Sputum Protein *N*-glycosylation in Cystic Fibrosis: a Sweet Response to Bacterial Colonisation

Vignesh Venkatakrishnan

Master of Biotechnology (Macquarie University, Australia)

Department of Chemistry and Biomolecular Sciences

Faculty of Science

Macquarie University, Australia

June 2014

This thesis presented for the degree of **Doctor of Philosophy**

Contents

Abstract v

Statement vii

Acknowledgement viii

Publications and conference proceedings x

Introduction 1

1.1 Cystic Fibrosis 2

1.1.1 Cystic fibrosis transmembrane conductance regulator 2

1.1.2 Mutation in CFTR 3

1.1.3 Hallmarks of CF 4

1.2 Microbiology of CF lungs 5

1.2.1 Bacterial infections 6

1.2.2 Fungal infections 8

1.2.3 Antimicrobial therapy 9

1.3 Sputum: a CF specimen 10

1.3.1 Sputum proteome 11

1.4 Protein *N*-glycosylation and analysis 13

1.4.1 Synthesis of *N*-linked glycans 13

1.4.2 Biological functions of *N*-glycans 15

1.4.3 Methodology involved in glycoprofiling 16

1.4.4 Mass spectrometry 21

1.5 Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis 24

1.6 Project aims 49

Cystic fibrosis and bacterial colonisation define the sputum *N*-glycosylation phenotype 50

2.1 Supplementary file 90

Investigating the link between clinical CF-typed *Pseudomonas aeruginosa* isolates and PAO1 laboratory strain with pauci-mannosidic *N*-glycans....163

3.1 Rationale....164

3.2 Introduction....166

3.3 Materials and methods....170

3.3.1 Isolation of *P. aeruginosa* strains from CF sputum....170
3.3.2 Identification of *P. aeruginosa* strains using 16S rRNA sequencing....170
3.3.3 Biolog Phenotypic microarray (PM) assay....171
3.3.4 *P. aeruginosa* challenge with glycoproteins....172
3.3.5 Extraction of *P. aeruginosa* whole cell proteome....172

3.4 Results....174

3.4.1 Isolation and identification of *P. aeruginosa* strains....174
3.4.2 Monosaccharides as a carbon source for *P. aeruginosa* growth....177
3.4.3 Possibility of exoglycosidase expression by *P. aeruginosa*....178
3.4.4 Qualitative mapping of whole cell *P. aeruginosa N*-glycans....186

3.5 Discussion....188

3.5.1 Absence of exoglycosidase activity secreted by *P. aeruginosa*....188
3.5.2 Utilisation of monosaccharides as carbon source by *P. aeruginosa*....189
3.5.3 *P. aeruginosa N*-glycans....191

3.6 Conclusion....192

Proteome and *N*-glycoproteome of Cystic Fibrosis sputum links the pauci-mannosidic *N*-glycans with abundant neutrophil proteins....193

4.1 Rationale....194

4.2 Introduction....195

4.3 Materials and methods....198

4.3.1 Sputum collection and patient information....198
4.3.2 Isolation of soluble sputum proteins....198

4.3.3 Peptide preparation and glycopeptide enrichment....198
4.3.4 *N*-glycan analysis of human neutrophils....199
4.3.5 Data processing and analysis....200
4.4 Results....201

4.4.1 Proteome analysis....201
4.4.2 Influx of proteins from neutrophil origin....207
4.4.3 Decrease in mucin concentrations in CF sputum....209
4.4.4 Contribution of proteases in sputum....210
4.4.5 Cluster analysis....212
4.4.6 *N*-glycan analysis of neutrophil cells....213
4.4.7 *N*-glycopeptide analysis of CF sputum....214
4.5 Discussion....221

4.5.1 Sputum complexity....221
4.5.2 Chronic infection and inflammation....221
4.5.3 Mucin secretion....224
4.5.4 Proposed new biosynthetic route for pauci-mannose *N*-glycan expression....224
4.6 Supplementary file....228

Conclusions and future directions....244

References....252

Appendix....269

Abstract

Cystic fibrosis (CF) is a prevalent autosomal recessive disease characterised by chronic infection and inflammation in the lungs. The damaged lungs secrete excess mucus that is expectorated as sputum. The lung mucus is the primary site of microbial infection, with *Pseudomonas aeruginosa* as one of the major infecting pathogens in CF patients. Alteration in glycosylation of mucins (heavily *O*-glycosylated proteins) has been shown to play a role in bacterial binding to the mucus in the lungs of CF patients. However, there has been little reported on the *N*-glycosylation of the protein complement of CF sputum.

We have characterized the total secreted protein *N*-glycome of sputum derived from five CF patients, two non-CF patients infected with pathogens (iNCF) and two pathogen free non-CF (NCF) patients. The glycosylation difference between CF and iNCF, compared to between iNCF and NCF, sputum separates the role played by the CF disease itself in regulating *N*-glycan expression as distinct from the effects of bacterial colonisation. A significant increase in unusual and truncated pauci-mannosidic *N*-glycans and a decrease in complex and hybrid *N*-glycans on the sputum proteins were observed in CF and iNCF sputum compared to NCF sputum, suggesting a common effect of infection.

To address the link between the variation in CF sputum *N*-glycans and bacterial colonisation, four *P. aeruginosa* clinical strains (PASS1, PASS2, PASS3 and PASS4) were isolated from sputum of CF patients. We showed that secreted bacterial exoglycosidases were not the cause of the trimming of the *N*-glycans to pauci-mannosidic structures and that the bacterial proteins do not carry pauci-mannose *N*-glycans. The presence of the pauci-mannose *N*-glycans on the sputum proteins were thus not directly of bacterial origin.

Proteomic analysis revealed an up-regulation of proteins involved in the immune response (specifically neutrophil proteins) in bacterial colonised (CF and iNCF) compared with non-infected (NCF) sputum. State-of-art site-specific *N*-glycoproteomic analysis of CF, iNCF and NCF sputum proteins showed that the abundant proteins of neutrophil origin, such as myeloperoxidase and azurocidin, carried these truncated pauci-mannosidic *N*-glycans.

Whilst abundant knowledge regarding mucin *O*-glycosylation is available in the literature, this study provides new structural *N*-glycan and site-specific *N*-glycoproteome information to aid in understanding the complex cellular and molecular environment of the CF affected respiratory tract.

Statement

I hereby certify that the work presented in this thesis titled "Sputum protein *N*-glycosylation in cystic fibrosis: a sweet response to bacterial colonisation" is the result of mu own work except where acknowledged and is not being submitted for a higher degree to any other University or Institution. Human ethics (approval number: 5201100423) and biosafety (approval number: VIV030712BHA) approval has been duly obtained for using human samples. I consent to a copy of this thesis being available in the University library for consultation, loan and photocopying forthwith.

Vignesh Venkatakrishnan

Acknowledgement

The last four years had been an unforgettable journey and would not have been easy without the love and support of friends and family. I take this opportunity to thank everyone involved and please accept my apology if I miss anyone. First and foremost, thanks to my supervisor Nicki who gave me this opportunity, guided me and forever grateful for the words of wisdom and constant support till the last minute. I owe a lot to Morten who played a major role in shaping up my thesis and thanks for every small meeting that taught me something new. Thank you Helena for being kind to allow me to attend the "EDGE" group meeting and all your inputs especially in microbiology helped me a lot.

Thanks to ever-smiling Robyn for helping me in critical revision of my thesis and to Liisa for making laboratory life easier for all the students (and for all the delicious cakes!). Their patience and enthusiasm has helped me a lot through my PhD life. My appreciation likewise extends to all the "Glyco@MQ" members particularly to Arun, Matthew, Chi-Hung, Jodie, Jenny, Katherine, Zeynep, Ling, Edwards, Ian, Aneesh, Yukie, Wei, Terry, Maja, Jasmine and Clara. Special thanks to Merrina for putting up with me in the same office for three and half years and for letting me play tamil songs. I owe a lot to all of you for all your support and constant encouragement and sorry for not acknowledging individually, this thesis is a result of all your suggestions and criticism.

Thanks to everyone in Super Science group, this brought students and professors from different field to single platform and thus contributing immensely to my understanding of other "OMICS" field. Thanks to the collaborators A/Prof Sharon Chen and A/Prof Wieland Meyer for donating us the sputum sample and to all patients who indirectly gave their sample to me. Having got the opportunity to spend few weeks in Dr.Niclas Karlsson's laboratory at University of Gothenburg, I'm grateful to him and Dr.Chunsheng Jin for having me there and expanding my expertise to analyse mucins.

There are countless number of other friends in and around the University to whom I would like to extend my sincerest gratitude particularly Karthikeyan, Jashan, Robert, Iniga/ Martin, Shabnam/ Masoud, Karthik, Sheemal, Rajeev, Sridevi, Prashanth, Mehdi, Andrew and Ishan.

Long distances never bothered as you guys (Kiran, Sarath, Pandia, Guru, Harish, Ganesh, Ramana, Karthi, Krishna, Su, Balaji and Ram) were a phone call away for all I needed and silently wishing me from different parts of the world.

"A friend in need is a friend indeed"- True to that saying, there were few friends without mentioning them this won't be over. Raghav/ Dharmini ka, Deepak/ Keerthi, Krishna/ Viji ka, Sibi/ Sumthy and Mohan anna you guys mean a lot to me. I have had some best and worst moments for the past three years and at every moment, Manveen thank you for standing next to me, helping me whenever I needed, for being the best stress buster and for all the coffee and lunch breaks. You all are my family away from home and your wishes, support and love will always be cherished by me.

Finally my family amma, appa, ahju, Ranjani, athimber, Venkatesh anna/ manni and everyone, I thank you all for persisting with me and believing me all through. I hope I won't shout at you ma if something here doesn't work and your struggles were a motivation for me and will make your dream come true soon. I dedicate this thesis to all the sacrifices you and appa have made to see me happy in my life.

Thank god for the coffee for keeping me up and fresh all the time and medicines for not letting me suffer for long. Final thank to almighty for giving me the patience and persistence at every bad (you could have been nice!) and good step. I thank you all (and anyone I forgot to mention), for putting up with me, guiding me all the way and teaching me something throughout :).

Publications and conference proceedings

Publication 1 (Chapter 1)

Venkatakrishnan V, Packer NH, Thaysen-Andersen M, Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis. ***Expert Review of Respiratory Medicine***, 2013, 7(5): 553-576.

Publication 2 (Chapter 2)

Venkatakrishnan V, Thaysen-Andersen M, Chen SCA, Nevalainen H, Packer NH, Cystic fibrosis and bacterial colonisation define the sputum *N*-glycosylation phenotype. ***Glycobiology***, September 2014 (Accepted).

Conference proceeding 1

Venkatakrishnan V*, Thaysen-Andersen M, Nevalainen H, Packer NH, November 2012. Bacteria change the *N*-glycan profile of the sputum proteins from cystic fibrosis patients. Proteomics & Beyond Symposium, Sydney, Australia.

*Oral presentation by Venkatakrishnan V

Conference proceeding 2

Venkatakrishnan V*, Thaysen-Andersen M, Nevalainen H, Packer NH, June 2013. Bacteria change the *N*-glycan profile of the sputum proteins from cystic fibrosis patients. 22nd International Symposium on Glycoconjugates, Dalian, China.

*Oral presentation by Venkatakrishnan V

Conference proceeding 3

Venkatakrishnan V*, Thaysen-Andersen M, Nevalainen H, Packer NH, January 2014. Detailed *N*-glycome characterisation of sputum reveals that cystic fibrosis patients carry a unique glycosylation phenotype. 27th International Carbohydrate Symposium, Bangalore, India.

*Oral presentation by Venkatakrishnan V

Conference proceeding 4

Penesyan A*,**Venkatakrishnan V**, Kumar S, Thaysen-Andersen M, Packer NH, Paulsen IT, September 2013. Freshly isolated Pseudomonas aeruginosa cystic fibrosis strains display phenotypic traits not observed in common laboratory strains. 14th International Conference on Pseudomonas, Lausanne, Switzerland.

*Poster presentation by Penesyan A

Conference proceeding 5

Thaysen-Andersen M*, **Venkatakrishnan V,** Nevalainen H, Packer NH, June 2014. Linking protein N-glycosylation and CF: malfunctioning CFTR and pulmonary infection yield aberrant N-glycosylation of sputum derived proteins of CF individuals. 37th European Cystic Fibrosis Conference, Gothenburg, Sweden.

*Poster presentation by Thaysen-Andersen M

Conference proceeding 6

Thaysen-Andersen M*, **Venkatakrishnan V**, Parker BL, Packer NH, June 2014. In-Depth Glycomics and Glycoproteomics of Sputum Proteins Reveal Unique N-glycosylation Signatures in Individuals with Cystic Fibrosis. 9th International Symposium on Glycosyltransferases, Porto, Portugal.

*Oral presentation by Thaysen-Andersen M

Chapter 1

Introduction

Introduction chapter contains a review article in the end (**Section 1.5**) published in Expert Review of Respiratory Medicine journal

Venkatakrishnan V*, Packer NH, Thaysen-Andersen M (2013), Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis, **Expert Review of Respiratory Medicine**, 7(5), 553-576.

*In this publication, I have contributed to the host-pathogen interaction section, qualitative compilation and comparison of glycan structures from sputum of cystic fibrosis and non-cystic fibrosis individuals (Figure 2, Table 2) and to the central figure (Figure 1) showing the simplified cellular and molecular cartoon of the host and lung environment in the presence of *Pseudomonas aeruginosa*.

1.1 Cystic Fibrosis

***"The disease has contributed much more to science than science has contributed to the disease"*- Riordan J. R., (1989)**

Cystic fibrosis (CF) is an autosomal recessive genetic disorder that was first identified as a disease in 1938; an autopsy report of a malnourished infant showed mucus plugging of the glandular ducts, a condition termed "cystic fibrosis of the pancreas" (Andersen 1938). The disease was then further characterised by malabsorption of fat and protein, growth failure, pulmonary infection and major damage to the pancreas resulting in pancreatic insufficiency (Craig et al. 1957). In addition to the pancreas, it was soon discovered that most of the exocrine glands were affected by the disease, which became classified as a "generalized excrinopathy" (Di Sant'Agnese 1956). According to the Cystic Fibrosis Foundation, USA, median survival age of CF patients in the 1940s and 1950s was less than one year but had gradually increased to 37 years in 2008 and continues to improve (CFFoundation 2008). One in 2500 live births among Caucasians are affected by CF and one in 25 people are carriers of the responsible genetic mutation (Massie et al. 2000; Massie et al. 2007).

Although CF affects each of the exocrine glands, the impact on the lungs and respiratory system is the most life threatening aspect of the disease. CF cannot be characterised by a single specific problem, but rather manifests in multiple complications ranging from airway obstruction, difficulty in breathing, inflammation and infection with a variety of bacterial and fungal pathogens. This literature survey presents the major research findings to date on the multiple causes and effects of the disease.

1.1.1 Cystic fibrosis transmembrane conductance regulator

More than 50 years after the identification of the CF disease, the CF gene was identified from cells isolated from the sweat duct of a CF patient by Riordan, Rommens and Tsui (Kerem et al. 1989; Riordan et al. 1989). The CF gene is located on the long arm of chromosome 7 and is approximately 250 kb in length with 27 exons and 26 introns (Riordan et al. 1989). It encodes the cystic fibrosis transmembrane conductance regulator protein (CFTR), which is a member of the ATP Binding Cassette (ABC) family and is defective in CF due to mutation. CFTR (Uniprot ID: P13569) is a glycoprotein that is 1480 amino acids long and approximately 168

kDa in molecular mass (Riordan et al. 1989). The main function of CFTR is the regulation of the ion channels, allowing sodium, potassium and bicarbonate ions to flow in and out of the cells to provide an electrolyte balance (Wine 1999; Quinton 2001). Bicarbonate ion channel facilitated by the functional CFTR appears to be crucial for mucus transport. One of the recent studies have shown that the absence of bicarbonate ions or the inhibition of bicarbonate transport decreases the amount of stimulated mucus released mainly MUC2 mucin in intestine (Garcia et al. 2009; Gustafsson et al. 2012).

1.1.2 Mutation in CFTR

Over 700 mutations have been identified in the CFTR gene that can be linked to the CF phenotype. CFTR mutations are broadly classified into five types. *Class I* mutations cause major defects in CFTR synthesis resulting in the absence of CFTR production. In *Class II* mutations, there is production of abnormal CFTR, which then fails to escape the endoplasmic reticulum. *Class III* mutations result in CFTR production and intracellular trafficking, but there is a disruption of activation and regulation at the cell membrane. In *Class IV* mutations, CFTR is expressed at the cell membrane but chloride conductance is reduced. Finally, *Class V* mutations decrease membrane CFTR function by decreasing splicing of normal CFTR (Doull 2001).

The most common mutation in the CFTR gene is ΔF508 (*Class II*), the deletion of phenylalanine at the 508th amino acid position resulting in the production of abnormal CFTR or the absence of CFTR. This mutation stops the biosynthetic processing of the CFTR protein, which is then retained and degraded in the endoplasmic reticulum (Cheng et al. 1990). CFTR is normally located at the apical surface of epithelial cells in sweat glands, intestines, pancreas and airways. Lack or loss of functional CFTR leads to an imbalance in the movement of electrolytes. In particular, the disruption of chloride and potassium ion channels stops Cl^- transport across the apical epithelial surface. Defective salt transport across the epithelium fails to hydrate the airway surface liquid (ASL) causing the mucus to be dehydrated. Dehydration of the mucus causes the cilia to stop beating and with arrested ciliary beating, mucociliary clearance becomes difficult (Pilewski et al. 1999).

1.1.3 Hallmarks of CF

Defective mucociliary clearance of thick stagnant mucus in the lung airways of a CF patient leads to infection and inflammation, which together further deteriorate the lung function. The severity of CF itself is characterised by the level of infection and inflammation, the major causes for increased morbidity and mortality rate among CF patients. The relationship between infection and inflammation in a disease like CF is complex, particularly in regards to cause and effect mechanisms i.e. whether infection leads to inflammation or vice versa. The relation between infection and inflammation can be argued in two different ways: 1) inflammation arises first due to lack of functional CFTR, which increases the activation of immune response (inflammatory cells). Infection amplifies the activation producing more neutrophils, which leads to tissue damage; and/ or 2) the thick viscous mucous secretions, caused by the lack of functional CFTR lead to bronchial obstruction, and help pathogens to infect and colonise. The infection causes inflammation which activates the increased proinflammatory cells followed by tissue damage.

Numerous studies have demonstrated that exacerbated inflammation can be an important primary component of CF airway disease. Clinical studies have shown the early presence of inflammation in the CF infant with an increased amount of proinflammatory chemokines such as interleukin (IL-8) in the lung intraluminal fluid irrespective of bacterial infection (Khan et al. 1995; Noah et al. 1997). Potential lung damage can also occur from the imbalance in protease and anti-protease activity in the lung, as demonstrated by Birrer *et al.*(1994) where an increased amount of neutrophil elastase was found in the respiratory epithelial lining fluid of CF children (Birrer et al. 1994). In other bronchoalveolar lavage fluid (BAL fluid) studies of CF, pathogen-free BAL samples from children contained an increased number of neutrophils as evidence for neutrophilic inflammation, suggesting inflammation need not be a response to infection in CF (Balough et al. 1995; Rosenfeld et al. 2001). The pro- and anti-inflammatory cytokines and IL-8 concentrations were increased in BAL fluid irrespective of whether CF pathogens were isolated. Using a mouse model, Zahm *et al.* (1997) showed airway inflammation in BAL, as assessed by an increase in IL-8 and neutrophil content, can occur irrespective of *P aeruginosa* infection (Zahm et al. 1997). In another study, lack of functional CFTR was shown to be responsible for the onset of airway inflammation in

the CF lung *in vitro;* inhibition of CFTR mimicked the CF inflammatory profile (Perez et al. 2007). Using a CF porcine model, Adam *et al.* (2013) has identified the presence of air trapping, airflow obstruction and airway size reduction in newborn piglets with CF before the onset of airway infection and mucus accumulation suggesting that CF impacts airway development and thus indirectly affects the onset of airway infection (Adam et al. 2013).

Despite the above evidence that inflammation alone can be a primary factor in CF, there are also many studies that have demonstrated that infection of the lungs can itself cause an immune response resulting in increased inflammation. In a comparison of CF and non-CF sputum, an increase in total cell counts, absolute neutrophil cell count, IL-8 level and neutrophil elastase activity were found in CF subjects, and in the same samples CF related bacterial pathogens were isolated (Sagel et al. 2001). Neutrophils, being the pro-inflammatory cells acting as a host response to defend against pathogenic invasion, can act as a double edged sword. Chronic inflammation and infection requires the host to keep defending, but overwhelming production of neutrophil products causes tissue damage (Weiss 1989). In a study by Nunley *et al.* (1998) CF and non-CF patients who had received lung transplants, an increased number of neutrophil or polymorphonuclear (PMNs) cells were found in the BAL fluid of all patients when patients with and without *P. aeruginosa* colonisation were compared (Nunley et al. 1998). Cigarette smoke-induced (non-CF) lung inflammation based on measurements of PMNs, macrophages, B cells, CD4 and CD8 lymphocytes, has also been shown to be amplified in severe emphysema associated with latent adenoviral infection (Retamales et al. 2001).

Understanding the role of pathogenic invasion in a disease like CF is important as it is also responsible for the increase in mortality rate. Colonisation with multiple microbial colonies eventually leads to the degradation of the lung architecture directly or indirectly. Thus it is necessary to identify the contribution to the disease of bacteria colonising CF lung airways.

1.2 Microbiology of CF lungs

"The microbiology of lung disease in patients with CF is a specialty unto itself"- (Stutman et al. 1987). Defective mucociliary clearance due to airway obstruction provides the right environment for a number of pathogens to infect and colonise the lung. The reports of

infection occurring in the respiratory tract of a CF patient date back to the early 1940s (Di et al. 1946). More recently, CF patients with numerous problems associated with the lungs or health in general, have been found to harbour hundreds of different pathogenic species in the lung (Yang et al. 2011).

The number of bacterial and fungal pathogens that have been isolated from the airways of CF patients is vast, but infections with certain pathogens are more serious and responsible for the increased mortality in CF. The major pathogen species that are associated with the CF lung include the bacteria *P. aeruginosa, Staphylococcus aureus, Burkholderia cepacia, Haemophillus influenza* and the fungi *Aspergillus fumigatus*, and *Scedosporium apiospermum* (Miller et al. 2003); regular identification of the microbial flora is a necessity as a part of designing antibiotic treatment. The pathogens exploit various resources available within the host for adhesion, infection and colonisation of the CF lung. The following sections focus on some of the common microbial infections in CF.

1.2.1 Bacterial infections

Understanding the bacterial ecology in the lung is of considerable importance in CF as chronic bacterial infection is one of the main causes of respiratory damage. Most of the bacterial species are acquired from the environment and in some cases person-to-person transmission occurs between CF patients (LiPuma et al. 1990; Goerke et al. 2000). According to the US Cystic Fibrosis Foundation National Patient Registry, 2004, prevalence of respiratory bacterial infections changes along with the age group (**Figure 1**). For example, the major lung coloniser in children is *S. aureus* and in adult CF lung is *P. aeruginosa* (CFFoundation 2004).

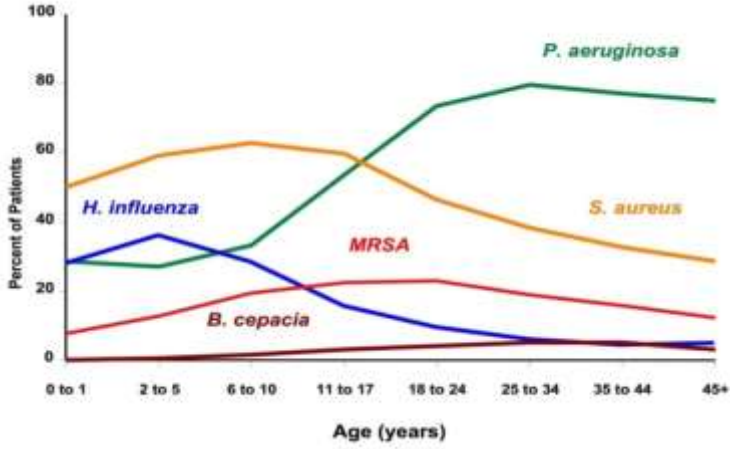


Figure 1: Bacterial colonisation progression of CF lungs with respect to increase in age. Infant and young CF patients are infected more frequently with *S. aureus* and *H. influenza* whereas adults are more likely to be infected with *P. aeruginosa*, which typically results in chronic infection. MRSA refers to methicillin resistant *S. aureus*. This figure was obtained from the Cystic Fibrosis foundation patient registry, 2004 (CFFoundation 2004).

1.2.1.1 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is one of the main bacterial pathogens responsible for the increased mortality and morbidity rate in CF. It is a gram negative, opportunistic pathogen present in a range of environmental conditions, capable of infecting the immunocompromised lung such as in CF. *P. aeruginosa* has been reported to be isolated at a very early age during infancy (Burns et al. 2001). Planktonic or non-mucoid *P. aeruginosa* isolates are converted into the most virulent form of mucoid *P. aeruginosa* as a result of chronic infection. Mucoid *P. aeruginosa* is generally non-motile and is capable of successful colonisation in the lung by forming a biofilm that is very difficult to eradicate. Mucoid strain biofilms comprise secreted polysaccharide material known as alginate that forms a mesh around the bacteria, which increases resistance to antibiotics and clearance by the immune system, thus resulting in acute chronic infection. The lung function becomes damaged at this stage, which is also due to the neutrophil derived proteases secreted as an immune response against the biofilms (Lyczak et al. 2002). *P. aeruginosa* infects more than 80% of adult CF patients and due to its increased prevalence in the CF lung and higher virulence, has become one of the most studied organisms in CF. The adhesion of *P. aeruginosa* to the human host and the adhesins utilised as the first stage of infection are explained in **Section 1.5 (Venkatakrishnan et al. 2013)**.

1.2.1.2 *Staphylococcus aureus*

Staphylococcus aureus is a gram positive, non-motile, non-spore forming commensal bacterium commonly found in the upper respiratory tract and nasal passages of healthy individuals and CF patients. It is one of the organisms that has been identified in young CF children at autopsy (Snelling et al. 1942; Pryles 1958). Often preceding *P. aeruginosa* infection, *S. aureus* is identified in more than 50% of patients affected by CF (CFFoundation 2002). The prevalence of *S. aureus* in CF patients who have not undergone antibiotic treatment is shown to be significantly higher than in treated CF patients and healthy controls (Goerke et al. 2000). Another study has shown increased adherence of a *S. aureus* strain isolated from a CF patient to bronchial and nasal epithelial cell lines compared to *S. aureus* isolated from non-CF subjects (Schwab et al. 1993). *S. aureus* also has affinity towards human bronchial mucins (Trivier et al. 1997). Asialoganglioside (aGM1), a bacterial receptor on bronchial epithelial cells also acts as a strong receptor to bind to *S. aureus* (Imundo et al. 1995).

1.2.2 Fungal infections

The identification of invasive fungal infections has dramatically increased over the past three decades with the newer and continuous monitoring of clinical samples (Alexander 2002). Incidence of fungal infections was once underestimated due to lack of sensitive culture techniques to isolate fungi from sputum, BAL fluid or any respiratory tract samples (Middleton et al. 2013). Accurate diagnosis of fungal infections is important for early provision of antifungal therapy to the CF patients. *Aspergillus fumigatus*, a spore-bearing filamentous fungus is the most identified fungal species in CF sputum. *A. fumigatus* is capable of causing pulmonary aspergillosis or allergic bronchopulmonary aspergillosis (ABPA), a condition that leads to inflammation and damaging of airway cells, occurring in 1 to 15% of patients (Knutsen et al. 1992; Stevens et al. 2003; Shoseyov et al. 2006).

Scedosporium species have also been linked with CF pathogenesis. *Scedosporium apiospermum*, is the second most common filamentous fungus associated with CF, based on the frequency of identification from bronchial secretions of CF patients (Cimon et al. 2000). *S. aurantiacum* is another representative of *Scedosporium* spp., capable of causing a range of serious infections, and has been recently isolated and identified specifically in Australian CF patients (Harun et

al. 2009; Harun et al. 2010). *Candida albicans*, a yeast, is another organism that has been frequently detected in the sputum of CF patients (Burns et al. 1998; Bakare et al. 2003).

1.2.3 Antimicrobial therapy

Development of antibiotics for treatment of multiple infections is one of the biggest achievements in medicine. Correspondingly, treatment with antibiotics to inhibit or kill the infecting pathogen is the most common therapy that is given to CF patients. In particular, treatment against multi-drug resistant *P. aeruginosa* is a regular practice for CF patients. Identification of *P. aeruginosa* during the initial colonisation is very important as antibiotic treatment during initial colonisation postpones chronic infection, which prevents worsening of pulmonary damage of lungs (Frederiksen et al. 1997). Established chronic infections, are difficult to clear, but aggressive antibiotic treatment with multiple combinations can temporarily stop the spreading of infection in its early stages (Valerius et al. 1991; Hoiby 1992). Mucoid *P. aeruginosa*, a sign of chronic infection, is more resistant to antibiotics by restricting the penetration of antibiotics with a viscous mesh formed by alginate and pus matrices (Cabral et al. 1987; Doring et al. 2000). Pulmonary aspergillosis, caused by chronic *A. fumigatus* infection, remains difficult to eradicate despite the introduction of new antifungal agents (Castric et al. 2001). Antifungal treatment with itraconazole against *A. fumigatus* infection was observed not to improve the pulmonary function in patients with CF in a recent clinical trial (Rorvig et al. 2013).

Multiple-combination bactericidal testing (MCBT) is one of the new options for directed antibiotic therapies, which is achieved by testing the bactericidal activities of various antibiotics against a range of bacteria, then using the best combination of antibiotics to treat the patients (Aaron et al. 2000; Lang et al. 2000; Saginur et al. 2006). However, despite the advances in this area, antibiotic treatment still fails to completely eradicate infecting pathogens in CF. Treatment of the early colonisation of infecting pathogens has given temporary relief to patients but the success rate of antibiotics against chronic infections declines, which makes antibiotic therapy insufficient for the effective treatment of CF patients in the long term.

To summarise, a vicious circle of airway obstruction, inflammation and infection presents many challenges to CF patients and clinicians. The identified origin of the disease is the defect due to a mutation in functional CFTR causing imbalance in the flow of electrolytes in and out of the cell. As a result, there is a build-up of thick, viscous and dehydrated mucus in the lung, followed by infection and inflammation leading to pulmonary damage. Chronic inflammation and infection, considered being the major characteristics of CF mortality and morbidity, complement each other and the question of which comes first has remained an unsolved puzzle. Mucus are a good habitat for multiple CF pathogens and exacerbate the problem of infection clearance. The continuous immune response against infection in turn affects the host by damaging the lung and other tissues. Antimicrobial therapies have advanced in recent years, but still fail to fully combat infection and mortality amongst the CF patient. Hence, research has been extended to further characterise and understand the relationship between CF and infection via proteomic and glycomic analyses of CF sputum, as detailed in **Section 1.3**.

1.3 Sputum: a CF specimen

Non-invasive expectoration of mucus by mere coughing or saline induction from a chronically inflammed lung is termed sputum. Sputum reflects the central lung airways and is widely recognised as a useful fluid for clinical diagnosis and monitoring of the disease. The non-invasive sampling method is also advantageous when using sputum as a disease model compared to BAL fluid and bronchial biopsy, which require bronchoscopy. Using a radiolabeled aerosol bolus delivery technique, the saline induced sputum derived from the central airway has been shown to contain little or no contribution from the peripheral airways (Alexis et al. 2001). Since sputum represents the major part of the contents of the lung airway, the information regarding the protein components, cellular compartments and microbial infection can potentially assist in the identification of biomarkers for severity of the disease and its progression.

Samples of sputum collected from a patient typically consist of mucins and non-mucin proteins, inflammatory cells, inorganic salts, DNA, lipids, actins, dead and live bacterial cells, pus and contaminants such as saliva (Matthews et al. 1963). Sputum is regularly used in hospitals for screening of microbial infections present in the airway of the CF lung.

Sputum collection from CF patients at regular time intervals helps in understanding the lung microbial ecology and the severity of the infection. Determination of microbial ecology in turn helps in setting up the antibiotic treatment for the patients.

Induced sputum is also analysed frequently to determine inflammatory markers in not only CF but other lung diseases such as asthma and bronchiectasis (Gray et al. 2008). The measurement of cellular profiles (neutrophils, eosinophils) and cytokine profiles of the sputum are also used as inflammation markers in lung diseases. One of the major characteristics of airway inflammation is the constant influx of polymorphonuclear leukocytes (PMNs) or the neutrophils into the airway by host and bacterial chemoattractants that include cytokines (Sagel et al. 2001). As such, neutrophil cells are inflammatory cells considered as a biomarker of CF airway inflammation. Elevated concentrations of pro-inflammatory cytokines such as tumour necrosis factor (TNF-alpha) and interleukins (IL-6 and IL-8) were observed in CF children compared to non-CF children with bronchiectasis (Osika et al. 1999). An increase in the cytokines was observed in CF sputum from chronic *P. aeruginosa* infected lung (Kronborg et al. 1993). Numerous studies in this way have shown the utilisation of induced sputum to assess the inflammation in the CF airway (Greally et al. 1993; Richman-Eisenstat et al. 1993; Koller et al. 1997).

1.3.1 Sputum proteome

Induced sputum reflects the composition of the lung airway, therefore characterisation of the whole sputum proteome has the potential to identify biomarkers of CF airway inflammation and infection. Knowledge of protein variation in CF and the subsequent post-translational modifications could play an important role in understanding the disease and also the pathogenesis. The sputum proteome can be divided into mucin and non-mucin proteins, in which the mucins constitute approximately 2-4% of the total weight of the sputum (Nicholas et al. 2009).

1.3.1.1 Mucins

Mucins are heavily *O*-glycosylated proteins usually with tandem repeats of serine/ threonine amino acid residues and a high-molecular mass. Mucins are the main protein component of the airway surface liquid (ASL) and are secreted mainly by goblet cells and submucosal

gland (SMG). Mucins are membrane bound to airway epithelial cells such as ciliated and non-ciliated epithelial cells and different mucins are secreted by each cell type in the airway. Mucins constitute the majority of the sputum and two of the most abundant secreted mucins in CF airway are MUC5AC and MUC5B (Schulz et al. 2007). Over the past 30 years, many studies have characterised the mucins from CF and non-CF sputum and extensive work has been performed to understand the role of mucins in CF disease pathogenesis. A detailed account of aberrant host mucin *O*-glycosylation (secreted and membrane bound mucins) and the role of mucins in CF pathogenesis is provided in our published review presented in **Section 1.5** (Venkatakrishnan et al. 2013)**.**

1.3.1.2 Non-mucin proteins

Apart from mucins, other proteins from different cellular origins such as inflammatory cells and epithelial cells are abundant in induced sputum and are referred to as non-mucin proteins. The number of studies focusing on the proteomic aspect of induced sputum is relatively few. One such study by Sloane *et al.* (2005), involved the proteomic analysis of sputum from CF adults and children and healthy control subjects. Proteins were analysed by two-dimensional (2-D) gel electrophoresis to determine the differentially expressed proteins, and protein identification was carried out by mass spectrometry (MS). A number of neutrophil- derived and inflammation associated proteins such as myeloperoxidase, IL-8 and α_1-antitrypsin were found to be differentially expressed in CF subjects compared to the control (Sloane et al. 2005).

An increasing number of studies have investigated the lung airways by performing shotgun proteomic analysis of different biofluids including sputum, BAL fluid, epithelial lining fluid, nasal lavage fluid and saliva (Noel-Georis et al. 2002; Huang 2004; Casado et al. 2005; Nicholas et al. 2006; Kipnis et al. 2008). A study by Nicholas *et al.* (2006), involved analysis of three induced sputum samples obtained from a healthy female with a smoking history. A combination of 2-D gel electrophoresis and GeLC-MS/MS, revealed a total of 191 confident matches to human proteins with most being neutrophil-derived proteins (Nicholas et al. 2006). The same study also compared the sputum proteome with the proteome of BAL fluid and saliva, which had been previously reported in 11 other studies. The similarities and

differences between the proteomes of the three bodily fluids is represented as a Venn diagram showing a large variation between the three origins (**Figure 2**) (Nicholas et al. 2006).

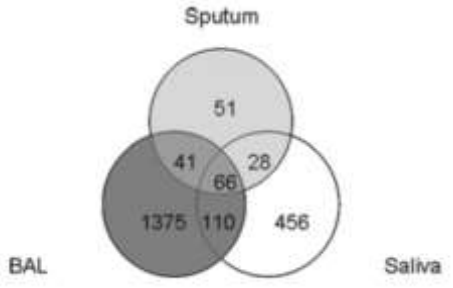


Figure 2: Venn diagram representing variation in the proteome obtained from sputum, BAL fluid and saliva (Nicholas et al. 2006). Although there is a subset of overlapping proteins between sputum, BAL fluid and saliva, there are number of proteins present specifically to each category indicating the overall difference between different sample types.

Although there are only a limited number of studies that have analysed the proteins in sputum samples from CF patients, a number of other studies list the proteins obtained from neutrophils alone (Lominadze et al. 2005; Rorvig et al. 2013). The major neutrophil-derived proteins are myeloperoxidase, matrix metalloproteinases (MMP), leukocyte elastase, complement C3 and α1-antitrypsin. These and other proteins in the sputum carry *N*-glycans but in comparison to the extensively studied mucin *O*-glycosylation (**Section 1.5**), the *N*-linked glycosylation of CF sputum has been understudied. Aberrant *N*-linked glycosylation has been shown to play a role in diseases involving inflammation and infection such as rheumatoid arthritis and systemic lupus erythematosus (Turner 1992; Gornik et al. 2008). Chronic inflammation and infection are the hallmarks of CF disease and sputum sampled from central airway of the lung constitutes of both infection and inflammatory markers. Therefore, understanding the sputum protein *N*-glycosylation will provide new insights into the mechanisms of inflammation and infection of the lungs of CF patients.

1.4 Protein *N*-glycosylation and analysis

1.4.1 Synthesis of *N*-linked glycans

N-linked glycosylation is a common post-translational modification of a protein in which the sugar molecules (*N*-glycans) are covalently attached to a protein at aspargine (Asn) residue in the amino acid sequon Asn-X-Ser/Thr (N-X-S/T) where X is any amino acid except proline. *N*-glycosylation is common to all eukaryotes and prokaryotes with specific structural

variations between species (Schwarz et al. 2011). The biosynthesis of *N*-glycans involves many enzymatic actions of glycosidases and glycosyltransferases (**Figure 3A**). It begins with the formation of a lipid linked oligosaccharide precursor, through the sequential addition of monosaccharides by the action of various glycosyltransferases to form the glycan precursor $Glc_3Man_9GlcNAc_2$. This is followed by *en bloc* transfer of the precursor to the proteins with polypeptide consensus sequence (N-X-S/T) in the endoplasmic reticulum (ER) catalysed by oligosaccharidetransferase (OST). The transferred 14 monosaccharide precursor undergoes further processing by exoglycosidases, which trim the glucose monosaccharide residues at the non-reducing end with glucosidase I and II. Glycans are trimmed down to mannose trisaccharide core by the action of several Golgi mannosidases, followed by the action of a series of *N*-acetyglucoaminyltransferases (GnTI, GnTII, GnTIV and GnTV) plus nucleotide sugars to form bi-, tri- and tetra- antennary glycans. The conjugated *N*-glycan can be further elaborated by the addition of galactose, fucose and sialic acid residues by the action of various galactosyltransferases (GalTs), fucosyltransferases (FUTs) and sialyltransferases (SiaTs) in the Golgi compartment. Depending on the linkage of monosaccharide residues to the tri-mannose core, the *N*-glycans are classified as high mannose, complex or hybrid type *N*-glycans (**Figure 3B**).Different modifications are also possible, such as addition of GlcNAc residue to the branching core mannose by β1-4 linkage, forming 'bisecting type' complex *N*-glycans.

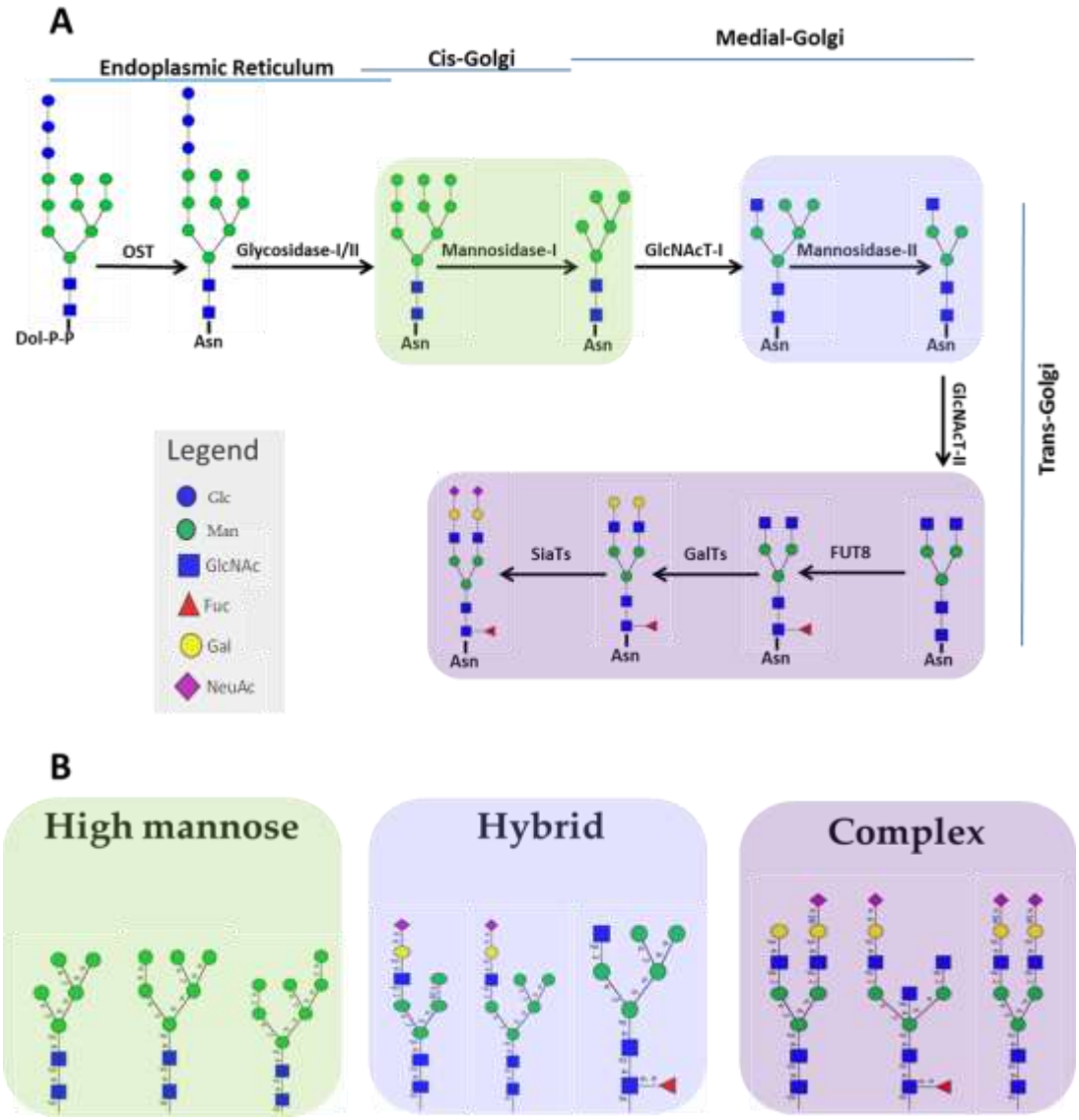


Figure 3: A) Pictorial representation of the *N*-glycan biosynthesis from dolichol precursor to high mannose, hybrid and complex *N*-glycans using the action of different enzymes; B) Three main classes of *N*-glycans synthesised in mammalians namely high mannose, hybrid and complex *N*-glycans.

1.4.2 Biological functions of *N*-glycans

Understanding the function of glycans in biological systems is one of the major aims in glycobiology. To investigate the role of *N*-glycans, the *N*-glycosylation machinery has been manipulated in many ways such as using genetic variation by gene manipulation or by eliminating specific protein glycosylation sites and by using inhibitors of *N*-glycan processing with tunicamycin (Stanley et al. 2009). Glycans are essential for the functioning and survival of the organism and the total inhibition of *N*-glycan synthesis can lead to cell death (Marek et al. 1999). Glycans are also involved in cell-cell interactions, for example L-selectins bind leukocytes involved in innate inflammatory response to endothelial cells

(Foxall et al. 1992). Moreover, glycans play a role in host pathogen interactions, for example sialylated glycans act as a host receptor for influenza virus (Nicholls et al. 2008). Lectins are glycan binding proteins utilised by bacteria for binding to a host. *P. aeruginosa* utilises lectins located on the outer membrane of the cell which act as adhesins to bind to the glycan epitopes in the human host (Venkatakrishnan et al. 2013). Many biological functions of *N*-glycans have been reported so far in the literature but still the knowledge is limited and the roles of many *N*-glycans are yet to be decoded.

Before understanding the function of *N*-glycans, it is important to characterise the glycans and glycoproteins from a given sample. Glycoprofiling of proteins can be performed by two different approaches; 1) glycomics and 2) glycoproteomics. Glycomics is the global characterisation of the glycans released from a complex mixture of glycoproteins. Glycoproteomics involves the simultaneous identification of the glycoproteins and the sites of glycosylation.

1.4.3 Methodology involved in glycoprofiling

1.4.3.1 Glycomics

1.4.3.1.1 Release of *N*-glycans from glycoproteins

One of the widely used methods to release *N*-glycans from glycoproteins is by using an enzyme peptide-N-glycosidase F (PNGase F). PNGase F is an aspargine deamidase with a specific cleavage site between the reducing end GlcNAc and the aspargine residue to which the *N*-glycan is attached (Maley et al. 1989). PNGase F does not cleave *N*-glycans that have fucose attached to the reducing end GlcNAc through an α1-3 linkage, a feature commonly observed in plants and some insect glycoproteins (Tretter et al. 1991); for these *N*-glycans, PNGase A can be used effectively. Other enzymes are also available for specific glycan release. Endoglycosidase H can specifically cleave high mannose and hybrid *N*-glycans from a protein whereas Endoglycosidase F can release biantennary *N*-glycans but not high mannose and hybrid *N*-glycans (Maley et al. 1989). Apart from the enzymatic method, chemical methods can also be used, such as hydrazinolysis, which cleaves the amide bond between the glycan and aspargine residue using hydrazine (Voisin et al. 2007).

1.4.3.1.2 Separation of released *N*-glycans

Detailed characterisation of *N*-glycans is often crucial for a molecular understanding of different biological processes. The main goal behind any glycomic anlaysis is to elucidate the detail on the types of glycan structures present, and their relative quantification. The detailed structural information of a complex glycan can be obtained through different analytical approaches such as nuclear magnetic resonance (NMR) and tandem mass spectrometry in combination with separation techniques such as liquid chromatography (LC) or capillary electrophoresis (CE). Different LC based separation systems used for the separation of released glycans include reverse-phase chromatography (RPC) (Delaney et al. 2001; Wuhrer et al. 2005), porous graphitised carbon chromatography (PGC) (Thomsson et al. 1999), hydrophilic interaction liquid chromatography (HILIC) (Hao et al. 2008; Wuhrer et al. 2009), high pH anion exchange chromatography (HPAEC) (Guignard et al. 2005) and size exclusion chromatography (SEC) (Ziegler et al. 2006).

Reverse-phase chromatography (RPC) is one of the most common separation methods in LC-MS. Due to poor retention of glycans on RP-LC column, increase in hydrophobicity of glycans by derivatising with a hydrophobic tag is required. Hydrophobicity of the tag helps in RP retention and the resolution is based on the glycan features (Charlwood et al. 2000). Several hydrophobic tags such as 2-aminobenzamide (2-AB), 2-aminopyridine (2-AP) and 2-aminobenzoic acid (2-AA) facilitate sensitive and selective detection either by UV absorption or fluorescence and influence the ionisation behaviour in subsequent MS (Wuhrer et al. 2005). Permethylation of mixture components subjected to reversed phase HPLC enables separation of glycans based on alpha and beta anomeric structures (Delaney et al. 2001). Released *N*-glycans from IgG antibodies derivatised with 2-AB and separated on a RP column with on-line fluorescence and MS detection has shown good resolution and segregation of different *N*-glycan types based on column retention, which also enhances quantification through fluorescent signals. The study also demonstrated good ionisation efficiency in electrospray and structural information obtained from the online MS detection (Chen et al. 2007).

Hydrophilic interaction liquid chromatography (HILIC) is another versatile technique for separating glycans that can be coupled with mass spectrometry, and is useful for structural

glycomics. HILIC is a variant of normal phase chromatography in which analytes interact and are retained in hydrophilic stationary phase by ionic interactions, dipole-dipole interactions and hydrogen bonding (Wuhrer et al. 2009). Different HILIC stationary phases are available, including bare silica and silica derivatised with different polar functional groups such as amine, amide, cyano and diol. Solvent systems or the mobile phase for HILIC are aqueous-organic mixtures such as water-methanol and water-acetonitrile. Hydrophilic partitioning of the analyte to the water-rich aqueous layer surrounding the hydrophilic stationary phase is described as the retention mechanism of pure HILIC. The retention of glycans on the neutral polar stationary phase is mainly dependant on the number of hydroxyl groups on the glycans and the larger the glycans the later they elute (Hao et al. 2008).

Unlike RP or HILIC techniques, **Porous graphitised carbon (PGC)** has an outstanding ability to separate non-derivatised oligosaccharides and has been a widely used LC method for separating glycans. The biggest advantage of using PGC compared to other chromatographic techniques is its ability to separate structural isomers with high resolution. PGC coupled with nano-LC/MS was shown to provide sufficient detection sensitivity for analysis of even fmol amounts of oligosaccharides with an increased sensitivity in using nano-scale graphitised carbon- high performance liquid chromatography (HPLC) compared to capillary scale graphitised carbon HPLC (Boyum et al. 1991; Karlsson et al. 2004; Henke et al. 2011; Jin et al. 2012). Microfluidic chip devices with graphitised carbon as the stationary phase enhance sensitivity and are commercially available (*e.g.* Agilent). The high shape-discrimination capability of PGC helps in separating isomeric and isobaric glycan structures, thereby establishing a structure- retention time correlation system (Hitchen et al. 2006). This PGC-LC retention behaviour coupled with tandem MS constitutes a most promising approach for glycomic analyses and thus PGC was the chosen chromatographic technique used in this thesis for separation of glycans.

Apart from these techniques, there are also other systems used for the separation of the released glycans. One of the methods is **capillary electrophoresis (CE)** which allows high resolution separation and reproducible quantification by on-column detection (Suzuki et al. 2001). Pre-capillary chemical derivatisation is required for CE with charged chromophore or

fluorophore such as disulfonated or trisulfonated aminonapthalenes and 8-aminopyrene-1,3,6-trisulfonic acid, which can be detected by UV absorption and/ or fluorescence intensity (Lee et al. 1992; Chiesa et al. 1993; Suzuki et al. 2001).

1.4.3.2 *N*-Glycoproteomics

The complexity involved in analysing site-specific glycosylation brings difficulty in glycoproteome analysis. The amount of information required to characterise a glycoprotein is much more than for unmodified proteins or chemically modified proteins (like phosphorylation and methylation. Glycoproteomics yields information about the protein carriers, the glycan attachment sites and the structure and occupancy of the glycan. The complexity of the glycopeptides due to the presence of two conjugated heterogeneous biomolecules (glycans and peptides) leads to the difficulty in analysing glycoproteome (Roy et al. 2014). Glycoproteins in a complex biological mixture require multiple strategies to enrich glyco-conjugated proteins or glycopeptides for analysis.

Lectin affinity chromatography (LAC) is a widely used method for crude protein prefractionation which utilises the affinity of lectins to bind to glycan epitopes. Lectins bound to membrane, agarose matrix or magnetic beads can be used to isolate and fractionate glycoproteins on the basis of their affinity to different glycan structures. Examples of lectins include Aleuriaaurantialectin (AAL), Canavalia ensiformis agglutinin (Con A), Jaqlin and wheat germ agglutinin (WGA), which have binding specificity towards glycan structures containing fucose, high mannose structures, galactose and α-linked sialic acid respectively (Kronis et al. 1985; Roque-Barreira et al. 1985; Scott et al. 1992; Wimmerova et al. 2003). LAC could be used for the enrichment of both glycoproteins and glycopeptides. Utilising lectins with broad specificity is preferred for achieving maximum glycoprotein recovery. Lectin affinity chromatography with multiple lectins can be used as serial- (SLAC) or multiple- (MLAC) lectin affinity chromatography. SLAC utilises different lectins in a serial pattern with varying glycan specificity to purify different components of a glycoprotein mixture, whereas MLAC utilises multiple lectins in a single step to maximise capture of the majority of glycoproteins present in a complex sample. The drawbacks of using LAC are that no single lectin has sufficient selectivity to cover the whole glycoproteome and it is difficult to eliminate non-specific binding and to enhance the binding capacity.

Hydrophilic interaction chromatography (HILIC) is one of the most favoured techniques used for glycopeptide enrichment; using a polar stationary phase and less polar mobile phase, it is capable of separating glycopeptides from the peptide mixture. HILIC is a variant of normal phase chromatography in which analytes interact and are retained on hydrophilic stationary phase by ionic interactions, dipole-dipole interactions and hydrogen bonding (Wuhrer et al. 2009). Stationary phases commonly employed for HILIC include diols, amines, amides, silanols and various compounds with anionic, cationic and zwitterionic functional groups (Alley et al. 2013). One of the advantages of HILIC is the better retention of glycans and glycoconjugates by hydrogen bonding, while contaminants including salts, detergent and non-glyco peptides have little or no retention (Hirche et al. 2005; Wuhrer et al. 2009).

Apart from techniques discussed above, glycopeptides can also be enriched using other chromatography techniques such as size exclusion chromatography (SEC) and reverse phase chromatography (RP). SEC, as reported by Alvarez-Manilla *et al*. (2006) provides more uniform enrichment on a Superdex Peptide column of the trypsin digested glycopeptides comprising only a small part (3-5%) in the peptide mixtures compared to non-glycosylated peptides (Henke et al. 2004). Sialylated glycopeptides can be enriched using titanium based enrichment analysis (Larsen et al. 2005). Of all the commonly used techniques for glycopeptide enrichment, HILIC is one of the preferred tools to enrich intact glycopeptides from complex peptide mixtures and it is been used as glycopeptide enrichment tool in this thesis.

1.4.3.2.2 LC separation of glycopeptides

Off- or on-line-LC coupled with tandem MS is one of the most used approaches for data acquisition and interpretation in a glycoproteome analysis. Different LC separation systems used for separating glycopeptides include RP-C_{18}, HILIC (Hirche et al. 2005; Pollard et al. 2005) and PGC (Borregaard 2010). Among these, RP separation system is the most favoured and established method in proteomics, which also makes it the most widely used method in glycoproteomics (Wuhrer et al. 2005). One of the advantages of RP-LC separation is that related glycopeptides sharing the same peptide moiety conjugated with a variety of neutral glycans elute closely together just before the non-glycosylated peptide variant (Belaaouaj et al. 1998). Elution can be achieved by increasing an acidic acetonitrile gradient with formic

acid or low concentrations of trifluoroacetic acid, in which the glycopeptides elute slightly earlier compared to non-glycosylated peptides. The hydrophilicity of the attached glycan on the glycosylated peptide helps in early elution compared to non-glycosylated peptides (Wuhrer et al. 2005). The RP system coupled with tandem MS results in detection and structural identification of intact glycopeptides.

1.4.4 Mass spectrometry

Glycan LC separation techniques coupled with mass spectrometry (MS) have become important tools in glycoprofiling. MS based analysis is well-suited for analysing glycoforms with characteristics such as high sensitivity, resolution, mass accuracy and dynamic range that vary between instruments. Tandem MS analysis is one of the most favoured methods for analysis of peptides and glycans with a reliable high throughput performance but unlike proteins, identification of glycans is complex. The MS technique is dependent mainly on the ionisation method, and three of the common methods used for ionisation are electrospray ionisation (ESI-MS) (Everest-Dass et al. 2013), matrix-assisted laser desorption ionisation (MALDI) (Domann et al. 2012) and fast atom bombardment MS (FAB-MS) (Thomsson et al. 1999).

ESI is a soft ionisation technique as the ionisation process imparts little excess energy. ESI involves ionising the liquid analyte by applying a high voltage that creates multiply charged ions during the relatively slow droplet desolvation process. An increase in the size of the molecule causes the ionisation efficiency to decrease, which limits the sensitivity to a limited mass range (Zaia 2010). ESI is sensitive to salts and other contaminants, and it is important to do a buffer exchange before introducing the analytes to mass spectrometry (Bielik et al. 2010). One of the main advantages of ESI is the ability to have online LC separation unit such as PGC coupled with tandem mass spectrometry for separation and analysis of glycoconjugates (Ruhaak et al. 2009).

Unlike ESI, MALDI is sensitive for higher molecular mass analyte and is been widely used for the analysis of glycoconjugates (Domann et al. 2012). Sample preparation is relatively easier with MALDI being tolerant to salts or contaminants. It requires mixing the analyte with an acidic mixture (matrix) forming crystals on a metal plate to which pulsed laser light

is targeted creating ions with analytes that are analysed (Zaia 2010). Choice of matrix includes 2,5-dihydroxybenzoic acid (DHB), 2,4,6-trihydroxyacetophenone (THAP) and α-cyano-4-hydroxycinnamic acid (CHCA).MALDI can result in loss of sialic acid, sulphate and phosphate residues from native glycoconjugates as they can dissociate during the MALDI process (Zaia 2010). Although sample preparation is laborious, ESI produces better resolved peaks for glycoconjugates than MALDI due to the absence of matrix adduct peaks and ESI-MS is well suited to be used for analysis of liquid chromatographic effluents (Satterfield et al. 2005; Zaia 2008).

1.4.4.1 Tandem mass spectrometry

In order to obtain a detailed characterisation of glycan structure, it is necessary to induce fragment ions by employing tandem mass spectrometry (MS^n). The nomenclature for glycan fragmentation was introduced by Domon and Costello, (1988) (**Figure 4**) (Domon 1988). Fragment ions from tandem MS that include the reducing end of the glycan are termed X, Y and Z, and the fragments that contain the non-reducing end of the glycans are termed A, B and C. Subscript numbers represent the location of the fragment from the reducing end whereas superscript numbers show the bond that is broken in the cross ring fragment (Domon 1988). Collision induced dissociation (CID) is a widely used fragmentation method to obtain MS/MS spectra of glycan analytes. CID fragmentation induces two types of bond cleavages; 1) glycosidic bond cleavage resulting from cleavage between two neighbouring residues and, 2) cross-ring cleavage resulting from fragmentation of a sugar ring. The glycosidic bond cleavage to monosaccharide masses on MS/MS spectra gives composition and oligosaccharide sequence information (Han et al. 2013).

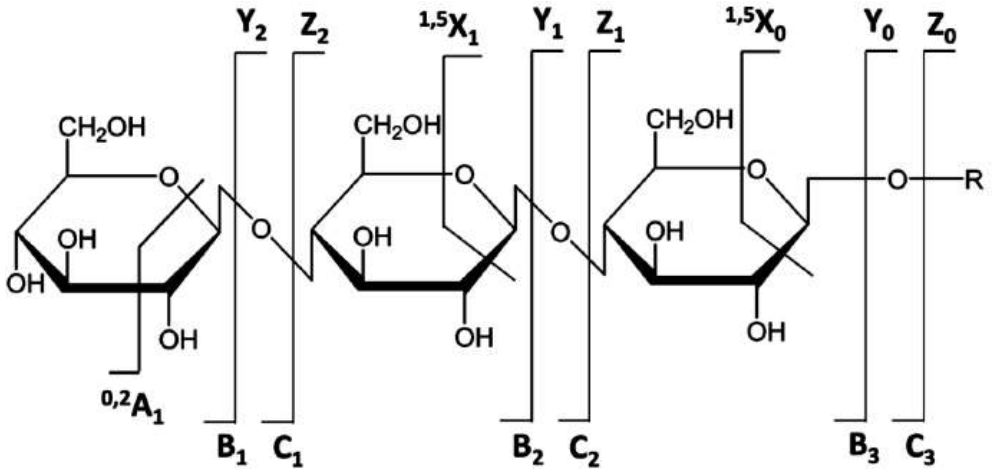


Figure 4: Nomenclature for glycan fragments observed in MS^n spectra. Figure is adapted from (Domon 1988). Fragment ions from tandem MS that include the reducing end of the glycan are termed X, Y and Z, and the fragments that contain the non-reducing end of the glycans are termed A, B and C. Subscript numbers represent the location of the fragment from the reducing end whereas superscript numbers show the bond that is broken in the cross ring fragmentation.

Higher energy collisional (C-trap) dissociation (HCD) fragmentation is another fragmentation technique extremely useful in generating information on the peptide backbone carrying the glycosylation site by producing a series of B- and Y- ions. A combination of CID/HCD enables the identification of glycan structure and the peptide backbone, allowing glycopeptide identification (Scott et al. 2011). Lower energy CID fragmentation predominantly cleaves the glycosidic linkage of the glycopeptides generating B/C- and Y/Z- ions, yielding information regarding the monosaccharide composition and general topology of the glycan (Gaut et al. 2001) whereas, higher energy HCD fragmentation generates abundant low-mass oxonium ions (*m*/*z* 204, *m*/*z* 366) and B/C ions from glycosidic cleavage. HCD also provides a series of b/y ions from peptide backbone without the glycan moiety and thus the combination of CID/HCD is more useful in glycopeptide identification (Brinkmann et al. 2004; Roy et al. 2014). Electron Transfer Dissociation (ETD) is another fragmentation technique, in which ETD induces cleavage of the N-Cα bonds of the peptide backbone to generate a series of c- and z- type fragment ions. Unlike CID, ETD preserves post translational modification and the amino acids (which are often labile upon CID), whilst conserving sequence information for peptide identification (Wuhrer et al. 2007; Wiesner et al. 2008).

To summarise, *N*-linked glycosylation is a common protein post-translational modification in eukaryotes and plays a biologically important role in various diseases. Glycoprofiling of any given biological sample involves characterising the glycan structures (glycomics) and the site-specific protein information (glycoproteomics). Analysis of *N*-linked glycans typically involves enzymatic cleavage of the glycans from proteins, followed by separation of glycans by various chromatographic techniques and structural determination by MS and MS/MS. Glycoproteome analysis involves enrichment of glycoproteins and/ or glycopeptides, followed by LC separation and detection and identification by MS MS/MS. Mapping the site-specific glycoproteome of a complex mixture is still technically difficult but will add to understanding of the role of glycosylation in CF.

1.5 Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis

Compared with *N*-glycosylation, host mucin *O*-glycosylation is the most widely studied protein glycosylation in CF and its role in pathogenesis is well explained. This section involves our published review article describing the aberrant mucin *O*-linked glycosylation observed in CF airways compared to the non-CF counterpart. The adhesion mechanisms employed by *P. aeruginosa* in utilising the host sugar epitopes on the mucins is also covered in this review.

Pages 25-48 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages.

Venkatakrishnan, V., Packer, N. H., & Thaysen-Andersen, M. (2013). Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis. *Expert Review of Respiratory Medicine*, *7*(5), 553-576.

DOI: 10.1586/17476348.2013.837752

1.6 Project aims

The wealth of knowledge available on host mucin glycosylation in CF is extensive, since over the past three decades research on sputum glycosylation mainly focussed on mucins and their role in CF pathogenesis. Although sputum is reported to contain many other glycoproteins, sputum *N*-glycosylation has not been characterised leaving a gap in the field of CF glycosylation research. Whilst abundant knowledge regarding mucin *O*-glycosylation is available in the literature, this study investigates structural *N*-glycan and site-specific N-glycoproteome information based on the hypothesis that *N*-glycans contribute to the complex cellular and molecular environment of the CF affected respiratory tract. The broad aim of the work presented in this thesis was to investigate the contribution of bacterial colonisation and genetic mutation to the CF *N*-glycosylation phenotype at both the glycomic and glycoproteomic level.

Specific aims were

1) To characterise and identify the variation in CF sputum *N*-glycans compared to those in non-CF sputum **(Chapter 2)**.
2) To assess if there was any direct link between bacterial colonisation and variation in CF sputum *N*-glycosylation **(Chapter 3)**.
3) To identify the variation in proteins between CF and non-CF sputum and to perform the first site-specific *N*-glycoproteome study on CF sputum **(Chapter 4)**.

Chapter 2

Cystic fibrosis and bacterial colonisation define the sputum *N*-glycosylation phenotype

This chapter focusses on characterising the sputum protein *N*-glycosylation of CF sputum and compares that with non-CF sputum obtained from patients with other lung conditions

Venkatakrishnan V, Thaysen-Andersen-M, Chen SCA, Nevalainen H, Packer NH (2014). Cystic fibrosis and bacterial colonisation define the sputum *N*-glycosylation phenotype, *Glycobiology*. (Accepted)

Contribution: In this publication, all glycan-based experiments analysing sputum samples, MS-structural glycan interpretation and manuscript preparation was performed by me.

Glycobiology Advance Access published September 4, 2014

Cystic fibrosis and bacterial colonisation define the sputum *N*-glycosylation phenotype

Vignesh Venkatakrishnan[1], Morten Thaysen-Andersen[1], Sharon C. A. Chen[2], Helena Nevalainen[1], and Nicolle H. Packer[1]

[1]Biomolecular Frontiers Research Centre, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia, 2109

[2]Centre for Infectious Diseases and Microbiology Laboratory Services, ICPMR-Pathology West, Westmead Hospital, Sydney, Australia 2006

Corresponding author: Prof. Nicolle H. Packer, Dept. of Chemistry and Biomolecular Sciences, Macquarie University, 2109 - North Ryde – Australia, Email: nicki.packer@mq.edu.au, Phone: (+61) 2 9850 8176, Fax: (+61) 2 9850 8313

Key words

Bacterial colonisation/ cystic fibrosis/ *N*-glycans/ paucimannose/ sputum

This is a pre-copyedited, author-produced version of an article accepted for publication in Glycobiology following peer review. The version of record Venkatakrishnan, V., Thaysen-Andersen, M., Chen, S. C. A., Nevalainen, H., & Packer, N. H. (2015). Cystic fibrosis and bacterial colonization define the sputum N-glycosylation phenotype. Glycobiology, 25(1), 88-100 is available online at: https://doi.org/10.1093/glycob/cwu092

Abstract

Although mucin *O*-glycosylation of sputum from individuals suffering from cystic fibrosis (CF) is known to be altered relative to their unaffected counterparts, protein *N*-glycosylation of CF sputum remains structurally and functionally under-characterised. We report the first *N*-glycome of soluble proteins in sputum derived from five CF patients, two pathogen-free and two pathogen-infected/colonised non-CF individuals suffering from other pulmonary conditions. *N*-glycans were profiled using porous graphitised carbon-liquid chromatography-negative ion-tandem mass spectrometry following enzymatic release from sputum proteins. The composition, topology and linkage isomers of 68 *N*-glycans were characterised and relatively quantified. Recurring structural features in all sputum *N*-glycomes were terminal α2,6-sialylation, α1,6-core fucosylation, β1,4-bisecting GlcNAcylation and compositions indicating paucimannosylation. Despite covering different genotypes, age, gender and microbial flora, the sputum *N*-glycomes showed little inter-patient and longitudinal variation within CF patients. Quantitative *N*-glycome inter-patient group comparisons revealed that lung infection/colonisation, in general, extensively enriches the CF sputum *N*-glycosylation phenotype with paucimannose with simultaneous over-sialylation/fucosylation and under-bisecting GlcNAcylation of complex/hybrid *N*-glycans. In contrast, the sputum from CF patients had only slightly increased abundance of paucimannose *N*-glycans relative to pathogen-infected/colonised non-CF individuals. Similar to mucin *O*-glycosylation, protein *N*-glycosylation appears to be regulated primarily as a secondary effect of bacterial infection and inflammation rather than the CF pathogenesis in sputum. This study provides new structural *N*-glycan information to help understand the complex cellular and molecular environment of the CF affected respiratory tract.

Introduction

Cystic Fibrosis (CF) is a prevalent autosomal recessive disorder affecting Caucasians more frequently than other populations (Heijerman 2005). CF is caused by a mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene located on the long arm of chromosome 7. Many mutations in the CFTR gene result in absence or malfunction of the cell surface CFTR protein, which functions as an important chloride and bicarbonate ion transporter in healthy individuals (Rommens et al. 1989; Quinton 2001; Reddy et al. 2001). Lack of functional CFTR leads, in turn, to a dysfunctional regulation of salt and water across epithelial cells of all affected tissues, in particular the exocrine glands and most dramatically, the respiratory tract (Mickle et al. 2000; Boucher 2007).

An important characteristic of CF pathogenesis is the build-up of mucous in the airways due to hyper-secretion and/or lower-than-normal mucous clearance. The issues are amplified further by the high viscosity of the mucous secretion caused by the salt and water imbalance across the epithelium, further reducing the mucocilliary clearance from the lungs (Voynow et al. 2009). This viscous mucous secretion, termed sputum when expectorated, facilitates increased propensity for adherence and colonisation of a range of bacterial and fungal pathogens which, as a consequence, infect CF individuals with higher prevalence. *Pseudomonas aeruginosa* is by far the most prevalent bacteria in CF whereas *Aspergillus fumigatus* is the major infecting fungi (Gilligan 1991; Lipuma 2010). In addition, the pathogenesis of CF is characterised by chronic lung inflammation, which together with the increased frequency of pathogen infections form the hallmarks of the CF lung disease related mortality (Gibson et al. 2003).

Sputum consists of many components such as lipids, carbohydrates, nucleic acids, pus, salts water, inflammatory cells, cellular debris and mucin and non-mucin proteins, all of which contribute to its highly complex biophysical and biochemical properties (Matthews et al. 1963). Mucins, which by definition are of high molecular mass and heavily glycosylated in defined tandem repeat regions, constitute the largest component of the proteins in sputum (Schulz et al. 2005). Airway mucins, which have multiple cellular origins and are either secreted (gel-forming) or membrane-bound to epithelial cell surfaces of the respiratory tract, are the main carriers of *O*-glycans in human lungs (Venkatakrishnan et al. 2013). A tremendous effort has been undertaken over the past three decades to document the structural alterations in *O*-linked glycosylation in CF with the aim of understanding their role in the disease (Venkatakrishnan et al. 2013). Although vastly simplified, the *O*-glycosylation alterations in CF can be summarised as follows: *Secreted airway mucins* have a higher degree of sialylation and sulfation in CF and a lower total *O*-glycosylation per protein weight (Davril et al. 1999; Xia et al. 2005; Schulz et al. 2007), whereas *membrane-bound airway mucins* have a higher degree of fucosylation and a lower degree of sialylation on their *O*-glycans in CF (Kube et al. 2001; Rhim et al. 2001) as compared to the non-CF (i.e. chronic bronchitis or bronchiectasis) equivalents. The unique sputum *O*-glycosylation phenotype, which is regulated primarily by inflammation rather than directly by the CF pathogenesis (Holmen et al. 2004; Schulz et al. 2005), enhances the capacity of adhesion and colonisation of bacterial pathogen(s) to the host airways (Carnoy et al. 1993; Carnoy et al. 1994; Lamblin et al. 2001; Venkatakrishnan et al. 2013). Surprisingly, despite its high abundance in most human tissues and bodily secretions including sputum, little has been reported on the *N*-glycosylation of sputum proteins and its possible regulation and functional impact in CF.

As the first, we have accurately profiled the total *N*-glycome derived from soluble sputum proteins from five CF patients (CF1-5), two pathogen-infected/colonised non-CF patients (iNCF1-2) with clinical upper respiratory tract infection (URTI) and two non-CF patients (NCF1-2) diagnosed with other lung diseases, but from whom no pathogens were isolated from sputum. This study was undertaken to investigate *in vivo* the importance of the absence of functional CFTR and other associated CF traits, in particular lung infection, on the sputum protein *N*-glycosylation. The heterogeneous sputum *N*-glycome profiles appeared uniform within all three patient groups, and importantly showed unique signatures upon inter-patient group comparisons. Our data confirms that both malfunctioning CFTR protein and the bacterial infection of the respiratory tract are factors affecting the sputum protein *N*-glycosylation in the complex molecular and cellular environment of the CF lung.

Results

Rationale and experimental design

The individuals included in this study covered both genders and a relative large age span i.e. age 19-69 (see **Table I** for patient specific information). The range of on-going antibiotic treatments of the CF1-5 individuals at the time of sample collection was targeted at the different microbial flora detected in their sputum. Four of the five CF patients were colonised with *P. aeruginosa* whereas *S. aureus* and *A. fumigatus* were detected in sputum of two patients. Comparing the sputum *N*-glycomes between CF, bacterial infected/ colonised non-CF (iNCF) and non-infected (NCF) non-CF patients enabled us to determine the effect

on the *N*-glycosylation phenotype of malfunctioning CFTR separately from the infection status. Hence, this study primarily focuses on comparisons between these three patient groups.

Secreted sputum proteins were isolated from whole sputum by lowering the sputum viscosity with subsequent elimination of intact cells, cellular debris and insoluble mucins by centrifugation, ensuring high recovery of the soluble *N*-glycoproteome. Using the same *N*-glycome analysis procedure, no significant *N*-glycans were detected in the discarded pellets (data not shown). The analysed soluble fraction can, as such, be considered as representative of the total sputum *N*-glycome.

Porous graphitised carbon (PGC)-LC-negative ion-MS/MS was used for detailed structural characterisation and relative quantitation of the reduced *N*-glycans following enzymatic release from their carrier sputum proteins. Analysis was performed in technical (LC-MS/MS) duplicates. In this approach, *N*-glycan structures and their sub-structural features can be characterised with high confidence based on their molecular mass, MS/MS fragmentation pattern and their relative and absolute PGC retention time including their monosaccharide composition and overall topology (Pabst et al. 2007; Jensen et al. 2012). In addition, partial information can be obtained on the linkage connectivity of the monosaccharide residues. Finally, the relative molar abundances of the individual *N*-glycans can be assessed with good accuracy using this method (Leymarie et al. 2013).

Overview of sputum N-glycans

In total, 56 *N*-glycan monosaccharide compositions were identified from the five CF and four non-CF LC-MS/MS datasets. From these monosaccharide compositions, a total of 105 *N*-

glycan structures with topology and linkage isomer variations were tentatively identified. Using monoisotopic molecular mass, the presence of high quality MS/MS fragmentation mass spectra and PGC retention time as minimum requirements for confident *N*-glycan structure identification, the data was filtered to identify a total of 68 well-defined *N*-glycan structures, **Table II**. The entire set of validating MS/MS spectra can be found in an annotated form in **Supplementary Figure S1** and has been entered into the UniCarb-DB MS/MS glycan database (http://www.unicarb-db.org/unicarb/show_data.action?id=618).

Some structural ambiguity, in particular for the connectivity of terminal monosaccharides including Lewis$^{a/c}$ and Lewis$^{b/y}$ type fucosylation, remains in some structures. In addition, some linkages of the chitobiose core and the initial arm extension were assumed from the established knowledge of the biosynthetic machinery i.e. GlcNAc-Man linkages on the 3' and 6' arm were assumed to be β1,2-linked for mono- and bi-antennary structures, whereas some linkage ambiguity was left for the third GlcNAc residue of tri-antennary structures (β1,4/6). The bisecting GlcNAc-Man linkage was assumed to be of the β1,4-type. The core fucosylation was assumed to be α1,6-linked. Although the Gal-GlcNAc linkage of the *N*-glycan arms can be β1,3 in humans, it is generally of the β1,4-linkage type, and hence illustrated as such. If required for later functional definition, further work will need to define the β1,3/4 Gal-GlcNAc distribution and validate the other assumed linkages.

For the characterised 68 *N*-glycans, a quantitative analysis was carried out to establish the relative molar abundances of the observed species. The lower abundant *N*-glycans, with predicted monosaccharide composition but without MS/MS confirmation data, constituted 2-3 % (mol/mol) of the entire pool of *N*-glycans and were omitted for quantitation (see **Supplementary Table S1** for the omitted compositions). In addition, we observed +80 Da

mass modifications on a very limited subset of *N*-glycans of the high mannose type (Man_{8}-Man_{7}). These, presumably phosphorylated *N*-glycans, corresponding to approximately 3% (mol/mol) of the entire set of sputum *N*-glycans, were also excluded in the quantitation due to the lack of solid structural evidence.

All *N*-glycan types were observed in the *N*-glycan profiles including high mannose, hybrid, complex and the less reported paucimannose type *N*-glycans. Compositions indicating paucimannose type structures dominated the CF and bacterial infected/ colonised non-CF (iNCF) sputum profiles whereas bisecting GlcNAcylation was an abundant feature of non-infected (NCF) sputum. Common structural features included α2,6-linked sialylation and α1,6-linked core fucosylation. The relative abundances of each of the *N*-glycans and their structural groups and sub-groups were calculated and inter-patient group comparisons were made to identify *N*-glycome similarities and differences.

No significant sputum-saliva cross-contamination

Sputum sampling is performed via non-invasive expectoration from patients where saliva cross-contamination potentially can be an issue. Saliva is rich in *N*-glycoproteins. To rule out significant salivary contamination in the sputum, the obtained *N*-glycomes were compared to our recently published salivary *N*-glycosylation profile (Everest-Dass et al. 2012). Of the top 10 most abundant salivary *N*-glycans, six of them, containing terminal Lewis-type determinants, were not observed in the sputum *N*-glycomes and the other four were only observed at trace levels i.e. 0.5-4% Hence, salivary contamination of the analysed sputum samples was evaluated to be negligible.

Limited intra-patient group sputum N-glycome variation

Figure 1 presents the average mass profiles of the reduced *N*-glycans released from the sputum proteins of five individuals with CF (CF1-5). The *m/z* and # numbers on the mass spectra correspond to *N*-glycans described in **Table II**, **Supplementary Table S1** and **Supplementary Figure S1**. The striking qualitative observation is that the *N*-glycans in the sputum from different individuals with CF (CF1-5) show a very high degree of similarity on the basis of compositional mass. This qualitative similarity was later confirmed at the quantitative and structural level (see next section). In addition, there was surprisingly little longitudinal variation in the *N*-glycome of sputum obtained from the same CF patient (CF5) in three samples collected over a period of 6 months. Uniform *N*-glycomes were also observed within the non-CF patient groups (data not shown) although the limited number of patients in these groups made a proper evaluation difficult.

Comparison of N-glycome of CF and infected non-CF (iNCF) sputum

The detailed structural *N*-glycomes of sputum proteins, as described in **Table II**, of CF patients (all of whom were colonised with pathogens) were compared with pathogen infected non-CF (iNCF) patient sputum to investigate whether the presence of the CFTR mutation has an effect on infected sputum protein *N*-glycosylation. Initially, the *N*-glycans were grouped and compared according to their types i.e. high mannose, hybrid, complex and paucimannose, **Figure 2**. As expected from their mass profiles (**Figure 1**), limited variation was observed across the patients in the CF (1-5) group showing dominating complex (42.8% ± 5.7%) and paucimannose (43.4% ± 4.1%) type *N*-glycans. High mannose

type *N*-glycans were of comparatively lower abundance (12.5% ± 3.2%) and hybrid type structures were barely present (1.2% ± 0.6%) in the total *N*-glycan pool. Although also a dominant structure, comparatively less paucimannose type *N*-glycans was observed in the sputum of the infected non-CF (iNCF) patient group (27.8% ± 4.4%, p=0.0027, n=14) compared to the CF group. A concomitant higher presence of complex type *N*-glycans were observed in iNCF (56.4% ± 1.8%, p=0.02, n=14) compared to CF sputum. The amount of *N*-glycans of the high mannose type (14.1% ± 6.3%, p=0.73, n=14) and hybrid type (1.7% ± 1.6 %, p=0.51, n=14) in iNCF sputum remained statistically similar to CF sputum, **Figure 2**.

The CF and iNCF sputum *N*-glycomes were investigated to identify further differences on the sub-structure level in the different glycan types. The distributions of the high mannose type *N*-glycans (Man_5-Man_9; Glycan #8a/b, #9, #10a/b, #11a/b, #12, respectively) in CF and iNCF sputum were generally similar, **Figure 3A**. The CF (Man_5: 21.2% ± 6.7%, Man_6: 49.1% ± 12.0%, Man_7: 14.3% ± 1.8%, Man_8: 8.2% ± 3.8% and Man_9: 7.8% ± 3% (mol/mol) of the total high mannose type *N*-glycans) and iNCF (Man_5: 13.3% ± 1.4%, Man_6: 49.1% ± 2.3%, Man_7: 17.0% ± 2.1%, Man_8: 6.6% ± 0.4% and Man_9: 14.1% ± 0.8%) sputum *N*-glycomes were both dominated by Man_6 in the distribution of high mannose structures.

The degree of total sialylation of sputum *N*-glycans, which was calculated as the abundance-weighted molar ratio of sialylated *N*-glycans irrespective of the NeuAc linkage of the total pool of sialylated *N*-glycans (complex/hybrid type), was also similar between CF (75.2% ± 4.7%) and iNCF (72.6% ± 4.9%, p=0.39, n=14), **Figure 3B**. α2,6-linked sialylation was the predominant NeuAc linkage type in both CF (83.3% ± 11.9%) and iNCF (83.8% ± 14.8%) sputum and no significant differences in the sialic acid subtypes were observed between the two patient groups, **Table III**.

The total fucosylation of the sputum *N*-glycome, which was calculated as the abundance-weighted molar ratio of all fucosylated *N*-glycans irrespective of their fucose linkage of the total pool of potentially fucosylated *N*-glycans (complex/hybrid/paucimannose type *N*-glycans), was also statistically similar between the CF (87.0% ± 4.4%) and iNCF (71.2% ± 14.9%, p=0.12, n=14), **Figure 3C**. Structures containing only α1,6-linked core fucosylation were predominant for both CF (67.8% ± 6.4%) and iNCF (68.0% ± 12.5%) sputum. *N*-glycans containing only Lewis-type fucosylation (predominantly Lewis$^{a/x}$) were of lower abundance than the core fucosylated *N*-glycans in all CF (27.9% ± 5.1%) and iNCF (25.8% ± 9.0%) samples, but yet were higher than structures containing both core and arm fucosylation (CF: 4.2% ± 2.7% and iNCF: 6.2% ± 3.8%). No significant differences in the linkage specific fucosylation were observed between CF and iNCF, **Table III**.

The abundance of bisecting β1,4-GlcNAc containing *N*-glycans, which was calculated as the -weighted molar ratio of bisecting GlcNAcylated *N*-glycans of the total pool of *N*-glycans (complex/hybrid type), was found to be relatively low and statistically similar between the CF (7.6% ± 5.0%) and iNCF (16.7% ± 11.7%, p= 0.22, n= 14) patient groups **Figure 3D**.

None of the six individual paucimannose *N*-glycans ($GlcNAc_2Man_3Fuc_1$ (M0F): glycan #1, $GlcNAc_2Man_1Fuc_1$ (M1F): glycan #2, $GlcNAc_2Man_2Fuc_0$ (M2): glycan #3, $GlcNAc_2Man_2Fuc_1$ (M2F): glycan #4, $GlcNAc_2Man_3Fuc_0$ (M3): glycan #5, $GlcNAc_2Man_3Fuc_1$ (M3F): glycan #6) were significantly altered in CF (0.8% ± 0.4%, 3.7% ± 1.4%, 2.1% ± 0.5%, 27.1% ± 4.4%, 2.3% ± 0.9% and 6.3% ± 2.2%) when compared to iNCF (0.7% ± 0.3%, 2.7% ± 3.0%, 1.7% ± 0.7%, 17.7% ± 7.2%, 1.6% ± 0.3%, 2.2% ± 0.8%). The dominance of low mass signals corresponding to compositions indicating paucimannose *N*-glycan (e.g. M2F, #4) in all CF sputum samples can be appreciated in the mass profiles in **Figure 1**.

Of the individual structures, eight *N*-glycans (glycan# 6, #7b, #16, #19, #20, #23b, #25a, #33a) were significantly more abundant, and three *N*-glycans (glycan# 27a, #39, #44) were of lower abundance, in CF relative to iNCF sputum. Some degree of relatedness was noted amongst these *N*-glycans i.e. truncated and mono-antennary *N*-glycans were found over-represented and bisecting GlcNAc containing *N*-glycans under-represented in CF.

Comparison of the N-glycomes of infected (iNCF) and pathogen-free (NCF) non-CF sputum

To understand the impact of infection alone on the *N*-glycosylation of sputum proteins, infected (iNCF) and non-infected (NCF) non-CF sputum *N*-glycomes were compared. The most striking difference was seen in the ten-fold lower molar abundance of compositions corresponding to paucimannose type *N*-glycans in NCF (3.3 ± 1.3%, $p=0.97 \times 10^{-4}$, n=8) compared to iNCF, **Figure 2**. As a consequence, significant increases in the relative molar abundance of complex type *N*-glycans (81.9% ± 1.8%, p=0.00050, n=8) and hybrid type *N*-glycans (4.2% ± 0.2%, p=0.021, n=8) were observed in the NCF sputum. The total amount of high mannose type *N*-glycans remained statistically similar in NCF (10.6% ± 0.2%, p=0.42, n=8) when compared to the iNCF group.

Non-infected (NCF) was found to have a different distribution of high mannose *N*-glycans (Man_5-Man_9) when compared to infected (iNCF) sputum (Man_5: 18.5% ± 16.2%, Man_6: 22.0% ± 3.8%, Man_7: 21.9% ± 4.2%, Man_8: 19.9% ± 5.1% and Man_9: 17.7% ± 10.6%), **Figure 3A**. Man_5 was significantly reduced (p=0.0025, n=14) and was no longer the dominating high mannose type species observed in the infected patient groups (both CF and non-CF).

The degree of total sialylation of sputum *N*-glycans was significantly reduced in NCF (44.3% ± 6.5%, p=0.00043, n=8) when compared to iNCF, **Figure 3B**. Total fucosylation was

statistically similar in NCF (75.2% ± 7.5%, p=0.65, n=8) when compared with iNCF, **Figure 3C**. Also, the breakdown of the total fucosylation to core and Lewis-type fucosylation is represented in **Table III** and is similar between the two non-CF groups.

Of the increased percentage of complex and hybrid structures in the iNCF group, bisecting GlcNAcylated *N*-glycans were more highly represented in uninfected NCF (48.5% ± 5.9%, p=0.0029, n=8), **Figure 3D**. This equates to a three-fold higher abundance of β1,4-linked bisecting GlcNAc containing structures in non-infected (NCF) compared to infected (iNCF) non-CF sputum.

Cluster analysis of sputum N-glycomes of the three patient groups

The relatedness between the sputum *N*-glycome profiles was further assessed by a cluster dendrogram using the relative molar abundances of the *N*-glycans, **Figure 4**. The limited intra-group *N*-glycome variations were illustrated by tight clustering of the *N*-glycomes within the patient groups. Interestingly, the infected iNCF1-2 group clustered more towards the CF group whereas the non-infected NCF1-2 clustered separately. This supports the previous observation suggesting that pathogen infection and possibly the associated inflammation plays the most crucial role in defining/regulating the *N*-glycosylation phenotype in CF sputum. In contrast, the actual CFTR genotype seems to influence the sputum *N*-glycosylation less than the presence of infection.

Paucimannose N-glycans are not bacterial degradation product

The possibility of degradation of the sputum *N*-glycans by the infecting bacteria such as the production of glycosidases by *P. aeruginosa* was investigated by a bacterial challenge assay where non-infected non-CF sputum (NCF) was incubated for 3 days at 37°C with pathogenic

P. aeruginosa isolated from a CF patient. Although the paucimannose *N*-glycans were increased after bacterial challenge, the sialo:asialo ratio of *N*-glycans as assessed by the *N*-glycan profiling method described in this study was found to be unchanged after incubation relative to the ratio before bacterial challenge (data not shown). Since removal of any monosaccharide would be the first step in degradation and would be expected to change the ratio of sialylated structures in the NCF sputum, this indicated that degradation of *N*-glycans by the infecting/colonising bacteria may not be the basis for the over-representation of the paucimannose in the CF and iNCF patients.

Also, from the data shown in this study it can be noted that the relative abundances of the high mannose *N*-glycans were not altered between CF, bacterial infected/ colonised (iNCF) non-CF and pathogen-free (NCF) non-CF sputum, which by itself indicates paucimannose is not a degraded product of high mannose *N*-glycans. Furthermore, there was no alteration in the ratio of core fucosylation before and after challenging pathogen-free NCF sputum with *P. aeruginosa*, supporting further that the paucimannose *N*-glycans are not bacterial degradation product.

Discussion

Mapping the biomolecular environment including the protein glycosylation of the respiratory tract and defining the molecular alterations that are associated with CF is a prerequisite for a better holistic understanding of the disease. Over the past three decades, protein (mucin) *O*-glycosylation has been studied in the context of CF to understand the role of the increased mucin *O*-glycans in the pathogenesis (Venkatakrishnan et al. 2013). We have

now characterised the *N*-glycosylation of secreted sputum proteins from CF patients and compared these with sputum *N*-glycomes of non-CF patients. Due to the difficulty in obtaining sputum non-invasively from healthy individuals, sputum from four non-CF individuals with other lung conditions were used.

Our aim in this work was partly to map the sputum *N*-glycosylation from cystic fibrosis and other pulmonary diseases since such information is currently unavailable in the literature and partly to investigate a mechanism of how *N*-glycosylation is regulated in CF disease. Although the number of patient samples used is relatively low (five CF vs four non-CF), we observe surprisingly little patient-to-patient variation within all three patient groups, which indicates that a relative low number of biological replicates can be used to establish an "average" *N*-glycome profile of sputum proteins.

Within the group of five CF patients there were variations in terms of age, gender, microbial population infecting the lung, antibiotic treatment and the severity of the disease (genotype dependent). In addition, the collected samples varied in terms of sputum colour and viscosity which could be due to the differential expression of alginate or the variable status of disease progression and infection (Gibson et al. 2003). However, counter-intuitive to these many patient/sample variables, the *N*-glycome profiles of the sputum proteins across the CF patients remained qualitatively and quantitatively remarkably uniform. In comparison, mucin *O*-glycosylation derived from sputum of CF patients has been previously shown to have large inter-individual variations, in particular in the terminal glycan determinants, which correlated with the severity of airway infection (Davril et al. 1999). Unlike aberrant mucin *O*-glycosylation, which is reported to be the effect of environmental factors inflammation (Holmen et al. 2004; Schulz et al. 2005), this study has suggested that the

altered global sputum *N*-glycosylation which is carried by many non-mucin proteins appears to be a direct effect of bacterial infection.

The single most abundant *N*-glycan (27.1% ± 4.4%) in CF sputum (*m/z* 895.4, glycan #4) appeared to be of the paucimannose type which is a class of *N*-glycans defined as the naked chitobiose core structures of the general monosaccharide composition $GlcNAc_2Man_{0-3}Fuc_{0-1}$ (Stanley et al. 2009). Paucimannose *N*-glycans are observed in high abundance in plants and invertebrates, but have rarely been reported in vertebrates (Altmann et al. 2001; Natsuka et al. 2002; Dam et al. 2013) as they are not a component of the known *N*-glycan biosynthetic pathway. Recent studies have observed the presence of paucimannose in cancer tissues, in kidney of a systemic lupus erythematosus mouse model and in saliva, albeit at much lower relative molar abundances (i.e. 5-10% of the total *N*-glycan pool) (Hashii et al. 2009; Balog et al. 2012; Everest-Dass et al. 2012) than seen in the infected sputum *N*-glycomes (i.e. CF and iNCF) in this study. The sputum of the infected CF and infected non-CF patients consistently carried more than a 30-40% relative molar abundance of paucimannose *N*-glycans, a ten-fold higher abundance than uninfected non-CF (NCF) sputum. All the types of micro-organisms hosted by the CF and iNCF patients (*P. aeruginosa*, *A. fumigatus*, *S. aureus* and *S. pneumoniae*) correlated with an increased relative abundance of paucimannose *N*-glycans, indicating that this signature was consequently not specific towards only *P. aeruginosa* infections. This suggests strongly that it is pathogen infection in general, and possibly the resulting inflammation, that plays a very important role in regulating the *N*-glycan phenotype in sputum, towards paucimannose-rich profiles, in both infected CF and non-CF lung pathologies. In support of this, a recent finding linked the appearance of paucimannose *N*-glycans to inflammation in human cancer tissues and mouse pancreatitis using a monoclonal

antibody called Mannitou compared to normal tissues or cells which exhibited weak or no staining (Zipser et al. 2012). We are currently validating the paucimannose structures further, mapping their determining their protein carriers to specify the molecular and cellular origin(s) and investigating the functional relationship between bacterial infection/inflammation of the respiratory tract and the high relative abundance of these truncated *N*-glycans.

Although the total proportion of high mannose type *N*-glycans of the total *N*-glycan pool were shown to be statistically unchanged between all three patient groups, the distribution of the individual high mannose type *N*-glycans differed. Specifically, Man_5 was significantly over-represented in the sputum of the two pathogen infected patient groups (i.e. CF and iNCF) compared to pathogen-free NCF sputum profiles. We could hypothesise that the possible importance of the higher abundance of Man_5 in the context of infection may be in its role of binding to mannan-binding lectin (MBL), a crucial carbohydrate binding protein in the complement pathway used by the host as a defence against infecting pathogens (Kilpatrick 2003; Fidler et al. 2009). Earlier studies have reported the binding affinity of MBL to oligomannosyl structures (Lee et al. 1991; Arnold et al. 2005). Also, MBL has been found to bind to IgM due to its high specificity for mannose which is found to be rich in the carbohydrate regions of IgM (Koppel et al. 2001; Arnold et al. 2005).

Sialic acids are terminal monosaccharide residues of glycans, which are of particular biological interest in a CF context as they have been shown to be involved in bacterial adhesion and cell-cell/cell-protein interactions (Ramphal et al. 1991; Plotkowski et al. 1993). The *O*-glycan structures on sputum proteins, as well as on the purified mucins (MUC5AC and MUC5B), are over-sialylated in CF compared to non-CF counterparts and correlate with

CF pathogenesis (Frates et al. 1983; Davril et al. 1999; Schulz et al. 2007). However, since the mucins generally seem to be under-glycosylated (under-occupied) in CF, the total sialylation of sputum proteins remains overall unchanged (Schulz et al. 2007). The connection between sialylation and infection was supported by our observation of an increase in sialylation per mole of (complex and hybrid) *N*-glycans in both infected CF and non-CF sputum compared to the pathogen-free non-CF sputum.

P. aeruginosa is known to have a sialic acid trimming enzyme, neuraminidase, which the bacterium is thought to use to uncover previously inaccessible galactose determinants for binding/adherence to the host with their well-characterised galactose-binding lectin, PA-IL (Avichezer et al. 1992; Chen et al. 1998). The molar over-sialylation of *N*- and *O*-glycosylation of proteins derived from *P. aeruginosa* infected CF and iNCF sputum indicates that *P. aeruginosa* does not appear to degrade the host *N*-glycoproteins to a great extent.

The degree of total fucosylated *N*-glycans was found to be statistically similar in CF, iNCF and the pathogen-free NCF. Unlike sputum *O*-glycosylation which shows a high abundance of terminal Lewis-type fucose containing determinants that play an important role in bacterial binding to the CF respiratory tract, such as via the *P. aeruginosa* fucose-binding lectin PA-IIL (Scharfman et al. 1999; Lillehoj et al. 2001; Xia et al. 2007), the *N*-glycans of sputum contained fucose predominantly in the α1,6-linked core linkage.

Another interesting *N*-glycan alteration was the significantly lower bisecting GlcNAcylation in the both infected patient groups (CF and iNCF) compared to pathogen-free NCF sputum. Synthesis of the β-1,4 bisecting GlcNAc directly on the mannose residue of the chitobiose core is catalysed by β-1,4-*N*-acetylglucosaminyltransferase III (GnT-III) expressed by the MGAT3 gene. Bisecting GlcNAc changes the structural conformation of the *N*-glycan

(Stanley 2002). The differences observed in bisecting GlcNAcylation may be explained by an infection-induced decrease in the enzyme level or activity of GnT-III, or alternatively over-expression of the competing glycosyltransferases (i.e. GnT-II, Gal-T1) for the *N*-glycan arm extension. Alternatively, the differences in bisecting GlcNAcylation may, in part, be explained by the presence of a different sputum proteome originating from cell types having other *N*-glycosylation synthetic machineries. In support of the latter, inflammatory cells including neutrophils and eosinophils which are highly abundant in both infected CF and non-CF lungs produce very little bisecting GlcNAc on their *N*-glycoproteins (Babu et al. 2009).

In depth LC-MS/MS based proteomics of the proteome isolated from the same CF and non-CF sputum samples is currently being carried out to identify which proteins carry the identified *N*-glycans and to determine whether the sputum proteomes are altered between the patient groups. This will yield insight into whether the *N*-glycome alterations observed in this study are driven by proteome alterations and/or by changes in the *N*-glycosylation machinery. In addition, we are currently performing large-scale *N*-glycoproteomics enabled by recent technology developments (Parker et al. 2013) to obtain direct evidence for the connectivity between the *N*-glycans and the individual sputum proteins. The sputum *N*-glycomes reported here can be expected to be carried by proteins derived from multiple origins from the complex cellular makeup of the respiratory tract including lung epithelial cells, submucosal glands, goblet cells and inflammatory cells and these changes to the *N*-glycosylation may only be biologically relevant at the individual glycoprotein level.

In conclusion, we are the first to report on the sputum *N*-glycome from CF and non-CF (pathogen-free and infected) patients. In total, 68 *N*-glycans were fully characterised and

their relative molar relationships described giving a detailed picture of the sputum *N*-glycomes. Comparison between patient groups revealed importantly that bacterial infection appeared to be the main effector of change in the protein *N*-glycosylation phenotype of CF patient lung sputum rather than the CF genotype. Infection, in general, produced paucimannose-rich *N*-glycomes in both CF and non-CF sputum which additionally were higher in complex-type sialylation and lower in bisecting GlcNAcylation than their pathogen-free sputum counterpart. The mechanism for the formation of the unusual paucimannose compositions appears to not be simply bacterial degradation and is being further investigated. This study provides much needed information on the complex cellular and molecular environment of the respiratory tract in a CF specific context.

Materials and methods

Sputum collection and patient information

Whole sputum samples (approximately 1 ml/patient) from five CF patients (called CF1-CF5), and four individuals who either had no underlying lung conditions or had other lung conditions (collectively called non-CF) were collected by non-invasive expectoration of sputum at Westmead Hospital, Sydney, Australia. The non-CF group was sub-divided according to the condition of their status: two individuals had clinical and radiological evidence of pneumonia (NCF1) or chronic obstructive pulmonary disease (COPD; NCF2) (collectively called NCF), but did not contain detectable pathogens in their sputum; and two pathogen-infected/colonised individuals (collectively called iNCF), who were diagnosed with an upper respiratory tract infection (URTI, iNCF1-2). Patient data including gender,

age, CF genotype, and microbiology culture results of sputum, were collected at the time of sputum sampling with informed consent of all patients (**Table I**). Sputum samples were stored at -20°C until used.

Isolation of soluble (secreted) sputum proteins

The sputum samples were thawed. To reduce the viscosity of the sputum, the protein disulfide bonds were reduced using 10 mM dithiothreitol (DTT) (final concentration) at 50°C for 1 hour and alkylated using 25 mM iodoacetamide (final concentration) at room temperature for 45 min in the dark. The samples were centrifuged for 30 min at 10,000 rpm to remove intact cells, cell debris and insoluble mucins. The supernatants containing the soluble secreted sputum proteins were transferred to new tubes and the protein concentrations were measured using Direct Detect protein quantitation system (Millipore). The pellets containing intact cells, cellular debris and insoluble mucins were found to be devoid of significant amounts of *N*-glycoproteins/*N*-glycans (data not shown) and, thus, were not integrated into the *N*-glycomics workflow and results.

Preparation and analysis of sputum N-glycans

N-glycans from sputum protein were released and processed essentially as described previously (Jensen et al. 2012). In brief, approximately 10 μg of soluble protein from each sputum sample was immobilised in separate spots on a PVDF membrane (Millipore). Proteins were stained using Direct Blue stain solution, excised and transferred to separate wells in a flat bottom polypropylene 96 well plate. *N*-glycans were released from the intact proteins using 5 U *N*-glycosidase F (PNGase F, *Flavobacterium meningosepticum*, Roche) in 10 μl water/well by overnight incubation at 37°C. The released *N*-glycans were reduced to

alditols by treatment with 0.5 M $NaBH_4$ in 50 mM KOH for 3 hours at 50°C. The reduction was quenched by the addition of 2 µl glacial acetic acid to the mixtures. The reduced *N*-glycans were desalted as the non-retained fraction of strong cation exchange resin packed as a micro-column on top of hydrophobic C18 material in solid phase extraction (SPE) formats, in order to remove cations and protein/peptide components. The *N*-glycans were further cleaned up with porous graphitised carbon (PGC) SPE to completely eliminate the presence of particulates and salts that could interfere with the downstream capillary LC-MS/MS. Glycans were eluted by 0.1% TFA (v/v) in 50% acetonitrile, dried by vacuum centrifugation and redissolved in 10 µl of water before transfer to glass vials (Waters) and LC-MS/MS analysis.

PGC-LC-MS/MS based N-glycomics

The global N-glycan profile of the sputum proteins was analysed in technical duplicates (two injections of the all sputum *N*-glycan samples in randomised order) using capillary LC-MS/MS on an Ultimate 3000 LC (Dionex) connected to an HCT 3-D ion trap (Bruker Daltonics) mass spectrometer. A HyperCarb PGC column (5 µm particle size, 180 µm internal diameter x 10 cm, Thermo Scientific) was used to separate the *N*-glycans using a constant flow rate of 2 µl/min. The column was equilibrated in 10 mM NH_4HCO_3 (aq) prior to sample injection (Solvent A). A 45 min linear gradient of 2-16% (v/v) of 10 mM NH_4HCO_3 in acetonitrile (solvent B) was applied, followed by a 20 min gradient from 16-45% solvent B before washing the column for 6 min at 45% solvent B and then re-equilibration in pure solvent A. ESI-MS was performed in negative ion polarity with the following scan events all acquired with a scan speed of 8,100 *m/z*/sec: MS full scan (*m/z* 100-2,200) and data-dependant

MS/MS scans after collision-induced dissociation (CID) of the top two most intense precursor ions with an absolute intensity threshold of 30,000 and relative intensity threshold of at least 5% intensity relative to the spectrum base peak. No dynamic exclusion was activated so that MS/MS triggering of closely eluting yet separated *N*-glycan isomers would occur. The following MS settings were used: drying gas temperature: 325°C, drying gas flow: 7 L/min, nebulizer gas: 18 psi, skimmer, trap drive and capillary exit were set at -40 V, -99.1 V and -166 V, respectively. Smart fragmentation was used with 30-200% (start-end) amplitude with 200 millisec maximum accumulation time and ion count target of 100,000 ions/scan event. The instrument was calibrated externally prior to acquisition. All LC-MS data was viewed and manually analysed using DataAnalysis 4.0 (Bruker Daltonics, Germany).

Determination of N-glycan structure and their relative molar distributions

The monosaccharide compositions and the individual *N*-glycan structures were manually determined using i) monoisotopic molecular mass, ii) MS/MS CID fragmentation profiles and iii) PGC LC retention behaviour. In order to confidently identify and characterise specific *N*-glycan structures high quality data for i)-iii) was considered a requirement. Only confidently characterised *N*-glycans were included in the relative quantitation (see below), the remaining were marked as 'trace/detectable'. Monoisotopic molecular masses, which generally were within 0.2 Da of the expected/theoretical monoisotopic masses were matched against possible/likely *N*-glycan monosaccharide compositions using GlycoMod (http://web.expasy.org/glycomod/). The proposed *N*-glycan monosaccharide compositions were confirmed manually by assignment of CID fragmentation MS/MS spectra and by qualitatively matching these spectra against assigned fragmentation spectra in Unicarb-DB

(http://unicarb-db.biomedicine.gu.se/) and other recently published *N*-glycome datasets (Parker et al. 2013; Sethi et al. 2013). In this process, GlycoWorkBench (http://www.glycoworkbench.org) was used iteratively to draw the proposed *N*-glycans and generate *in silico* glycosidic (B/Y- and C/Z-type) and cross-ring (A/X-type) glycan fragments to assist in the structural interpretation. In addition, substantial in-house knowledge of the absolute and relative PGC retention time of mammalian *N*-glycans, in particular for isobaric linkage-type *N*-glycan isomers, was used to cross-validate and to increase the confidence of the reported structures.

Some structural ambiguity, in particular for the connectivity of terminal monosaccharides including Lewis$^{a/x}$ and Lewis$^{b/y}$ type fucosylation, remains for some structures. In addition, some linkages of the chitobiose core and the initial arm extension were assumed from the established knowledge of the biosynthetic machinery i.e. GlcNAc-Man linkages on the 3' and 6' arm were assumed to be β1,2-linked for mono- and bi-antennary structures, whereas some linkage ambiguity was left for the third GlcNAc residue of tri-antennary structures (β1,4/6). The bisecting GlcNAc-Man linkage was assumed to be of the β1,4-type. The core fucosylation was assumed to be α1,6-linked. Although the Gal-GlcNAc linkage of the *N*-glycan arms can be β1,3 in humans, it is generally of the β1,4-linkage type, and hence illustrated as such. If required for later functional definition, further work will need to define the β1,3/4 Gal-GlcNAc distribution and validate the other assumed linkages.

Statistical assessment of N-glycome data

The structural *N*-glycan features of interest were compared between the three patient groups using all data points in the individual groups (i.e. two technical LC-MS/MS replicates for each biological (patient) replicate). Two-tailed Student's T-test, type 2 and 3 were applied

using MicroSoft Excel to assess the significance of the obtained *N*-glycome data between patient groups since the molecular change was not hypothesised and tested for. As such the generic null-hypotheses used in this study can be described as follows: The specific *N*-glycosylation structural feature being tested for does not change significantly between the relevant patient groups. P values less than 0.05 indicated that the null hypothesis was not supported. This null hypothesis rejection, in turn, means that the specific *N*-glycosylation structural feature being tested was significantly different between the relevant patient groups.

Cluster analysis

Cluster analyses were performed in order to check the relatedness of the sputum samples (CF1-5, iNCF1-2 and NCF1-2) using the relative abundances of the 68 *N*-glycans characterised from all the sputum samples. Hierarchical clustering was performed using the "hclust" implementation from R statistical package, with Euclidean distance and complete linkage on the log-transformed data.

Acknowledgements

We thank Professor Peter Middleton, Head of Cystic Fibrosis unit, Westmead clinical school for allowing us to study his patients, Shilpa Patel for collecting the sputum and patients, nurses in the CF unit and Dr. Dana Pascovici for her help in statistical analysis. Human research ethics approval was obtained from the Sydney West Area Health service (HREC/10/WMEAD/180). Authors declare no conflict of interests. This project was funded

by Macquarie University Postgraduate research fund and a Super Science Program from the Australian Research Council (FS110200026).

Abbreviations

CF: cystic fibrosis; CFTR: cystic fibrosis transmembrane conductance regulator; CID: collision-induced dissociation; COPD: chronic obstructive pulmonary disease; GnT-III: β-1,4-*N*-acetylglucosaminyltransferase III; iNCF: infected non-CF patient group; LC: liquid chromatography; MS: mass spectrometry; MS/MS: tandem mass spectrometry; NCF: pathogen-free non-CF patient group; non-CF: non-cystic fibrosis; PGC: porous graphitised carbon; PMN: polymorphonuclear leukocytes; SPE: solid phase extraction; URTI: upper respiratory tract infection

References

Altmann, F., Fabini, G., Ahorn, H. and Wilson, I. B. 2001. Genetic model organisms in the study of N-glycans. *Biochimie.* 83(8): 703-712.

Arnold, J. N., Wormald, M. R., Suter, D. M., Radcliffe, C. M., Harvey, D. J., Dwek, R. A., Rudd, P. M. and Sim, R. B. 2005. Human serum IgM glycosylation: identification of glycoforms that can bind to mannan-binding lectin. *J Biol Chem.* 280(32): 29080-29087.

Avichezer, D., Katcoff, D. J., Garber, N. C. and Gilboa-Garber, N. 1992. Analysis of the amino acid sequence of the Pseudomonas aeruginosa galactophilic PA-I lectin. *J Biol Chem.* 267(32): 23023-23027.

Babu, P., North, S. J., Jang-Lee, J., Chalabi, S., Mackerness, K., Stowell, S. R., Cummings, R. D., Rankin, S., Dell, A. and Haslam, S. M. 2009. Structural characterisation of neutrophil glycans by ultra sensitive mass spectrometric glycomics methodology. *Glycoconj J.* 26(8): 975-986.

Balog, C. I., Stavenhagen, K., Fung, W. L., Koeleman, C. A., McDonnell, L. A., Verhoeven, A., Mesker, W. E., Tollenaar, R. A., Deelder, A. M. and Wuhrer, M. 2012. N-glycosylation of colorectal cancer tissues: a liquid chromatography and mass spectrometry-based investigation. *Mol Cell Proteomics.* 11(9): 571-585.

Boucher, R. C. 2007. Evidence for airway surface dehydration as the initiating event in CF airway disease. *J Intern Med.* 261(1): 5-16.

Carnoy, C., Ramphal, R., Scharfman, A., Lo-Guidice, J. M., Houdret, N., Klein, A., Galabert, C., Lamblin, G. and Roussel, P. 1993. Altered carbohydrate composition of salivary mucins from patients with cystic fibrosis and the adhesion of Pseudomonas aeruginosa. *Am J Respir Cell Mol Biol.* 9(3): 323-334.

Carnoy, C., Scharfman, A., Van Brussel, E., Lamblin, G., Ramphal, R. and Roussel, P. 1994. Pseudomonas aeruginosa outer membrane adhesins for human respiratory mucus glycoproteins. *Infect Immun.* 62(5): 1896-1900.

Chen, C. P., Song, S. C., Gilboa-Garber, N., Chang, K. S. and Wu, A. M. 1998. Studies on the binding site of the galactose-specific agglutinin PA-IL from Pseudomonas aeruginosa. *Glycobiology.* 8(1): 7-16.

Dam, S., Thaysen-Andersen, M., Stenkjaer, E., Lorentzen, A., Roepstorff, P., Packer, N. H. and Stougaard, J. 2013. Combined N-glycome and N-glycoproteome analysis of the Lotus japonicus seed globulin fraction shows conservation of protein structure and glycosylation in legumes. *J Proteome Res.* 12(7): 3383-3392.

Davril, M., Degroote, S., Humbert, P., Galabert, C., Dumur, V., Lafitte, J. J., Lamblin, G. and Roussel, P. 1999. The sialylation of bronchial mucins secreted by patients suffering from cystic fibrosis or from chronic bronchitis is related to the severity of airway infection. *Glycobiology.* 9(3): 311-321.

Everest-Dass, A. V., Jin, D., Thaysen-Andersen, M., Nevalainen, H., Kolarich, D. and Packer, N. H. 2012. Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by Candida albicans. *Glycobiology.* 22(11): 1465-1479.

Fidler, K. J., Hilliard, T. N., Bush, A., Johnson, M., Geddes, D. M., Turner, M. W., Alton, E. W., Klein, N. J. and Davies, J. C. 2009. Mannose-binding lectin is present in the infected airway: a possible pulmonary defence mechanism. *Thorax.* 64(2): 150-155.

Frates, R. C., Jr., Kaizu, T. T. and Last, J. A. 1983. Mucus glycoproteins secreted by respiratory epithelial tissue from cystic fibrosis patients. *Pediatr Res.* 17(1): 30-34.

Gibson, R. L., Burns, J. L. and Ramsey, B. W. 2003. Pathophysiology and management of pulmonary infections in cystic fibrosis. *Am J Respir Crit Care Med.* 168(8): 918-951.

Gilligan, P. H. 1991. Microbiology of airway disease in patients with cystic fibrosis. *Clin Microbiol Rev.* 4(1): 35-51.

Hashii, N., Kawasaki, N., Itoh, S., Nakajima, Y., Kawanishi, T. and Yamaguchi, T. 2009. Alteration of N-glycosylation in the kidney in a mouse model of systemic lupus erythematosus: relative quantification of N-glycans using an isotope-tagging method. *Immunology*. 126(3): 336-345.

Heijerman, H. 2005. Infection and inflammation in cystic fibrosis: a short review. *J Cyst Fibros*. 4 Suppl 2: 3-5.

Holmen, J. M., Karlsson, N. G., Abdullah, L. H., Randell, S. H., Sheehan, J. K., Hansson, G. C. and Davis, C. W. 2004. Mucins and their O-Glycans from human bronchial epithelial cell cultures. *Am J Physiol Lung Cell Mol Physiol*. 287(4): L824-834.

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(7): 1299-1310.

Kilpatrick, D. C. 2003. Introduction to mannan-binding lectin. *Biochem Soc Trans*. 31(Pt 4): 745-747.

Koppel, R. and Solomon, B. 2001. IgM detection via selective recognition by mannose-binding protein. *J Biochem Biophys Methods*. 49(1-3): 641-647.

Kube, D., Adams, L., Perez, A. and Davis, P. B. 2001. Terminal sialylation is altered in airway cells with impaired CFTR-mediated chloride transport. *Am J Physiol Lung Cell Mol Physiol*. 280(3): L482-492.

Lamblin, G., Degroote, S., Perini, J. M., Delmotte, P., Scharfman, A., Davril, M., Lo-Guidice, J. M., Houdret, N., Dumur, V., Klein, A. and Rousse, P. 2001. Human airway mucin glycosylation: a combinatory of carbohydrate determinants which vary in cystic fibrosis. *Glycoconj J*. 18(9): 661-684.

Lee, R. T., Ichikawa, Y., Fay, M., Drickamer, K., Shao, M. C. and Lee, Y. C. 1991. Ligand-binding characteristics of rat serum-type mannose-binding protein (MBP-A). Homology of binding site architecture with mammalian and chicken hepatic lectins. *J Biol Chem*. 266(8): 4810-4815.

Leymarie, N., Griffin, P. J., Jonscher, K., Kolarich, D., Orlando, R., McComb, M., Zaia, J., Aguilan, J., Alley, W. R., Altmann, F., Ball, L. E., Basumallick, L., Bazemore-Walker, C. R., Behnken, H., Blank, M. A., Brown, K. J., Bunz, S. C., Cairo, C. W., Cipollo, J. F., Daneshfar, R., Desaire, H., Drake, R. R., Go, E. P., Goldman, R., Gruber, C., Halim, A., Hathout, Y., Hensbergen, P. J., Horn, D. M., Hurum, D., Jabs, W., Larson, G., Ly, M., Mann, B. F., Marx, K., Mechref, Y., Meyer, B., Moginger, U., Neusubeta, C., Nilsson, J., Novotny, M. V., Nyalwidhe, J. O., Packer, N. H., Pompach, P., Reiz, B., Resemann, A., Rohrer, J. S., Ruthenbeck, A., Sanda, M., Schulz, J. M., Schweiger-Hufnagel, U., Sihlbom, C., Song, E., Staples, G. O., Suckau, D., Tang, H., Thaysen-Andersen, M., Viner, R. I., An, Y., Valmu, L., Wada, Y., Watson, M., Windwarder, M., Whittal, R., Wuhrer, M., Zhu, Y. and Zou, C. 2013. Interlaboratory study on differential analysis of protein glycosylation by mass spectrometry: the ABRF glycoprotein research multi-institutional study 2012. *Mol Cell Proteomics*. 12(10): 2935-2951.

Lillehoj, E. P., Hyun, S. W., Kim, B. T., Zhang, X. G., Lee, D. I., Rowland, S. and Kim, K. C. 2001. Muc1 mucins on the cell surface are adhesion sites for Pseudomonas aeruginosa. *Am J Physiol Lung Cell Mol Physiol*. 280(1): L181-187.

Lipuma, J. J. 2010. The changing microbial epidemiology in cystic fibrosis. *Clin Microbiol Rev*. 23(2): 299-323.

Matthews, L. W., Spector, S., Lemm, J. and Potter, J. L. 1963. Studies on Pulmonary Secretions. I. The over-All Chemical Composition of Pulmonary Secretions from Patients with Cystic Fibrosis, Bronchiectasis, and Laryngectomy. *Am Rev Respir Dis*. 88: 199-204.

Mickle, J. E. and Cutting, G. R. 2000. Genotype-phenotype relationships in cystic fibrosis. *Med Clin North Am*. 84(3): 597-607.

Natsuka, S., Adachi, J., Kawaguchi, M., Nakakita, S., Hase, S., Ichikawa, A. and Ikura, K. 2002. Structural analysis of N-linked glycans in Caenorhabditis elegans. *J Biochem*. 131(6): 807-813.

Pabst, M., Bondili, J. S., Stadlmann, J., Mach, L. and Altmann, F. 2007. Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Anal Chem*. 79(13): 5051-5057.

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(12): 5791-5800.

Plotkowski, M. C., Bajolet-Laudinat, O. and Puchelle, E. 1993. Cellular and molecular mechanisms of bacterial adhesion to respiratory mucosa. *Eur Respir J.* 6(6): 903-916.

Quinton, P. M. 2001. The neglected ion: HCO3. *Nat Med.* 7(3): 292-293.

Ramphal, R., Carnoy, C., Fievre, S., Michalski, J. C., Houdret, N., Lamblin, G., Strecker, G. and Roussel, P. 1991. Pseudomonas aeruginosa recognizes carbohydrate chains containing type 1 (Gal beta 1-3GlcNAc) or type 2 (Gal beta 1-4GlcNAc) disaccharide units. *Infect Immun.* 59(2): 700-704.

Reddy, M. M. and Quinton, P. M. 2001. Selective activation of cystic fibrosis transmembrane conductance regulator Cl- and HCO3- conductances. *JOP.* 2(4 Suppl): 212-218.

Rhim, A. D., Stoykova, L., Glick, M. C. and Scanlin, T. F. 2001. Terminal glycosylation in cystic fibrosis (CF): a review emphasizing the airway epithelial cell. *Glycoconj J.* 18(9): 649-659.

Rommens, J. M., Iannuzzi, M. C., Kerem, B., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N. and et al. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science.* 245(4922): 1059-1065.

Scharfman, A., Degroote, S., Beau, J., Lamblin, G., Roussel, P. and Mazurier, J. 1999. Pseudomonas aeruginosa binds to neoglycoconjugates bearing mucin carbohydrate determinants and predominantly to sialyl-Lewis x conjugates. *Glycobiology.* 9(8): 757-764.

Schulz, B. L., Sloane, A. J., Robinson, L. J., Prasad, S. S., Lindner, R. A., Robinson, M., Bye, P. T., Nielson, D. W., Harry, J. L., Packer, N. H. and Karlsson, N. G. 2007. Glycosylation of sputum mucins is altered in cystic brosis patients. *Glycobiology.* 17(7): 698-712.

Schulz, B. L., Sloane, A. J., Robinson, L. J., Sebastian, L. T., Glanville, A. R., Song, Y., Verkman, A. S., Harry, J. L., Packer, N. H. and Karlsson, N. G. 2005. Mucin glycosylation changes in cystic fibrosis lung disease are not manifest in submucosal gland secretions. *Biochem J.* 387(Pt 3): 911-919.

Sethi, M. K., Thaysen-Andersen, M., Smith, J. T., Baker, M. S., Packer, N. H., Hancock, W. S. and Fanayan, S. 2013. Comparative N-Glycan Profiling of Colorectal Cancer Cell Lines Reveals Unique Bisecting GlcNAc and alpha-2,3-Linked Sialic Acid Determinants Are Associated with Membrane Proteins of the More Metastatic/Aggressive Cell Lines. *J Proteome Res.*

Stanley, P. 2002. Biological consequences of overexpressing or eliminating N-acetylglucosaminyltransferase-TIII in the mouse. *Biochim Biophys Acta.* 1573(3): 363-368.

Stanley, P., Schachter, H. and Taniguchi, N. (2009). N-Glycans. Essentials of Glycobiology. C. R. Varki A, Esko JD, et al., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press;.

Venkatakrishnan, V., Packer, N. H. and Thaysen-Andersen, M. 2013. Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis. *Expert Rev Respir Med.* 7(5): 553-576.

Voynow, J. A. and Rubin, B. K. 2009. Mucins, mucus, and sputum. *Chest.* 135(2): 505-512.

Xia, B., Royall, J. A., Damera, G., Sachdev, G. P. and Cummings, R. D. 2005. Altered O-glycosylation and sulfation of airway mucins associated with cystic fibrosis. *Glycobiology.* 15(8): 747-775.

Xia, B., Sachdev, G. P. and Cummings, R. D. 2007. Pseudomonas aeruginosa mucoid strain 8830 binds glycans containing the sialyl-Lewis x epitope. *Glycoconj J.* 24(1): 87-95.

Zipser, B., Bello-DeOcampo, D., Diestel, S., Tai, M.-H. and Schmitz, B. 2012. Mannitou Monoclonal Antibody Uniquely Recognizes Paucimannose, a Marker for Human Cancer, Stemness, and Inflammation. *Journal of Carbohydrate Chemistry.* 31(4-6): 504-518.

Figure legends

Figure 1. PGC-LC-negative ion-MS/MS mass profiles of reduced *N*-glycans derived of soluble sputum proteins from five individuals with CF (CF1-5). Similarities between the CF glycomes were identified. In addition, three similar mass profiles corresponding to the sputum *N*-glycome from the same patient (CF5) at three different time points are presented denoted as CF5, CF5- 3 Months and CF5- 6 Months. Top five abundant *N*-glycans across all CF sputum is also represented. *represents contaminanting signals of unknown origin and #n corresponds to the structure numbers in Table 2.

Figure 2. Abundance of the identified *N*-glycans grouped into four *N*-glycan types including complex, high mannose, hybrid and paucimannose of sputum proteins derived from five CF (CF1-5), two infected non-CF (iNCF1-2) and two pathogen-free non-CF (NCF1-2) patients. Two-tailed Student's t-tests showed significant differences in the complex, hybrid and paucimannose type *N*-glycans between CF and non-CF *N*-glycomes, whereas no significant differences were observed for *N*-glycans of the high mannose type.

Figure 3. A) Pie-chart illustrating the distribution of the high-mannose type *N*-glycans (Man_5-Man_9) released from CF (CF1-5), infected non-CF (iNCF1-2) and pathogen-free non-CF (NCF1-2) sputum proteins. While non-CF shows a relative equal distribution of the high-mannose type *N*-glycans, CF and iNCF have a comparative higher proportion of Man_5 *N*-glycan. Box-plots showing that the degree of sialylation (B), fucosylation (C) and bisecting GlcNAcylation (D) of sputum *N*-glycans are altered in CF compared with non-CF sputum *N*-glycans. These values are calculated as a abundance weighted ratio of the modifed *N*-glycans out of the total weighted subset of potentially modifed *N*-glycan.

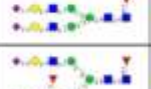
Figure 4. Cluster dendrogram of the sputum *N*-glycomes of patients from the three catagories (CF1-5, iNCF1-2 and NCF1-2) measured using the relative abundances of the observed *N*-glycans. The *N*-glycomes within each CF patient groups cluster well together and *N*-glycomes from the infected non-CF groups clustered stronger with the CF profiles than the pathogen-free non-CF suggesting that the malfunction of CFTR influence the sputum *N*-glycosylation less in CF than the bacterial infection.

Table I: Clinical and laboratory characteristis of the study patients

Patient No.	Age	Sex	Co-morbidity	CF genotype	Bacteria/ fungi cultured	Predominant antibiotic therapy
CF1	27	M	Cystic Fibrosis (CF)	ΔF508 / ΔF508	Mucoid *P. aeruginosa*	Piperacillin/tazobactam ciprofloxacin, polymxin, inhaled tobramycin
CF2	23	M	CF	ΔF508 / ΔF508	Non-mucoid *P. aeruginosa*	Clarithromycin, moxifloxacin
CF3	19	F	CF	ΔF508 / ΔF508	*S. aureus*	Piperacillin/tazobactam, azithromycin, itraconazole, inaled tobramycin and bronchitol,
CF4	23	F	CF	ΔF508 / ΔF508	Mucoid and non-mucoid *P. aeruginosa*	Piperacillin/tazobactam, inhaled polymyxin, azithromycin, itraconazole
CF5	40	F	CF	ΔF508 / ΔD1507	*P. aeruginosa & A. fumigatus*	Piperacillin/tazobactam, ciproflaxin, inhaled tobramycin and polymyxin
iNCF1	28	F	Upper Respiratory Tract Infection (URTI)	N/A	*S. pneumonia*	None
iNCF2	29	M	(URTI	N/A	*S. aureus*	None
NCF1	18	F	Pneumonia	N/A	No pathogens isolated	Ceftriaxone, azithromycin,
NCF2	69	F	Chronic Pulmonary Obstructive Disorder (COPD)	N/A	No pathogens isolated	Cetriaxone

Table II: Overview of the 68 *N*-glycan structures and their relative abundances in sputum from the CF, iNCF and NCF patient groups with the average relative abundances in each category.

#	Structure	Av. Relative abundance ± St. Dev (%)			#	Structure	Av. Relative abundance ± St. Dev (%)		
		CF	iNCF	NCF			CF	iNCF	NCF
1		0.8 ± 0.4	0.7 ± 0.3	0	2		3.7 ± 1.4	2.7 ± 3	0
3		2.1 ± 0.5	1.7 ± 0.7	0.1 ± 0	4		27.1 ± 4.4	17.7 ± 7.2	0.7 ± 0.5
5		2.3 ± 0.9	1.6 ± 0.3	0.6 ± 0.5	6		6.3 ± 2.2	2.2 ± 0.8	1.8 ± 1.3
7a		0.4 ± 0.2	0.4 ± 0.3	0.1 ± 0.1	7b		0.5 ± 0.4	0.2 ± 0.2	0
8a		1.3 ± 0.6	1.4 ± 1.1	0.1 ± 0.1	8b		1.2 ± 0.5	0.8 ± 0.8	1.9 ± 1.5
9		6.3 ± 3.1	6.1 ± 4.9	2.3 ± 0.5	10a		0.9 ± 0.5	1.5 ± 0.5	1.4 ± 0.2
10b		0.9 ± 0.8	1.0 ± 0.6	0.9 ± 0.3	11a		0.7 ± 0.4	0.6 ± 0.8	1.9 ± 0.6
11b		0.3 ± 0.4	0.7 ± 0.9	0.2 ± 0.1	12		0.9± 0.3	2.0 ± 1.3	1.9 ± 0.9
13		0.2 ± 0.3	0.3 ± 0.3	0.9 ± 0.3	14a		0.5 ± 0.5	0.7 ± 0.8	1.1 ± 1.0
14b		0.1 ±0.1	0	0.8 ± 1.3	15		0.5 ± 0.3	0.8 ± 0.3	1.4 ± 0.4
16		0.8 ± 0.2	0.3 ± 0.1	0.1 ± 0.1	17a		0.5 ± 0.5	0.8 ± 1.0	4.1 ± 2.6
17b		0.9 ± 1.1	0.6 ± 0.7	0.5 ± 0.2	18		0.3 ± 0.3	0.6 ± 0.6	9.4 ± 3.7
19		0.8 ± 0.3	0.5 ± 0.1	0	20		0.5 ± 0.5	0	0.7 ± 0.8
21a		0.5 ± 0.3	0.9 ± 0.3	0.7 ± 0.2	21b		0.2 ± 0.2	0.6 ± 0.6	0.2 ± 0.2
22a		2.0 ± 1.6	4.3 ± 2.3	22.1±8.1	22b		0	0	0.8 ± 0.6
23a		9.2 ± 1.2	7.2 ± 6.3	1.1 ± 1.0	23b		1.1 ± 0.5	0.5 ± 0.4	2.1 ± 2.3

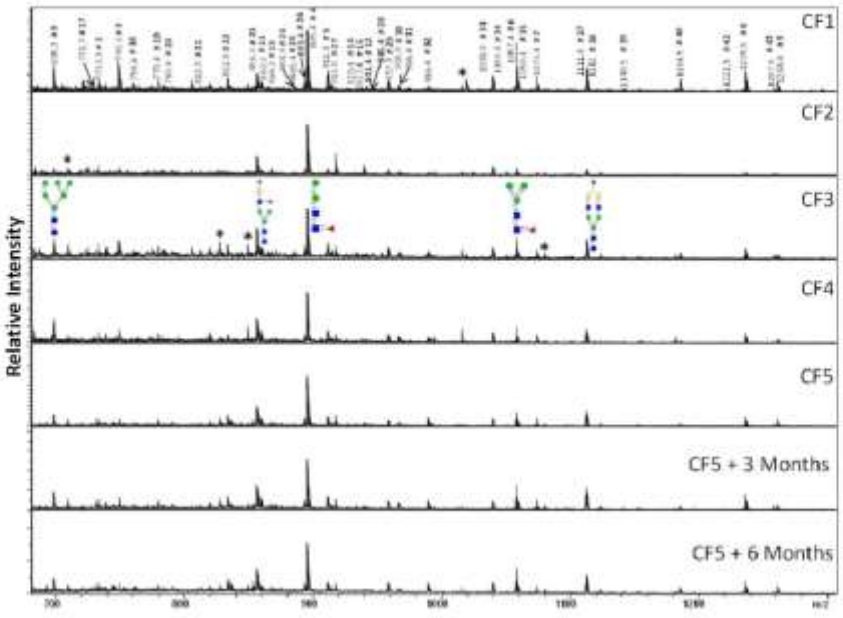
23c		0	0	0.2 ± 0.3	23d		0	0	0.1 ± 0.1
24a		0.3 ± 0.6	7.8 ± 9.9	1.2 ± 0.6	24b		0	2.1 ± 2.7	0.7 ± 1.1
25a		1.0 ± 0.8	0.3 ± 0.3	0.7 ± 0.9	25b		0.6 ± 1.6	0	0.6 ± 0.5
26		0.8 ± 0.9	0.7 ± 0.2	0.3 ± 0.1	27a		0.5 ± 0.5	1.6 ± 0.7	0.1 ± 0.1
27b		0.3 ± 0.4	1.0 ± 1.4	6.6 ± 2.8	28a		0.4 ± 0.4	0.2 ± 0.3	0.9 ± 0.5
28b		0	0	0.7 ± 0.3	28c		0	0	1.0 ± 1.0
29a		1.4 ± 1.7	2.1 ± 2.4	1.0 ± 1.6	29b		0.1 ± 0.1	0.1 ± 0.1	0.5 ± 0.6
30a		1.1 ± 0.6	1.6 ± 0.8	0.2 ± 0.4	30b		0.1 ± 0.6	1.5 ± 2.4	0.7 ± 1.4
31		1.2 ± 2.2	0.1 ± 0.1	0.2 ± 0.4	32a		0	0.7 ± 0.9	1.6 ± 0.5
32b		0	0.3 ± 0.4	0	33a		2.1 ± 0.9	1.0 ± 0.6	0.2 ± 0.3
33b		0.8 ± 1.4	0	2.7 ± 0.2	34		2.2 ± 2.1	3.2 ± 2.4	0.3 ± 0.4
35		0.2 ± 0.3	1.2 ± 1.4	4.3 ± 0.2	36		0	0.2 ± 0.2	1.4 ± 2.0
37a		2.5 ± 2.5	3.7 ± 4.4	0.7 ± 0.5	37b		0.1 ± 0.3	0.2 ± 0.2	0.2 ± 0.2
38a		5.2 ± 3.0	4.3 ± 1.1	1.9 ± 1.0	38b		0.9 ± 1.3	0	2.7 ± 2.4
39		0.2 ± 0.2	0.5 ± 0.2	1.5 ± 0.8	40		0.9 ± 0.7	0.8 ± 0.7	0.7 ± 0.1
41a		1.2 ± 0.7	0.6 ± 0.6	0.7 ± 0.1	41b		0.4 ± 0.4	0.8 ± 1.1	0.3 ± .02
42		0.1 ± 0.2	0	0	43a		0.2 ± 0.1	1.0 ± 0.8	0.3 ± 0.2

34

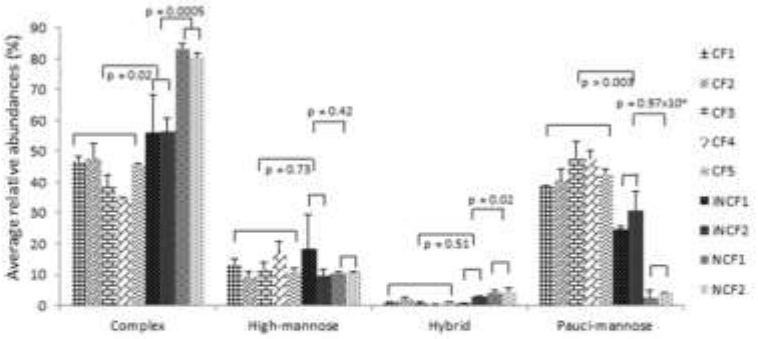
43b		0.1 ± 0.2	0.1 ± 0.1	0.3 ± 0.6	44		0	0.2 ± 0.1	0.1 ± 0.1

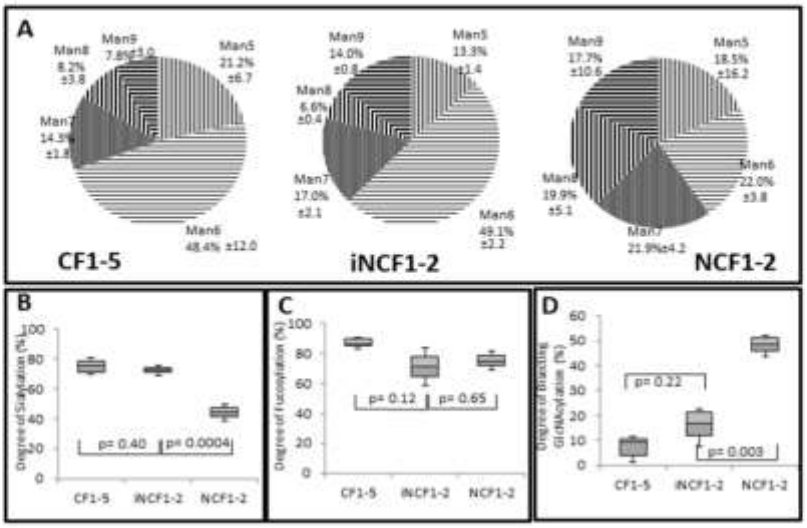
Table III: Distribution of linkage specific sialylation and fucosylation measured as an average percentage (± standard deviation) of the total sialylation of complex and hybrid type *N*-glycans from sputum of the three patient groups. P of all distributions using Student's T-test indicated that neither the linkage specific sialylation nor fucosylation are significantly different when comparing the three patient groups.

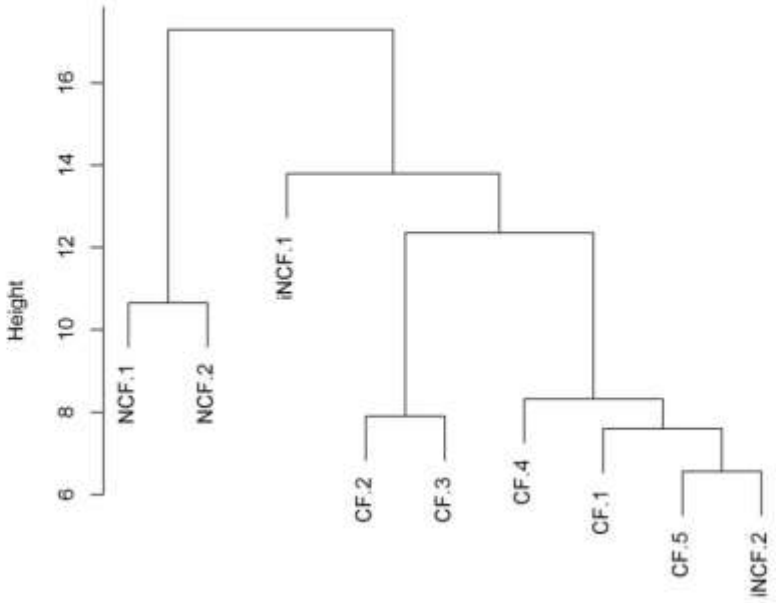
Feature	Linkage type	Relative Abundances (%)			p Values	
		CF	iNCF	NCF	CF vs iNCF	iNCF vs NCF
Sialylation (NeuAc)	α2,6	83.3 ± 11.9	83.8 ± 14.8	88.1 ± 3.3	0.96	0.60
	α2,3	15.6 ± 11.9	13.7 ± 11.9	9.8 ± 2.3	0.80	0.56
	α2,6 and α2,3	1.0 ± 0.7	2.5 ± 3.2	2.1 ± 1.6	0.44	0.86
Fucosylation (Fucose)	α1,6 (Core)	67.8 ± 6.4	68 ± 12.5	80.7 ± 2.6	0.98	0.13
	α1,3/4 (Arm)	27.9 ± 5.1	25.8 ± 9	17.8 ± 2.6	0.68	0.17
	Core and Arm	4.2 ± 2.7	6.2 ± 3.8	1.5 ± 0.3	0.39	0.09



37







2.1 Supplementary file

Supplementary file for

Cystic fibrosis and bacterial colonisation define the sputum *N*-glycosylation phenotype

Vignesh Venkatakrishnan[1], Morten Thaysen-Andersen[1], Sharon C.- A. Chen[2], Helena Nevalainen[1], Nicolle H. Packer[1]

[1]Biomolecular Frontiers Research Centre, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia 2109

[2]Centre for Infectious Diseases and Microbiology Laboratory Services, ICPMR-Pathology West, Westmead Hospital, Sydney, Australia 2006

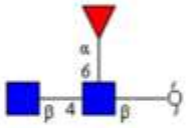
Supplementary Table S1: Monosaccharide compositions of the *N*-glycans without quality MS/MS which were not used for relative quantitative analysis. Presence or absence of each *N*-glycan in CF, iNCF and NCF is also represented.

Glycan #	m/z	charge	Obs. Mass (M)	Theo mass (M)	Hex	HexNAc	dHex	NeuAc	Phosphate	Type	CF	iNCF	NCF
S2	667.3	1	668.3	668.2	1	2			1	Pauci-mannose	Y	Y	N
S2	911.4	1	912.4	912.3	3	2				Pauci-mannose	Y	Y	N
S3	1073.4	1	1074.4	1074.4	4	2				Pauci-mannose	Y	Y	N
S4-a	1153.4	1	1154.4	1154.4	4	2			1	Pauci-mannose	Y	Y	N
S4-b	1153.4	1	1154.4	1154.4	4	2			1	Pauci-mannose	Y	Y	N
S5-a	738.3	2	1478.6	1478.5	6	2			1	High Mannose	Y	Y	Y
S5-b	738.3	2	1478.6	1478.5	6	2			1	High Mannose	Y	Y	N
S6-a	819.3	2	1640.6	1640.5	7	2			1	High Mannose	Y	Y	Y
S6-b	819.9	2	1640.6	1640.5	7	2			1	High Mannose	Y	N	Y
S7	859.3	2	1720.6	1720.5	7	2			2	High Mannose	Y	Y	Y
S8	1235.5	1	1236.5	1236.4	5	2				High Mannose	N	Y	N
S9	1315.5	1	1316.5	1316.4	5	2			1	High Mannose	Y	Y	N
S10	937.4	2	1876.8	1876.7	5	3	1	1		Hybrid	N	N	Y
S11-a	783.3	2	1568.6	1568.5	4	3		1		Complex	N	N	Y
S11-b	783.3	2	1568.6	1568.5	4	3		1		Complex	N	N	Y
S12	812.3	2	1626.6	1626.6	4	4	1			Complex	N	N	Y
S13	863.8	2	1729.6	1729.6	4	3	2		2	Complex	Y	Y	N
S14-a	957.3	2	1917.8	1917.7	4	4	1	1		Complex	N	N	Y
S14-b	957.3	2	1917.8	1917.7	4	4	1	1		Complex	N	N	Y
S15	966.4	2	1934.8	1934.7	5	4	2			Complex	Y	Y	Y

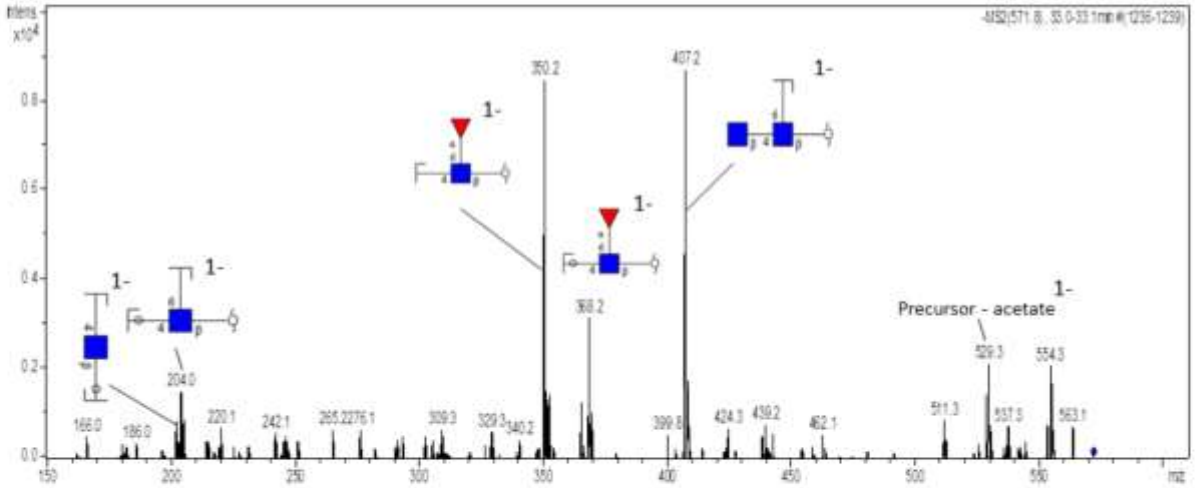
S16	986.4	2	1974.8	1974.7	4	5		1		Complex	N	Y	N
S17	1038.9	2	2079.8	2079.7	5	4	1	1		Complex	N	N	N
S18	1039.4	2	2080.8	2080.8	5	4	3			Complex	Y	Y	Y
S19	1098.4	1	1099.4	1099.4	2	3	1			Complex	Y	Y	N
S20	1112	2	2226.0	2225.8	5	4	2	1		Complex	N	N	Y
S21-a	1132.5	2	2267.0	2266.8	4	5	2	1		Complex	Y	Y	Y
S21-b	1132.5	2	2267.0	2266.8	4	5	2	1		Complex	N	N	N
S22	1140.5	2	2283.0	2282.8	5	5	1	1		Complex	N	N	Y
S23	1184.5	2	2371.0	2370.8	5	4	1	2		Complex	Y	Y	Y
S24-a	1257.5	2	2517.0	2516.9	5	4	2	2		Complex	Y	Y	Y
S24-b	1257.5	2	2517.0	2516.9	5	4	2	2		Complex	N	N	Y
S25	1260.5	1	1261.5	1261.5	3	3	1			Complex	Y	Y	Y
S26	1330.5	2	2663.0	2662.9	5	4	3	2		Complex	Y	Y	Y
S26	1367	2	2736.0	2736.0	6	5	1	2		Complex	Y	N	N
S27	1367.5	2	2737.0	2737.0	6	5	3	1		Complex	Y	Y	Y
S28	1439.5	2	2881.0	2881.0	6	5		3		Complex	Y	Y	N
S29	1512.5	2	3027.0	3027.1	6	5	1	3		Complex	Y	Y	N

Supplementary Figure S1: The entire set of validated MS/MS spectra in an annotated form of each of the 68 identified *N*-glycans is given

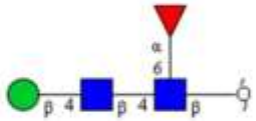
Glycan #1



m/z 571.3 (1-), RT: 33.0 min
$[M-H]^-$ = 571.3

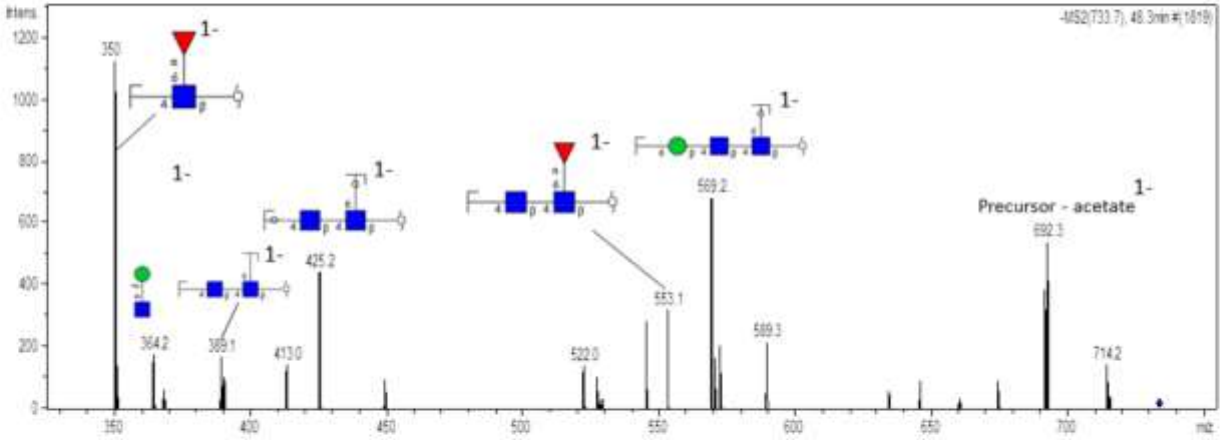


Glycan #2



m/z 733.4 (1-), RT: 48.3 min

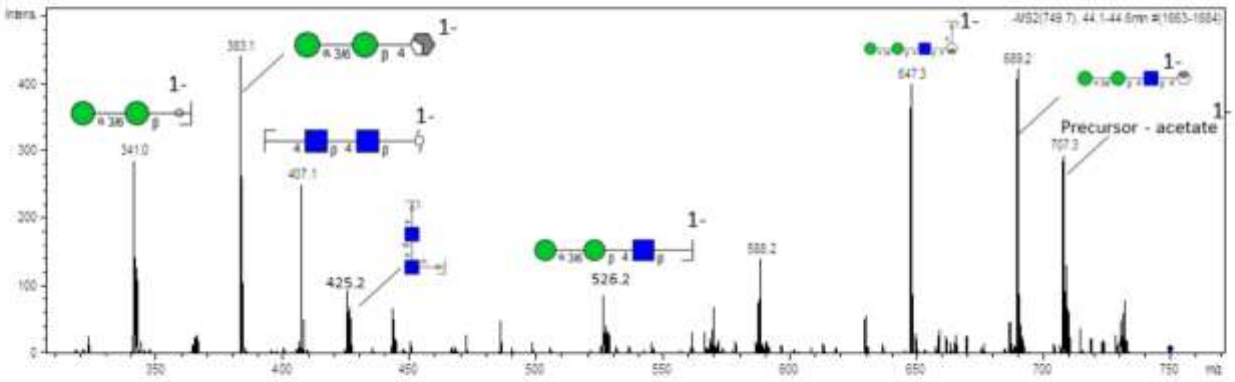
$[M\text{-}H]^-$ = 733.4



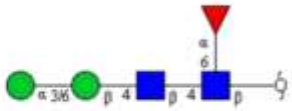
Glycan #3

m/z 749.4 (1-), RT: 44.4 min

$[M-H]^-$ = 749.4

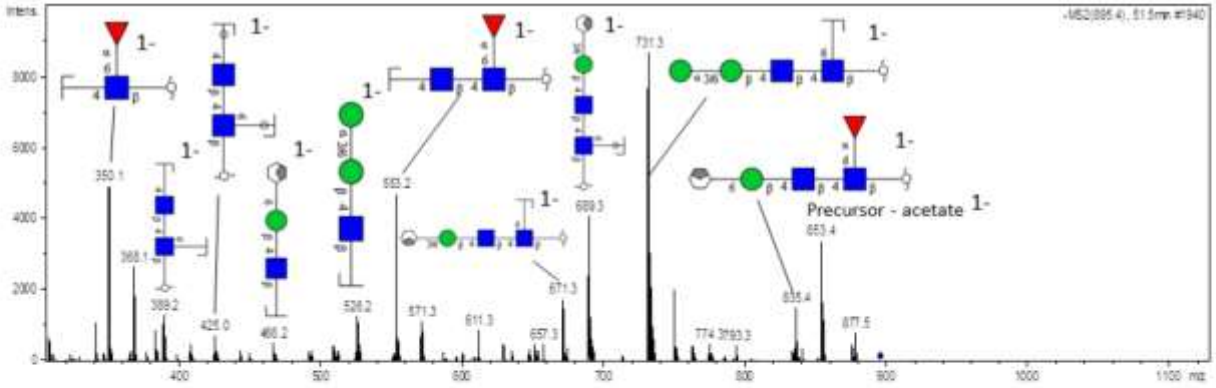


Glycan #4

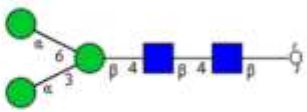


m/z 895.4 (1-), RT: 51.5 min

$[M\text{-}H]^-$ = 895.4

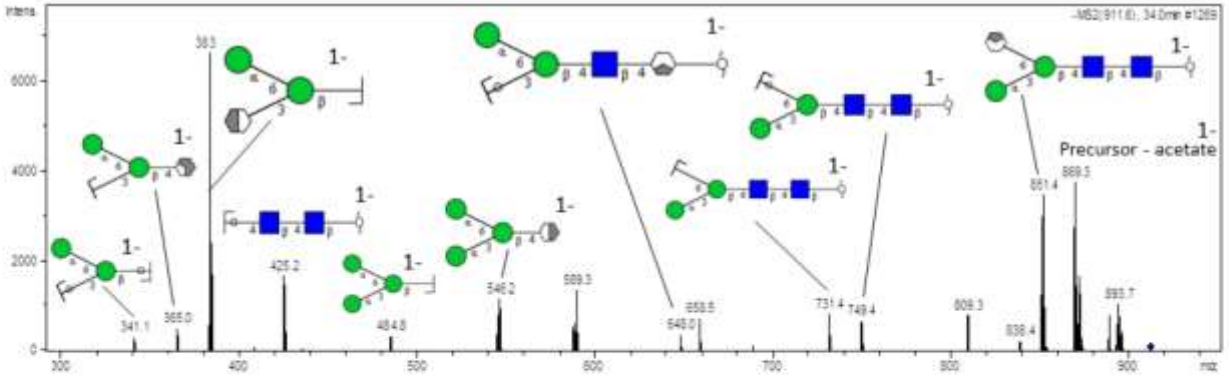


Glycan #5

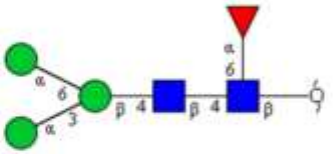


m/z 911.4 (1-), RT: 34.0 min

$[M\text{-}H]^-$ = 911.4

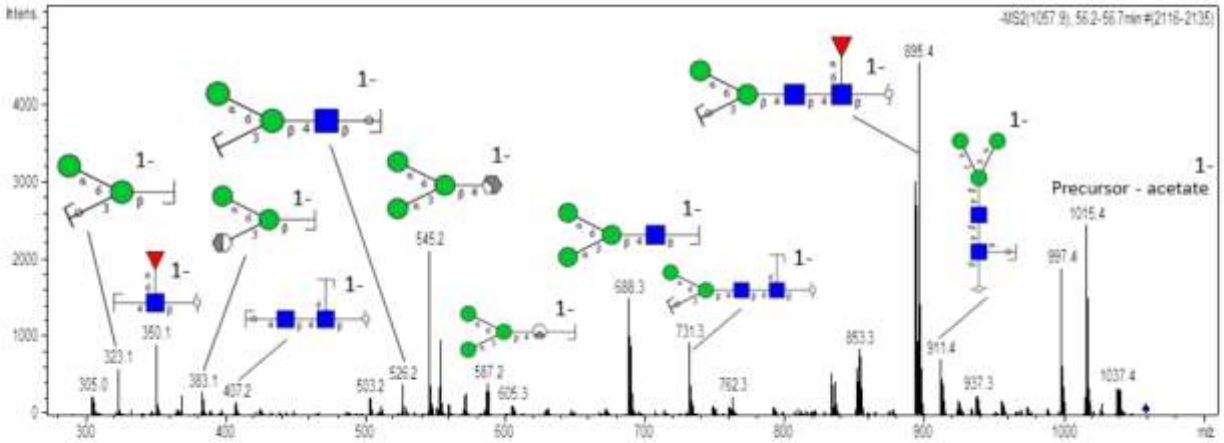


Glycan #6

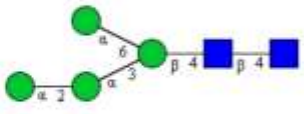


m/z 1057.4 (1-), RT: 56.4 min

$[M\text{-}H]^-$ = 1057.4

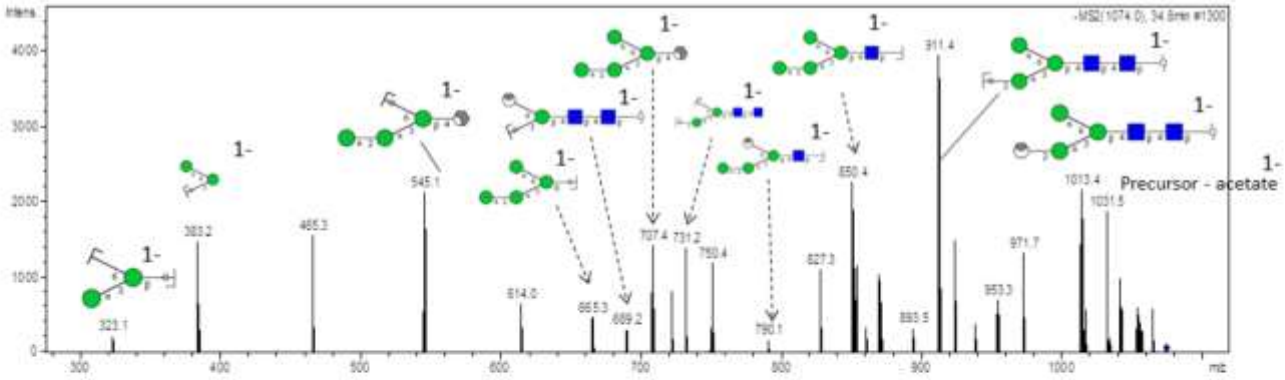


Glycan #7a

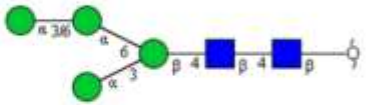


m/z 1073.4 (1-), RT: 34.8 min

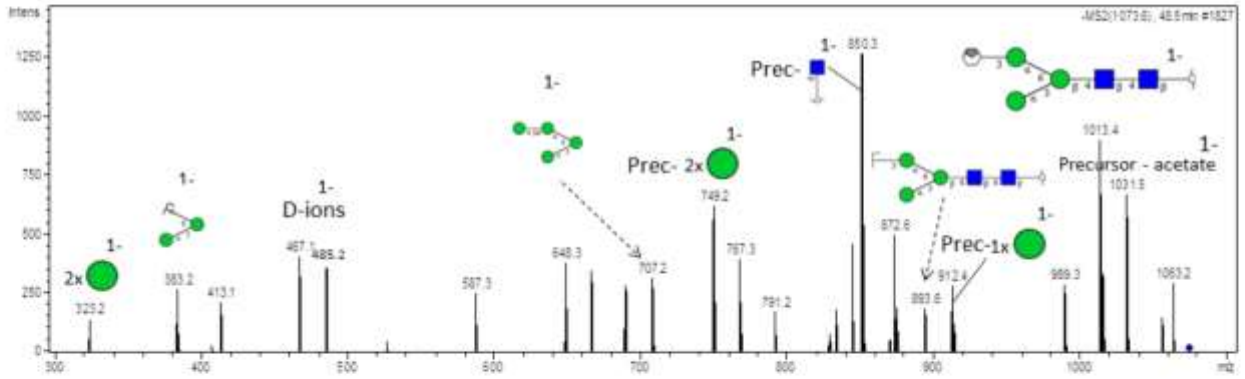
$[M\text{-}H]^-$ = 1073.4



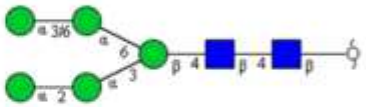
Glycan #7b



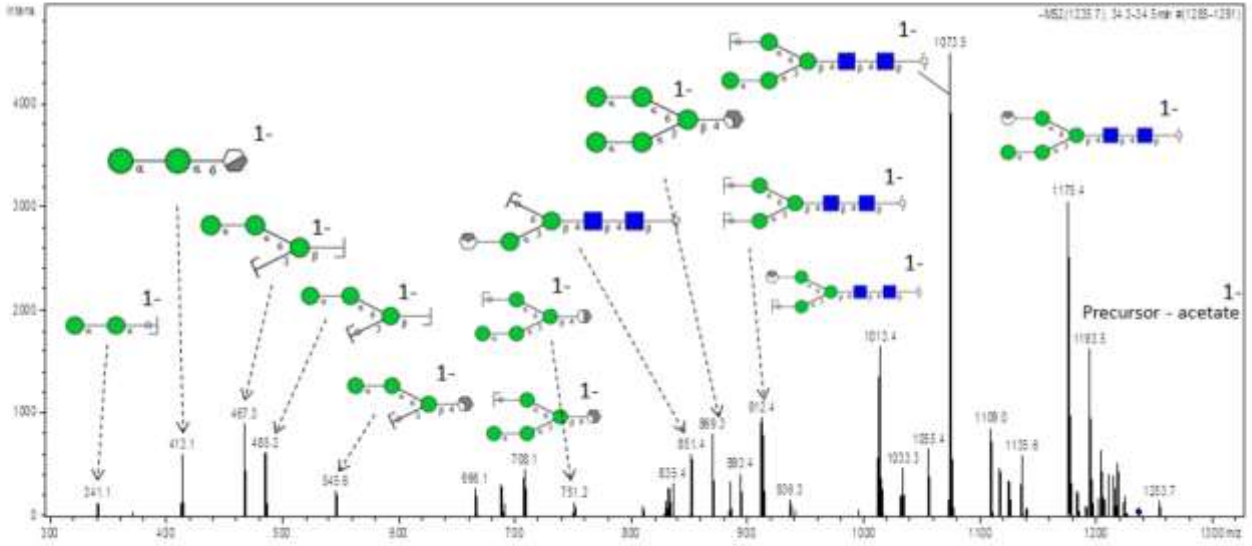
m/z 1073.4 (1-), RT: 48.5 min
$[M-H]^-$ = 1073.4



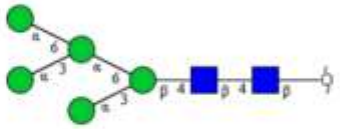
Glycan #8a



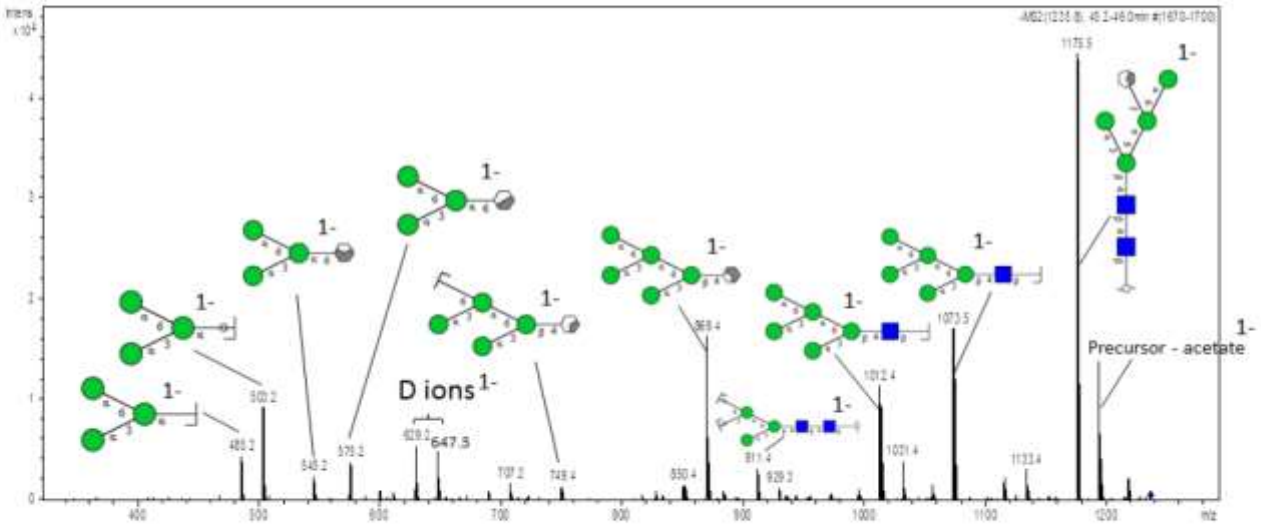
m/z 1235.4 (1-), RT: 34.4 min
$[M\text{-}H]^-$ = 1235.4



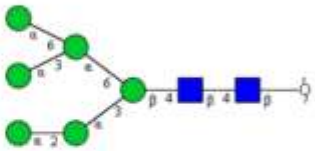
Glycan #8b



m/z 1235.4 (1-), RT: 45.2 min
$[M\text{-}H]^-$ = 1235.4

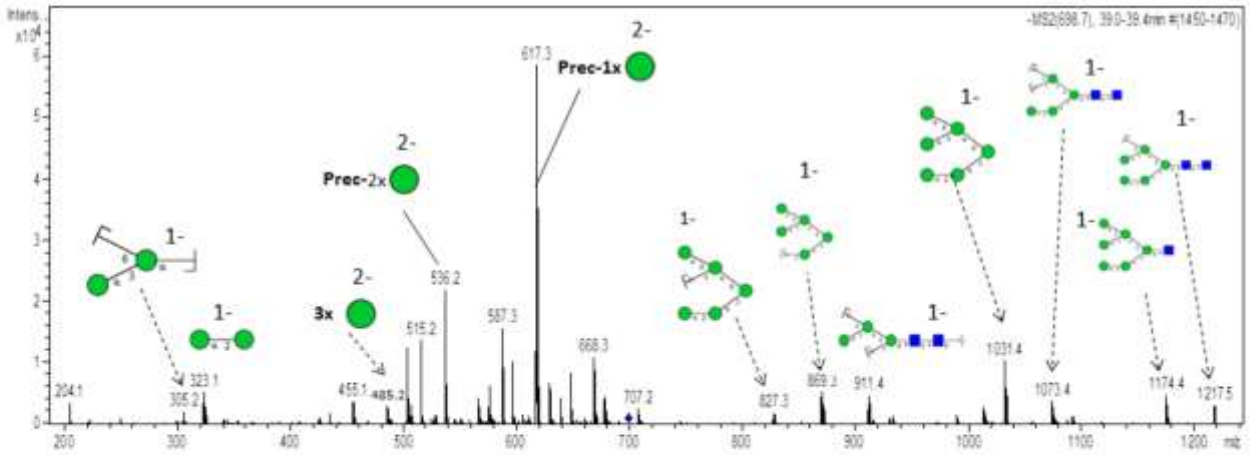


Glycan #9

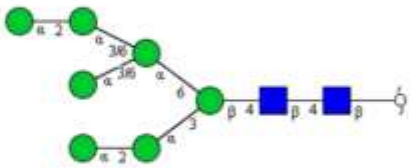


m/z 698.4 (2-), RT: 39.2 min

$[M-H]^-$ = 1397.8

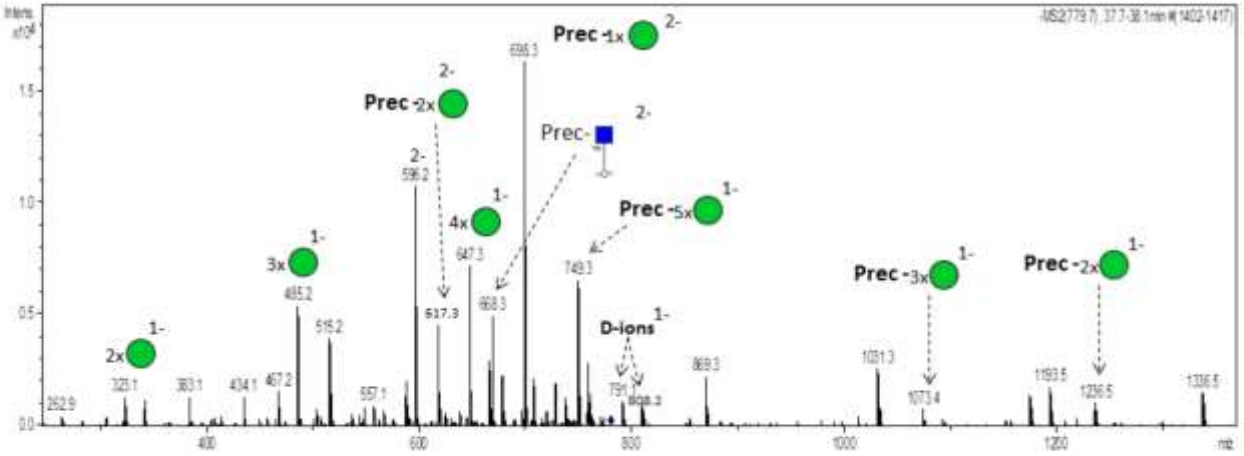


Glycan #10a

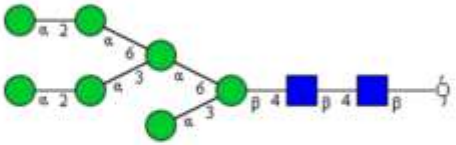


m/z 779.4 (2-), RT: 37.9 min

$[M-H]^-$ = 1559.8

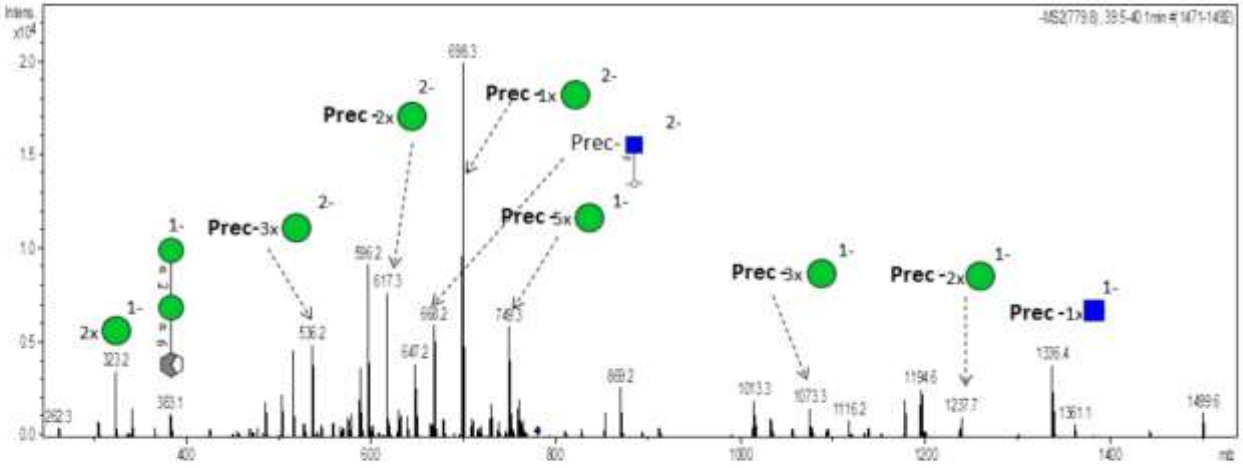


Glycan #10b

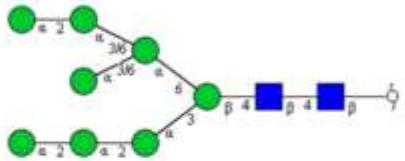


m/z 779.4 (2-), RT: 39.7 min

$[M\text{-}H]^-$ = 1559.8

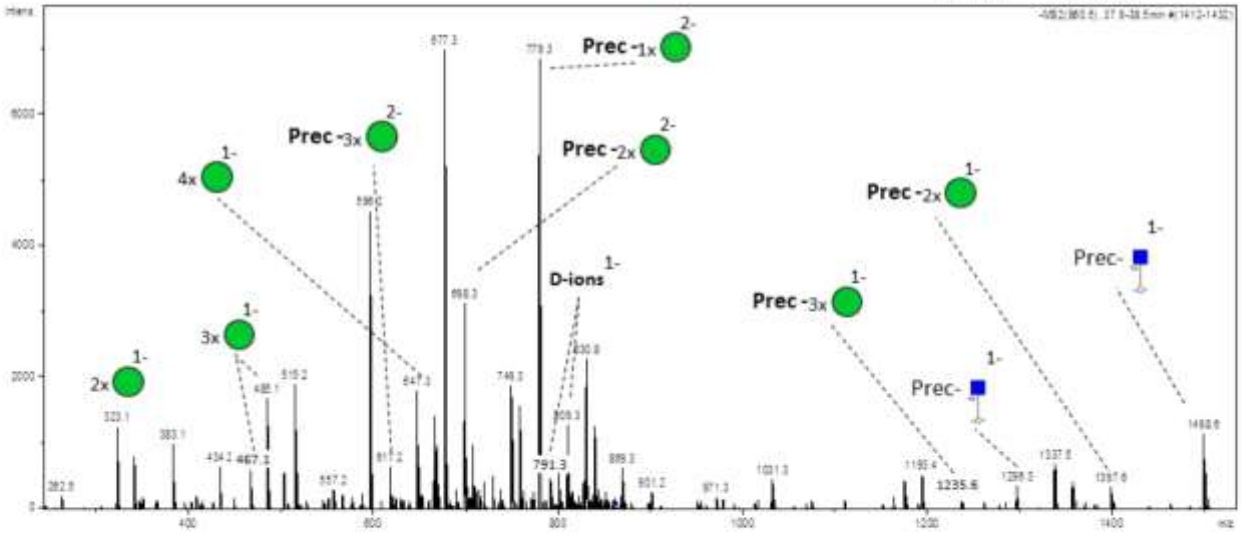


Glycan #11a

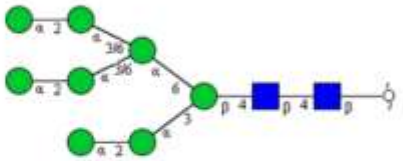


m/z 860.4 (2-), RT: 38.3 min

$[M-H]^-$ = 1721.8

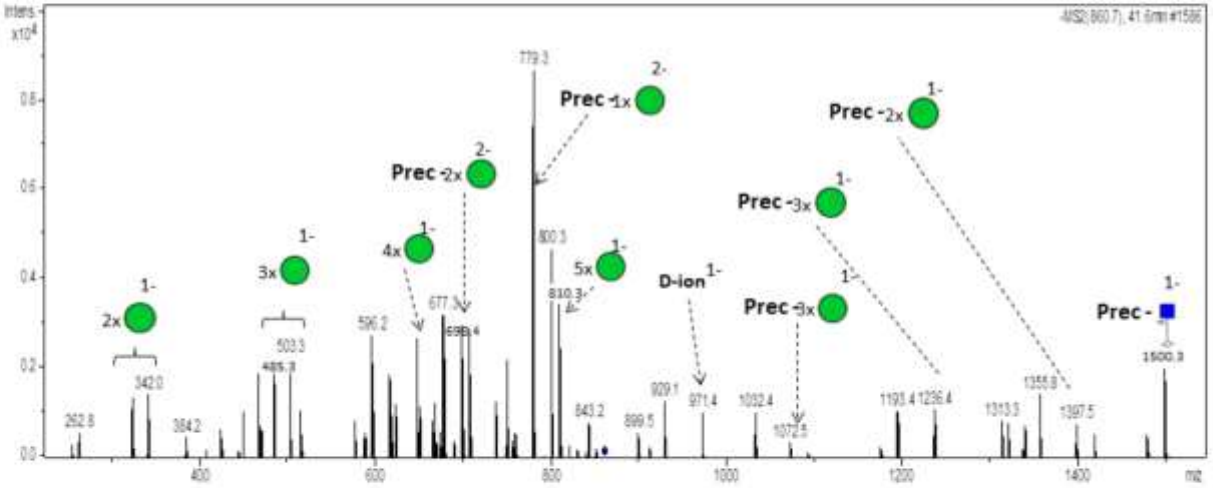


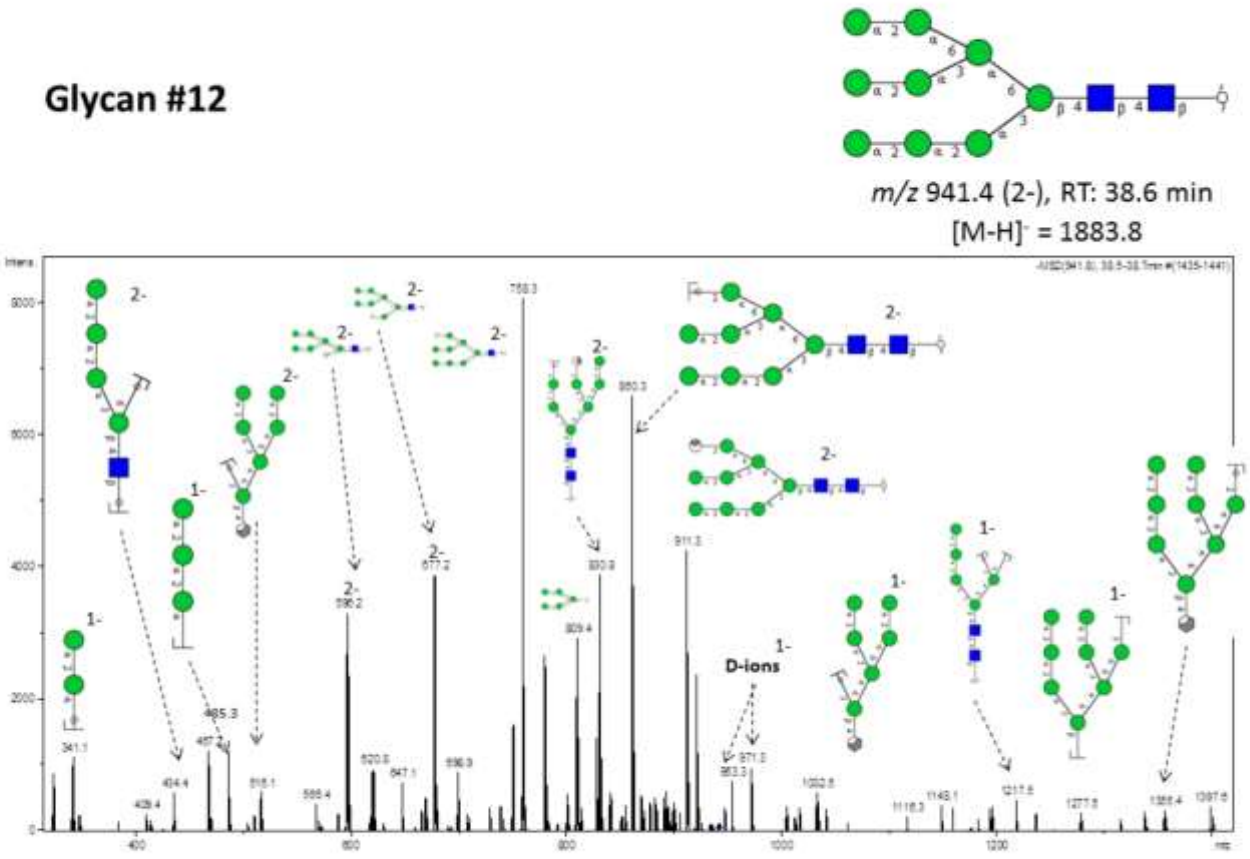
Glycan #11b

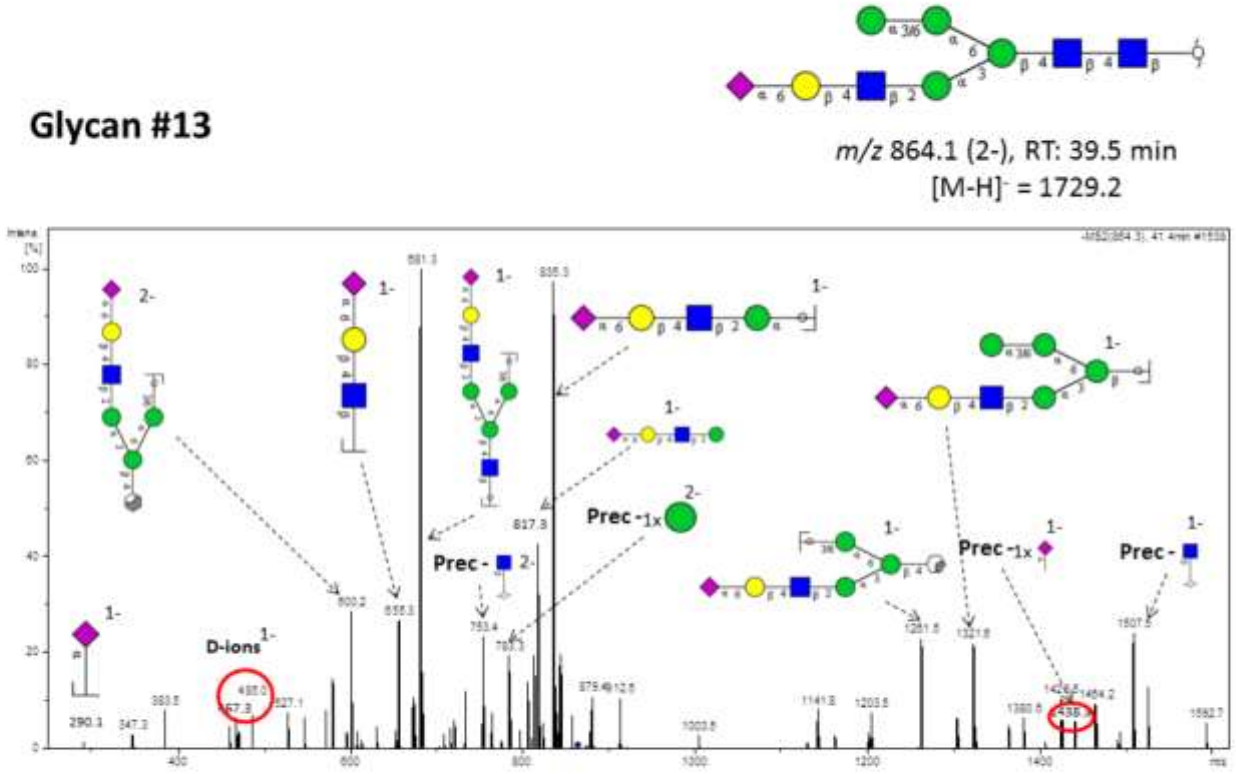


m/z 860.4 (2-), RT: 41.6 min

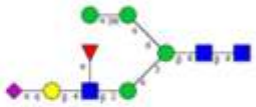
$[M\text{-}H]^-$ = 1721.8



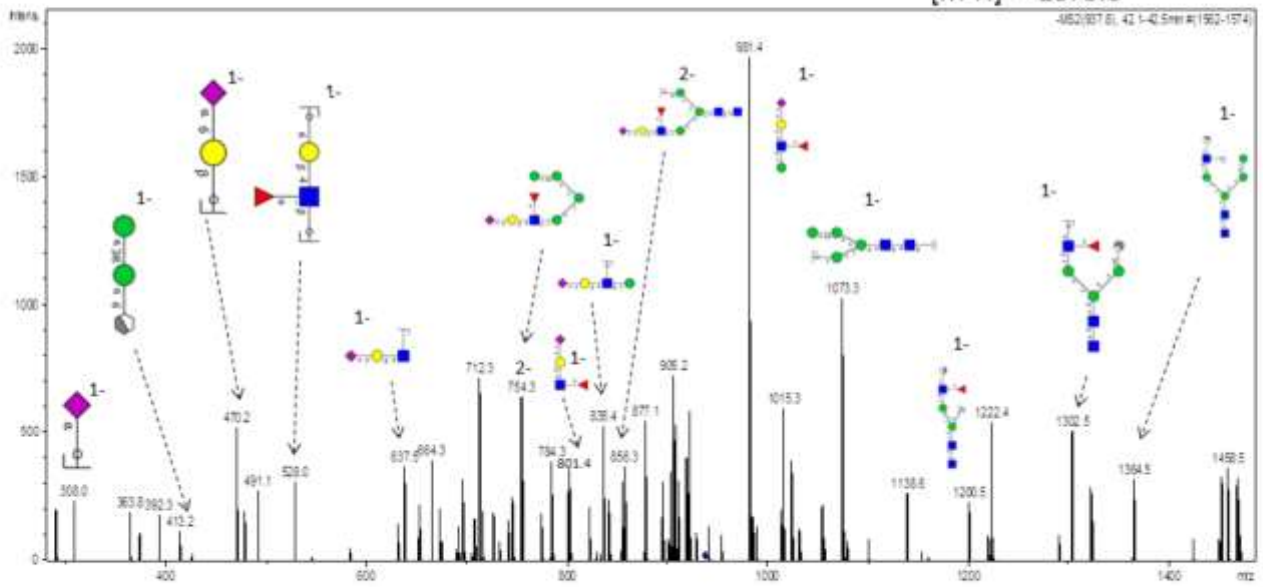

Glycan #12
m/z 941.4 (2-), RT: 38.6 min
[M-H]⁻ = 1883.8
D-ions

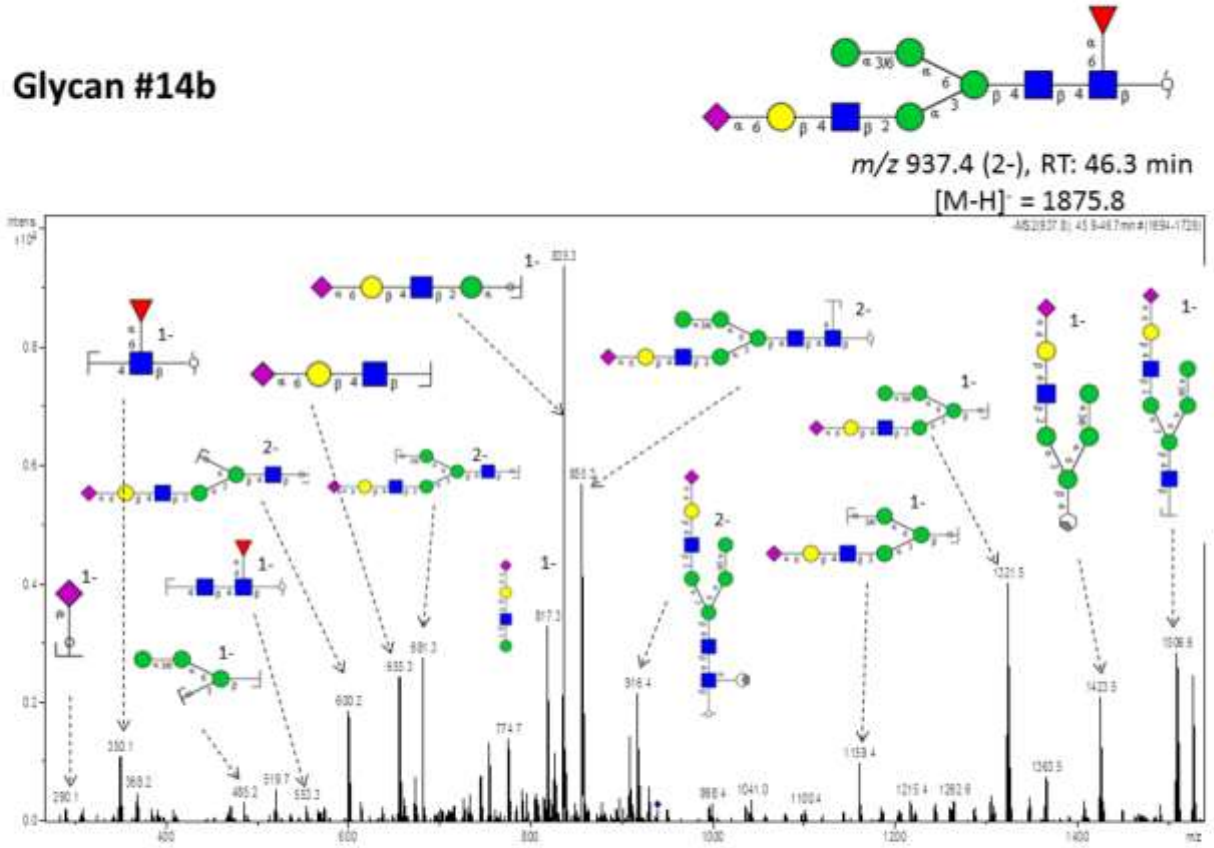

Glycan #13
m/z 864.1 (2-), RT: 39.5 min
[M-H]- = 1729.2
D-ions
Prec -1x
Prec -1x
Prec -

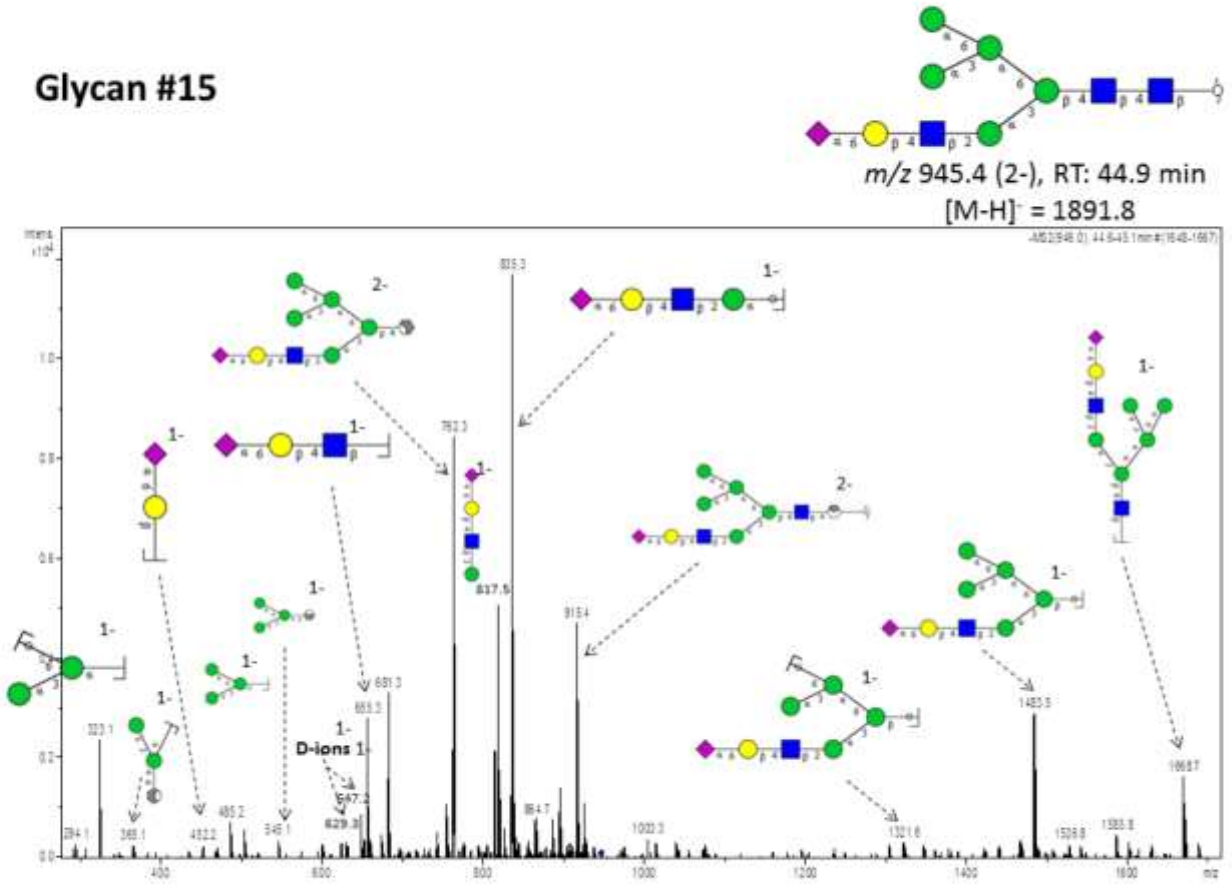
Glycan #14a



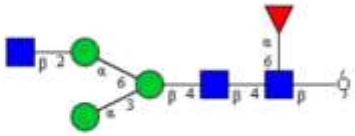
m/z 937.4 (2-), RT: 42.3 min
$[M-H]^-$ = 1875.8



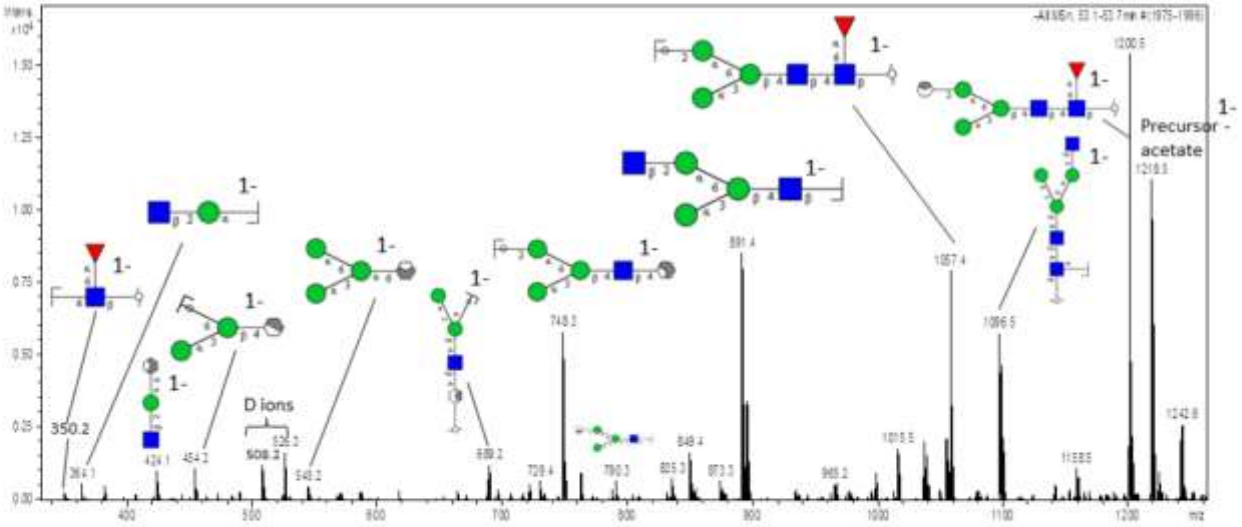

Glycan #14b
m/z 937.4 (2-), RT: 46.3 min
[M-H]⁻ = 1875.8


Glycan #15
m/z 945.4 (2-), RT: 44.9 min
[M-H]⁻ = 1891.8
D-ions 1-

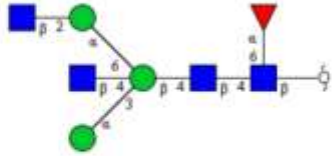
Glycan #16



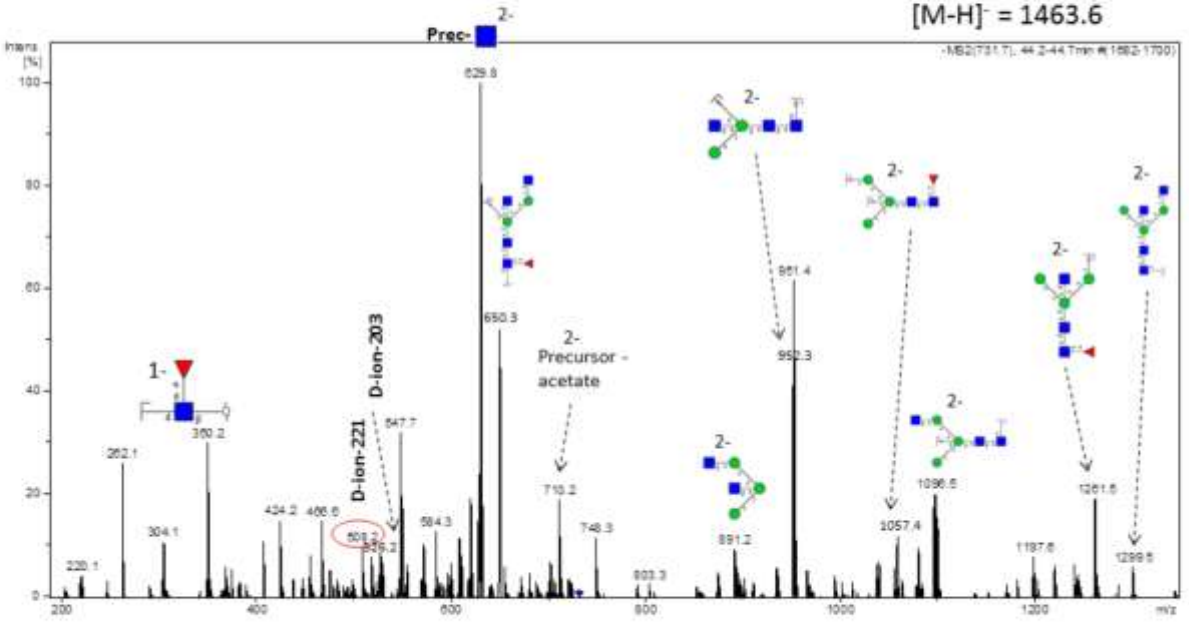
m/z 1260.4 (1-), RT: 53.3 min
$[M-H]^-$ = 1235.4

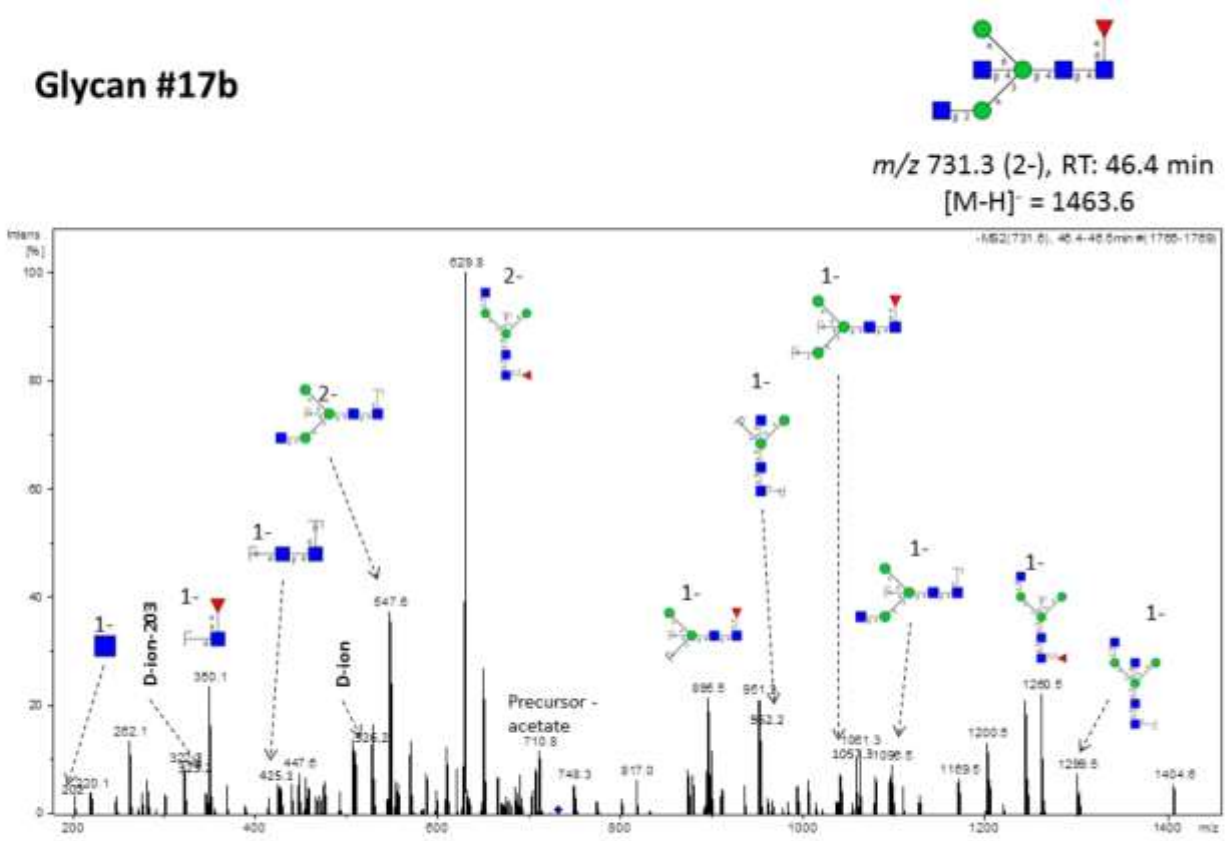


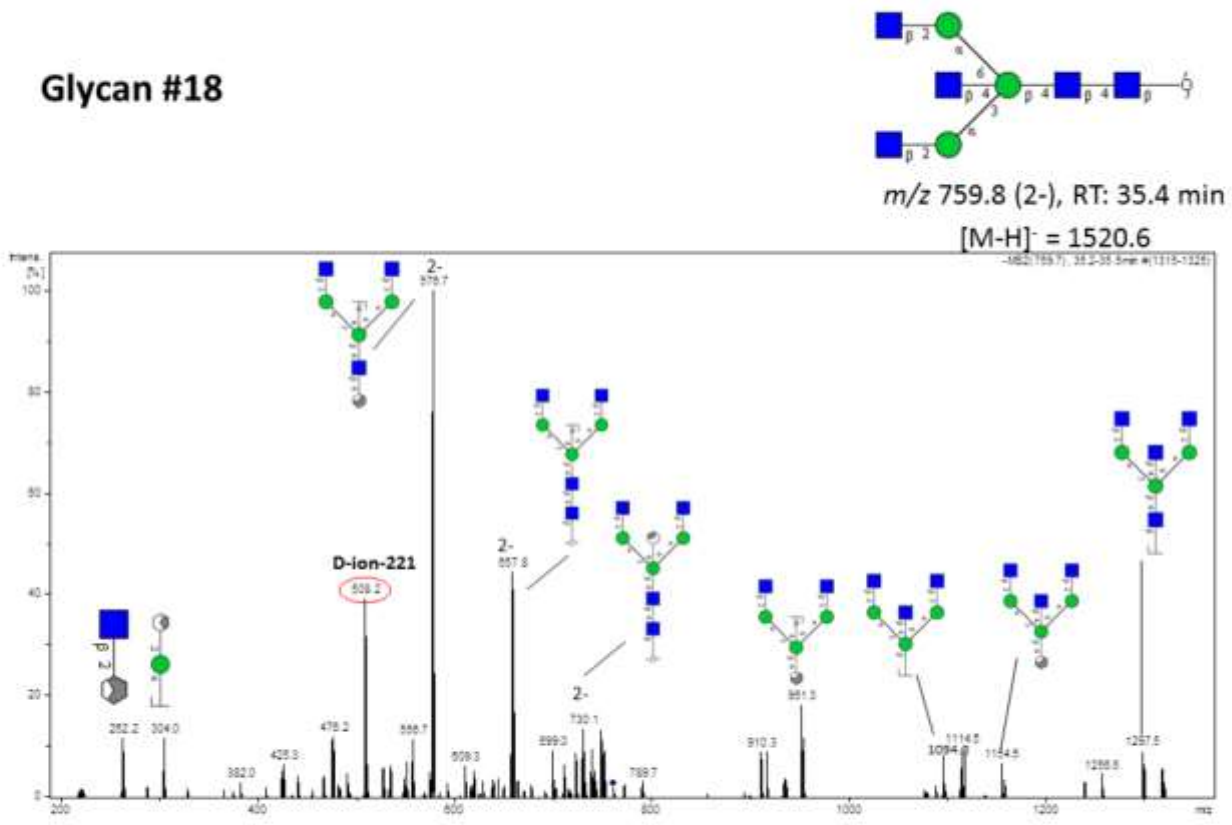
Glycan #17a



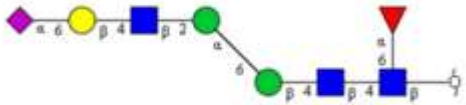
m/z 731.3 (2-), RT: 44.4 min
$[M\text{-}H]^-$ = 1463.6




Glycan #17b
m/z 731.3 (2-), RT: 46.4 min
[M-H]⁻ = 1463.6
D-ion-203
D-ion
Precursor - acetate
Intens. [%]
m/z

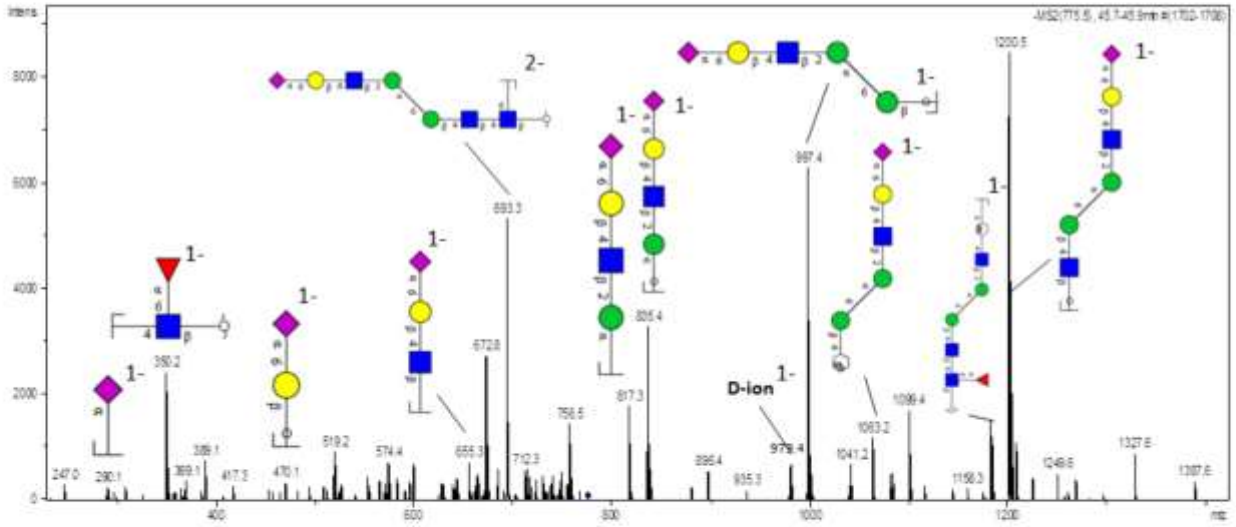

Glycan #18
m/z 759.8 (2-), RT: 35.4 min
[M-H]- = 1520.6
D-ion-221
2-
2-
2-

Glycan #19

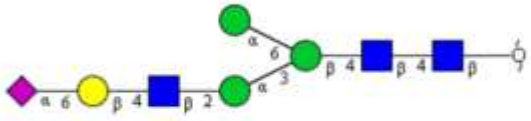


m/z 775.4 (2-), RT: 45.8 min

$[M\text{-}H]^-$ = 1551.8

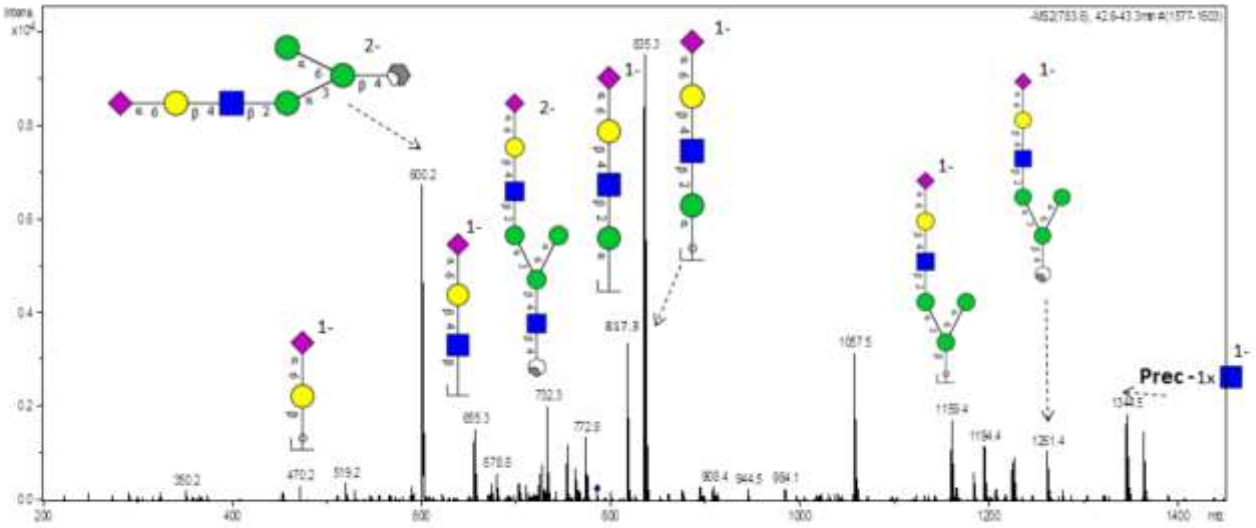


Glycan #20

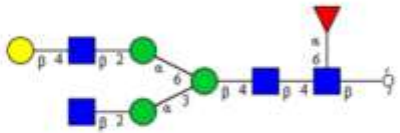


m/z 783.4 (2-), RT: 43.0 min

$[M\text{-}H]^-$ = 1567.8

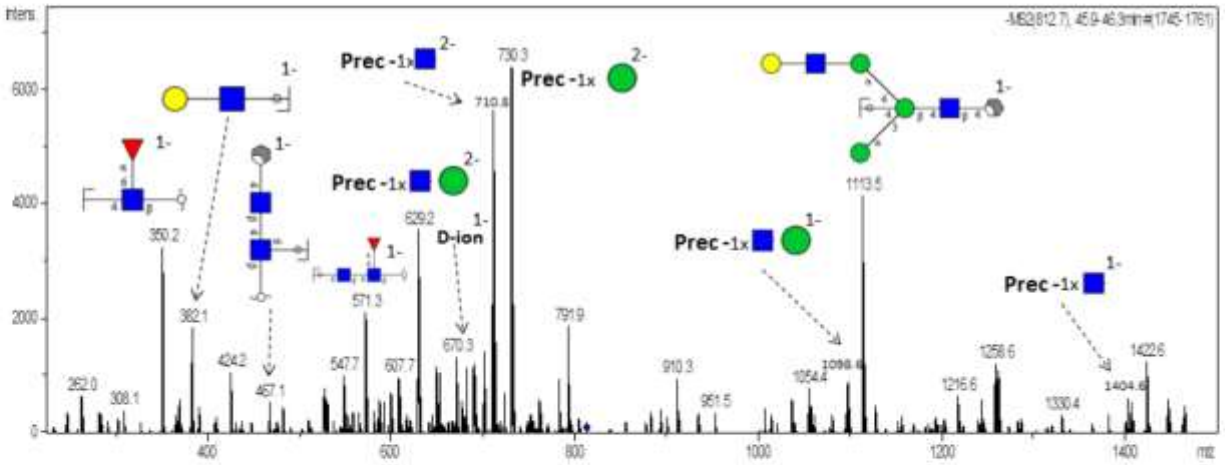


Glycan #21a

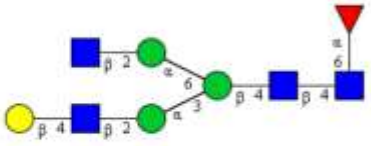


m/z 812.4 (2-), RT: 46.1 min

$[M\text{-}H]^-$ = 1625.8

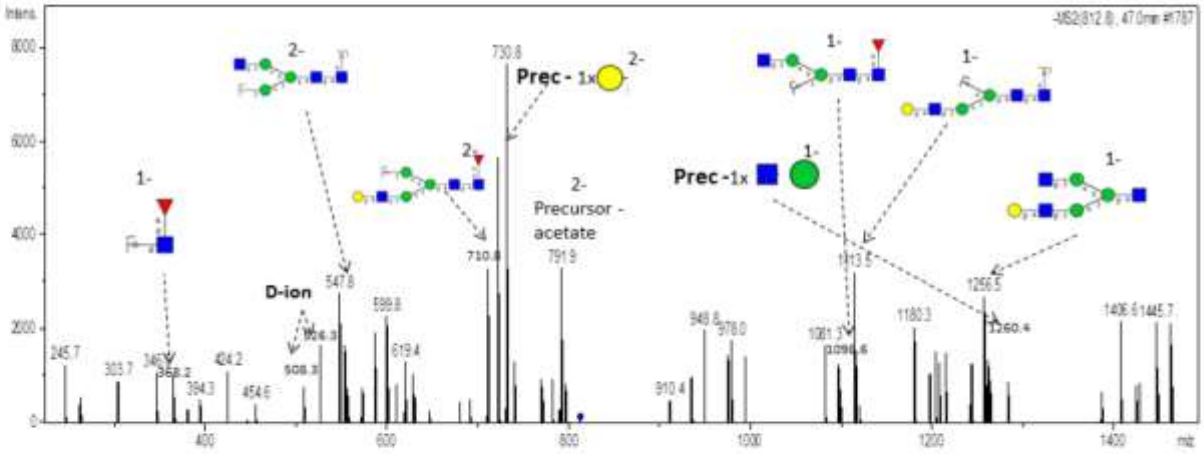


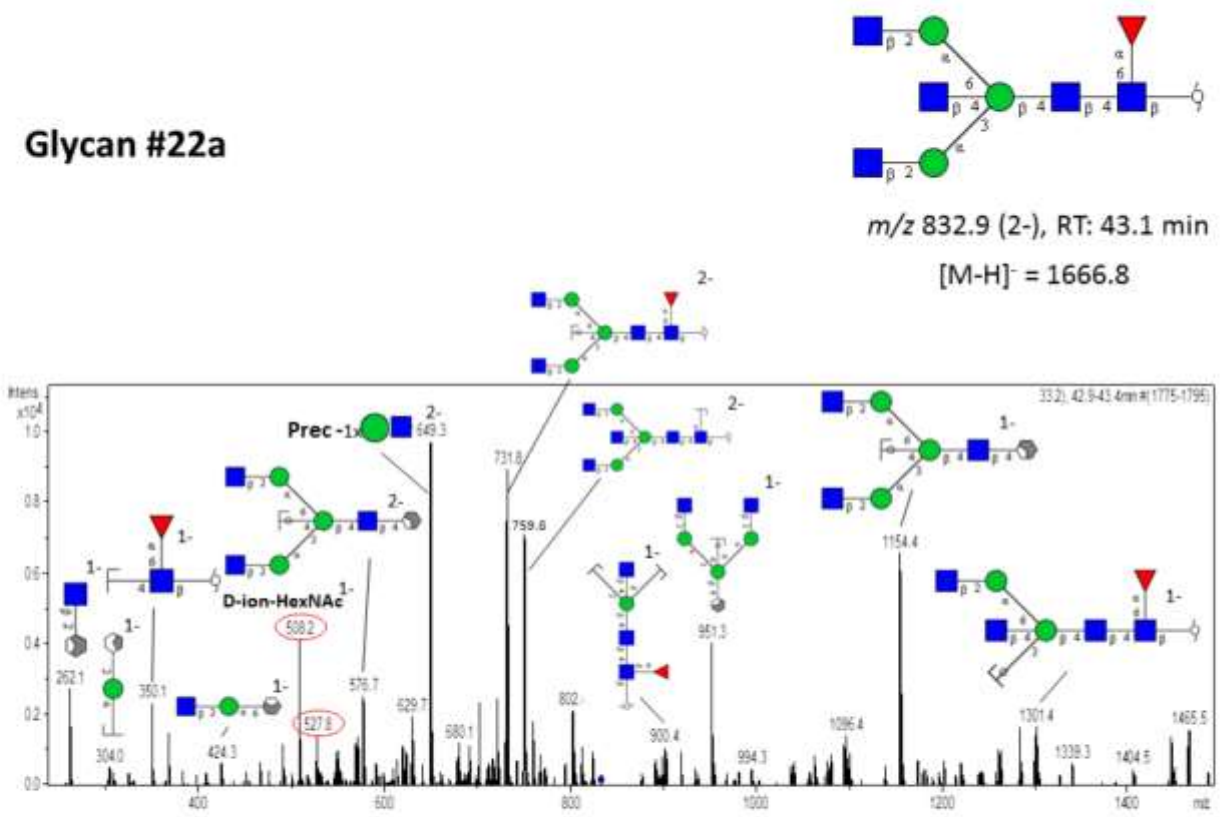
Glycan #21b



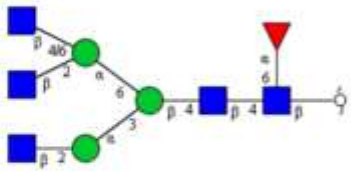
m/z 812.4 (2-), RT: 47.0 min

$[M-H]^-$ = 1625.8



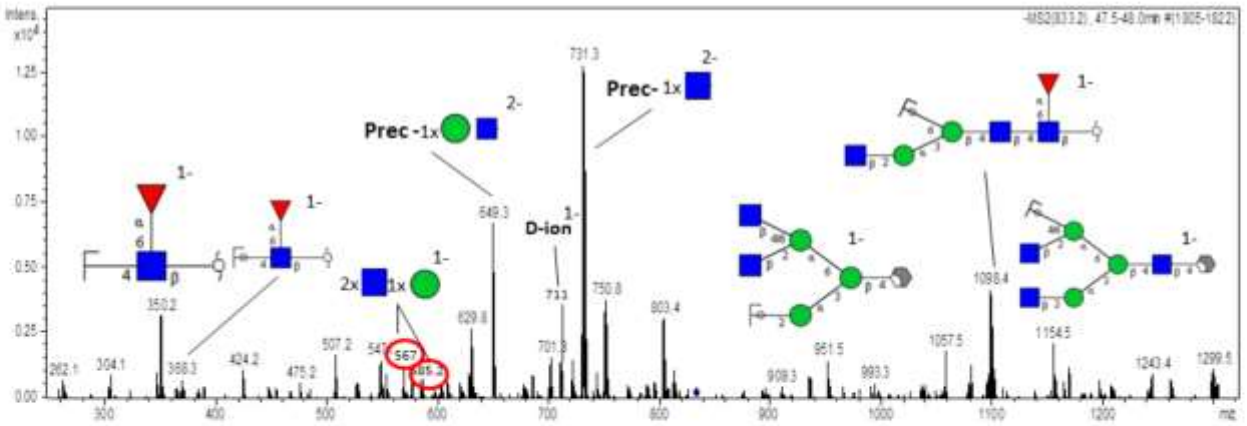

Glycan #22a
m/z 832.9 (2-), RT: 43.1 min
[M-H]- = 1666.8
Prec -1x
2-
649.3
D-ion-HexNAc
508.2
527.8
759.6
731.8
1154.4
951.3
900.4
1301.4
1465.5
1096.4
994.3
1338.3
1404.5
576.7
629.7
680.1
424.3
350.1
304.0
262.1
1-
2-
Intens.
x10^4
m/z

Glycan #22b

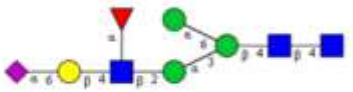


m/z 832.9 (2-), RT: 47.7 min

$[M-H]^-$ = 1666.8

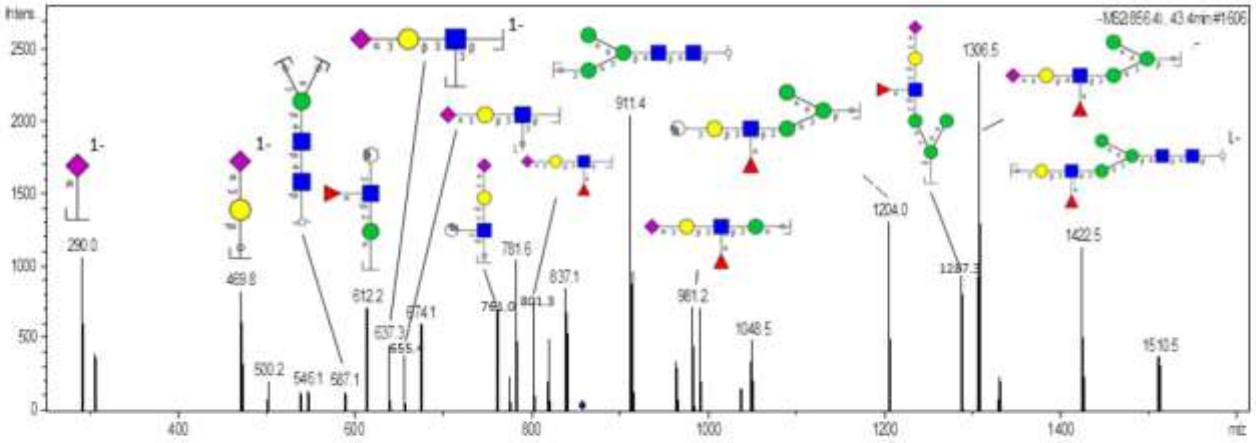


Glycan #23a

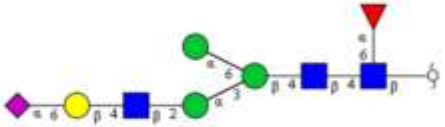


m/z 856.4 (2-), RT: 43.4 min

$[M-H]^-$ = 1713.8

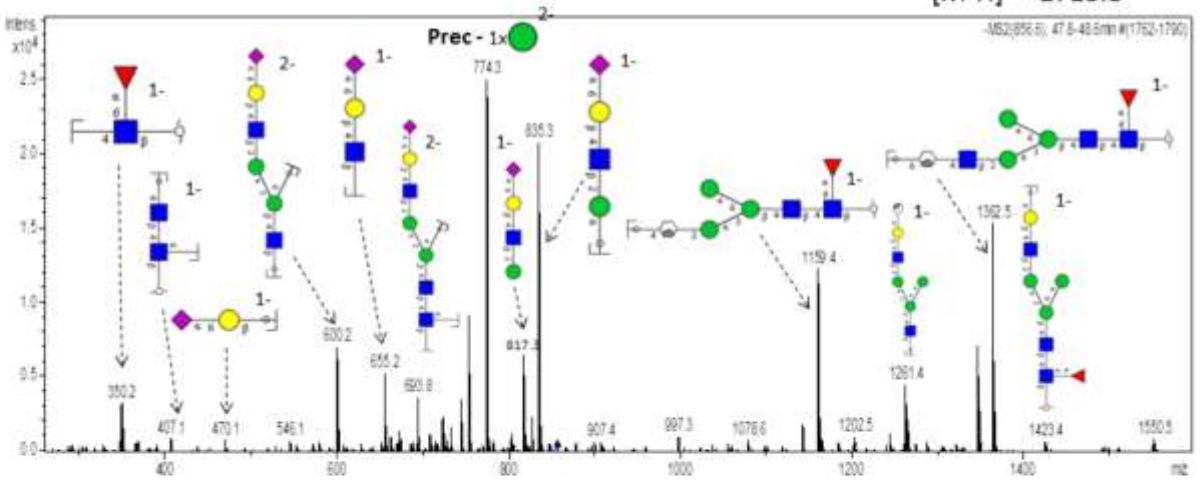


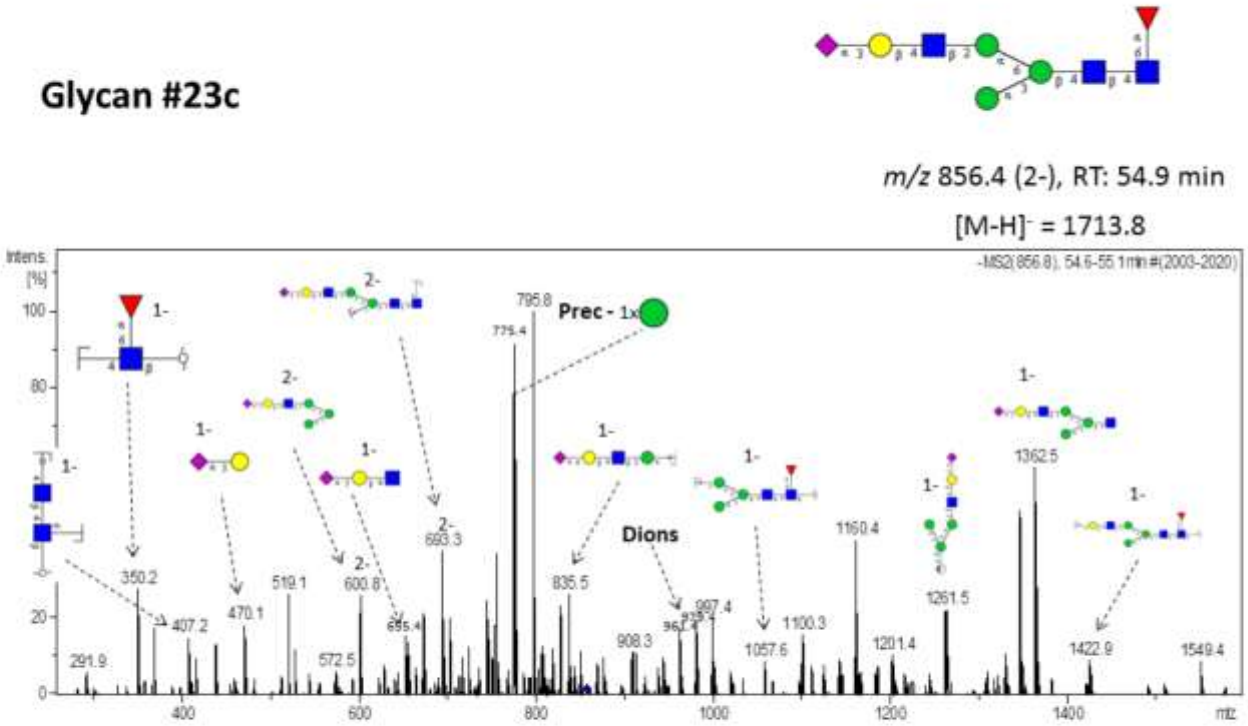
Glycan #23b



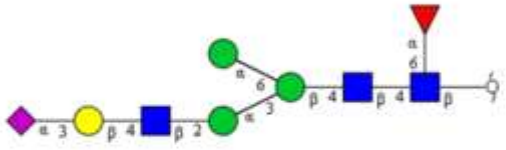
m/z 856.4 (2-), RT: 48.2 min

$[M\text{-}H]^-$ = 1713.8



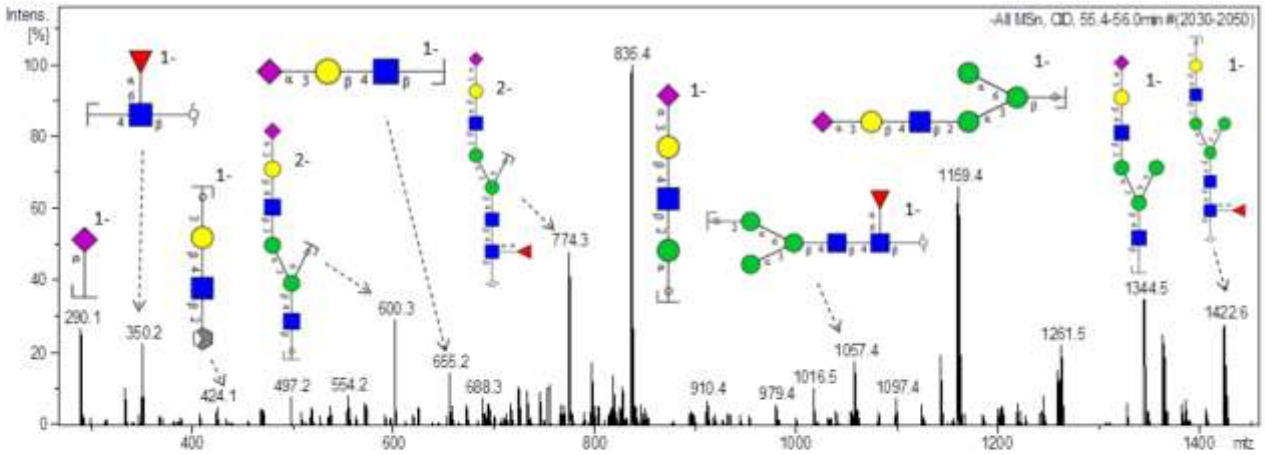

Glycan #23c
m/z 856.4 (2-), RT: 54.9 min
[M-H]- = 1713.8
Intens. [%]
-MS2(856.8), 54.6-55.1min #(2003-2020)
Prec - 1x
Dions
350.2
407.2
470.1
519.1
572.5
600.8
693.3
775.4
795.8
835.5
908.3
997.4
1057.6
1100.3
1160.4
1201.4
1261.5
1362.5
1422.9
1549.4
291.9
mz

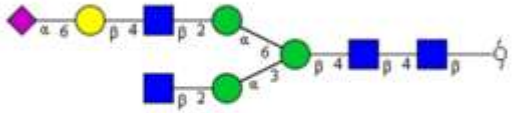
Glycan #23d



m/z 856.4 (2-), RT: 55.7 min

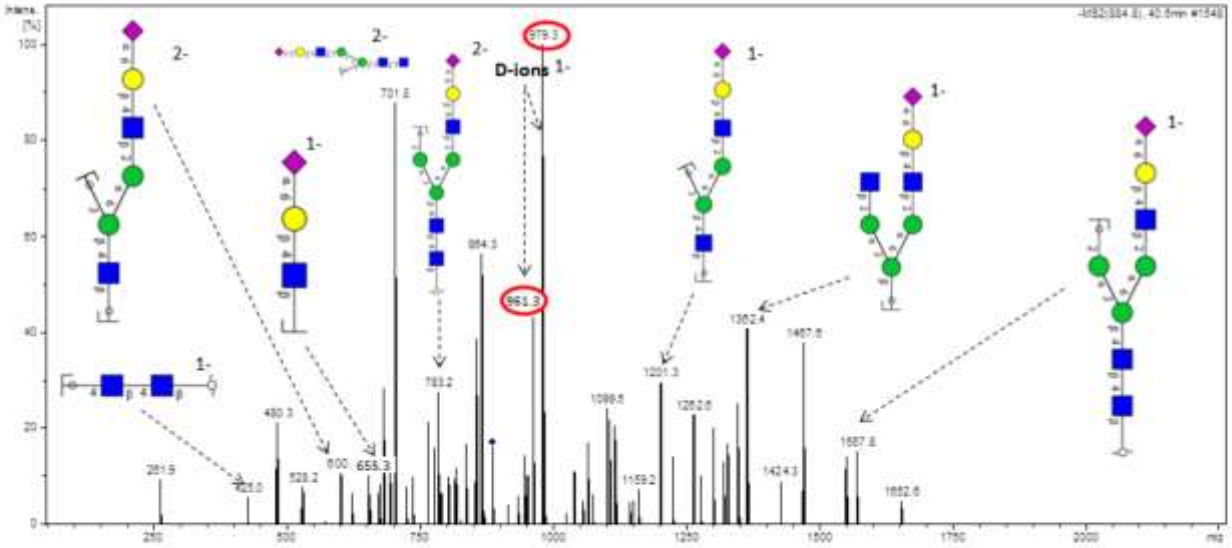
$[M-H]^-$ = 1713.8

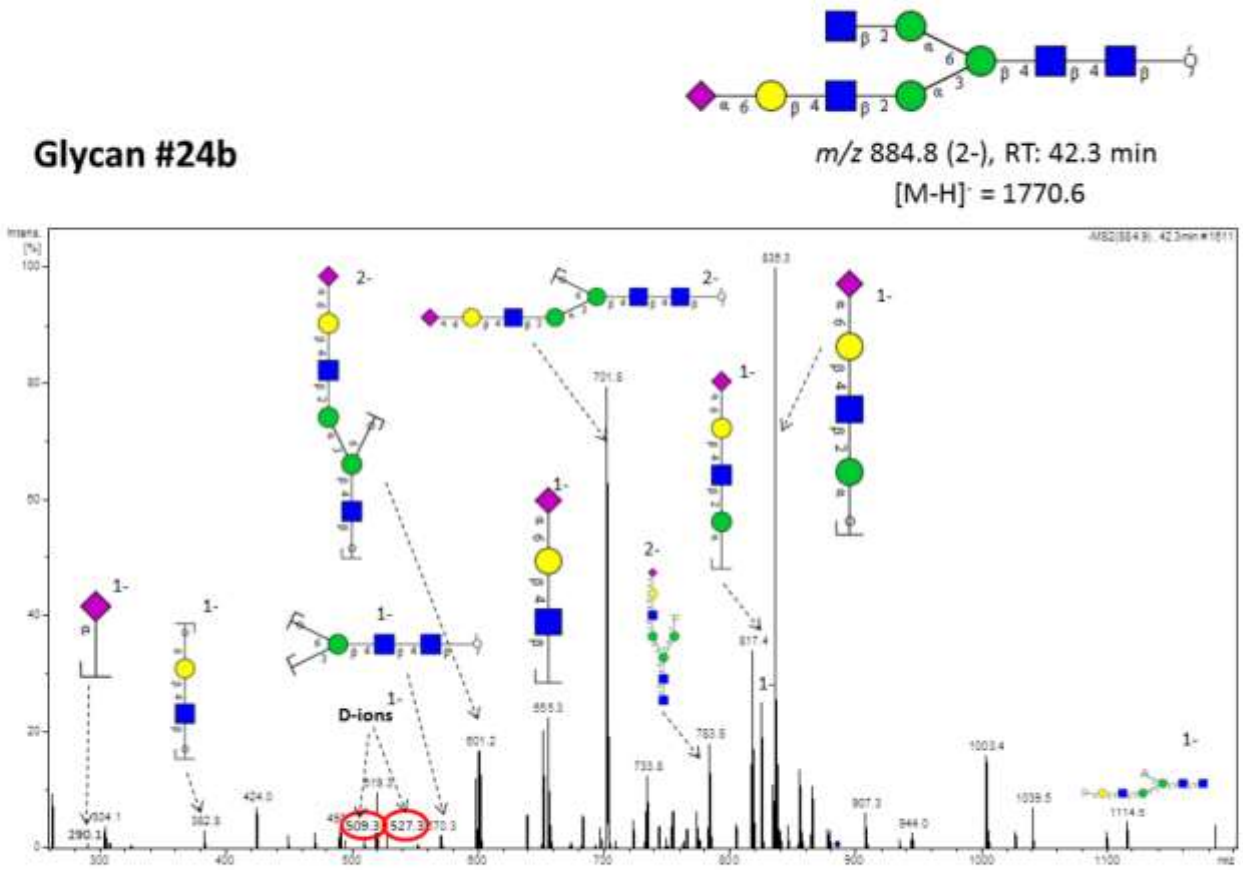


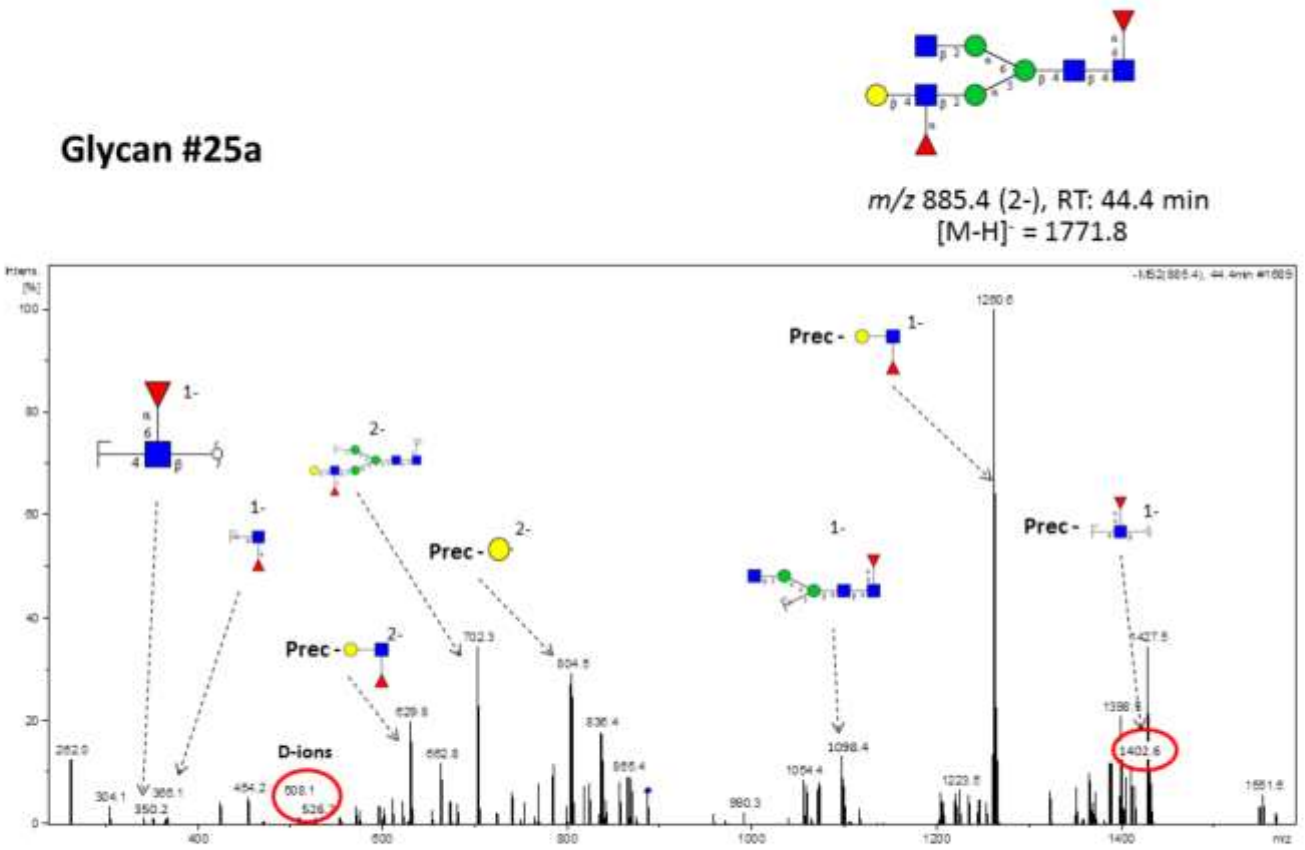


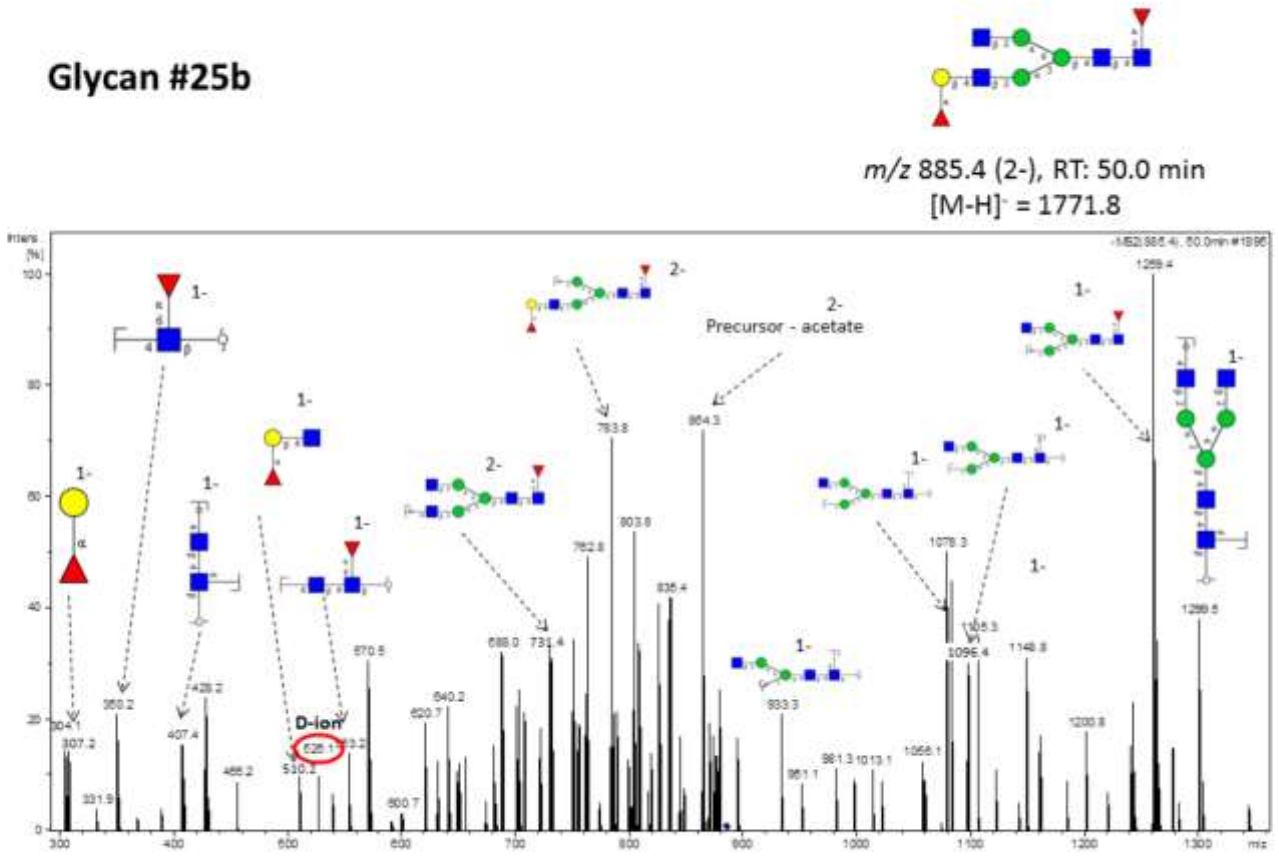
Glycan #24a

m/z 884.8 (2-), RT: 40.6min
$[M-H]^-$ = 1770.6

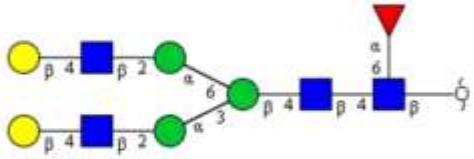



Glycan #24b
m/z 884.8 (2-), RT: 42.3 min
[M-H]⁻ = 1770.6
D-ions
509.3
527.3

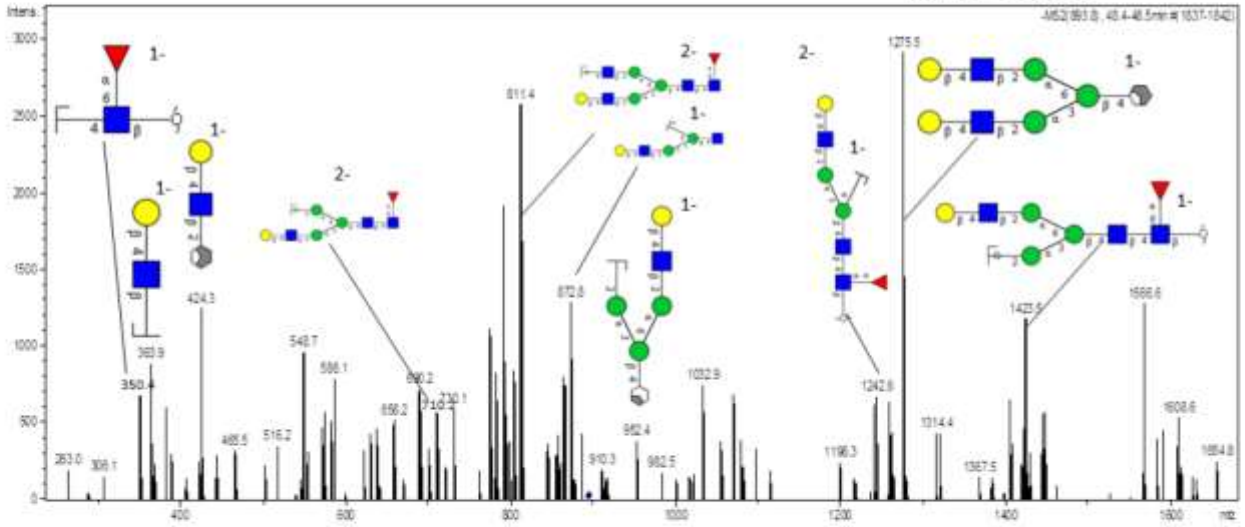

Glycan #25a
m/z 885.4 (2-), RT: 44.4 min
[M-H]- = 1771.8
Prec -
D-ions
1-
2-


Glycan #25b
m/z 885.4 (2-), RT: 50.0 min
[M-H]⁻ = 1771.8
Precursor - acetate
D-ion

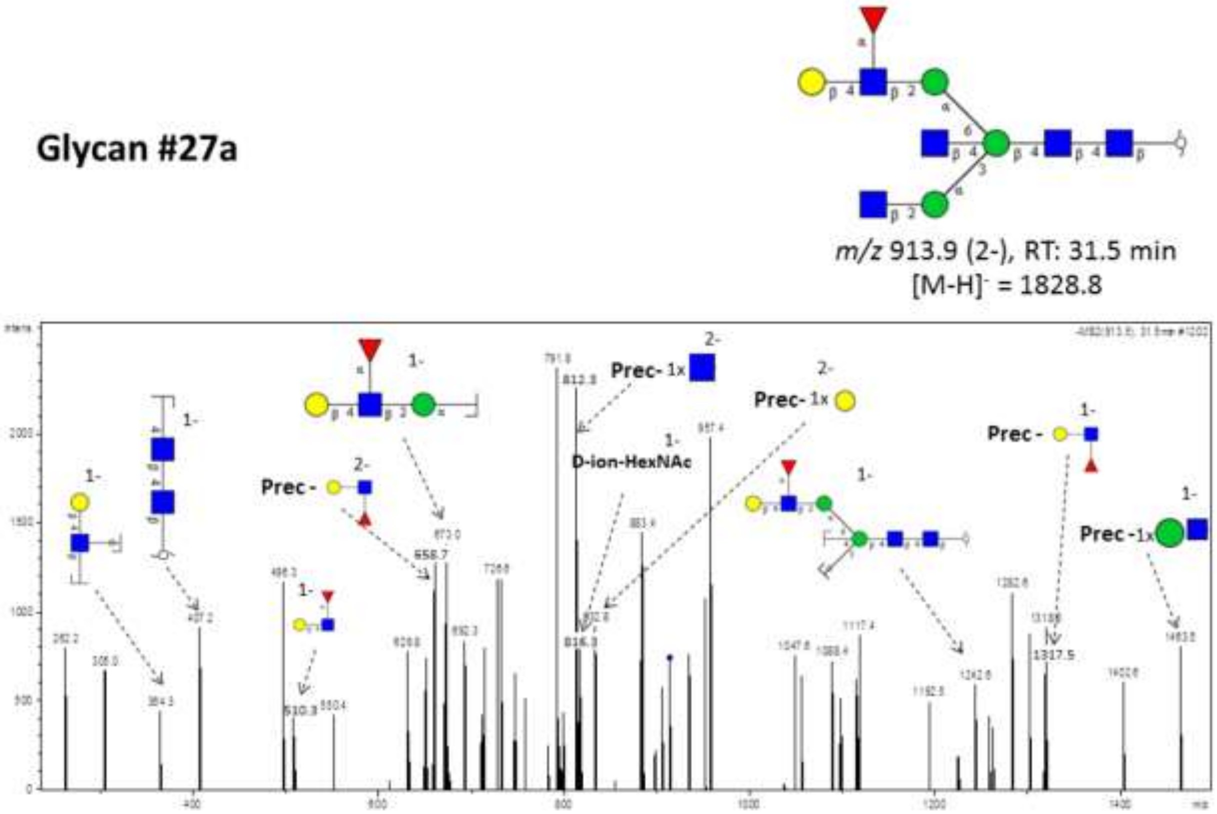
Glycan #26

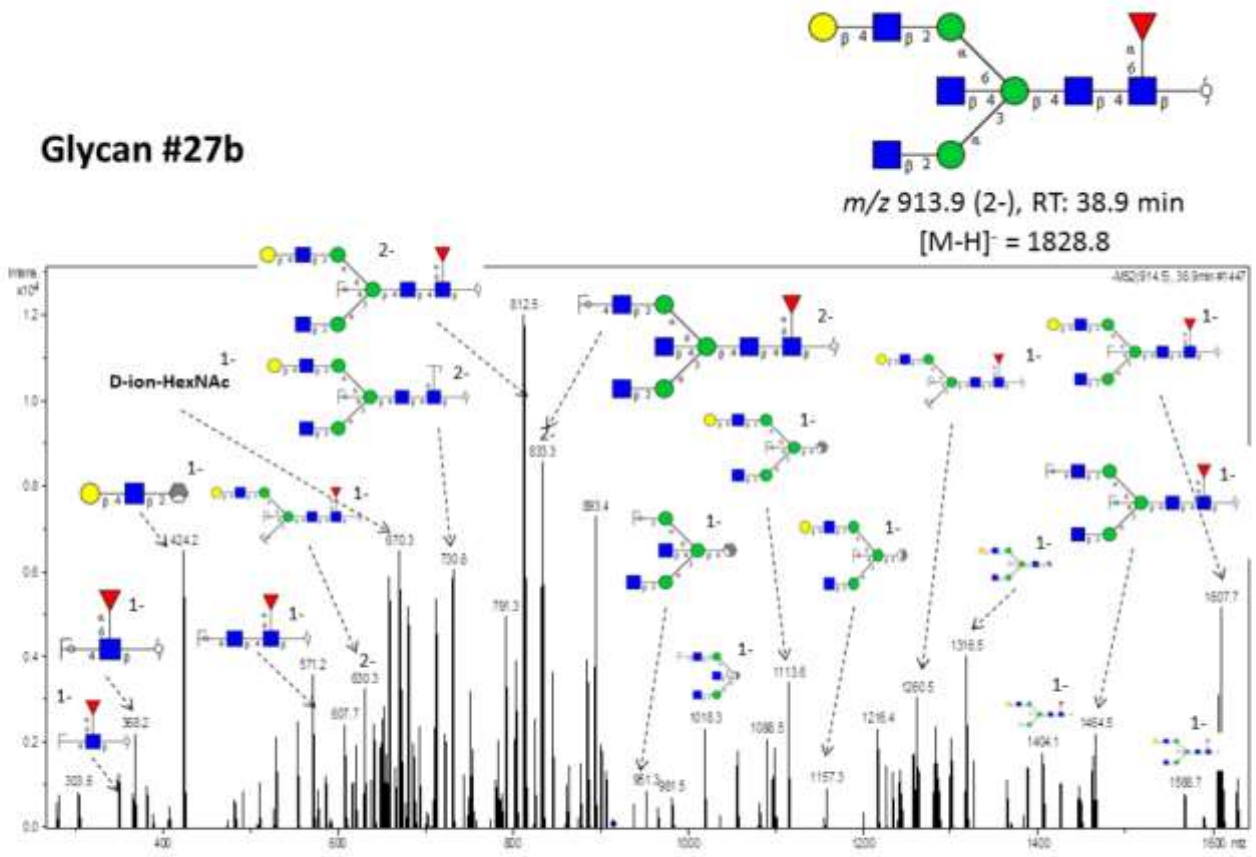


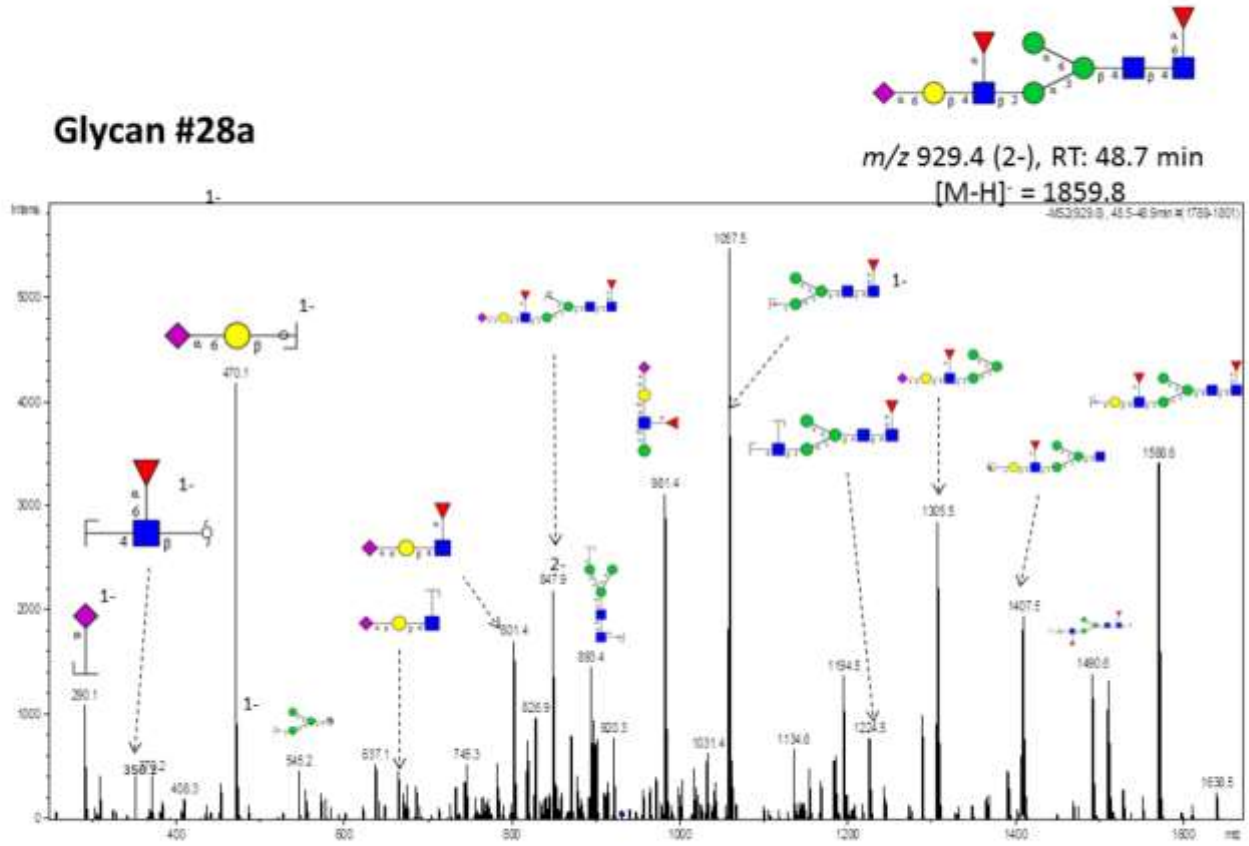
m/z 893.4 (2-), RT: 48.4 min
$[M-H]^-$ = 1787.8

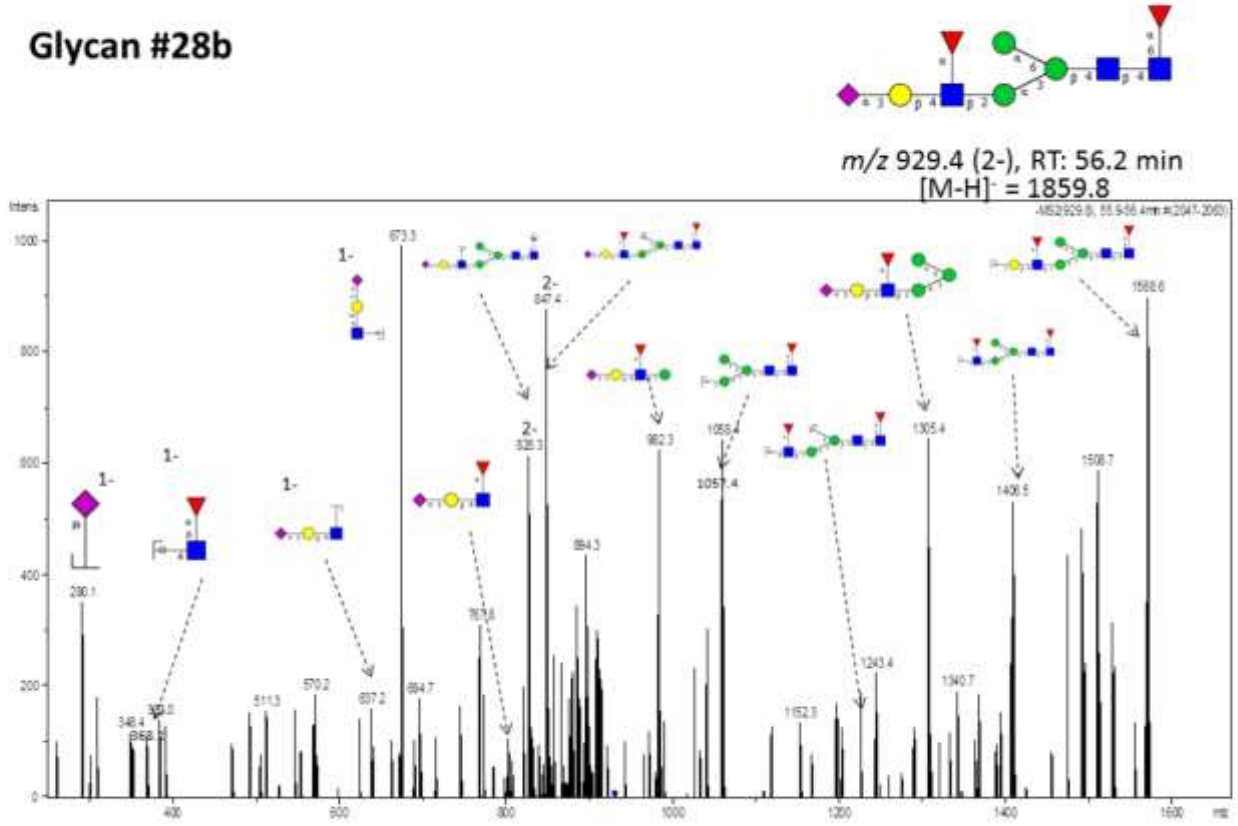


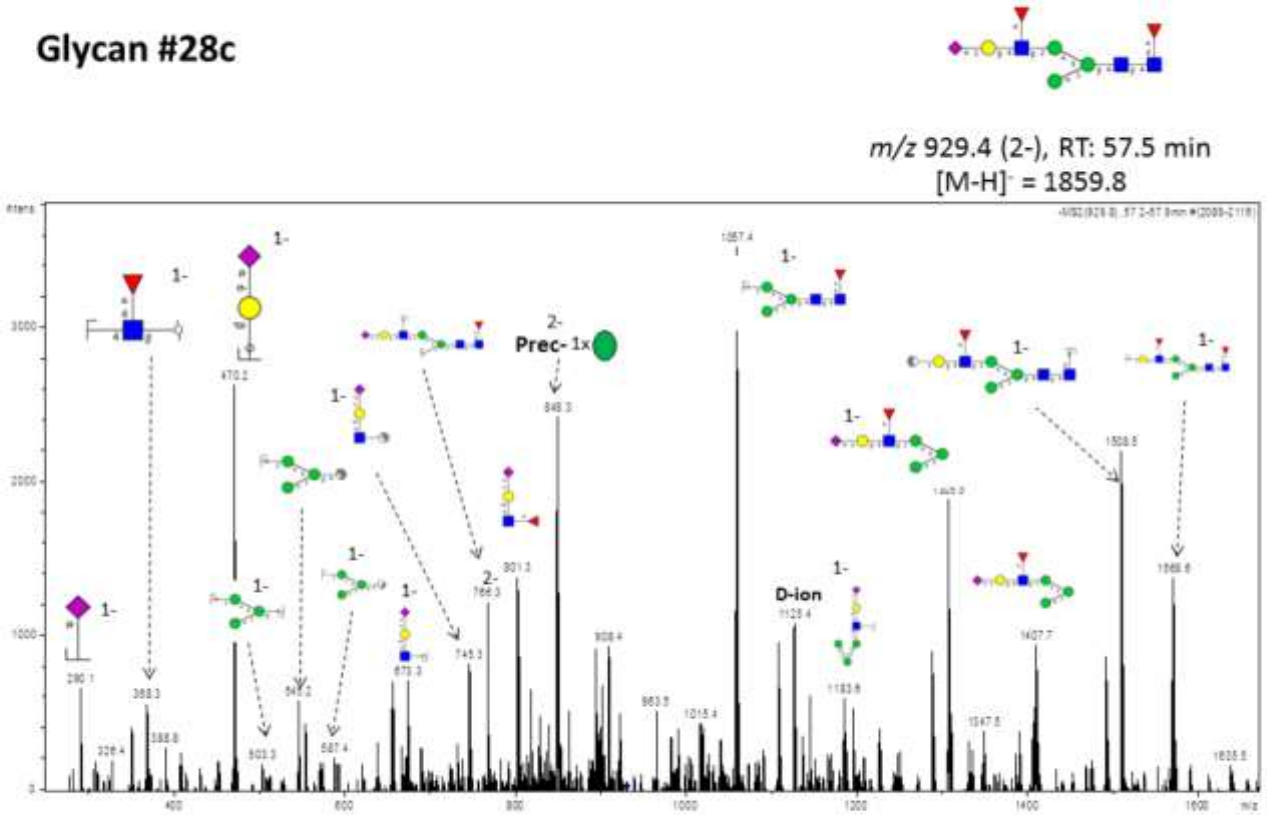
Glycan #27a

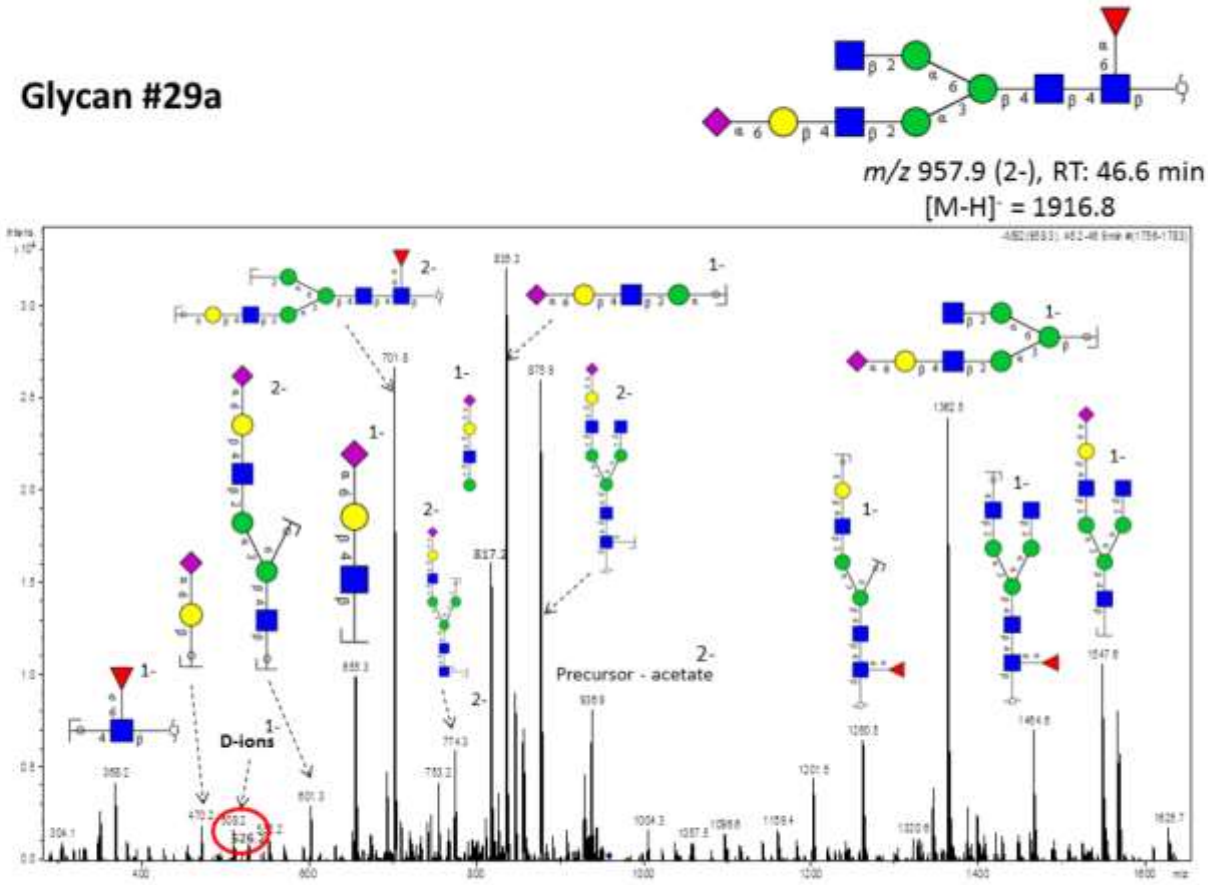

m/z 913.9 (2-), RT: 31.5 min
[M-H]- = 1828.8
Prec- 1x
Prec-
D-ion-HexNAc
658.7
816.3
510.3
1317.5

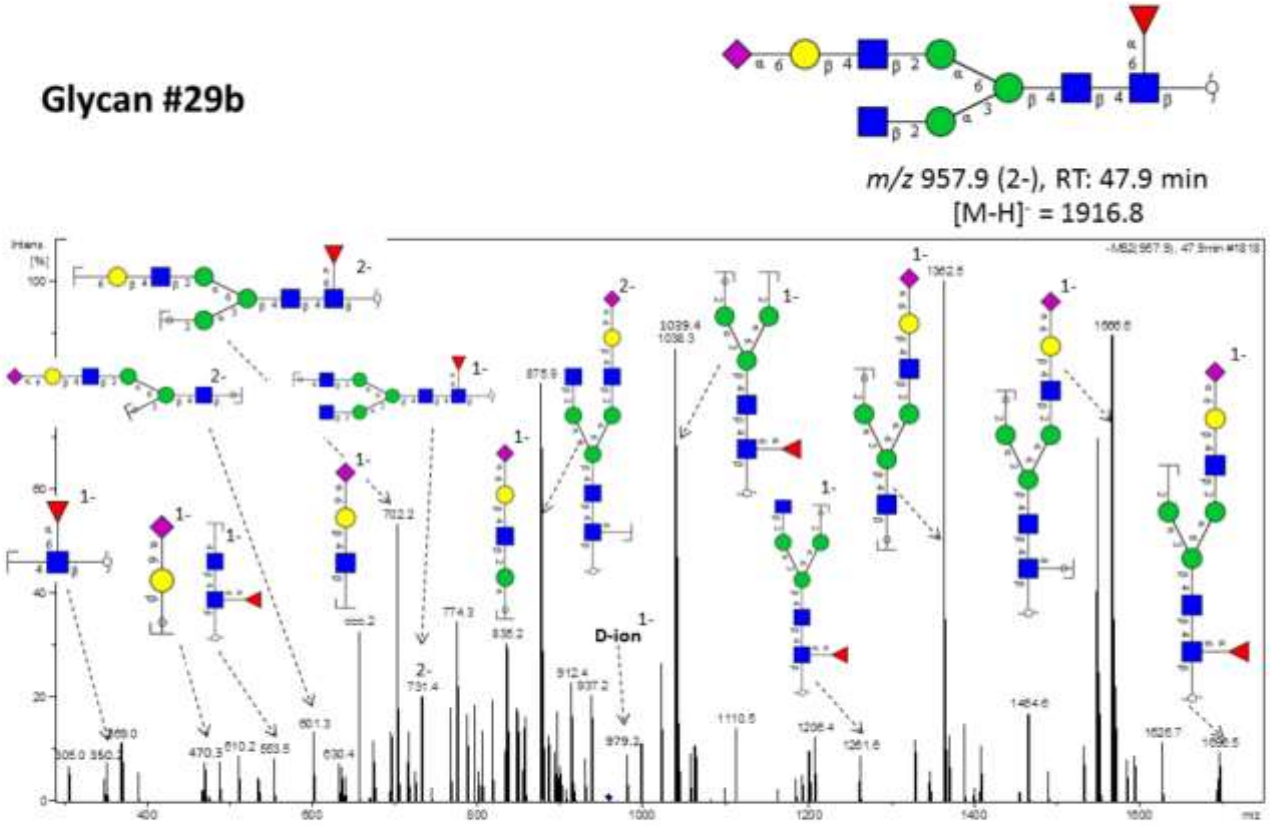

Glycan #27b
m/z 913.9 (2-), RT: 38.9 min
[M-H]- = 1828.8
D-ion-HexNAc

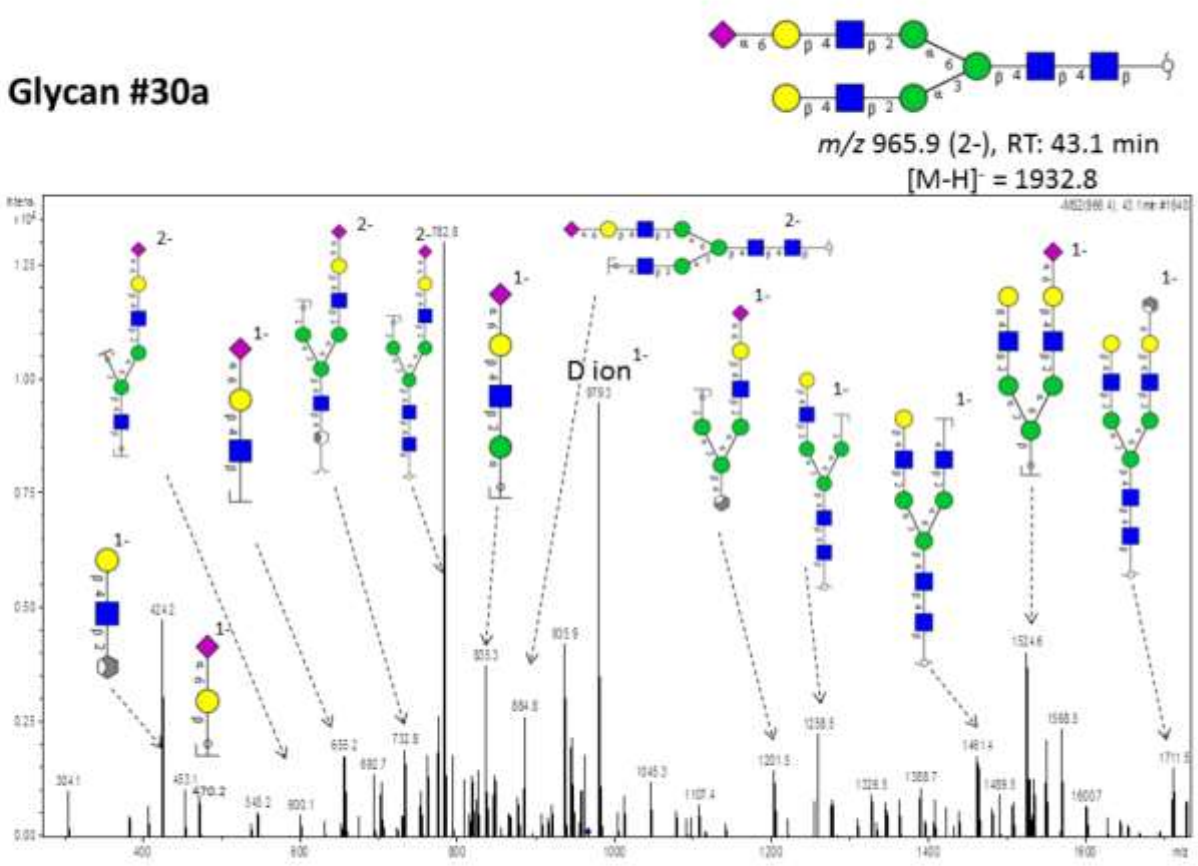

Glycan #28a
m/z 929.4 (2-), RT: 48.7 min
[M-H]⁻ = 1859.8

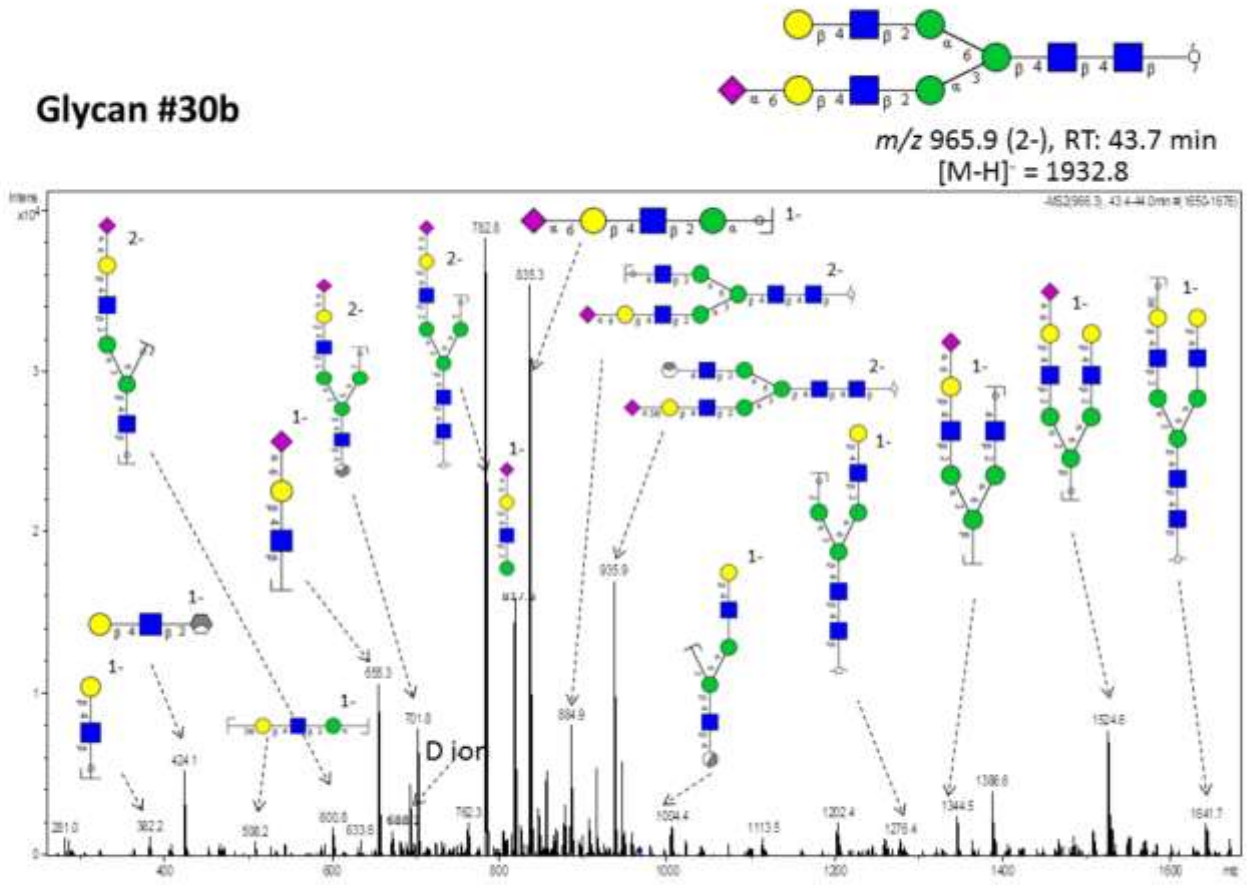

Glycan #28b
m/z 929.4 (2-), RT: 56.2 min
[M-H]- = 1859.8

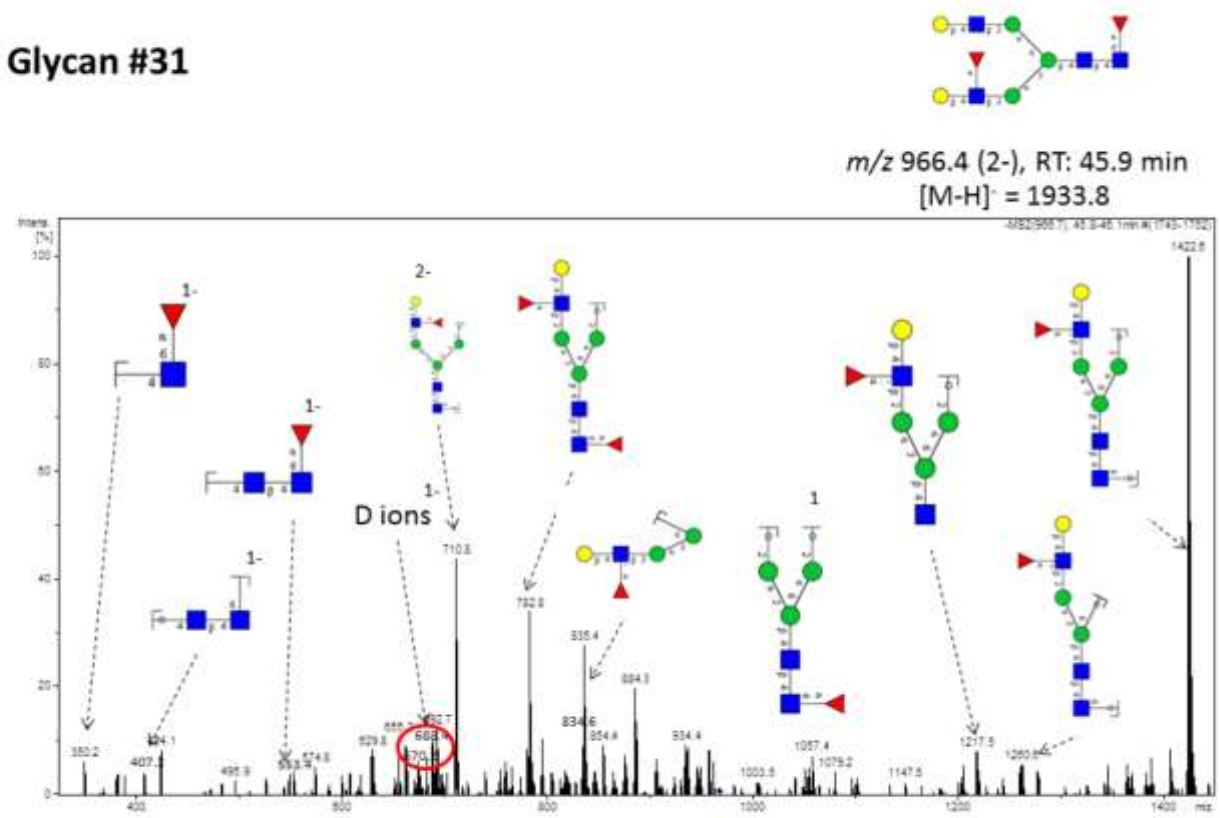

Glycan #28c
m/z 929.4 (2-), RT: 57.5 min
[M-H]⁻ = 1859.8
Prec- 1x
D-ion

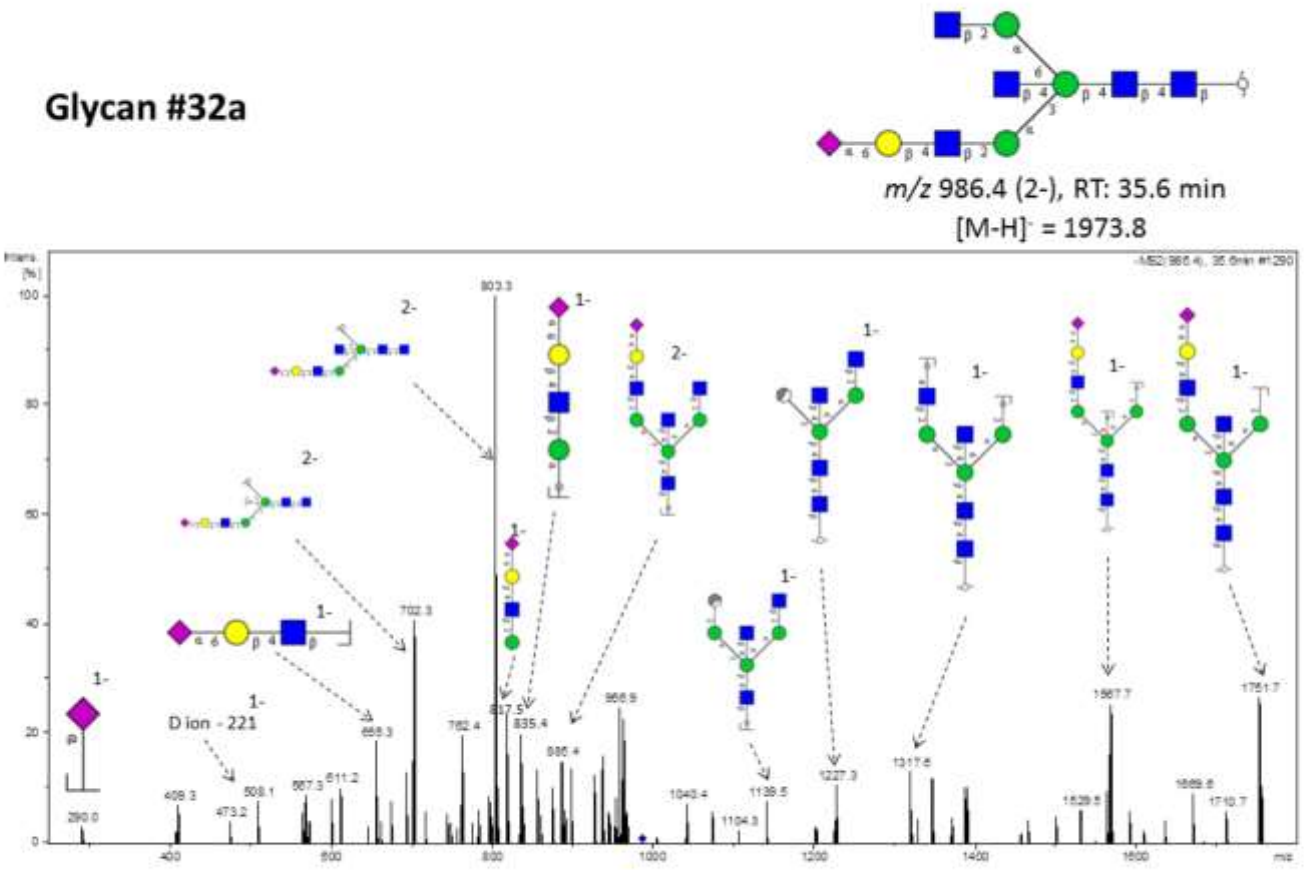

Glycan #29a
m/z 957.9 (2-), RT: 46.6 min
[M-H]- = 1916.8
D-Ions
Precursor - acetate

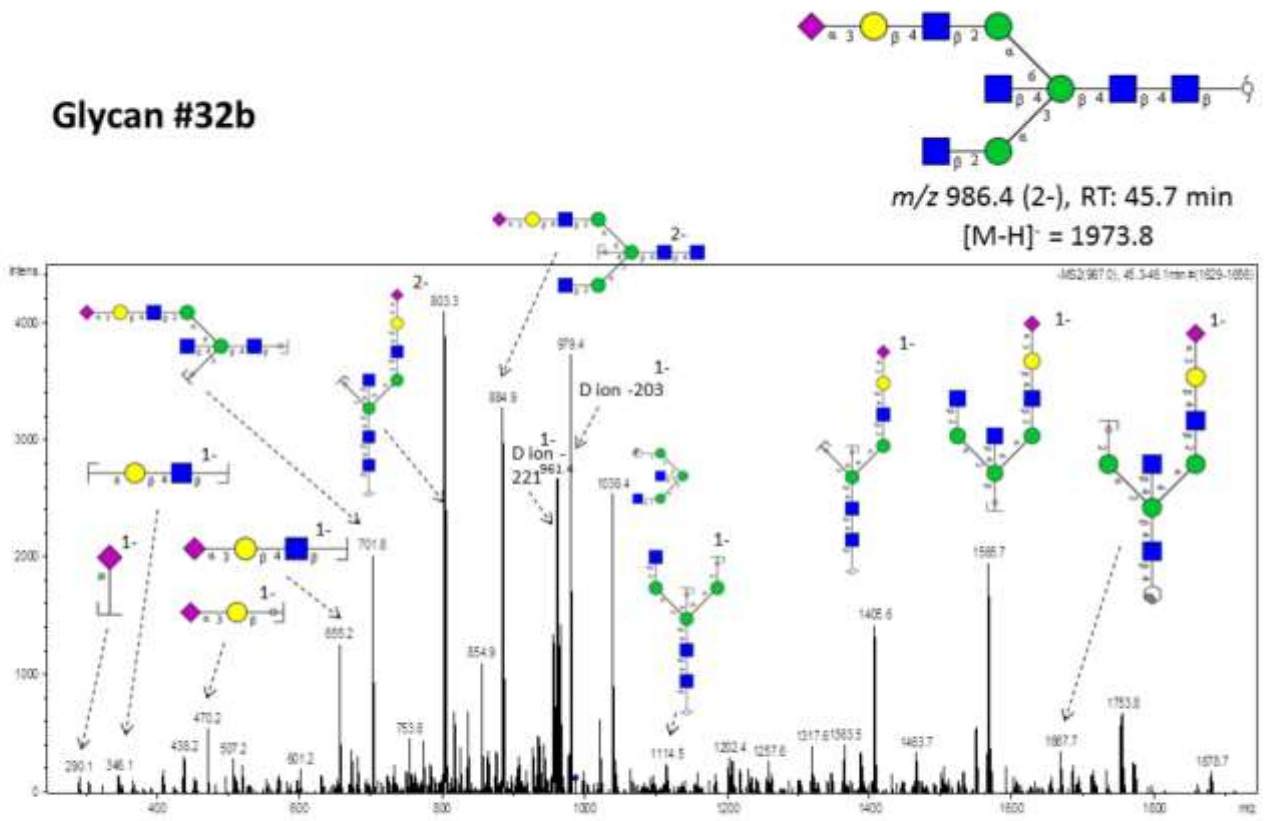

Glycan #29b
m/z 957.9 (2-), RT: 47.9 min
[M-H]- = 1916.8
D-ion

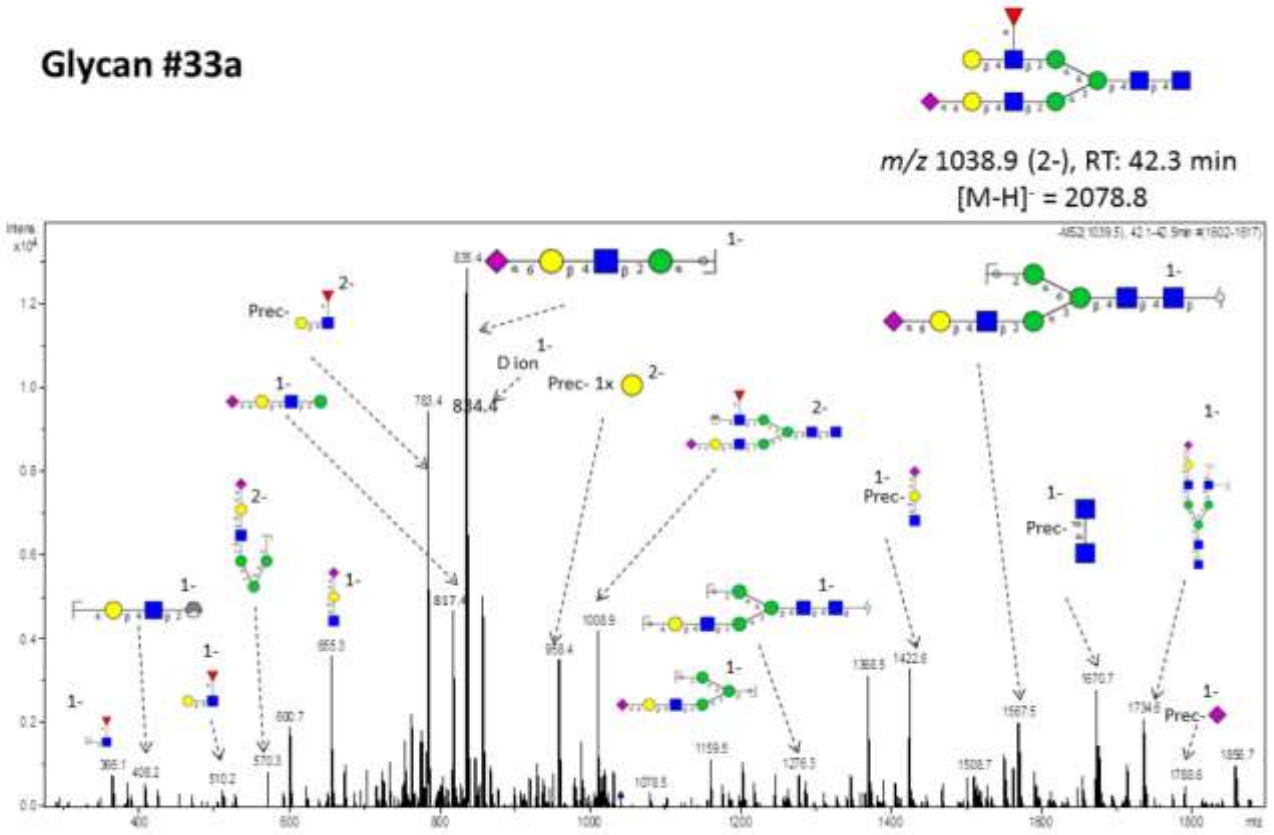

Glycan #30a
m/z 965.9 (2-), RT: 43.1 min
[M-H]- = 1932.8
D ion

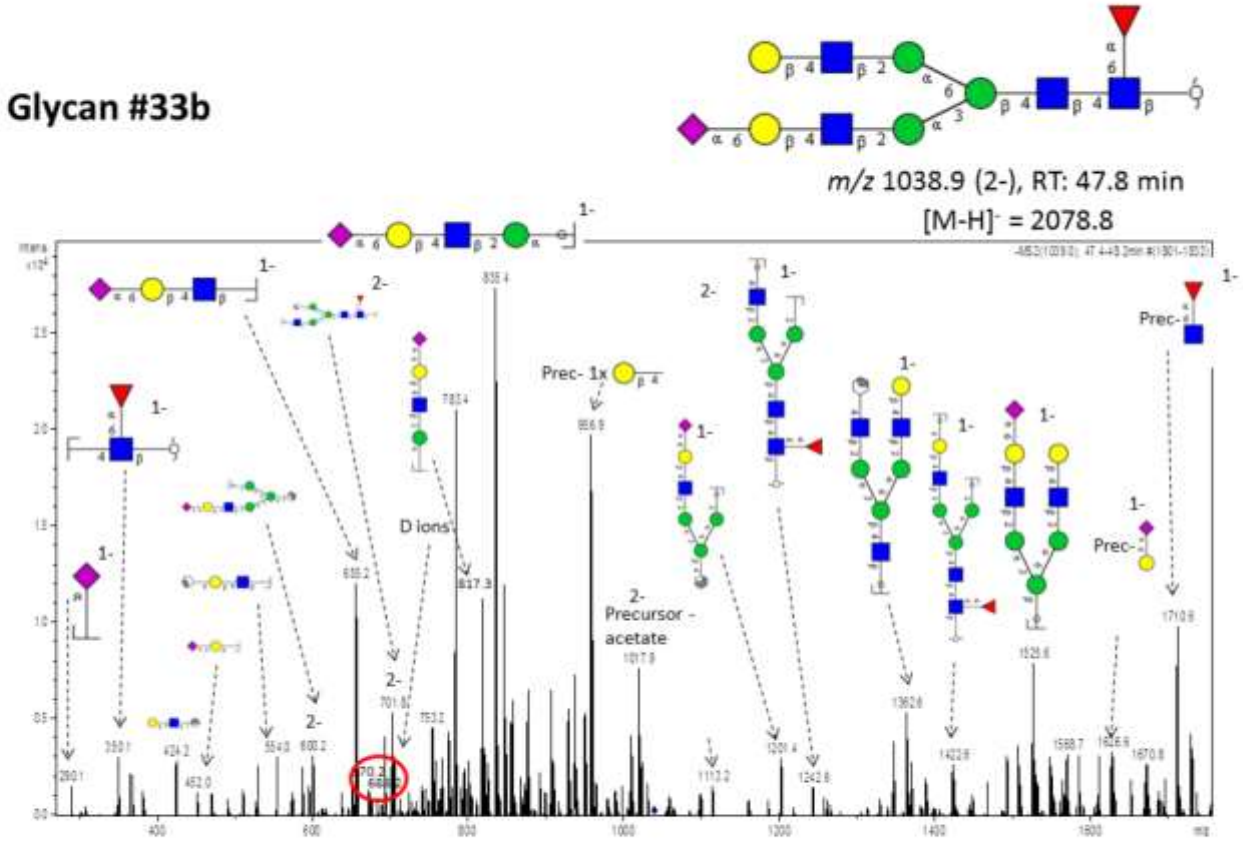

Glycan #30b
m/z 965.9 (2-), RT: 43.7 min
[M-H]- = 1932.8


Glycan #31
m/z 966.4 (2-), RT: 45.9 min
[M-H]⁻ = 1933.8
D ions
1-
2-
1
Intens. [%]
m/z

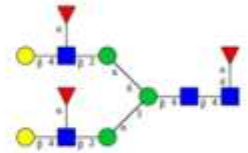

Glycan #32a
m/z 986.4 (2-), RT: 35.6 min
[M-H]⁻ = 1973.8
D ion - 221


Glycan #32b
m/z 986.4 (2-), RT: 45.7 min
[M-H]- = 1973.8
D ion -203
D ion -221
Intens.
m/z


Glycan #33a
m/z 1038.9 (2-), RT: 42.3 min
[M-H]- = 2078.8
Prec-
D ion
Prec- 1x
884.4
1038.9
1422.6
1567.5
1734.6
1856.7

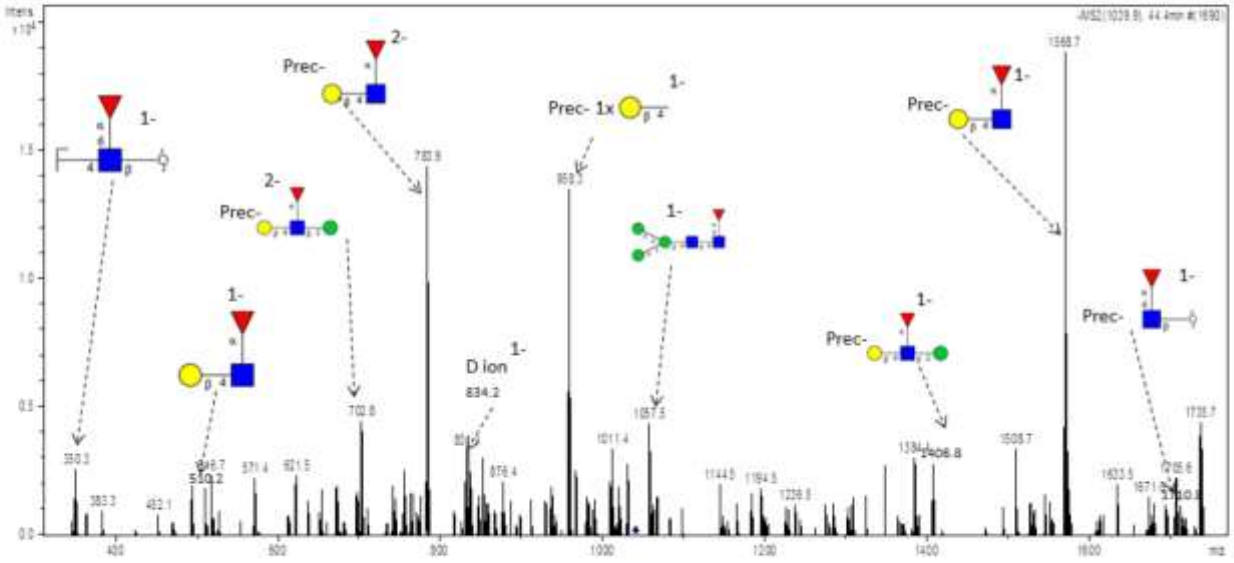

Glycan #33b
m/z 1038.9 (2-), RT: 47.8 min
[M-H]- = 2078.8
D ions
Prec- 1x
2-
Precursor -
acetate
Prec-

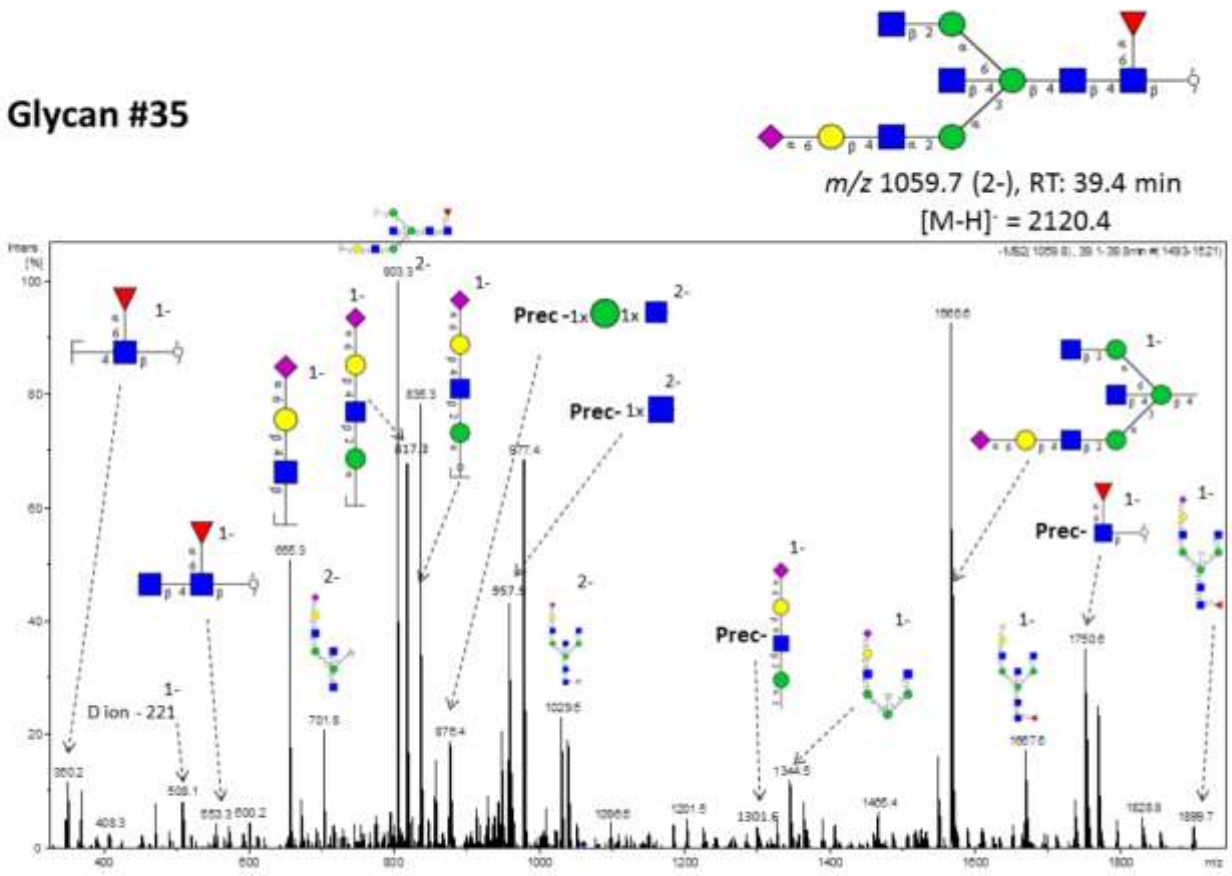
Glycan #34

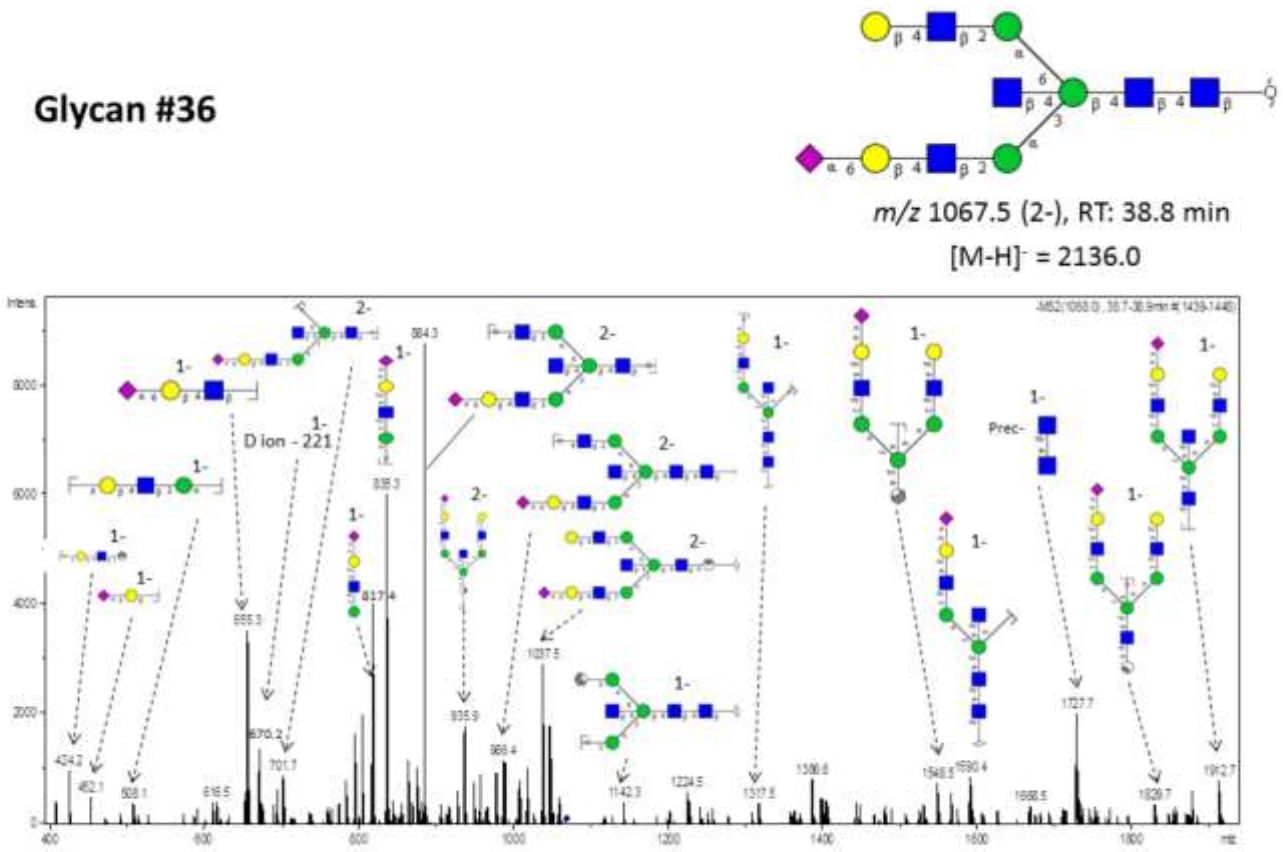


m/z 1039.4 (2-), RT: 44.4 min

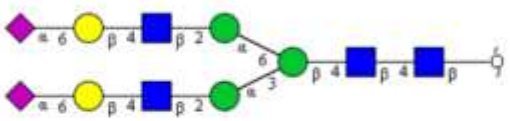
$[M-H]^-$ = 2078.8




Glycan #35
m/z 1059.7 (2-), RT: 39.4 min
[M-H]- = 2120.4
Prec -1x 1x
Prec- 1x
Prec-
D ion - 221
Intens. [%]
m/z

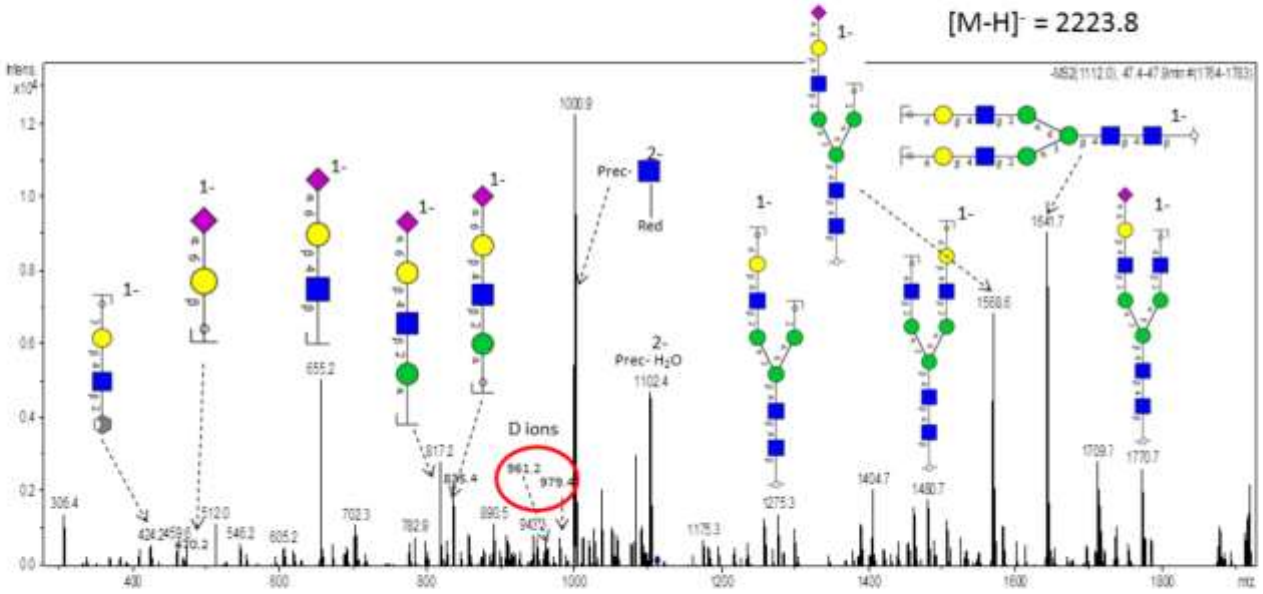

Glycan #36
m/z 1067.5 (2-), RT: 38.8 min
[M-H]- = 2136.0
D ion - 221
Prec-

Glycan #37a

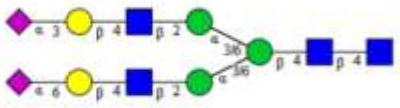


m/z 1111.4 (2-), RT: 47.6 min

$[M-H]^-$ = 2223.8

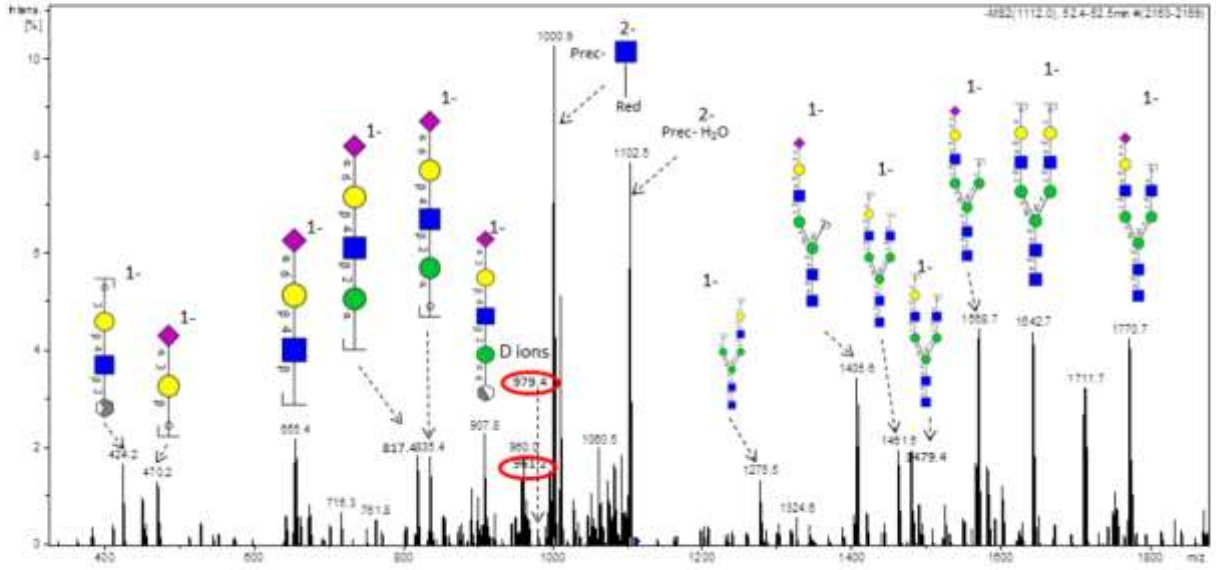


Glycan #37b

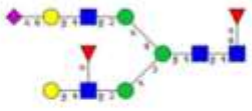


m/z 1111.4 (2-), RT: 52.5 min

$[M\text{-}H]^-$ = 2223.8

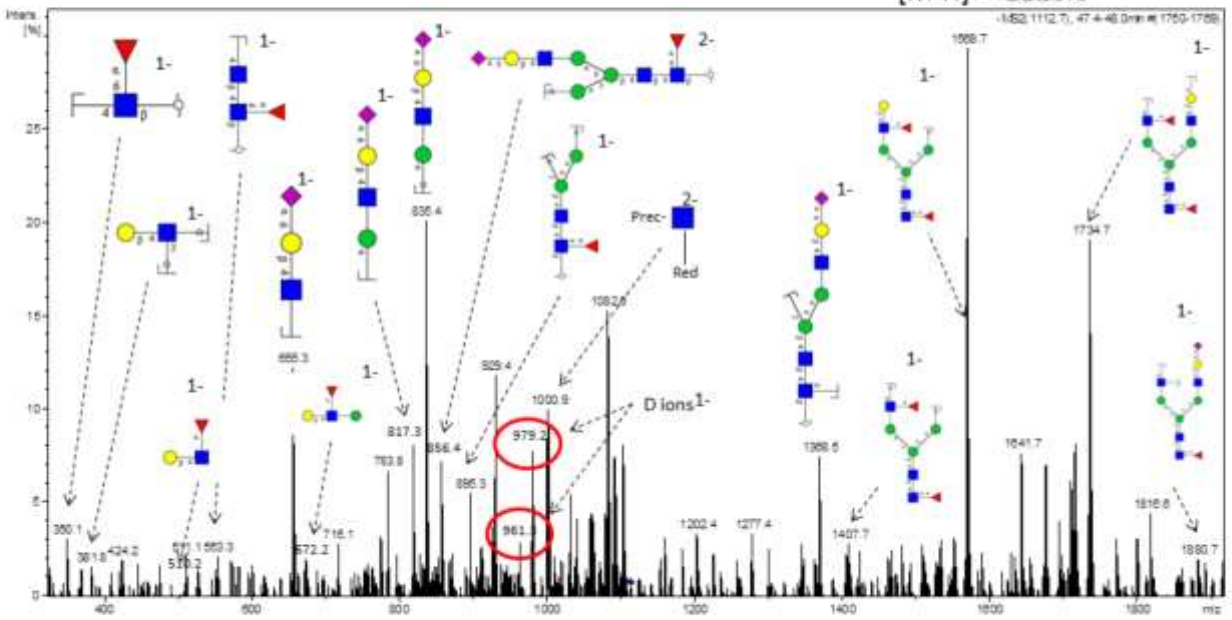


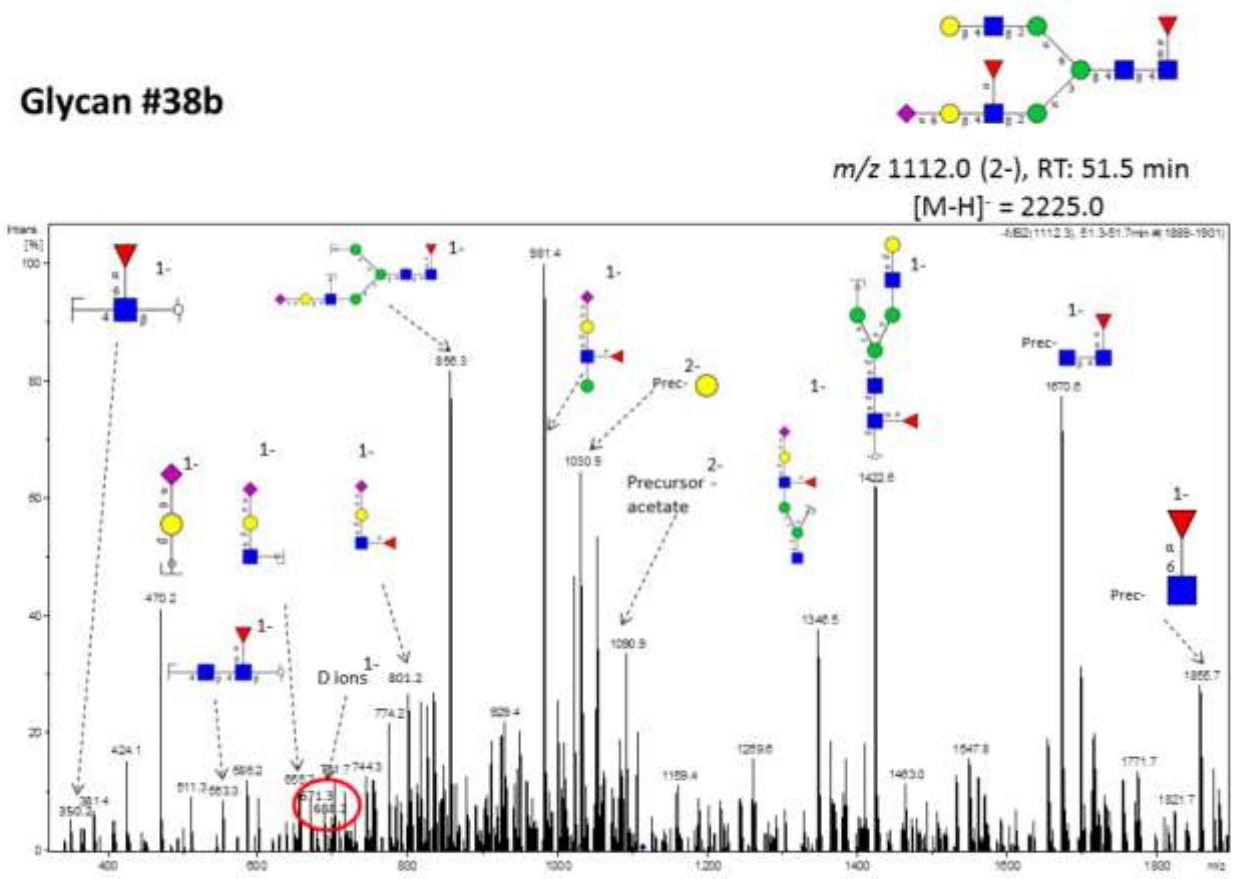
Glycan #38a



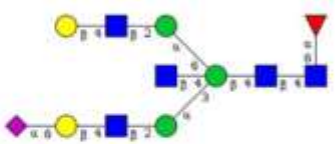
m/z 1112.0 (2-), RT: 47.7 min

$[M-H]^-$ = 2225.0

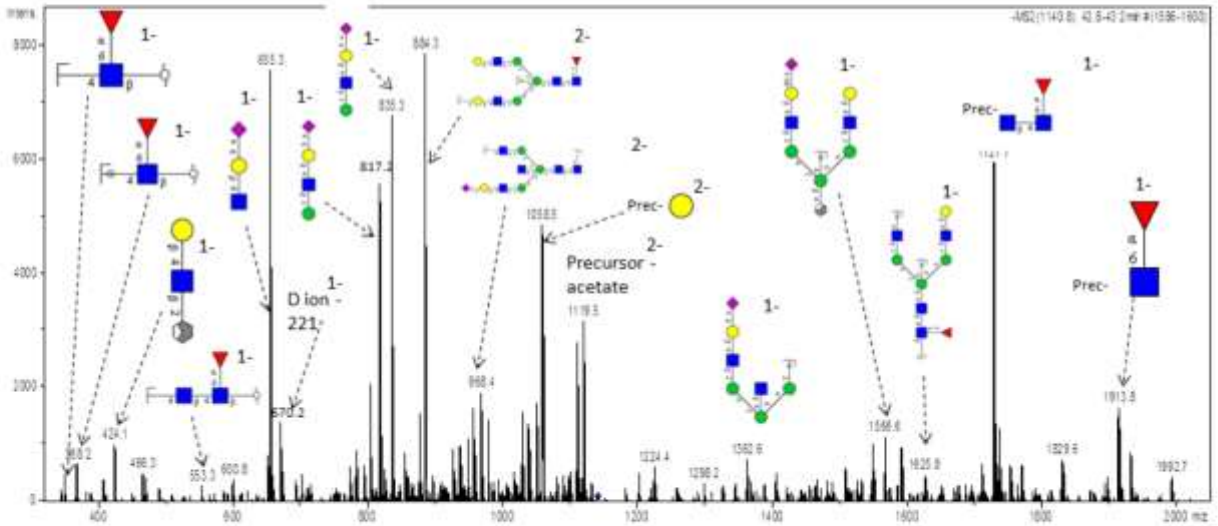



Glycan #38b
m/z 1112.0 (2-), RT: 51.5 min
[M-H]- = 2225.0
D ions
Precursor - acetate
Prec-

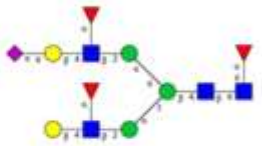
Glycan #39



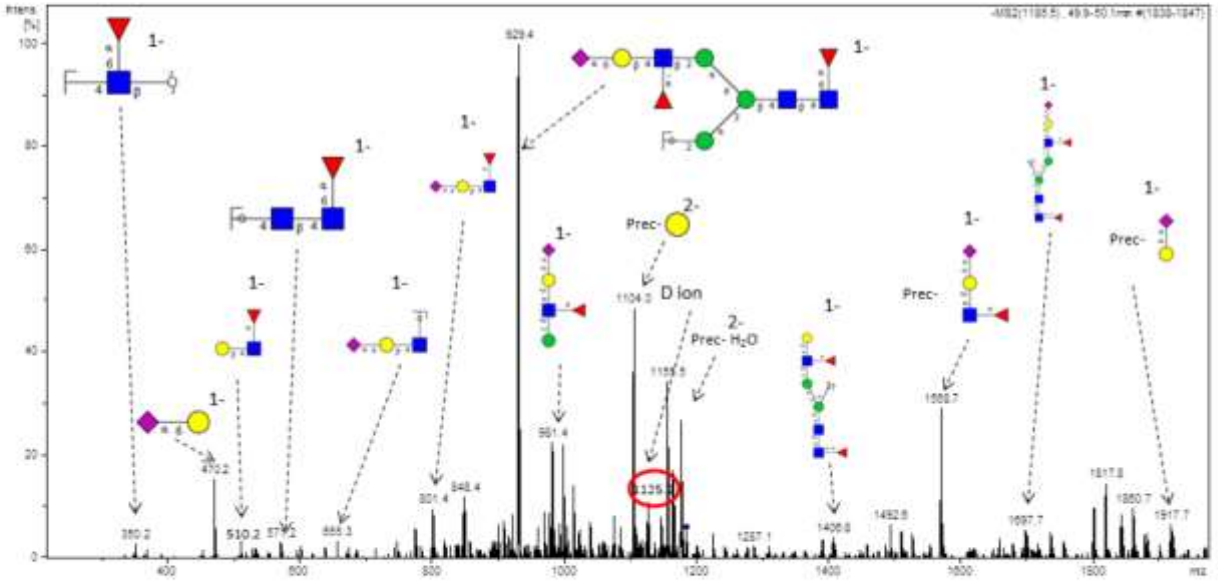
m/z 1140.5 (2-), RT: 43.0 min
$[M-H]^-$ = 2282.0



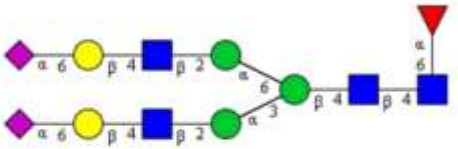
Glycan #40



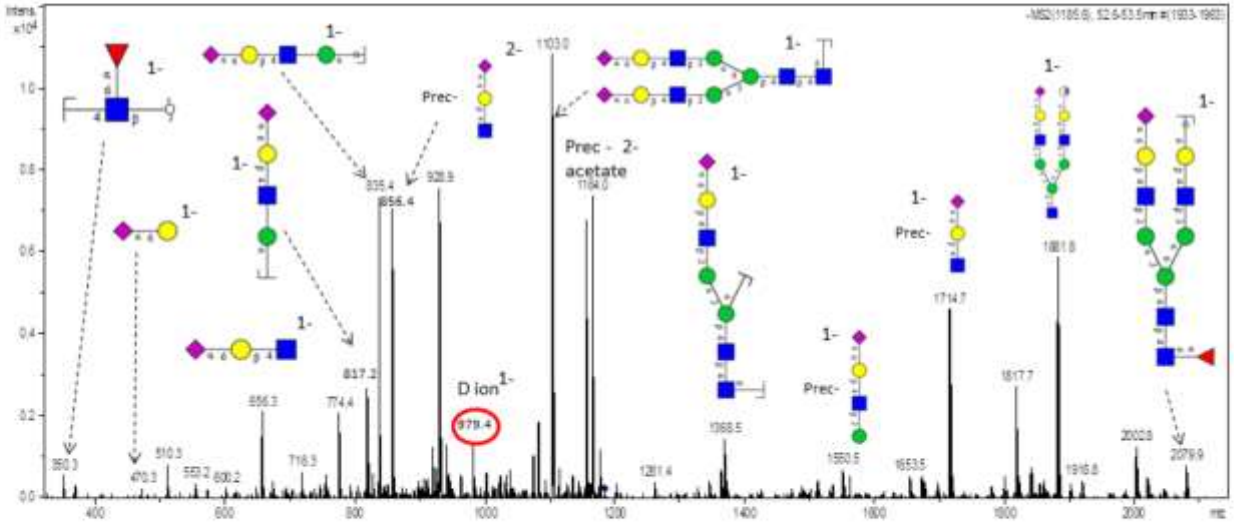
m/z 1185.0 (2-), RT: 50.2 min
$[M-H]^-$ = 2371.0



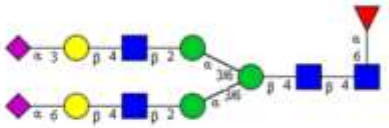
Glycan #41a



m/z 1184.5 (2-), RT: 53.0 min
$[M-H]^-$ = 2370.0

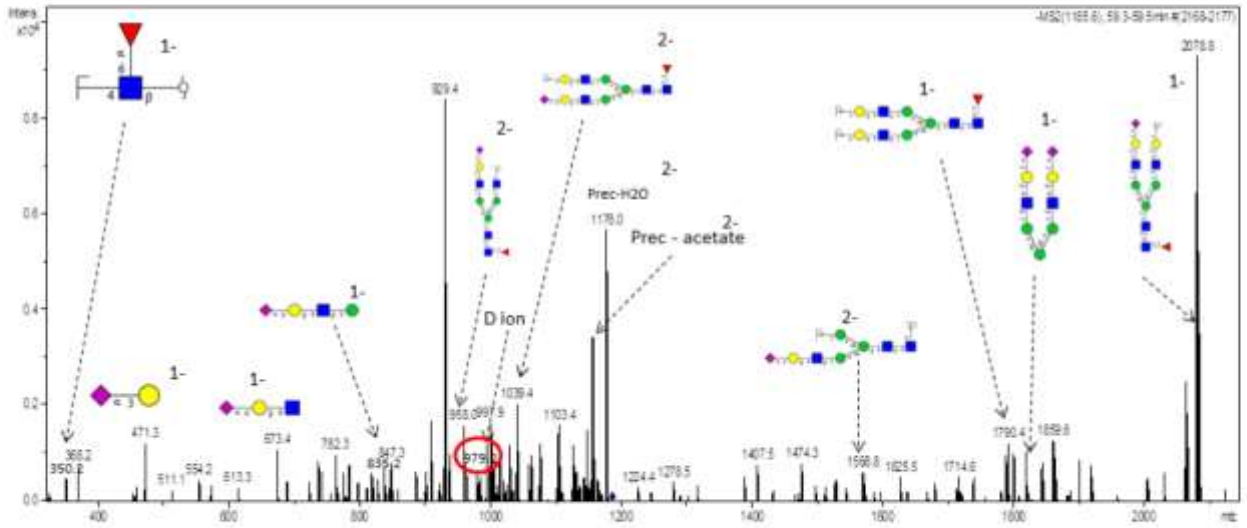


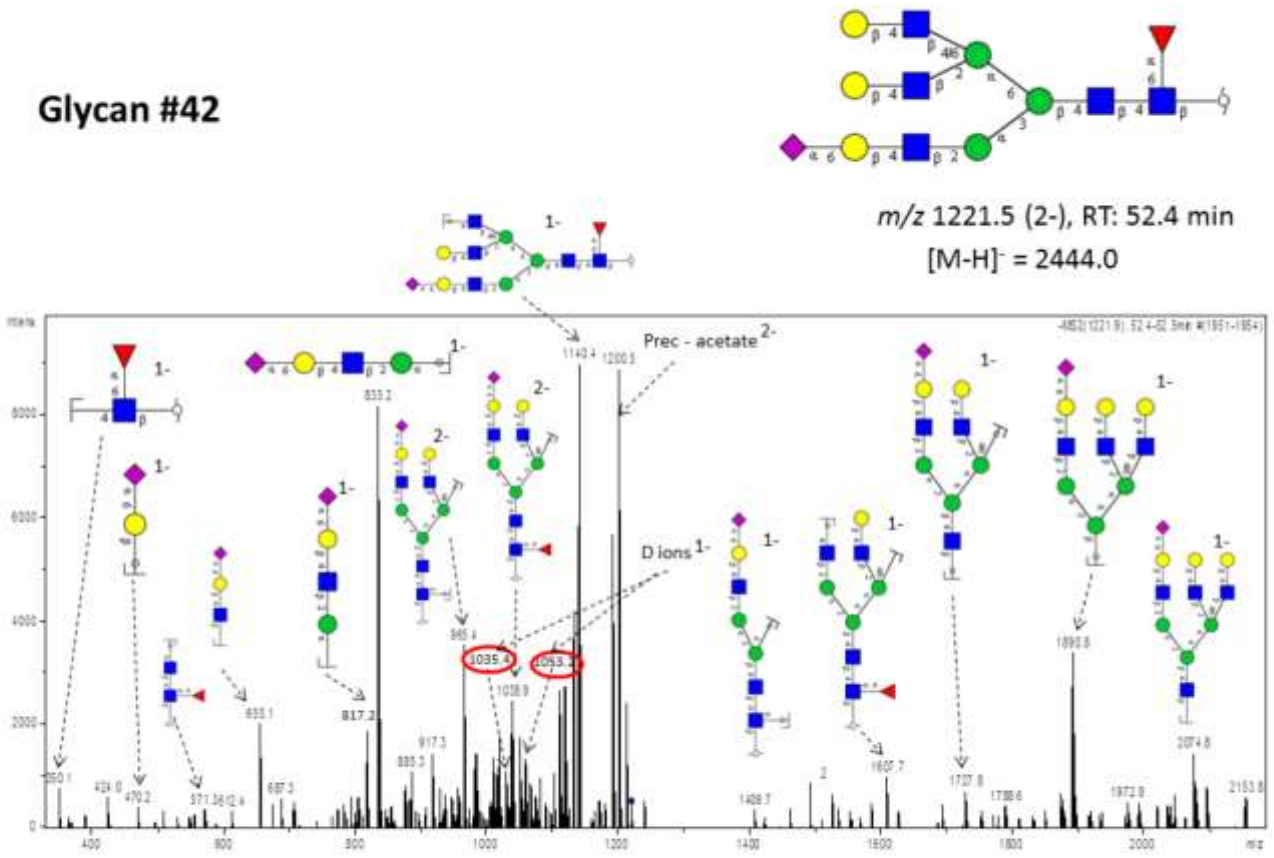
Glycan #41b



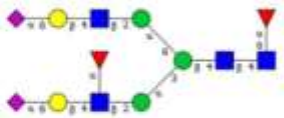
m/z 1184.5 (2-), RT: 59.4 min

$[M-H]^-$ = 2370.0

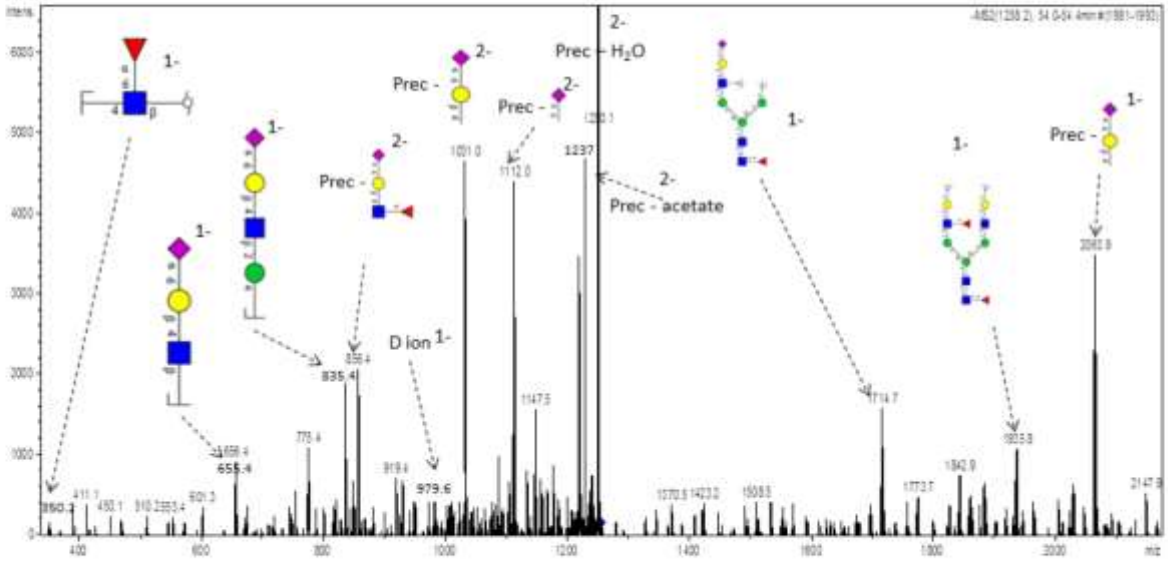



Glycan #42
m/z 1221.5 (2-), RT: 52.4 min
[M-H]- = 2444.0
Prec - acetate
D ions
1035.4
1053.4
817.2

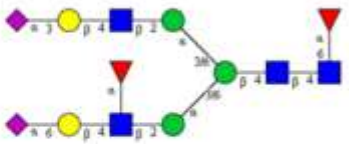
Glycan #43a



m/z 1257.5 (2-), RT: 54.2 min
$[M-H]^-$ = 2516.0

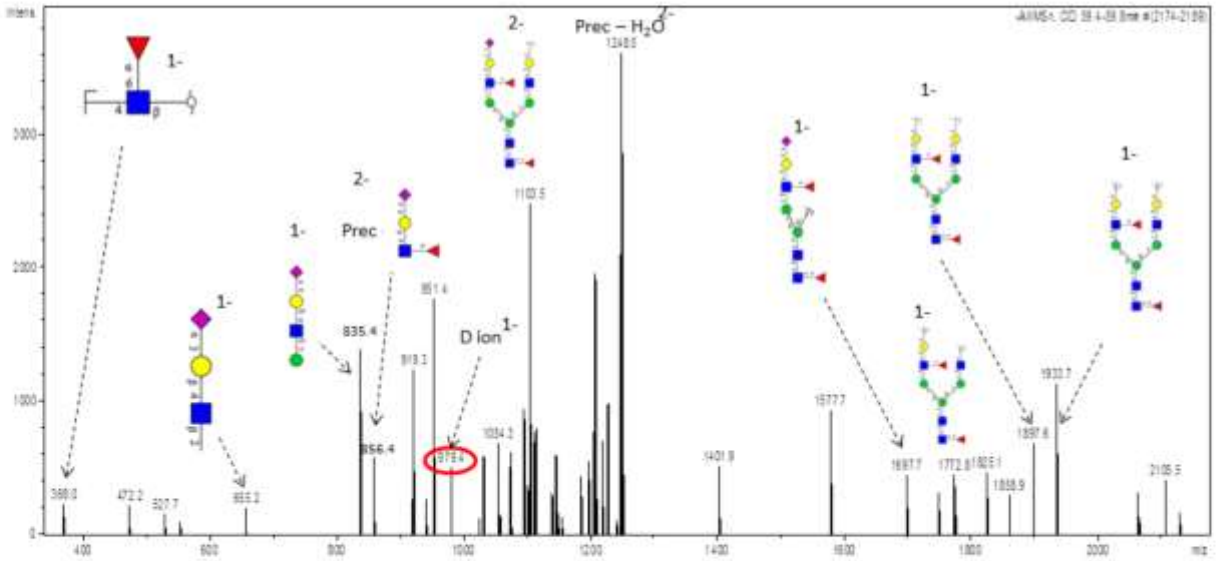


Glycan #43b

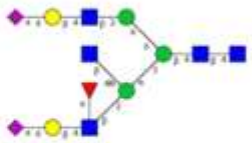


m/z 1257.5 (2-), RT: 59.6 min

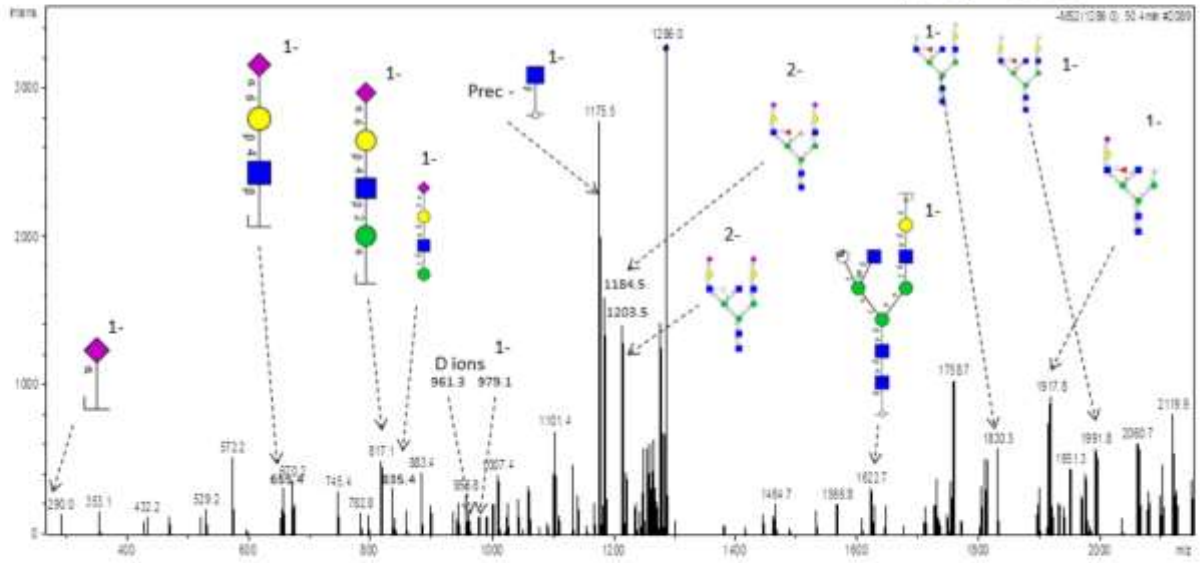
$[M-H]^-$ = 2516.0



Glycan #44



m/z 1286.0 (2-), RT: 50.4 min
$[M-H]^-$ = 2573.0



Chapter 3

Investigating the link between clinical CF-typed *Pseudomonas aeruginosa* isolates and PAO1 laboratory strain with pauci-mannosidic *N*-glycans

Contribution: In this chapter, all experimental work and analysis was performed by me.

3.1 Rationale

The first complete characterisation of the sputum *N*-glycome showed a higher abundance (>40% of the total *N*-glycans) of unique pauci-mannosidic *N*-glycans present in CF sputum, which was a 10-fold increase compared to non-infected non-CF sputum (**Chapter 2**). These pauci-mannose glycans were also present in abundance, at 25-30% of the total *N*-glycans, in infected non-CF sputum. This implies that pauci-mannose *N*-glycan structures are associated with bacterial colonisation.

The aim of the work described in this chapter was to investigate the association between the dominant bacterial species in CF lungs (*Pseudomonas aeruginosa*) and the presence of pauci-mannosidic *N*-glycans in sputum. In particular, the possibility that the infective bacteria may be degrading the glycans on the sputum proteins to produce the truncated pauci-mannose structures was investigated.

The increase in the pauci-mannose *N*-glycans was observed in bacterial colonised/ infected CF and iNCF sputum irrespective of the type of bacterium colonising the lung. Out of five CF patients, four were colonised with *P. aeruginosa*, one of the patients was infected with *Staphylococcus aureus* and infected non-CF (iNCF) patients were colonised with *S. aureus* and *Streptococcus pneumoniae*. Thus, the increase in pauci-mannose phenotype is not specific to any particular microorganism but to microbial colonisation in general. Since *P. aeruginosa* is the most dominant pathogen observed in CF patients (**Chapter 2, Table 1**), it was used as a model organism to investigate the relationship between bacterial colonisation and increased pauci-mannose *N*-glycans.

P. aeruginosa strains were isolated from the sputum of CF patients and identified using 16S rRNA sequencing. These isolates were used to determine whether the bacteria have exoglycosidase activities such as neuraminidase, galactosidase, hexosaminidase and mannosidase, which are necessary to trim down a complex, hybrid or high mannose type *N*-glycan to a pauci-mannosidic *N*-glycan. The actions of these fresh clinical isolates were compared against the common laboratory grown, cultured and maintained *P. aeruginosa* strain PAO1.

In addition, the potential of *P. aeruginosa* to utilise any free monosaccharides released from trimmed glycans as a sole source of carbon for cell respiration *in vitro* was assessed using the Biolog phenotype microarray platform. The *N*-glycan profile of the *P. aeruginosa* strains was also examined to determine whether the unusual and unique pauci-mannosidic structures were a bacterial product.

3.2 Introduction

Understanding the microbial flora of the cystic fibrosis (CF) respiratory tract is of considerable importance, considering the mortality rate caused by chronic infections in the lungs of CF patients (CF lung) (Harrison 2007). The CF airway is a complex microbial ecosystem with the presence of a heterogeneous microbial community including bacteria, fungi and viruses (Harrison 2007). Colonisation with new pathogens and the interactions between various microbial strains complicates the physiology of the CF lung, and thereby increases the difficulty in treating lung infections of a CF patient. The susceptibility of CF patients to microbial infection has been reported since the earliest description in 1940s and the microbiology of the CF airways is said to be a speciality unto itself (Stutman et al. 1987). The major pathogens that are involved with CF pathogenicity include mucoid and non-mucoid *Pseudomonas aeruginosa, Staphylococcus aureus, Burkholderia cepacia* and *Aspergillus fumigatus* (Miller et al. 2003). Other pathogenic bacteria that have been isolated from CF lungs are *Strenotrophomonas maltophilia* and *Alcaligens xylosoxidans* (Goss et al. 2002; Lyczak et al. 2002; Goss et al. 2004; Lambiase et al. 2006).

The importance of understanding the ecological and evolutionary processes shaping these microbial communities has become well recognised over the past decades in order to control infection of the CF lung. Sputum from CF patients is regularly used in hospitals as a part of screening procedures to monitor the microbial flora of the lung airways. Isolation of multiple microbes from a single patient's sputum proves the coexistence of bacteria, fungi and viruses in airways. Numerous studies have shown coinfection and communities of two or more bacterial species isolated from one airway and have also started to elucidate the ways pathogens might interact *in vivo* (Hoiby 1974; Petersen et al. 1981; Burns et al. 1998; Rogers et al. 2004; Wahab et al. 2004). A review article by Harrison (Harrison 2007) in 2007 detailed the microbial ecology of the CF lung, and the community interactions of pathogens in co-infections. The prevalence or colonisation of infecting bacteria in the CF lung changes depending on the age group of the patient. In 2004, Cystic Fibrosis Foundation National Patient Registry (Cystic Fibrosis Foundation, 2004) reported that *S. aureus* is the most common colonising bacterium in a child CF lung whereas *P. aeruginosa* colonises more than 80% of adult CF lung (CFFoundation 2004). Thus, *P. aeruginosa* is one of the most widely

studied, and reported to be the most destructive organism in terms of mortality and morbidity rate among CF patients.

P. aeruginosa is capable of colonisation by forming biofilms in the environment and in lungs. Production of polysaccharides, mainly alginate, and the utilisation of available extracellular DNA are very important for the *P. aeruginosa* species colonisation and biofilm formation by facilitating aggregation and adherence (Osman et al. 1986; Gal et al. 2003; Laue et al. 2006). Alginate production also favours conversion of non-mucoid to mucoid forms of *P. aeruginosa*. Bacteria in general are capable of dispersing the biofilm in unfavourable conditions by down-regulating the extracellular matrix components and producing enzymes that cleave matrix components (Gjermansen et al. 2005; Gjermansen et al. 2010). Eradication or dispersion of *P. aeruginosa* biofilm in CF lung is one of the most difficult aspects in treating CF patients and recent developments in dispersing biofilms with glycopeptide dendrimers shows a promising future in CF glycotherapeutics (**Chapter 1**).

Rapid identification of microbial flora and the dominance of pathogens isolated in a patient's sputum sample plays a major role in the choice of antibiotic treatment. Earliest methods included cultivation-based detection on selective media and serological assays, which are time consuming and have a low success rate in the identification of multiple pathogens at once (Rennie et al. 1978). Also, mucoid *P. aeruginosa* has been reported to dominate the agar growth plates, which limits the isolation of other organisms and produces substances that inhibit the growth of other bacteria (van Belkum et al. 2000). The conventional culture-based methods have now been more commonly replaced with molecular methods such as PCR amplification of fragments of ribosomal bacterial DNA (van Belkum et al. 2000). The PCR-based amplification of ribosomal genes has been able to reveal the complex bacterial populations in various clinical samples such as sputum and faecal samples (Millar et al. 1996; Bittar et al. 2008). One recent method (Nazaret et al. 2009) involved Ribosomal Intergenic Spacer Analysis (RISA) coupled with HPLC for the detection of lung bacterial colonisers in the lungs of children with CF and shows great potential for the detection of novel or emerging pathogens in CF lungs.

Bacteria can be distinguishable from one another by their growth on different agar media containing variations in carbon and nitrogen sources (Wong et al. 1984). Different Carbon

and Nitrogen growth phenotypes can define specific bacterial species using selective media (Bochner 2009). However, for screening and identification of emerging pathogens, a high-throughput technique becomes an absolute must to measure hundreds of phenotypes at a time. Recently, Biolog Systems developed phenotype MicroArrays (PM) for this purpose. The Microarray is a respiration based 96-well plate assay to simultaneously screen up to 2000 phenotypic traits including an array of sugar and nitrogen sources (Bochner et al. 2001; Bochner 2003). The assay utilises the reduction of the tetrazolium dye to colorimetrically detect the respiration of cells on each individual carbon and nitrogen source. Utilisation of the Carbon or Nitrogen source (growth phenotype) causes the irreversible conversion of dye colour to purple under physiological conditions over a period of hours and measures the degree of respiration. OmniLog instrument (developed by Biolog Systems) was incorporated along with this technology to read and record the colour change in PM assays (Bochner et al. 2001; Johnson et al. 2008).

Bacterial cleavage of oligosaccharides and the utilisation of the different monosaccharides as carbon source has been widely studied in *Bifidobacterium* species, which is found in gastrointestinal microbiota and plays an important role in maintaining human health. This bacterial species in the infant gut is capable of utilising complex oligosaccharides in human milk as a carbon and energy source (Marcobal et al. 2010). They are capable of degrading the human milk oligosaccharides with the increased activity of secreted exoglycosidases such as fucosidase, sialidase and galactosidase (Sela et al. 2011; Sela et al. 2012; Yoshida et al. 2012). These probiotic bacteria such as *Bifidobacterium* and *Lactobacillus* are often used as dietary supplements to prevent a range of human diseases including diarrhoea (D'Souza et al. 2002).

The physiology and pathology of the gut and respiratory tract are closely related (Mortaz et al. 2013). In addition, sputum *N*-glycosylation (**Chapter 2**) has similar glycan epitopes to human milk oligosaccharides such as Lewis epitopes, terminal sialylation and so on. *P. aeruginosa* has been suggested to secrete neuraminidase which cleaves the sialic acid capped oligosaccharides on mucus and on the epithelial cell layer. This in turn helps in unmask the galactosylated epitopes and improves the adhesion of pathogens to the cell surface using the bacterial lectins and other adhesins (**Chapter 1**) (Venkatakrishnan et al. 2013). Although the enhancement of the adhesion of *P. aeruginosa* to host mucin/ epithelial cell surface by

secreting neuraminidase is known, but whether *P. aeruginosa* utilises the released sialic acids as carbon source has not been investigated.

To explain the abundance of pauci-mannose structures in infected sputum as being bacterially produced would require 1) that *P. aeruginosa* utilises or expresses the exoglycosidases necessary to trim *N*-glycans all the way to make the unusual pauci-mannose type *N*-glycans, and 2) that they use the then available released free sugars as a carbon source for respiration and/or 3) that the bacterial glycoproteins carry the pauci-mannose structures.

To address the first option clinical *P. aeruginosa* strains and laboratory strain PAO1 were incubated with standard purified glycosylated proteins such as fetuin, IgG and RNaseB and complex non-infected (NCF) non-CF sputum, that contain different classes of complex and high mannose glycans, and the glycosylation profiles of these proteins before and after incubation with bacteria were examined by PGC-LC-ESI-MS/MS. Secondly, the utilisation of the types of monosaccharides that would be released as monosaccharides from the trimming of *N*-glycans was tested by the Biolog PM assay for their capacity to be used as carbon source by the different *P. aeruginosa* strains. Finally, the *N*-glycan profile of *P. aeruginosa* membranes was analysed for the presence of pauci-mannose structures.

3.3 Materials and methods

3.3.1 Isolation of *P. aeruginosa* strains from CF sputum

Non-invasively obtained sputum samples collected from CF patients (CF1, CF2, CF4 and CF5) as described in **Chapter 2**, were used for isolation of *P. aeruginosa* strains. Information regarding the CF patients including age, gender and CF phenotype are given in **Chapter 2**, **Table 1**. To reduce the viscosity, sputum samples were reduced using 10 mM DTT (final concentration) and serial dilutions of homogenised sputum were prepared using phosphate-buffered saline solution. Fifty µl of the diluted sputum were streaked on Horse blood agar (Edwards Instruments Co, New South Wales, Australia) and incubated overnight at 37°C. API 20NE test kit (BioMerieux, UK) was used initially to screen 12 individual colonies obtained on Horse blood agar plates. The test kit included a strip containing 20 different biochemical tests to which the isolates were inoculated and incubated at 30°C for 24 hours. The identification of the isolates was carried out as per API 20NE instruction manual. Four strains from the different sputum samples which showed positive identity for *P. aeruginosa* in the API 20NE test were picked for identification by 16S rRNA sequencing and stored at -20°C in glycerol stocks.

3.3.2 Identification of *P. aeruginosa* strains using 16S rRNA sequencing

The four isolated *P. aeruginosa* strains were plated on Luria-Bertani agar (Luria et al. 1957) plates and incubated overnight at 37°C. A colony of cells were then inoculated into 5 ml liquid LB medium, incubated overnight at 37°C and the bacteria collected from the culture medium by centrifugation at 4000g for 10 minutes. The supernatant was discarded and the cell pellet resuspended in 2 ml XS buffer (0.5 g potassium ethyl xanthogenate (Fluka, 140-89-6), 10 ml 4M ammonium acetate, 5 ml 1M Tris-HCl pH 7.4, 2 ml 0.45M EDTA, 2.5 ml 20%(w/v) SDS to 50 ml). The resuspended pellet was incubated for 2 hours at 65°C, vortexed for 10 seconds and incubated on ice for 30 minutes. After further centrifugation for 10 minutes at 14000g, the supernatant was collected in a new tube and an equal volume of 100% isopropanol was added, incubated at room temperature for 5 minutes and centrifuged at 14000g for 10 minutes. The pellet was washed once with 70% (v/v) ethanol and resuspended in 100 µl of nuclease free water. After diluting 10 times, the concentration of the extracted

DNA was determined with a Nanodrop ND1000 spectrophotometer (NanoDrop Technologies).

The identification of the bacterial strains was performed using 16S rRNA sequencing. For PCR, 1 µl of extracted DNA from each strain was mixed with 22 µl PCR master mix (Promega, WI, USA) and 0.4 µl of forward primer and 0.4 µl of reverse primer. Thermal cycle parameters used for PCR run are given in **Table 1**. To analyse the completed PCR reaction and to check for any contamination present, 5 µl of the amplified DNA was loaded on 1% agarose gel with molecular size markers (Invitrogen). The amplified 16S region was sequenced by in-house Macquarie University DNA Analysis Facility (Sydney, Australia). The resultant sequences were searched against *Pseudomonas* species (taxid: 286) using the NCBI BLASTN 2.2.29+ program.

Table 1: Parameters that were used in polymerase chain reaction (PCR) for amplification of 16S rRNA region of the DNA.

Cycle	Temperature (°C)	Time	# of Cycles
Initial denaturation	95	2 minutes	1
Denaturation	95	20 Seconds	30
Annealing	52	30 Seconds	30
Extension	68	2 minutes 30 Seconds	30
Final extension	68	1 minute	1
Cooling	4		

3.3.3 Biolog Phenotypic microarray (PM) assay

Identified clinical isolates of *P. aeruginosa* strains and the laboratory *P. aeruginosa* strain PAO1 were plated on LB agar plates and incubated overnight at 37°C. The bacterial cells were harvested using a sterile swab and inoculated into a test tube containing 10 ml IF-0a solution (Biolog). The suspension was mixed thoroughly and 42% transmittance was achieved as determined by a turbidimeter. An appropriate volume of the prepared suspension was

added to IF-0a plus dye mix solution (Biolog) to give 85% transmittance cell density. 100 µl of this suspension was added to each well of Biolog PM Plates 1 and 2A containing 190 different carbon sources and water was used as two negative controls. The PM plates were incubated in the Omnilog automated plate reader for 72 hours at 37°C and change in dye colour due to reduction of tetrazolium dye as a result of bacterial respiration was recorded at 590 nm absorbance after every 15 minutes. Data analysis was performed using Kinetic and Parametric software (Biolog). Detailed protocols for phenotypic characterisation of microbial cells using Biolog phenotype microarrays are given in (Mackie et al. 2014).

3.3.4 *P. aeruginosa* challenge with glycoproteins

Three clinical *P. aeruginosa* strains and PAO1 were harvested from 5 ml of LB broth grown overnight at 37°C. Culture medium was centrifuged at 4000g for 20 minutes and pellets were further washed twice with 1 ml phosphate-buffered saline (PBS) solution and centrifuged at 4000g for 5 minutes. The cells were then diluted with PBS solution to an OD_{600} of approximately 1.0 identified using a Biophotometer. 25 µl aliquots of 50 µg purified glycoproteins (RNase B, immunoglobulin G and fetuin) and non-infected (NCF) non-CF sputum were inoculated with 10 µl of each type of *P. aeruginosa* and incubated overnight at 37°C. Glycoproteins were incubated at 37°C without bacterial inoculation as controls. 10 µg of protein was dot blotted on PVDF membrane and the glycans were released from the protein as described in **Chapter 2, materials and methods section**.

3.3.5 Extraction of *P. aeruginosa* whole cell proteome

Clinical *P. aeruginosa* strains and PAO1 strain were grown overnight in LB broth, centrifuged and the pellet washed as above. The cells were then diluted to an OD_{600} of 1.0 identified using Biophotometer which is approximately 10^7 cells per ml. The approximate yield of bacterial cells was determined by dry weight and an equal weight of acid washed glass beads (Sigma) was added. A further 500 µl of ice-cold PBS (pH 7.4) was added to submerge the cells and beads and the mixture was vortexed for 30 seconds. Three cycles of bead beating with intensity of 6.5 for duration of 20 seconds was performed with intermittent cooling on ice for 10 minutes. The unbroken cells and debris were removed by centrifugation at 2500g at 4°C for 8 minutes. Acetone precipitation of the proteins in the supernatant was

performed by adding ice-cold acetone at 6 times the volume of the supernatant followed by incubation at -20°C for an hour. After centrifugation at 2500g at 4°C for 15 minutes, the pellet was air dried and reconstituted in 4 M urea. Protein concentration was determined using the Direct Detect Spectrometer (Millipore). 10 μg of protein was dot blotted on PVDF membrane and the glycans were released from the protein as described in **Chapter 2, materials and methods section**.

3.4 Results

3.4.1 Isolation and identification of *P. aeruginosa* strains

To study whether *P. aeruginosa* contributes to the *N*-glycome variation observed in CF sputum, four strains were isolated from sputum collected from four different CF (CF1, CF2, CF4 and CF5) patients (**Chapter 2, Table 1**). The CF sputum samples were reduced and plated on Horse Blood agar plates. Multiple colonies from the different sputum samples were screened for the identification of *P. aeruginosa* strains using the API 20NE test. Four strains from four different sputum samples were identified to be different *P. aeruginosa* strains. For further confirmation of the identity of the isolated strains, 16S rRNA sequencing was carried out. The genomic DNA was extracted using Xanthogenate-SDS (XS) DNA extraction protocol which worked well with polysaccharide rich organisms. Final concentrations of the extracted DNA for the four isolates from which the DNA was amplified are given in **Table 2**.

Table 2: The concentration of the four extracted DNAs determined with a Nanodrop ND1000 spectrophotometer**.**

Strain	Concentration (ng/ µl)	Std. Dev
1	35.05	±0.35
2	34.15	±0.78
3	35.6	±3.14
4	38.15	±0.28

After PCR, a portion of the amplified DNA was analysed on an agarose gel to assess fragment size and purity. The gel image of the PCR product showed a clear DNA band at the expected size of approximately 1060 bp (16S rRNA region) and no contaminants were visualised in the gel lane (**Figure 1**). The 16S rRNA sequences from each isolate were searched against *Pseudomonas* species (taxid: 286) using NCBI BLASTN 2.2.29+ program (Zhang et al. 2000). The BLAST search found a high confidence match against the *P. aeruginosa* sequence. The top two hits for search results for each of the four isolates along with query coverage, identity and NCBI accession are given in **Table 3**. The four fresh

isolates from clinical sputum samples were named PASS1, PASS2, PASS3 and PASS4 for future investigations.

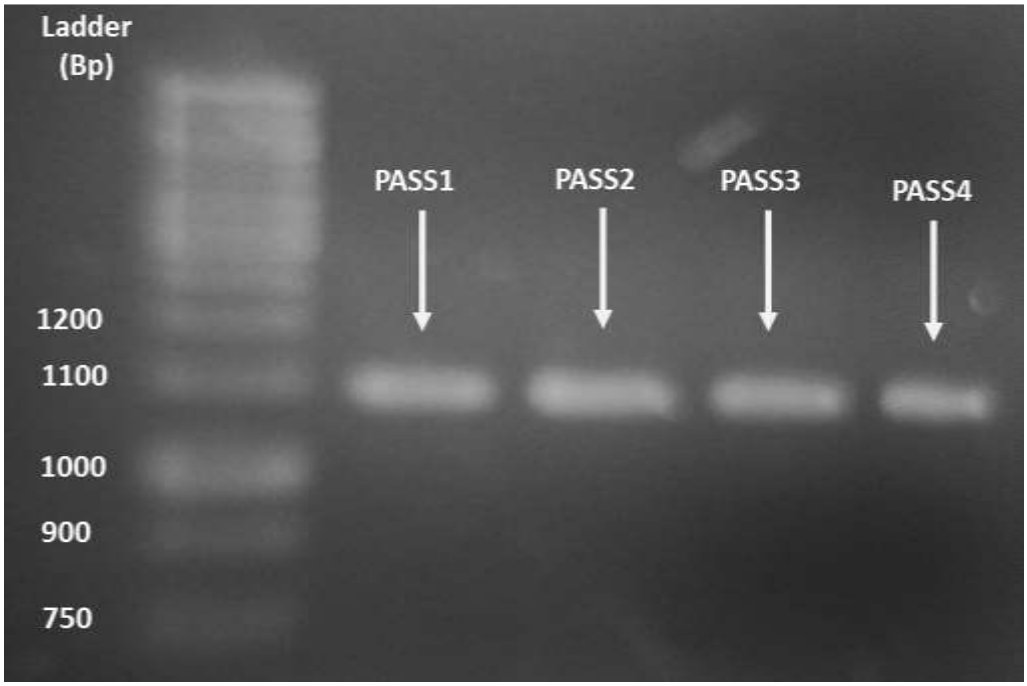


Figure 1: 1% Agarose gel that was run after the DNA amplification to confirm the purity of the amplified 16S rDNA of expected size of approximately 1060 BP. Lane 1 represents the DNA molecular size marker (Invitrogen), lane 2, 3 and 4, the amplified 16SDNA product of PASS1, PASS2, PASS3 and PASS4 *P. aeruginosa* clinical isolates respectively.

Table 3: Top two hits in BLAST search for the species identification of the four sputum isolates along with their % query coverage, % Identity and NCBI accession number.

Clinical isolate #	Description of organism matched	Query coverage (%)	Identity (%)	NCBI accession
Isolate 1	*Pseudomonas aeruginosa* LESlike4 sequence	100	99	CP006985.1
Isolate 1	*Pseudomonas aeruginosa* LESlike1 sequence	100	99	CP006984.1
Isolate 2	*Pseudomonas* sp. MH1(2014) 16S ribosomal RNA gene, partial sequence	100	100	KJ081228.1
Isolate 2	*Pseudomonas aeruginosa* strain JAY2N 16S ribosomal RNA gene, partial sequence	100	100	KF952246.1
Isolate 3	*Pseudomonas* sp. MH1(2014) 16S ribosomal RNA gene, partial sequence	100	100	KJ081228.1
Isolate 3	*Pseudomonas aeruginosa* LESlike4 sequence	100	100	CP006985.1
Isolate 4	*Pseudomonas aeruginosa* strain hqd-01 16S ribosomal RNA gene, partial sequence	100	100	KC959478.1
Isolate 4	*Pseudomonas aeruginosa* strain Z91 16S ribosomal RNA gene, partial sequence	100	100	KC212090.1

3.4.2 Monosaccharides as a carbon source for *P. aeruginosa* growth

CF and infected non-CF sputum were shown to have a high abundance of protein glycans that had apparently been trimmed (**Chapter 2**). Phenotype microarrays were used to determine whether *P. aeruginosa* could use the monosaccharides that would result from the possible trimming by the bacteria of the host cell glycans, as a carbon source. A total of 190 different carbon sources including various monosaccharides, as well as amino acids and carboxylic acids were tested individually in two 96-well plates (PM1 and PM2A). The four freshly isolated *P. aeruginosa* strains were compared against the laboratory *P. aeruginosa* strain PAO1 in utilising various carbon sources for respiration. Water was used as negative control to represent no *P. aeruginosa* cellular respiration with lack of carbon source.

In this study we wanted to specifically test whether *P. aeruginosa* can use the seven monosaccharides that are predominantly seen in human *N*-glycans, namely glucose, galactose, mannose, sialic acid, fucose, *N*-acetyl Glucosamine (GlcNAc) and *N*-acetyl Galactosamine (GalNAc) as sole carbon source. **Figure 2** shows the cellular respiration levels of the different *P. aeruginosa* strains with the monosaccharides as the only provided source of carbon. It is interesting to observe that all five strains did not utilise galactose, mannose, sialic acid, fucose and GalNAc as sole carbon source for respiration. Glucose was widely utilised by all the strains as expected, except PASS4 strain, which did not utilise any of the seven monosaccharides as a sole carbon source. Also, PASS2 shows a slower respiration rate in utilising glucose as carbon source than the other three strains PASS1, PASS3 and PAO1. Only one *N*-glycan constituent, GlcNAc, was able to be used for growth by PASS1, PASS3 and PAO1 *P. aeruginosa* strains (**Figure2**). The PASS2 strain could not utilise GlcNAc for respiration. These data suggest that, *P. aeruginosa* does not utilise the trimmed monosaccharides as carbon source for cellular respiration except glucose and GlcNAc, which are utilised by PASS1, PASS3 and PAO1 *P. aeruginosa* strains.

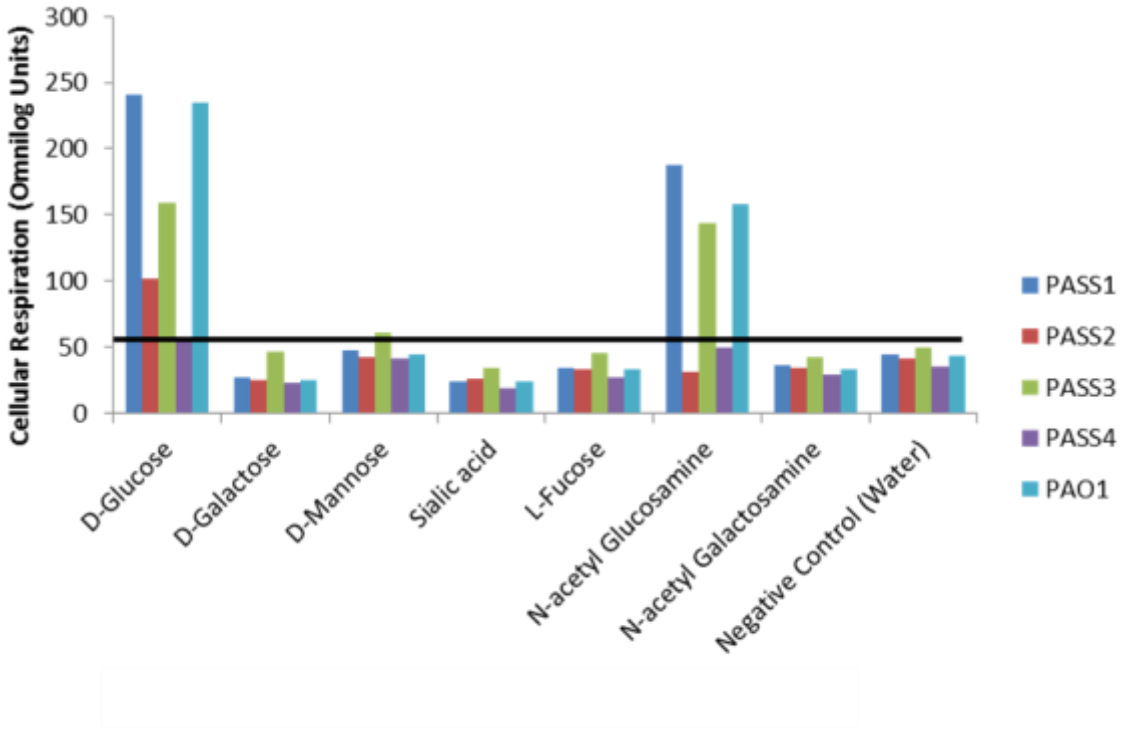


Figure 2: Cellular respiration level of five *P. aeruginosa* strains using monosaccharides as the sole carbon source were analysed in a single experiment on a Biolog phenotype microarray. Black line shows the threshold of carbon source respiration. Above the line shows the monosaccharides utilised by *P. aeruginosa* strains for cell respiration with respect to negative control (water).

3.4.3 Possibility of exoglycosidase expression by *P. aeruginosa*

In **chapter 2**, an increase in the relative abundance of unique pauci-mannosidic *N*-glycans and decrease in complex and hybrid glycans were evident on both bacterial colonised/ infected CF and iNCF sputum proteins. To further test whether the variation was a result of bacterial exoglycosidase activity, the isolated *P. aeruginosa* strains PASS1, PASS2, PASS3 and laboratory strain PAO1 were incubated with standard purified glycoproteins and non-CF (NCF) sputum. Since, PASS4 *P. aeruginosa* strain does not utilise any of the monosaccharides as carbon source including glucose, it was not used to test the possibility of bacterial exoglycosidase expression. Proteins for which the *N*-glycosylation is well characterised (Takahashi et al. 1987; Green et al. 1988; Fu et al. 1994) and that carry different classes of *N*-glycans were chosen: fetuin (highly sialylated complex), IgG (neutral complex), RNase B (high mannose) and non-infected (NCF) non-CF sputum (**Chapter 2**). Each of the four *P. aeruginosa* strains was inoculated into a water solution containing the pure glycoproteins separately and NCF sputum and incubated at 37°C overnight. Two controls were used in this

study, one was a "no inoculation/ no incubation control" in which neither bacteria were added nor incubation of glycoprotein and sputum at 37°C was performed, and the second was the "incubation control" in which the bacteria was not added but the glycoprotein and sputum were incubated at 37°C. It was considered that a resultant trimming of the monosaccharides from the *N*-glycan structures, as determined by PGC-LC-ESI-MS/MS, could indicate the presence of specific exoglycosidases secreted by the bacteria in the sputum.

The possible action of bacterial neuramindase on complex sialylated *N*-linked glycoproteins was tested by incubating clinical and laboratory *P. aeruginosa* strains with standard fetuin protein. Fetuin has six known *N*-linked glycan compositions, namely bi-antennary mono-sialylated, bi-antennary di-sialylated, tri-antennary mono-sialylated, tri-antennary di-sialylated, tri-antennary tri-sialylated and tri-antennary tetra sialylated (Green et al. 1988). Each of the glycans has a terminal sialylated residue(s) and any action of neuraminidase would cleave the sialic acids and result in non/less sialylated *N*-glycans. *N*-glycans were released from fetuin after incubation with the *P. aeruginosa* strains and from the control samples. The *N*-glycans were separated on a porous graphitised carbon column (PGC) and the monoisotopic MS profile was used for monosaccharide compositional analysis. Analysis was performed in technical (LC-MS/MS) duplicates. Peak area based relative quantitation was performed to calculate the ratio of sialo:asialo *N*-glycan compositions.

The "no inoculation/ no incubation" control and "incubation control" did not change the sialo:asialo ratio from 100:0 based on the relative abundances of the observed released *N*-glycans from fetuin. There was also no change in the ratio between sialo and asialo *N*-glycans after incubating with clinical *P. aeruginosa* strains PASS1, PASS2 and PASS3 and the laboratory strain PAO1. This can be appreciated at a qualitative level in **Figure 3i and ii** in which the extracted ion chromatogram peaks of the bi-antennary di-sialylated and bi-antennary mono-sialylated glycans after incubation in the presence of bacteria were seen to be still present in the released glycans of fetuin. Concomitantly, the possible neuraminidase treated product (bi-antennary galactosylated glycan) of these glycans was not observed (**Figure 3iii**). This suggests that the reduced sialic acid in the CF sputum *N*-glycans was probably not a result of a neuraminidase secreted by *P. aeruginosa*. Interestingly, there is a change in ratio of disialylated *N*-glycan isoforms (**Figure 3i**) on treatment with the three

clinical isolates PASS1 (**Figure 3i-D**), PASS2 (**Figure 3i-E**) and PASS3 (**Figure 3i-F**), in which there is a decrease in α2,6-2,6 sialylation and a simultaneous increase in α2,6-2,3 sialylation. This is intriguing and was observed in triplicate. We speculate that, this may possibly be an action of trans-sialidase enzyme, which takes off a specific linked sialic acid and puts it back with a different linkage in order to change the binding epitopes. Trans-sialidase activity has so far only been identified on a protozoan *Trypanosoma cruzi* (Buschiazzo et al. 2012) but not on *P. aeruginosa*. Thus, further study needs to be done to investigate the change in the ratio of disialylated glycans and is outside the scope of this thesis.

Neutral *N*-glycans terminating with galactose or GlcNAc residues are also present in sputum proteins as described in the previous chapter (**Chapter 2**). To check if *P. aeruginosa* secretes the galactosidase or hexosaminidase needed to produce the observed pauci-mannose structures, purified IgG protein containing neutral *N*-glycans were incubated with the three clinical (PASS1, PASS2 and PASS3) and PAO1 *P. aeruginosa* strains. The extracted ion chromatogram (EIC) (**Figure 4i, ii and iii**) shows that three neutral *N*-glycan structures on IgG remain after incubation with the bacteria indicating a lack of any exoglycosidase activity. In addition, the mass of the possible resultant pauci-mannosidic *N*-glycan was absent (**Figure 4iV**) supporting the lack of galactosidase and hexosaminidase activity in the clinical or laboratory *P. aeruginosa* cells. To confirm the qualitative data with statistical significance, the relative abundances of IgG *N*-glycans were divided based on 1) galactose terminating, 2) GlcNAc terminating and 3) sialic acid terminating *N*-glycans. The average relative abundance and the standard deviation of those three categories present in controls and *P. aeruginosa* challenged IgG based on technical duplicates is given in **Table 4**. Although there are variations in the relative abundances of each class, one-way ANOVA test showed there was no statistically significant difference with p value greater than 0.05 in each of the three categories across the six samples, thus supporting the notion that bacterial galactosidase and hexosaminidase activity do not contribute to the changes in the CF sputum *N*-glycome.

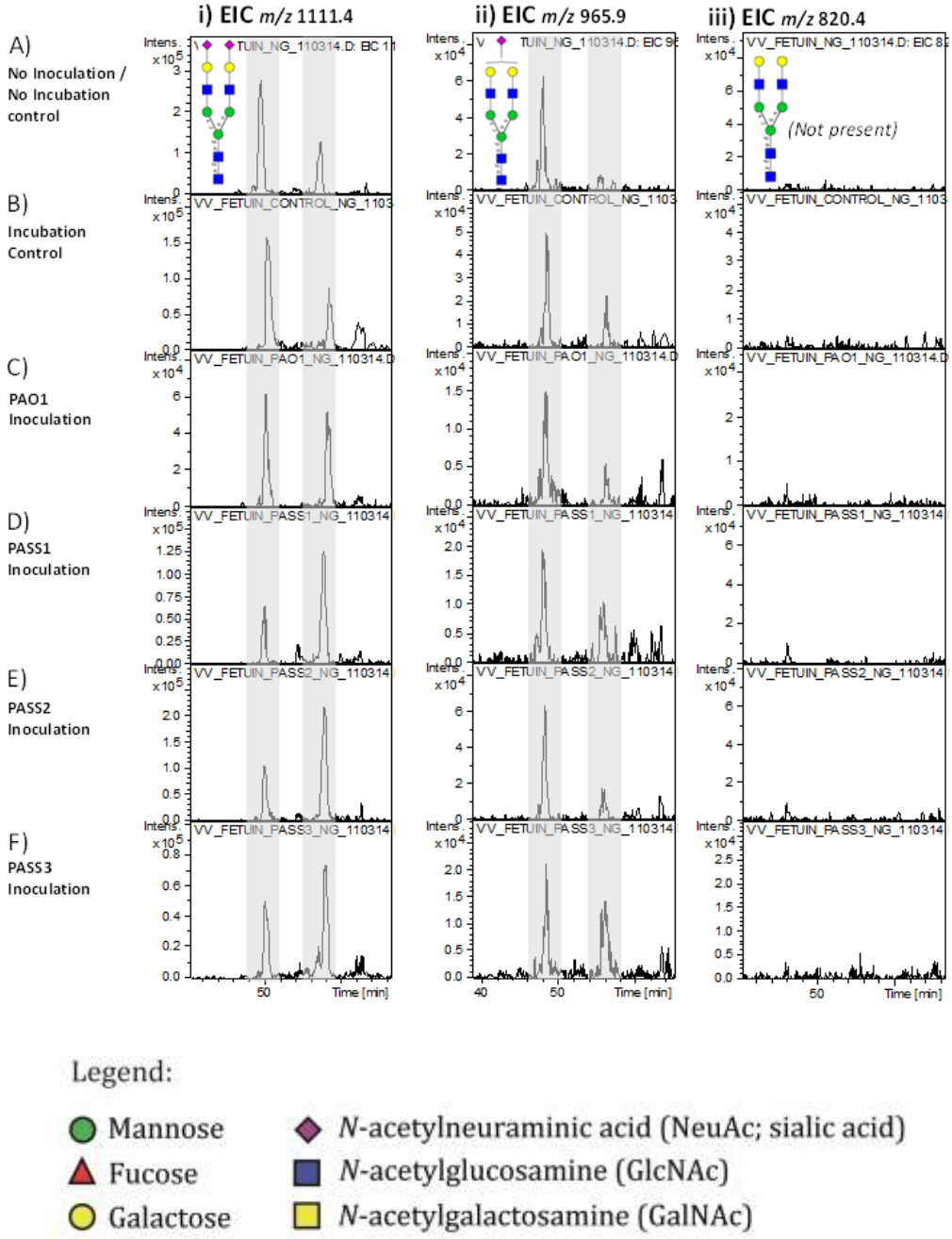


Figure 3: Extracted Ion Chromatogram (EIC) (*m*/*z* 1111.4, *m*/*z* 965.9 and *m*/*z* 820.4) of three different *N*-glycans released from purified fetuin protein in two controls and four samples incubated overnight with *P. aeruginosa* PASS1, PASS2, PASS3 and PAO1. Column i and ii show the EIC (*m*/*z* 1111.4 and *m*/*z* 965.9) of two masses of *N*-glycans released from fetuin. The lack of peak in Column iii (EIC of *m*/*z* 820.4), the asialo form, suggests the absence of possible neuraminidase enzymatic action.

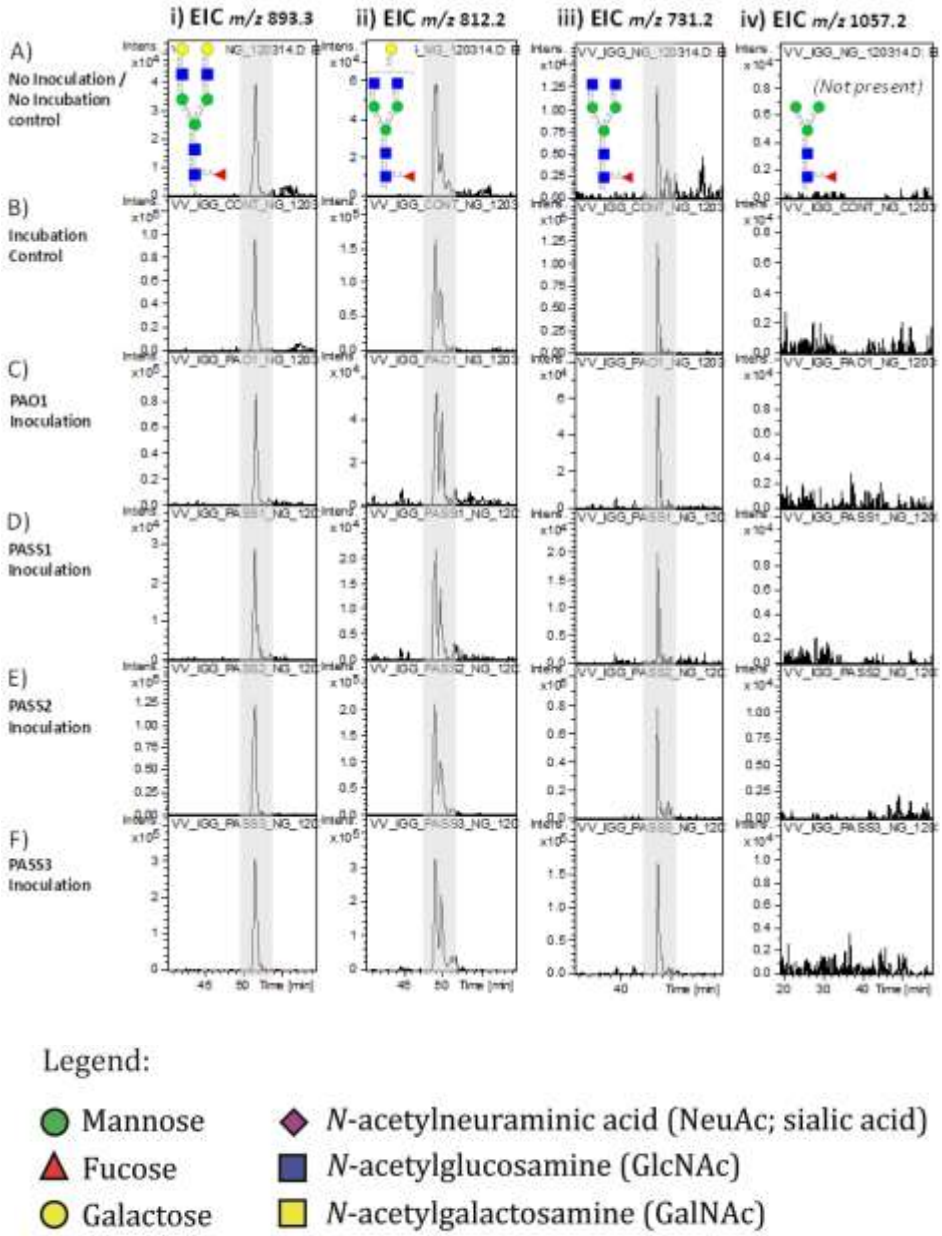


Figure 4: Extracted Ion Chromatogram (EIC) of four different *N*-glycans released from purified IgG protein incubated with *P. aeruginosa* strains PASS1, PASS2, PASS3 and PAO1 overnight. Columns i, ii and iii show the EIC of three masses (*m*/*z* 893.3, *m*/*z* 812.2 and *m*/*z* 731.2) released from IgG. The lack of peak in Column iv (EIC of *m*/*z* 1057.2) suggests the absence of possible galactosidase and hexosaminidase enzymatic actions.

Table 4: Average relative abundances (%) and standard deviation of galactose, GlcNAc and Sialic acid terminating *N*-glycans obtained from IgG (Control), IgG incubated at 37°C (Incubation control) and IgG incubated at 37°C and inoculated with PASS1, PASS2, PASS3 and PAO1 *P. aeruginiosa* strains. P-value was calculated by one-way ANOVA test showing no statistical variation ($p < 0.05$) among different complex *N*-glycan classes. Data measured in duplicates for control (n=2) and *P. aeruginosa* challenged strains (n=4).

	Control (Relative abundances)		**Incubation Control (Relative abundances)**		**PASS1 (Relative abundances)**		**PASS2 (Relative abundances)**		**PASS3 (Relative abundances)**		**PAO1 (Relative abundances)**		**One-way ANOVA ($p < 0.05$)**
	Average (%)	**SD**	**Average (%)**	**SD**	**Average (%)**	**SD**	**Average (%)**	**SD**	**Average (%)**	**SD**	**Average (%)**	**SD**	
Galactose terminating	64.21	10.7	57.26	4.92	48.51	10.1	61.35	17.68	63.67	4.81	46.27	9.48	0.452
GlcNAc Terminating	17.71	6.49	30.77	5.44	35.63	19.33	24.56	16.01	23.6	7.21	33.04	8.46	0.667
Sialic acid terminating	18.08	4.24	11.97	0.52	15.86	9.23	14.1	1.67	12.73	12.02	20.69	1.03	0.748

Table 5: Average relative abundances (%) and standard deviation of Man5 to Man9 *N*-glycans obtained from RNase B (Control), RNase B incubated at 37°C (Incubation control) and RNase B incubated at 37°C and inoculated with PASS1, PASS2, PASS3 and PAO1 *P. aeruginiosa* strains. P-value was calculated by one-way ANOVA test showing no statistical variation ($p < 0.05$) among different mannose type *N*-glycans. Data measured in duplicates for control (n=2) and *P. aeruginosa* challenged strains (n=4).

	Control		**Incubation Control**		**PASS1**		**PASS2**		**PASS3**		**PAO1**		**One-way ANOVA ($p < 0.05$)**
	Average (%)	**SD**	**Average (%)**	**SD**	**Average (%)**	**SD**	**Average (%)**	**SD**	**Average (%)**	**SD**	**Average (%)**	**SD**	
Man5	33.63	6.80	38.04	0.51	24.63	4.18	28.09	2.09	30.88	3.37	24.12	2.74	0.061
Man6	26.43	8.00	20.96	0.23	17.29	0.38	24.19	3.24	24.75	1.40	20.00	6.44	0.411
Man7	7.93	1.22	6.56	0.30	9.61	0.05	8.62	1.94	6.81	2.54	11.93	0.33	0.06
Man8	20.78	2.49	21.36	7.15	31.97	0.83	26.41	3.58	23.23	3.87	26.64	12.10	0.534
Man9	11.22	2.51	8.08	1.12	16.48	5.45	12.70	0.48	14.33	3.29	17.31	2.59	0.139

To further investigate whether *P. aeruginosa* contributes to the trimming of the sputum *N*-glycans by secreting a mannosidase enzyme, the bacterial cells were incubated with purified RNase B which contains Man5 to Man9 high mannose *N*-glycans. After incubating *P. aeruginosa* strains with RNase B, the average relative abundances (%) of Man5 to Man9 *N*-glycans were calculated in technical duplicates (**Table 5**). One-way ANOVA analysis of each *N*-glycan (Man5 -Man9) showed there was no statistically significant difference in the high mannose *N*-glycans between incubation with any of the *P. aeruginosa* strains. In **chapter 2**, it was seen that the relative ratios of the high mannose *N*-glycans were not altered between CF, pathogen-colonised iNCF and pathogen-free NCF sputum proteins which by itself indicates no mannosidase activity by the presence of pathogens.

Apart from the purified glycoproteins such as fetuin, IgG and RNase B, *P. aeruginosa* cells were incubated with the non-infected (NCF) non-CF sputum which contains higher abundances (~82%) of complex type *N*-glycans and only traces (~3%) of pauci-mannosidic type *N*-glycans (**Chapter 2**). Bacterial cells were incubated with NCF sputum to investigate if bacteria trim the elongated complex, hybrid or highmannose type *N*-glycans in sputum and in turn increases the abundances of pauci-mannosidic *N*-glycans in *P. aeruginosa* challenged NCF sputum. The average mass spectra (retention time: 30-60 minutes) of "no inoculation/ no incubation" control sputum (**Figure 5A**), incubation control sputum (**Figure 5B**) and different *P. aeruginosa* challenged sputum (**Figure 5C, D, E and F**) is represented in **Figure 6**. It could be appreciated from **Figure 5C, D, E and F**, that neither there is an increase in the abundances of the pauci-mannosidic *N*-glycans after incubating the NCF sputum with PASS1, PASS2, PASS3 and PAO1 *P. aeruginosa* strains, nor a decrease in other abundant complex glycans, suggesting lack of exoglycosidase activity expressed by bacteria to trim the *N*-glycans in NCF sputum. Taken together, it appears that the trimmed *N*-glycans in infected sputum are not a result of the secretion of exoglycosidases by the colonising bacteria.

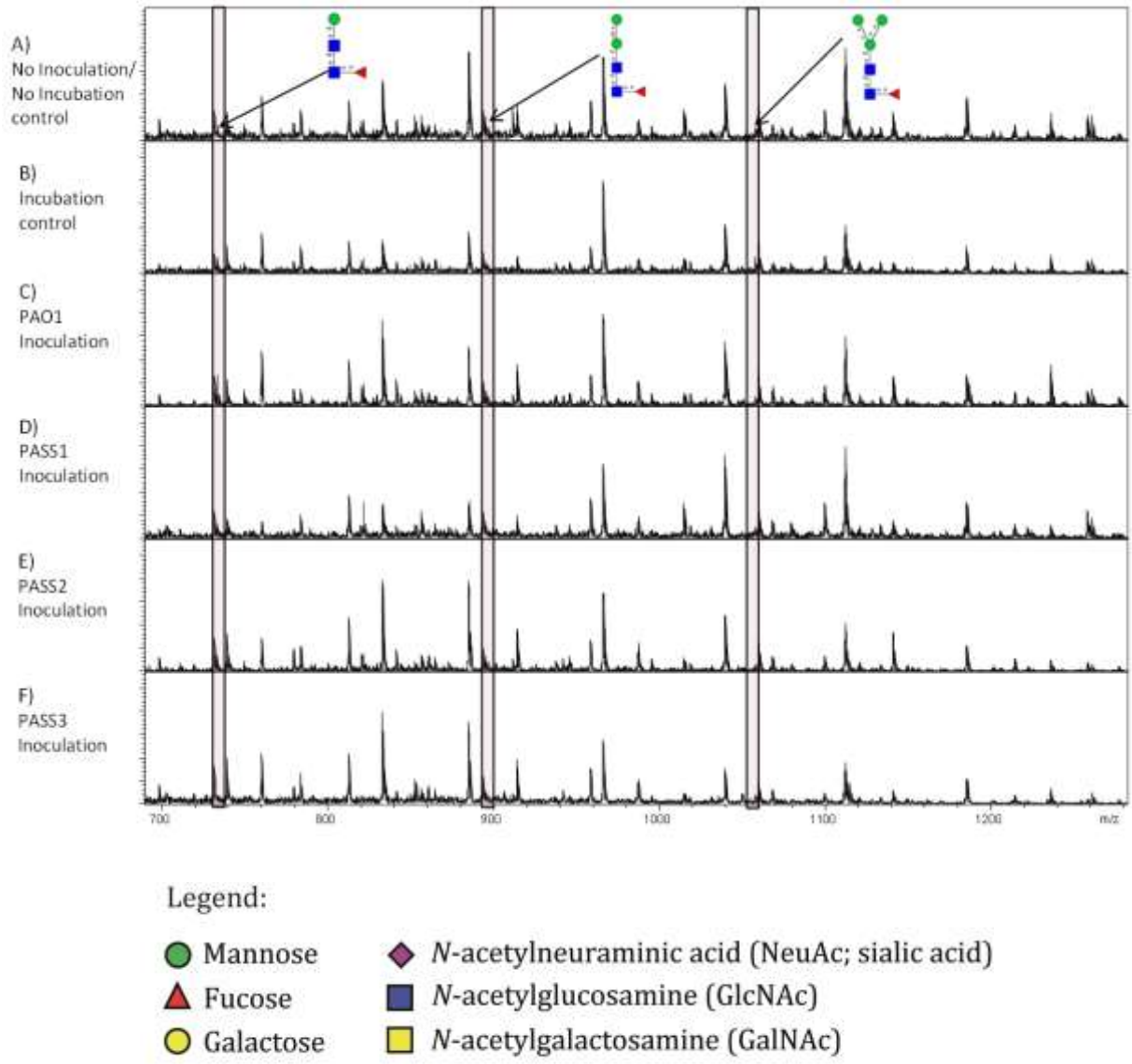


Figure 5: Average mass spectra (retention time: 30- 60 minutes) obtained from non-infected (NCF) non-CF sputum incubated with *P. aeruginosa* strains PASS1, PASS2, PASS3 and PAO1 overnight. Rows A and B, represents the average mass spectra of two controls and rows C, D, E and F, represents the average mass spectra of *P. aeruginosa* inoculated NCF sputum.

3.4.4 Qualitative mapping of whole cell *P. aeruginosa N*-glycans

Another possible explanation for the increased abundance of pauci-mannose structures in CF sputum is the possibility that the bacterial glycoproteins carry these *N*-glycans. The proteins of *P. aeruginosa* were obtained by lysing approximately 10^7 cells by bead beating and acetone precipitation. *N*-glycans were released from 10 μg of proteins using PNGase F with analysis by PGC-LC-ESI-MS/MS. The amount of bacterial protein used to release the *N*-glycans is the same as the amount of proteins used to analyse *N*-glycans from sputum. **Figure 6** shows the chromatogram of the *N*-glycans released from PAO1, PASS1 and PASS2 proteins. The three

chromatograms were compared with the sputum *N*-glycome obtained from a CF patient on a qualitative level to confirm if there were any common *N*-glycans present between *P. aeruginosa* and sputum. There were no common *N*-glycans observed in CF sputum and *P. aeruginosa* and the most abundant *N*-glycan seen in CF sputum (*m/z* 895.4, $Fuc_1Man_2GlcNAc_2$) was not present (**Figure 6**). Thus, the presence of the abundant pauci-mannosidic *N*-glycans in bacterial colonised sputum is not directly from the bacterial origin.

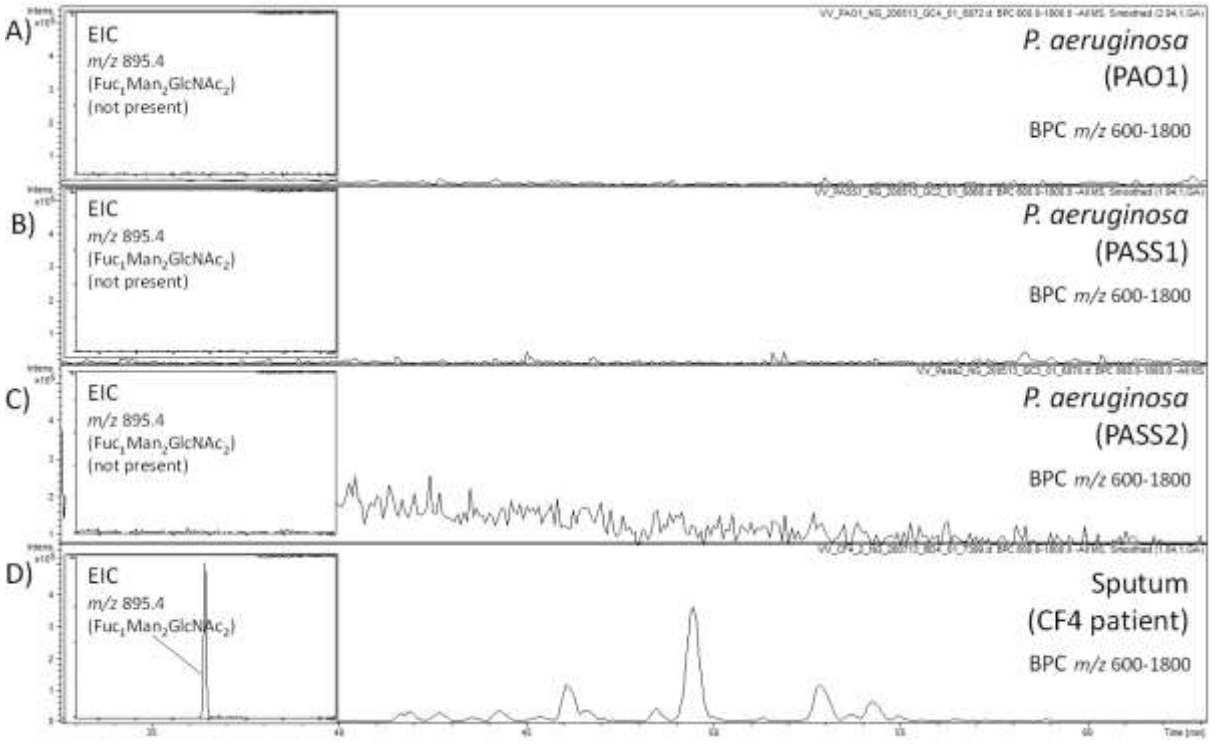


Figure 6: Base peak chromatogram (BPC) of the PNGase F released *N*-glycans of *P. aeruginosa* strains PAO1 (A), PASS1 (B) and PASS2 (C) compared with the CF sputum *N*-glycome (D) as a qualitative comparison suggesting the glycomes of sputum has not been contaminated with bacterial *N*-glycans. EIC of the most abundant pauci-mannose *N*-glycan in CF sputum ($GlcNAc_2Man_2Fuc_1$, m/z 895.4) is completely absent in *P. aeruginosa N*-glycans.

3.5 Discussion

The role of infection and pathogenicity of the infecting/ colonising pathogens are very important research areas for a disease like CF. Though mutation in the *CFTR* gene is the starting point of CF, inflammation and infection, especially by *P. aeruginosa*, increases the mortality and morbidity rate among CF patients. Even in co-infections, *P. aeruginosa* is the most dominant pathogen in CF airways, specifically in CF adult lung and colonises the lung by forming biofilms. This dominance was reflected in the primary isolation of *P. aeruginosa* from CF sputum samples and the four isolated clinical strains were compared in this study with the classic PA01 laboratory strain.

3.5.1 Absence of exoglycosidase activity secreted by *P. aeruginosa*

For *P. aeruginosa* to be involved directly in production of pauci-mannosidic *N*-glycans in infected human sputum, the secretion of at least four exoglycosidases is needed. Neuraminidase, β-galactosidase, β-hexosaminidase and α- and β-mannosidase are required to cleave off monosaccharides from terminally sialylated complex *N*-glycans to form pauci-mannosidic *N*-glycans. The *in vitro* assays involving incubation of *P. aeruginosa* strains with the standard *N*-glycan carrying glycoproteins fetuin, IgG and RNase B and non-infected (NCF) non-CF sputum, did not indicate the secretion of active exoglycosidases by any of the *P. aeruginosa* strains.

Previous studies have shown the presence of neuraminidase in bacterial species including *P. aeruginosa, S. pneumonia* (Camara et al. 1994; King et al. 2004) and *Hemophilus influenza* (Vimr et al. 2002) by isolating and characterising the neuraminidase protein responsible for neuraminidase activity. The purified PAO1 neuraminidase was shown to be more active than the *Clostridium perfiringens* neuraminidase in cleaving sialic acid off epithelial cells to expose the asialo receptors (Cacalano et al. 1992). Construction of a mutant *P. aeruginosa* strain by deletion of the Δ2794 neuraminidase locus caused the mutant *P. aeruginosa* strain PAO1 to be unable to colonise the respiratory tract or form biofilm *in vitro* compared to the wild type (Soong et al. 2006).

In contrast to the above studies, there was no significant neuraminidase activity secreted by *P. aeruginosa* clinical strains and the PAO1 strain in the current study. Previous studies

showing neuraminidase activity were carried out with a partially-purified neuraminidase or by constructing a mutant by deleting the neuraminidase gene locus, and the effect on *N*-glycoproteins compared to *O*-glycoproteins (Venkatakrishnan et al. 2013) was not differentiated. Possibly there is a limited or low activity of neuraminidase in *P. aeruginosa in vivo* that was not detected in our *in vitro* study. Nevertheless, to obtain the pauci-mannose *N*-glycan structures observed in the infected sputum in this study, would require other exoglycosidases as well. The expression of other exoglycosidases such as galactosidase, hexosaminidase and mannosidase has not been reported in any pathogenic bacteria infecting the respiratory tract. This is consistent with the findings in this study where there was no trimming of the standard purified glycoproteins that would reflect the activity of these enzymes. The presence of other bacterial exoglycosidase such as *S. aureus* could be ruled out, as most of the CF sputum used in this study were colonised with *P. aeruginosa* and increase in the pauci-mannosidic *N*-glycans is observed in iNCF sputum colonised with other bacteria. This suggests the presence of abundant pauci-mannose in sputum is more specific to infection and not specific to any single bacterium. Thus, it appears unlikely that bacterial exoglycosidases trim the complex *N*-glycans to produce the pauci-mannosidic *N*-glycans in bacterial colonised/ infected CF and iNCF sputum (**Chapter 2**).

3.5.2 Utilisation of monosaccharides as carbon source by *P. aeruginosa*

It is known that sputum is a complex mixture of host derived substances that provide substrates and nutrition for bacterial growth (Ohman et al. 1982; Son et al. 2007). Within the CF lung airway, a large proportion of infecting bacteria is able to proliferate and grow in dense communities (Costerton et al. 1999; Singh et al. 2000). Individual carbon sources have also been shown to modulate *P. aeruginosa* surface motility and *in vitro* biofilm formation (Klausen et al. 2003; Shrout et al. 2006).

This study indicates that most of the sugars or monosaccharides comprising the *N*-glycans and also the *O*-glycans on mucins are not used by clinical and laboratory *P. aeruginosa* strains when provided as a sole carbon source for cellular respiration. Glucose was used by all strains except PASS4 as a sole carbon source as it provides the carbon in most living organisms for growth and metabolism. It is involved in various metabolic pathways as metabolic intermediate and used as a secondary source of energy. Thus, utilisation of glucose

as a sole carbon source was well documented and an expected outcome. This was further verified in our laboratory by Karthik Kamath (unpublished observation) that, *P. aeruginosa* strains PASS1, PASS3 and PAO1 had faster growth rate in minimal medium supplemented with glucose as compared to PASS2 strain, which had slower growth in the same media (data not shown). *P. aeruginosa* strain PASS4 on the other hand did not grow in minimal medium supplemented with glucose and from Biolog phenotype microarray, it could be noted that it did not utilise any of the monosaccharides as a carbon source including glucose.

Amongst the seven monosaccharide building blocks only *N*-acetylglucosamine (GlcNAc) could be used as a carbon source for cellular respiration by PASS1, PASS3 and PAO1. Earlier studies have revealed that genes responsible for the catabolism of GlcNAc were highly up-regulated during *in vitro* growth in CF sputum (Palmer et al. 2005). An *in silico* analysis of the *P. aeruginosa* strain PA14 genome (Mikkelsen et al. 2011) showed that the *P. aeruginosa* GlcNAc catabolism gene was organised as a part of an operon. The same research group have also shown that certain virulence factor encoding genes in *P. aeruginosa* were up-regulated during the growth on GlcNAc proving the involvement of GlcNAc in inducing these genes (Korgaonkar et al. 2011). Similar kinds of virulence factors inducing activity of GlcNAc metabolism and the role in cell signalling were also observed in the human fungal pathogen *Candida albicans* (Naseem et al. 2012). This suggests, that GlcNAc utilised by *P. aeruginosa* as a carbon source could be involved in some kind of bacterial virulence.

The biofilm formation of the clinical *P. aeruginosa* (PASS1, PASS2, PASS3 and PASS4) and laboratory strain PAO1 was assessed in our laboratory by Dr.Anahit Penesyan (unpublished observation), to investigate the virulence level of the bacteria. It was performed using a flow cell system and visualised on laser scanning confocal microscope after 48 hours of growth in LB medium. *P. aeruginosa* strains PASS1, PASS3 and PASS4, as well as PA01, were capable of biofilm formation whereas PASS2 did not form biofilm. It is interesting to note that the utilisation of GlcNAc by PASS1, PASS3 and PAO1 correlated with the capacity of biofilm formation, which may be linked with an induced virulence factor. The role of GlcNAc in pathogenicity and virulence of the fresh isolates requires further study. Although PASS4 strain did not utilise any of the monosaccharide as sole carbon source, it had higher respiration levels in utilising adenosine and inosine (DNA components) compared to other

P. aeruginosa strains and the capability of PASS4 to utilise adenosine and inosine could be linked with the biofilm formation. Extracellular DNA has been shown to play a role in *P. aeruginosa* biofilm as cell-to-cell interconnecting compound and DNA is reported to make up a substantial fraction of the matrix material in *P. aeruginosa* biofilms (Whitchurch et al. 2002; Matsukawa et al. 2004; Allesen-Holm et al. 2006).

In summary, Biolog phenotype microarray data suggest that *P. aeruginosa* is not using the trimmed monosaccharides that would result from a trimming of the sputum *N*-glycans as a sole source of carbon for respiration. The possibility of these monosaccharides enhancing established bacterial growth cannot be ruled out.

3.5.3 *P. aeruginosa N*-glycans

The *N*-glycans were released from *P. aeruginosa* PASS1, PASS2 and PAO1 strains using PNGase F enzyme to make sure the truncated pauci-mannosidic *N*-glycans were not of bacterial origin. Glycosylation was previously considered to be restricted to eukaryotes but advances in analytical techniques for structural elucidation and genome sequencing have resulted in increased reports of both *N*- and *O*-linked glycosylation in bacteria. *N*-linked protein glycosylation in bacteria and mainly in the mucosal pathogens has been hypothesised to be involved in lipopolysaccharide biosynthesis (Fry et al. 1998). Glycosylation in bacteria including *P. aeruginosa* is shown to be different to that of human glycosylation (Castric et al. 2001; Hitchen et al. 2006; Voisin et al. 2007). In a widely studied organism C*ampylobacter jejuni*, alterations in *N*-linked glycosylation have been shown to influence adhesion, invasion and colonisation of the bacteria (Szymanski et al. 1999; Szymanski et al. 2002). The *C. jejuni N*-glycosylation involves unusual bacillosamine (2,4-diacetamido-2,4,6-trideoxy-d-glucopyranose) as the core monosaccharide residue linked to the *N*-glycosylation consensus sequence Asn-X-Ser/Thr unlike the eukaryote Asn which is linked to GlcNAc (Young, Brisson et al. 2002; Amin, Ishiwata et al. 2006). The specificity of PNGase F enzyme does not allow the enzyme to cleave the sugars that does not contain GlcNAc attached to the glycopeptides, which could be the reason for not observing any possible *P. aeruginosa N*-glycan in this study. A review on protein glycosylation in bacterial mucosal pathogens has summarised the role of bacterial glycosylation in pathogenesis of the infecting bacteria with the human host (Szymanski et al. 2005). The absence of any *N*-glycans

released by PNGase F from the *P. aeruginosa* establishes that the pauci-mannosidic *N*-glycans observed in the bacterial colonised/ infected sputum were not derived either from *P. aeruginosa* or bacterial proteins.

3.6 Conclusion

In the previous Chapter (**Chapter 2**) we observed a significantly increased abundance of pauci-mannosidic *N*-glycans in sputum samples from lungs that were colonised or infected by bacteria. The main focus of this chapter was to examine whether infective bacteria are involved directly in the formation of these unusual *N*-glycans on the sputum proteins. The data in this chapter have suggested that the clinical isolates of bacteria from these sputa appear not be directly involved with the production of pauci-mannosidic *N*-glycans in terms of exoglycosidase secretion, carbon usage or bacterial protein glycosylation.

Chapter 4

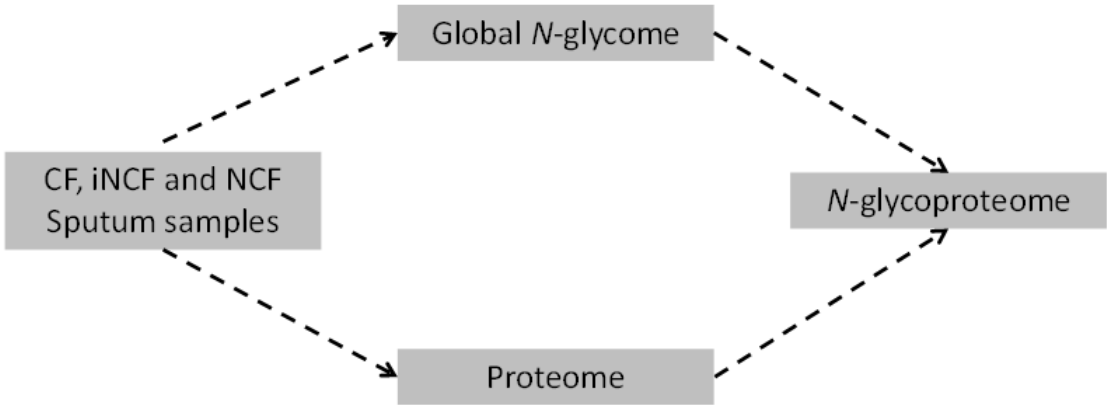
Proteome and *N*-glycoproteome of Cystic Fibrosis sputum links the pauci-mannosidic *N*-glycans with abundant neutrophil proteins

Contribution: In this chapter, all experimental work and data analysis was performed by me.

4.1 Rationale

A unique *N*-glycosylation phenotype was observed in bacterial colonised CF and infected non-CF (iNCF) sputum. Increase in the truncated pauci-mannosidic *N*-glycans and decrease in complex and hybrid type *N*-glycans in CF and iNCF compared to pathogen free non-CF (NCF) sputum suggested a possible role played by bacteria in regulating the sputum *N*-glycosylation (**Chapter 2**). Pauci-mannose formation as a result of trimming by bacterial exoglycosidases such as neuraminidase, galactosidase and hexosaminidase was subsequently ruled out in the work described in the previous chapter (**Chapter 3**). Also, the whole cell *N*-glycome of clinical *P. aeruginosa* isolates (PASS1, PASS2) and laboratory *P. aeruginosa* isolate (PAO1) suggested pauci-mannosidic *N*-glycans were not of bacterial origin.

The main focus of this chapter was to perform site-specific *N*-glycoproteome analysis to identify the proteins carrying the unique pauci-mannosidic *N*-glycans in bacterial colonised/ infected sputum using the already available knowledge of sputum *N*-glycome (**Chapter 2**) and the here presented global proteome analysis of CF, iNCF and NCF sputum.



4.2 Introduction

Proteomics is the large scale determination of proteins as part of the study of gene and cellular functions at the protein level. The important feature of proteomics is its ability to identify each protein expressed by a cell or a tissue and, further, to understand the properties of those proteins. Proteomic analysis helps in investigating the protein profiles of different cells and biofluids such as epithelial cells and saliva, identification of novel biomarkers for disease diagnosis and also drug discovery for targeted drug approaches (MacGregor et al. 2008; Chen et al. 2010; Ali et al. 2011; Charro et al. 2011; Ohlmeier et al. 2012). The developments in analysing complex protein samples together with the continuous advancement in the field of mass spectrometry over the past two decades have favoured proteomics for biomarker discovery.

Limited numbers of proteomic studies have focused on analysing samples from patients with airway diseases including cystic fibrosis, chronic obstructive pulmonary disorder (COPD), asthma and upper respiratory tract infection (URTI) compared to other diseases such as cancer. The analysis of proteins identified from airway samples such as sputum, lung tissues and bronchoalveolar lavage (BAL) fluid, might lead to understanding the airway inflammation and the progression of the disease.

Sputum sampling, as a non-invasive method that closely reflects the central lung airway secretions including epithelial, submucosal gland and mucous secretions, makes it an ideal choice for proteomic analysis of the lungs (Sagel et al. 2001; Nicholas et al. 2006). Sputum is rich in high molecular weight mucin (MUC5B and MUC5AC) glycoproteins (Schulz et al. 2007). Apart from mucins, there are other proteins such as inflammatory and epithelial proteins in the sputum proteome that could be used as disease markers (Nicholas et al. 2006). The sputum proteome of CF or other airway diseases is not a widely studied area in the field of proteomics and the literature available is limited.

Proteomic analysis of induced sputum in CF and other lung diseases has shown an up-regulation of inflammatory proteins and proteins induced by the immune response (Sagel et al. 2001). Most of the studies that analysed the proteins from sputum have utilised the conventional 2-D gel electrophoresis (2-DE) method for the separation of proteins. 2-DE was

an effective separation technique for arraying and characterising the protein expression in BAL fluid and sputum (Griese et al. 2001; von Bredow et al. 2001). The first study to report the differential sputum proteome between CF and non-CF subjects used 2-DE for separation and identified differentially expressed neutrophil proteins including myeloperoxidase and interleukin-8 as biomarkers of inflammation relating to pulmonary exacerbations (Sloane et al. 2005).

Amongst the limited number of studies that have explored the CF proteome, none have included an analysis of the site-specific *N*-glycosylation of the identified proteins. Having identified the *N*-glycosylation variation in CF sputum compared to NCF sputum in **Chapter 2**, the proteins to which the changes are associated are of interest to be investigated. Performing a standard release and characterisation of glycans or identifying the glycoproteins and the possible sites of attachment from a biological sample does not provide site-specific information on protein glycosylation (Parker et al. 2013). The main task of a global glycoproteome analysis is to connect the possible glycan attachment site(s) of a protein with the specific glycosylation counterpart. Unlike purified glycoprotein analysis, analysing complex protein samples like sputum becomes much more difficult (Kolarich et al. 2006; Korekane et al. 2011; Sumer-Bayraktar et al. 2011). The challenge is mainly associated with detection, quantitative measurements and structural characterisation of glycopeptides which are in low abundance in the total proteolytic digest commonly used in proteomic analysis.

With recent technological advancements in MS and tandem MS fragmentation, the glycoproteome analysis of a complex protein mixture is becoming relatively easier. One of the recent studies by Parker *et al.* (2013), involved a site-specific glycopeptide analysis of rat brain membrane extracts using a combination of glycomics and glycoproteomics and confidently identified 863 unique intact *N*-linked glycopeptides from 161 glycoproteins employing CID/HCD, ETD fragmentation (Parker et al. 2013). Another study by Yin *et al.* (2013) performed glycoproteomic analysis of the human endothelial cell secretome further demonstrating the potential of HCD-ETD fragmentation (Yin et al. 2013). Thus, the combination of different fragmentation techniques has been shown to be useful in the structural characterisation of the glycoproteome.

The main aim of the work described in this chapter was to identify the major glycoproteins in CF and non-CF sputum that carry pauci-mannosidic *N*-glycans and to investigate whether the alteration in *N*-glycosylation profile observed in **Chapter 2** correlated with specific proteins in the proteome profile. The work involved a shotgun proteomic analysis to identify the differentially regulated proteins between CF, iNCF and NCF sputum, and an *N*-glycoproteome analysis of the abundant glycoproteins in the sputum using the knowledge gained from *N*-glycome characterisation (**Chapter 2**).

4.3 Materials and methods

4.3.1 Sputum collection and patient information

Sputum collection and patient data including gender, age, CF genotype and microbiology culture results of sputum are provided in **Chapter 2, Table 1**. To summarise, whole sputum samples from five CF patients (CF1-CF5), and four individuals who either had no underlying lung conditions or had other lung conditions (collectively called non-CF) were collected by non-invasive expectoration of sputum at Westmead Hospital, Sydney, Australia. The non-CF group was sub-divided according to their clinical presentation and pathology: two individuals had clinical and radiological evidence of pneumonia (NCF1) or chronic obstructive pulmonary disease (COPD; NCF2) (collectively called NCF), but did not contain detectable pathogens in their sputum; and two pathogen-infected/colonised individuals (collectively called iNCF) were diagnosed with an upper respiratory tract infection (URTI, iNCF1-2).

4.3.2 Isolation of soluble sputum proteins

Initial sample preparation involving isolation of soluble sputum proteins is described in the material and methods section of **Chapter 2, materials and methods section**. In short, sputum proteins were reduced, alkylated and centrifuged to remove intact cells and cell debris. The protein concentrations were measured using the Direct Detect protein quantitation system (Millipore).

4.3.3 Peptide preparation and glycopeptide enrichment

Glycopeptide enrichment from tryptic digests of sputum proteins was performed as described previously (Parker et al. 2013). One milligram of soluble sputum protein from nine sputum samples was digested using 25 mM ammonium bicarbonate (final concentration) and 0.5% (w/w) trypsin, incubated for 2 hours at 37°C. Then another round 0.5% (w/w) trypsin was added and incubation continued for 12 hours at 37°C. $MgCl_2$ was added to a final concentration of 1 mM and the digested peptides were dephosphorylated with 5 U/µl antartic phosphatase (New England, Biolabs) for 3 hours at 30°C. The samples were acidified to 2% (v/v) formic acid. The samples were desalted with hydrophilic lipophilic balance solid

phase extraction (Waters) and the peptides were eluted using 80% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid (TFA). Approximately 15 μg of peptide mixture from each sputum sample was aliquoted, dried by vacuum centrifugation for proteomic analysis. Samples were taken up in 10μl 0.1% (v/v) formic acid and analysed on RP-LC-MS/MS.

Remaining peptides were loaded onto an in-house zwitterionic Zic-HILIC SPE cartridge containing 80 mg of Zic-HILIC material (Sequant/Merck) packed onto a C8 disk (Empore). The flow-through was collected and loaded again through the column. The column was then washed twice with 1 ml of 80% (v/v) acetonitrile and 1% (v/v) TFA. Enriched glycopeptides were eluted with 1 ml 0.1% (v/v) TFA followed by 100 μl of 25 mM ammonium bicarbonate and finally 100 μl of 50% (v/v) acetonitrile and dried by vacuum centrifugation. The samples were taken up in 10 μl of 0.1% (v/v) formic acid and 5 μl was injected for LC-MS/MS analysis. Intact *N*-glycopeptides were separated by reverse phase chromatography on an C-18 Easy-spray column (2 μm, 100Å, 75 μm x 500 mm) with an integrated PicoFrit ESI emitter (New Objective) using HPLC (Easy LC-II). The HPLC gradient was 0-40% solvent B (solvent A = 0.1% formic acid; solvent B = 90% acetonitrile and 0.1% formic acid) over 165 minutes at a flow of 250 nl/minute. Parameters used were as follows: capillary temperature = 300°C, source voltage = 1.7 kv, and S-lens RF level = 69%. MS detection was achieved using an LTQ-Orbitrap Elite (Thermo scientific). An MS scan (*m*/*z* 700-2500) was recorded in Orbitrap set at a resolution of 60,000 at *m*/*z* 400 followed by data-dependant CID (detection in the ion trap) and HCD (detection in the Orbitrap) MS/MS analysis of the seven most intense precursors. Parameters for acquiring CID were: activation time = 10 ms, and normalised collision energy = 35 and parameters for acquiring HCD were: activation time = 0.1 ms, normalised collision energy = 35, and resolution = 15000.

4.3.4 *N*-glycan analysis of human neutrophils

Resting neutrophils were kindly donated by Dr. Niclas G. Karlsson, University of Gothenburg, Sweden. Neutrophils were isolated to high purity from buffy coat of healthy individuals as described in (Boyum et al. 1991; Jin et al. 2012). In brief, the neutrophils were obtained using dextran sedimentation first at 1 x g to remove RBCs, followed by hypotonic lysis to remove remaining RBCs and platelets and finally centrifugation in a Ficoll-Paque gradient, in which neutrophils were isolated from the bottom of the differing Ficoll gradient.

Obtained neutrophils were washed twice in Krebs-Ringer phosphate buffer. The concentration of neutrophil cells were counted using a hemocytometer and were resuspended in Krebs-Ringer phosphate buffer and stored on ice as resting cells until use. Approximately 50 µl of the neutrophil cells (2.2 x 10^7/ml) were centrifuged at 4000 x g for 5 minutes to pellet the neutrophil cells, which was followed by cell lysing with 300 µl of 6M urea and 0.1% SDS. The lysed cells were then mixed thoroughly using a rotary shaker in a cold room for 2 hours and the soluble proteins were collected after centrifugation for 10 minutes at 4000 x g. 15 µg of soluble whole cell proteins were spotted on PVDF membrane and the *N*-glycans were released enzymatically using PNGase F and analysed using PGC-ESI-MS/MS as described in **Chapter 2, materials and methods section**.

4.3.5 Data processing and analysis

The proteomic data analysis of the raw data was performed with MaxQuant software (version 1.4.1.2), supported by Andromeda as a database search engine for peptide identification as explained in (Cox et al. 2011). Time-dependent mass calibration was performed on the raw MS files from nine sputum samples (5 CF, 2 iNCF and 2 NCF) with technical duplicates and the resulting accurate precursor masses were searched against the IPI human database (version 3.68, 87, 061 entries), complemented with the standard MaxQuant contaminant database. The Andromeda search included variable modifications of methionine oxidation and N-terminal acetylation and fixed modification of cysteine carbamidomethylation. Enzyme specificity was set as C-terminal to Arg and Lys and allowing a maximum of two mass cleavages. The false discovery rate (FDR) was set at 1% for peptides and proteins and the minimum peptide length to 6 amino acids. In case the identified peptides of two proteins were shared by two homologous proteins, the two proteins were reported as one by MaxQuant as one protein group. The normalised intensity value for each of the identified peptides obtained using MaxQuant were used for statistical analysis using Perseus statistical analysis software (Version 1.4.1.3) to identify the regulated proteins from different sputum groups based on student t-test (p-value < 0.05).

The list of all the identified regulated proteins, p-values (<0.05) and fold change (±1.5 fold) was uploaded to Ingenuity Pathway Analysis (IPA) to map proteins into biological networks and to retrieve functions and pathways. In IPA, the regulated proteins are mapped to genetic

networks linked to the Ingenuity database and ranked based on p score. The significance values for the network and pathway analyses were calculated using right-tailed Fisher's exact test. Panther classification system (free tool) was used to classify the identified proteins according to the biological processes, molecular functions and protein classes. The Panther website includes a suite of tools to analyse large-scale experimental data with statistical tests. Finally the coefficient of determination (R^2) was calculated using Microsoft Excel to determine the relationship between the relative abundances of *N*-glycans obtained from CF sputum and neutrophil proteins. A higher coefficient value (>0.90) is an indicator of better goodness of fit for the observations.

4.4 Results

4.4.1 Proteome analysis

4.4.1.1 Rationale and experimental design

The aim of this experiment was to identify the soluble sputum proteins from individuals with and without CF. Non-CF sputum was classified into two categories, pathogen infected/ colonised (iNCF) sputum and pathogen-free (NCF) sputum. Soluble sputum proteins were isolated from whole sputum and the proteins were trypsin digested and desalted using HLB solid phase extraction. Reversed phase (RP)-LC-MS/MS was used for the separation and analysis of the peptides and the peptides were searched against the human protein database using the Andromeda search engine integrated with MaxQuant quantitation software. Each of the samples were analysed in technical duplicates by MS for peptide identification. Normalised intensity value calculated based on area under the curve for each of the identified peptides was obtained using MaxQuant and further statistical analysis was performed using the Perseus statistical tool.

4.4.1.2 Overview of the sputum proteomes of CF and non-CF patients

In total, 721 non-redundant proteins were identified using the IPI human database from five CF and four non-CF sputum MS datasets. The identified proteins were filtered based on their presence in each of the sputum categories *i.e.* CF, iNCF and NCF. The filtering strategy employed here was to include only those proteins that were identified at least twice in at least one of the sputum types. Thus, a protein identified at least twice from the five CF (CF 1-

5) or two iNCF (iNCF 1-2) or two NCF (NCF 1-2) sputum proteomes was considered a hit and thus included for analysis. After filtering, the number of non-redundant proteins identified from CF, iNCF and NCF sputum was reduced to 438 non-redundant proteins and the entire lists of 438 non-redundant, identified from nine sputum samples used for quantitation can be found in **Supplementary table 1**.

The proteins identified in each of the sputum categories were compiled and compared using a Venn diagram to represent the commonalities and uniqueness among CF, iNCF and NCF sputum proteins (**Figure 1**). The Venn diagram was created using an online tool (http://bioinfogp.cnb.csic.es/tools/venny/index.html). A total of 85, 8 and 84 proteins were found to be uniquely identified in CF, iNCF and NCF sputum proteomes, respectively, and 147 proteins were found to be common between all three sputum categories (**Figure 1**).

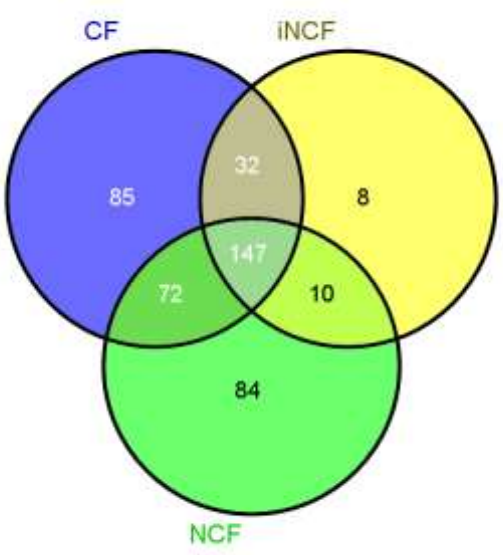


Figure 1: Venn diagram representing the number of common and unique proteins identified from CF, iNCF and NCF sputum.

4.4.1.3 Comparison of the proteomes of CF and infected non-CF (iNCF) sputum

The first comparison was between the CF and iNCF sputum proteomes. As mentioned in Chapter 2, the difference between CF and iNCF sputum identifies the role played by the CF disease in regulating the sputum proteome. A similar filtering criterion as discussed above was used for identifying a protein hit, *i.e.* the protein must be present twice out of five CF

sputum samples or present in both the iNCF sputum samples. A total of 354 non-redundant proteins were identified from the CF and iNCF sputum proteomes combined, of which 336 proteins were identified in CF and 197 proteins were identified in iNCF sputum.

To identify the proteins that were differentially regulated between CF and iNCF, the normalised intensity value of each identified peptides calculated by MaxQuant quantitation analysis software was used and the regulated proteins were identified using Perseus statistical analysis software. To remove any bias, the intensities were $\log_2$ transformed and those $\log_2$ transformed values for each protein were used to identify the regulated proteins and their fold change by performing two sample t-test. Those p values that were less than 0.05 were considered statistically significant and minimum fold change values of plus or minus two were considered to be regulated. A total of 16 proteins were regulated between CF and iNCF sputum groups, out of those, seven proteins were up-regulated and nine proteins were down-regulated in CF sputum relative to iNCF. The list of regulated proteins, Uniprot ID, gene name, p value and fold change value are given in **Table 1**. Similar to that observed in the analysis of the sputum *N*-glycomes, there was little difference between CF and iNCF sputum proteomes.

Table 1: Proteins that are up- and down-regulated in CF sputum compared to iNCF sputum.

#	Uniprot ID	Protein names	Gene ID	Fold Change	Glyco-protein
1	P16930	Fumarylactoacetase	FAH	25.04	N
2	Q08188	Protein-glutamine gamma-glutamyltransferase E	TGM3	24.47	N
3	Q70J99	Protein unc-13 homolog D	UNC13D	23.71	N
4	P05089	Arginase-1	ARG1	4.67	N
5	P06744	Glucose-6-phosphate isomerase	GPI	3.59	N
6	Q6P4A8	Phospholipase B-like 1	PLBD1	3.44	Y
7	P01766	Ig heavy chain V-III region BRO		2.53	Y
8	P08238	Heat shock protein HSP 90-beta	HSP90AB1	-2.96	Y
9	P07237	Protein disulfide-isomerase	P4HB	-3.74	N
10	P04632	Calpain small subunit 1	CAPNS1	-18.72	N
11	Q5CZC0	Fibrous sheath-interacting protein 2	FSIP2	-19.35	N
12	P25705	ATP synthase subunit alpha, mitochondrial	ATP5A1	-23.20	N
13	P02749	Beta-2-glycoprotein 1	APOH	-23.28	Y
14	P01019	Angiotensinogen	AGT	-23.41	Y
15	P07858	Cathepsin B	CTSB	-23.82	Y

16	P08575	Receptor-type tyrosine-protein phosphatase C	PTPRC	-24.00	Y

Datasets representing proteins with altered expression obtained by Perseus statistical analysis software were imported into the Ingenuity Pathway Analysis (IPA) tool. The list of differentially expressed proteins in CF and iNCF sputum were analysed by IPA tool and inflammatory response was suggested to be the main disease and disorder linked with the regulated proteins.

4.4.1.4 Comparison of the proteomes of iNCF and NCF sputum

The second level of comparison was to identify the variation between infected (iNCF) and non-infected (NCF) non-CF sputum, which investigates the role played by the bacterial infection/ colonisation in regulating sputum proteome. A similar filtering strategy was used, *i.e.* a protein had to be identified twice in either iNCF (iNCF 1-2) sputum or NCF (NCF 1-2) sputum to be considered as a hit. A total of 353 non-redundant proteins were identified from iNCF and NCF sputum combined, with 157 identified proteins common to both the sputum types. In NCF sputum (NCF 1-2), 313 proteins were identified in total compared to 197 proteins in iNCF sputum.

In total, 90 proteins were identified to be differentially regulated between iNCF and NCF sputum, with 22 proteins up-regulated and 68 proteins down-regulated in iNCF. The differentially regulated proteins are listed in **Table 2**, along with their Uniprot ID, gene name and fold change. The list of differentially expressed proteins in iNCF and NCF were analysed by IPA tool and *inflammatory disease and inflammatory response* were suggested to be the top diseases and disorders involving the differentially expressed proteins. IPA analysis here and in comparison between CF and iNCF, supports the expected outcome of sputum proteins being involved in the immune response to infection and inflammation.

Table2: Proteins that are up- and down-regulated in iNCF sputum compared to NCF sputum.

#	Uniprot ID	Gene ID	Protein names	Fold Change	Glyco-protein
1	O75131	CPNE3	Copine-3	28.63	N
2	J3QSF7	MPO	Myeloperoxidase	28.12	Y
3	P06733-2	ENO1	Alpha-enolase	24.81	N
4	P11678	EPX	Eosinophil peroxidase	24.76	Y

5	P16949	STMN1	Stathmin	24.48	N
6	P48637	GSS	Glutathione synthetase	24.42	N
7	Q9H4A4	RNPEP	Aminopeptidase B	24.35	Y
8	P08575	PTPRC	Receptor-type tyrosine-protein phosphatase C	24.00	Y
9	P16035	TIMP2	Metalloproteinase inhibitor 2	23.40	N
10	E7ER45	MGAM	Maltase-glucoamylase, intestinal	23.40	N
11	O95881	TXNDC12	Thioredoxin domain-containing protein 12	22.93	N
12	P04632	CAPNS1	Calpain small subunit 1	22.93	N
13	P01031	C5	Complement C5	22.05	Y
14	P02042	HBD	Hemoglobin subunit delta	22.04	N
15	K7EJH8	ACTN4	Alpha-actinin-4	21.85	N
16	P27797	CALR	Calreticulin	6.78	Y
17	P15144	ANPEP	Aminopeptidase N	5.00	Y
18	Q14019	COTL1	Coactosin-like protein	4.57	N
19	P20160	AZU1	Azurocidin	4.43	Y
20	P52209	PGD	6-phosphogluconate dehydrogenase, decarboxylating	3.30	N
21	P06731	CEACAM5	Carcinoembryonic antigen-related cell adhesion molecule 5	2.77	Y
22	P07900	HSP90AA1	Heat shock protein HSP 90-alpha	2.34	N
23	Q6P5S2	C6orf58	UPF0762 protein C6orf58	-3.31	Y
24	Q8TDL5	BPIFB1	BPI fold-containing family B member 1	-3.58	Y
25	P61626	LYZ	Lysozyme C	-4.64	Y
26	P01033	TIMP1	Metalloproteinase inhibitor 1	-4.67	Y
27	P01833	PIGR	Polymeric immunoglobulin receptor	-5.73	Y
28	P50395	GDI2	Rab GDP dissociation inhibitor beta	-19.48	N
29	P01611		Ig kappa chain V-I region Wes	-19.86	N
30	Q9NZD2	GLTP	Glycolipid transfer protein	-21.11	N
31	B3EWG6	FAM25G	Protein FAM25G	-21.13	N
32	P02545	LMNA	Prelamin-A/C	-21.18	N
33	H7C1X2	GGT5	Gamma-glutamyltransferase 5	-21.25	N
34	P22392	NME2	Nucleoside diphosphate kinase B	-21.43	N
35	P13798	APEH	Acylamino-acid-releasing enzyme	-21.48	N
36	Q9BZG9	LYNX1	Ly-6/neurotoxin-like protein 1	-21.61	Y
37	A8K7I4	CLCA1	Calcium-activated chloride channel regulator 1	-21.77	Y
38	B7Z7A9	PGK1	Phosphoglycerate kinase	-21.86	N
39	P15814	IGLL1	Immunoglobulin lambda-like polypeptide 1	-21.96	N
40	Q6UWP8	SBSN	Suprabasin	-22.05	N
41	Q6UXB3	LYPD2	Ly6/PLAUR domain-containing protein 2	-22.05	Y
42	P23396	RPS3	40S ribosomal protein S3	-22.08	N

43	P30044	PRDX5	Peroxiredoxin-5, mitochondrial	-22.19	N
44	Q96BQ1	FAM3D	Protein FAM3D	-22.21	Y
45	P00747	PLG	Plasminogen	-22.21	Y
46	P04220		Ig mu heavy chain disease protein	-22.25	Y
47	Q92743	HTRA1	Serine protease HTRA1	-22.26	N
48	P35325	SPRR2B	Small proline-rich protein 2B	-22.56	N
49	P15328	FOLR1	Folate receptor alpha	-22.69	Y
50	P01763		Ig heavy chain V-III region WEA	-22.70	Y
51	Q99102	MUC4	Mucin-4	-22.76	Y
52	P01708		Ig lambda chain V-II region BUR	-22.76	N
53	P01597		Ig kappa chain V-I region DEE	-22.78	N
54	E7EVJ5	CYFIP2	Cytoplasmic FMR1-interacting protein 2	-22.82	N
55	P08311	CTSG	Cathepsin G	-23.03	Y
56	P02647	APOA1	Apolipoprotein A-I	-23.19	Y
57	A0M8Q6	IGLC7	Ig lambda-7 chain C region	-23.32	N
58	P60903	S100A10	Protein S100-A10	-23.38	N
59	Q9NQ38	SPINK5	Serine protease inhibitor Kazal-type 5	-23.40	N
60	H0Y7V6	SFTPB	Pulmonary surfactant-associated protein B	-23.53	N
61	P19438-5		Tumor necrosis factor receptor superfamily member 1A	-23.62	Y
62	Q9UBX7	KLK11	Kallikrein-11	-23.84	Y
63	O14745	SLC9A3R1	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	-23.85	N
64	P18510	IL1RN	Interleukin-1 receptor antagonist protein	-23.87	Y
65	P16402	HIST1H1D	Histone H1.3	-23.87	N
66	P06331		Ig heavy chain V-II region ARH-77	-23.90	N
67	P01771		Ig heavy chain V-III region HIL	-23.92	N
68	Q9NZT1	CALML5	Calmodulin-like protein 5	-23.92	N
69	O95274	LYPD3	Ly6/PLAUR domain-containing protein 3	-23.93	Y
70	P02652	APOA2	Apolipoprotein A-II	-24.29	Y
71	Q99935	PROL1	Proline-rich protein 1	-24.44	Y
72	P01779		Ig heavy chain V-III region TUR	-24.49	N
73	P22528	SPRR1B	Cornifin-B	-24.59	N
74	Q8NBJ4	GOLM1	Golgi membrane protein 1	-24.75	Y
75	P13987	CD59	CD59 glycoprotein	-24.88	Y
76	P04155	TFF1	Trefoil factor 1	-24.90	N
77	P63313	TMSB10	Thymosin beta-10	-24.92	N
78	P29508	SERPINB3	Serpin B3	-25.04	N
79	P28325	CST5	Cystatin-D	-25.05	N
80	P18065	IGFBP2	Insulin-like growth factor-binding protein 2	-25.18	Y

81	P01700		Ig lambda chain V-I region HA	-25.28	N
82	P27482	CALML3	Calmodulin-like protein 3	-25.59	N
83	P01702		Ig lambda chain V-I region NIG-64	-25.84	N
84	D6RHX1	MUC7	Mucin-7	-26.02	Y
85	Q08188	TGM3	Protein-glutamine gamma-glutamyltransferase Eglutamyltransferase E 27 kDa non-catalytic chain	-26.10	N
86	Q96QR1	SCGB3A1	Secretoglobin family 3A member 1	-26.55	N
87	Q9UBG3	CRNN	Cornulin	-26.89	N
88	P08118	MSMB	Beta-microseminoprotein	-27.22	N
89	P01011	SERPINA3	Alpha-1-antichymotrypsin	-27.76	Y
90	P35321	SPRR1A	Cornifin-A	-28.18	N

4.4.2 Influx of proteins from neutrophil origin

The high number of inflammatory response and disease related proteins that were identified in sputum collectively suggested a high influx of the most abundant white blood cell, neutrophils. Increased neutrophil cells were observed previously in sputum obtained from patients suffering from CF, COPD and asthma and are reported to be the most dominant inflammatory cell types identified in sputum (Jatakanon et al. 1999; Peleman et al. 1999; Downey et al. 2009). Rorvig *et al.* (2013), performed the most recent large scale proteomic profiling of human neutrophils, isolated from a freshly drawn peripheral blood from a healthy donor and identified more than 1200 unique proteins (Rorvig et al. 2013).

The proteome lists from CF, iNCF and NCF sputum were matched and compared with the neutrophil proteins identified in Rorvig *et al.*, (2013) study. All the proteins that were identified in sputum were classified into neutrophil and non-neutrophil proteins and the total number of neutrophil proteins identified (as a percentage of the total identified proteins) in CF, iNCF and NCF are represented in **Table 3**. Approximately 50% of the identified proteins in all of CF, iNCF and NCF sputum were of neutrophil origin. The abundance of neutrophil proteins in each sputum was also calculated based on the intensity as determined by MaxQuant of the total intensity of the identified neutrophil proteins as a percentage of the total intensity (**Table 3**). According to the abundance, 65% of the proteins in bacterial infected/ colonised CF and infected (iNCF) non-CF sputum was of neutrophil origin; surprisingly, a still high, but relatively lower percentage (46.8%) of the proteins in

non-infected non-CF (NCF) sputum was of neutrophil origin (**Table 3**). The sputum proteins identified to be from neutrophils, along with their Uniprot ID and gene name are listed in **Supplementary table 2**.

Table 3: The total neutrophil proteins identified in CF, iNCF and NCF sputum and the abundance of the identified neutrophil proteins in sputum.

	No. of neutrophil proteins identified in sputum	**Neutrophil protein abundance in sputum (%)**
CF	173	66.4
iNCF	120	65.2
NCF	153	46.8

Six of the most abundant neutrophil proteins identified along with the log_2 transformed intensity in CF, iNCF and NCF sputum are represented in **Figure 2**. A general trend could be observed from the graph that the intensity of neutrophil proteins in CF sputum are all comparatively higher than in both iNCF and NCF sputum. Except for neutrophil gelatinase-associated lipocalin, myeloperoxidase, azurocidin, catalase, phospholipase B-like 1 and matrix-metalloproteinase 9 were the neutrophil glycoproteins proteins identified to be statistically significantly up-regulated in CF compared to NCF sputum (**Figure 2**).

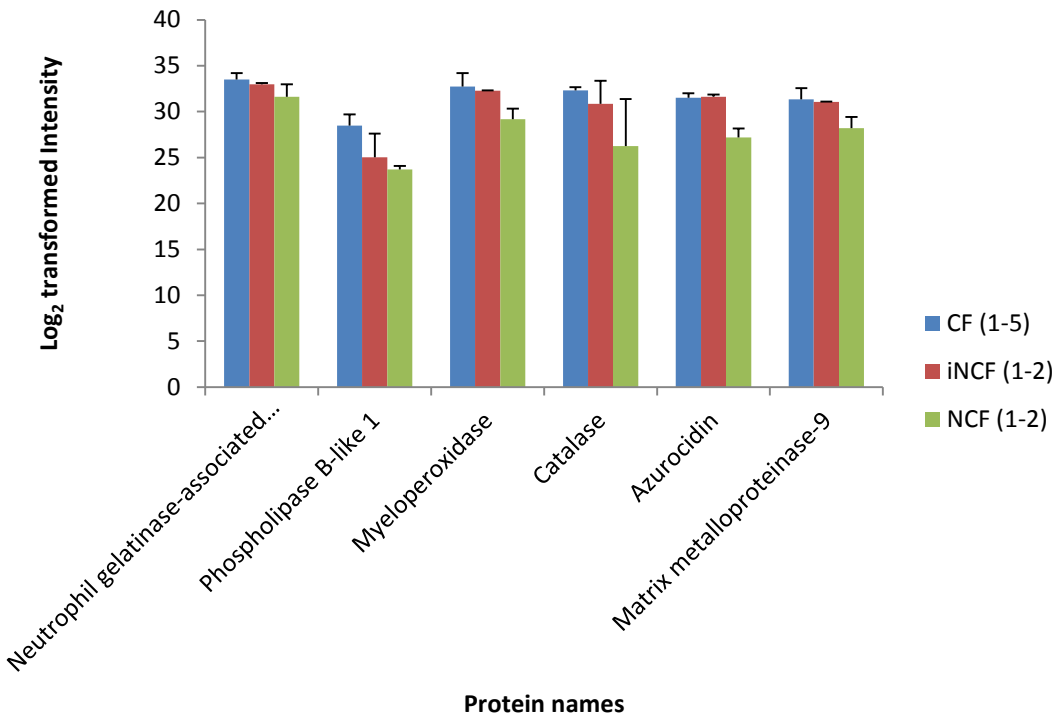


Figure 2: The most abundant neutrophil proteins identified in CF sputum were plotted against their log_2 transformed intensity. The log_2 transformed intensity of the same protein from iNCF and NCF sputum is also represented. Except for neutrophil gelatinase-associated lipocalin, every other protein marked in this graph are statistically up-regulated in CF compared to NCF.

4.4.3 Decrease in mucin concentrations in CF sputum

Mucous hypersecretion is one of the hallmarks of airway inflammation disease such as CF, asthma and chronic bronchitis. Goblet cells and SMG are the major secretors of mucin glycoproteins, which are commonly the most abundant glycoproteins identified in sputum. The mucus layer acts as the first barrier in defence against invading pathogens (Roy et al. 2014). Five mucins were identified in the proteome analysis of CF, iNCF and NCF sputum namely MUC5AC, MUC5B, MUC4, MUC7 and MUC16. Two of the most abundant mucin proteins, MUC5B and MUC5AC were up-regulated in non-infected (NCF) non-CF sputum compared to CF proteome with a fold change of 3.1 and 3.4 respectively (**Figure 3**). Overall abundances of all mucin proteins in NCF sputum is high compared to the abundance in CF and iNCF sputum.

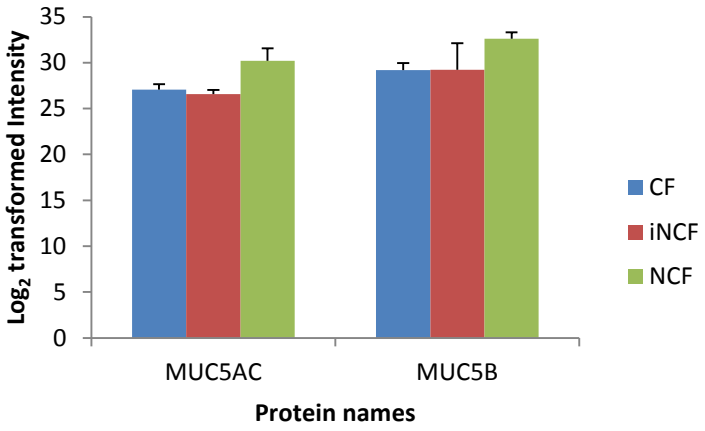


Figure 3: Log2 transformed intensity of the two most abundant sputum mucins (MUC5AC and MUC5B), representing an up-regulation of mucins in non-infected (NCF) sputum compared to bacterial infected CF sputum.

4.4.4 Contribution of proteases in sputum

Panther classification tool (version 9.0), which is freely available, was used to classify the sputum proteins into different classes. The whole list of 438 non-redundant proteins identified in the CF, iNCF and NCF sputum were imported onto the Panther tool and one of the most prominent protein classes found were the proteases. In terms of the number of identified proteins, proteases corresponded to 10%, and in terms of protein abundance the identified proteases in CF, iNCF and NCF sputum categories accounted for 12.2% of the total protein content. Serine proteases were the most commonly observed proteases, including cathepsin G, azurocidin, lactotransferrin and neutrophil elastase. Metalloproteases including matrix metalloproteinase-9, neutrophil collagenase, aminopeptidase N and aminopeptidase B, were identified to be abundant proteins in all the sputum samples. Most of the proteases were identified as neutrophil proteins with antimicrobial properties and were up-regulated in CF sputum compared to non-infected (NCF) non-CF sputum. The whole list of protease enzymes along with their gene name and Uniprot ID is given in **Table 5**.

Table 5: Lists of proteases identified in CF, iNCF and NCF sputum.

#	Protein names	Gene IDs	Uniprot IDs
1	Acylamino-acid-releasing enzyme	APEH	P13798
2	Aminopeptidase B	RNPEP	Q9H4A4
3	Aminopeptidase N	ANPEP	P15144
4	Apoptosis-associated speck-like protein containing a CARD	PYCARD	Q9ULZ3
5	Azurocidin	AZU1	P20160
6	Beta-2-glycoprotein 1	APOH	P02749
7	C4b-binding protein alpha chain	C4BPA	P04003
8	Calpain small subunit 1	CAPNS1	P04632
9	Calpain-1 catalytic subunit	CAPN1	P07384
10	Cathepsin B;Cathepsin B light chain	CTSB	P07858
11	Cathepsin D;Cathepsin D light chain	CTSD	P07339
12	Cathepsin G	CTSG	P08311
13	Cathepsin S	CTSS	P25774
14	Ceruloplasmin	CP	P00450
15	Complement factor H	CFH	P08603
18	Gamma-glutamyl hydrolase	GGH	Q92820
19	Haptoglobin	HP;HPR	P00738
20	Hemopexin	HPX	P02790
21	Kallikrein-1	KLK1	P06870
22	Kallikrein-11	KLK11	Q9UBX7
23	Lactotransferrin	LTF	P02788
24	Leukotriene A-4 hydrolase	LTA4H	P09960
25	Lysosomal Pro-X carboxypeptidase	PRCP	P42785
26	Matrix metalloproteinase-9	MMP9	P14780
27	Myeloblastin	PRTN3	P24158
28	Neutrophil collagenase	MMP8	P22894
29	Neutrophil elastase	ELANE	P08246
30	Plasminogen;Plasmin heavy chain A	PLG	P00747
31	Proteasome subunit alpha type-1	PSMA1	P25786
32	Proteasome subunit alpha type-4	PSMA4	P25789
33	Proteasome subunit alpha type-5	PSMA5	P28066

34	Proteasome subunit alpha type-7	PSMA7	O14818
35	Proteasome subunit beta type-1	PSMB1	P20618
36	Proteasome subunit beta type-2	PSMB2	P49721
37	Proteasome subunit beta type-3	PSMB3	P49720
38	Proteasome subunit beta type-8	PSMB8	P28062
39	Proteasome subunit beta type-9	PSMB9	P28065
40	Protein DJ-1	PARK7	Q99497
41	Serine protease HTRA1	HTRA1	Q92743
42	Xaa-Pro dipeptidase	PEPD	P12955

4.4.5 Cluster analysis

The relatedness between the CF, iNCF and NCF sputum proteomes was further investigated by a cluster dendrogram using the normalised relative intensity of each protein obtained from MaxQuant search. Similar to that which was observed in the sputum *N*-glycomes, the proteome of pathogen colonised (iNCF 1-2) non-CF sputum clustered with the proteome of CF (CF1-5) sputum whereas the non-infected (NCF 1-2) non-CF sputum proteome clustered separately (**Figure 4**). This supports the similarity found between the number of regulated proteins between CF and iNCF sputum compared with those between iNCF and NCF sputum. The clustering of CF with infected non-CF proteomes also correlates with a greater number of immune response related proteins found in the proteomes of CF and iNCF sputum. It is also noted that NCF sputum clustered separately from the bacterial colonised iNCF sputum. It thus appears that a combination of infection and inflammation in CF and iNCF may trigger more immune response proteins in the sputum compared to the NCF sputum in which only inflammation is the predominant clinical factor.

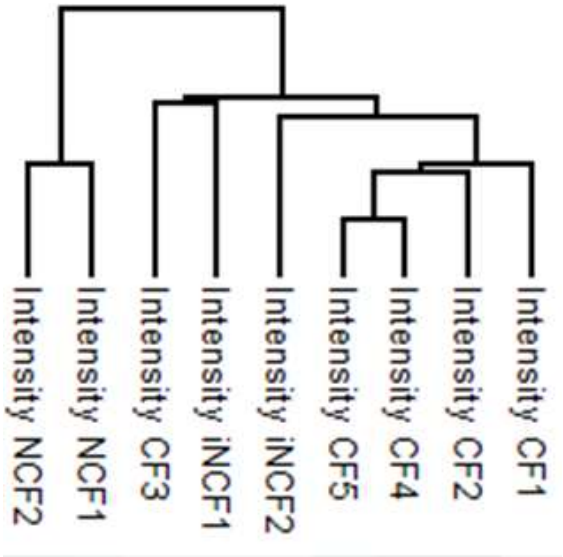


Figure 4: Cluster analysis of CF, iNCF and NCF sputum proteome, clustering iNCF and CF sputum together and NCF sputum clusters separately.

4.4.6 *N*-glycan analysis of neutrophil cells

The higher abundance of neutrophil proteins in CF and infected (iNCF) non-CF sputum (two-thirds of the sputum proteome) in comparison to non-infected (NCF) non-CF sputum suggested that the differences in the sputum *N*-glycomes (**Chapter 2**) could be specifically attributable to the *N*-glycosylation of the neutrophil proteins. To investigate this, *N*-glycans were enzymatically released from the solubilised whole cell lysate proteins obtained from neutrophil cells isolated from the blood of a healthy donor. The relative abundances of the identified *N*-glycans from neutrophil proteins were compared with the *N*-glycans obtained from CF sputum **(Figure 5**). As shown in **Figure 5**, the *N*-glycan profile of CF sputum and neutrophils were remarkably similar to each other, which was verified statistically by calculating coefficient of determination giving an R^2 value of 0.9. The total *N*-glycans released from neutrophils had pauci-mannose *N*-glycans in relative abundance of around 43%**,** which is similar to the abundance of pauci-mannose identified in CF sputum (**Chapter 2**) with a R^2 value of 0.96. Specifically, the most abundant *N*-glycan in CF sputum ($GlcNAc_2Man_2Fuc_1$, M2F, *m*/*z* 895.4) was also the most abundant *N*-glycan identified on the neutrophil proteins. This strongly suggests that the most abundant and unique pauci-mannose *N*-glycans identified in sputum originate from the neutrophil proteins in the CF lung.

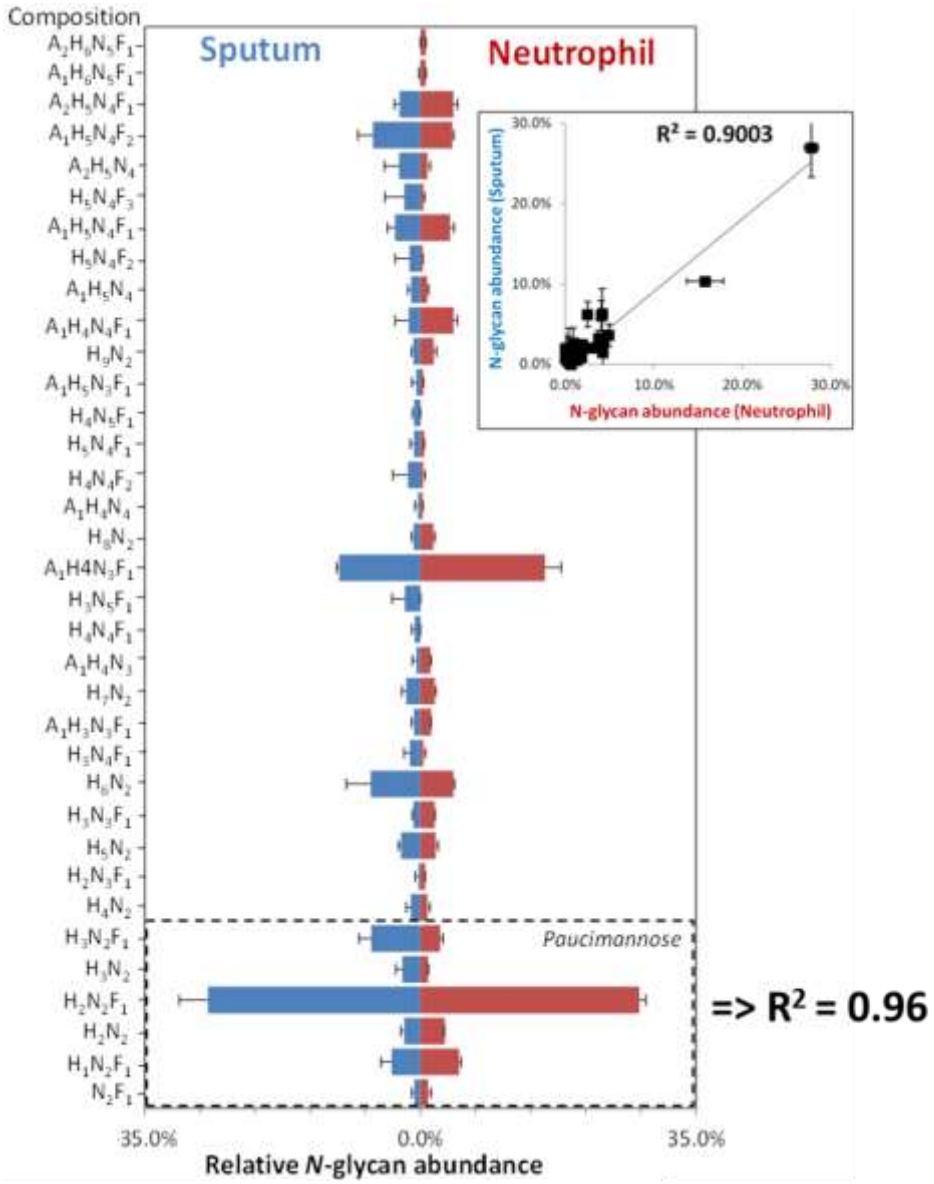


Figure 5: Neutrophil and CF sputum *N*-glycome comparison showing a similar *N*-glycan profile between sputum and neutrophils, particularly in the abundances of pauci-mannosidic *N*-glycans. Glycan composition key: H- Hexose, N- HexNAc, F- Fucose and A- Sialic acid.

4.4.7 *N*-glycopeptide analysis of CF sputum

Proteomic analysis of sputum revealed an increased abundance of neutrophil associated proteins in bacterial colonised/ infected CF and non-CF (iNCF) sputum compared to non-infected (NCF) non-CF sputum. Also, *N*-glycans released from neutrophil whole cell proteins mimicked the sputum protein *N*-glycosylation profile with higher and similar abundances of pauci-mannose *N*-glycans. Since the particular focus of this project was on the characterisation and role of glycosylation in cystic fibrosis, further investigation was

conducted specifically on the CF sputum to determine the identity of the proteins carrying the pauci-mannose glycans and the specific sites to which they are attached.

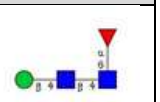
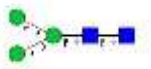
A targeted analysis was performed to identify the site *N*-glycosylation of the up-regulated abundant neutrophil glycoproteins in CF sputum. Myeloperoxidase (Uniprot ID: P05164), a neutrophil protein that was found to be up-regulated in CF compared to NCF sputum with a p value of 0.0008, was identified with 20 non-redundant *N*-glycopeptides comprising three *N*-glycosylation sites identified in this study out of six sites. Pauci-mannosidic *N*-glycans were identified on each of the sites, along with high mannose and complex *N*-glycans. A total of seven *N*-glycans were identified on Asn323, five *N*-glycans on Asn355 and eight *N*-glycans on Asn483. The individual glycan structures, with glycan and glycopeptide mass are presented in **Table 6**. Putative structures are drawn based on the *N*-glycan mass and the detailed structures of the *N*-glycans profiled in the global *N*-glycan analysis of CF sputum (**Chapter 2**).

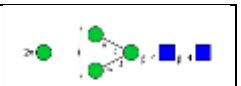
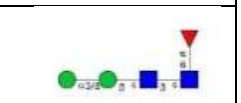
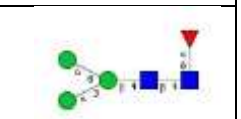
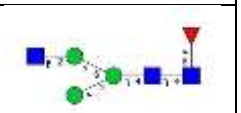
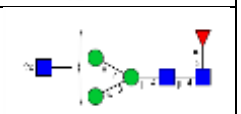
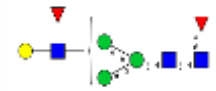
Apart from myeloperoxidase, other CF up-regulated sputum neutrophil glycoproteins were also identified to be carrying pauci-mannose *N*-glycans; azurocidin (Uniprot ID: P20160) and phospholipase B-like 1 (Uniprot ID: Q6P4A8). Only one site out of three possible sites (Asn171) and one glycoform, with *N*-glycan M2F ($GlcNAc_2Man_2Fuc_1$), was identified in azurocidin. Phospholipase B-like 1 glycoprotein was also identified with one site out of five sites (Asn71) and five glycoforms including pauci-mannose and high mannose structures (**Table 6**).

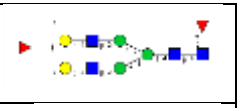
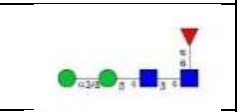
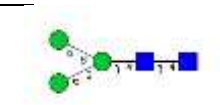
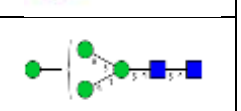
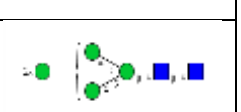
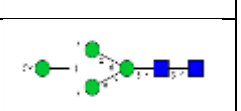
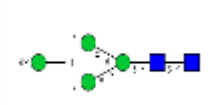
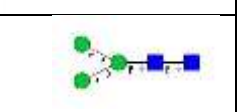
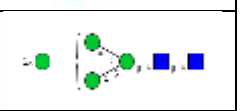
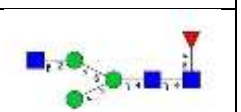
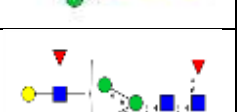
Other *N*-glycopeptides were identified from other neutrophil glycoproteins including neutrophil gelatinase- associated lipocalin (Uniprot ID: P80188) and complement C3 (Uniprot ID: P01024), which were not differentially regulated in CF compared to NCF sputum. Lipocalin glycopeptides were mainly found to carry complex *N*-glycans, with one pauci-mannose and high mannose *N*-glycan on a single site Asn85 **(Table 6)**. In addition, a glycopeptide from a non-neutrophil glycoprotein, Ig gamma-2 chain C (Uniprot ID: P01859) was identified, with two GlcNAc terminating *N*-glycans on a single glycosylation site Asn176, (**Table 6**).

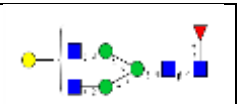
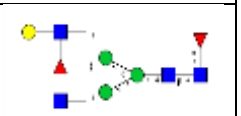
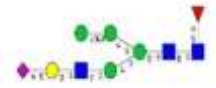
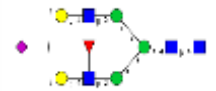
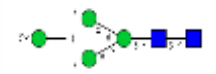
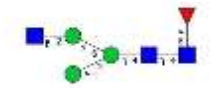
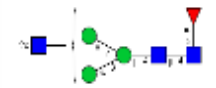
The results of the site-specific analysis on this set of abundant proteins demonstrated that pauci-mannose glycans are carried on sites of human neutrophil proteins identified in sputum and provided further evidence that the higher abundance of pauci-mannose in bacterial colonised CF sputum is not a product of bacterial degradation but is rather a result of the high neutrophil cell content of CF sputum. Neutrophil proteins have been shown to play an important role in host defence against pathogen infection, as discussed further in **Section 4.5** (Hirche et al. 2005).

Table 6: Site-specific *N*-glycoproteome analysis by HILIC glycopeptide enrichment and RP-LC-MS/MS of six neutrophil glycoproteins and Ig gamma-2 chain C glycoprotein, with peptide sequence and corresponding *N*-glycan structure. The proteins marked with an asterisk (*) are the neutrophil glycoproteins that are up-regulated in CF sputum compared to non-infected (NCF) non-CF sputum. Glycosylation site is indicated in bold italics in the peptide sequence.

#	Protein Name	Uniprot ID	No. of *N*-glycosylation site identified/ total number of sites	Site	Peptide Sequence	Peptide mass	Glycan mass	Type	Observed Glycopeptide mass (m/z)	Charge state	*N*-glycan structure
1	Myeloperoxidase (*)	P05164	3 /6	Asn323	SCPACPGS***N***ITIR	1434.6	714.3	Pauci-mannose	1073.97	2	
							730.3	Pauci-mannose	1081.97	2	
							876.3	Pauci-mannose	1154.99	2	
							1054.4	High mannose	1244.02	2	
							1216.4	High mannose	1325.05	2	
							1378.5	High mannose	1406.07	2	
							1540.5	High mannose	1487.60	2	
				Asn355	***N***MSNQLGLLAV NQR	1577.8	892.3	Pauci-mannose	822.71	3	
							1054.4	High mannose	1314.60	2	

							1216.4	High mannose	1395.62	2	
							1378.5	High mannose	1476.65	2	
							1540.5	High mannose	1557.68	2	
				Asn483	SY***N***DSVDPR	1054.5	876.3	Pauci-mannose	964.90	2	
							1038.4	Pauci-mannose	1045.92	2	
							1216.4	High mannose	1134.95	2	
							1378.5	High mannose	1215.97	2	
							1241.5	Complex	1147.46	2	
							1444.5	Complex	1249	2	
							1549.5	Complex	1301.52	2	
							1606.6	Complex	1330.03	2	

2	Matrix metalloproteinase 9 (*)	P14780	1/3	Asn38	T***N***LTDRQLAEEY LYR	721.4	1914.7	Complex	1317.54	2	
3	Azurocidin (*)	P20160	1/3	Asn171	FV***N***VTVPEDQCR	1566.8	876.3	Pauci-mannose	1221.04	2	
4	Phospholipase B-like 1 (*)	Q6P4A8	1/5	Asn71	NVMDKNNGDA AYGFY***N***NSVK	1450.7	892.3	Pauci-mannose	1170.98	2	
							1054.4	High mannose	1252.01	2	
							1216.4	High mannose	1090.45	3	
							1378.5	High mannose	1414.06	2	
							1540.5	High mannose	1495.09	2	
5	Neutrophil gelatinase-associated lipocalin	P80188	1/1	Asn85	(EDK)SY***N***VTSVL FR(K)	1561.8	892.3	Pauci-mannose	817.37	3	
							1216.4	High mannose	1201.53	2	
							1241.5	Complex	1214.04	2	
							1549.6	Complex	1367.59	2	

							1606.6	Complex	1396.61	2	
							1752.6	Complex	1104.15	3	
							1768.6	Complex	1109.48	3	
							1856.7	Complex	1138.82	3	
							1913.7	Complex	1076.47	3	
							2205.8	Complex	973.68	4	
6	Complement C3	P01024	1/3	Asn85	TVLTPATNHMG ***N***VTFIPANR	2275.2	1378.5	High mannose	1217.27	3	
							1540.5	High mannose	1271.23	3	
7	Ig gamma-2 chain C	P01859	1/1	Asn176	EEQF***N***STFR	1159.5	1241.5	Complex	1199.99	2	
							1444.5	Complex	1301.53	2	

4.5 Discussion

4.5.1 Sputum complexity

One of the challenges of identifying manifestations of a disease through analysis of sputum is the nature of sputum itself. Not only is sputum a highly viscous and complex mixture of compounds that presents practical difficulties for analysis, it is also not easy to have a "healthy" sample of sputum for comparison since sputum itself is mainly produced in response to respiratory inflammation and infection. Therefore in this project, the sputum proteome and *N*-glycome obtained from CF patients was compared to that of non-CF counterparts in which patients had other lung related conditions such as asthma, COPD or URTI other than CF. The CF, bacterial colonised/ infected non-CF (iNCF) and non-infected non-CF (NCF) sputum all featured "inflammation" as a common feature of the conditions associated with sputum production. In addition, CF and iNCF/ URTI sputum were also affected by bacterial colonisation, whereas NCF sputum was not. This made it possible to draw some associations from the results of the sputum analysis and bacterial infection status.

4.5.2 Chronic infection and inflammation

Chronic inflammation being one of the major hallmarks of CF disease has an increased neutrophil count in the lung tissues as compared to the healthy lungs. High neutrophil count in CF lung or in patients with other lung conditions like asthma, pneumonia or COPD is a sign of chronic lung inflammation and infection (Jatakanon et al. 1999). As demonstrated in this study, infection together with inflammation can result in production of immune response related proteins, which are mainly carried by the most abundant white blood cells, neutrophils. The abundance of inflammatory cells or neutrophils identified in sputum could vary between each patient and also between different lung conditions. Increase in the level of inflammation and infection could lead to an increase in amount of neutrophil cells/ proteins observed in sputum. Thus, quantitative proteomic analysis of sputum as carried out in this work serves a valuable role in identifying the abundance of neutrophil proteins.

Polymorphonuclear leukocytes (PMNs) or neutrophils provide the first line of defence of the innate immune system against pathogenic infection (Watt et al. 2005; Borregaard 2010).

Neutrophils are granulated cells, which are produced and recruited from the blood circulation to the inflamed or injured tissues. Granules are classified mainly into three types namely primary or azurophilic, secondary or specific and tertiary or gelatinase granules, which store proteins that help the host in defence against pathogens (Borregaard 2010). Neutrophils are recruited from blood circulation to the infection site as a response to different signalling molecules including microbial molecules and cytokines or chemokines. Neutrophils utilise various mechanisms to respond to bacterial invasion such as degranulation (Stenson et al. 1980; Jaovisidha et al. 1999), phagocytosis (Urban et al. 2006) and neutrophil extracellular traps (NETs) (Kaplan et al. 2012; Obermayer et al. 2014).

Pathogenic infection plays a role in activation and degranulation of the neutrophil cells, which in turn act against the pathogen invasion. The resting neutrophil cells are activated by inflammatory stimuli; the neutrophils degranulate, release antimicrobial factors including various antimicrobial proteins and proteases to inactivate the bacteria, and can then engulf the bacteria through phagocytosis (Brinkmann et al. 2004). In a study by Wark *et al.* (2002), viral infection in acute asthma was associated with increased neutrophil degranulation in sputum and increased cell lysis, and linked the degranulation event to clinical severity (Wark et al. 2002). The antimicrobial proteins that are secreted by the neutrophil granules during degranulation (or granule bursts) include myeloperoxidase, cathepsin G (Bangalore et al. 1990), lactotransferrin (Sanchez et al. 1992; Yamauchi et al. 1993) and neutrophil elastase (Belaaouaj et al. 1998; Mayer-Scholl et al. 2004).

Activated neutrophils secrete myeloperoxidase, one of the most abundant proteins synthesised and stored in azurophilic granules, that is responsible for oxygen dependant bactericidal activity. Myeloperoxidase produces an array of potent toxic oxidants involved in killing pathogens by using hydrogen peroxide and chloride to generate hypochlorus acid, a potent bactericidal oxidant (Nauseef et al. 1983; Dri et al. 1985; Gaut et al. 2001). Apart from the oxidative system, the non-oxidative system of killing the invading pathogen by neutrophils involves the secretion of various serine proteases such as cathepsin G, neutrophil elastase and azurocidin. Serine proteases are involved in disrupting the structural integrity of the bacterial cell and thus inactivate the invading pathogen (Ganz 1999). The antimicrobial

property of these neutrophil proteins is well characterised *in vitro* but the role *in vivo* and their biological relevance in controlling infections is less clear (Bardoel et al. 2014).

The *N*-glycosylation profile released from the whole cell proteins of resting neutrophil cells contained an abundance of pauci-mannosidic *N*-glycans and the overall *N*-glycan profile, including that of pauci-mannose structures, closely matched the CF sputum *N*-glycome. As mentioned in **Chapter 2**, pauci-mannose is not commonly identified in mammalian systems and there are reports suggesting vertebrates cannot synthesise pauci-mannose structures (Schachter 2009). Over the last few years, a few studies have shown pauci-mannose expression in mammalian systems including cancer cells and tissues, human buccal cells and saliva (Balog et al. 2012; Everest-Dass et al. 2012; Sethi et al. 2014). However, significantly, the work carried out for this thesis revealed that pauci-mannose glycans are present in CF sputum and generally on neutrophil glycoproteins, demonstrating that humans can express pauci-mannosidic *N*-glycans in high abundance in specific cells.

Furthermore, targeted glycoproteome analysis carried out in this work specifically identified the neutrophil proteins myeloperoxidase and azurocidin as carriers of pauci-mannose *N*-glycans. In another study, intact glycopeptide analysis of myeloperoxidase purified from healthy donors, showed the presence of pauci-mannose *N*-glycans on four of the five identified sites using MALDI-MS/MS (Ravnsborg et al. 2010). Pauci-mannose and high mannose dominated the number of glycoforms identified on the myeloperoxidase, providing further evidence that this abundant neutrophil protein carries pauci-mannosidic *N*-glycan structures. In that study, the truncated pauci-mannose *N*-glycans were reported as high mannose or as complex glycans (if containing a core fucose residue), and the significance of pauci-mannose identification on myeloperoxidase was not emphasised.

In another study, abundant pauci-mannosidic *N*-glycans on *C. elegans* proteins were shown to be involved in response to bacterial pathogenic invasion, suggesting that the pauci-mannose structures are components of the worm's innate immune system (Shi et al. 2006). The high abundance of pauci-mannose *N*-glycans on proteins from human innate immune cells (neutrophils) that defend against pathogenic invasion suggests that the pauci-mannose may play a similar immune-response role in humans, but this has yet to be determined.

4.5.3 Mucin secretion

Overall mucin abundances in CF sputum is less than the abundance identified in non-infected (NCF) non-CF sputum, especially the two major mucins MUC5AC and MUC5B, which are down-regulated in CF sputum. As this is the most counter-intuitive result (since mucous secretion increases in CF) this discrepancy should be clearly acknowledged. The contribution of proteases that degrade the non-glycosylated portion of the mucins as shown by Schulz *et al*. (2007) would also serve to reduce the identification of mucins by the tryptic digestion commonly used for protein identification since these are the only regions accessible to trypsin and thus no/few peptides would result for identification purposes (Schulz et al. 2007). Also, an earlier study has shown a lower abundance of MUC5AC and MUC5B in CF sputum compared to non-CF bronchitis patients. One of the proposed causes for the lower abundance of mucins in CF is the presence of neutrophil related proteases or serine proteases such as neutrophil elastase that may degrade the airway mucins in cystic fibrosis (Schulz et al. 2007). CF sputum with no pathogen colonisation has been shown to have relatively more mucin compared to CF sputum with intermittent infection, and sputum from CF patients with chronic infection has the least mucin (MUC5AC and MUC5B) (Henke et al. 2011). The increase in neutrophil proteases observed in the CF sputum correlates well with the decrease in overall mucin protein identification. Also, recent papers have shown that sample handling and post expectoration degradation can contribute to the mucin degradation and that mucin content is underestimated by MS dependent on trypsin digestion (Henderson et al. 2014; Horsley et al. 2014). Thus, it can be assumed that the amount of mucin determined in this study may be under-represented.

4.5.4 Proposed new biosynthetic route for pauci-mannose *N*-glycan expression

The *N*-glycan biosynthetic machinery has a standard known pathway, in which α-glucosidase I/II and α-mannosidase act on the *N*-glycoprotein precursor forming a Man5 structure. The enzyme GlcNAc transferase (GnT-I) acts on the Man5 structure forming a hybrid glycan, and with the action of α-mannosidase, branching mannose residues are trimmed which leads to the structure $GlcNAc_3Man_3$. The glycan can then be either core fucosylated or not depending of the action of the enzyme FUT8. In the established *N*-glycosylation pathway synthesis, action of galactosyl transferase (GalT) takes place forming

hybrid *N*-glycans and subsequent action of GnT-I, GalT and sialyl transferase (SiaT) enzymes takes place to form complex sialylated *N*-glycans.

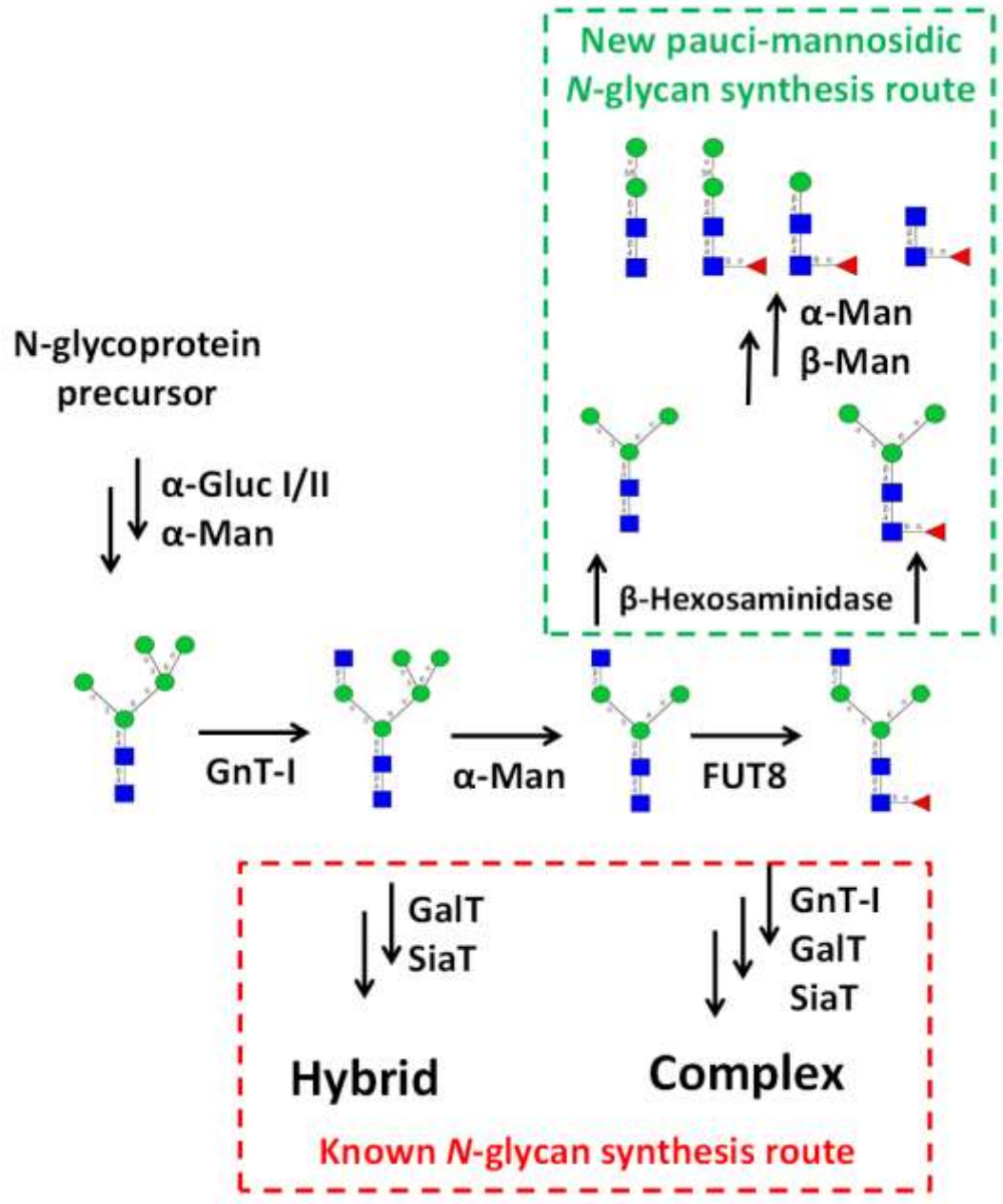


Figure 6: Proposed alternate biosynthetic pathway for the synthesis of pauci-mannosidic *N*-glycans. Increased β-hexosaminidase activity in neutrophils would increase the expression of pauci-mannosidic *N*-glycans, as observed in infected CF and non CF sputum, and corresponds to a decrease in hybrid and complex *N*-glycans.

In neutrophil cells, we propose an alternate biosynthetic route for the synthesis of pauci-mannosidic *N*-glycans, involving a possible enzymatic action of β-hexosaminidase A/B (Hex A or Hex B) on either $GlcNAc_3Man_3$ (M3) or $GlcNAc_3Man_3Fuc_1$ (M3F), cleaving the terminal GlcNAc and forming pauci-mannosidic *N*-glycans. Further action of α- and β-mannosidase

on the already truncated pauci-mannose might synthesise the smaller pauci-mannosidic *N*-glycans. The whole proposed pathway is represented in **Figure 6**. The proposed pathway is underpinned by other studies that have reported the activity of a β-hexosaminidase in neutrophil cells isolated from healthy blood samples in relation to its action on GM2 gangliosides (Pennybacker et al. 1996; Casal et al. 2003; Casal et al. 2005; Wendeler et al. 2009). The work demonstrates that there is a β- hexosaminidase enzyme in neutrophils capable of trimming a terminal GlcNAc residue, and thus indicates that the synthesis of pauci-mannose *N*-glycans is possible via a different biosynthetic route to the known *N*-glycan synthesis. In infected CF and iNCF sputum, consisting of an abundance of neutrophil proteins, there was an up-regulation of pauci-mannose and down-regulation of hybrid and complex *N*-glycans compared to non-infected (NCF) sputum, thus suggesting an overexpression of β- hexosaminidase and supporting the existence of the alternate pathway, suggesting the regulation of pauci-mannose in infected CF and iNCF sputum is due to the infected status and not due to the CF disease mutation itself.

In conclusion, proteomic analysis of CF and iNCF sputum revealed that about two-thirds of the sputum proteins were neutrophil proteins, compared to non-infected (NCF) non-CF sputum, which consisted of less than 50% of neutrophil proteins. Infected non-CF (iNCF) patient sputum also showed an elevated neutrophil protein abundance. Cluster analysis of proteins obtained from CF, iNCF and NCF sputum linked the infected iNCF sputum with the CF sputum whereas NCF (non-infected) sputum clustered separately to be different from both of the infected sputum status. Neutrophil proteins are rich in pauci-mannosidic *N*-glycans, which correlated with the high abundance of pauci-mannose *N*-glycans on CF sputum proteins. Site-specific *N*-glycoproteome analysis showed that three of the most abundant neutrophil proteins, myeloperoxidase, azurocidin and phospholipase B-like 1 proteins carry pauci-mannoses *N*-glycans. Neutrophil proteins, being an immune response against pathogenic invasion, thus carry the pauci-mannose *N*-glycans, suggesting that they are the basis of the pauci-mannose *N*-glycan increase in bacterial infected/colonised CF and iNCF sputum that arises from the inflamed central lung airways. The proposed new biosynthetic pathway of pauci-mannosidic *N*-glycans (**Figure 6**) involves a previously reported neutrophil β-hexosaminidase that cleaves the terminal GlcNAc residue from GM2 gangliosides. The action of this enzyme on the $GlcNAc_2Man_3GlcNAc_1$–protein intermediate

of the major *N*-glycan biosynthetic pathway would result in the observed down regulation of the hybrid and complex *N*-glycans and up-regulation of the pauci-mannosidic *N*-glycans in CF and infected non CF lungs.

4.6 Supplementary file

Supplementary table 1: The lists of 438 non-redundant proteins identified in CF, iNCF and NCF sputum along with the protein's gene ID and Uniprot ID.

#	Protein names	Gene ID	Uniprot ID
1	10 kDa heat shock protein, mitochondrial	HSPE1	P61604
2	14-3-3 protein beta/alpha	YWHAB	P31946
3	14-3-3 protein eta	YWHAH	Q04917
4	14-3-3 protein sigma	SFN	P31947
5	14-3-3 protein zeta/delta	YWHAZ	P63104
6	40S ribosomal protein S3	RPS3	P23396
7	60S acidic ribosomal protein P2	RPLP2	P05387
8	6-phosphogluconate dehydrogenase, decarboxylating	PGD	P52209
9	6-phosphogluconolactonase	PGLS	O95336
10	78 kDa glucose-regulated protein	HSPA5	P11021
11	Acid ceramidase	ASAH1	Q13510
12	Actin, alpha skeletal muscle	ACTA1	P68133
13	Actin, cytoplasmic 1	ACTB	P60709
14	Actin-related protein 10	ACTR10	Q9NZ32
15	Actin-related protein 2/3 complex subunit 4	ARPC4	P59998
16	Activated RNA polymerase II transcriptional coactivator p15	SUB1	P53999
17	Acylamino-acid-releasing enzyme	APEH	P13798
18	Acyl-CoA-binding protein	DBI	P07108
19	Acyl-coenzyme A thioesterase 4	ACOT4	Q8N9L9
20	Adenylyl cyclase-associated protein 1	CAP1	Q01518
21	ADP-ribosyl cyclase 2	BST1	Q10588
22	A-kinase anchor protein 9	AKAP9	Q99996
23	Alpha-1-acid glycoprotein 1	ORM1	P02763
24	Alpha-1-acid glycoprotein 2	ORM2	P19652
25	Alpha-1-antichymotrypsin	SERPINA3	P01011
26	Alpha-1-antitrypsin	SERPINA1	P01009
27	Alpha-1B-glycoprotein	A1BG	P04217
28	Alpha-2-HS-glycoprotein	AHSG	P02765
29	Alpha-2-macroglobulin	A2M	P01023
30	Alpha-2-macroglobulin-like protein 1	A2ML1	A8K2U0
31	Alpha-actinin-1	ACTN1	P12814
32	Alpha-actinin-1	ACTN1	G3V2X9
33	Alpha-actinin-4	ACTN4	O43707
34	Alpha-actinin-4	ACTN4	K7EJH8
35	Alpha-amylase 1	AMY1A	P04745
36	Alpha-enolase	ENO1	P06733
37	Alpha-enolase	ENO1	P06733-2
38	Aminopeptidase B	RNPEP	Q9H4A4

39	Aminopeptidase N	ANPEP	P15144
40	Angiotensinogen	AGT	P01019
41	Annexin A1	ANXA1	P04083
42	Annexin A2	ANXA2	P07355
43	Annexin A3	ANXA3	P12429
44	Annexin A6	ANXA6	P08133
45	Anterior gradient protein 2 homolog	AGR2	O95994
46	Antileukoproteinase	SLPI	P03973
47	Antithrombin-III	SERPINC1	P01008
48	Apolipoprotein A-I	APOA1	P02647
49	Apolipoprotein A-II	APOA2	P02652
50	Apolipoprotein A-IV	APOA4	P06727
51	Apolipoprotein B-100	APOB	P04114
52	Apolipoprotein D	APOD	P05090
53	Apolipoprotein E	APOE	P02649
54	Apoptosis-associated speck-like protein containing a CARD	PYCARD	Q9ULZ3
55	Arginase-1	ARG1	P05089
56	Aspartate aminotransferase, cytoplasmic	GOT1	P17174
57	ATP synthase subunit alpha, mitochondrial	ATP5A1	P25705
58	Azurocidin	AZU1	P20160
59	Bactericidal permeability-increasing protein	BPI	P17213
60	Basic salivary proline-rich protein 3	PRB3	Q04118
61	Beta-2-glycoprotein 1	APOH	P02749
62	Beta-2-microglobulin	B2M	P61769
63	Beta-microseminoprotein	MSMB	P08118
64	BPI fold-containing family A member 1	BPIFA1	Q9NP55
65	BPI fold-containing family A member 2	BPIFA2	Q96DR5
66	BPI fold-containing family B member 1	BPIFB1	Q8TDL5
67	BPI fold-containing family B member 2	BPIFB2	Q8N4F0
68	Brain acid soluble protein 1	BASP1	P80723
69	C4b-binding protein alpha chain	C4BPA	P04003
70	Calcineurin subunit B type 1	PPP3R1	P63098
71	Calcium-activated chloride channel regulator 1	CLCA1	A8K7I4
72	Calcyphosin	CAPS	Q13938
73	Calmodulin	CALM1	P62158
74	Calmodulin-like protein 3	CALML3	P27482
75	Calmodulin-like protein 5	CALML5	Q9NZT1
76	Calnexin	CANX	P27824
77	Calpain small subunit 1	CAPNS1	P04632
78	Calpain-1 catalytic subunit	CAPN1	P07384
79	Calreticulin	CALR	P27797
80	Calreticulin	CALR	K7EJB9
81	Carbonic anhydrase 1	CA1	P00915
82	Carbonic anhydrase 6	CA6	P23280
83	Carcinoembryonic antigen-related cell adhesion molecule 5	CEACAM5	P06731

84	Carcinoembryonic antigen-related cell adhesion molecule 8	CEACAM8	P31997
85	Catalase	CAT	P04040
86	Cathelicidin antimicrobial peptide	CAMP	P49913
87	Cathepsin B	CTSB	P07858
88	Cathepsin D	CTSD	P07339
89	Cathepsin G	CTSG	P08311
90	Cathepsin S	CTSS	P25774
91	CD59 glycoprotein	CD59	P13987
92	CD9 antigen	CD9	P21926
93	Ceruloplasmin	CP	P00450
94	Chitinase-3-like protein 1	CHI3L1	P36222
95	Chitotriosidase-1	CHIT1	Q13231
96	Chromodomain-helicase-DNA-binding protein 6	CHD6	Q8TD26
97	Clusterin	CLU	P10909
98	Coactosin-like protein	COTL1	Q14019
99	Cofilin-1	CFL1	P23528
100	COMM domain-containing protein 2	COMMD2	Q86X83
101	Complement C3	C3	P01024
102	Complement C4-A	C4A	P0C0L4
103	Complement C5	C5	P01031
104	Complement factor B	CFB	B4E1Z4
105	Complement factor H	CFH	P08603
106	Copine-3	CPNE3	O75131
107	Cornifin-A	SPRR1A	P35321
108	Cornifin-B	SPRR1B	P22528
109	Cornulin	CRNN	Q9UBG3
110	Coronin-1A	CORO1A	P31146
111	Corticosteroid-binding globulin	SERPINA6	P08185
112	Cystatin-A	CSTA	P01040
113	Cystatin-B	CSTB	P04080
114	Cystatin-C	CST3	P01034
115	Cystatin-D	CST5	P28325
116	Cystatin-S	CST4	P01036
117	Cystatin-SA	CST2	P09228
118	Cystatin-SN	CST1	P01037
119	Cysteine-rich secretory protein 3	CRISP3	P54108
120	Cytidine deaminase	CDA	P32320
121	Cytochrome c	CYCS	P99999
122	Cytoplasmic FMR1-interacting protein 2	CYFIP2	E7EVJ5
123	D-dopachrome decarboxylase	DDT	P30046
124	Deleted in malig0.5t brain tumors 1 protein	DMBT1	Q9UGM3
125	Delta-aminolevulinic acid dehydratase	ALAD	P13716
126	Desmoglein-3	DSG3	P32926
127	Dynein heavy chain 12, axonemal	DNAH12	Q6ZR08
128	Dynein heavy chain 17, axonemal	DNAH17	Q9UFH2-4

129	EF-hand domain-containing protein D2	EFHD2	Q96C19
130	Elongation factor 1-alpha 1	EEF1A1	P68104
131	Endoplasmin	HSP90B1	P14625
132	Eosinophil cationic protein	RNASE3	P12724
133	Eosinophil peroxidase	EPX	P11678
134	Epididymal secretory protein E1	NPC2	P61916
135	Ester hydrolase C11orf54	C11orf54	Q9H0W9
136	Ezrin	EZR	P15311
137	F-actin-capping protein subunit alpha-1	CAPZA1	P52907
138	Fatty acid-binding protein, epidermal	FABP5	Q01469
139	Ferritin heavy chain	FTH1	P02794
140	Ferritin light chain	FTL	P02792
141	Fibrinogen alpha chain	FGA	P02671
142	Fibrinogen beta chain	FGB	P02675
143	Fibrinogen gamma chain	FGG	P02679
144	Fibronectin	FN1	P02751
145	Fibrous sheath-interacting protein 2	FSIP2	Q5CZC0
146	Filamin-A	FLNA	P21333
147	Folate receptor alpha	FOLR1	P15328
148	Fructose-bisphosphate aldolase A	ALDOA	P04075
149	Fructose-bisphosphate aldolase C	ALDOC	P09972
150	Fumarylacetoacetase	FAH	P16930
151	Galectin-3	LGALS3	P17931
152	Galectin-3-binding protein	LGALS3BP	Q08380
153	Gamma-glutamyl hydrolase	GGH	Q92820
154	Gamma-glutamyltransferase 5	GGT5	H7C1X2
155	Gelsolin	GSN	P06396
156	Glucose-6-phosphate isomerase	GPI	P06744
157	Glutamate dehydrogenase 1, mitochondrial	GLUD1	P00367
158	Glutaredoxin-1	GLRX	P35754
159	Glutathione peroxidase 3	GPX3	P22352
160	Glutathione reductase, mitochondrial	GSR	P00390
161	Glutathione S-transferase P	GSTP1	P09211
162	Glutathione synthetase	GSS	P48637
163	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	P04406
164	Glycogen phosphorylase, liver form	PYGL	P06737
165	Glycolipid transfer protein	GLTP	Q9NZD2
166	Glyoxalase domain-containing protein 4	GLOD4	B7Z403
167	Golgi membrane protein 1	GOLM1	Q8NBJ4
168	Granulins	GRN	P28799
169	Haptoglobin	HP	P00738
170	Haptoglobin	HP	J3KSV1
171	Heat shock 70 kDa protein 1A/1B	HSPA1A	P08107
172	Heat shock 70 kDa protein 6	HSPA6	P17066
173	Heat shock cognate 71 kDa protein	HSPA8	P11142

174	Heat shock protein beta-1	HSPB1	P04792
175	Heat shock protein HSP 90-alpha	HSP90AA1	P07900
176	Heat shock protein HSP 90-beta	HSP90AB1	P08238
177	Hemoglobin subunit alpha	HBA1	P69905
178	Hemoglobin subunit alpha	HBA2	G3V1N2
179	Hemoglobin subunit beta	HBB	P68871
180	Hemoglobin subunit delta	HBD	P02042
181	Hemopexin	HPX	P02790
182	Hepatoma-derived growth factor	HDGF	P51858
183	Heterogeneous nuclear ribonucleoprotein K	HNRNPK	P61978
184	Histidine-rich glycoprotein	HRG	P04196
185	Histone H1.0	H1F0	P07305
186	Histone H1.2	HIST1H1C	P16403
187	Histone H1.3	HIST1H1D	P16402
188	Histone H1.4	HIST1H1E	P10412
189	Histone H1.5	HIST1H1B	P16401
190	Histone H2A type 2-type 1-H	HIST2H2AA3	Q6FI13
191	Histone H2B type 1-N	HIST1H2BN	Q99877
192	Histone H3.2	HIST2H3A	Q71DI3
193	Histone H4	HIST1H4A	P62805
194	Ig alpha-1 chain C region	IGHA1	P01876
195	Ig alpha-2 chain C region	IGHA2	P01877
196	Ig gamma-1 chain C region	IGHG1	P01857
197	Ig gamma-2 chain C region	IGHG2	P01859
198	Ig gamma-3 chain C region	IGHG3	P01860
199	Ig gamma-4 chain C region	IGHG4	P01861
200	Ig heavy chain V-I region HG3		P01743
201	Ig heavy chain V-I region V35		P23083
202	Ig heavy chain V-II region ARH-77		P06331
203	Ig heavy chain V-III region BRO		P01766
204	Ig heavy chain V-III region BUT		P01767
205	Ig heavy chain V-III region GA		P01769
206	Ig heavy chain V-III region GAL		P01781
207	Ig heavy chain V-III region HIL		P01771
208	Ig heavy chain V-III region JON		P01780
209	Ig heavy chain V-III region KOL		P01772
210	Ig heavy chain V-III region LAY		P01775
211	Ig heavy chain V-III region TIL		P01765
212	Ig heavy chain V-III region TUR		P01779
213	Ig heavy chain V-III region VH26		P01764
214	Ig heavy chain V-III region WEA		P01763
215	Ig kappa chain C region	IGKC	P01834
216	Ig kappa chain V-I region AG		P01593
217	Ig kappa chain V-I region DEE		P01597
218	Ig kappa chain V-I region EU		P01598

219	Ig kappa chain V-I region HK102	IGKV1-5	P01602
220	Ig kappa chain V-I region Kue		P01604
221	Ig kappa chain V-I region Ni		P01613
222	Ig kappa chain V-I region Wes		P01611
223	Ig kappa chain V-II region RPMI 6410		P06310
224	Ig kappa chain V-II region TEW		P01617
225	Ig kappa chain V-III region CLL		P04207
226	Ig kappa chain V-III region HAH		P18135
227	Ig kappa chain V-III region NG9		P01621
228	Ig kappa chain V-III region VG		P04433
229	Ig kappa chain V-III region WOL		P01623
230	Ig kappa chain V-IV region	IGKV4-1	P06312
231	Ig kappa chain V-IV region Len		P01625
232	Ig lambda chain V-I region HA		P01700
233	Ig lambda chain V-I region NEWM		P01703
234	Ig lambda chain V-I region NIG-64		P01702
235	Ig lambda chain V-I region WAH		P04208
236	Ig lambda chain V-II region BUR		P01708
237	Ig lambda chain V-III region LOI		P80748
238	Ig lambda chain V-III region SH		P01714
239	Ig lambda chain V-IV region Hil		P01717
240	Ig lambda-2 chain C regions	IGLC2	P0CG05
241	Ig lambda-7 chain C region	IGLC7	A0M8Q6
242	Ig mu chain C region	IGHM	P01871
243	Ig mu heavy chain disease protein		P04220
244	IgGFc-binding protein	FCGBP	Q9Y6R7
245	Immunoglobulin J chain	IGJ	P01591
246	Immunoglobulin lambda-like polypeptide 1	IGLL1	P15814
247	Immunoglobulin lambda-like polypeptide 5	IGLL5	B9A064
248	Insulin-like growth factor-binding protein 2	IGFBP2	P18065
249	Insulin-like growth factor-binding protein 7	IGFBP7	Q16270
250	Integrin alpha-M	ITGAM	P11215
251	Integrin beta-2	ITGB2	P05107
252	Interleukin-1 receptor antagonist protein	IL1RN	P18510
253	Kallikrein-1	KLK1	P06870
254	Kallikrein-11	KLK11	Q9UBX7
255	Keratocan	KERA	O60938
256	Kinesin-like protein KIF16B	KIF16B	Q96L93
257	Kininogen-1	KNG1	P01042
258	Lactoperoxidase	LPO	P22079
259	Lactotransferrin	LTF	P02788
260	Lamin-B1	LMNB1	P20700
261	Leucine-rich alpha-2-glycoprotein	LRG1	P02750
262	Leukocyte elastase inhibitor	SERPINB1	P30740
263	Leukotriene A-4 hydrolase	LTA4H	P09960

264	Lipase member H	LIPH	H7BZL3
265	Lipocalin-1	LCN1	P31025
266	L-lactate dehydrogenase A chain	LDHA	P00338
267	L-lactate dehydrogenase A chain	LDHA	F5GZQ4
268	L-lactate dehydrogenase B chain	LDHB	P07195
269	Ly-6/neurotoxin-like protein 1	LYNX1	Q9BZG9
270	Ly6/PLAUR domain-containing protein 2	LYPD2	Q6UXB3
271	Ly6/PLAUR domain-containing protein 3	LYPD3	O95274
272	Lysosomal Pro-X carboxypeptidase	PRCP	P42785
273	Lysozyme C	LYZ	P61626
274	Macrophage migration inhibitory factor	MIF	P14174
275	Macrophage-capping protein	CAPG	P40121
276	Malate dehydrogenase, cytoplasmic	MDH1	P40925
277	Malate dehydrogenase, mitochondrial	MDH2	P40926
278	Maltase-glucoamylase, intestinal	MGAM	E7ER45
279	Matrix metalloproteinase-9	MMP9	P14780
280	Metalloproteinase inhibitor 1	TIMP1	P01033
281	Metalloproteinase inhibitor 2	TIMP2	P16035
282	Moesin	MSN	P26038
283	Mucin-16	MUC16	Q8WXI7
284	Mucin-4	MUC4	Q99102
285	Mucin-5AC	MUC5AC	P98088
286	Mucin-5B	MUC5B	Q9HC84
287	Mucin-7	MUC7	D6RHX1
288	Myeloblastin	PRTN3	P24158
289	Myeloperoxidase	MPO	P05164
290	Myeloperoxidase	MPO	J3QSF7
291	Myosin light polypeptide 6	MYL6	P60660
292	Myosin-9	MYH9	P35579
293	Myristoylated alanine-rich C-kinase substrate	MARCKS	P29966
294	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	SLC9A3R1	O14745
295	N-acetylgalactosamine-6-sulfatase	GALNS	P34059
296	N-alpha-acetyltransferase 38, NatC auxiliary subunit	NAA38	O95777
297	Neutrophil collagenase	MMP8	P22894
298	Neutrophil cytosol factor 2	NCF2	P19878
299	Neutrophil defensin 3	DEFA3	P59666
300	Neutrophil elastase	ELANE	P08246
301	Neutrophil gelatinase-associated lipocalin	LCN2	P80188
302	Non-histone chromosomal protein HMG-17	HMGN2	P05204
303	Non-secretory ribonuclease	RNASE2	P10153
304	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	NUCKS1	Q9H1E3
305	Nucleobindin-1	NUCB1	Q02818
306	Nucleophosmin	NPM1	P06748
307	Nucleoside diphosphate kinase B	NME2	P22392
308	Olfactomedin-4	OLFM4	Q6UX06

309	Pantetheinase	VNN1	O95497
310	Peptidoglycan recognition protein 1	PGLYRP1	O75594
311	Peptidyl-prolyl cis-trans isomerase A	PPIA	P62937
312	Peptidyl-prolyl cis-trans isomerase B	PPIB	P23284
313	Peptidyl-prolyl cis-trans isomerase FKBP1A	FKBP1A	P62942
314	Peroxiredoxin-1	PRDX1	Q06830
315	Peroxiredoxin-5, mitochondrial	PRDX5	P30044
316	Phosphatidylethanolamine-binding protein 1	PEBP1	P30086
317	Phosphoglycerate kinase	PGK1	B7Z7A9
318	Phosphoglycerate mutase 1	PGAM1	P18669
319	Phospholipase B-like 1	PLBD1	Q6P4A8
320	Phospholipid transfer protein	PLTP	P55058
321	Pigment epithelium-derived factor	SERPINF1	P36955
322	Plasma protease C1 inhibitor	SERPING1	P05155
323	Plasminogen	PLG	P00747
324	Plastin-2	LCP1	P13796
325	Pleckstrin	PLEK	P08567
326	Polymeric immunoglobulin receptor	PIGR	P01833
327	Polyubiquitin-B	UBB	P0CG47
328	PR domain zinc finger protein 5	PRDM5	Q9NQX1
329	Prelamin-A/C	LMNA	P02545
330	Proactivator polypeptide	PSAP	P07602
331	Pro-cathepsin H	CTSH	P09668
332	Profilin-1	PFN1	P07737
333	Prolactin-inducible protein	PIP	P12273
334	Proline-rich protein 1	PROL1	Q99935
335	Proline-rich protein 4	PRR4	Q16378
336	Prominin-1	PROM1	O43490
337	Proteasome subunit alpha type-1	PSMA1	P25786
338	Proteasome subunit alpha type-4	PSMA4	P25789
339	Proteasome subunit alpha type-5	PSMA5	P28066
340	Proteasome subunit alpha type-7	PSMA7	O14818
341	Proteasome subunit beta type-1	PSMB1	P20618
342	Proteasome subunit beta type-2	PSMB2	P49721
343	Proteasome subunit beta type-3	PSMB3	P49720
344	Proteasome subunit beta type-8	PSMB8	P28062
345	Proteasome subunit beta type-9	PSMB9	P28065
346	Protein disulfide-isomerase	P4HB	P07237
347	Protein disulfide-isomerase A3	PDIA3	P30101
348	Protein disulfide-isomerase A4	PDIA4	P13667
349	Protein DJ-1	PARK7	Q99497
350	Protein FAM25G	FAM25G	B3EWG6
351	Protein FAM3D	FAM3D	Q96BQ1
352	Protein NipSnap homolog 3A	NIPSNAP3A	Q9UFN0
353	Protein S100-A10	S100A10	P60903

354	Protein S100-A11	S100A11	P31949
355	Protein S100-A12	S100A12	P80511
356	Protein S100-A4	S100A4	P26447
357	Protein S100-A6	S100A6	P06703
358	Protein S100-A7	S100A7	P31151
359	Protein S100-A8	S100A8	P05109
360	Protein S100-A9	S100A9	P06702
361	Protein S100-P	S100P	P25815
362	Protein unc-13 homolog D	UNC13D	Q70J99
363	Protein-glutamine gamma-glutamyltransferase E	TGM3	Q08188
364	Prothymosin alpha	PTMA	P06454
365	Pulmonary surfactant-associated protein B	SFTPB	H0Y7V6
366	Purine nucleoside phosphorylase	PNP	P00491
367	Pyruvate kinase PKM	PKM	P14618
368	Rab GDP dissociation inhibitor beta	GDI2	P50395
369	Ras GTPase-activating-like protein IQGAP1	IQGAP1	P46940
370	Ras-related C3 botulinum toxin substrate 1	RAC1	P63000
371	Receptor-type tyrosine-protein phosphatase C	PTPRC	P08575
372	Receptor-type tyrosine-protein phosphatase eta	PTPRJ	Q12913
373	Rho GDP-dissociation inhibitor 2	ARHGDIB	P52566
374	Ribonuclease T2	RNASET2	O00584
375	Ribose-5-phosphate isomerase	RPIA	P49247
376	Ribulose-phosphate 3-epimerase	RPE	Q96AT9
377	Ryanodine receptor 2	RYR2	Q92736
378	Salivary acidic proline-rich phosphoprotein 1/2	PRH1	P02810
379	Secretoglobin family 3A member 1	SCGB3A1	Q96QR1
380	Serine protease HTRA1	HTRA1	Q92743
381	Serine protease inhibitor Kazal-type 5	SPINK5	Q9NQ38
382	Serotransferrin	TF	P02787
383	Serpin B10	SERPINB10	P48595
384	Serpin B3	SERPINB3	P29508
385	Serum amyloid P-component	APCS	P02743
386	SH3 domain-binding glutamic acid-rich-like protein	SH3BGRL	O75368
387	SH3 domain-binding glutamic acid-rich-like protein 3	SH3BGRL3	Q9H299
388	Small proline-rich protein 2B	SPRR2B	P35325
389	Small proline-rich protein 2E	SPRR2E	P22531
390	Small proline-rich protein 3	SPRR3	Q9UBC9
391	Stathmin	STMN1	P16949
392	Submaxillary gland androgen-regulated protein 3B	SMR3B	P02814
393	Sulfhydryl oxidase 1	QSOX1	O00391
394	Superoxide dismutase [Cu-Zn]	SOD1	P00441
395	Superoxide dismutase [Mn], mitochondrial	SOD2	P04179
396	Suprabasin	SBSN	Q6UWP8
397	Synaptic vesicle membrane protein VAT-1 homolog	VAT1	Q99536
398	Syntenin-1	SDCBP	O00560

399	Talin-1	TLN1	Q9Y490
400	Thioredoxin	TXN	P10599
401	Thioredoxin domain-containing protein 12	TXNDC12	O95881
402	Thymosin beta-10	TMSB10	P63313
403	Thymosin beta-4	TMSB4X	P62328
404	Titin	TTN	Q8WZ42
405	Transaldolase	TALDO1	P37837
406	Transcobalamin-1	TCN1	P20061
407	Transgelin-2	TAGLN2	P37802
408	Transketolase	TKT	P29401
409	Translin	TSN	Q15631
410	Translin-associated protein X	TSNAX	Q99598
411	Transthyretin	TTR	P02766
412	Trefoil factor 1	TFF1	P04155
413	Trefoil factor 3	TFF3	Q07654
414	Triosephosphate isomerase	TPI1	P60174
415	Tropomyosin alpha-3 chain	TPM3	P06753
416	Tubulin alpha-1B chain	TUBA1B	P68363
417	Tubulin beta-4B chain	TUBB4B	P68371
418	Tumor necrosis factor receptor superfamily member 1A	TNFRSF1A	P19438-5
419	Twinfilin-2	TWF2	Q6IBS0
420	Tyrosine-protein phosphatase non-receptor type substrate 1	SIRPA	P78324
421	Uncharacterized protein C5orf64	C5orf64	Q2M2E5
422	UPF0556 protein C19orf10	C19orf10	Q969H8
423	UPF0762 protein C6orf58	C6orf58	Q6P5S2
424	Uteroglobin	SCGB1A1	E9PN95
425	UTP--glucose-1-phosphate uridylyltransferase	UGP2	Q16851
426	Vascular non-inflammatory molecule 2	VNN2	O95498
427	Vasodilator-stimulated phosphoprotein	VASP	P50552
428	VEGF co-regulated chemokine 1	CXCL17	Q6UXB2
429	Vimentin	VIM	P08670
430	Vinculin	VCL	P18206
431	Vitamin D-binding protein	GC	P02774
432	Vitronectin	VTN	P04004
433	Voltage-gated potassium channel subunit beta-2	KCNAB2	Q13303
434	WAP four-disulfide core domain protein 2	WFDC2	Q14508
435	WASH complex subunit FAM21C	FAM21C	Q9Y4E1
436	Xaa-Pro dipeptidase	PEPD	P12955
437	Zinc-alpha-2-glycoprotein	AZGP1	P25311
438	Zymogen granule protein 16 homolog B	ZG16B	Q96DA0

Supplementary table 2: The lists of non-redundant neutrophil proteins identified in CF, iNCF and NCF sputum.

#	Protein names	Gene IDs	Uniprot IDs
1	40S ribosomal protein S3	RPS3	P23396
2	60S acidic ribosomal protein P2	RPLP2	P05387
3	6-phosphogluconate dehydrogenase, decarboxylating	PGD	P52209
4	6-phosphogluconolactonase	PGLS	O95336
5	78 kDa glucose-regulated protein	HSPA5	P11021
6	Acid ceramidase	ASAH1	Q13510
7	Adenylyl cyclase-associated protein 1	CAP1	Q01518
8	Alpha-1-acid glycoprotein 2	ORM2	P19652
9	Alpha-1-antichymotrypsin	SERPINA3	P01011
10	Alpha-1-antitrypsin	SERPINA1	P01009
11	Alpha-1B-glycoprotein	A1BG	P04217
12	Alpha-2-HS-glycoprotein	AHSG	P02765
13	Alpha-actinin-1	ACTN1	P12814
14	Alpha-actinin-4	ACTN4	O43707
15	Alpha-enolase	ENO1	P06733
16	Alpha-enolase	ENO1	P06733-2
17	Aminopeptidase N	ANPEP	P15144
18	Angiotensinogen	AGT	P01019
19	Annexin A1	ANXA1	P04083
20	Annexin A2	ANXA2	P07355
21	Annexin A3	ANXA3	P12429
22	Annexin A6	ANXA6	P08133
23	Antileukoproteinase	SLPI	P03973
24	Antithrombin-III	SERPINC1	P01008
25	Apolipoprotein A-I	APOA1	P02647
26	Apolipoprotein A-II	APOA2	P02652
27	Apolipoprotein A-IV	APOA4	P06727
28	Apolipoprotein B-100	APOB	P04114
29	Apoptosis-associated speck-like protein containing a CARD	PYCARD	Q9ULZ3
30	Arginase-1	ARG1	P05089
31	ATP synthase subunit alpha, mitochondrial	ATP5A1	P25705
32	Azurocidin	AZU1	P20160
33	Bactericidal permeability-increasing protein	BPI	P17213
34	Beta-2-glycoprotein 1	APOH	P02749

35	Brain acid soluble protein 1	BASP1	P80723
36	Calcineurin subunit B type 1	PPP3R1	P63098
37	Calmodulin	CALM1	P62158
38	Calmodulin-like protein 5	CALML5	Q9NZT1
39	Calnexin	CANX	P27824
40	Calpain-1 catalytic subunit	CAPN1	P07384
41	Calreticulin	CALR	P27797
42	Carcinoembryonic antigen-related cell adhesion molecule 8	CEACAM8	P31997
43	Catalase	CAT	P04040
44	Cathelicidin antimicrobial peptide	CAMP	P49913
45	Cathepsin B	CTSB	P07858
46	Cathepsin D	CTSD	P07339
47	Cathepsin G	CTSG	P08311
48	Cathepsin S	CTSS	P25774
49	CD59 glycoprotein	CD59	P13987
50	Ceruloplasmin	CP	P00450
51	Chitinase-3-like protein 1	CHI3L1	P36222
52	Chitotriosidase-1	CHIT1	Q13231
53	Clusterin	CLU	P10909
54	Coactosin-like protein	COTL1	Q14019
55	Cofilin-1	CFL1	P23528
56	Complement C3chain	C3	P01024
57	Coronin-1A	CORO1A	P31146
58	Cystatin-A	CSTA	P01040
59	Cystatin-B	CSTB	P04080
60	Cystatin-C	CST3	P01034
61	Cysteine-rich secretory protein 3	CRISP3	P54108
62	Cytidine deaminase	CDA	P32320
63	Cytochrome c	CYCS	P99999
64	D-dopachrome decarboxylase	DDT	P30046
65	EF-hand domain-containing protein D2	EFHD2	Q96C19
66	Elongation factor 1-alpha 1	EEF1A1	P68104
67	Endoplasmin	HSP90B1	P14625
68	Eosinophil cationic protein	RNASE3	P12724
69	Eosinophil peroxidase	EPX	P11678
70	Epididymal secretory protein E1	NPC2	P61916
71	Ezrin	EZR	P15311

72	F-actin-capping protein subunit alpha-1	CAPZA1	P52907
73	Fatty acid-binding protein, epidermal	FABP5	Q01469
74	Ferritin heavy chain	FTH1	P02794
75	Ferritin light chain	FTL	P02792
76	Filamin-A	FLNA	P21333
77	Folate receptor alpha	FOLR1	P15328
78	Fructose-bisphosphate aldolase A	ALDOA	P04075
79	Fructose-bisphosphate aldolase C	ALDOC	P09972
80	Galectin-3	LGALS3	P17931
81	Galectin-3-binding protein	LGALS3BP	Q08380
82	Gamma-glutamyl hydrolase	GGH	Q92820
83	Gamma-glutamyltransferase 5	GGT5	H7C1X2
84	Gelsolin	GSN	P06396
85	Glucose-6-phosphate isomerase	GPI	P06744
86	Glutamate dehydrogenase 1, mitochondrial	GLUD1	P00367
87	Glutaredoxin-1	GLRX	P35754
88	Glutathione reductase, mitochondrial	GSR	P00390
89	Glutathione S-transferase P	GSTP1	P09211
90	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	P04406
91	Glycogen phosphorylase	PYGL	P06737
92	Granulins	GRN	P28799
93	Haptoglobin	HP	P00738
94	Heat shock 70 kDa protein 1A/1B	HSPA1A	P08107
95	Heat shock 70 kDa protein 6	HSPA6	P17066
96	Heat shock cognate 71 kDa protein	HSPA8	P11142
97	Heat shock protein HSP 90-alpha	HSP90AA1	P07900
98	Heat shock protein HSP 90-beta	HSP90AB1	P08238
99	Hemoglobin subunit alpha	HBA1	P69905
100	Hemoglobin subunit beta	HBB	P68871
101	Heterogeneous nuclear ribonucleoprotein K	HNRNPK	P61978
106	Integrin alpha-M	ITGAM	P11215
107	Integrin beta-2	ITGB2	P05107
108	Interleukin-1 receptor antagonist protein	IL1RN	P18510
109	Lactotransferrin	LTF	P02788
110	Lamin-B1	LMNB1	P20700
111	Leucine-rich alpha-2-glycoprotein	LRG1	P02750
112	Leukocyte elastase inhibitor	SERPINB1	P30740

113	Leukotriene A-4 hydrolase	LTA4H	P09960
114	L-lactate dehydrogenase B chain	LDHB	P07195
115	Lysozyme C	LYZ	P61626
116	Macrophage migration inhibitory factor	MIF	P14174
117	Malate dehydrogenase, cytoplasmic	MDH1	P40925
118	Malate dehydrogenase, mitochondrial	MDH2	P40926
119	Maltase-glucoamylase, intestinal	MGAM	E7ER45
120	Matrix metalloproteinase-9	MMP9	P14780
121	Metalloproteinase inhibitor 2	TIMP2	P16035
122	Moesin	MSN	P26038
123	Myeloblastin	PRTN3	P24158
124	Myeloperoxidase	MPO	P05164
125	Myosin light polypeptide 6	MYL6	P60660
126	Myosin-9	MYH9	P35579
127	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	SLC9A3R1	O14745
128	N-acetylgalactosamine-6-sulfatase	GALNS	P34059
129	Neutrophil collagenase	MMP8	P22894
130	Neutrophil cytosol factor 2	NCF2	P19878
131	Neutrophil defensin 3	DEFA3	P59666
132	Neutrophil elastase	ELANE	P08246
133	Neutrophil gelatinase-associated lipocalin	LCN2	P80188
134	Non-secretory ribonuclease	RNASE2	P10153
135	Nucleobindin-1	NUCB1	Q02818
136	Olfactomedin-4	OLFM4	Q6UX06
137	Pantetheinase	VNN1	O95497
138	Peptidoglycan recognition protein 1	PGLYRP1	O75594
139	Peptidyl-prolyl cis-trans isomerase A	PPIA	P62937
140	Peptidyl-prolyl cis-trans isomerase B	PPIB	P23284
141	Peroxiredoxin-1	PRDX1	Q06830
142	Peroxiredoxin-5, mitochondrial	PRDX5	P30044
143	Phosphatidylethanolamine-binding protein 1	PEBP1	P30086
144	Phosphoglycerate mutase 1	PGAM1	P18669
145	Phospholipase B-like 1	PLBD1	Q6P4A8
146	Plastin-2	LCP1	P13796
147	Pleckstrin	PLEK	P08567
148	Polymeric immunoglobulin receptor	PIGR	P01833
149	Pro-cathepsin H	CTSH	P09668

150	Profilin-1	PFN1	P07737
151	Prolactin-inducible protein	PIP	P12273
152	Proteasome subunit alpha type-1	PSMA1	P25786
153	Proteasome subunit alpha type-4	PSMA4	P25789
154	Proteasome subunit alpha type-5	PSMA5	P28066
155	Proteasome subunit alpha type-7	PSMA7	O14818
156	Proteasome subunit beta type-1	PSMB1	P20618
157	Proteasome subunit beta type-2	PSMB2	P49721
158	Proteasome subunit beta type-3	PSMB3	P49720
159	Protein disulfide-isomerase	P4HB	P07237
160	Protein disulfide-isomerase A3	PDIA3	P30101
161	Protein disulfide-isomerase A4	PDIA4	P13667
162	Protein DJ-1	PARK7	Q99497
163	Protein NipSnap homolog 3A	NIPSNAP3A	Q9UFN0
164	Protein S100-A11	S100A11	P31949
165	Protein S100-A12	S100A12	P80511
166	Protein S100-A4	S100A4	P26447
167	Protein S100-A6	S100A6	P06703
168	Protein S100-A7	S100A7	P31151
169	Protein S100-A8	S100A8	P05109
170	Protein S100-A9	S100A9	P06702
171	Protein S100-P	S100P	P25815
172	Protein unc-13 homolog D	UNC13D	Q70J99
173	Protein-glutamine gamma-glutamyltransferase E	TGM3	Q08188
174	Purine nucleoside phosphorylase	PNP	P00491
175	Pyruvate kinase PKM	PKM	P14618
176	Ras GTPase-activating-like protein IQGAP1	IQGAP1	P46940
177	Ras-related C3 botulinum toxin substrate 1	RAC1	P63000
178	Receptor-type tyrosine-protein phosphatase C	PTPRC	P08575
179	Receptor-type tyrosine-protein phosphatase eta	PTPRJ	Q12913
180	Rho GDP-dissociation inhibitor 2	ARHGDIB	P52566
181	Serotransferrin	TF	P02787
182	Serpin B10	SERPINB10	P48595
183	Serpin B3	SERPINB3	P29508
184	SH3 domain-binding glutamic acid-rich-like protein	SH3BGRL	O75368
185	Sulfhydryl oxidase 1	QSOX1	O00391
186	Superoxide dismutase [Mn], mitochondrial	SOD2	P04179

187	Suprabasin	SBSN	Q6UWP8
188	Synaptic vesicle membrane protein VAT-1 homolog	VAT1	Q99536
189	Syntenin-1	SDCBP	O00560
190	Talin-1	TLN1	Q9Y490
191	Thioredoxin	TXN	P10599
192	Thioredoxin domain-containing protein 12	TXNDC12	O95881
193	Transaldolase	TALDO1	P37837
194	Transcobalamin-1	TCN1	P20061
195	Transgelin-2	TAGLN2	P37802
196	Transketolase	TKT	P29401
197	Translin	TSN	Q15631
198	Translin-associated protein X	TSNAX	Q99598
199	Transthyretin	TTR	P02766
200	Triosephosphate isomerase	TPI1	P60174
201	Tropomyosin alpha-3 chain	TPM3	P06753
202	Tubulin beta-4B chain	TUBB4B	P68371
203	Twinfilin-2	TWF2	Q6IBS0
204	UPF0556 protein C19orf10	C19orf10	Q969H8
205	UTP--glucose-1-phosphate uridylyltransferase	UGP2	Q16851
206	Vascular non-inflammatory molecule 2	VNN2	O95498
207	Vasodilator-stimulated phosphoprotein	VASP	P50552
208	Vimentin	VIM	P08670
209	Vinculin	VCL	P18206
210	Vitamin D-binding protein	GC	P02774
211	Zinc-alpha-2-glycoprotein	AZGP1	P25311

Chapter 5

Conclusions and future directions

5.1 Conclusions

5.1.1 Cystic Fibrosis – a disease characterised by genetic abnormality and infection

Cystic fibrosis (CF) is an autosomal recessive disorder affecting each of the exocrine glands, but the most life threatening aspect of the disease is the impact on the lungs and respiratory system. CF is due to a mutation of the cystic fibrosis transmembrane conductance regulator (CFTR) gene disrupting the regulation of sodium and potassium ion channels (Wine 1999). Disruption in the ion channels causes excessive and viscous mucous secretion in the lung airways and the defective salt transport dehydrates the mucus, causing difficulty in mucociliary clearance (Pilewski et al. 1999).

Defective mucociliary clearance in the CF lung airways leads to infection and inflammation, which are the hallmarks of CF disease. Chronic infection and inflammation together deteriorate the lung function and characterise the severity of the disease. The thick and viscous mucus also harbours a variety of different pathogenic microorganisms in the lung including bacteria and fungi. *P. aeruginosa* and *S. aureus* are the major bacterial species and *A. fumigatus*, the major fungal species infecting the CF lung (Yang et al. 2011).

5.1.2 Involvement of protein glycosylation in CF

Mucins are the most abundant high-molecular mass proteins in mucus, which are heavily *O*-glycosylated on the tandem repeats of serine/threonine residues. Host mucin *O*-glycosylation in CF has been well-studied in the literature for more than three decades and has been shown to be correlated with, but not caused by, the CFTR mutation. Mucin acts as a first line of defence and plays a protective role in the innate immune system of healthy individuals, but in viscous mucus in CF, mucins play a role in bacterial adhesion – the first step in pathogenic infection. For example, *P. aeruginosa* utilises membrane bound lectins such as PA-IL and PA-IIL that bind to galactose and fucose residues respectively, and has other adhesins like flagellin that bind to different glycan receptors such as galactose, fucose, lewis^{x} and sialyl-lewis^{x} on the mucins (Avichezer et al. 1992; Chen et al. 1998; Lillehoj et al. 2002; Sabin et al. 2006; Marotte et al. 2007; Venkatakrishnan et al. 2013). Although aberrant mucin *O*-glycosylation and its role in CF pathogenesis are well documented in the literature, the

characterisation and role of *N*-protein glycosylation in CF had not been investigated prior to the commencement of this thesis.

The study presented in this thesis is thus the first to characterise the *N*-glycosylation of CF sputum compared to that of non-CF sputum. Since sputum represents the major part of the contents of the central lung airway, the information regarding the protein components, cellular compartments and microbial infection can potentially assist in the identification of biomarkers for severity of the disease and its progression. Samples of sputum were obtained from nine patients, including five CF patients with bacterial infection of the lungs/ colonised CF (CF1-5), and four non-CF patients, two of which had pathogen infected/colonised (iNCF1-2) sputum and two that were pathogen free (NCF1-2).

5.1.3 The *N*-glycome of CF and infected non-CF (iNCF) sputum is characterised by the unusual incidence of pauci-mannose glycans

In each of the samples of bacterial infected sputum, whether CF or infected non-CF, there was an increase in the relative abundance of the unusual pauci-mannosidic *N*-glycan, a truncated *N*-glycan that has a monosaccharide composition equal to or less than that of the chitobiose core. The pauci-mannose *N*-glycans were the most abundant *N*-glycans in CF and iNCF sputum irrespective of patient's age, gender, genotype and the microbial flora colonising the lung airway. Thus the increase in pauci-mannosidic *N*-glycans in sputum was not patient specific or disease specific but could be associated with pathogen infection in general.

Pauci-mannose on proteins is not a part of the normal biosynthetic pathway of *N*-glycans in mammals and to date it has been mainly reported to occur on plants and invertebrate proteins (Schachter 2009; Dam et al. 2013). In *C. elegans*, it has been reported that pauci-mannose glycans are necessary for the survival of the organism and also play a role in the innate immunity of the worm (Shi et al. 2006; Schachter 2009). Whether, or how, mammals synthesise pauci-mannose is not known. Hence, the finding of abundant pauci-mannose glycans on the proteins of sputum from the human CF lung reported in this work is of significance. The role of pauci-mannose glycans in humans is yet to be determined.

In CF and iNCF sputum, pauci-mannose *N*-glycans were found to be at a level of approximately 40% and 30% respectively of the total *N*-glycans, which was 10-fold and 7-fold more than in non-infected non-CF sputum. The increase in pauci-mannose *N*-glycans in CF and iNCF sputum was accompanied by a two-fold decrease in complex and hybrid *N*-glycans relative to the abundance of high mannose *N*-glycans that stayed statistically constant between pathogen colonised CF sputum and pathogen free non-CF sputum. These observations lead to the possibility that either the complex or hybrid *N*-glycans (but not the high mannose *N*-glycans) have been trimmed by the infecting bacteria to form pauci-mannose *N*-glycans following secretion of the glycoproteins, or that the pauci-mannose structures were synthesised in the host cells prior to secretion of the proteins into the sputum mucus.

5.1.4 No evidence for the direct generation of pauci-mannose *N*-glycans by *P. aeruginosa*

To address whether the pauci-mannose glycans in the sputum could be produced by the action of bacterial exoglycosidases, four clinical *P. aeruginosa* bacterial strains were isolated from CF sputum, namely PASS1, PASS2, PASS3 and PASS4, that were identified as *Pseudomonas* strains using 16S rRNA sequencing. One purpose of trimming of monosaccharides from secreted glycoproteins by bacteria could be to utilise the sugars as a carbon source for metabolism. However, the Biolog phenotype microarray study determined that the monosaccharide residues that comprise the complex and hybrid *N*-glycans were not utilised (at least as a sole carbon source) by bacteria for cellular respiration. In addition, there was no evidence of enzymatic cleavage of complex, hybrid or high mannose *N*-glycans by any bacterial exoglycosidase activity (such as neuraminidase, galactosidase, hexosaminidase and mannosidase) to produce pauci-mannosidic *N*-glycans on three purified glycoproteins (fetuin, IgG and RNaseB) or on non-CF sputum glycoproteins. Bacterial glycoproteins are reported to carry different sugar residues to mammalian proteins, but as a further check on the possibility that the pauci-mannose glycans in the sputum were directly of bacterial origin, *N*-glycans were released from the proteins of a *P. aeruginosa* whole cell lysate to investigate if the bacterial proteins carry the truncated *N*-linked pauci-mannose. Pauci-mannosidic *N*-glycans were not found in the *P. aeruginosa* glycome thus eliminating the

possibility of bacterial origin of these structures. The question of other bacterial glycosidases other than from P. aeruginosa glycosidase may theoretically be possible in the production of pauci-mannose in infected sputum. However, the fact that we found pauci-mannose exclusively on neutrophil proteins strongly suggests, it is not a degradation product.

5.1.5 Pauci-mannose glycans in sputum are found to be of neutrophil origin

Since direct generation of the pauci-mannose glycans by the bacteria present in the sputum was not supported, the alternate possibility that humans synthesise pauci-mannose structures on glycoproteins and also whether the variation observed in sputum *N*-glycosylation is a result of protein changes in different sputum category was investigated. As the first step to identify any differences in the glycoproteins present in CF and non-CF sputum, a global proteome analysis was carried out. The most remarkable immediate difference was the increased abundance of immune response related neutrophil proteins in bacterial colonised CF and iNCF sputum in comparison to non-infected sputum. Abundant neutrophil glycoproteins in the CF sputum were identified as myeloperoxidase, azurocidin, neutrophil elastase, and matrix metalloproteinase-9, all of which have previously been shown to have antibacterial activity by killing the bacteria *in vitro* (Bangalore et al. 1990; Nauseef 2001; Hirche et al. 2005).

Analysis of the *N*-glycans released from whole cell proteins of neutrophils isolated from blood identified abundant pauci-mannosidic *N*-glycans, supporting the contention that human neutrophil proteins carry these unusual structures. Moreover, the *N*-glycan profile of the neutrophils closely matched the CF sputum *N*-glycosylation profile further suggesting that the abundant pauci-mannose glycans are synthesised in neutrophils. The final proof that the pauci-mannose structures characterised in the infected sputum samples (both CF and infected nonCF) were indeed of neutrophil origin was determined by site-specific *N*-glycoproteome analysis. The abundant CF neutrophil proteins myeloperoxidase, azurocidin and phospholipase B-like proteins were found to carry pauci-mannosidic glycans at different *N*-linked glycosylation sites.

5.1.6 A proposed biosynthetic pathway of pauci-mannose glycans in human neutrophils

We propose that a modified *N*-glycan biosynthetic (**Chapter 4, Figure 6**) pathway occurs in neutrophils, with the alternate action of a β-hexosaminidase enzyme found to be in neutrophils and has been characterised for its function in glycolipid degradation. The β-hexosaminidase enzyme cleaves the terminal GlcNAc residue from the $GlcNAc_2Man_3GlcNAc_1$ *N*-glycan structure that is found on proteins as the first step towards the common synthesis of hybrid *N*-glycans. This cleavage is followed by the activity of α- and β-mannosidases, which cleave the mannose residues down to the chitobiose core or further to the Man2, Man1 and Man0 level, thus synthesising the truncated pauci-mannosidic type *N*-glycans. As such, these pauci-mannose *N*-glycans should be acknowledged as a fourth major class of *N*-glycans in humans, alongside high mannose, hybrid and complex *N*-glycans.

5.1.7 Pauci-mannose – a new indicator of lung inflammation and infection in CF and other respiratory diseases

Direct bacterial generation of pauci-mannose glycans on the proteins in infected sputum was effectively discounted in this work. Nevertheless, an increase in the abundance of pauci-mannosidic *N*-glycans correlated well with the level of pathogenic infection. CF patients with chronic infection exhibited a higher abundance of pauci-mannosidic *N*-glycans on their sputum proteins as compared to sputum of patients with upper respiratory tract infection (URTI, iNCF), who would be expected to have a lower level of bacterial colonisation. Correspondingly, uninfected sputum contained very few pauci-mannose *N*-glycans. This work also established the origin of the pauci-mannose as being on neutrophil proteins known to have antimicrobial activity. Thus, to conclude, albeit on a limited number of samples, the *N*-glycosylation on CF sputum proteins appears to be indirectly related to bacterial colonisation/infection. Moreover, a sputum *N*-glycome that contains abundant pauci-mannose may be a general indicator of a neutrophilic immune response defending the host against pathogenic invasion. To have clinical relevance, clearly more samples from patients with and without infection needs to be investigated.

5.2 Future directions

The role of *N*-glycosylation in a disease like CF is yet to be solved. In all samples of lung sputum we have seen an increase in an unusual type of protein *N*-glycan class called pauci-mannose, which is clearly synthesised by humans despite previous reports to the contrary. Determining the functional role of these cell-specific unusual pauci-mannose *N*-glycans is very important. From this study (**Chapter 4**), we have shown that the pauci-mannose structures are carried on neutrophil glycoproteins and these proteins have antimicrobial activity. The bactericidal activity of these glycoproteins has been studied *in vitro* (Bardoel et al. 2014) but the mechanism *in vivo* is still unknown and the role of the identified pauci-mannose *N*-glycans in the antimicrobial activity of these proteins is of interest.

Pauci-mannose *N*-glycans seem to be important for *C. elegans* in defence against invading bacterial pathogens and in innate immune response of the worm. Possibly, a similar kind of role can be played by human pauci-mannose *N*-glycans in defending the host against invading pathogens. One such example may involve the mannose-binding lectin (MBL) that is involved in the innate immunity of individuals by activating the lectin complement pathway which triggers defence against the pathogens (Kilpatrick 2002; Kilpatrick 2002). A question to be answered is whether the terminal mannose residues of pauci-mannose can bind to MBL and take part in this immune defense pathway.

Neutrophils are granulated cells with primary, secondary and tertiary granules, each of the granules playing a specific role in innate immunity and defence. It is important to investigate whether pauci-mannose *N*-glycans are granule specific. Myeloperoxidase and azurocidin for example, are typical primary or azurophilic granule proteins and degranulation of the azurophilic granule releases biomolecules with antibacterial activity (Rorvig et al. 2013). As mentioned earlier, antimicrobial activity of neutrophil proteins such as myeloperoxidase, azurocidin and cathepsin G have been studied *in vitro* but, it will be necessary to see whether deglycosylating the neutrophil proteins changes the bactericidal activity.

Only a handful of studies recently have started to report *N*-linked reported pauci-mannose occurring at low abundance (< 10% of total *N*-glycans) in mammalian system (Balog et al. 2012; Everest-Dass et al. 2012; Sethi et al. 2014). From this study, we now know that humans

can synthesise pauci-mannose in high abundance (40%-45% of total *N*-glycans) in specific secretions, as we have seen in sputum of CF patients. It is thus important not to neglect (as in (Babu et al. 2009)) the presence of pauci-mannose in *N*-glycan MS profiles simply because the mass of these structures falls below the usual mass cut-off (*m*/*z* 1237.4 in +ve ion MS and *m*/*z* 1235.4 in –ve ion MS) of the core *N*-glycan. The role of pauci-mannose proteins and functions is expected to be more widespread than just CF and sputum.

In terms of CF as a disease and glycosylation, mucin *O*-glycosylation has been extensively studied in terms of bacterial adhesion and its exact contribution is still a subject of debate. Studying the *N*-glycoproteome and its role further enhances our understanding of the biomolecular environment of the CF lung. We have done an in-depth analysis of the glycosylation of infected sputum and have shown that sputum *N*-glycosylation in CF is mainly driven by bacterial colonisation/infection; there are other CF specimens like BAL fluid and lung epithelial cells in which the glycosylation is still unknown. Thus, as a result of this (initial) study, it is clear that future research should focus not only on understanding the general role of *N*-glycosylation, but also specifically that of the pauci-mannose *N*-glycans in different diseases including, but not limited to cystic fibrosis.

Chapter 6

References

Aaron, S. D., Ferris, W., Henry, D. A., Speert, D. P. and Macdonald, N. E. 2000. Multiple combination bactericidal antibiotic testing for patients with cystic fibrosis infected with Burkholderia cepacia. *Am J Respir Crit Care Med.* 161(4 Pt 1): 1206-1212.

Adam, R. J., Michalski, A. S., Bauer, C., Abou Alaiwa, M. H., Gross, T. J., Awadalla, M. S., Bouzek, D. C., Gansemer, N. D., Taft, P. J., Hoegger, M. J., Diwakar, A., Ochs, M., Reinhardt, J. M., Hoffman, E. A., Beichel, R. R., Meyerholz, D. K. and Stoltz, D. A. 2013. Air trapping and airflow obstruction in newborn cystic fibrosis piglets. *Am J Respir Crit Care Med.* 188(12): 1434-1441.

Alexander, B. D. 2002. Diagnosis of fungal infection: new technologies for the mycology laboratory. *Transpl Infect Dis.* 4 Suppl 3: 32-37.

Alexis, N. E., Hu, S. C., Zeman, K., Alter, T. and Bennett, W. D. 2001. Induced sputum derives from the central airways: confirmation using a radiolabeled aerosol bolus delivery technique. *Am J Respir Crit Care Med.* 164(10 Pt 1): 1964-1970.

Ali, M., Lillehoj, E. P., Park, Y., Kyo, Y. and Kim, K. C. 2011. Analysis of the proteome of human airway epithelial secretions. *Proteome Sci.* 9: 4.

Allesen-Holm, M., Barken, K. B., Yang, L., Klausen, M., Webb, J. S., Kjelleberg, S., Molin, S., Givskov, M. and Tolker-Nielsen, T. 2006. A characterization of DNA release in Pseudomonas aeruginosa cultures and biofilms. *Mol Microbiol.* 59(4): 1114-1128.

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(4): 2668-2732.

Andersen, D. H. 1938. Cystic fibrosis of the pancreas and its relation to celiac disease: A clinical and pathologic study. *American Journal of Diseases of Children.* 56(2): 344-399.

Avichezer, D., Katcoff, D. J., Garber, N. C. and Gilboa-Garber, N. 1992. Analysis of the amino acid sequence of the Pseudomonas aeruginosa galactophilic PA-I lectin. *J Biol Chem.* 267(32): 23023-23027.

Babu, P., North, S. J., Jang-Lee, J., Chalabi, S., Mackerness, K., Stowell, S. R., Cummings, R. D., Rankin, S., Dell, A. and Haslam, S. M. 2009. Structural characterisation of neutrophil glycans by ultra sensitive mass spectrometric glycomics methodology. *Glycoconj J.* 26(8): 975-986.

Bakare, N., Rickerts, V., Bargon, J. and Just-Nubling, G. 2003. Prevalence of Aspergillus fumigatus and other fungal species in the sputum of adult patients with cystic fibrosis. *Mycoses.* 46(1-2): 19-23.

Balog, C. I., Stavenhagen, K., Fung, W. L., Koeleman, C. A., McDonnell, L. A., Verhoeven, A., Mesker, W. E., Tollenaar, R. A., Deelder, A. M. and Wuhrer, M. 2012. N-glycosylation of colorectal cancer tissues: a liquid chromatography and mass spectrometry-based investigation. *Mol Cell Proteomics.* 11(9): 571-585.

Balough, K., McCubbin, M., Weinberger, M., Smits, W., Ahrens, R. and Fick, R. 1995. The relationship between infection and inflammation in the early stages of lung disease from cystic fibrosis. *Pediatr Pulmonol.* 20(2): 63-70.

Bangalore, N., Travis, J., Onunka, V. C., Pohl, J. and Shafer, W. M. 1990. Identification of the primary antimicrobial domains in human neutrophil cathepsin G. *J Biol Chem.* 265(23): 13584-13588.

Bardoel, B. W., Kenny, E. F., Sollberger, G. and Zychlinsky, A. 2014. The Balancing Act of Neutrophils. *Cell Host Microbe.* 15(5): 526-536.

Belaaouaj, A., McCarthy, R., Baumann, M., Gao, Z., Ley, T. J., Abraham, S. N. and Shapiro, S. D. 1998. Mice lacking neutrophil elastase reveal impaired host defense against gram negative bacterial sepsis. *Nat Med.* 4(5): 615-618.

Bielik, A. M. and Zaia, J. 2010. Historical overview of glycoanalysis. *Methods Mol Biol.* 600: 9-30.

Birrer, P., McElvaney, N. G., Rudeberg, A., Sommer, C. W., Liechti-Gallati, S., Kraemer, R., Hubbard, R. and Crystal, R. G. 1994. Protease-antiprotease imbalance in the lungs of children with cystic fibrosis. *Am J Respir Crit Care Med.* 150(1): 207-213.

Bittar, F., Richet, H., Dubus, J. C., Reynaud-Gaubert, M., Stremler, N., Sarles, J., Raoult, D. and Rolain, J. M. 2008. Molecular detection of multiple emerging pathogens in sputa from cystic fibrosis patients. *PLoS One.* 3(8): e2908.

Bochner, B. R. 2003. New technologies to assess genotype-phenotype relationships. *Nat Rev Genet.* 4(4): 309-314.

Bochner, B. R. 2009. Global phenotypic characterization of bacteria. *FEMS Microbiol Rev.* 33(1): 191-205.

Bochner, B. R., Gadzinski, P. and Panomitros, E. 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* 11(7): 1246-1255.

Borregaard, N. 2010. Neutrophils, from marrow to microbes. *Immunity.* 33(5): 657-670.

Boyum, A., Lovhaug, D., Tresland, L. and Nordlie, E. M. 1991. Separation of leucocytes: improved cell purity by fine adjustments of gradient medium density and osmolality. *Scand J Immunol.* 34(6): 697-712.

Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D. S., Weinrauch, Y. and Zychlinsky, A. 2004. Neutrophil extracellular traps kill bacteria. *Science.* 303(5663): 1532-1535.

Burns, J. L., Emerson, J., Stapp, J. R., Yim, D. L., Krzewinski, J., Louden, L., Ramsey, B. W. and Clausen, C. R. 1998. Microbiology of sputum from patients at cystic fibrosis centers in the United States. *Clin Infect Dis.* 27(1): 158-163.

Burns, J. L., Gibson, R. L., McNamara, S., Yim, D., Emerson, J., Rosenfeld, M., Hiatt, P., McCoy, K., Castile, R., Smith, A. L. and Ramsey, B. W. 2001. Longitudinal assessment of Pseudomonas aeruginosa in young children with cystic fibrosis. *J Infect Dis.* 183(3): 444-452.

Buschiazzo, A., Muia, R., Larrieux, N., Pitcovsky, T., Mucci, J. and Campetella, O. 2012. Trypanosoma cruzi trans-sialidase in complex with a neutralizing antibody: structure/function studies towards the rational design of inhibitors. *PLoS Pathog.* 8(1): e1002474.

Cabral, D. A., Loh, B. A. and Speert, D. P. 1987. Mucoid Pseudomonas aeruginosa resists nonopsonic phagocytosis by human neutrophils and macrophages. *Pediatr Res.* 22(4): 429-431.

Cacalano, G., Kays, M., Saiman, L. and Prince, A. 1992. Production of the Pseudomonas aeruginosa neuraminidase is increased under hyperosmolar conditions and is regulated by genes involved in alginate expression. *J Clin Invest.* 89(6): 1866-1874.

Camara, M., Boulnois, G. J., Andrew, P. W. and Mitchell, T. J. 1994. A neuraminidase from Streptococcus pneumoniae has the features of a surface protein. *Infect Immun.* 62(9): 3688-3695.

Casado, B., Pannell, L. K., Iadarola, P. and Baraniuk, J. N. 2005. Identification of human nasal mucous proteins using proteomics. *Proteomics.* 5(11): 2949-2959.

Casal, J. A., Cano, E. and Tutor, J. C. 2005. Beta-hexosaminidase isoenzyme profiles in serum, plasma, platelets and mononuclear, polymorphonuclear and unfractionated total leukocytes. *Clin Biochem.* 38(10): 938-942.

Casal, J. A., Chabas, A. and Tutor, J. C. 2003. Thermodynamic determination of beta-hexosaminidase isoenzymes in mononuclear and polymorphonuclear leukocyte populations. *Am J Med Genet A.* 116A(3): 229-233.

Castric, P., Cassels, F. J. and Carlson, R. W. 2001. Structural characterization of the Pseudomonas aeruginosa 1244 pilin glycan. *J Biol Chem.* 276(28): 26479-26485.

CFFoundation. 2002. Patient registry 2001 annual report. *Cystic Fibrosis Foundation, Washington, D.C.*

CFFoundation. 2004. Cystic Fibrosis Foundation, Patient Registry 2004 Annual Report. *Bethesda, Maryland.*

CFFoundation. 2008. Cystic Fibrosis Foundation, Patient Registry 2008 Annual Report.

Charlwood, J., Birrell, H., Gribble, A., Burdes, V., Tolson, D. and Camilleri, P. 2000. A probe for the versatile analysis and characterization of N-linked oligosaccharides. *Anal Chem.* 72(7): 1453-1461.

Charro, N., Hood, B. L., Faria, D., Pacheco, P., Azevedo, P., Lopes, C., de Almeida, A. B., Couto, F. M., Conrads, T. P. and Penque, D. 2011. Serum proteomics signature of cystic fibrosis patients: a complementary 2-DE and LC-MS/MS approach. *J Proteomics.* 74(1): 110-126.

Chen, C. P., Song, S. C., Gilboa-Garber, N., Chang, K. S. and Wu, A. M. 1998. Studies on the binding site of the galactose-specific agglutinin PA-IL from Pseudomonas aeruginosa. *Glycobiology.* 8(1): 7-16.

Chen, H., Wang, D., Bai, C. and Wang, X. 2010. Proteomics-based biomarkers in chronic obstructive pulmonary disease. *J Proteome Res.* 9(6): 2798-2808.

Chen, X. and Flynn, G. C. 2007. Analysis of N-glycans from recombinant immunoglobulin G by on-line reversed-phase high-performance liquid chromatography/mass spectrometry. *Anal Biochem.* 370(2): 147-161.

Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., O'Riordan, C. R. and Smith, A. E. 1990. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell.* 63(4): 827-834.

Chiesa, C. and Horvath, C. 1993. Capillary zone electrophoresis of malto-oligosaccharides derivatized with 8-aminonaphthalene-1,3,6-trisulfonic acid. *J Chromatogr.* 645(2): 337-352.

Cimon, B., Carrere, J., Vinatier, J. F., Chazalette, J. P., Chabasse, D. and Bouchara, J. P. 2000. Clinical significance of Scedosporium apiospermum in patients with cystic fibrosis. *Eur J Clin Microbiol Infect Dis.* 19(1): 53-56.

Costerton, J. W., Stewart, P. S. and Greenberg, E. P. 1999. Bacterial biofilms: a common cause of persistent infections. *Science.* 284(5418): 1318-1322.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. and Mann, M. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res.* 10(4): 1794-1805.

Craig, J. M., Haddad, H. and Shwachman, H. 1957. The pathological changes in the liver in cystic fibrosis of the pancreas. *AMA J Dis Child.* 93(4): 357-369.

D'Souza, A. L., Rajkumar, C., Cooke, J. and Bulpitt, C. J. 2002. Probiotics in prevention of antibiotic associated diarrhoea: meta-analysis. *BMJ.* 324(7350): 1361.

Dam, S., Thaysen-Andersen, M., Stenkjaer, E., Lorentzen, A., Roepstorff, P., Packer, N. H. and Stougaard, J. 2013. Combined N-glycome and N-glycoproteome analysis of the Lotus japonicus seed globulin fraction shows conservation of protein structure and glycosylation in legumes. *J Proteome Res.* 12(7): 3383-3392.

Delaney, J. and Vouros, P. 2001. Liquid chromatography ion trap mass spectrometric analysis of oligosaccharides using permethylated derivatives. *Rapid Commun Mass Spectrom.* 15(5): 325-334.

Di Sant'Agnese, P. A. 1956. Fibrocystic disease of the pancreas, a generalized disease of exocrine glands. *J Am Med Assoc.* 160(10): 846-853.

Di, S. A. P. E. and Andersen, D. H. 1946. Celiac syndrome; chemotherapy in infections of the respiratory tract associated with cystic fibrosis of the pancreas; observations with penicillin and drugs of the sulfonamide group, with special reference to penicillin aerosol. *Am J Dis Child.* 72: 17-61.

Domann, P., Spencer, D. I. and Harvey, D. J. 2012. Production and fragmentation of negative ions from neutral N-linked carbohydrates ionized by matrix-assisted laser desorption/ionization. *Rapid Commun Mass Spectrom.* 26(4): 469-479.

Domon, B., and Costello, C. E. 1988. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates *Glycoconjugate Journal.* 5: 397-409.

Doring, G., Conway, S. P., Heijerman, H. G., Hodson, M. E., Hoiby, N., Smyth, A. and Touw, D. J. 2000. Antibiotic therapy against Pseudomonas aeruginosa in cystic fibrosis: a European consensus. *Eur Respir J.* 16(4): 749-767.

Doull, I. J. 2001. Recent advances in cystic fibrosis. *Arch Dis Child.* 85(1): 62-66.

Downey, D. G., Bell, S. C. and Elborn, J. S. 2009. Neutrophils in cystic fibrosis. *Thorax.* 64(1): 81-88.

Dri, P., Soranzo, M. R., Cramer, R., Menegazzi, R., Miotti, V. and Patriarca, P. 1985. Role of myeloperoxidase in respiratory burst of human polymorphonuclear leukocytes. Studies with myeloperoxidase-deficient subjects. *Inflammation.* 9(1): 21-31.

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(6): 895-906.

Everest-Dass, A. V., Jin, D., Thaysen-Andersen, M., Nevalainen, H., Kolarich, D. and Packer, N. H. 2012. Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by Candida albicans. *Glycobiology.* 22(11): 1465-1479.

Foxall, C., Watson, S. R., Dowbenko, D., Fennie, C., Lasky, L. A., Kiso, M., Hasegawa, A., Asa, D. and Brandley, B. K. 1992. The three members of the selectin receptor family recognize a common carbohydrate epitope, the sialyl Lewis(x) oligosaccharide. *J Cell Biol.* 117(4): 895-902.

Frederiksen, B., Koch, C. and Hoiby, N. 1997. Antibiotic treatment of initial colonization with Pseudomonas aeruginosa postpones chronic infection and prevents deterioration of pulmonary function in cystic fibrosis. *Pediatr Pulmonol.* 23(5): 330-335.

Fry, B. N., Korolik, V., ten Brinke, J. A., Pennings, M. T., Zalm, R., Teunis, B. J., Coloe, P. J. and van der Zeijst, B. A. 1998. The lipopolysaccharide biosynthesis locus of Campylobacter jejuni 81116. *Microbiology.* 144 (Pt 8): 2049-2061.

Fu, D., Chen, L. and O'Neill, R. A. 1994. A detailed structural characterization of ribonuclease B oligosaccharides by 1H NMR spectroscopy and mass spectrometry. *Carbohydr Res.* 261(2): 173-186.

Gal, M., Preston, G. M., Massey, R. C., Spiers, A. J. and Rainey, P. B. 2003. Genes encoding a cellulosic polymer contribute toward the ecological success of Pseudomonas fluorescens SBW25 on plant surfaces. *Mol Ecol.* 12(11): 3109-3121.

Ganz, T. 1999. Oxygen-independent microbicidal mechanisms of phagocytes. *Proc Assoc Am Physicians.* 111(5): 390-395.

Garcia, M. A., Yang, N. and Quinton, P. M. 2009. Normal mouse intestinal mucus release requires cystic fibrosis transmembrane regulator-dependent bicarbonate secretion. *J Clin Invest.* 119(9): 2613-2622.

Gaut, J. P., Yeh, G. C., Tran, H. D., Byun, J., Henderson, J. P., Richter, G. M., Brennan, M. L., Lusis, A. J., Belaaouaj, A., Hotchkiss, R. S. and Heinecke, J. W. 2001. Neutrophils employ the myeloperoxidase system to generate antimicrobial brominating and chlorinating oxidants during sepsis. *Proc Natl Acad Sci U S A.* 98(21): 11961-11966.

Gjermansen, M., Nilsson, M., Yang, L. and Tolker-Nielsen, T. 2010. Characterization of starvation-induced dispersion in Pseudomonas putida biofilms: genetic elements and molecular mechanisms. *Mol Microbiol.* 75(4): 815-826.

Gjermansen, M., Ragas, P., Sternberg, C., Molin, S. and Tolker-Nielsen, T. 2005. Characterization of starvation-induced dispersion in Pseudomonas putida biofilms. *Environ Microbiol.* 7(6): 894-906.

Goerke, C., Kraning, K., Stern, M., Doring, G., Botzenhart, K. and Wolz, C. 2000. Molecular epidemiology of community-acquired Staphylococcus aureus in families with and without cystic fibrosis patients. *J Infect Dis.* 181(3): 984-989.

Gornik, O. and Lauc, G. 2008. Glycosylation of serum proteins in inflammatory diseases. *Dis Markers.* 25(4-5): 267-278.

Goss, C. H., Mayer-Hamblett, N., Aitken, M. L., Rubenfeld, G. D. and Ramsey, B. W. 2004. Association between Stenotrophomonas maltophilia and lung function in cystic fibrosis. *Thorax.* 59(11): 955-959.

Goss, C. H., Otto, K., Aitken, M. L. and Rubenfeld, G. D. 2002. Detecting Stenotrophomonas maltophilia does not reduce survival of patients with cystic fibrosis. *Am J Respir Crit Care Med.* 166(3): 356-361.

Gray, R. D., MacGregor, G., Noble, D., Imrie, M., Dewar, M., Boyd, A. C., Innes, J. A., Porteous, D. J. and Greening, A. P. 2008. Sputum proteomics in inflammatory and suppurative respiratory diseases. *Am J Respir Crit Care Med.* 178(5): 444-452.

Greally, P., Hussein, M. J., Cook, A. J., Sampson, A. P., Piper, P. J. and Price, J. F. 1993. Sputum tumour necrosis factor-alpha and leukotriene concentrations in cystic fibrosis. *Arch Dis Child.* 68(3): 389-392.

Green, E. D., Adelt, G., Baenziger, J. U., Wilson, S. and Van Halbeek, H. 1988. The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy. *J Biol Chem.* 263(34): 18253-18268.

Griese, M., von Bredow, C. and Birrer, P. 2001. Reduced proteolysis of surfactant protein A and changes of the bronchoalveolar lavage fluid proteome by inhaled alpha 1-protease inhibitor in cystic fibrosis. *Electrophoresis.* 22(1): 165-171.

Guignard, C., Jouve, L., Bogeat-Triboulot, M. B., Dreyer, E., Hausman, J. F. and Hoffmann, L. 2005. Analysis of carbohydrates in plants by high-performance anion-exchange chromatography coupled with electrospray mass spectrometry. *J Chromatogr A.* 1085(1): 137-142.

Gustafsson, J. K., Ermund, A., Ambort, D., Johansson, M. E., Nilsson, H. E., Thorell, K., Hebert, H., Sjovall, H. and Hansson, G. C. 2012. Bicarbonate and functional CFTR channel are required for proper mucin secretion and link cystic fibrosis with its mucus phenotype. *J Exp Med.* 209(7): 1263-1272.

Han, L. and Costello, C. E. 2013. Mass spectrometry of glycans. *Biochemistry (Mosc).* 78(7): 710-720.

Hao, Z., Xiao, B. and Weng, N. 2008. Impact of column temperature and mobile phase components on selectivity of hydrophilic interaction chromatography (HILIC). *J Sep Sci.* 31(9): 1449-1464.

Harrison, F. 2007. Microbial ecology of the cystic fibrosis lung. *Microbiology.* 153(Pt 4): 917-923.

Harun, A., Perdomo, H., Gilgado, F., Chen, S. C., Cano, J., Guarro, J. and Meyer, W. 2009. Genotyping of Scedosporium species: a review of molecular approaches. *Med Mycol.* 47(4): 406-414.

Harun, A., Serena, C., Gilgado, F., Chen, S. C. and Meyer, W. 2010. Scedosporium aurantiacum is as virulent as S. prolificans, and shows strain-specific virulence differences, in a mouse model. *Med Mycol.* 48 Suppl 1: S45-51.

Henderson, A. G., Ehre, C., Button, B., Abdullah, L. H., Cai, L. H., Leigh, M. W., DeMaria, G. C., Matsui, H., Donaldson, S. H., Davis, C. W., Sheehan, J. K., Boucher, R. C. and Kesimer, M. 2014. Cystic fibrosis airway secretions exhibit mucin hyperconcentration and increased osmotic pressure. *J Clin Invest.* 124(7): 3047-3060.

Henke, M. O., John, G., Rheineck, C., Chillappagari, S., Naehrlich, L. and Rubin, B. K. 2011. Serine proteases degrade airway mucins in cystic fibrosis. *Infect Immun.* 79(8): 3438-3444.

Henke, M. O., Renner, A., Huber, R. M., Seeds, M. C. and Rubin, B. K. 2004. MUC5AC and MUC5B Mucins Are Decreased in Cystic Fibrosis Airway Secretions. *Am J Respir Cell Mol Biol.* 31(1): 86-91.

Hirche, T. O., Gaut, J. P., Heinecke, J. W. and Belaaouaj, A. 2005. Myeloperoxidase plays critical roles in killing Klebsiella pneumoniae and inactivating neutrophil elastase: effects on host defense. *J Immunol.* 174(3): 1557-1565.

Hitchen, P. G. and Dell, A. 2006. Bacterial glycoproteomics. *Microbiology.* 152(Pt 6): 1575-1580.

Hoiby, N. 1974. Epidemiological investigations of the respiratory tract bacteriology in patients with cystic fibrosis. *Acta Pathol Microbiol Scand B Microbiol Immunol.* 82(4): 541-550.

Hoiby, N. 1992. Prevention and treatment of infections in cystic fibrosis. *Int J Antimicrob Agents.* 1(5-6): 229-237.

Horsley, A., Rousseau, K., Ridley, C., Flight, W., Jones, A., Waigh, T. A. and Thornton, D. J. 2014. Reassessment of the importance of mucins in determining sputum properties in cystic fibrosis. *J Cyst Fibros.* 13(3): 260-266.

Huang, C. M. 2004. Comparative proteomic analysis of human whole saliva. *Arch Oral Biol.* 49(12): 951-962.

Imundo, L., Barasch, J., Prince, A. and Al-Awqati, Q. 1995. Cystic fibrosis epithelial cells have a receptor for pathogenic bacteria on their apical surface. *Proc Natl Acad Sci U S A.* 92(7): 3019-3023.

Jaovisidha, P., Peeples, M. E., Brees, A. A., Carpenter, L. R. and Moy, J. N. 1999. Respiratory syncytial virus stimulates neutrophil degranulation and chemokine release. *J Immunol.* 163(5): 2816-2820.

Jatakanon, A., Lalloo, U. G., Lim, S., Chung, K. F. and Barnes, P. J. 1999. Increased neutrophils and cytokines, TNF-alpha and IL-8, in induced sputum of non-asthmatic patients with chronic dry cough. *Thorax.* 54(3): 234-237.

Jatakanon, A., Uasuf, C., Maziak, W., Lim, S., Chung, K. F. and Barnes, P. J. 1999. Neutrophilic inflammation in severe persistent asthma. *Am J Respir Crit Care Med.* 160(5 Pt 1): 1532-1539.

Jin, C., Ekwall, A. K., Bylund, J., Bjorkman, L., Estrella, R. P., Whitelock, J. M., Eisler, T., Bokarewa, M. and Karlsson, N. G. 2012. Human synovial lubricin expresses sialyl Lewis x determinant and has L-selectin ligand activity. *J Biol Chem.* 287(43): 35922-35933.

Johnson, D. A., Tetu, S. G., Phillippy, K., Chen, J., Ren, Q. and Paulsen, I. T. 2008. High-throughput phenotypic characterization of Pseudomonas aeruginosa membrane transport genes. *PLoS Genet.* 4(10): e1000211.

Kaplan, M. J. and Radic, M. 2012. Neutrophil extracellular traps: double-edged swords of innate immunity. *J Immunol.* 189(6): 2689-2695.

Karlsson, N. G., Wilson, N. L., Wirth, H. J., Dawes, P., Joshi, H. and Packer, N. H. 2004. Negative ion graphitised carbon nano-liquid chromatography/mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. *Rapid Commun Mass Spectrom.* 18(19): 2282-2292.

Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M. and Tsui, L. C. 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science.* 245(4922): 1073-1080.

Khan, T. Z., Wagener, J. S., Bost, T., Martinez, J., Accurso, F. J. and Riches, D. W. 1995. Early pulmonary inflammation in infants with cystic fibrosis. *Am J Respir Crit Care Med.* 151(4): 1075-1082.

Kilpatrick, D. C. 2002. Mannan-binding lectin and its role in innate immunity. *Transfus Med.* 12(6): 335-352.

Kilpatrick, D. C. 2002. Mannan-binding lectin: clinical significance and applications. *Biochim Biophys Acta.* 1572(2-3): 401-413.

King, S. J., Hippe, K. R., Gould, J. M., Bae, D., Peterson, S., Cline, R. T., Fasching, C., Janoff, E. N. and Weiser, J. N. 2004. Phase variable desialylation of host proteins that bind to Streptococcus pneumoniae in vivo and protect the airway. *Mol Microbiol.* 54(1): 159-171.

Kipnis, E., Hansen, K., Sawa, T., Moriyama, K., Zurawel, A., Ishizaka, A. and Wiener-Kronish, J. 2008. Proteomic analysis of undiluted lung epithelial lining fluid. *Chest.* 134(2): 338-345.

Klausen, M., Heydorn, A., Ragas, P., Lambertsen, L., Aaes-Jorgensen, A., Molin, S. and Tolker-Nielsen, T. 2003. Biofilm formation by Pseudomonas aeruginosa wild type, flagella and type IV pili mutants. *Mol Microbiol.* 48(6): 1511-1524.

Knutsen, A. and Slavin, R. G. 1992. Allergic bronchopulmonary mycosis complicating cystic fibrosis. *Semin Respir Infect.* 7(3): 179-192.

Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. and Altmann, F. 2006. Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics.* 6(11): 3369-3380.

Koller, D. Y., Nething, I., Otto, J., Urbanek, R. and Eichler, I. 1997. Cytokine concentrations in sputum from patients with cystic fibrosis and their relation to eosinophil activity. *Am J Respir Crit Care Med.* 155(3): 1050-1054.

Korekane, H., Korekane, A., Yamaguchi, Y., Kato, M., Miyamoto, Y., Matsumoto, A., Hasegawa, T., Suzuki, K., Taniguchi, N. and Ookawara, T. 2011. N-Glycosylation

profiling of recombinant mouse extracellular superoxide dismutase produced in Chinese hamster ovary cells. *Glycoconj J.* 28(3-4): 183-196.

Korgaonkar, A. K. and Whiteley, M. 2011. Pseudomonas aeruginosa enhances production of an antimicrobial in response to N-acetylglucosamine and peptidoglycan. *J Bacteriol.* 193(4): 909-917.

Kronborg, G., Hansen, M. B., Svenson, M., Fomsgaard, A., Hoiby, N. and Bendtzen, K. 1993. Cytokines in sputum and serum from patients with cystic fibrosis and chronic Pseudomonas aeruginosa infection as markers of destructive inflammation in the lungs. *Pediatr Pulmonol.* 15(5): 292-297.

Kronis, K. A. and Carver, J. P. 1985. Wheat germ agglutinin dimers bind sialyloligosaccharides at four sites in solution: proton nuclear magnetic resonance temperature studies at 360 MHz. *Biochemistry.* 24(4): 826-833.

Lambiase, A., Raia, V., Del Pezzo, M., Sepe, A., Carnovale, V. and Rossano, F. 2006. Microbiology of airway disease in a cohort of patients with cystic fibrosis. *BMC Infect Dis.* 6: 4.

Lang, B. J., Aaron, S. D., Ferris, W., Hebert, P. C. and MacDonald, N. E. 2000. Multiple combination bactericidal antibiotic testing for patients with cystic fibrosis infected with multiresistant strains of Pseudomonas aeruginosa. *Am J Respir Crit Care Med.* 162(6): 2241-2245.

Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P. and Jorgensen, T. J. 2005. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics.* 4(7): 873-886.

Laue, H., Schenk, A., Li, H., Lambertsen, L., Neu, T. R., Molin, S. and Ullrich, M. S. 2006. Contribution of alginate and levan production to biofilm formation by Pseudomonas syringae. *Microbiology.* 152(Pt 10): 2909-2918.

Lee, K. B., Desai, U. R., Palcic, M. M., Hindsgaul, O. and Linhardt, R. J. 1992. An electrophoresis-based assay for glycosyltransferase activity. *Anal Biochem.* 205(1): 108-114.

Lillehoj, E. P., Kim, B. T. and Kim, K. C. 2002. Identification of Pseudomonas aeruginosa flagellin as an adhesin for Muc1 mucin. *Am J Physiol Lung Cell Mol Physiol.* 282(4): L751-756.

LiPuma, J. J., Dasen, S. E., Nielson, D. W., Stern, R. C. and Stull, T. L. 1990. Person-to-person transmission of Pseudomonas cepacia between patients with cystic fibrosis. *Lancet.* 336(8723): 1094-1096.

Lominadze, G., Powell, D. W., Luerman, G. C., Link, A. J., Ward, R. A. and McLeish, K. R. 2005. Proteomic analysis of human neutrophil granules. *Mol Cell Proteomics.* 4(10): 1503-1521.

Luria, S. E. and Burrous, J. W. 1957. Hybridization between Escherichia coli and Shigella. *J Bacteriol.* 74(4): 461-476.

Lyczak, J. B., Cannon, C. L. and Pier, G. B. 2002. Lung infections associated with cystic fibrosis. *Clin Microbiol Rev.* 15(2): 194-222.

MacGregor, G., Gray, R. D., Hilliard, T. N., Imrie, M., Boyd, A. C., Alton, E. W., Bush, A., Davies, J. C., Innes, J. A., Porteous, D. J. and Greening, A. P. 2008. Biomarkers for cystic fibrosis lung disease: application of SELDI-TOF mass spectrometry to BAL fluid. *J Cyst Fibros.* 7(5): 352-358.

Mackie, A. M., Hassan, K. A., Paulsen, I. T. and Tetu, S. G. 2014. Biolog Phenotype MicroArrays for Phenotypic Characterization of Microbial Cells. *Methods Mol Biol.* 1096: 123-130.

Maley, F., Trimble, R. B., Tarentino, A. L. and Plummer, T. H., Jr. 1989. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal Biochem.* 180(2): 195-204.

Marcobal, A., Barboza, M., Froehlich, J. W., Block, D. E., German, J. B., Lebrilla, C. B. and Mills, D. A. 2010. Consumption of human milk oligosaccharides by gut-related microbes. *J Agric Food Chem.* 58(9): 5334-5340.

Marek, K. W., Vijay, I. K. and Marth, J. D. 1999. A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality. *Glycobiology.* 9(11): 1263-1271.

Marotte, K., Preville, C., Sabin, C., Moume-Pymbock, M., Imberty, A. and Roy, R. 2007. Synthesis and binding properties of divalent and trivalent clusters of the Lewis a disaccharide moiety to Pseudomonas aeruginosa lectin PA-IIL. *Org Biomol Chem.* 5(18): 2953-2961.

Massie, J., Forbes, R., Dusart, D., Bankier, A. and Delatycki, M. B. 2007. Community-wide screening for cystic fibrosis carriers could replace newborn screening for the diagnosis of cystic fibrosis. *J Paediatr Child Health.* 43(11): 721-723.

Massie, R. J., Olsen, M., Glazner, J., Robertson, C. F. and Francis, I. 2000. Newborn screening for cystic fibrosis in Victoria: 10 years' experience (1989-1998). *Med J Aust.* 172(12): 584-587.

Matsukawa, M. and Greenberg, E. P. 2004. Putative exopolysaccharide synthesis genes influence Pseudomonas aeruginosa biofilm development. *J Bacteriol.* 186(14): 4449-4456.

Matthews, L. W., Spector, S., Lemm, J. and Potter, J. L. 1963. Studies on Pulmonary Secretions. I. The over-All Chemical Composition of Pulmonary Secretions from Patients with Cystic Fibrosis, Bronchiectasis, and Laryngectomy. *Am Rev Respir Dis.* 88: 199-204.

Mayer-Scholl, A., Averhoff, P. and Zychlinsky, A. 2004. How do neutrophils and pathogens interact? *Curr Opin Microbiol.* 7(1): 62-66.

Middleton, P. G., Chen, S. C. and Meyer, W. 2013. Fungal infections and treatment in cystic fibrosis. *Curr Opin Pulm Med.* 19(6): 670-675.

Mikkelsen, H., McMullan, R. and Filloux, A. 2011. The Pseudomonas aeruginosa reference strain PA14 displays increased virulence due to a mutation in ladS. *PLoS One.* 6(12): e29113.

Millar, M. R., Linton, C. J., Cade, A., Glancy, D., Hall, M. and Jalal, H. 1996. Application of 16S rRNA gene PCR to study bowel flora of preterm infants with and without necrotizing enterocolitis. *J Clin Microbiol.* 34(10): 2506-2510.

Miller, M. B. and Gilligan, P. H. 2003. Laboratory aspects of management of chronic pulmonary infections in patients with cystic fibrosis. *J Clin Microbiol.* 41(9): 4009-4015.

Mortaz, E., Adcock, I. M., Folkerts, G., Barnes, P. J., Paul Vos, A. and Garssen, J. 2013. Probiotics in the management of lung diseases. *Mediators Inflamm.* 2013: 751068.

Naseem, S., Parrino, S. M., Buenten, D. M. and Konopka, J. B. 2012. Novel roles for GlcNAc in cell signaling. *Commun Integr Biol.* 5(2): 156-159.

Nauseef, W. M. 2001. Contributions of myeloperoxidase to proinflammatory events: more than an antimicrobial system. *Int J Hematol.* 74(2): 125-133.

Nauseef, W. M., Metcalf, J. A. and Root, R. K. 1983. Role of myeloperoxidase in the respiratory burst of human neutrophils. *Blood.* 61(3): 483-492.

Nazaret, S., Assade, F., Brothier, E., Freydiere, A. M., Bellon, G. and Cournoyer, B. 2009. RISA-HPLC analysis of lung bacterial colonizers of cystic fibrosis children. *J Microbiol Methods.* 76(1): 58-69.

Nicholas, B. and Djukanovic, R. 2009. Induced sputum: a window to lung pathology. *Biochem Soc Trans.* 37(Pt 4): 868-872.

Nicholas, B., Skipp, P., Mould, R., Rennard, S., Davies, D. E., O'Connor, C. D. and Djukanovic, R. 2006. Shotgun proteomic analysis of human-induced sputum. *Proteomics.* 6(15): 4390-4401.

Nicholls, J. M., Chan, R. W., Russell, R. J., Air, G. M. and Peiris, J. S. 2008. Evolving complexities of influenza virus and its receptors. *Trends Microbiol.* 16(4): 149-157.

Noah, T. L., Black, H. R., Cheng, P. W., Wood, R. E. and Leigh, M. W. 1997. Nasal and bronchoalveolar lavage fluid cytokines in early cystic fibrosis. *J Infect Dis.* 175(3): 638-647.

Noel-Georis, I., Bernard, A., Falmagne, P. and Wattiez, R. 2002. Database of bronchoalveolar lavage fluid proteins. *J Chromatogr B Analyt Technol Biomed Life Sci.* 771(1-2): 221-236.

Nunley, D. R., Grgurich, W., Iacono, A. T., Yousem, S., Ohori, N. P., Keenan, R. J. and Dauber, J. H. 1998. Allograft colonization and infections with pseudomonas in cystic fibrosis lung transplant recipients. *Chest.* 113(5): 1235-1243.

Obermayer, A., Stoiber, W., Krautgartner, W. D., Klappacher, M., Kofler, B., Steinbacher, P., Vitkov, L., Grabcanovic-Musija, F. and Studnicka, M. 2014. New aspects on the structure of neutrophil extracellular traps from chronic obstructive pulmonary disease and in vitro generation. *PLoS One.* 9(5): e97784.

Ohlmeier, S., Mazur, W., Linja-Aho, A., Louhelainen, N., Ronty, M., Toljamo, T., Bergmann, U. and Kinnula, V. L. 2012. Sputum proteomics identifies elevated PIGR levels in smokers and mild-to-moderate COPD. *J Proteome Res.* 11(2): 599-608.

Ohman, D. E. and Chakrabarty, A. M. 1982. Utilization of human respiratory secretions by mucoid Pseudomonas aeruginosa of cystic fibrosis origin. *Infect Immun.* 37(2): 662-669.

Osika, E., Cavaillon, J. M., Chadelat, K., Boule, M., Fitting, C., Tournier, G. and Clement, A. 1999. Distinct sputum cytokine profiles in cystic fibrosis and other chronic inflammatory airway disease. *Eur Respir J.* 14(2): 339-346.

Osman, S. F., Fett, W. F. and Fishman, M. L. 1986. Exopolysaccharides of the phytopathogen Pseudomonas syringae pv. glycinea. *J Bacteriol.* 166(1): 66-71.

Palmer, K. L., Mashburn, L. M., Singh, P. K. and Whiteley, M. 2005. Cystic fibrosis sputum supports growth and cues key aspects of Pseudomonas aeruginosa physiology. *J Bacteriol.* 187(15): 5267-5277.

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(12): 5791-5800.

Peleman, R. A., Rytila, P. H., Kips, J. C., Joos, G. F. and Pauwels, R. A. 1999. The cellular composition of induced sputum in chronic obstructive pulmonary disease. *Eur Respir J.* 13(4): 839-843.

Pennybacker, M., Liessem, B., Moczall, H., Tifft, C. J., Sandhoff, K. and Proia, R. L. 1996. Identification of domains in human beta-hexosaminidase that determine substrate specificity. *J Biol Chem.* 271(29): 17377-17382.

Perez, A., Issler, A. C., Cotton, C. U., Kelley, T. J., Verkman, A. S. and Davis, P. B. 2007. CFTR inhibition mimics the cystic fibrosis inflammatory profile. *Am J Physiol Lung Cell Mol Physiol.* 292(2): L383-395.

Petersen, N. T., Hoiby, N., Mordhorst, C. H., Lind, K., Flensborg, E. W. and Bruun, B. 1981. Respiratory infections in cystic fibrosis patients caused by virus, chlamydia and mycoplasma--possible synergism with Pseudomonas aeruginosa. *Acta Paediatr Scand.* 70(5): 623-628.

Pilewski, J. M. and Frizzell, R. A. 1999. Role of CFTR in airway disease. *Physiol Rev.* 79(1 Suppl): S215-255.

Pollard, H. B., Ji, X. D., Jozwik, C. and Jacobowitz, D. M. 2005. High abundance protein profiling of cystic fibrosis lung epithelial cells. *Proteomics.* 5(8): 2210-2226.

Pryles, C. V. 1958. Staphylococcal pneumonia in infancy and childhood; an analysis of 24 cases. *Pediatrics.* 21(4): 609-623.

Quinton, P. M. 2001. The neglected ion: HCO3. *Nat Med.* 7(3): 292-293.

Ravnsborg, T., Houen, G. and Hojrup, P. 2010. The glycosylation of myeloperoxidase. *Biochim Biophys Acta.* 1804(10): 2046-2053.

Rennie, R. P., Nord, C. E., Sjoberg, L. and Duncan, I. B. 1978. Comparison of bacteriophage typing, serotyping, and biotyping as aids in epidemiological surveillance of Klebsiella infections. *J Clin Microbiol.* 8(6): 638-642.

Retamales, I., Elliott, W. M., Meshi, B., Coxson, H. O., Pare, P. D., Sciurba, F. C., Rogers, R. M., Hayashi, S. and Hogg, J. C. 2001. Amplification of inflammation in emphysema and its association with latent adenoviral infection. *Am J Respir Crit Care Med.* 164(3): 469-473.

Richman-Eisenstat, J. B., Jorens, P. G., Hebert, C. A., Ueki, I. and Nadel, J. A. 1993. Interleukin-8: an important chemoattractant in sputum of patients with chronic inflammatory airway diseases. *Am J Physiol.* 264(4 Pt 1): L413-418.

Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L. and et al. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science.* 245(4922): 1066-1073.

Rogers, G. B., Carroll, M. P., Serisier, D. J., Hockey, P. M., Jones, G. and Bruce, K. D. 2004. characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16s ribosomal DNA terminal restriction fragment length polymorphism profiling. *J Clin Microbiol.* 42(11): 5176-5183.

Roque-Barreira, M. C. and Campos-Neto, A. 1985. Jacalin: an IgA-binding lectin. *J Immunol.* 134(3): 1740-1743.

Rorvig, S., Ostergaard, O., Heegaard, N. H. and Borregaard, N. 2013. Proteome profiling of human neutrophil granule subsets, secretory vesicles, and cell membrane: correlation with transcriptome profiling of neutrophil precursors. *J Leukoc Biol.* 94(4): 711-721.

Rosenfeld, M., Gibson, R. L., McNamara, S., Emerson, J., Burns, J. L., Castile, R., Hiatt, P., McCoy, K., Wilson, C. B., Inglis, A., Smith, A., Martin, T. R. and Ramsey, B. W. 2001. Early pulmonary infection, inflammation, and clinical outcomes in infants with cystic fibrosis. *Pediatr Pulmonol.* 32(5): 356-366.

Roy, M. G., Livraghi-Butrico, A., Fletcher, A. A., McElwee, M. M., Evans, S. E., Boerner, R. M., Alexander, S. N., Bellinghausen, L. K., Song, A. S., Petrova, Y. M., Tuvim, M. J., Adachi, R., Romo, I., Bordt, A. S., Bowden, M. G., Sisson, J. H., Woodruff, P. G., Thornton, D. J., Rousseau, K., De la Garza, M. M., Moghaddam, S. J., Karmouty-Quintana, H., Blackburn, M. R., Drouin, S. M., Davis, C. W., Terrell, K. A., Grubb, B.

R., O'Neal, W. K., Flores, S. C., Cota-Gomez, A., Lozupone, C. A., Donnelly, J. M., Watson, A. M., Hennessy, C. E., Keith, R. C., Yang, I. V., Barthel, L., Henson, P. M., Janssen, W. J., Schwartz, D. A., Boucher, R. C., Dickey, B. F. and Evans, C. M. 2014. Muc5b is required for airway defence. *Nature.* 505(7483): 412-416.

Ruhaak, L. R., Deelder, A. M. and Wuhrer, M. 2009. Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry. *Anal Bioanal Chem.* 394(1): 163-174.

Sabin, C., Mitchell, E. P., Pokorna, M., Gautier, C., Utille, J. P., Wimmerova, M. and Imberty, A. 2006. Binding of different monosaccharides by lectin PA-IIL from Pseudomonas aeruginosa: thermodynamics data correlated with X-ray structures. *FEBS Lett.* 580(3): 982-987.

Sagel, S. D., Kapsner, R., Osberg, I., Sontag, M. K. and Accurso, F. J. 2001. Airway inflammation in children with cystic fibrosis and healthy children assessed by sputum induction. *Am J Respir Crit Care Med.* 164(8 Pt 1): 1425-1431.

Saginur, R., Stdenis, M., Ferris, W., Aaron, S. D., Chan, F., Lee, C. and Ramotar, K. 2006. Multiple combination bactericidal testing of staphylococcal biofilms from implant-associated infections. *Antimicrob Agents Chemother.* 50(1): 55-61.

Sanchez, L., Calvo, M. and Brock, J. H. 1992. Biological role of lactoferrin. *Arch Dis Child.* 67(5): 657-661.

Satterfield, M. B. and Welch, M. J. 2005. Comparison by LC-MS and MALDI-MS of prostate-specific antigen from five commercial sources with certified reference material 613. *Clin Biochem.* 38(2): 166-174.

Schachter, H. 2009. Paucimannose N-glycans in Caenorhabditis elegans and Drosophila melanogaster. *Carbohydr Res.* 344(12): 1391-1396.

Schulz, B. L., Sloane, A. J., Robinson, L. J., Prasad, S. S., Lindner, R. A., Robinson, M., Bye, P. T., Nielson, D. W., Harry, J. L., Packer, N. H. and Karlsson, N. G. 2007. Glycosylation of sputum mucins is altered in cystic fibrosis patients. *Glycobiology.* 17(7): 698-712.

Schwab, U. E., Wold, A. E., Carson, J. L., Leigh, M. W., Cheng, P. W., Gilligan, P. H. and Boat, T. F. 1993. Increased adherence of Staphylococcus aureus from cystic fibrosis lungs to airway epithelial cells. *Am Rev Respir Dis.* 148(2): 365-369.

Schwarz, F. and Aebi, M. 2011. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol.* 21(5): 576-582.

Scott, J. K., Loganathan, D., Easley, R. B., Gong, X. and Goldstein, I. J. 1992. A family of concanavalin A-binding peptides from a hexapeptide epitope library. *Proc Natl Acad Sci U S A.* 89(12): 5398-5402.

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(2): M000031-MCP000201.

Sela, D. A., Garrido, D., Lerno, L., Wu, S., Tan, K., Eom, H. J., Joachimiak, A., Lebrilla, C. B. and Mills, D. A. 2012. Bifidobacterium longum subsp. infantis ATCC 15697 alpha-fucosidases are active on fucosylated human milk oligosaccharides. *Appl Environ Microbiol.* 78(3): 795-803.

Sela, D. A., Li, Y., Lerno, L., Wu, S., Marcobal, A. M., German, J. B., Chen, X., Lebrilla, C. B. and Mills, D. A. 2011. An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides. *J Biol Chem.* 286(14): 11909-11918.

Sethi, M. K., Thaysen-Andersen, M., Smith, J. T., Baker, M. S., Packer, N. H., Hancock, W. S. and Fanayan, S. 2014. Comparative N-glycan profiling of colorectal cancer cell lines reveals unique bisecting GlcNAc and alpha-2,3-linked sialic acid determinants are associated with membrane proteins of the more metastatic/aggressive cell lines. *J Proteome Res.* 13(1): 277-288.

Shi, H., Tan, J. and Schachter, H. 2006. N-glycans are involved in the response of Caenorhabditis elegans to bacterial pathogens. *Methods Enzymol.* 417: 359-389.

Shoseyov, D., Brownlee, K. G., Conway, S. P. and Kerem, E. 2006. Aspergillus bronchitis in cystic fibrosis. *Chest.* 130(1): 222-226.

Shrout, J. D., Chopp, D. L., Just, C. L., Hentzer, M., Givskov, M. and Parsek, M. R. 2006. The impact of quorum sensing and swarming motility on Pseudomonas aeruginosa biofilm formation is nutritionally conditional. *Mol Microbiol.* 62(5): 1264-1277.

Singh, P. K., Schaefer, A. L., Parsek, M. R., Moninger, T. O., Welsh, M. J. and Greenberg, E. P. 2000. Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature.* 407(6805): 762-764.

Sloane, A. J., Lindner, R. A., Prasad, S. S., Sebastian, L. T., Pedersen, S. K., Robinson, M., Bye, P. T., Nielson, D. W. and Harry, J. L. 2005. Proteomic analysis of sputum from adults and children with cystic fibrosis and from control subjects. *Am J Respir Crit Care Med.* 172(11): 1416-1426.

Snelling, C. E. and Erb, I. H. 1942. Cystic fibrosis of the pancreas. *Arch Dis Child.* 17(92): 220-226.

Son, M. S., Matthews, W. J., Jr., Kang, Y., Nguyen, D. T. and Hoang, T. T. 2007. In vivo evidence of Pseudomonas aeruginosa nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients. *Infect Immun.* 75(11): 5313-5324.

Soong, G., Muir, A., Gomez, M. I., Waks, J., Reddy, B., Planet, P., Singh, P. K., Kaneko, Y., Wolfgang, M. C., Hsiao, Y. S., Tong, L. and Prince, A. 2006. Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production. *J Clin Invest.* 116(8): 2297-2305.

Stanley, P., Schachter, H. and Taniguchi, N. (2009). N-Glycans. Essentials of Glycobiology. C. R. Varki A, Esko JD, et al., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press;.

Stenson, W. F. and Parker, C. W. 1980. Monohydroxyeicosatetraenoic acids (HETEs) induce degranulation of human neutrophils. *J Immunol.* 124(5): 2100-2104.

Stevens, D. A., Moss, R. B., Kurup, V. P., Knutsen, A. P., Greenberger, P., Judson, M. A., Denning, D. W., Crameri, R., Brody, A. S., Light, M., Skov, M., Maish, W. and Mastella, G. 2003. Allergic bronchopulmonary aspergillosis in cystic fibrosis--state of the art: Cystic Fibrosis Foundation Consensus Conference. *Clin Infect Dis.* 37 Suppl 3: S225-264.

Stutman, H. R. and Marks, M. I. 1987. Pulmonary infections in children with cystic fibrosis. *Semin Respir Infect.* 2(3): 166-176.

Sumer-Bayraktar, Z., Kolarich, D., Campbell, M. P., Ali, S., Packer, N. H. and Thaysen-Andersen, M. 2011. N-glycans modulate the function of human corticosteroid-binding globulin. *Mol Cell Proteomics.* 10(8): M111 009100.

Suzuki, S., Tanaka, R., Takada, K., Inoue, N., Yashima, Y., Honda, A. and Honda, S. 2001. Analysis of sialo-N-glycans in glycoproteins as 1-phenyl-3-methyl-5-pyrazolone derivatives by capillary electrophoresis. *J Chromatogr A.* 910(2): 319-329.

Szymanski, C. M., Burr, D. H. and Guerry, P. 2002. Campylobacter protein glycosylation affects host cell interactions. *Infect Immun.* 70(4): 2242-2244.

Szymanski, C. M. and Wren, B. W. 2005. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol.* 3(3): 225-237.

Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. and Guerry, P. 1999. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol.* 32(5): 1022-1030.

Takahashi, N., Ishii, I., Ishihara, H., Mori, M., Tejima, S., Jefferis, R., Endo, S. and Arata, Y. 1987. Comparative structural study of the N-linked oligosaccharides of human normal and pathological immunoglobulin G. *Biochemistry.* 26(4): 1137-1144.

Thomsson, K. A., Karlsson, N. G. and Hansson, G. C. 1999. Liquid chromatography-electrospray mass spectrometry as a tool for the analysis of sulfated oligosaccharides from mucin glycoproteins. *J Chromatogr A.* 854(1-2): 131-139.

Tretter, V., Altmann, F. and Marz, L. 1991. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. *Eur J Biochem.* 199(3): 647-652.

Trivier, D., Houdret, N., Courcol, R. J., Lamblin, G., Roussel, P. and Davril, M. 1997. The binding of surface proteins from Staphylococcus aureus to human bronchial mucins. *Eur Respir J.* 10(4): 804-810.

Turner, G. A. 1992. N-glycosylation of serum proteins in disease and its investigation using lectins. *Clin Chim Acta.* 208(3): 149-171.

Urban, C. F., Reichard, U., Brinkmann, V. and Zychlinsky, A. 2006. Neutrophil extracellular traps capture and kill Candida albicans yeast and hyphal forms. *Cell Microbiol.* 8(4): 668-676.

Valerius, N. H., Koch, C. and Hoiby, N. 1991. Prevention of chronic Pseudomonas aeruginosa colonisation in cystic fibrosis by early treatment. *Lancet.* 338(8769): 725-726.

van Belkum, A., Renders, N. H., Smith, S., Overbeek, S. E. and Verbrugh, H. A. 2000. Comparison of conventional and molecular methods for the detection of bacterial pathogens in sputum samples from cystic fibrosis patients. *FEMS Immunol Med Microbiol.* 27(1): 51-57.

Venkatakrishnan, V., Packer, N. H. and Thaysen-Andersen, M. 2013. Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis. *Expert Rev Respir Med.* 7(5): 553-576.

Vimr, E. and Lichtensteiger, C. 2002. To sialylate, or not to sialylate: that is the question. *Trends in microbiology.* 10(6): 254-257.

Voisin, S., Kus, J. V., Houliston, S., St-Michael, F., Watson, D., Cvitkovitch, D. G., Kelly, J., Brisson, J. R. and Burrows, L. L. 2007. Glycosylation of Pseudomonas aeruginosa strain Pa5196 type IV pilins with mycobacterium-like alpha-1,5-linked d-Araf oligosaccharides. *J Bacteriol.* 189(1): 151-159.

von Bredow, C., Birrer, P. and Griese, M. 2001. Surfactant protein A and other bronchoalveolar lavage fluid proteins are altered in cystic fibrosis. *Eur Respir J.* 17(4): 716-722.

Wahab, A. A., Janahi, I. A., Marafia, M. M. and El-Shafie, S. 2004. Microbiological identification in cystic fibrosis patients with CFTR I1234V mutation. *J Trop Pediatr.* 50(4): 229-233.

Wark, P. A., Johnston, S. L., Moric, I., Simpson, J. L., Hensley, M. J. and Gibson, P. G. 2002. Neutrophil degranulation and cell lysis is associated with clinical severity in virus-induced asthma. *Eur Respir J.* 19(1): 68-75.

Watt, A. P., Courtney, J., Moore, J., Ennis, M. and Elborn, J. S. 2005. Neutrophil cell death, activation and bacterial infection in cystic fibrosis. *Thorax.* 60(8): 659-664.

Weiss, S. J. 1989. Tissue destruction by neutrophils. *N Engl J Med.* 320(6): 365-376.

Wendeler, M. and Sandhoff, K. 2009. Hexosaminidase assays. *Glycoconj J.* 26(8): 945-952.

Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C. and Mattick, J. S. 2002. Extracellular DNA required for bacterial biofilm formation. *Science.* 295(5559): 1487.

Wiesner, J., Premsler, T. and Sickmann, A. 2008. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics.* 8(21): 4466-4483.

Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C. and Imberty, A. 2003. Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for Aleuria aurantia lectin. *J Biol Chem.* 278(29): 27059-27067.

Wine, J. J. 1999. The genesis of cystic fibrosis lung disease. *J Clin Invest.* 103(3): 309-312.

Wong, K., Roberts, M. C., Owens, L., Fife, M. and Smith, A. L. 1984. Selective media for the quantitation of bacteria in cystic fibrosis sputum. *J Med Microbiol.* 17(2): 113-119.

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(1-2): 115-128.

Wuhrer, M., de Boer, A. R. and Deelder, A. M. 2009. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. *Mass Spectrom Rev.* 28(2): 192-206.

Wuhrer, M., Deelder, A. M. and Hokke, C. H. 2005. Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci.* 825(2): 124-133.

Yamauchi, K., Tomita, M., Giehl, T. J. and Ellison, R. T., 3rd. 1993. Antibacterial activity of lactoferrin and a pepsin-derived lactoferrin peptide fragment. *Infect Immun.* 61(2): 719-728.

Yang, L., Jelsbak, L. and Molin, S. 2011. Microbial ecology and adaptation in cystic fibrosis airways. *Environ Microbiol.* 13(7): 1682-1689.

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(4): 956-978.

Yoshida, E., Sakurama, H., Kiyohara, M., Nakajima, M., Kitaoka, M., Ashida, H., Hirose, J., Katayama, T., Yamamoto, K. and Kumagai, H. 2012. Bifidobacterium longum subsp. infantis uses two different beta-galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. *Glycobiology.* 22(3): 361-368.

Zahm, J. M., Gaillard, D., Dupuit, F., Hinnrasky, J., Porteous, D., Dorin, J. R. and Puchelle, E. 1997. Early alterations in airway mucociliary clearance and inflammation of the lamina propria in CF mice. *Am J Physiol.* 272(3 Pt 1): C853-859.

Zaia, J. 2008. Mass spectrometry and the emerging field of glycomics. *Chem Biol.* 15(9): 881-892.

Zaia, J. 2010. Mass spectrometry and glycomics. *OMICS.* 14(4): 401-418.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol.* 7(1-2): 203-214.

Ziegler, A. and Zaia, J. 2006. Size-exclusion chromatography of heparin oligosaccharides at high and low pressure. *J Chromatogr B Analyt Technol Biomed Life Sci.* 837(1-2): 76-86.

Appendix removed from Open Access version as they may contain sensitive/confidential content.