N-Glycan analysis and engineering

Anders Holmggaard Hansen


The Novo Nordisk Foundation Center for Biosustainability,
Technical University of Denmark, Kgs. Lyngby, Denmark

Supervisors:
Robert Madsen
Scott James Harrison
Helene Fastrup Kildegaard
**Front cover picture:**

Human alpha-1-antitrypsin cartoon structure with the three A2G2S2 N-glycans attached at positions 70, 107 and 271. Notice the three N-linked glycans represent 7000 Dalton of the 52000 Dalton total weight (13 % is the three N-glycans). Picture is made with PyMol.
Table of Contents

Acknowledgements .......................................................................................................... 5
Abstract ........................................................................................................................... 6
Dansk Resume .................................................................................................................. 7
Abbreviations ................................................................................................................... 8
Disclaimer of work ......................................................................................................... 10
List of publications: ........................................................................................................ 13
Aim of thesis .................................................................................................................. 15
Introduction: .................................................................................................................. 16
Glycobiology .................................................................................................................. 16
  Sugar monomers ................................................................................................................. 20
  Terminology in glycans: .................................................................................................... 21
  Types of glycosylation ....................................................................................................... 21
  N-Glycans .......................................................................................................................... 22
Organization and presentation of glycans ........................................................................... 28
Immune response to glycans ........................................................................................... 30
Choosing a production cell line for pharmaceutical products ............................................. 31
The Erythropoietin case story ........................................................................................... 33
  Glyco-engineering for enhanced erythropoietin N-glycan branching, sialylation and modification ......................................................................................................................... 33
The analysis of N-glycans: ............................................................................................... 35
  Fluorescence detection of labeled N-glycans ................................................................ 35
  Mass spectrometers in glycan analysis: .......................................................................... 38
  Common mass analyzers ................................................................................................. 39
  Glycoproteomics: ........................................................................................................... 41
Results: .......................................................................................................................... 42
Chapter 1: The analysis of N-glycans ............................................................................. 42
  Ion trap mass spectrometry ............................................................................................. 42
  Space-Charge problem with ion-trap MS during analysis: ................................................. 43
  Choosing the right HPLC gradient for LC-MS analysis ..................................................... 46
  Example of HPLC-MS method: ...................................................................................... 46
  Annotation of N-glycans ................................................................................................ 50
Chapter 2: Publication I: Fluorine modifications to 2-AA improves mass spectrometer signal of labeled N-Glycans .............................................................................................................. 53
Chapter 3: Glycans in product quality ............................................................................. 57
Acknowledgements

I dedicate this work to my family, friends and colleagues. Namely my wife Mette for her endless support and trust in me. My parents Bjarne and Elsebet whom I owe much respect and gratitude for love, guidance, patience, encouragement and for always believing in me even when I don’t. To my two children Erik and Signe, now two and four years old, both a happy daily reminder to play, imagine, stay curious and never forget to dream with no borders.

To Scott Harrison, who welcomed me to this laboratory and helped me on my first steps on this PhD. Without his gentle pressure this thesis would never have been accomplished and a big thanks for aid and his and Jen’s hospitality during my visit in his laboratory in Birmingham.

To Hanne Bjerre Christensen, for her solid support and never letting me forget that I needed to write. To my gentle giant of a boss, Bjørn Gunnar Voldborg for the resources to make all this happen. Helene Fastrup Kildegaard for great discussions about science and glycans over the years, I sincerely miss our discussions. To Tune Wulff for the patience it possibly requires to share an office with me. To all the great people I am working with in the CHO group and in the analytics group especially, Thomas Amann, Mette Kristensen, Sara Petersen Bjørn, Tae Kwang Ha, Lars Schrübers, Lasse Ebdrup Pedersen, Andreas Klitgaard, Eric Yuzhou Fan, Helle Munck Petersen, Johnny Arnsdorf, Zulfiya Sukhova, Karen Kathrine Brøndum, Karoline Schousboe Fremming, Marianne Decker, Sanne Schoffelen, Lars Boje Petersen, Mikkel Schubert, Hooman Hefzi, Stefan Kol, Eva Balslev Jørgensen, Charlotte Brøchner, Lizzie Eriksen, Kristian Lund Jensen, Dorthe Charlotte Bonefeldt, Elham Maria Javidi, Sarah Dina Blomquist, Christian Oscar Wistrøm and Michael Horsted Pfeiffer for good talks, laughs and great collaboration. You have all of you helped me with understanding analytical chemistry or mammalian cell biology in praxis or in theory and together I believe we have created great science! Also a big thanks to Stine Clemmensen and Magdalena Skrzypczak for great collaboration on glycoengineering the S2 cell line. Finally a big thank you to The Novo Nordisk Foundation and Grundforskningsfonden for the financial support.
Abstract

The structural elucidation of N-glycans represents an analytical challenge as N-glycans are diverse with more than 100,000 reported structures. N-Glycans are complex molecules, up to 5000 Da in size, and their functions are determined by their unique chemical structures. Effective and robust analytical methods are required to properly elucidate glycan structures, as even minute changes in branching or changing a single sugar moiety can have profound biological impacts. Rigorous analysis of the glycans on biopharmaceuticals is therefore essential to ensure therapeutic quality and safety. To aid research in this field the toolbox of methods for analysis of N-glycans needs to be expanded. The glycobiology field has been severely hampered by a lack of accurate and solid analytical methods. I here demonstrate a new halogenated compound for labeling of N-glycans for analysis by Liquid chromatography-mass spectrometry (LC-MS). This method is 2-3 fold more sensitive than the previous state of the art using 2-Aminobenzamide (2-AB) or 2-Aminobenzoic acid (2-AA). This method is more sensitive and more precise, whether for monitoring glycosylation or for identifying unknown N-glycans.

Glycoengineering using rational design of glycans creates immense possibilities. The possibility to make hormones, plasma proteins and antibodies with human glycosylation will make the next generation of biopharmaceutical drugs safer and more efficient. Here we demonstrate the entirely humanized N-glycans profile of two plasma proteins, which were made recombinantly in a Chinese hamster ovary (CHO) cell line. Both of these could potentially be used as a replacement in patients lacking a functional plasma protein. We furthermore demonstrate how glycans change by rational choice of cell line, how glycosylation changes over time in a fed-batch reactor and the combinatorial knock out of galactoses on N-glycans on proteins from a CHO cell. This has been possible because of the establishment of a stable and accurate analysis platform of N-glycans. The platform utilizes high performance liquid chromatography to separate the N-glycans and their isomers with on-line mass spectrometry to verify the mass of the glycans to aid identification.

The coming together of state of the art analytical methods with modern understanding and engineering possibilities in molecular biology will be of great importance in the years to come.
Dansk Resume

Den strukturelle analyse af N-glykaner repræsenterer en analytisk-kemisk udfordring, da N-glykaner er særdeles uensartede med mere end 100.000 rapporterede strukturer. De er komplekse molekyler, op til 5000 Da i størrelse, og deres funktioner bestemmes af deres unikke kemiske strukturer. Effektive og robuste analytiske metoder er nødvendige for korrekt at opklare glykanstrukturer, da selv små ændringer kan have stor biologisk betydning. Grundig analyse af glykaner fra biofarmaceutiske lægemidler er derfor afgørende for at sikre kvaliteten. For at støtte forskningen på området skal værktøjskassen til analyse af N-glykaner udvides, da glykobiology-feltet er blevet hæmmet af manglen på præcise og solide analytiske metoder. Vi demonstrerer her en ny halogeneret forbindelse til mærkning af N-glykaner og efterfølgende analyse med LC-MS. Denne metode er 2-3 gange mere følsom end den tidligere teknologi med anvendelse af 2-AB eller 2-AA. Denne metode er mere følsom og mere præcis, hvad enten det gælder glykosylerings profiler eller identifikation af ukendte N-glykaner.


Samlet set vil sammenkomsten af moderne analytiske metoder med moderne forståelses- og tekniske muligheder inden for molekylærbiologi være af stor betydning i de kommende år.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-AA</td>
<td>Antranillic acid, 2-Aminobenzoic acid</td>
</tr>
<tr>
<td>2-AB</td>
<td>2-Aminobenzamide, Anthranilamide</td>
</tr>
<tr>
<td>4-TFB</td>
<td>2-Amino-4-(trifluoromethyl)benzoic acid</td>
</tr>
<tr>
<td>AAT</td>
<td>Alpha 1-antitrypsin, SerpinA1</td>
</tr>
<tr>
<td>AAVHSC</td>
<td>Hematopoietic stem cell-derived adeno-associated virus vector</td>
</tr>
<tr>
<td>ADC</td>
<td>Antibody drug conjugate</td>
</tr>
<tr>
<td>ADCC</td>
<td>Antibody-dependent cell-mediated cytotoxicity</td>
</tr>
<tr>
<td>alpaga-l</td>
<td>Galactose-α1,3-galactose</td>
</tr>
<tr>
<td>APCI</td>
<td>Atmospheric-pressure chemical ionization</td>
</tr>
<tr>
<td>B4galt</td>
<td>β-1,4-Galactosyltransferase</td>
</tr>
<tr>
<td>B3gnt</td>
<td>βGal β-1,3-N-Acetylglucosaminyltransferase</td>
</tr>
<tr>
<td>C1inh</td>
<td>Plasma protease C1 inhibitor, SerpinG1</td>
</tr>
<tr>
<td>CAD</td>
<td>Charged aerosol detector</td>
</tr>
<tr>
<td>Cas9</td>
<td>CRISPR Associated protein 9</td>
</tr>
<tr>
<td>CE</td>
<td>Capillary electrophoresis</td>
</tr>
<tr>
<td>CFB</td>
<td>Novo Nordisk Foundation, Center for Biosustainability at DTU</td>
</tr>
<tr>
<td>CFG</td>
<td>Consortium for Functional Glycomics</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary</td>
</tr>
<tr>
<td>CID</td>
<td>Collision-induced dissociation</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTU</td>
<td>Technical University of Denmark</td>
</tr>
<tr>
<td>EI</td>
<td>Electron ionization</td>
</tr>
<tr>
<td>EMA</td>
<td>European Medicines Agency</td>
</tr>
<tr>
<td>EPO</td>
<td>Erythropoietin</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>ESI</td>
<td>Electro spray ionization</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration, USA</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Fut8</td>
<td>Fucosyltransferase 8</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>Glu</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-Acetylglucosamine</td>
</tr>
<tr>
<td>GU</td>
<td>Glucose units</td>
</tr>
<tr>
<td>HCD</td>
<td>Higher-energy collisional dissociation</td>
</tr>
<tr>
<td>HILIC</td>
<td>Hydrophilic interaction chromatography</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>KI</td>
<td>Knock in</td>
</tr>
<tr>
<td>KO</td>
<td>Knock out</td>
</tr>
<tr>
<td>LacNAc</td>
<td>N-acetyllactosamine, Galactose-β-1,4-N-Acetylglucosamin</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>Mgat</td>
<td>N-Acetylglucosaminyltransferase</td>
</tr>
<tr>
<td>MRM</td>
<td>Multiple reaction monitoring</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry/mass spectrometer</td>
</tr>
<tr>
<td>m/z</td>
<td>mass-to-charge ratio</td>
</tr>
<tr>
<td>NeuGC</td>
<td>N-Glycolylneuraminic acid</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PNGase</td>
<td>Peptide-N-Glycosidase</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>S2</td>
<td><em>Drosophila melanogaster</em> Schneider 2 cells</td>
</tr>
<tr>
<td>St3gal</td>
<td>β-Galactoside α-2,3-Sialyltransferase</td>
</tr>
<tr>
<td>St6gal</td>
<td>β-Galactoside α-2,6-Sialyltransferase</td>
</tr>
<tr>
<td>TALEN</td>
<td>Transcription activator-like effector nuclease</td>
</tr>
<tr>
<td>UHPLC</td>
<td>Ultra high performance liquid chromatography</td>
</tr>
<tr>
<td>UPLC</td>
<td>Ultra performance liquid chromatography</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
<tr>
<td>XIC</td>
<td>Extracted-ion chromatogram</td>
</tr>
</tbody>
</table>
Disclaimer of work

For all data presented in this thesis I have carried out the work related to the glycan analysis. Most of the work builds on a collaboration with groups at DTU (Technical University of Denmark) CFB (Center for Biosustainability), DTU Bioengineering and Expres2ion Biotechnologies doing the cell work and I conducted the analysis of glycans and the interpretation of the data. I have set up the entire LC-MS (Liquid chromatography-mass spectrometry) platform including testing multiple columns, developing gradients, setting up MS (Mass spectrometer), tuning gas-flows, ion source settings, setting up the software integration settings. I have screened carbohydrate cleaving enzymes and label reagents from a multitude of suppliers including Sigma (enzymes and label reagents), Thermo Fisher (enzymes and label reagents), ProZyme (purification columns, enzymes and label reagents), Perkin Elmer (enzymes and label reagents), New England Biolabs (enzymes) and finally Waters (96 well purification plates, enzymes and label reagents). Before deciding on first ProZyme 2-AB that was the most versatile and with highest sensitivity at that time. Subsequently I changed to Waters when the new RapiFluor label came on the market in early 2016 because of the increase in sensitivity.

Most proteins in this thesis have been purified by Helle Munck Petersen, Sanne Schoffelen or Stefan Kol at the CHO (Chinese hamster ovary) protein platform. The construction of DNA/vectors etc. was done by Sara Petersen Bjørn and Patrice Ménard. The cell work was performed by The CHO Core cell lab by Johnny Arnsdorf, Karen Kathrine Brøndum, Zulfiya Sukhova or Karoline Schousboe Fremming. All work with bioreactors have been performed by Marianne Decker, Kristian Lund Jensen and Elham Maria Javidi.

Specifically for Chapter 1, the data were obtained in collaboration with the CHO group: The examples in this Chapter about annotating glycan data, samples derive from CHO cell KO cell lines coming from the CHO-Core cell laboratory.
Specifically for Chapter 2, the concept of adding halogens to increase the ionization capabilities was conceived by Scott Harrison.

Specifically for Chapter 3, the data were obtained in collaboration with the CHO group: The examples in this Chapter about annotating glycan data, all samples derives from CHO cell KO cell lines coming from the CHO-Core cell laboratory.

Specifically for Chapter 4 the data were obtained in collaboration with the CHO cell line engineering CLED group. Cell work and titer measurements of the CHO-CS13 cell line was performed by Tae Kwang Ha and Stefan Kol, respectively. The concept of adding Baicalein to reduce stress was conceived by Tae Kwang. The cell work related to the comparison between CHO-K1, DG44 and CHO-S was performed by Yuzhou Fan.

Specially for Chapter 5, I contributed to the design/selection of the experiments including the selection of the B4galt1,2,3,4 genes to knock out of the KO cell line and the work with the glycan analysis LC-MS, annotation of spectra, quantification and interpretation of data. Thomas Amann carried out all the cell work and metabolite data related to this paper.

Bioreactor runs and growth data from them are by the curtesy of the CHO core fermentation team Marianne Decker, Kristian Lund Jensen, Elham Maria Javidi. Protein purifications by Helle Munck Petersen, Stefan Kol at the CHO core protein platform.

Specially for Chapter 6, I contributed to the design/selection of the experiments and the work with the glycan analysis LC-MS, annotation of spectra, quantification and interpretation of data. Thomas Amann made the cell work, metabolite data and sample preparation of glycans related to this paper.
Specially for Chapter 7, I contributed to the design/selection of the experiments including the nine of the ten genes to knock out of the KO cell line (except Glul) and the work with the glycan analysis LC-MS, annotation of spectra, quantification and interpretation of data. Thomas Amann made all the cell work, metabolite data and sample preparation of glycans related to this paper.

Specifically for Chapter 8. I contributed with the work related to the LC-MS analysis of N-glycans including alterations to the LC method needed to analyze the shorter insect glycans. N-glycan samples were obtained from ExpreS²ion Biotechnologies. Data and similar figure are in the PhD thesis by Stine Clemmensen and in submitted (awaiting patent before submitting) manuscript: Tailored N-glycosylation in Drosophila S2 cells aimed at: humanization, de-fucosylation, and increased mannosylation. The cell work and protein purification of Drosophila S2 and derived proteins was performed by Stine Clemmensen.
List of publications

The following work is included in this thesis:

**Glyco-engineered CHO cell lines producing alpha-1-antitrypsin and C1 esterase inhibitor with fully humanized N-glycosylation profiles**

**N-Glycan derivatization with halogen-containing compounds can improve mass-spectrometry signal**
Hansen, A. H.; Harrison, S. J.; (Submitted manuscript)

**CRISPR/Cas9-Multiplexed Editing of Chinese Hamster Ovary B4Gal-T1, 2, 3, and 4 Tailors N-Glycan Profiles of Therapeutics and Secreted Host Cell Proteins**

The following work is included partly in this thesis:

**Baicalein Reduces Oxidative Stress in CHO Cell Cultures and Improves Recombinant Antibody Productivity**

**N-glycosylation toolbox in Drosophila S2 cells: humanization, de-fucosylation, and increased mannosylation**
S. Clemmensen; A. Hansen; M. Skrzypczak; P. Szlarski; I. Nielsen; T. Søgaard; L. Poulsen, B. Voldborg; M. Andersen; H. Fastrup; M. Nielsen; W. Jongh (Manuscript in preparation)

In addition to the work included in this thesis I have been involved in the following publications during the thesis period:

**2018:**

Bjørn Gunnar Voldborg, Stefan Kol, Anders Holmgaard Hansen, Helene Fastrup Kildegaard. Filed; 11 July 2018
2017:

**Predictive glycoengineering of biosimilars using a Markov chain glycosylation model**

**De-bugging and maximizing plant cytochrome P450 production in Escherichia coli with C-terminal GFP fusions**

**Building a fructan LC–MS2 library and its application to reveal the fine structure of cereal grain fructans**

2016:


**A Markov chain model for N-linked protein glycosylation - towards a low-parameter tool for model-driven glycoengineering**

2015:

**One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment**

**LC-MS analysis reveals the presence of graminan- and neo-type fructans in wheat grains**
Aim of thesis

In 2014 when this work was initiated the state of the art for N-glycan analysis was 2-AB or 2-AA labeling of released N-glycan pool by robotics pipetting for accuracy and reproducibility. The labeled N-glycans were subsequently analyzed by HPLC-MS with fluorescence detection. Other methods like, HPLC only, MS only, tandem MS, capillary electrophoresis and high-performance anion exchange chromatography with pulsed amperometric detection were used for analysis of glycans in specific cases. All of these have drawbacks and do not have the versatility and/or sensitivity of HPLC-MS with labeled glycans. The biggest drawback for the 2-AB or 2-AA method was the sensitivity of the mass spectrometer. Thus the aim of this thesis was to design a strategy to enhance the mass spectrometry sensitivity part by adding halogens to the 2-AA labeling compound. This design should then increase the mass spectrometer signal without hampering the fluorescence signal, as well as improve the signal to noise and limit of detection of the method.

With this new more sensitive and accurate method we would be able to assist in the glycoengineering effort to map out the exact function of the glycosyltransferases in the CHO cell line. This thesis will describe the analysis of N-glycans and how that supports glycoengineering in the CHO cell line.
Introduction

Glycobiology

Glycobiology is a growing field as the importance of complex sugars becomes more known to researchers and scientists around the world. A pubmed search of the word glycobiology in the year 2017 yields over 745 hits on manuscripts whereas in the 2008 only 296 hits were obtained i.e. a 2.5 factor increase over nine years. Glycobiology as a term covers structure, function and the biology of oligosaccharides/carbohydrates, in this thesis, generally called glycans. Glycans are present in and on every known living cell and organism in the world.

Glycobiology is especially relevant to biopharmaceuticals. In mammalian cells more than 60% of the proteins are post-translationally modified (PTM) and the most common modification is glycosylation.

N- and O-Glycans are often involved in functions that are of great concern to us all like safety, half-life and efficacy of pharmaceutical drugs and how viruses can infect us.

Almost half of all new drug approvals by FDA (Food and drug administration) and EMA (European medicines agency) were biopharmaceuticals in 2018\(^1\) compared to about a quarter in 2014\(^2\). This trend seems to continue as for example Novo Nordisk chose recently to discontinue the work with small molecule drugs to focus on biopharmaceutical products\(^3\). Most biopharmaceuticals contain N-glycan modifications and these are important for the safety and efficacy of the drug – a so called critical quality attribute. During the production of the biopharmaceuticals one have to be able to ensure the structures of the N-glycans do not change from batch to batch. This is however not an easy task. The analysis of the N-glycans is laborious and expensive. Shajahan et al. notes in a recent review\(^4\) that:

>>The structural analysis of glycoproteins is a challenging endeavor... but only a very limited number of labs have the expertise required to accomplish this task.<<

The potential however is to be able to control the N-glycans on the cell line level and thereby making the N-glycans by design – A quality by design more than a quality driven by
process parameters. With new gene editing technologies it is now possible to start engineering glycosylation to increase the safety and efficacy of pharmaceutical drugs.

In nature glycans are critical in a large variety of functions. In the seminal paper by Varki: Biological roles of oligosaccharides: All of the theories are correct⁵, the emphasis on glycans in a multitude of functions is still valid. Examples of glycan importance in biology range from cancer, virus, cell development, autoimmune diseases and drugs to protein stability. This is exemplified in cancer by the truncation of glycans⁶–⁸, from this it follows that the glycans are obvious antibody targets. The overexpression of a glycosyltransferase is linked to cancer⁹. For viral infection, branching of terminal sialic acids are important for how virus infect birds, swine and humans¹⁰. It has been shown that glycan profiles can be used as biomarkers for inflammatory bowel diseases¹¹. It has been reported that in cell development glycans regulate the stem cell status and thus the fate of the developing cells¹². Stability of proteins are dependent on glycans and glycans protect against aggregation¹³ ¹⁴. Glycosylation defines the adhesive properties of proteins and cells exemplified with the glycocalyx – glycans covering all cells¹⁵. In high shear environments like the bloodstream the cells are heavily glycosylated to avoid cell-cell adherence and to protect from fluid shear stress¹⁶ ¹⁷.

Considering this central role that glycans play in the previously mentioned fields, glycoproteins are becoming the target of next-generation bio-therapeutics and diagnostics.

The logical question to ask at this point is: “So why has the field taken so long to get to where we are today with design of N-glycans in the production of recombinant glycoproteins?” As cited below, in 1989, thirty years ago, Lee et al. made a cell line where β-Galactoside α2,6-Sialyltransferase was knocked in into Chinese Hamster Ovary (CHO) cells. They foresaw that N-glycans will be designed by engineering the glycosyltransferases to produce the desired form. Lee et al:  

>>>...the ability to engineer cell lines with altered glycosylation machinery may provide additional flexibility to the design and production of recombinant glycoproteins in the future<<¹⁸

Three major things have changed since 1989;

• Genomes are publicly available
• Gene editing tools have been developed
• Analysis of glycans have improved

Since the first full genomes became publicly available in the 1990s many others have followed. With the knowledge of the DNA sequence one can start editing the genome in a targeted fashion.

Gene editing tools have improved dramatically with the introduction of zinc fingers\textsuperscript{19,20}, TALENs\textsuperscript{21,22} (Transcription activator-like effector nuclease) and especially CRISPR/Cas\textsuperscript{23–25} (Clustered regularly interspaced short palindromic repeats / CRISPR Associated protein 9), making gene-editing feasible, fast and cheap enough for most laboratories.

In 2008 the state of the art N-glycan analysis was a three day sample preparation including overnight trypsination and overnight PNGase F treatment before HPLC or HPLC-MS could be performed. Many laboratories could not perform this procedure, as they were unable to achieve stable and reproducible results and on top of that the PNGase F enzyme is quite expensive. The gold standard for derivatizing the N-glycans were 2-Antranillic acid or 2-Aminobenzamide that only have a low extinction coefficient compared to Tamra or coumarin type dyes, thus hampering the analysis because relatively large sample quantities were needed to perform the identification and annotation.

Major improvements in all of these three areas have made it possible to tailor-make a wide variety of N-glycans on recombinant glycoproteins recently exemplified in Yang et al.\textsuperscript{26}, Schulz et al.\textsuperscript{27} Luo et al.\textsuperscript{28} and Amann et al.\textsuperscript{29}. Smaller but important improvements are a common nomenclature for glycans in 2009 with the Oxford\textsuperscript{30} and CFG (Consortium for Functional Glycomics) standard\textsuperscript{30,31} of drawing the complex structures to a form where all the important information is still present, but immensely faster to read and understand. Throughout this thesis predominantly CFG nomenclature will be used as that is the output of Glycoworkbench drawing software\textsuperscript{32}. Oxford nomenclature may be more accurate as the lines and position designates linkage and position respectively. The reason being that if not the specific branching is known then there is no advantage of Oxford nomenclature and the colorful CFG is faster to read.

These problems are in part summed up by Sheridan et al.:
But development of the field has been hampered by its sheer complexity and a lack of tools comparable to those available for genomics and proteomics.\(^{33}\)

And Shubhakar et al.:

Automated interpretation is mostly performed assisted by a predefined set of possibilities, or by matching experimental data to a database\(^{34}\)

Even though such databases exist today\(^{35-37}\) and have since around 2010, they are either restricted to Waters™ equipment until 2018 or not easily searchable. Now these databases are accessible in open source format accessible via www.glycostore.org and www.glytoucan.org, the latter previously called GlycomeDB before changing the name into GlyTouCan, a part of the portal glycosmos.org. On the commercial software side PREMIER Biosoft/Thermo Scientific has software designed to handle the characterization of N-glycans by MS/MS. The software, SimGlycan™ predicts structure of glycans from the MS/MS but have no retention time information to aid structural information and therefore it relies solely on the MS and MS/MS data. Similarly for the open source and free GlycanMass at expasy.org/glycanmass/ that is a tool to calculate the mass of a glycan structure. The problem with both of these pieces of software, for our use on iontrap MS, is that without high resolution MS there will often be a number of false positives as many glycans have same or similar mass and the fact that software does not perform a sanity check if the calculated result can be true for the specific sample in terms of species, retention time or rarity of the database-hit. This problem increases if algorithms include theoretical structures as well as chemically synthesized structures that may not exist in nature. As databases become larger/better, includes MS/MS data and retention time together with the functionality of selecting species/cell line as seen in proteomic software, automated annotation of N-glycans will improve greatly.

A recently published example demonstrates native mass-spectrometry of glycoproteins that is able to identify N- and O-glycans but note that: unambiguous assignment of N-glycan structures is still not possible at this level\(^{38}\) as sialic acids needs to be removed prior to analysis and because no retention time from HLPC is acquired. An example of the complexity is shown by the Erythropoietin (EPO) native mass-spectrum exhibiting hundreds
of different peaks due to the heterogeneity on its three N- and one O-glycosylation sites. This makes analysis of such data very time consuming.

This thesis will focus on N-glycans. O-Glycans are equally important for many biological aspects but the analysis of O-glycans are lacking because of the problem of not having a single enzyme to cleave it from the protein. This leaves only chemical release usually by beta-elimination that risk pealing of sugars. One of the most interesting biological functions of O-glycans are the involvement in sperm-egg interaction:\(^{39,40}\):

>>>removal of O-linked oligosaccharides destroys its sperm receptor activity<<

and thus links O-glycans directly to fertility of humans.

Sugar monomers

The glycans consist of sugar monomers to ease the reading of larger structures, graphical representations will be used. Here in Figure 1 is the sugar monomers and their CFG graphical representation shown.

![Figure 1: Sugar monomers; their graphical representation, names and shorthand](image)

- □ = N-Acetylglucosamine (GlcNAc)
- ● = Mannose (Man)
- △ = Fucose (Fuc)
- ○ = Galactose (Gal)
- ▶ = Sialic Acid (N-Acetyl Neuraminic Acid) (Sia)
- ★ = Xylose (Xyl)
Terminology in glycans

Branching  When more than one sugar-monomer is attached to the same sugar-monomer and thus causing a branch-point in the structure

Glycan length  Short- or long-chain oligosaccharides

Glycan composition  Types of individual sugar-monomers that are linked together to make a particular N-glycan structure

Glycan structure  Divided into multi-branched, branched or unbranched chains

Macroheterogeniety  Glycosylation sites, and occupancy of the possible sites

Microheterogeniety  The profile or fingerprint of glycans attached to one or more sites on a glycoprotein

Glycosidic bond  Covalent bonds between monosaccharides

Reducing end  The monosaccharide of a glycan that have a free anomeric carbon, not involved in a glycosidic bond and capable of converting to the open ring form

Glycan presentation  How the glycan is presented to target cell, receptor etc. Includes the surrounding protein and composition of the glycan

Types of glycosylation

There are several different types of Glycosylation reported in the literature and these include:

N-linked  Binds to the amino acid asparagine,

Subgroups: High mannose, Hybrid, Complex
O-linked: Binds to the hydroxyl group of serine or threonine amino acid

Glycation: Non-enzymatic attachment of glucose occurs on lysine \textit{in vivo}

Glypiation: Glycan core links a phospholipid and a protein

C-linked: Mannose binds to the indole ring of tryptophan

Phosphoglycosylation: Glycan binds to serine via phosphodiester bond

\textbf{N-Glycans}

N-Glycans consist of a series of monosaccharides, for N-glycans in mammals common monosaccharides are: galactose, mannose, N-acetylglucosamine and sialic acid that are typically forming a branched structure with very specific branching see Figure 1 and Figure 2.

N-Glycans consist of a core region with two N-acetylglucosamine and three mannose units with two of the mannoses splitting from the third. This core region is conserved across species and can be found in mammals, insects, plants and fungi. See Figure 3 for graphical representation of the core structure. From each of the mannoses in the core multiple antennae can be attached, either a N-acetylglucosamine or mannose depending on the type of N-glycan. To ease the “reading” of glycans we often apply cartoon structures so specific attributes can quickly be seen (Figure 2).
Teaching of molecular/cellular biology and biochemistry today focuses on Crick’s 1958 ‘central dogma’ of molecular biology that ‘DNA makes RNA (Ribonucleic acid) makes protein’. This linear approach is not applicable in the subject of N-glycans that are not made in this linear fashion. Whereas DNA and RNA are sugars, they are under strict biological/template control to be branched the same way and only be linear structures. Glycans on the other hand are built in a non-template driven fashion. The construction of a given N-glycan is thus dependent on a series of glycosyltransferases present in specific
compartments of the Golgi building the N-glycan stepwise. This form of non-template driven construction is thus error-prone and often results in truncated forms in immortalized or cancer cell lines.

Intriguingly the majority of DNA in the genomes of higher organisms does not encode for proteins but are non-coding so-called “junk” DNA. Recently Kindgren et al. have found a compelling example of gene regulation by this RNA transcribed from non-coding DNA, thus revealing a new level of complexity on the DNA level as well thus a good example that also for DNA are there a more complexity than previously described.

N-Glycans beyond the core M3 (Man3, GlcNAc2Man3) are built in this non-template driven fashion. Any addition to this core structure (the core part of the structure, is the structure shown in Figure 3) will only have guidelines not rules because of the lack of a template. The core fucose can be α1,3 branched as seen in plants and insect cells or α1,6 as in mammalian systems. The first N-acetylglucosamine on both mannose branches are most commonly across species β1,2 branched (See Figure 2).

![Diagram of N-Glycan Core Structures](image)

**Figure 3:** Generally accepted nomenclature variants of “Man3”, “M3” N-glycan core graphical structure. Last three structures include information on branching (Oxford notation: The angle designates position, have dotted lines for α and solid lines for β linkages)
To date there are hundreds of recombinant glycoproteins approved\textsuperscript{1,2} for use in humans and these are most commonly produced in CHO cells. The CHO cells have been in use since 1947 and a wide variety exist of both adhesive and suspension types. CHO cells are mainly used because they are easy and efficient to cultivate and because it is \textit{relatively} easy to achieve titers above 1g/L of recombinant glycoproteins. An important feature for the pharmaceutical industry is that CHO cells are generally regarded as safe for production of pharmaceutical products. One of the reasons for that is that the glycans produced by CHO cells are relatively similar in structure to those in human\textsuperscript{42}.

The N-glycans are attached to an asparagine residue\textsuperscript{43} of an amino acid sequence of the protein having the general pattern: Asn-X- Ser/Thr or N-X-S/T, where X may be any amino acid except for proline. Not all the possible sites with this motif are N-glycan sites though. For glycosylation to occur the site has to be on the surface of the partially folded protein and even then, not all sites will be occupied. When attempting to engineer new N-glycan sites into a protein the success-rate is 30\% or lower even with rationally chosen sites. This was demonstrated in Flintegaard \textit{et al.} where 4 out of 15 of carefully selected and engineered sites turned out to be glycosylated\textsuperscript{44} \textit{in vivo}.

Remarkable new data suggest that more than one motif for N-glycan sites is possible. In Sun \textit{et al.} they demonstrate that the atypical glycosite motif N-X-V is an N-glycan site in two cases\textsuperscript{45}. Including O-glycans nine amino acids are known to be glycosylated\textsuperscript{46}; Asn, Arg, Ser, Thr, Tyr, Trp, Cys, hydroxylysine, and hydroxyproline.

\textbf{N-glycans can be divided into three subgroups:}

- N-glycans of the high-mannose type, in which multiple mannose residues are added to the core structure
- N-glycans of the Hybrid type, where there is attachment of one N-acetylglucosamine residues to one of the mannoses of the core structure and the other mannose can be elongated with more mannoses.
- Complex-type glycans, where there is the attachment of two or more N-acetylglucosamine residues to each of the mannoses of the core structure.
See Figure 4 for graphical representations of the subgroups.

Figure 4: From left to right; A) High mannose type, B) hybrid and C) complex type

In mammalian cells, the M9 plus three glucose units are first linked to a long chain lipid, dolichol phosphate, and then transferred on bloc to the nascent peptide backbone. This occurs in the endoplasmic reticulum (ER). After folding of the protein, the three glucose and 3-4 of the mannoses are cleaved off and the glycoprotein is transferred to Golgi, where the rest of the processing into complex forms occurs. The later stages of the assembly of complex N-glycans with glycosyltransferases are described in the introduction section of Chapter 5.

See Figure 5 for a schematic overview of the build up of a complex N-glycan from GlcNAc monomer attached to the lipid-linked, phosphorylated oligosaccharide dolichol into the finished complex N-glycan. Schematic figures showing Golgi are usually divided into three compartments; cis, medial, trans, here Golgi is shown as only two because of space restriction.
Looking at the complexity of N-glycans across species the differences become clear. The representative structures are from the GlyTouCan database\textsuperscript{37} to represent the various species. See Figure 6.
Organization and presentation of glycans

Glycan organization and presentation relates to how the glycans are presented on the surface of cells, coating the cell membrane. Presentation can also be glycan binding proteins or lectins that binds very specifically to some structures but not to others. For cell-cell interaction it is very important how the specific glycan is organized and/or presented on the cell surface. That includes both the neighboring sugar-monomers and possible branches but also the surrounding surfaces of the protein. Arthur et al. showed that the presentation of a small sugar moiety could vary a lot depending on how it was presented\(^5\). When looking at site occupancy for specific proteins the surrounding 3D surfaces of the protein must have an influence on the micro-heterogeneity of that site\(^5\).

Sialic acids can be present in various ways on N-glycans or O-glycans and branching of sialic acids as \(\alpha2,3\) or \(\alpha2,6\) and they have an impact on serum half-life or the infection rate of a virus. Exemplified in the plasma protein C1 inhibitor having a significantly prolonged serum half-life in rats with a higher sialic acid content\(^5\). The influenza virus from the avian and swine flu usually binds sialic acids via \(\alpha2,3\) rather than \(\alpha2,6\)\(^10,\)\(^5\), and thus it is not as infectious to humans as it is to birds and swine.
As noted in the introduction, several papers on the KI of st6gal1 to produce humanized glycoproteins have been made. The cells used as the parent cell line already had sialic acids of the other branching variant, the α2,3, instead of the α2,6. This shows that the presentation/branching of a single sugar monomer can be of importance.

The position of a fucose can vary as the antennae type fucose can make a Lewis X or sialyl Lewis X motif whereas the core type fucosylation, as an example, is involved in Antibody-dependent cell-mediated cytotoxicity (ADCC) effector function on IgG’s. Consequently, the presence of the fucose is important for function/effect of the biopharmaceutical drug and the position is important as the α1,3 position is potentially immunogenic.

N-Glycans are organism specific, cell type specific, protein specific, site specific and changes on the protein backbone can alter the N-glycan composition on site specific N-glycans. Thus in theory it would be possible to engineer the glycoform by changing the peptide backbone instead of knocking out glycosyltransferases, if a maintained function of the glycoprotein is un-altered. Other approaches involve media additions and growth conditions in general. Figure 7 captures the most important features that result in a specific N-glycan on a specific site of a glycoprotein. Changing any of the parameters might cause the resulting structure to be a completely different one.

![Figure 7: N-glycan structure is specified by the properties of the biological context](image)

Presentation of glycans (how glycans, or just parts of glycans, are presented to the lectin or glycan-binding protein) is important. As an example the glycan-binding protein Galectin-1 binds strongly to α2,3-Sialylated and non-sialylated terminal N-Acetyllactosamine units but not at all to α2,6-Sialylated terminal N-Acetyllactosamine units, demonstrating how tiny changes in the glycan structure can affect binding. Similarly, Song et al. shows that just a
methyl group on a sialic acid is an important feature in a virus binding to glycan structure. See Figure 8.

**Figure 8:** Figure from Song et al.\(^65\). Virus binds very specifically to only some glycan structures. A single methyl group on NeuGC terminal sialic acid is the difference between structure 48 (Neu5Gcα3Galβ4GlcNAcβ3Galβ4Glc) with almost no detected binding and structure 49 (Neu5GcMeα3Galβ4GlcNAcβ3Galβ4Glc)

**Immune response to glycans**

Plants have the capability to add a core fucose linked α1,3 whereas the mammalian type core fucose is linked α1,6. This seems unimportant at first but because the PNGase F enzyme only can cleave structures with the core α1,6 fucose this poses a problem with analyzing N-glycans from plants and other species with the capability to add the α1,3 type and it is a potential problem in therapeutics containing this specific fucose linked α1,3 branching variant, as humans do not have it.

Specifically for replacement therapy treatment with taliglucerase alfa, a version of human glucocerebrosidase produced in plant cells, no immune response is observed after 30 months\(^61\)\(^62\), still 11% of clinical trial patients were considered to have antibodies to the plant specific glycans on taliglucerase alfa after long term treatment. These patients are excluded from receiving more of the drug.

The most well studied case is the Galactose-α1,3-galactose (alpha-gal) glycan moiety (see also Figure 9) on the therapeutic Cetuximab, where patients receiving the drug potentially develops an immune response to the drug\(^68\)\(^–\)\(^70\).
This suggests that anti-plant glycan antibodies occurs more rarely than anti-alpha-gal glycan antibodies.

Galactose-α1,3-galactose is specifically a α1,3 branched galactose bound to another galactose, see Figure 9 below. It can be attached to either branch and both branches can simultaneously be occupied by an alpha-galactose moiety.

Choosing a production cell line for pharmaceutical products

As stated previously the structure and location of the glycans on a protein are important to the function of a cell line. To maintain a desired glycan profile it is important to control the bio-process and post-process stability for the manufacturing of pharmaceutical biologics. Due to the complexity of the glycoproteins, the synthesis of complex N-glycans are only seen in eukaryotic systems. This does include yeast, as *Pichia pastoris* can be engineered to make humanized N-glycans with impressive results. The wild type yeast however, does not make the complex human forms but mostly high mannose and hyper mannose forms and smaller quantities of hybrid type N-glycans. Similarly, plants and microalgae are also incapable of making humanized complex glycosylation without engineering, as plants synthesize glycans carrying β1,2-xylose and α1,3-fucose that may induce immune responses in humans. Nevertheless only few drugs produced in non-mammalian cell systems have
been approved until now, see Walsh\textsuperscript{2}. In the case of insect cells, it produces small complex N-glycans, hybrid forms and high mannose forms (see Chapter 8 for glyco-engineering of an S2 insect cell line). In mammalian cells, mannose (M5 and M6) structures are produced as an intermediary step in the ER before the construction of complex N-glycans occur in the Golgi.

Assurance of high quality products over the years has been important for the manufacturing of biopharmaceuticals. N-Glycans are an important part of this as a critical quality component\textsuperscript{75} or critical quality attribute\textsuperscript{76} especially because of ADCC in monoclonal antibodies where the presence or lack of fucose alters the efficacy of the drug. This can in part be solved as quality by design on N-glycan level but that has historically not been the case and several lessons have been learned over time with the most important ones being the creation of afucosylated cell lines\textsuperscript{77–80}, alpha-gal\textsuperscript{68,81,82} where patients receiving the biopharmaceutical developed anaphylaxis and antibodies specific for the alpha-gal moiety. Similarly N-glycolyl neuraminic acid (Neu5Gc)\textsuperscript{83–85} is un-wanted because of the potential for hypersensitivity reactions to recombinant therapeutic glycoproteins containing Neu5GC. Cases like these are the reason to the tight control from FDA/EMA on biopharmaceutical products today.

From the pharmaceutical point of view, it is important that the biopharmaceutical product can pass the FDA/EMA safety criteria. No glycans can be recognized as non-human and secondly, glycans are preferred to be as close to the human counterpart as possible. This is because of the risk of the patient, receiving the drug, developing antibodies against the non-human glycan over time. The distinction here is important, as CHO cells produce almost entirely human like N-glycans, but not necessarily the actual specific human form. This will be discussed further in the AAT case later in Chapter 7. To ensure both the safety and efficacy of biopharmaceutical products the FDA/EMA requires that the N-glycan profile of the product is very similar from batch to batch. From this it follows, that the fewer N-glycan species in the profile, the easier the monitoring of the batches will be, and subsequently the risk of batch failure is decreased. As a result, very homogeneous, human identical or human like, glycosylation profiles are highly desirable for the biopharmaceutical companies.
The Erythropoietin case story

Even slight (Planned or accidental) modification of the EPO N-glycans in recombinant systems often results in a decrease or a complete loss of *in vivo* biologic activity compared to the reference; a natural, human, sialylated, multi-antennary version. This is because the several features of this EPO N-glycan are needed for specific purposes. If the sialic acids are removed, EPO loses its *in vivo* biologic activity\(^9\), whereas the amount of antennas (degree of branching) does not correlate to *in vivo* activity\(^86\). This loss of *in vivo* biologic activity of de-sialylated EPO could be explained by a rapid clearing from the circulation mediated by a galactose-binding protein\(^9\), which again results in hepatic uptake and degradation in the lysosome. Importantly the opposite is the case when looking at *in vitro* activity; the non-, mono- and di-sialylated N-glycans significantly and positively correlated with *in vitro* activity\(^86\). Interestingly EPO containing LacNAc elongations seems be cleared faster\(^87\). Suggesting that reducing LacNAc elongations on therapeutic recombinant glycoproteins will improve their half-life and efficacy *in vivo* and emphasizes that that it is not just size of the glycan that is important but the specific structure of the glycan.

If the N-glycans are *smaller* eg. containing fewer antennas thus containing more bi-antennary and tri-antennary structures than the reference, but are still sialylated, are cleared three times faster from the blood and thus having a much lower *in vivo* biologic activity over time than the version with more antennas\(^88\).

Glyco-engineering for enhanced erythropoietin N-glycan branching, sialylation and modification

One of the most studied glyco-proteins are EPO. Yin *et al.* demonstrated the production of a humanized erythropoietin by Knock in (KI) of the glycosyltransferases: Mgat4, Mgat5 and St6gal1\(^89\) and thus by overexpression of these three glycosyltransferase genes in combination to produce tailor-made complex N-glycan structures on a therapeutic
glycoprotein. Similarly in a study by Zon et al. where they demonstrate enhanced sialylation of recombinant human erythropoietin in CHO cells by combinatorial engineering of selected genes\textsuperscript{90}. This is more or less applicable to pharmaceutical product in the bloodstream as improved half-life can be achieved for a range of glycoproteins.

The body’s control system, which surveys the glycoprotein and clears it if not properly sialylated can be bypassed by the addition of a polyethylene glycol (PEG) -linker to the protein. The added PEG-linker changes the pharmacokinetics of PEG-rhEPO produced in glyco-engineered \textit{Pichia pastoris} when compared to non-engineered versions. This even to a degree where PEG-Yeast produced has an even longer half-life than Aranesp, a EPO version with an engineered extra N-glycan site\textsuperscript{91}.

Similarly the attachment of a PEG linker can mimic and extend the effect of a large branched, sialylated N-glycan on blood factors as well as EPO\textsuperscript{91–93}.

Other ways of adding the PEG linker is via a modified sialic acid, PEG-ylated at the 5’-amino position on the CMP-nucleotide. The PEG-linked CMP-nucleotide can thus be transferred to a glycan on a glycoprotein \textit{in vitro} by a sialyltransferase\textsuperscript{92}.

Surprisingly the KO or knockdown of LacNAc via B3gnt2 seems to increase the antennae in EPO by removing the precursor feedback inhibition, here CMP-Neu5Ac.\textsuperscript{9}
The analysis of N-glycans

Fluorescence detection of labeled N-glycans

In 2014 the state of the art methods for analysis of N-glycans were either through CE\textsuperscript{94–96}, HPLC\textsuperscript{97 98}, HPLC-MS\textsuperscript{99} or a proteomics-LC-MS approach\textsuperscript{100 101}, however all of these were too complicated to be used routinely or did not have the capability to annotate unusual N-Glycans. The industry standard was to analyze the pool of N-glycans enzymatically released from the protein backbone with Peptide: N-glycosidase F (PNGase F) and subsequently label these with the fluorescent 2-aminobenzamide (2-AB) or 2-aminobenzoic acid (2-AA), catalyzed by use of a reducing agent. As 2-AA is a regulated compound because it is easily converted to a common illicit drug, most laboratories uses 2-AB. 2-AB is also the first label broadly accepted by the field with a commercial kit from Ludger already in 2001\textsuperscript{102}. Importantly, the toxic reductant NaBH$_3$CN in the labeling procedure could be exchanged with the much less toxic picoline borane in 2010\textsuperscript{103} making the overall reaction conditions much less toxic.

Although released, un-derivatized N-glycans can be analyzed by either charged aerosol detector (CAD) or direct “native” ESI-MS of the glycan with no derivatization. For routine analysis, detection of the native N-glycan is not realistic even with modern HPLC-MS equipment due to limited sensitivity for these compounds. In theory CAD detection\textsuperscript{104} and highly sensitive HPLC-MS\textsuperscript{105 106} equipment could in special cases be used for native N-glycan analysis. This has been shown by Zaytseva et al. for the analysis of IgG N-glycans\textsuperscript{107} but as the setup is targeted and thus only detects the targeted analytes, it is not commonly used.

A more sensitive method of N-glycan analysis is by combining fluorescence derivatization of the N-glycans to analysis by MS. Fluorescent labeling methods are currently the best for quantitative N-glycomics\textsuperscript{108}, with both high sensitivity and wide dynamic range. This increases the signal to noise of the analyte up to a factor of 1,000 improving detection limits greatly. The state-of-the-art for N-glycan analysis have been 2-AB derivatization since 1995 with the Bigge et al. paper\textsuperscript{41} and more so after the Ludger commercial kit from 2001.
Derivatization methods can be broadly divided into three categories: derivatizing the reducing end of a released and reduced glycan, derivatizing the NH$_2$ group left by the PNGase F enzyme on the glycan, and finally derivatization of all functional groups. For both of the reducing end derivatizing types, commonly used tagging reagents include 2-AB, 2-AA, p-aminobenzoic acid ethyl ester, 2-AP, Instant-2-AB, Procainamide and RapiFluor. This type of derivatization facilitates purification, and enhances useful MS fragmentation data as the reducing end is now labeled.

For derivatization of all functional groups e.g. hydroxyl groups, permethylation is by far the most important type of full derivatization used for glycans. Unfortunately it destroys acetyl esters and other base-labile functional groups during the derivatization process.

N-Glycans are sometimes present only in very small quantities on glycoproteins. Example: IgG titer in WT CHO corresponds to about 80 mg/L, so in a shakeflask with 60 mL media = 4.8 mg IgG protein, whereas about 2% of the mass are glycans. This gives 0.096 mg (96 ug) total N-glycans produced, and only a small portion can realistically be analyzed after loss in purification and that not all can be injected in the LC-MS. As another example, the cell line in Amann et al. producing AAT$^{109}$, achieved a titer of 130 mg/L in a shakeflask with 60 mL media = 7.8 mg AAT in total produced).

To carry-out the analysis and annotation of a single N-glycan species from the IgG example above, corresponding to under 2% of total glycan content, then requires accurate analysis of single glycan species in the ~2 ug scale. Still assuming no loss during purification, injection and detection. The first RapiFluor paper from Waters uses 15 ug IgG$^{110}$ and that is the recommended amount according to the protocol$^{111}$.

Selected fluorescence compounds for N-glycan analysis with fluorescently labelled N-glycans are 2-AB, 2-AA, Procainamide and RapiFluor. See paper I in appendix A for more.

Interestingly, the label compound from the new RapiFluor method from Waters derives from 6-aminoquinolyl-N-hydroxysuccinimidy carbamate$^{112}$ originally invented by Millipore. Millipore and Waters patented the compound together in 1992 for an amino acid analysis procedure that was later commercialized by Waters under the name AccQ•Tag. The good fluorophore from the original Millipore/Waters label has gone off patent and has been
picked up by Waters to be used in N-glycan labeling today. Similarly, Prozyme, which produces and sells instant-AB and Instant procainamide, both widely used labeling compounds, is in the process of being bought by Agilent to strengthen their glycan portfolio in the competition with Waters. Both companies now deliver complete Glycan analysis platforms including consumables, LC-MS instrumentation and software.

The latest of the combinatorial approaches is an elaborate sample preparation protocol that uses multiple, parallel rounds of exoglycosidase enzymes and subsequent array lectin profiling and mass spectrometry. The algorithms then incorporate the glycan-array data to the mass spectrometry data to unambiguous annotation of the glycans in an automated setup. The obvious advantage here is the automatic annotation and linkage information captured, by either the lectin array data or the data from the exoglycosidase cleavages. However, the problem is the vast amount of preparation for each individual sample and the cost of exoglycosidases and lectin array added to the price of running the mass spectrometer. Jensen et al. have published a workflow to analyze both N- and O-glycans from a given glycoprotein in a four day procedure, the data requiring largely manual interpretation afterwards.

An emerging method for analyzing N-glycans is bottom up glyco-proteomics where an enzyme, typically trypsin, cleaves the glycoprotein into glycopeptides by proteolytic digestion, up-concentrated and subsequently analyzed these by mass spectrometry, Yang et al. shows a good example of the analysis of complex N-glycans from EPO being analysed by this method. This approach is clearly feasible for IgG’s and EPO where the range of possible glycans are limited but analyzing samples where the glycans are unknown, the inherent complexity of data analysis can become a big challenge. However, when routinely analyzing known glycoproteins with a known glycoprofile this is a powerful method for capturing protein sequence, potential amino acid modifications, glycan site-occupancy and N-glycan profile in a single setup. For a setup like this, a high-resolution accurate mass setup is required.
Mass spectrometers in glycan analysis

A mass spectrometer is a combination of: A part that ionizes the sample, a mass analyzer and a detector. Common mass analyzers are: quadrupole, iontrap or time-of-flight (TOF) or orbitrap.

Overall a mass-spectrometer provides mass over charge value \((m/z)\). In larger, more expensive instruments capable of fragmentation two values can typically be obtained: the intact mass of the analyte and the masses of the fragments. However, if fragmented in either an iontrap, the second quadrupole of a triple quadrupole or in the HCD cell of a quadrupole/orbitrap system, the output is still \(m/z\). The value of that \(m/z\) output is however directly related to composition of the molecule in question. The possibility to determine the mass of a molecule and its fragments present in only nano- or pico-mole amounts is of great value when the analyte is hard to obtain in larger quantities. Nuclear magnetic resonance instruments (NMR) can also prove solid data and unsurpassed structural information on molecules but is orders of magnitude less sensitive.

In LC-MS systems with fairly high flow rates \((0.3-0.7 \text{ mL/min})\), electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI) are the most commonly applied sources to convert the liquid into gaseous form and in the process ionizing the analytes. ESI is especially useful in producing ions from macromolecules because it allows ionization without fragmenting the molecule in the ionization process. As N-glycans can be over 5000u this is an advantage. APCI allows for the higher flow rates and the ionization of non-polar compounds but risks the unwanted source fragmentation of the glycans.
### Common mass analyzers

**Table 1: Mass analyzers used in N-glycan analysis**

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Glycan analysis</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single quadrupole</td>
<td>Cheapest, low mass accuracy and low resolution.</td>
<td>+</td>
<td>Simple and stable. Low sensitivity in scanning mode.</td>
</tr>
<tr>
<td>Triple quadrupole</td>
<td>Most sensitive and a quantitative instrument, still low mass accuracy and low resolution.</td>
<td>+</td>
<td>Great for targeted analysis. Low sensitivity in scanning mode. Overly complex for glycan analysis.</td>
</tr>
<tr>
<td>TOF</td>
<td>Capable of very high mass range. High mass accuracy and high resolution.</td>
<td>++</td>
<td>A good choice for glycan analysis, the fast scanning and accurate mass makes it a good match. Modern instruments are also very sensitive.</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Capable of very high mass range and 2D scanning.</td>
<td>+</td>
<td>A good instrument for glycan analysis that does not require chromatographic separation. Not possible to combine on-line with HPLC. Risk of losing signal of charged or non-charged glycans dependent on method.</td>
</tr>
<tr>
<td>TOF-TOF</td>
<td>Capable of very high mass range. High mass accuracy and high resolution.</td>
<td>+</td>
<td>Mostly seen in combination with MALDI.</td>
</tr>
<tr>
<td>Quadrupole-TOF:</td>
<td>Capable of selecting and fragmenting ion precursors and subsequently measure the masses of the fragments. High mass accuracy and high resolution.</td>
<td>+++</td>
<td>A good choice for glycan analysis, the fast scanning and accurate mass makes it a good match. The optional fragmentation is good for structure-resolution.</td>
</tr>
<tr>
<td>Instrument</td>
<td>Description</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Iontrap</td>
<td>Very sensitive in scanning over a mass range. Have the capability to do MS^n. Low mass accuracy and low resolution.</td>
<td>++</td>
<td>The very high sensitivity in scanning mode is a big advantage in glycan analysis. The optional fragmentation is good for structure-resolution. The low mass accuracy is only drawback.</td>
</tr>
<tr>
<td>Fourier transform mass spectrometer (FTMS)</td>
<td>Extremely high mass accuracy and resolution. Rare.</td>
<td>+++</td>
<td>Running cost would make it unsuitable for routine analysis.</td>
</tr>
<tr>
<td>Quadrupole-Orbitrap</td>
<td>Capable of selecting and fragmenting ion precursors and subsequently measure the masses of the fragments. High mass accuracy and high resolution.</td>
<td>+++</td>
<td>A good choice for glycan analysis, it is fast enough scanning and the very accurate mass accuracy makes it a good match. The optional fragmentation is good for structure-resolution.</td>
</tr>
<tr>
<td>Quadrupole-iontrap-Orbitrap</td>
<td>Capable of selecting and fragmenting ion precursors and subsequently measure the masses of the fragments. Only instrument that have the capability to do MS^n and accurate mass of the fragments. High mass accuracy and high resolution.</td>
<td>+++</td>
<td>A good choice for glycan analysis, it is fast enough scanning and the very accurate mass accuracy makes it a good match. The optional fragmentation is good for structure-resolution. The capability to do MS^n is rarely used. Overly complex for glycan analysis.</td>
</tr>
</tbody>
</table>
Glycoproteomics:

A way to achieve high throughput N- and O-glycan analysis, that covers site occupancy, would be to analyze the N-glycans in a glyco-proteomic setup. This is still not possible because of lack of sensitivity of MS and lack of software to handle the data sets. As the chromatography is now on the protein part it can no longer be used to separate the glycans based on type of sugars as was possible in HILIC type chromatography. As the mass spectrometer cannot distinguish between GlcNAc and GalNAc (as they are isobaric) and they have to be referred to as unspecified HexNAc. Similarly, all hexose residues; mannose, galactose and glucose all have to be referred to as unspecified hexoses.

Site specific-, sensitive and quantitative glycosylation analysis is not well developed but do exist for IgG as described by Hong et al. using multiple reaction monitoring (MRM) on a triple quadropole system\textsuperscript{114}. The obvious drawback here is that because of the nature of a multiple reaction monitoring method that it will not detect unforeseen glycans or glycan sites whereas an untargeted proteomics method would capture them. Another MRM method of N-glycosylation sites, loose the information of the glycan attached to the site of interest leaving only the information that the site was occupied\textsuperscript{115}.
Results

Chapter 1: The analysis of N-glycans

The N-glycan data in this thesis is produced by releasing the pool of N-glycans enzymatically from the protein backbone with PNGase F and subsequently derivatize with fluorescent 2-AB, 2-AA, derivatives hereof or RapiFluor and subsequent analysis by iontrap LC-MS. In this chapter the development of the HPLC method and the setup of the iontrap MS will be described.

Ion trap mass spectrometry

As an iontrap MS is not a high resolution instrument setting the resolution as high as possible for the instrument is important. Iontrap instruments are also known to have space-charge problems and both resolution and the space charge effect was investigated.

Resolution

The linear iontrap is a fast and versatile mass analyzer when scanning over a 1500 m/z range as often done in N-glycan analysis. Choosing resolution mode can either greatly improve the accuracy of the ion trap mass spectrometer or increase sensitivity but not both. Running the Thermo Iontrap instrument in higher-resolution-modes, “Enhanced” and “Zoom” mode, the relative signal is much lower compared to “normal” resolution mode. This sensitivity reduction renders the high resolution function unrealistic for glycan analysis where the target is in very small quantities. Only metal ion adducts, in this case K⁺ adducts, seems to be stable enough to cope with the longer time in the trap needed for the improved resolution in high resolution mode (data not shown).

For some unknown reason there is a higher response if the gradient slope is flat compared to a steeper slope. In theory, the steeper the slope of the gradient, the sharper the peak and...
thus more ions per second to be detected. This should result in higher signal if the mass analyzer can cope with the ions without loss or space-charge problem. As the manual annotation of glycan masses rely on averaging over 50-100 scans this effect could result as an artifact of the statistical analysis with the narrower peaks giving a smaller sample size resulting in a higher standard deviation and hence a loss of signal to noise. A possible secondary affect is a kind of matrix effect that disappears when looking at a sum of multiple scans.

**Space-Charge problem with ion-trap MS during analysis**

The possible problem with space-charge on iontrap mass spectrometry occurs if too many ions are introduced in to the iontrap at the same time. As like charged ions repel each other and there is not infinite space in the focused beam of the iontrap. The iontrap collects a number of ions in the trap before pushing them to the detector in “packages” dependent on timing of scan rates in the settings. Even if the package size-settings are set to “automatic” and thus controlled by the software/instrument to avoid too many ions in the trap at any given time, we wanted to investigate if this would be an issue with fast gradients where the analytes elute in very sharp peaks and many ions are sent to the iontrap at the same time.

To test if the space-charge problem also occurs in this LC-MS setup, we made an infusion of a high concentration samples of N-glycans from Rituximab. This was also tested with a smaller scan range to see if that could improve the signal by removing the background ions (data not shown). In Figure 10 the extracted ion chromatograms of the two runs show that with a 0.7 mL/min fast gradient (A) compared to 0.5 mL/min (B), with the same amount of sample injected. The gradient used in (B) is similar to the one showed in Figure 11. The glycan elutes earlier, as expected with the higher flow, the peak was sharper and the area of the peak was only slightly higher. The similar, increased area of the peak suggests that no signal was lost due to the space charge problem, possibly because of Thermos dual-pressure ion trap cell in the Velos model.
Notice that the signal to noise ratio was better at a flow rate of 0.5 mL/min compared to the fast gradient with S/N: 605 and S/N: 650, respectively. It was also clear that the baseline is a lot less noisy in the 0.5 mL/min flow conditions.

The conclusion was that under normal run conditions there should be no time-space problem with the ion-trap under these conditions with an amount of ions in this magnitude. Source setting were not optimized for the high flow rate to keep the setup comparable and simple. Therefore, it is possible that better signal and better signal to noise could have been achieved if the source settings had been optimized specifically for the high flow rate.
**Figure 10:** Extracted ion chromatograms of a glycan moiety under two different run conditions. A: 0.7 mL/min flow rate and B: 0.5 mL/min flow rate
Choosing the right HPLC gradient for LC-MS analysis

An HPLC gradient is always, a tradeoff between speed, separation of analytes, flow rates, and versatility. Generally the gradient used for N-glycan analysis here consists of a short initial slope to elute impurities separately from the actual glycans and allow the analytes to settle on the head of the column. In the end of the initial slope the very smallest of the N-glycans elute at around 8 min. The longer separation slope starts at 8 min. and in this segment separation of most N-glycans are achieved. Next segment is the cleaning, with a high water ratio any potential polar compounds (including glycans) are washed off the column. The high water ratio in this segment also reduces the risk of carry-over between samples. A subsequent re-equilibration segment is the last step before next injection. However, I have never observed any other compounds than glycans eluting in the cleaning step in either the fluorescence- or MS data.

For some of these gradients the flow rate starts higher as the high percentage solvent lowers the backpressure. During the high water content especially during the washing step the flow rate is reduced to avoid overpressure of the system.

Example of HPLC-MS method

The N-glycan analysis was performed on an Ultimate 3000 UHPLC system equipped with RS Fluorescence detector. The HPLC system is connected online to a Thermo Fisher Velos pro iontrap with electrospray ionization source (all Thermo Scientific).

Up to 40µL (full loop) labeled N-glycans were loaded onto an HILIC-type column (Waters Acquity UPLC Glycan BEH Amide 130 Å, Waters, 2.1 mm × 150 mm, 1.7 µm, Waters Corp.).

Fluorescence detector conditions for 2-AB labeled N-glycans: Flow cell temperature 50° C, Excitation 360 nm, emission 428 nm, sensitivity 6, data acquisition at 5 Hz and response time 1 seconds, data collected in one segment with a start delay of 4 min from injection.
Fluorescence detector conditions for RapiFluor labeled N-glycans: Flow cell temperature 50°C, Excitation 265 nm, emission 425 nm, sensitivity 6, data acquisition at 5 Hz and response time 2 seconds, data collected in one segment with a start delay of 7 min from injection.

Mass spectrometer conditions were: Capillary temperature 275 °C, Sheath gas flow rate 38, Aux gas flow rate 14, Sweep gas flow rate 1, Spray voltage 3.5 kV, S-Lens RF level 60%. Mass range 500–2000 m/z. Mass range setting: Normal, Polarity: Positive, no source fragmentation (0V). Data collected in one segment with a start delay of 3 min from injection.

For 2-AB labeled N-glycans the following gradient is a useful generic gradient (Figure 11):

A: Acetonitrile, B: 50mM Ammonium formate pH 4.4, General flow rate 0.5 mL/min, 1 min 31% B, 8 min 39% B, 24 min 47% B, 25 min 55% B, 25.5 min 55% B, 26 min 31% B, 30 min 31% B.

![Figure 11: Relatively simple HPLC gradient for 2-AB labeled N-glycans](image)

Notice the two-slope separation gradient, as there are fewer of the small glycan species they can be separated with a steeper slope.

For RapiFluor labeled N-glycans the following gradient is a useful generic gradient (Figure 12 top):

A: Acetonitrile, B: 50mM Ammonium formate pH 4.4, Starting flow rate 0.55 mL/min, 0 min 25% B, 8,5 min 25% B, , 9 min 27% B, 38 min 43% B flow rate 0.5 mL/min, 39 min 75% B flow
rate 0.3 mL/min, 40 min 75% B flow rate 0.25 mL/min, 41 min 25% B flow rate 0.25 mL/min, 46 min 25% B flow rate 0.55 mL/min.

The flow rates in this method have been adjusted so there are no sudden pressure spikes and the pressure curve is relatively flat even as the gradient increases the aqueous buffer at 39 min (Figure 12 bottom).

**Figure 12**: Top: Gradient for RapiFluor labeled glycans. Generic gradient that fits most separations. Bottom: Pressure during gradient
Specialized for separating smaller, RapiFluor labeled, N-glycans, the initial slope needs to have a lower starting point (Figure 13):

A: Acetonitrile, B: 50mM Ammonium formate pH 4.4, Starting flow rate 0.55 mL/min, 0 min 18% B, 38 min 43% B flow rate 0.4 mL/min, 39 min 75% B flow rate 0.15 mL/min, 40 min 75% B flow rate 0.15 mL/min, 41 min 18% B flow rate 0.25 mL/min, 46 min 18% B flow rate 0.5 mL/min.

Figure 13: Gradient for RapiFluor labeled glycans. Specialized to include separation of small glycans.
Annotation of N-glycans

The N-glycans analysis platform set up for the present study is based on released, derivatized N-glycans analyzed with LC-MS. However, the specific annotation of N-glycan species is not trivial. Because of the complexity of the N-glycan molecules with respect to branching and composition of sugars many pieces of information combined can give a solid annotation of a glycan species (Figure 14).

![Diagram](image)

Figure 14: Combinatorial approach to how multiple pieces of information links to the final annotation or strengthen the confidence in the annotation

The spiking of known glycan species (standards) could be needed as well as exoglycosidase treatment\textsuperscript{116,117} to gain better understanding of branching. This approach is hampered by the lack of specific enzymes that only cleave one bond and one bond only, as many of the exoglycosidases cleave multiple branching variants. Another problem is that exoglycosidases are not always branch specific, so only the linkage will be known, but not what branch it was attached to. It can in theory be semi-automated as several laboratories have described\textsuperscript{117–119} but more often the sample is split and exoglycosidase digestion is performed in 96 well format or similar.
Permethylation is a way to achieve information on branching but it requires a skilled analytical chemist and often quite a lot of time to resolve even simple structures. On top of these problems the mass spectrometer data is not always trivial to annotate to a specific mass. As N-glycans carry no easily charged moieties, they are detected as a multitude of adducts that change from day to day and differ from small to larger N-glycans. As examples in figure 15 and 18 shows that the mass spectrum of a single N-glycan-species split over a multitude of peaks with different adducts. Notice in figure 15 that the mass seems about 0.5 m/z off, but this is because the sugar structure contains so many C13 atoms that the major peak is the second isotope peak. The major peak is also not always mono-charged with a proton [M+H]⁺ but often only seen as di-charged [M+2H] ++ or triply-charged [M+3H] ++++. The main adduct is also unpredictable as it is either [M+H]⁺, [M+NH₃]⁺, [M+Na]⁺, [M+K]⁺ or the di-charged equivalents as [M+H+NH₃] ++, [M+2Na] ++ or other combinations. This makes automated peak annotation difficult unless the software is able to recognize both charge state, adduct from the pattern and match the retention time. Such a software combining these features is not known yet even though many exist that can perform some of these functionalities but rarely combined and often limited to specific vendors, high resolution accurate mass instruments or complicated workflow.
Figure 15: Two examples of analyte adduct pattern. Top: Example of MS data of a specific N-glycan structure with the theoretical monoisotopic mass [M+2H]^{++} = 1305.0045 m/z including the RapiFluor label. Bottom: Example of MS data of a specific N-glycan structure with the theoretical monoisotopic mass [M+2H]^{++} = 1097.9169 m/z including the RapiFluor label.

See also Chapter 3: Branching detail with a case study on annotating a branching variant.
Chapter 2: Publication I: Fluorine modifications to 2-AA improves mass spectrometer signal of labeled N-Glycans

Submitted paper. Full paper is attached in appendix A.

Sensitivity, for released N-glycans labeled with 2-AA or 2-AB, analyzed with mass spectrometry is relatively low. To improve this, we looked into how to improve the ionization and thus the sensitivity in N-glycan analysis with released labeled N-glycans. For 2-AA and 2-AB we achieved the best signal to noise results in negative ion mode mass spectrometry. The idea was to add one or more halogen atoms that are known to improve ionization in negative mode mass spectrometry, while not disturbing the conjugated double bond system in the compound that contributes with the fluorescent properties. In figure 16 the labeling compounds used in this thesis are shown.
From figure 16 it is also clear that the RapiFluor compound (Figure 16, D) published by Lauber et al. in 2015\textsuperscript{110}, contains a much larger conjugated system than either of the three 2-AA based compounds and as a result is much more fluorescent. The tertiary amine in the
RapiFluor compound is very good for ionization in positive mode making the RapiFluor both very fluorescent and with good mass spectrometry ionization properties.

Using the 4-TFB as a label for N-glycan analysis proved successful. In figure 17 the data shows a factor 2-3 increase in signal, compared to 2-AA and 2-AB, for a single N-glycan moiety as the high mannose structure M6.

![Figure 17: Bar chart of the summed peak area from the extracted ion chromatogram for the M6 moiety released from RNase B. (Picture and figure text is from submitted paper in appendix A)](image)

The fluorescence signal for the 4-TFB labeled N-glycans did however go down compared to 2-AB and even to be a factor of 2 lower compared to 2-AA even when the conjugated system of 2-AA should be intact (Figure 19).
**Figure 18:** Detection (fluorescence) of glycan moieties released from RNase B. Isomers are collapsed into one peak for simplicity. Peak areas are represented here as bars. (Figure and figure text from publication in Appendix A)

Overall the 4-TFB compound have improved ionization properties compared to both the 2-AB and 2-AA in negative mode MS but at the cost of lower fluorescence properties compared to 2-AA.
Chapter 3: Glycans in product quality

Biopharmaceutical companies spend a considerable amount of work, and multiple orthogonal methods, in assuring that their products fall within the pre-defined ranges, originally approved by licensing bodies such as FDA and EMA. For glycans on biopharmaceuticals there is an inherent problem in reproducing the pattern or fingerprint of complex glycans of recombinant glycoproteins from batch to batch. Especially for biosimilars that have to match the original profile of glycans from the originally registered substance. With the complexity of producing biopharmaceuticals, even the smallest changes in media, pH, dissolved O₂ or many other factors can have marked effects on the glycoprofile. The FDA and EMA use consistency in glycoprofile as an indirect measure of the quality of process control in production. This all leads to the industry wanting a better understanding and control over the glycosylation.

Product quality and controlling glycosylation

In CHO cells the glycosylation profile of a recombinant glycoprotein is generally very heterogeneous with respect to occupancy and structural variation. IgGs are one of the few exceptions with low heterogeneity as it in CHO normally only produces 4-5 major N-glycan structures, because of steric hindrance. As the glycosylation is closely related to efficacy the product, quality is linked to the glycosylation. The possibility to design the glycosylation profile is therefore considered quality by design as it is built into the cell line and thus moved from bioprocess to strain engineering. For the pharmaceutical industry the challenge has previously been to control this heterogeneity and to push the glycoprofile in a desired direction. To ensure consistency from batch to batch it is a requirement to have a stable bioprocess. The bioprocess must also be designed to maximize the desired biological properties including glycans. The tools here have been media additions, modeling and process parameters.
Recently the development of biosimilars has been of great interest to the pharmaceutical industry as several of the IgG blockbusters have gone off patent. There is therefore a big interest in making generics of these, but because of the complexity of a big molecule as IgG including the heterogeneity of the N-glycan profile the new term “Biosimilar” has been introduced as a generic IgG can never be truly identical as the originally registered drug (originator) which itself varies from batch to batch\textsuperscript{134}.

“\textit{the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components},” FDA, Section 351 of the PHS Act

For biosimilars analytical methods are of exceptional importance as the similarity has to be proven by rigorous structural and functional comparison between originator and biosimilar. In Chapter 5, 6 and 7 examples of glyco-engineering are shown in detail.

**Changes in glycosylation profiles over time in a batch- and fed-batch bioreactor**

As glycans, especially the degree of sialylation of the glycans are often important for product quality as described earlier. We know that the glycan profile changes if the cells are not grown under optimal conditions but the extend was unknown. We therefore investigated the extent of the changes in glycosylation during a batch- and fed-batch bioreactor-run, respectively. Only the fully sialylated glycan moieties are monitored as quantitative areas could not be achieved for the other structures because of overlapping peaks.

From the data of the monitored glycans in the batch run (Figure 19) the glycan profile of the seed train at day 0 (P0) seems to match the pattern of day 5 of the experiment. This is to be expected as the seed train has been growing for about four days in a shakeflask prior to seeding the bioreactor. Whereas the quality of the seed train for the fed batch experiment was of poorer quality with many smaller N-glycan structures, the reason for this is not known.
From the data it is can be seen that some glycan moieties vary a factor of two during both batch- and fed-batch bioreactor-runs. Excluding the P0 time point that shows the seed train glycans and not the glycans produced in the actual bioreactor, there is a clear tendency towards larger glycans during the run. The profile with the largest structures can be found at P7 in the fed batch while the cells are in the start of the stationary phase. The fed batch outperforms the batch in glycan size and sialic acid content early on and are still good at P12.

In conclusion we have demonstrated the variance of N-glycan moieties over time. As it can be important for a biopharmaceutical product to have a specific N-glycan profile it is here demonstrated how important the time of harvest can be. As a result of this we standardized
how the seeding and the time for harvest was performed to ensure data could be compared from experiment to experiment.

Branching details

Even though CHO cells are the FDA/EMA preferred cell line because its reliable human-like glycosylation pattern, a smaller issue with CHO cells is that they sometimes make unusual branching. NeuSGC or Alpha-gal are examples of unwanted unusual branching or sugar-monomer substitution. These two are rarely produced in detectable amounts in CHO cells but sometimes Gal-β1,3-GlcNAc branching can occur according to the GlyTouCan database.

In our data we have a case of a glyco-engineered cell line produces a peak that correspond to the mass of A2G2 structure was split into two peaks eluting very close to each other, (Figure 20).

Figure 20: Unexpected shouldering peak on a simple A2G2 structure, shoulder of peak of interest marked by red. Mass/EIC is the same for main peak and the shoulder-peak.
Other cell lines with the same structure in the same batch run did not have this shouldering peak (data not shown). Extracted-ion chromatogram (XIC) data shows that the mass of the shoulder was the same as for the main peak. Both major peaks relate to the same glycan species with the theoretical monoisotopic mass of 1951.77, as they are both doubly charged ions with two protons and a proton and a potassium ion respectively (Figure 21).

![Figure 21: Mass spectrum of the possible A2G2 structure with the theoretical monoisotopic mass [M+2H]** = 976.8907 including the RapiFluor label.](image)

Annotating the main peak, including branching, can be done via the mass, the retention time and the biology of this cell line. For CHO cells the branching is well defined in the literature where CHO cells produce $\beta_{1,2}$-GlcNAc on the A2G2 glycan under normal growth conditions. A basic assumption is that the core region GlcNAc$_2$Man$_3$ is conserved being controlled by oligosaccharyl transferase in ER. This leaves the possibilities that either at least one expected sugar monomer has been exchanged with another isobaric sugar or that the branching of either the GlcNAc or the Galactose is not as it is normally. An exchange of sugar monomers is very rarely seen in normal conditions with serum free feed so that is less likely. The branching under normal growth conditions should correspond to the first of the three structures (Figure 22) with $\beta_{1,2}$-GlcNAc and $\beta_{1,4}$-Galactoses bound to the core GlcNAc$_2$Man$_3$. 
Figure 22: Three of the possible branching variants of a A2G2 structure, all with the same mass \([M+2H]^{2+} = 976.8907\) including the RapiFluor label.

Many more branched structures are theoretically possible, as every sugar moiety in theory can be orientated \(\alpha\) or \(\beta\) and the linkage can in theory be on any of the six carbons. As there are four sugars added to the core GlcNAc\(_3\)Man\(_3\) with six positions each with the \(\alpha\) or \(\beta\) orientation possible. However, most of these theoretical branching forms have never been reported, e.g. from the literature the GlcNAc usually occurs in \(\beta1,2\), \(\beta1,4\) or \(\beta1,6\) and the galactose can be \(\beta1,4\) or \(\beta1,3\) configuration\(^{35,135}\), making only a subset of combinations possible.

The GlycoStore/GlycoBase public database\(^{35}\) have a confirmed structure where one of the two galactoses is \(\beta1,3\) linked instead of the normal \(\beta1,4\) (Table 2). The gly cose unit (GU) retention time marker-value is listed to 7.101 and 7.133 respectively. Slightly more than the 0.02 GU value that separates the two resolved peaks M5 from A2G1 with retention times of 15.08 min and 15.37 min respectively. This leads to that the 0.032 GU units corresponds to
about 0.4 min a little more than seen here. Overall is the $\beta_1,3$ linked galactose instead of the normal $\beta_4$ linked a possible annotation for this shouldering peak.

**Table 2:** Data from Campbell et al.\textsuperscript{35} Showing a potential hit for the unknown structure

<table>
<thead>
<tr>
<th>Glycan Name</th>
<th>GU value</th>
<th>Formula</th>
<th>Monoisotopic Mass (No Label)</th>
<th>RF Neutral</th>
<th>RF [M+H]+</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2G(4)1Ga(3)1</td>
<td>7.101</td>
<td>C62H104N4O46</td>
<td>1640.592</td>
<td>1951.767</td>
<td>1952.775</td>
</tr>
<tr>
<td>A2G(4)2</td>
<td>7.133</td>
<td>C62H104N4O46</td>
<td>1640.592</td>
<td>1951.767</td>
<td>1952.775</td>
</tr>
</tbody>
</table>
Variance in glycosylation profile between CHO strains

The N-glycosylation profiles of these three widely used industrial CHO cell lines vary, where K1 is the oldest, and possibly closest to the original CHO cell line. In figure 23 the secretome N-glycan profiles of CHO K1 vs DG44 vs CHO-S is shown.

**Figure 23:** Annotated fluorescence trace of secreted proteins from CHO-S WT, DG44 WT and CHO-K1 WT cell lines. Glycan profiles are from day 4

From this data it is quite clear that the composition of the secretome or the native glycosylation machinery of these three cell lines have changed from the ancestral. CHO-S
WT has higher level of elongated and tetra-antennary structures and CHO-K1 WT has a higher abundance of terminal sialylation. DG44 WT cells are in between the two. CHO-K1 is the oldest of the three but may have many more passages than the other two so the effect is probably more from selection than from number of passages. Nevertheless as the passage numbers for these are not known no clear conclusion can be made. It is interesting to note that the profiles differ this much between non-engineered CHO cell lines. There is a better annotated CHO-S WT (Figure 24).

**CHO-S WT heterogeneous N-Glycan profile**

The CHO-S cell line is used because it has the capacity to make a wide variety of mammalian type N-glycans. When using a pool of proteins, as shown here it is easy to see the wide heterogeneity of possible structures from FM3 to a doubly LacNAc elongated FA6G6S4 structure. If a purified glycoprotein like EPO was shown only the bigger N-glycans would be seen and showing a purified IgG would show only the smaller structures. Here the pool of CHO secreted proteins are shown - 26 N-glycans structures were assigned whereas eight of them have multiple branching variants, capable of separation in the LC-gradient, bringing the total number of assigned N-glycan species to 35 (Figure 24).
Figure 24: Annotated fluorescence trace of secreted proteins from CHO-S WT cell line.

With this wide capacity for N-glycosylating proteins it makes sense that CHO cells are a good production host for glycosylated recombinant glycoproteins.

About ADCC, IgG and Fc-receptor

As discussed previously the specific glycosylation state of an antibody is critical for its functionality. The Fcγ-receptor binds to the FC-stem region of the antibody and thus recruiting an effector cell to lyse the target cell (Figure 25). The ADCC is triggered by the binding of Fcγ receptors, on natural killer cells, to the antibody stem Fc region\textsuperscript{136,137}. The binding is sterically hindered by a core α1,6 fucose often present on the antibody stem. CHO cells are almost entirely core fucosylated on IgG’s suggesting a glycoengineered
afucosylated CHO cell line is desirable. It is well established that therapeutic antibodies without core fucose show more potency than their fucosylated counterparts\textsuperscript{57,138}. More specifically the removal of core fucose, causes the ADCC enhancement through improved FcGammaRIIIa binding thus causing the improved ADCC. The pharmaceutical industry have as a consequence modified the existing cell lines producing the new products or simply changed into new cell lines producing afucosylated N-glycans. Now more than 2/3 of the glycoengineered cells listed in Pereira \textit{et al.}\textsuperscript{78} are marked as afucosylated.

\textbf{Figure 25.} Antibody dependent cell mediated cytotoxicity. Figure is adapted from\textsuperscript{139}.

The most well reported and important example of an engineered anti-cancer drug, IgG antibodies such as Rituximab and Herceptin lacking core fucose have enhanced ADCC both \textit{in vitro} and \textit{in vivo}, an attribute desirable for a therapeutic mAb\textsuperscript{28,57}. This is not the case for a conjugated mAb, where the Fc region is optional and sometimes even removed. Non-fucosylated IgG’s can be generated by knocking down expression of GDP-fucose 4,6-dehydratase\textsuperscript{140}. Alternatively the FX knockout CHO cell line is capable of expressing IgGs with either fully fucosylated or fully afucosylated glycan profiles via media addition\textsuperscript{79}. Yet another variant is the KI of GDP-6-deoxy-D-lyxo-4-hexulose reductase that deflects an intermediate in the de novo synthesis of fucose to a dead-end, resulting in the production of non-fucosylated IgGs\textsuperscript{80}.
While the effect of the fucose in IgG is well established and seems to have the same effect for all IgGs with respect to ADCC the effect of other modifications as bisecting GlcNAc by Mgat3 is disputed and may be confused by the effect of lower fucose content in Mgat3 KI cell lines Umaña et al. writes that:

>>>We show that by producing chCE7 with increasing amounts of bisected complex oligosaccharides, we can obtain significant ADCC activity<<<\(^{141}\)

Showing that with Mgat3 KI in CHO cells the ADCC activity was improved. Whereas Shinkawa et al. concludes the opposite\(^{138}\), that Mgat3 KI would lower the ADCC activity. However, with the KI of Mgat3 the N-glycan becomes first modified by Mgat3 glycosyltransferase and the resulting N-glycan with bisecting GlcNAc can no longer be a biosynthetic substrates for the fucosyltransferase. So the reported ADCC effect for bisecting GlcNAc could be a direct result of the lower fucose content, not the bisecting GlcNAc.

Similarly to the importance of the IgG glycoforms, the antibody binding levels decreased for deglycosylated Fc receptors, suggesting that not only is the N-glycan of the IgG important but also the N-glycan on the receptor is important for binding the IgG to the receptor\(^{142}\).

Since 2003 a range of publications demonstrating very impressive data on yeast as a production host for recombinant proteins\(^{143}\) producing IgG\(^{36,119,120}\) and EPO\(^{73,91}\) as biopharmaceuticals. Gerngross et al. (2006) showed that a yeast-produced homogeneously N-glycosylated form of the IgG Rituximab is up to 10-fold more active in binding its receptor than the Rituximab drug\(^{72}\) as it does not contain the core fucose, see Chapter 3. A hypothesis for the lack of any yeast produced drugs on the market are that small quantities of yeast type N- or O-glycans are still present and the final product must contain much less than 0.2% before it can be used in humans. This results in extensive purification and the
following loss of product before the needed product quality is achieved and clinical trials can be initiated.
Chapter 4: Publication II: Baicalein Reduces Oxidative Stress in CHO Cell Cultures and Improves Recombinant Antibody Productivity

Full paper is attached in appendix A

Media additions can improve productivity of CHO cells

Figures and data from Ha et al.\textsuperscript{146}

Media compositions can be used to either change IgG N-glycosylation pattern as demonstrated in Kildegaard \textit{et al.}\textsuperscript{129} or in Zhang \textit{et al.}\textsuperscript{147} Media addition can aid the production or productivity of a glycoprotein. Here in figure 26 and 27, an example is shown on how reduced stress improves titer of a recombinant antibody without affecting the glycosylation profile, by the addition of Baicalein to the media.

\textbf{Figure 26: Profiles of mAb concentration of a CHO-CS13-1.00 cell line producing mAb in six-well plated cultures with various chemical reagents. No chemical reagent (closed circle), 50mM BHA (closed square), 10mM NAC (closed triangle), 10mM berberine chloride (closed diamond), 100mM baicalein (open square), 10mM kaempferol (open triangle), and 10mM apigenin (open diamond). Figure and figure text is from Ha et al.\textsuperscript{146}}
Figure 27: Profiles of N-linked glycosylation of a rCHO cell line producing mAb during shake flask cultures with baicalein addition. Culture supernatants were harvested on days 6 and 9. No baicalein on day 6 (white), No baicalein on day 9 (dashed white), DMSO on day 6 (light gray), DMSO on day 9 (dashed light gray), baicalein on day 6 (gray), and baicalein on day 9 (dashed gray). G0, G1 and G2 represent G0+G0F, G1F+G1S1F, and G2F+G2S1F+G2S2F respectively. The error bars represent the standard deviations calculated from two independent experiments. *P<0.05 (Figure and figure text is from Ha et al.146)

Not only can the media compositions change the glycan profile, other parameters that cause stress to the cells could be too low oxygen levels or physical shear on the cells from the stirring and affect the glycan profile. In this case of an upscaling attempt from 15 mL scale to 250 mL scale, the cells had lower viability during the exponential growth phase than expected 92-95% (Figure 28, B) versus 97-98% in the smaller scale (Figure 28, A). The N-glycan profile is much more heterogeneous than expected at harvest, just under 80% viability. The presence of 5% of the 5-mannose structure furthermore suggests that the cells are stressed or that dead cells have lysed and released unprocessed protein from the ER. The cause of the stress in this case is unknown, but the N-glycan profile is clearly affected (See Figure 28).
Figure 28: Annotated fluorescence trace of N-glycans from AAT derived from CHO-S cell line producing AAT.
Chapter 5: Glycoengineering

To tailor-make a given N-glycoprofile it is necessary to trim down the WT N-glycoprofile by the combinatorial KO of the glycosyltransferases making un-wanted additions to the target glycoform. Searching the literature on mammalian glycosyltransferases gives the overview seen in Figure 29. From the figure it is clear that some glycosyltransferases are very specific as the Mgat family or Fut8 whereas the B4galt and the St3gal families are less specific and the transferases can add the sugar to any of the branches.
Figure 29: Overview of glycosyltransferase gene names and the glycosyl bond it makes.

Table 3: List of important glycosyltransferase and related genes in CHO and humans, genes in parenthesis are genes with little or disputed effect

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mgat1</td>
<td>First antennae, β1,2 GlcNAc added to α1,3 mannose</td>
</tr>
<tr>
<td>Mgat2</td>
<td>Second antennae, β1,2 GlcNAc added to α1,6 mannose</td>
</tr>
<tr>
<td>Mgat3</td>
<td>Bisecting antennae, β1,4 added to first mannose, Not naturally active in CHO or mouse</td>
</tr>
<tr>
<td>Mgat 4A/4B/(4C)</td>
<td>Third antennae, β1,4 GlcNAc added to α1,3 mannose</td>
</tr>
<tr>
<td>Mgat 5/(5B):</td>
<td>Fourth antennae, β1,6 GlcNAc added to α1,6 mannose</td>
</tr>
</tbody>
</table>
**Mgat N-glycan engineering in CHO cells**

Figure 30 shows the glycan profile produced by a CHO KO cell line where MGAT4A, MGAT4B and MGAT5 have been knocked out. The following is the procedure that leads to the verification of the knock outs.

The DNA analysis from MiSeq was performed twice and the first round was used for picking the clones. The data here in table 4 shows that there is a possibility of an in-frame -3 base pair shift in the MGAT4A gene. However the fraction is small so it is possibly an artifact from the analysis and thus a false positive. Later this false positive was verified by a second round of sequencing that shows that both the +10 and the -3 indel was artifacts. Note that the 2’ round is usually done after the clone have been analyzed for glycoforms, and thus have the N-glycan-phenotype established.

**Table 4: MiSeq data**

<table>
<thead>
<tr>
<th>Knock Out Genes</th>
<th>1’st round, Picking</th>
<th>2’ Round, Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGAT4A</td>
<td>-1/10/-3</td>
<td>84/9.6/5</td>
</tr>
<tr>
<td>MGAT4B</td>
<td>1</td>
<td>86.9</td>
</tr>
<tr>
<td>MGAT5</td>
<td>1</td>
<td>94.6</td>
</tr>
</tbody>
</table>
The first positive sequence result leads to picking of clones and subsequent N-glycan analysis to prove the glycan phenotype. To get useful glycan analysis the cells must be above 90% viable or there will be too many impurities to annotate the glycan data properly. In this case impurities can be seen between 10-12 minutes of figure 30 marked with a question mark. Notice that peaks eluting later than the main peak at 24 minutes are all assigned to LacNAc-dimer elongated structures as B3gnt2 has not been knocked out in the particular cell line. How the elongation is attached is not known and the structures showed here are not confirmed. According to the literature\textsuperscript{148} the preferred branch is $\beta1,6\text{GlcNAc-}\alpha1,6\text{Man}$ but as that branch is not present in the KO cell line as a substrate, the annotation is speculative.
The annotations of the large N-glycans in figure 30 is also an example of how the analysis depend on knowledge of the biology of the cell line including genes that have been knocked out. In this specific case the lacNAc elongations would have been incorrectly annotated as antennas as the retention time shift would probably have been unnoticed. It is also an example of when the natural substrate is no longer present unexpected structures may appear. In the case when the natural substrate for LacNAc elongations are gone, the transferase is clearly able to add the LacNAc to a bi-antennary substrate, although the branching specificity is now unknown. As a result, new bi-antennary structures with LacNAc elongations cannot be annotated branch specific and the elongation on the bi-antennary structure can be on either branch or be distributed on both branches. In this case Tandem MS would be a good tool to solve the structure by looking for a poly-LacNAc fragment that could solve the question of distribution.
B4galt glycan engineering

To make agalactosylated glycans, KO of the β-1,4-Galactosyltransferase genes B4galt1, B4galt2 and B4galt3 produces almost entirely agalactosylated glycans. B4galt1 are the most active followed by B4galt3 and in some cases B4galt4 as well. In figure 31 secretome data from a triple KO of B4galt1, B4galt2 and B4galt3 is shown. First of all it is clear that only three structures, five isomers in total, are annotated as having a galactose moiety. Logically the annotated A2FG1S1 structure probably also has an isomer in the small un-annotated peaks in the 18-20 min retention time area, but it could not be detected/confirmed by mass spectrometry.

These combinatorically knocked out B4galt cell lines are also elaborated in the Amann et al. paper\textsuperscript{29} and in Chapter 6.
Intriguingly when knocking out these three B4galt genes the resulting N-glycan profile have some unusual branching variants marked by asterisks. The A2F structure splits into three isomers separated by the LC, marked with one asterisks, and all with the same mass corresponding to a A2F structure. Biologically it should only produce Mgat1 and Mgat2 β-1,2 branching on both of the α-1,3 and α-1,6 mannose, see first structure of Figure 32.
Figure 32: Three of the possible branching variants of a A2F structure, all with the same mass \([M+2H]^{++} = 887.8668\) including the RapiFluor label.

As it splits into three isomers the hypothesis is that without the galactoses the Mgat4a, Mgat4b and Mgat5, N-Acetylglucosaminyltransferases can produce the non-standard branched structures shown in Figure 33, that it would not normally be seen in wild type systems. It is likely that this results from changes in the enzyme kinetics.

The same for the A3F structure that splits into three isomers separated by the LC, marked with two asterisks, and all with the same mass corresponding to a A3F structure. Here the potential biology behind the three isomers is not as clear. The two first structures could hypothetically be the branching variants from Mgat1+2+4a+4b or Mgat1+2+5, see Figure 33. However the third isomer requires the glycosyltransferase to either add the GlcNAc monomer to the wrong branch resulting in the third structure in figure 33 or adding the GlcNAc in another unknown position. The fourth structure in figure 33 is a predicted database hit in both Center for Functional Glycomics and GlycomeDB- databases.
The structure includes the possibility of the GlcNAc to be attached bisecting to the first mannose as well as attached to any of the two GlcNAc monomers. The bisecting GlcNAc from the first mannose would normally be the product of Mgat3 glycosyltransferase but that gene is silenced in the CHO-S cell line where this sample originates.

*Figure 33:* Four of the possible branching variants of a A3F structure, all with the same mass \([M+2H]^{++} = 989.4065\) including the RapiFluor label.
In table 5, a list of glyco-engineered cell lines with distinctly different profiles is shown. Using the literature and later the knowledge from Yang et al. we have created a panel of 26 cell lines that produce a wide range of N-glycan profiles. Creating the panel depends on the combinatorial KO of the following glycosyltransferase genes: Mgat1, Mgat2, Mgat4A/4B/(4C), Mgat 5/(5B), B4galt1/2/3/4, B3gnt2, St3Gal(1)/(2)/(3)/4/(5)/6, St6gal1/2 and Fut8 (N-glycan data not shown).

**Table 5: Glyco-engineered cell lines. X denotes that the cell line is created and N-glycan phenotype is verified. ND = Not done.**

<table>
<thead>
<tr>
<th></th>
<th>Mono-antennary</th>
<th>Bi-antennary</th>
<th>Tri-antennary</th>
<th>Tetra-antennary</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>No Core Fucose</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>No Sialic acids</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>No Galactoses</td>
<td>ND</td>
<td>X</td>
<td>ND</td>
<td>X</td>
</tr>
<tr>
<td>No Fucose/sialic acids</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>No Fucose/galactoses</td>
<td>X</td>
<td>X</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Human sialic acids (St6gal1 KI)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>(X) Not verified</td>
</tr>
<tr>
<td>No Fucose + Human sialic acids (St6gal1 KI)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>(X) Not complete</td>
</tr>
</tbody>
</table>
A major change in the N-glycan profile would be to engineer a cell line that contained prevalently glycosylates proteins with agalactosylated structures. Based on the works of Zhang et al.\textsuperscript{26} we wanted to study the combinatorial effect of the B4galt genes. The hope was to be able to also identify if any of the B4galt genes had preferences for a specific branch or if they did not have any branch specificity. The data showing the effect on secretome of the combinatorial KO can be found in Figure 34. In summary B4galt1 and B4galt3 double KO makes about 70% agalactosylated N-glycan structures and the further KO of B4galt2 only adds 10% agalactosylated N-glycan structures to a total of \textasciitilde80%. This is however with the cost of more unwanted high mannose structures.
Figure 34: Secretome N-glycan profile of generated B4Gal-T-KO clones. N-glycan secretome analysis from batch cultivation of parental cell lines and KO cell lines harvested after 5 days of cultivation and normalized to area under the curve of total agalactosylated (G0), mono-galactosylated (G1), bi-galactosylated (G2), tri-galactosylated (G3), tetra-galactosylated (G4), and high-mannose (HM) N-glycan peaks per cell line. Increase of G0-proportion is given in % after additional B4Gal-T2-KO in T2-3-KO and T1-2-3-KO compared to T3-KO A and T1-3-KO, respectively. Where present, error bars indicate SD of three clones (T1-2-3-4 KO, T2-3-4-KO, and T1-2-KO) or four clones (T1-2-3-KO) (Figure and figure text is from Amann et al.

When producing recombinant proteins in the B4galt KO cell lines similar results are achieved compared to the secretome. In figure 36 data from cell lines with B4galt KO combinations, producing Rituximab and EPO, shows that either the B4galt1 and B4galt3 double KO or the triple KO of B4galt1, B4galt2 and B4galt3 is needed to achieve agalactosylation.
Figure 35: Rituximab and EPO N-glycosylation profiles in WT and B4Gal-T KO cell lines after transient transfection. A) Comparison of rituximab N-glycans purified out of pooled supernatants within shake flask duplicates from CHO-SWT, T3-KO A, T2-3-KO, T1-2-3-KO A, and T1-3-KO with N-glycan proportions of agalactosylated (G0), mono-galactosylated (G1), bi-galactosylated (G2), and high-mannose structures (HM) normalized to AUC of total N-glycan peaks per clone. B) Detailed N-glycan profiles of rituximab purified out of pooled supernatants within shake flask duplicates from T1-2-3-KO A (orange line) and T1-3-KO (black line) after HPLC histogram annotation via MS. C) Comparison of EPO N-glycans purified out of pooled supernatants within shake flask duplicates from CHO-S WT, T3-KO A, T2-3-KO, T1-2-3-KO A, and T1-3-KO with N-glycan proportions of agalactosylated (G0), mono- (G1), bi- (G2), tri- (G3) and greater or equal tetra-galactosylated structures ($\geq$G4) normalized to AUC of total N-glycan peaks per clone. D) Detailed N-glycan profile of EPO purified out of pooled supernatants within shake flask duplicates from T1-2-3-KO A. (Figure and figure text is from Amann et al.29)
Chapter 7: AAT CHO N-Glycan engineering

The human plasma protein Alpha 1-Antitrypsin in healthy individuals shows a very homogeneous N-glycan profile (Figure 36). Notice that the HPLC column was worn during sample A and thus the retention times cannot be compared between these two runs and peaks are much better resolved and separated in B compared to A.

![Figure 36: Annotated fluorescence trace of purified AAT from; A: Human Serum purified at DTU and B: The biopharmaceutical drug Prolastina™. Both proteins shows very high degree of homogeneity and both have the A2G2S2 structure as the main peak.](image)

This very homogeneous profile makes a good case of how glycoengineering can aid in producing next-generation-humanized biopharmaceuticals.

Producing active AAT recombinantly is feasible and several papers and patents are published on how that can be done\textsuperscript{149 150 151 152 153 154 152} and on top of that an impressive 21 attempts from various non CHO cell lines are listed in Karnaukhova et al.\textsuperscript{155}. It is important to note that active AAT can be produced even in \textit{Escherichia coli} and yeast but that serum
half-life is significantly lower for AAT from yeast compared to human serum AAT\textsuperscript{156}. This leads to the obvious question of why AAT is cleared so fast in the recombinant version? Human AAT is annotated in Uniprot database\textsuperscript{157} as only having one a few types of post translational modifications:

- Three complex type N-glycans attached to positions 70, 107 and 271
- Two Phosphoserines at position 38 and 383
- One S-cysteinyl cysteine modification at position 256

N-glycan modifications are known to have effect on half-life or in vivo activity of other biopharmaceutical drugs as seen on rhEPO\textsuperscript{158, 159} suggesting this could also be the case for rhAAT.

In industry the quality of the biopharmaceutical product often depend heavily on the product having the correct pattern of N-glycosylation:

>>The carbohydrate content (neutral sugars, amino sugars and sialic acids) should be determined. In addition, the structure of the carbohydrate chains, the oligosaccharide pattern (antennary profile), the glycosylation site(s) and occupancy should be analysed.<<\textsuperscript{160}

To ensure a homogeneous N-glycan profile it is important to harvest before the viability drops too low and the dead cells start to break and glycoprotein with un-processed N-glycans from the ER and Golgi are released. The homogeneous profile would be an example of the importance of monitoring the N-glycan profile in the end of a bioreactor run to ensure quality and homogeneity of the N-glycans. Quality is here defined as low quantities of high mannose structures and non- or partly sialylated structures. In this example N-glycans were analyzed after harvest on day 13 (28.6.2018).
The harvested product at day 13 does not show as uniform N-glycosylation probably because viability was too low and thus large numbers of dead/lysed cells are present. These lysed cells contribute with partly processed glycoforms from the ER and Golgi. The presence of high mannose structures, in this case a Man₅GlcNAc₂ at 15 min is usually a clear marker of stressed cells that does not process N-glycans correctly. The faster clearing of high mannose IgG was demonstrated by Kanda et al.¹⁶¹.
In conclusion, even if high titers can be reached in late phase of fed-batch production, the product quality will not be as good as earlier where viability is higher. Specifically in this case the AAT product harvested at 82% viability at day 11 in another bioreactor shows a very nice homogeneous N-glycan profile with predominantly one A2G2S2 structure, whereas the product harvested at 60% viability at day 13 is much less homogeneous with three other main N-glycan structures, see Figure 38.
Figure 38: Annotated fluorescence trace of N-glycans from purified AAT from two separate experiments. A is harvested at 60% viability and B is harvested at 82% viability. It is clear that harvesting a fed batch before viability drops too much under 80% is important for N-glycan homogeneity.
Cartoon visualization of the human AAT 3D structure, the fold of the AAT backbone derives from the PDB entry 1ATU X-ray crystallography at 2.70Å resolution and covering the aminoacids 45-418 (no signal peptide). The three most common N-glycans (A2G2S2) are attached at positions 70, 107 and 271, respectively, see Figure 39. This clearly shows the relative spacial size of the three glycans on AAT and they seem to take up more space than the ca. 13% of molecular weight they correspond to.

**Figure 39:** AAT protein backbone cartoon structure with A2G2S2 N-glycans at positions 70, 107 and 271. Notice that the N-glycosylation is just under 7000dalton of the 52000dalton total weight (13% is the three N-glycans). Figure is made with PyMol162
Publication IV: Glyco-engineered CHO cell lines producing alpha-1-antitrypsin and C1 esterase inhibitor with fully humanized N-glycosylation profiles

Full paper is attached in appendix A, Paper is also included in the PhD thesis by Thomas Amann.

The combination of the sensitive N-glycan analysis developed and the knowledge of the glycosyltransferases made it possible to humanize the N-glycan profile for two recombinantly produced human glycoproteins. As seen in figure 24 and 40 the WT CHO N-glycan profile is highly heterogeneous. With the combined KO of the eight glycosyltransferases Mgat4a, 4b, 5 + St3gal3, 4, 6 + B2gnt2 + Fut8 and Spil3, the profile becomes much more homogeneous (compare figure 40 B CHO-S WT with 10x KO A and 10x KO B with about 80% A2G2 moiety).
Figure 40: Growth and N-glycan structure analysis of CHO-S WT and 10× KO cell lines. (A) Viable cell density (VCD) and viability of batch cultures of CHO-S WT and two clonal cell lines (10× KO A and 10× KO B) with indels in eight glycosyltransferases as well as Glul and Sppl3. Error bars indicate the standard deviation of triplicate parallel cultures. (B) N-Glycan profiling of total secreted proteins from CHO-S WT and the 10× KO A and 10× KO B clones. In the chromatogram, elution time indicated on the x-axis and y-axis represents signal intensity normalized to highest peak. (C) Proportion of non-fucosylated, biantennary N-glycans with
terminal galactose (A2G2) in total secreted proteins from CHO-S WT and the 10× KO A and 10× KO B clones. Figure and figure text is from Amann et al.109)

Since the KO of the ten targets did not seem to interfere with cell culture performance, we could continue to co-express of ST6GAL1 and rhA1AT or ST6GAL1 and rhC1INH respectively, in the 10× KO-derived clone B. The choice of clone B was based on the slightly higher viable cell count not because of the N-glycan profile as the two clones have an almost identical N-glycan profile.

After a second round of single cell screening for FITC-SNA lectin positive clones producing rhA1AT or rhC1INH the clones produced a very homogeneous N-glycan profile from the two purified glycoproteins (Figure 41).

![Figure 41: Characterization of purified rhA1AT and rhC1INH. N-glycan structures annotated from rhA1AT and rhC1INH produced in clones A1-1 and C1-1. (Figure and figure text is adapted from Amann et al.109)](image)

The two drugs Prolastina and Cinryze are both purified from pools of human plasma, the N-glycan profile here is therefore the target for the glyco-engineering effort. For the
comparison of the N-glycan profile of our engineered version with the N-glycan profile of the two commercial drugs with the target N-glycan profile see Figure 42.

Figure 42: Characterization of purified A1AT and C1INH. N-Glycan structures annotated from rhA1AT and rhC1INH produced in clones A1-1, A1-2, C1-1 and C1-2. As well as the two commercial drugs Prolastina and Cinryze. (Figure and figure text is adapted from the supplementary information from Amann et al.109)
Chapter 8: Publication V: N-glycan engineering in *Drosophila S2* cells

(Paper will be submitted, include in part only)

Other organisms used for biopharmaceutical production can also be glyco-engineered. The *Drosophila S2* cell line is another industry relevant cell line that can be engineered to have the core α1,6 fucose removed with KO of the *FucT6* gene equivalent to the human *Fut8* gene (see Figure 43). Both genes catalyze the addition of fucose by a α1-6 linkage to the first of the core GlcNAc residues. The S2 cell line does have a GlcNAc transferase that add a GlcNAc but in WT S2 the GlcNAc is removed later by the fused lobes (*fdl*) gene.

*Figure 43: Simplified biosynthetic pathway to S2 N-glycosylation.*

Looking at the N-glycan profile of secreted host cell proteins from the S2 cell line, the most dominant N-glycan is the FM3 (Figure 44, A). By KO of the *FucT6* gene all the fucosylated structures are gone (Figure 44, B), successfully creating an afucosylated cell line.
Figure 44: Annotated fluorescence trace of N-glycans from host cell proteins from S2 cell line. A is WT and B is a knock out cell line with the FucT6 removed.

Engineering the S2 cell line to produce more humanized complex N-glycans requires that the high mannose structures should be removed and two more GlcNAc sugars should be attached to the Man$_3$GlcNAc$_2$ core to produce the A2G2 structure. As knocking out $fdl$ is not enough to make FA1 glycan (data not shown) KI of insect variants of Mgat1 (equivalent to the human Mgat1) is needed to produce the FA1 structure. For further making the FA2 glycan KI of insect variants Mgat2 (equivalent to the human Mgat2) is needed. Figure 45 shows the successful KO of $fdl$ and subsequent KI of Mgat1 and Mgat2 in cell line producing the truncated version of the placental malaria antigen (VAR2CSA), which spans from ID1 to ID2a. The retention times cannot be compared from Figure 45 A versus B as purified ID1-ID2a protein from S2 WT cell line was labelled with 2-AB and ID1-ID2a protein from S2 cell line with $\Delta fdl$ and Mgat1 and Mgat2 KI was labeled with Rapifluor.

This glycoengineering effort resulted in a cell line that increased FA2 glycan structure from 0% to >64% on the ID1 to ID2a protein.
Figure 45: Annotated fluorescence trace of N-glycans from purified ID1-ID2a protein from S2 WT cell line (A).

Purified ID1-ID2a protein from S2 cell line with Δfdl and Mgat1 and Mgat2 KI (B).
Discussion

Analyzing N-glycans

The perfect method for analyzing N-glycans derived from glyco-proteins would be on either the intact glyco-protein by native top-down glyco-proteomics or directly on released N-glycans without any labeling. This will probably happen as instrument vendors and software companies keep updating the current platforms. I hope that the future instruments and software will be able to have the sensitivity to perform the analysis of un-labeled N-glycans. The challenges are the following:

- Sensitivity
- Low false negative rate
- Structural annotation
- Low price (cost of analysis)

Current setups for analyzing N-glycans without any label are at least an order of magnitude less sensitive than if a label is applied. With a fluorescent label CE can be so sensitive that some samples need to be diluted before analysis thus exceeding the need for sensitivity. Without label the CE or HPLC would be for separation only relying on the mass spectrometer for identification. One could envision a system where for released N-glycans LC-MS or CE-MS could be used to separate the N-glycans. The software could then efficiently use the retention time to include or exclude potential N-glycan hits with a database. This requires the database to exceed 500 structurally annotated N-glycans as this is the amount of N-glycans that is expected from a range of glyco-proteins from mammalian cells. This is where current software is lacking, and the databases are still not containing enough well-defined structures to be able to do this kind of annotation unless the sample derives from an IgG or another simple glyco-protein.

Tandem MS could also be utilized to include or exclude potential N-glycan hits with a database containing fragmentation data. The potential problem here is that only electron ionization (EI) fragmentation is comparable across platforms whereas collision-induced
dissociation (CID) or Higher-energy collisional dissociation (HCD) does not necessarily give the same fragment pattern across platforms.

For the Glyco-proteomic approach the main issue is sensitivity for rare glycan moieties. With a very heterogeneous sample containing over 20 N-glycan structures on a given glycan site the mass spectrometer struggles with the signal intensity of the smaller N-glycan moieties being suppressed by the ions from the more dominant species. This could possibly be solved with 2 dimensional separation where the peptides are separated with C18 column and the glyco-peptides separated by glycan on HILIC type column. This 2D column setup could thus ensure different elution times for the glyco-peptides. The advantage is that both N- and O-glycans could be analyzed with one method, that the glycan site is verified, no bias with sterical hindrance for PNgase release and that no labeling step is required prior to analysis. Expanding this would be to do injection of the intact native glyco-protein and do top-down glyco-proteomics. The danger here is that the conditions in a proteomic setup is too harsh for the glycans thus risking peeling of the glycans especially the loss of sialic acids if the ionization is too hard.

Structural information is harder to achieve with glycoproteomics if only separated by C18 or top-down. There would be no structural information from HILIC type separation about branching variants, size or retention time helping to solve the structure. For glyco-proteomics structural information would rely on tandem MS breakdown of the glycans and thus the risk of not acquiring the fragment or cross ring fragment needed for accurate structural assignment. With no HILIC separation the difference between hexoses (mannose and galactoses) is also a problem.

Price is also a factor in glyco-proteomics. High resolution accurate mass instruments are very expensive to buy and still requires experts to run.

CE-MS is a potentially very interesting platform for the analysis of N-glycans. The very low flow rates that is an advantage in mass spectrometry as no heated source is needed as in nano spray LC-MS. The good separation of the CE is a good match to the higher sensitivity of
the mass spectrometer with the low flow rate. Historically there have been stability issues with coupled CE-MS systems that is probably the cause of their current rarity.

**Glycosylation profiles changes over time and by glycoengineering**

In our study of changes in glycosylation profiles over time in batch- and fed-batch bioreactors we found a surprisingly large variation of selected N-glycan structures over time. Some glycan moieties vary more than a factor of two during batch- and fed-batch bioreactor-runs. We know that industry monitors the glycan profile during production but we did not know the extent of the changes.

For glycoengineering it is important that some glycosyltransferases are very specific as the Mgat family or Fut8 where the transferred sugar is added very specifically to one Antennae/position/branching only. Whereas the B4galt and the St3gal families are less specific and the transferases can add the sugar to any of the antennae/branches although still in one specific orientation β1,4 and α2,3 respectively. With that we can engineer structures that are very branch specific containing defined antennae and with and without the core fucose. The next positions the galactoses and the sialic acids cannot be defined to a specific antennae and are thus only possible to engineer as present or not present.

With the study of combinations of B4galt KO’s we had hoped to be able to identify branching specificity of the individual B4galt transferases. But unfortunately, in all cases we could not identify a B4galt that had a clear branch specificity, since all transferases seem to be able to add a galactose to any of the four branches. We did however find that the B4galt transferases are more or less protein specific. KO of just B4galt1 is enough to get almost zero galactoses on a IgG and further KO of more B4galt’s does not change much. When looking at secretome or EPO a double KO of B4galt1 and B4galt3 is needed to make mostly agalactosylated N-glycan structures and the further KO of B4galt2 to a triple KO adds about 10% agalactosylated N-glycan structures. This difference from protein to protein suggests that it is not the transferases themselves that are branch specific, but that the transferases recognizes the protein or a combination of protein and glycan in the case of B4galt.
However the reason why B4gal3 is not able to act on IgG but works well with EPO and other glycoproteins is still unknown. It is worth noting that especially the triple KO contains more high mannose structures compared to the single and double KO possibly because the lack of galactoses puts stress on the cells. Surprisingly we were not able to produce entirely agalactosylated glycans even with a B4galt1, B4galt2, B4galt3, B4galt4 tetra KO.

With the glycoengineering of the cell line used for recombinant production of A1AT and C1inhibitor (Chapter 7) we achieved a very homogeneous A2G2S2 N-glycan. We had not expected to achieve such homogeneous profile as we normally see some amount of partially finished or partially degraded N-glycans in the N-glycan profile of recombinant proteins. The cause of this very homogeneous profile could be that we have chosen proteins that have a great capacity for being N-glycosylated, but as the wild type cell line producing AAT shows a wide range of N-glycan structures in the profile it must be something we engineered into the cell. On top of that it is not just clonal variation – a perfect clone as we have multiple clones with and without the Glu1 KO that all show a very homogeneous profile. This leads to the hypothesis that this homogeneous profile is a result of the nine glycosyltransferases knocked out making room in the Golgi for more of the remaining glycosyltransferases. This could increase the capacity for N-glycan processing as there are more transferases in the Golgi to process the Glycoprotein. This could be tested by re-introducing Mgat4a and Mgat4b in the cell line and see if the profile shifts to be a homogeneous A3G3S3 structure.

Even though glycans are now used as bio-markers for diseases, it is still not known if it is the cause, effect or non-related \textsuperscript{163–165}. With glyco-engineering it could be possible to make the specific glycoproteins and test these in animal models to elucidate if they are the cause of the disease or the effect. In some cases the glycan biomarker is site specific. This is a level of complexity that we have no control over in mammalian cell systems and as far as we know no-one have proposed a hypothesis of how even full organisms are able to control the micro-heterogeneity of glycans in a site specific manor. The task of taking control of the micro-heterogeneity of glycans in single cell systems is even more daunting.

From glycoengineering, as shown in this thesis, it is possible to trim down the N-glycans to the desired form with KO of selected glycosyltransferases. With overexpression and media composition it is possible to some degree to reduce the amount of smaller un-finished N-
glycans. However there are changes that are not easily performed as adding a third or fourth antennae to the N-glycans on an IgG.

The type of analysis used in this thesis relies on the release of N-glycans with PNGase F. Studies on Caenorhabditis elegans shows that much more complex core structures are possible\textsuperscript{166} and that they are only poorly or not at all released with the PNGase F used in this thesis. Therefore it is possible that small quantities of unusual N-glycans can be present in mammalian samples without these being detected, as glycoproteomics is not sensitive enough and hydrazine release is rare because of the more labor intensive preparation, peeling of sugar-residues and de-acylation that occurs.
Conclusions

A robust, sensitive and versatile platform for the analysis of N-glycans has been set up and used for a multitude of projects including the analysis of glycoproteins from Humans, mammalian (CHO), filamentous fungi (Aspergillus niger), insect cells (S2) and synthetically made glycans. The LC-MS platform is capable of screening and structurally assign N-glycans for routine and discovery work with a proven throughput of 1,000-1,500 samples a year.

We here described a 2-3 fold more sensitive method for the analysis of N-glycans that could be implemented in other laboratories much cheaper than using the expensive new methods as RapiFluor even if RapiFluor is even more sensitive.

With the accurate and sensitive platform we were able to engineer an entirely humanized version of two plasma proteins AAT and C1inh with very homogeneous N-glycan profile consisting of 90% A2G2S2 N-glycans and glycoprofile that is a very close match to one present in healthy humans.

With our cell lines with combinatorial KO of galactosyl transferases, we are now able to produce non-galactosidated glycoproteins as IgG’s with a very homogeneous N-glycan profile consisting of more than 90% A2F or with further Fut8 KO A2 glycans only.

With the panel of glycoengineered cell lines shown in table 4 we are now able to produce a given recombinant protein with a range of distinctly different glycoprofiles to either match a pre-defined glycoprofile or to screen for efficacy, half-life, receptor binding etc. We believe that the next generation of biopharmaceutical drugs should be produced in such a panel and be screened for the glycoprofile that have an improved function half-life and/or efficacy.

Finally this thesis contains a range of examples on how the bioreactor conditions, such as the addition of Baicalein, the choice of cell line and the time point of harvest, impact N-glycosylation.
Future work

For analysis and especially for the databases of glycans the introduction of ion mobility and thus the collisional cross section as a fixed number for every glycan is a very interesting addition to the glycan mass-spectrometry-analysis toolbox. Ion mobility could possibly be used for annotating LacNac moieties to either antennae or as elongations.

Glycoproteomics is also a field in growth driven by the wishes from the pharmaceutical industry that would like to have the combined peptide and glycosylation analysis done in one single method. Waters, Thermo and Agilent suppliers of mass spectrometers now have a biopharma option that should be able to deliver the combined analysis of peptides and glyco-peptides. It will be interesting to see if the instruments and the software is capable of capturing the details of glyco-profiling directly on the glyco-peptide level.

Characterization of a few unusual N-glycans, annotated as the acetylated versions of A2G2S2(ac) and A2FG2S2(ac) found in CHO KO cell lines should be done. This will require tandem MS to prove that the extra acetyl group can be assigned to the sialic acid.

Further interesting topics for optimizing analysis:

- Exoglycosidase verification of branching
- Re-optimize mass spectrometer and source settings in higher resolution mode to possibly get enough sensitivity for analysis – result would be a lot more accurate mass of the ion trap instrument.

It will be interesting to see how a heavily engineered CHO cell line with specifically engineered N-glycans (as Fut8, St3gals, B4galts KO) will perform with media additions like Baicalein.

Further interesting topics for engineering could be:

- Overexpress the corresponding glycosyltransferases to the structures not completed (incomplete branching and capping with sialic acids)
• Alternative KO route to KO NeuSGC and Alpha-gal in CHO (the obvious route is patented by Sigma)

Most approved biotherapeutics in 2018 are still IgG’s and with only minor changes to their glycosylation, mainly the KO of Fut8. Now with the works of Yang et al. and present work the possibility to analyze the healthy human variant of a hormone, plasma protein, blood factor etc. is the obvious way forward for future biopharmaceuticals.
References


(30) Harvey, D. J.; Merry, A. H.; Royle, L.; Campbell, M. P.; Dwek, R. A.; Rudd, P. M. Proteomics 2009, 9 (15), 3796.


(35) Campbell, M. P.; Royle, L.; Radcliffe, C. M.; Dwek, R. A.; Rudd, P. M. Bioinformatics 2008, 24 (9), 1214.


(102) Garner, B.; Merry, A. H.; Royle, L.; Harvey, D. J.; Rudd, P. M.; Thillet, J. J. Biol. Chem. 2001, 276 (25), 22200.


(111) Waters. Waters - GlycoWorks Rapi Fluor-MS N -Glycan Kit [ CARE AND USE MANUAL ].


(114) Hong, Q.; Lebrilla, C. B.; Miyamoto, S.; Ruhaak, L. R. Anal. Chem. 2013, 85 (18), 1.


(162) Schrödinger, L. November 2015, .


Appendix A: Publications

Publication I: Fluorine modifications to 2-AA improves mass spectrometer signal of labeled N-Glycans
Fluorine modifications to 2-AA improves mass spectrometer signal of labeled N-Glycans

Anders Holmgaard Hansen1,a, Scott James Harrisona

a: The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark.

1: Corresponding author, ahoha@biosustain.dtu.dk

Abstract

2-Aminobenzamide (2-AB) and Antranillic acid (2-AA) is widely used as a dyes for derivatizing N-glycans enabling them to be analyzed by high performance liquid chromatography (HPLC) with fluorescence detection and possibly on-line mass spectrometry (MS). Although new derivatization reagents as procainamide and rapiflour are now emerging, they are very expensive and cannot be bought cheaply for research purposes as 2-AB and 2-AA. We have screened 24 halogen containing compounds for useful properties as derivatization agent for N-Glycan analysis. We were looking for the following properties:

- Comparable or better fluorescent signal intensities
- Similar excitation and emission wavelengths
- Able to react under the same conditions as 2-AB
- Better ionization/signal in negative mode MS
- Similar retention times in HILIC type liquid chromatography
In 2-Amino-4-(trifluoromethyl)benzoic acid (TFB) we have found a compound with these specific qualities.

**Keywords:**

2-Aminobenzamide, 2-Antranillic acid, Procainamide, Glycoprotein, N-Glycan, Carbohydrate, Sugar, Mass spectrometry, PTM, 2-AB, 2-AA, derivatization

1. **Introduction**

Post-translational modifications (PTMs) of proteins are important for the majority of life forms[1] and modulate factors such as the stability[2], activity and/or half-life[3] of the protein. Glycosylation of proteins is a type of PTM and it is acknowledged as one of the most important. Glycosylation is the most abundant PTM and have been shown to be present on at least 50% of all known proteins[4]. The types of protein glycosylation are traditionally divided into two groups; O-glycans where the polysaccharide chains are attached to serine and/or threonine amino acids and N-glycans in which the polysaccharide chains attached solely to asparagine amino acids. N-Glycans are further divided into three main types; High mannose, Hybrid and Complex. For recombinant glycoproteins are synthetic[5] or PEG-modified glycans[6] also seen.

The N-glycan structure is a critical quality attribute for the production of therapeutic proteins in the pharma industry because even small changes can severely change the physiological, therapeutic and physical properties of a glycoprotein. N-glycans have been shown to have effect on folding, function, degradation, half-life and activity. Exemplified by α1-antitrypsin protein folding [7]. Additional glycan can result in dysfunctional enzyme
activity of C1 inhibitor [8]. A missing or incorrect glycan pattern on alkaline phosphatase result in protein degradation instead of protein secretion [9] [10]. Missing galactoses on the glycans attached to IgG can be pathogenic [11]. The EPO and IgG-antibody glycoforms influence glycoprotein activity [12][13] and these effects are considered applicable to all glyco-proteins, resulting in that the accurate analysis of N-glycans is critical in the production of recombinant proteins.

Current screening methods for detection and characterization of N-glycans are based on retention time of enzymatically released glycans, which then are derivatized, separated and detected using fluorescence. Separation are based on HILIC-HPLC, or HILIC-ultra-high performance liquid chromatography (UHPLC). This fluorescence based analysis yields gives data that allow relative quantitation of peaks but inadequate characterization of the glycan pattern to allow unequivocal identification[14]. Combining liquid-chromatography / fluorescence detection with mass spectrometry (LC/MS) is more definitive, but the commonly used derivatization agents either have low fluorescence or do not ionize well in the mass spectrometer source, making analysis of small amounts of glycans difficult. The current industry standard method utilizes reductive amination derivatization of 2-Aminobenzamide (2-AB) [15], 2-Amino benzoic acid (2-AA) and Instant AB [16], none of these perform well in mass spectrometry. Other sensitive, non-common reported methods utilizing compounds such as TAMRA [17] and APTS [18] are all highly sensitive in capillary electrophoresis and MS. However, these have poor separation in liquid chromatography, making them unsuitable for the LC-MS that is now gaining more widespread use in laboratories as it is a more stable platform than CE-MS for routine analysis and allows for the MS to be online coupled to the separation. Several vendors now provide kit solutions to
the dye procainamide (PCA) where studies suggest 10-50x improved MS signal[19][20]
compared to 2-AB and it is starting to see more use lately.

2. Materials and Methods

2.1 De-glycosylation of glycoproteins:
Bovine RNase B (Sigma Aldrich) was de-glycosylated with PNase F (New England Biolabs)
according to the manufacturer’s specification.

Derivatization of oligosaccharides:
Slightly modified from Bigge et al. [22] using picoline borane as reductant [23].
Oligosaccharide standards, N-glycans from RNase B, were labeled according to the following
procedure. To make the reductive solution, 9.6mg of picoline borane (Sigma), 56 µL DMSO
and 24 µL acetic acid were mixed at room temperature. Next, 15 µL of this mixture was
added to 1 mg of 2-Amino-4-(trifluoromethyl)benzoic acid. 5 µL of the obtained labeling
solution is added to an Eppendorf vial, together with N-glycans from 25µg RNase B as dried
oligosaccharide. The vial is then vigorously vortexed for 10 sec. and then placed in a
thermostated mixer for 2.5 hours at 65 °C at 2000 rpm. Once the reaction is finished, the
labeled sugars are purified with CU cleanup cartridges (Prozyme), according to the supplier’s
instructions. The final purified sugars in 30 µL water are diluted with 50 µL of 100%
Acetonitrile and analyzed using LC-MS.

2.2 LC-MS analysis:
The LC-MS analysis were performed on an Ultimate 3000 (Thermo Fischer) HPLC coupled with a Velos Pro Linear Iontrap MS (Thermo Fischer). The column was a Waters Glycan BEH 1.7 µm particle size 50 x 2.1 mm. LC-Gradient is A: 50mM Ammonium formate (Sigma Aldrich), B: Acetonitrile (Merck Hypergrade LC-MS). Gradient starting at 16% A, 1 min. 16% A, 8 min. 24% A, 24min. 32% A, 25 min. 40% A, 25.5 min. 40% A, 26 min. 16% A, 30 min. 16% A. MS settings were: 700mz to 2000mz window; fullscan, negative mode; source fragmentation at 35V; HESI capillary temperature: 275°C. Sweep gas: 5, Sheet Gas: 50, Aux gas: 20. Data collected in profile mode from 0 - 25 min.
3. Results

3.1

In the current study, we have identified a new derivatization agent for the sensitive identification and characterization of glycans. The compound has the following advantages compared to the standard derivatization agent 2-AB:

- The compound maintains good fluorescence spectrometry properties.
- The compounds ionize better, which gives a higher ion current and better MS sensitivity.
- Due to higher ion current, tandem mass-spectrometry can be used for rapid and reliable identification.
- Due to improved mass-spectrometry signal, small glycan moieties can be accurately identified with MS or MS/MS.
- A compound with superior MS signal could work as drop-in substitute compounds in laboratories that currently use 2-AB (2-Aminobenzamide) or 2-AA (2-Aminobenzamide) on a routine basis.

3.2 Fluorescence

For sensitive detection of labeled glycans on HPLC the intensity of the fluorescence peaks are important. Figure 1 shows a comparison of the industry standard 2-AB compared to TFB. Both compounds were run in quadruplicates throughout the procedure and standard deviation is calculated and shown as error-bar.
Figure 1: Detection (fluorescence) of glycan moieties released from RNase B. Isomers are collapsed into one peak for simplicity. Peak areas are represented here as bars.

3.3 MS signal

We chose negative ion mode for analysis as although the signal intensity is lower than in positive ion mode, the chemical noise is significantly lower thus resulting in a better signal to noise in negative mode (data not shown). With this in mind, the compounds used as derivatizing agents (dyes) were tailored to stabilize a negative charge. As previously described the improved MS signal is great for identifying the individual glycans. Here we show the 2 fold increase in signal of our compound compared to 2AB and 2AA. Man6 peak was chosen because it was structure that flew as a single ion in both cases, allowing direct comparison. Notice that we did not optimize the MS conditions to TFB but used the same settings we use with routine 2AB analysis in our lab.
Figure 2: Bar chart of the summed peak area from the extracted ion chromatogram for the man6 moiety released from RNase B.

From this data it is clear that 2-Amino-4-(trifluoromethyl)benzoic acid is slightly less fluorescent than the reference 2-AB but the MS signal is improved by at least a factor of 2.

In figure 3A it is shown that the overall fingerprint-pattern of the RNase B N-glycans does not change between the labeling compounds, so the normalized results would be identical. TFB has a fluorescence signal comparable with 2-AB.
Figure 3:

A: Fluorescence chromatogram of 2-Amino-4-(trifluoromethyl)benzoic acid derivatized oligosaccharides (solid line) compared to reference 2-Aminobenzamide (dotted line).

B: MS extracted ion chromatogram for the man6 moiety of 2-Amino-4-(trifluoromethyl)benzoic acid derivatized oligosaccharides (solid line) compared to reference 2-Aminobenzamide (dotted line).
4. Discussion

4.1

Comparing compounds for labeling of oligosaccharides is complicated by the fact that the reductive amination reaction not 100% reproducible day to day or laboratory to laboratory. Some laboratories claim very high sensitivity when using a specific compound that others cannot reproduce. It is commonly accepted that to compare different runs of analysis, peak areas must be normalized as the intensity vary from day to day and reaction to reaction. We speculate that minute variations in ratios of the substances involved can influence the yield of the reaction combined with purification efficiency may vary. Moreover, a cleanup step is usually performed and that increases the risk of loss of labeled oligosaccharide. As an example the published increased fluorescence of 2-AA, by a factor of three to 44 times [24] [25] depending on the labeling conditions, compared to 2-AB have we not been able to reproduce in our lab. This is supported by Pabst et al.[26] who achieves similar fluorescence from 2-AA and 2-AB. 2-AA is sometimes advantageous as it separates isomers that 2-AB does not[21].

4.2

New compounds for this type of derivatization have emerged, such as procainamide, instant procainamide and RapiFluor. It will be interesting to see if they prove to be as good as described and are accepted by the scientific community. The only drawback seems to be the relatively low signal in negative mode MS. We find it interesting that unless the big vendors sell these labeling agents as a kit they do not get widely accepted in the scientific
community. As an example Procainamide was reported as a great compound for glycan labeling in MALDI systems in 2000 [27] and again by Pabst in 2009 [26]. To our knowledge procainamide is not used by many. Results from [27] suggests that electrospray MS signal from procainamide is comparable to that of 2-AB but the MALDI signal is greatly increased. Other fluorescent labelling reagents as f-mono [28] SPOT [29] and F-mocCL [30] have emerged but never reached a broad recognition as 2-AB and 2-AA has even with clearly superior properties.

4.3

We focused on negative mode MS because we believe that especially negative mode MS is useful when a deeper structural information of the oligosaccharide is required. In negative mode, the possibility exists of cross-ring breakage into A-Fragments and X-fragments (Domon, Costello nomenclature[31]) of complex structures. These are rarely seen in positive mode. Our results show the MS signal in negative mode of 2-Amino-4-(trifluoromethyl)benzoic acid is much higher than the reference 2-AB.

4.4

Optimal conditions for derivatization with each compound were not determined in this study. The final signal strength in MS and fluorescence reflect factors such as product yield and losses during cleanup as well as mass spectrometric ionization efficiency. Neither is the instrument settings as fluorescence detector sensitivity or source conditions in MS, optimized for any of the compounds.

5. Conclusion
5. 1

This new compound has all the desired traits we were initially looking for: Comparable fluorescence signal intensities, Similar excitation and emission wavelengths, Able to react under the same conditions as 2-AB, Excellent signal in negative mode MS, Similar retention times in HILIC type liquid chromatography. Finally no changes is needed to current setups already using 2-AB as TMB is having similar excitation and emission spectra, both runs in negative mode MS and retention times only shift slightly, with no change in pattern. Thus making the TFB compound a true drop in for all labs now using a 2-AB setup for the analysis of glycans. The relatively small addition of a trifluoro group to a known dye gave a significant increase in MS signal. We hope that this knowledge can be combined with compounds more fluorescent than the 2-AA this is based on, to produce new compounds with even better fluorescent and MS capabilities.

6. Acknowledgments

6. 1

We would like to thank the Novo Nordisk Foundation for the generous funding.

7. Conflict of Interest:

7. 1

The authors are inventors on a patent on the use of the TFB and similar compounds owned by the Technical University of Denmark.
References


[7] L.M. Powell, R.H. Pain, Effects of glycosylation on the folding and stability of human,


Comparison of procainamide


Publication II: Baicalein Reduces Oxidative Stress in CHO Cell Cultures and Improves Recombinant Antibody Productivity
Baicalein Reduces Oxidative Stress in CHO Cell Cultures and Improves Recombinant Antibody Productivity

Tae Kwang Ha, Anders Holmgaard Hansen, Stefan Kol, Helene Fastrup Kildegaard,* and Gyun Min Lee*

Oxidative stress that naturally accumulates in the endoplasmic reticulum (ER) as a result of mitochondrial energy metabolism and protein synthesis can disturb the ER function. Because ER have a responsibility on the protein synthesis and quality control of the secreted proteins, ER homeostasis has to be well maintained. When H₂O₂, an oxidative stress inducer, is added to recombinant Chinese hamster ovary (rCHO) cell cultures, it reduced cell growth, monoclonal antibody (mAb) production, and galactosylated form of mAb in a dose-dependent manner. To find an effective antioxidant for rCHO cell cultures, six antioxidants (hydroxyniis, N-acetylcysteine, baicalein, berberine chloride, kaempferol, and apigenin) with various concentrations were examined individually as chemical additives to rCHO cell cultures producing mAb. Among these antioxidants, baicalein shows the best mAb production performance. Addition of baicalein significantly reduced the expression level of BIP and CHOP along with reduced reactive oxygen species level, suggesting oxidative stress accumulated in the cells can be relieved using baicalein. As a result, addition of baicalein in batch cultures resulted in 1.7–1.8-fold increase in the maximum mAb concentration (MMC), while maintaining the galactosylation of mAb. Likewise, addition of baicalein in fed-batch culture resulted in 1.6-fold increase in the MMC while maintaining the galactosylation of mAb. Taken together, the results obtained here demonstrate that baicalein is an effective antioxidant to increase mAb production in rCHO cells.

1. Introduction

Chinese hamster ovary (CHO) cells are the most widely used mammalian host cell lines for the commercial production of recombinant therapeutic proteins, including monoclonal antibodies (mAbs). Over the past two decades, the growing demand for therapeutic mAbs has provided a challenge for the process of developing mass production of high quality mAbs.

Endoplasmic reticulum (ER), the central part of the secretory pathways in eukaryotic cells, is responsible for controlling the quality of secreted and resident proteins through the regulation of protein translocation, protein folding, and early post-translational modifications. A number of physiological conditions such as oxidative stress, hypoglycemia, acidosis, and thermal instability can disturb the ER functions, which triggers ER stress. Prolonged ER stress induces apoptotic cell death.

Oxidative stress refers to an imbalance between the production of reactive oxygen species (ROS) – such as hydrogen peroxide (H₂O₂), superoxide anion, and hydroxyl radicals – and their destruction by the antioxidant defense system. ROS accumulate endogenously as by-products of oxidative phosphorylation and energy metabolism within mitochondria and protein folding and disulfide bond formation in the ER during cell cultures. Antioxidants reduce the oxidative stress level and suppress the apoptotic cell death by scavenging oxygen free radicals, inhibiting chain reaction of oxidation, and detoxifying peroxide.

In order to restore the redox balance, antioxidants such as L-ascorbic acid 2-phosphate (VCP), the reduced form of glutathione (GSII), N-acetylcysteine (NAC), and N-acetylimidazole (NAI) have been used in rCHO cell cultures. Addition of NAC inhibited apoptotic cell death, resulting in increased production of erythropoietin (EPO) and human interferon-β-1a. Also, addition of BHA reduced apoptotic cell death and improved coagulation factor VII (FVIII). Likewise, addition of VCP or GSII increased tissue plasminogen activator (tPA) production. However, despite the importance of mass production of mAbs, studies on the effect of antioxidants on the production and quality of mAbs in rCHO cell cultures have not been fully substantiated.

In this study, to find a more effective antioxidant in CHO cell cultures, six different antioxidants including baicalein, which have been used widely in mammalian cell cultures, were evaluated as chemical supplements with two different rCHO cell lines producing the same mAb in six-well plates. Then, batch and
fed-batch cultures were performed in shake flasks with the supplementation of baicalein, which showed the best effect on culture performance among the six antioxidants. The ROS and ER stress levels were measured to study the effect of baicalein on mAb production and quality.

2. Experimental Section

2.1. Cell Line and Cell Maintenance

The two mAb-producing CHO cell lines with different cloned gene dosages (CS13-0.02 and CS13-1.00) were used in this study. The CS13-0.02 and CS13-1.00 cell lines were established through the co-transfection of light and heavy chain vectors containing dihydrofolate reductase into CHO DG44 cells and subsequently selected at 20 nM and 1 μM methotrexate (MTX, Sigma-Aldrich, St. Louis, MO), respectively. Cells were adapted to grow in a serum-free suspension culture and maintained in 125 ml Erlenmeyer flasks (Corning, Corning, NY) with 30 ml of PowerCHO2CD (Lonza, Basel, Switzerland) supplemented with 4 mM glutamine (Lonza), 2 μM L-anti-clumping agent (Thermo Scientific, Rockford, IL), and 20 nM or 1 μM MTX in a humidified incubator at 120 rpm, 37 °C, and 5% CO2.

2.2. Antioxidant Screening

Exponentially growing cells were inoculated at 3 x 10^5 cells/mL into 6-well plates containing 3 mL of the maintenance medium and the plates were incubated in a humidified incubator at 120 rpm, 37 °C, and 5% CO2. BHA, NAC, baicalein, berberine chloride, kaempferol, and apigenin were dissolved in DMSO at a concentration of 500, 250, 100, 50, and 50 μM, respectively. After 3 days of cultivation, each antioxidant was individually added to the cultures at various concentrations (10–500 μM). All chemicals were purchased from Sigma-Aldrich unless otherwise stated. Cell cultures without an antioxidant and with addition of 30 μL DMSO were also performed as controls. Samples were harvested every other day for measuring the viable cell concentration and mAb concentration. Culture supernatants were aliquotted and kept frozen at −70 °C for further analysis.

2.3. Batch and Fed-Batch Cultures

Exponentially growing cultures were inoculated at 3 x 10^5 cells/mL into 125 mL Erlenmeyer flasks containing 30 mL of maintenance medium and the flasks were then incubated in a humidified shaking incubator at 120 rpm, 37 °C, and 5% CO2. For batch cultures, baicalein was added to the cultures of CS13-0.02 and CS13-1.00 cells on day 3 at 100 and 200 μM, respectively. Batch cultures without addition of baicalein and with addition of 60 or 30 μL DMSO were also performed as controls. For fed-batch cultures, baicalein was added to the cultures of CS13-1.00 cells at 100 μM on day 3 and thereafter, CHO CD EfficientFeed™ B (Invitrogen, Carlsbad, CA) was added to the cultures daily from day 4 to 8 at 5% of v/v ratio. Fed-batch cultures without addition of baicalein were also performed as a control. Samples were harvested every day for measuring the viable cell concentration and mAb concentration. Culture supernatants were aliquotted and kept frozen at −70 °C for further analysis.

2.4. Viable Cell Concentration and mAb Concentration

The viable cell concentration was estimated with the Nucleo-Counter NC-200 cell counter (Chemometec, Allerod, Denmark). The mAb concentration was measured with an Octet RED96 (Pall, Menlo Park, CA) as described previously. The specific mAb productivity (μgA/mg) was evaluated from a plot of the mAb concentration against the time integral values of the viable cell concentration.

2.5. Western Blot Analysis

Western blot analysis was performed as described previously. Antibodies used for the Western blot analysis were anti-GRP78/Bip, anti-C180, anti-cleaved-caspase 3, and anti-BAX. Anti-vinculin was used as a loading control. All antibodies were purchased from Cell Signaling (Cell Signaling Technology, Beverly, MA).

2.6. Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

Two key pathway genes for ER stress (GRP78/Bip and CHOP) were evaluated by qRT-PCR as described previously. The qRT-PCR was run in an Mx3005P (Agilent Technologies, Santa Clara, CA) using Brilliant III Ultra-Fast SYBR® Green master mix (Agilent Technologies).

2.7. Measurement of Intracellular ROS

Cells were harvested from the cultures with or without addition of antioxidants and plated into 96-well optical bottom plates (Nunc, Thermo Scientific). Cells were then stained with CM-H2DCFDA (Molecular Probes, Eugene, OR) for 60 min in the dark according to the manufacturer’s protocol. The ROS level of approximately 10 000–15 000 cells was analyzed by a Celigo Cell Imaging Cytometer (Nexelom Bioscience, Lawrence, MA) with the expression analysis application, as described previously.

2.8. Purification and Glycan Analysis of mAb

Cell culture samples were taken from the cultures. After centrifugation and filtration to remove the cells and cell debris, the secreted mAbs in the culture supernatants were purified by protein A affinity chromatography (recombinant protein A agrose, Pierce, Rockford, IL), according to the manufacturer’s protocol. Purified mAbs were fluorescently labeled with GlykoPrep Rapid N-Glycan kit (ProZyme, Hayward, CA), according to the manufacturer’s protocol. N-linked glycan
analysis was performed by LC-MS system using a Thermo Ultimate 3000 HPLC with fluorescence detector coupled on-line to a Thermo Velos Pro Iontrap MS, as described previously.22

2.9. Statistical Analysis

Reported values are expressed as mean ± standard deviation, unless otherwise noted. The data were analyzed using a two-tailed Student’s t-test. The differences between the means were considered significant at P < 0.05.

3. Results

3.1. Oxidative Stress Negatively Affects mAb Production and Galactosylation in rCHO Cell Cultures

To determine the effect of oxidative stress on mAb production and glycosylation, CS13-1.00 cells were cultivated in shake flasks. H2O2 was used as an oxidative stress inducer and added to the cultures on day 3 at various concentrations (0–10 mM). Cultures were performed three separate times.

As expected, H2O2 addition significantly reduced cell growth, viability, and mAb production in a dose-dependent manner (Figure S1A–C, Supporting Information). Furthermore, H2O2 addition reduced G1 and G2 phase of glycans in a dose-dependent manner (Figure S1D, Supporting Information). The ROS level of cells significantly increased along with addition of 100 μM H2O2, indicating that H2O2 addition imposed oxidative stress on the cells (P < 0.05) (Figure S2A, Supporting Information). In addition, the mRNA and protein expression levels of BiP and CHOP, which are two of the best studied genes transcriptionally induced by ER stress, were significantly upregulated with H2O2 addition (P < 0.05) (Figure S2B–D, Supporting Information).

To generalize the effect of oxidative stress on mAb production and galactosylation, the same sets of experiments were performed with CS13-0.02 cells. As observed in the cultures of CS13-1.00 cells, H2O2 addition significantly reduced cell growth, viability, and mAb production in a dose-dependent manner (data not shown). Thus, oxidative stress negatively affects cell growth and mAb production and galactosylation in rCHO cell cultures.

3.2. Among Antioxidants Tested, Baicalein Had the Highest Maximum mAb Concentration (MMC)

To find an effective antioxidant for rCHO cell cultures, six antioxidants (50 μM BHA, 10 μM NAC, 100 μM baicalein, 10 μM berberine chloride, 10 μM kaempferol, and 10 μM apigenin) were examined. CS13-1.00 cells were cultivated in six-well plates and each antioxidant was individually added to the cultures on day 3. In parallel, cells were cultivated without any antioxidant as a control. The concentration of each antioxidant used in this study was based on the literature reports and then optimal concentration was experimentally determined (Figure S3, Supporting Information). Cultures were performed three separate times.

Figure 1 shows the profiles of cell growth, viability, and mAb concentration during the cultures. Antioxidants showed different effects on cell growth (Figure 1A). Compared to the control culture, only BHA and baicalein suppressed cell growth. However, cells in the cultures with addition of baicalein were viable for a longer period (Figure 1B). Despite reduced cell growth rate, baicalein showed the highest MMC (659.8 ± 49.8 μg/mL), which was approximately 1.3-fold higher than the control culture (Figure 1C, P < 0.05). The time integral of viable cell concentration (IVCC) of the cultures with baicalein was lower than that in the control cultures, suggesting that dnAB with baicalein is significantly higher than that in the control cultures.

The dnAB with baicalein was 26.1 ± 2.5 pg/cell/day, which was approximately 1.2-fold higher than that in the control culture. Among the antioxidants tested, baicalein showed the highest reduction of ROS level (P < 0.01). When the same sets of experiments were performed with CS13-0.02 cells, baicalein addition also resulted in the highest MMC (data not shown).

Baicalein was thus chosen for further analysis as an efficient antioxidant for improved mAb production.

3.3. Baicalein Improves mAb Production in Batch Cultures

To further investigate the potential of baicalein as an antioxidant for improving mAb production, CS13-1.00 cells were cultivated in the shake flasks and 100 μM baicalein was added to the cultures on days 3. In parallel, cells were cultivated without any antioxidant as a control. As another control, cells were also cultivated with addition of 300 μL DMSO on day 3, which was used for dissolving baicalein.

Figure 2 shows the profiles of cell growth, viability, mAb concentration, and ROS level during the cultures. As observed in 6-well plate cultures, baicalein suppressed cell growth and extended culture longevity (Figure 2A and B). The maximum viable cell concentration (MVCC) and specific growth rate (μ) in the control cultures (2.6 ± 0.2 10^6 cells/mL and 0.34 ± 0.01 day^-1) decreased to 2.3 ± 0.2 10^6 cells/mL and 0.29 ± 0.01 day^-1 in the cultures with addition of baicalein, respectively. However, the IVCC of the cultures with baicalein was similar to that in the control cultures due to prolonged culture duration. Addition of DMSO only rapidly decreased cell viability, suggesting that the beneficial effect of baicalein on cell viability outweighed the cytotoxicity of DMSO.

Addition of baicalein dissolved in DMSO significantly increased dnAB. Therefore, despite the inhibited cell growth by DMSO, the MMC in the cultures with baicalein (902.3 ± 32.0 μg/mL) was 1.7-fold higher than that in the control cultures (Figure 2C). The μ, MVCC, dnAB, and MMC in the cultures shown in Figure 2 were summarized in Table 1.

The ROS level increased rapidly during the control cultures and was not significantly affected by DMSO addition (Figure 2D, P > 0.05). Addition of baicalein significantly reduced the ROS level during the cultures, suggesting that decreased ROS level by baicalein contributed in part to improved cell viability and mAb production.

To determine the effect of baicalein on galactosylation pattern of mAbs, mAbs in the culture supernatants harvested on days 6 and 9 were purified by protein A affinity chromatography. Figure 3 shows
Figure 1. Profiles of (A) cell growth, (B) viability, (C) mAb concentration, and (D) ROS level of a CS13-1.00 cell line producing mAb during six-well plated cultures with various chemical reagents. A–C No chemical reagent (closed circle), 50 μM BHA (closed square), 10 μM NAC (closed triangle), 10 μM berberine chloride (closed diamond), 100 μM baicalein (open square), 10 μM kaempferol (open triangle), and 10 μM apigenin (open diamond). D No chemical reagent (white), 50 μM BHA (light gray), 10 μM NAC (dashed light gray), 100 μM baicalein (gray), 10 μM berberine chloride (dashed gray), 10 μM kaempferol (dark gray), and 10 μM apigenin (dashed dark gray). The error bars represent the standard deviations calculated from three independent experiments. P < 0.05, P < 0.01.

the profiles of galactosylated glycan proportion of mAbs. During the control cultures, G0 form increased from 67.6 ± 5.9% on day 6 to 76.5 ± 4.7% on day 9 with a concomitant decrease in G1 and G2 forms. In contrast, G0 form did not increase significantly in the cultures with addition of baicalein as well as in the cultures with addition of DMSO only (P > 0.05). G0 forms on day 9 in the cultures with addition of baicalein and DMSO only were 66.2 ± 2.3% and 68.9 ± 5.9%, respectively, which is significantly lower than those obtained from the control cultures (P < 0.05). Thus, baicalein addition helped to maintain the proportion of galactosylated form of mAb during the cultures, which is likely due in part to DMSO used for dissolving baicalein.

To confirm the potential of baicalein as an efficient antioxidant for improved mAb production in rCHO cell cultures, the same sets of experiments, except for baicalein concentration, were performed with CS13-0.02 cells (data not shown). The optimal concentration of baicalein for mAb production for CS13-0.02 cells was 200 μM.

As observed in the cultures of CS13-1.00 cells, baicalein suppressed cell growth, while extending culture longevity (Figure 4A and B). Due to increased q_{mAb} (5.4 ± 0.8 pg/cell/day) by addition of baicalein, the MMC in the cultures with baicalein (171.1 ± 29.3 μg/mL) was 1.8-fold higher than that in the control cultures (Figure 4C). Addition of baicalein significantly reduced the ROS level during the cultures (Figure 4D). ER stress level and apoptotic markers were also significantly decreased by addition of baicalein (data not shown). In addition, baicalein addition did not negatively affect galactosylation of mAb during the cultures (Figure S4, Supporting Information). Thus, these results support the potential of baicalein as an efficient antioxidant for improved mAb production in rCHO cell cultures.

3.4. Baicalein Reduces ER Stress in Batch Cultures

Baicalein is known to reduce the ER stress through the reduction of oxidative stress, which is one of the main factors for apoptosis induction. To determine the baicalein-mediated ER stress level and anti-apoptotic effect, the mRNA and protein levels of genes
related to ER stress and/or apoptosis were investigated by qRT-PCR and Western blot, respectively, with cells harvested on day 5, 7, and 9 of the cultures shown in Figure 2.

Figure 5A and B shows the mRNA expression level of two key pathway genes for ER stress, BIP, and CHOP, respectively. GAPDH was used as a reference gene. In the control cultures including the culture with addition of DMSO only, the expression level of both BIP and CHOP kept increasing rapidly during the cultures, suggesting that the ER stress naturally accumulates in the cells during the cultures. In contrast, baicalein addition significantly decreased the expression level of both genes along with reduction of ROS level (P < 0.05, Figure 2D). These results suggest that oxidative stress is an important factor for inducing the ER stress, and that the ER stress can be relieved using baicalein.

Figure 5C shows the Western blot results of the proteins related to the ER stress, as well as apoptosis. Like the mRNA expression levels, the protein expression level of both BIP and CHOP also significantly increased during the control cultures. Baicalein addition, however, significantly reduced the expression level of both BIP and CHOP. In particular, CHOP expression was almost blocked by addition of baicalein. Because CHOP has a

<table>
<thead>
<tr>
<th>C</th>
<th>0.29</th>
<th>0.01</th>
<th>2.62</th>
<th>0.23</th>
<th>17.49</th>
<th>1.37</th>
<th>520.1</th>
<th>9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-</td>
<td>0.18</td>
<td>0.02</td>
<td>1.18</td>
<td>0.19</td>
<td>25.05</td>
<td>2.23</td>
<td>465.5</td>
<td>93.9</td>
</tr>
<tr>
<td>T</td>
<td>0.21</td>
<td>0.01</td>
<td>2.35</td>
<td>0.06</td>
<td>22.13</td>
<td>2.20</td>
<td>901.3</td>
<td>112.9</td>
</tr>
</tbody>
</table>

No baicalein (C), DMSO (C-), and 100 µM baicalein (T). Values are means ± standard deviations of three independent experiments.
role in the apoptosis induction, the expression levels of BAX and cleaved-caspase-3 which are well-known apoptosis markers were also measured. Baicalin addition significantly reduced the expression of BAX and cleaved-caspase-3. Thus, antioxidant effect of baicalin relieved the ER stress, which led to decreased apoptotic cell death.

3.5. Baicalin Improves mAb Production in Fed-Batch Cultures

In fed-batch culture, culture duration is extended by nutrient feeding. As observed in batch culture, the ROS level necessarily increased as a result of cell metabolism during the culture. Therefore, the ROS level in fed-batch culture with longer culture duration is expected to be higher than that in batch culture, which negatively affects mAb production and galactosylation.

To investigate the potential of baicalin as an antioxidant for improving mAb production in fed-batch cultures, CS13-3.00 cells were cultivated in the shake flasks and 100 μM baicalin was added to the cultures on day 3, followed by daily feeding of nutrient cocktails from day 4 to 8. In parallel, cells were cultivated without addition of baicalin as a control.

Cultures were performed three separate times.

Figure 6 shows the profiles of cell growth, viability, mAb concentration, and ROS levels during the fed-batch cultures. Viable cell concentration, viability, and mAb concentrations were estimated and plotted daily before feeding nutrient cocktails, to avoid obtaining excessively complex profiles. The feedings of nutrient cocktails increased the MVCC and extended culture duration compared to the batch cultures shown in Figure 2 (Figure 6A and B). As observed in batch cultures, baicalin suppressed cell growth and extended culture longevity in fed-batch cultures (Figure 6A and B). Due to increased ϕ3HABA, the MMC in the fed-cultures with baicalin (1666.8 143.7 μg mL−1) was 1.6-fold higher than that in the control fed-batch cultures (Figure 6C). The μ, MVCC, ϕ3HABA, and MMC in the cultures shown in Figure 6 were summarized in Table 2.

The ROS level increased rapidly during the control fed-batch cultures and addition of baicalin significantly reduced the ROS level during the fed-batch cultures (P < 0.05, Figure 6D). In addition, G0 form of mAb in control fed-batch culture increased from 4.0 to 0.3% on day 9 to 10.9 to 1.3% on day 12, with a concomitant decrease in G1 and G2 forms. In contrast, galactosylated form of mAb did not change significantly from day 9 to day 12 during fed-batch cultures with baicalin (Figure 5S, Supporting Information). Thus, these results suggest that baicalin can be used as an efficient antioxidant for improved mAb production in fed-batch cultures of rCHO cells.

4. Discussion

Oxidative stress occurs when the balance between antioxidants and ROS shifts in favor of ROS due to either depletion of antioxidants or accumulation of ROS. ROS accumulate endogenously in living cells as a result of normal cellular metabolism and high concentrations of ROS can disturb ER functions and induction of ER stress. Under prolonged ER stress, cells eventually go through apoptotic cell death. Because ER has an important role in the protein synthesis and quality control through the regulation of proper folding of proteins, ER has to maintain an oxidizing and high calcium environment.

Previously, the negative effect of oxidative stress induced by H2O2 addition on cell growth and viability was observed in mammalian cell lines such as hybridoma, HEK293, HeLa cells, and CHO cells.[26–29] Among ROS, H2O2 is an important contributor to oxidative stress, which is converted from superoxide that leaks from mitochondria.[30] H2O2 addition significantly increased cell growth and viability of mammalian cell lines by inducing ER-meditated apoptosis.[25,28,29] In this study, we also observed that H2O2 addition in rCHO cell cultures increased the ROS level and ER stress, resulting in decreased cell growth and cell viability in a dose-dependent manner. Furthermore, it decreased not only mAb production, but also galactosylated form of mAb, suggesting that the maintenance of the redox balance is critical for high-quality mAb production in rCHO cell cultures.

A chemical approach to reduce the oxidative stress in rCHO cell cultures by medium supplementation is an efficient means that can be easily implemented in industrial processes. Antioxidants such as NAC and GSH have been used to improve therapeutic protein production in rCHO cell cultures.[26–29] In this study, we found a more effective antioxidant for rCHO cell cultures, six antioxidants (BHA, NAC, baicalin, berberine chloride, kaempferol, and apigenin) known to have an effect on...
ROS scavenging and reduction of oxidative stress were examined with mAb producing rCHO cell lines. To generalize the effect of antioxidants on cell growth and mAb production, the same sets of experiments were performed with two mAb producing rCHO cell lines (CS13-1.00 and CS13-0.02). The two rCHO cell lines produce the same mAb, but have different q_{mAb}. The q_{mAb} of CS13-1.00 is approximately 7.1 times higher than that of CS13-0.02. The effect of antioxidants on cell growth and mAb production was the same for both cell lines regardless of their different q_{mAb}. All six chemicals examined reduced the ROS level. However, their effect on cell growth and mAb production differed significantly among them. Regarding cell growth, berberine chloride showed a positive effect, whereas BH4 and baicalin showed a negative effect. Regarding mAb production, only baicalin increased MMC, demonstrating its potential as an effective antioxidant for improved mAb production. Baicalin decreased μ and MVCC, but increased culture duration and q_{mAb}. The beneficial effect of baicalin on culture duration and q_{mAb} outweighed its detrimental effect on μ and MVCC, resulting in significantly increased MMC (Figures 2, 4, and 6). The beneficial effect of baicalin on mAb production was further generalized by performing the same sets of experiments with Rituximab-producing CHO-K1 cell line. Addition of baicalin in batch cultures resulted in 2.2-fold increase in MMC, while maintaining the galactosylation of mAb (Figure S6A–E, Supporting Information). In addition, baicalin showed a higher reduction of ROS in shake flasks than in 6-well plates, suggesting that its effect also depends on culture type. Baicalin (5,6,7-trihydroxyflavone) is a well-known flavonoid that is originally isolated from the roots of Scutellaria baicalensis Georgi. Baicalin, which has been shown to have multiple biological activities including anti-inflammatory, anti-carcinogenic, and anti-HIV properties, has a strong antioxidant activity toward ROS. Among the various antioxidants, baicalin has attracted considerable attention because it has several interesting functions. As a polyphenol, the flavone backbone of baicalin carries three hydroxyl groups linked with the aromatic lipophilic structure, which makes it a strong free radical scavenger. In addition, baicalin, being free from sugar moieties, is a more lipid...
soluble, and may be able to penetrate membranes more easily. This may explain the reason why baicalin showed the best antioxidant effectiveness among the tested antioxidants.

For use of baicalin in rCHO cell cultures, baicalin was dissolved in DMSO. When only DMSO was added to the cultures of CS13.1.00 at a concentration used for dissolving baicalin, it was cytotoxic, but it increased qERAB while maintaining galactosylated form of mAb (Figure 3). DMSO has been applied for enhancing the production of recombinant proteins in CHO cell cultures. It acts as a chemical chaperone that is known to improve the folding capacity of ER, facilitate the protein folding in ER, and enhance the secretion of proteins. In addition, it is known to induce cell cycle arrest at G1 phase, which elevates the expression level of many genes related to the ribosomal biosynthesis along with larger cell size and metabolically more active cells. The cytotoxic effect of DMSO was relieved significantly by baicalin and its beneficial effect on qERAB and galactosylation of mAb was carried over in the baicalin solution.

When cells are exposed to ER stress, cells activate a series of complementary adaptive mechanisms known as the unfolded protein response (UPR) to buffer the ER stress. The UPR is regulated by three sensor proteins: inositol-requiring 1a, double-stranded RNA-dependent protein kinase-like ER kinase, and activating transcription factor 6. Upon ER stress, BiP is released to activate the three sensors and downstream signaling that block the protein translation, upregulate the expression of ER chaperones, and degrade the unfolded and misfolded proteins. If these signaling pathways fail to recover ER homeostasis, apoptotic cell death is induced by CHOP. As shown in Figure 5, baicalin addition reduced the mRNA and protein expression level of both BiP and CHOP. These results imply that the reduction of ROS blocked the accumulation of ER stress, which makes the activation of UPR unnecessary and extends the culture duration through the reduction of apoptotic cell death. Along with low ER stress, improved cell viability by baicalin may also help to maintain the highly galactosylated form of mAb. Cell viability is one of the critical factors affecting glycosylation, because proteases and glycosidases released from dead cells accumulate in the culture medium and remove glycan structures of the recombinant proteins. As a result, baicalin addition increased mAb production, while maintaining galactosylated form of mAb.
Figure 6. Profiles of (A) cell growth, (B) viability, (C) mAb concentration, and (D) ROS level of a C513-1.00 cell line producing mAb during shake flask cultures with fed-batch and baicalein addition. A–C) Fed-batch without baicalein (open square) and fed batch culture with 100 μM baicalein (closed circle). D) Fed-batch culture without baicalein (white) and fed-batch culture with 100 μM baicalein (gray). The error bars represent the standard deviations calculated from three independent experiments. \( P < 0.05, \ D < 0.01. \)

For large-scale commercial production of mAbs, fed-batch culture has been widely used because of its operational simplicity and high-titer. The ROS level in fed-batch culture with extended culture duration by nutrient feeding was higher than that in batch culture (data not shown). Therefore, cells in fed-batch culture were exposed to higher oxidative stress for a longer period, compared with those in batch culture. As observed in batch cultures, baicalein addition in fed-batch cultures significantly reduced the ROS level during the cultures. As a result, it increased mAb production as well as galactosylated form of mAb (Figure 6), demonstrating its potential as an effective antioxidant for improved mAb production in fed-batch culture.

In conclusion, oxidative stress negatively affected the production and galactosylation of mAb in rCHO cell cultures. Among the various antioxidants tested in this study, baicalein showed the best mAb production performance in both batch and fed-batch cultures of rCHO cells. Baicalein addition significantly enhanced mAb production while maintaining galactosylated forms of mAb. Thus, baicalein is an effective antioxidant for use in rCHO cell cultures for improved mAb production.

Table 2. The \( \mu \), MVCC, \( \eta_{\text{mAb}} \), and MMC with or without baicalein addition during fed-batch cultures.

<table>
<thead>
<tr>
<th></th>
<th>( \mu ) (day(^{-1}))</th>
<th>MVCC (10(^6) cells/mL)</th>
<th>( \eta_{\text{mAb}} ) (pg/mL/day)</th>
<th>MMC (μg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fed batch</td>
<td>0.31</td>
<td>0.01</td>
<td>2.79</td>
<td>17.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1063.5</td>
</tr>
<tr>
<td>Fed batch w/ baicalein</td>
<td>0.29</td>
<td>0.02</td>
<td>2.37</td>
<td>21.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1666.8</td>
</tr>
</tbody>
</table>

Values are means ± standard deviations of three independent experiments.
Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
This research was supported by Danish Council for Independent Research – Technology and Production Sciences (FTP), The Novo Nordisk Foundation, and the Ministry of Science, ICT and Future Planning for Basic Core Technology Development Program for the Oceans and the Polar Regions of the NSF (NRF-2016M1A3A2010313). The authors thank Helle Munck Petersen for technical assistance with the protein purification.

Conflict of Interest
The authors declare no commercial or financial conflict of interest.

Keywords
antioxidant, baicalin, CHO cell, ER stress, galactosylation, oxidative stress

Received: June 11, 2017
Revised: November 3, 2017
Published online:

Publication III: CRISPR/Cas9-Multiplexed Editing of Chinese Hamster Ovary B4Gal-T1, 2, 3, and 4 Tailors N-Glycan Profiles of Therapeutics and Secreted Host Cell Proteins
CRISPR/Cas9-Multiplexed Editing of Chinese Hamster Ovary B4Gal-T1, 2, 3, and 4 Tailors N-Glycan Profiles of Therapeutics and Secreted Host Cell Proteins

Thomas Amann, Anders Holmgaard Hansen, Stefan Kol, Gyun Min Lee, Mikael Rørdam Andersen, and Helene Fastrup Kildegaard

In production of recombinant proteins for biopharmaceuticals, N-glycosylation is often important for protein efficacy and patient safety. IgG withagalactosylated (G0)-N-glycans can improve the activation of the lectin-binding complement system and be advantageous in the therapy of lupus and virus diseases. In this study, the authors aimed to engineer CHO-S cells for the production of proteins with G0-N-glycans by targeting B4Gal-T isoform genes with CRISPR/Cas9. Indel mutations in genes encoding B4Gal-T1, -T2, and -T3 with and without a disrupted B4Gal-T4 sequence resulted in only ≈1% galactosylated N-glycans on total secreted proteins of 3-4 clones per genotype. The authors revealed that B4Gal-T4 is not active in N-glycan galactosylation in CHO-S cells. In the triple-KO clones, transiently expressed erythropoietin (EPO) and rituximab harbored only ≈6% and ≈3% galactosylated N-glycans, respectively. However, simultaneous disruption of B4Gal-T1 and -T3 may decrease cell growth. Altogether, the authors present the advantage of analyzing total secreted protein N-glycans after disrupting galactosyltransferases, followed by expressing recombinant proteins in selected clones with desired glycosylation profiles at a later stage. Furthermore, the authors provide a cell platform that prevalently glycylates proteins with G0-N-glycans to further study the impact of galactosylation on different in vitro and in vivo functions of recombinant proteins.

1. Introduction
Chinese hamster ovary (CHO)-derived cells are the major workhorses within mammalian cell lines and represent the cell platform in which >50% of the marketed recombinant proteins are produced.14 Thereof, recombinant monoclonal antibodies (mAbs) are the main product subclass and are utilized for the treatment of cancer and various inflammatory diseases.15 As a result of post-translational protein processing, mAbs harbor two predominantly bi-antennary N-glycans, one on each heavy chain at Asn297 or Asn297. In contrast to mAbs, erythropoietin (EPO) has three N-glycosylation sites occupied by predominantly tri- and tetra-antennary structures.16 In general, N-glycosylation can impact protein folding, immunoregulation, cellular homeostasis and the biological half-life of proteins.17 Within mAbs, the fragment crystallizable (Fc) N-glycans at Asn297 have a strong influence on anti-inflammatory properties, antibody-dependent cell-mediated cytotoxicity and complement-dependent cytotoxicity.18

The heterogeneous N-glycan profile of glycoproteins produced in CHO is one of the major factors that causes mAb heterogeneity and can be further optimized regarding core-fucosylation, galactosylation, antennarity, and terminal capping by sialic acids. Rituximab is an immunoglobulin G (IgG) 1-class molecule, one of the recombinant glycoproteins produced in CHO, and exceeds annual revenues of USD 7 billion.19 Rituximab targets the B-cell surface antigen CD20 in B-cell lymphoma and is predominantly N-glycosylated by A2FG0 and A2FG1 structures when produced in non-glycoengineered CHO cells.20 Since several studies revealed nonfucosylated IgGs have significantly higher binding affinity for the Fc-gamma receptor IIa (FcyRIIa) than fucosylated IgG versions,21 different approaches successfully removed the core-fucose by knockout of alpha-(1,6)-fucosyltransferase (FUT8) or tissue-specific transplantation antigen P35B (TSTA3) in IgG-expressing CHO cell lines.22-24

Additionally, galactosylated IgG1 variants with terminal N-Acetylgalactosamine (GlcNAc) (referred to as G9 glycoforms)
can increase the binding to RR11a13 and are accessible for the mannose-binding protein. They can therefore promote activation of the lectin-binding complement system136 without impacting in vivo clearance.12–16 Furthermore, HIV patients with high viral inhibition displayed an increased proportion of galactosylated N-glycans on global serum IgG, suggesting that galactosylated IgG variants may have antiviral activity.20 Interestingly, lupus patients showed improved disease symptoms after treatment with galactosylated antibodies.21 These G0-IgG variants can be obtained by sequential treatment of wild type (WT)-IgG with neuraminidase and galactosidase or by supplementing the cultivation medium with galactose analogues to block cellular B4Gal-T7.22 Nevertheless, fewer cell engineering attempts were initiated to produce G0-IgG1 compared to engineering nonfucosylated IgG1 variants.

As the CHO genome sequence became publicly available,23 CHO cell engineering is no longer performed in a "black box," which shortens cell line development and empowers a targeted approach for the engineering of a G0 CHO cell line. The classes of glycosyltransferases are made of homologous gene families, where the class of β-1,4-galactosyltransferases (B4GalTs) consists of seven members, B4Gal-T1–T7, which all transfer galactose from uridine diphosphate galactose (UDP-Gal) to GlcNAc and GlcNAc-terminated oligosaccharides (EC 2.4.1.138).24–26 B4Gal-T5 and -T6 are described to mainly function in O-glycosylation,24,25 whereas B4Gal-T7 transfers UDP-Gal within glycosaminoglycan biosynthesis and, therefore, is not involved in the N-glycosylation of proteins.27 A further study indicated that B4Gal-T1, -T2, -T3, and -T4 perform N-glycan galactosylation more efficiently than B4Gal-T5 and -T6 and suggested different brand preferences for the family members of β-1,4-galactosyltransferases.28 In addition, B4Gal-T4 is reported to also be active in the glycosylation of mucin-type core 2 branch in the O-glycosylation pathway.29 Furthermore, B4Gal-T1-KO mutants are described to have dramatically reduced galactosylation in secreted host cells (secretome) N-glycans and reduced growth of mice.22,29 In a previous study performed with CHO-K1 cells, disruption of B4Gal-T1, -T2, and -T3 led to almost fully agalactosylated EPO and rituximab.30 However, the impact of B4Gal-T disruptions on cell growth of more than one was not performed. The CHO-K1 study included single-KO of the B4Gal-T isofoms targeted in our work. In contrast, we aimed to study the N-glycosylation activity of B4Gal-T1, -T2, -T3, and -T4 after combinatorial KO in the industrially relevant CHO-S cell line. Especially, B4Gal-T2 and -T4 in CHO-S and the effect of B4Gal-T1 KO on cell growth, both with respect to clonal variation, were the driving motives of this work. Therefore, we applied clustered regularly interspaced short palindromic repeats (CRISPR-associated protein 9 (CRISPR/Cas9) to multiplex B4Gal-T disruptions in CHO-S cells. Cell growth and N-glycosylation profiles of multiple clones for each triple- and quadruple-KO combinations were analyzed to additionally examine clonal variation. N-glycosylation analysis of total secreted proteins, as well as transiently expressed rituximab and EPO (representing dissimilar N-glycan profiles), in the B4Gal-T edited cell lines was performed. The analysis demonstrated that N-glycans can be tailored for a greater variety of secreted glycoproteins, as represented by more than 250 proteins within the CHO-S secretome31 in addition to EPO and rituximab. With this, we found that screening the secretome N-glycans of our engineered clones is a promising strategy toward the expression of rituximab and EPO with G0 N-glycans in selected clones.

2. Experimental Section

2.1. sgRNA and GFP_2A_Cas9 Plasmid Design

GFP_2A_Cas9 and single-guide RNA (sgRNA) plasmids were constructed as previously described.13 The sgRNA target design for B4Gal-T1, B4Gal-T2, B4Gal-T3, and B4Gal-T4 was performed using CRISPry.13 The target sites for the mentioned genes and the oligos for sgRNA cloning are listed in Tables S1 and S2, Supporting Information, respectively.

2.2. Cell Cultivation and Transfection for Multiplexed Genome Editing

CHO-S suspension cells (Life Technologies, Carlsbad, CA) were cultivated in a CD CHO medium supplemented with 8 mM l-glutamine and 1 μM L-1 anticlumping agent (Life Technologies). Cells were incubated in a humidified incubator at 120 rpm, 37 °C and 5% CO₂. Cell passaging was conducted every 2–3 days at 3 × 10⁶ cells/mL after measuring viable cell densities (VCDs) and viabilities with the NucleoCounter NC-200 Cell Counter (ChemoMetec, Allerod, Denmark). One day prior transfection with CRISPR reagents, the anti-clumping agent was removed by centrifugation and 5 to 6 × 10⁶ cells/mL were seeded in a six well plate (BD Biosciences, San Jose, CA) for each transfection. At the day of transfection, each sample was seeded at 1 × 10⁶ cells/mL, and a total DNA load of 3.5 μg was transfected with FuGENE® HD transfection reagent (Promega, Madison, WI) and OptiPro SFM medium (Life Technologies), according to the manufacturer’s recommendations. The GFP_2A_Cas9 sgRNA plasmid ratios for each sample are presented in Table S3, Supporting Information. To measure transfection efficiency, pmaxGFP® vector (Lonza, Basel, Switzerland) transfection was performed. Cells were harvested for fluorescence-activated cell sorting (FACS) 48 h after transfection.

2.3. Single-Cell Cloning Using FACS

Before FACS, cells were filtered through a 40 μm cell strainer into a FACS-compatible tube. Operating a FACSJazz (BD Biosciences), single fluorescent-positive cells were sorted into 384-well plates (Corning, New York, NY) already containing 30 μL CD CHO medium supplemented with 8 mM l-glutamine, 1.5% HEPES buffer and 1% Antibiotic-Antimycotic (Gibco, Waltham, MA) per well. For cell sorting, fluorescent-positive cell populations were gated based on non-transfected WT CHO-S cells. Two weeks after cell sorting the clones were moved to 96-well flat-bottom plates (BD Biosciences) and expanded for deep sequencing analysis and batch cultivation.
2.4. Deep Sequencing Analysis

Confluent colonies from 96-well flat-bottom replicate plates were harvested for genomic DNA extraction. DNA extraction was performed using QuickExtract DNA extraction solution (Epi-centre, Illumina, Madison, WI) according to the manufacturer’s instructions. The library preparation was based on Illumina 16S Metagenomic Sequencing Library Preparation, and deep sequencing was carried out on a MiSeq Benchtop Sequencer (Illumina, San Diego, CA). The protocol for amplifying the targeted genomic sequences, Amplicon purification, adapter-PCR and following quality analysis was based on a previously published work.\[^{13}\] PCR primers are presented in Table S4. Supporting Information.

2.5. Batch Cultivation to Study Cell Growth and Secretome N-Glycans

For batch cultivation and secretome analysis, cells were seeded at 3.0 × 10^5 cells mL\(^{-1}\) in Corning vent cap shake flasks (Sigma–Aldrich, St. Louis, MO) as duplicates in 30 mL CD CHO medium supplemented with 8 mM L-glutamine and 1 μM L-anti-clumping agent (Life Technologies). Cells were incubated in a humidified incubator at 120 rpm, 37°C and 5% CO\(_2\). Cell densities and viabilities were determined once per day using the NucleoCounter NC-250 Cell Counter (ChemoMetec). Based on cell densities from days 0 to 7 we calculated the integral of viable cells (IVC). The IVC was used to statistically determine differences in cell growth between the generated clones. Using a two-tailed, unpaired t-test with Prism\(^{7}\) software, we grouped the clones into two sets. One set with (n = 16) and one set without combinatorial disruption of B4Gal-T1 and -T3 (n = 16). Combining all IVC values of the clones in each set, we determined if a set of clones had a significant change of IVC compared to CHO-S WT and WT ctr cells. Secretome sample volume was calculated to harbor 20 × 10^8 cells and harvested 5 days after seeding to be pooled within biological replicates.

2.6. Batch Cultivation for Transient Rituximab/EPO Transfection and Rituximab/EPO N-Glycan Analysis

For transient expression of rituximab and EPO, cells were seeded in Corning vent cap shake flasks (Sigma–Aldrich) as duplicates with cell densities ≈1 × 10^6 cells mL\(^{-1}\) in 60 mL CD CHO medium supplemented with 8 mM L-glutamine (Life Technologies). Cells were incubated in a humidified incubator at 120 rpm, 37°C and 5% CO\(_2\) and transfected with 75 μg of rituximab or EPO encoding plasmid for each flask using Freestyle\(^{TM}\) MAX reagent together with OptiPRO SFM medium (Life Technologies) according to the manufacturer’s recommendations. A total of 1 μM L-anti-clumping agent was added 24 h after transfection. pmaxGFP\(^{\beta}\) vector (Lonza) transfection was performed to measure transfection efficiencies. Cell densities and viabilities were determined once per day using the NucleoCounter NC-250 Cell Counter (ChemoMetec). To purify rituximab and EPO, the supernatants of the transfected clones were harvested 3 days after transfection and pooled within duplicates.

2.7. Rituximab and EPO Purification

For rituximab purification, supernatant samples were centrifuged (1000 g, 5 min, 4°C) and afterwards filtered (pore size = 0.22 μm) to remove cells and cell debris. Rituximab was purified by protein A affinity chromatography (MabSelect, GE Healthcare, Uppsala, Sweden) according to the manufacturer’s protocol. Human protein C4 (HPC4)-tagged EPO was purified from supernatants using Anti-Protein C Affinity Matrix from Roche (Basel, Switzerland, Cat. Nr. 11815024001) as per the instructions of the manufacturer.

2.8. N-Glycan Analysis

Sample preparation for N-glycan analysis was performed with GlycoWorks RapiFlour-MS N-Glycan Kit ( Waters, Milford, MA) according to the manufacturer’s instructions. A total of 12 μg purified protein or 12 μL of 10× concentrated (Amicon Ultra-15, Merck, Darmstadt, Germany) secretome sample were used for each sample. Labeled N-Glycans were analyzed by a LC-MS system using a Thermo Ultimate 3000 HPLC with fluorescence detector coupled online to a Thermo Velos Pro Iontrap MS, as described previously with minor modifications.\[^{13}\] Separation gradient was 30% to 43% buffer, and MS was run in positive mode. The amount of N-Glycan was measured by integrating the areas under the normalized fluorescence spectrum peaks with Thermo Xcalibur software (Thermo Fisher Scientific, Waltham, MA) giving the normalized, relative amount of the glycans.

3. Results

3.1. Generation of Engineered CHO-S Cell Lines With Combinations of Indels in Multiple B4Gal-T Genes

To investigate the exact impact of B4Gal-T1, -T2, -T3, and -T4-KO on N-glycan galactosylation, we aimed to generate clones with insertion or deletion (indel) mutations in several of the genes. To get these combinations in a minimal number of operations, we co-transfected Cas9 (GFP_2A_Cas9) with sgRNAs against B4Gal-T1, -T2, and -T3 in the first transfection (Table S3, Supporting Information). After single-cell cloning, we carried out deep sequencing to identify clones with exclusively out-of-frame indels in the targeted sequences. In a second round of transfections, we aimed to generate clones with indels in additional B4Gal-T target genes. Therefore we co-transfected GFP_2A_Cas9 with sgRNAs against B4Gal-T1 and -T4 into a clone with confirmed indels in B4Gal-T2 and -T3 (Table 1). In our study, a total of 109 potential deletion clones were deep sequenced for genomic indels in the targeted regions (Table S5, Supporting Information). Thereof, clones with in-frame indel or indel frequency <98% were discarded. We expanded clear single- and multi-KO clones of I–4 targets. Next, we isolated multiple independent clones for each genotype to study true
biological replicates of the phenotypes, in total 17 clones (Table 1). Two control clones (WT ctr and T2-3-KO), which showed no additional insertion or deletion after each round of transfection, were additionally selected to investigate the impact of transfection and subcloning on growth and N-glycan profile.

3.2. Effect on Growth From Different B4Gal-T-KOs and Indel Combinations

The overall aim of our study is to provide a CHO-S platform to produce recombinant proteins with agalactosylated N-glycans. Engineering cells toward G0-glycans leads to altered N-glycans on the recombinant protein and also on total host cell glycoproteins. As cell growth performance is a substantial factor for industrial protein production platforms, we first evaluated whether decreased N-glycan galactosylation influences CHO cell growth. During shake flask batch experiment, the WT ctr clone was identified to have similar growth and viability compared to CHO-S WT (Figure 1). Double-KO of B4Gal-T1 and -T3 (T1-3-KO) indicated decreased growth (Figure 1A), whereas the two clones with frame-shifts in B4Gal-T3 were not influenced in cell growth (Figure 1A). Compared to CHO-S WT and WT ctr, T2-3-KO clone revealed similar growth and T2-3- KO ctr clone exhibited decreased growth (Figure 1B). The four triple-KO clones with frame-shifts in B4Gal-T1, -T2, and -T3 (T1-2-3-KO) and the three T1-2-3-4-KO mutants had decreased growth compared to CHO-S WT (Figure 1B and C). For the three T2-3-4-KO and three T1-2-KO clones, we observed a high diversity in growth between the clones (Figure 1D and E). We additionally calculated the IVC for the different clones over the 7 days of the batch experiment. Compared to CHO-S WT and WT ctr, averages of clone groups without combinatorial disruption of B4Gal-T1 and -T3 had no change in IVC (Figure S4, Supporting Information). In contrast, the IVC from clones harboring a combination of disrupted B4Gal-T1 and -T3 (T1-2-3-KO, T1-3-KO, T1-2-3-4-KO) was slightly decreased.

3.3. Effects of B4Gal-T-KOs on Secretome N-Glycan Profiles

To examine the contribution of the targeted B4Gal-Ts on galactosylation of the different N-glycan branches, we analyzed the remaining levels of N-glycan galactosylation on total secreted proteins in 17 clones with combinatorial disruption of B4Gal-Ts. As presented in Figure S1, Supporting Information, the complex bi-antennary di-sialylated N-glycan structure (A2F2G2S2) was the major structure within the CHO-S WT secretome. Notably, in CHO-S WT cells, only one minor peak (0.7%) of G0-N-glycans could be annotated (Figure S1, Supporting Information). T3-KO A, T2-3-KO, and T2-3-4-KO clones showed a N-glycan pattern with minor differences compared to CHO-S WT, and G0 structures were only present in T2-3-KO and T2-3-4-KO clones (Figure 2 and Figure S1, Supporting Information). In contrast,
the T3-KO clone exhibited a total of \( \approx 65\% \) G0 structures and \( \approx 10\% \) G1 N-glycans (Figure 2). Indels in B4Gal-T1 and -T2 resulted in the absence of G4 forms, reduced G3 and G2 forms, and increased G1 and G0 proportions (Figure 2). Additionally, we could only annotate \( \approx 1\% \) galactosylated N-glycan structures in the secretomes of T1-2-3-KO and T1-2-3-4-KO clones (Figure 2). The major N-glycan structures of T1-2-3-KO and T1-2-3-4-KO clones were A2FG0, A3FG0 and A4FG0 (Figure S1, Supporting Information). Furthermore, the additional B4Gal-T4 indel in T1-2-3-4-KO clones did not increase G0 proportions or eliminate G1 N-glycans when compared to T1-2-3-KO cell lines (Figure 2). Altogether, disruption of B4Gal-T2 in conjunction with B4Gal-T1 and -T3 decreased the galactosylated secretome N-glycan proportion from \( \approx 10\% \) (T1-3-KO) down to \( \approx 1\% \) (T1-2-3-KOS) with A2FG0 as the dominating Nglycan structure (Figure S1, Supporting Information). The role of B4Gal-T2 in N-glycan galactosylation has previously not been studied in exact terms. To address this, we compared sets of two clones differing in their genotype by the KO of B4Gal-T2 (Figure 2). The occurrence of agalactosylation without (clone T3-KO A and T1-3-KO) and with additional KO of B4Gal-T2 (clones T1-2-3-KO, T2-3-KO, and T2-3-KO ctr) was analyzed. We conclude that the additional disruption of B4Gal-T2 stacked on T3-KO or T1-3-KO increased the proportion of G0-N-glycans by \( \approx 3-10\% \) (Figure 2).

3.4. Tailored Rituximab and EPO N-Glycosylation After B4Gal-T-Double and Triple-KOS

To investigate if engineered secretome N-glycans will also be represented on selected therapeutic proteins, we transiently expressed rituximab and EPO in CHO-S WT and KO clones T3-KO A, T2-3-KO, T1-3-KO, and T1-2-3-KO A and analyzed the resulting N-glycan structures upon purification. CHO-S WT,
clone T3-KO A, and T2-3-KO displayed comparable rituximab N-glycan profiles with G0 and G1 as prevalent structures with both ≈40% of total rituximab N-glycans (Figure 3A and Figure S2, Supporting Information). In contrast, rituximab purified from clones T1-2-3-KO A and T1-3-KO clones was mostly N-glycosylated by bi-antennary G0 sugars. As presented in Figure 3B, double-KO of B4Gal-T1 and -T3 (T1-3-KO) resulted in higher G0-N-glycan proportions on rituximab (≈84%) than in clone T1-2-3-KO A (≈68%). In both clones, the bi-antennary, G0-nonglycosylated A2FG0 was clearly the main structure. However, HM, A2G0 and A2FG1 N-glycans were also annotated, where T1-2-3-KO A revealed higher HM proportions than other cell lines. Cell growth after rituximab transfection was comparable between CHO-S WT, WT ctrl, T3-KO A, and T1-3-KO (Figure S2, Supporting Information), whereas clones T2-3-KO and T2-1-2-3-KO A revealed increased viable cell concentrations on day 3. For transiently expressed EPO, the N-glycan profiles of CHO-S WT, T3-KO A, and T2-3-KO are similar where annotated N-glycan structures predominantly harbor ≥4 galactose residues; however, G0 forms are not present in EPO from CHO-S WT (Figure 3C and Figure S3, Supporting Information). In contrast, double-KO of B4Gal-T1 and -T3 resulted in increased G0 proportions (≈72%), whereas G3- and G4-glycans could not be identified. Analyzing N-glycan structures of EPO from the triple-KO clone T1-2-3-KO A, we could only annotate agalactosylated and mono-galactosylated N-glycans (Figure S3, Supporting Information). Overall, disruption of B4Gal-T1 and -T3 with or without additional disruption of B4Gal-T2 resulted in rituximab with ≈2–3% galactosylated N-glycans. Single disruption of B4Gal-T3 or disruption of both B4Gal-T2 and -T3, did not change rituximab N-glycosylation. However, disruption of B4Gal-T2 in addition to indels in B4Gal-T1 and -T3 increased the G0 N-glycan proportion of transiently expressed EPO from ≈72% to ≈91%.

4. Discussion

Targeting multiple genes in one transfection with CRISPR/Cas9 is a time-saving method to generate clones with different indel combinations in several genes.13 However, clones often have in-frame indels, which may not disrupt the gene(s).16 First, we co-transfected with sgRNAs against a combination of B4Gal-T1, -T2, and -T3. In a second round of transfection, we built up triple-KO (T1-2-3-KO and T2-3-4-KO) and quadruple KO clones (T1-2-3-4-KO) based on transfections of the T2-3-KO cell line. Although it is faster, a limitation of this multiplexing method is that not all desired KO combinations might appear after deep sequencing of single-cell clones. An alternative approach would be to use two sgRNAs per target to remove major parts of target DNA sequences from the genome. The double-cuts approach is in general less efficient than single cut and furthermore complicates multiplexing. However, in-frame indels become less of a concern.15

The effects of B4Gal-T disruptions on cell growth and the glycosylation of total secreted proteins have to our knowledge not been studied in details previously. In a previous study, cell growth was investigated for one clone with a combinatorial disruption of B4Gal-T1 and FUTN.18 Here, we aimed to assess the impact of B4Gal-T indels on cell growth and N-glycosylation in groups of clones with the same combination of indels to additionally address clonal variation. After disrupting the four targets, we observed that clones with indels in B4Gal-T1 and -T3 have decreased IVC when compared to CHO-S WT and WT ctrl (Figure 1 and Figure S4, Supporting Information). The reduced IVC in T1-3-KO, T1-2-3-KO, and T1-2-3-4-KO could be associated to the high G0-N-glycan proportions of their secreteme (Figure 2) or be linked to clonal variation, which is known to be challenging when working with CHO cells.18 However, glycosylation plays a main role in cell-cell communication via, for example, endocytosis, receptor activation, and cell adhesion,19 and glycosylation engineering, therefore, might impact cultivation performance. We also report heterogeneous cell growth of clones within the generated indel combination groups. This might also be a result of clonal variation after subcloning or due to off-target effects. While subcloning did not influence growth of the WT ctrl clone, subcloning of T2-3-KO lead to decreased growth of the T2-3-KO ctrl (Figure 1).

We investigated if disruption of B4Gal-T1, -T2, and -T3 in CHO-S cells is sufficient to produce predominantly agalactosylated proteins and if additional disruption of B4Gal-T4 is of any
benefit for decreased galactosylation. In contrast to a previous study, which suggested B4Gal-T1–4 to all be active in N-glycan galactosylation,[38] our results indicate that B4Gal-T1, -T2, and -T3 are the most active B4Gal-Ts in the N-glycosylation pathway of CHO-S cells and that B4Gal-T4 has very little or no contribution to galactosylation of N-glycans in CHO-S cells. The lack of N-glycosylation activity of B4Gal-T4 in our work supports another study where B4Gal-T4 was reported to be active in the galactosylation of mucin-type core 2 branching in the O-glycosylation pathway.[33] Furthermore, B4Gal-T5, -T6, and -T7 (and potentially unknown B4Gal-Transferases) in sum contribute only up to ~3% N-glycan galactosylation of the secretome, as seen in Figure 2. Our results indicate that subcloning had no impact on secretome N-glycosylation as the WT ctrl and T2-3-KO ctrl clones showed comparable N-glycan structures to their parental cell lines in the batch cultivation (Figure 2).

As the T2-3-KO clone still produced G1, G2, G3, and G4 structures and all KO cell lines with indels in B4Gal-T1 lack G4 N-glycans, B4Gal-T1 is very likely capable of transferring galactose to all four branches. Therefore, we suggest that B4Gal-T1 is the most active N-glycan processing B4Gal-T within the family of β1,4-galactosyltransferases of CHO-S cells, which is in line with previous work in another CHO cell line.[33] We expected to reveal branch specificities for all four targeted B4Gal-Ts. Due to low galactosylation activities of B4Gal-T2, -T3, and -T4, we can only conclude that B4Gal-T1 can galactosylate all four antennas (Figure 2) and that its branch preference needs to be explored further.

Moreover, we analyzed that B4Gal-T2 activity contributes to ~3–10% of N-glycan galactosylation (Figure 2). Since single B4Gal-T3-KO did not decrease galactosylation (Figure 2), we suggest that B4Gal-T3 has only a minor role in CHO-S N-glycosylation or that its disrupted N-glycan transferase
function can be compensated by B4Gal-T1 and -T2 activity in the T3-KO clone.

The remaining level of rituximab galactosylation of the CHO-S-derived clone T1-3-KO (n=2-3%) is comparable, yet slightly higher to another study where decreased rituximab galactosylation (n=1%) was achieved by knocking out B4Gal-T1 and -T3 in CHO-K1-derived cell lines. This difference in remaining N-glycan galactosylation could be due to differences in the N-glycan pathways of the cell lines used (CHO-S versus CHO-K1) or due to clonal variation. Within our triple-KO cell line T1-2-3-KO-A, we also noticed a significant amount of high-mannose (HM) structures on transiently expressed rituximab (Figure 3A and B). HM structures are a critical quality attribute within biopharmaceutical protein production and can accumulate during cell culture performance. Process design and genetic engineering could be two possibilities to overcome accumulated HM structures, which might represent proteins accumulated in the Golgi-situated N-glycan machinery after disrupting Golgi-residing B4Gal-T1, -T2, and -T3. This disruption might cause increased traffic and residence time of secretome proteins in the Golgi lumen without being further processed by glycosyltransferases. Recent studies displayed increased processing of N-glycans after overexpression of Mga4 and Mga5, which could result in lower HM proportions on glycoproteins.

For glycoproteins harboring tri- or tetra-antennary N-glycans, as is the case for EPO, KO of B4Gal-T1 and -T3 is not sufficient to produce mainly agalactosylated glycoproteins (Figure 3B with n=20% EPO galactosylation in T1-3-KO). In comparison, rituximab expressed in clone T1-3-KO resulted in only 3% galactosylated structures (Figure 3A). Therefore, we propose that bi-antennary N-glycosylated proteins as rituximab can be produced with mostly agalactosylated N-glycans after double-KO of B4Gal-T1 and -T3. However, tri- and tetra-antennary N-glycosylated secretome proteins as EPO additionally need KO of B4Gal-T2 to be predominantly agalactosylated. For transiently expressed EPO in CHO-K1-derived cells with triple-KO of T1-2-3-KO, -T2, and -T3 the proportions of galactosylated N-glycans were found to be n=4% in an earlier study. In our study, we annotated n=6% galactosylated N-glycans on transiently expressed EPO from the CHO-S-derived triple-KO T1-2-3-KO A (Figure 3C). Although these results indicate similar effects on galactosylation of EPO after disruption of two identical gene targets, deviations could be related to differences between CHO-K1 and CHO-S expression levels of nontargeted B4Gal-T isomers.

Therefore, we suggest that engineering cells with nongalactosylated N-glycans on a secretome level in CHO-S WT is a promising strategy toward producing G0-IgG1 and G0-EPO at a later stage. Despite the divergent gene expression levels between different CHO cell lines this engineering strategy is suitable for both CHO-K1 and CHO-S-derived cell lines as utilized in our work. In the presented study the triple-KO with n=1% galactosylated structures on the secretome also showed predominantly agalactosylated N-glycans on transiently expressed rituximab with only 3% galactosylated N-glycans and on transiently expressed EPO with remaining n=6% galactosylated N-glycan structures.

In summary, our study presents the necessity of disrupting the three genes, B4Gal-T1, -T2, and -T3, to produce predominantly agalactosylated secretome proteins, rituximab and EPO in CHO-S cells. The possibility to engineer tri- and tetra-antennary G0 N-glycans, which are naturally not produced in CHO-S WT cells (Figure S1, Supporting Information), is presented. Prior engineering of secretome N-glycans in a WT cell gives rise to the flexibility of expressing several different model proteins in the engineered cell line at a later stage. Such model proteins might include already marketed antibodies or other therapeutic proteins. With our cell platform that prevalently glycosylates proteins with G0-N-glycans, we demonstrate an alternative to galactosidase treatment of recombinant proteins to investigate further beneficial in vitro and in vivo characteristics based on tailored G0 N-glycosylation profiles.

Abbreviations

Asn, Asparagine; AUC, area under curve; B4Gal-T, β-1,4-galactosyltransferase; Cas9, CRISPR-associated protein 9; CHO, Chinese hamster ovary; CRISPR, clustered regularly interspaced short palindromic repeats; EPO, erythropoietin; FACS, fluorescence-activated cell sorting; Fc, fragment crystallizable; FcyRIIIa, Fc-gamma receptor III a; FUT8, alpha (1-6)-fucosyltransferase; G0, agalactosylated; GlcNAc, N-Acetylgalactosamine; HM, high-mannose; HPC4, human protein C4; IgG, immunoglobulin G; indel, insertion or deletion; mAb, monoclonal antibody; sgRNA, single guide RNA; TSTA3, tissue-specific transplantation antigen P53B; UDP-Gal, uridine diphosphate galactose; VCD, viable cell density; WT, wild type.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

T.A. and A.H.H. contributed equally to this publication. The authors thank Sara Petersen Bjørn, Bjørn Voldborg, Johnny Åmssdorff, Yuzhou Fan, and Patrice Menard for valuable guidance and support. The authors thank Karen Katrine Brendum, Nachon Charanyanonda Petersen, Karoline Schousboe Flemming, and Zulfiya Suhkova for excellent technical assistance with the FACS and MiSeq library preparation; Helle Munck Petersen for assistance with the protein purification; Anna Koza and Mads Valdemar Anderson for assistance with the MiSeq analysis; and Johan Rockberg (KTH, Sweden) for providing the HPC-4 sequence. The Novo Nordisk Foundation (NNF10CC1016517) supported this work. T.A., H.F.K., and M.R.A. are receiving funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 642663. A.H.H., M.R.A., H.F.K., and T.A. planned the experiments. T.A. performed the experimental work and wrote the manuscript. A.H.H. performed the N-glycan analysis, and S. K. conducted the protein purifications. A.H.H., M.R.A., H.F.K., S.K., and G.M.L. guided the project, contributed to experimental design, and commented and corrected the manuscript.

Conflict of Interest

The authors declare no commercial or financial conflict of interest.
Supplementary Material for

CRISPR/Cas9-multiplexed editing of Chinese hamster ovary B4Gal-T1, 2, 3 and 4 tailors N-glycan profiles of therapeutics and secreted host cell proteins

Thomas Amann¹, Anders Holmgaard Hansen¹, Stefan Kol¹, Gyun Min Lee¹², Mikael Rørdam Andersen³, Helene Fastrup Kildegaard¹

¹The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark
²Department of Biological Sciences, KAIST, Daejeon, Republic of Korea
³Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark

Correspondence: Helene Fastrup Kildegaard, The Novo Nordisk Foundation Center for Biosustainability, Kemitorvet, Building 220, 2800 Kgs. Lyngby, Denmark

E-mail: hef@biosustain.dtu.dk

<table>
<thead>
<tr>
<th>Section, figures and tables</th>
<th>Page number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Figure 1</td>
<td>2</td>
</tr>
<tr>
<td>Supplementary Figure 2</td>
<td>3</td>
</tr>
<tr>
<td>Supplementary Figure 3</td>
<td>4</td>
</tr>
<tr>
<td>Supplementary Figure 4</td>
<td>5</td>
</tr>
<tr>
<td>Supplementary Table S1</td>
<td>6</td>
</tr>
<tr>
<td>Supplementary Table S2</td>
<td>7</td>
</tr>
<tr>
<td>Supplementary Table S3</td>
<td>8</td>
</tr>
<tr>
<td>Supplementary Table S4</td>
<td>9</td>
</tr>
<tr>
<td>Supplementary Table S5</td>
<td>10-12</td>
</tr>
<tr>
<td>Supplementary Table S6</td>
<td>13</td>
</tr>
</tbody>
</table>
Supplementary Figure 1. Detailed HPLC histograms of analyzed N-glycan samples from parental and generated KO cell lines. Structures from MS-annotated peaks are indicated with N-glycan cartoons and represent N-glycans from total secreted cell protein.

Supplementary Figure 2: Rituximab N-glycan profiles and cell growth after transient transfection. N-glycan analysis of rituximab purified from different B4Gal-T-KO clones and CHO-S-WT. Rituximab was harvested three days after transient transfection for purification, N-glycan labeling and analysis. Annotated N-glycan structures point at corresponding histogram peaks. VCD and cell viabilities are presented in bottom right panel.
Supplementary Figure 3: EPO N-glycan profiles and cell growth after transient transfection. N-glycan analysis of EPO purified from different B4Gal-T-KO clones and CHO-S-WT. EPO was harvested three days after transient transfection for purification, N-glycan labeling and analysis. Annotated N-glycan structures point at corresponding histogram peaks. VCD and cell viabilities are indicated in bottom right panel.
Supplementary Figure 4: IVC comparison of different cell lines during batch experiment. Two major groups of genotypes with (T1-2-3-KO, T1-2-3-4-KO, T1-3-KO) and without combinatorial disruption of B4Gal-T1 and -T3 are compared to IVC of WT and WT ctr clone.
Supplementary Table S1: sgRNA target sequences. The bases in red mark the PAM site.

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Target sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B4GalT1</td>
<td>TATCCCATTTGCAACCGGCGG</td>
</tr>
<tr>
<td>B4GalT2</td>
<td>GGACACCACCTAGCTATGG</td>
</tr>
<tr>
<td>B4GalT3</td>
<td>TGATGCTCGGGGCACGATGG</td>
</tr>
<tr>
<td>B4GalT4</td>
<td>TGGAGGCCGATCACCCTGG</td>
</tr>
</tbody>
</table>
Supplementary Table S2: Oligos for sgRNA expression vector cloning.

<table>
<thead>
<tr>
<th>Oligo name</th>
<th>Oligo sequence (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>gRNA_B4galt1_NW_003615120.1_283507_fwd</td>
<td>GGAAAGGACGAAAACACCTATCCCTACATTTTGCACGGCCTTTTAGAGCTAGAAAT</td>
</tr>
<tr>
<td>gRNA_B4galt1_NW_003615120.1_283507_rev</td>
<td>CTAAAACGCCGTGCACCAATTGGGATAGGGTGTCTTGCTCTTTCCACAAAGATAT</td>
</tr>
<tr>
<td>gRNA_B4galt2_NW_003613906.1_163085_fwd</td>
<td>GGAAAGGACGAAACACCGGGAGCAACCTAGCTATGTCTGTATAGAGCTAGAAAT</td>
</tr>
<tr>
<td>gRNA_B4galt2_NW_003613906.1_163085_rev</td>
<td>CTAAAACACATAGGGTGGGTGGCTCCCGGTGTTTCTCTTCCACAAAGATAT</td>
</tr>
<tr>
<td>gRNA_B4galt3_NW_003614301.1_233251_fwd</td>
<td>GGAAAGGACGAAACACCGGTGATGTCTGGGACCAGATTTTAGAGCTAGAAAT</td>
</tr>
<tr>
<td>gRNA_B4galt3_NW_003614301.1_233251_rev</td>
<td>CTAAAACACATCGTGCACGGAGCAACCTACGCTGTTCCTCTTCCACAAAGATAT</td>
</tr>
<tr>
<td>gRNA_B4galt4_NW_003614660.1_444021_fwd</td>
<td>GGAAAGGACGAAACACCGGGAGCAACCTAGCTATGTCTGTATAGAGCTAGAAAT</td>
</tr>
<tr>
<td>gRNA_B4galt4_NW_003614660.1_444021_rev</td>
<td>CTAAAACACATCGTGCACGGAGCAACCTAGCTATGTCTGTATAGAGCTAGAAAT</td>
</tr>
</tbody>
</table>
Supplementary Table S3: Cas9-2A_GFP and sgRNA plasmid ratios for transfections.

<table>
<thead>
<tr>
<th>Transfection #</th>
<th>Amount of transfected plasmid [μg]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GFP_2A_Cas9</td>
</tr>
<tr>
<td>1.1</td>
<td>1.9</td>
</tr>
<tr>
<td>1.2</td>
<td>1.9</td>
</tr>
<tr>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>1.9</td>
</tr>
<tr>
<td>2.2</td>
<td>1.9</td>
</tr>
<tr>
<td>2.3</td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Table S4**: Primer list for deep sequencing (MiSeq). The primers contain overhang sequences compatible with Illumina Nextera XT indexing (forward primer overhang: TCGTCGGCACGGTCGATGTATAAGAGACAG, reverse primer overhang: GTCTCGTGGGCTCGAGATGTGATGATAAGAGACAG).

<table>
<thead>
<tr>
<th>Primer Code</th>
<th>NCBI Reference</th>
<th>Oligo Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiSeq_B4gah1</td>
<td>NW_003615120.1</td>
<td>283507_fwd: TCGTCGGCACGGTCGATGTATAAGAGACAGAGCATTCTGTAACGAGT</td>
</tr>
<tr>
<td>MiSeq_B4gah1</td>
<td>NW_003615120.1</td>
<td>283507_rev: GTCTCGTGGGCTCGAGATGTGATGATAAGAGACAGAGCATTCTGTAACGAGT</td>
</tr>
<tr>
<td>MiSeq_B4gah2</td>
<td>NW_003613906.1</td>
<td>163085_fwd: TCGTCGGCACGGTCGATGTATAAGAGACAGAGCATTCTGTAACGAGT</td>
</tr>
<tr>
<td>MiSeq_B4gah2</td>
<td>NW_003613906.1</td>
<td>163085_rev: GTCTCGTGGGCTCGAGATGTGATGATAAGAGACAGAGCATTCTGTAACGAGT</td>
</tr>
<tr>
<td>MiSeq_B4gah3</td>
<td>NW_003614301.1</td>
<td>233251_fwd: TCGTCGGCACGGTCGATGTATAAGAGACAGAGCATTCTGTAACGAGT</td>
</tr>
<tr>
<td>MiSeq_B4gah3</td>
<td>NW_003614301.1</td>
<td>233251_rev: GTCTCGTGGGCTCGAGATGTGATGATAAGAGACAGAGCATTCTGTAACGAGT</td>
</tr>
<tr>
<td>MiSeq_B4gah4</td>
<td>NW_003614660.1</td>
<td>444821_fwd: TCGTCGGCACGGTCGATGTATAAGAGACAGAGCATTCTGTAACGAGT</td>
</tr>
<tr>
<td>MiSeq_B4gah4</td>
<td>NW_003614660.1</td>
<td>444821_rev: GTCTCGTGGGCTCGAGATGTGATGATAAGAGACAGAGCATTCTGTAACGAGT</td>
</tr>
</tbody>
</table>
Supplementary Table S5:

Deep sequencing results of single cell derived clones. Detailed sequencing analysis results for B4Gal-T1, -T2, -T3 and –T4 of generated clones. Clones selected for batch analysis are marked in red with out-of-frame indels in blue.
<table>
<thead>
<tr>
<th>#:</th>
<th>10-06-05</th>
<th>Biden</th>
<th>#:</th>
<th>10-06-05</th>
<th>Biden</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>18:07:34</td>
<td>18:07:34</td>
<td>4</td>
<td>18:07:34</td>
<td>18:07:34</td>
</tr>
<tr>
<td>5</td>
<td>18:58:12</td>
<td>18:58:12</td>
<td>7</td>
<td>18:58:12</td>
<td>18:58:12</td>
</tr>
<tr>
<td>9</td>
<td>20:05:00</td>
<td>20:05:00</td>
<td>11</td>
<td>20:05:00</td>
<td>20:05:00</td>
</tr>
<tr>
<td>23</td>
<td>00:05:34</td>
<td>00:05:34</td>
<td>25</td>
<td>00:05:34</td>
<td>00:05:34</td>
</tr>
<tr>
<td>24</td>
<td>00:22:45</td>
<td>00:22:45</td>
<td>26</td>
<td>00:22:45</td>
<td>00:22:45</td>
</tr>
<tr>
<td>25</td>
<td>00:39:57</td>
<td>00:39:57</td>
<td>27</td>
<td>00:39:57</td>
<td>00:39:57</td>
</tr>
<tr>
<td>26</td>
<td>00:57:08</td>
<td>00:57:08</td>
<td>28</td>
<td>00:57:08</td>
<td>00:57:08</td>
</tr>
<tr>
<td>27</td>
<td>01:14:20</td>
<td>01:14:20</td>
<td>29</td>
<td>01:14:20</td>
<td>01:14:20</td>
</tr>
<tr>
<td>30</td>
<td>02:05:54</td>
<td>02:05:54</td>
<td>32</td>
<td>02:05:54</td>
<td>02:05:54</td>
</tr>
<tr>
<td>31</td>
<td>02:23:06</td>
<td>02:23:06</td>
<td>33</td>
<td>02:23:06</td>
<td>02:23:06</td>
</tr>
<tr>
<td>32</td>
<td>02:40:17</td>
<td>02:40:17</td>
<td>34</td>
<td>02:40:17</td>
<td>02:40:17</td>
</tr>
<tr>
<td>33</td>
<td>02:57:29</td>
<td>02:57:29</td>
<td>35</td>
<td>02:57:29</td>
<td>02:57:29</td>
</tr>
<tr>
<td>34</td>
<td>03:14:40</td>
<td>03:14:40</td>
<td>36</td>
<td>03:14:40</td>
<td>03:14:40</td>
</tr>
<tr>
<td>35</td>
<td>03:31:52</td>
<td>03:31:52</td>
<td>37</td>
<td>03:31:52</td>
<td>03:31:52</td>
</tr>
<tr>
<td>37</td>
<td>04:05:75</td>
<td>04:05:75</td>
<td>39</td>
<td>04:05:75</td>
<td>04:05:75</td>
</tr>
<tr>
<td>40</td>
<td>04:57:09</td>
<td>04:57:09</td>
<td>42</td>
<td>04:57:09</td>
<td>04:57:09</td>
</tr>
</tbody>
</table>
**Supplementary Table S6: Gene-ID overview of targeted B4Gal-T-Isoforms.**

<table>
<thead>
<tr>
<th>Target</th>
<th>Gene-ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>B4Gal-T1</td>
<td>100689430</td>
</tr>
<tr>
<td>B4Gal-T2</td>
<td>100689434</td>
</tr>
<tr>
<td>B4Gal-T3</td>
<td>100689346</td>
</tr>
<tr>
<td>B4Gal-T4</td>
<td>100689435</td>
</tr>
</tbody>
</table>
Publication IV: Glyco-engineered CHO cell lines producing alpha-1-antitrypsin and C1 esterase inhibitor with fully humanized N-glycosylation profiles
Glyco-engineered CHO cell lines producing alpha-1-antitrypsin and C1 esterase inhibitor with fully humanized N-glycosylation profiles

Thomas Amann, Anders Holmgaard Hansen, Stefan Kol, Henning Gram Hansen, Johnny Arndorf, Saranya Nallaparedy, Bjorn Voldborg, Gyu Min Lee, Mikael Rørdam Andersen, Helene Fastrup Kildegaard

The Novo Nordisk Foundation Center for BioSustainability, Technical University of Denmark, Kgs. Lyngby, Denmark
Department of Biological Sciences, KAIST, Daejeon, Republic of Korea
Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark

ARTICLE INFO

Keywords
Chinese hamster ovary (CHO) cells
CRISPR/Cas9
Glyco-engineering
Biotechnology
Multiplying
Plasma proteins

ABSTRACT

Recombinant Chinese hamster ovary (CHO) cells are able to provide biopharmaceuticals that are essentially free of human viruses and have N-glycosylation profiles similar, but not identical, to humans. Due to differences in N-glycan moieties, two members of the serpin superfamily, alpha-1-antitrypsin (AAAT) and plasma protein C1 inhibitor (C1INH), are currently derived from human plasma for treating AAAT and C1INH deficiency. Deriving therapeutic proteins from human plasma is a cost-intensive process and also harbors a risk of transmitting infectious particles. Recombinantly produced AAAT and C1INH (rAAAT, rhC1INH) decorated with humanized N-glycans is therefore of clinical and commercial interest.

Here, we present engineered CHO cell lines producing rAAAT or rhC1INH with fully humanized N-glycosylation profiles. This was achieved by combining CRISPR/Cas9-mediated disruption of 10 gene targets with overexpression of human ST6GAL1. We were able to show that the N-linked glyco-structures of rAAAT and rhC1INH are homogeneous and similar to the structures obtained from plasma-derived AAAT and C1INH, marketed as Protasi®-C and Cinryze®, respectively. rAAAT and rhC1INH produced in our glyco-engineered cell line showed no detectable differences to their plasma-purified counterparts on SDS-PAGE and had similar enzymatic activity. The work presented here shows the potential of expanding the glyco-engineering toolbox for CHO cells to produce a wider variety of glycoproteins with fully humanized N-glycan profiles. We envision replacing plasma-derived AAAT and C1INH with recombinant versions and thereby decreasing our dependence on human donor blood, a limited and possibly unsafe protein source for patients.

1. Introduction

Chinese hamster ovary (CHO) cells serve an important role in the biotechnology industry as the primary workhorse for the production of recombinant protein therapeutics (Zhu, 2012). Many of these therapeutics are glycoproteins that contain one or more N-glycan and/or O-glycan chains. As N-glycans can potentially affect protein folding, immune regulation, cellular homeostasis and the biological half-life of proteins (Dalziel et al., 2014; Jennewein and Alter, 2017), it is considered a critical quality attribute and much effort has been put forth to...
improve features of protein N-glycosylation. The production of diverse N-glycan structures is a major contributor to the heterogeneity of protein products derived from CHO cells. The inherent heterogeneity of CHO N-glycan profiles is especially a drawback when one distinct N-glycan structure is desired on the protein product. Two examples of human proteins with distinct, homogeneous N-glycan structures are found within the serpin superfamily, alpha-1-antitrypsin (A1AT) and C1 esterase inhibitor (C1INH) (Clerc et al., 2015). Patients with the genetic disorders alpha-1-antitrypsin deficiency (AATD) or hereditary angioedema (HAE-C1INH) have decreased plasma levels of functional A1AT or C1INH, respectively, and are currently treated with prophylactic augmentation therapy of plasma purified A1AT or C1INH (Stoller and Abousouan, 2005; Feusner et al., 2014). Augmentation therapy is cost intensive (Stoller and Abousouan, 2005), and C1INH purified from pooled donor plasma has been associated with hepatitis C virus infections prior to the introduction of virucidal methods (Mannucci et al., 1995). Despite current dedicated virus inactivation steps, cases of Hepatitis G transmission have been reported (Filippi et al., 1998) and non-enveloped viruses can still be transmitted via plasma-derived products (Souci et al., 2013). Nevertheless, approved C1INH formulas are concentrates purified from human donors (Berinert®, Cinryze) despite containing undesired protein impurities identified as α-1-antichymotrypsin, ceruloplasmin and Factor C3 (Feusner et al., 2014).

Both native human plasma A1AT (pA1AT) and C1INH (pC1INH) possess an N-glycan profile with ~60-80% diantennary, disialylated, non-fucosylated (A2G2S2) structures with human-like α2-6-linked sialic acids. pA1AT has three N-glycosylation sites (Clerc et al., 2015), while natural A2G2S2 structures are not essential for biological activity but enhance in vivo half-life and in vitro protein stability (Karnasheva et al., 2006; Kwon and Yu, 1993). pC1INH is thought to be one of the most heavily glycosylated plasma proteins and harbors ten O-linked and six N-linked glycan structures (Stavenhagen et al., 2017). The six N-linked glycan moieties with A2G2S2 as predominant structure have been shown to be increase serum half-life and are reported to increase in vivo efficacy (Minta, 1981; Longhurst, 2008; van Doorn et al., 2005). The ability to generate high A2G2S2 N-glycan proportions on rA1AT/ rC1INH may be critical for improving product quality. Although efforts have been reported, generating higher proportions of A2G2S2 N-glycans has not to our knowledge been published. For instance, CHO cells lack active Sβ6 beta galactoside alpha-2,6-sialyltransferase 1 (ST6GAL1) to cap N-glycans with alpha-2,6-linked sialic acids (Lewis et al., 2013). rA1AT and rC1INH were produced in various platforms (Wright et al., 1991; Ziemek, 1998; Moir and Dumas, 1987; Khutsi et al., 2017; Jaberie and Ngehibalinamahia, 2016; Castilho et al., 2014; Niklas et al., 2013; Blanchard et al., 2011; Paterson et al., 1994; Lee et al., 2013; Bos et al., 2003; Lamark et al., 2001; Wolff et al., 2001). However, these approaches revealed low productivity or the N-glycosylation was far from the profile of pA1AT or pC1INH and therefore clinical trials from the human blood making intravenous administration impractical.

Commercially available rC1INH from transgenic rabbits shows activity similar to pC1INH and has decreased virus transmission risk. However, it differs in N-glycosylation profiles from pC1INH and therefore reveals a risk of allergy, a dissatisfactory pharmacokinetic profile and consequently is unlikely to be of use in prophylaxis (Frank, 2010). Glycosylation-engineering in primary human cells aimed to mimic O-glycan profiles of pC1INH. However, complete sialylation of rC1INH N-glycan structures was not achieved (Wissing et al., 2015).

To address this, we aimed to engineer the heterogeneous CHO-S N-glycan profile towards a predominantly non-fucosylated biantennary A2G2 structure by combining the publicly available CHO-K1 genome sequence (Xu et al., 2011), clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 (CRISPR/Cas9) for multiplexing gene editing (Grav et al., 2015), and reported CHO glycosyltransferases (Yang et al., 2015). To this end, we made functional knockouts of the Glu-gene and nine glycosylation-gene targets (10× KO, Suppl. Table 1). We hypothesized that rHA1AT and rHC1INH produced in this genetic background with parallel co-expression of ST6GAL1 would display an N-glycan profile similar to pHA1AT and pHC1INH with predominant A2G2S2 N-glycan structures. We present CHO clones producing rHA1AT/rHC1INH similar to pHA1AT/pHC1INH when assessed by N-glycan analysis, protein activity, SDS-PAGE and isoelectric focusing. CHO derived rHA1AT and rHC1INH with fully humanized N-glycan profiles have the potential to replace the cost-intensive and possibly unsafe plasma-based augmentation therapy of AATD and HAE-C1INH patients without compromising activity and N-glycosylation.

2. Materials and methods

2.1. sgRNA, GFP, 2A,Cas9 and A1AT/C1INH-ST6GAL1 Glul plasmid design

GFP, 2A,Cas9 and single guide RNA (sgRNA) plasmids were constructed as previously described (Grav et al., 2015). The sgRNA target sequences for Mga4A, Mga4B, Mga5, S3ga3A, S3ga4A, S3ga6A, B3m2, Fust, Splp3l and Glui was performed using "CRISPY" (Ronda et al., 2014). The target sites for the mentioned genes and the oligos for sgRNA cloning are listed in Suppl. Table S1 and Table S2, respectively.

Plasmids for co-expression of A1AT/C1INH and ST6GAL1 were constructed with well-specific excision reagent cloning method as previously described (Pristovsek et al., 2018; Lund et al., 2014) (Suppl. Fig. 1). The DNA sequences of the plasmids are listed in Suppl. Table S5.

2.2. Cell cultivation and transfection for genome editing

CHO-S suspension cells were incubated in a humidified incubator at 120 rpm, 37 °C, 5% CO2, passed to 2–3 × 106 cells/mL every 2–3 days and transfected in 6-well plates (BD Biosciences, San Jose, CA) as described previously (Grav et al., 2015). The GFP, 2A,Cas9/sgRNA plasmid ratios for each transfection was 1:1 of which the plasmid load of sgRNA was divided equally by the amount of different sgRNAs used per transfection (Suppl. Table S4). To measure FACS sorting efficiency, pmxGFP vector (Lonza, Basel, Switzerland) transfection was performed as well. Cells were harvested for fluorescence-activated cell sorting (FACS) 48 h post transfection.

2.3. Single cell cloning of genome edited cells using FACS

Prior to FACS, cells were filtered through a 40 µm cell strainer into a FACS-compatible tube.

Single fluorescent-positive (GFP) cells were sorted into 384-well plates (Corning, New York, NY) containing 30 µL CD CHO medium supplemented with 8 mM l-alanine, 1.5% HEPES buffer and 1% Antibiotic-Antimycotic ( Gibco, Waltham, MA) per well as described previously (Hansen et al., 2016). For cell sorting, fluorescent-positive cell populations were gated based on non-transfected WT CHO-S cells. Two weeks after cell sorting cell colonies were moved to 96-well flat-bottom plates (BD Biosciences) and expanded for deep sequencing analysis and batch cultivation.

2.4. Deep sequencing analysis

Confluent colonies from 96-well flat-bottom replicate plates were harvested for genomic DNA extraction. DNA extraction was performed using QuickExtract DNA extraction solution (Epicentre, Illumina, Madison, WI) according to the manufacturer's instructions. The library preparation was based on Illumina 16s Metagenomic Sequencing Library Preparation and deep sequencing was carried out on a MiSeq Benchtop Sequencer (Illumina, San Diego, CA). The protocol for amplifying the targeted genomic sequences, amplicon purification, adapter-PCR and following quality analysis was based on previously
2.5. Transfection and expression in polycyonal cell lines by applying MSX-selection

Cells were seeded in 250 mL Corning vent cap shake flasks (Sigma-Aldrich) as duplicates with cell densities \( \sim 1 \times 10^5 \) cells/mL in 60 mL CD CHO medium supplemented with 8 mM \( \gamma \)-glutamine (Life Technologies) and transfected with 75 \( \mu \)g of pAAT-Gal-STG6GAL1 plasmid or 75 \( \mu \)g of rCINH-Gal-STG6GAL1 plasmid (Suppl. Fig. 1) using Freestyle® MAX reagent together with OptiPRO SFM medium (Life Technologies) according to the manufacturer's recommendations. 1 \( \mu \)L/ml anti-clumping agent was added 24 h after transfection. pmaxGFP* vector (Lonza) transfection was performed to measure transfection efficiencies. Two days after transfection, cells were transferred into 60 mL CD CHO medium lacking \( \gamma \)-glutamine (Life Technologies) and supplemented with 1 \( \mu \)L/ml anti-clumping agent and 0 \( \mu \)M, 10 \( \mu \)M, 30 \( \mu \)M or 50 \( \mu \)M MSX (EMD Millipore, Billerica, MA).

Cell densities and viabilities were determined once per day using the NucleoCounter NC-250 Cell Counter (ChemoMetec). The cells were passaged in fresh selection medium every 2-3 days until viability and doubling time reached stable values. Polyclonal cell lines (pools) were seeded in duplicates at \( \sim 1 \times 10^5 \) cells/mL with corresponding MSX concentrations. Cell densities and viabilities were determined once per day and supernatants of the pools were harvested three days after seeding and pooled within duplicates for purification of rAAT and rCINH.

2.6. Single cell cloning of cells from polycyonal cell pools using FACs

Non-stained single cells were sorted from pools as described above. For cell sorting, all viable cells were gated for sorting into 384-well plates with \( \gamma \)-glutamine-free medium. Two weeks after cell sorting the clones were moved to 96-well flat-bottom plates (BD Biosciences) and expanded to shake flask format in CD CHO medium supplemented with 1 \( \mu \)L/mL anti-clumping agent, 25 \( \mu \)M MSX and lacking \( \gamma \)-glutamine.

2.7. Screening cell pools and single cell clones for human-like a2,6-sialic acid linkage formation with lectin staining

For lectin staining of cells, triplicates of 10,000 cells per sample were diluted in 200 \( \mu \)L of 0.22 \( \mu \)m pore size filtered CD CHO medium (Life Technologies) supplemented with 5 \( \mu \)g/mL Hoechst 33342 (Merck, Darmstadt, Germany) and 1 \( \mu \)g/mL Fluorescein isothiocyanate (FITC) labeled Sambucus nigra agglutinin (SNA) lectin (Biomers, Hamburg, Germany). After 60 min incubation in the dark at 37 \(^\circ\)C and 5% CO\(_2\), the cells were washed with 200 \( \mu \)L CD CHO medium and then washed twice with 200 \( \mu \)L phosphate buffered saline (PBS) (300 g, 5 min, RT). The samples were resuspended in 200 \( \mu \)L PBS and transferred to 96-well plate for final centrifugation at 300 g for one minute. The percentage of FITC SNA positive cells was determined in a 96-well optical-bottom microplate (Greiner Bio-One, Frickenhausen, Germany) using a Geligo Imaging Cell Cytometer (Nexcelom Bioscience, Lawrence, MA). Cells were identified using the blue channel (Hoechst-positive cells), and the green channel (FITC SNA-positive cells) was used to detect cells with alpha-2,6-sialic acid linkage. A Hoechst/FITC SNA-stained CHO-S WT sample was gated to distinguish between FITC-positive and FITC-negative cells.

2.8. Batch cultivation: cell growth analysis and N-glycosylation profiling

For batch cultivation and N-glycan analysis, cells were seeded at 0.4 \( \times \) 10^5 cells/mL in 250 mL Corning vent cap shake flasks (Sigma-Aldrich, St. Louis, MI) as duplicates in 60 mL CD CHO medium supplemented with 1 \( \mu \)L/ml anti-clumping agent (Life Technologies). CHO-S WT and non-producing parental 10 \( \times \) KO cell lines were additionally supplemented with 8 mM \( \gamma \)-glutamine. rAAT/rCINH producing clones were cultivated in \( \gamma \)-glutamine-free medium at all times and passaged in medium containing 25 \( \mu \)M MSX until the batch cultivation was initiated. Cell densities and viabilities were determined once per day using the NucleoCounter NC-250 Cell Counter (ChemoMetec) until the viability was < 70%, at which point the culture was terminated. Supernatant samples with total secreted protein (secretome) from CHO-S WT and parental, non-producing 10 \( \times \) KO cell lines were taken five days after seeding and pooled within biological replicates. The volume for secretome samples was calculated to harvest 20 \( \times \) 10^5 cells. For all shake flasks, additional supernatant samples were taken by centrifuging 1 mL of cell suspension for 5 min at 1000 g and storage of supernatant at -80 \(^\circ\)C until further analysis.

2.9. rAAT and rCINH purification

rAAT and rCINH were purified using CaptureSelect affinity resins (Thermo Fisher Scientific) according to the manufacturer's instructions. rAAT was further purified by size exclusion chromatography on a Superdex 200 increase 10/300 GL column (GE Healthcare) equilibrated in PBS.

2.10. Titre assessment of rAAT/rCINH producing clones

rAAT and rCINH titers were determined using biolayer interferometry on an Octet RED96 ( Pall, Menlo Park, CA, USA) as described previously for AAT (Noth et al., 2017). After hydration in PBS, streptavidin biosensors (18-5021, Fortebio, Pall) were functionalized with CaptureSelect biotin anti-AAT conjugate or CaptureSelect biotin anti-CINH conjugate (Thermo Fisher Scientific) at 5 \( \mu \)g/mL in PBS, and blocked in PBS containing 1 \( \mu \)g/mL biotin (600 and 300 s incubation steps, respectively). Standards were prepared in spent CHO-S medium using plasma-derived AAT (Athens Research & Technology) at 100, 50, 25, 12.5, 6.3, 3.1, and 1.6 \( \mu \)g/mL or CINH (R&D systems) at 40, 20, 10, 5, 2.5, 1.25, and 0.625 \( \mu \)g/mL. Samples and standards were diluted two-fold and contained 0.1% BSA w/v, 0.1% tween-20 w/v, and 500 mM NaCl. When needed, samples were further diluted to fall within the range of the standard dilution series. After equilibration in spent CHO-S medium (120 s), samples and standards were measured for 300 s with a shaking speed of 1000 rpm at 30 °C. Regeneration was performed with 50 mM TRIS, 2 M MgCl\(_2\), pH 7.5. Assays were performed in 96-well black microplates (Greiner Bio-One, Kremsmunster, Austria). Octet System Data Analysis 7.1. software was used to calculate binding rates and absolute AAT and CINH concentrations.

2.11. SDS-PAGE, isoelectric focusing and PNGase treatment

SDS-PAGE was performed on Novex 4–12% Tris-Glycine mini gels and isoelectric focusing (IEF) was performed on Novex pH 3–10 IEF gels (Thermo Fisher Scientific) as per the manufacturer's instructions. Deglycosylation with PNGase F was performed according to the manufacturer's instructions (New England Biolabs, Ipswich, MA).

2.12. Activity assays

AAT inhibitory activity was determined using the EnzChek Elastase Assay Kit (Molecular Probes, Eugene, OR) according to the manufacturer's instructions. In short, AAT (8.0, 4.0, 2.0, 1.0, 0.5, 0.25, 0.13, and 0.06 \( \mu \)M) was incubated with purified active porcine pancreatic elastase and fluorescein labeled substrate (DQ elastin). Measurement of fluorescence was performed after 45 min at room temperature (Excitation: 485 nm, slit width 9.0 nm; Emission: 530 nm, slit width 13.5 nm).

CINH inhibitory activity was determined using the Technochrom CINH Assay Kit (Technoclone, Vienna, Austria). In short, plasma
containing CIINH activity (120%, 60%, 30%) and samples (~0.25 μM) were incubated with substrate-buffer mixture for 3 min at room temperature, after which 50% acetic acid was added. Extinction was measured at 405 nm.

2.13. N-glycan analysis

N-glycans were derivatized with GlycoWorks RapiFluor-MS N-Glycan Kit (Waters, Milford, MA) according to the manufacturer’s instruction. Briefly, 12 μL purified protein or 12 μL of 10× concentrated (Amicon Ultra-15, Merck) secretion sample was used for each sample. Labeled N-Glycans were analyzed by LC-MS as described previously (Grav et al., 2015). Separation gradient from 30% to 43% 50 mM ammonium formate buffer and MS were run in positive mode. Amount of N-Glycan was measured by integrating the peaks with Thermo Xcalibur software (Thermo Fisher Scientific, Waltham, MA) giving the normalized, relative amount of the glycans.

2.14. Proteomics

Preparation of protein extract from CHO cells were done as previously described in Bonde et al. (2016). Liquid chromatography was performed on a Cap-LC system (Thermo scientific) coupled to an 75 μm × 15 cm 2 μm C18 easy spray column (Thermo Scientific). The flow rate was 1.2 μL/min and a stepped gradient, going from 4% to 40% acetonitrile water over 50 min. The samples were sprayed into an Orbitrap Q Exactive HF-X mass spectrometer (Thermo Scientific). MS-level scans with resolution set to 60,000; Data dependent MS2 selection with HCD collision energy set to 28%. The resulting data were analyzed using MaxQuant. Data was searched against the Chinese hamster database retrieved from Uniprot with proteome ID UP000001075.

3. Results

3.1. Growth profile and N-glycan profile of clonal 10× KO cell lines

The aim of our study was to produce rhA1AT and rhCIINH in CHO cells with N-glycans profiles similar to human pA1AT and pCIINH. Our approach was to engineer the heterogenous N-glycan profile of CHO-S WT cells towards a homogeneous A2G2S2 N-glycan structure, which is the predominant N-glycan on pA1AT and pCIINH. To this end, we generated out-of-frame insertions or deletions (indels) in eight glycosyltransferase genes (Mgat4A, Mgat4B, Mgat5, St3Gal3, St3Gal4A, St3Gal4B, B3gnt2, Fut8) as well as in the genes Ssp3 and Glu (Suppl. Table S4) over four successive rounds of multiplexed CRISPR/Cas9 gene editing. Two clones with indels in the targeted genes were subjected to growth analysis and N-glycan profiling.

Two clones (10× KO A and 10× KO B) with out-of-frame indels in all ten gene targets were obtained and both showed a pronounced increase in batch culture longevity when compared to the parental CHO-S WT cell line (Fig. 1A).

CHO-S WT reached maximal viable cell density of ~6 × 10⁶ cells/ml on day five and cell viability declined rapidly to ~50% on day 6. In contrast, the 10× KO A and 10× KO B clones had cell viabilities >75% until day 10 of the batch cultivation and reached higher maximal viable cell density than CHO-S WT.

N-glycan analysis of the CHO-S WT secretome resulted in more than 14 annotated N-glycan structures (Fig. 1B) where the A2G2S2 structure, predominantly found on pA1AT and pCIINH, was not detected. The majority of CHO-S WT N-glycans contained core-fucosylation. The N-glycans produced by CHO-S WT cells appear diverse and comprise high-mannose structures as well as non-galactosylated, fully and partially sialylated di-, tri- and tetra-antennary structures (all with alpha-2,3-linked sialic acids). A2G2S2 was found as the main N-glycan on total secreted proteins of CHO-S WT. In contrast, the N-glycan profiles of 10× KO A and 10× KO B are more homogenous (Fig. 1B) with all structures lacking core-fucosylation. In addition, only relatively small amounts of CHO-specific alpha-2,3-linked sialylation were present.

After disruption of the targeted genes, the proportion of A2G2 within N-glycan structures of total secreted proteins was increased from 3.5% (CHO-S WT) to 79% in both 10× KO clones (Fig. 1C). We concluded that the 10× KO A and B clones were suitable host cell lines in our effort to generate humanized N-glycans.

3.2. Introducing human-like sialylation in 10× KO cell lines

On the basis of A2G2 secretome N-glycan structures of clone 10× KO B, we aimed to develop clonal cell lines expressing St6GAL1 and rhCIINH or St6GAL1 and rhA1AT. We envisioned that such cell lines are capable to produce rhA1AT or rhCIINH with predominant A2G2S2 N-glycan structures as found on pA1AT and pCIINH. The functional Glu-KO selection system was confirmed by MSx-dosage dependent recovery times of cell viabilities from transfected cell pools (Suppl. Fig. 3A). Passageing of the different transfection pools was performed until viability and doubling times were stable. We then conducted FACs-based single cell cloning with the 50 μM MSx-selected cells. During the expansion of the generated clones, only clones exhibiting predominant FITC-SNA staining and detectable levels of rhA1AT/rhCIINH in supernatants on coomassie-stained SDS-PAGE gels were selected (Suppl. Fig. 2). Based on these criteria, two rhA1AT (A1-1 and A1-2) and two rhCIINH (C1-1 and C1-2) producing clones were selected for further characterization.

SNA lectins are reported to bind predominantly to sialic acids of N-glycans linked to the galactose residue in a human-like alpha-2,6-sialylation. Analyzing FITC-SNA-stained CHO-S WT, we found relatively low levels of alpha-2,6-sialylation (Fig. 2A). To determine the proportion of cells with human-like sialylation, FITC-SNA stained CHO-S WT cell lines were used to gate between FITC-positive and FITC-negative cells (Suppl. Fig. 2A). Within the 50 μM MSx-selected polyclonal cell lines, <30% of the cells were found to comprise alpha-2,6-linked sialic acids on N-glycans of cell surface proteins (Fig. 2B). In comparison, 82–90% of the cells in the populations of the selected four clones (A1-1, A1-2, C1-1 and C1-2) had the desired alpha-2,6-linked sialic acids on their N-glycans.

3.3. Extended culture longevity retained in rhA1AT and rhCIINH-producing clones

Both the two rhA1AT- and the two rhCIINH-producing clones showed the extended culture longevity as well as maximum VCD similar to the parental 10× KO B cell line (Fig. 3A). Viabilities of all clones were >75% until day 9 and clone C1-2 maintained cell viability >75% until day 11, similar to the non-producing parental 10× KO B. In both rhCIINH-producing clones, rhCIINH titers increased from day 0 until day 5, but then stagnated at ~40 μg/ml until the end of the batch cultures (Fig. 3B). In comparison, rhA1AT titers from clones A1-1 and A1-2 increased continuously to 123 μg/ml and 117 μg/ml, respectively. Despite increasing numbers of viable cells, the stagnation of CIINH titers at ~40 μg/ml for clones C1-1 and C1-2 in the second half of the batch cultures leads to the assumption that rhCIINH is unstable in the cell culture. By SDS-PAGE gel analysis of late phase supernatant samples we observed protein bands migrating just below rhCIINH (~70-100 kDa) which were not pre(Suppl. Fig. 4A).

On the basis of rhA1AT/rhCIINH concentration in the cultivation supernatant and integral of viable cells (IVC), we determined the average specific productivity of the four clones during day 2–5 and day 6–9 (Fig. 3C). In the early phase, the specific productivity of rhA1AT for clones A1-1 and A1-2 was 5.8 and 4.0 pg/cell*day, respectively, decreasing to 2–3 pg/cell*day in the late phase (day 6–9). On the contrary, the two rhCIINH-producing clones expressed rhCIINH at ~5 pg/cell*day in the early phase, whereas their specific productivity in the
Fig. 1. Growth and N-glycan structure analysis of CHO-S WT and 10x KO cell lines. (A) Viable cell density (VCD) and viability of batch cultures of CHO-S WT and two clonal cell lines (10x KO A and 10x KO B) with indels in eight glycosyltransferases as well as Ghb and Sspt. Error bars indicate the standard deviation of three replicate parallel cultures. (B) N-glycan profiling of total secreted proteins from CHO-S WT and the 10x KO A and 10x KO B clones. In the chromatogram, elution time indicated on the x-axis and y-axis represents signal intensity normalized to highest peak. (C) Proportion of non-fucosylated, biantennary N-glycans with terminal galactose (A2G2) in total secreted proteins from CHO-S WT and the 10x KO A and 10x KO B clones.
late phase decreased to ~0 pg/cell·day. Overall, cell growth of all four clones was comparable to the parental cell line whereas product titers of A1AT-producing clones were increased compared to C1INH-producers.

3.4. Activity and N-glycosylation profile of CHO-produced rhA1AT and rhC1INH are similar to plasma-derived products

We purified and characterized rhA1AT and rhC1INH to investigate the impact of our N-glycosylation engineering approach on product quality and protein activity. Therefore we compared the protein products produced in clonal cell lines derived from 10 × KO B (rha1AT and rhC1INH) to the CHO-S WT and plasma-derived counterparts (pLAI1AT and pC1INH).

SDS-PAGE gel analysis revealed that purified rhA1AT and rhC1INH produced in the four clones seem to have hydrodynamic volumes (molecular weight) similar to pLAI1AT and pC1INH without detectable impurities as seen in pC1INH (Fig. 4A). rhA1AT and rhC1INH produced

Fig. 2. FITC-SNA lectin staining of selected poly- and monoclonal cell lines. (A) Fluorescent images of CHO-S WT and A1-1 cell line. Cells were stained for alpha-2,6-sialic acid linkage with FITC-SNA (green) and for nuclei with Hoechst (blue). The bottom right corner bar displays a length of 500 µm. (B) Comparison of FITC-SNA positive cells. FITC-SNA lectin staining of CHO-S WT, two 50 µM MIAA polyclonal cell lines and four selected clones. Bars indicate the proportion of cells with positive FITC signal due to SNA lectin binding on alpha-2,6-linked sialic acids on the cell surface. Error bars represent standard deviation of three individual measurements per sample.

Fig. 3. Growth profiles, product titers and specific productivities of selected producing and non-producing clones. (A) Viability cell densities and cell viabilities of CHO-S WT, 10 × KO B the rha1AT (A1-1 and A1-2) and rhC1INH (C1-1 and C1-2) producing clonal cell lines measured in batch cultures. Error bars indicate range of duplicate parallel cultures. (B) rhA1AT and rhC1INH titer in supernatants during the batch culture experiment. Error bars indicate standard deviation of three individual measurements from two shake flasks per clone. (C) Specific productivities of the rha1AT and rhC1INH-producing clonal cell lines in the batch culture experiment. Average specific productivity was calculated from day 2 to 5 and from day 6 to 9. Colored symbols represent average measured specific productivity for shake flask duplicates. Black lines show the average specific productivity based on the three measurements of shake flask duplicates.
in CHO-S WT background did not co-migrate with plA1AT and plC1INH, respectively. However, after deglycosylation with PNGaseF, all re-combinantly produced proteins aligned with corresponding bands of plA1AT and plC1INH with the exception of rC1INH produced in a CHO-S WT background displayed an additional protein band at ~65 kDa.

To further characterize the CHO produced rHA1AT and rHC1INH, we performed IEF gel analysis (Fig. 4B). rHA1AT from clones A1-1 and A1-2 manifested in two bands with isoelectric points (pl) at ~pH 4.5 similar to plA1AT. In contrast, rHA1AT produced in a CHO-S WT background displayed more than nine detectable isomers with pl ranging between pH 4 - 5.

IEF gel analysis of rHC1INH produced in a CHO-S WT background resulted in isomers with pl ranging from pH 4 to 5.5. A high degree of heterogeneity was also found in purified rHC1INH produced in clone C1-1. However, rHC1INH produced in clone C1-2 was less heterogeneous with pl ~ 3.5 similar to pC1INH. In N-glycan analysis of purified rHA1AT and rHC1INH from CHO-S WT cells we detected a higher degree of heterogeneity compared to N-glycan structure of rHA1AT and rHC1INH from polyclonal 10 × KO cell pools (Suppl. Fig. 3B). The polycyonal cell lines revealed two predominant sugar structures on both proteins (A2G2 and A2G2S2 N-glycans), whereas we could not detect the A2G2S2 structure on products from CHO-S WT. Moreover, the amount of predominant N-glycan structures on rHA1AT and rHC1INH was decreased from two (polyclonal pools) to one (monoclonal producers), identified as A2G2S2 N-glycan (Fig. 4C).

All four 10 × KO-derived monoclonal cell lines produced rHA1AT and rHC1INH with higher proportion of A2G2S2 structures than plA1AT and plC1INH (Suppl. Fig. 2C and Suppl. Fig. 3C). The proportion of A2G2S2 in rHA1AT and rHC1INH was approximately 88-92% and 84%, respectively, and 82% for plHA1AT and 66% for plC1INH.

Finally, we investigated the activity of purified rHA1AT and rHC1INH activity was determined by its inhibitory function of elastase activity (Fig. 4D, Suppl. Fig. 4B). Similar to plA1AT and rHA1AT from CHO-S WT, a decrease in elastase activity was detected at A1AT concentrations 0.1 μM for rHA1AT from clones A1-1 and A1-2. In addition, 50% of elastase inhibition was reached at ~0.3 μM A1AT for plA1AT as well as rHA1AT. In vitro activity of purified rHC1INH produced by clones C1-1 and C1-2 was similar or higher compared to pC1INH.

4. Discussion

We aimed to produce rHA1AT and rHC1INH in CHO-S with N-glycan profiles similar to plHA1AT and pC1INH, which to our knowledge has not yet been achieved by recombinant expression. First, the heterogeneous N-glycan profile of CHO-S WT cells was changed to more homogenous profiles in bespoke cell lines with predominant A2G2 N-glycan structures (Fig. 1B). Disrupting nine N-glycosylation-related genes increased the A2G2 proportion on total secreted protein from 3.5% in CHO-S WT-derived cells to ~80% in 10 × KO cell lines (Fig. 1C). This supports the previously suggested strategy to decrease N-glycan branching and alpha-2,3-sialylation by disrupting MgtAT4A, MgtAT4B, MgtAT5, St3gaI3, St3gaP4 and St3gaI6 (Yang et al., 2015). The impact of gene disruptions on cell culture performance was assessed in batch cultures. Interestingly, the monoclonal cell lines with disruption in ten gene targets showed enhanced growth characteristics compared to CHO-S WT cells (Fig. 1A). Overexpression of the G418 gene has previously been found to decrease ammonia levels, which might explain improved CHO cell growth of the four characterized producer clones in l-glutamine-free medium (Noh et al., 2017). However, the cause for the boosted cell growth of G418-lacking 10 × KO cell lines in l-glutamine-supplemented medium (Fig. 1A) remains to be explored in further studies. Additionally, when CHO-S cells were subjected to a single cell sorting-based selection procedure, we suspect that the protocol generates monoclonal cell lines with an extended culture longevity phenotype compared to the parental cell line. Thus, we do not suggest that knocking out the 10 gene targets leads to the observed extended culture longevity phenotype.

Since the disruption of the ten targets did not seem to interfere with cell culture performance, we performed co-expression of ST6GAL1 and rHA1AT or ST6GAL1 and rHC1INH in the 10 × KO-derived clone B. After transfection, we observed a MSX-concentration dependent recovery of the transfected cell pools and successful killing of untransfected 10 × KO B after 5 days of growth in l-glutamine-free medium similar to a previous study (Suppl. Fig. 3A) (Noh et al., 2018). However, untransfected CHO-S WT cells were also able to recover up to the highest MSX-concentration of 50 μM, which is in accordance to previous work (Pristovéck et al., 2018). As shown in the killing curve of the untransfected 10 × KO cells, the advantage of the G418 KO system here seems to be the elimination of untransfected cells.

Surprisingly, after selection at 50 μM MSX, the polyclonal cell lines did not show the desired predominant A2G2S2 glycosylation of purified rHA1AT and rHC1INH as we found incomplete sialylation on both proteins (Suppl. Fig. 3B). Similar lack of sialylation in the polyclonal cell lines was found after FITC-SNA lectin staining where only <30% of cells were identified to have alpha-2,6-sialylation (Fig. 2B). This might be due to incomplete vector integration into the genomic DNA or chromosomal instability leading to a heterogeneous cell population with reduced stability of the integrated elements as reported earlier (Jun et al., 2006). The cell pools most likely harbor a mixture of cells which do not all express the ST6GAL1-protein successfully. However, we were able to discard clones with incomplete sialylation by single cell cloning and screening for FITC-SNA lectin positive clones producing rHA1AT or rHC1INH.

Although the stagnating rHC1INH titer at day 5, the reported rHC1INH titer here is higher than previously reported production platforms with maximum titers of only 6 μg/mL in insect cells (Wolff et al., 2001) and 30 μg/mL in P. pastoris (Bos et al., 2003) albeit the media we used was without any optimization for protein production. Further investigation needs to be carried out to understand this stagnation of rHC1INH titers. However, it can be speculated that the rate of rHC1INH production is similar to the rate of rHC1INH degradation, preventing an increase of rHC1INH titer in the cultivation supernatant. In comparison to pC1INH (Cárayez), the rHC1INH from C1-1 and C1-2 did not show α1-antichymotrypsin, ceruloplasmin and Factor C impurities on SDS-PAGE gels (Filippi et al., 1998) (Fig. 4A), implying that CHO-based cell platforms possibly can supply HAE-C1INH patients with higher purity than human plasma.

With titers between 300 and 400 μg/mL, human neuronal cell lines produce rHA1AT with higher titers than our clones A1-1 and A1-2. However, rHA1AT from these neuronal cells exhibits core-fucosylation,
is not fully sialylated, and therefore differs largely from plAIAT N-glycosylation (Blanchard et al., 2011). Similarly, earlier studies expressed rhAIAT in CHO with titers of up to 1.15 g/L, though differing from plAIAT by revealing core-fucosylation and alpha-2,3-sialylation (Paterson et al., 1994; Lee et al., 2013; Chin et al., 2013).

In contrast to the production platforms listed earlier, rhAIAT and rhC1INH produced in our 10× KO cells are not only exceeding sialylation levels of plAIAT and pC1INH (Suppl. Fig. 2C), but also reveal human-like alpha-2,6-sialylation instead of alpha-2,3-sialylation. Previous work reported only 2.6% A2G2S2 structures on rhAIAT expressed in CHO-K1, however with 2,3- sialic-acid linkage (Lee et al., 2013).

Interestingly, the increased sialylation of rhAIAT from the two clones had no impact on in vitro activity (Fig. 4D). This is in accordance with previous work, which showed that AIAT activity is not linked to its N-glycosylation and CHO WT produced rhAIAT has similar activity to plAIAT (Karnaukhova et al., 2006; Hansen et al., 2016). Furthermore, differences of CHO-S WT- and 10× KO-derived rhAIAT were made visible using IEF gel analysis, where rhAIAT from the two clones revealed similar patterns to plAIAT (Fig. 4B).

As presented in Fig. 4B, in contrast to rhAIAT, rhC1INH differed partially in IEF gel analysis profile from pC1INH. Increased rhC1INH sialylation (Suppl. Fig. 2C) might lead to altered charge distribution and consequently cause changes in IEF gel pattern. For rhC1INH from clone C1-1, we observed a double-band in SDS-PAGE analysis although N-glycans were removed by PNGaseF treatment (Fig. 4A). This might indicate that O-glycan charge variants are responsible for the heterogeneity observed in IEF gel analysis and SDS-PAGE of rhC1INH from C1-1 (Figs. 4A and 4B) as described previously (Hansen et al., 2015).

In summary, our work describes a strategy to successfully engineer the heterogeneous N-glycosylation profile of CHO-S WT cells to the specific A2G2S2 N-glycan structure with the purpose of producing rhAIAT and rhC1INH with N-glycan profiles similar to human plasma-derived products. We used CRISPR/Cas9 to disrupt ten genes and then overexpressed rhAIAT or rhC1INH on a STSGLA1- and Glu-encoding plasmid. After selection with MSX and single cell cloning, we identified clones expressing rhAIAT or rhC1INH with titers of up to 124 µg/mL and 42 µg/mL, respectively (Fig. 3B). Purified rhAIAT and rhC1INH were similar to the plasma-derived counterparts judged by SDS-PAGE analysis (Fig. 4A), degree and type of sialylation (Fig. 2B, Suppl. Fig. 2C, Suppl. Fig. 3C) and in vitro activity (Fig. 4D). Thus, the work presented shows the promise and potential of replacing cost-intensive and possibly unsafe plasma-derived augmentation therapy for AATD and C1INH-HAE patients by CHO-produced rhAIAT and rhC1INH. This strategy is in compliance with the Medical and Scientific Advisory Council (MASAC) recommendation of replacing plasma-derived products with recombinant products for treatment of diseases (National Hemophilia Foundation, 2014).

Acknowledgement

The authors thank Søren Petersen Bjørn, Yuuhou Fan and Patrice Menard for valuable guidance and support. The authors thank Karen Katrine Brendum, Nachon Charnnyanonda Peterson, Karoline Schousboe Frenming and Zafiya Sukhova for excellent technical assistance with the FACS, MiSeq library preparation and cell cultivation, Helle Munck Petersen for assistance with the protein purification, Anna Koza and Madis Valdemar Anderson for assistance with the MiSeq analysis and Tune Wuill for proteomics data and support with the N-glycan analysis. This work was supported by the Novo Nordisk Foundation, Denmark (NNF10CC1016517). T.A., H.F.K. and M.R.A. are receiving funding from the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie EU Grant agreement no. 642663.

Author contributions


Competing interests

A patent based on this work has been filed. The authors on the patent are A. H. H., S. K., B. V., H. F. K. The International Patent Application No. is EP12720407. The remaining authors declare no competing financial interests.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jmb.2018.11.014.

References


Supplementary Material for

Glyco-engineered CHO cell lines producing alpha-1-antitrypsin and C1 esterase inhibitor with fully humanized N-glycosylation profiles

Thomas Amann1,*, Anders Holmgaard Hansen1,*,#, Stefan Kol1, Henning Gram Hansen1, Johnny Arnsdorf1, Saranya Nallapareddy1, Bjørn Voldborg1, Gyun Min Lee1,2, Mikael Rørdam Andersen3, Helene Fastrup Kildegaard1,#

1The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

2Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

3Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark

* These authors contributed equally to this publication

# Shared correspondence

Correspondence:

Anders Holmgaard Hansen, The Novo Nordisk Foundation Center for Biosustainability, Kemitorvet, Building 220, 2800 Kgs. Lyngby, Denmark

E-mail: ahoha@biosustain.dtu.dk
<table>
<thead>
<tr>
<th>Tables and Figures</th>
<th>Page number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suppl. Fig. 1</td>
<td>2</td>
</tr>
<tr>
<td>Suppl. Fig. 2</td>
<td>3</td>
</tr>
<tr>
<td>Suppl. Fig. 3</td>
<td>4</td>
</tr>
<tr>
<td>Suppl. Fig. 4</td>
<td>5</td>
</tr>
<tr>
<td>Table S1</td>
<td>6</td>
</tr>
<tr>
<td>Table S2</td>
<td>7</td>
</tr>
<tr>
<td>Table S3</td>
<td>8</td>
</tr>
<tr>
<td>Table S4</td>
<td>9</td>
</tr>
<tr>
<td>Table S5</td>
<td>10</td>
</tr>
<tr>
<td>Table S6</td>
<td>13</td>
</tr>
</tbody>
</table>
Suppl. Fig. 1: Vector design for overexpression. (A) rhC1INH and ST6GAL1 are under the control of mCMV-hEF-1a-5’ consisting of an mCMV enhancer element, a hEF-1a promoter element, and a 5’ HTLV untranslated region. The plasmid backbone consists of an ampicillin resistance cassette and SV40 early promoter in front of the Glul (GS) sequence. Poly-A sequences (purple) terminate Glul, rhA1AT and rhC1INH cassettes. (B) rhA1AT and ST6GAL1
are under the control of mCMV-hEF-1a-5'. The plasmid backbone consists of an ampicillin resistance cassette and SV40 early promoter in front of the Glul (GS) sequence. Poly-A sequences (purple) terminate Glul, rhA1AT and rhC1INH cassettes.

Suppl. Fig. 2: Screening for FITC SNA positive and rhA1AT producing clones. (A) Supernatant of monoclonal cell lines (clone 1 - 5) on SDS PAGE to screen for rhA1AT producing clones. Control samples consist of plA1AT and supernatant from non-producing CHO-S WT cells. (B) Cells of CHO-S WT, the 50 µM MSX A1AT polyclonal cell line and the clones A1-1 and A1-2 were stained with FITC-SNA. The histograms show the number of cells on y-axis and FITC mean intensity on the x-axis. CHO-S WT cells were used for gating between FITC-positive and -negative cells (red line). (C) N-glycan analysis of purified A1AT and C1INH versions. A2G2S2 proportions of purified rhA1AT and rhC1INH from different clones compared to plA1AT and plC1INH.
Suppl. Fig. 3: Viability of cells during MSX-based selection and N-glycan analysis of A1AT and C1INH derived from plasma as well as polyclonal and monoclonal cell lines. (A) Cell viability of non-transfected and transfected cell lines during selection with different concentrations of MSX. (B) N-glycan histograms of purified A1AT and C1INH from transiently transfected CHO-
S WT and 50 µM MSX-selected polyclonal cell lines in comparison to Cinryze and Prolastin-C.

(C) N-glycan analysis of purified rhA1AT and rhC1INH from four selected monoclonal producing cell lines.

Suppl. Fig. 4: Analysis of cell culture supernatants and protein activity after protein purification. (A) SDS-PAGE gel analysis of supernatants from C1INH-producing clones and the 10x KO B parental cell line during batch cultivation. 10 µL of supernatant from day 8 and 10 of triplicate cultures as well as p/C1INH at different concentrations were analyzed by SDS-PAGE under reducing conditions. Intact C1INH is migrating slightly above 100 kDa as indicated by an asterisk. (B) A1AT activity assay of plA1AT, rhA1AT transiently expressed in CHO-S WT and rhA1AT from stable clones A1-1 and A1-2 (left panel) as well as C1INH activity assay of pl/C1INH, rhC1INH transiently expressed in CHO-S WT and rhC1INH from stable clones C1-1 and C1-2 (right panel).
Table S1: sgRNA target sequences. The bases in red mark the PAM site site

<table>
<thead>
<tr>
<th>Gene name of target</th>
<th>hypothesized KO effect</th>
<th>Target sequence (5’→ 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGAT4A</td>
<td>decreased branching</td>
<td>GTCTACATTCGTCACTGTCG</td>
</tr>
<tr>
<td>MGAT4B</td>
<td>decreased branching</td>
<td>GCTTCAGTCGGGATCCTCTGGG</td>
</tr>
<tr>
<td>MGAT5</td>
<td>decreased branching</td>
<td>GGATGGCTACCCCCACTGCCAGG</td>
</tr>
<tr>
<td>ST3GAL3</td>
<td>decreased sialylation</td>
<td>GATCTAGCCCACCTTCGAAAGG</td>
</tr>
<tr>
<td>ST3GAL4</td>
<td>decreased sialylation</td>
<td>GTGTCGCTTGTGTGGTGGTGG</td>
</tr>
<tr>
<td>ST3GAL6</td>
<td>decreased sialylation</td>
<td>GGAGTTGATCATTTGTAGCGG</td>
</tr>
<tr>
<td>B3GNT2</td>
<td>decreased elongation</td>
<td>GTTGGGCAAGACGCCCGAGG</td>
</tr>
<tr>
<td>FUT8</td>
<td>no core-fucosylation</td>
<td>GTCAGAGCCTGACAAGTGGG</td>
</tr>
<tr>
<td>SPPL3</td>
<td>hyper-glycosylation</td>
<td>AGAGAGACGGACGTCAATAGG</td>
</tr>
<tr>
<td>GLUL*</td>
<td>Gln-dependent growth</td>
<td>TCCCAATCAGCAAACAGACTGG</td>
</tr>
</tbody>
</table>

*the Glul sgRNA efficiency during KO-generation of the presented sequence was very low compared to other target sgRNAs and we recommend the usage of a different design
Table S2: Oligos for sgRNA expression vector cloning.

<table>
<thead>
<tr>
<th>Oligo Name</th>
<th>Oligo sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gRNA_MGAT4A_411545_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTAGAGCTAGAAAT</td>
</tr>
<tr>
<td>gRNA_MGAT4A_411545_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_MGAT4B_1280368_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_MGAT4B_1280368_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_MGAT5_327084_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_MGAT5_327084_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_ST3GAL3_244730_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_ST3GAL3_244730_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_ST3GAL4_964386_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_ST3GAL4_964386_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_ST3GAL6_1812502_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_ST3GAL6_1812502_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_B3GNT2_1273293_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_B3GNT2_1273293_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_FUT8_681494_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_FUT8_681494_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_SPPL3_213040_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_SPPL3_213040_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_GLUL_941540_fwd</td>
<td>GGAAAGGACGAAACACCCGCTCATCATTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
<tr>
<td>gRNA_GLUL_941540_rev</td>
<td>CTAACACACGACGTAGACAGAACACTGCTCGTTTTCCCGTACAGAAAT</td>
</tr>
</tbody>
</table>
Table S3: Primer list for deep sequencing (MiSeq). The primers contain overhang sequences compatible with Illumina Nextera XT indexing
(forward primer overhang: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG,
reverse primer overhang: GTCTCGTGGGCTCGAGATGTGTATAAGAGACAG).

<table>
<thead>
<tr>
<th>primer name</th>
<th>sequence (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiSeq_MGAT4A_411545_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAGACAGAAGGCAAATCTACG</td>
</tr>
<tr>
<td>MiSeq_MGAT4A_411545_rev</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGTTAACAGCTACACAGGAAGAGCA</td>
</tr>
<tr>
<td>MiSeq_MGAT4B_1280368_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGATGGGGTGTATGGAGGT</td>
</tr>
<tr>
<td>MiSeq_MGAT4B_1280368_rev</td>
<td>GTCTCGTGGGCTCGAGATGTGTATAAGAGACAGTTGCAGACTGCTCTCCTTTG</td>
</tr>
<tr>
<td>MiSeq_MGAT5_327084_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATGAATCTCATGGTTTCCTTTTG</td>
</tr>
<tr>
<td>MiSeq_MGAT5_327084_rev</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGGCTTCAAGACTCAACTCTTTCCC</td>
</tr>
<tr>
<td>MiSeq_ST3GAL3_244730_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGAAACAGCATGGGCAAAC</td>
</tr>
<tr>
<td>MiSeq_ST3GAL3_244730_rev</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGACTGGAATGTGGATGGTGGCC</td>
</tr>
<tr>
<td>MiSeq_ST3GAL4_964386_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCACTGTCTTACTACCCACAGAGA</td>
</tr>
<tr>
<td>MiSeq_ST3GAL4_964386_rev</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGTCCTTTCATTATATTCAAGAGCCAC</td>
</tr>
<tr>
<td>MiSeq_ST3GAL6_1812502_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATGGTTTACCTGTTTTC</td>
</tr>
<tr>
<td>MiSeq_ST3GAL6_1812502_rev</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGAAGTGGTGAGTGTGTCCTGT</td>
</tr>
<tr>
<td>MiSeq_B3GNT2_1273293_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCACCCACAGGCAAACAG</td>
</tr>
<tr>
<td>MiSeq_B3GNT2_1273293_rev</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGCAGCTGAAGCCACATGAGAAA</td>
</tr>
<tr>
<td>MiSeq_FUT8_681494_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCGTGGGAGTGTGTTTACTGAGGATA</td>
</tr>
<tr>
<td>MiSeq_FUT8_681494_rev</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGTCTGCTGAGGTGTGTTGCTGT</td>
</tr>
<tr>
<td>MiSeq_SPPL3_213040_fwd</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGCGTGGGAGTGTGTTTACTGAGGATA</td>
</tr>
<tr>
<td>MiSeq_SPPL3_213040_rev</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGAAGTGGTGAGTGTGTTGCTGT</td>
</tr>
<tr>
<td>MiSeq_GLUL_941540_fwd2</td>
<td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGCGTGGGAGTGTGTTTACTGAGGATA</td>
</tr>
<tr>
<td>MiSeq_GLUL_941540_rev2</td>
<td>GTCCTCGTGGGCTCGAGATGTGTATAAGAGACAGCAGCGTGGGAGTGTGTTTACTGAGGATA</td>
</tr>
</tbody>
</table>
### Table S4: Indels generated in ten targeted genes by CRISPR/Cas9 multiplexing.

<table>
<thead>
<tr>
<th>Multiplexing round</th>
<th>Gene</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MGAT4A</td>
<td>MGAT4B</td>
<td>MGAT5</td>
<td>ST3GAL4</td>
<td>ST3GAL6</td>
</tr>
<tr>
<td>10x KO clone A</td>
<td>+2</td>
<td>-1</td>
<td>+1</td>
<td>-5/+1</td>
<td>+1</td>
</tr>
<tr>
<td>10x KO clone B</td>
<td>+2</td>
<td>-1</td>
<td>+1</td>
<td>-5/+1</td>
<td>+1</td>
</tr>
</tbody>
</table>

### Table S5: Nucleotide sequences for overexpression vectors.

<table>
<thead>
<tr>
<th>vector</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SerpinG1 plasmid</td>
<td>agacgtcaTGCAGGAAAAGACATGTAAGCACAAGAGGCCAGAAGAGCCGAGTTTGCCTGCGGCTTTCTTCCATAGG</td>
</tr>
<tr>
<td></td>
<td>CTCGGCCCCCCTGAGCAGACACTAACAAAATTCAGGCCTAAGTGCAGAGGTGGTGGCTACACTGCACTAGAGCACTAGTGGTCTGCTCCTCTGGTGACAGCCCAGAGCTGTTCTTCGAGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGTGTGAGTCCTCTGATTCGGATGTCGAATAAAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGACAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGC</td>
</tr>
</tbody>
</table>
C1INH/ SerpinG1

BGHpA

mCMV-

hEF-1a-5'

promoter
AmpR promoter

SV40pA

GLUL

Page | 191
Table S6: Proteomics-based identification of proteins in whole cell lysates from CHO KO and CHO WT cells. Of the KO’s only Glul could be detected. KO of Glul is verified as no hits in KO cell line.

<table>
<thead>
<tr>
<th>Protein IDs</th>
<th>Fasta headers</th>
<th>Intensity 10x KO cell line_1</th>
<th>Intensity 10x KO cell line_2</th>
<th>Intensity WT_1</th>
<th>Intensity WT_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G3HLB3</td>
<td>tr</td>
<td>G3HLB3</td>
<td>G3HLB3_CRIGR Glutamine synthetase OS=Cricetulus griseus</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G3HG36</td>
<td>tr</td>
<td>G3HG36</td>
<td>G3HG36_CRIGR Glutamine synthetase OS=Cricetulus griseus</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Publication V: Tailored N-glycosylation in Drosophila S2 cells aimed at: humanization, de-fucosylation, and increased mannosylation
Title: Tailored N-glycosylation in *Drosophila* S2 cells aimed at: humanization, de-fucosylation, and increased mannosylation

Stine Broch Clemmensena¹
Anders Holmgaaard Hansenb
Magdalena Skrzypczakc
Piotr Szlarski³
Ida Busch Nielsena
Teit Max Moscote Søgaarda
Lars Poulsen³
Mikael Rørdam Andersed
Helene Fastrup Kildegard³
Morten Agertougn Nielsen³
Willem Adriaan de Jongh³

²ExpresS²ion Biotechnologies, SCION-DTU Science Park, Hørsholm, Denmark.
³Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark
³Centre for Medical Parasitology at the Department of Immunology and Microbiology, University of Copenhagen, and Department of Infectious Diseases, Copenhagen University Hospital, Denmark
²Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark

¹To whom correspondence should be addressed. Email: scl@expresS²ionbio.com
**Keywords:** *Drosophila* S2 cells, N-glycosylation, humanization, core fucose, high-mannose

Abstract

The effect of N-glycosylation on protein pharmaceuticals has been studied extensively, most notably in monoclonal antibody (mAB) therapies, but also for virus-like particle vaccines. These studies have highlighted the importance of terminal sialic acid for serum half-life, and also a striking increase in effector functions for mABs lacking core fucose. The objective of this study was to engineer a Drosophila S2 cell-toolbox enabling further study glycans’ effect on vaccine and therapeutic protein efficacy. Specifically, we aimed to 1) Achieve a more humanized glycan structure through disruption of fused lobes gene and insertions of N-acetylglucosaminyl (GlcNAc) transferases, 2) Remove core fucosylation through disruption of α1,6-fucosyltransferase (FucT6) gene, and 3) Generate a “high-mannose” glycan structure by disrupting α-mannosidase gene (α-Man-Ia). Our efforts towards humanizing the glycan structure resulted in a cell line that increased double-GlcNAc-capped structures from 0% to >64%. Disruption of the FucT6 gene resulted in a cell line that showed no fucosylation. Disruption of the α-Man-Ia gene resulted in a cell line that significantly increased the levels of “high-mannose”.

This study successfully derived three new S2 cell lines and confirmed their modified glycan patterns, and can now be used for studying the effect of specific glycan patterns.
**Introduction**

N-glycosylation is an important consideration in production of pharmaceuticals. Most drugs intended for human use show better efficacy, half-life and pharmacokinetics when carrying a humanized and complex N-glycan structure.\[^{[1]}\] On the other hand, short and simple insect glycans could be advantageous in a vaccine where an immune response is desired. Addition of immunogenic glycans to recombinant antigens has the possibility to further enhance immunogenicity. A lot of effort has been put into uncovering and understanding the different N-linked glycan structures on current drugs and vaccines and also on new glycoprotein compounds.\[^{[2–5]}\]

One example of this importance has been shown in the case of Cetuximab, a monoclonal antibody (mAb) produced in a murine myeloma cell line, which is used to treat various cancer types. Cetuximab can induce serious side effects, allergic responses and anaphylactic shock.\[^{[5]}\] The reason hereof lies in a non-human glycan structure called α1,3-galactose (α-gal), which is present on the recombinant antibody, but not on human produced proteins. Murine cell lines such as the one used to produce Cetuximab, and other non-human cells can attach α-gal to N-glycans,\[^{[5]}\] but human cells lack the responsible galactosyltransferase and do therefore not carry α-gal. Adding to the case, 1% of circulating antibodies in humans are directed towards α-gal,\[^{[6]}\] which causes the human immune system to react against Cetuximab and other recombinantly produced α-gal-carrying proteins. Other known immunogenic glycan structures are Neu5Gc,\[^{[7]}\] β1,2-xylose\[^{[8]}\] and α1,3-fucose,\[^{[9]}\] which are produced in mammalian cells, plants and some insect cells, respectively. As in the case of Cetuximab, these glycan structures would most likely not be beneficial if attached to a therapeutic mAb, however, they could possibly improve immunogenicity when carried on a vaccine antigen.

It has been shown that high-mannose and pauci-mannose N-glycans are advantageous in a vaccine against HIV compared with complex N-glycan structures.\[^{[4]}\] Kong et al found that antigens expressed in *Spodoptera frugiperda* 9 (Sf9) and Sf9 Mimic™ cells induced significantly greater
antigen-specific serum IgG responses compared to 293F-expressed HIV gp120.\textsuperscript{[4]} The titers of antigen-specific IgG produced against HIV gp120 produced in insect cells were significantly higher than those against the gp120s produced in mammalian cells. Their conclusion was that insect-cell-derived gp120 is more immunogenic than its mammalian-cell-expressed counterpart irrespective of the type of adjuvant used or its mode of action.\textsuperscript{[4]} It is also known mannose is advantageous in the case of a virus-like particle (VLP) antigen delivery system.\textsuperscript{[10]} Al-Barwani et al found that Rabbit hemorrhagic disease virus VLP conjugated to mannoside exhibited significantly enhanced binding and internalization by murine dendritic cells, macrophages and B cells as well as human dendritic cells and macrophages.\textsuperscript{[10]}

Cerezyme\textsuperscript{®} is an enzyme replacement therapy for the lysosomal storage disease named Gauchers disease. Cerezyme\textsuperscript{®} undergoes glyco-processing where the complex N-glycans are trimmed down to Pauci-mannose structures prior to being sold. This is due to enhanced uptake in antigen presenting cells by mannose receptors.\textsuperscript{[11]} It is not only the presence of the correct N-glycosylation that benefit in a vaccine or drug setting, also the lack of specific N-glycans can be important. It has been shown that monoclonal antibodies lacking core $\alpha$1,6-fucose have improved antibody-dependent cell-mediated cytotoxicity via improved FcγRIIIa binding without it altering antigen binding or complement-dependent cytotoxicity.\textsuperscript{[12–14]}

It is tempting to conclude that an immunogenic glycosylation is advantageous when producing a vaccine. However, for the influenza hemagglutinin (HA) antigen, de Vries et al showed that a humanized glycan structure allow better up-take of APCs and therefore a better immune response. While antibody titers are higher with the insect derived HA proteins, the neutralization and HA inhibition titers were higher with the mammalian cell produced HA proteins.\textsuperscript{[15]} Recombinant HA proteins containing tri- or tetra-antennary complex, terminaly sialylated and asialyated-glactose type N-Glycans induced better protective immunity in mice to lethal challenge.\textsuperscript{[16]} Cases like these
emphasize the importance of carefully choosing an expression system when it comes to drugs and vaccine targets.

Insect cells can be a good choice for expressing recombinant proteins that require correct folding and post-translational modifications. They are relatively cheap and cost efficient compared to mammalian expression systems. Insect cells naturally glycosylate in a pauci-mannosidic manner, often with a α1,6-fucose attached to the core N-acetylglucosamine (GlcNAc) and to some degree also high-mannoses. The glycosylation pathway in insect cells differ from the one of mammalian cells, as insect cells have a hexoaminidase encoded by the fused lobes gene (fdl) which encodes a β-N-acetyl-D-hexosaminidase that cleaves off a terminal GlcNAc on the A1 structure. This leads to a M3, or pauci-mannose, structure. Mammalian cells retain the GlcNAc and add further carbohydrates to achieve complex glycan structures.

Substantial optimization of the glycosylation pathway has been done in the baculovirus insect cell system line Spodoptera frugiperda 9 (Sf9), less so in Drosophila S2 (S2) cells.

Disruption of gene function used to be a tedious and expensive procedure. With the discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and the Cas9 enzyme, new possibilities of gene disruptions were made possible. Previously, fdl was inhibited by addition of GlcNAcase inhibitor (2-acetamido-1,2-dideoxynojirimycin), by RNA interference, or simply by ignoring it and inserting glycosyltransferases. Recently with the advantages and accessibility of disrupting genes using CRISPR/Cas9 more work on disrupting fdl has been done. Mabashi-Asazuma et al disrupted fdl in S2R+ cells using CRISPR/Cas9 and recently CRISPR/Cas9 was also used to disrupte fdl in Sf9 cells. Various glycosyltransferases intended to humanize the N-glycan profile have been inserted into the genome of the insect cell lines S2, Sf9, and E4a. Geisler et al exchanged the membrane part of β1,4-galactosyltransferase to
allow better retention in the Golgi membrane, which resulted in relatively high amounts of G2 in recombinant human erythropoietin (hEPO) and could serve the basis for further sialylation of N-glycans in Sf9 cells.\textsuperscript{[33]} The closest to a humanized structure reached in insect cell culture is that of Toth et al who managed to achieve 39\% terminally sialylated N-glycans on purified hEPO.\textsuperscript{[34]} In an effort to humanize the N-glycosylation of S2 cells, Kim et al RNAi suppressed \textit{fdl} and inserted the $\beta$1,4-galactosyltransferase into the genome. In this cell line they reached small amounts of G2 on purified recombinant hEPO.\textsuperscript{[21]}

So far no efforts have been put into either the BEVS or S2 system to increase immunogenicity of the glycosylation. The insect cell line High Five\textsuperscript{TM} is known to natively express some amounts of the immunogenic 1,3-fucose\textsuperscript{[35]}, but no active attempts to achieve a more immunogenic structure in insect cells have been reported.

Generally, there is still a lot of glyco-engineering to be done in insect cells, but especially in S2 cells in order to express proteins with more or less immunogenic glycosylation structures. On that account, we set out to establish three glyco-modified S2 cell lines. The goal of the first cell line was to achieve a more humanized glycan structure. This was done by knocking out \textit{fdl} and afterwards inserting the genes for N-acetylglucosaminyltransferase I and II (\textit{Mgat1} and \textit{Mgat2}) to achieve A2. This work was done in a monoclonal cell line expressing the recombinant part of VAR2CSA involved in placental malaria and in addition binds onco fetal chondroitin sulfate A present on cancer cells, ID1-ID2a.\textsuperscript{[36]} Secondly, we disrupted the $\alpha$1,6-fucosyltransferase gene, \textit{FucT6} (a homolog to the human \textit{fut8} gene) to achieve a S2 cell line that does not attach a core $\alpha$1,6-fucose on the glycan structure to allow for production of antigens and Fc-fusion proteins. This was done in a wild type cell line. Lastly, we established a cell line that glycosylates in a high-mannose manner, by disrupting the $\alpha$-\textit{Man-la} gene, also in a wild type cell line. This enzyme is responsible for trimming
high-mannose structures to M5. Our goal was to enhance vaccine antigen immunogenicity through addition of immunogenic highly mannosylated glycans.

The construction of these three new cell lines expands the S2 N-glycosylation toolbox and allows expression of many different glycoprotein targets, whether they are intended as therapeutics, antibodies or vaccine targets.

Examples of nomenclature as by GlycoStore.org. All glycans analyzed in this work are exemplified in Supporting Information “Glycan percentages”.

![Diagram of glycan structures](image-url)
Materials and Methods

Plasmid Construction

The online E-CRISP tool (www.e-crisp.org; German Cancer Research Center) was used to identify CRISPR/Cas9 sequences within the Drosophila melanogaster genome that target fdl, FucT6, and α-Man-Ia. sgRNA target sequences were selected as 20 nt sequences preceding an NGG PAM sequence in the genome. The oligonucleotide pairs XX-F and XX-R were used to construct DNA fragments consisting of each targeting sequence with overhangs to enable their subcloning into pExpreS2-CRISPR (ExpreS2ion Biotechnologies, Hørsholm, Denmark). The sequence of each synthetic oligo is given in Supporting Information “Oligo sequences for disruptions of fdl, α-Man-Ia and FucT6.”.

The N-acetylglucosaminyl transferase I and II were constructed by ordering Mgat1 (NP_525117, NCBI) and Mgat2 (GenBank: CAC83074.1) online (GeneArt, ThermoFisher) and inserting them into pExpreS2-PAC (ExpreS2ion Biotechnologies, Hørsholm, Denmark) using EcoRI and NotI.

S2 cell culture, transfection, and cloning

ExpreS2 Cells (ExpreS2ion Biotechnologies, Hørsholm, Denmark), hereafter “S2 cells” were routinely maintained at 25°C and 130rpm in suspension in shake flasks in culture medium EX-CELL 420 Serum-Free Medium for Insect Cells (Sigma-Aldrich, cat. Nr. 14420C, Steinheim, Germany) supplemented with 100 units/mL penicillin and 0.1 mg/mL streptomycin (Pen-Strep Solution, Biological Industries, cat. 03-031B, Cromwell, CT, USA), hereafter “culture medium”. The S2 cells were counted with a CASY® Cell Counter every 3-4 days and passaged by centrifugation or dilution to 8E6c/ml.
For transfection, the S2 cells were passaged to 8E6c/ml in shake flasks in culture medium and transfected the following day, by splitting the cells to 2E6c/ml and mixing with first 50µl ExpreS2 Insect- TRx5 (ExpreS2ion Biotechnologies, Horsholm, Denmark) transfection reagent and second with 12,5µg of plasmid DNA. The transfected cells were then transferred to a 25cm² Tissue Culture flask (In Vitro, Fredensborg, Denmark). A polyclonal cell line was selected herein for 21 days in culture medium supplemented with 10% fetal bovine serum (FBS) (Fischer Scientific, Roskilde, Denmark) and 1,5mg/ml zeocin (Thermo Fisher, Hvidovre, Denmark) or 4,0mg/ml geneticin (InvivoGen, Toulouse, France), by dilution to 1E6c/ml every 3-4 days or whenever the cells reached a density higher than 1,5E6c/ml.

For selection with puromycin (Mgat1 and Mgat2) the cells were transfected and grown in a 50ml vented Falcon Tube in 8ml of culture medium. The cells were transfected by adding and swirling 200µl ExpreS2 Insect- TRx5 Transfection Reagent and 20ng of plasmid DNA. After 2-4 hours FBS was added to a concentration of 10%. Puromycin was added after 24 hours to a final concentration of 100µg/ml. The cells were counted and split by centrifugation every 3-4 days in culture medium + 10% FBS + 100µg/ml puromycin for 2 weeks. Hereafter the stable cell lines were transferred to a 125ml Shake Flask and passaged as described previously.[38]

The monoclonal cell line expressing ID1-ID2a was established previously.[39]

Monoclonal cell lines were obtained by limited dilution in 96 well plates (In Vitro, cat. GR-655180, Fredensborg, Denmark) using non-transfected S2 feeder cells at 0,6E6c/ml. Stably transfected polyclonal cells were seeded out at concentrations of 100 cells/mL, 30 cells/mL or 10 cells/mL. The total volume in wells was 150µl culture medium supplemented with 10% fetal bovine serum and 1,5mg/ml zeocin or 4,0mg/ml geneticin. Over 2-3 weeks the 96-well plates were inspected regularly and monoclonal cell lines were identified and the monoclonal cell lines were expanded from 96well plates to 250mL shake flasks (Sigma-Aldrich, cat. CLS431255, St. Louis, MO, USA).
For production of ID1-ID2a (cell line described previously\cite{ref}) the cells were expanded to a total volume of 2L and grown in a 5L Thomson Optimum Growth Flask (Thomson, Sittingbourne, England) and 200µL/L PD30 was added. The cell culture supernatant was harvested by centrifugation at 4400xg, filtered through a 0.22µm filter, concentrated 1:4 times, and exchanged into 20mM Phosphate, pH 6.6, using a tangential flow filtration device (Pall, NY, USA). Protein purification of ID1-ID2a proceeded through ion exchange chromatography (HiTrap SP Sepharose FF, GE Healthcare Life Sciences) and was eluted step-wise (6%, 15%, and 22%) in 20mM Phosphate and 1M NaCl pH 6.6. The purified protein was diluted 1:4 in 20mM Phosphate pH 7.0 and run on a HiTrap Capto adhere column (GE Healthcare Life Sciences, Brøndbyvester, Denmark) for removal of contaminants, and was eluted step-wise (13%, 30%, 80%) in 20mM Phosphate, 1M NaCl pH 7.0 and was aliquoted and then snap-frozen in liquid nitrogen and stored at -80°C.

**Indel Detection by Amplicon Analysis (IDAA)**

IDAA was carried out as previously described\cite{ref}. Briefly, genomic DNA was extracted using PureLinkTM Genomic DNA Mini Kit (Fisher Scientific, cat. nr. K182001, Roskilde, Denmark). Primers were designed using the online prediction tool “Primer3”: 

- **FucT6** (F:TTCGCAAGGAACGGGGCTCCGAAC, R:GCAAGGAACGGGGCTCCGAACGTT)
- **α-Man-Ia** (F:CAACGTTTGGAGCAAAAGATTC, R:AACCACCTACCTCTTTGACCTTC)

PCR was performed Phusion High-Fidelity PCR kit (1x HF buffer, 0.2 mM dNTP, 0.25U Phusion polymerase, 0.025µM forward primer, 0.25µM reverse primer and 0.25µM 6-FAM 5’-labelled universal primer (Thermofisher, cat. nr. F553S, Hvidovre, Denmark). The PCR-amplicons were analyzed by Fragment Length Analysis (FLA) (Eurofins, Glostrup, Denmark).
**SDS-PAGE and Western Blotting**

Supernatant samples were analyzed by SDS-PAGE and Western Blot analysis. Briefly, proteins were resolved by 10% SDS-PAGE and then transferred to a nitrocellulose membrane. Non-specific binding was blocked by incubating the membrane in Carbo-Free™ Blocking Solution (VectorLabs, Cat. No. SP-5040, Burlingame, CA, USA) for 30 min at room temperature. Then the blot was incubated for 30 minutes in PBS with 10µg/ml biotinylated lectin and washed in PBS + 0.2% Tween 20™. The secondary antibody was HRP-conjugated Streptavidin (Fisher Scientific, Roskilde, Denmark) diluted 1:5000 for 45 min followed by a wash in PBS + 0.2% Tween 20™. Novex® ECL (WP20005, Fisher Scientific, Roskilde, Denmark) was used for detection. Lectins used: for recognition of α1,6-fucose: Biotinylated Lens culinaris (LCA); high-mannose: Biotinulated Musa Paradisiaca (BanLec) (both from VectorLabs, Burlingame, CA, USA).

**Mass Spectrometry**

Glycoprofiling was performed as previously described[41]. Briefly, supernatants were filtered and proteins contained in the sample were concentrated by centrifugation using Amicon Ultra columns (Merck Millipore, Merck KGaA, Darmstadt, Germany) with 3000 Da cutoff. N-glycans from retained proteins were released and fluorescently labeled with GlycoPrep Rapid 2-AB N-Glycan kit (ProZyme Inc., Hayward, CA) or GlycoWorks RapiFluor-MS N-Glycan Kit (Waters, Elstree, UK). Labeled N-glycans were analyzed by LC-MS on a Thermo Ultimate 3000 HPLC with fluorescence detector coupled on-line to a Thermo Velos Pro Ion Trap MS. Glycan abundance was measured by integrating the areas under normalized fluorescence spectrum peaks with Xcalibur software (Thermo Fisher Scientific, Hvidovre, Denmark) giving the relative amount of the glycans. All
annotated sugar structures are peaks with correct mass and at least a signal to noise value of 10:1 as calculated with Xcalibur.

**Results**

**Humanization: Disruption of *fdl* and insertion of *Mgat1* and *Mgat2**

The work on a humanized glycan structure was carried out in a monoclonal S2 cell line that recombinantly expresses the placental malaria protein VAR2CSA, truncation variant ID1-ID2a. Since this is an antigen for a vaccine, it would make sense to add immunogenic glycans to it to improve the vaccine properties of the protein. However, ID1-ID2a has also been shown to bind to many types of cancer cells and when a toxin is attached to ID1-ID2a, it can reduce or clear the cancer cells efficiently.[36] One of the ways to optimize this protein for future cancer treatment could be to add human-like glycans to it, so it will become less immunogenic. Therefore, we worked towards humanizing the glycan structure on ID1-ID2a. The glycosylation of this protein when expressed in wild-type S2 cells was analyzed on Liquid Chromatography – Mass Spectrometry (LC-MS) prior to any modifications and showed two main peaks of 36% M3 and 58% FM3. See Figure 1.

These are the two most prevalent glycan structures found in insect cells[42]. Next we wanted to confirm that the glycosylation profile did not change significantly under different growth conditions and purification protocols. We therefore analyzed purified ID1-ID2a from 3 day vs. 4 days batches, from shake flasks and bioreactors, and during different steps of the purification. The results were consistent: M3 with or without core fucose in a 40:60 ratio, very similar to the profile seen in Figure 1.
The first modification towards a more humanized glycan structure involved a knock out of \textit{fdl}.\textsuperscript{[19]} For this, two different target sgRNAs were chosen (Se Supporting Information “Oligo sequences for disruptions of \textit{fdl}, \textit{\(\alpha\)-Man-Ia} and \textit{FucT6}”) and designated \(\Delta/\textit{fdl}3\) and \(\Delta/\textit{fdl}12\). Purified ID1-ID2a from these two polyclonal pools was analyzed on LC-MS, but no difference in the glycan profile compared to the wild type protein were detected (data not shown). \(\Delta/\textit{fdl}3\) was cloned and \(\Delta/\textit{fdl}3\) Clone 1 showed an increase of FA1/A1 from 0% to 58%, which was expected. The second step towards humanizing this cell line was to transfecit with \textit{Mgat1} and \textit{Mgat2}, which were responsible for attaching GlcNAcs to the mannose tree. Purified ID1-ID2a here from was analyzed on LC-MS. See Figure 2.

The LC-MS differentiates glycan structures by size and hydrophilicity on a hilic type column. The difference between fucosylated and non-fucosylated glycans are easily distinguished by this method. However, since we were comparing the effect of \textit{fdl} disruption and insertion of \textit{Mgat1} and \textit{Mgat2} the percentages with and without fucose were pooled. The label “other” refers to other glycans structures that were present in the glycan pool, but that we deemed irrelevant for the analysis of \textit{fdl} disruption and insertion of \textit{Mgat1} and \textit{Mgat2} - for example higher mannose structure and a few more complex (See Supporting Information for all glycans “LC-MS chromatograms”). In the WT ID1-ID2a cell line there was >97% FM3/M3 glycans. After disruption of \textit{fdl} and re-cloning there was a significant decrease from >97% FM3/M3 to <39% and an increase of FA1/A1 to 58%. This was as expected. According to a suggested glycosylation pathway in insect cells by Shi et al, it should only be necessary to insert the N-acetylglucosaminyl transferases II (\textit{Mgat2}), since the insect cells already harbor the \textit{Mgat1}.\textsuperscript{[20]} However, since we still saw >38% FM3/M3 after the disruption of \textit{fdl}, we incorporated both \textit{Mgat1} and \textit{Mgat2} into the genome of this monoclonal cell line and re-cloned. Hereafter, we found <10% FM3/M3, <22% FA1/A1, and >64% FA2/A2 on purified ID1-ID2a as depicted in Figure 2. The “other” section of the bar graphs in Figure 2
represents glycans that occur in <2% of the total glycans. They can all be seen in Supporting Information “LC-MS chromatograms”.

This Δfdl + Mgat1 and Mgat2 ID1-ID2a cell line is a solid foundation for building the next steps in constructing humanized glycosylation on ID1-ID2a.

Glycan analysis by LC-MS of the impact of the FucT6 disruption

We were also interested in constructing a cell line that does not attach core α1,6-fucose to the glycan tree. We did this by disrupting FucT6, which encodes a α1,6-fucosyltransferase. The work on disrupting the FucT6 gene was conducted in a wild type cell line (WT-S2). We designed three CRISPR/Cas9 sgRNA target sequences for the disruption of the FucT6 gene (See Supporting Information “Oligo sequences for disruptions of fdl, α-Man-Ia and FucT6:”). These were transfected and selected for as described in Materials and Methods. The secretomes of S2-WT and the ΔFucT6 with a polyclonal disruption were analyzed on LC-MS. In Figure 3 the data is represented as bar graphs. Full annotations can be found in Supporting Information “LC-MS chromatograms”.

Opposed to the fdl disruption in the WT-ID1-ID2a cell line described above, it was possible to see an effect of the disruption on a polyclonal level. The S2-WT cell line showed approximately >64% of FM3 and <12% of M3. After disruption of the FucT6 gene the polyclonal glycan profile shifted and showed an increase in M3 to >39% and a decrease in FM3 to <37% of the total glycans. This indicated that the disruption was successful. However, to establish a cell line with no display of fucose further sub-cloning was done. A total of 39 ΔFucT6 clones were obtained by limited dilution. These were screened by a lectin blot (See Supporting Information “Lectin blot for screening of ΔFucT6 clones.”). Based on the lectin blot two clones were chosen for LC-MS analysis, and both showed a similar secretome glycan patterns with 0% fucose on the glycans upon
LC-MS. One example is shown in Figure 3. This clone shows a complete phenotypical knockout of the fucosyltransferase encoded by \textit{FucT6}. This also indicated that there is no $\alpha$1,3-fucose in S2 cells, which has been reported in other insect cells.\cite{35} Normally the glycans that have $\alpha$1,3-fucose also have the $\alpha$1,6-fucose.\cite{43} However, this glycan combination would be detectable by mass spectrometry. To further investigate if we had any $\alpha$1,3-fucose in the glycan pool even though we saw no di-fucosylated structures on LC-MS, we performed a Western Blot with a 1,3-fucose specific antibody and found no such structures (data not shown). This monoclonal cell line with a disrupted \textit{FucT6} gene can now serve the basis of future antibody and Fc-fusion protein expression in S2 cells.

\textbf{LC-MS analysis of the disruption of $\alpha$-\textit{man-Ia}}

S2 cells can be used for expressing recombinant vaccine antigens. As vaccines are supposed to stimulate the immune system and thereby raise antibodies against the antigen, we set out to construct a cell line that glycosylates in a more immunogenic manner. It has previously been shown that mannoses give an advantage in several immune setting\cite{10,44,45} and as S2 cells naturally glycosylate with M3 structures, we wanted to make a cell line that produced proteins with even more mannose, namely the high-mannose structures. The pathway of glycan trimming in insect cells involves several $\alpha$-mannosidases that trim existing mannose structures. We wanted to disrupt one of these and analyze the outcome on the glycan-level. The disruption of $\alpha$-\textit{Man-Ia} was carried out in a wild type S2 cell line and three target sgRNAs for CRISPR/Cas9 were chosen (See Supporting Information “Oligo sequences for disruptions of \textit{fdl}, $\alpha$-\textit{Man-Ia} and \textit{FucT6}: ”) and transfected and selected for as described in Materials and Methods. Two out of the three targets did not show any effect (data not shown), however, one target showed a shift towards higher mannose glycan structures by LC-MS analysis on a polyclonal level, see Figure 4. This polyclonal cell line
was cloned by limited dilution and the monoclonal cell lines obtained hereby were initially analyzed for genotypic indels by IDAA and by lectin blot (See Supporting Information “Lectin blot for screening of Δα-Man-Ia clones using BanLec”).

Based on the IDAA and the lectin analysis, we analyzed 3 secretome samples on LC-MS, see Figure 4. The LC-MS analysis confirmed the IDAA and lectin blot and we now have three monoclonal cell lines with an α-Man-Ia disruption. Furthermore, the LC-MS analysis provided a much more detailed picture of the distribution of what we refer to as “high-mannose”. The WT carried >77% FM3/M3 and <23% high-mannose. For Δα-Man-Ia Clone A, B, and C there is a significant shift to structures of “high-mannose” character, where these represent around 87%, 75%, and 85% for Clone A, B, and C respectively. However, there is some FM3/M3 left in all three clones. It is interesting to note that the ratio between the high-mannose structures vary within the clones. Clone A and C have around 30% M8, whereas Clone B only shows around 17% of M8. See chromatograms with all annotated glycans in Supporting Information “LC-MS chromatograms”. These three cell lines expand our toolbox as we now can produce recombinant proteins with high-mannose glycan structures on.
Discussion
The objective of this study was to modify the N-glycosylation in S2 cells to develop a toolbox for the study of the effect of glycosylation on therapeutic protein and vaccine immunogenicity, serum half-life, and pharmacokinetics. Three new cell lines were constructed 1) Initial steps towards humanization of ID1-ID2a by disruption of \textit{fdl} and insertion of \textit{Mgat1} and \textit{Mgat2}, 2) Disruption of the \textit{FucT6} gene to produce antigens and Fc-fusion proteins, and 3) by disruption of the \textit{α-Man-Ia} gene to construct a cell line that glycosylates with a more immunogenic pattern for production of vaccine antigens.

The first glycan-modification we did was disruption of \textit{fdl} and insertion of \textit{Mgat1} and \textit{Mgat2} to achieve a more human-like glycosylation on ID1-ID2a. CRISPR/Cas9 disruptions have not been published in S2 cells, but in S2R+ cells and Sf9 cells it has been shown to work on a polyclonal level.\textsuperscript{[30,46]} In S2 cells \textit{fdl} has been silenced with RNA to some extent.\textsuperscript{[23]} On a monoclonal level we saw that the CRISPR/Cas9 editing of the \textit{fdl} locus reduced or limited the FDL function. Although the disruption strategy worked, around 38\% of the native FM3/M3 glycan structures remain on ID1-ID2a expressed in \textit{Δfdl} 3-1 clone (Figure 3). One could argue that it is only a limited edit of the \textit{fdl} and that the gene got disrupted in a way where parts of the enzyme are still active. However, similar levels of residual pauci-mannose have been seen before; also in a \textit{Drosophila} fly where there was a complete disruption of the \textit{fdl} gene.\textsuperscript{[19]} Mabashi-Asazuma et al saw a similar pattern in S2R+ cells, a derivative of S2 cells, with \textit{fdl} disruption and speculated that the reason hereof lies in a Golgi \textit{α}-mannosidaseIII that can remove two terminal mannose structures from the \textit{α}-branch of the upstream M5 precursor, independently of \textit{Mgat1} and FDL, to produce pauci-mannosidic products.\textsuperscript{[30]} This has been shown to be the case in another insect cell line; \textit{Sf9} cells.\textsuperscript{[47]} Furthermore, Mabashi-Asazuma et al identified an ortholog of this gene in the \textit{Drosophila} genome. We speculate this is also the case in our \textit{fdl} disruption and this could be tested by disrupting the suspected Golgi \textit{α}-mannosidaseIII.
Furthermore, after addition of \textit{Mgat1} and \textit{Mgat2} and after cloning FA2/A2 was present >64% on purified ID1-ID2a. This is the highest percentage FA2/A2 reported on any insect cell produced protein.\textsuperscript{[21,23,30–32,48]}

Kim et al also found that by addition of \(\beta1,4\)-galactosyltransferase without the addition of \textit{Mgat1} and \textit{Mgat2} there was an increase in the level of G1 on hEPO.\textsuperscript{[23]} We only found these glycan structures present on ID1-ID2a after insertion of \textit{Mgat1} and \textit{Mgat2} and cloning (see Supporting Information “Purified \(\Delta fdl + Mgat1\) and \textit{Mgat2} produced ID1-ID2a.”). This could both be due to differences in glycosylation sites in the recombinant proteins and differences in the S2 cell lines as these adapt, grow and glycosylate differently over time in different laboratories. The glycan profile for the \(\Delta fdl + Mgat1\) and \textit{Mgat2} ID1-ID2a cell line also included small amounts of a glycan with m/z: 1203.5. (See Supporting Information “Purified \(\Delta fdl + Mgat1\) and \textit{Mgat2} produced ID1-ID2a.”). This mass corresponds to a FG2 with a glycolyl deoxynonulosonate on the one branch. This has not previously been reported in S2 cells or other insect cells.

The next steps for this cell line towards a more human-like glycan structure will be to express the gene for a \(\beta1,4\)-galactosyltransferase. Furthermore, after obtaining a G2 glycan structure, then additionally express the genes for \(\alpha2,3\)-sialyltransferase and \(\alpha2,6\)-sialyltransferase, and sialic acid transporter proteins, to achieve fully humanized glycan structures. It could also be of value to express \textit{mgat4} and \textit{magt5} to make tri- and tetra-antennary structures. However, already now it would be interesting to see how ID1-ID2a with >64% FA2/A2 behaves in a mouse cancer model\textsuperscript{[36]} compared to native ID1-ID2a.

The second modification we did was the disruption of the \textit{FucT6} gene. The \textit{FucT6} gene was successfully disrupted in a WT S2 cell line by the use of CRISPR/Cas9. The polyclonal pool showed an effect and the analyzed clone showed 100% absence of fucose. Some insect cell lines,
such as the High Five cell line, are known to express the immunogenic α1,3-fucose.\textsuperscript{[49]} Even though there has been found small amounts of di-fucosylated glycans and even less with only the α1,3-fucose the \textit{Drosophila} embryo\textsuperscript{[50,51]} we found no 1,3-fucose or double-fucosylated glycans in the WT-S2 cells or the ΔFucT6 clone.

A disruption of \textit{FucT6} has never been published in S2 cells or in other insect cell lines before and this unique cell line creates a strong foundation for the expression of antibodies in S2 cells and for expressing Fc-fusion proteins due to the enhanced effector functions that the fucose-lacking glycan structures facilitate.

The third modification we did was disruption of α-Man-Ia to construct a cell line that is capable of adding high-mannose glycan structures to the recombinant proteins. The disruption of α-Man-Ia has not previously been described in S2 or other insect cells. In this study the disruption of α-Man-Ia in S2 cells resulted in clones with >75% high-mannose glycans. However, in all three obtained and LC-MS analyzed clones there were still significant levels of FM3 and M3. From a regulatory perspective it is advantageous with a uniform glycan pattern and the next step towards this would be to disrupt other mannosidases. Jarvis et al suggested a glycosylation pathway for insect cells in 2013 where the α-man-Ia in the endoplasmic reticulum and Golgi trims M9 structures to M5\textsuperscript{[20]}. Our data support the fact that α-man-Ia plays an important role in trimming down the high-mannose structures to M3, although it is not the only enzyme responsible for the trimming. Mammalian cells have several mannosidases\textsuperscript{[52,53]} and a quick search in the \textit{Drosophila} genome reveals at least four other interesting mannosidase targets.\textsuperscript{[54]} Even with significant future opportunities for further glyco-engineering, the “high-mannose” cell lines established in this work can already now be applied to the expression of recombinant antigens where a high-mannose glycan pattern is desired, such as HIV or other vaccine antigens.
In summary, we successfully established a toolbox for expression of various glycoproteins in S2 cells. We modified the N-glycosylation of S2 cells in three different directions; humanization, defucosylation and added mannosylation. First, we constructed a monoclonal S2 cell line with insertions of ID1-ID2a, Mgat1, and Mgat2, and a fdl disruption. On purified ID1-ID2a here from, the percentage of FA2/A2 is >64%. This is not a fully human glycan structures, but serves a solid foundation for further humanization work. Secondly, we constructed a monoclonal S2 cell line that has no core-fucosylation capabilities by disruption of FucT6. This cell line can now serve as a tool for expression of antibodies and Fc-fusion proteins in S2 cells. Lastly, we constructed a cell lines that glycosylate with a “high-mannose” pattern, which can be used for expression of vaccine antigens.

Acknowledgements

The authors are grateful for assistance of Zhang Yang, (Center for Glycomics, University of Copenhagen).

References


[33] Geisler C, Mabashi-asazuma H, Kuo C, Khoo K, Jarvis DL. Engineering beta1, 4-galactosyltransferase I to reduce secretion and enhance N-glycan elongation in insect cells.


[56] Abrahams J. Building a PGC_LC_MS N-glycan retention library and elution mapping
Figure legends

Figure 1. Chromatogram of LC-MS analysis of purified ID1-ID2a. Full annotations can be found in Supporting Information “LC-MS chromatograms”. Square: GlcNAc, circle: mannose, triangle: fucose.

Figure 2. Glycan profile determined by LC-MS of purified ID1-ID2a in three different cell lines; WT, fdl3 Clone 1, and fdl3 Clone 1 Mgat1 and Mgat2.

Figure 3. Bar-graph representation of LC-MS analysis of glycosylation of the secretome of WT-S2, polyclonal ΔFucT6 (P:ΔFucT6), and monoclonal ΔFucT6 (M:ΔFucT6).

Figure 4. LC-MS analysis of glycans on supernatant glycoproteins from S2-WT, Clone A, B, and C presented as percentages.

Supporting Information

Oligo sequences for disruptions of *fdl*, *α-Man-Ia* and *FucT6*:

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence of oligo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fdl3-F</td>
<td>TTCGCCGCGCAGCGATACAGCCA</td>
</tr>
<tr>
<td>Fdl3-R</td>
<td>AACTGGCTGTATCGCTGCGCCGC</td>
</tr>
<tr>
<td>Fdl12-F</td>
<td>TTCGCTGGGCCTTGTTGACTCTCT</td>
</tr>
<tr>
<td>Fdl12-R</td>
<td>AACAGGAGTCACCAAGGCCCAGC</td>
</tr>
<tr>
<td>FucT6_36-F</td>
<td>TTCGTATCGCCGATCGAGTTGGCC</td>
</tr>
<tr>
<td>FucT6_36-R</td>
<td>AACGGCCAACCTCGATCGGCGATAC</td>
</tr>
<tr>
<td>FucT6_71-F</td>
<td>TTCGAGTTAATTGAGACTATGCAC</td>
</tr>
<tr>
<td>FucT6_71-R</td>
<td>AACGTGCATAGTCTCAATTAACCT</td>
</tr>
<tr>
<td>FucT6_56-F</td>
<td>TTCGCAAGGAACGGGGCTCGGAAC</td>
</tr>
<tr>
<td><strong>FucT6_56-R</strong></td>
<td>AACGTTCGGAGCCCGCTCCCTTGC</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>alpha-Man-I_138- F</strong></td>
<td>TTCGTGATCAGGCAGCCGCAGCACAC</td>
</tr>
<tr>
<td><strong>alpha-Man-I_138- R</strong></td>
<td>AACGTGTGTCCGCGCCCTGACATCAC</td>
</tr>
<tr>
<td><strong>alpha-Man-I_166 -F</strong></td>
<td>TTCGCCTCTGCGGCAGTCGCAGCCGC</td>
</tr>
<tr>
<td><strong>alpha-Man-I_166 -R</strong></td>
<td>AACGCGGCTGATTCGCGCCAGAGCGC</td>
</tr>
<tr>
<td><strong>alpha-Man-I_520- F</strong></td>
<td>TTCGAATATCGCGAGGGTCGCGAT</td>
</tr>
<tr>
<td><strong>alpha-Man-I_520- R</strong></td>
<td>AACATCGCGACCCTCAGCAGTCGATATTTC</td>
</tr>
</tbody>
</table>

**Lectin blot for screening of ΔFucT6 clones.**
As LC-MS is costly and time-consuming we wanted to screen these clones for possible *FucT6* disruption. The DNA was extracted and the clones were analyzed on a genetic level by Indel Detection by Amplicon Analysis (IDAA) (Results not shown). From this we selected 14 clones for further screening. The next screening was done on a phenotypic level by Western Blot of supernatant using a α1,6-fucose specific lectin, *Lens Culinaris* (LCA), see Figure 5. The LCA binds and gives signal on Clone 18, 19, 22, 24, 25, 30-32 suggesting that Clone 20, 21, 28, 29, 34, and 35 all have a disruption in the *FucT6* gene. On the lectin blot there was a clear difference between clones where the lectin bound and clones where it bound with less specificity.

![Figure 5. Western Blot with LCA to detect clones with a FucT6 disruption. E. coli lysate was used as a negative control](image)
Lectin blot for screening of Δα-Man-Ia clones using BanLec

Figure 6. Lectin Blot of supernatant samples of monoclonal cell lines with disruption of α-Man-Ia

We used a mannose specific lectin, BanLec, which distinguishes between M3 and high-mannose. In this case the lectin binds to high-mannose structures on proteins below the size of 62kDa and we therefore got a clear picture of which clones might carry the disruption. See Figure 6.

The IDAA analysis revealed that a disruption of α-Man-Ia had occurred in 7 out of 14 clones (data not shown). Based on this, three clones with the expected disruption and one without was further analyzed on LC-MS.

LC-MS chromatograms
Below is original LC-MS chromatograms of all analysis done in this work.
Purified WT-S2 produced ID1-ID2a

Purified Δfdl3-1 Clone 1 produced ID2-ID2a

Purified Δfdl + Mgat1 and Mgat2 produced ID1-ID2a.
Secretome of WT-S2
Secretome of polyclonal ΔFucT6

Secretome of Monoclonal ΔFucT6
<table>
<thead>
<tr>
<th>Glycan Structure</th>
<th>Name (as by GlycoStore.org)</th>
<th>WT-ID1-ID2a [%]</th>
<th>Δfdl ID1-ID2a [%]</th>
<th>Δfdl + Mgat1 and Mgat2 ID1-ID2a [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>M3</td>
<td>37,6</td>
<td>15,3</td>
<td>5,6</td>
<td></td>
</tr>
<tr>
<td>FM3</td>
<td>59,5</td>
<td>22,9</td>
<td>4,2</td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>-</td>
<td>20,6</td>
<td>11,6</td>
<td></td>
</tr>
<tr>
<td>FA1</td>
<td>-</td>
<td>37,5</td>
<td>10,8</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>-</td>
<td>0,0</td>
<td>20,3</td>
<td></td>
</tr>
<tr>
<td>FA2</td>
<td>-</td>
<td>0,8</td>
<td>44,2</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>1,1</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>1,8</td>
<td>0,9</td>
<td>0,5</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>-</td>
<td>-</td>
<td>0,5</td>
<td></td>
</tr>
<tr>
<td>M6</td>
<td>-</td>
<td>1,2</td>
<td>0,7</td>
<td></td>
</tr>
<tr>
<td>M7</td>
<td>-</td>
<td>0,9</td>
<td>0,4</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>-</td>
<td>-</td>
<td>0,5</td>
<td></td>
</tr>
<tr>
<td>FG1</td>
<td>0,0</td>
<td>-</td>
<td>0,5</td>
<td></td>
</tr>
<tr>
<td>A2G(4)2Ga1S1 or Deoxynonulosonate-FG2 or M5A2G1S1</td>
<td>-</td>
<td>-</td>
<td>0,2</td>
<td></td>
</tr>
</tbody>
</table>
At 21 minutes on chromatogram “Purified Δfdl + Mgat1 and Mgat2 produced ID1-ID2a.” a tiny peak that corresponds to 0.2% of the total glycan pool elutes. For the glycans herein the m/z: 1203.5 and this mass could theoretically be different structures. We suggest three possible structures that have the correct mass, however, further analysis is needed to determine if it is one of the suggested. This observation is interesting as it shows that S2 cells are capable of adding galactose to their glycans without transfection of galactosyltransferase genes.

Percentages of glycans in secretomes of S2-WT, P:ΔFucT6, M:ΔFucT6, MΔα-Man-Ia Clone A, B and C.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Glycan names (as by GlycoStore.org)</th>
<th>S2-WT [%]</th>
<th>P:ΔFucT6 [%]</th>
<th>M:ΔFucT6 [%]</th>
<th>M:Δα-Man-Ia Clone A [%]</th>
<th>M:Δα-Man-Ia Clone B [%]</th>
<th>M:Δα-Man-Ia Clone C [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>M3</td>
<td></td>
<td>12,0</td>
<td>39,1</td>
<td>41,9</td>
<td>3,4</td>
<td>4,7</td>
<td>3,3</td>
</tr>
<tr>
<td>FM3</td>
<td></td>
<td>65,1</td>
<td>37,8</td>
<td>-</td>
<td>8,8</td>
<td>18,3</td>
<td>11,4</td>
</tr>
<tr>
<td>A1</td>
<td></td>
<td>-</td>
<td>-</td>
<td>6,1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FA1</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0,9</td>
<td>1,6</td>
<td>-</td>
</tr>
<tr>
<td>M4</td>
<td></td>
<td>-</td>
<td>-</td>
<td>6,4</td>
<td>4,9</td>
<td>4,0</td>
<td>1,8</td>
</tr>
<tr>
<td>M5</td>
<td></td>
<td>4,0</td>
<td>6,1</td>
<td>3,7</td>
<td>6,6</td>
<td>19,6</td>
<td>9,0</td>
</tr>
<tr>
<td></td>
<td>M6</td>
<td>M7</td>
<td>M8</td>
<td>M9</td>
<td>M9Glc1</td>
<td>M9Glc2</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5,0</td>
<td>3,7</td>
<td>4,6</td>
<td>4,2</td>
<td>0,7</td>
<td>0,6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,1</td>
<td>3,9</td>
<td>3,2</td>
<td>4,3</td>
<td>0,8</td>
<td>0,7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8,8</td>
<td>2,1</td>
<td>6,7</td>
<td>6,6</td>
<td>3,2</td>
<td>1,3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12,0</td>
<td>19,7</td>
<td>30,7</td>
<td>9,1</td>
<td>2,8</td>
<td>1,2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16,1</td>
<td>17,4</td>
<td>16,9</td>
<td>1,4</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15,3</td>
<td>24,7</td>
<td>28,8</td>
<td>5,6</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

*This structure was present, but not possible to quantify, as there was a large polymer peak at this peak also, marked by “?” on the chromatograms “Secretome of WT-S2” and “Secretome of polyclonal ΔFucT6”. These peaks were not included in the percentage calculations.*
Figures

Figure 1

![Graph showing data over time with relative absorbance on the y-axis and time in minutes on the x-axis.]

Figure 2

![Bar chart showing comparisons between WT ID1-ID2a, Δfdl ID1-ID2a, and Δfdl+Mgat1 and Mgat2 ID1-ID2a with categories Other, FA2/A2, FA1/A1, and FM3/M3.]

Page 236