Development of robust workflows for quantitative therapeutic protein characterization using multi-attribute methodology (MAM)

by

Magdalena Widgren Sandberg, M. Sc.

Dissertation

For the acquisition of the academic degree

Doctor rerum naturalium

Dr. rer. nat.

at the

University of Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Department of Chemistry

June 24, 2021

Referees:

1. Prof. Dr. Hartmut Schlüter
2. Prof. Dr. rer. nat. habil. Markus Perbandt

Date of oral defense:

15 October 2021

The research for this thesis was conducted at the company Alphalyse A/S in Odense, Denmark and was funded by The Horizon 2020 Marie Sklodowska-Curie Action ITN 2017 of the European Commission (H2020-MSCA-ITN-2017) through the Analytics for Biologics (A4B) project. A manuscript based on the work described in chapter 5.2.1 has been accepted for publication in the journal *Amino Acids* in a special issue on PTMs.

I. Table of contents

Contents

II. Abbreviations and naming conventions

Abbreviation	Name
Attribute	A specific protein modification located at a defined amino acid within the sequence of a proteoform
CEX	Cation exchange chromatography
CHO cells	Chinese hamster ovary cells
CQA	Critical quality attribute
ER	Endoplasmic reticulum
GM-CSF	Granulocyte-macrophage colony-stimulating factor
HILIC	Hydrophilic interaction liquid chromatography
HPLC	High pressure liquid chromatography
IE	Ionization efficiency
IgG	Immunoglobulin G
m/z	Mass-to-charge ratio
MAM	Multi-attribute method/methodology
MS	Mass spectrometry
PAP	Pulmonary alveolar proteinosis
Pharmacodynamics	The study of how the drug affects the organism
Pharmacokinetics	The study of how the organism affects the drug
Proteoform	Different molecular forms of a protein arising from the same gene are termed proteoforms of that protein (1)
PTM	Post translational modification
QbD	Quality by design
QC	Quality control
rCE-SDS	Reduced capillary electrophoresis sodium dodecyl sulphate
RP	Reversed phase

1. Zusammenfassung

Bei der Herstellung von therapeutischen Proteinen entstehen Proteoformen, die die Wirksamkeit und Sicherheit des therapeutischen Proteinpräparats beeinträchtigen können. Daher ist es von großer Bedeutung, solche kritischen Proteoformen zu charakterisieren, zu quantifizieren und zu entfernen, sodass ihr Anteil niedrig gehalten werden kann. Quantitative Bottom-Up-Proteomik mit Flüssigchromatographie gekoppelt an Tandem-Massenspektrometrie (LC-MS/MS) ist eine leistungsfähige Methode, um solche Proteoformen mit hoher Selektivität und Sensitivität zu erfassen und zu quantifizieren. Dies ermöglicht es, spezifische chemische Zusammensetzungen, die Proteoformen beinhalten (Attribute), zu charakterisieren und gegebenenfalls auch mit funktionellen Eigenschaften zu verbinden. Obwohl solche Methoden, die in der pharmazeutischen Industrie gemeinhin als Multi-Attribut-Methoden (MAM) bezeichnet werden, einen hohen Informationsgehalt über das therapeutische Protein-Präparat liefern, gibt es dennoch einige Aspekte, die von den herkömmlichen MAMs nicht abgedeckt werden. Dies sind zum Beispiel Analysen von nicht gereinigten Proben, bei denen die Beeinflussung des Signals durch Wirtszellproteine ein Problem darstellen können, sowie strukturelle Aspekte einiger Proteoformen. Das Hauptziel dieser Arbeit war es, aktuelle Bottom-Up-basierte Proteomik-Ansätze für die Quantifizierung von niedrig abundanten Proteoformen in therapeutischen Proteinen im Hinblick auf die Selektivität und Sensitivität für nicht gereinigte Proben und die Art der Proteinmodifikationen, die abgedeckt werden, zu verbessern. Es wurde ein MAM entwickelt, das auf der Quantifizierung durch Messungen im datenunabhängigen Aufnahmemodus (DIA) basiert, um Signalstörungen durch Wirtszellproteine zu umgehen und die Identifizierung und Quantifizierung definierter Modifikationen an einzelnen definierten Aminosäuren zu gewährleisten. Die Methode wurde für die Quantifizierung von Proteoformen eines Biosimilars des therapeutischen monoklonalen Antikörpers (IgG) Adalimumab optimiert und umfasst eine DIA-Spektrenbibliothek, die aus 15 von 20 Asparagin-Deamidierungsstellen, 2 von 29 Glutamin-Deamidierungsstellen, 5 von 5 Methionin-Oxidationsstellen und 12 N-Glykanvarianten besteht. Ein IgG enthaltenes Zellkulturfiltrat wurde mittels LC-MS/MS analysiert und die Quantifizierung der Modifikationsgrade wurde mit der DIA-basierten Methode oder einer auf dem datenabhängigen Aufnahmemodus (DDA) basierenden Methode verglichen, wobei letztere der konventionell verwendete MAM-Ansatz ist. Die mit der DIA-basierten Methode erzielten quantitativen Ergebnisse unterschieden sich bei Modifikationsgraden von > 2 % um nicht mehr als 20 % von denen, die mit DDA erzielt wurden. Die beiden Ansätze lieferten eine ähnliche Präzision für die Quantifizierung des Modifikationsgrades in technischen sowie biologischen Replikaten, wobei der gleiche Cutoff-Wert verwendet wurde. Signalinterferenzen in den Vorläuferspektren wurden bei drei deamidierten Peptiden mit einem Modifikationsgrad unter 2 % festgestellt. Dies führte zu einer Überschätzung des Deamidierungsgrades einer Position und möglicherweise zu den hohen CV-Werten, die für die beiden anderen Positionen gemessen wurden. Es wurde gezeigt, dass die Signalinterferenz bei allen drei Peptiden vermieden werden konnte, indem die Quantifizierung auf extrahierten Ionenchromatogrammen (XICs) von Fragment-Ionen aus der DIA-basierten MS-Methode anstelle der Vorläufer-Ionen aus einer DDA-basierten MS-Methode basierte. Als die beiden Quantifizierungsansätze auf deamidierte Peptide in geringer Menge in einer gereinigten IgG-Probe angewendet wurden, die zwei Wochen lang einem forcierten Abbau bei pH 9 und 37 ℃ unterzogen worden war, zeigte der DDA-basierte Ansatz im Gegenteil eine höhere Präzision als DIA. Dies könnte darauf zurückzuführen sein, dass die Vorläufer-Ionen eine höhere Assay-Empfindlichkeit als die Fragment-Ionen aufweisen, wenn keine Signalinterferenzen auftreten. Des weiteren wurde die Linearität des DIA-basierten MAM für die Quantifizierung der Methionin-Oxidation demonstriert und erwies sich mit R2-Werten > 0,99 als hoch, wenn die für die Quantifizierung verwendeten Fragmente optimiert wurden. In dieser Studie wurde außerdem eine neuartige Modifikation von +70 Da

beobachtet, die in Proben von in E. coli exprimiertem rhGM-CSF identifiziert wurde. Die Modifikation konnte am N-Terminus des Proteins und an Lysinresten lokalisiert werden. Außerdem enthielt die Modifikation wahrscheinlich eine Carbonylgruppe, da ihre Masse durch Reaktion mit dem Boran-Pyridin-Komplex um 2 Da zunahm und sie mit 2,4-Dinitrophyenylhydrazin reagierte. C4H6O wurde als elementare Zusammensetzung der Modifikation vorgeschlagen und das Reaktionsprodukt von Crotonaldehyd durch Michael-Addition als chemische Struktur. Das Vorhandensein eines Carbonylanteils deutete darauf hin, dass die Modifikation für die therapeutische Wirkung des Proteins entscheidend sein könnte, da die Proteincarbonylierung mit verschiedenen Krankheiten und dem Proteinabbau in Verbindung gebracht wurde. Darüber hinaus wurde eine MAM-Methode für die Quantifizierung von Disulfid-Scrambling in rhGM-CSF optimiert, indem proteolytische Enzyme ausgewählt wurden, die disulfidverknüpfte Peptide mit wenigen Disulfidbindungen und mit geeigneten Peptidgrößen für die Bottom-up-Proteomik generieren. Durch die Kombination der Enzyme GluC und Chymotrypsin wurden hauptsächlich disulfidverknüpfte Peptide mit einer Disulfidbindung und mit Peptidgrößen von 9-15 Aminosäuren erzeugt. Mit der MAM-Methode wurde das Disulfid-Scrambling in Proben, die einem forcierten Abbau unterzogen wurden, in einer unter reduzierenden Bedingungen gelagerten Rückfaltungsprobe und in zwei Entwicklungsproben des Produktionsprozesses quantifiziert. Die Protein-Deamidierung wurde in Proben, die einem forcierten Abbau unterzogen wurden, über einen Zeitraum von 3 Wochen verfolgt. Der Grad der +70 Da modifizierten Positionen wurde in den Prozessentwicklungsproben quantifiziert, zusammen mit der Aminosäuresubstitution von Isoleucin zu Valin und der Methionin-Oxidation. Die quantitative Genauigkeit des Disulfid-Scrambling-Assays kann durch unterschiedliche Ionisierungseffizienzen der verschiedenen disulfidverknüpften Peptide beeinflusst werden. Zusammenfassend wurden in dieser Arbeit mehrere MAMs entwickelt, die Aspekte abdecken, die von den konventionellen quantitativen Bottom-Up-Ansätzen nicht erfasst werden, und die das Potenzial haben, die Qualität von Produktionsprozessen für therapeutische Proteine zu verbessern. Darüber hinaus wurde eine neuartige Modifikation in rhGM-CSF, das in E. coli produziert wurde, identifiziert, die für die therapeutische Wirkung des Proteins entscheidend sein könnte.

2. Abstract

During therapeutic protein production, proteoforms are formed that can affect the efficacy and safety of the drug. It is therefore of high importance to characterize, quantify and remove such critical proteoforms so that their levels may be maintained low. Quantitative bottom-up proteomics using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) is a powerful method to characterize and quantify such proteoforms with high sensitivity. This enables connecting specific modifications located at defined amino acids within the sequence of a proteoform (attributes) to functional effects on the therapeutic protein. Although such methods, commonly called multi-attribute methods (MAMs) in the pharmaceutical industry, yield a high amount of information about the therapeutic protein, there are still some aspects that are not covered by conventional MAMs. Those are for example measurements in non-purified samples where signal interference from host cell proteins is an issue and the structural aspects of some proteoforms. The main aim of this thesis was to improve current bottom-up based proteomics approaches for quantification of low abundant proteoforms in therapeutic proteins with regard to the assay performance in non-purified samples and the kind of protein modifications that are covered. A MAM based on data-independent acquisition (DIA) quantification mode was developed to circumvent signal interference from host cell proteins and to assure identification and quantification of specific amino acid modifications located to single defined amino acid sites. The method was optimized for the quantification of proteoforms of a biosimilar of the therapeutic monoclonal antibody (IgG) adalimumab including a DIA-spectral library consisting of 15 out of 20 asparagine deamidation sites, 2 out of 29 glutamine deamidation sites, 5 out of 5 methionine oxidation sites and 12 N-glycan variants. A cell culture filtrate of the IgG was analyzed by LC-MS/MS and quantification of the modification degrees was compared using the DIA based method or a data-dependent acquisition (DDA) based method, the latter being the conventionally used MAM approach. The quantitative results obtained with the DIA based method did not differ from those obtained by DDA by more than 20% for modification degrees measuring > 2%. The two approaches gave similar precision for quantification of the modification degree in technical as well as biological replicates, applying the same cutoff level. Signal interference in the precursor spectra was identified in three deamidated peptides measuring under 2%. This led to overestimation of the deamidation degree of one site and possibly to the high CV values measured for the other two sites. It was demonstrated that the signal interference could be avoided in all three peptides by basing the quantification on extracted ion chromatograms (XICs) from fragment ions from the DIA based MS method instead of the precursor ions from a DDA based MS method. When the two quantification approaches were applied on low-abundance deamidated peptides in a purified IgG sample which had been subjected to forced degradation at pH 9 and 37 ℃ for two weeks, the DDA based approach did, on the contrary, show a higher precision than DIA. This could be attributed to the precursor ions giving a higher assay sensitivity than the fragment ions, when signal interference was not an issue. Further, the linearity of the DIA based MAM was demonstrated for quantification of methionine oxidation and was found to be high with R^2 values > 0.99 when the fragments used for quantification had been optimized. In this study, in addition, a novel modification of +70 Da identified in samples of rhGM-CSF expressed in *E. coli* was characterized by chemical assays in combination with MS. The modification could be located at the protein N-terminal and at lysine residues. Further, the modification likely contained a carbonyl group since its mass increased by 2 Da by reaction with borane pyridine complex, and it reacted with 2,4-dinitrophyenylhydrazine. C_4H_6O was suggested as elemental composition of the modification, and the reaction product of crotonaldehyde through Michael addition as chemical structure. The presence of a carbonyl moiety indicated that the modification may be critical for the protein's therapeutic effect since protein carbonylation has been connected to various diseases and protein degradation. Furthermore, a MAM method was optimized for quantification of disulfide

scrambling in rhGM-CSF by selection of proteolytic enzymes generating disulfide linked peptides with few disulfide bonds and with appropriate peptide sizes for bottom-up proteomics. By combining the enzymes GluC and Chymotrypsin, mainly disulfide linked peptides with one disulfide bond were generated with peptide sizes of 9-15 amino acids. With the MAM method, disulfide scrambling was quantified in samples subjected to forced degradation, in a refolding sample stored under reducing conditions and in two production process development samples. Protein deamidation was followed in samples subjected to forced degradation over a 3-week period. The degree of +70 Da modified protein sites was quantified in the process development samples, together with amino acid substitution from isoleucine to valine and methionine oxidation. The quantitative accuracy of the disulfide scrambling assay may be affected by differences in ionization efficacies between the different disulfide linked peptides. In summary, several MAMs were developed in this work covering aspects which are not covered by conventional quantitative bottom-up approaches and which has the potential to improve the quality of therapeutic protein production processes. Further, a novel modification was identified in rhGM-CSF produced in *E. coli* which may be critical for the protein's therapeutic effect.

3. Introduction

3.1 Therapeutic proteins

Proteins are large biomolecules with a broad versatility of functions in the body, including catalyzing metabolic reactions, transporting molecules, and providing structure to cells and organisms. About 20,000-25,000 proteins are thought to be coded for by the human genome but the whole human proteome, including proteins with different amino acid sequence variations and modifications, has been estimated to be up to 100 times that size. (2,3) This high number of proteins with their own specific function poses a huge challenge when it comes to localizing causes of medical conditions. On the other side, it creates a great opportunity for the discovery of new protein-based drugs.

There are several advantages of using protein-based therapeutics as compared to small molecule-based drugs. Proteins usually have a highly complex set of functions that cannot be mimicked by a small chemical compound (3). This also makes therapeutic proteins less likely to interfere with other biological processes. Since many proteins used as therapeutics are naturally produced by the body, they may also be accompanied by a decreased immunological response. Moreover, protein therapeutics offers the possibility to replace proteins which are naturally produced by a gene that is deleted in the patient (4).

From an economic point of view, it is often easier to obtain far-reaching patent protection for therapeutic proteins than for small molecule-based drugs (3). The main reason is that therapeutic proteins are so unique in their function and form. The global therapeutic proteins market was in 2019 valued about $97.9 billion and is expected to grow to $170.6 billion through 2023 (5).

This background sets the ground for why it is worth investing the time and money in developing new protein-based therapeutics as well as assays to evaluate these therapeutics.

3.1.1 Recombinant protein expression

Very few therapeutic proteins are purified from their native source. Instead, they are commonly expressed by recombinant DNA technology (3). This means that the therapeutic protein's DNA is inserted into a DNA vector using specific enzymes. The vector is then transferred into a host cell which uses the vector to express the protein. The therapeutic protein can then be produced in high yield by promoting cell growth and expression of the vector. Recombinant DNA technology is advantageous in comparison to protein purification from the natural source, both from an economic perspective and from a functional perspective. High yields may be produced at lower cost and the drug becomes more homogenous and can therefore have a higher specific activity and a lower risk of developing immunogenic reactions. Moreover, the risk of transmitting diseases between humans and animals is eliminated.

There are a number of different expression cell systems available including bacteria, yeast, insect cells, mammalian cells and transgenic animals and plants. The choice of expression system depends on several factors from which cost of production and which protein modifications that are required for the protein's biological function are important.

3.1.2 Therapeutic antibodies

Monoclonal antibodies make up a significant amount of the therapeutic proteins approved for clinical use. Antibodies are Y-shaped proteins which are used by the immune system to detect and protect the body from foreign objects (6), see Figure 1. It contains one region on each of its arms that recognizes a specific foreign molecule, the antigen. The antigen binding sites on the antibody are produced with a wide variety of specificities in order to increase the chances that a foreign molecule

will be recognized by the immune system. It is therefore possible to raise antibodies against specific disease related molecules to be used in protein-based therapies.

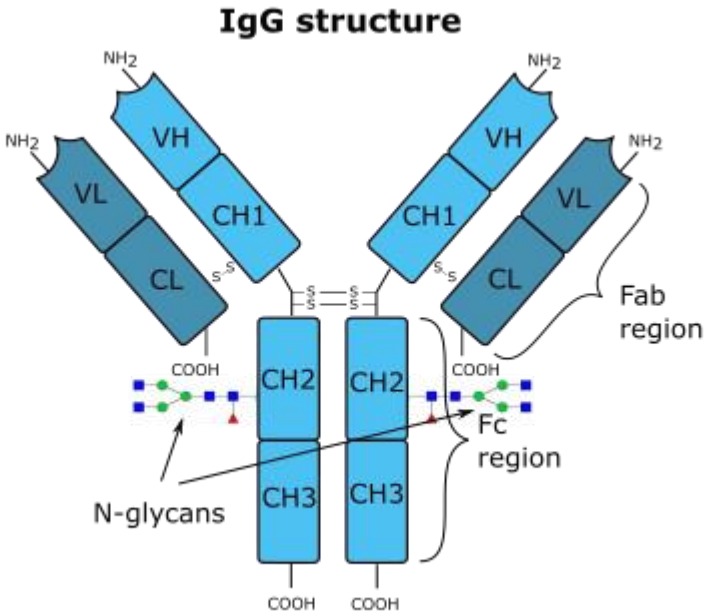


Figure 1: Overview image of the functional regions of an IgG antibody. VL = variable region, light chain; CL = constant region, light chain; VH = variable region, heavy chain; CH = constant region, heavy chain. The structure of the depicted N-glycans is just an example and were drawn using GlycoWorkbench (www.glycoworkbench.org).

The majority of the monoclonal antibodies approved for clinical use are of the class IgG. They are large, globular proteins weighing around 150 kDa and contain two identical heavy chains and two light chains being bound together by disulfide bonds, see Figure 1. The two arms of the antibody are called the Fab regions and contain the antigen binding sites, while the trunk of the Y-structure is called the Fc region and has the role to activate effector functions of the immune response upon Fab-antigen interaction. The Fc region consists of two constant regions from each of the two heavy chains and has essentially the same amino acid sequence in all antibodies from the same subclass. The Fab region consists of one variable region and one constant region from the heavy chain and from the light chain. The amino acid sequence of the variable regions varies with the target antigen of the antibody.

Therapeutic antibodies are generally designed to work by one, or several, out of four modes of action (7). They can work by neutralizing the target, meaning that the antibody binds to a ligand or receptor on a cell surface which results in that a cell signaling pathway becomes blocked. Antigen binding can also trigger antibody-dependent cell-mediated cytotoxicity (ADCC) or complement-dependent cytotoxicity (CDC). The antibody-antigen binding triggers the ADCC by binding immune effector cells to the Fc subunit and the CDC by binding the C1 protein following initiation of the classical complement pathway and finally leading to lysis of the target cell in both cases. The antibody can also function as a very specific delivery system for drugs to the target cell, delivering for example radioisotopes, toxins, drugs, or cytokines (8).

Genetically identical antibodies, or monoclonal antibodies, can today be produced by recombinant DNA technology. Monoclonal antibodies are usually produced in eukaryotic expression systems, with Chinese hamster ovary (CHO) cells being one of the most prevalently used system. One of the main reasons for this is that antibodies are dependent on their glycan structures to function properly, which are carbohydrate chains attached to the protein, and CHO cells produce very similar glycan structures

to those observed in humans. Protein glycosylation will be described in more detail in chapter 3.1.4.2. A therapeutic IgG recombinantly expressed by CHO cells is one out of two target proteins of this thesis.

3.1.3 Granulocyte macrophage colony stimulating factor

Granulocyte macrophage colony stimulating factor (GM-CSF) is a monomeric protein which is secreted by various immune cells to function as a cytokine and a white blood cell growth factor (9). Production of autoantibodies against GM-CSF in the lungs is a condition called autoimmune pulmonary alveolar proteinosis (PAP) (10). The condition causes disruption of the lungs' host defense and of the homeostasis of the lung surfactant and will finally lead to breathing issues and higher risk of infection. Biologically active forms of recombinantly expressed GM-CSF have been produced to be administered to patients suffering from PAP and other GM-CSF deficiency related diseases. Various expression systems have been employed for this purpose including plant and mammalian cells, yest and *E. coli* (11–13). From these expression systems, *E. coli* is perhaps the most employed system since it has the ability to grow fast at a high density and on cheap substrates. The mature form of GM-CSF is a 127 amino acid long protein with two N-glycosylation sites and several O-glycosylation sites (14). Its three-dimensional structure is partly defined by four cysteine residues connected by disulfide linkages between sites C56 and C96 and sites C88 and C121 respectively. However, the presence of the GM-CSF glycans have not been found to affect the protein activity. It is therefore possible to express GM-CSF in *E. coli* despite the fact that the host does not possess a system to add protein glycosylation (15). This fact may even be advantageous since it means that the number of possible proteoforms is reduced. GM-CSF is the second of two target proteins of this thesis.

3.1.4 Proteoforms in therapeutic proteins

Although recombinant protein expression usually yields more homogenous products than purification directly from the host, therapeutic proteins are still large molecules produced by highly complex cellular systems. As a consequence, there are many parameters during the protein production that may affect the final drug product. The production of a therapeutic protein starts by protein expression in the fermentor (16). This is followed by protein harvest, purification, polishing and formulation. If the production parameters are not carefully monitored through all of these steps it may lead to the formation of various proteoforms which may affect the potency of the drug as well as its pharmacokinetics and pharmacodynamics.

3.1.4.1 Amino acid substitution

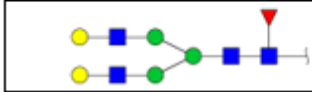
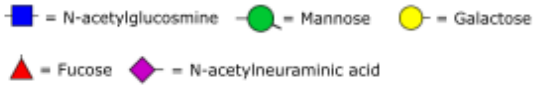
Proteoforms where one or several amino acids have been substituted to other amino acids are not uncommonly found in protein production batches (17). These sequence variants arise during protein expression and may be formed on the genetic level or on the translational level. Genetic sequence variants may for example come from error-prone DNA repair mechanisms, rearrangements, deletions and frame shifting. Sequence variants arising on the translational level is a result of misincorporation of amino acids. This may be an effect of an unbalanced cell culture leading to the depletion of one amino acid which may then be substituted by a similar amino acid. If the sequence variant is incorporated at a place in the amino acid sequence which is involved in the protein activity, this may have deleterious effects for the protein activity.

3.1.4.2 Glycosylation

After the DNA has been translated into a protein most expression systems have functions to perform post translational modifications (PTMs) to the proteins. These are reversible or irreversible chemical changes that are being made to the proteins following translation. Protein glycosylation is one of the most common PTMs introduced by eucaryotic expression systems. Glycosylation means the enzymatic addition of oligosaccharide molecules, glycans, to the protein. The most common types of

glycosylation in therapeutic proteins are N-linked and O-linked glycosylation. N-glycans are glycan structures attached to the nitrogen atom of an asparagine residue and O-glycans are linked to the oxygen atom of serine or threonine residues. While O-glycans are less commonly found among IgG-type antibodies, basically all IgGs contain one N-glycan on each of its heavy chains in the Fc portion of the protein. The N-glycans are produced in the endoplasmic reticulum in eucaryotic cells and are further processed in the Golgi apparatus (18). The most common N-glycans found in monoclonal antibodies are built from the oligosaccharides N-acetyl glucosamine, mannose, galactose, fucose and N-acetyl neuraminic acid and a few common structures can be seen in Table 1.

Table 1: Common N-glycan structures in monoclonal antibodies. The naming convention is based on the simplest bisecting glycan structure G0 to which monosaccharides are added or subtracted according to the format X(n). G: galactose, F: fucose, M: mannose, N: N-acetyl-glucosamine, S: N-acetylneuraminic acid (sialic acid). The high mannose structures Man5 and Man6 are exceptions and are based on the number of mannose residues added to the two core N-acetylglucosamine residues. The glycan structures were drawn using GlycoWorkbench (www.glycoworkbench.org). The table was inspired by (18).

	G0F		G0
	G1F		G1
	G2F		G2
	G0F-N		Man5
	G1F-N		Man6
	G1FS		

The N-glycans are highly important for the structure and function of the IgG. The role of the Fc-region of the antibody is to modulate the immune cell activity by binding to effector molecules, following binding of the Fab-region to the antigen. Changes in the N-glycan structure may lead to conformational changes of the Fc region which could change the binding affinity to the Fc receptors and subsequent changes in the effector functions (19). These include the complement-dependent cytotoxicity (CDC), the antibody dependent cell-mediated cytotoxicity (ADCC) and the antibody dependent cell-mediated phagocytosis (ADCP). It is well known that the absence of Fc N-glycans dramatically effects the binding to the different Fc receptors. Further, afucosylation, *i.e.* the absence of a fucose moiety on the N-acetyl glucosamine linked to the protein, has in several studies proved to give a better binding of the Fc region to some of the Fc receptors modulating the effector functions (18). It has also been shown that the Fc glycosylation may affect the protein clearance. IgGs with high amounts of terminal mannose are rapidly cleared out from the circulation upon administration. Further, glycan structures that are not recognized by the immune system may trigger an immune response (20).

Bacterial expression systems do not contain functions for addition of glycan structures. It may therefore be advantageous to use such systems for recombinant protein expression if glycosylation is

not important for the protein's mode of action, since this will reduce the number of possible proteoforms. One such example is production of therapeutic GM-CSF, see chapter 3.1.3.

3.1.4.3 Disulfide bonds

The three-dimensional structure of many proteins is achieved and stabilized by the formation of inter- and/or intrachain disulfide bonds (21). These are covalent bonds formed between the sulfur atoms of cysteine residues in the amino acid sequence. In for example the IgG molecule, the four IgG subunits are all held together by disulfide bonds and disruption of these would break the molecule apart (22). The formation of disulfide bonds is a PTM which in eukaryotic cells take place in the endoplasmic reticulum (ER). The proteins that require disulfide bonds for their function are directed to the ER for processing by a signal peptide, which is then cleaved off after processing (21). In bacteria, such as *E. coli*, formation of disulfide bonds takes instead place in the periplasm. Upon entering the periplasm, the protein is subjected to a number of enzymes that oxidizes the cysteine residues in an ordered fashion to form disulfide bonds. The yields of correctly folded proteins from recombinant expression in *E. coli* are however highly unpredictable and are very dependent on the expression conditions and the protein being expressed. A number of mutant strains of *E. coli* has, therefore, been developed with superior systems for disulfide folding. An alternative strategy is refolding of the recombinantly expressed protein purely by chemical strategies. During high-level expression of many recombinant proteins in *E. coli* the proteins start to form insoluble aggregates, called inclusion bodies (23). These may be separated from the majority of the rest of the bacteria by cell lysis and centrifugation during protein harvest. The pellets are then solubilized using detergents and chemicals that reduce the disulfide bonds, followed by controlled protein refolding and formation of disulfide bonds using agents with a combination of reduced and oxidized thiols. These different folding strategies all require highly optimized conditions to achieve high yields of correctly folded proteins. Further, the purification and formulation steps following the refolding may contribute to mis-assembly or degradation of disulfide bonds (24). Elevated pH and temperatures have been found to induce rearrangements of disulfide bonds. So has also common formulation additives like for protein albumin and thiol containing molecules like cysteine/cystine and glutathione. Scrambling of disulfide bonds may change the three-dimensional structure of therapeutic protein, leading to changed binding characteristics. Further, free cysteine residues are highly reactive and may for example form cysteine bonds to other proteins, leading to protein aggregation and compromising the drug safety. Disulfide bond formation is therefore a PTM which is highly important to monitor during protein manufacturing.

3.1.4.4 Deamidation

Another PTM that commonly occurs in therapeutic proteins is asparagine and glutamine deamidation. Deamidation can occur as a non-enzymatic reaction and starts with the peptide bond in the amino acid at position N+1 attacking the asparagine or glutamine carbonyl carbon to release ammonia, see Figure 2 (25). This results in the formation of a ring-shaped intermediate called succinimide, which is hydrolyzed to either isoaspartic acid/isoglutamic acid or aspartic acid/glutamic acid.

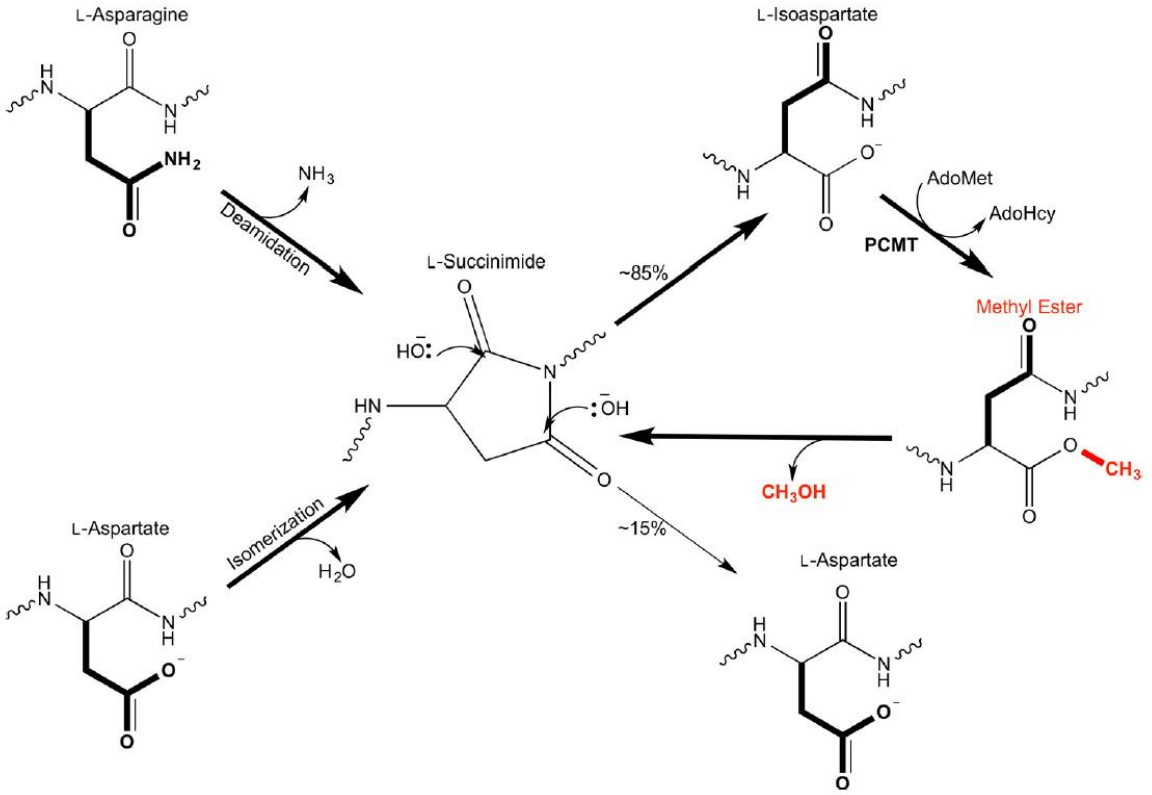


Figure 2: Deamidation of asparagine residues and isomerization of aspartic acid residues in proteins. The restoration of isoaspartic acid by the repair enzyme protein isoaspartyl methyltransferase (PIMPT or PCMT) is also shown. Taken from (26).

Asparagine residues are typically deamidated at a much higher rate than glutamine residues. Deamidation may have two large implications on the tertiary structure of the protein (27). Firstly, the neutral or slightly basic asparagine residue is converted to an acidic residue which may change the hydrogen bonding structure. Secondly, the formation of the isoaspartate residue involves incorporation of an extra methylene unit into the peptide backbone. Both these changes will affect the local environment of the initial residue and may have a large impact on the folding structure of the protein. Deamidation is also commonly known as one of the processes by which protein degradation is controlled in vivo.

Non-enzymatic deamidation is affected by a number of parameters such as pH, temperature and ionic strength (28). Alkaline conditions are for example one of the factors that are known to increase the deamidation rate. A low pH may, however, induce conversion of aspartate residues to isoaspartic acid. The amino acids proximal to the asparagine/glutamine have also proven to be important for the deamidation rate, with asparagine with a glycine residue at position N+1 being one of the most susceptible combinations. Deamidation may be induced during all steps of the protein production and during protein storage as well. It is therefore an important PTM to monitor in the drug product.

3.1.4.5 Oxidation

Methionine oxidation is one of the most common PTMs identified in therapeutic proteins. Oxidation of the methionine thiol group leads to formation of methionine S-oxide, see Figure 3 (29). This product may then be further oxidized to methionine S,S-dioxide.

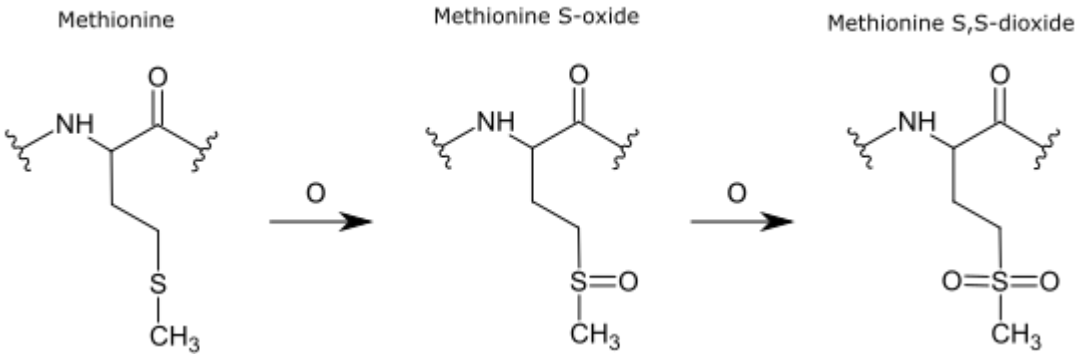


Figure 3: Methionine oxidation, first to S-oxide and then to methionine S,S-dioxide.

Protein oxidation is related to protein degradation, and many age-related diseases are associated with the accumulation of methionine S-oxide. In therapeutic IgG, methionine oxidation in the Fc region resulted in weaker binding to Protein A, a protein often used in affinity purification of the recombinantly expressed IgG (30). It further led to decreased binding affinity to the neonatal Fc receptor and to the Fc gamma receptor, leading to a reduced immune response, and to a shorter antibody half-life. Other therapeutic proteins, such as the Granulocyte colony-stimulating factor (G-CSF), have also shown reduced therapeutic activity upon oxidation. Further studies have shown that oxidation may lead to changes in the protein secondary and tertiary structure and to the formation of protein aggregates which may compromise the drug safety. Besides methionine, amino acid oxidation may also occur of cysteine, histidine, phenylalanine, tryptophan, and tyrosine residues.

Like deamidation, amino acid oxidation may occur through all steps of the protein production process. An imbalance in the levels of dissolved oxygen in the expression medium during protein expression has proven to lead to oxidation products. Metal ions is another potent source to oxidative stress. These may be present for example as impurities in buffers or additives, at surfaces or through metal affinity chromatography. Oxidative impurities, such as formaldehyde and hydrogen peroxide, has also been identified in polymeric excipients which are added to the protein formulation. These polymeric excipients, like polyethylene glycol (PEG) or polysorbate, may also oxidize spontaneously in aerobic environments generating several peroxides. To limit protein oxidation, the addition of different antioxidants have been suggested (31). Because of this high number of oxidative sources, amino acid oxidation remains one of the most important PTMs of therapeutic proteins to monitor but also one of the most difficult.

3.2 Characterization methods

3.2.1 Critical quality attribute characterization

In development of therapeutic proteins, it is important to identify which product components that may have a negative effect on the performance of the product. A critical quality attribute (CQA) is defined as a physical, chemical, biological, or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality (32). This includes proteoforms containing protein modifications, such as those described in the previous chapter, which may affect the efficacy or safety of the drug. Regulatory authorities recommend using a quality by design (QbD) approach during development of therapeutic proteins. This means that the product quality should be ensured by having a thorough understanding of the product and the processes by which it is produced, so that any risks that may affect the drug quality can be avoided. The QbD principles involves the development of a quality target product profile which identifies CQAs and appropriate strategies to make sure the CQAs are maintained within acceptable levels. In order

to achieve an accurate quality target product profile, it is important to use accurate and specific tools to characterize and quantify the potential CQAs.

3.2.2 High performance liquid chromatography

Proteoforms of therapeutic proteins may be characterized by a range of different analytical techniques. Analysis by high pressure/ performance liquid chromatography (HPLC) is one of the most employed techniques and has been widely used since its introduction in the 1960s. HPLC is based on a liquid containing the analytes of interest being passed with a high pressure over a column filled with a solid material with adsorbent properties. The analytes contained in the sample will move with different velocities through the column because of differences in their interaction with the solid material. The components will therefore be recorded as separate peaks when they elute from the column. Proteoforms may be separated from each other by HPLC based on various properties depending on the chemical groups linked to the solid material, usually attributed the solid phase, and depending on the composition of the liquid, usually attributed the running phase. For example, the native protein may be separated from the more acidic deamidated proteoforms and from the more basic oxidized proteoforms by using a charged stationary phase. This technique is called ion exchange chromatography. Proteoforms may also be separated based on their hydrophobicity. In reversed phase chromatography (RP) the proteins are passed through a hydrophobic stationary phase using a polar liquid phase. The most hydrophilic molecules will pass fastest through the column while the most hydrophobic molecules will require a less polar solvent for elution. RP chromatography is therefore usually performed using a gradient of a polar liquid phase, such as water, and a non-polar liquid phase, such as acetonitrile. The chromatographic performance may be adjusted through a number of parameters, such as selection of chemistry in the stationary or running phase, the particle size of the stationary phase and the column dimensions.

3.2.3 Spectrophotometric methods

When the proteoforms have been chromatographically separated, a detector is required to pick up the signal form the protein. One of the most common detectors used for quality control of therapeutic proteins are spectrophotometric detectors in which light is used for analyte quantification. Proteins absorb light of wavelength 280 nm particularly well, because of the presence of the aromatic amino acid residues tryptophan, tyrosine and phenylalanine. The degree of absorbed light at this wavelength may therefore be used to quantify the proteoforms eluting from the HPLC. This technique requires prior knowledge or further analysis of the composition of the detected peaks. It is also highly dependent on the chromatographic resolution of all relevant proteoforms. The QbD strategy, recommended by regulatory authorities, requires information on the drug product on the molecular level (33). These profile-based methods do therefore not align with the QbD principles.

3.2.4 Mass spectrometry

Mass spectrometry (MS) has become a valuable tool for detection and characterization of proteoforms (22,25,34). Commonly the proteins are being separated by RP HPLC, as described in chapter 3.2.2, before being injected into the MS (35). There the proteins are ionized, and the ion intensities are detected as a function of their mass-to-charge ratio (m/z). The protein masses can then be calculated from these spectra. This analysis strategy is commonly called LC-MS, liquid chromatography coupled to mass spectrometry. Other chromatographic techniques than RP may be employed (36). However, injection of non-volatile salts into the mass spectrometer will lead to loss of analyte signal. Therefore, only volatile salts can be used for MS analysis, such as ammonium bicarbonate or ammonium acetate. Two different strategies may be employed to achieve molecular information about the protein structure. These techniques are called bottom-up and top-down proteomics (37). The most traditionally used approach is bottom-up which involves the proteins being

digested by a proteolytic enzyme into smaller peptides (38), see Figure 4. The digest is then separated by RP HPLC coupled online to a mass spectrometer. The peptides are ionized in the MS source and their m/z values are being detected by a first mass analyzer (MS1). Selected peptides are then sent into a collision cell where they are being fragmented, so that the peptide bonds break at several places, followed by detection by a second mass analyzer (MS2). These spectra can then be used to achieve information about the peptide sequence as well as the location of potential modifications. The first detection step is usually called the MS level (MS1) and the detection after the peptide fragmentation is called the MS/MS level (MS2). This data acquisition strategy is therefore commonly called LC-MS/MS or tandem mass spectrometry, since sequential mass analyzers are being used. The peptides and potential modifications in the sample are commonly identified by matching the recorded MS/MS spectra against theoretical spectra from the expected proteins in the sample (39). A search engine is fed with information about the sample, including a database containing the sequences of the expected proteins, the digestion conditions, expected modifications and the expected precision of the instrument. Each recorded MS/MS spectra is then matched against all theoretically calculated MS/MS spectra, the best match is found, and the match is given a score based on the probability that the observed match is a random event. If enough peptides from a certain protein are identified with an acceptable score, that protein can be considered identified.

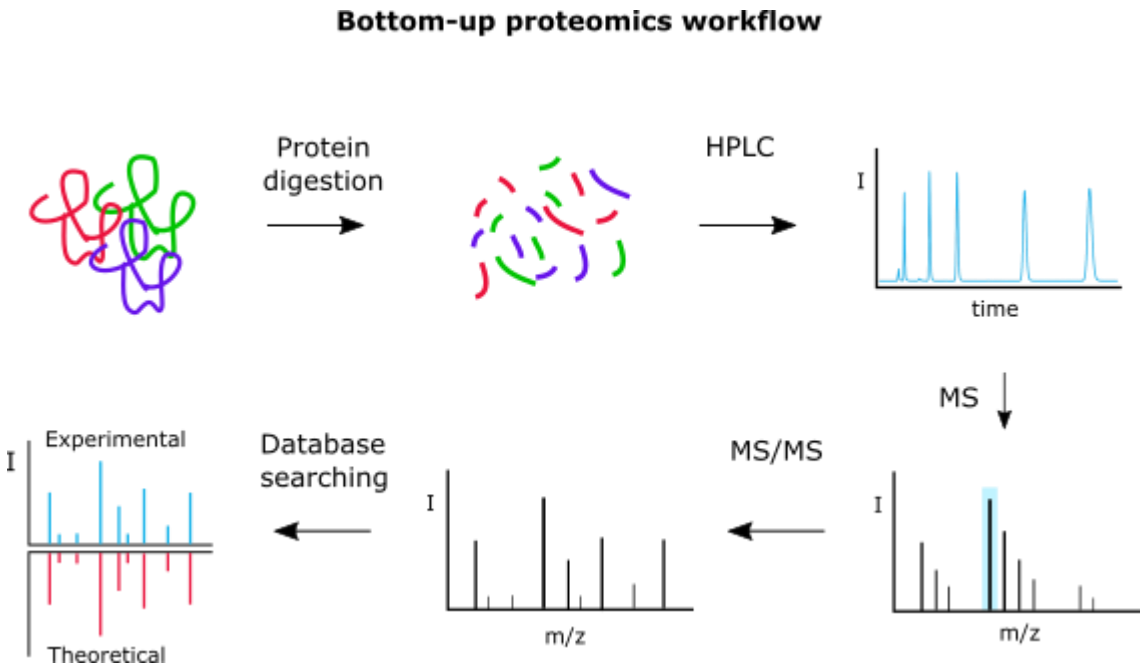


Figure 4: In bottom-up proteomics the protein is digested with an enzyme into peptides. These peptides are then separated by RP HPLC followed by ionization and detection in the MS. The peptides are fragmented individually in a collision cell, a step called MS/MS, followed by detection. The blue box in the MS spectrum exemplifies selection of a peptide with a certain m/z value for MS/MS fragmentation. The MS/MS spectra are then matched against theoretical peptide spectra using a protein database search engine to identify the peptide amino acid sequence, the originating protein, and potential modifications.

Top-down proteomics is a newer technique than bottom-up and involves analysis of intact proteins by MS/MS (34). The advantage of using top-down is that information can be achieved on the complete protein composition of a specific proteoform as compared to bottom-up approaches when this information is lost. Drawbacks are that it is less sensitive to detection of low abundant proteoforms, it produces very complex data to analyze and requires the use of high-resolution MS instrumentation. In this thesis only bottom-up methods were applied.

3.2.4.1 Electrospray ionization (ESI)

The standard technique for ionization of peptides analyzed by LC-MS/MS is electrospray ionization (ESI) (35). ESI is a soft ionization technique, meaning that it mainly leaves the peptides intact upon ionization. It is also compatible with peptides being continuously introduced into the MS by RP HPLC. In ESI the liquid sample is introduced into the source region through a small steel or glass capillary under atmospheric pressure (40). A strong electric potential is applied between the capillary and a counter electrode. This leads to the formation of a jet of charged droplets which are being evaporated as they travel down a pressure gradient and finally only the charged analyte will be left. The process of formation of the charged analyte is highly complex and will not be described in further detail here (41). Peptides and proteins are usually analyzed in positive polarity mode meaning that the analytes are sprayed at low pH in order to improve the formation of positively charged ions. ESI normally produces multiply charged analytes and the number of charged residues is related to the number of basic amino acids in the molecule. In all MS approaches applied in this work, ESI was used for peptide ionization.

3.2.4.2 Peptide fragmentation by MS/MS

Peptide fragmentation by MS/MS may be performed by different strategies from which the most common strategy is collision induced dissociation (CID). In CID the peptides are collided with uncharged gas atoms (42). This will result in the transfer of kinetic energy to vibrational energy, finally resulting in the breaking of peptide bonds. The m/z values of the fragments can then be used to determine the amino acid sequence and eventual modifications of the fragmented peptide, commonly called the precursor, since the mass of the precursor is known. Several bonds in the peptide backbone may break during fragmentation being the Cα-C, the C-N or the N-Cα bond. Further, the fragment may contain either the C-terminal or the N-terminal of the peptide. A naming nomenclature has therefore been developed for the different possible fragments (43), see Figure 5.

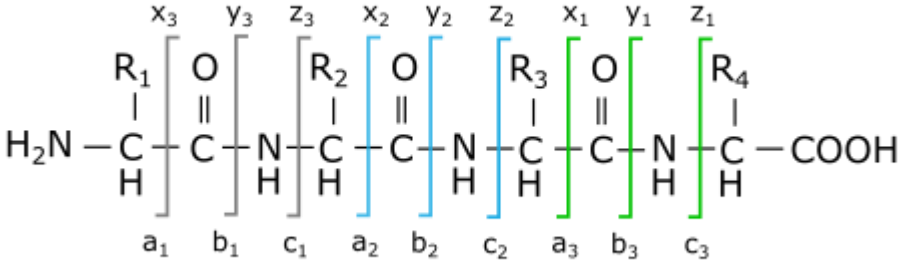


Figure 5: Nomenclature for peptide fragments including the peptide N-terminal and C-terminal respectively. Image inspired by(43).

Fragments involving the N-terminal are named a_n, b_n and c_n and represent the cleavage of the three different peptide backbone bonds in the order mentioned above. In the same way fragments involving the peptide C-terminal are named x_n, y_n and z_n and represents the cleavage of the same three peptide backbone bonds. The superscript *n* indicates the number of amino acid residues in the fragment. Fragmentation by CID mainly produces b and y type ions.

3.2.4.3 MS data acquisition

There are different strategies to acquire the MS/MS data for bottom-up analysis of proteins. Data dependent acquisition (DDA) is one widely used strategy. During DDA all peptides within a certain m/z range are detected at the MS level (37). Then certain peptides within this range are selected for

fragmentation by a defined strategy, which commonly involves selection of the most intense ions. Since the time between two MS scans, commonly called the cycle time, is defined in the MS method there will only be enough time to fragment a limited number of peptides within one cycle. This means that if the same ions have the highest intensities throughout several cycles these would be fragmented multiple times leaving other, lower abundant ions, unfragmented. Therefore, exclusion lists are commonly used in DDA based MS methods, meaning that when an ion with a certain m/z value has been fragmented it will not be fragmented again for a defined time period (44). This is motivated since the fragmentation is usually used only for identification of the peptide while the precursor signal is used for quantification. However, not the same ions will always be selected for fragmentation between replicate runs. This may pose a problem if, for example, a new modified peptide is identified in a new production batch of a therapeutic protein. It may then be difficult to judge if this is a new modification or if the peptide was simply not selected for fragmentation during the analysis of the previous batch.

As a complement to DDA based strategies, data independent acquisition (DIA) based strategies perform fragmentation of all precursor ions within the selected m/z range in each cycle. When the DIA strategy SWATH-MS is applied, the precursors are fragmented in sequential windows (45). As an example, the m/z range 300-1700 may be divided into 20 windows á 70 m/z. The peptides are then first detected at the MS level. This is followed by window fragmentation, first 300-370 m/z, then 370-440 m/z and so on, covering the whole m/z range. This should be compared to DDA where individual precursor peaks are selected for fragmentation. Figure 6 illustrates the difference between DIA and DDA. The advantage of DIA is that all precursor ions are fragmented independently of the intensity of the other detected ions. This improves the reproducibility of the identified peptides between runs (47). It also enables quantification from the fragment ion signal. A disadvantage is that several precursor ions may be fragmented in one window, leading to higher complexity of the MS/MS spectra. This is usually overcome by building an ion library typically based on DDA. The DDA based library contains the fragment m/z values from each precursor ion subjected to MS/MS and can be used to search the SWATH window covering the right precursor ion for the matching fragments ions.

Alternative strategies to monitor peptide fragmentation reproducibly is selective reaction monitoring (SRM) or multiple reaction monitoring (MRM) in which defined precursor ion-fragment ion pairs are monitored (37). These methods do however require prior knowledge of all ions of interest and the number of pairs that are possible to monitor is limited.

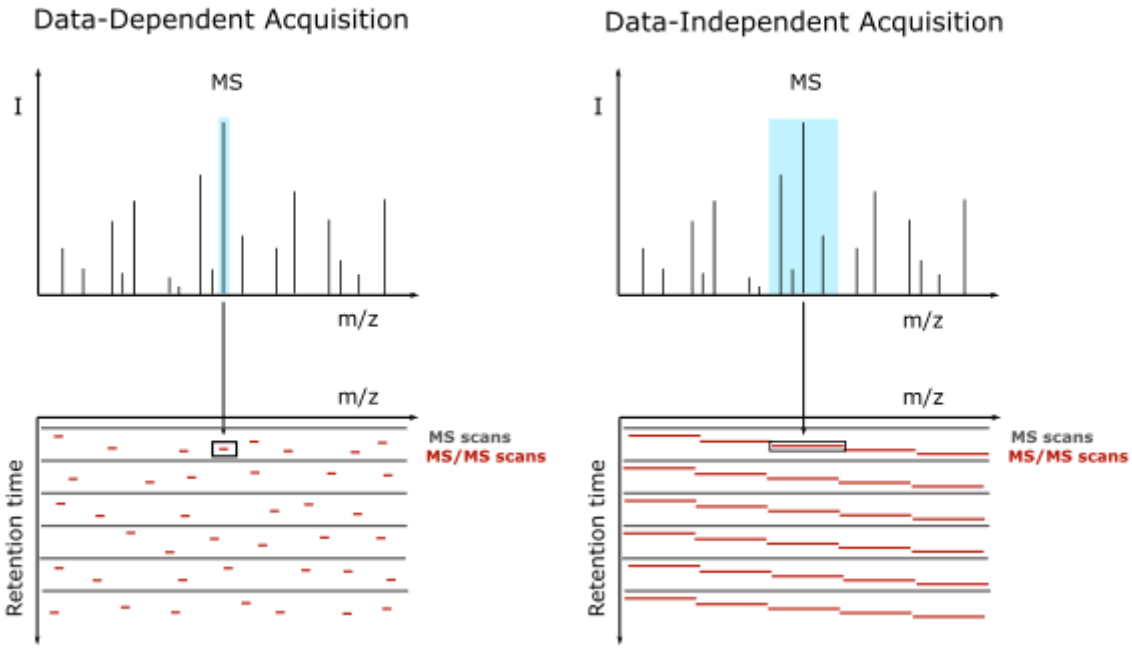


Figure 6: In both DDA and DIA the precursor ions are detected on the MS level. This is, during DDA, followed by MS/MS fragmentation of some precursor ions, based on predefined selection parameters. When the DIA strategy SWATH-MS is applied, all precursor ions are instead fragmented in sequential m/z windows. This means that all precursor ions will be fragmented in each cycle during DIA, as compared to DDA when they are not. Image inspired by (46).

3.2.4.4 Quantitative MS data analysis

When the LC-MS/MS data has been acquired, the peptides in the sample are identified by searching the data against a database search engine, as described in the introduction to chapter 3.2.4. These search results are then used to build a spectral library, which is a non-redundant collection of annotated and curated LC-MS/MS peptide spectra (48). The sample data is searched against this library to extract the ion intensities over time, called extracted ion chromatograms (XICs). These chromatograms are then used to perform quantitative peptide measurements (49). The spectral libraries are commonly built from DDA data but recently some software have also introduced spectral library building from DIA data (45).

3.2.5 Multi- attribute methodology (MAM)

3.2.5.1 What is MAM?

Bottom-up proteomics workflows have been suggested for use in quality control (QC) of therapeutic proteins (50–52). By applying bottom-up based approaches, several modification types, such as charge, glycan and clipped variants, can be relatively quantified in the same assay and in a site-specific manner. This has the advantage that both information about the kind of modification and its location within the amino acid sequence can be used to determine its effect on the therapeutic protein, *i.e.* if it is a CQA. It does also align with the QbD principles recommended by regulatory authorities. These bottom-up based approaches can be compared to conventional QC approaches in which a number of assays, such as hydrophilic interaction chromatography (HILIC), cation exchange chromatography (CEX) and reduced capillary electrophoresis-sodium dodecyl sulphate (rCE-SDS), are applied on the intact protein level, and from which the site-specific information cannot be achieved (53). The bottom-up based quantitative approaches described have by the industry been called the multi-attribute methodology/method (MAM), since several protein modifications located at defined amino acids,

attributes, may be quantified in the same method. The MAM workflow usually includes protein denaturation and reduction of the disulfide bonds followed by alkylation of the cysteines. The protein is then digested into peptides by a protease. The most employed protease is Trypsin since it has a high specificity for digestion C-terminally to lysine and arginine residues which reduces the data complexity compared to less specific proteases. Both lysine and arginine are also basic amino acids which ensures good ionization of the peptides. The data from the digest is then acquired by LC-MS/MS by DDA which is used to build a spectral library as described above. The data is searched against this spectral library and the XICs from all the identified precursor ions are extracted. The relative modification degree is then quantified from the peptide with the best signal and covering the amino acid site of interest. The XIC areas from the precursor ions of the modified variant of this peptide are summed and divided by the summed precursor XIC areas from the modified and non-modified variants of the peptide. This strategy has been established since the difference in ionization efficiency between the modified and non-modified variant of a peptide is negligible for most protein modifications. Several studies have shown the applicability of MAMs to follow the formation of CQAs such as deamidation, oxidation and N-glycan variants (52,54,55). Robust and accurate MAMs have also been demonstrated for transition between labs (56,57). Although there seems to be a large interest in using MAMs for quality control of therapeutic proteins, not so many such methods have been introduced as release assays yet (58). In this thesis two MAMs were developed for two different therapeutic proteins with the aim of solving aspects that are not yet solved with conventional MAMs. It should be highlighted that information about the complete composition of specific proteoforms are lost in bottom-up proteomics. It may therefore provide additional information to the QC assay to also include a method based on intact protein analysis.

3.2.5.2 Sample preparation induced modifications

Some modifications in therapeutic proteins, such as deamidation and oxidation, may be induced while preparing the sample to be monitored by the MAM. It is therefore difficult to monitor the true quantities of these modifications from just one protein sample. The sample preparation conditions should be optimized so that the formation of these modifications is kept as low as possible, and the quantities are usually monitored compared to a reference standard (56).

The chemicals used for the sample preparation may also cause modifications of the protein. As an example, the commonly used denaturing agent urea may form isocyanic acid at elevated temperatures which may react with primary amines to form carbamyl adducts (59). Further iodoacetamide, which is commonly used to alkylate reduced cysteine residues, may also alkylate other functional groups such as lysine residues and the protein N-terminal. It is important to keep these modification levels low in order to avoid loss of peptide signal as well as reducing the risk of false peptide identifications (60).

3.2.5.3 MAM in complex samples

Today, most MAMs are based on DDA (53,54,56). In DDA based MS strategies the peptides are identified on the MS2 level while quantification is performed on the MS1 level. This approach has been proven successful for quantification of modifications in purified therapeutic proteins. These methods are however rarely used to quantify modifications in complex samples, such as samples taken directly form the bioreactor, even though such studies have also been demonstrated (61). A reason for this may be the risk of signal interference from host cell proteins (HCPs). It would be highly advantageous if protein modifications could be quantified through all steps of the protein production process since it would make it easier to identify critical process steps where the modifications may be introduced. This would also reduce production process development times and costs since expensive and tedious purification steps could be avoided for many cases. Another drawback of DDA based

MAMs is that if novel modifications are discovered at a later stage in the protein production process, it may not be possible to identify these in earlier batches if they are low in abundance, since they may not have been selected for fragmentation.

An alternative strategy could be to acquire the MAM data by DIA. In DIA based approaches the fragmentation data from all detected precursor ions are acquired throughout the run. This opens for using the fragmentation ions for quantification in a MAM. MS/MS fragmentation of peptides usually results in a range of fragment ions from each precursor ion. It may therefore be possible to find a number of unique fragment ions that are not interfering, even if there is signal interference in the precursor spectra. Acquisition by DIA would also enable identification of novel CQAs in earlier batches since fragmentation data would be available for all precursor ions. In this work a MAM based on DIA was developed and benchmarked against a DDA based MAM with the aim to circumvent above mentioned limitations of DDA based approaches.

3.2.5.4 Identification and characterization of CQAs

To continuously improve analytical methods for therapeutic proteins, it is important that all identified, unknown proteoforms are evaluated for if their presence may affect the therapeutic protein's effect. Many programs developed specifically for data analysis of MAMs contain a new peak detection (NPD) function which identifies unmatched peaks (58). The use of such functions will not be further discussed in this work. However, an unknown proteoform of the protein GM-CSF, recombinantly expressed in *E. coli*, was identified in process development samples of GM-CSF during quality control by intact protein LC-MS analysis. The analysis was performed at the institution before this PhD was initiated and has not been published. In this work chemical assays in combination with MS analysis was applied to determine the chemical structure and source of this attribute to further be able to make a judgement if the modification may be a CQA.

3.2.5.5 Disulfide bond characterization

Since the breakthrough of mass spectrometry, LC-MS and LC-MS/MS has been widely used to map disulfide bonds (22). This is most often done by bottom-up based approaches employed on proteins with non-reduced disulfide bonds. The quantitative power of these methods is, however, often poor, which is probably why disulfide mapping is not commonly introduced into MAMs. One reason for the poor quantitative power is that the enzymatic digestion often results in very complex bundles of peptides connected by disulfide bonds, if more than one cysteine residue is present in the enzymatic peptides. The size and stereochemistry of these bundles make them difficult to fragment, and it is therefore not be possible to identify the disulfide linkages within these bundles unambiguously. Further, very large bundles of peptides will not be covered by the m/z range of common bottom-up MS methods. A second reason why quantitative measurements of disulfide bonds are commonly not performed by peptide mapping is that the amino acid composition as well as the peptide length determines how well the peptide is ionized. It has for example been found that bulky and hydrophobic amino acids as well as basic amino acids ionize better than acidic amino acids. MAMs are generally based on the assumption that the ionization efficiency of the peptide with and without that modification is similar. This will hence not always be true when the studied modification is another peptide, as will be the case for mapping of disulfide bonds. A third reason why disulfide bond mapping is not commonly introduced into MAMs is the lack of software for data analysis. Until recently most of the commonly used database search engines, such as the Mascot or MaxQuant, have not been supporting searches including disulfide linked peptides. However, recently Mascot did introduce support for identification of crosslinked peptides. This has made identification of disulfide-linked peptides much easier and has opened for the possibility of including disulfide linked peptides when searching for other common protein modifications, such as deamidation and oxidation.

In order to overcome the loss of quantitative power of bottom-up approaches on disulfide-linked peptides, attempts have been done to separate and quantify the proteoforms on the intact level, followed by identification by peptide mapping. For example, different proteoforms of an IgG was separated by SDS-PAGE, followed by bottom-up analysis of the individual bands (62). Although the method proved to be successful for quantification of IgG proteoforms with different interchain connections, it is highly dependent on the separation power of the SDS-PAGE. The method further requires long analysis times since each protein sample will result in several proteolytic digests. One strategy to improve the quantitative power of bottom-up based approaches on disulfide-linked peptides could be by selecting enzymes for the proteolytic digest that will only generate peptides with one single cysteine. This would reduce the complexity of the data significantly since each peptide could only be linked to one other peptide and would also improve the fragmentation of the peptides. Although trypsin is the most commonly employed enzyme in bottom-up analysis, there are a number of other commercially available proteolytic enzymes with other amino acid specificities that could be evaluated for use in a MAM protocol covering disulfide scrambling. The difference in ionization efficacy between peptides may still be present. However, if the majority of the disulfide linked peptides could be assigned, quantification based on XIC areas could still give a good indication on the degree of disulfide scrambling in the protein. In this work a MAM including quantitative analysis of disulfide scrambling was developed. The focus during method development was on appropriate enzyme selection to generate peptides of appropriate sizes and disulfide linked peptides with low data complexity to enable quantitative bottom-up protein analysis.

3.2.5.6 Software selection

There are several alternatives for software selection for MAM data analysis. Several of the MS instrument vendors have developed their own software to perform standard MAM quantification. In this thesis the MAMs were performed using mass spectrometers from SCIEX AB, who provides the program BioPharmaView (BPV) for MAM data processing. BPV can, however, not perform quantification from MS/MS data. Therefore, the DIA data acquired for this thesis was quantified in the open-source program Skyline (MacCross Lab) (49). The spectral library was created in Skyline from the search results of DDA data in Mascot (Matrix Science). Skyline is also a useful software for MAM data processing acquired by DDA since peptide identification can be performed based only on the precursor mass and retention time without MS/MS fragmentation data, if the peptide is already present in the spectral library. This provides more consistent results with DDA based MS methods than software which requires MS/MS fragmentation for peptide identification because the fragmentation of low-abundance peptides is often inconsistent. A new version of BPV has been developed, which enables searching for disulfide linked peptides, called BPV Flex. This program was evaluated and compared to Skyline for quantification of disulfide linked peptides and peptides with a free cysteine. Although the programs gave similar quantification results it was found that BPV Flex gave slightly higher values for the shorter peptides than Skyline did. This could be attributed to the fact that BPV Flex only uses the monoisotopic peak for quantification while in Skyline the peaks used for quantification can be manually adjusted, *e.g.,* all isotopes measuring over a certain cutoff intensity. Skyline was therefore found to be better suited for quantification of disulfide bonds since the isotopic distribution may shift between the different disulfide linked and non-linked peptides. The use of solely the monoisotopic peptide for quantification of the modification degree may be better suited for the standard MAM workflow, when the modified and the non-modified peptide have the same amino acid sequence and hence very similar isotopic distributions.

4. Aim of the thesis

Bottom-up proteomics offers the possibility to quantify low-abundance proteoforms of therapeutic proteins in just one method and providing site-specific information on the protein modifications. Such methods are therefore commonly in the biopharma industry called Multi-Attribute Methods (MAMs). Although reproducible and sensitive MAMs have been demonstrated, there are still some aspects that are not covered by conventional MAMs, such as measurements in non-purified samples where signal interference from host cell proteins is an issue and the structural aspects of some proteoforms. The main aim of this thesis was to improve current bottom-up based proteomics approaches for quantification of low abundant proteoforms in therapeutic proteins with regard to the assay performance in non-purified samples and the kind of protein modifications that are covered. This aim was split into three sub-aims.

Commonly, MAM workflows are based on DDA based mass spectrometry methods by which peptide identification is performed on the MS2 level and relative quantification is performed on the MS1 level. These methods may suffer from signal interference when applied on complex samples, such as cell culture samples, and the possibility to backtrack modifications discovered in later process development is limited because of the use of exclusion lists for fragmentation. In the first project of this thesis, it was suggested that these issues could be overcome by basing MAMs on DIA. The quantification of the modification degree could then be based on fragment ions which would give more opportunities to find XICs without signal interference, since the total number of fragment ions generated per peptide is higher than the total number of precursor ions. The possibility to backtrack newly identified modifications would also be ensured since all precursor ions are recorded by DIA methods. The first sub-aim of this thesis was therefore to develop a MAM employing DIA data for quantification. The method should be applied on a cell culture filtrate and on purified samples of a biosimilar of the commercially available monoclonal antibody adalimumab. The performance of the DIA based assay should then be compared to that of a DDA based assay with regard to the two parameters accuracy and linearity, defined in the ICH Topic Q6B (63), in order to answer the scientific question: Can a MAM based on DDA be exchanged for one based on DIA without losing precision and linearity when signal interference in the precursor spectra is an issue?

An important part of MAM assay development is to identify protein modifications that affect the efficacy or safety of the therapeutic protein, what in the biopharma industry is termed critical quality attributes (QCAs), so that these may be introduced into MAMs. A novel modification of +70 Da was identified in process development samples of human GM-CSF recombinantly expressed in *E. coli* (rhGM-CSF) during quality control. The second sub-aim of this thesis was to identify the chemical structure of this novel modification by combining chemical assays with mass spectrometry, to identify its source and determine if it may be a CQA.

Disulfide bond scrambling is a potentially CQA that is not commonly introduced into MAM assays. One of the major reasons for this is that proteolytic digests of proteins with non-reduced disulfide bonds often results in large peptide bundles of several disulfide linked peptides. These bundles are difficult to fragment, and the data analysis is often very complex. The third sub-aim of this thesis was to develop a MAM for rhGM-CSF targeting disulfide bond characterization as well as other proteoforms with focus on proper selection of proteolytic enzymes in order to decrease peptide size and data complexity. The method should then be applied on samples subjected to forced degradation as well as production process development samples in order to demonstrate its applicability to quantify the targeted modification types.

5. Results

MAM assays requires optimization based on the target protein's amino acid sequence as well as the target modifications. One of the main reasons to this is that different proteins will result in different sets of peptides with a given sample preparation protocol. These peptides will in turn have different sizes, ionization efficiencies and fragmentation patterns which will affect how well they are detected by the mass spectrometer. In this thesis two different MAMs were optimized and evaluated for characterization of two different therapeutic proteins in order to fulfill the aims of the thesis.

5.1 Evaluating the potential of using DIA in a MAM for a therapeutic IgG

The first scientific question that we wanted to answer was: Can a MAM based on DDA be exchanged to one based on DIA without losing precision and linearity when signal interference in the precursor spectra is an issue? A scheme was set up describing the different steps that should be preceded in order to answer this question, see Figure 7. The scheme involved eight steps from which five steps described optimization of a MAM assay based on DIA data acquisition, step 6-7 involved validating the assay and in the eighth step the assay was applied on a number of samples. The optimization steps involved were 1) selecting which modification types the method should cover, 2) optimization of the sample preparation protocol, 3) generation of a spectral library containing the amino acid sites and modifications of interest, 4) optimization of the DIA based mass spectrometry method and 5) selection of which MS/MS fragments that should be used for the quantification. Steps 6-7 treated validation of the assay regarding the two assay characteristics precision and linearity. These are two out of nine assay characteristics suggested by the European Medicines Agency to test when validating analytical procedures for testing biotechnical or biological products (63). The DIA based MAM was benchmarked against a MAM based on DDA.

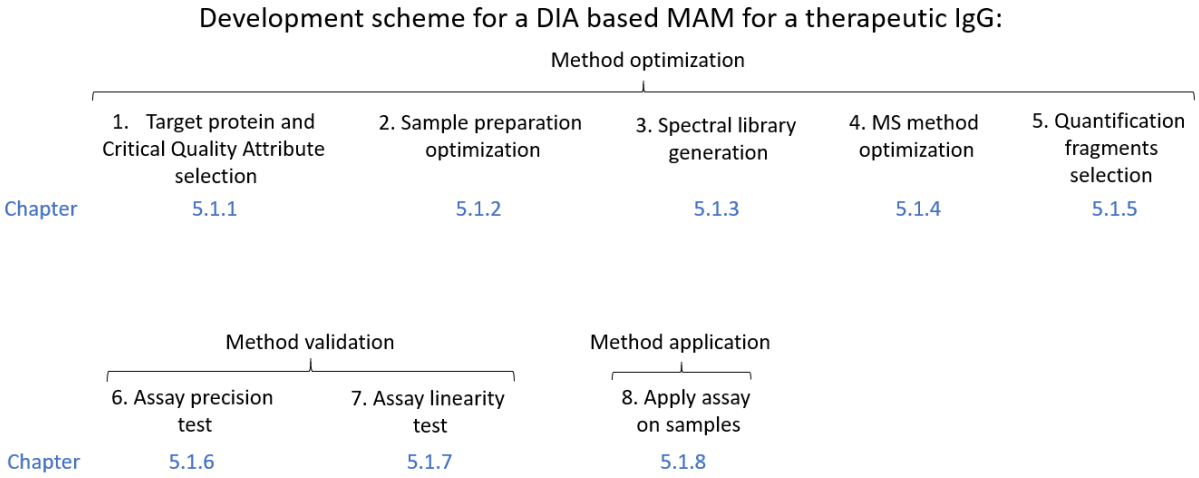


Figure 7: Scheme describing the steps taken to optimize and test a DIA based MAM for a biosimilar of the monoclonal IgG adalimumab.

5.1.1 Target protein and modification type selection for the DIA based MAM

IgGs constitute a large part of the therapeutic proteins on the market today, see chapter 3.1.2. Through a collaboration with a bioprocess lab at the University of Natural Resources and Life Science in Vienna (BOKU) we could get access to a biosimilar of the commercially available monoclonal antibody (IgG) adalimumab. It was decided that the assay should be optimized for this protein, since our collaborators could provide us with samples of this biosimilar from different stages of the production and purification process. This would give us the opportunity to test the DIA based MAM

on complex samples, with possible signal interference from surrounding host cell proteins, and compare its performance to that of a DDA based assay. Asparagine and glutamine deamidation, methionine oxidation and N-glycosylation are three of the most commonly occurring PTMs observed in therapeutic IgGs and have all been related to various deleterious effects on the therapeutic IgGs, see chapters 3.1.4.4, 3.1.4.5 and 3.1.4.2. These three modification types were selected to be targeted by the MAM.

5.1.2 Sample preparation protocol development for an adalimumab biosimilar

The purpose of developing a MAM is to be able monitor the relative amounts of CQAs in the protein so that a decision can be made regarding whether the batch fulfills the requirements specified by the quality target product profile or not, see chapter 3.2.1. Some CQAs, like deamidation and oxidation, may, however, be introduced by the MAM assay itself, see chapter 3.2.5.2. It is therefore important to optimize the assay so that the degree of assay induced modification is kept as low as possible and to know what modification degree that is induced by the assay. It is also important to have a good signal from the peptides containing the target modification sites to be able to detect low abundant proteoforms. This means both having a high signal and a good peak shape that can be used for quantification.

The sample preparation protocol development was based on common MAMs from the literature. These protocols commonly include denaturation and reduction of the protein, to reduce the cysteine linkages, followed by alkylation of the free cysteines (52,64). This is followed by a buffer exchange using a desalting/ buffer exchange column, to remove the denaturing agent as well as the reducing and alkylating agents, which otherwise may inhibit the enzyme used for digestion in the following step. Enzymatic digestion is then performed, most commonly using Trypsin. This is followed by sample cleanup from salts which may otherwise interfere with the subsequent LC-MS analysis. For the developed method we chose to use guanidine hydrochloride (GdnHCl) as the denaturing agent, since IgG is a large molecule that needs a strong denaturing agent in order to open up the whole structure for efficient reduction, alkylation and digestion. This can be compared to another commonly used denaturing agent, urea, which is a weaker denaturing agent and may therefore be more useful for denaturing smaller proteins (65). For reduction of cysteine bridges and the following alkylation of the cysteines, DTT and iodoacetamide were used respectively, which are two common agents used for these purposes. A few options for buffer exchange were evaluated. The strategy that would interact the least with the IgG would be to avoid a buffer exchange but instead dilute the GdnHCl to a lower concentration. This strategy was tested but resulted in a high degree of missed cleavages and low MS signal from the less abundant glycopeptide variants, probably due to enzyme inhibition by denaturation by GdnHCl (data not shown). Buffer exchange on centrifugal concentrators from Sartorius (Vivaspin series) were therefore evaluated but resulted in large sample loss and polymer peaks detected by LC-MS analysis using some batches of the Vivaspin columns (data not shown). Zeba Spin Desalting columns from Thermo Fisher, which uses a size-exclusion chromatography resin for protein desalting, were also evaluated for the purpose. Compared to the Vivaspin columns, the Zeba Spin Columns were found to give a high protein yield with low degree of polymer in the sample. These columns were therefore selected for the buffer exchange. For protein digestion, a combination of LysC and Trypsin was used, in order to limit the number of missed cleavage sites C-terminally to lysine. Digestion in 1 M urea further reduced the number of missed cleavages, compared to digestion without any denaturing agent, and did not introduce peptide carbamylation over 1% (data not shown). For selection of digestion buffer, a sodium phosphate buffer with neutral pH was found to limit the degree of sample preparation induced deamidation. This could be compared to digestion in the MS compatible buffer ammonium bicarbonate, which induced high degrees of deamidation (data not shown).

A number of other sample preparation protocols from the literature were also tested but discarded because of either too many missed cleavages or too low sample yield (54,66–68).

There are several options available to clean the digest from salt ions. One common strategy is to perform a solid phase extraction (SPE). SPE is based on a resin that will only bind the peptides but not the salt ions under certain conditions. The resin is placed in, for example, a pipet tip and after binding the peptides, the salts may be washed away, and the peptides would subsequently be eluted in a solvent which attracts the peptides better than the resin does. Three different SPE protocols were evaluated.

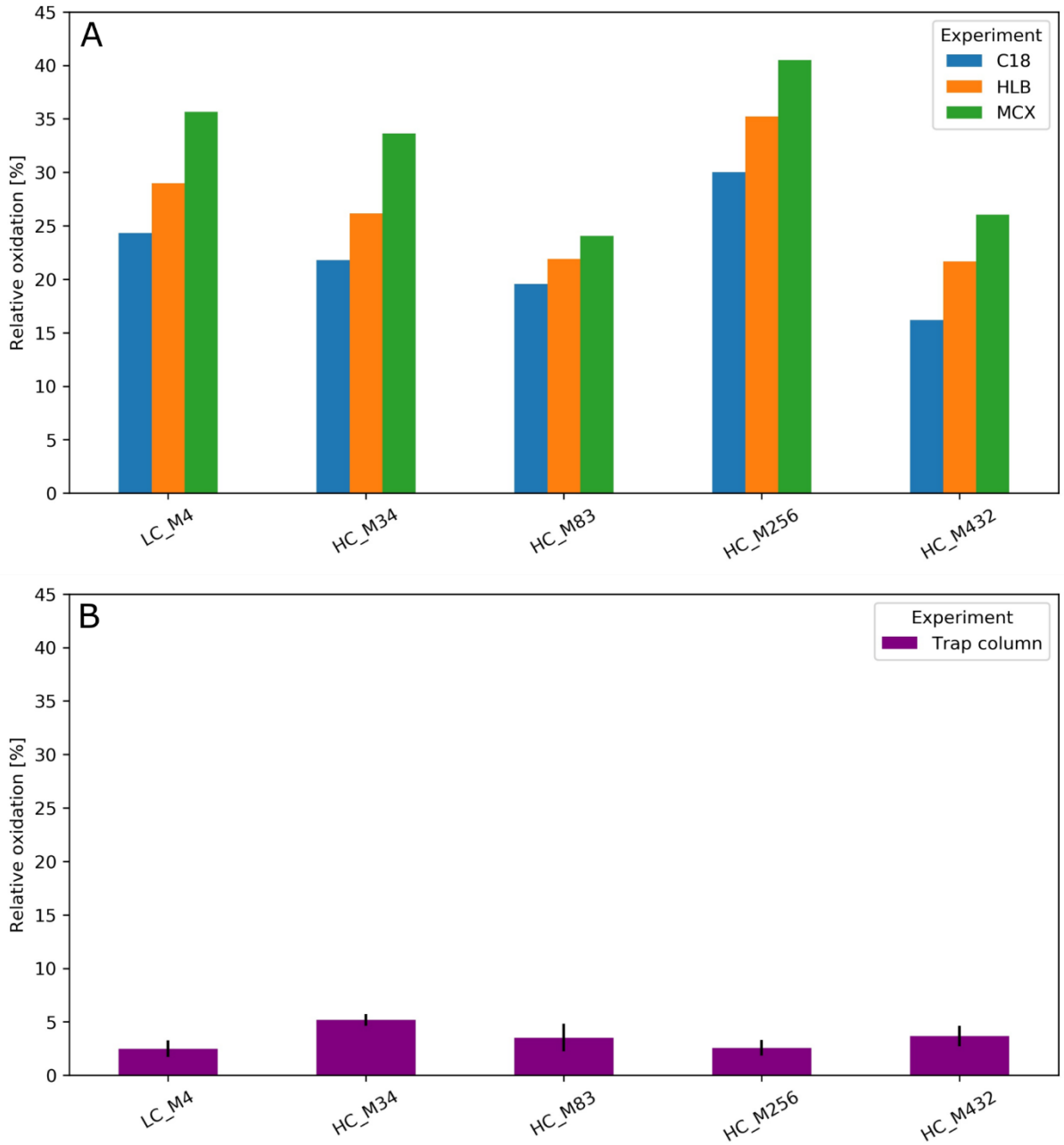


Figure 8: Relative quantification of methionine oxidation in the five peptides covering the methionine sites of the adalimumab biosimilar. A digest of the Protein A purified IgG was cleaned up before RP-LC-MS analysis (a) by three different SPE techniques: C18, HLB and MCX or (b) on-line using a trap column-analytical column setup. In b) the bars represent mean values and error bars the standard deviation, N = 2. MS acquisition was performed by DDA and quantification was based on the precursor XIC areas. See Table 2 for the amino acid sequences of the abbreviated peptides.

Table 2: Peptides used for quantification of methionine oxidation in the adalimumab biosimilar. The "amino acid site" corresponds to the amino acid site for which the modification was quantified. The "amino acid sequence" shows which modified peptide that was used to quantify the modification degree of that specific site together with the same peptide without the modification.

Amino acid site	**Chain**	**Amino acid sequence**
LC_M4	Light chain	DIQ**M[+16]**TQSPSSLSASVGDR
HC_M34	Heavy chain	LSCAASGFTFDDYA**M[+16]**HWVR
HC_M83	Heavy chain	NSLYLQ**M[+16]**NSLR
HC_M256	Heavy chain	DTL**M[+16]**ISR
HC_M432	Heavy chain	WQQGNVFSCSV**M[+16]**HEALHNHYTQK

The two first resins were both reversed phase resins, a C18 resin from OMIX and Waters HLB resin. The peptides were bound under acidic, aqueous conditions and eluted in an organic solvent, 0.1% FA in acetonitrile for C18 and 100% methanol for HLB. The third resin was an MCX resin from Waters, which is a mixed mode resin with both reversed phase and cation exchange properties. The peptides were bound under acidic, aqueous conditions and eluted in 5% ammonium hydroxide in methanol. See chapter 7.2.6 for detailed protocols. All three protocols resulted in high recovery of peptides. However, all three protocols also resulted in high degrees of sample preparation induced methionine oxidation, ranging from 16% to 41% oxidation of the 5 methionine containing peptides, see Figure 8a. A certain degree of sample preparation induced oxidation may be acceptable, which may be subtracted by using a control sample with a known degree of oxidation. However, since it should be possible to monitor oxidation degrees of a few percentages with the assay, it was reasoned that the oxidation degree induced by the SPE protocols was too high to be able to subtract it without inducing a significant quantification error. An alternative strategy to remove contaminants from samples before MS analysis is by using a trap column. This is a smaller column which is connected online with the analytical column with the possibility of switching valves between the columns. The sample can then be loaded onto the trap column in running phase A and be washed from contaminants by a few column volumes while leading the flow to waste, before applying a gradient of running phase B while letting the sample enter the analytical column. Besides offering sample cleanup, the trap column may act as a sample concentrator as well as a guard column, hindering larger particles from entering the analytical column and thereby prolonging its lifetime. The performance of two different trap columns were evaluated together with the currently used analytical column, a nanoEase M/Z Peptide CSH C18 column with a pore size of 130 Å and a bead size of 1.7 µm from Waters. The first evaluated trap column contained the same stationary phase from Waters as the analytical column but with a larger bead size (5µm). The second trap column contained a C18 stationary phase from YMC. Both columns gave very poor chromatography when connected with the analytical column (data not shown). Therefore, a trap-analytical column combination which was recommended to us by AB Sciex was evaluated instead, consisting of a ChromXP C18CL column with a pore size of 120 Å from Eksigent in both the trap and the analytical column but with a larger bead size in the trap (5 µm) than in the analytical column (3 µm). See chapter 7.2.7 for the detailed protocol. This setup gave a much better chromatography than the CSH column did in combination with either of the tested trap columns (data

not shown). The peak shape and separation of some peptides was not as good as when the CSH column was used alone. However, the chromatography of the majority of the peptides, including the peptides of highest interest for the assay, were considered being good. The main parameter to consider when selecting what resin to use for the trap column is the resin's retention capacity. If the system has negligible dead volume between the trap and the analytical column, the phases of the two columns can be the same (69). Otherwise, the retention capacity of the trap column should be lower than that of the analytical column, so that the peptides can be refocused on the analytical column and retain a high peak capacity. A high dead volume may be the reason for the poor chromatography when the CSH C18 trap-analytical column setup was used. No comparison of the relative column retentivities between the CSH C18 resin and the C18 resin from YMC has been done to our knowledge. It has, however, been shown that it may be tricky to find a good trap column match for the CSH C18 analytical column because of its low retentivity (70). Sample analysis using the trap-analytical column setup from Eksigent led to much lower levels of sample preparation induced oxidation than any of the SPE methods, between 2.49% ± 0.76 and 5.17% ± 0.54, see Figure 8b. This trap-analytical column setup was therefore selected for the sample cleanup step in the MAM assay.

Most MAM workflows in literature are optimized for purified protein samples. Since one of the main purposes of this protocol was to be able to monitor protein modifications in samples taken directly form the cell culture, an additional, initial, protein cleanup step was needed, prior to protein reduction and alkylation. The usage of spin columns for buffer exchange, like the ZebaSpin desalting columns which were used for the buffer exchange prior to digestion, has a few disadvantages. First, a significant degree of sample loss could be seen when using these columns, especially from samples subjected to forced degradation. Implementing two buffer exchange steps in the sample preparation protocol would lead to an even larger sample loss and subsequently lower signals from the low abundant peptide species, which are often the peptides of highest interest for the MAM. Secondly, ZebaSpin columns are quite expensive and introducing a second column desalting step would increase the price of the assay. A common method for cleaning proteins from contaminants is by precipitating the protein in an organic solvent. Addition of a high concentration of an organic solvent, like acetone, leads to protein aggregation by electrostatic and dipole forces. After precipitation in -20 ˚C overnight, the proteins can be dissolved in a denaturing agent followed by reduction and alkylation. An initial protein precipitation step was introduced into the sample preparation protocol after having confirmed that it introduced low levels of oxidation and deamidation, see Figure 8b. The final sample preparation protocol is described in chapter 7.2.3.

The protocol was tested on the adalimumab biosimilar purified on Protein A and on the biosimilar filtrated directly from the cell culture and was analyzed by LC-MS/MS acquired by DDA. The sequence coverage of the Protein A purified IgG from the analysis was 100% for the LC and 82% for the HC, see Figure 9a.

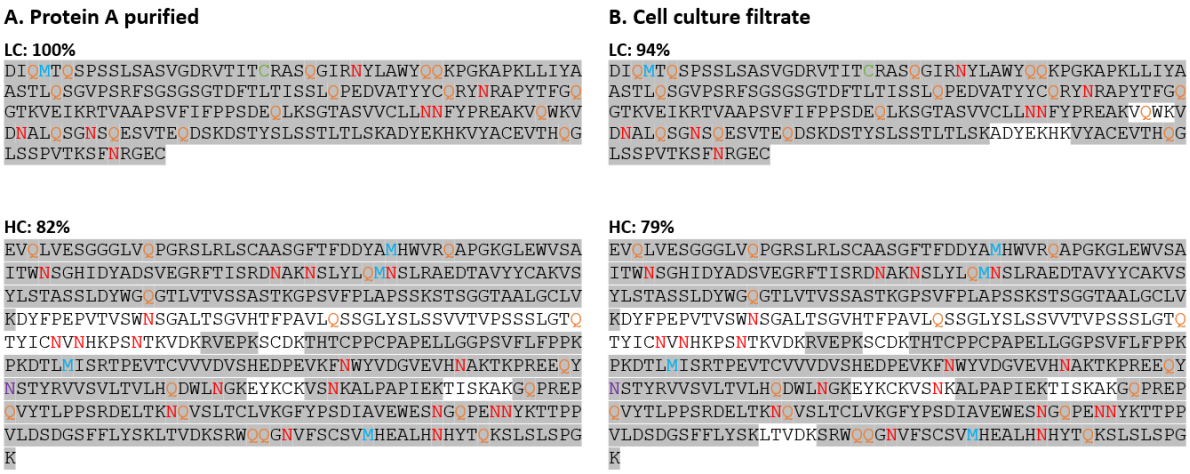


Figure 9: Sequence coverage of the adalimumab biosimilar (a) purified on Protein A or (b) filtrated from the cell culture. The samples were prepared and digested according to the optimized protocol and analyzed by RP LC-MS coupled online to a trap column. The sample data was acquired by DDA. The data was searched in the Mascot database search engine and was imported to Skyline using a library built from the Mascot search results. Marked grey = the sequence coverage reported by Skyline. The letters for the asparagine (red), methionine (blue) and glutamine (orange) residues are marked with individual colors and the glycosylated asparagine have its own color (purple).

The sections lost by the assay were one very long peptide, amino acid sites 152-209, and a few very short peptides. The long peptide was probably not detected because its most abundant charge state has an m/z value over 1700, which is the highest m/z value that the assay covers. The short peptides were probably not detected because their most abundant charge states have m/z values lower than 300, the lowest m/z value that the assay covers. The sequence coverage of the IgG in the cell culture filtrate was slightly lower than that of the purified protein, 94% for the LC and 79% for the HC, Figure 9b. The background to this may be signal interference in the mass spectrometer, leading to suppression of some peptide signals. It may also be that the protein is more completely digested in the cell culture filtrate compared to the purified sample, since the sections of the protein sequence that are lost come from peptides shorter than 7 amino acids. This may be explained by difficulties in measuring the protein concentration with good precision in the cell culture filtrate, resulting in slightly different enzyme:protein ratios in the two samples. The lost sequence sections led to a loss of four asparagine sites and two glutamine sites in the purified sample and five asparagine sites and three glutamine sites in the cell culture filtrate. This sequence coverage was considered acceptable, and the scheme of this final sample preparation protocol can be seen in Figure 10.

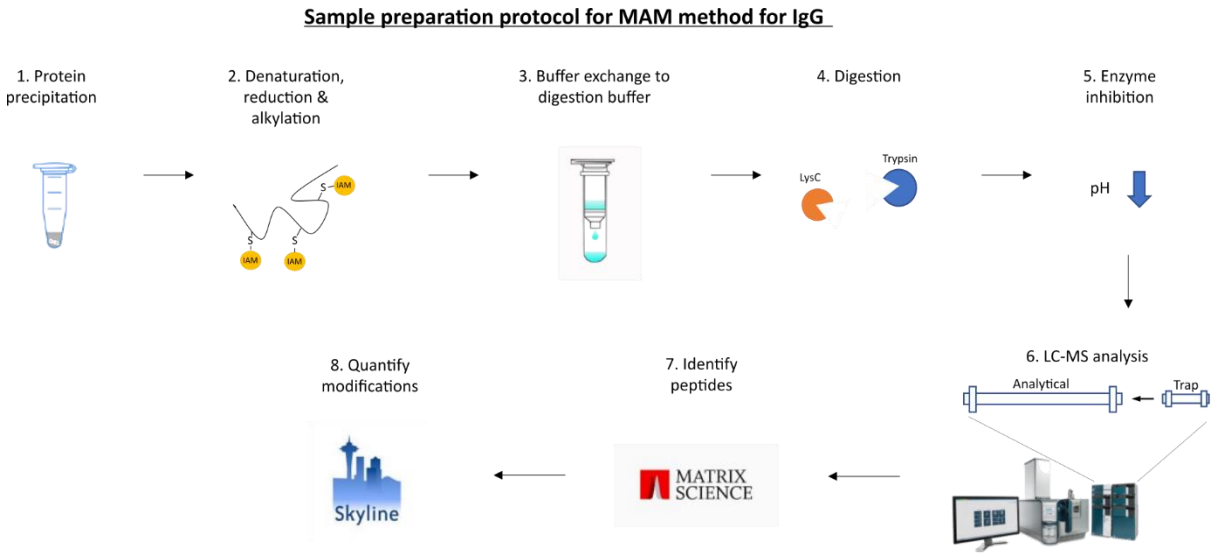


Figure 10: Scheme of the workflow for the MAM optimized for the adalimumab biosimilar. 1) Protein buffer exchange by acetone precipitation, 2) denaturation and cysteine reduction and alkylation with iodoacetamide, 3) buffer exchange on Zeba Spin Desalting columns to digestion buffer, 4) protein digestion with Trypsin and LysC and 5) enzyme inhibition by addition of 1% TFA. 6) The digested sample was then cleaned online on a trap column followed by analysis by LC-MS/MS. 7) The peptides were identified using the Mascot database search engine and 8) a library was built in Skyline from the Mascot search results. The sample data was then searched using this library and the relative peptide modification degree was quantified from XIC areas in Skyline.

5.1.3 Spectral library generation

To search DIA data a spectral library is required, see chapter 3.2.4.4. The spectral library was generated in two different ways, one for the deamidated and oxidized peptides and one for glycosylated peptides.

5.1.3.1 Spectral library generation for deamidation and oxidation prone amino acid sites

The spectral library containing peptides with deamidation and oxidation prone amino acid sites was created using spectral data from digests of the adalimumab biosimilar acquired by DDA. The DDA data was searched in the Mascot database search engine for peptide identifications and the search results were used to create a spectral library in Skyline. Peptide species which are too low in intensity or are co-eluting with many other, more abundant, peptides may not be selected for fragmentation by the DDA method and will thereby not be included in the library. This may typically be the case for very complex samples. Therefore, in order to make sure that spectral data from all deamidation prone sites was included in the library, the protein A purified adalimumab biosimilar was subjected to forced degradation by incubation at pH 9 at 37 ˚C for 1 and 2 weeks. See chapter 7.2.4 for a detailed protocol. The DDA spectral data from the resulting deamidated peptides was then included in the library. Out the 20 asparagine sites covered by the assay, 15 sites were identified in its deamidated form in the 2 weeks stressed sample. From the 29 glutamine sites 2 sites were identified as deamidated in the same sample. This included all asparagine and glutamine sites in peptides that were found deamidated, even if the peptide was only identified with deamidation at one site but contained several deamidation sites. The reason for this was that the peaks from the same peptide with deamidation at two different sites may not be chromatographically resolved. Some deamidated peptides did elute as two peaks, probably corresponding to isoaspartate and aspartate. The presence of and aspartate or isoaspartate residue could, however, not be distinguished by the methods used. The asparagine site containing the N-glycan was not included in the library of deamidated peptides. The deamidation sites were contained within 19 peptides and are presented in Table 3. These peptides were included in the

spectral library. The successive increase of relative deamidation percent for those amino acids that were affected by the stressing can be seen in Figure 11, data acquired by DDA.

Table 3: Peptides used for quantification of asparagine and glutamine deamidation in the adalimumab biosimilar. The "amino acid site" corresponds to the amino acid site for which the modification was quantified. The "amino acid sequence" shows which modified peptide that was used to quantify the modification degree of that specific site together with the same peptide without the modification. Some peptides were identified at two different retention times corresponding to the two isoforms isoaspartate and aspartate.

Amino acid site	**Chain**	**Amino acid sequence**	**Isoform**
LC_N92_iso1	Light chain	YN[+1]RAPYTFGQGTK	1
LC_N92_iso2	Light chain	YN[+1]RAPYTFGQGTK	2
LC_N137	Light chain	SGTASVVCLL**N[+1]**NFYPR	-
LC_N138	Light chain	SGTASVVCLLN**N[+1]**FYPR	-
LC_N210	Light chain	**SFN[+1]**RGEC	-
HC_N74_iso1	Heavy chain	FTISRD**N[+1]**AK	1
HC_N74_iso2	Heavy chain	FTISRD**N[+1]**AK	2
HC_N77	Heavy chain	**N[+1]**SLYLQMNSLR	-
HC_Q82	Heavy chain	NSLYL**Q[+1]**MNSLR	-
HC_N84_iso1	Heavy chain	NSLYLQM**N[+1]**SLR	1
HC_N84_iso2	Heavy chain	NSLYLQM**N[+1]**SLR	2
HC_N290_iso1	Heavy chain	FNWYVDGVEVH**N[+1]**AK	1
HC_N290_iso2	Heavy chain	FNWYVDGVEVH**N[+1]**AK	2
HC_N319_iso1	Heavy chain	VVSVLTVLHQDWL**N[+1]**GK	1
HC_N319_iso2	Heavy chain	VVSVLTVLHQDWL**N[+1]**GK	2
HC_N365	Heavy chain	**N[+1]**QVSLTCLVK	-
HC_Q366	Heavy chain	N**Q[+1]**VSLTCLVK	-
HC_N393	Heavy chain	GFYPSDIAVEWESNGQPE**N[+1]**NYK	-
HC_N425	Heavy chain	SRWQQG**N[+1]**VFSCSVMHEALHNHYTQK	

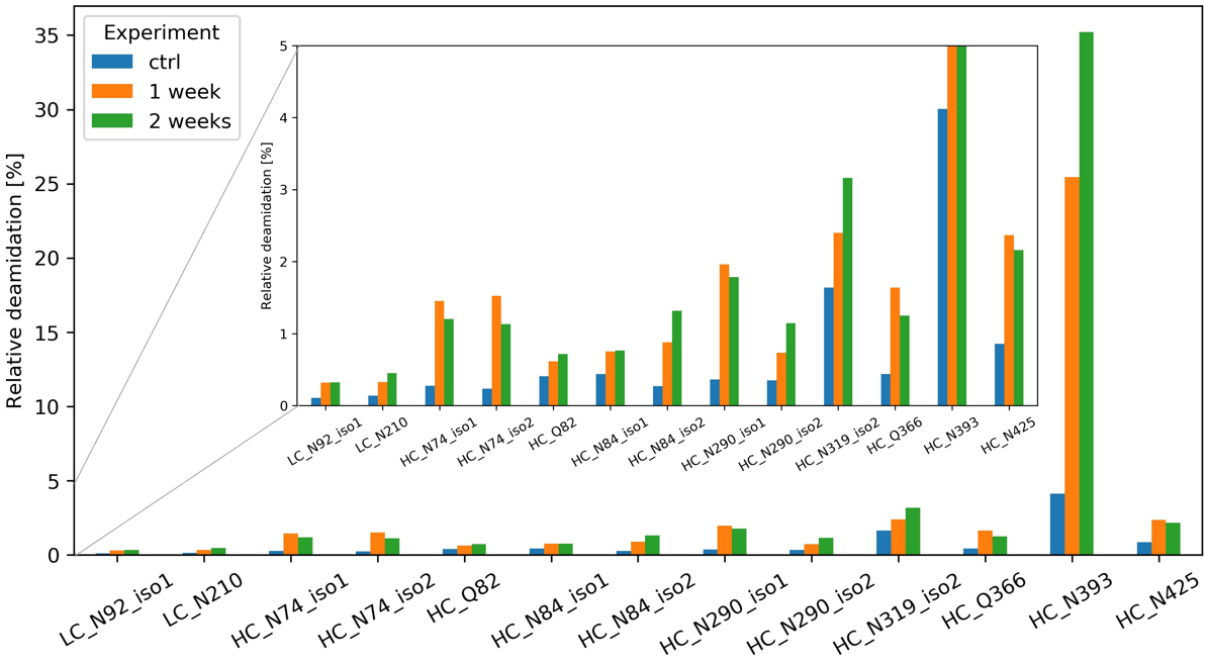


Figure 11: Quantification of deamidation in Protein A purified adalimumab biosimilar stressed at pH 9 at 37 ˚C for 0, 1 or 2 weeks. The stressed sample was prepared and digested according to the optimized protocol and analyzed by RP LC-MS. The identified deamidated peptides were then included in a spectral library to use for identification of deamidated species. The data shown is based on DDA. Relative quantification of the modification degree was performed using the precursor ions for each peptide. Only peptides in which the measured deamidation increased upon stressing was included in this graph. See Table 3 for the amino acid sequences of the abbreviated peptides.

13 different peptides were found to be deamidation prone. For some of the sites the deamidation increased successively over the two weeks while for some sites the deamidation increased from the control sample to the 1-week stressed sample but not further when the sample was stressed for 2 weeks. The site found to be most affected by the stressing was site HC_N393 in peptide *GFYPSDIAVEWESNGQPENNYK*, with a deamidation degree of 35% in the 2 weeks stressed sample. It can however not be ruled out that this deamidated peak also contained variants that were deamidated at other sites in the peptide and which were not retention time separated, since the peptide contained three asparagine residues and one glutamine residue. This is not unlikely since this peptide contains the deamidation hotspot N388 with the deamidation prone *NG* motif. The other peptides found to be deamidation prone under the stressing conditions were not found with deamidation degrees higher than 4% in the 2 weeks stressed sample. Two sites were found with deamidation degrees over 2%: site HC_N319_iso2 with 3.2% and site HC_N425 with 2.2% deamidation, "_iso2" meaning the second out of two isoforms. The other 10 peptides contained deamidation degrees between 0.3% and 1.8%.

In order to include spectral data from peptides containing the five oxidized methionine residues, the Protein A purified IgG was incubated with 0.1% H_2O_2 at denaturing conditions for 1 h, see chapter 7.2.5 for a detailed protocol. This resulted in 100% oxidation of the peptides covering all five methionine sites. These were included in the spectral library.

5.1.3.2 Peptide library generation for N-glycan species

Adalimumab is an IgG1 and has two glycosylation sites, one N-glycan located in the hinge region on each of the two heavy chains, see chapters 3.1.2 and 3.1.4.2. Identification of the protein's glycovariants directly by peptide mapping using reversed phase LC-MS/MS analysis is challenging. The reason to this is first because the glycan composition does not affect the peptide hydrophobicity much

compared to the amino acid composition. The different glycoforms of the same glycopeptide are therefore not separated very well on the RP column. Secondly, fragmentation of glycopeptides is more complex than that of non-glycosylated peptides, since both the peptide chain and the glycan chain may fragment. The carbohydrate bonds are less stable than peptide bonds and you may end up with peptides where only the glycan is fragmented and not the peptide backbone. Thirdly, there is a very limited selection of software available for identification of glycopeptides. Therefore, the glycan profile of the adalimumab biosimilar was identified by released glycan analysis. The Protein A purified IgG was subjected to PNGase F, an enzyme that is specific to digestion between the innermost GlcNAc of the N-glycan and the asparagine residue. The released glycans were labelled with the fluorophore Rapifluor (Waters) and analyzed by HILIC coupled to a fluorescence detector or to ESI-MS/MS. See chapter 7.2.9 for a detailed protocol. The identified glycan profile recorded by fluorescence detection can be seen in Figure 12 with the structure identified from MS data assigned to the peaks. A list of the names of all the glycan structures is shown in Table 4.

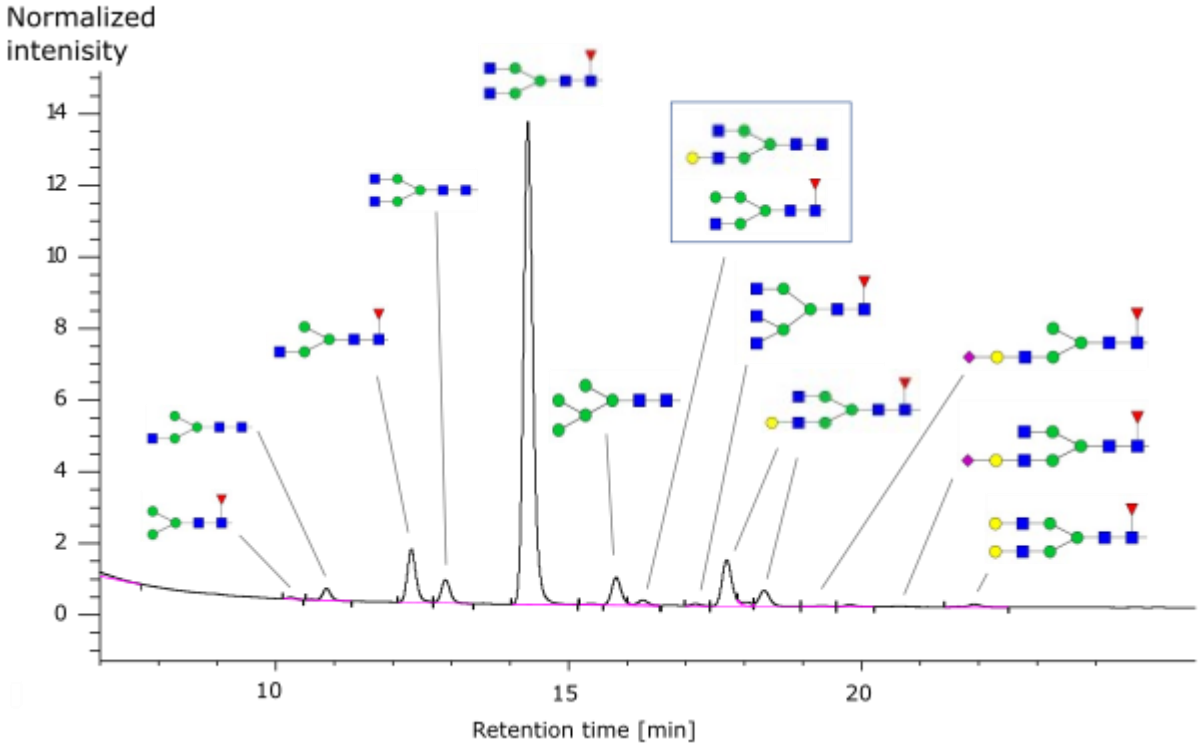
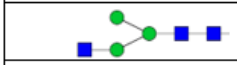
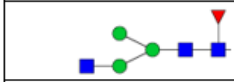
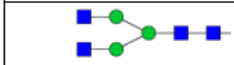
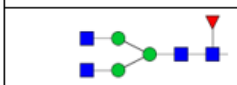
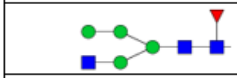
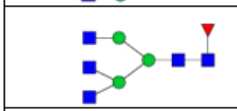
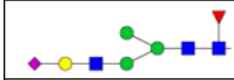
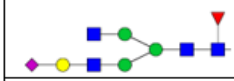


Figure 12: Released N-glycans from the protein A purified adalimumab biosimilar analyzed by HILIC coupled to fluorescence detection or ESI-QTOF-MS. The glycans were released by PNGaseF digestion and labelled with the fluorophore RapiFluor prior to LC-MS analysis. Green circle = mannose, blue box = N-acetylglucosamine, yellow circle = galactose, red triangle = fucose and purple diamond = N-acetylneuraminic acid. Common names of the glycan structures are shown in Table 4.

The G0F species was clearly identified as the most abundant N-glycan in the sample. This was expected since this is often one of the most abundant N-glycan species in monoclonal IgGs. The G1F species was also present with two clear peaks, representing the two isomers, together with the G0, G0F-N and G0-N species. The high mannose species M5 was also identified. High mannose levels have been linked to protein clearance and is therefore particularly important to be able to quantify in therapeutic IgGs. Some other low-abundance glycovariants were also identified. All identified N-glycan species were expected from eukaryotic protein expression.

Table 4: N-glycans released from the adalimumab biosimilar and identified by HILIC-HPLC coupled to ESI-MS. The naming convention of the "Common name" is based on the simplest bisecting glycan structure G0 to which monosaccharides are added or subtracted according to the format X(n). G: galactose, F: fucose, M: mannose, N: N-acetyl-glucosamine, S: N-acetylneuraminic acid (sialic acid). The high mannose structures Man3F and Man5 are exceptions and are based on the number of mannose residues added to the two core N-acetylglucosamine residues. The "Composition based name" reports the number of each monosaccharide that the glycan contains. The glycan structures were drawn using GlycoWorkbench (www.glycoworkbench.org). See Table 1 for a translation of the glycan building blocks.

Glycan structure	Common name	Composition based name
	Man3F	H3N2F1
	G0-N	H3N3
	G0F-N	H3N3F1
	G0	H3N4
	G0F	H3N4F1
	Man5	H5N2
	G1	H4N4
		H4N3F1
		H3N5F1
	G1F	H4N4F1
	G1S-N	H4N3F1S1
	G1FS	H4N4F1S1
	G2F	H5N4F1

A peptide library, containing possible glycopeptides with the G0F glycan, was built manually in Skyline and was used to search the DDA data of the Protein A purified IgG analyzed with the developed MAM assay for the precursor ions from these glycopeptides. The peptide *TKPREEQYNSTYR*, containing two missed cleavages, was found to be the most abundant glycopeptide in the sample. A peptide library was therefore built based on this peptide including all glycan species identified by the released glycan assay. This was not a true spectral library, since it did not contain any MS/MS spectral information from the glycopeptides.

5.1.4 Mass spectrometry method optimization

The carbohydrate bonds are generally less stable than the peptide bonds and may therefore break when subjected to high voltage or temperatures. If some glycans would fragment already in the ion source this could lead to false measured degrees of the different glycovariants. Therefore, to make sure that no glycans were fragmented in the ion source, oxonium ions from the glycans were searched

for in the MS1 spectra from the digested IgG analyzed by LC-MS/MS. See Table 5 for a list of the oxonium ions included in the search.

Table 5: Common glycopeptide oxonium ions which were searched for in the MS1 data to detect in-source fragmentation *(71)*.

Glycan	**m/z**
HexNAc-2H_2O-CH_2O	138
Hex	163
HexNAc	204
NeuAc-H_2O	274
NeuAc	292
HexNAc+Hex	366
Hex+HexNAc+dHex	512
NeuAc+Hex+HexNAc	657

A DDA method recording MS1 signal in the range 130-1700 m/z was used to test the effect of different settings of collision energy (CE), declustering potential (DP) and source temperature (T) on in-source glycan fragmentation. Figure 13a shows that a strong signal from the HexNAc oxonium ion (pink) and a slightly weaker signal from the HexNAc + Hex oxonium ion (red) was observed when a CE of 10 V, a DP of 80 V and a source temperature of 100 ˚C was used for the MS1 ionization. These signals aligned well with that from the glycopeptide with a G0F glycan attached (blue). When the CE was decreased to 6 V and the DP to 10 V the signals from the glycan oxonium ions were no longer visible. Decreasing the source temperature from 100 ˚C to 50 ˚C and 30 ˚C when using the low voltage settings did not change the signal from the oxonium ions any further. Figure 13b shows the signal from the HexNAc oxonium ions and Figure 13c the signal from the HexNAc + Hex when using the four different methods with 1) CE = 10 V, DP = 80 V, T = 100 ˚C (green), 2) CE = 6 V, DP = 10 V, T = 100 ˚C (red), 3) CE = 6 V, DP = 10 V, T = 50 ˚C (pink) and 4) CE = 6 V, DP = 10 V, T = 30 ˚C (blue). Characteristic oxonium ions from sialic acid were also searched for but were not found. The reason may be that the glycovariants containing sialic acid residues were present in very low amounts in the Humira biosimilar. Therefore, the signal intensities from the sialic acid containing glycopeptides were instead compared between the four different MS methods. No clear difference in signal intensity was found. The method using CE 6 V, DP 10 V and source temperature 100 ˚C was therefore selected for the MAM MS1 method. See chapter 7.2.7 for the final method.

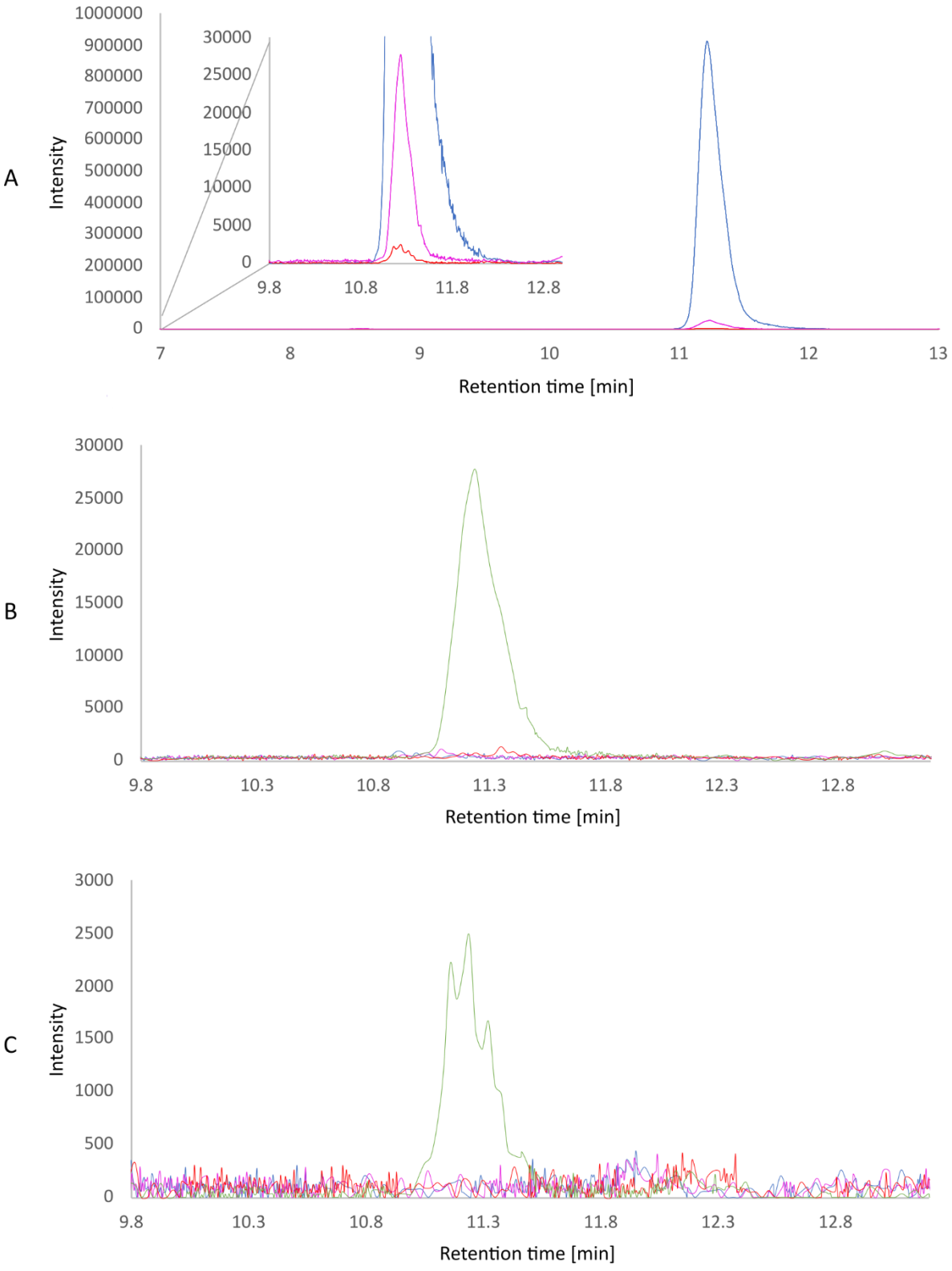


Figure 13: A digest of the adalimumab biosimilar purified on Protein A was analyzed by RP LC-MS in DDA mode covering precursor ions in the range 130-1700 m/z to identify in-source fragmentation of glycopeptides. (a) Data was acquired using MS parameters CE: 10 V, DP: 80 V, T:100 ˚C and the XICs from the 4+ charged precursor ion of the glycopeptide TKPREEQYNSTYR with the glycan G0F (blue), the HexNAc oxonium ion (pink) and the HexNAc + Hex oxonium ion (red) was overlaid. Data was then acquired using four different MS methods with parameters: CE: 10 V, DP: 80 V, T: 100 ˚C (green), CE: 6 V, DP: 10 V, T: 100 ˚C (red), CE: 6 V, DP: 10 V, T: 50 ˚C (pink), CE: 6 V, DP: 10 V, T: 30 ˚C (blue) and the XICs from (b) the HexNAc oxonium ion and (c) the HexNAc + Hex oxonium ion were plotted. CE= collision energy, DP = declustering potential, T = source temperature.

The MS2 signal from the DIA based MS method was optimized by allowing a certain time for the entire MS2 acquisition and then experimenting with different number of SWATH windows. First, a total cycle time was decided upon which gave approximately 10 points over the peaks. This was considered appropriate for quantification. The cycle time 1.7 s was found to fulfil this requirement for most peaks. The cycle time minus the MS1 accumulation time was then the time available for accumulation of MS2 data. The required MS1 range was selected based on the spectral library ions and was found to be 300-1700 m/z. Windows with variable sizes within this range were calculated using a variable window calculator provided by AB Sciex. The calculator was fed with the MS1 data from a digest of a cell culture filtrate of the adalimumab biosimilar. Windows with variable sizes were calculated based on this data so that the number of spectra was evenly distributed between the windows. This was done to reduce signal interference in the complex sample. Four DIA methods with 53, 41, 37 and 34 SWATH windows respectively were compared for analysis of the adalimumab biosimilar cell culture filtrate. For some peptides it was clear that decreasing the number of windows led to an improved fragment signal, since this allowed for an increased signal accumulation time, see Figure 14.

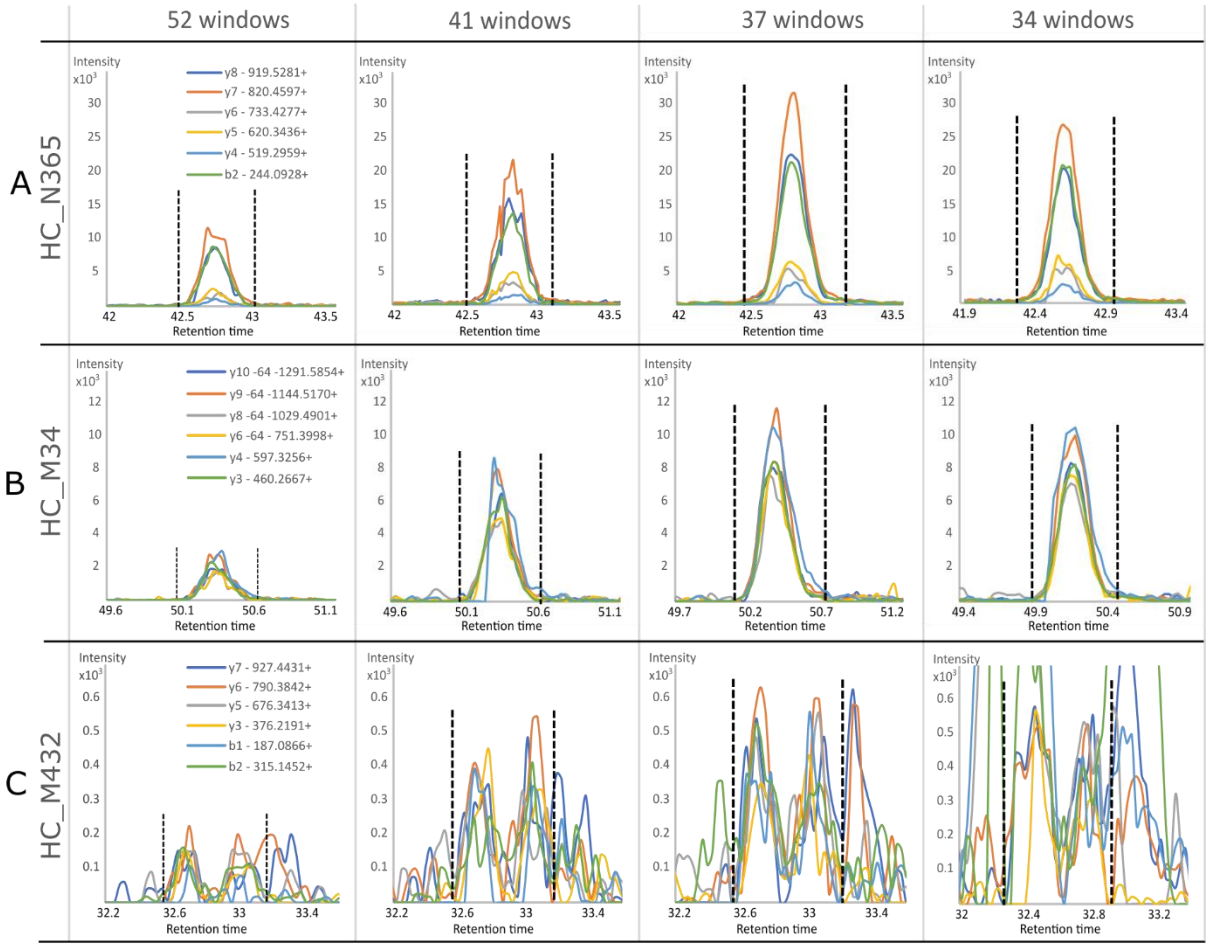


Figure 14: The number of SWATH windows were varied together with the window accumulation time, while keeping the total window accumulation time constant. Four DIA methods with 53, 41, 37 and 34 SWATH windows respectively were tested, shown as columns. The XICs for the 6 fragment ions with the largest area were extracted for the precursor ions of (a) the deamidated peptide HC_N365, charge state +4, (b) the oxidized peptide HC_M34, charge state +3 and (c) the oxidized peptide HC_M432, charge state +4. The last peptide showed peak splitting. The methods were tested on a digest of the adalimumab biosimilar cell culture filtrate. Dashed lines are showing peak integration bars. See Table 2 and 3 for the amino acid sequences of the abbreviated peptides.

This was for example the case for the peptide with deamidation at site HC_N365 and the two peptides with oxidation at sites HC_M34 and HC_M432. The peptide containing site HC_M432 showed peak splitting, a phenomenon that can appear because of formation of diastereomers upon oxidation. Decreasing the number of windows from 53 to 41 windows improved the fragment signals a lot for these three peptides and so did also going from 41 to 37 windows. However, when the number of windows was decreased further to 34 windows, the signal did not improve more for these fragments. Moreover, signals from coeluting peptides started to interfere with those from the peptides of interest for the site HC_M432. It was therefore decided that the method with 37 SWATH windows should be used for the MAM assay. See chapter 7.2.7.2 for the final method.

The purpose of setting up a MAM based on DIA was to make it easier to avoid signal interference from peptides from other proteins in complex samples. While optimizing the DIA based method it was observed that the precursor ions from the peptide containing site LC_N138 seemed to elute as two peaks with very similar retention times, see Figure 15a.

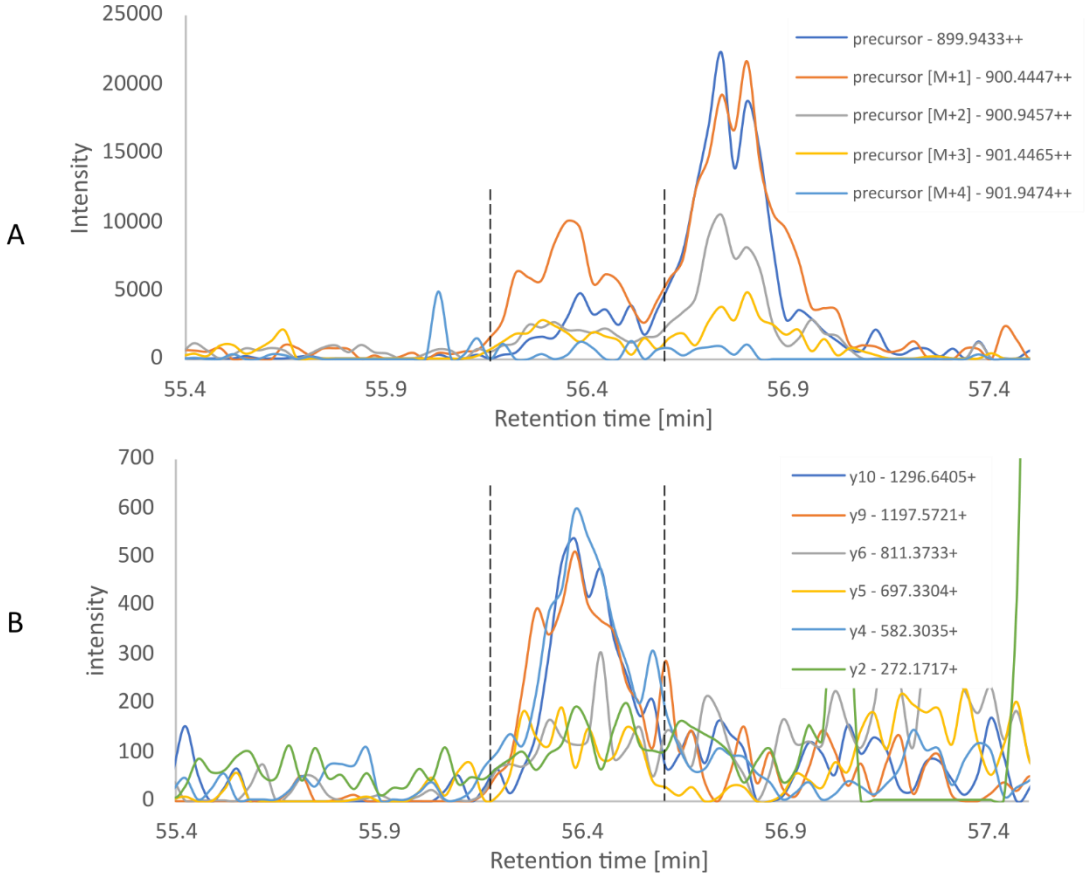


Figure 15: The peptide with deamidation at site LC_N138 showed (a) two peaks with poor peak separation in the XICs of the 2+ charged precursor while (b) the XICs from fragment ions from the same precursor could confirm that the first peak was the correct assignment. The data was recorded from a digest of the adalimumab biosimilar cell culture filtrate analyzed by RP LC-MS acquired by DIA.

When the fragment ions from the peptide were searched for in the DIA data, only the smaller of the two peaks was found to belong to the peptide of interest, see Figure 15b. This indicates that the assignment of peptide peaks may be improved by basing the quantification on XICs from fragment ions instead of precursor ions.

5.1.5 Selection of fragment ions for quantification

It was decided that the most stable fragment ions should be used for the quantification using DIA data, since assay accuracy is highly important and one of the characteristics that the assay would be tested for. The 20 fragments with the highest XIC areas were therefore extracted for all deamidated peptides identified in a digest of the 2 weeks stressed sample. The variation in the relative areas of these fragments were then evaluated between 4 technical replicates of the same digest acquired on separate days. It was found that a higher relative XIC area was generally related to a lower coefficient of variation (CV) between the areas from the different runs. Therefore, the 5 fragments with the highest intensities for each precursor ion were selected to use for quantification, excluding fragments with obvious signal interference in the 2 weeks stressed sample or in the cell culture filtrate. One exception to this was the peptide *GFYPSDIAVEWESNGQPENNYK*, which contains the deamidation prone asparagine residues HC_N388 and HC_N393. Only one deamidated peak of this peptide could be identified from the MS1 data, see Figure 16a. However, the fragments y4, y6 and y8 from the DIA data could be used to distinguish between the peptide with deamidation at site HC_N388 (the SNG site) and site HC_N393 (the ENN site). This was possible since these fragments differ by 1 Da, i.e., one deamidation, between the two peptides and these fragments were found to be slightly retention time separated on the RP column, see Figure 16b and c. One drawback of using this method is that the degree of deamidation at site HC_N393 may be overestimated since the peptide with deamidation at this site has almost the same m/z value as the first isotope of the peptide with deamidation at site HC_N388. This together with the fact that the two fragment peaks are not completely separated may lead to overestimation of the deamidation degree at site HC_N393. The fragments y4, y6 and y8 were however the most abundant fragments from the two peptides and changing to using lower abundant fragments that have the same m/z value for both peptides could decrease the signal sensitivity. The fragments y4, y6 and y8 was therefore selected for quantification of deamidation at sites HC_N388 and HC_N393. See chapter 7.2.8.1 for a detailed description of the fragment selection procedure.

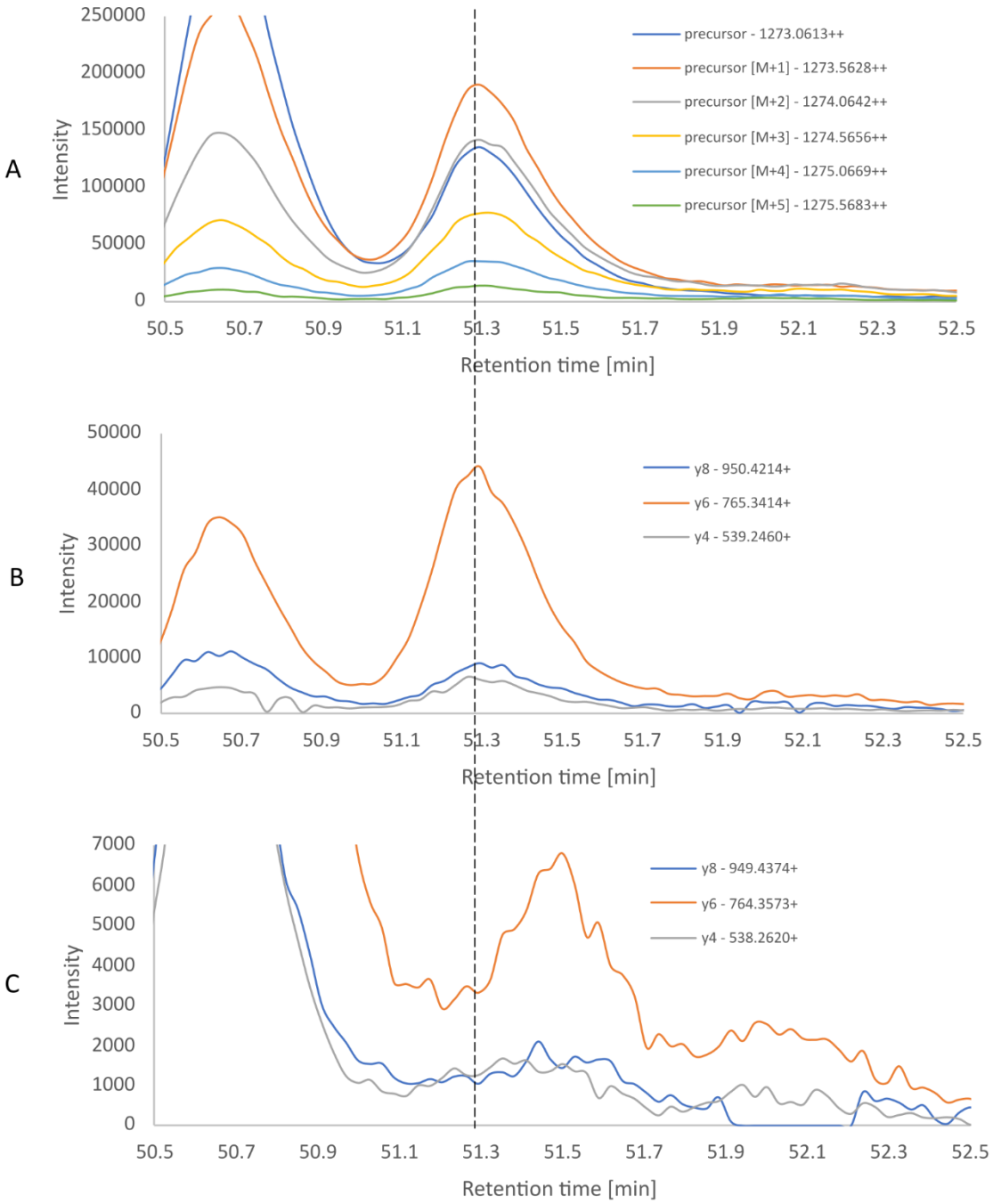


Figure 16: XICs of (a) the +2 charged precursor ion of peptide GFYPSDIAVEWESNGQPENNYK with one deamidation and the corresponding fragment ions y4, y6 and y8 of the deamidated peptides (b) GFYPSDIAVEWESNGQPEN[+1]NYK (deamidation site HC_N393) and (c) GFYPSDIAVEWESN[+1]GQPENNYK (deamidation site HC_N388). Peak separation between the two peptides could be observed on the fragment level but not on the precursor level. The peptides were identified in a digest of protein A purified adalimumab biosimilar stressed for 2 weeks at pH 9 and 37 ˚C. Data was acquired by DIA.

The ambition with the DIA method was to quantify the modification degree of all three types of modification: deamidation, oxidation and N-glycan variants, using MS2 data. However, after careful consideration it was decided that quantification of the glycovariants should be performed on the MS1 level instead. The reason for this was that the different glycovariants of the same peptide are not well resolved on a RP column. These different glycopeptides do also give very similar fragments when subjected to MS/MS fragmentation and it would therefore be hard to distinguish between them if they were fragmented in the same SWATH window. A high number of SWATH windows would then

be required in the MS method since the masses between some of the co-eluting glycopeptides are very small. As an example, the precursors ions of the glycopeptides containing G1F and G0F would have been fragmented in the same window, even if as many as 70 windows had been used, calculating window sizes using the variable window calculator described in chapter 5.1.4. This would have led to very short MS2 data accumulation times compromising the assay sensitivity. Since no signal interference was identified in the MS1 signals from the glycopeptides in the cell culture sample, it was decided that such compromises were not motivated. An MS1 accumulation time of 150 ms was found to give good glycopeptide signal enabling glycopeptide quantification down to at least 1%. See chapter 7.2.10 for a detailed description of the quantification strategy for the glycosylated peptides.

5.1.6 Assay precision using the DIA based MAM

In order to measure the precision of the assay, two samples of the adalimumab biosimilar were used. The first sample was the cell culture filtrate from expression of the IgG, which was selected to examine the precisions of the DIA and DDA based MAMs in a complex mixture. The second sample was the IgG purified on Protein A and stressed for 2 weeks at pH 9 and 37 °C, which was selected in order to cover most of the potential deamidation sites. The two IgG samples are described in chapters 7.2.1 and 7.2.4 respectively. The assay precision was monitored on four different levels: 1) precision between the DIA and DDA method, 2) precision when the same digest was acquired in replicates the same day, 3) precision when the same digest was acquired in replicates over several, individual days with other samples run in between, and 4) precision between different digests of the same sample. See chapter 7.2.11 for a detailed description of the experiment.

First, the quantification of deamidation, oxidation and N-glycosylation was studied in a digest of the cell culture filtrate. The modification degree from deamination and oxidation was quantified from fragment ions and N-glycan variants were quantified from the precursor ions when the DIA based method was used, while all three kinds of modifications were quantified from the precursor ions when the DDA based method was used. Sites with modification degrees measuring > 2% by at least one of the assays were included in the comparison. Figure 17a shows the mean quantification values using the DIA based method (blue bars) and the DDA based method (orange bars) in three technical replicates acquired the same day. The data is also displayed in Appendix table 1. The deamidation degree for the four sites included in the comparison measured between 1.60% ± 0.00060 and 3.68% ± 0.038 while the oxidation degree for all five sites measured between 10.41% ± 0.10 and 17.46% ± 0.037. Quantification of the N-glycan variants showed that the most abundant peak came from the G0F species, contributing to around 65% of the glycopeptides. The other glycovariants measuring more than 2% in the sample were the G0F-N, G1F, G0 and M5 species. This result agreed with the data from the released glycan analysis (Figure 12).

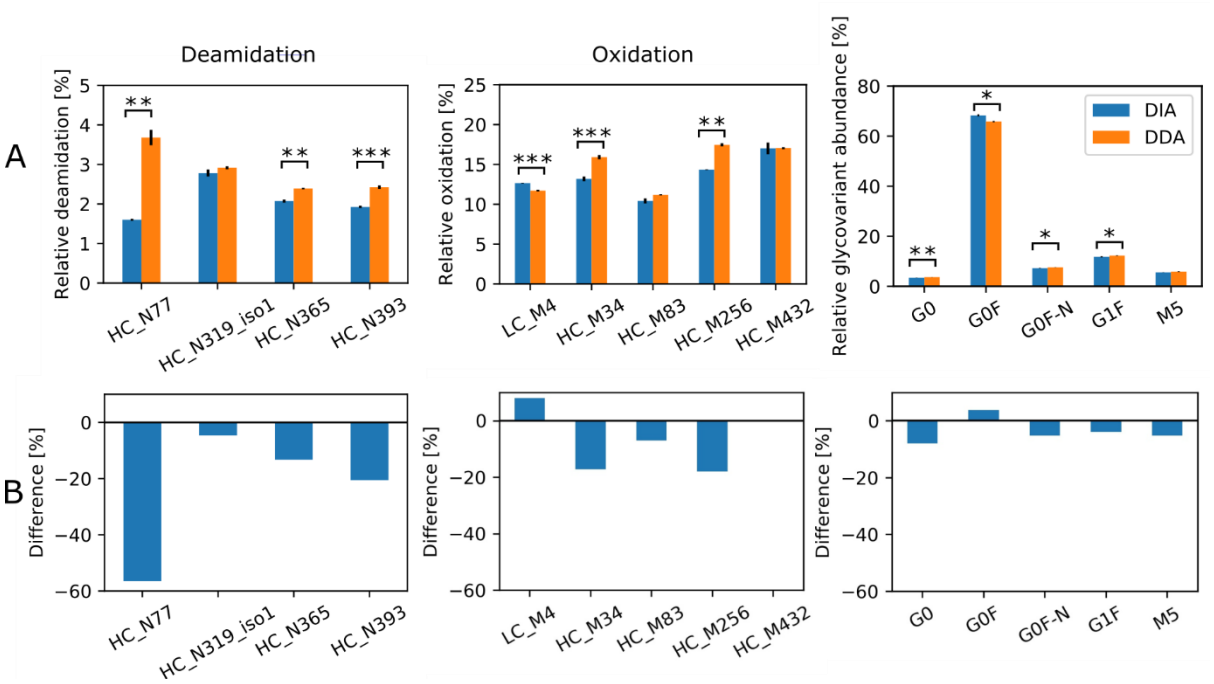


*Figure 17: Quantification of deamidation, oxidation and N-glycoforms in a digest of the adalimumab biosimilar cell culture filtrate using MS data acquired by DDA (orange) or DIA (blue). (a) Mean values and standard deviations as error bars from quantification in the same digest acquired on the same day, N=3. An independent samples t-test was performed on each peptide between the two acquisition methods, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (b) Difference between the values calculated using the DIA based method as compared to the DDA based method ($mean_{DIA}$/ $mean_{DDA}$). Modification degrees >2% were included. The raw data can be found in Appendix table 1. Oxidation and deamidation degrees were calculated from the XIC areas from fragment ions from the DIA data and from precursor ions from the DDA data. N-glycoform degrees were calculated from XICs from precursor ions from both the DIA and DDA data. See Table 2 and Table 3 for the amino acid sequences of the abbreviated peptides. The amino acid sequence of the glycopeptide is TKPREEQYNSTYR and compositional figures of the N-glycan variants can be found in Table 4.*

Figure 17b shows the difference between the quantitative values acquired by the two methods, calculated by dividing the value measured by the DIA based method with that measured by the DDA based method. The values did generally not differ by more than 20%. One exception to this was site HC_N77, for which the DDA based method gave a significantly higher deamidation degree which was more than twice of that measured by the DIA based method ($mean_{DIA}$ = 1.60% ± 0.00060, $mean_{DDA}$ = 3.68% ± 0.038, p = 0.0030). The difference was investigated by performing an independent samples t-test between the values obtained with DIA and DDA based methods, as described in chapter 7.2.8.4. Upon studying the spectra from the precursor ions and the fragment ions from the peptide containing deamidation at this site, signal interference was identified in the fourth isotope of the precursor ion. This signal interference could be attributed to the high value measured by the DDA based method, see Figure 18. This is another example showing that quantification using fragmentation data can be used as a tool to avoid signal interference from coeluting peptide precursor ions. Further, a trend could be observed that the quantification values from the DIA data were slightly lower than those from the DDA data.

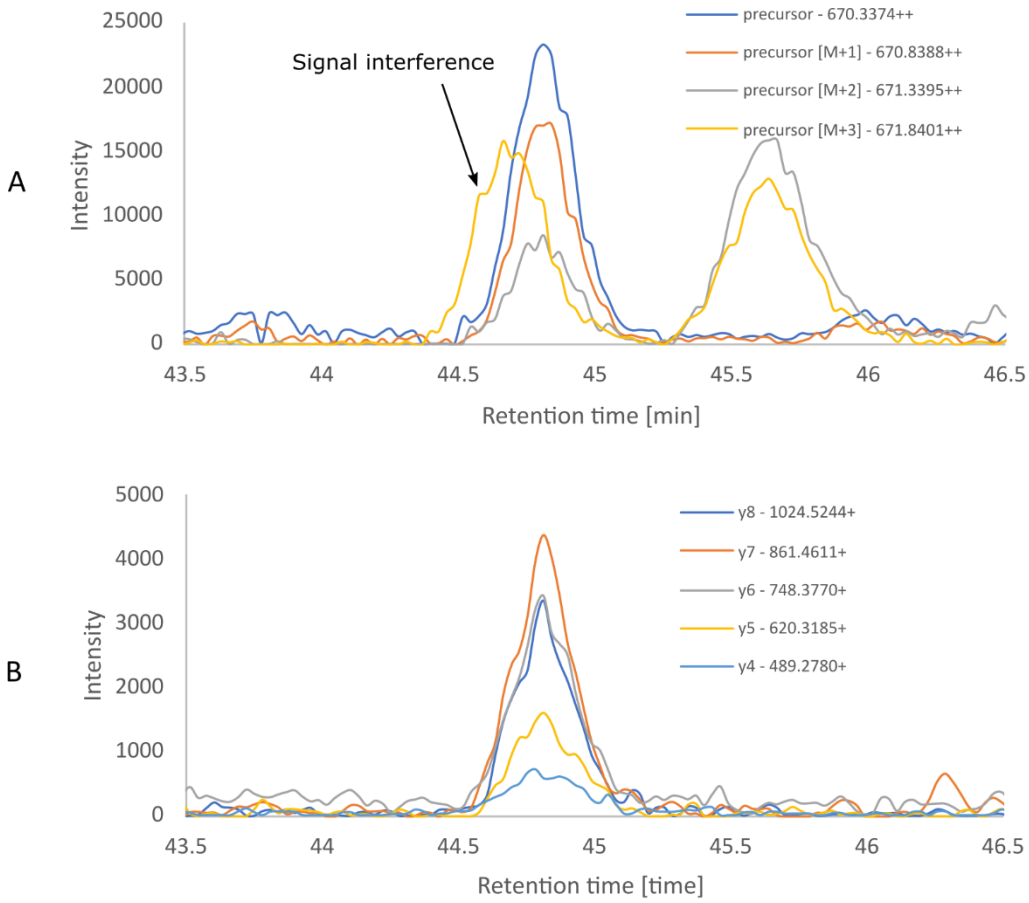


Figure 18: The peptide with deamidation at site HC_N77, showed (a) signal interference in the XICs of isotope [M+3] of the 2+ charged precursor ion but (b) no signal interference in the fragment XICs from the same precursor ion. The data was recorded in a digest of the adalimumab biosimilar cell culture filtrate analyzed by RP LC-MS acquired by DIA. See Table 3 for the amino acid sequence of the abbreviated peptide.

The assay precision within each acquisition method was then studied. First, the cell culture filtrate was digested and acquired with the DIA or the DDA based method in three technical replicates acquired after each other. Figure 19a shows the coefficient of variation (CV) between the replicate analyses when the same peptides as in Figure 17 were quantified. CV is a kind of normalized standard deviation, being the standard deviation divided by the mean value. The CVs for quantification of deamidation and oxidation events in the sample was under 0.05 between the three technical replicates acquired on the same day. The CVs for quantification of the glycovariants were even lower, under 0.03. In a second precision test, three technical replicates of the digested cell culture filtrate were acquired on three individual days with other samples analyzed in between. The CVs for quantification of deamidation and oxidation events within the replicate runs increased then slightly but not more than to 0.1, see Figure 19b. For quantification of the glycovariants the CVs were, however, still under 0.03. In a third precision test the cell culture filtrate was digested in five replicates started on five individual days. The CVs for quantification of deamidation events and glycoforms were then still under 0.1, Figure 19c. However, the CVs for quantification of oxidation events in the sample was as high as 0.2. This is much more than the technical variation, illustrated by the two first replicate levels, and may be attributed to variation in the sample preparation induced oxidation, which was already found to be problematic during optimization of the sample preparation protocol, see chapter 5.1.2. Testing the assay precisions on these three levels illustrated the increasing degree of variation that was introduced when going from one level to another. The mean values and CV values for all identified

peptides in the cell culture filtrate analyzed on the three replicate levels are also displayed in Appendix table 2.

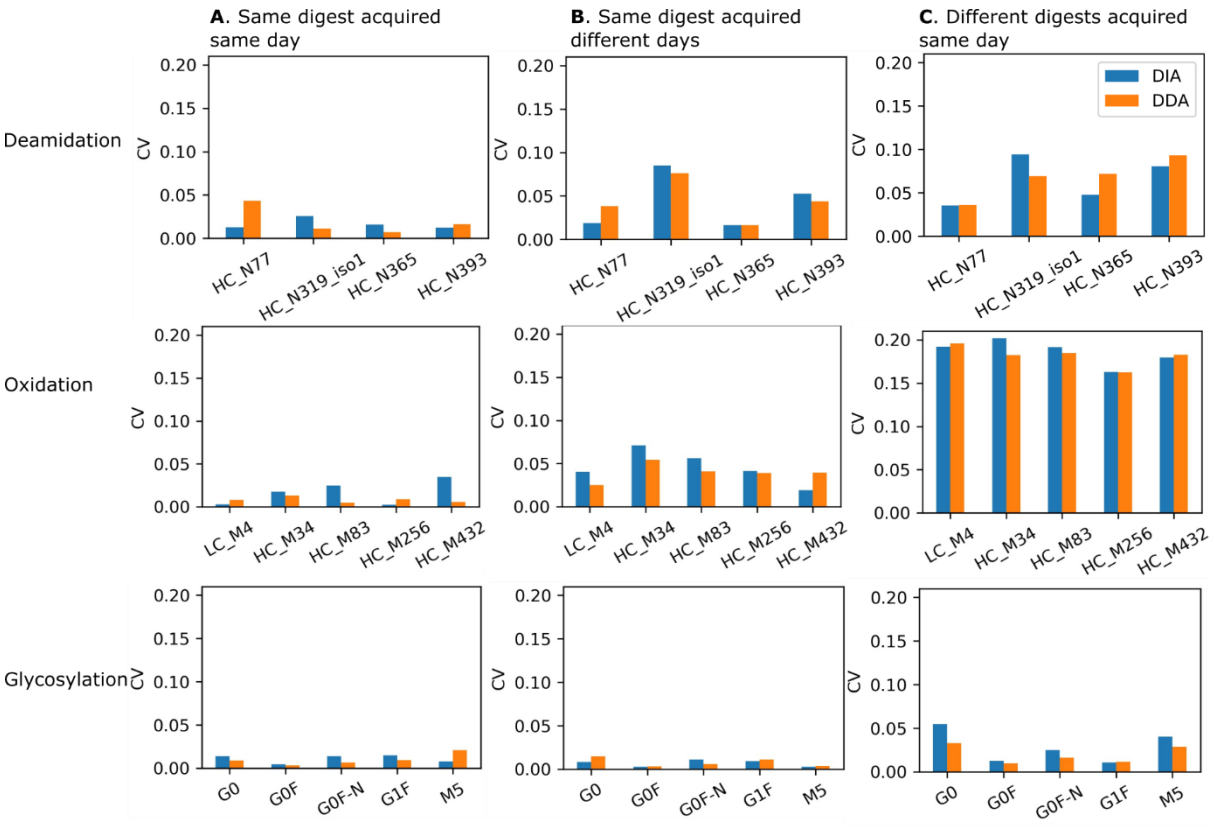


Figure 19: Coefficients of variation (CVs) for quantification of deamidation (top), oxidation (middle) and N-glycoforms (bottom) between replicate analyses of a digested adalimumab biosimilar cell culture filtrate. The CVs were measured on three replicates levels: (a) the same digest acquired on the same day, N = 3, (b) the same digest acquired on different days, N = 3 and (c) different digests performed on different days and acquired on the same day, N = 5. The digest data was acquired by both DDA (orange) and DIA (blue). Modification percentages measuring > 2% were included. The raw data can be found in Appendix table 2. Oxidation and deamidation degrees were calculated from the XIC areas from fragment ions from the DIA data and from precursor ions from the DDA data. N-glycoform degrees were calculated from XICs from precursor ions from both the DIA and DDA data. See Table 2, 3 for the amino acid sequences of the abbreviated peptides. The amino acid sequence of the glycopeptide is TKPREEQYNSTYR and compositional figures of the N-glycan variants can be found in Table 4.

It was hypothesized that the DDA based method would show lower precision than the DIA based method when applied on the cell culture filtrate because of signal interference in the precursor spectra that may be avoided by using DIA. No such trend could be observed when a modification degree cutoff of 2% was used, see Figure 19. However, when all identified deamidation events were included, also lower abundant events than 2%, the DIA based method showed higher precision than DDA between replicate digests for some sites, see Appendix figure 1. One such example was the site LC_N138 which measured around 0.5% deamidation in the sample and gave a CV value of 0.35 by the DDA based method and 0.03 by the DIA based method. The deamidated form of this peptide was identified with signal interference in its precursor ion spectra, see Figure 15. Signal interference could also be identified in the deamidated peptide containing site HC_N84 which also had a higher CV by the DDA based method than the DDA based method, 0.20 compared to 0.12 (data not shown).

The deamidation degree and assay precisions of the DIA and DDA based methods were then studied when applied on the Protein A purified and 2 weeks stressed IgG sample. Figure 20a shows the mean

deamidation values with error bars from four replicate digests as measured by the DIA (blue) and DDA (orange) based method with a cut-off value of 2%. The raw data is displayed in Appendix table 3.

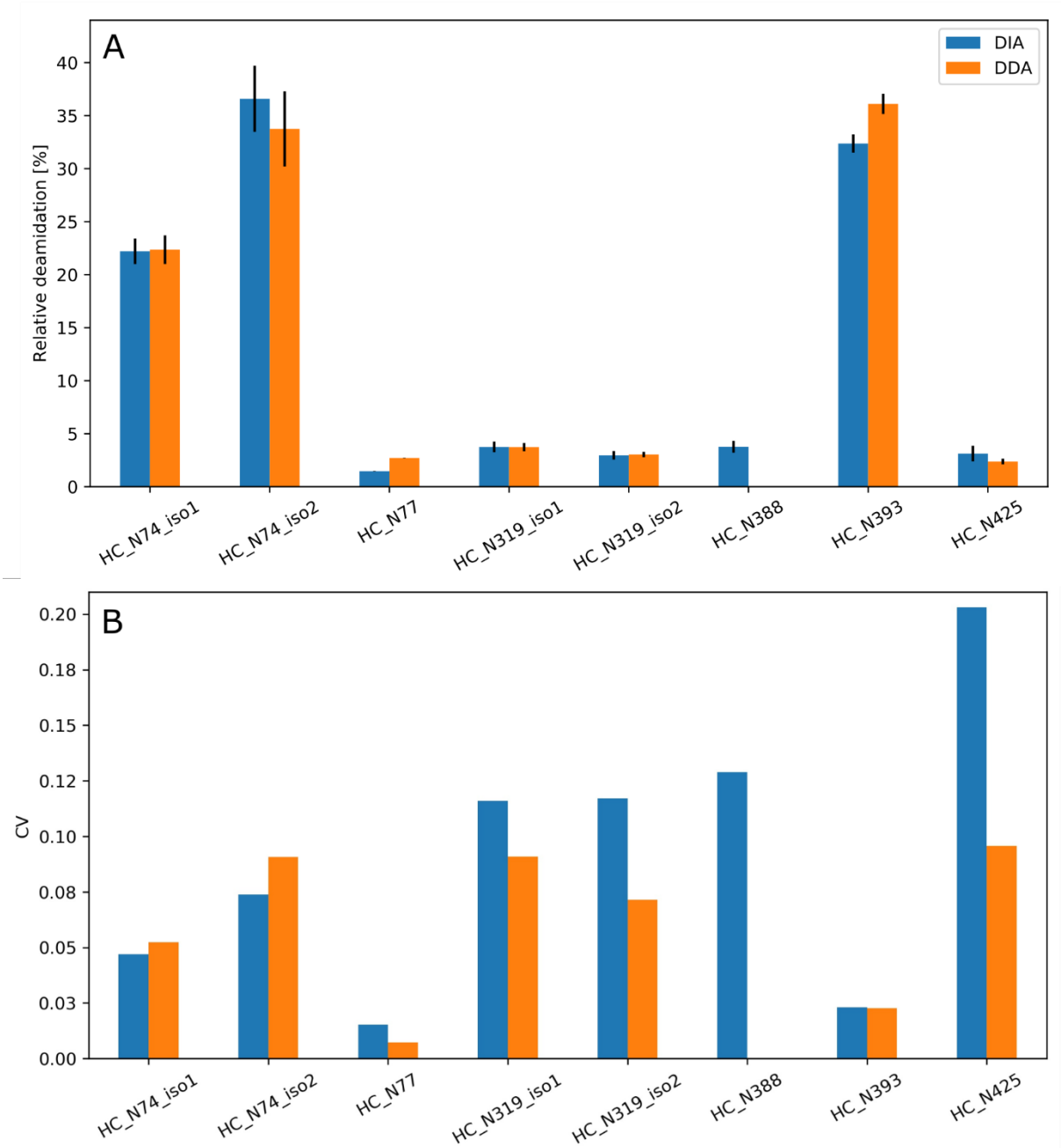


Figure 20: The Protein A purified adalimumab stressed at pH 9 and 37 ˚C for 2 weeks was digested in replicates, N = 4, and analyzed with the DDA based MAM assay (orange) and the DIA based MAM assay (blue). (a) Mean values for quantification of the deamidation degree at the amino acid sites, with error bars showing standard deviation. (b) CVs between the deamidation quantification values. Quantities were calculated from fragment XIC areas from the DIA data and from precursor XICs from the DDA data. Modification percentages measuring > 2% were included. The raw data can be found in Appendix table 3. See Table 3 for the amino acid sequences of the abbreviated peptides.

Site HC_N74, including two isoforms, and site HC_N393 showed high deamidation degree with values between 20% and 40%. The other five deamidation sites, HC_N77, HC_N319_iso1, N319_iso2, HC_N388 and HC_N425, showed values under 5%. In Figure 20b the CVs for deamidation quantification between the replicate digests are presented. No trend for higher precision, *i.e.* lower CVs, of the DIA based method compared to the DDA based method could be observed for the purified

and stressed IgG based on Figure 20b. The peptide containing site HC_N425 showed on the contrary much higher CV when the deamidation degree was quantified by the DIA based method compared to DDA, 0.20 compared to 0.10. Figure 21 shows the XICs from the deamidated variant of the peptide containing site HC_N425 in four replicate digests of the purified and stressed IgG from a) the +5 charged precursor ion and b) the five most abundant fragment ions from the same precursor ion.

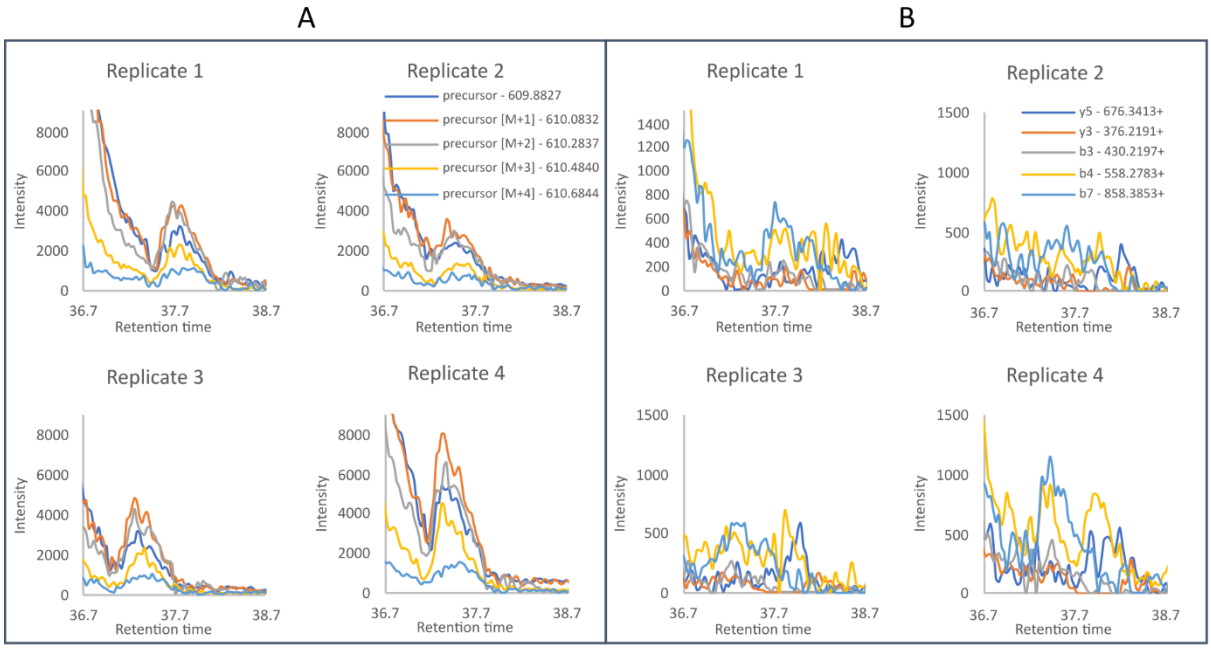


Figure 21: XICs from the 5+ charged peptide containing site HC_N425 from four replicate digests of the adalimumab biosimilar purified on Protein A and stressed for 2 weeks. (a) shows the precursor XICs acquired by DDA and (b) shows the fragment XICs acquired by DIA.

The signals from the precursor ions seem to be more stable than those from the fragment ions, based on visual inspection of the data illustrated in the figure, indicating that the sensitivity for this peptide is better when using the precursor ions for quantification than the fragment ions. The peptide containing site HC_N425 had a low signal, compared to the other peptides included in the study (data not shown), and a poor elution profile. The precision was therefore lower for deamidation quantification of this site compared to the other sites when applying the DIA based method, despite the deamidation degree being higher that 2%. When including deamidation abundances lower than 2% in the study, see Appendix figure 2, it could be seen that also site LC_N138, which measured around 0.5% deamidation, had a higher CV when measured by DIA, 0.20, than by DDA, 0.08. 0.20 is a much higher CV for the DIA based method and this site compared to when the same method was applied on the cell culture filtrate, by which the CV was 0.03 as mentioned above. This could be explained by the peptide signals generally being lower in the purified and stressed IgG than in the cell culture filtrate (data not shown), hence the higher CVs in the purified and stressed IgG when applying the DIA based method. Figure 22 shows the CVs plotted as a function of the total deamidated peptide signal for peptides quantified by the precursor ions (DDA) or by the fragment ions (DIA).

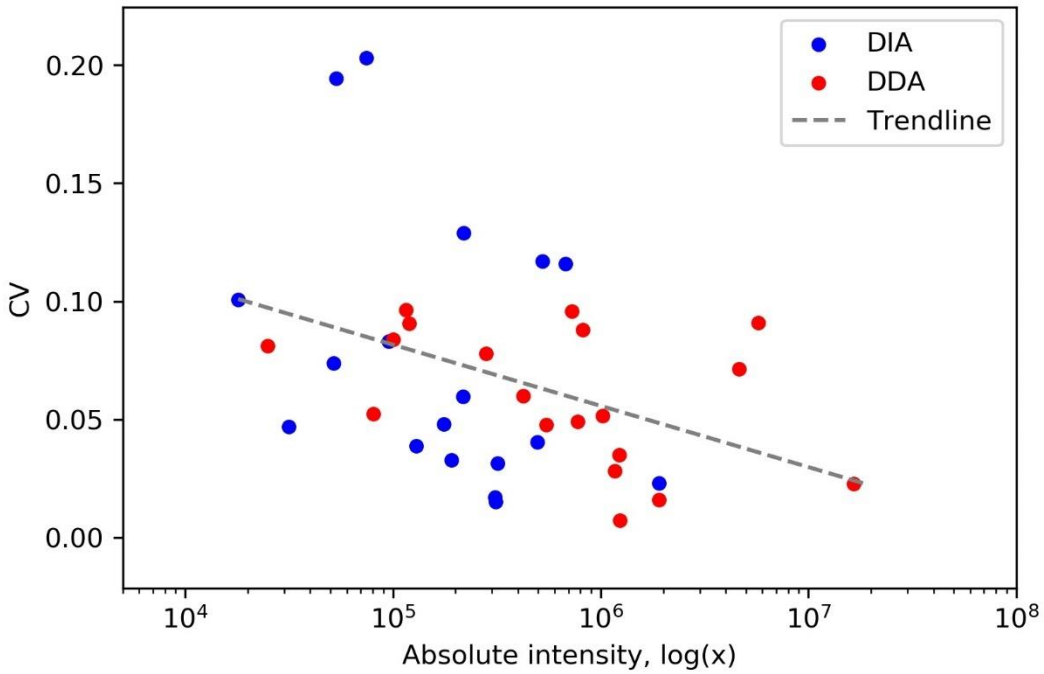


Figure 22: CVs between the deamidation quantification values in the Protein A purified and 2 weeks stressed IgG between replicate digests plotted as a function of the mean total XIC area from the deamidated peptide, N = 4. The areas are divided into two groups based on which method the sample was quantified by: the DIA based or the DDA based method. A trendline based on the whole data series is plotted having the function y = -0.0113ln(x) + 0.211 and the R^2 = 0.149.

From the data there appears to be a trend for a higher CV when the peptide signal is lower, as expected. By visual inspection, the peptides quantified by the DDA based method also appears to be centered around a higher peptide signal than those measured by DIA, which may be the reason for why more peptides have a high CV when quantified by the DIA based method. To conclude, the sensitivity of the DDA based MAM appeared to be better than that of the DIA based MAM when signal interference was not an issue. This led to a higher precision of the DDA based method than of the DIA based method for low abundant peptides in the purified and stressed IgG sample.

5.1.7 Assay linearity using the DIA based MAM

The assay linearity was investigated for quantification of methionine oxidation. Methionine oxidation was induced in the Protein A purified adalimumab biosimilar by incubating the protein under denaturing conditions in 0.1% hydrogen peroxide. This resulted in 100% oxidation of all five methionine residues, as quantified by the DDA based assay. The stressed and the non-stressed IgG samples were then digested separately, and the digests were mixed in 10 combinations so that the degree of stressed IgG was 0%, 1%, 2.5%, 5%, 7.5%, 10%, 20%, 40%, 60% and 80% in the samples. The samples were acquired using the DIA and the DDA based MS method and the oxidation degree was quantified and plotted against the theoretical oxidation degree. See chapter 7.2.12 for a detailed description of the experiment. Figure 23a shows the linearity of oxidation degree quantification in the peptide containing site LC_M4 (left) in the range 0-80% and that of site HC_M83 in the range 0-20% (right) using the DDA based MAM.

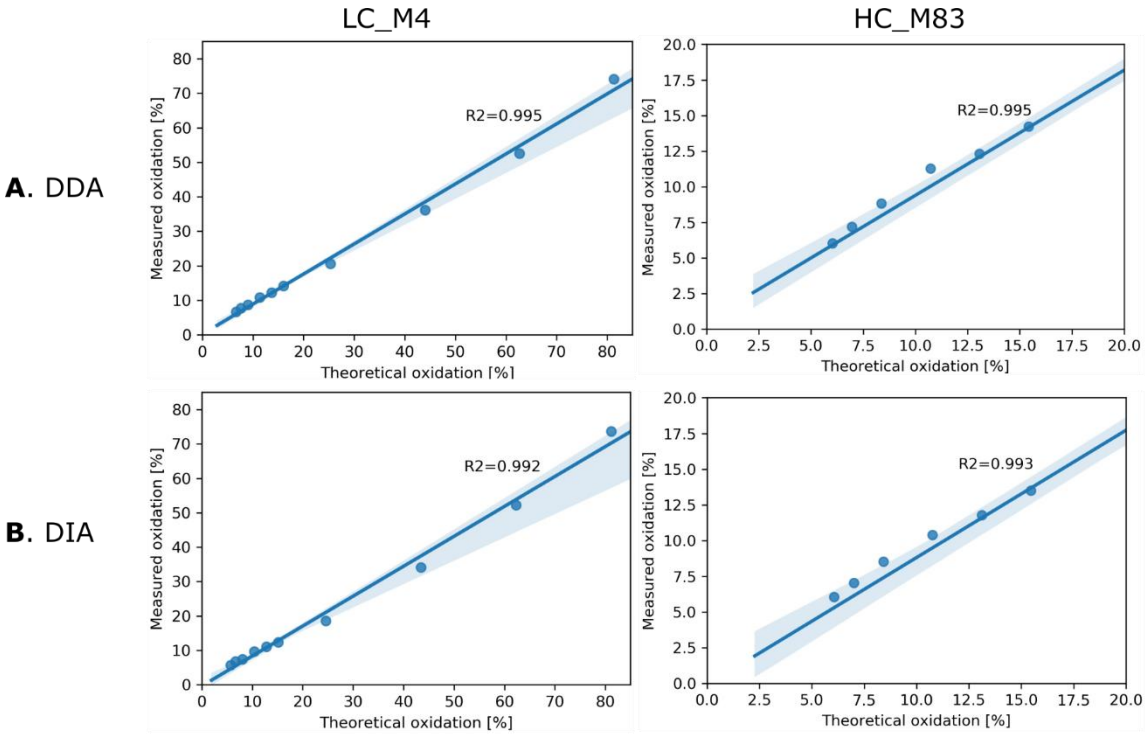


Figure 23: Linearity of quantification of methionine oxidation degree using the DDA and the DIA based MAMs. Two digests of Protein A purified adalimumab biosimilar, one with no induced oxidation and one with 100% induced oxidation, were mixed in different ratios. The linearity of oxidation degree quantification was compared for the peptides containing site LC_M4 in the range 0-80% and site HC_M83 in the range 0-20% using (a) the DDA based MAM and (b) the DIA based MAM. See Table 2 for the amino acid sequences of the abbreviated peptides.

The linearity of the DDA based assay was good for quantification both in the lower range and the upper range with an R^2 value > 0.99. The assay linearity was on the contrary poor when the DIA based assay was applied, and quantification was performed using the 5 fragments with the highest XIC areas for each peptide (data not shown). Another strategy for quantification was therefore developed for these peptides. The five fragments with the largest XIC areas from the oxidized peptide, excluding fragments from neutral loss, were used for quantification of both the oxidized and the non-oxidized peptides. For the oxidized peptides, the fragments with and without neutral loss were included. This strategy was selected since fragments may ionize differently depending on their amino acid composition. If the same fragments are selected for both the modified and the non-modified peptide this effect may be minimized. The oxidized methionine often loses the molecule CH_3SOH upon fragmentation. It seems therefore motivated to include both the fragment with and without this neutral loss in order to achieve an accurate quantification. Further, fragments which showed signal interference in the cell culture filtrate were not used. Another adjustment was that fragment b_3 was excluded from quantification of the peptide containing site LC_M4 since this fragment was much stronger in the non-modified than in the modified peptide. The reason for the decreased favorability of fragmentation between the modified amino acid residue and the residue at position N-1 may be an altered local electrochemical environment upon oxidation of the residue. This strategy of fragment selection resulted in good linearity of the DIA based method with R^2 values > 0.99, see Figure 23b. These fragments were therefore also used in the precision study in chapter 5.1.6 for quantification of the oxidation events. The two peptides containing the methionine sites HC_M34 and HC_M432 did not show good linearity when measured by the DDA based assay, see Appendix figure 3. The reason was probably that the concentration of oxidized peptides in the stressed sample and that of non-oxidized peptides in the non-stressed sample was not the same because of formation of other

oxidation products in the stressed sample. These may for example be dioxidized methionine residues and oxidation of other oxidation prone amino acids, such as histidine, tyrosine, tryptophan and phenylalanine. The linearity when quantifying the oxidation degree of these peptides was therefore not studied using the DIA based method.

5.1.8 Application of the DIA based MAM on stressed IgG samples

In summary, the optimized MAM based on DIA did perform similarly to a MAM based on DDA regarding assay precision. However, signal interference in the precursor spectra of a few low abundance deamidated peptides led overestimation of the deamidation degree of one site and possibly to the higher CV values measured for two other sites. The DDA based method did perform better in purified samples because of better assay sensitivity. Further, the assay linearity for quantification of the oxidation degree was comparable when using the DIA or the DDA based MAM when the fragments used for quantification from DIA data was properly selected. Regarding the overall performance of the optimized MAM, the CVs were not higher from replicate digests than from technical replicates analyzed on different days when N-glycoforms and deamidations were quantified. However, for quantification of the oxidation degree the CVs were about twice as high between replicate digests than between the technical replicates analyzed on different days.

The DIA based MAM was then tested on five samples from the same batch of the adalimumab biosimilar which had been treated in different ways. The following samples were analyzed: 1) an untreated cell culture filtrate, 2) the cell culture filtrate which had been incubated at pH 3.5 for 1 h followed by neutralization to pH 7, 3) the IgG purified by Protein A, 4) the IgG purified by Protein A which had been incubated at pH 3.5 for 1 h followed by neutralization to pH 7 and 5) the IgG purified by Protein A which had been stressed for 2 weeks at pH 9 and 37 °C. The four first samples were sent to us by our collaborators in Vienna while the fifth sample was sample 3) that had been stressed by us. All samples were analyzed from 5 digest replicates, besides the IgG stressed for 2 weeks which was digested in 4 replicates. See chapter 7.2.13 for a detailed description of the experiment.

The samples that were stressed at pH 3.5 for 1 h did not show different deamidation levels compared to the corresponding samples which had not been stressed, see Figure 24a. This was valid both for the cell culture filtrate and the Protein A purified IgG. This conclusion was based on the fact that all the measured sites had overlapping error bars, which will also be the strategy used to determine differences in modification degrees in the following results. However, the deamidation level of site HC_N77 was slightly higher in the cell culture filtrate than in the Protein A purified IgG, showing 1.6% deamidation compared to 1.2% respectively. This may be a result of the purification. The deamidation levels of sites LC_137, HC_319_iso1 and HC_365 were around 2%, 3% and 2% respectively in all measured samples. It is difficult do know if this is a true deamidation that appeared already during sample production or if the deamidation levels are sample preparation induced. The deamidation degree was much higher for several sites in the 2 weeks stressed sample than in the other samples, as expected. For example, site HC_N393 showed a particularly high deamidation degree in the 2 weeks stressed sample, measuring 32%, compared to around 2% in the other samples. This peptide contains the deamidation prone *NG* motif (site HC_N388) but, interestingly, it was the asparagine site located between a glutamine and another asparagine residue that was found to be deamidation prone during the applied stress conditions.

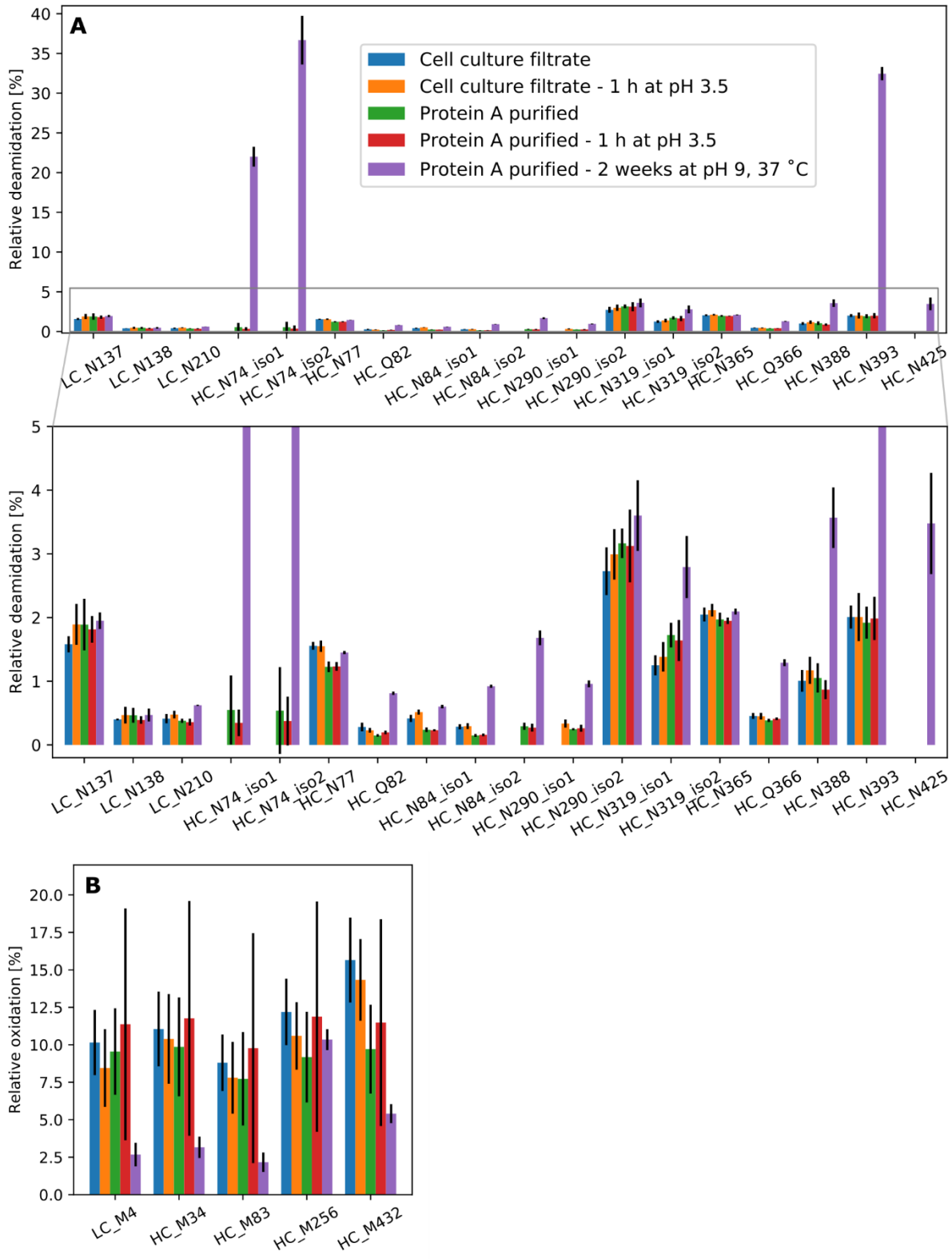


Figure 24: Quantification of (a) deamidation and (b) oxidation using the DIA based MAM in five different adalimumab biosimilar samples. The samples were: 1) untreated cell culture filtrate, 2) cell culture filtrate which had been left at pH 3.5 for 1 h followed by neutralization to pH 7, 3) the IgG purified on Protein A, (4) the IgG purified on Protein A which had been left at pH 3.5 for 1 h followed by neutralization to pH 7 and 5) the IgG purified on Protein A which had been left at pH 9 and 37 ˚C for 2 weeks. The graph shows mean values from replicate digests (sample 1-4: N = 5, sample 5: N = 4) with error bars showing standard deviation. 40 µg IgG was prepared of samples 1-4 and 10 µg of sample 5. See Table 2 and Table 3 for the amino acid sequences of the abbreviated peptides.

Another site that was found with a high deamidation degree in the 2 weeks stressed sample was site HC_N74, for which two deamidated isoforms were present in 22% and 37% respectively in the stressed sample while both isoforms measured around 0.5% in the other samples. The deamidation degree at this site may be overestimated in this assay though, since the deamidation degree was much lower when 40 µg of the same sample was used for the assay, around 1%, see Figure 11, than when 10 µg was used, as in this assay. This result will be further discussed in chapter 6.1. See Appendix figure 4 for the same figure as Figure 24, with the change that the digested IgG mass is 40 µg for all samples and no replicates are present for the 2 weeks stressed IgG. Other examples of sites were the deamidation degree was higher in the 2 weeks stressed sample were sites HC_290_iso1, HC_319_iso2, HC_N388 and HC_N425 measuring 1.7%, 2.8%, 3.6% and 3.5% respectively in the stressed sample as compared to 0.3%, 1.7%, 1.1% and 0% respectively in the non-stressed Protein A purified sample.

Quantification of methionine oxidation in the four first samples resulted in mean values between 10-15%, see Figure 24b. The standard deviations between the five replicates were, however, too high to determine any differences in oxidation levels. The oxidation level in the 2 weeks stressed sample seemed to be lower than in the other samples for all methionine sites besides site HC_M256. This is not so probable, since this is the Protein A purified IgG that have been stressed and it seems unlikely that the oxidation level would decrease during temperature stressing. The 2 weeks stressed sample was, however, digested in 10 µg while the other samples were digested in 40 µl. Perhaps this may have had an effect on the sample-preparation induced oxidation levels.

5.2 Discovery of a new potential CQA in rhGM-CSF and optimization of a MAM

An important part of MAM assay development is to evaluate new potential CQAs in therapeutic proteins and then be able to include these in a MAM, see chapter 3.2.5.4. During quality control of fermentation batches of human granulocyte macrophage colony stimulating factor (GM-CSF, see chapter 3.1.3), recombinantly expressed in E. coli, a modification of +70 Da was discovered in the protein. Since this modification has not been well documented in literature, a project was defined involving characterization of the modification. This led to the definition of the second sub-aim of this thesis: to identify the chemical structure of this novel modification by combining chemical assays with mass spectrometry in order to identify its source and determine if it may be a CQA. If the modification was considered a CQA it should then be included in a MAM for the protein.

Another CQA that is not commonly introduced into MAM assays is disulfide scrambling. One reason for this is that the data may be very complex if the digestion is incomplete or if the selection of enzymes is not well planned. The third sub-aim of this thesis was: to develop a MAM for rhGM-CSF targeting disulfide bond characterization as well as other CQAs with focus on proper selection of proteolytic enzymes in order to decrease peptide size and data complexity. The method should then be applied on samples subjected to forced degradation as well as process samples in order to demonstrate its applicability to quantify the targeted modification types in the GM-CSF.

Figure 25 describes the steps to be taken to fulfill sub-aims two and three. These involve 1) characterization of a novel modification of +70 Da in rhGM-CSF, 2) selection of target modification types for the MAM, 3) optimization of the sample preparation method for the target protein and modifications and 4) testing the assay on representative samples.

Development scheme for DDA based MAM for GM-CSF:

	1. Novel attribute characterization	2. MAM method attribute selection	3. Sample preparation optimization	4. Test assay
Chapter	5.2.1	5.2.2	5.2.3	5.2.4

Figure 25: Scheme describing the steps taken to set up and test a MAM for the therapeutic protein rhGM-CSF.

5.2.1 Characterization of a novel +70 Da modification in rhGM-CSF produced in *E. coli*

When a fermentation batch of rhGM-CSF was analyzed intact by RP HPLC coupled to UV detection and ESI-MS it was found that the main protein peak contained several species upon deconvolution of its charge state envelope, Figure 26a.

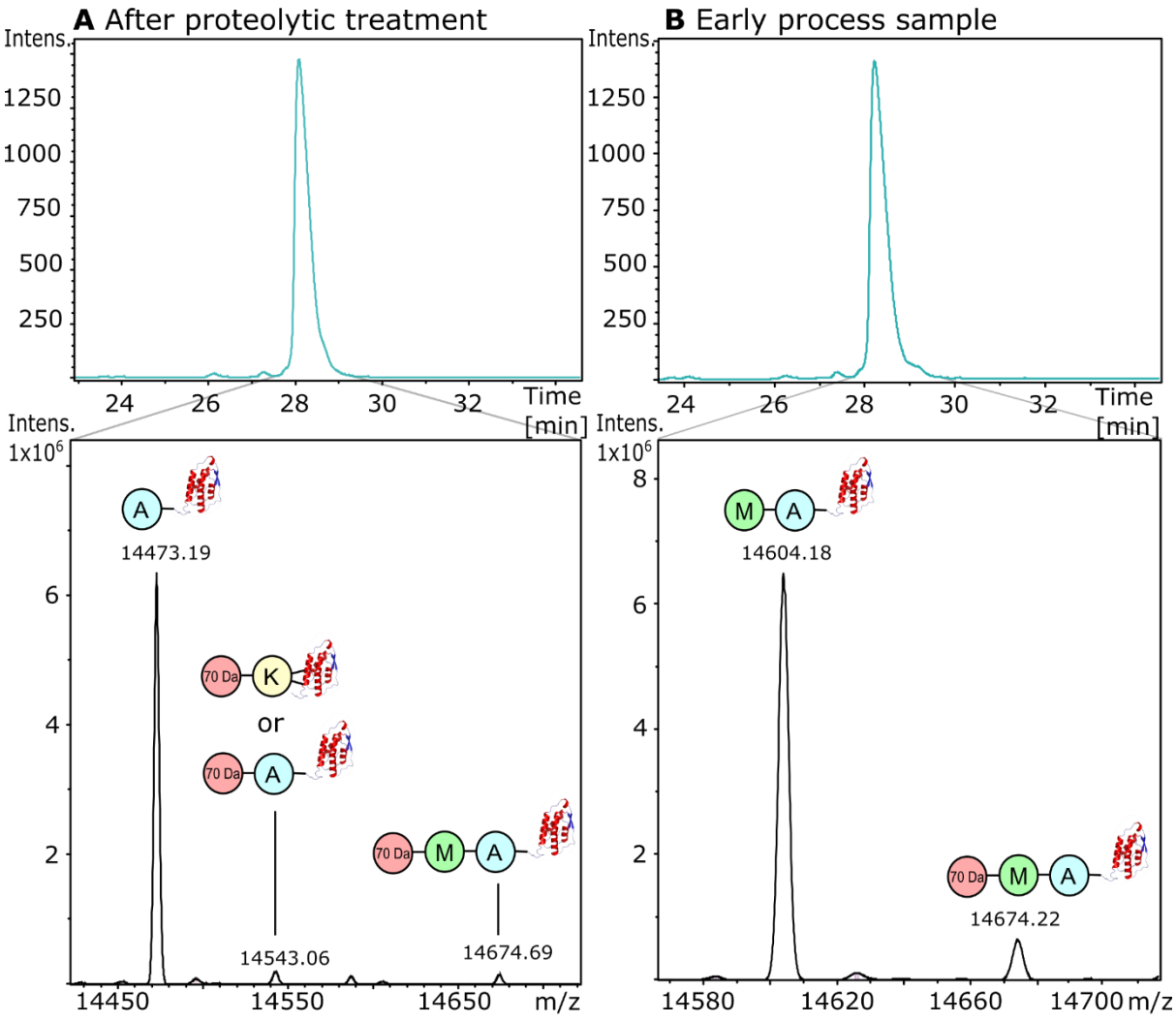


Figure 26: Intact analysis of two process samples of GM-CSF, recombinantly expressed in E. coli, by RP-LC coupled to UV (upper picture) and ESI-MS led to identification of a modification of +70 Da in the protein. Deconvolution of the main UV peak from analysis of the two process samples (lower picture) showed that a) the early process sample, which had not been proteolytically digested to remove the N-terminal methionine, contained two proteoforms: the methionylated protein and the same protein with a +70 Da modification. (b) The later process sample, which had been proteolytically digested, contained mainly the expected product but also low abundance of the protein with a +70 Da modification and the protein with an N-terminal methionine and a +70 Da modification. The location of the modification in the drawn picture is an educated guess since intact analysis cannot confirm the location of the modifications in the protein sequence.

The main species could be attributed to the native protein, with the mass 14473.57 Da. Two low abundance proteoforms were also present in the sample, with the masses 14532.76 Da and 14674.85 Da. The first of the two proteoforms had the same mass as the native protein with an additional mass of +70 Da. The second of the proteoforms had the same mass as the native protein with and extra methionine residue and an additional mass of +70 Da. When eukaryotic proteins are recombinantly expressed in E. coli, protein translation is initiated by a methionine residue. This residue is then digested by bacterial proteases after expression. When proteins are over-expressed in *E. coli* the bacterial proteases can often not keep up with the rate of protein expression which results in a certain degree of methionylated proteins in the product. The protein product is therefore often subjected to proteolytic digestion in vitro, after purification from the host. It is therefore not unlikely to find a protein with an extra methionine residue in the sample. The additional mass of +70 Da could, however, not be attributed to any modification commonly found in protein samples. To confirm that the second of the low abundance proteoforms did contain a methionine residue, an early process sample of the same batch but which had not yet been subjected to proteolytic treatment was analyzed by the same method. The major proteoform in this sample was the native protein with an extra methionine residue, with the mass 14604.73 Da, Figure 26b. The same proteoform that was present in the proteolytically treated sample, corresponding to the methionylated protein with a +70 Da modification, mass 14674.86 Da, was also present in high abundance in the sample. This confirmed that the proteolytically processed sample contained a low abundance of the native protein and the methionylated protein, both with an additional mass of +70 Da. This data led to the formation of the second sub-aims of the thesis: to identify the chemical structure of this novel modification by combining chemical assays with mass spectrometry and through this identify its source and determine if it may be a CQA. See chapter 7.3.6 for a detailed description of the MS analysis.

To locate the modification in the protein sequence, the methionylated early process sample was digested using the enzymes GluC and LysC. See chapter 7.3.2 for the digestion protocol. These specific enzymes were selected to achieve full sequence coverage. Trypsin, which is commonly used for protein digestion, could not be used since GM-CSF contains an arginine reside at the fourth site from the protein N-terminal. The digested protein was acquired by LC-MS/MS using DDA, see chapter 7.3.7. The peptides in the sample were identified using the Mascot database search engine and the peak areas were analyzed using Skyline. The +70 Da modification was found to be mainly located on the protein N-terminal and to a lower degree on lysine residues (data not shown).

To determine the elemental composition of the modification, low m/z MS/MS fragments from the modified and non-modified N-terminal peptide were used, see Figure 27a and Figure 27b.

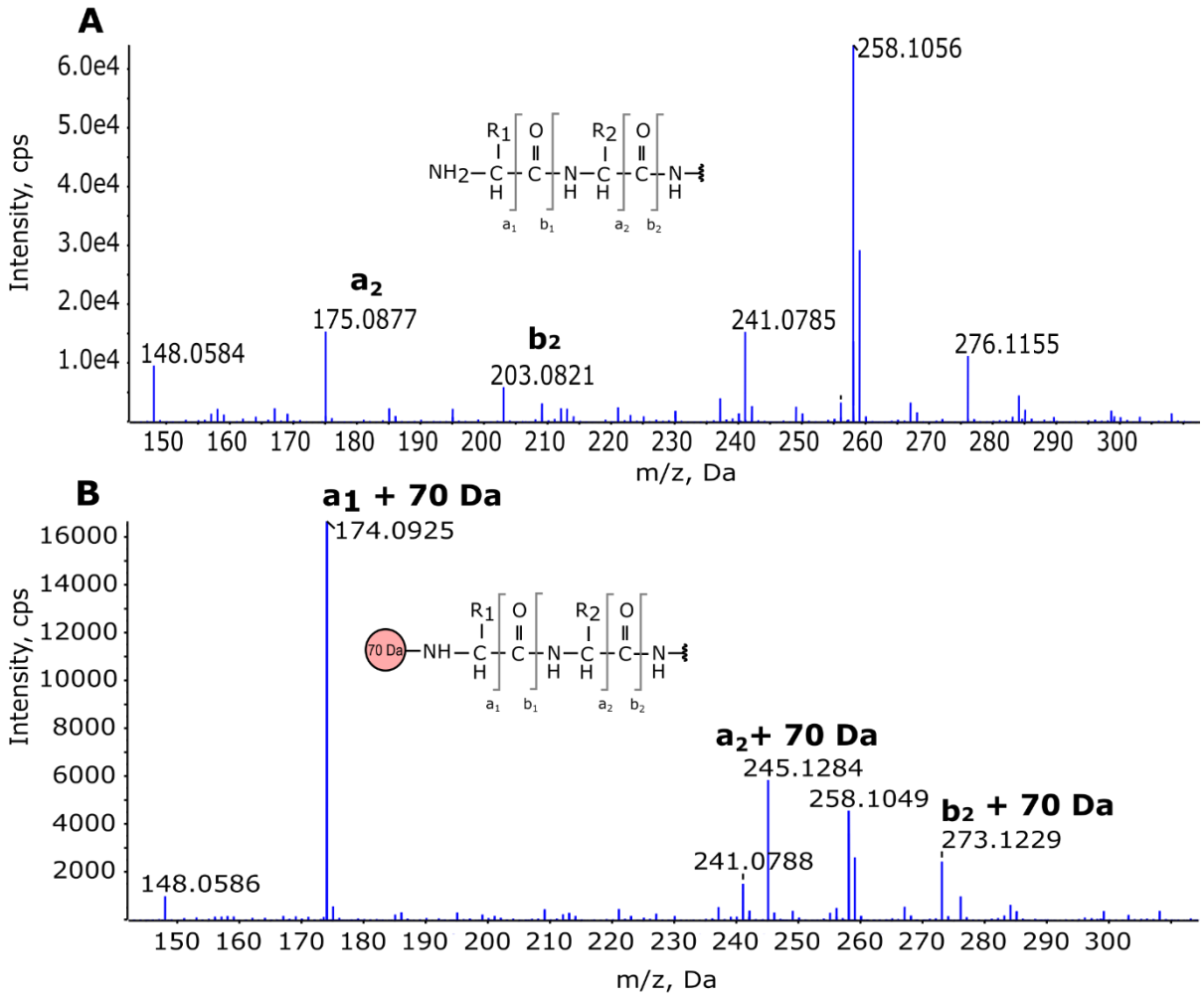


Figure 27: The methionylated early process sample was analyzed by bottom-up proteomics using proteases GluC and LysC for digestion. The masses of the low m/z MS/MS fragmentation ions from the protein N-terminal peptide MAPARSPSPSTQPWEHVNAIQE with charge state +3, (a) without and (b) with the +70 Da modification, were used to calculate the mass of the modification with high accuracy. The mass difference between the a_1, a_2 and b_2 fragment ions for the two peptides were used to calculate the mass.

The mass difference between the a_1, a_2 and b_2 fragment ions from the two peptides were calculated to determine the mass of the modification with high accuracy and the mean value from these calculations was found to be 70.0424 m/z. The theoretical value of the a_1 fragment from the non-modified peptide was used since only m/z values above 130 were recorded. This value was compared to the theoretical mass of a number of possible elemental compositions, Table 6. The elemental composition that had the lowest mass deviation from the experimental mass was C_4H_6O.

Table 6: Chemical compositions of theoretical molecular groups with a mass deviation of +/- 0.05 Da from the experimental value of the +70 Da modification identified in GM-CSF. The experimental value of the modification was 70.042 Da. The theoretical mass, mass deviation and source, if documented in literature, of the molecular groups are included. The calculation of the theoretical masses is described in chapter 7.3.9.

Elemental composition	Theoretical mass [Da]	Mass deviation [ppm]	Source
C2NO2	69.992904	706.18	
CN3O	70.004137	545.81	
C3H2O2	70.0054792	526.65	Pyruvic acid[1]
N5	70.01537	385.43	
C2H2N2O	70.0167122	366.27	
CH2N4	70.0279452	205.90	
C3H4NO	70.0292874	186.73	
C2H4N3	70.0405204	26.36	
C4H6O	**70.0418626**	**7.20**	Crotonaldehyde[2], methyl vinyl ketone[3], butyryl-CoA[4]
C3H6N2	70.0530956	153.18	
C4H8N	70.0656708	332.71	
C5H10	70.078246	512.25	Pentanal[5]

A thorough literature study was performed to investigate if other studies have identified 70 Da modifications in proteins. In a study by Ichihashi et al (73) it was found that the reactive 2-alkenal crotonaldehyde derivatizes with nucleophilic amino acid side chains, such as lysine and histidine, through Michael addition under physiological conditions. The result was the addition of a C_4H_6O moiety, with the mass +70 Da, Figure 28a.

[1] (72)
[2] (73)
[3] (74)
[4] (75)
[5] (76)

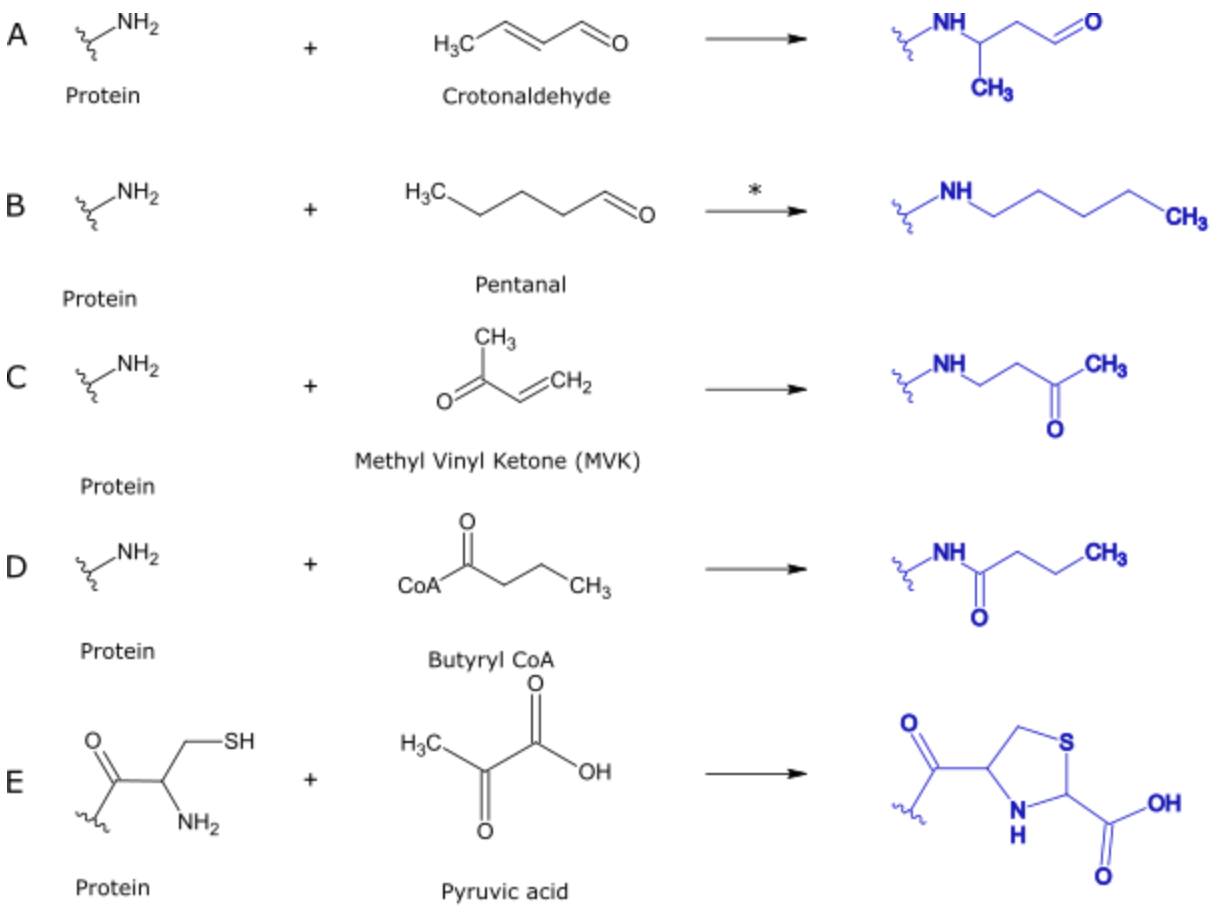


Figure 28: Five sources to protein primary amine modification resulting in addition of 70 Da documented in literature. (a) (73), (b) (76), the reduced Schiff base product is shown, (c) (74), (d) (75), (e) (72).

In another study the aldehyde pentanal was found to react with lysine residues to add the molecule C_5H_{10} after reduction of the aldehyde double bond (76), Figure 28b. The reaction proceeded through the formation of a Schiff's base and added a mass of +70 Da. Further, ketones may react with protein side chains. In two separate studies the compounds methyl vinyl ketone (MVK) and ethyl vinyl ketone were found to form derivates with a protein N-terminal valine of hemoglobin through Michael addition (74,77). Addition of MVK led the same added elemental composition as crotonaldehyde described above, Figure 28c. Studies have also shown that lysine residues may be butyrylated through the compound butyryl-CoA in the same fashion as acetylation may proceed through the compound acetyl-CoA (75,78). The resulting elemental composition is then also the same as from addition of a crotonaldehyde, described above, Figure 28d. Further, an adduct from what was thought to be pyruvic acid has been described and identified on the N-terminal cysteine of a proteins recombinantly expressed in *E. coli* (72), Figure 28e. The elemental compositions and theoretical masses from the compounds identified in the literature study are also summarized in Table 6.

From the protein modifications identified in the literature study, the elemental composition C_4H_6O from derivatization with crotonaldehyde, MVK or butyryl-CoA, had the lowest mass deviation from its theoretical value: 7.20 ppm. Both the elemental composition from addition of pyruvic acid and pentanal resulted in a mass deviation of around 500 ppm. This was considered being outside the range of the instrument's precision. The conclusion was therefore that C_4H_6O was the most probable elemental composition of the modification.

To determine the chemical structure of the modification, the methionylated GM-CSF was subjected to two chemical assays. Both the modifications from crotonaldehyde, MVK and butyryl-CoA contains a carbonyl moiety. To confirm that the modification did contain a carbonyl moiety the protein was reacted with 2,4-dinitrophenylhydrazine (DNPH), a classic carbonyl reagent, in the first assay. The protein was digested by GluC and Lys C followed by incubation overnight in 100 mM DNPH in acidic conditions. The protein was then analyzed by LC-MS/MS acquisition. See chapter 7.3.4 for a detailed protocol of the experiment. Figure 29a and Figure 29b shows the XICs from the protein N-terminal peptide and a lysine containing peptide with the +70 Da adduct before incubation with DNPH.

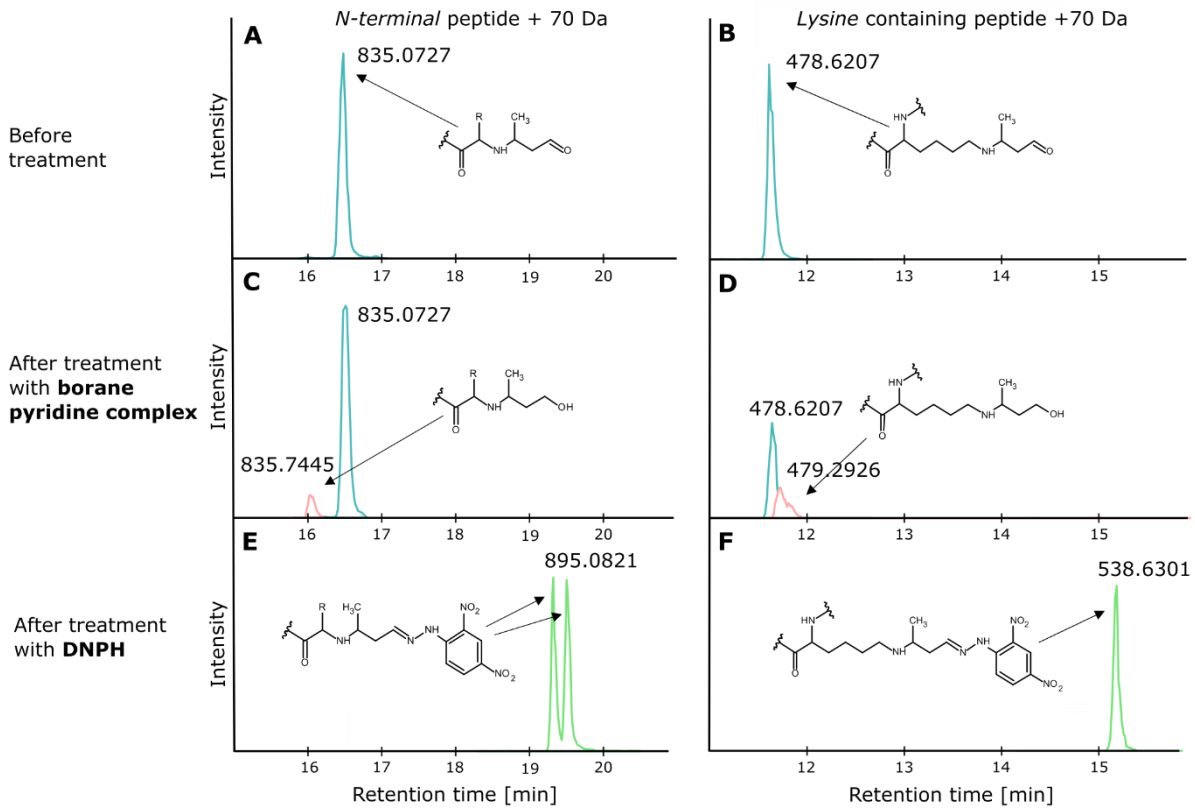


Figure 29: The GM-CSF early process sample containing the +70 Da modification was characterized by two chemical assays in combination with bottom-up MS analysis. The XIC from (a) the protein N-terminal peptide M[+70]APARSPSPSTQPWEHVNAIQE with charge state +3 and (b) the lysine containing peptide LYK[+70]QGLRGSLTK with charge state +3, without any chemical treatment. The XIC from (c) the N-terminal peptide and (c) the lysine containing peptide in the protein sample after treatment with borane pyridine complex. The XIC from (e) the N-terminal peptide and (f) the lysine containing peptide in the protein sample after treatment with DNPH.

Figure 29e and Figure 29f show the XICs from the same two peptides after incubation with DNPH. Both the +70 Da modification in the N-terminal peptide and in the lysine containing peptide were fully converted to the adduct +250 Da, corresponding to a DNPH moiety. The peak splitting of the N-terminal peptide is probably due to the formation of diasteromers, leading to changes in the protein secondary structure, upon derivatization with the aldehyde oxygen (29). The second assay was performed to possibly exclude the modification by butyryl-CoA. In the modification from butyryl-CoA the carbonyl carbon is involved in a strong amide bond. A weak reducing agent like borane pyridine complex would not be able to reduce a carbonyl being part of such a strong bond. 100 mM borane pyridine complex was therefore reacted with the intact GM-CSF early process sample overnight at denaturing conditions. The protein was analyzed by digestion with GluC and LysC and acquisition by LC-MS/MS. See chapter 7.3.3 for a detailed protocol of the experiment. Figure 29c and Figure 29d show the XICs from the N-terminal peptide and the lysine containing peptide from Figure 29a and

Figure 29b after the reduction. In both peptides two hydrogen molecules were incorporated in the +70 Da modification upon reduction with borane pyridine. Butyryl-CoA could thereby be excluded as a possible candidate to have induced to modification.

To investigate closer if the +70 Da modification could come from modification by MVK or crotonaldehyde, a late process sample of GM-CSF with a low degree of +70 Da modification was derivatized with these compounds. Commercially available MVK and crotonaldehyde was bought and left to react with the GM-CSF for 24 h at pH 8.5. The protein was then analyzed by digestion with GluC and LysC followed by acquisition by LC-MS/MS. See chapter 7.3.5 for a detailed protocol of the experiment. Incubation of GM-CSF with MVK led to the formation of adducts of +70 Da at some lysine sites. Figure 30c shows the XICs from peptide *Q[+17]GLRGSLTK[+70]LK* from incubation of GM-CSF with 1 µM, 10µM and 100 µM MVK.

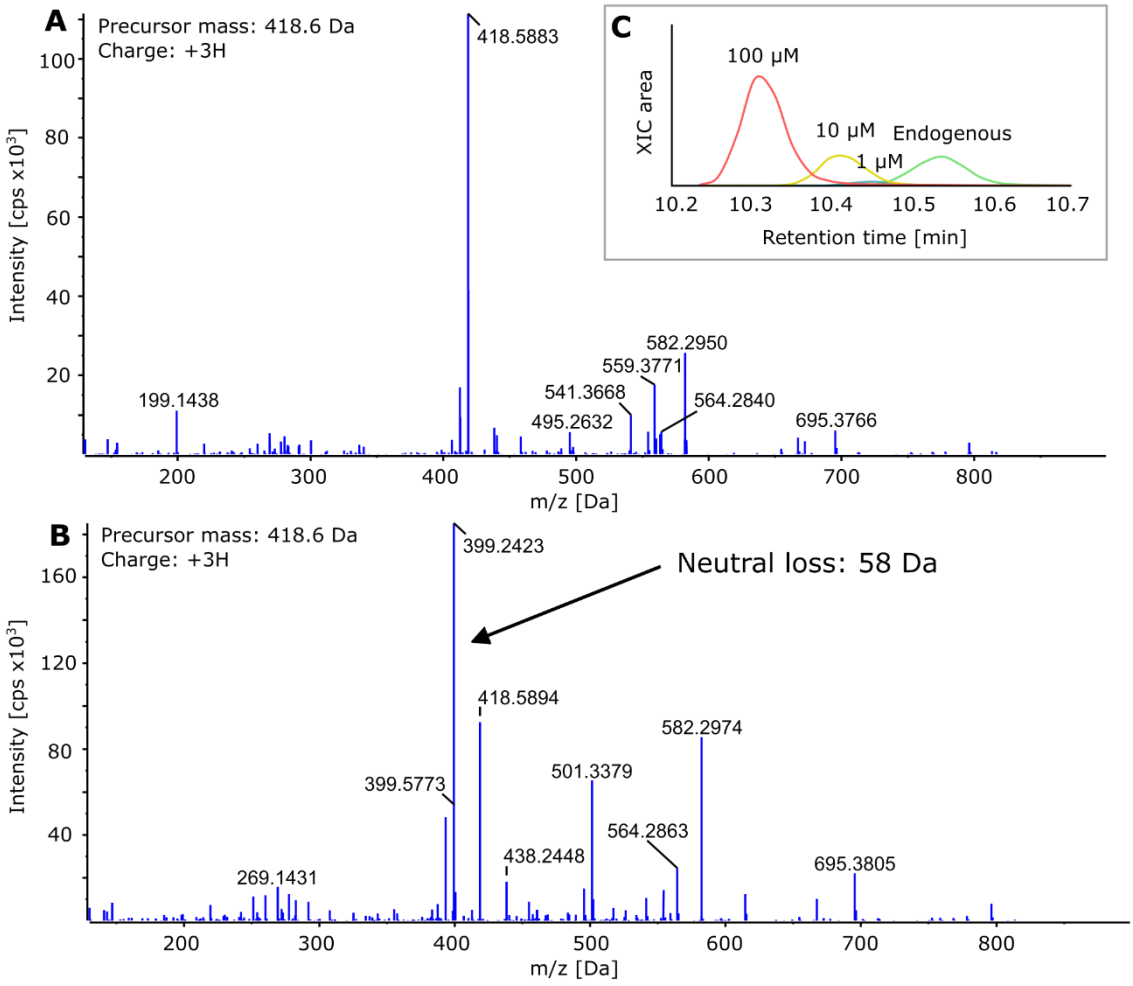


Figure 30: A GM-CSF sample with very low degrees of the +70 Da modification was derivatized with MVK and analyzed by bottom-up together with the methionylated sample containing the endogenous +70 Da modification. The MS/MS fragmentation spectra from the peptide Q[+17]GLRGSLTK[+70]LK with charge state +3 from (a) the peptide with the endogenous modification and (b) the peptide with the MVK derivatized modification were compared. (c) The retention time of the peptide with the endogenous modification (green) was also compared with that of the same peptide in samples derivatized with 1 µM (blue), 10 µM (yellow) and 100 µM (red) MVK.

The figure also shows the XIC from this peptide in the early process sample containing the endogenous modification. It can be seen from the figure that the retention time of the peptide with endogenous modification is about 0.2 min later that that with MVK. Figure 30a and Figure 30b show the MS/MS

fragmentation spectra from the same peptide with the endogenous modification and with the modification from MVK. The fragmentation spectra from the peptide with the endogenous modification showed a neutral loss of 58 Da that could not be observed from the peptide with the endogenous modification. This neutral loss seemed to be a trend for all peptides derivatized by the MVK modification. It could therefore be concluded that the modification did not come from MVK. When the GM-CSF late process sample was reacted with crotonaldehyde, no +70 Da adduct formation could be identified at the protein N-terminal or lysine residues (data not shown). Modification of histidine residues, which is also a good nucleophile, was however identified.

Other proteoforms identified in the same batch as the +70 Da modification contained the amino acid substitution from isoleucine to valine. This substitution has been metabolically linked to oxygen depletion, leading to an overexpression of valine through the intermediate pyruvate. Since the metabolic link to the +70 Da modification arising from a pyruvate group was very close, derivatization with pyruvic acid was also explored. This was done despite the fact that the pyruvate mass differed by ~500 ppm by calculation form low m/z fragment masses, Table 6. It was found that pyruvate reacted with the protein to form an adduct of +70 Da by intact mass analysis by RP HPLC coupled to UV detection and ESI-MS (data not shown). The modification was, however, not detected by peptide mapping. This points in the direction of that either the modification reacted with other amino acid groups than the protein N-terminal, lysine and histidine, or the modification was not stable enough to be detected by bottom-up analysis. In either case the conclusion was that the modification from pyruvate does not possess the same properties as the endogenous modification.

Since it could not be demonstrated that reacting GM-CSF with crotonaldehyde resulted in a +70 Da adduct, it was investigated if there could be any metabolic connection to the formation of crotonaldehyde. The recording of a number of parameters during production of the specific batch was therefore investigated. It was found that the feeding of glucose during production of the batch had escalated far above normal values. This in turn had resulted in depletion of the oxygen levels in the batch. The formation of crotonaldehyde could be metabolically linked to this oxygen depletion, see discussion in chapter 6.2. Carbonylation of proteins has further been related to protein dysfunction. The modification was therefore considered a possible CQA which should be included in a MAM assay for rhGM-CSF.

5.2.2 Selection of target modification types for a MAM for rhGM-CSF

The third sub-aim of this thesis involved developing a MAM for rhGM-CSF targeting disulfide bond characterization as well as other CQAs, see chapter 3.2.5.5. In the previous section, it was found that the +70 Da modification is possibly a CQA. This modification was therefore also included in the assay. Further, amino acid substitution from isoleucine to valine has been identified in the same batch as the +70 Da modification (data not shown). The source to the substitution could be traced to the oxygen depletion in the batch described above, leading to accumulation of valine residues and thereby to a shift in the translation equilibrium (79). The substitution of isoleucine to valine may have implications on the protein binding and folding and this kind of protein modification was therefore also included in the assay. Finally, asparagine deamidation and methionine oxidation were included in the assay since it has been well documented in literature that these two modification types may have several negative effects on therapeutic proteins, see chapters 3.1.4.4 and 3.1.4.5. These five types of modifications were targeted when a MAM for GM-CSF was optimized and are summarized in Table 7.

Table 7: Target attributes of the MAM optimized for GM-CSF.

	GM-CSF MAM target modification types
1	Disulfide scrambling and free cysteine residues
2	+70 Da modification
3	Isoleucine to valine amino acid substitution
4	Asparagine deamidation
5	Methionine oxidation

5.2.3 Optimization of a MAM sample preparation protocol for quantification of disulfide scrambling together with four other CQAs in rhGM-CSF

There are a few reasons for why disulfide scrambling is not so commonly included in MAMs. One reason is that common programs for processing of peptide mapping data have not been supporting searching for crosslinked peptides. This has, however, been largely overcome by the fact that many of these programs have started to introduce support for searching crosslinked data. This is for example the case for the database search engine Mascot and the software Skyline, useful for quantitative analysis of peptide mapping data. Another reason is that data from protein digests where the cysteine residues have not been reduced, will easily get very complex. Each cysteine residue may link to another cysteine residue. Hence, if a peptide contains several cysteine residues it may link to several peptides and can create very large peptide complexes. Moreover, it is harder to achieve efficient digests of proteins that have not been opened up by cysteine bond reduction. These peptide complexes may have too large m/z values to be covered by the peptide mapping method. Moreover, these complexes do often not ionize as efficiently as other peptides and it may therefore be difficult to identify all disulfide connections. To overcome this issue the assay optimization was largely focused on finding appropriate enzymes for the assay. The enzyme or enzyme combination should mainly generate peptides containing only one cysteine. If this was achieved, it should be possible to only search for disulfide linked peptides involving two peptides. The two linked peptides should also not be too large in order to be covered by the peptide mapping method.

GM-CSF contains four cysteine residues which are linked by site C54-C96 and site C88-C121 in the native protein, see Figure 31 for the amino acid sequence of rhGM-CSF.

APARSPSPSTQPWEHVNAIQEARRLLNLSRDTAAEMNETVEVISEMFDLQ
EPTCLQTRLELYKQGLRGSLTKLKGPLTMMASHYKQHCPPTPETSCATQI
ITFESFKENLKDFLLVIPFDCWEPVQE

Figure 31: Amino acid sequence of rhGM-CSF expressed in E. coli. Target amino acids are highlighted as follows. C = red, K = green, I = blue, N = yellow, M = purple

These cysteine sites should be separated into individual peptides by the assay. The assay should also cover the protein N-terminal, the lysine residues, the isoleucine residues, the asparagine residues and the methionine residues. The general workflow for enzyme selection for the assay was as follows. The

non-reduced GM-CSF was digested by the enzyme followed by enzyme inhibition and reduction of protein's cysteine residues using TCEP in a fraction of the digest. The reduced digest was then analyzed by LC-MS/MS by DDA. The data was searched for peptide identifications using the Mascot database search engine and the identified peptides were quantified in Skyline. If the digest was not found to generate homogenous pools of peptides containing only one cysteine for the four cysteine sites, then the digest protocol was discarded. Otherwise, the non-reduced digest was searched for possible disulfide scrambling events. In theory the more specific enzymes would generate more homogenous pools of peptides. GluC is a good candidate enzyme to use in the digestion protocol for GM-CSF. The enzyme is specific for digestion C-terminally to a glutamate residue and should theoretically generate peptides where all four cysteine residues are separated upon digestion with GM-CSF. The workflow described above was used to test the performance of GluC in the digestion protocol. The enzyme was found to generate close to homogenous peptide pools for all four cysteine sites. When the disulfide linked peptides were analyzed, however, the very long peptide containing site C88 (residues 61-93) was found to give very poor chromatographic resolution (data not shown). This digestion protocol was therefore discarded.

Other specific enzymes than GluC are for example Trypsin, LysC and AspN. None of these enzymes can separate the cysteines at the positions 88 and 96. Therefore a combination of enzymes was employed to trim down the long C88 peptide generated by digestion with GluC. The enzyme LysC was combined with GluC in an overnight digestion at physiological pH. This also generated quite homogenous peptide pools for all four cysteine sites, see Figure 32.

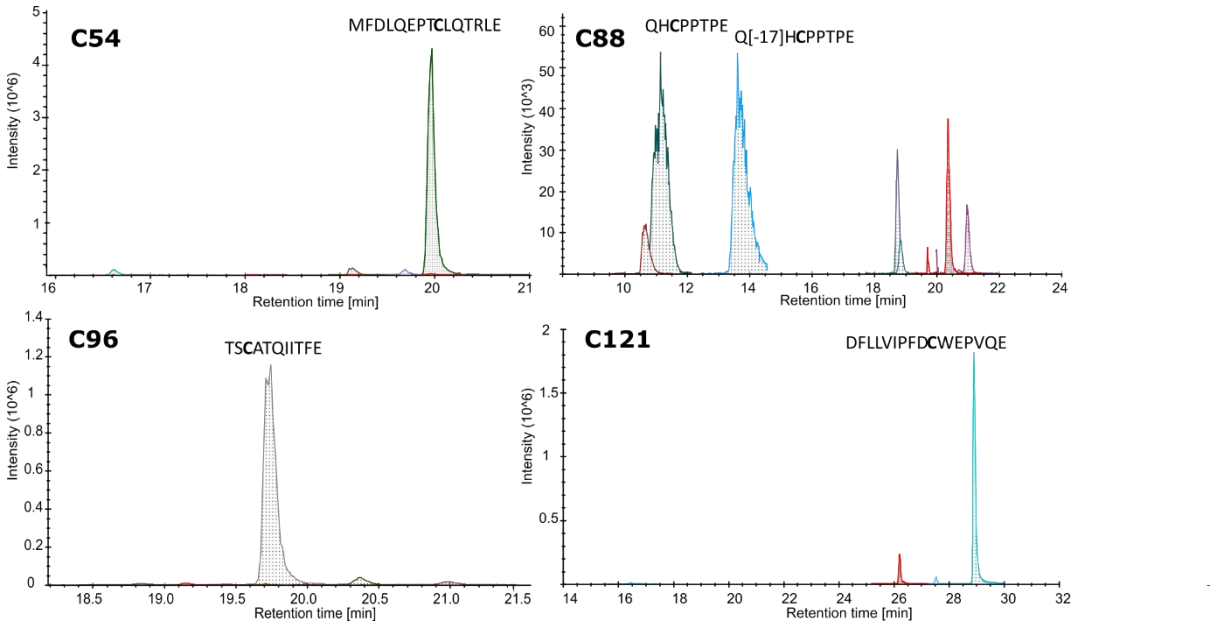


Figure 32: GM-CSF was digested non-reduced with GluC+LysC followed by enzyme inhibition using TFA. A fraction of the sample was reduced with TCEP and analyzed by LC-MS. The figure shows the precursor XICs from all the reduced cysteine containing peptides. The peptides are grouped according to the four cysteine sites of GM-CSF: C54, C88, C96 and C121.

The site C88 did yield the least homogenous pool of peptides, with the peptide QHCPPTPE, with and without N-terminal pyroglutamate formation, as the main peptide variants and some lower abundant peptides containing both the C88 and C96 site. The overall yield of cysteine separated peptides was however considered good. The assay was therefore evaluated for quantification of disulfide linked peptides. Since the presence of free cysteine residues is also an important protein modification, GM-CSF was first alkylated with iodoacetamide, to catch any free cysteine resides. This was followed by digestion using LysC and GluC which was terminated by lowering the pH, see chapter 7.4.4. A peptide

library containing possible disulfide linked peptides was manually generated in Skyline by combining the most abundant cysteine containing peptides discovered by the reduced assay. These disulfide-linked peptides were not identified by Mascot since Mascot had not added support to search for disulfide liked peptides yet at this point. Peptide variants with an alkylated cysteine were also included in the library.

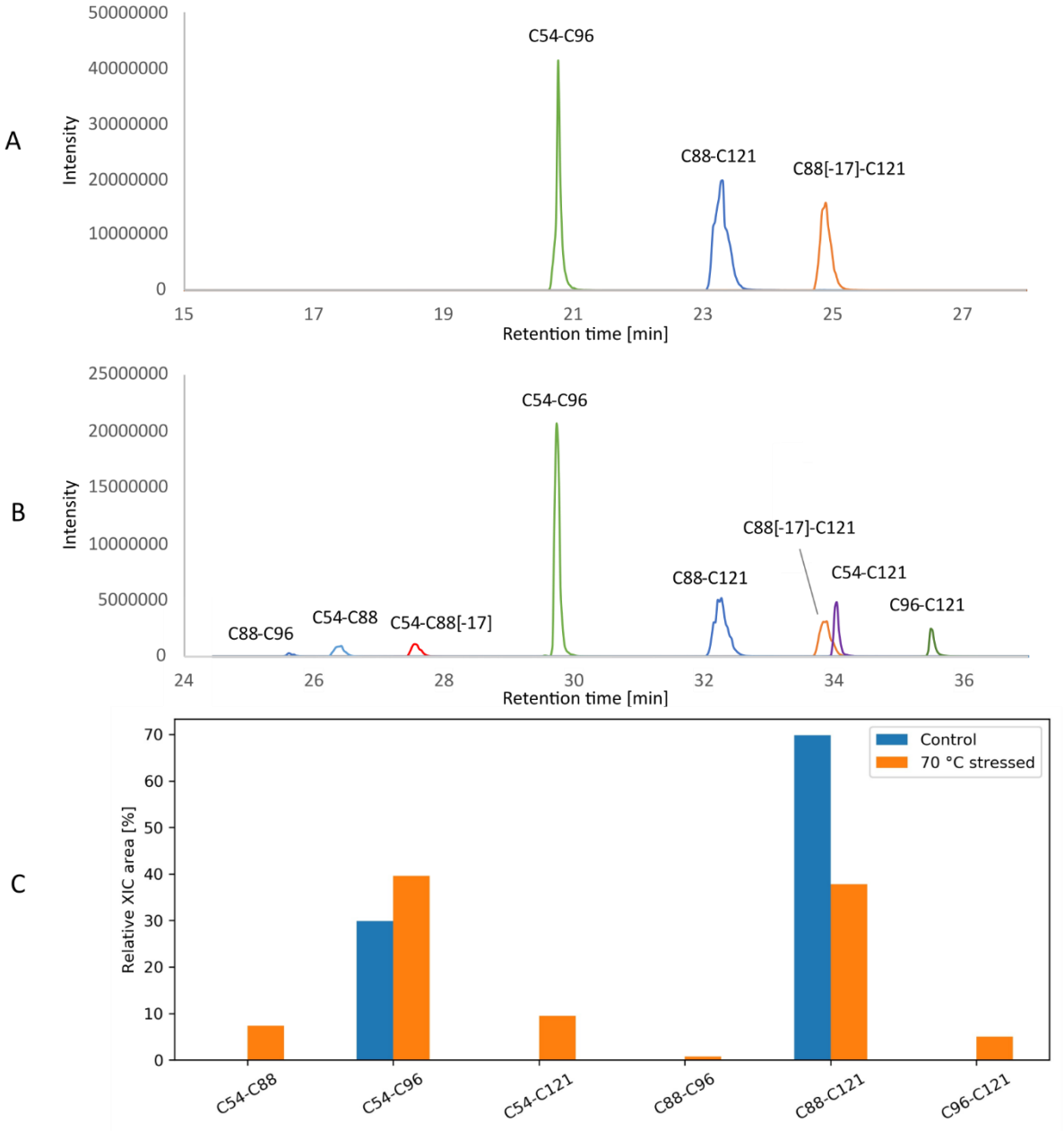


Figure 33: The levels of disulfide scrambling were monitored in GM-CSF using the LysC + GluC based MAM assay. A spectral library was built in Skyline containing all disulfide linked peptide combinations based on the most abundant peptides from the reduced digest. Digests of GM-CSF, (a) untreated and (b) stressed at 70 ˚C for 3h, were searched using the library and the precursor XICs were overlaid. (c) The relative abundances of the crosslinked events were quantified and plotted. Quantification was performed by summing the precursor XIC areas from the peptides containing the linked event of interest and dividing that by the total precursor XIC area of identified disulfide-linked peptides.

The precursor data from the non-reduced digest was then searched against this library. The data from the reduced digest was used to confirm the identifications of the disulfide linked peptides in the non-

reduced digest, see chapter 7.4.10 for the disulfide bond reduction protocol. The precursor XICs from the identified disulfide linked peptides are shown in Figure 33a. The two expected disulfide linked events, C54-C96 and C88-C121, were identified in the GM-CSF sample. The peptide containing site C88 was found both with and without N-terminal pyroglutamate formation. The same sample was then stressed to induce disulfide scrambling for 3 h at 70 ˚C and analyzed using the same workflow, see chapter 7.4.3. All possible disulfide scrambling events could be identified in the sample, as shown in Figure 33b. The relative amount of the different disulfide events as well as free cysteines were then quantified in the two samples. This was done by summing all precursor XIC areas from one disulfide event and then dividing this number by the sum of all precursor XIC areas from all disulfide-linked peptides as well as peptides with an alkylated cysteine. See chapter 7.4.12.1 for a detailed description of the disulfide quantification procedure. The result can be seen in Figure 33c. Only the two expected disulfide events were found in the control sample, and these had an intensity distribution of 30:70. In the stressed sample the main disulfide events were the expected ones but the scrambled events C54-C88, C54-C121 and C96-C121 measured close to 10% by the assay. The free cysteine levels measured in the samples were not over 1% (data not shown).

The MAM assay was then tested on the other modification types that the assay should target. The degree of +70 Da modification at the lysine sites and the protein N-terminal was quantified in two GM-CSF samples. The samples were the methionylated early process sample, which was found to contain a high degree of +70 Da modification in chapter 5.2.1 (Figure 26b), and a control sample, which was a late process sample of GM-CSF with no or very low degrees of the +70 Da modification as determined by intact LC-MS analysis. The modification degree was quantified in the reduced sample by summing the precursor XIC areas from all peptides containing a modification at the site of interest and dividing that by the total precursor XIC area from all peptides containing the site, see chapter 7.4.12.3. Figure 34a, blue bars, shows relative quantification of the +70 Da modification in the early process sample (left) and the control sample (right). The data is also presented in Appendix table 4.

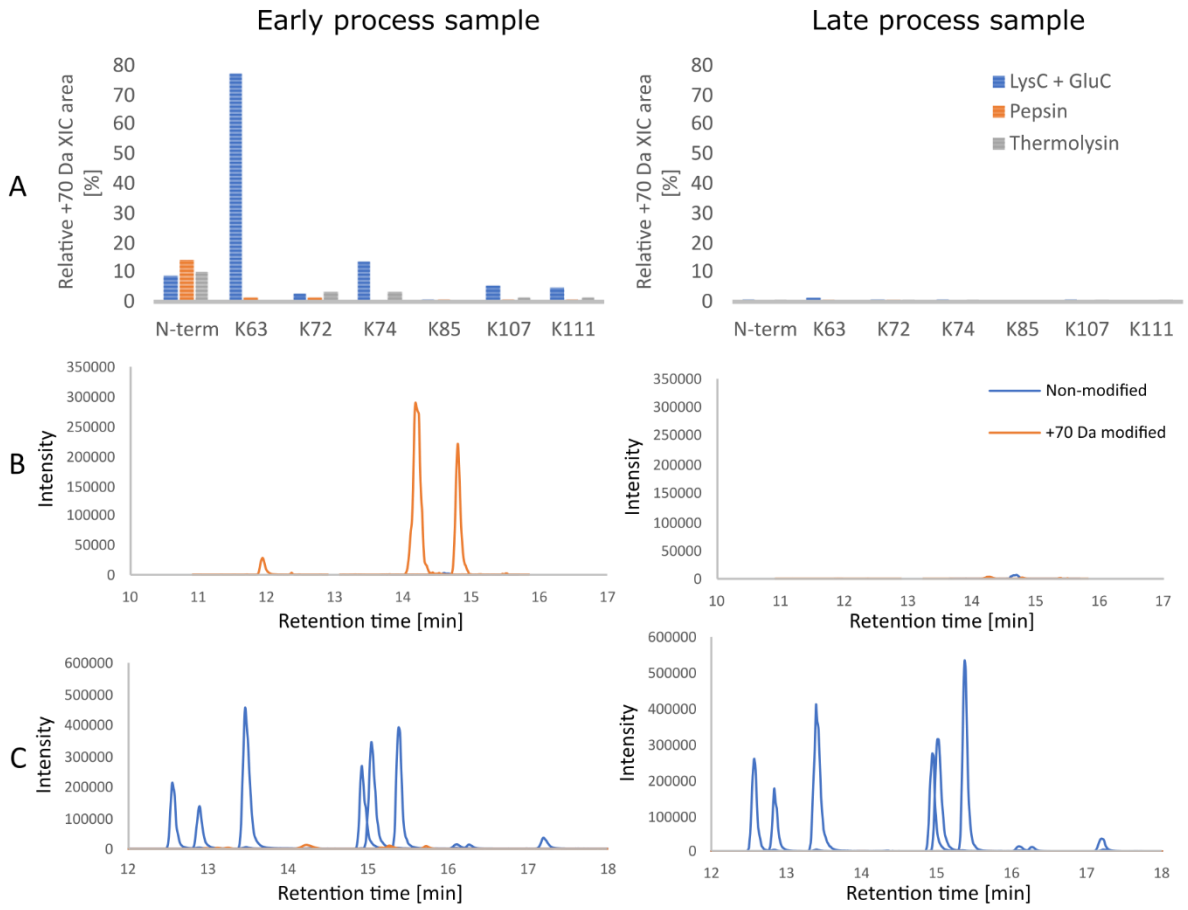


Figure 34: (a) The MAM assay based on LysC + GluC digestion was evaluated for quantification of +70 Da modification degree in GM-CSF. The assay was tested on the methionylated early process sample with +70 Da modification from chapter 5.2.1 and on a GM-CSF control sample, which was a late process sample of GM-CSF with no or very low degrees of the +70 Da modification determined by intact LC-MS analysis. The assay values were compared to similar assays based on the enzymes Pepsin and Thermolysin, to evaluate if the quantitative values did agree. The precursor XICs are shown for all peptides containing site K74 (b) from digestion with LysC + GluC and (c) from digestion with Pepsin, color code by if they did contain a +70 Da modification or not.

The control sample did contain a very low modification degree as measured by the assay, with the highest degree being 1.8% modification of site K63. Some sites in the early process sample were found to contain very high modification degrees by the assay. These were for example the protein N-terminal, site K63 and site K74, with 9%, 77% and 14 % modification degree, respectively. Based on the analysis by intact mass from Figure 26b, the total modification degree should not be more than around 10-20%, by visual inspection. The assay based on digestion by GluC and LysC was therefore benchmarked against two other peptide mapping protocols, using Pepsin or Thermolysin respectively, see chapters 7.4.5 and 7.4.6. These two protocols gave about similar quantitative results when applied on the early process sample, which was around 10% modification of the protein N-terminal and a few percentages modification of most of the six lysine residues, see Figure 34a, red and grey bars. The conclusion was that the assay by GluC and LysC overestimated the modification degree of the lysine residues K63, K74, K107 and K111. The peptides used for quantification of the lysine sites by the three assays were studied closer to investigate the background to the overestimation. Figure 34b shows the XIC area from all peptides containing site K74 from digestion with GluC and LysC, identified in the early process sample, left, and the control sample, right. The blue chromatogram shows the total intensity

from peptides without the +70 Da modification and the red chromatogram from peptides with the modification. Almost all the XIC area in the early process sample come from peptides with a +70 Da modified lysine residue. In the control sample, which should contain a very low degree of +70 Da modification, the total XIC area from this site is much lower than in the early process sample. Figure 34c shows the XIC area from all peptides containing site K74 from digestion with Pepsin identified in the early process sample, left, and the control sample, right, categorized by non-modified peptides (blue) and +70 Da modified peptides (red). In both samples the majority of the total XIC area comes from non-modified peptides. The total XIC area is also similar in the two process samples. Site K74 is located just two amino acids C-terminally from site K72. That means that all K74 residues will be present in a two amino acids long peptide if the digest is complete by digestion with LysC, and will not be detectable by LC-MS/MS. If the K74 residue is modified there is a high possibility that LysC will not recognize the amino acid and hence will not be able to digest. The result is that the majority of the peptides available to use for quantification of the modification degree will be peptides with a modified lysine residue, which, most possibly, will result in inaccurate quantitative results. The same reasoning could be used to understand why the modification degree was overestimated on the other lysine sites as well. The conclusion was that a lysine specific enzyme should not be used for quantification of modifications located on lysine residues.

Other enzymes specific for one or two amino acids were explored that could trim down the peptide containing site C88 from the GluC digestion to a shorter peptide, but no enzyme with a suitable specificity was found. Therefore, enzymes with lower specificity were explored using the same strategy as for the GluC + LysC assay described above. Appendix table 4 summarizes all enzymes and digestion protocols investigated. Chymotrypsin is an enzyme that hydrolyzes peptide bonds C-terminally to the aromatic amino acids tyrosine, phenylalanine and tryptophan specifically, but has also a lower specificity for C-terminal digestion of leucine, methionine, alanine, aspartate and glutamate residues. It was found that by combining GluC with Chymotrypsin in a 2 h digestion, the peptide containing site C88 could be further trimmed down. Figure 35a shows the precursor XICs from all peptides containing a cysteine residue from a digest of GM-CSF with GluC and Chymotrypsin, after disulfide bond reduction. The peptides are grouped by cysteine site. The peptide pools from all four sites seemed to be quite homogenous, with one peptide species containing one single cysteine mainly representing each of the sites. To test the assay's ability to quantify disulfide linked peptides and free cysteine residues the GM-CSF sample was alkylated with iodoacetamide and then digested with GluC and Chymotrypsin. The digestion was terminated with 1% TFA and analyzed by LC-MS/MS in DDA mode. See chapter 7.4.8 for a protocol description. The data was then searched for peptide identifications through the Mascot databased search engine using a method including disulfide linked peptides. This was followed by a quantitative data search in Skyline using a library created manually from the Mascot search results. The identities of the disulfide linked peptides were confirmed by comparison with the reduced sample. Quantification of the degree of disulfide scrambling and free cysteine residues was then performed by summing all precursor XIC areas corresponding to each disulfide linked or free cysteine event and dividing that area with the precursor XIC area from all disulfide linked and alkylated cysteine peptides, see 7.4.12.2.

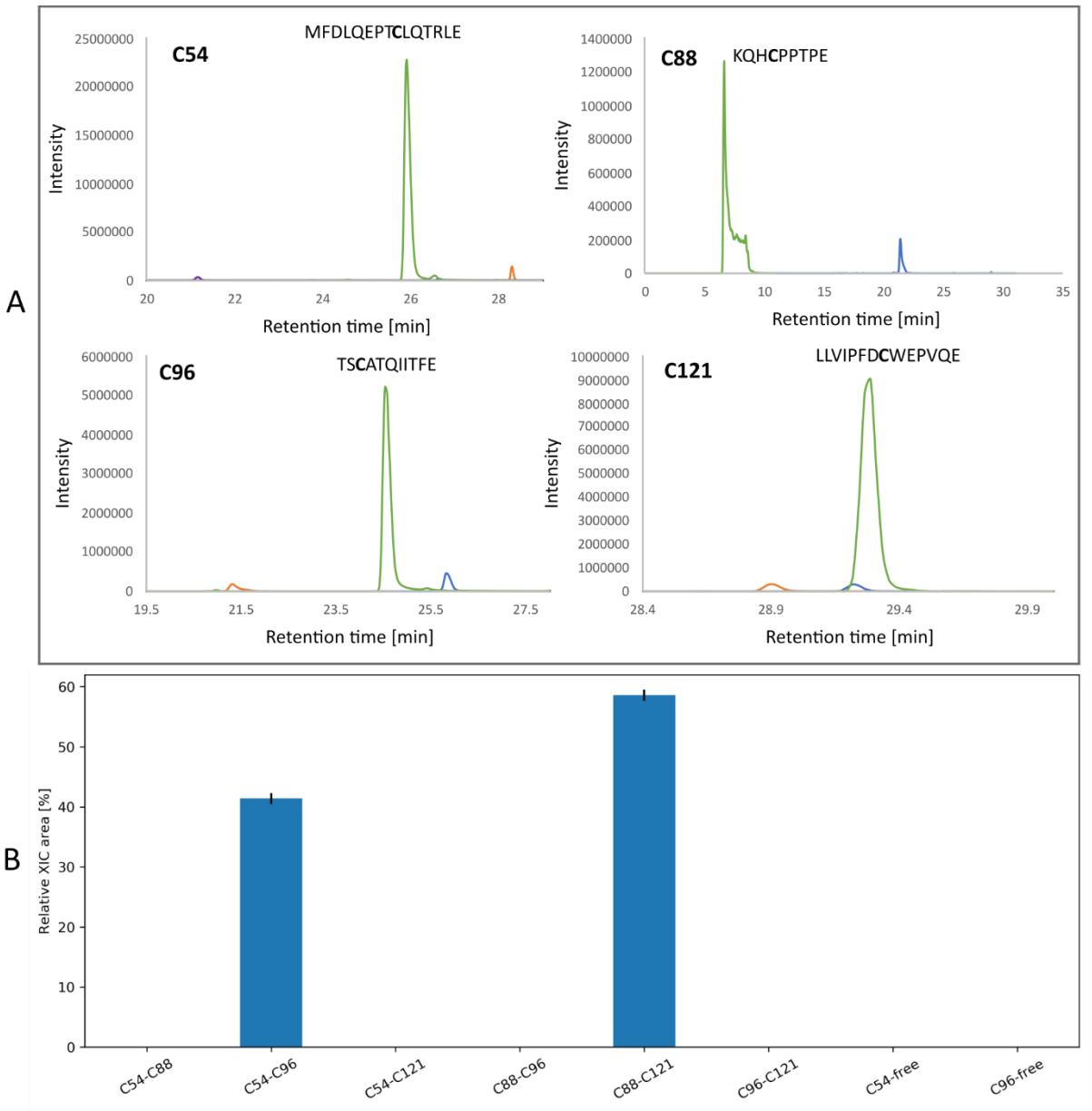


Figure 35: A GM-CSF control sample was digested non-reduced with GluC + Chymotrypsin followed by enzyme inhibition using TFA. A fraction of the sample was reduced with TCEP and both the non-reduced and reduced sample were analyzed by RP LC-MS by DDA. (a) The precursor XICs from all the reduced cysteine containing peptides. The peptides are grouped according to the four cysteine sites of GM-CSF: C54, C88, C96 and C121. (b) Quantification of the different scrambling events in the sample, bars showing mean values and error bars the standard deviation, N = 2. Quantification was performed by summing the precursor XIC areas from the peptides containing the linked event of interest and dividing that by the total precursor XIC area of identified disulfide-linked peptides. Disulfide linked peptides were identified using the Mascot databases search engine and the XIC areas were calculated by Skyline using a spectral library based on the search in Mascot.

Figure 35b shows the degree of the different disulfide linked events and free cysteine residues in the GM-CSF sample. Only the two expected disulfide linked events, C54-C96 and C88-C121, could be identified. The relative intensity distribution between these disulfide-linked events was ~40% of the C54-C96 event and ~60% of the C88-C121 event. Ideally this distribution should be 50:50 if the assay was truly quantitative, since the protein should contain the same number of cysteines from all four sites. It can however not be assumed that this will be the case when quantification is performed based on mass spectrometry data, since different peptides will ionize differently. The fact that the intensity distribution was 40:60, which is close to 50:50, however, indicated that the assay may give good

quantitative results. The assay based in GluC and Chymotrypsin was therefore evaluated for quantification of the targeted modification types and amino acid sites using different representative samples, as will be described in the next section. An overview of the MAM based on GluC and Chymotrypsin digestion is described in Figure 36.

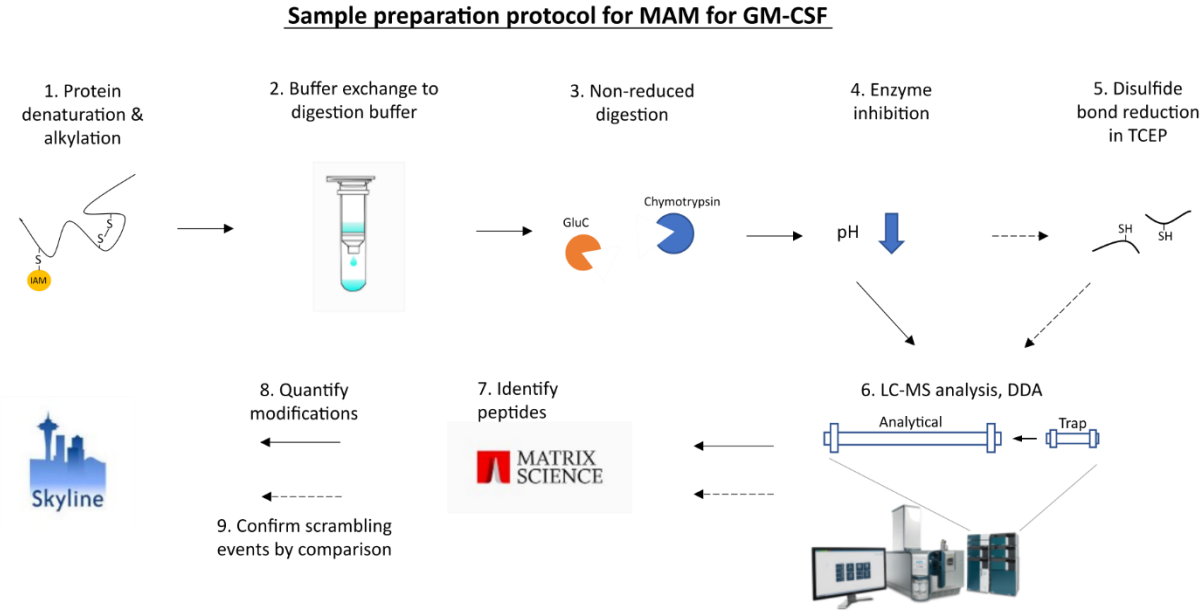


Figure 36: Scheme of the workflow for the MAM optimized for GM-CSF. 1) The protein was denatured and unbound cysteines were alkylated, 2) the buffer was changed to digestion buffer, 3) the non-reduced protein was digested with the enzymes GluC and Chymotrypsin and 4) the enzyme activity was inhibited. 5) A fraction of the sample was then reduced and 6) both samples were analyzed by LC-MS by DDA. 7) The peptides were identified in the non-reduced sample using the Mascot database search engine (Matrix Science) and 8) a library was created in Skyline from these results. 9) The non-reduced and reduced data was searched in Skyline and the relative modification degrees at the target amino acid sites were quantified from the precursor XIC areas, as described in section "Methods". The reduced data was used to confirm disulfide events in the non-reduced data.

5.2.4 Evaluation of a MAM for GM-CSF based on the enzymes GluC and Chymotrypsin

The MAM based on GluC and Chymotrypsin digestion described in Figure 36 was first evaluated for quantification of disulfide scrambling and free cysteine events in seven samples, see Figure 37. The analyzed samples were 1) a GM-CSF control sample, which was a late process sample of GM-CSF with no or very low degrees of the +70 Da modification determined by intact LC-MS analysis, 2), 3), 4) the GM-CSF control sample which had been subjected to forced degradation in ammonium bicarbonate buffer with pH 8.3 at 40 ˚C for 1, 2 and 3 weeks 5) the methionylated early process sample of GM-CSF in which the +70 Da modified protein was identified in chapter 5.2.1, 6) the early process sample which had been incubated at 70 ˚C for 3 hours to induce disulfide scrambling and 7) a refolded product of GM-CSF stored in guanidinium hydrochloride and DTT and which contained around 50% host cell proteins (HCP)s (data not shown). Sample descriptions can be found in chapters, 7.4.1, 7.4.2 and 7.4.3. All samples but the GM-CSF stressed for 3 h at 70 ˚C were analyzed in two replicates. The spectral library built in Skyline, which was used for quantification of the disulfide scrambling events, was based on the search results from all the analyzed samples when searched using the Mascot search engine, see chapter 7.4.12.2.

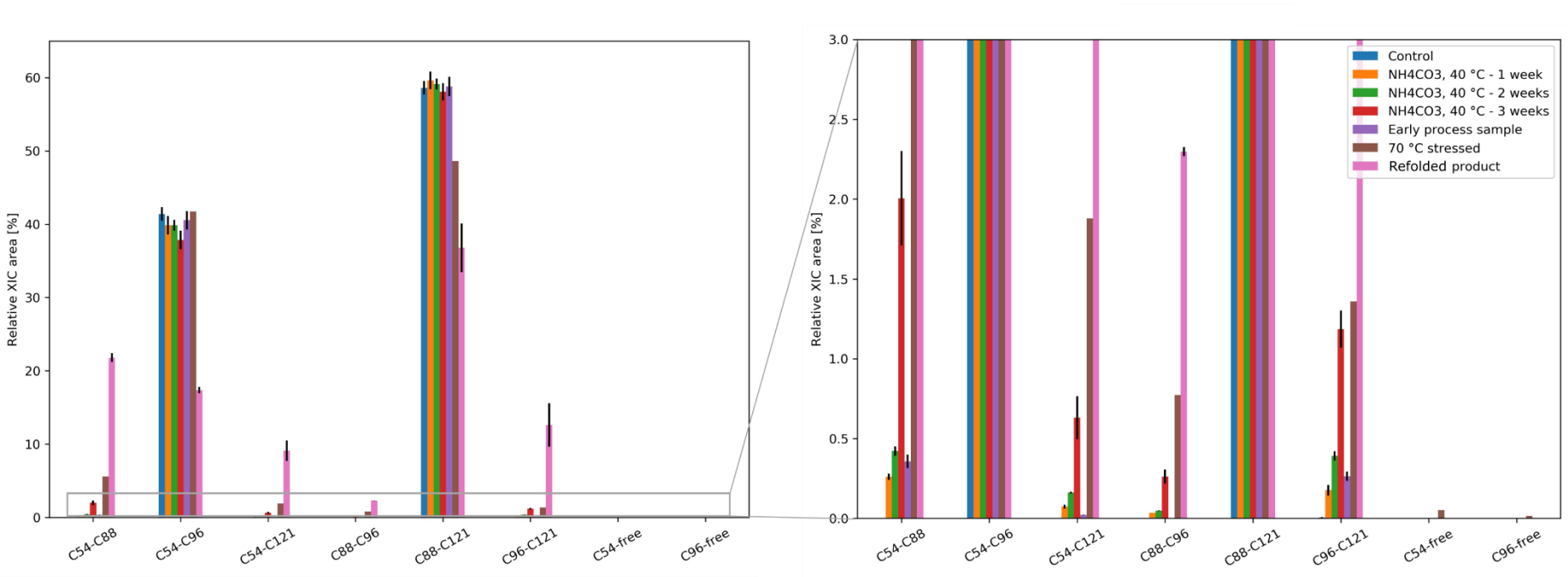


Figure 37: The levels of disulfide scrambling, and free cysteines were monitored in seven different GM-CSF samples using the MAM optimized for GM-CSF. The samples monitored were 1) a control sample, 2, 3) ,4) the control sample stored in NH_4CO_3 buffer and 40 ˚C for 1, 2 or 3 weeks, 5) the early process sample from chapter 5.2.1 with +70 Da modification, 6) the early process sample stressed at 70 ˚C for 3 h and 7) a refolded product stored in DTT. A library built from the identified disulfide scrambling events when all seven samples were searched in the Mascot database search engine was used for quantification of the disulfide scrambling. All values besides those from the 70 ˚C stressed sample were based on two replicate digests.

Quantification of the scrambling events using the assay showed that the control GM-CSF sample contained no detectable disulfide scrambling or free cysteine residues before treatment. However, when the sample was subjected to forced degradation in ammonium bicarbonate at 40 ˚C the scrambling increased successively over the three weeks. The scrambling event that increased most under the stressing conditions was the linkage C54-C88, which measured 0.3% in GM-CSF stressed for one week and 2% in GM-CSF stressed for 3 weeks. It was not unexpected that scrambling did increase in this sample since both high pH and elevated temperatures are known factors that may increase disulfide scrambling in protein samples. When disulfide scrambling was measured in the early process sample a few scrambling events could be detected with low intensity. These were mainly linkages C54-C88 and C96-C121 and measured 0.4% and 0.3% respectively. When this sample was stressed at 70 ˚C all four possible disulfide scrambling events were detected in the sample. The main event was linkage C54-C88, which was present in 6%, followed by linkages C54-C121 and C96-C121 which were present in 1.9% and 1.4%. The sites C54 and C96 were also identified as alkylated peptides, representing free cysteine residues, in quantities 0.05% and 0.02% respectively. This indicates that it is not so favorable for the cysteines in the GM-CSF sequence to appear in its unbound form. Finally, a refolded product of GM-CSF which had been stored in guanidinium hydrochloride and DTT was analyzed. This sample contained high levels of scrambling with the linkage C54-C88 being present in 22%, linkage C96-C121 in 13% and C54-C121 in 9%. This high degree of scrambling was expected since the sample had been stored at reducing conditions and so the disulfide bonds had not been reformed during controlled conditions. No signal interference from the HCPs were identified in the precursor ion spectra from this sample, despite the HCP degree being around 50%, which is considered high. Therefore, no mass spectrometry method based on DIA was introduced into the MAM but instead the DDA based method was used. Overall, the degrees of disulfide scrambling detected in the seven GM-CSF samples were expected. The refolded product, which had been stored at reducing conditions, contained the highest degree of scrambling, followed by the sample stressed at 70 ˚C and then the samples stressed over several weeks at high pH and 40 ˚C. The disulfide scrambling event C54-C88 seemed to be the most favorable event while C88-C96 seemed to be the least favorable event. This may however also be explained by the first peptide combination being more efficiently ionized in the mass spectrometer than second peptide combination. Free cysteine events were only detected in the sample stressed at 70 ˚C and in levels under 0.1%. In summary the MAM including quantification of disulfide scrambling was found to give predictive quantitative results for the GM-CSF samples tested by the assay.

The MAM was then used to follow the formation of asparagine deamidation in the three samples subjected to forced degradation described above. The control GM-CSF sample had been incubated in ammonium bicarbonate with pH 8.3 at 40 ˚C for 1, 2 or 3 weeks followed by buffer exchange before applying the MAM, see chapter 7.4.2. The peptide with the highest precursor XIC area for each asparagine site from the digest was used for quantification. See Table 8 for the amino acid sequences. Quantification was performed by calculating the XIC area ratio between the modified variant of the peptide and the modified plus the non-modified variant of the peptide. The value from the control sample was then subtracted from the measured modification degree, see chapter 7.4.12.3. Figure 38a shows how the successive deamidation of the four asparagine sites in GM-CSF could be followed over the three weeks of forced degradation using the MAM assay.

Table 8: Peptides used for quantification of deamidation, +70 Da modification, Ile to Val substitution and Met oxidation. The "amino acid site" corresponds to the amino acid site for which the modification was quantified. The "amino acid sequence" shows which modified peptide that was used to quantify the modification degree of that specific site together with the same peptide without the modification. A hyphen corresponds to a disulfide bond.

Amino acid site	**Amino acid sequence**
N17	APARSPSPSTQPWEHV**N[+1]**AIQE
N27 or N37	**N[+1]**LSRDTAAEMNE & NLSRDTAAEM**N[+1]**E
N109	SFKE**N[+1]**LKDF
N-terminal	**M[+70]**APARSPSPSTQPWEHVNAIQE & **A[+70]**PARSPSPSTQPWEHVNAIQE
K63	LY**K[+70]**QGLRGSL
K72 or K74	T**K[+70]**LKGPLTMMASHY & TKL**K[+70]**GPLTMMASHY
K85	**K[+70]**QHCPPTPE-LLVIPFDCWEPVQE
K107 or K111	SF**K[+70]**ENLKDF & SFKENL**K[+70]**DF
I19	EHVNA**V**QE
I43	NLSRDTAAEMNETVEV**V**SE
I100 or I101	MFDLQEPTCLQTRLE-TSCATQ**V**ITFE & MFDLQEPTCLQTRLE-TSCATQI**V**TFE
I100 & I101	MFDLQEPTCLQTRLE-TSCATQ**VV**TFE
I117	KQHCPPTPE-SFKENLKDFLLV**V**PFDCWEPVQE
M36	NLSRDTAAE**M[+16]**NE
M46	**M(+16)**FDLQEPTCLQTRLETSCATQIITFE
M79 or M80	KGPLT**M[+16]**MASHY & KGPLTM**M[+16]**ASHY

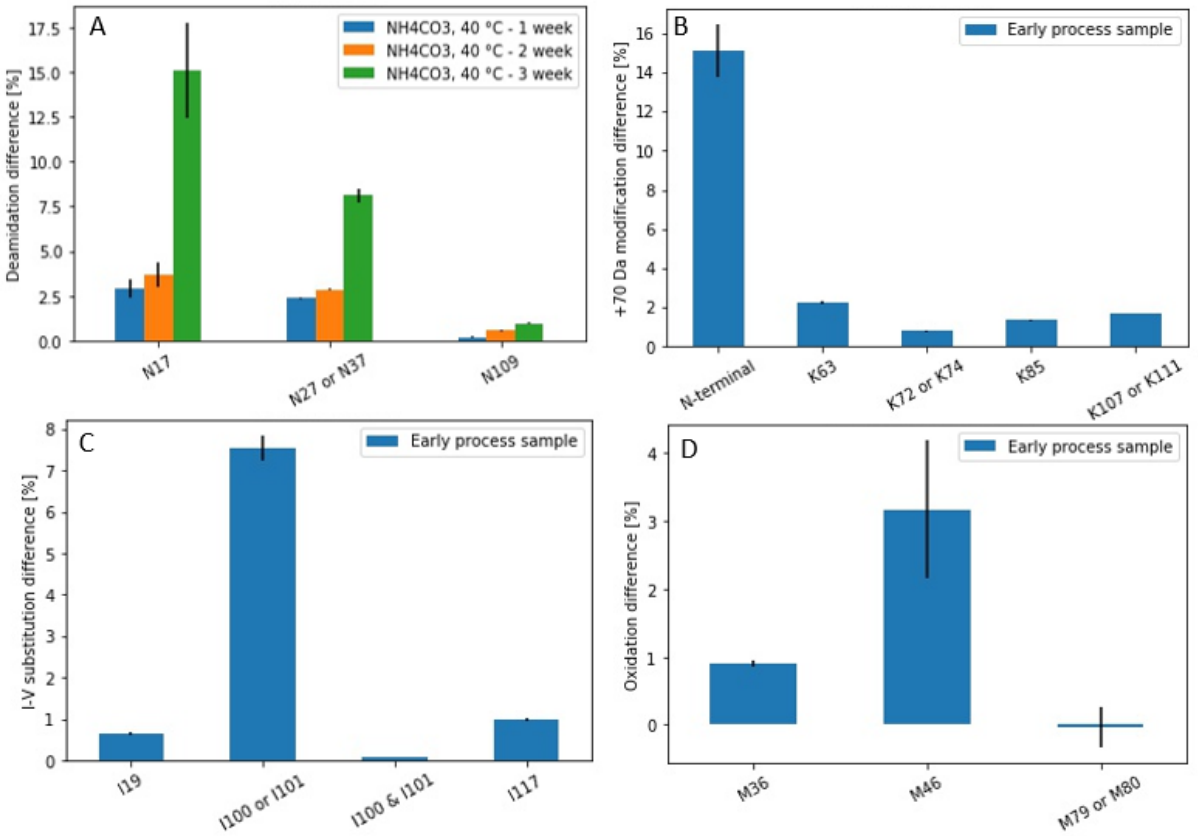


Figure 38: The MAM assay optimized for GM-CSF was tested for quantification of four CQAs. (a) Deamidation degree was quantified in a GM-CSF control sample subjected to forced degradation in NH_4CO_3 buffer at 40 ˚C for 0, 1, 2 or 3 weeks. The degree of (b) +70 Da modification, (c) Ile to Val substitution and (d) Met oxidation was quantified in the early process sample from chapter 5.2 and in the control sample. Quantification of the modification degree in the samples was performed based on the precursor XICs from the peptides shown in Table 8. The modification difference was then calculated compared to the control sample. Mean values and bars for standard deviation are based on modification differences from replicate digests (a-c: N = 2, d: N = 3). MS data was acquired by DDA.

Site N17 seemed to be most susceptible to deamidation with 2.9% deamidation in the 1-week sample and 15% deamidation in the 3 weeks stressed sample. The sites N27 and N37 were present in the same peptide and were together 8.1% deamidated after 3 weeks of stressing. Site N109 seemed to be least susceptible with 1.0% deamidation after 3 weeks stressing.

The degree of +70 Da modification on the protein N-terminal and on lysine residues were then monitored using the MAM. The modification degree was monitored in the early process sample containing +70 Da modified protein, which was characterized in chapter 5.2.1, and the GM-CSF control sample. The peptides to use for quantification were selected in the same way as for the deamidation quantification, choosing the peptides with the highest XIC areas for each amino acid site. The values from the control sample were then subtracted from those from the early process sample, in order to identify any differences in modification degrees. See Table 8 for the amino acid sequences. Both the N-terminal peptide variants with and without the methionine were included in the assay since the early process sample contained a high degree of protein with non-cleaved N-terminal methionine. Site K85 was monitored using the disulfide bound peptide [KQHCPPTPE-LLVIPFDCWEPVQE]. Figure 38b shows the quantification of the +70 Da modification degree on the protein N-terminal and on the six lysine residues in the early process sample. The modification degrees measured by the MAM were comparable to those measured by the assays based on Pepsin and Thermolysin. The degree of modified protein N-terminal in the early process sample was 15% as measured by the MAM. This can

be compared to the levels measured by the Pepsin based assay, 14%, and the Thermolysin based assay, 11%, see Figure 34a. The lysine sites were found to be modified in levels of 0.8%-2.3% as measured by the MAM. This included also the disulfide linked peptide containing site K85, which was found to contain 0.3% +70 Da modification. This could be compared to the modification degrees measured by the Pepsin based assay, 0-1.9%, and by the Thermolysin based assay, 0-2.2%. It is difficult to compare the modification levels of the exact sites directly between the three assays since some sites were quantified using the same peptide and these were not the same in all three assays. For example, the sites K72 and K74 as well as sites K107 and K111 were quantified using the same peptide in the developed MAM. The modification degree in the control sample was not above 0.1% for any of the sites as measured by the MAM assay (data not shown).

The amino acid substitution from isoleucine to valine residues was then quantified using the MAM. The MAM was tested on the early process sample from chapter 5.2.1, containing proteoforms modified with the +70 Da modification and in which the amino acid substitution had been identified, and on the GM-CSF control sample. Peptides were selected for quantification and the modification degree was quantified using the same strategy as for as for quantification of deamidation and the +70 Da modification. From the five isoleucine sites, three were quantified from disulfide linked peptides, namely sites I100, I101 and I117, see Table 8 for the amino acid sequences. In the early process sample the amino acid substitution was highest for the peptide containing both site I100 and I101, measuring around 7.5%, see Figure 38c. This was followed by the substitution of site I117 and site I19, measuring 1% and 0.7% respectively. No amino acid substitution could be identified at site I43. The intensity of the non-modified peptide containing this site was however lower than the intensity of the peptides covering the other sites, reducing the assay sensitivity for this site. In the control sample the highest degree of amino acid substitution was 0.1% identified in site I100/ I101 (data not shown).

Finally, the degree of modified methionine residues was quantified in the GM-CSF early process sample and the control sample. The methionine oxidation was quantified in three replicate digests since sample preparation induced oxidation was found to vary highly based on the results from the IgG MAM from chapter 5.1.6. The quantification was performed using the same strategy as for quantification of deamidation, +70 Da modification and isoleucine to valine substitution. From the four methionine sites, sites M36 and M46 were found to be more oxidized in the early process sample than in the control sample, see Figure 38d. This conclusion was drawn from the fact that the standard deviation of the oxidation difference between the samples was smaller than the mean oxidation difference for these sites. The site that was found to be most oxidized was site M46, with an oxidation degree difference of 3.2% compared to the control sample. This was also the only site quantified from a disulfide linked peptide, see Table 8 for the amino acid sequences. Site M36 was found to be 0.9% more oxidized than the control sample while no difference could be observed for site M79/ M80 between the two samples.

6. Discussion

The overall aim of this thesis was to explore if common MAMs may be improved by using alternative mass spectrometry acquisition strategies or digestion protocols. This aim was targeted through three different projects.

6.1 Influence of DIA on MAM accuracy and linearity

When quantifying proteoforms in complex samples using DDA based MAMs, signal interference and identification of low abundance proteoforms may be an issue. In the first project it was therefore asked if a MAM based on DDA may be exchanged for a method based on DIA without losing assay

precision and linearity, see chapter 5.1. Protein deamidation, oxidation and N-glycosylation was quantified in a biosimilar of the therapeutic IgG adalimumab using bottom-up proteomics workflows based on DIA and on DDA. The precisions of the two methods were fist investigated by replicate analyses of a cell culture filtrate of the IgG using CV between replicate analyses as metrics. The precision was investigated on three replicate levels: technical replicates acquired the same day, technical replicates acquired different days and replicate digests. The limit of quantification when using MAMs has previously been defined as the minimum modification degree with a CV < 0.1 (56,80). A modification degree cutoff value of 2% was applied in this study for comparison of the precision at the three replicate levels which resulted in all technical replicates having a CV < 0.1. The DIA and DDA based MAMs did perform similarly with regard to precision between technical and biological replicates of the cell culture filtrate, applying this cutoff level (Figure 19). Technical replicates showed CVs < 0.05 for all identified modification events over this cutoff value when acquired on the same day and CVs < 0.1 when acquired on different days. Biological replicates showed CVs < 0.1 for deamidation events and N-glycan variants and CVs < 0.2 for oxidation events. These three replicate levels illustrated nicely the increasing degree of variation that is induced in the analytical results both on the instrument level, when the time and number of samples analyzed between two samples increases, and when the biological variation from the digest is introduced. The high CVs from quantification of the oxidation degree between the biological replicates was believed to arise from a higher variation in the degree of sample preparation induced oxidation, which will be further discussed in chapter 6.4. It was however still possible to compare the assay precision between the DIA and DDA based methods from the technical replicates and the quantification of oxidation was not found to behave any differently than quantification of the other modifications targeted by the method.

For most modification events the quantitative results obtained by the DIA based method did not differ by more than 20%, compared to the DDA based method (Figure 17). However, the deamidation degree of site HC_N77 was found to be overestimated about twice by the DDA based method because of signal interference in the precursor spectra. This could be resolved by applying the DIA based method and basing the relative quantification on the XICs from MS/MS fragmentation instead of XICs from the precursor ions and resulted in a deamidation degree of 1.6% instead of 3.7%. When the two methods were applied also on deamidation events measuring under 2% in the cell culture filtrate, two deamidated peptides were identified which had signal interference in their precursor spectra. These also showed higher CVs when quantified by the DDA based method than by the DIA based method. DIA based MS methods have been reported to provide good assay precision in proteomics studies (81). In a study by Gillet *et al.* the sensitivity of quantifying peptides spiked into a yeast digest using the precursor or fragment ions acquired by a DIA (SWATH-MS) based MS method was compared (82). It was found that about half of the spiked in peptides showed higher sensitivity when quantified by the fragment ions. In our study we found tendencies to a similar phenomenon, as described above. However, not enough peptides with signal interference were identified to be able to determine a clear trend.

When the deamidation degree was quantified in the purified and stressed IgG some low abundant deamidated peptides showed instead higher CVs when measured by the DIA based MAM than by DDA (Figure 22). For example, the peptide with deamidation at site HC_N425, measuring 2.4% by the DDA based assay, had a twice as high CV by the DIA based method (0.20) compared to the DDA based method (0.10), (Figure 22). The lower assay precision of the DIA based method was suggested to be related to a lower assay sensitivity of the MS/MS fragment signal than that of the precursor signal when precursor signal interference was not an issue. The precision achieved with the DDA based strategy in the purified sample was similar to what has previously been described by Zhang et al. using a similar acquisition strategy and instrument (80), resulting in CVs from biological replicates under or

around 0.1 for modification events measuring over 1%, see Appendix table 3. Similar accuracies have been reported by other studies with other instrument setups (57,83). A CV cutoff of 2% was used for comparison of technical and biological replicates in this current study and was chosen so that the CVs from technical replicates of the cell culture filtrate would be < 0.1, as was mentioned above. If a cutoff level of 1% had been applied, the deamidation site HC_N319_iso2, which had a mean deamidation degree of 1.6% by the DDA based strategy in technical replicates of the cell culture filtrate analyzed on different days, would for example also have been included in the comparison. This site had a CV of 0.04 by the DDA based strategy and 0.14 by DIA. The lower performance of the DIA based approach with regard to precision could be improved by improving the peptide signal. For example, the peptide containing deamidation site HC_N425 had a poor peak separation between the deamidated and the non-modified species and consequently a low peptide signal (Figure 21). The peak separation may be improved by using a column with a chemistry that gives a better peak separation. In this study an LC gradient of 68 minutes was applied in order to achieve a good peak separation between deamidated and non-deamidated peptides. An analytical column with a CSH chemistry was found to give good peak separation with a 28 min gradient and also allowed for higher sample loading than 1 µg. If, for example, a suitable trap could be found for this column, this may improve the total assay sensitivity. Another alternative could be to use a mass spectrometer with a built-in switch between waste and injection and thereby circumvent the need to use a trap. This may, however, increase the wear on the analytical column. A better peak separation could also open up for shorter runtimes, which would be favorable both from an economical perspective and from an assay sensitivity perspective. It could be seen that the XIC signals from some of the low abundant peaks were not very stable (Figure 21). This may be helped by running the assay in a normal flow setup (84), instead of the microflow setup that was employed in this work. A suitable assay precision test should, however, always be performed in connection with running the assay in order to know what precision to expect, since the daily form of the instruments and reagents may change.

Quantification of the deamidation degree at the site HC_N74, isoform 1 and 2, showed much higher values in Figure 20a, around 20% and 35% respectively, than in Figure 11, where both isoforms showed around 1% deamidation. An explanation to this could be that the sample mass that the assay was applied on was different in the two figures. The samples used for Figure 20a contained a total IgG mass of 10 µg while the samples used for Figure 11, contained a total mass of 40 µg. One theory for why the samples showed different deamidation degrees is that the quantification is based on the peptide *FTISRDN[+1]AK,* containing one missed cleavage at the N-1 position from an aspartate residue. This residue may have isomerized to isoaspartate in the 2 weeks stressed sample, making the site less prone to digestion by Trypsin. If the digestion conditions ended up being somewhat different in the two samples, then the sample with the more advantageous digestion conditions may have been more efficiently digested than the other. This could, for example, be the result if the yield from the buffer exchange was higher when a higher protein mass was added to the buffer exchange column. Then the sample with the more favorable digestion conditions would have generated a lower degree of missed cleavage peptides without a deamidation than the sample with the less favorable digestion conditions. The deamidation degree of the missed cleavage peptide would then be higher in the sample with the favorable digestion conditions than in the sample with the less favorable conditions. This example shows the importance of verifying the quantitative performance of bottom-up based methods against other validated methods during method development. This could for example be against methods based on intact protein analysis which generally involve less sample handling that may affect the protein and yield information on entire proteoforms.

The developed MAM based on DIA did not use MS/MS data for quantification of the N-glycan species. This choice was motivated by the fact that it would have required a very high number of SWATH

windows, which would have reduced the sensitivity of the assay, see chapter 5.1.5. Further, no signal interference in the precursor spectra of the glycopeptides could be observed in the samples that the assay was tested on. If signal interference does not seem to be an issue during upstream sample analysis it may be motivated to only use the precursor data for the N-glycan quantification. However, the fragmentation pattern contains valuable data that may be important for identification of a new peak as a glycopeptide. Different approaches could be taken to optimize the LC-MS method so that the glycopeptides are fragmented in separate SWATH windows. One such approach could be to build the variable SWATH windows based on the m/z values from the glycopeptide precursor ions instead of on all recorded precursor m/z values, as has been done previously for glycoproteomic analysis (85). A challenge for MAMs is, though, that it is important to avoid signal interference in all peptide fragmentation spectra, not only in those from the glycopeptides. This is especially important when dealing with complex samples with a high amount of signal interference. A combined approach may be possible that both prioritizes separation of precursor ions from glycopeptides as well as the total precursor ion signals when building the variable SWATH widows. An alternative strategy could be to separate the glycopeptides chromatographically, using for example a HILIC based resin. While this would improve the separation of the peptide glycovariants, it would however also lead to poor separation of peptides with different amino acid sequence. An approach based in two-dimensional chromatography could also be employed to improve separation of both peptides and glycovariants (86).

It has previously been shown that MAMs based on DDA provide good assay linearity with R^2 values > 0.99 (54,87,88). Further, quantification of protein levels in proteomics studies using DIA based methods have shown to provide good assay linearity (89,90). In this work the linearity of the developed DIA based MAM was evaluated and compared to that of a DDA based method for quantification of the methionine oxidation degree in a Protein A purified adalimumab biosimilar. A sample with induced oxidation was spiked into a sample with no oxidation in different ratios and the oxidation degree was quantified using the two methods. It was found that the assay linearity of the DIA based method was poor when the five most intense fragment ions from each peptide was used for quantification. Instead, if the same fragments were used for both the modified and the non-modified peptide and both fragments with and without neutral loss were included for the modified species, the assay linearity improved significantly (Figure 23). Further, fragmentation events next to a methionine residue was in some cases less favorable when the methionine residue was oxidized than when it was not and were therefore not used for quantification. The assay linearity is dependent on the non-modified and modified peptides giving similar signal intensities. If the signal from for example a modified peptide is lower per peptide amount than from the non-modified variant of that peptide this will result is a non-linear regression curve. It is therefore of highest importance that fragments which give similar summed intensities for non-modified and modified peptides are selected for use in DIA based MAMs if assay linearity shall be assumed. For future studies it would be interesting to study how other modifications, such as deamidation, affects the fragmentation pattern and thereby the linearity of quantification using DIA based methods.

An alternative MS strategy to DIA to avoid signal interference could for example be multiple-reaction monitoring (MRM). In MRM based methods predefined precursor and fragment ions are monitored sequentially. This could increase the assay sensitivity compared to the DIA based method since signal interference from co-fragmenting ions in the SWATH windows would not be an issue. In a study by Zhang et al. different MS instruments and strategies were compared for use in a MAM for an IgG2 expressed in CHO cells (80). It was found that from all identified protein modification events in the sample that may increase as part of the sample preparation, such as deamidation and oxidation, modifications measuring down to 0.1% by an MRM method on a TripleTOF instrument had CVs under

0.1. This could be compared to when the same protein modifications were quantified on an orbitrap instrument when instead modifications measuring down to 0.3% had a CV under 0.1, despite the higher resolution of the instrument. The use of and MRM based method would, however, eliminate the possibility to identify new modifications in old batch runs. A recently developed successor to the DIA technology SWATH-MS, which was used in this study, is scanning SWATH (91). In scanning SWATH, the precursor ions are selected by using a sliding isolation window, instead of the stepwise window acquisition that is used in conventional SWATH, while the fragmentation spectra are continuously recorded by the TOF analyzer. One result of this is that the duty cycle can be completed in a shorter time since there is no need to empty the collision cell between the steps. Another result is that a time dependent dimension is introduced into the duty cycle which can be used to assign the fragment spectra to the proper precursor ion more accurately than by conventional SWATH. Scanning SWATH may therefore be used to reduce co-fragmentation of precursor ions with similar m/z values and retention times. This could make it much easier to build well performing MAMs based on quantification from MS/MS data in the future.

In summary a MAM based on DIA was developed for quantification of deamidation, oxidation and N-glycan variants. The DIA based method could be used to avoid signal interference in the precursor spectra but had tendencies towards lower sensitivity than the DDA based MAM when signal interference was not an issue. The assay also showed good linearity when the precursor ions were selected with care. Another big advantage of quantification from DIA based methods compared to DDA is that the peptide identity is confirmed in each run while low abundant peptides may not be identified by the DDA based method. If the developed MAM will always be applied on very pure samples the advantage of using a DIA based MAM may not be as high as when applied on more crude samples, since the retention time and the isotopic distribution of the precursor ion may be enough for peptide identification. However, the possibility of identifying novel modifications in previous runs will still be a great advantage with the DIA based MAM. It is also possible to use a combined approach in the DIA based MAM, using precursor ions for some sites and fragment ions for others, as was also demonstrated in this study.

6.2 Investigating if a novel proteoform of rhGM-CSF with +70 Da may be a CQA

The second sub-aim of this thesis was to characterize the chemical structure of a novel modification of +70 Da identified in production process development samples of rhGM-CSF to be able to determine its source and if it may be a CQA, see chapter 5.2.1. The elemental composition of the modification was found to be C_4H_6O (Table 6) and it was found to contain a carbonyl group which was reducible by the weak reducing agent borane pyridine complex (Figure 29). Two sources to a modification fulfilling these requirements were identified in literature, MVK and crotonaldehyde. *In situ* recreation of the modification with MVK showed that this adduct did not possess the same fragmentation properties as the endogenous modification (Figure 30). However, the modification could not be recreated with crotonaldehyde. It was later found that the batch in which the modification had been identified had been a subject to overfeeding of glucose, which in turn had led to oxygen depletion in the cell culture media. It is likely that the formation of the +70 Da modification can be related to these changed expression conditions. A low oxygen pressure may induce an imbalance in the redox potential in biological systems (92–94). This in turn may lead to the formation of reactive oxygen species through lipid peroxidation pathways. Crotonaldehyde is one of the reaction products from lipid peroxidation (95). This compound is stable enough to diffuse smaller distances to cause the formation of protein adducts. It is therefore likely that the overfeeding of glucose may have resulted in the formation of crotonaldehyde reaction products in the protein batch. There may be several explanations to why commercial crotonaldehyde did not react with GM-CSF to form the modification identified in the

analyzed fermentation batch. For example, we might not have managed to reproduce the reaction conditions during which the modification was formed. Several parameters may have affected the reaction like for example various enzymes and metabolites present in the expression media. It may also be that the wrong substrate was used for the reaction. For example, only the trans isomer of crotonaldehyde is commercially available and therefore reaction with the cis isomer could not be investigated. In summary, crotonaldehyde was found to be the most probable source to the +70 Da modification through a metabolic linkage and through exclusion of other candidates.

For further studies it would be interesting to investigate the molecular structure of the modification by a technique such as NMR to have conclusive results on its structure. The use of NMR requires, however, highly concentrated samples. This would therefore require that a technique was found that could separate this proteoform from other proteoforms, since this was not achieved by RP HPLC. It would further be of high interest for the biopharma industry to investigate in closer detail which expression parameters that affects the formation of this modification of +70 Da. It also remains to be studied if the modification is specific to the protein or the expression system or if it is a general modification.

Though it could not be exclusively proven that the source to the +70 Da modification was crotonaldehyde, it could be shown that the modification contained a protein carbonyl moiety. Protein carbonylation has been metabolically linked to aging and to various diseases, such as Alzheimer's disease and atherosclerosis (96). The introduction of a carbonyl moiety may lead to protein dysfunction, either by blocking of the interaction sites or by conformational changes. Conformational changes may lead to protein aggregation and clearance. Carbonylation has also generally been related to initiation of protein degradation. The conclusion was therefore that the +70 Da modification may be a CQA. However, structure function studies are required in order to achieve conclusive results.

6.3 Introducing characterization of disulfide variants into MAMs

In the third project the question was asked if quantification of disulfide scrambling may be introduced in a MAM for GM-CSF together with other modification types by appropriate enzyme selection? A MAM based on digestion with the enzymes GluC and Chymotrypsin was optimized and evaluated for quantification of protein modifications from disulfide scrambling, free cysteine residues, +70 Da modification, deamidation, isoleucine to valine substitution and methionine oxidation, see chapter 5.2.3 and 5.2.4. Quantification of the different disulfide events and free cysteine residues was performed based on the precursor XIC areas from all cysteine containing peptides identified in a non-reduced digest. The assay was specifically tested for quantification of disulfide scrambling and free cysteine residues using a set of GM-CSF samples which had been subjected to conditions such as high pH, elevated temperatures and reducing conditions and was found to give predictive results. It is not obvious that quantitative results can be obtained from mass spectrometry data when the different protein modifications included in the relative quantification are contained in different peptides. One reason for this is that the proteolytic digestion is often not complete, resulting in the sites of interest being contained in a range of different peptides with missed cleavage sites. Therefore, a more specific enzyme than Chymotrypsin would have been more ideal to use for the MAM. This issue was, however, targeted by the developed assay in two ways. First, the digestion conditions were optimized so that the pools of the peptides containing each cysteine site were as homogenous as possible, see Figure 35. Secondly, all the disulfide linked peptides identified by Mascot in all the samples were included in the quantification of each disulfide event. Peptides not covered by the assay would be those with too short amino acid sequence to be captured by the MS method or with more than one disulfide linkage. In the way the assay was optimized, the degree of peptides with more than one cysteine residue was, however, very low. The second reason why quantitative results cannot necessarily be achieved from

mass spectrometry analysis of protein modifications contained in different peptides is that the amino acids ionize differently well. In a study by Liigand et al. the ionization efficiencies (IE) of 21 amino acids were measured (97). It was found that large and hydrophobic amino acids as well as basic amino acids ionize better than hydrophilic and acidic amino acids. The amino acids leucine, phenylalanine and histidine were found to ionize especially well. Another aspect that should be taken into account for disulfide linked peptides is that the tertiary structure induced by the disulfide bond may lead to shielding of amino acids. The shielding could make amino acids with high IE less accessible to solvent and thereby reduce the overall IE of the peptide. In GM-CSF there is, for example, a leucine residue located at position N+1 from the cysteine residue C54, which may be shielded by the disulfide linkage. This peptide does, on the other side, contain four residues with high IE, one phenylalanine and three leucine residues, which may cover for this ionization loss. It was further found in the study by Liigand et al. that, as the oligopeptide becomes longer the effect of the IE of the individual amino acids on the overall oligopeptide IE decreases, an effect that starts already after 6 amino acid residues. A study by Loo et al. has further shown that the effect of disulfide bond reduction on the overall IE is most pronounced for proteins weighing more than 10 kDa (98). Based on these studies the ionization effects may not be that big for relative quantification of disulfide linked peptides. Quantification of the free cysteine residues may then be more affected by the difference in amino acid IEs. Free cysteine residues are quantified from peptides which are not disulfide linked and since these peptides are shorter, the IEs of the individual amino acid residues may have a larger effect on the total IE in those peptides. However, acceptable proteoform cutoff values must be determined individually for each MAM and therapeutic protein and must be based on proper activity as well as pharmacokinetics and pharmacodynamics studies. It may therefore be possible to adjust the cutoff values for the free cysteine degree according to the different IEs. An alternative strategy could be to normalize the quantitative assay against the different peptide IEs. In the study by Liigand et al. a model to predict the IEs of different oligopeptides with up to 14 amino acid residues was developed. If a similar model could be developed for disulfide linked peptides this could be used to further increase the quantitative accuracy of the developed MAM assay.

A large number of the recently developed methods for analysis of disulfide bonds in therapeutic proteins are based on peptide mapping analysis by LC-MS/MS. These methods usually focus on the identification of disulfide variants in the protein samples but not on the relative quantification of the different variants. The reason for this is often the high complexity of the data, arising from the presence of several cysteine residues in the peptide sequences. One strategy to overcome this problem is to separate the different isoforms on the intact level, where quantitative analysis is performed, followed by peptide mapping of the different isoforms. This has for example been performed by SDS-PAGE or CE and is useful for quantification of for example proteoforms of IgG with different combinations of linkages to the different protein subunits (62). The method is, however, dependent on the chromatographic separation of all proteoforms, which might not be achievable when the scrambling comes from intra-chain linkages. The analysis is also time consuming, since each proteoform requires a separate digest. An alternative strategy to map the disulfide linkages in peptides have been proposed by Albert et al. (99). In their study the peptides were immobilized on a solid phase followed by a partial reduction of the disulfide bonds followed by alkylation and washing. This was then followed by reduction of the next set of disulfide linkages by increasing the reducing agent concentration further followed by alkylation with another alkylating agent. By performing this partial reduction in a stepwise manner, the different linkages in a peptide with several disulfide linkages could be mapped. One strategy to make the MAM developed in this thesis for characterization of GM-CSF even more quantitatively reliable could be to introduce this strategy into the sample preparation protocol. If a sequential reduction could be performed of the different

disulfide linkage combinations in GM-CSF it may be possible to determine the different linkage combinations from one single peptide representing each disulfide site. The relative degree of different alkylating agents could then be used to determine the degree of the different scrambling events and quantification could be performed using the same strategy as for the other protein modification types quantified by the MAM. This strategy would however require careful optimization and evaluation since it has to our knowledge only been employed on pure peptides. The effect on sample preparation induced modifications, would also require careful evaluation. SPE was in this work, for example, identified as a prominent source to formation of sample preparation induced oxidation.

Another aspect of disulfide-scrambling quantification is that it is not always possible to separate all cysteine residues in a protein by enzymatic digestion. This may for example be the case when the protein has a high content of cysteine residues and may lead to the formation of highly complex disulfide linked peptides. In these cases, fragmentation by CID may not provide enough data to identify all disulfide-linked peptides. An alternative fragmentation technique is ETD, which mainly results in cleavage of disulfide bonds together with a lower degree of z and c ions. This technique may be used to determine the structure of for example cyclized peptides and may be combined with CID to solve the structure of complex disulfide bound peptides (62). Optimization of peptide fragmentation techniques may be used in combination with optimization of the sample preparation protocol to achieve the most information about the protein's disulfide bonds. In this work the focus was on optimization of the sample preparation protocol.

6.4 Oxidation quantification in MAMs

Sample preparation induced modification is a common issue when quantifying the oxidation degree by bottom-up methods (100). The general assumption that is often being made is that the same degree of oxidation is induced in all samples prepared at the same time. When the methionine oxidation degree was quantified in the GM-CSF early process sample and the control sample from three replicate digests, the modification degree in the control sample was subtracted from that in the early process sample for each replicate. The mean and standard deviation for the difference in modification degree was then calculated and if the standard deviation for the difference was lower than the mean difference, the modification degree was considered different between the two samples. This was the case for two out of three methionine sites (Figure 38d). This strategy could not be applied when the IgG MAM was used. This could be seen from the fact that when the difference in oxidation degree was calculated between the IgG samples, this value varied a lot between the replicate digests. For example, in the cell culture filtrate the oxidation of site LC_M4 was 10.6% in replicate 1 and 6.2% in replicate 2 while the modification degree for the same site in the Protein A purified IgG was 6.3% in replicate 1 and 12.4% in replicate 2 (data not shown). Therefore, the mean of the absolute oxidation values between the replicate digests were plotted in for example Figure 24b and not the mean between the oxidation differences, as was done when the GM-CSF MAM was used. Already during the optimization of the sample preparation protocol, it could be observed that high oxidation degrees may be induced by for example SPE of the digested protein, see chapter 5.1.2. This oxidation degree could be reduced by performing the cleanup online with the LC-MS analysis using a trap column (Figure 8). One big difference between the sample preparation protocols for the two proteins was that a protein cleanup step by acetone precipitation was included in the beginning of the IgG protocol which was not present in the protocol for GM-CSF. Protein precipitation is a common strategy for changing the protein solution content in proteomics (101). This involves the proteins being aggregated upon addition of an organic solvent from the introduction of electrostatic and dipole forces. The organic solvent can then be removed by pipetting and evaporation of the last volume of organic solution, and the protein may then be dissolved in the desired solvent. One explanation to the

high variation in oxidation degree between the IgG samples from the same preparation could be that they were exposed to air in various degrees. The removal of the organic solvent is performed manually and will result in slightly different volumes left in the vial for evaporation. Some precipitates may have become drier than others upon evaporation and hence resulted in higher oxidation degrees. It is, however, not completely fair to compare the variation in modification degree between the two MAMs since the method for GM-CSF was only tested in three replicate digests while the method for the IgG was tested in five replicates. Nevertheless, it may be concluded from these studies that sample preparation induced oxidation is an artifact that complicates measurement of the true oxidation degree. It is therefore of high importance to know the variation in the sample preparation induced oxidation levels of a specific assay when the assay is used to study protein oxidation levels.

6.5 Improving MAMs for improved therapeutic protein proteoform characterization

The main aim of this thesis was to improve current bottom-up based proteomics approaches for quantification of low abundant proteoforms in therapeutic proteins with regard to the assay performance in non-purified samples and the kind of protein modifications that are covered. From performing the first sub-aim it was found that signal interference in the precursor spectra of complex protein samples may be avoided by basing the MAM on DIA instead of DDA. It was shown that this could be accomplished without losing assay linearity and also, to a certain degree, without losing assay precision. However, the MAM based on DIA showed lower sensitivity than that based on DDA which led to a lower precision of the DIA based method for low abundance signals. It will therefore be a trade-off if the loss of signal sensitivity may be motivated for the way the assay will be used. From performing the second sub-aim it was found that overfeeding of glucose to recombinant expressions systems based on *E. coli* is a probable source to a protein modification of +70 Da identified in rhGM-CSF. This modification was found to be a possible CQA and should therefore be included in MAMs for this protein and expression system. It remains to be studied if the modification should be included in MAMs for all therapeutic proteins expressed in recombinant expression systems. From performing the third sub-aim it was found that quantification of disulfide scrambling events may be included in a MAM by proper selection of proteolytic enzymes. It was shown that the relative amount of different disulfide bond linkages could be followed in a predictive way in rhGM-CSF samples subjected to forced degradation and in production process development samples by using a MAM based on proteolytic digestion with GluC and Chymotrypsin. Caution should however be taken when using the quantitative values as absolute values since the amino acid sequence as well as the three-dimensional structure of the different disulfide linked peptides may affect their ionization efficiency. In summary, several strategies to improve current bottom-up based proteomics approaches have been demonstrated in this thesis. These methods may be used to speed up production process development strategies for therapeutic proteins and thereby reduce the production costs so that the drugs may reach the society earlier and be sold for a cheaper price. The developed strategies may also improve the knowledge on therapeutic proteins before they are being introduced on the market, leading to safer protein-based drugs.

7. Materials and methods

7.1 Chemicals

Urea, sodium phosphate dibasic dihydrate, sodium phosphate monobasic dihydrate, 1,4-Dithiothreitol (DTT), iodoacetamide (IAM), N-ethylmaleimide (NEM), borane pyridine complex, 4-Vinylpyridine, triethylammonium bicarbonate buffer (TEAB), crotonaldehyde, 3-buten-2-one (MVK), ammonium hydroxide (NH_4OH), ammonium formate, phosphoric acid, ammonium bicarbonate (NH_4CO_3), Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), hydrogen peroxide (H_2O_2), MES hydrate and formic

acid (FA) were all purchased from Sigma-Aldrich. Acetonitrile (ACN), acetone, ethanol (EtOH), acetic acid and methanol (MeOH) were all purchased from Chemsolute. Trifluoroacetic acid (TFA) (Acros Organics), DNPH (Tokyo Chemical Industry Co., Ltd.), guanidinium hydrocholoride (GdnHCl) (Thermo Fisher), GlycoWorks RapiFluorMS Basic Kit-24 (Waters).

7.2 Methods used in chapter 5.1 "Evaluating the potential of using DIA in a MAM for a therapeutic IgG"

7.2.1 Adalimumab biosimilar sample information

A batch of an adalimumab biosimilar was produced by PhD students at the University of Natural Resources and Life Sciences (BOKU) in Vienna by fed-batch culture. Four samples from the batch was analyzed in this study, 1) an untreated cell culture filtrate, total protein concentration: 8.7 µg/µl, IgG concentration: 0.61, 2) the cell culture filtrate which had been left at pH 3.5 for 1 h followed by neutralization to pH 7, total protein concentration: 5.8 µg/µl, IgG concentration: 0.35, 3) the IgG purified on Protein A (MabSelect SuRe), IgG concentration: 13.43 µg/µl and 4) the protein purified on Protein A (MabSelect SuRe) which had been left at pH 3.5 for 1 h followed by neutralization to pH 7, IgG concentration: 7.1 µg/µl. The total protein concentration in the cell culture filtrates were measured by amino acid analysis, see chapter 7.2.2, and the IgG concentrations were measured by integration of the peak area from Protein A analysis (performed at BOKU).

7.2.2 Amino acid analysis

Amino acid analysis was performed as a protein characterization service by Alphalyse A/S. Amino acid analysis is a method to determine the absolute amounts of individual amino acids in a sample. The amino acids are separated and quantified using ion exchange chromatography and post-column derivatization with ninhydrin. The total amount of proteins with unknown sequences is calculated as the sum of the amino acids.

The acid hydrolysis was performed for 20 hours at 110°C, in 6 M hydrochloric acid, 0.1% phenol, 0.1% thioglycolic acid under reduced pressure in an atmosphere of nitrogen. Identification and quantification of the amino acids was performed on a BioChrom 30+ amino acid analyzer using ion exchange chromatography, post-column derivatization with ninhydrin and detection at two wavelengths, 570 nm and 440 nm. A known amount of the unusual amino acid Sarcosine (Sar) was added as an internal control standard. 18 of the common 20 amino acids were determined, since Tryptophan is degraded during the hydrolysis, and the yield of Cysteine is so variable that it is not calculated. The Asparagine was determined with Aspartic acid (Asx) and Glutamine with glutamic acid (Glx). The total protein content was calculated as the sum of the individual 18 amino acids.

7.2.3 IgG precipitation, denaturation, reduction, alkylation and digestion

Adalimumab biosimilar samples were diluted in 100 mM sodium phosphate buffer, pH 7.0. The Protein A purified IgGs were diluted to 0.1 µg/µl and the cell culture filtrate samples to 1 µg/µl, total protein concentration. Acetone stored in -20 ˚C was added to the samples in a sample to acetone ratio of 1:4 and were stored in -20 ˚C overnight. Samples were vortexed for a few seconds before centrifugation for 10 minutes at 4 ˚C at 14,000 g. The supernatant was removed, and the pellet washed twice with 300 µl acetone/EtOH/H2O 2:2:1, spinning the pellet down for 10 minutes at 4 ˚C at 14,000 g before removing the supernatant. The pellet was left to dry for ~15 minutes.

The pellet was dissolved in 6 M GdnHCl, 6 mM DTT and 20 mM sodium phosphate buffer and the sample was denatured and reduced for 1 h at 56 ˚C. This was followed by alkylation in 12 mM iodoacetamide for 45 min at room temperature in dark. The sample buffer was changed to 1 M urea in 100 mM phosphate buffer using Zeba Spin Desalting Columns, 40K MWCO, 0.5 mL (Thermo Fisher),

according to the product protocol, followed by digestion with LysC (Lysyl EndopeptidaseR; FUJIFILM Wako Pure Chemical Corporation), 0.02 AU enzyme/ 50 µg protein, and Trypsin (Sequencing grade modified trypsin; Promega) using 1:50 enzyme to protein ratio in 30 ˚C overnight. The digest was stopped with 1% TFA.

7.2.4 Forced IgG degradation at high pH and elevated temperature

10 µl 0.8% NH_4OH was added to 10 µl 10 µg/µl Protein A purified adalimumab biosimilar. The sample was incubated at 37 ˚C for 1 or 2 weeks. The sample pH was then adjusted to pH 6.5 by adding 10 µl 1.3 M MES buffer with pH 6.5 and 70 µl ultra-high quality (UHQ) water. The samples were then prepared according to the protocol described in chapter 7.2.3.

7.2.5 Forced IgG methionine oxidation in hydrogen peroxide

1 µl 2% H_2O_2 and 15 µl 8 M GdHCl was added to 1 µl 10 µg/µl Protein A purified adalimumab. The sample was incubated at 56 ˚C for 1 h. The sample pH was then adjusted to pH 6.5 by adding 10 µl 1.3 M MES buffer with pH 6.5 and 70 µl UHQ water. The samples were then prepared according to the protocol described in chapter 7.2.3.

7.2.6 Peptide cleanup methods by SPE

7.2.6.1 Peptide cleanup by C18 OMIX tips

A 100 µl C18 OMIX tip (Agilent Technologies) was applied to a pipet. The resin was wet 3 times by pipetting and then dispensing to waste 100 µl of 50% ACN. The resin was equilibrated 3 times by pipetting and then dispensing to waste 100 µl of 0.1% TFA. The sample was bound to the resin by pipetting and dispensing for 15 times. The resin was rinsed 3 times by pipetting and then dispensing to waste 100 µl of 0.1% TFA. The peptides were eluted in 30 µl 0.1% FA in 50% ACN once and in 30 µl 0.1% FA in 70% ACN twice. The liquid was removed by vacuum centrifugation.

7.2.6.2 Peptide cleanup by MCX µElution Plate

The samples were diluted 1:1 in 4% phosphoric acid. The sample was loaded slowly on an Oasis prime MCX µElution Plate (Oasis) by applying vacuum. The peptides were washed with 200 µl 100 mM Ammonium Formate with 2% FA by applying vacuum and then by 200 µl 100% MeOH by applying vacuum followed by centrifugation for 1 min at 1,700 rpm. The peptides were eluted twice in 30 µl 5% NH_4OH in MeOH prepared fresh, by centrifugation for 1 min at 1,700 rpm. The liquid was removed by vacuum centrifugation.

7.2.6.3 Peptide cleanup by HLB µElution Plate

The samples were diluted 1:1 in 4% phosphoric acid. The wells of a HLB µElution Plate (Oasis) were conditioned with 200µl 100% methanol by applying vacuum. The wells were equilibrated with 200µl UHQ water by applying vacuum. Samples were loaded slowly by applying vacuum followed by centrifugation for 1 min at 1,700 rpm. The peptides were washed in 5% methanol by applying vacuum followed by centrifugation for 1 min at 1,700 rpm. The peptides were eluted twice in 30 µl methanol by centrifugation for 1 min at 1,700 rpm. The liquid was removed by vacuum centrifugation.

The peptides purified by SPE were analyzed by peptide mapping according to chapter 7.2.7 but evading the cleanup on trap and using the analytical column NanoEase M/Z CSH130, 1.7 µm, 300 µm x 150 mm (Waters) at a column oven temperature of 60 °C.

7.2.7 Peptide cleanup on trap coupled online to RP-LC-ESI-TripleTOF-MS analysis

10 µl 0.1 µg/µl digested adalimumab biosimilar was loaded on a 0.3 mm trap cartridge, ChromXP, C18CL, 5µm, 120Å (Eksigent) and was washed for 6 min in phase A (0.1% FA in UHQ water) at a flow rate of 5 µl/min online before eluting the sample on the analytical column.

LC-MS acquisition of the digested IgG was performed using an Eksigent system coupled to a SCIEX TripleTof 6600 mass spectrometer. The digest was eluted from the trap column onto the analytical column 3C18-CL-120, 3µm, 120Å, 0.3 x 150 mm (Eksigent) at 30 °C column oven temperature. Elution was performed at a flow rate of 5 µL/min with solvent A (0.1% FA in UHQ) and solvent B (0.1% FA in ACN). The column was equilibrated for 2 min at 5% solvent B before applying a linear gradient of 5-29% solvent B over 58 min. This was followed by a steeper gradient of 29-55% solvent B in 5 min, up to 75% solvent B in 5 min followed by column washing and reconditioning.

Two other trapping columns were also evaluated: NanoEase M/Z CSH C18 Trap Column, 130 Å, 5 µm, 3000 µm x 50 mm (custom made from Waters) and YMC-Pack Pro C18, 12 nm, S-3 µm, 5 x 0.5 mm ID, 1/16" (YMC) in combination with the analytical column NanoEase M/Z CSH130, 1.7 µm, 300 µm x 150mm (Waters) at 60 °C column oven temperature. Both setups gave poor chromatography and were therefore never used for the MAM.

Mass spectrometry acquisition was performed in positive polarity mode using declustering potential 20 V, collision energy 6 V and source temperature 100 ˚C.

7.2.7.1 Data dependent acquisition

The MS1 data from DDA was recorded in the range 300-1700 m/z with an accumulation time of 0.25 s and using 4-time bins to sum. MS2 acquisition was performed in information dependent mode (IDA) on charge states 2-5 exceeding 100 cps on a maximum of 25 candidate ions and excluding former candidate ions for 5 s after 1 occurrence, MS/MS scan range 130-2000 m/z. The total cycle time was 1.5 s.

7.2.7.2 Data independent acquisition (SWATH-MS)

Data independent acquisition was performed using the SWATH-MS strategy. MS1 data was recorded in the range 300-1700 m/z with an accumulation time of 0.15 s and using 4-time bins to sum. MS2 acquisition was performed on precursors using 37 variable SWATH windows (Table 9) calculated using a variable window calculator provided by AB Sciex. The window sizes were calculated so that the precursors from a DDA analysis of a digest of the cell culture filtrate were evenly distributed between the windows. Fragmentation data was recorded in the range 130-2000 m/z with 41 ms accumulation time and using 8 bins to sum. The total cycle time was 1.7 s.

Table 9: SWATH windows used for the DIA based MS method.

Precursor ion start mass	Precursor ion stop mass	Declustering potentioal (DP)	Collision energy (CE)	CE spread	Ion release delay	Ion release width
299.5	350.2	20	16.3	5	30	15
349.2	385.2	20	19.26	5	30	15
384.2	406.2	20	21.28	5	30	15
405.2	421.6	20	22.53	5	30	15
420.6	436.3	20	23.48	5	30	15
435.3	449.6	20	24.39	5	30	15
448.6	463.6	20	25.23	5	30	15
462.6	476.2	20	26.08	5	30	15
475.2	488.8	20	26.87	5	30	15
487.8	502.1	20	27.67	5	30	15
501.1	517.5	20	28.53	5	30	15
516.5	533.6	20	29.5	5	30	15

532.6	549	20	30.49	5	30	15
548	563.7	20	31.45	5	30	15
562.7	577.7	20	32.36	5	30	15
576.7	592.4	20	33.24	5	30	15
591.4	607.8	20	34.17	5	30	15
606.8	622.5	20	35.12	5	30	15
621.5	637.9	20	36.05	5	30	15
636.9	653.3	20	37.01	5	30	15
652.3	670.8	20	38	5	30	15
669.8	689.7	20	39.11	5	30	15
688.7	710.7	20	40.32	5	30	15
709.7	733.8	20	41.66	5	30	15
732.8	759.7	20	43.14	5	30	15
758.7	788.4	20	44.79	5	30	15
787.4	817.1	20	46.59	5	30	15
816.1	845.8	20	48.38	5	30	15
844.8	872.4	20	50.15	5	30	15
871.4	899.7	20	51.82	5	30	15
898.7	928.4	20	53.54	5	30	15
927.4	959.9	20	55.37	5	30	15
958.9	1,002.60	20	57.48	5	30	15
1,001.60	1,064.90	20	60.38	5	30	15
1,063.90	1,149.60	20	64.54	5	30	15
1,148.60	1,285.40	20	70.41	5	30	15
1,284.40	1,699.10	20	81.59	5	30	15

7.2.8 IgG data analysis

7.2.8.1 Deamidation quantification using data from DIA (SWATH-MS)

The sample subjected to 2 weeks forced degradation was digested and acquired on four different days by the DDA and DIA based MS methods. The .wiff files from the DDA were converted to .mgf files using ProteoWizard's MS convert program (version 3.0.18204 64-bit). These were searched in the Mascot probability-based search engine (version 2.7.0.1) using a database containing only the adalimumab biosimilar sequence. Carbamidomethyl on cysteines was added as fixed modification and asparagine and glutamine deamidation, glutamine and glutamic acid pyroglutamination and methionine oxidation were added as variable modifications. Trypsin was added as digestion enzyme, up to 3 missed cleavages were accepted, peptide charges from +2 to +4 were included, peptide tolerance settings were ± 15 ppm, MS/MS tolerance settings were 0.05 Da and ESI-QUAD-TOF was selected as instrument. A library was built in Skyline (version 19.1.0.193) using the DIA workflow from .dat files created from the Mascot search results. A cut-off score of 0.95 was used to build the library. This library was used to extract precursor ions with charge states from 2-5, fragment ions with charge states from 1-2 and y and b ions as well as precursor ions. The 20 most intense MS/MS fragment ions were extracted from filtered ion charges and types using an ion match tolerance of 0.05 m/z. Product ions from ion 3 to the last ion were extracted and DIA precursor windows were used for exclusion. For MS1 filtering, peaks with a minimum of 5% of base peak were included, a centroid mass analyzer was used and a mass accuracy of 20 ppm. For MS/MS filtering, DIA was selected as acquisition method,

the isolation windows used for the DIA method were imported, a centroid mass analyzer was used and a mass accuracy of 30 ppm. Precursor ions were added manually for those peptides where the most intense precursor ion was not identified by MS/MS, so that the charge states making up the majority of the precursor intensity were included. The same charge states were added for both the modified and the non-modified variant of the same peptide sequence. For the precursor ions that were not present in the spectral library the fragment ions with the highest XIC area and good peak shape by manual inspection were selected. The data was exported from Skyline and the relative XIC areas from the fragment ions from each precursor ion were calculated. The CV between the relative fragment areas from the four runs was calculated for each fragment. The CVs were found to be generally lower for higher relative fragment areas. Therefore, the five fragments with the highest relative mean areas were selected for quantification of each precursor. The only exception to this was the peptide containing sites HC_N388 and HC_N393 which was quantified from fragments y4, y6 and y8 using both the deamidated and the non-modified peptide, see the reasoning for this in chapter 5.1.5. The peptide library in Skyline was adjusted according to these criteria and the DIA acquisition data from the samples of interest were searched against this library. The modification degree was quantified by extracting the data from Skyline and calculating the relative area of the summed fragment XIC areas from the modified peptide in relation to the summed fragment XIC areas from the non-modified and modified peptides.

7.2.8.2 Oxidation quantification using data from DIA (SWATH-MS)

The sample subjected to induced methionine oxidation stress was digested and acquired by the DDA and DIA based MS methods described above. The .wiff files from the DDA were converted to .mgf files using ProteoWizard's MS convert program (version 3.0.18204 64-bit), and these were searched in the Mascot probability-based search engine (version 2.7.0.1) as described in chapter 7.2.8.1. A library was built in Skyline (MacCross lab, version 19.1.0.193) using the DIA workflow from .dat files generated from the Mascot search results. A cut-off score of 0.95 was used to build the library. This library was used to extract the 20 most intense fragment ions from each precursor ion from the DIA data as described in chapter 7.2.8.1. The five fragment ions with the highest XIC areas from each precursor ion from the oxidized peptides, were selected for quantification of both the oxidized and the non-oxidized variant of the peptide, excluding ions from neutral loss. When the fragment ions had been selected the neural loss variant of the selected fragments were also included in the peptide library. Fragment ions with signal interference in the cell culture filtrate sample were not selected for the library. Further, fragment b_3 was excluded from quantification of the peptide containing site LC_M4, see explanation in chapter 5.1.7. This quantification strategy was developed since the strategy used to quantify deamidated peptides, described above, gave non-linear results. This was tested using the linearity assay described below. The peptide library in Skyline was adjusted according to these criteria and DIA data from the samples of interest were searched against this library. The modification degree was quantified by extracting the data from Skyline and calculating the relative area of the summed fragment XIC areas from the modified peptide in relation to the summed fragment XIC areas from the non-modified and modified peptides.

7.2.8.3 Deamidation and oxidation quantification using data from DDA

A spectral library was built in Skyline in a similar fashion as for quantification of DIA data, see chapter 7.2.8.1, but the DDA workflow in Skyline was used instead for building the library. The same precursor ions as were used for quantification of the DIA data were included in the spectral library. The .wiff files from the DDA were searched using the library. The modification degree was quantified by extracting the data from Skyline and calculating the relative area of the summed precursor XIC areas from the

modified peptide in relation to the summed precursor XIC areas from the modified and non-modified peptide. The precursor areas were filtered using an idot product cutoff of 0.7.

7.2.8.4 Independent samples t-test of quantitative results from DIA or DDA based MAM

Independent samples t-tests were performed to determine the statistical significance of the differences between the quantitative results obtained by the DIA and the DDA based strategies. Deamidation, oxidation and N-glycan variant degrees were quantified from three replicate injections of the cell culture filtrate on the MS using the DIA or DDA based strategy. The differences in quantitative values between the DIA and DDA-based methods were evaluated for each peptide using independent samples t test. The t-test was performed in Excel using the function "t-test: Two-sample Assuming unequal variances". Unequal variances were assumed since the standard deviation from the results for the two methods differed by more than two times for several peptides.

7.2.9 Released IgG glycan analysis

7.2.9.1 Glycan release and fluorescence labeling using GlycoWorks RapiFluor-MS N-Glycan Kit

7.5 µl 1.3 µg/µl Protein A purified adalimumab was denatured in 5 µl 3% RapiGest in 5x GlycoWorks Rapid Buffer for 3 min at 90 ˚C. The sample was then digested with 5 µl GlycoWorks Rapid PNGase F for 5 min at 50 ˚C. Glycans were tagged with 5 µl RapiFluor-MS Reagent Solution for 5 min at room temperature. The sample was then diluted with 180 µl ACN followed by cleanup on a GlycoWorks HILIC µElution Plate. The plate columns were conditioned with 200 µl UHQ water by applying vacuum followed by equilibration with 200 µl 85% ACN in the same way. The samples were loaded by slowly applying vacuum. The columns were washed twice with 600 µl 1% FA in 90%. The labelled glycans were eluted twice in 22.5 µl GlycoWorks SPE Elution Buffer by centrifugation for 1 min at 1,700 rpm. The eluted glycans were diluted in 155 µl GlycoWorks Sample Diluent-DMF/ACN before analysis by HILIC-fluorescence detection and HILIC-MS.

7.2.9.2 Released glycan acquisition by HILIC-fluorescence detection

The released and labelled glycans were analyzed on an Agilent 1290 Infinity II HPLC system coupled to fluorescence detection with excitation wavelength 265 nm and emission wavelength 425 nm at 2 Hz frequency. 40 µl sample was loaded on an Acquity UPLC Glycan BEH amide 120 Å, 1.7 µm, 2.1 x 150 mm analytical column (Waters) held at 45 ˚C. A gradient of phase A (50 mM ammonium formate, pH 4.5) and phase B (100% ACN) was applied at 0.5 µl/min flow rate, going from 20% to 27% phase A in 3 min, then to 37% phase A in 2 minutes, and then to 100% phase A in 31.5 min, followed by wash and equilibration.

7.2.9.3 Released glycan acquisition by HILIC-ESI-QTOF-MS

The released and labelled glycans were analyzed on an Agilent 1290 Infinity II HPLC system coupled to a Bruker Maxis Impact Q-TOF MS. 80 µl sample was loaded on the system using the HILIC method described in chapter 7.2.9.2. Mass spectrometry acquisition was performed in positive mode. MS1 and MS2 data were acquired in the range 50-3000 m/z with 2 Hz scan rate and a total cycle time of 1.5 s. MS2 acquisition was performed in data dependent mode with an absolute threshold of 300 counts per 1000 summations, excluding the precursor ion after 1 spectrum for 2 s while reconsidering it if the current intensity/ the previous intensity was above 2.5.

7.2.9.4 Released glycan identification in ProteinScape and Skyline

The released glycan data was searched in ProteinScape using a glycan classification method, to identify glycan oxonium ions, and a glycan analysis method, to identify glycopeptide ions. All identified glycan species with a score > 10, by the glycan analysis method, and a glycan classification, by the glycan classification method, were included in a spectral library in Skyline using the proteomics interface. Each RapiFluor labelled glycan was included in the library by adding the glycan as a modification to a

fictional peptide with the same molecular composition as the RapiFluor molecule, charge states 2-5 were included. The fictional peptide was NSTF[-$C_3H_6O_6$]. The released glycan data was searched against this library and all glycan identifications with an idot product > 0.8 and a clear peak shape by visual inspection were selected for the glycopeptide library.

7.2.10 Quantification of glycosylated peptides

A peptide library containing all possible glycovariants of the peptide TKPREEQYN[+glycan]STYR was built in Skyline by adding the glycopeptides identified by released glycan analysis as modifications to the peptide. This peptide was selected since it was found to give the highest precursor intensity when the Protein A purified adalimumab biosimilar was analyzed. The precursor ions from the protein digests acquired by DDA or DIA were searched against this library, including charge states 2-4. The relative amount of each glycovariant in the sample was then calculated by summing the precursor XIC areas from one glycopeptide variant and calculating the relative area compared to the sum of all glycopeptide variant areas in the sample.

7.2.11 MAM precision study

40 µg of the cell culture filtrate (described in chapter 7.2.1) was prepared in five replicates according to the sample preparation protocol described in chapter 7.2.3. In the same way 10 µg of the 2 weeks stressed IgG (described in chapter 7.2.4) was prepared in four replicates and one sample of 40 µg. One of the cell culture filtrate digests and the 40 µg digest of the 2 weeks stressed IgG were acquired with DDA and DIA in three technical replicates on the same day and in three technical replicates acquired on three separate days. The replicate digests were also acquired with both DDA and DIA. Data treatment and modification degree quantification was performed as described above.

7.2.12 MAM linearity study for quantification of methionine oxidation

A digest of the adalimumab biosimilar sample subjected to induced methionine oxidation stress (described in chapter 7.2.5) was mixed with a digest of the non-stressed sample in the following percentages: 0%, 1%, 2.5%, 5%, 7.5%, 10%, 20%, 40%, 60% and 80%. The samples were acquired by DDA and DIA. The sample data was searched in a Skyline library as described above in the section 7.2.8.2 and 7.2.8.3. The measured oxidation percentage was quantified by extracting the data from Skyline and calculating the relative area of the summed XIC areas from the modified peptide in relation to the summed XIC areas from the non-modified and modified peptides. The theoretical oxidation percentages in the samples were corrected using the measured oxidation percentage in sample with 0% spiked in IgG with forced oxidation.

7.2.13 MAM assay testing

40 µg of 1) the untreated cell culture filtrate, 2) the cell culture filtrate stressed at pH 3.5, 3) the Protein A purified IgG and 4) the Protein A purified IgG stressed at pH 3.5 (described in chapter 7.2.1) were prepared in five replicates according to the sample preparation protocol described in chapter 7.2.3. In the same way 10 µg of the 2 weeks stressed IgG (described in chapter 7.2.4) was prepared in four replicates and one sample of 40 µg. All samples were acquired by DIA. Data treatment and modification degree quantification was performed as described above.

7.3 Methods used in chapter 5.2.1 “Characterization of a novel +70 Da modification in rhGM-CSF produced in *E. coli*”

7.3.1 GM-CSFs sample description

Three samples of rhGM-CSFG expressed in E. coli were used in this study. 1) An early process sample from batch 3 before protease treatment to remove N-terminal methionine (DSP3-SEP46), protein concentration: 7.8 µg/µl, 2) a late process sample from batch 3 (DSP3-SEP99B), protein concentration:

1.0 µg/µl, 3) a late process sample from batch 5(DSP5-SEP99), protein concentration: 2.23 µg/µl. Protein concentrations were measured by amino acid analysis in the intact protein sample (see chapter 7.2.2).

7.3.2 Protein digestion

An early process sample of GM-CSF (DSP3-SEP46) was denatured in 6 M urea in 50 mM sodium phosphate (NaP) buffer, pH 7 for 1 h. The sample was then diluted to 0.8 M urea in NaP buffer and digested overnight at 30 ℃ with LysC and GluC (sequencing grade; Promega, Madison, WI) in enzyme to protein ratios 1:10 and 1:25 respectively. The digestion was stopped with 1% TFA.

7.3.3 Reduction of GM-CSF carbonyl groups with borane pyridine complex

The GM-CSF early process sample was denatured and reduced in 6 M urea and 100 mM borane pyridine complex in NaP buffer with pH 7 at ambient temperature overnight. The sample buffer was changed on Vivaspin centrifugal concentrators with 5 kDa molecular weight cutoff (MWCO) filters (Sartorius) to 6 M urea in NaP buffer. The disulfide bonds were reduced with 5 mM DTT for 1 h at 30 ℃ followed by alkylation in 10 mM 4-Vinylpyridine for 45 min at ambient temperature. The sample was then digested for 2 h at 30℃ with LysC, 1:10 enzyme to protein ratio, followed by sample dilution in NaP buffer to 0.8 M urea and digestion overnight at 30 ℃ with GluC, 1:25 enzyme to protein ratio. The digestion was stopped with 1% TFA. The samples were then analyzed by RP-LC-ESI-QTOF-MS as described in chapter 7.3.7.

7.3.4 Derivatization of GM-CSF carbonyls with DNPH

7.5 µg of the digested early process sample, as described in chapter 7.3.2, was dried using a vacuum centrifuge. 25 µl of 100 mM DNPH in 0.5% TFA in DMSO was added to the dried digest and was incubated in a shaker overnight at ambient temperature. The samples were then analyzed by RP-LC-ESI-QTOF-MS as described in chapter 7.3.7.

7.3.5 GM-CSF derivatization with MVK and crotonaldehyde

A late process sample of GM-CSF with very low levels of +70 Da modification (DSP5-SEP99) was incubated with 1 µM, 10 µM, 100 µM, 1 mM or 10 mM MVK or crotonaldehyde in 100 mM Triethylammonium bicarbonate (TEAB) for 25 h at 37 ℃. The sample buffer was changed on Vivaspin centrifugal concentrators with 5 kDa MWCO filters to 6 M urea in 50 mM NaP buffer, pH 7. The samples were then digested according to the protocol described in chapter 7.3.2. After digestion the samples were cleaned up prior to LC-MS analysis on a HLB µElution plate as described in chapter 7.2.6.3. The digests were then analyzed by RP-LC-ESI-TripleTOF-MS as described in chapter 7.3.8.

7.3.6 Intact GM-CSF analysis by RP-LC-UV/ESI-MS

The early process sample (DSP3-SEP46) and a late process sample from the same batch (DSP3-SEP99B) were analyzed using an Agilent 1290 Infinity II LC system with a variable wavelength detector coupled to a Bruker Maxis Impact mass spectrometer with ESI-QTOF capabilities. The samples were separated on a an ACQUITY UPLC Protein BEH C4 column with pore size 300Å, particle size 1.7 µm and column dimensions 2.1 mm x 150 mm (Waters), operated at 60 °C column oven temperature. A linear gradient of solvent A (0.1% FA in UHQ water) and solvent B (0.1% FA in ACN) was applied for 30 min from 36-56% phase B at a flow rate of 0.2 mL/ min followed by column washing at high percentage solvent B and column reconditioning. MS data was acquired in the range 500-3000 m/z and the data was processed in DataAnalysis (Bruker) using the MaxEnt algorithm for deconvolution.

7.3.7 Peptide mapping by RP-LC-ESI-QTOF-MS

The digests of the GM-CSF early process sample were analyzed using an Exion LC system coupled to a SCIEX x500b mass spectrometer with ESI-QTOF capabilities. 1.2 µg GM-CSF was loaded of the non-

treated and borane pyridine complex treated samples. 5 µg GM-CSF was loaded of the DNPH treated samples. The samples were separated on an Xselect CSH C18 XP column with pore size 130Å, particle size 2.5 µm and column dimensions 2.1 mm x 150 mm (Waters), operated at 60 °C column oven temperature. Solvent A (0.1% FA in UHQ water) and B (0.1% FA in ACN) were used to first wash the digest in 1% solvent B for 6 min, letting the flow thorough go to wase, before eluting the peptides into the mass spectrometer using a linear gradient going to 50% solvent B in 26 min, using a flow of 0.2 mL/ min. This was followed by column washing at high percentage solvent B and column reconditioning. The mass spectrometer was operated using positive polarity with a cycle time of 1.2 s. MS data was acquired in the range 300-1800 m/z with 0.5 s accumulation time. MS/MS data was acquired in the range 130-2000 m/z using IDA mode on a maximum of 13 candidate ions with charge states 2-5 and with intensities exceeding 200 cps.

7.3.8 Peptide mapping by RP-LC-ESI-TripleTOF-MS

Digests of the MVK or crotonaldehyde derivatized GM-CSF were analyzed using an Eksigent LC system coupled to a SCIEX 6600 mass spectrometer with ESI-TripleTOF capabilities. 1 µg GM-CSF was loaded. The samples were separated on a nanoEase M/Z CSH column with pore size 130Å, particle size 1.7 µm and column dimensions 300 µm x 150 mm (Waters), operated at 60 °C column oven temperature. Solvent A (0.1% FA in UHQ water) and B (0.1% FA in ACN) were used to first equilibrate the column for 2 min in 5 % solvent B before eluting the peptides into the mass spectrometer using a linear gradient going to 27% phase B in 23 min, using a flow of 5 µL/ min. This was followed by column washing at high percentage phase B and column reconditioning. The mass spectrometer was operated using positive polarity with a cycle time of 1.3 s. MS data was acquired in the range 300-1700 m/z with 0.2 s accumulation time. MS/MS data was acquired in the range 130-2000 m/z using IDA mode on a maximum of 25 candidate ions with charge states 2-5 and with intensities exceeding 100 cps. Former candidate ions were excluded for 3 s after 1 occurrence.

7.3.9 Peptide mapping data processing

The raw files were converted to .mgf files using ProteoWizard's program MS convert (version 3.0.18204 64-bit). Peptides were identified by searching the .mgf files using the Mascot probability-based search engine (version 2.4) against a database containing 10 sequence variants of GM-CSF. Variable modifications included were C_4H_6O, C_4H_8O, $C_{10}H_{10}N_4O_4$, $C_3H_2O_2$, C_3H_4O and $C_9H_{16}O_2$ on amino acids Cys, His, Lys and peptide N-terminal, peptide N-terminal Glu pyroglutamination and Met oxidation. LysC + GluC were added as digestion enzymes including C-terminal digestion of the amino acids Lys, Asp and Glu which were not proceeded by a Pro residue. Up to 4 missed cleavages were accepted, peptide charges from +2 to +4 were included, peptide tolerance settings were ± 15 ppm, MS/MS tolerance settings were 0.05 Da and ESI-QUAD-TOF was selected as instrument. The data was then quantitatively analyzed in Skyline (MacCross lab, version 19.1.0.193) using a spectral library built from .dat files generated from the Mascot search results.

All masses in Table 6 were calculated using the "Molecular Weight Calculator" provided by the Pacific Northwest National Laboratory website (https://omics.pnl.gov/software/molecular-weight-calculator).

7.4 Methods used in chapters 5.2.3 and 5.2.4 to set up a MAM for rhGM-CSF

7.4.1 GM-CSF sample description

The GM-CSF samples used for the study were 1) a late process sample of GM-CSF with no or very low degrees of the +70 Da modification determined by intact LC-MS (DSP5-SEP99 described in chapter 7.3.1), 2) an early process sample of GM-CSF in which the +70 Da modified protein was identified in chapter 5.2.1 (DSP3-SEP46 described in chapter 7.3.1), 3) a GM-CSF refolded product stored in 0.6 M

GdnHCl, DTT, L-arginine, EDTA, Tris, Cysteine/ cystine and HCl with 57% host cell protein content according to mass spectrometry-based analysis (GEMA-010-6B1), protein concentration: 0.48 µg/µl, as measured by amino acid analysis (described in chapter 7.2.2).

7.4.2 Forced GM-CSF degradation at high pH and elevated temperature

The buffer was changed on 300 µg of the late process sample of GM-CSF (DSP5-SEP99) to 50 mM ammonium bicarbonate, pH 8.3 using Vivaspin 500 concentrators, 5 kDa MWCO according to the product protocol and the volume was adjusted to 300 µl. The sample was split into three vials which were incubated in a heating block held at 40 ˚C for 1, 2 or 3 weeks. The samples were then prepared according to the protocol described in chapter 7.4.8.

7.4.3 Inducing disulfide scrambling by incubation at 70 ˚C

10 µg GM-CSF (DSP3-SEP46 or DSP5-SEP99) in 100 mM phosphate buffer, pH 7, was incubated for 3 h in a heating block held at 70 ˚C. The stressed sample was then prepared according to chapter 7.4.4 when digested with GluC and LysC and according to chapter 7.4.7 when digested with GluC and Chymotrypsin.

7.4.4 Sample preparation for MAM assay using GluC and LysC for digestion

10 µg of GM-CSF (DSP3-SEP46 and DSP5-SEP99) was denatured and alkylated in 6 M urea and 5 mM iodoacetamide in 100 mM phosphate buffer, pH 7, for 1h at 30 ˚C protected from light. The sample was diluted to 0.8 M urea in phosphate buffer, followed by digestion with GluC (sequencing grade, Promega), using 1:25 enzyme to protein ratio, and LysC (Lysyl EndopeptidaseR; FUJIFILM Wako Pure Chemical Corporation), using 1:10 enzyme to protein ratio, in 30 ˚C overnight. The digest was stopped with 1% TFA and data was acquired according to protocol 7.3.7.

7.4.5 Sample preparation by Pepsin digestion

10 µg of GM-CSF (DSP3-SEP46 and DSP5-SEP99) was denatured and reduced in 6 M urea and 5 mM DTT in 50 mM NaP buffer with pH 7.0 for 1 h at 30℃. The free cysteines were then alkylated by addition of 10 mM iodoacetamide followed by incubation for 45 min at room temperature protected from light. The sample was diluted to 0.8 M urea by addition of UHQ water and 1% acetic acid was added resulting in a pH of 3. The protein was then digested with a Pepsin (Promega) to protein ratio of 1:50 for 2 h at 30℃. Sample data was acquired according to the protocol described in chapter 7.3.7.

7.4.6 Sample preparation by Thermolysin digestion

50 µg of GM-CSF (DSP3-SEP46 and DSP5-SEP99) was denatured and reduced in 6 M urea and in 100 mM Sodium Acetate buffer with pH 5.0 for 1 h at 30℃. The free cysteines were then alkylated by addition of 5 mM N-ethylmaleimide followed by incubation for 45 min at room temperature. The sample was diluted to 1.5 M urea by addition of buffer followed by protein digestion using a Thermolysin (Sigma Aldrich) to protein ratio of 1:25 for 2 h at 30℃. The digestion was terminated by the addition of 1% TFA and sample data was acquired according to protocol described in chapter 7.3.7.

7.4.7 Preparation of sample with induced disulfide scrambling using enzymes GluC and Chymotrypsin for digestion

The GM-CSF (DSP3-SEP46) stressed according to chapter 7.4.3 was denatured in 6.4 M urea in 100 mM phosphate buffer, pH 7, for 1h at 30 ˚C. This was followed by alkylation of free cysteines in 5 mM iodoacetamide for 1 h at room temperature protected from light. The sample was diluted in buffer to 1.1 M urea before precipitating the protein in -20 ˚C acetone, with a sample to acetone ratio of 1:4, overnight at -20 ˚C. Samples were vortexed for a few seconds before centrifugation for 10 minutes at 4 ˚C at 14,000 g. The supernatant was removed, and the pellet washed twice with 300 µl acetone/EtOH/H2O 2:2:1, spinning the pellet down for 10 minutes at 4 ˚C at 14,000 g before removing

the supernatant. The pellet was left to dry for ~15 minutes. The pellet was dissolved in 9 µl of 8 M urea and 91 µl phosphate buffer, followed by digestion with GluC, using 1:25 enzyme to protein ratio, and Chymotrypsin (Roche), using 1:100 enzyme to protein ratio, in 30 ˚C for 2 h. The digest was stopped with 1% TFA.

7.4.8 Sample preparation for MAM assay using GluC and Chymotrypsin for digestion

10 µg of GM-CSF was denatured and alkylated in 6 M urea and 5 mM iodoacetamide in 100 mM phosphate buffer, pH 7, for 1h at 30 ˚C protected from light. The sample buffer was then changed to 0.8 M urea in 100mM phosphate buffer using Zeba Spin Desalting Columns, 7K MWCO, 0.5 mL (Thermo Fisher), according to the product protocol, followed by digestion with GluC, using 1:25 enzyme to protein ratio, and Chymotrypsin, using 1:100 enzyme to protein ratio, in 30 ˚C for 2 h. The digest was stopped with 1% TFA and data was acquired according to chapter 7.4.11.

7.4.9 Sample preparation of a refolded GM-CSF product

The buffer of 35 µg (total protein mass) of the refolded GM-CSF product (GEMA-010-6B1) was changed to the denaturation buffer, 6 M urea in 100 mM NaP, pH 7, using Zeba Spin desalting columns before the sample preparation was performed according to the protocol described in chapter 7.4.8. 1 µg total protein was then analyzed by RP-LC-ESI-TripleTOF-MS.

7.4.10 Disulfide bond reduction

A fraction of the digests was reduced in 110 mM TCEP incubated at 95 ˚C for 5 min.

7.4.11 Sample cleanup on trap coupled online to peptide mapping by RP-LC-ESI-TripleTOF-MS

LC-MS acquisition was performed on an Eksigent system coupled to a SCIEX TripleTof 6600 mass spectrometer. 10 µl of 0.1 µg/µl digested GM-CSF was loaded on a 0.3 mm trap cartridge, ChromXP, C18CL, 5µm 120Å (Eksigent) coupled online to the analytical column 3C18-CL-120, 3µm, 120Å, 0.3 x 150 mm (Eksigent). The sample was washed for 6 min on the trap in phase A (0.1% FA in UHQ water) at flow rate 5 µl/min and room temperature before being eluted on the analytical column held at 30°C at the same flow rate. The analytical column was equilibrated for 2 min at 5% solvent B before applying a linear gradient of 5-27% solvent B over 23 min. This was followed by a steeper gradient of 27-50% solvent B in 3 min, up to 80% solvent B in 1.5 min followed by column washing and reconditioning.

DDA MS1 data was recorded in the range 300-1700 m/z with an accumulation time of 0.20 s and using 4-time bins to sum. MS2 acquisition was performed in information dependent mode (IDA) on charge states 2-5 exceeding 100 cps on a maximum of 25 candidate ions and excluding former candidate ions for 3 s after 1 occurrence, MS/MS scan range 130-2000 m/z. The total cycle time was 1.3 s.

7.4.12 Data analysis

7.4.12.1 Data analysis and quantification of disulfide scrambling in samples digested with GluC and Chymotrypsin

.wiff files from non-reduced samples acquired by DDA were converted to .mgf files with assigned precursor charge states files using ProteinPilot (version 5.0.1). The Mascot probability-based search engine (version 2.7.0.1) was then used to search the .mgf files against a protein database containing only the GM-CSF sequence (https://www.uniprot.org/uniprot/P04141) without the signal peptide. The database was searched using a crosslinking method, allowing disulfide linkages between maximum two peptides with at least two amino acids, both intra-links and loop links included. Carbamidomethyl on Cys, Gln pyroglutamination and Met oxidation were added as variable modifications. LysC + Chymotrypsin were added as digestion enzymes including C-terminal digestion of the amino acids Glu, Phe, Leu, Trp and Tyr which were not proceeded by a Pro residue. Up to 9 missed cleavages were accepted, peptide charges from +2 to +4 were included, peptide tolerance

settings were ± 15 ppm, MS/MS tolerance settings were 0.05 Da and ESI-QUAD-TOF was selected as instrument. All disulfide linked peptides and peptides with a carbamidomethylated cysteine residue identified by Mascot were included in a peptide library built manually in Skyline including charge states 2-6. The precursor ions from the .wiff files from the non-reduced and reduced samples were searched against this library. Peaks that were minimum 5% of the base peak were included and a centroid mass analyzer and a mass accuracy of 20 ppm were used for MS1 filtering. The disulfide linked events identified in the non-reduced samples were confirmed by an absence of the peak in the reduced data, as well as by manual inspection. The data was extracted from Skyline and the relative degree of each scrambling event was quantified by summing the XIC areas from all peptides including that scrambling event and calculating the ratio between this area and the summed XIC areas from all disulfide linked peptides. An idot product cutoff value of 0.8 was used for the quantification.

7.4.12.2 Data analysis and quantification of disulfide scrambling in samples digested with GluC and LysC

.wiff files from reduced samples acquired by DDA were converted to .mgf files using ProteoWizard's MS convert program (version 3.0.18204 64-bit). The Mascot search engine (version 2.4) was used to search the .mgf files against a protein database containing 10 sequence variants of the GM-CSF protein including Ile-Val substitution and sequences including a translation initiating N-terminal Met residue. Variable modifications included were carbamidomethyl on Cys, C_4H_6O on Lys and peptide N-terminals, Gln pyroglutamination and Met oxidation. LysC + GluC were added as digestion enzymes including C-terminal digestion of amino acids Lys, Asp and Glu which were not proceeded by a Pro residue. Up to 4 missed cleavages were accepted, peptide charges from +2 to +4 were included, peptide tolerance settings were ± 15 ppm, MS/MS tolerance settings were 0.05 Da and ESI-QUAD-TOF was selected as instrument. A library was built in Skyline using the DDA workflow from .dat files generated from the Mascot search results and this was used to search the .wiff files from the reduced digests. A peptide library of possible disulfide linked peptides was then built manually in Skyline by selecting the cysteine containing peptide with the highest XIC area for each site from the reduced digest and pair these in all possible combinations of two peptides, charge states 2-6 were included. This library was used to search the .wiff files from the non-reduced digests. Peaks that were minimum 5% of the base peak were included, a centroid mass analyzer and a mass accuracy of 10 ppm was used for MS1 filtering. The data was extracted from Skyline and the relative degree of each scrambling event was quantified by summing the precursor XIC areas from all peptides including that scrambling event and calculating the ratio between this area and the summed precursor XIC areas from all disulfide linked peptides. An idot product cutoff value of 0.8 was used for peptide filtration.

This workflow was used initially since at the time when this project started, Mascot had not yet introduced the possibility to search for cross-linking events.

7.4.12.3 Data analysis and quantification of the +70 Da modification, Ile to Val substitution, Asn deamidation and Met oxidation

.mgf files from non-reduced samples acquired by DDA were searched in the Mascot search engine (version 2.7.0.1) against a protein database containing 10 sequence variants of the GM-CSF protein including Ile-Val substitution and sequences including a translation initiating N-terminal Met residue. Variable modifications included were carbamidomethyl on Cys, C_4H_6O on Lys and peptide N-terminals, Gln pyroglutamination, Asn deamidation and Met oxidation. LysC + GluC or Chymotrypsin + GluC were added as digestion enzymes according to the enzyme specifications and number of missed cleavages described in chapter 7.4.12.1 and 7.4.12.2. The rest of the search parameters were set as described in chapter 7.4.12.2. A library was then built in Skyline using the DDA workflow from .dat files from the Mascot search results using a cut-off score of 0.95. Modifications contained in disulfide linked peptides were added manually and were verified from the crosslinking search in Mascot described in chapter

7.4.12.1. Precursor ion charge states of 2-6 were included. The .wiff files from the non-reduced digests were searched against this library. For MS1 filtering, peaks that were minimum 5% of the base peak were included and a centroid mass analyzer and a mass accuracy of 20 ppm was used. For each modification site of interest, the peptide with the highest XIC area containing the site and with a good peak shape by visual inspection was selected for quantification of the modification degree. Peptides with as few modification sites as possible were also aimed for. All peptide sequences selected for quantification are summarized in Table 8. Precursor ions with one charge more and less than the identified charge state were added manually for better confirmation of the peptide identification and it was made sure that the same charge states were included for both the modified and the non-modified variant of the same peptide sequence. The modification degree was quantified by extracting the data from Skyline and calculating the relative area of the summed precursor XIC areas from the modified peptide in relation to the summed precursor XIC areas from the non-modified and modified peptides. An idot product cutoff value of 0.8 was used for the quantification.

In GM-CSF digested by GluC and LysC the +70 Da modification degree was quantified by summing the precursor XICs from the peptide with the highest XIC area containing the site and dividing by the total precursor XIC area of all peptides containing the site. The reason for this was that for some sites only the modified variant could be found for some peptide sequences while the site could be found non-modified in other peptide sequences. The reason for this was inhibition of digestion at modified sites and it was initially thought that the quantification issue that followed from this could be overcome by just including all peptides containing the site in the calculation. In the end it was however found that the signal loss from short peptides without the modification was so high that the quantification was not accurate using this strategy.

8. References

1. The Consortium for Top Down Proteomics, Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. Nat Methods. 2013 Mar;10(3):186–7.

2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004 Oct 21;431:931–45.

3. Leader B, Baca QJ, Golan DE. Protein therapeutics: a summary and pharmacological classification. Nat Rev Drug Discov. 2008 Jan;7(1):21–39.

4. Soukharev S, Hammond D, Ananyeva NM, Anderson JAM, Hauser CAE, Pipe S, et al. Expression of Factor VIII in Recombinant and Transgenic Systems. Blood Cells, Molecules, and Diseases. 2002 Mar;28(2):234–48.

5. The Business Research Company. Therapeutic Proteins Global Market Report 2020-30: COVID-19 Impact and Recovery [Internet]. 2020 Oct. Report No.: ID: 4871552. Available from: https://www.researchandmarkets.com/reports/4871552

6. Shire SJ. Introduction. In: Monoclonal Antibodies [Internet]. Elsevier; 2015 [cited 2021 Feb 22]. p. 1–15. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780081002964000014

7. Suzuki M, Kato C, Kato A. Therapeutic antibodies: their mechanisms of action and the pathological findings they induce in toxicity studies. J Toxicol Pathol. 2015;28(3):133–9.

8. Zafir-Lavie I, Michaeli Y, Reiter Y. Novel antibodies as anticancer agents. Oncogene. 2007 May;26(25):3714–33.

9. Francisco-Cruz A, Aguilar-Santelises M, Ramos-Espinosa O, Mata-Espinosa D, Marquina-Castillo B, Barrios-Payan J, et al. Granulocyte–macrophage colony-stimulating factor: not just another haematopoietic growth factor. Med Oncol. 2014 Jan;31(774):1–14.

10. Tazawa R, Trapnell BC, Inoue Y, Arai T, Takada T, Nasuhara Y, et al. Inhaled Granulocyte/Macrophage–Colony Stimulating Factor as Therapy for Pulmonary Alveolar Proteinosis. Am J Respir Crit Care Med. 2010 Jun 15;181(12):1345–54.

11. Wong GG, Witek JS, Temple PA, Wilkens KM, Leary AC, Luxenberg DP, et al. Human GM-CSF: Molecular Cloning of the Complementary DNA and Purificaton of the Natural and Recombinant Proteins. 1985 May;228:810–5.

12. Forno G, Bollati Fogolin M, Oggero M, Kratje R, Etcheverrigaray M, Conradt HS, et al. N- and O-linked carbohydrates and glycosylation site occupancy in recombinant human granulocyte-macrophage colony-stimulating factor secreted by a Chinese hamster ovary cell line: N- and O-glycosylation of rhGM-CSF. European Journal of Biochemistry. 2004 Feb 9;271(5):907–19.

13. Zhou F, Wang M-L, Albert HH, Moore PH, Zhu YJ. Efficient transient expression of human GM-CSF protein in Nicotiana benthamiana using potato virus X vector. Appl Microbiol Biotechnol. 2006 Oct;72(4):756–62.

14. Schwanke RC, Renard G, Chies JM, Campos MM, Junior ELB, Santos DS, et al. Molecular cloning, expression in Escherichia coli and production of bioactive homogeneous recombinant human granulocyte and macrophage colony stimulating factor. International Journal of Biological Macromolecules. 2009 Aug;45(2):97–102.

15. Sahdev S, Khattar SK, Saini KS. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem. 2007 Nov 29;307(1–2):249–64.

16. Marichal-Gallardo PA, Álvarez MM. State-of-the-art in downstream processing of monoclonal antibodies: Process trends in design and validation. Biotechnol Progress. 2012 Jul;28(4):899–916.

17. Lin TJ, Beal KM, Brown PW, DeGruttola HS, Ly M, Wang W, et al. Evolution of a comprehensive, orthogonal approach to sequence variant analysis for biotherapeutics. mAbs. 2019 Jan 2;11(1):1–12.

18. Cymer F, Beck H, Rohde A, Reusch D. Therapeutic monoclonal antibody N-glycosylation – Structure, function and therapeutic potential. Biologicals. 2018 Mar;52:1–11.

19. Liu L. Antibody Glycosylation and Its Impact on the Pharmacokinetics and Pharmacodynamics of Monoclonal Antibodies and Fc-Fusion Proteins. Journal of Pharmaceutical Sciences. 2015 Jun;104(6):1866–84.

20. van Beers MMC, Bardor M. Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. Biotechnology Journal. 2012 Dec;7(12):1473–84.

21. Berkmen M. Production of disulfide-bonded proteins in Escherichia coli. Protein Expression and Purification. 2012 Mar;82(1):240–51.

22. Lakbub JC, Shipman JT, Desaire H. Recent mass spectrometry-based techniques and considerations for disulfide bond characterization in proteins. Anal Bioanal Chem. 2018 Apr;410(10):2467–84.

23. Thomson CA, Olson M, Jackson LM, Schrader JW. A Simplified Method for the Efficient Refolding and Purification of Recombinant Human GM-CSF. Abrams WR, editor. PLoS ONE. 2012 Nov 14;7(11):1–6.

24. Weinfurtner D. Analysis of Disulfide Bond Formation in Therapeutic Proteins. In: Oxidative Folding of Proteins: Basic Principles, Cellular Regulation and Engineering [Internet]. Royal Society of Chemistry; 2018. p. 81–98. (Chemical Biology). Available from: https://pubs.rsc.org/en/content/chapter/bk9781782629900-00081/978-1-78262-990-0

25. Hao P, Adav SS, Gallart-Palau X, Sze SK. Recent advances in mass spectrometric analysis of protein deamidation. Mass Spec Rev. 2017 Nov;36(6):677–92.

26. Patananan AN, Capri J, Whitelegge JP, Clarke SG. Non-repair Pathways for Minimizing Protein Isoaspartyl Damage in the Yeast Saccharomyces cerevisiae. Journal of Biological Chemistry. 2014 Jun;289(24):16936–53.

27. O'Connor PB, Cournoyer JJ, Pitteri SJ, Chrisman PA, McLuckey SA. Differentiation of aspartic and isoaspartic acids using electron transfer dissociation. J Am Soc Mass Spectrom. 2006 Jan;17(1):15–9.

28. Zhong X, Wright JF. Biological Insights into Therapeutic Protein Modifications throughout Trafficking and Their Biopharmaceutical Applications. International Journal of Cell Biology. 2013;2013:1–19.

29. Lao YW, Gungormusler-Yilmaz M, Shuvo S, Verbeke T, Spicer V, Krokhin OV. Chromatographic behavior of peptides containing oxidized methionine residues in proteomic LC–MS experiments: Complex tale of a simple modification. Journal of Proteomics. 2015 Jul;125:131–9.

30. Torosantucci R, Schöneich C, Jiskoot W. Oxidation of Therapeutic Proteins and Peptides: Structural and Biological Consequences. Pharm Res. 2014 Mar;31(3):541–53.

31. Lam XM, Yang JY, Cleland JL. Antioxidants for Prevention of Methionine Oxidation in Recombinant Monoclonal Antibody HER2. Journal of Pharmaceutical Sciences. 1997 Nov;86(11):1250–5.

32. European Medicines Agency. ICH guideline Q11 on development and manufacture of drug substances (chemical entities and biotechnological/biological entities), Step 3 [Internet]. 2011. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-guideline-q11-development-manufacture-drug-substances-chemical-entities-biotechnological/biological-entities_en.pdf

33. Beck A, Wagner-Rousset E, Ayoub D, Van Dorsselaer A, Sanglier-Cianférani S. Characterization of Therapeutic Antibodies and Related Products. Anal Chem. 2013 Jan 15;85(2):715–36.

34. Schaffer LV, Millikin RJ, Miller RM, Anderson LC, Fellers RT, Ge Y, et al. Identification and Quantification of Proteoforms by Mass Spectrometry. Proteomics. 2019 May;19(10):1800361.

35. Hoffmann E de, Stroobant V. Mass spectrometry: principles and applications. 3rd ed. Chichester, West Sussex, England; Hoboken, NJ: J. Wiley; 2007. 305–324 p.

36. Le-Minh V, Halgand F, Van der Rest G, Taverna M, Smadja C. Conformation assessment of therapeutic monoclonal antibodies by SEC-MS: Unravelling analytical biases for application to quality control. Journal of Pharmaceutical and Biomedical Analysis. 2020 Jun;185:113252.

37. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature. 2016 Sep;537(7620):347–55.

38. Dupree EJ, Jayathirtha M, Yorkey H, Mihasan M, Petre BA, Darie CC. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. Proteomes. 2020 Jul 6;8(3):14.

39. Kertesz-Farkas A, Reiz B, P. Myers M, Pongor S. Database Searching in Mass Spectrometry Based Proteomics. CBIO. 2012 May 1;7(2):221–30.

40. Fenn J, Mann M, Meng C, Wong S, Whitehouse C. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989 Oct 6;246(4926):64–71.

41. Kebarle P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. J Mass Spectrom. 2000;15.

42. Shukla AK, Futrell JH. Tandem mass spectrometry: dissociation of ions by collisional activation. J Mass Spectrom. 2000;1069–90.

43. Roepstorff P, Fohlman J. Letter to the editors. Biol Mass Spectrom. 1984 Nov;11(11):601–601.

44. Johnson D, Boyes B, Fields T, Kopkin R, Orlando R. Optimization of Data-Dependent Acquisition Parameters for Coupling High-Speed Separations with LC-MS/MS for Protein Identifications. J Biomol Tech. 2013 Jul;jbt.13-2402-003.

45. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Molecular Systems Biology. 2018 Aug;14(8):e8126.

46. Pappireddi N, Martin L, Wühr M. A Review on Quantitative Multiplexed Proteomics. ChemBioChem. 2019 May 15;20(10):1210–24.

47. Hu A, Noble WS, Wolf-Yadlin A. Technical advances in proteomics: new developments in data-independent acquisition. F1000Res. 2016 Mar 31;5:419.

48. Lam H, Aebersold R. Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics. Methods. 2011 Aug;54(4):424–31.

49. Schilling B, Rardin MJ, MacLean BX, Zawadzka AM, Frewen BE, Cusack MP, et al. Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline. Molecular & Cellular Proteomics. 2012 May;11(5):202–14.

50. Xu X, Qiu H, Li N. LC-MS multi-attribute method for characterization of biologics. Journal of Applied Bioanalysis. 2017 Apr 3;3(2):21–5.

51. Rogers RS, Abernathy M, Richardson DD, Rouse JC, Sperry JB, Swann P, et al. A View on the Importance of “Multi-Attribute Method” for Measuring Purity of Biopharmaceuticals and Improving Overall Control Strategy. AAPS J. 2018 Jan;20(1):7.

52. Rogers RS, Nightlinger NS, Livingston B, Campbell P, Bailey R, Balland A. Development of a quantitative mass spectrometry multi-attribute method for characterization, quality control testing and disposition of biologics. mAbs. 2015 Sep 3;7(5):881–90.

53. Rogstad S, Yan H, Wang X, Powers D, Brorson K, Damdinsuren B, et al. Multi-Attribute Method for Quality Control of Therapeutic Proteins. Anal Chem. 2019 Nov 19;91(22):14170–7.

54. Wang Y, Li X, Liu Y-H, Richardson D, Li H, Shameem M, et al. Simultaneous monitoring of oxidation, deamidation, isomerization, and glycosylation of monoclonal antibodies by liquid chromatography-mass spectrometry method with ultrafast tryptic digestion. mAbs. 2016 Nov 16;8(8):1477–86.

55. Buettner A, Maier M, Bonnington L, Bulau P, Reusch D. Multi-Attribute Monitoring of Complex Erythropoetin Beta Glycosylation by GluC Liquid Chromatography–Mass Spectrometry Peptide Mapping. Anal Chem. 2020 Jun 2;92(11):7574–80.

56. Zhang Z, Shah B, Guan X. Reliable LC-MS Multiattribute Method for Biotherapeutics by Run-Time Response Calibration. Anal Chem. 2019 Apr 16;91(8):5252–60.

57. Millán-Martín S, Jakes C, Carillo S, Buchanan T, Guender M, Kristensen DB, et al. Inter-laboratory study of an optimised peptide mapping workflow using automated trypsin digestion for monitoring monoclonal antibody product quality attributes. Anal Bioanal Chem. 2020 Oct;412(25):6833–48.

58. Mouchahoir T, Schiel JE, Rogers R, Heckert A, Place BJ, Ammerman A, et al. New Peak Detection Performance Metrics from the MAM Consortium Interlaboratory Study. J Am Soc Mass Spectrom. 2021 Mar 12;jasms.0c00415.

59. Sun S, Zhou J-Y, Yang W, Zhang H. Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. Analytical Biochemistry. 2014 Feb;446:76–81.

60. Dong Q, Liang Y, Yan X, Markey SP, Mirokhin YA, Tchekhovskoi DV, et al. The NISTmAb tryptic peptide spectral library for monoclonal antibody characterization. mAbs. 2018 Apr 3;10(3):354–69.

61. Liu Y, Fernandez J, Pu Z, Zhang H, Cao L, Aguilar I, et al. Simultaneous Monitoring and Comparison of Multiple Product Quality Attributes for Cell Culture Processes at Different Scales Using a LC/MS/MS Based Multi-Attribute Method. Journal of Pharmaceutical Sciences. 2020 Nov;109(11):3319–29.

62. Wang Y, Lu Q, Wu S-L, Karger BL, Hancock WS. Characterization and Comparison of Disulfide Linkages and Scrambling Patterns in Therapeutic Monoclonal Antibodies: Using LC-MS with Electron Transfer Dissociation. Anal Chem. 2011 Apr 15;83(8):3133–40.

63. European Medicines Agency. ICH Topic Q 6 B, Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products. 1999.

64. Mouchahoir T, Schiel JE. Development of an LC-MS/MS peptide mapping protocol for the NISTmAb. Analytical and Bioanalytical Chemistry. 2018 Mar;410(8):2111–26.

65. Parker MJ, Spencer J, Clarke AR. An Integrated Kinetic Analysis of Intermediates and Transition States in Protein Folding Reactions. Journal of Molecular Biology. 1995 Nov;253(5):771–86.

66. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods. 2014 Mar;11(3):319–24.

67. Overgaard M, Cangemi C, Jensen ML, Argraves WS, Rasmussen LM. Total and isoform-specific quantitative assessment of circulating fibulin-1 using selected reaction monitoring MS and time-resolved immunofluorometry. Proteomics Clin Appl. 2015;9(7–8):767–75.

68. Moggridge S, Sorensen PH, Morin GB, Hughes CS. Extending the Compatibility of the SP3 Paramagnetic Bead Processing Approach for Proteomics. J Proteome Res. 2018;17(4):1730–40.

69. Meiring HD, van der Heeft E, ten Hove GJ, de Jong APJM. Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. J Sep Sci. 2002;12.

70. Lauber MA, Koza SM, Fountain KJ. Optimizing Peak Capacity in Nanoscale Trap-Elute Peptide Separations with Differential Column Heating. 2014 May;(Application Note No 720005047). Available from: https://www.waters.com/content/dam/waters/en/app-notes/2014/720005047/720005047-en.pdf

71. Yu J, Schorlemer M, Gomez Toledo A, Pett C, Sihlbom C, Larson G, et al. Distinctive MS/MS Fragmentation Pathways of Glycopeptide-Generated Oxonium Ions Provide Evidence of the Glycan Structure. Chem Eur J. 2016 Jan 18;22(3):1114–24.

72. Liu Y-H, Wylie D, Zhao J, Cure R, Cutler C, Cannon-Carlson S, et al. Mass spectrometric characterization of the isoforms in Escherichia coli recombinant DNA-derived interferon alpha-2b. Analytical Biochemistry. 2011 Jan;408(1):105–17.

73. Ichihashi K, Osawa T, Toyokuni S, Uchida K. Endogenous Formation of Protein Adducts with Carcinogenic Aldehydes. J Biol Chem. 2001 Jun 29;276(26):23903–13.

74. von Stedingk H, Davies R, Rydberg P, Törnqvist M. Methyl vinyl ketone—Identification and quantification of adducts to N-terminal valine in human hemoglobin. Journal of Chromatography B. 2010 Oct;878(27):2491–6.

75. Chen Y, Sprung R, Tang Y, Ball H, Sangras B, Kim SC, et al. Lysine Propionylation and Butyrylation Are Novel Post-translational Modifications in Histones. Molecular & Cellular Proteomics. 2007 May;6(5):812–9.

76. Afonso CB, Sousa BC, Pitt AR, Spickett CM. A mass spectrometry approach for the identification and localization of small aldehyde modifications of proteins. Archives of Biochemistry and Biophysics. 2018 May;646:38–45.

77. Carlsson H, Motwani HV, Osterman Golkar S, Törnqvist M. Characterization of a Hemoglobin Adduct from Ethyl Vinyl Ketone Detected in Human Blood Samples. Chem Res Toxicol. 2015 Nov 16;28(11):2120–9.

78. Xu J-Y, Xu Z, Liu X, Tan M, Ye B-C. Protein Acetylation and Butyrylation Regulate the Phenotype and Metabolic Shifts of the Endospore-forming *Clostridium acetobutylicum*. Mol Cell Proteomics. 2018 Jun;17(6):1156–69.

79. Soini J, Falschlehner C, Liedert C, Bernhardt J, Vuoristo J, Neubauer P. Norvaline is accumulated after a down-shift of oxygen in Escherichia coli W3110. Microb Cell Fact. 2008;7(1):30.

80. Zhang Z, Chan PK, Richardson J, Shah B. An evaluation of instrument types for mass spectrometry-based multi-attribute analysis of biotherapeutics. mAbs. 2020 Jan 1;12(1):1783062.

81. Huang Q, Yang L, Luo J, Guo L, Wang Z, Yang X, et al. SWATH enables precise label-free quantification on proteome scale. Proteomics. 2015 Apr;15(7):1215–23.

82. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Molecular & Cellular Proteomics. 2012 Jun;11(6):O111.016717.

83. Xu W, Jimenez RB, Mowery R, Luo H, Cao M, Agarwal N, et al. A Quadrupole Dalton-based multi-attribute method for product characterization, process development, and quality control of therapeutic proteins. mAbs. 2017 Oct 3;9(7):1186–96.

84. Schmidt A, Karas M, Dülcks T. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? J Am Soc Mass Spectrom. 2003 May;14(5):492–500.

85. Zhou C, Schulz BL. Glycopeptide variable window SWATH for improved Data Independent Acquisition glycoproteomics. Analytical Biochemistry. 2020 May;597:113667.

86. Roca LS, Gargano AFG, Schoenmakers PJ. Development of comprehensive two-dimensional low-flow liquid-chromatography setup coupled to high-resolution mass spectrometry for shotgun proteomics. Analytica Chimica Acta. 2021 Apr;1156:338349.

87. Qian C. Fully automated peptide mapping multi-attribute method by liquid chromatographymass spectrometry with robotic liquid handling system. Journal of Pharmaceutical and Biomedical Analysis. 2021;9.

88. Cao M. An Automated and Qualified Platform Method for Site-Specific Succinimide and Deamidation Quantitation Using Low-pH Peptide Mapping. Journal of Pharmaceutical Sciences. 2019;10.

89. Liu Y, Brunner R, Navarro P, Aebersold R. Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. 2013;11.

90. Wu W, Dai R-T, Bendixen E. Comparing SRM and SWATH Methods for Quantitation of Bovine Muscle Proteomes. J Agric Food Chem. 2019 Feb 6;67(5):1608–18.

91. Messner CB, Demichev V, Bloomfield N, Yu JSL, White M, Kreidl M, et al. Ultra-fast proteomics with Scanning SWATH. Nat Biotechnol [Internet]. 2021 Mar 25 [cited 2021 Jun 7]; Available from: http://www.nature.com/articles/s41587-021-00860-4

92. Clanton TL. Hypoxia-induced reactive oxygen species formation in skeletal muscle. Journal of Applied Physiology. 2007 Jun;102(6):2379–88.

93. Yajima D, Motani H, Hayakawa M, Sato Y, Sato K, Iwase H. The relationship between cell membrane damage and lipid peroxidation under the condition of hypoxia-reoxygenation: analysis of the mechanism using antioxidants and electron transport inhibitors. Cell Biochem Funct. 2009 Aug;27(6):338–43.

94. Schulte M, Frick K, Gnandt E, Jurkovic S, Burschel S, Labatzke R, et al. A mechanism to prevent production of reactive oxygen species by Escherichia coli respiratory complex I. Nat Commun. 2019 Dec;10(2551):1–9.

95. Esterbauer H, Schaur RJ, Zollner H. Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. Free Radical Biology and Medicine. 1991 Jan;11(1):81–128.

96. Dalle-Donne I, Aldini G, Carini M, Colombo R, Rossi R, Milzani A. Protein carbonylation, cellular dysfunction, and disease progression. Journal of Cellular and Molecular Medicine. 2006 Apr;10(2):389–406.

97. Liigand P, Kaupmees K, Kruve A. Influence of the amino acid composition on the ionization efficiencies of small peptides. J Mass Spectrom. 2019 Jun;54(6):481–7.

98. Loo JA, Edmonds CG, Udseth HR, Smith RD. Effect of reducing disulfide-containing proteins on electrospray ionization mass spectra. Anal Chem. 1990 Apr 1;62(7):693–8.

99. Albert A, Eksteen JJ, Isaksson J, Sengee M, Hansen T, Vasskog T. General Approach To Determine Disulfide Connectivity in Cysteine-Rich Peptides by Sequential Alkylation on Solid Phase and Mass Spectrometry. Anal Chem. 2016 Oct 4;88(19):9539–46.

100. Tajiri-Tsukada M, Hashii N, Ishii-Watabe A. Establishment of a highly precise multi-attribute method for the characterization and quality control of therapeutic monoclonal antibodies. Bioengineered. 2020 Jan 1;11(1):984–1000.

101. Pérez-Rodriguez S, Ramírez OT, Trujillo-Roldán MA, Valdez-Cruz NA. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis of Chinese hamster ovary cell homogenates. Electronic Journal of Biotechnology. 2020 Nov;48:86–94.

102. Merck's website [Internet]. Available from: https://www.sigmaaldrich.com/SE/en

103. Waters' Safety Data Sheets search engine [Internet]. Available from: https://www.waters.com/waters/en_SE/Safety-Data-Sheets/nav.htm?cid=134855469&locale=en_SE

104. GHS Classification [Internet]. Available from: https://pubchem.ncbi.nlm.nih.gov/ghs/

9. Appendix

Appendix table 1: Quantification of deamidation, oxidation and N-glycoforms in a digest of the adalimumab biosimilar cell culture filtrate using mass spectrometry data acquired by DDA or DIA. Mean values and standard deviations from quantification in the same digest acquired on the same day, N = 3. P-values from an independent samples t-test are displayed and was performed on each peptide measured between the two acquisition methods, assuming unequal variances. Modification percentages measuring >2% were included. Oxidation and deamidation quantities were calculated from the fragment XIC areas from the DIA data and from the precursor XICs from the DDA data. N-glycoform quantities were calculated from precursor XICs from both DIA data and DDA data. See Table 2 and Table 3 for the amino acid sequences of the abbreviated peptides. The amino acid sequence of the glycopeptide is TKPREEQYNSTYR and compositional figures of the N-glycan variants can be found in The G0F species was clearly identified as the most abundant N-glycan in the sample. This was expected since this is often one of the most abundant N-glycan species in monoclonal IgGs. The G1F species was also present with two clear peaks, representing the two isomers, together with the G0, G0F-N and G0-N species. The high mannose species M5 was also identified. High mannose levels have been linked to protein clearance and is therefore particularly important to be able to quantify in therapeutic IgGs. Some other low-abundance glycovariants were also identified. All identified N-glycan species were expected from eukaryotic protein expression.

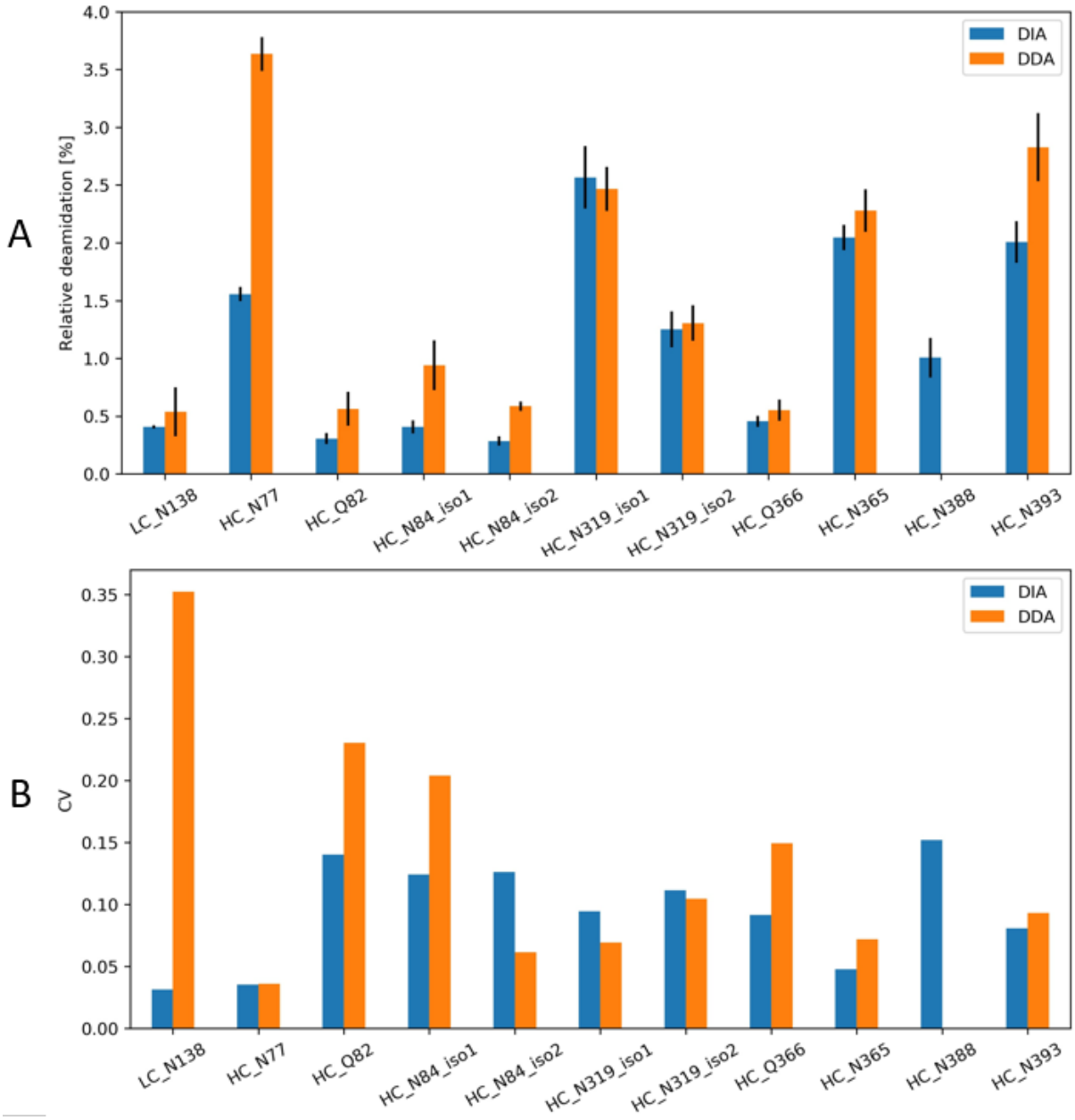
Table 4.

		Mean (%) ± SD DIA	**Mean (%) ± SD DDA**	**Statistical significance**	**$mean_{DIA}$ / $mean_{DDA}$ (%)**
Deamidation site	HC_N77	1.60 ± 0.00060	3.68 ± 0.038	p = 0.0030	-56.51
	HC_N319_iso 1	2.78 ± 0.0076	2.92 ± 0.0016	p = 0.094	-4.6
	HC_N365	2.07 ± 0.0016	2.39 ± 0.00042	p = 0.0012	-13.26
	HC_N393	1.93 ± 0.00083	2.43 ± 0.0023	p = 0.00059	-20.58
Oxidation site	LC_M4	12.66 ± 0.0019	11.71 ± 0.013	p = 0.00091	8.1
	HC_M34	13.18 ± 0.082	15.91 ± 0.067	p = 0.00026	-17.15
	HC_M83	10.41 ± 0.10	11.19 ± 0.0040	p = 0.053	-6.95
	HC_M256	14.33 ± 0.0016	17.46 ± 0.037	p = 0.0013	-17.93
	HC_M432	17.01 ± 0.54	17.04 ± 0.015	p = 0.94	-0.2
N-glycan variant	G0	3.30 + 0.0034	3.70 + 0.0016	p = 0.0019	-8.02
	G0F	68.30 + 0.15	65.81 + 0.067	p = 0.027	3.79
	G0F-N	7.21 + 0.015	7.61 + 0.0034	p = 0.015	-5.23
	G1F	11.80 + 0.045	12.27 + 0.019	p = 0.047	-3.88
	M5	5.57 + 0.0028	5.87 + 0.023	p = 0.081	-5.17

Appendix table 2: Mean values and coefficients of variation (CVs) for quantification of deamidation, oxidation and N-glycoforms between replicate analyses of a digested adalimumab biosimilar cell culture filtrate. The CVs were measured on three replicates levels: 1) the same digest acquired on the same day, N = 3, 2) the same digest acquired on different days, N = 3 and 3) different digests performed on different days and acquired on the same day, N = 5. The digest data was acquired by both DIA and DDA.

		Same digest acquired same day				**Same digest acquired different days**				**Different digests acquired different days**			
		DIA		**DDA**		**DIA**		**DDA**		**DIA**		**DDA**	
		Mean (%)	**CV**	**Mean (%)**	**CV**	**Mean (%)**	**CV**	**Mean (%)**	**CV**	**Mean (%)**	**CV**	**Mean (%)**	**CV**
Deamidation site	LC_N138	0.409	0.052	0.630	0.343	0.420	0.052	0.460	0.301	0.407	0.032	0.538	0.352
	HC_N77	1.601	0.013	3.681	0.043	1.582	0.018	3.721	0.038	1.556	0.035	3.634	0.036
	HC_Q82	0.292	0.097	0.481	0.118	0.281	0.191	0.494	0.019	0.306	0.140	0.564	0.230
	HC_N84_iso1	0.459	0.024	1.002	0.012	0.447	0.032	0.904	0.307	0.408	0.124	0.940	0.204
	HC_N84_iso2	0.267	0.092	0.608	0.059	0.277	0.160	0.607	0.073	0.285	0.126	0.586	0.061
	HC_N319_iso1	2.783	0.026	2.917	0.011	2.922	0.085	2.625	0.076	2.566	0.094	2.467	0.069
	HC_N319_iso2	1.379	0.058	1.465	0.033	1.181	0.142	1.618	0.043	1.251	0.112	1.305	0.105
	HC_Q366	0.490	0.046	0.575	0.161	0.459	0.197	0.598	0.185	0.455	0.091	0.551	0.149
	HC_N365	2.074	0.016	2.390	0.007	2.131	0.016	2.263	0.017	2.047	0.048	2.279	0.072
	HC_N388	1.014	0.164	-	-	1.358	0.110	-	-	1.007	0.152	-	-
	HC_N393	1.926	0.012	2.425	0.016	2.403	0.052	3.236	0.044	2.008	0.081	2.828	0.093
Oxidation site	LC_M4	12.658	0.003	11.710	0.008	14.591	0.041	13.903	0.025	10.672	0.192	10.285	0.196
	HC_M34	13.183	0.018	15.912	0.013	15.589	0.071	18.136	0.054	11.053	0.202	13.585	0.182
	HC_M83	10.412	0.025	11.190	0.005	12.442	0.056	12.786	0.041	8.802	0.192	9.442	0.185
	HC_M256	14.330	0.002	17.461	0.009	15.687	0.041	18.710	0.039	12.191	0.163	14.659	0.162
	HC_M432	17.011	0.035	17.044	0.006	16.815	0.019	19.001	0.040	15.514	0.180	13.674	0.183
N-glycan variant	G0	3.400	0.014	3.697	0.009	3.170	0.008	3.645	0.015	3.259	0.054	3.677	0.033
	G0F	68.302	0.005	65.810	0.003	69.289	0.003	66.698	0.003	68.698	0.012	66.104	0.010
	G0F-N	7.209	0.014	7.607	0.006	7.218	0.011	7.682	0.006	7.108	0.025	7.539	0.016

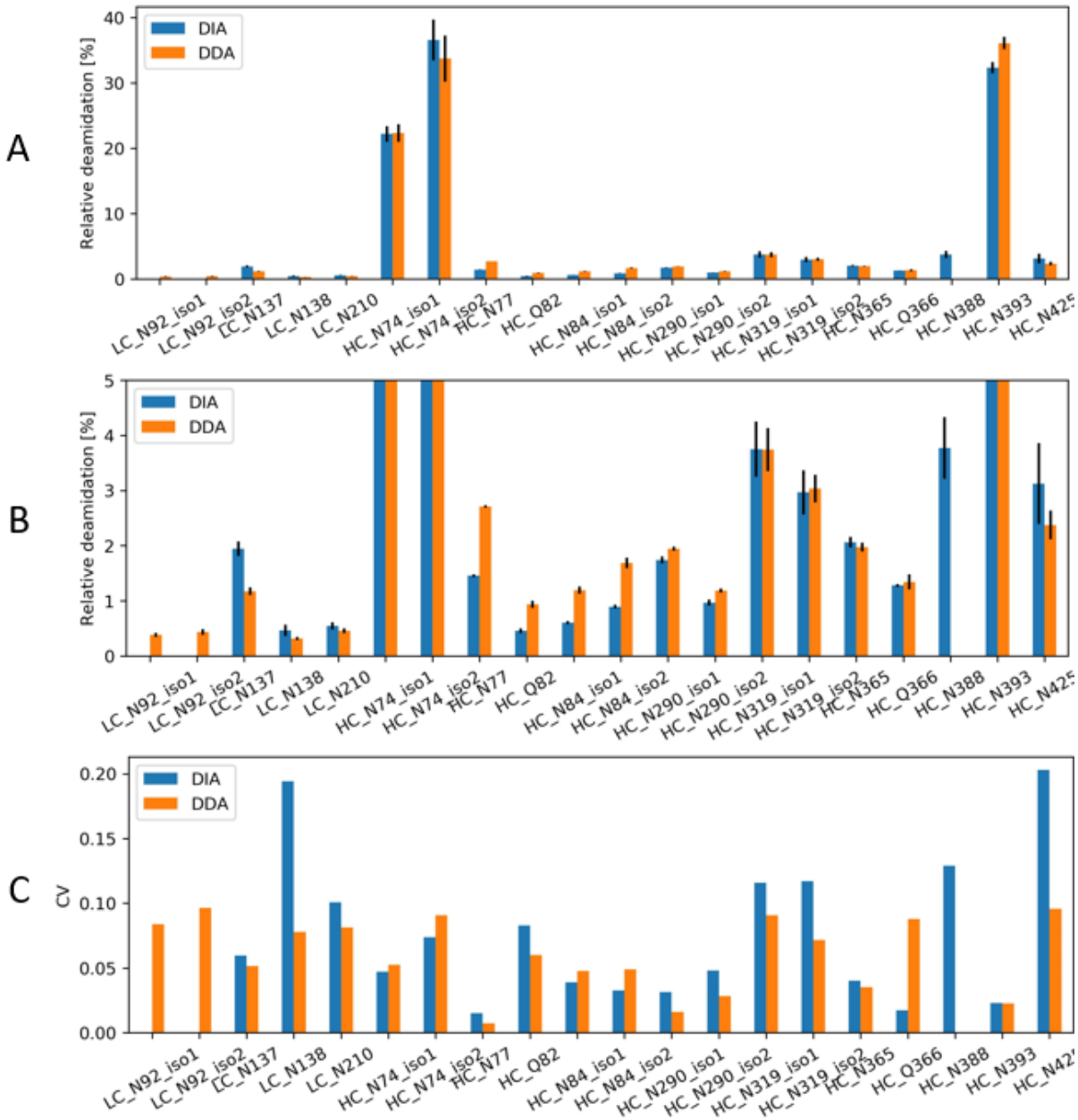
	G1F	11.797	0.015	12.273	0.009	12.019	0.009	12.616	0.011	11.790	0.010	12.246	0.012
	H3N2F1	0.416	0.082	0.459	0.077	-	-	-	-	0.323	0.214	0.476	0.089
	H3N3	1.434	0.032	1.704	0.020	1.366	0.018	1.590	0.005	1.380	0.037	1.592	0.065
	H3N5F1	0.248	0.052	0.373	0.100	0.236	0.019	0.362	0.034	0.251	0.139	0.355	0.169
	H4N3F1	0.556	0.053	0.689	0.078	0.514	0.043	0.656	0.054	0.565	0.061	0.698	0.035
	H4N3S1F1	0.209	0.206	0.343	0.060	-	-	-	-	0.237	0.061	0.349	0.083
	H4N4	0.278	0.083	0.385	0.049	0.278	0.116	0.415	0.107	0.286	0.064	0.392	0.073
	H5N4F1	0.580	0.025	0.786	0.032	0.584	0.026	0.672	0.069	0.554	0.062	0.719	0.093
	M5	5.570	0.008	5.873	0.021	5.325	0.003	5.665	0.004	5.550	0.040	5.852	0.029



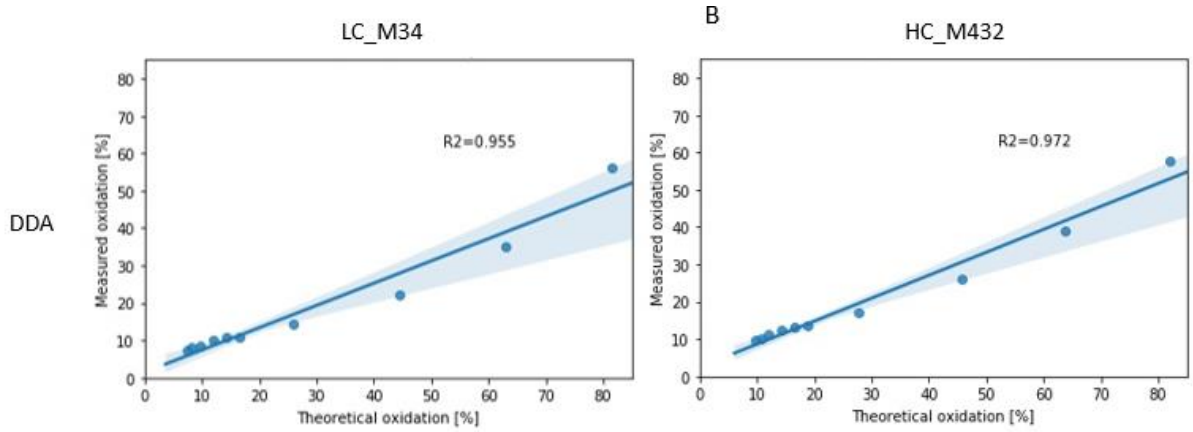
Appendix figure 1: The cell culture filtrate was digested in replicates, N = 5, and analyzed with the MAM assay and acquired by DDA (orange) or DIA (blue). (a) Mean values for quantification of the deamidation degree at the sites, with error bars showing standard deviation. (b) CVs between the deamidation quantification values. Quantities were calculated from fragment XIC areas from the DIA data and from precursor XICs from the DDA data. See Table 3 for the amino acid sequences of the abbreviated peptides.

Appendix table 3: Mean, standard deviation and CV values for quantification of deamidation degrees in digest replicates, N = 4, of the Protein A purified and stressed adalimumab biosimilar at pH 9, 37 ˚C for 2 weeks. The data was acquired and quantified by the DIA or DDA based MS method.

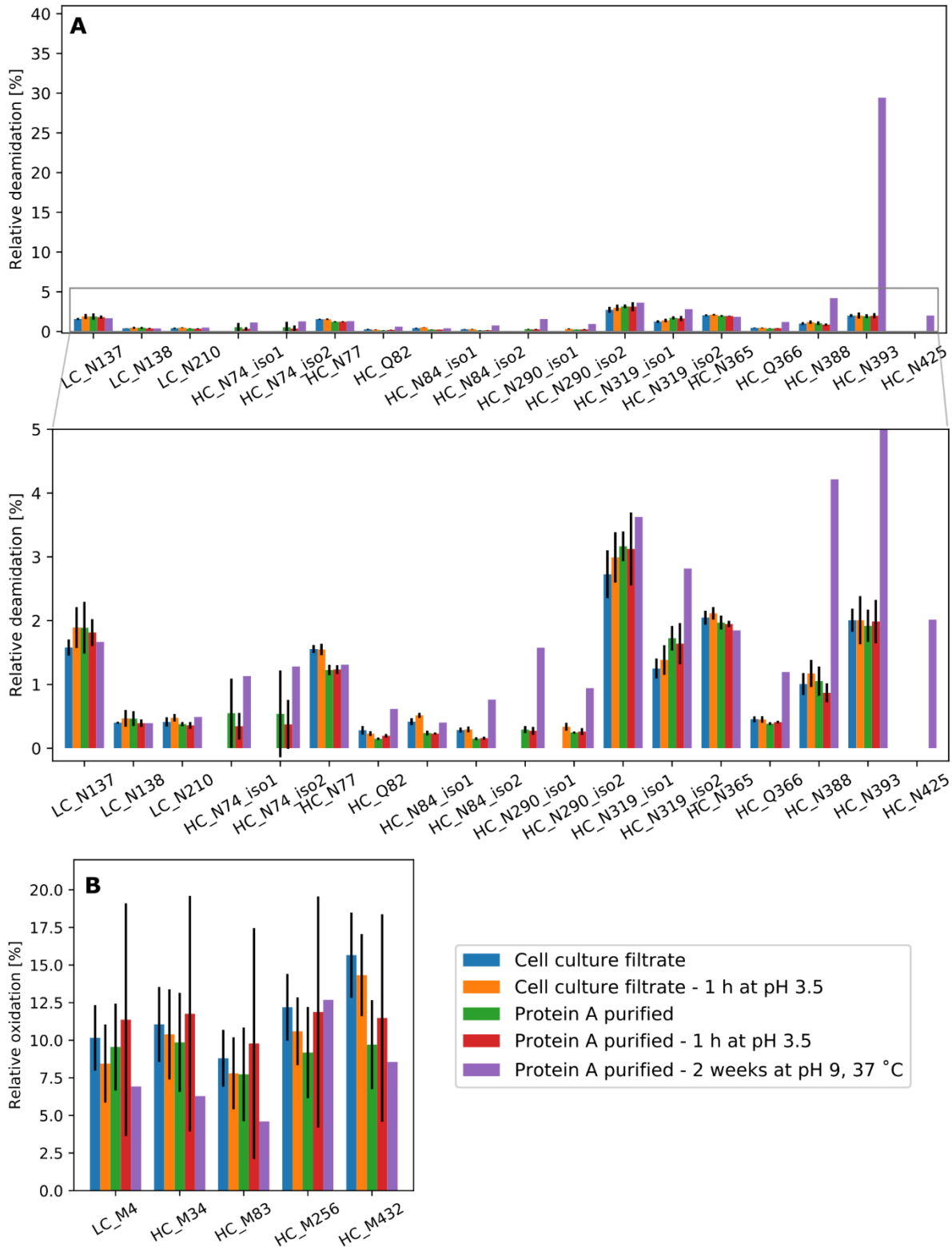
	DIA			DDA		
Deamidation site	**Mean (%)**	**Std. (%)**	**CV**	**Mean (%)**	**Std. (%)**	**CV**
LC_N92_iso1	-	-	-	0.383	0.037	0.084
LC_N92_iso2	-	-	-	0.436	0.048	0.096
LC_N137	1.944	0.134	0.060	1.174	0.070	0.052
LC_N138	0.466	0.104	0.194	0.319	0.029	0.078
LC_N210	0.547	0.064	0.101	0.460	0.043	0.081
HC_N74_iso1	22.206	1.206	0.047	22.355	1.352	0.052
HC_N74_iso2	36.591	3.117	0.074	33.749	3.538	0.091
HC_N77	1.456	0.026	0.015	2.709	0.023	0.007
HC_Q82	0.456	0.044	0.083	0.939	0.065	0.060
HC_N84_iso1	0.606	0.027	0.039	1.192	0.066	0.048
HC_N84_iso2	0.895	0.034	0.033	1.690	0.096	0.049
HC_N290_iso1	1.740	0.063	0.032	1.943	0.036	0.016
HC_N290_iso2	0.969	0.054	0.048	1.186	0.039	0.028
HC_N319_iso1	3.750	0.502	0.116	3.739	0.392	0.091
HC_N319_iso2	2.964	0.401	0.117	3.032	0.250	0.071
HC_N365	2.059	0.096	0.040	1.976	0.080	0.035
HC_Q366	1.281	0.025	0.017	1.343	0.136	0.088
HC_N388	3.768	0.561	0.129	-	-	-
HC_N393	32.370	0.864	0.023	36.106	0.946	0.023
HC_N425	3.123	0.732	0.203	2.373	0.262	0.096



Appendix figure 2: The Protein A purified adalimumab stressed at pH 9 and 37 ˚C for 2 weeks was digested in 4 replicates and analyzed with the MAM assay and acquired by DDA (orange) or DIA (blue). (a) shows the mean values for quantification of the deamidation degree at the sites, with error bars showing standard deviation. (b) shows a zoom in of (a) with mean values 0%-5%, (c) shows the CVs between the deamidation quantification values. Quantities were calculated from fragment XIC areas from the DIA data and from precursor XICs from the DDA data. See Table 3 for the amino acid sequences of the abbreviated peptides.



Appendix figure 3: Linearity of quantification of methionine oxidation using the MAM assay. Two digests of Protein A purified adalimumab biosimilar, one with no induced oxidation and one with 100% induced oxidation, were mixed in different ratios. The linearity of oxidation quantification at sites HC_M34 and HC_M432 was not so good. Data was acquired using DDA. See Table 2 for the amino acid sequences of the abbreviated peptides.



Appendix figure 4: Quantification of (a) deamidation and (b) oxidation using the DIA based MAM in five different adalimumab biosimilar samples. The samples were: 1) untreated cell culture filtrate, 2) cell culture filtrate which had been left at pH 3.5 for 1 h followed by neutralization to pH 7, 3) the IgG purified on Protein A, (4) the IgG purified on Protein A which had been left at pH 3.5 for 1 h followed by neutralization to pH 7 and 5) the IgG purified on Protein A which had been left at pH 9 and 37 ˚C for 2 weeks. The graph shows mean values from replicate digests (sample 1-4: N = 5, sample 5: N = 1) with error bars showing standard deviation. 40 µg IgG was prepared of all samples. See Table 2 and Table 3 for the amino acid sequences of the abbreviated peptides.

Appendix table 4: Quantification of the degree of +70 Da modification (%) at the protein N-terminal and at the lysine residues in two samples of rhGM-CSF by peptide mapping using 1) LysC + GluC, 2) Pepsin or 3) Thermolysin for digestion. Data was acquired by RP LC-MS by DDA.

	Enzyme	N-term	K63	K72	K74	K85	K107	K111
Early process sample	LysC + GluC	9.05	77.4	2.65	13.79	0.56	5.78	4.97
	Pepsin	14.42	1.90*	1.90*	0	1.01	1.12*	1.12*
	Thermolysin	10.5	-	3.35*	3.35*	-	1.69	1.84
Late process sample	LysC + GluC	0.03	1.83	0.05	0.55	0	0.18	0
	Pepsin	0	0.03	0.03	0	0	0	0
	Thermolysin	0.05	-	0.09*	0.09*	-	0.05	0.05

*The site was quantified by the same peak as another site

Appendix table 5: Enzyme combinations and digestion conditions evaluated for GM-CSF MAM assay.

	Enzyme:Protein ratio	Digestion time	Digestion conditions	Temperature
LysC (FUJIFILM Wako Pure Chemical Corporation) + Trypsin (Promega)	1:50	2h & ON	1M urea in 100mM NaP, pH 7.2	30 ˚C
Trypsin (Promega)	1:50	2h	1.5M urea in 50mM NH4CO3 buffer	30 ˚C
Chymotrypsin (Sigma Aldrich)	1:50	2h & ON	1M urea in 100mM NaP, pH 7.2	30 ˚C
GluC (Roche)	1:25	2h & ON	0.5M urea in 100mM NaP, pH 7.2	30 ˚C
Pepsin (Promega)	1:50	2h	1.4M urea in 1%TFA, pH 1	30 ˚C
Thermolysin (Sigma Aldrich)	1:25	2h	1.5M urea in 50mM NaAc, pH 5	30 ˚C
GluC (Promega)	1:25	2h & ON	0.8M urea in 50mM NaP, pH 6 and pH 7	30 ˚C

ProAla (Promega)	1:50	2h	0.8M urea in 33mM HCl, pH 1.5	30 ˚C
Protease from Rhizopus species (Sigma Aldrich)	1:5	2h	0.8M urea in 1% acetic acid, pH 3	30 ˚C
Pepsin (Fluka)	1:50	2h	0.8M urea in 1% acetic acid, pH 3	30 ˚C
GluC (Promega)+LysC (FUJIFILM Wako Pure Chemical Corporation)	1:25 + 1:5, 1:10, 1:25	ON	0.8M urea in 100mM NaP, pH 6 and pH 7	30 ˚C
GluC (Promega)+Trypsin (Promega)	1:25 + 1:25 & 1:50	ON	0.8M urea in 100mM NaP, pH 6 & 8	30 ˚C
GluC (Promega)+Pepsin (Fluka)	1:25 + 1:50	2h + ON	0.8M urea in 100mM NaP, pH 7 + 1% acetic acid, pH 3	30 ˚C
GluC (Promega) + Chymotrypsin (Roche)	1:25 + 1:100	2h, 4h, 6h & ON	0.8M urea in 100mM NaP, pH 7	30 ˚C
Proteinase K (Sigma Aldrich)	1:50	2h & ON	0.8M urea in 100mM Tris, pH 7.5, 1mM CaCl2	30 ˚C

10. Risk and safety statements

The following potentially hazardous substances, according to the Globally harmonized system of classification and labelling of chemicals, were used for this work (102–104).

Compound	GHS symbol	GHS hazard	Hazard statement codes	Precautionary statement codes
1,4-Dithiothreitol (DTT)		GHS05, GHS07	H302, H315, H318	P264, P270, P280, P301+P312, P302+P352, P305+P351+P338
2,4-Dinitrophenylhydrazine (DNPH)		GHS02, GHS07	H206 - H302	P210 - P212 - P230 - P233 - P280 - P370 + P380 + P375 - P501
3-buten-2-one (methyl vinyl ketone (MVK))		GHS02, GHS05, GHS06, GHS09	H225 - H300 + H310 + H330 - H314 - H317 - H410	P210 - P280 - P301 + P310 + P330 - P303 + P361 + P353 - P304 + P340 + P310 - P305 + P351 + P338 + P310
4-Vinylpyridine		GHS02, GHS05, GHS06, GHS09	H226 - H301 + H311 + H331 - H314 - H317 - H411	P210 - P273 - P280 - P303 + P361 + P353 - P304 + P340 + P310 - P305 + P351 + P338
Acetic acid		GHS02, GHS05	H226 - H314	P210 - P280 - P303 + P361 + P353 - P305 + P351 + P338 + P310
Acetone		GHS02, GHS07	H225 - H319 - H336	P210 - P233 - P240 - P241 - P242 - P305 + P351 + P338

Acetonitrile (ACN)		GHS02, GHS07	H225 - H302 + H312 + H332 - H319	P210 - P280 - P301 + P312 - P303 + P361 + P353 - P304 + P340 + P312 - P305 + P351 + P338
Ammonium bicarbonate		GHS07	H302	P301 + P312 + P330
Ammonium formate		GHS07	H319	P264 - P280 - P305 + P351 + P338 - P337 + P313
Ammonium hydroxide (NH_4OH)		GHS05, GHS07, GHS09	H302 - H314 - H335 - H400	P261 - P273 - P280 - P301 + P312 - P303 + P361 + P353 - P305 + P351 + P338
Borane pyridine complex		GHS02, GHS06	H225 - H301 - H310 + H330 - H315 - H319	P210 - P260 - P280 - P284 - P301 + P310 - P302 + P350
Crotonaldehyde		GHS02, GHS05, GHS06, GHS08, GHS09	H225 - H301 + H311 - H315 - H318 - H330 - H335 - H341 - H373 - H400	P201 - P210 - P280 - P301 + P310 + P330 - P304 + P340 + P310 - P305 + P351 + P338 + P310
Ethanol (EtOH)		GHS02, GHS07	H225 - H319	P210 - P233 - P240 - P241 - P242 - P305 + P351 + P338

Formic acid (FA)		GHS02, GHS05, GHS06	H226 - H302 - H314 - H331	P210 - P280 - P301 + P312 - P303 + P361 + P353 - P304 + P340 + P310 - P305 + P351 + P338
Glycoworks RapiFluor-MS Labeling Module		GHS02, GHS07, GHS08	H302, H312, H319, H332, H360D, H226	P321, P201, P280, P308 + P313, P202, P370 + P378, P210
GlycoWorks Sample Diluent-DMF/ACN		GHS02, GHS07, GHS08	H302, H312, H319, H332, H360D, H225	P280, P321, P201, P202, P280, P308+P313
Guanidinium hydrocholoride (GdnHCl)		GHS07	H302 + H332 - H315 - H319	P261 - P264 - P301 + P312 - P302 + P352 - P304 + P340 + P312 - P305 + P351 + P338
Hydrogen peroxide (H_2O_2)		GHS05, GHS07	H302 - H318 - H412	P273 - P280 - P301 + P312 + P330 - P305 + P351 + P338 + P310
Iodoacetamide (IAM)		GHS06	H301 - H315 - H317 - H319 - H335	P261 - P264 - P280 - P301 + P310 - P302 + P352 - P305 + P351 + P338

Methanol (MeOH)		GHS02, GHS06, GHS08	H225 - H301 + H311 + H331 - H370	P210 - P233 - P280 - P301 + P310 - P303 + P361 + P353 - P304 + P340 + P311
N-ethylmaleimide (NEM)		GHS05, GHS06	H300 - H311 - H314 - H317	P260 - P270 - P280 - P303 + P361 + P353 - P304 + P340 + P310 - P305 + P351 + P338
Phosphoric acid		GHS05, GHS07	H290 - H302 - H314	P280 - P301 + P312 + P330 - P301 + P330 + P331 - P303 + P361 + P353 - P305 + P351 + P338 + P310
Trifluoroacetic acid (TFA)		GHS05, GHS07	H314 - H332 - H412	P261 - P273 - P280 - P303 + P361 + P353 - P304 + P340 + P310 - P305 + P351 + P338
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)		GHS05	H314	P260 - P280 - P301 + P330 + P331 - P303 + P361 + P353 - P304 + P340 + P310 - P305 + P351 + P338

Hazard statement code	**Hazard statement**
H206	Fire, blast or projection hazard; increased risk of explosion if desensitizing agent is reduced
H225	Highly Flammable liquid and vapor
H226	Flammable liquid and vapor
H290	May be corrosive to metals
H300	Fatal if swallowed
H301	Toxic if swallowed
H302	Harmful if swallowed
H311	Toxic in contact with skin
H312	Harmful in contact with skin
H314	Causes severe skin burns and eye damage
H315	Causes skin irritation

H317	May cause an allergic skin reaction
H318	Causes serious eye damage
H319	Causes serious eye irritation
H330	Fatal if inhaled
H331	Toxic if inhaled
H332	Harmful if inhaled
H335	May cause respiratory irritation
H336	May cause drowsiness or dizziness
H341	Suspected of causing genetic defects
H360D	May damage the unborn child
H370	Causes damage to organs
H373	Causes damage to organs through prolonged or repeated exposure
H400	Very toxic to aquatic life
H410	Very toxic to aquatic life with long lasting effects
H411	Toxic to aquatic life with long lasting effects
H412	Harmful to aquatic life with long lasting effects
H310+H330	Fatal in contact with skin or if inhaled
H300+H310+H330	Fatal if swallowed, in contact with skin or if inhaled
H301+H311	Toxic if swallowed or in contact with skin
H301+H311+H331	Toxic if swallowed, in contact with skin or if inhaled
H302+H332	Harmful if swallowed or if inhaled
H302+H312+H332	Harmful if swallowed, in contact with skin or if inhaled

Precautionary statement code	**Precautionary statement**
P201	Obtain special instructions before use.
P202	Do not handle until all safety precautions have been read and understood.
P210	Keep away from heat, hot surface, sparks, open flames and other ignition sources. - No smoking.
P212	Avoid heating under confinement or reduction of the desensitized agent.
P230	Keep wetted with ...
P233	Keep container tightly closed.
P240	Ground/bond container and receiving equipment.
P241	Use explosion-proof [electrical/ventilating/lighting/.../] equipment.
P242	Use only non-sparking tools.
P260	Do not breathe dust/fume/gas/mist/vapors/spray.
P261	Avoid breathing dust/fume/gas/mist/vapors/spray.
P264	Wash ... thoroughly after handling.
P270	Do not eat, drink or smoke when using this product.
P273	Avoid release to the environment.
P280	Wear protective gloves/protective clothing/eye protection/face protection.
P284	[In case of inadequate ventilation] Wear respiratory protection.
P308+P313	IF exposed or concerned: Get medical advice/attention
P321	Specific treatment (see ... on this label).
P370+P378	In case of fire: Use dry sand, dry chemical or alcohol-resistant foam for extinction

P301+P310	IF SWALLOWED: Immediately call a POISON CENTER or doctor/ physician.
P301+P310+P330	IF SWALLOWED: Immediately call a POISON CENTER or doctor/ physician. Rinse mouth.
P301+P312	IF SWALLOWED: call a POISON CENTER/doctor/... IF you feel unwell.
P301+P312+P330	IF SWALLOWED: call a POISON CENTER/doctor/... IF you feel unwell. Rinse mouth.
P301+P330+P331	IF SWALLOWED: Rinse mouth. Do NOT induce vomiting.
P302+P350	IF ON SKIN: Gently wash with plenty of soap and water.
P302+P352	IF ON SKIN: wash with plenty of water.
P303+P361+P353	IF ON SKIN (or hair): Take off Immediately all contaminated clothing. Rinse SKIN with water [or shower].
P304+P340+P310	IF INHALED: Remove victim to fresh air and keep at rest in a position comfortable for breathing. Immediately call a POISON CENTER or doctor/physician.
P304+P340+P311	IF INHALED: Remove victim to fresh air and keep at rest in a position comfortable for breathing. Call a POISON CENTER or doctor/…
P304+P340+P312	IF INHALED: Remove victim to fresh air and keep at rest in a position comfortable for breathing. Call a POISON CENTER or doctor/... if you feel unwell.
P305+P351+P338	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses if present and easy to do - continue rinsing.
P305+P351+P338+P310	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses if present and easy to do - continue rinsing. Immediately call a POISON CENTER or doctor/physician.
P308+P313	IF exposed or concerned: Get medical advice/attention.
P337+P313	IF eye irritation persists: Get medical advice/attention.
P370+P380+P375	In case of fire: Evacuate area. Fight fire remotely due to the risk of explosion.
P501	Dispose of contents/container to an approved waste disposal plant.

11. Acknowledgements

I would like to thank Alphalyse so much for letting me conduct my PhD work at their facility and for being so optimistic and supportive all the way through this thesis. I would especially like to thank my supervisors at Alphalyse: Thomas Kofoed, Jakob Bunkenborg and Stine Thyssen for always taking the time to listen, explain and discuss my projects. I would also like to thank all the employees at Alphalyse for all the good times, both outside and inside the lab, and for all the interesting discussions.

I would like to thank my supervisor at University of Hamburg, Prof. Hartmut Schlüter, for taking the time to read through my work and for all the valuable feedback.

I would also like to thank Gregory Silva Dutra at the University of Natural Resources and Life Sciences (BOKU) for providing me with the antibody samples used in this thesis.

Further, I would like to thank the Horizon 2020 Marie Sklodowska-Curie Action ITN 2017 of the European Commission (H2020-MSCA-ITN-2017) for providing the funding for this PhD thesis. I would also like to thank everyone involved in the Analytics for Biologics training school for all the wonderful knowledge and memories that we have built together.

I would like to aim a sincere thank you to my friends Jakob Willforss and Maria Ekemohn for taking the time to read through my thesis, giving valuable feedback and for taking the time to discuss the data interpretation.

I would also like to thank my wonderful family for all the patience and mental support. Finally, a huge thank you goes to my boyfriend Fredrik Svensson for all the help with all sorts of things, such as programming issues, proofreading and mental support when times have been heavy.

12. Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.

_____

Lund, den 24.06.2021