**, 14-19

27. Hollingsworth, M. A., and Swanson, B. J. (2004) Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer* **4**, 45-60

28. Hattrup, C. L., and Gendler, S. J. (2008) Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol.* **70**, 431-457

29. Zanin, M., Baviskar, P., Webster, R., and Webby, R. (2016) The Interaction between Respiratory Pathogens and Mucus. *Cell Host Microbe* **19**, 159-168

30. Tian, E., and Ten Hagen, K. G. (2009) Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconj. J.* **26**, 325-334

31. Joshi, H. J., Steentoft, C., Schjoldager, K. T.-B. G., Vakhrushev, S. Y., Wandall, H. H., and Clausen, H. (2015) Protein O-GalNAc Glycosylation: Most Complex and Differentially Regulated PTM. in *Glycoscience: Biology and Medicine* (Taniguchi, N., Endo, T., Hart, G. W., Seeberger, P. H., and Wong, C.-H. eds.), Springer Japan. pp 1049-1064

32. Plomp, R., Dekkers, G., Rombouts, Y., Visser, R., Koeleman, C. A., Kammeijer, G. S., Jansen, B. C., Rispens, T., Hensbergen, P. J., Vidarsson, G., and Wuhrer, M. (2015) Hinge-Region O-Glycosylation of Human Immunoglobulin G3 (IgG3). *Mol. Cell. Proteomics* **14**, 1373-1384

33. Schjoldager, K. T., and Clausen, H. (2012) Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochim. Biophys. Acta* **1820**, 2079-2094

34. Xue, J., Zhu, L.-P., and Wei, Q. (2013) IgG-Fc N-glycosylation at Asn297 and IgA O-glycosylation in the hinge region in health and disease. *Glycoconj. J.* **30**, 735-745

35. Le Pendu, J., Nystrom, K., and Ruvoen-Clouet, N. (2014) Host-pathogen co-evolution and glycan interactions. *Curr. Opin. Virol.* **7**, 88-94

36. Banerjee, N., and Mukhopadhyay, S. (2016) Viral glycoproteins: biological role and application in diagnosis. *VirusDisease* **27**, 1-11

37. Bossart, K. N., Fusco, D. L., and Broder, C. C. (2013) Paramyxovirus entry. *Adv. Exp. Med. Biol.* **790**, 95-127

38.   Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T. M., Sasisekharan, V., and Sasisekharan, R. (2008) Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat. Biotechnol.* **26**, 107-113

39.   Jardetzky, T. S., and Lamb, R. A. (2014) Activation of paramyxovirus membrane fusion and virus entry. *Curr. Opin. Virol.* **5**, 24-33

40.   Vigerust, D. J., and Shepherd, V. L. (2007) Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol.* **15**, 211-218

41.   Balzarini, J. (2007) Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nat. Rev. Microbiol.* **5**, 583-597

42.   Yoneda, M., Georges-Courbot, M. C., Ikeda, F., Ishii, M., Nagata, N., Jacquot, F., Raoul, H., Sato, H., and Kai, C. (2013) Recombinant measles virus vaccine expressing the Nipah virus glycoprotein protects against lethal Nipah virus challenge. *PLoS One* **8**, e58414

43.   Goff, P. H., Krammer, F., Hai, R., Seibert, C. W., Margine, I., Garcia-Sastre, A., and Palese, P. (2013) Induction of cross-reactive antibodies to novel H7N9 influenza virus by recombinant Newcastle disease virus expressing a North American lineage H7 subtype hemagglutinin. *J. Virol.* **87**, 8235-8240

44.   Olofsson, S., Blixt, O., Bergstrom, T., Frank, M., and Wandall, H. H. (2015) Viral O-GalNAc peptide epitopes: a novel potential target in viral envelope glycoproteins. *Rev. Med. Virol.* **26**, 34-48

45.   Suen, K. F., Turner, M. S., Gao, F., Liu, B., Althage, A., Slavin, A., Ou, W., Zuo, E., Eckart, M., Ogawa, T., Yamada, M., Tuntland, T., Harris, J. L., and Trauger, J. W. (2010) Transient expression of an IL-23R extracellular domain Fc fusion protein in CHO vs. HEK cells results in improved plasma exposure. *Protein Expr. Purif.* **71**, 96-102

46.   Croset, A., Delafosse, L., Gaudry, J. P., Arod, C., Glez, L., Losberger, C., Begue, D., Krstanovic, A., Robert, F., Vilbois, F., Chevalet, L., and Antonsson, B. (2012) Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J. Biotechnol.* **161**, 336-348

47.   Sriwilaijaroen, N., Kondo, S., Yagi, H., Wilairat, P., Hiramatsu, H., Ito, M., Ito, Y., Kato, K., and Suzuki, Y. (2009) Analysis of N-glycans in embryonated chicken egg chorioallantoic and amniotic cells responsible for binding and adaptation of human and avian influenza viruses. *Glycoconj. J.* **26**, 433-443

48.   An, Y., Rininger, J. A., Jarvis, D. L., Jing, X., Ye, Z., Aumiller, J. J., Eichelberger, M., and Cipollo, J. F. (2013) Comparative glycomics analysis of influenza hemagglutinin (H5N1) produced in vaccine relevant cell platforms. *J. Proteome Res.* **12**, 3707-3720

49. Doores, K. J., Bonomelli, C., Harvey, D. J., Vasiljevic, S., Dwek, R. A., Burton, D. R., Crispin, M., and Scanlan, C. N. (2010) Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13800-13805

50. Garcia-Beato, R., Martinez, I., Franci, C., Real, F. X., Garcia-Barreno, B., and Melero, J. A. (1996) Host cell effect upon glycosylation and antigenicity of human respiratory syncytial virus G glycoprotein. *Virology* **221**, 301-309

51. Lamb, R. A., and Parks, G. D. (2013) Paramyxoviridae. in *Fields virology* (Knipe, D. M., and Howley, P. M. eds.), 6 Ed., Lippincott Williams & Wilkins, Philadelphia, PA. pp 957-995

52. Stein, C. E., Birmingham, M., Kurian, M., Duclos, P., and Strebel, P. (2003) The global burden of measles in the year 2000-a model that uses country-specific indicators. *J. Infect. Dis.* **187 (Suppl 1)**, S8-14

53. Rubin, S., Eckhaus, M., Rennick, L. J., Bamford, C. G., and Duprex, W. P. (2015) Molecular biology, pathogenesis and pathology of mumps virus. *J. Pathol.* **235**, 242-252

54. Gentile, A., Bhutta, Z., Bravo, L., Samy, A. G., Garcia, R. D., Hoosen, A., Islam, T., Karimi, A., Salem, M., Simasathien, S., Sohail, A., Watanaveeradej, V., Wiedenmayer, K., and Schmitt, H. J. (2010) Pediatric disease burden and vaccination recommendations: understanding local differences. *Int. J. Infect. Dis.* **14**, 649-658

55. Lee, B., and Ataman, Z. A. (2011) Modes of paramyxovirus fusion: a Henipavirus perspective. *Trends Microbiol.* **19**, 389-399

56. Aljofan, M. (2013) Hendra and Nipah infection: Emerging paramyxoviruses. *Virus Res.* **177**, 119-126

57. Le Bayon, J. C., Lina, B., Rosa-Calatrava, M., and Boivin, G. (2013) Recent developments with live-attenuated recombinant paramyxovirus vaccines. *Rev. Med. Virol.* **23**, 15-34

58. Shaw, C. A., Ciarlet, M., Cooper, B. W., Dionigi, L., Keith, P., O'Brien, K. B., Rafie-Kolpin, M., and Dormitzer, P. R. (2013) The path to an RSV vaccine. *Curr. Opin. Virol.* **3**, 332-342

59. Mazur, N. I., Martinon-Torres, F., Baraldi, E., Fauroux, B., Greenough, A., Heikkinen, T., Manzoni, P., Mejias, A., Nair, H., Papadopoulos, N. G., Polack, F. P., Ramilo, O., Sharland, M., Stein, R., Madhi, S. A., Bont, L., and Respiratory Syncytial Virus, N. (2015) Lower respiratory tract infection caused by respiratory syncytial virus: current management and new therapeutics. *Lancet Respir. Med.* **3**, 888-900

60. Halpin, K., and Rota, P. (2015) A Review of Hendra Virus and Nipah Virus Infections in Man and Other Animals. in *Zoonoses - Infections Affecting Humans and Animals: Focus on Public Health Aspects* (Sing, A. ed.), Springer Netherlands, Dordrecht. pp 997-1012

61. Gould, A. R., Kattenbelt, J. A., Selleck, P., Hansson, E., Della-Porta, A., and Westbury, H. A. (2001) Virulent Newcastle disease in Australia: molecular epidemiological analysis of viruses isolated prior to and during the outbreaks of 1998-2000. *Virus Res.* **77**, 51-60

62. Wang, L.-F., Collins, P. L., Fouchier, R. A. M., Kurath, G., Lamb, R. A., Randall, R. E., and Rima, B. K. (2012) Family Paramyxoviridae. in *Virus Taxonomy Ninth Report of the International Committee on Taxonomy of Viruses* (King, A. M. Q., Adams, M. J., Carstens, E. B., and Lefkowitz, E. J. eds.), Elsevier, San Diego. pp 672-685

63. Bose, S., Jardetzky, T. S., and Lamb, R. A. (2015) Timing is everything: Fine-tuned molecular machines orchestrate paramyxovirus entry. *Virology* **479-480**, 518-531

64. Chang, A., and Dutch, R. E. (2012) Paramyxovirus fusion and entry: multiple paths to a common end. *Viruses* **4**, 613-636

65. Takimoto, T., and Portner, A. (2004) Molecular mechanism of paramyxovirus budding. *Virus Res.* **106**, 133-145

66. Horvath, C. M., and Lamb, R. A. (1992) Studies on the fusion peptide of a paramyxovirus fusion glycoprotein: roles of conserved residues in cell fusion. *J. Virol.* **66**, 2443-2455

67. Poor, T. A., Jones, L. M., Sood, A., Leser, G. P., Plasencia, M. D., Rempel, D. L., Jardetzky, T. S., Woods, R. J., Gross, M. L., and Lamb, R. A. (2014) Probing the paramyxovirus fusion (F) protein-refolding event from pre- to postfusion by oxidative footprinting. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2596-2605

68. Horvath, C. M., Paterson, R. G., Shaughnessy, M. A., Wood, R., and Lamb, R. A. (1992) Biological activity of paramyxovirus fusion proteins: factors influencing formation of syncytia. *J. Virol.* **66**, 4564-4569

69. Villar, E., and Barroso, I. (2006) Role of sialic acid-containing molecules in paramyxovirus entry into the host cell: A minireview. *Glycoconj. J.* **23**, 5-17

70. Bose, S., Song, A. S., Jardetzky, T. S., and Lamb, R. A. (2014) Fusion activation through attachment protein stalk domains indicates a conserved core mechanism of paramyxovirus entry into cells. *J. Virol.* **88**, 3925-3941

71. Biacchesi, S., Pham, Q. N., Skiadopoulos, M. H., Murphy, B. R., Collins, P. L., and Buchholz, U. J. (2005) Infection of nonhuman primates with recombinant human metapneumovirus lacking the SH, G, or M2-2 protein categorizes each as a nonessential accessory protein and identifies vaccine candidates. *J. Virol.* **79**, 12608-12613

72. Karron, R. A., Buonagurio, D. A., Georgiu, A. F., Whitehead, S. S., Adamus, J. E., Clements-Mann, M. L., Harris, D. O., Randolph, V. B., Udem, S. A., Murphy, B. R., and Sidhu, M. S. (1997) Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vitro: Clinical evaluation and molecular characterization of a cold-

passaged, attenuated RSV subgroup B mutant. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13961-13966

73. Kwilas, S., Liesman, R. M., Zhang, L., Walsh, E., Pickles, R. J., and Peeples, M. E. (2009) Respiratory syncytial virus grown in Vero cells contains a truncated attachment protein that alters its infectivity and dependence on glycosaminoglycans. *J. Virol.* **83**, 10710-10718

74. Alexander, D. J. (2000) Newcastle disease and other avian paramyxoviruses. *Rev. Sci. Tech.* **19**, 443-462

75. Lancaster, J. (1976) A History of Newcastle disease with comments on its economic effects. *Worlds Poult. Sci. J.* **32**, 167-175

76. Munir, M., Shabbir, M. Z., Akhtar, S., Tang, Y., Yaqub, T., Ahmad, A., Mustafa, G., Alam, M. A., Santhakumar, D., and Nair, V. (2016) Infectivity of Wild Bird Origin Avian Paramyxovirus Serotype 1 and Vaccine Effectiveness in Chickens. *J. Gen. Virol.* **Ahead of print**

77. Ganar, K., Das, M., Sinha, S., and Kumar, S. (2014) Newcastle disease virus: Current status and our understanding. *Virus Res.* **184**, 71-81

78. Nagai, Y. (1993) Protease-dependent virus tropism and pathogenicity. *Trends Microbiol.* **1**, 81-87

79. Moscona, A. (2005) Neuraminidase inhibitors for influenza. *N. Engl. J. Med.* **353**, 1363-1373

80. Fournier, P., and Schirrmacher, V. (2013) Oncolytic Newcastle Disease Virus as Cutting Edge between Tumor and Host. *Biology* **2**, 936-975

81. Zamarin, D., and Palese, P. (2012) Oncolytic Newcastle disease virus for cancer therapy: old challenges and new directions. *Future Microbiol.* **7**, 347-367

82. Bukreyev, A., and Collins, P. L. (2008) Newcastle disease virus as a vaccine vector for humans. *Curr. Opin. Mol. Ther.* **10**, 46-55

83. He, D., Sun, L., Li, C., Hu, N., Sheng, Y., Chen, Z., Li, X., Chi, B., and Jin, N. (2014) Anti-Tumor Effects of an Oncolytic Adenovirus Expressing Hemagglutinin-Neuraminidase of Newcastle Disease Virus in Vitro and in Vivo. *Viruses* **6**, 856-874

84. Murawski, M. R., McGinnes, L. W., Finberg, R. W., Kurt-Jones, E. A., Massare, M. J., Smith, G., Heaton, P. M., Fraire, A. E., and Morrison, T. G. (2010) Newcastle disease virus-like particles containing respiratory syncytial virus G protein induced protection in BALB/c mice, with no evidence of immunopathology. *J. Virol.* **84**, 1110-1123

85. Motamedi, M., Amani, J., Shahsavandi, S., and Salmanian, A. (2014) In Silico Design of Multimeric HN-F Antigen as a Highly Immunogenic Peptide Vaccine Against Newcastle Disease Virus. *Int. J. Pept. Res. Ther.* **20**, 179-194

86.   McGinnes, L. W., Pantua, H., Laliberte, J. P., Gravel, K. A., Jain, S., and Morrison, T. G. (2010) Assembly and biological and immunological properties of Newcastle disease virus-like particles. *J. Virol.* **84**, 4513-4523

87.   Toyoda, T., Sakaguchi, T., Imai, K., Inocencio, N. M., Gotoh, B., Hamaguchi, M., and Nagai, Y. (1987) Structural comparison of the cleavage-activation site of the fusion glycoprotein between virulent and avirulent strains of newcastle disease virus. *Virology* **158**, 242-247

88.   Gorman, J. J., Nestorowicz, A., Mitchell, S. J., Corino, G. L., and Selleck, P. W. (1988) Characterization of the sites of proteolytic activation of Newcastle disease virus membrane glycoprotein precursors. *J. Biol. Chem.* **263**, 12522-12531

89.   Nagai, Y., Klenk, H.-D., and Rott, R. (1976) Proteolytic cleavage of the viral glycoproteins and its significance for the virulence of Newcastle disease virus. *Virology* **72**, 494-508

90.   Zhao, W., Zhang, Z., Zsak, L., and Yu, Q. (2013) Effects of the HN gene C-terminal extensions on the Newcastle disease virus virulence. *Virus Genes* **47**, 498-504

91.   Yuan, P., Paterson, R. G., Leser, G. P., Lamb, R. A., and Jardetzky, T. S. (2012) Structure of the Ulster strain newcastle disease virus hemagglutinin-neuraminidase reveals auto-inhibitory interactions associated with low virulence. *PLoS Pathog.* **8**, e1002855

92.   Kim, S. H., Xiao, S., Paldurai, A., Collins, P. L., and Samal, S. K. (2014) Role of C596 in the C-terminal extension of the haemagglutinin-neuraminidase protein in replication and pathogenicity of a highly virulent Indonesian strain of Newcastle disease virus. *J. Gen. Virol.* **95**, 331-336

93.   Sánchez-Felipe, L., Villar, E., and Muñoz-Barroso, I. (2012) α2-3- and α2-6- N-linked sialic acids allow efficient interaction of Newcastle Disease Virus with target cells. *Glycoconj. J.* **29**, 539-549

94.   Sergel, T., McGinnes, L. W., Peeples, M. E., and Morrison, T. G. (1993) The Attachment Function of the Newcastle Disease Virus Hemagglutinin-Neuraminidase Protein Can Be Separated from Fusion Promotion by Mutation. *Virology* **193**, 717-726

95.   Bose, S., Zokarkar, A., Welch, B. D., Leser, G. P., Jardetzky, T. S., and Lamb, R. A. (2012) Fusion activation by a headless parainfluenza virus 5 hemagglutinin-neuraminidase stalk suggests a modular mechanism for triggering. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2625-2634

96.   Yuan, P., Swanson, K. A., Leser, G. P., Paterson, R. G., Lamb, R. A., and Jardetzky, T. S. (2011) Structure of the Newcastle disease virus hemagglutinin-neuraminidase (HN) ectodomain reveals a four-helix bundle stalk. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14920-14925

97. Bose, S., Welch, B. D., Kors, C. A., Yuan, P., Jardetzky, T. S., and Lamb, R. A. (2011) Structure and mutagenesis of the parainfluenza virus 5 hemagglutinin-neuraminidase stalk domain reveals a four-helix bundle and the role of the stalk in fusion promotion. *J. Virol.* **85**, 12855-12866

98. Mirza, A. M., and Iorio, R. M. (2013) A mutation in the stalk of the newcastle disease virus hemagglutinin-neuraminidase (HN) protein prevents triggering of the F protein despite allowing efficient HN-F complex formation. *J. Virol.* **87**, 8813-8815

99. Melanson, V. R., and Iorio, R. M. (2006) Addition of N-glycans in the stalk of the Newcastle disease virus HN protein blocks its interaction with the F protein and prevents fusion. *J. Virol.* **80**, 623-633

100. Melanson, V. R., and Iorio, R. M. (2004) Amino acid substitutions in the F-specific domain in the stalk of the newcastle disease virus HN protein modulate fusion and interfere with its interaction with the F protein. *J. Virol.* **78**, 13053-13061

101. Pitt, J. J., Da Silva, E., and Gorman, J. J. (2000) Determination of the disulfide bond arrangement of Newcastle disease virus hemagglutinin neuraminidase. Correlation with a beta-sheet propeller structural fold predicted for paramyxoviridae attachment proteins. *J. Biol. Chem.* **275**, 6469-6478

102. McGinnes, L., Sergel, T., Reitter, J., and Morrison, T. (2001) Carbohydrate modifications of the NDV fusion protein heptad repeat domains influence maturation and fusion activity. *Virology* **283**, 332-342

103. Samal, S., Khattar, S. K., Kumar, S., Collins, P. L., and Samal, S. K. (2012) Coordinate deletion of N-glycans from the heptad repeats of the fusion F protein of Newcastle disease virus yields a hyperfusogenic virus with increased replication, virulence, and immunogenicity. *J. Virol.* **86**, 2501-2511

104. Diabate, S., Geyer, R., and Stirm, S. (1984) Structure of the major oligosaccharides in the fusion glycoprotein of Newcastle disease virus. *Eur. J. Biochem.* **139**, 329-336

105. Lopaticki, S., Morrow, C. J., and Gorman, J. J. (1998) Characterization of pathotype-specific epitopes of Newcastle disease virus fusion glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and post-source decay sequencing. *J. Mass Spectrom.* **33**, 950-960

106. McGinnes, L. W., and Morrison, T. G. (1995) The role of individual oligosaccharide chains in the activities of the HN glycoprotein of Newcastle disease virus. *Virology* **212**, 398-410

107. Panda, A., Elankumaran, S., Krishnamurthy, S., Huang, Z., and Samal, S. K. (2004) Loss of N-linked glycosylation from the hemagglutinin-neuraminidase protein alters virulence of Newcastle disease virus. *J. Virol.* **78**, 4965-4975

108. Crennell, S., Takimoto, T., Portner, A., and Taylor, G. (2000) Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. *Nat. Struct. Biol.* **7**, 1068-1074

109. McGinnes, L. W., and Morrison, T. G. (1997) Disulfide bond formation is a determinant of glycosylation site usage in the hemagglutinin-neuraminidase glycoprotein of Newcastle disease virus. *J. Virol.* **71**, 3083-3089

110. Hall, C. B., Weinberg, G. A., Iwane, M. K., Blumkin, A. K., Edwards, K. M., Staat, M. A., Auinger, P., Griffin, M. R., Poehling, K. A., Erdman, D., Grijalva, C. G., Zhu, Y., and Szilagyi, P. (2009) The burden of respiratory syncytial virus infection in young children. *N. Engl. J. Med.* **360**, 588-598

111. Nair, H., Nokes, D. J., Gessner, B. D., Dherani, M., Madhi, S. A., Singleton, R. J., O'Brien, K. L., Roca, A., Wright, P. F., Bruce, N., Chandran, A., Theodoratou, E., Sutanto, A., Sedyaningsih, E. R., Ngama, M., Munywoki, P. K., Kartasasmita, C., Simões, E. A. F., Rudan, I., Weber, M. W., and Campbell, H. (2010) Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *The Lancet* **375**, 1545-1555

112. Taylor, S., Taylor, R. J., Lustig, R. L., Schuck-Paim, C., Haguinet, F., Webb, D. J., Logie, J., Matias, G., and Fleming, D. M. (2016) Modelling estimates of the burden of respiratory syncytial virus infection in children in the UK. *BMJ Open* **6**, e009337

113. Falsey, A. R., and Walsh, E. E. (2000) Respiratory syncytial virus infection in adults. *Clin. Microbiol. Rev.* **13**, 371-384

114. Anderson, L. J., Dormitzer, P. R., Nokes, D. J., Rappuoli, R., Roca, A., and Graham, B. S. (2013) Strategic priorities for respiratory syncytial virus (RSV) vaccine development. *Vaccine* **31 Suppl 2**, B209-215

115. Collins, P. L., and Melero, J. A. (2011) Progress in understanding and controlling respiratory syncytial virus: still crazy after all these years. *Virus Res.* **162**, 80-99

116. Acosta, P. L., Caballero, M. T., and Polack, F. P. (2015) Brief History and Characterization of Enhanced Respiratory Syncytial Virus Disease. *Clin. Vaccine Immunol.* **23**, 189-195

117. Kim, H. W., Canchola, J. G., Brandt, C. D., Pyles, G., Chanock, R. M., Jensen, K., and Parrott, R. H. (1969) Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. *Am. J. Epidemiol.* **89**, 422-434

118. McLellan, J. S., Ray, W. C., and Peeples, M. E. (2013) Structure and function of respiratory syncytial virus surface glycoproteins. *Curr. Top. Microbiol. Immunol.* **372**, 83-104

119. Bose, M. E., He, J., Shrivastava, S., Nelson, M. I., Bera, J., Halpin, R. A., Town, C. D., Lorenzi, H. A., Noyola, D. E., Falcone, V., Gerna, G., De Beenhouwer, H., Videla, C., Kok,

T., Venter, M., Williams, J. V., and Henrickson, K. J. (2015) Sequencing and analysis of globally obtained human respiratory syncytial virus A and B genomes. *PLoS One* **10**, e0120098

120. Johnson, P. R., Spriggs, M. K., Olmsted, R. A., and Collins, P. L. (1987) The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 5625-5629

121. Gorman, J. J., Ferguson, B. L., Speelman, D., and Mills, J. (1997) Determination of the disulfide bond arrangement of human respiratory syncytial virus attachment (G) protein by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Protein Sci.* **6**, 1308-1315

122. Shingai, M., Azuma, M., Ebihara, T., Sasai, M., Funami, K., Ayata, M., Ogura, H., Tsutsumi, H., Matsumoto, M., and Seya, T. (2008) Soluble G protein of respiratory syncytial virus inhibits Toll-like receptor 3/4-mediated IFN-beta induction. *Int. Immunol.* **20**, 1169-1180

123. Hallak, L. K., Spillmann, D., Collins, P. L., and Peeples, M. E. (2000) Glycosaminoglycan sulfation requirements for respiratory syncytial virus infection. *J. Virol.* **74**, 10508-10513

124. Techaarpornkul, S., Collins, P. L., and Peeples, M. E. (2002) Respiratory syncytial virus with the fusion protein as its only viral glycoprotein is less dependent on cellular glycosaminoglycans for attachment than complete virus. *Virology* **294**, 296-304

125. Johnson, S. M., McNally, B. A., Ioannidis, I., Flano, E., Teng, M. N., Oomens, A. G., Walsh, E. E., and Peeples, M. E. (2015) Respiratory syncytial virus uses CX3CR1 as a receptor on primary human airway epithelial cultures. *PLoS Pathog.* **11**, e1005318

126. Tayyari, F., Marchant, D., Moraes, T. J., Duan, W., Mastrangelo, P., and Hegele, R. G. (2011) Identification of nucleolin as a cellular receptor for human respiratory syncytial virus. *Nat. Med.* **17**, 1132-1135

127. Gonzalez-Reyes, L., Ruiz-Arguello, M. B., Garcia-Barreno, B., Calder, L., Lopez, J. A., Albar, J. P., Skehel, J. J., Wiley, D. C., and Melero, J. A. (2001) Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9859-9864

128. Zimmer, G., Budz, L., and Herrler, G. (2001) Proteolytic activation of respiratory syncytial virus fusion protein. Cleavage at two furin consensus sequences. *J. Biol. Chem.* **276**, 31642-31650

129. Rawling, J., Cano, O., Garcin, D., Kolakofsky, D., and Melero, J. A. (2011) Recombinant Sendai viruses expressing fusion proteins with two furin cleavage sites mimic the syncytial

and receptor-independent infection properties of respiratory syncytial virus. *J. Virol.* **85**, 2771-2780

130.   Krzyzaniak, M. A., Zumstein, M. T., Gerez, J. A., Picotti, P., and Helenius, A. (2013) Host cell entry of respiratory syncytial virus involves macropinocytosis followed by proteolytic activation of the F protein. *PLoS Pathog.* **9**, e1003309

131.   Swanson, K. A., Settembre, E. C., Shaw, C. A., Dey, A. K., Rappuoli, R., Mandl, C. W., Dormitzer, P. R., and Carfi, A. (2011) Structural basis for immunization with postfusion respiratory syncytial virus fusion F glycoprotein (RSV F) to elicit high neutralizing antibody titers. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9619-9624

132.   Day, N. D., Branigan, P. J., Liu, C., Gutshall, L. L., Luo, J., Melero, J. A., Sarisky, R. T., and Del Vecchio, A. M. (2006) Contribution of cysteine residues in the extracellular domain of the F protein of human respiratory syncytial virus to its function. *Virol J.* **3**, 34

133.   Wertz, G. W., Krieger, M., and Ball, L. A. (1989) Structure and cell surface maturation of the attachment glycoprotein of human respiratory syncytial virus in a cell line deficient in O glycosylation. *J. Virol.* **63**, 4767-4776

134.   Gruber, C., and Levine, S. (1985) Respiratory syncytial virus polypeptides. IV. The oligosaccharides of the glycoproteins. *J. Gen. Virol.* **66**, 417-432

135.   McDonald, T. P., Jeffree, C. E., Li, P., Rixon, H. W., Brown, G., Aitken, J. D., MacLellan, K., and Sugrue, R. J. (2006) Evidence that maturation of the N-linked glycans of the respiratory syncytial virus (RSV) glycoproteins is required for virus-mediated cell fusion: The effect of alpha-mannosidase inhibitors on RSV infectivity. *Virology* **350**, 289-301

136.   Wertz, G. W., Collins, P. L., Huang, Y., Gruber, C., Levine, S., and Ball, L. A. (1985) Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 4075-4079

137.   Fuentes, S., Coyle, E. M., Golding, H., and Khurana, S. (2015) Nonglycosylated G-Protein Vaccine Protects against Homologous and Heterologous Respiratory Syncytial Virus (RSV) Challenge, while Glycosylated G Enhances RSV Lung Pathology and Cytokine Levels. *J. Virol.* **89**, 8193-8205

138.   Sparer, T. E., Matthews, S., Hussell, T., Rae, A. J., Garcia-Barreno, B., Melero, J. A., and Openshaw, P. J. M. (1998) Eliminating a region of respiratory syncytial virus attachment protein allows induction of protective immunity without vaccine-enhanced lung eosinophilia. *The Journal of Experimental Medicine* **187**, 1921-1926

139.   Zimmer, G., Trotz, I., and Herrler, G. (2001) N-glycans of F protein differentially affect fusion activity of human respiratory syncytial virus. *J. Virol.* **75**, 4744-4751

140. Lambert, D. M. (1988) Role of oligosaccharides in the structure and function of respiratory syncytial virus glycoproteins. *Virology* **164**, 458-466

141. Brooks, S. A. (2009) Strategies for analysis of the glycosylation of proteins: current status and future perspectives. *Mol. Biotechnol.* **43**, 76-88

142. Meyer, B., and Möller, H. (2007) Conformation of Glycopeptides and Glycoproteins. in *Glycopeptides and Glycoproteins: Synthesis, Structure, and Application* (Wittmann, V. ed., Springer Berlin, Heidelberg

143. Guttman, M., Weinkam, P., Sali, A., and Lee, K. K. (2013) All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosylated proteins. *Structure* **21**, 321-331

144. Zaia, J. (2004) Mass spectrometry of oligosaccharides. *Mass Spectrom. Rev.* **23**, 161-227

145. Wuhrer, M., Catalina, M. I., Deelder, A. M., and Hokke, C. H. (2007) Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **849**, 115-128

146. Dalpathado, D. S., and Desaire, H. (2008) Glycopeptide analysis by mass spectrometry. *Analyst* **133**, 731-738

147. Pan, S., Chen, R., Aebersold, R., and Brentnall, T. A. (2011) Mass spectrometry based glycoproteomics-from a proteomics perspective. *Mol. Cell. Proteomics* **10**, R110.003251

148. Schiel, J. E. (2012) Glycoprotein analysis using mass spectrometry: unraveling the layers of complexity. *Anal. Bioanal. Chem.* **404**, 1141-1149

149. Zauner, G., Kozak, R. P., Gardner, R. A., Fernandes, D. L., Deelder, A. M., and Wuhrer, M. (2012) Protein O-glycosylation analysis. *Biol. Chem.* **393**, 687-708

150. Alley, W. R., Jr., Mann, B. F., and Novotny, M. V. (2013) High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem. Rev.* **113**, 2668-2732

151. Fenn, L. S., and McLean, J. A. (2013) Structural separations by ion mobility-MS for glycomics and glycoproteomics. *Methods Mol. Biol.* **951**, 171-194

152. Zauner, G., Selman, M. H., Bondt, A., Rombouts, Y., Blank, D., Deelder, A. M., and Wuhrer, M. (2013) Glycoproteomic analysis of antibodies. *Mol. Cell. Proteomics* **12**, 856-865

153. Khoo, K.-H. (2014) From Mass Spectrometry-Based Glycosylation Analysis to Glycomics and Glycoproteomics. in *Glycobiology of the Nervous System* (Yu, R. K., and Schengrund, C.-L. eds.), Springer New York, NY

154. Thaysen-Andersen, M., and Packer, N. H. (2014) Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim. Biophys. Acta* **1844**, 1437-1452

155. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198-207

156. Banerjee, S., and Mazumdar, S. (2012) Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* **2012**, 282574

157. Kraj, A., Desiderio, D. M., and Nibbering, N. M. (2008) Mass spectrometer's building blocks in *Mass spectrometry: instrumentation, interpretation, and applications* (Desiderio, D. M., and Nibbering, N. M. eds.), John Wiley & Sons, Hoboken, New Jersey pp 15-71

158. Mann, M., and Kelleher, N. L. (2008) Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18132-18138

159. Zubarev, R. A., and Makarov, A. (2013) Orbitrap mass spectrometry. *Anal. Chem.* **85**, 5288-5296

160. Senko, M. W., Remes, P. M., Canterbury, J. D., Mathur, R., Song, Q., Eliuk, S. M., Mullen, C., Earley, L., Hardman, M., Blethrow, J. D., Bui, H., Specht, A., Lange, O., Denisov, E., Makarov, A., Horning, S., and Zabrouskov, V. (2013) Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal. Chem.* **85**, 11710-11714

161. Louris, J. N., Cooks, R. G., Syka, J. E. P., Kelley, P. E., Stafford, G. C., and Todd, J. F. J. (1987) Instrumentation, applications, and energy deposition in quadrupole ion-trap tandem mass spectrometry. *Anal. Chem.* **59**, 1677-1685

162. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **4**, 709-712

163. 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. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9528-9533

164. Steen, H., and Mann, M. (2004) The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5**, 699-711

165. Medzihradszky, K. F., and Chalkley, R. J. (2015) Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spectrom. Rev.* **34**, 43-63

166. Paizs, B., and Suhai, S. (2005) Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **24**, 508-548

167. Roepstorff, P., and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601

168. Biemann, K. (1990) Appendix 5. Nomenclature for peptide fragment ions (positive ions). in *Methods Enzymol.*, Academic Press. pp 886-887

169. Biemann, K. (1992) Mass spectrometry of peptides and proteins. *Annu. Rev. Biochem.* **61**, 977-1010

170. Michalski, A., Neuhauser, N., Cox, J., and Mann, M. (2012) A systematic investigation into the nature of tryptic HCD spectra. *J. Proteome Res.* **11**, 5479-5491

171. Xia, Y., Liang, X., and McLuckey, S. A. (2006) Ion trap versus low-energy beam-type collision-induced dissociation of protonated ubiquitin ions. *Anal. Chem.* **78**, 1218-1227

172. Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E., Rad, R., and Gygi, S. P. (2011) Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol. Cell. Proteomics* **10**, M111 009910

173. Sweredoski, M. J., Pekar Second, T., Broeker, J., Moradian, A., and Hess, S. (2015) High resolution data-independent acquisition with electron transfer dissociation mass spectrometry: Multiplexed analysis of post-translationally modified proteins. *Int. J. Mass. Spectrom.* **390**, 155-162

174. Coon, J. J., Syka, J. E. P., Schwartz, J. C., Shabanowitz, J., and Hunt, D. F. (2004) Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *Int. J. Mass. Spectrom.* **236**, 33-42

175. Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J., and Hunt, D. F. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta* **1764**, 1811-1822

176. Zhurov, K. O., Fornelli, L., Wodrich, M. D., Laskay, U. A., and Tsybin, Y. O. (2013) Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem. Soc. Rev.* **42**, 5014-5030

177. Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **120**, 3265-3266

178. 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*. *Anal. Chem.* **77**, 1831-1839

179. Hunt, D. F., Shabanowitz, J., and Bai, D. L. (2015) Peptide sequence analysis by electron transfer dissociation mass spectrometry: A web-based tutorial. *J. Am. Soc. Mass Spectrom.* **26**, 1256-1258

180. Good, D. M., Wenger, C. D., McAlister, G. C., Bai, D. L., Hunt, D. F., and Coon, J. J. (2009) Post-acquisition ETD spectral processing for increased peptide identifications. *J. Am. Soc. Mass Spectrom.* **20**, 1435-1440

181. Ma, B., and Johnson, R. (2012) De novo sequencing and homology searching. *Mol. Cell. Proteomics* **11**, O111 014902

182. Zaia, J. (2008) Mass spectrometry and the emerging field of glycomics. *Chem. Biol.* **15**, 881-892

183. Raman, R., Raguram, S., Venkataraman, G., Paulson, J. C., and Sasisekharan, R. (2005) Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Meth* **2**, 817-824

184. Rakus, J. F., and Mahal, L. K. (2011) New technologies for glycomic analysis: toward a systematic understanding of the glycome. *Annu. Rev. Anal. Chem. (Palo Alto Calif.)* **4**, 367-392

185. Ruhaak, L. R., Zauner, G., Huhn, C., Bruggink, C., Deelder, A. M., and Wuhrer, M. (2010) Glycan labeling strategies and their use in identification and quantification. *Anal. Bioanal. Chem.* **397**, 3457-3481

186. Harvey, D. J. (2015) Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2011-2012. *Mass Spectrom. Rev.*, 1-168

187. Carr, S. A., and Roberts, G. D. (1986) Carbohydrate mapping by mass spectrometry: a novel method for identifying attachment sites of Asn-linked sugars in glycoproteins. *Anal. Biochem.* **157**, 396-406

188. Christiansen, M. N., Kolarich, D., Nevalainen, H., Packer, N. H., and Jensen, P. H. (2010) Challenges of determining O-glycopeptide heterogeneity: a fungal glucanase model system. *Anal. Chem.* **82**, 3500-3509

189. Harvey, D. J. (2005) Fragmentation of Negative Ions from Carbohydrates: Part 3. Fragmentation of Hybrid and Complex N-Linked Glycans. *J. Am. Soc. Mass Spectrom.* **16**, 647-659

190. Jensen, P. H., Karlsson, N. G., Kolarich, D., and Packer, N. H. (2012) Structural analysis of N- and O-glycans released from glycoproteins. *Nat. Protoc.* **7**, 1299-1310

191. Thobhani, S., Yuen, C.-T., Bailey, M. J. A., and Jones, C. (2009) Identification and quantification of N-linked oligosaccharides released from glycoproteins: An inter-laboratory study. *Glycobiology* **19**, 201-211

192. Moh, E. S., Thaysen-Andersen, M., and Packer, N. H. (2015) Relative versus absolute quantitation in disease glycomics. *Proteomics Clin. Appl.* **9**, 368-382

193. Rosati, S., Yang, Y., Barendregt, A., and Heck, A. J. R. (2014) Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat. Protoc.* **9**, 967-976

194.	Guerrero, A., Lerno, L., Barile, D., and Lebrilla, C. B. (2015) Top-down analysis of highly post-translationally modified peptides by Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **26**, 453-459

195.	Halim, A., Carlsson, M. C., Madsen, C. B., Brand, S., Moller, S. R., Olsen, C. E., Vakhrushev, S. Y., Brimnes, J., Wurtzen, P. A., Ipsen, H., Petersen, B. L., and Wandall, H. H. (2015) Glycoproteomic analysis of seven major allergenic proteins reveals novel post-translational modifications. *Mol. Cell. Proteomics* **14**, 191-204

196.	Sumer-Bayraktar, Z., Nguyen-Khuong, T., Jayo, R., Chen, D. D. Y., Ali, S., Packer, N. H., and Thaysen-Andersen, M. (2012) Micro- and macroheterogeneity of N-glycosylation yields size and charge isoforms of human sex hormone binding globulin circulating in serum. *Proteomics* **12**, 3315-3327

197.	Medzihradszky, K. F., Gillece-Castro, B. L., Settineri, C. A., Reid Townsend, R., Masiarz, F. R., and Burlingame, A. L. (1990) Structure determination of O-linked glycopeptides by tandem mass spectrometry. *Biol. Mass Spectrom.* **19**, 777-781

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

199.	Huddleston, M., Bean, M., and Carr, S. (1993) Collisional fragmentation of glycopeptides by ESI LC/MS and LC/MS/MS: Methods for selective detection of glycopeptides in protein digests. *Anal. Chem.* **65**, 877-884

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

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

202.	Hogan, J. M., Pitteri, S. J., Chrisman, P. A., and McLuckey, S. A. (2005) Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J. Proteome Res.* **4**, 628-632

203.	Segu, Z. M., and Mechref, Y. (2010) Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun. Mass Spectrom.* **24**, 1217-1225

204.	Hart-Smith, G., and Raftery, M. J. (2012) Detection and characterization of low abundance glycopeptides via higher-energy C-trap dissociation and orbitrap mass analysis. *J. Am. Soc. Mass Spectrom.* **23**, 124-140

205. Alley, W. R., Jr., Mechref, Y., and Novotny, M. V. (2009) Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun. Mass Spectrom.* **23**, 161-170

206. Scott, N. E., Parker, B. L., Connolly, A. M., Paulech, J., Edwards, A. V., Crossett, B., Falconer, L., Kolarich, D., Djordjevic, S. P., Hojrup, P., Packer, N. H., Larsen, M. R., and Cordwell, S. J. (2011) Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobacter jejuni. *Mol. Cell. Proteomics* **10**, M000031-MCP000201

207. Catalina, M. I., Koeleman, C. A. M., Deelder, A. M., and Wuhrer, M. (2007) Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain. *Rapid Commun. Mass Spectrom.* **21**, 1053-1061

208. Viner, R. I., Zhang, T., Second, T., and Zabrouskov, V. (2009) Quantification of post-translationally modified peptides of bovine alpha-crystallin using tandem mass tags and electron transfer dissociation. *J. Proteomics* **72**, 874-885

209. Zhao, P., Viner, R., Teo, C. F., Boons, G. J., Horn, D., and Wells, L. (2011) Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J. Proteome Res.* **10**, 4088-4104

210. Halim, A., Ruetschi, U., Larson, G., and Nilsson, J. (2013) LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J. Proteome Res.* **12**, 573-584

211. Myers, S. A., Daou, S., Affar el, B., and Burlingame, A. (2013) Electron transfer dissociation (ETD): the mass spectrometric breakthrough essential for O-GlcNAc protein site assignments-a study of the O-GlcNAcylated protein host cell factor C1. *Proteomics* **13**, 982-991

212. Williams, J. P., Pringle, S., Richardson, K., Gethings, L., Vissers, J. P., De Cecco, M., Houel, S., Chakraborty, A. B., Yu, Y. Q., Chen, W., and Brown, J. M. (2013) Characterisation of glycoproteins using a quadrupole time-of-flight mass spectrometer configured for electron transfer dissociation. *Rapid Commun. Mass Spectrom.* **27**, 2383-2390

213. Trinidad, J. C., Schoepfer, R., Burlingame, A. L., and Medzihradszky, K. F. (2013) N- and O-glycosylation in the murine synaptosome. *Mol. Cell. Proteomics* **12**, 3474-3488

214. Hinneburg, H., Stavenhagen, K., Schweiger-Hufnagel, U., Pengelley, S., Jabs, W., Seeberger, P. H., Silva, D. V., Wuhrer, M., and Kolarich, D. (2016) The art of destruction:

Optimizing collision energies in Quadrupole-Time of Flight (Q-TOF) instruments for glycopeptide-based glycoproteomics. *J. Am. Soc. Mass Spectrom.* **27**, 507-519

215. Kolli, V., and Dodds, E. D. (2014) Energy-resolved collision-induced dissociation pathways of model N-linked glycopeptides: implications for capturing glycan connectivity and peptide sequence in a single experiment. *Analyst* **139**, 2144-2153

216. Cao, Q., Zhao, X., Zhao, Q., Lv, X., Ma, C., Li, X., Zhao, Y., Peng, B., Ying, W., and Qian, X. (2014) Strategy integrating stepped fragmentation and glycan diagnostic ion-based spectrum refinement for the identification of core fucosylated glycoproteome using mass spectrometry. *Anal. Chem.* **86**, 6804-6811

217. Domon, B., and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5**, 397-409

218. Vékey, K., Ozohanics, O., Tóth, E., Jekő, A., Révész, Á., Krenyácz, J., and Drahos, L. (2013) Fragmentation characteristics of glycopeptides. *Int. J. Mass. Spectrom.* **345–347**, 71-79

219. Yu, J., Schorlemer, M., Gomez Toledo, A., Pett, C., Sihlbom, C., Larson, G., Westerlind, U., and Nilsson, J. (2016) Distinctive MS/MS Fragmentation Pathways of Glycopeptide-Generated Oxonium Ions Provide Evidence of the Glycan Structure. *Chemistry* **22**, 1114-1124

220. Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Ruetschi, U., Brinkmalm, G., Sihlbom, C., Lengqvist, J., Larson, G., and Nilsson, J. (2014) Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome Res.* **13**, 6024-6032

221. Marino, F., Bern, M., Mommen, G. P., Leney, A. C., van Gaans-van den Brink, J. A., Bonvin, A. M., Becker, C., van Els, C. A., and Heck, A. J. (2015) Extended O-GlcNAc on HLA Class-I-Bound Peptides. *J. Am. Chem. Soc.* **137**, 10922-10925

222. Nilsson, J. (2016) Liquid chromatography-tandem mass spectrometry-based fragmentation analysis of glycopeptides. *Glycoconj. J.* **33**, 261-272

223. Perdivara, I., Petrovich, R., Allinquant, B., Deterding, L. J., Tomer, K. B., and Przybylski, M. (2009) Elucidation of O-glycosylation structures of the beta-amyloid precursor protein by liquid chromatography-mass spectrometry using electron transfer dissociation and collision induced dissociation. *J. Proteome Res.* **8**, 631-642

224. Horn, D. M., Ge, Y., and McLafferty, F. W. (2000) Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal. Chem.* **72**, 4778-4784

225. 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. *Anal. Chem.* **79**, 477-485

226. Frese, C. K., Altelaar, A. F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A. J., and Mohammed, S. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **84**, 9668-9673

227. Parker, B. L., Thaysen-Andersen, M., Fazakerley, D. J., Holliday, M., Packer, N. H., and James, D. E. (2015) Terminal galactosylation and sialylation switching on membrane glycoproteins upon TNF-alpha-induced insulin resistance in adipocytes. *Mol. Cell. Proteomics* **15**, 141-153

228. Saba, J., Dutta, S., Hemenway, E., and Viner, R. (2012) Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *Int. J. Proteomics* **560391**, 1-7

229. Singh, C., Zampronio, C. G., Creese, A. J., and Cooper, H. J. (2012) Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J. Proteome Res.* **11**, 4517-4525

230. Yin, X., Bern, M., Xing, Q., Ho, J., Viner, R., and Mayr, M. (2013) Glycoproteomic analysis of the secretome of human endothelial cells. *Mol. Cell. Proteomics* **12**, 956-978

231. Wu, S. W., Pu, T. H., Viner, R., and Khoo, K. H. (2014) Novel LC-MS(2) product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal. Chem.* **86**, 5478-5486

232. Stavenhagen, K., Hinneburg, H., Thaysen-Andersen, M., Hartmann, L., Varon Silva, D., Fuchser, J., Kaspar, S., Rapp, E., Seeberger, P. H., and Kolarich, D. (2013) Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J. Mass Spectrom.* **48**, 627-639

233. Moh, E. S., Lin, C. H., Thaysen-Andersen, M., and Packer, N. H. (2016) Site-specific N-glycosylation of recombinant pentameric and hexameric human IgM. *J. Am. Soc. Mass Spectrom.* **27**, 1143-1155

234. Ongay, S., Boichenko, A., Govorukhina, N., and Bischoff, R. (2012) Glycopeptide enrichment and separation for protein glycosylation analysis. *J. Sep. Sci.* **35**, 2341-2372

235. Larsen, M. R., Jensen, S. S., Jakobsen, L. A., and Heegaard, N. H. (2007) Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. *Mol. Cell. Proteomics* **6**, 1778-1787

236. Palmisano, G., Lendal, S. E., Engholm-Keller, K., Leth-Larsen, R., Parker, B. L., and Larsen, M. R. (2010) Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **5**, 1974-1982

237. Darula, Z., Sherman, J., and Medzihradszky, K. F. (2012) How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol. Cell. Proteomics* **11**, O111.016774

238. Noborn, F., Gomez Toledo, A., Sihlbom, C., Lengqvist, J., Fries, E., Kjellen, L., Nilsson, J., and Larson, G. (2015) Identification of chondroitin sulfate linkage region glycopeptides reveals prohormones as a novel class of proteoglycans. *Mol. Cell. Proteomics* **14**, 41-49

239. Darula, Z., Sarnyai, F., and Medzihradszky, K. F. (2016) O-glycosylation sites identified from mucin core-1 type glycopeptides from human serum. *Glycoconj. J.* **33**, 435-445

240. Mysling, S., Palmisano, G., Hojrup, P., and Thaysen-Andersen, M. (2010) Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal. Chem.* **82**, 5598-5609

241. Selman, M. H., Hemayatkar, M., Deelder, A. M., and Wuhrer, M. (2011) Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. *Anal. Chem.* **83**, 2492-2499

242. Jensen, P. H., Mysling, S., Hojrup, P., and Jensen, O. N. (2013) Glycopeptide enrichment for MALDI-TOF mass spectrometry analysis by hydrophilic interaction liquid chromatography solid phase extraction (HILIC SPE). *Methods Mol. Biol.* **951**, 131-144

243. Tang, H., Mayampurath, A., Yu, C. Y., and Mechref, Y. (2014) Bioinformatics protocols in glycomics and glycoproteomics. *Curr. Protoc. Protein Sci.* **76**, 2 15 11-17

244. Dallas, D. C., Martin, W. F., Hua, S., and German, J. B. (2013) Automated glycopeptide analysis-review of current state and future directions. *Brief. Bioinform.* **14**, 361-374

245. Aoki-Kinoshita, K. F. (2013) Introduction to informatics in glycoprotein analysis. *Methods Mol. Biol.* **951**, 257-267

246. Baycin Hizal, D., Wolozny, D., Colao, J., Jacobson, E., Tian, Y., Krag, S. S., Betenbaugh, M. J., and Zhang, H. (2014) Glycoproteomic and glycomic databases. *Clin. Proteomics*. 10.1186/1559-0275-11-15

247. Liu, G., and Neelamegham, S. (2015) Integration of systems glycobiology with bioinformatics toolboxes, glycoinformatics resources, and glycoproteomics data. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **7**, 163-181

248. Hu, H., Khatri, K., Klein, J., Leymarie, N., and Zaia, J. (2016) A review of methods for interpretation of glycopeptide tandem mass spectral data. *Glycoconj. J.* **33**, 285-296

249. Lee, L. Y., Moh, E. S., Parker, B. L., Bern, M., Packer, N. H., and Thaysen-Andersen, M. (2016) Toward Automated N-Glycopeptide Identification in Glycoproteomics. *J. Proteome Res.* **15**, 3904-3915

250. Desaire, H., and Hua, D. (2009) When can glycopeptides be assigned based solely on high-resolution mass spectrometry data? *Int. J. Mass. Spectrom.* **287**, 21-26

251. Medzihradszky, K. F., Kaasik, K., and Chalkley, R. J. (2015) Characterizing sialic acid variants at the glycopeptide level. *Anal. Chem.* **87**, 3064-3071

252. Darula, Z., and Medzihradszky, K. F. (2014) Glycan side reaction may compromise ETD-based glycopeptide identification. *J. Am. Soc. Mass Spectrom.* **25**, 977-987

253. Darula, Z., and Medzihradszky, K. F. (2015) Carbamidomethylation Side Reactions May Lead to Glycan Misassignments in Glycopeptide Analysis. *Anal. Chem.* **87**, 6297-6302

254. Medzihradszky, K. F. (2014) Noncovalent dimer formation in liquid chromatography-mass spectrometry analysis. *Anal. Chem.* **86**, 8906-8909

255. Deshpande, N., Jensen, P. H., Packer, N. H., and Kolarich, D. (2010) GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J. Proteome Res.* **9**, 1063-1075

256. Keil, B. (1992) Essential Substrate Residues for Action of Endopeptidases. in *Specificity of Proteolysis*, 1 Ed., Springer-Verlag Berlin-Heidelberg-New York. pp 43-229

257. Gorman, J. J. (1987) Fluorescent labeling of cysteinyl residues to facilitate electrophoretic isolation of proteins suitable for amino-terminal sequence analysis. *Anal. Biochem.* **160**, 376-387

258. Gorman, J. J., Corino, G. L., and Mitchell, S. J. (1987) Fluorescent labeling of cysteinyl residues. *Eur. J. Biochem.* **168**, 169-179

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

260. Dave, K. A., Headlam, M. J., Wallis, T. P., and Gorman, J. J. (2011) Preparation and analysis of proteins and peptides using MALDI TOF/TOF mass spectrometry. *Curr. Protoc. Protein Sci.* **63**, 16.13.11-16.13.21

261. Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: Advanced Peptide and Protein Identification Software. *Curr. Protoc. Bioinformatics* **UNIT 13.20**, 13.20.11-13.20.14.

262. He, L., Xin, L., Shan, B., Lajoie, G. A., and Ma, B. (2014) GlycoMaster DB: software to assist the automated identification of N-linked glycopeptides by tandem mass spectrometry. *J. Proteome Res.* **13**, 3881-3895

263. Chandler, K. B., Pompach, P., Goldman, R., and Edwards, N. (2013) Exploring site-specific N-glycosylation microheterogeneity of haptoglobin using glycopeptide CID tandem mass spectra and glycan database search. *J. Proteome Res.* **12**, 3652-3666

264. Cooper, C. A., Gasteiger, E., and Packer, N. H. (2001) GlycoMod – A software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **1**, 340-349

265. Bern, M., Cai, Y., and Goldberg, D. (2007) Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal. Chem.* **79**, 1393-1400

266. Campbell, M. P., Peterson, R., Mariethoz, J., Gasteiger, E., Akune, Y., Aoki-Kinoshita, K. F., Lisacek, F., and Packer, N. H. (2014) UniCarbKB: building a knowledge platform for glycoproteomics. *Nucleic Acids Res.* **42**, D215-221

267. Gomez Toledo, A., Nilsson, J., Noborn, F., Sihlbom, C., and Larson, G. (2015) Positive mode LC-MS/MS analysis of chondroitin sulfate modified glycopeptides derived from light and heavy chains of the human inter-alpha-trypsin inhibitor complex. *Mol. Cell. Proteomics* **14**, 3118-3131

268. Parker, B. L., Thaysen-Andersen, M., Solis, N., Scott, N. E., Larsen, M. R., Graham, M. E., Packer, N. H., and Cordwell, S. J. (2013) Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. *J. Proteome Res.* **12**, 5791-5800

269. Everest-Dass, A. V., Abrahams, J. L., Kolarich, D., Packer, N. H., and Campbell, M. P. (2013) Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J. Am. Soc. Mass Spectrom.* **24**, 895-906

270. Patiny, L., and Borel, A. (2013) ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. *J. Chem. Inf. Model.* **53**, 1223-1228

271. Zhu, Z., Hua, D., Clark, D. F., Go, E. P., and Desaire, H. (2013) GlycoPep Detector: a tool for assigning mass spectrometry data of N-linked glycopeptides on the basis of their electron transfer dissociation spectra. *Anal. Chem.* **85**, 5023-5032

272. Strum, J. S., Nwosu, C. C., Hua, S., Kronewitter, S. R., Seipert, R. R., Bachelor, R. J., An, H. J., and Lebrilla, C. B. (2013) Automated assignments of N- and O-site specific glycosylation with extensive glycan heterogeneity of glycoprotein mixtures. *Anal. Chem.* **85**, 5666-5675

273. Woodin, C. L., Hua, D., Maxon, M., Rebecchi, K. R., Go, E. P., and Desaire, H. (2012) GlycoPep grader: a web-based utility for assigning the composition of N-linked glycopeptides. *Anal. Chem.* **84**, 4821-4829

274. Simmons, G. C. (1967) The isolation of Newcastle disease virus in Queensland. *Aust. Vet. J.* **43**, 29-30

275. Chen, L., Colman, P. M., Cosgrove, L. J., Lawrence, M. C., Lawrence, L. J., Tulloch, P. A., and Gorman, J. J. (2001) Cloning, expression, and crystallization of the fusion protein of Newcastle disease virus. *Virology* **290**, 290-299

276. Chen, L., Gorman, J. J., McKimm-Breschkin, J., Lawrence, L. J., Tulloch, P. A., Smith, B. J., Colman, P. M., and Lawrence, M. C. (2001) The structure of the fusion glycoprotein of Newcastle disease virus suggests a novel paradigm for the molecular mechanism of membrane fusion. *Structure* **9**, 255-266

277. Webster, A. F., Taylor, J. H., and Barnes, J. M. (1970) The efficiency of Australian Newcastle disease virus for vaccine production. *Aust. Vet. J.* **46**, 540-541

278. UniProt-Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204-212

279. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539

280. Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951-960

281. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372

282. Yusoff, K., Tan, W. S., Lau, C. H., Ng, B. K., and Ibrahim, A. L. (1996) Sequence of the haemagglutinin-neuraminidase gene of the Newcastle disease virus oral vaccine strain V4(UPM). *Avian Pathol.* **25**, 837-844

283. Kattenbelt, J. A., Meers, J., and Gould, A. R. (2006) Genome sequence of the thermostable Newcastle disease virus (strain I-2) reveals a possible phenotypic locus. *Vet. Microbiol.* **114**, 134-141

284. Good, D. M., Wirtala, M., McAlister, G. C., and Coon, J. J. (2007) Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* **6**, 1942-1951

285. Hodder, A. N., Selleck, P. W., White, J. R., and Gorman, J. J. (1993) Analysis of pathotype-specific structural features and cleavage activation of Newcastle disease virus membrane glycoproteins using antipeptide antibodies. *J. Gen. Virol.* **74** 1081-1091

286. Hodder, A. N., Liu, Z. Y., Selleck, P. W., Corino, G. L., Shiell, B. J., Grix, D. C., Morrow, C. J., and Gorman, J. J. (1994) Characterization of Field Isolates of Newcastle Disease Virus Using Antipeptide Antibodies. *Avian Dis.* **38**, 103-118

287. Khatri, K., Klein, J. A., White, M. R., Grant, O. C., Leymarie, N., Woods, R. J., Hartshorn, K. L., and Zaia, J. (2016) Integrated omics and computational glycobiology reveal structural basis for influenza A virus glycan microheterogeneity and host interactions. *Mol. Cell. Proteomics* **15**, 1895-1912

288. An, Y., and Cipollo, J. F. (2011) An unbiased approach for analysis of protein glycosylation and application to influenza vaccine hemagglutinin. *Anal. Biochem.* **415**, 67-80

289. Blake, T. A., Williams, T. L., Pirkle, J. L., and Barr, J. R. (2009) Targeted N-linked glycosylation analysis of H5N1 influenza hemagglutinin by selective sample preparation and liquid chromatography/tandem mass spectrometry. *Anal. Chem.* **81**, 3109-3118

290. Crispin, M., and Doores, K. J. (2015) Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design. *Curr. Opin. Virol.* **11**, 63-69

291. Schwarzer, J., Rapp, E., Hennig, R., Genzel, Y., Jordan, I., Sandig, V., and Reichl, U. (2009) Glycan analysis in cell culture-based influenza vaccine production: influence of host cell line and virus strain on the glycosylation pattern of viral hemagglutinin. *Vaccine* **27**, 4325-4336

292. Welch, B. D., Yuan, P., Bose, S., Kors, C. A., Lamb, R. A., and Jardetzky, T. S. (2013) Structure of the parainfluenza virus 5 (PIV5) hemagglutinin-neuraminidase (HN) ectodomain. *PLoS Pathog.* **9**, e1003534

293. Morrison, T. G., McQuain, C., O'Connell, K. F., and McGinnes, L. W. (1990) Mature, cell-associated HN protein of Newcastle disease virus exists in two forms differentiated by posttranslational modifications. *Virus Res.* **15**, 113-133

294. Ferreira, L., Villar, E., and Munoz-Barroso, I. (2004) Gangliosides and N-glycoproteins function as Newcastle disease virus receptors. *Int. J. Biochem. Cell Biol.* **36**, 2344-2356

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

296. Paulson, J. C., Weinstein, J., Dorland, L., van Halbeek, H., and Vliegenthart, J. F. (1982) Newcastle disease virus contains a linkage-specific glycoprotein sialidase. Application to the localization of sialic acid residues in N-linked oligosaccharides of alpha 1-acid glycoprotein. *J. Biol. Chem.* **257**, 12734-12738

297. Smith, G. W., and Hightower, L. E. (1983) Biological consequences of neuraminidase deficiency in Newcastle disease virus. *J. Virol.* **47**, 385-391

298. Yoshima, H., Nakanishi, M., Okada, Y., and Kobata, A. (1981) Carbohydrate structures of HVJ (Sendai virus) glycoproteins. *J. Biol. Chem.* **256**, 5355-5361

299. Sakaguchi, T., Toyoda, T., Gotoh, B., Inocencio, N. M., Kuma, K., Miyata, T., and Nagai, Y. (1989) Newcastle disease virus evolution. I. Multiple lineages defined by sequence variability of the hemagglutinin-neuraminidase gene. *Virology* **169**, 260-272

300. Kattenbelt, J. A., Stevens, M. P., and Gould, A. R. (2006) Sequence variation in the Newcastle disease virus genome. *Virus Res.* **116**, 168-184

301. Brindley, M. A., Suter, R., Schestak, I., Kiss, G., Wright, E. R., and Plemper, R. K. (2013) A stabilized headless measles virus attachment protein stalk efficiently triggers membrane fusion. *J. Virol.* **87**, 11693-11703

302. Bensink, Z., and Spradbrow, P. (1999) Newcastle disease virus strain I2 – a prospective thermostable vaccine for use in developing countries. *Vet. Microbiol.* **68**, 131-139

303. Spradbrow, P. B., MacKenzie, M., and Grimes, S. E. (1995) Recent isolates of Newcastle disease virus in Australia. *Vet. Microbiol.* **46**, 21-28

304. Wambura, P., Meers, J., and Spradbrow, P. (2006) Development of a cell culture method for quantal assay of strain I-2 of Newcastle disease virus. *Vet. Res. Commun.* **30**, 689-696

305. Morrison, T. G. (1988) Structure, function, and intracellular processing of paramyxovirus membrane proteins. *Virus Res.* **10**, 113-135

306. Panda, A., Huang, Z., Elankumaran, S., Rockemann, D. D., and Samal, S. K. (2004) Role of fusion protein cleavage site in the virulence of Newcastle disease virus. *Microb. Pathog.* **36**, 1-10

307. de Leeuw, O. S., Koch, G., Hartog, L., Ravenshorst, N., and Peeters, B. P. (2005) Virulence of Newcastle disease virus is determined by the cleavage site of the fusion protein and by both the stem region and globular head of the haemagglutinin-neuraminidase protein. *J. Gen. Virol.* **86**, 1759-1769

308. Toyoda, T., Sakaguchi, T., Hirota, H., Gotoh, B., Kuma, K., Miyata, T., and Nagai, Y. (1989) Newcastle disease virus evolution. II. Lack of gene recombination in generating virulent and avirulent strains. *Virology* **169**, 273-282

309. Prehm, P., Scheid, A., and Choppin, P. W. (1979) The carbohydrate structure of the glycoproteins of the paramyxovirus SV5 grown in bovine kidney cells. *J. Biol. Chem.* **254**, 9669-9677

310. Olsen, J. V., de Godoy, L. M. F., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010-2021

311. Nagai, Y., Ogura, H., and Klenk, H.-D. (1976) Studies on the assembly of the envelope of Newcastle disease virus. *Virology* **69**, 523-538

312. Stone-Hulslander, J., and Morrison, T. G. (1997) Detection of an interaction between the HN and F proteins in Newcastle disease virus-infected cells. *J. Virol.* **71**, 6287-6295

313. Morrison, T., Ward, L. J., and Semerjian, A. (1985) Intracellular processing of the Newcastle disease virus fusion glycoprotein. *J. Virol.* **53**, 851-857

314. Anderson, K., Stott, E. J., and Wertz, G. W. (1992) Intracellular processing of the human respiratory syncytial virus fusion glycoprotein: amino acid substitutions affecting folding, transport and cleavage. *J. Gen. Virol.* **73**, 1177-1188

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

316. Diabaté, S., Geyer, R., and Stirm, S. (1983) Separation and sugar component analysis of the oligosaccharides in the surface glycoproteins of Newcastle Disease Virus. *Arch. Virol.* **76**, 321-334

317. Morrison, T., and Portner, A. (1991) Structure, function, and intracellular processing of the glycoproteins of paramyxoviridae. in *The Paramyxoviruses* (Kingsbury, D. W. ed.), Springer US, Boston, MA. pp 347-382

318. Bagai, S., and Lamb, R. A. (1995) Individual roles of N-linked oligosaccharide chains in intracellular transport of the paramyxovirus SV5 fusion protein. *Virology* **209**, 250-256

319. Segawa, H., Yamashita, T., Kawakita, M., and Taira, H. (2000) Functional analysis of the individual oligosaccharide chains of sendai virus fusion protein. *J. Biochem.* **128**, 65-72

320. Alkhatib, G., Shen, S. H., Briedis, D., Richardson, C., Massie, B., Weinberg, R., Smith, D., Taylor, J., Paoletti, E., and Roder, J. (1994) Functional analysis of N-linked glycosylation mutants of the measles virus fusion protein synthesized by recombinant vaccinia virus vectors. *J. Virol.* **68**, 1522-1531

321. von Messling, V., and Cattaneo, R. (2003) N-linked glycans with similar location in the fusion protein head modulate paramyxovirus fusion. *J. Virol.* **77**, 10202-10212

322. Moll, M., Kaufmann, A., and Maisner, A. (2004) Influence of N-glycans on processing and biological activity of the nipah virus fusion protein. *J. Virol.* **78**, 7274-7278

323. Carter, J. R., Pager, C. T., Fowler, S. D., and Dutch, R. E. (2005) Role of N-linked glycosylation of the Hendra virus fusion protein. *J. Virol.* **79**, 7922-7925

324. Swanson, K., Wen, X., Leser, G. P., Paterson, R. G., Lamb, R. A., and Jardetzky, T. S. (2010) Structure of the Newcastle disease virus F protein in the post-fusion conformation. *Virology* **402**, 372

325. Yin, H. S., Paterson, R. G., Wen, X., Lamb, R. A., and Jardetzky, T. S. (2005) Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. *Proc. Natl. Acad. Sci. U. S. A.* **102**

326. McLellan, J. S., Yang, Y., Graham, B. S., and Kwong, P. D. (2011) Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. *J. Virol.* **85**, 7788-7796

327. El Najjar, F., Schmitt, A. P., and Dutch, R. E. (2014) Paramyxovirus glycoprotein incorporation, assembly and budding: a three way dance for infectious particle production. *Viruses* **6**, 3019-3054

328. McLellan, J. S., Chen, M., Leung, S., Graepel, K. W., Du, X., Yang, Y., Zhou, T., Baxa, U., Yasuda, E., Beaumont, T., Kumar, A., Modjarrad, K., Zheng, Z., Zhao, M., Xia, N., Kwong, P. D., and Graham, B. S. (2013) Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113-1117

329. Wong, J. J., Paterson, R. G., Lamb, R. A., and Jardetzky, T. S. (2016) Structure and stabilization of the Hendra virus F glycoprotein in its prefusion form. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1056-1061

330. Collins, P. L., and Mottet, G. (1991) Post-translational processing and oligomerization of the fusion glycoprotein of human respiratory syncytial virus. *J. Gen. Virol.* **72**, 3095-3101

331. Hanisch, F. G., and Breloy, I. (2009) Protein-specific glycosylation: signal patches and cis-controlling peptidic elements. *Biol. Chem.* **390**, 619-626

332. McLellan, J. S., Chen, M., Joyce, M. G., Sastry, M., Stewart-Jones, G. B., Yang, Y., Zhang, B., Chen, L., Srivatsan, S., Zheng, A., Zhou, T., Graepel, K. W., Kumar, A., Moin, S., Boyington, J. C., Chuang, G. Y., Soto, C., Baxa, U., Bakker, A. Q., Spits, H., Beaumont, T., Zheng, Z., Xia, N., Ko, S. Y., Todd, J. P., Rao, S., Graham, B. S., and Kwong, P. D. (2013) Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* **342**, 592-598

333. Geoghegan, K. F., Hoth, L. R., Tan, D. H., Borzilleri, K. A., Withka, J. M., and Boyd, J. G. (2002) Cyclization of N-terminal S-carbamoylmethylcysteine causing loss of 17 Da from peptides and extra peaks in peptide maps. *J. Proteome Res.* **1**, 181-187

334. Olsen, J. V., Ong, S. E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **3**, 608-614

335. Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L., and Gygi, S. P. (2015) A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat. Biotechnol.* **33**, 743-749

336. Reimer, J., Shamshurin, D., Harder, M., Yamchuk, A., Spicer, V., and Krokhin, O. V. (2011) Effect of cyclization of N-terminal glutamine and carbamidomethyl-cysteine (residues) on the chromatographic behavior of peptides in reversed-phase chromatography. *J. Chromatogr. A* **1218**, 5101-5107

337. Krokhin, O. V., Ens, W., and Standing, K. G. (2003) Characterizing degradation products of peptides containing N-terminal Cys residues by (off-line high-performance liquid chromatography)/matrix-assisted laser desorption/ionization quadrupole time-of-flight measurements. *Rapid Commun. Mass Spectrom.* **17**, 2528-2534

338. Gotoh, M., Sato, T., Kiyohara, K., Kameyama, A., Kikuchi, N., Kwon, Y.-D., Ishizuka, Y., Iwai, T., Nakanishi, H., and Narimatsu, H. (2004) Molecular cloning and characterization of β1,4-N-acetylgalactosaminyltransferases IV synthesizing N,N′-diacetyllactosediamine. *FEBS Lett.* **562**, 134-140

339. Hanisch, F. G., Ragge, H., Kalinski, T., Meyer, F., Kalbacher, H., and Hoffmann, W. (2013) Human gastric TFF2 peptide contains an N-linked fucosylated N,N'-diacetyllactosediamine (LacdiNAc) oligosaccharide. *Glycobiology* **23**, 2-11

340. Zeck, A., Pohlentz, G., Schlothauer, T., Peter-Katalinic, J., and Regula, J. T. (2011) Cell type-specific and site directed N-glycosylation pattern of FcγRIIIa. *J. Proteome Res.* **10**, 3031-3039

341. Teng, M. N., and Collins, P. L. (1998) Identification of the respiratory syncytial virus proteins required for formation and passage of helper-dependent infectious particles. *J. Virol.* **72**

342. Battles, M. B., Langedijk, J. P., Furmanova-Hollenstein, P., Chaiwatpongsakorn, S., Costello, H. M., Kwanten, L., Vranckx, L., Vink, P., Jaensch, S., Jonckers, T. H. M., Koul, A., Arnoult, E., Peeples, M. E., Roymans, D., and McLellan, J. S. (2016) Molecular mechanism of respiratory syncytial virus fusion inhibitors. *Nat. Chem. Biol.* **12**, 87-93

343. Khoo, K.-H., and Yu, S.-Y. (2010) Mass Spectrometric Analysis of Sulfated N- and O-Glycans. *Methods Enzymol.* **478**, 3-26

344. Zhang, Y., Jiang, H., Go, E. P., and Desaire, H. (2006) Distinguishing phosphorylation and sulfation in carbohydrates and glycoproteins using ion-pairing and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **17**, 1282-1288

345. Bonar, D., and Hanisch, F. G. (2014) Trefoil factor family domains represent highly efficient conformational determinants for N-linked N,N'-di-N-acetyllactosediamine (LacdiNAc) synthesis. *J. Biol. Chem.* **289**, 29677-29690

346. Miller, E., Fiete, D., Blake, N. M., Beranek, M., Oates, E. L., Mi, Y., Roseman, D. S., and Baenziger, J. U. (2008) A necessary and sufficient determinant for protein-selective glycosylation in vivo. *J. Biol. Chem.* **283**, 1985-1991

347. Skelton, T. P., Kumar, S., Smith, P. L., Beranek, M. C., and Baenziger, J. U. (1992) Pro-opiomelanocortin synthesized by corticotrophs bears asparagine-linked oligosaccharides terminating with SO4-4GalNAc beta 1,4GlcNAc beta 1,2Man alpha. *J. Biol. Chem.* **267**, 12998-13006

348. Mengeling, B. J., Manzella, S. M., and Baenziger, J. U. (1995) A cluster of basic amino acids within an alpha-helix is essential for alpha-subunit recognition by the glycoprotein hormone N-acetylgalactosaminyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 502-506

349. Dharmesh, S. M., Skelton, T. P., and Baenziger, J. U. (1993) Co-ordinate and restricted expression of the ProXaaArg/Lys-specific GalNAc-transferase and the GalNAc beta 1,4GlcNAc beta 1,2Man alpha-4-sulfotransferase. *J. Biol. Chem.* **268**, 17096-17102

350. Green, E. D., van Halbeek, H., Boime, I., and Baenziger, J. U. (1985) Structural elucidation of the disulfated oligosaccharide from bovine lutropin. *J. Biol. Chem.* **260**, 15623-15630

351. Fiete, D., Beranek, M., and Baenziger, J. U. (2012) Molecular basis for protein-specific transfer of N-acetylgalactosamine to N-linked glycans by the glycosyltransferases beta1,4-N-acetylgalactosaminyl transferase 3 (beta4GalNAc-T3) and beta4GalNAc-T4. *J. Biol. Chem.* **287**, 29194-29203

352. Woodworth, A., Fiete, D., and Baenziger, J. U. (2002) Spatial and temporal regulation of tenascin-R glycosylation in the cerebellum. *J. Biol. Chem.* **277**, 50941-50947

353. Fiete, D., Mi, Y., Oats, E. L., Beranek, M. C., and Baenziger, J. U. (2007) N-linked oligosaccharides on the low density lipoprotein receptor homolog SorLA/LR11 are modified with terminal GalNAc-4-SO4 in kidney and brain. *J. Biol. Chem.* **282**, 1873-1881

354. Dell, A., Morris, H. R., Easton, R. L., Panico, M., Patankar, M., Oehniger, S., Koistinen, R., Koistinen, H., Seppala, M., and Clark, G. F. (1995) Structural analysis of the oligosaccharides derived from glycodelin, a human glycoprotein with potent immunosuppressive and contraceptive activities. *J. Biol. Chem.* **270**, 24116-24126

355. Cheng, K., Chen, R., Seebun, D., Ye, M., Figeys, D., and Zou, H. (2014) Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. *J. Proteomics* **110**, 145-154

356. Manzella, S. M., Hooper, L. V., and Baenziger, J. U. (1996) Oligosaccharides containing β1,4-Linked N-acetylgalactosamine, a paradigm for protein-specific glycosylation. *J. Biol. Chem.* **271**, 12117-12120

357. Krarup, A., Truan, D., Furmanova-Hollenstein, P., Bogaert, L., Bouchier, P., Bisschop, I. J., Widjojoatmodjo, M. N., Zahn, R., Schuitemaker, H., McLellan, J. S., and Langedijk, J. P. (2015) A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism. *Nat. Commun.* **6**, 8143

358. Prasanphanich, N. S., Luyai, A. E., Song, X., Heimburg-Molinaro, J., Mandalasi, M., Mickum, M., Smith, D. F., Nyame, A. K., and Cummings, R. D. (2014) Immunization with recombinantly expressed glycan antigens from Schistosoma mansoni induces glycan-specific antibodies against the parasite. *Glycobiology* **24**, 619-637

359. Naus, C. W., van Remoortere, A., Ouma, J. H., Kimani, G., Dunne, D. W., Kamerling, J. P., Deelder, A. M., and Hokke, C. H. (2003) Specific antibody responses to three schistosome-related carbohydrate structures in recently exposed immigrants and established residents in an area of Schistosoma mansoni endemicity. *Infect. Immun.* **71**, 5676-5681

360. Luyai, A. E., Heimburg-Molinaro, J., Prasanphanich, N. S., Mickum, M. L., Lasanajak, Y., Song, X., Nyame, A. K., Wilkins, P., Rivera-Marrero, C. A., Smith, D. F., Van Die, I., Secor, W. E., and Cummings, R. D. (2014) Differential expression of anti-glycan antibodies in schistosome-infected humans, rhesus monkeys and mice. *Glycobiology* **24**, 602-618

361. van Die, I., van Vliet, S. J., Nyame, A. K., Cummings, R. D., Bank, C. M. C., Appelmelk, B., Geijtenbeek, T. B. H., and van Kooyk, Y. (2003) The dendritic cell–specific C-type lectin DC-SIGN is a receptor for Schistosoma mansoni egg antigens and recognizes the glycan antigen Lewis x. *Glycobiology* **13**, 471-478

362. Walther, T., Karamanska, R., Chan, R. W., Chan, M. C., Jia, N., Air, G., Hopton, C., Wong, M. P., Dell, A., Malik Peiris, J. S., Haslam, S. M., and Nicholls, J. M. (2013) Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. *PLoS Pathog.* **9**, e1003223

363. Boyce, T. G., Mellen, B. G., Mitchel, E. F., Jr., Wright, P. F., and Griffin, M. R. (2000) Rates of hospitalization for respiratory syncytial virus infection among children in medicaid. *J. Pediatr.* **137**, 865-870

364. Pasquato, A., Ramos da Palma, J., Galan, C., Seidah, N. G., and Kunz, S. (2013) Viral envelope glycoprotein processing by proprotein convertases. *Antiviral Res.* **99**, 49-60

365. Stansell, E., Panico, M., Canis, K., Pang, P. C., Bouche, L., Binet, D., O'Connor, M. J., Chertova, E., Bess, J., Lifson, J. D., Haslam, S. M., Morris, H. R., Desrosiers, R. C., and Dell, A. (2015) Gp120 on HIV-1 virions lacks O-linked carbohydrate. *PLoS One* **10**, e0124784

366. Yang, W., Shah, P., Toghi Eshghi, S., Yang, S., Sun, S., Ao, M., Rubin, A., Jackson, J. B., and Zhang, H. (2014) Glycoform analysis of recombinant and human immunodeficiency

virus envelope protein gp120 via higher energy collisional dissociation and spectral-aligning strategy. *Anal. Chem.* **86**, 6959-6967

367.   Gorman, J. J., McKimm-Breschkin, J. L., Norton, R. S., and Barnham, K. J. (2001) Antiviral activity and structural characteristics of the nonglycosylated central subdomain of human respiratory syncytial virus attachment (G) glycoprotein. *J. Biol. Chem.* **276**, 38988-38994

368.   Chirkova, T., Lin, S., Oomens, A. G., Gaston, K. A., Boyoglu-Barnum, S., Meng, J., Stobart, C. C., Cotton, C. U., Hartert, T. V., Moore, M. L., Ziady, A. G., and Anderson, L. J. (2015) CX3CR1 is an important surface molecule for respiratory syncytial virus infection in human airway epithelial cells. *J Gen Virol* **96**, 2543-2556

369.   Jeong, K. I., Piepenhagen, P. A., Kishko, M., DiNapoli, J. M., Groppo, R. P., Zhang, L., Almond, J., Kleanthous, H., Delagrave, S., and Parrington, M. (2015) CX3CR1 Is Expressed in Differentiated Human Ciliated Airway Cells and Co-Localizes with Respiratory Syncytial Virus on Cilia in a G Protein-Dependent Manner. *PLoS One* **10**, e0130517

370.   Teng, M. N., Whitehead, S. S., and Collins, P. L. (2001) Contribution of the respiratory syncytial virus G glycoprotein and its secreted and membrane-bound forms to virus replication in vitro and in vivo. *Virology* **289**, 283-296

371.   Whelan, J. N., Reddy, K. D., Uversky, V. N., and Teng, M. N. (2016) Functional correlations of respiratory syncytial virus proteins to intrinsic disorder. *Mol. Biosyst.* **12**, 1507-1526

372.   Bukreyev, A., Yang, L., and Collins, P. L. (2012) The secreted G protein of human respiratory syncytial virus antagonizes antibody-mediated restriction of replication involving macrophages and complement. *J. Virol.* **86**, 10880-10884

373.   Collins, P. L., and Mottet, G. (1992) Oligomerization and post-translational processing of glycoprotein G of human respiratory syncytial virus: altered O-glycosylation in the presence of brefeldin A. *J. Gen. Virol.* **73**, 849-863

374.   Bagdonaite, I., Norden, R., Joshi, H. J., King, S. L., Vakhrushev, S. Y., Olofsson, S., and Wandall, H. H. (2016) Global Mapping of O-Glycosylation of Varicella Zoster Virus, Human Cytomegalovirus, and Epstein-Barr Virus. *J. Biol. Chem.* **291**, 12014-12028

375.   Colgrave, M. L., Snelling, H. J., Shiell, B. J., Feng, Y. R., Chan, Y. P., Bossart, K. N., Xu, K., Nikolov, D. B., Broder, C. C., and Michalski, W. P. (2012) Site occupancy and glycan compositional analysis of two soluble recombinant forms of the attachment glycoprotein of Hendra virus. *Glycobiology* **22**, 572-584

376.   Stone, J. A., Nicola, A. V., Baum, L. G., and Aguilar, H. C. (2016) Multiple novel functions of Henipavirus O-glycans: The first O-glycan functions identified in the paramyxovirus family. *PLoS Pathog.* **12**, e1005445

377. Bradel-Tretheway, B. G., Liu, Q., Stone, J. A., McInally, S., and Aguilar, H. C. (2015) Novel Functions of Hendra Virus G N-Glycans and Comparisons to Nipah Virus. *J. Virol.* **89**, 7235-7247

378. Tran, D. T., and Ten Hagen, K. G. (2013) Mucin-type O-glycosylation during development. *J. Biol. Chem.* **288**, 6921-6929

379. Fiete, D., Beranek, M., and Baenziger, J. U. (2012) Peptide-specific transfer of N-acetylgalactosamine to O-Linked glycans by the glycosyltransferases β1,4-N-acetylgalactosaminyl transferase 3 (β4GalNAc-T3) and β4GalNAc-T4. *J. Biol. Chem.* **287**, 29204-29212

380. Tetaert, D., Ten Hagen, K. G., Richet, C., Boersma, A., Gagnon, J., and Degand, P. (2001) Glycopeptide N-acetylgalactosaminyltransferase specificities for O-glycosylated sites on MUC5AC mucin motif peptides. *Biochem. J.* **357**, 313-320

381. Jensen, P. H., Kolarich, D., and Packer, N. H. (2010) Mucin-type O-glycosylation-putting the pieces together. *FEBS J.* **277**, 81-94

382. Thaysen-Andersen, M., Wilkinson, B. L., Payne, R. J., and Packer, N. H. (2011) Site-specific characterisation of densely O-glycosylated mucin-type peptides using electron transfer dissociation ESI-MS/MS. *Electrophoresis* **32**, 3536-3545

383. Bagdonaite, I., Norden, R., Joshi, H. J., Dabelsteen, S., Nystrom, K., Vakhrushev, S. Y., Olofsson, S., and Wandall, H. H. (2015) A strategy for O-glycoproteomics of enveloped viruses--the O-glycoproteome of herpes simplex virus type 1. *PLoS Pathog.* **11**, e1004784

384. Brautigam, J., Scheidig, A. J., and Egge-Jacobsen, W. (2013) Mass spectrometric analysis of hepatitis C viral envelope protein E2 reveals extended microheterogeneity of mucin-type O-linked glycosylation. *Glycobiology* **23**, 453-474

385. Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Levery, S. B., and Clausen, H. (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8**, 977-982

386. Corry, J., Johnson, S. M., Cornwell, J., and Peeples, M. E. (2016) Preventing Cleavage of the Respiratory Syncytial Virus Attachment Protein in Vero Cells Rescues the Infectivity of Progeny Virus for Primary Human Airway Cultures. *J. Virol.* **90**, 1311-1320

387. Elango, N., Prince, G. A., Murphy, B. R., Venkatesan, S., Chanock, R. M., and Moss, B. (1986) Resistance to human respiratory syncytial virus (RSV) infection induced by immunization of cotton rats with a recombinant vaccinia virus expressing the RSV G glycoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 1906-1910

388. Tytgat, K. M., Swallow, D. M., Van Klinken, B. J., Buller, H. A., Einerhand, A. W., and Dekker, J. (1995) Unpredictable behaviour of mucins in SDS/polyacrylamide-gel electrophoresis. *Biochem. J.* **310**, 1053-1054

389. Altgarde, N., Eriksson, C., Peerboom, N., Phan-Xuan, T., Moeller, S., Schnabelrauch, M., Svedhem, S., Trybala, E., Bergstrom, T., and Bally, M. (2015) Mucin-like Region of Herpes Simplex Virus Type 1 Attachment Protein Glycoprotein C (gC) Modulates the Virus-Glycosaminoglycan Interaction. *J. Biol. Chem.* **290**, 21473-21485

390. Dube, D. H., and Bertozzi, C. R. (2005) Glycans in cancer and inflammation - potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4**, 477-488

391. Iversen, M. B., Reinert, L. S., Thomsen, M. K., Bagdonaite, I., Nandakumar, R., Cheshenko, N., Prabakaran, T., Vakhrushev, S. Y., Krzyzowska, M., Kratholm, S. K., Ruiz-Perez, F., Petersen, S. V., Goriely, S., Bibby, B. M., Eriksson, K., Ruland, J., Thomsen, A. R., Herold, B. C., Wandall, H. H., Frische, S., Holm, C. K., and Paludan, S. R. (2016) An innate antiviral pathway acting before interferons at epithelial surfaces. *Nat. Immunol.* **17**, 150-158

392. Mi, Y., Lin, A., Fiete, D., Steirer, L., and Baenziger, J. U. (2014) Modulation of mannose and asialoglycoprotein receptor expression determines glycoprotein hormone half-life at critical points in the reproductive cycle. *J. Biol. Chem.* **289**, 12157-12167

393. Stewart-Jones, G. B., Thomas, P. V., Chen, M., Druz, A., Joyce, M. G., Kong, W. P., Sastry, M., Soto, C., Yang, Y., Zhang, B., Chen, L., Chuang, G. Y., Georgiev, I. S., McLellan, J. S., Srivatsan, S., Zhou, T., Baxa, U., Mascola, J. R., Graham, B. S., and Kwong, P. D. (2015) A cysteine zipper stabilizes a pre-fusion F glycoprotein vaccine for respiratory syncytial virus. *PLoS One* **10**, e0128779

394. Okamoto, N., Uchida, A., Takakura, K., Kariya, Y., Kanzaki, H., Riittinen, L., Koistinen, R., Seppala, M., and Mori, T. (1991) Suppression by human placental protein 14 of natural killer cell activity. *Am. J. Reprod. Immunol.* **26**, 137-142

395. Larranaga, C. L., Ampuero, S. L., Luchsinger, V. F., Carrion, F. A., Aguilar, N. V., Morales, P. R., Palomino, M. A., Tapia, L. F., and Avendano, L. F. (2009) Impaired immune response in severe human lower tract respiratory infection by respiratory syncytial virus. *Pediatr. Infect. Dis. J.* **28**, 867-873

396. Welliver, T. P., Reed, J. L., and Welliver, R. C., Sr. (2008) Respiratory syncytial virus and influenza virus infections: observations from tissues of fatal infant cases. *Pediatr. Infect. Dis. J.* **27**, S92-96

397. Kaiko, G. E., Phipps, S., Angkasekwinai, P., Dong, C., and Foster, P. S. (2010) NK cell deficiency predisposes to viral-induced Th2-type allergic inflammation via epithelial-derived IL-25. *J. Immunol.* **185**, 4681-4690

398. Zarbock, A., Ley, K., McEver, R. P., and Hidalgo, A. (2011) Leukocyte ligands for endothelial selectins: specialized glycoconjugates that mediate rolling and signaling under flow. *Blood* **118**, 6743-6751

399. Kuebler, W. M. (2006) Selectins revisited: the emerging role of platelets in inflammatory lung disease. *J. Clin. Invest.* **116**, 3106-3108

400. Benam, K. H., Villenave, R., Lucchesi, C., Varone, A., Hubeau, C., Lee, H. H., Alves, S. E., Salmon, M., Ferrante, T. C., Weaver, J. C., Bahinski, A., Hamilton, G. A., and Ingber, D. E. (2016) Small airway-on-a-chip enables analysis of human lung inflammation and drug responses in vitro. *Nat. Methods* **13**, 151-157

401. Wang, S. Z., Smith, P. K., Lovejoy, M., Bowden, J. J., Alpers, J. H., and Forsyth, K. D. (1998) Shedding of L-selectin and PECAM-1 and upregulation of Mac-1 and ICAM-1 on neutrophils in RSV bronchiolitis. *Am. J. Physiol.* **275**, L983-989

402. Mohammed, R. N., Watson, H. A., Vigar, M., Ohme, J., Thomson, A., Humphreys, I. R., and Ager, A. (2016) L-selectin Is Essential for Delivery of Activated CD8(+) T Cells to Virus-Infected Organs for Protective Immunity. *Cell Rep* **14**, 760-771

403. Wedepohl, S., Kaup, M., Riese, S. B., Berger, M., Dernedde, J., Tauber, R., and Blanchard, V. (2010) N-glycan analysis of recombinant L-Selectin reveals sulfated GalNAc and GalNAc-GalNAc motifs. *J. Proteome Res.* **9**, 3403-3411

404. Wedepohl, S., Beceren-Braun, F., Riese, S., Buscher, K., Enders, S., Bernhard, G., Kilian, K., Blanchard, V., Dernedde, J., and Tauber, R. (2012) L-selectin-a dynamic regulator of leukocyte migration. *Eur. J. Cell Biol.* **91**, 257-264

405. Gupta, G. S. (2012) L-Selectin (CD62L) and Its Ligands. in *Animal Lectins: Form, Function and Clinical Applications* (Gupta, G. S. ed., Springer-Verlag Wien, Vienna

406. Grinnell, B. W., Hermann, R. B., and Yan, S. B. (1994) Human Protein C inhibits selectin-mediated cell adhesion: role of unique fucosylated oligosaccharide. *Glycobiology* **4**, 221-225

407. Mowery, P., Yang, Z. Q., Gordon, E. J., Dwir, O., Spencer, A. G., Alon, R., and Kiessling, L. L. (2004) Synthetic glycoprotein mimics inhibit L-selectin-mediated rolling and promote L-selectin shedding. *Chem. Biol.* **11**, 725-732

408. Malhotra, R., Ward, M., Bright, H., Priest, R., Foster, M. R., Hurle, M., Blair, E., and Bird, M. (2003) Isolation and characterisation of potential respiratory syncytial virus receptor(s) on epithelial cells. *Microbes Infect.* **5**, 123-133

409. Harms, G., Kraft, R., Grelle, G., Volz, B., Dernedde, J., and Tauber, R. (2001) Identification of nucleolin as a new L-selectin ligand. *Biochem. J.* **360**, 531-538

Appendices

Appendix A

Parameters used for analysis of NDV V4-VAR HN tryptic peptides and glycopeptides

Ms Parameters	V4-VAR tryptic sample: Orbitrap (OT) HCD	V4-VAR tryptic sample: OT HCD-pd-OT ETD	V4-VAR tryptic sample: OT HCD-pd-IT CID	V4-VAR tryptic/PNGase F sample: OT HCD-pd-OT ETD
Polarity	Positive	Positive	Positive	Positive
MS^1 Detector Type	Orbitrap	Orbitrap	Orbitrap	Orbitrap
MS^1 Resolution at 200 (m/z)	120K	120K	120K	120K
MS^1 Scan Range (m/z)	400-1500	380-2000	400-1800	400-1800
MS^1 ACG Target	400,000	200,000	200,000	200,000
MS^1 Injection Time (ms)	50	50	50	50
MS^2 Precursor Selection Criteria	Most intense	Highest charge, lowest m/z	Most intense	Highest charge, lowest m/z
MS^2 Precursor Intensity Threshold	5000	5,000	5,000	5,000
MS^2 Precursor Scan Range Selection	400-1500	380-1800	400-1500	550-1500
MS^2 Precursor Charge State Selection	2-7	2-8	2-7	3-7
Exclude After (n times)	1	1	1	2
Exclusion Duration (s)	25	20	20	20
MS^2 Precursor Isolation Window (m/z)	1.6	2	3	3
MS^2 Detector Type	Orbitrap	Orbitrap	Orbitrap	Orbitrap
MS^2 Resolution	60K	30K	60K	60K
MS^2 AGC Target	50,000	50,000	100,000	100,000
MS^2 Maximum Injection Time	60	60	120	120
MS^2 Activation Type	HCD	HCD	HCD	HCD
MS^2 Normalised Collision Energy (CE)	30%	30%	35%	30%
Product Dependant (pd) Ion Trigger (m/z) 138.0545, 204.0867, 366.1396 & 366.1396	N/A	Within top 20 product ions	Within top 20 product ions	Within top 30 product ions including 292.1027
pd-MS^2 Detection	N/A	Orbitrap	Ion trap	Orbitrap
pd-MS^2 Resolution	N/A	60K	N/A	60K
pd-MS^2 ACG Target	N/A	100,000	10,000	100,000
pd-MS^2 Maximum Injection Time (Ms)	N/A	250	150	150
pd-MS^2 Activation Type	N/A	ETD	CID	ETD
pd- MS^2 Fragmentation Parameters	N/A	Charge dependant ETD with one microscan	30% NCE	Charge dependant ETD with two microscans

Appendix B

NDV HN sequence alignment

Multiple sequence alignment of NDV HN was implemented using fifteen annotated NDV HN sequences from the UniProt website. A list of the Uniprot entries is provided in the table below. Within the sequence alignment sequence identifiers are listed on the left and the NDV Queensland/66 strain (UniProt entry P13850) is the first identifier. All N-linked consensus (N-X-S/T) sites have been highlighted in yellow.

Uniprot Entry	Entry Name	Isolate name	Length (amino acids)
P13850	HN_NDVQ	Newcastle disease virus (strain Queensland/66) (NDV)	616
P12558	HN_NDVU	Newcastle disease virus (strain Chicken/Northern Ireland/Ulster/67) (NDV)	616
P12555	HN_NDVD	Newcastle disease virus (strain D26/76) (NDV)	616
Q9Q2W5	HN_NDVK	Newcastle disease virus (strain Kansas) (NDV)	577
P32884	HN_NDVB	Newcastle disease virus (strain Beaudette C/45) (NDV)	577
P35743	HN_NDVL	Newcastle disease virus (strain Chicken/United States/LaSota/46) (NDV)	577
P12559	HN_NDVH4	Newcastle disease virus (strain B1-Hitchner/47) (NDV)	577
P12553	HN_NDVTG	Newcastle disease virus (strain Chicken/United States(TX)/GB/48) (NDV)	577
Q91UL0	HN_NDVB1	Newcastle disease virus (strain Chicken/United States/B1/48) (NDV)	577
P12556	HN_NDVI	Newcastle disease virus (strain Italien/45) (NDV)	571
P35742	HN_NDVJ	Newcastle disease virus (strain Iba/85) (NDV)	571
P35741	HN_NDVH3	Newcastle disease virus (strain Her/33) (NDV)	571
P35740	HN_NDVC	Newcastle disease virus (strain Chi/85) (NDV)	571
P12557	HN_NDVM	Newcastle disease virus (strain Miyadera/51) (NDV)	571
P12554	HN_NDVA	Newcastle disease virus (strain Chicken/Australia-Victoria/32) (NDV)	570

```
SP|P13850|HN_NDVQ    MDRAVSQVALENDEREAKNTWRLVFRIAILLSTVVTLAISAAALAYSMEASTPSDLVGIP 60
SP|P12558|HN_NDVU    MDRAVSQVALENDEREAKNTWRLVFRIAILLLTVVTLAISAAALAYSMEASTPSDLIGIP 60
SP|P12555|HN_NDVD    MDRAVSQVALENDEREAKNTWRLVFRIAILLLTVVTLAISAAALAYSMEASTPSDLVGIP 60
SP|Q91UL0|HN_NDVB1   MDRAVSQVALENDEREAKNTWRLIFRIAILFLTVVTLAISVASLLYSMGASTPSDLVGIP 60
SP|Q9Q2W5|HN_NDVK    MDRAVSQVALENDEREAKNTWRLIFRIAILLLTVVTLATSVASLVYSMGASTPSDLVGIP 60
SP|P35743|HN_NDVL    MDRAVSQVALENDEREAKNTWRLIFRIAILFLTVVTLAISVASLLYSMGASTPSDLVGIP 60
SP|P32884|HN_NDVB    MDRAVSQVALENDEREAKNTWRLIFRIAILLLTVVTLATSVASLVYSMGASTPSDLVGIP 60
SP|P12553|HN_NDVTG   MDRAVSQVALENDEREAKNTWRLIFRIAILLLTVVTLATSVASLVYSMGASTPSDLVGIP 60
SP|P12559|HN_NDVH4   MDRAVSQVALENDEREAKNTWRLIFRIAILFLTVVTLAISVASLLYSMGASTPSDLVGIP 60
SP|P35741|HN_NDVH3   MDRAVSRVALENEEREAKNTWRFVFRIAILLLIVITLAISAAALVYSMEASTPGDLVGIP 60
SP|P35742|HN_NDVJ    MDRAVSRVVLENEEREAKNTWRFVFRIAVLLLIVMTLAISAAALVYSMGASTPRDLASIS 60
SP|P12557|HN_NDVM    MDRTVNQVALENDEREAKNTWRLVFRIATLLLIVMTLAFSAAALAYSMEASTPGDLVGIP 60
SP|P12556|HN_NDVI    MDRAVGRVALENEEREAKNTWRFVFRIAIFLLIVITLAISAAALVYSMEASTPGDLVGIP 60
SP|P35740|HN_NDVC    MDRAVNRVVLENEEREAKNTWRLVFRIAVLLLMVMTLAISAAALVYSMGASTPRDLAGIS 60
SP|P12554|HN_NDVA    MNRAVCQVALENDEREAKNTWRLVFRIAILLLTVMTLAISAAALAYSMEASTPGDLVSIP 60
                     *:*:* :*.***:*********::**** ::  *:*** *.*:* *** **** ** .*


SP|P13850|HN_NDVQ    TAISRAEEKITSALGSNQDVVDRIYKQVALESPLALLNTESTIMNAITSLSYRINGAANS 120
SP|P12558|HN_NDVU    TAISRAEEKITSALGSNQDVVDRIYKQVALESPLALLNTESTIMNAITSLSYQINGAANS 120
SP|P12555|HN_NDVD    TAISRTEEKITSALGSNQDVVDRIYKQVALESPLALLNTESTIMNAITSLSYQINGAANS 120
SP|Q91UL0|HN_NDVB1   TRISRAEEKITSTLGSNQDVVDRIYKQVALESPLALLNTETTIMNAITSLSYQINGAANN 120
SP|Q9Q2W5|HN_NDVK    TRISRAEEKITSALGSNQDVVDRIYKQVALESPLALLNTETTIMNAITSLSYQINGAANN 120
SP|P35743|HN_NDVL    TRISRAEEKITSTLGSNQDVVDRIYKQVALESPLALLKTETTIMNAITSLSYQINGAANN 120
SP|P32884|HN_NDVB    TRISRAEEKITSALGSNQDVVDRIYKQVALESPLALLNTETTIMNAITSLSYQINGAANN 120
SP|P12553|HN_NDVTG   TRISRAEEKITSALGSNQDVVDRIYKQVALESPLALLNTETTIMNAITSLSYQINGAANN 120
SP|P12559|HN_NDVH4   TRISRAEEKITSTLGSNQDVVDRIYKQVALESPLALLNTETTIMNAITSLSYQINGAANN 120
SP|P35741|HN_NDVH3   TVISRAEEKITSALSSNQDVVDRIYKQVALESPLALLNTESVIMNAITSLSYQINGAANN 120
SP|P35742|HN_NDVJ    TAISKMEDKITSSLSSNQDVVDRIYKQVALESPLALLNTESIIMNAITSLSYQINGAANN 120
SP|P12557|HN_NDVM    TAISRAEEKITSALGSNQDVVDRIYKQVALESPLALLNTESIIMNAITSLSYQINGAANN 120
SP|P12556|HN_NDVI    TVISRAEEKITSALSSNQDVVDRIYKQVALESPLALLNTESVIMNAITSLSYQINGAANN 120
SP|P35740|HN_NDVC    TVISKTEDKVTSLLSSKQDVIDRIYKQVALESPLALLNTESIIMNAITSLSYQINGAANN 120
SP|P12554|HN_NDVA    TAISRAEGKITSALGSNQDVVDRIYKQVALESPLALLNTESIIMNAITSLSYQINGAANN 120
                     * **: * *:** *.*:***:****************:**: **********:******.


SP|P13850|HN_NDVQ    SGCGAPIHDPDYIGGIGKELIVDDASDVTSYYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P12558|HN_NDVU    SGCGAPIHDPDYIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P12555|HN_NDVD    SGCGAPIHDPDYIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|Q91UL0|HN_NDVB1   SGWGAPIHDPDYIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|Q9Q2W5|HN_NDVK    SGWGAPIHDPDFIGGIGKELIVDNASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P35743|HN_NDVL    SGWGAPIHDPDYIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P32884|HN_NDVB    SGWGAPIHDPDFIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P12553|HN_NDVTG   SGWGAPIHDPDFIGGIGKELIVDDASDVTSFYPSAFQEHHNFIPAPTTGSGCIRIPSFDM 180
SP|P12559|HN_NDVH4   SGWGAPIHDPDYIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P35741|HN_NDVH3   SGCGAPVHDPDYIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDI 180
SP|P35742|HN_NDVJ    SGCGAPVHDPDYIGGIGKELIVDDTSDVTSFYPSAYQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P12557|HN_NDVM    SGCGAPVHDPDYIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P12556|HN_NDVI    SGCGAPVHDPDYIGGIGKELIVDDASDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDI 180
SP|P35740|HN_NDVC    SGCGEPVHDPDYIGGIGKELIVDDISDVTSFYPSAYQEHLNFIPAPTTGSGCTRIPSFDM 180
SP|P12554|HN_NDVA    SGCGAPVHDPDYIGGIGKELIVDDTSDVTSFYPSAFQEHLNFIPAPTTGSGCTRIPSFDM 180
                     ** * *:****:***********: *****:****:*** ************ ******:


SP|P13850|HN_NDVQ    SATHYCYTHNVILSGCRDHSHSHQYLALGVLRTSATGRVFFSTLRSINLDDTQNRKSCSV 240
SP|P12558|HN_NDVU    SATHYCYTHNVILSGCRDHSHSHQYLALGVLRTSATGRVFFSTLHSINLDDTQNRKSCSV 240
SP|P12555|HN_NDVD    SATHYCYTHNVILSGCRDRSHSHQYLALGVLRTSATGRVFFSTLRSINLDDTQNRKSCSV 240
SP|Q91UL0|HN_NDVB1   SATHYCYTHNVILSGCRDHSHSYQYLALGVLRTSATGRVFFSTLRSINLDDTQNRKSCSV 240
SP|Q9Q2W5|HN_NDVK    SATHYCYTHNVILSGCRDHSHSHQYLALGVLRTTATGRIFFSTLRSISLDDTQNRKSCSV 240
SP|P35743|HN_NDVL    SATHYCYTHNVILSGCRDHSHSYQYLALGVLRTSATGRVFFSTLRSINLDDTQNRKSCSV 240
SP|P32884|HN_NDVB    SATHYCYTHNVILSGCRDHSHSHQYLALGVLRTTATGRIFFSTLRSISLDDTQNRKSCSV 240
SP|P12553|HN_NDVTG   SATHYCYTHNIISSGCRDHSHSYQYLALGVLRTSATGRIFFSTLRSINLDDTQNRKSCSV 240
SP|P12559|HN_NDVH4   SATHYCYTHNVILSGCRDHLHSHQYLALGVLRTSATGRVFFSTLRSINLDDTQNRKSCSV 240
SP|P35741|HN_NDVH3   SATHYCYTHNVILSGCRDHSHSHQYLALGVLRTSATGRVFFSTLRSINLDDNQNRKSCSV 240
SP|P35742|HN_NDVJ    SATHYCYTHNVILSGCRDHSHSHQYLALGVLRTSATGKVFFSTLRSINLDDTQNRKSCSV 240
SP|P12557|HN_NDVM    SATHYCYTHNVILSGCRDHSQSHQYLALGVLRTSATGRVFFSTLRSINLDDTQNRKSCSV 240
SP|P12556|HN_NDVI    SATHYCYTHNVILSGCRDHSHSHQYLALGVLRTSATGRVFFSTLRSINLDDNQNRKSCSV 240
SP|P35740|HN_NDVC    STTHYCYTHNVILSGCRDHSHSHQYLALGVLRTSATGRVFFSTLRSINLDDTQNRKSCSV 240
SP|P12554|HN_NDVA    SATH-CYTHNVIFSGCRDHSHSHQYLALGVLRTSATGRVFFSTLRSINLDDTQNRKSCSV 239
                     *:** *****:* *****: :*:***********:***::*****:**.***.********
```

```
SP|P13850|HN_NDVQ   SATPLGCDMLCSKVTETEEEDYNSAIPTSMVHGRLGFDGQYHEKDLDVTTLFEDWVANYP 300
SP|P12558|HN_NDVU   SATPLGCDMLCSKVTETEEEDYNSAVPTSMVHGRLGFDGQYHEKDLDVTTLFEDWVANYP 300
SP|P12555|HN_NDVD   SATPLGCDMLCSKVTETEEEDYNSAIPTSMVHGRLGFDGQYHEKDLDVTTLFEDWVANYP 300
SP|Q91UL0|HN_NDVB1  SATPLGCDMLCSKATETEEEDYNSAVPTRMVHGRLGFDGQYHEKDLDVTTLFGDWVANYP 300
SP|Q9Q2W5|HN_NDVK   SATPLGCDMLCSKVTETEEEDYNSAVPTLMAHGRLGFDGQYHEKDLDVTTLFEDWVANYP 300
SP|P35743|HN_NDVL   SATPLGCDMLCSKVTETEEEDYNSAVPTRMAHGRLGFDGQYHEKDLDVTTLFGDWVANYP 300
SP|P32884|HN_NDVB   SATPLGCDMLCSKVTETEEEDYNSAVPTLMAHGRLGFDGQYHEKDLDVTTLFEDWVANYP 300
SP|P12553|HN_NDVTG  SATPLGCDMLCSKVTETEEEDYNSAVPTLMVHGRLGFDGQYHEKDLDVTTLFEDWVANYP 300
SP|P12559|HN_NDVH4  SATPLGCDMLCSKATETEEEDYNSAVPTRMVHGRLGFDGQYHEKDLDVTTLFGDWVANYP 300
SP|P35741|HN_NDVH3  SATPLGCDMLCSKITETEEEDYSSVTPTSMVHGRLGFDGQYHEKDLDVITLFKDWVANYP 300
SP|P35742|HN_NDVJ   SATPLGCDMLCSKVTETEEEDYKSVTPTSMVHGRFRFDGQYHEKDSDRTTLFKDWVANYP 300
SP|P12557|HN_NDVM   SATPLGCDMLCSKVTETEEEDYNSVTPTSMVHGRLGFDGQYHEKDLDVTTLFGDWVANYP 300
SP|P12556|HN_NDVI   SATPLGCDMLCSKITETEEEDYSSVTPTSMVHGRLGFDGQYHEKDLDVITLFKDWVANYP 300
SP|P35740|HN_NDVC   SATPLGCDMLCSKVTETEEEDYKSVTPTSMVHGRLGFDGQYHEKDLDTTVLFKDWVANYP 300
SP|P12554|HN_NDVA   SATPLGCDMLCSKVTETEEEDYNSVIPTSMVHGRLGFDGQYHEKDLDVTTLFGDWVANYP 299
                    ************* ********.*. ** *.***: ********* *   .** *******


SP|P13850|HN_NDVQ   GVGGGSFIDNRVWFPVYGGLKPNSPSDTAQEGKYVIYKRYNDTCPDEQDYQIQMAKSSYK 360
SP|P12558|HN_NDVU   GVGGGSFIDNRVWFPVYGGLKPNSPSDTAQEGKYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P12555|HN_NDVD   GVGGGSFIDNRVWFPVYGGLKPNSPSDTAQEGKYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|Q91UL0|HN_NDVB1  GVGGGSFIDSRVWFSVYGGLKPNSPSDTVQEGKYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|Q9Q2W5|HN_NDVK   GVGGGSFIDGRVWFSVYGGLKPNSPSDTVQEGKYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P35743|HN_NDVL   GVGGGSFIDSRVWFSVYGGLKPNSPSDTVQEGKYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P32884|HN_NDVB   GVGGGSFIDGRVWFSVYGGLKPNSPSDTVQEGKYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P12553|HN_NDVTG  GVGGGSFIDSRVWFSVYGGLKPNSPSDTVQEEKYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P12559|HN_NDVH4  GVGGGSFIDSRVWFSVYGGLKPNTPSDTVQEGKYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P35741|HN_NDVH3  GVGGGSFIDNRVWFPVYGGLKPNSPSDTVQEGRYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P35742|HN_NDVJ   GVGGGSFIDDRVWFPIYGGLKPNSPSDIAQEGKYVIYKRYNNTFPDKQDYQIRMAKSSYK 360
SP|P12557|HN_NDVM   GVGGGSFIDSRVWFPIYGGLKPNSPSDTAQEGRYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P12556|HN_NDVI   GVGGGSFIDNRVWFPVYGGLKPNSPSDTAQEGRYVIYKRYNDTCPDEQDYQIRMAKSSYK 360
SP|P35740|HN_NDVC   GVGGGSFIDDRVWFPVYGGLKPNSPSDTAQEGKYVIYKRYNNTCPDEQDYQIRMAKSSYK 360
SP|P12554|HN_NDVA   GVGGGSFIDNRVWFPVYGGLKPSSPSDTGQEGRYVIYKRYNDTCPDEQDYQIRMAKSSYK 359
                    ********* **** :******.:*** ** :*********:* **:*****:*******


SP|P13850|HN_NDVQ   PGRFGGKRVQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRVLTVGTSHFLYQRGSSY 420
SP|P12558|HN_NDVU   PGRFGGKRVQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRVLTVGTSHFLYQRGSSY 420
SP|P12555|HN_NDVD   PGRFGGKRVQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRVLTVGTSHFFYQRGSSY 420
SP|Q91UL0|HN_NDVB1  PGRFGGKRIQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRILTVGTSHFLYQRGSSY 420
SP|Q9Q2W5|HN_NDVK   PGRFGGKRIQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRILTVGTSHFLYQRGSSY 420
SP|P35743|HN_NDVL   PGRFGGKRIQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRILTVGTSHFLYQRGSSY 420
SP|P32884|HN_NDVB   PGRFGGKRIQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRILTVGTSHFLYQRGSSY 420
SP|P12553|HN_NDVTG  PGRFGGKRIQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRILTVGTSHFLYQRGSSY 420
SP|P12559|HN_NDVH4  PGRFGGKRIQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRILTVGTSHFLYQRGSSY 420
SP|P35741|HN_NDVH3  PGRFGGKRVQQAILSIKVSTSLGEDPVLTIPPNTVTLMGAEGRVLTVGTSHFLYQRGSSY 420
SP|P35742|HN_NDVJ   PGRFGGKRVQQAILSIKVSTSLGEDPVLTVPPNTITLMGAEGRVLTVGTSHFLYQRGSSY 420
SP|P12557|HN_NDVM   PRRFGGKRVQQAILSIKVSTSLGEDPVLTVPPNTVTLMGAEGRVLTVGTSHFLYQRGSSY 420
SP|P12556|HN_NDVI   PGRFGGKRVQQAILSIKVSTSLGEDPVLTVPPNTVTLMGPEGRVLTVGTSHFLYQRGSSY 420
SP|P35740|HN_NDVC   PGRFGGKRVQQAILSIKVSTSLGEDPVLTIPPNTITLMGAEGRVLTVGTSHFLYQRGSSY 420
SP|P12554|HN_NDVA   PGRFGGKRVQQAILSIKVSTSLGEDPVLTIPPNTVTLMGAEGRVLTVGTSHFLYQRGSSY 419
                    *  *****:********************:****:**** ***:********:*******


SP|P13850|HN_NDVQ   FSPALLYPMIVSNKTATLHSPYTFNAFTRPGSVPCQASARCPNSCVTGVYTDPYPLVFYR 480
SP|P12558|HN_NDVU   FSPALLYPMTVSNKTATLHSPYTFDAFTRPGSVPCQASARCPNSCVTGVYTDPYPLVFYR 480
SP|P12555|HN_NDVD   FSPALLYPMTVSNKTATLHSPYTFNAFTRPGSVPCQASARCPNSCVTGVYTDPYPLVFYR 480
SP|Q91UL0|HN_NDVB1  FSPALLYPMTVSNKTATLHSPYTFNAFTRPGSIPCQASARCPNSCVTGVYTDPYPLIFYR 480
SP|Q9Q2W5|HN_NDVK   FSPALLYPMTVSNKTATLHSPYTFNAFTRPGSIPCQASARCPNSCVTGVYTDPYPLIFYR 480
SP|P35743|HN_NDVL   FSPALLYPMTVSNKTATLHSPYTFNAFTRPGSIPCQASARCPNPCVTGVYTDPYPLIFYR 480
SP|P32884|HN_NDVB   FSPALLYPMTVSNKTATLHSPYTFNAFTRPGSIPCQASARCPNSCVTGVYTDPYPLIFYR 480
SP|P12553|HN_NDVTG  FSPALLYPMTVSNKTATLHSPYTFNAFTRPGSIPCQASARCPNSCVTGVYTDPYPLIFYR 480
SP|P12559|HN_NDVH4  FSPALLYPMTVSNKTATLHSPYTFNAFTRPGSIPCQASARCPNSCVTGVYTDPYPLIFYR 480
SP|P35741|HN_NDVH3  FSPALLYPMTVNNKTATLHSPYTFNAFTRPGSVPCQASARCPNSCVTGVYTDPYPLIFHR 480
SP|P35742|HN_NDVJ   FSPALLYPMTVYQQTATLHSPYTFNAFTRPGSVPCQASARCPNSCITGVYTDPYPLVFHR 480
SP|P12557|HN_NDVM   FSPALLYPMTVNNKTATLHSPYTFNAFTRPGSVPCQASARCPNSCVTGVYTDPYPLVFHA 480
SP|P12556|HN_NDVI   FSPALLYPMTVNNKTATLHSPYTFNAFTRPGSVPCQASARCPNSCVTGVYTDPYPLVFHR 480
SP|P35740|HN_NDVC   FSPALLYPMTVNNKTATLHSPYTFNAFTRPGSVPCQASARCPNSCITGVYTDPYPLIFHR 480
SP|P12554|HN_NDVA   FSPALLYPMTVNNNTATLHSPYTFNAFTRPGSVPCQASARCPNSCVTGVYTDPYPLVFHR 479
                    ********* * ::*********:*******:********** *:***********:*:
```

```
SP|P13850|HN_NDVQ  NHTLRGVFGTMLDDKQARLNPVSAVFDSISRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|P12558|HN_NDVU  NHTLRGVFGTMLDDKQARLNPVSAVFDSISRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|P12555|HN_NDVD  NHTLRGVFGTMLDDEQARLNPVSAVFDSISRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|Q91UL0|HN_NDVB1 NHTLRGVFGTMLDGEQARLNPASAVFDSTSRSRITRVSSSSIKAAYTTSTCFKVVKTNKT 540
SP|Q9Q2W5|HN_NDVK  NHTLRGVFGTMLDSEQARLNPASAVFDSTSRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|P35743|HN_NDVL  NHTLRGVFGTMLDGVQARLNPASAVFDSTSRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|P32884|HN_NDVB  NHTLRGVFGTMLDSEQARLNPTSAVFDSTSRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|P12553|HN_NDVTG NHTLRGVFGTMLDGEQARLNPASAVFDSTSRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|P12559|HN_NDVH4 NHTLRGVFGTMLDGEQARLNPASAVFDSTSRSRITRVSSSSIKAAYTTSTCFKVVKTNKT 540
SP|P35741|HN_NDVH3 NHTLRGVFGTMLDDGQARLNPVSAVFDNISRSRITRVSSRTKAAYTTSTCFKVVKTNKT 540
SP|P35742|HN_NDVJ  NHTLRGVFGTMLDDEQARLNPVSAVFDNISRSRVTRVSSSSTKAAYTTSTCFKVVKTSKA 540
SP|P12557|HN_NDVM  NHTLRGVFGTMLDDERARLNPVSAVFDNVSRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|P12556|HN_NDVI  NHTLRGVFGTMLDDKQARLNPVSAVFDNISRSRITRVSSSSTKAAYTTSTCFKVVKTNKT 540
SP|P35740|HN_NDVC  NHTLRGVFGTMLDDEQARLNPVSAVFDNISRSRVTRVSSSSTKAAYTTSTCFKVVKTNKA 540
SP|P12554|HN_NDVA  NHTLRGVFGTMLDDEQARLNLVSAVFDNISRSRITRVSSRTKAAYTTSTCFKVVKTNKT 539
                   *************  :****  .*****. ****:******  **************.*:


SP|P13850|HN_NDVQ  YCLSIAEISNTLFGEFRIVPLLVEILKDDGVREARSSRLSQLREGWKDDIVSPIFCDAKN 600
SP|P12558|HN_NDVU  YCLSIAEISNTLFGEFRIVPLLVEILKDDGVREARAGRLSQLREGWKDDIVSPIFCDAKN 600
SP|P12555|HN_NDVD  YCLSIAEISNTLFGEFRIVPLLVEILKDDGVREARSGRLSQLQEGWKDDIVSPIFCDAKN 600
SP|Q91UL0|HN_NDVB1 YCLSIAEISNTLFGEFRIVPLLVEILKDDGVREARSG---------------------- 577
SP|Q9Q2W5|HN_NDVK  YCLSIAEISNTLFGEFRIVPLLVEILKNDGVREARSG---------------------- 577
SP|P35743|HN_NDVL  YCLSIAEISNTLFGEFRIVPLLVEILKDDGVREARSG---------------------- 577
SP|P32884|HN_NDVB  YCLSIAEISNTLFGEFRIVPLLVEILKNDGVREARSG---------------------- 577
SP|P12553|HN_NDVTG YCLSIAEISNTLFGEFRIVPLLVEILKNDGVREARSG---------------------- 577
SP|P12559|HN_NDVH4 YCLSIAEISNTLFGEFRIVPLLVEILKDDGVREARSG---------------------- 577
SP|P35741|HN_NDVH3 YVLSIAEISNTLFGEFRIVPLLVEILKNDGV---------------------------- 571
SP|P35742|HN_NDVJ  YCLSIAEISNTLFGEFRIVPLLVEILKDDRV---------------------------- 571
SP|P12557|HN_NDVM  YCLSIAEISNTLFGEFRIVPLLVEILKDDKV---------------------------- 571
SP|P12556|HN_NDVI  YCLSIAEISNTLFGEFRIVPLLVEILKEDGV---------------------------- 571
SP|P35740|HN_NDVC  YCLSIAEISNTLFGEFRIVPLLVEILKDDRV---------------------------- 571
SP|P12554|HN_NDVA  YCLSIAEISNTLFGEFRIVPLLVEILKDDGV---------------------------- 570
                   *  ***********************:* *


SP|P13850|HN_NDVQ  QTEYRRELESYAASWP 616
SP|P12558|HN_NDVU  QTEYRRELESYAASWP 616
SP|P12555|HN_NDVD  QTEYRRELESYAASWP 616
SP|Q91UL0|HN_NDVB1 ----------------
SP|Q9Q2W5|HN_NDVK  ----------------
SP|P35743|HN_NDVL  ----------------
SP|P32884|HN_NDVB  ----------------
SP|P12553|HN_NDVTG ----------------
SP|P12559|HN_NDVH4 ----------------
SP|P35741|HN_NDVH3 ----------------
SP|P35742|HN_NDVJ  ----------------
SP|P12557|HN_NDVM  ----------------
SP|P12556|HN_NDVI  ----------------
SP|P35740|HN_NDVC  ----------------
SP|P12554|HN_NDVA  ----------------
```

*   Positions have a single, fully conserved residue.

:   Conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix.

•   Conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix.

Appendix C

## HCD MS/MS of NDV V4-VAR HN peptide Q87-K138

Amino acid sequence changes in V4-VAR from QLD/66 (UniProt entry P13850) have been highlighted in red in the sequence schematic. They represent sequence changes R113Q, N115S and N119S. All b- and y-ions indicated in the schematic were observed in the MS/MS spectrum of the precursor at *m/z* 1319.660 (4+) with a precursor mass tolerance of 0.45 ppm. Not all ions have been labelled on the spectrum for ease of interpretation. Within the continuous y-ion series y-18 to y-35, ions y-19, y-20 and y-23 to y-26 (highlighted red in the spectrum) confirm the sequence changes.
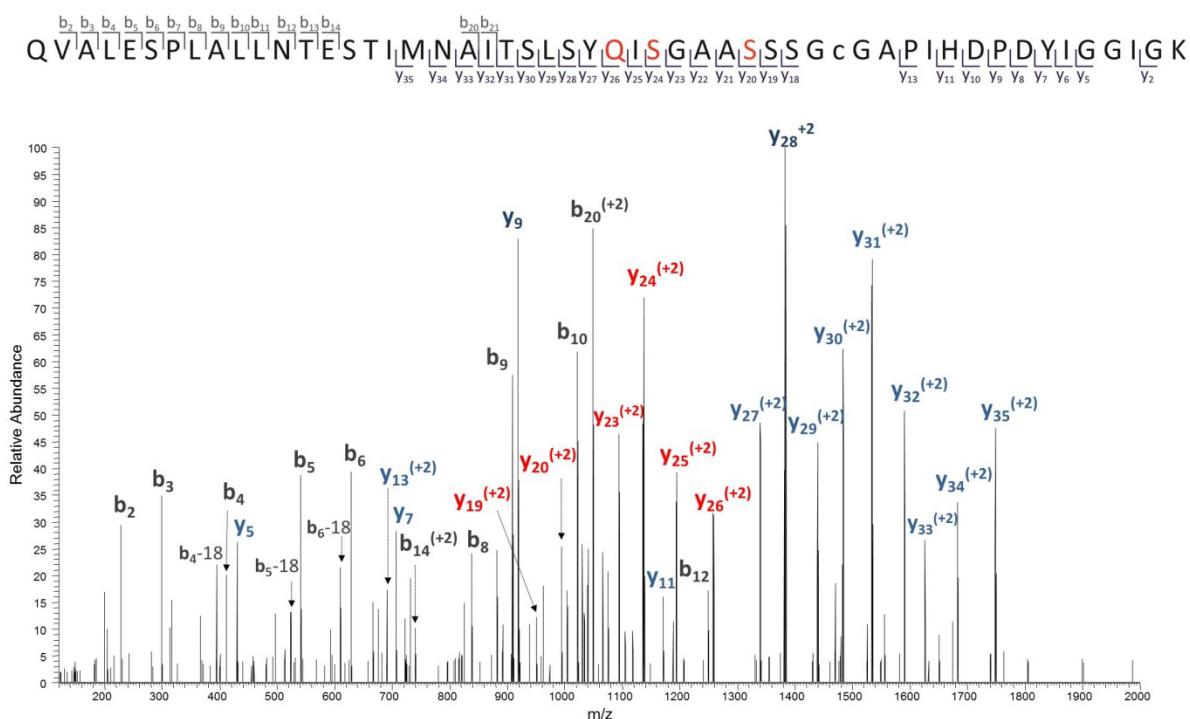
HCD MS/MS of NDV V4-VAR HN peptide M175-R197

The amino acid sequence change I175M in V4-VAR from QLD/66 (UniProt ID P13850) has been highlighted in red in the sequence schematic. All b- and y-ions indicated in the schematic were observed in the MS/MS spectrum of the precursor at *m/z* 916.399 (3+) with a precursor mass tolerance of -0.78 ppm. Not all ions have been labelled on the spectrum for ease of interpretation. Full sequence coverage was obtained through y-series ions and y-22 (highlighted red in the spectrum) confirms the mass of methionine at the N-terminus of the peptide.
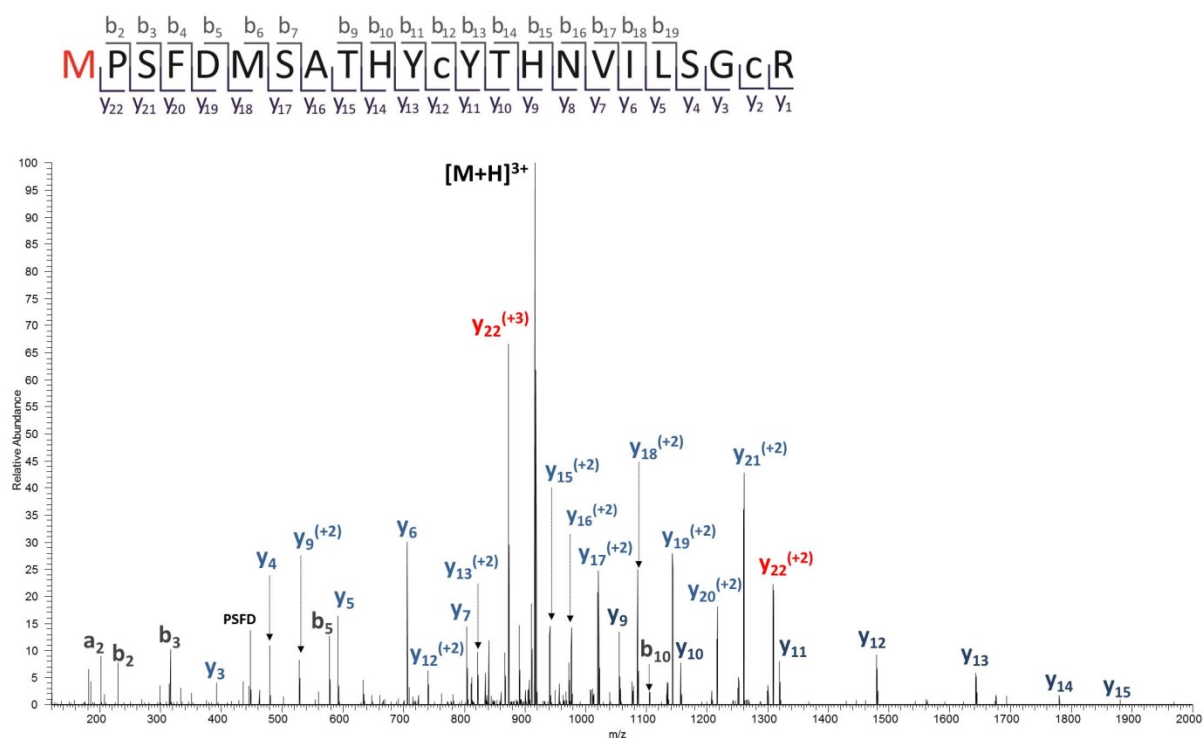
Appendix E

Protein sequence coverage of hRSV sF based on identified peptides from Mascot searches after HCD MS/MS

(a) RSV sF was digested with trypsin and a NCE of 30% was used for the MS analysis. Sequence coverage of was 84.86%. (b) RSV sF was digested with trypsin and a stepped (25±5%) NCE was used for the MS analysis. Sequence coverage of was 85.74%. (c) RSV sF was digested with trypsin then PNGase F and a stepped (25±5%) NCE was used for the MS analysis. Sequence coverage of was 80.11%.

(a)

(b)

(c)

```
          1        11         21         31         41         51         61         71         81         91
  1                        G                                                                                G        G
                           DD          D  C                                                   G        D     D  OD
      MELLILKANA ITTILTAVTF CFASGQNITE EFYQSTCSAV SKGYLSALRT GWYTSVITIE LSNIKENKCN GTDAKVKLIK QELDKYKNAV TELQLLMQST

101           DD           OD  D D    D                            C          D          D          D
      PATNNRARRE LPRFMNYTLN NAKKTNVTLS KKRKRRFLGF LLGVGSAIAS GVAVCKVLHL EGEVNKIKSA LLSTNKAVVS LSNGVSVLTF KVLDLKNYID

201   G    D G C                              D            O D       D O   D D   O DD               C
      KQLLPILNKQ SCSISNIETV IEFQQKNNRL LEITREFSVN AGVTTPVSTY MLTNSELLSL INDMPITNDQ KKLMSNNVQI VRQQSYSIMC IIKEEVLAYV

301              C          C  D      D C       C  D        C          C  O D      D  CD
      VQLPLYGVID TPCWKLHTSP LCTTNTKEGS NICLTRTDRG WYCDNAGSVS FFPQAETCKV QSNRVFCDTM NSLTLPSEVN LCNVDIFNPK YDCKIMTSKT

401              C              D C    D        D    D                                              D  D
      DVSSSVITSL GAIVSCYGKT KCTASNKNRG IIKTFSNGCD YVSNKGVDTV SVGNTLYYVN KQEGKSLYVK GEPIINFYDP LVFPSDEFDA SISQVNEKIN

501   D
      QSLAFIRKSD ELLSAIGGYI PEAPRDGQAY VRKDGEWVLL STFLGGLVPR GSHHHHHHSA WSHPQFEK
```

Retention time profiles of putative N-linked glycopeptides containing site N70 from HCD

(NCE of 25±5%) MS/MS analysis of RSV sF digested by trypsin

(a) The extracted ion chromatogram (EIC) at m/z 204.0867 indicated the presence of putatively glycosylated peptides. (b) The EIC for the calculated Y1 ions of glycopeptides containing N-linked sites N70 with carbamidomethyl Cys residues (^{69}C(+57.0215)NGTDAK^{75}) and pyro-carbamidomethyl (^{69}C(+39.9949)NGTDAK^{75}) delineate the retention times of the glycopeptides. The observed masses of the Y1 ions can be seen in the EIC above some of the more intense Y1 ions.