

Improvement of liquid chromatography for analysis and purification  
of proteoforms via rational protein purification parameter screening  
and sample displacement chromatography

**Dissertation**

for the acquisition of the academic degree

**Doctor rerum naturalium (Dr. rer. nat.)**

at the

Faculty of Mathematics, Informatics and Natural Science

Department of Chemistry

University of Hamburg

Submitted by

**Siti Nurul Hidayah, M. Sc.**

From Magelang, Indonesia

Hamburg

2021

**Supervisor**

Prof. Dr. Hartmut Schlüter

**Recommendation of the evaluators for the dissertation:**

Prof. Dr. Hartmut Schlüter

Prof. Dr. rer. nat habil. Markus Perbandt

**Date of disputation:**

03 September 2021

The research for this thesis was carried out in the Section Mass Spectrometry Proteomics, headed by Prof. Dr. Hartmut Schlüter, Diagnostic Center at the University Medical Centre Hamburg-Eppendorf (UKE) from 1 April 2018 to May 2021.

## List of Publications

1. Heikaus L, **Hidayah S**, Gaikwad M, Kotasinska M, Richter V, Kwiatkowski M, et al. Sample displacement batch chromatography of proteins. *Methods Mol Biol.* 2021;2178:285–99.
2. Jeong K, Kim J, Gaikwad M, **Hidayah SN**, Heikaus L, Schlüter H, et al. FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-Down Proteomics. *Cell Syst.* 2020;10(2):213-218.e6.
3. **Hidayah SN**, Gaikwad M, Heikaus L, Schlüter H., *Proteoforms 2019*, IntechOpen, DOI: 10.5772/intechopen.89644.
4. Fazel R, Guan Y, Vaziri B, Krisp C, Heikaus L, Saadati A, **Hidayah SN**, Gaikwad M, Schlüter H. Structural and in vitro functional comparability analysis of altebrel™, a proposed etanercept biosimilar: Focus on primary sequence and glycosylation. *Pharmaceuticals.* 2019;12(1):1–16.

## List of Posters Publications

1. International mass spectrometry conference (IMSC) 2018: Cation Exchange Chromatography in Displacement Mode for Online Analysis of Intact Proteins by Mass Spectrometry.
2. German Society for Mass Spectrometry (DGMS) 2020: Application of Sample Displacement Batch Chromatography and Intact Protein Mass Spectrometry Analysis for Enrichment and Identification of Proteoforms.

## Table of Contents

List of Abbreviations .....	7
Abstract .....	9
Zusammenfassung.....	10
Introduction.....	12
Proteoforms.....	12
Proteoforms in therapeutic proteins.....	13
Proteform purification .....	15
Mass spectrometry for proteoforms .....	18
Aim.....	20
Results.....	21
<b>Method Optimisation for Protein Purification Parameter Screening (PPS) and Sample Displacement Batch Chromatography (SDBC).....</b>	21
<i>The choice of device for sample displacement chromatography.....</i>	21
<i>Sample selection and calculation of binding capacity .....</i>	22
<i>Optimisation of incubation time &amp; incubation method .....</i>	25
<i>Optimisation of sample preparation for top-down mass spectrometry .....</i>	26
<b>Purification Parameter Screening (PPS) for Ovalbumin Proteoforms.....</b>	30
<i>Screening of pH and resin for ovalbumin proteoforms purification.....</i>	31
<i>The effect of additional NaCl in sample application buffer .....</i>	34
<i>Analysis of ovalbumin proteoforms in PPS optimal condition and original sample.....</i>	35
<i>Identification of ovalbumin proteoforms with mass-matching .....</i>	37
<b>Sample Displacement Batch Chromatography (SDBC) for Ovalbumin Proteoforms .....</b>	40
<b>Application of PPS and SDBC for Monoclonal Antibody (mAb).....</b>	49
Discussion .....	56
Materials and Reagents .....	64
Methods.....	66
Cation exchange chromatography of ovalbumin .....	66
Chromatography resin preparation.....	66
Protein purification parameter screening (PPS) .....	66
Sample displacement batch chromatography (SDBC) .....	68
Sample preparation for mass spectrometry .....	69
Flow injection analysis – mass spectrometry (FIA-MS).....	69
Data analysis.....	70
N-glycan release, purification, and permethylation .....	70

<b>Nano LC-MS/MS for permethylated N-glycans .....</b>	<b>71</b>
<b>Identification for differential N-glycomics .....</b>	<b>71</b>
<b>References .....</b>	<b>73</b>
<b>Supplements .....</b>	<b>80</b>
<b>Risk and Safety .....</b>	<b>89</b>
<b>Acknowledgements .....</b>	<b>91</b>
<b>Declaration .....</b>	<b>92</b>

## List of Abbreviations

Abbreviation	Meaning
A	Acetylation
ABC	Ammonium bicarbonate
ACN	Acetonitrile
AEX	Anion exchange chromatography
AGC	Automatic gain control
BCA	Bicinchoninic acid
CEX	Cation exchange chromatography
CID	Collision induced dissociation
DC	Displacement chromatography
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ESI	Electrospray ionisation
FA	Formic acid
FIA	Flow injection analysis
Fuc	Fucose
G	N-glycosylation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hex	Hexose
HexNac	N-acetyl hexosamine
HMW	High molecular weight
HPLC	High performance liquid chromatography
IAA	Iodoacetamide
IEX	Ion exchange chromatography
LC	Liquid chromatography
LMW	Low molecular weight
m/z	Mass to charge ratio
mAbs	Monoclonal antibodies
MALDI	Matrix-assisted laser desorption/ionisation
MES	2-(N-morpholino) ethanesulfonic acid
min	Minute
MS	Mass spectrometry
NaCl	Sodium chloride
Neu5Gc	N-glycolylneuraminic acid
P	Phosphorylation
pI	Isoelectric point
PPS	Protein purification parameter screening

PTMs	Post translational modifications
qTOF	Quadrupole time of flight
rhEPO	Recombinant erythropoietin
RNA	Ribonucleic acid
RPLC	Reversed-phase chromatography
RVs	Reaction vials
S	Succinimide
S/N	Signal to noise ratio
SAX	Strong anion exchange
SCX	Strong cation exchange
SDBC	Sample displacement batch chromatography
SDC	Sample displacement chromatography
SEC	Size exclusion chromatography
TNF	Tumour necrosis factor
TOF	Time of flight
UPLC	Ultra-high-performance liquid chromatography

---



## Abstract

In contrast to small molecule drugs, therapeutic proteins are characterised by their composition of many very similar but not identical species (proteoforms) regarding their composition of atoms. This heterogeneity may arise due to alternatively spliced RNA transcripts, genetic variations, or post translational modifications (PTMs), thus there are many proteoforms encoded by a single gene. The presence of some proteoforms may reduce efficacy of therapeutic proteins and may even induce harmful side effects. Therefore, the analysis and purification of proteoforms are essential. However, the fractionation and purification of proteoforms is highly challenging because many proteoforms are physically and chemically very similar and often present in a very low abundance. Sample displacement chromatography (SDC) in previous studies was described as more efficient compared to other chromatography modes. This thesis aims to answer the question whether SDC is suitable for separating proteoforms of the same protein and to develop a strategy for the purification of proteoforms using sample displacement batch chromatography (SDBC). Ovalbumin from chicken egg was used as a model sample.

Protein purification parameter screening (PPS) was conducted to obtain the most optimal parameters for SDBC, yielding a two-fold increase in the number of detected proteoforms in comparison to the original sample. A significant enrichment of low-abundant basic ovalbumin proteoforms was obtained which gives a proof that SDBC is effective even for the separation of closely related proteoforms. SDBC was also able to separate ovalbumin proteoforms into two categories: highly enriched proteoforms in the SDBC early fractions and highly enriched proteoforms in the later fractions. Low-abundant basic proteoforms were well enriched in the early fractions, while in the later fractions, less basic proteoforms were present. In addition, with SDBC a complete removal of non-ovalbumin proteins such as ovomucoid, ovotransferin, and other ovalbumin related proteins, being impurities in the original sample, was achieved. The application of SDBC for separation of proteoforms of the recombinant monoclonal antibody adalimumab showed comparably well results. In conclusion, the results obtained in this study show the utility and possibility of SDBC for simple and effective separation of proteoforms, which in the future can be used for removal of harmful proteoforms and tool for enrichment of low abundant proteoforms for further analysis.

## Zusammenfassung

Ein Medikament auf Basis eines kleinen organischen Moleküls ist charakterisiert durch eine genau definierte Zusammensetzung seiner Atome. Dagegen beinhaltet ein Medikament, das zu den therapeutischen Proteinen zählt, nicht ein einziges Molekül mit genau definierter Zusammensetzung seiner Atome sondern eine Vielzahl von Spezies, auch Proteoformen genannt, die zum Teil sehr ähnliche aber nicht identische Zusammensetzung ihrer Atome haben, aber auch große Unterschiede aufweisen können. Alle Proteoformen eines Proteins werden von ein und demselben Gen kodiert. Die Heterogenität therapeutischer Proteine kann durch alternativ gespleißte RNA-Transkripte, genetische Variationen oder posttranslationale Modifikationen (PTMs) entstehen. Einige Proteoformen therapeutischer Proteine haben in Abhängigkeit ihrer Zusammensetzung eine verringerte Wirksamkeit, andere können sogar unerwünschte Nebenwirkungen hervorrufen. Daher ist die Analyse der genauen Identität und Quantität der Proteoformen in therapeutischen Protein-Präparaten sowie die Entfernung von Proteoformen mit geringer Wirksamkeit oder Nebenwirkungen wichtig. Die Reinigung von Proteoformen sind jedoch sehr anspruchsvoll, da ein Teil der Proteoformen eines Proteins in ihren physikalischen und chemischen Eigenschaften sehr ähnlich sind. Proben-Verdrängungs-Chromatographie (sample displacement chromatography, SDC) ermöglicht eine unkomplizierte Anreicherung und Reinigung von Molekülen und führt im Vergleich zu anderen Chromatographie-Modi zu höheren Ausbeuten und besseren Trennergebnissen. Ziel dieser Arbeit war es 1. zu prüfen, ob die SDC für die Reinigung von Proteoformen eines Proteins geeignet ist und 2. eine Strategie für die Anwendung der SDC zur Proteoform-Reinigung zu entwickeln.

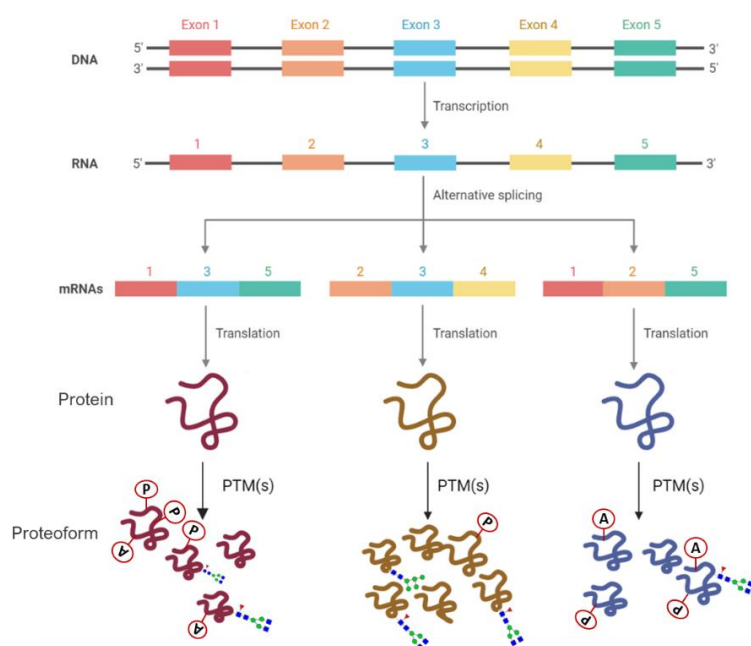
Zur Bestimmung von Parametern zur Erzielung optimaler Trennergebnisse mit der SDC wurde ein Proteinreinigungparameter-Screening (PPS) durchgeführt. Als Modellprotein wurde Ovalbumin aus Hühnerei genutzt. Mit dem per PPS-ermittelten Parametersatz wurde eine SDC zur Trennung von Proteoformen von Ovalbumin im Batch-Modus (SDBC) durchgeführt. In den resultierenden SDBC Fraktionen konnte eine im Vergleich zur Originalprobe doppelte Zahl individueller Proteoformen nachgewiesen werden, was ein deutlicher Hinweis darauf ist, dass die Heterogenität durch die SDBC in den resultierenden Fraktionen deutlich verringert wurde. Eine signifikante Anreicherung von basischen Ovalbumin-Proteoformen in den frühen

Fraktionen mit geringer Abundanz wurde erreicht, was zeigt, dass die SDBC für die Trennung von Proteoformen geeignet ist. In den Fraktionen der SDBC konnten ausschließlich die Ovalbumin-Proteoformen nachgewiesen werden. Die Begleit-Proteine in der originalen kommerziellen Ovalbumin Fraktion, wie Ovomuroid und Ovatransferin, wurden im Durchfluß der SDBC detektiert. Diese Beobachtung unterstreicht die Effektivität der SDBC. Mit der SDBC gelang auch die Trennung von Proteoformen des rekombinanten monoklonalen Antikörpers Adalimumab. Zusammenfassend konnte in der vorliegenden Arbeit der Nutzen der SDBC für die Trennung von Proteoformen demonstriert und eine Strategie zum Einsatz der SDBC für die Fraktionierung von Proteoformen eines Proteins entwickelt werden. Die SDBC ist geeignet die Komplexität der Zusammensetzung von Proteoformen in Proteinproben deutlich zu reduzieren sowie niedrig abundante Proteoformen anzureichern. Damit ist in Zukunft die Charakterisierung von Proteoformen vereinfacht.

## Introduction

### Proteoforms

Proteoforms, formerly protein species (1,2) are molecules arising from the same gene. The difference in the composition of atoms of proteoforms can be caused by alternative spliced ribonucleic acid (RNA) transcript, post-translational modifications (PTMs), and other less-explored molecular events (3). **Figure 1** shows the generation of proteoforms, which starts from transcription process of deoxyribonucleic acid (DNA) to precursor messenger RNA (mRNA) into mature mRNA, followed by the translation of mRNA into protein. During the splicing, different alternatives of exons arrangement can happen, resulting in different mature mRNA which then translated into different proteins. After translation, protein can be further modified by PTMs (4). In human proteome study, there are 19,587 to 20,245 protein-coding genes. If one gene codes one protein around 2,000 proteoforms are estimated. This number may rise to around 70,000 if many genes are transcribed with splice variants and go up to hundreds of thousands if some PTMs are added (5).



**Figure 1.** Schematic overview of the generation of many different proteoforms from a single gene starting from transcription of DNA to the addition of various possible post-translational modifications (PTMs). Figure was created using Biorender.com

PTMs can occur at the N- and C- terminal of proteins or on the amino acid side chains by modifying the functional groups or adding a new one such as carbohydrate molecules or phosphate. There are more than 200 chemical moieties that can be enzymatically conjugated to amino acids in a protein (6). Nucleophilic functional group in amino acid side chains like thiol in cysteine, amine in lysine, arginine and histidine, and hydroxyl group of threonine and serin are most susceptible to modification (6–8). PTMs vary in size and complexity, ranging from simple methionine oxidation or cysteine oxidation, to complex form of disulfide linkage, to the attachment of heavy oligosaccharide units (9). Glycosylation is the most abundance, heterogenous, and complex PTMs as its biosynthesis is template free-driven process. It results in various forms of branching, glycan composition, and other types of isomerism (10). In addition to that, there are possibilities of post-glycosylation modifications on the glycans, such as, phosphorylation, methylation, acetylation, and sulfation which further increase the diversity of glycosylation (11).

### **Proteoforms in therapeutic proteins**

Therapeutic proteins (TPs) are protein-based drugs that are genetically engineered in the laboratory. They are characterised by their composition of highly heterogenous proteoforms. TPs possess several advantages over small molecule drugs due to their higher binding selectivity and specificity. This makes therapeutic proteins able to target specific steps in disease pathology (12). One of the first therapeutic proteins is recombinant human insulin which is used to improved glycemic control for patients with diabetic (13). Since then, different types of therapeutic proteins have been approved for several different treatments such as recombinant erythropoietin (rhEPO), a growth factor used in the treatment of anemia (14), Filgrastim, a recombinant human methionyl granulocyte colony stimulating factor (G-CSF) used to treat low neutrophil count, recombinant human interferon- $\beta$ -1a (rhIFN- $\beta$ ) which is used for treatment of multiple sclerosis (15), and monoclonal antibodies (mAbs) which target a specific antigen in the body and used for the treatment of cancer, Crohns's disease, rheumatoid arthritis, transplant rejection, multiple sclerosis, and systemic lupus erythematosus. Currently, more than 40% of approved therapeutic proteins are monoclonal

antibodies (mAbs), including antibody fragment antigen binding and antibody drug conjugate (16).

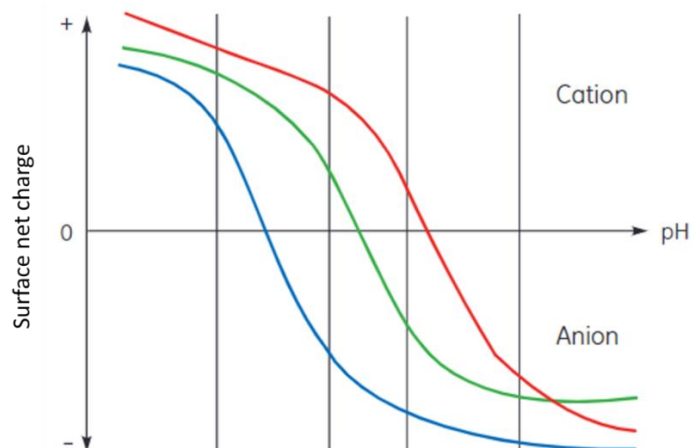
In TPs, the number of proteoforms are not as overwhelming as in human proteins. There are many PTMs that are not associated with currently marketed therapeutic proteins. Nevertheless, the heterogeneity of recombinant therapeutic proteins shall not be underestimated. In monoclonal antibodies (mAbs) for example, several modifications are possible including N-terminal modification, deamidation, isomerisation, oxidation, glycosylation, glycation, C-terminal modification, and cysteine related modification (17). Those modifications result in the formation of various charge variants and heterogeneities. During the production of TPs, the choice of organisms, cells, and conditions may affect the characteristic of the final products. Glycan analysis of some products produced in HEK293 cells and CHO cells showed differences in molecular weight, isoelectric point (pI), and glycoprotein structure. CHO-derived proteins showed to have less complex glycoprotein profile but more sialylated compared to HEK293-derived proteins (18). Furthermore, as the synthesised of the proteins finished, harvesting and several purification processes follow might potentially induce modification in the protein such as oxidation of the methionine residues or deamidation of the asparagine residues. As product themselves, they may undergo product degradation which leads to more differ variants, such as deamidation which can occur during biopharmaceutical manufacture and storage (19).

The presence of some proteoforms might affect the efficacy of TPs and some can induce negative side-effect for the patients. Deamidation impact the functional and pharmacological changes such as changes in drug activity, changes to target antigen recognition (20), alter product's half-life (21), and potential effect on immunogenicity (22). The present of certain glycoforms like N-glycolylneuraminic acid (Neu5Gc) in therapeutic glycoproteins is critically reducing the drug efficacy (10). Removal of any n-glycosylation site on recombinant EPO significantly reduces its in vivo activity as measured by stimulation of EPO in polycythemic mice and recombinant EPO administered intravenously to rodents (23,24). Considering the complexity of proteoforms and their effect in TPs, identification of these undesired proteoforms and their removal is an important issue. However, the analysis and separation

of proteoforms are highly challenging due to their complexity and their physical and chemical similarities. Moreover, proteoforms often present in a low abundance which makes the detection, analysis, and identification of proteoforms more challenging. Analytics for biologics (A4B) consortium has dedicated a collaborative research aiming to tackle the challenge in proteoforms analysis, particularly in TPs. The first step for the identification of undesirable proteoforms is to enrich their abundance so that the low-abundant proteoforms can be easily detected and further characterised. After that, focus is needs to separate of those proteoforms from other proteoforms, and finally to remove them.

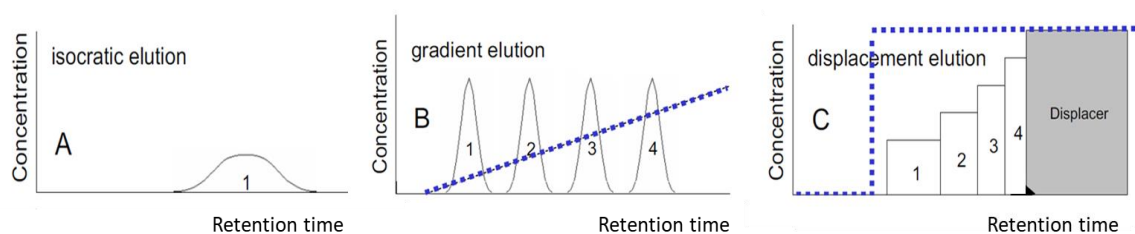
### **Proteoform purification**

Liquid Chromatography (LC) is the method of choice for separation and purification of molecules. Most of LC techniques such as *affinity chromatography* (25), *size exclusion chromatography* (SEC) (26), *reversed-phase chromatography* (RPLC) (27), and hydrophilic interaction chromatography (HILIC) (28) still focus on the separation and purification of protein from a mixture. *Ion exchange chromatography* (IEX) is an LC technique that have been used for proteoforms separation and purification as it separates molecules based on their net surface charge (29–31). The existence of PTMs in different proteoforms will contribute to the change of the isoelectric point (pI) of the protein (32). Therefore, different proteoforms will have different pI value and ideally can be separated using this IEX technique. Proteoforms behaviour in solution theoretically will be similar to the behaviour of protein in solution in which protein that have no net charge at a pH equivalent to its isoelectric point (pI) will not interact with a charged medium. However, at a pH above its pI, protein will bind to a positively charged medium or anion exchanger (AEX) and at a pH below its pI, protein will bind to a negatively charged medium or cation exchanger (CEX) (**figure 2**) (33). Several studies have aimed to separate proteoforms in monoclonal antibodies (more widely known as charge variants) using ion exchange chromatography either in salt gradient mode or pH gradient mode (34–38).



**Figure 2.** Theoretical protein titration curve, showing how surface charge varies with pH (33).

Apart from gradient elution mode, some studies have reported the use of displacement mode in IEX for the separation of proteoforms (39,40). Displacement chromatography (DC) is one of chromatography modes like frontal, isocratic, and gradient elution mode (**figure 3**) (41). In DC, the separation and elution are based on a competitive binding of the sample components themselves and an additional molecule called displacer which has the highest affinity toward the chromatography matrix. It was first developed in 1943 for the separation of amino acids by Tiselius (42). In DC, sample components are resolved into consecutive 'rectangular' zones of highly concentrated substances rather than 'narrow gaussian peaks' found in gradient elution (43). Displacement elution can be incorporated in any type of chromatography such as reversed-phase, ion exchange, hydrophilic interaction, or affinity chromatography (39,44–48).

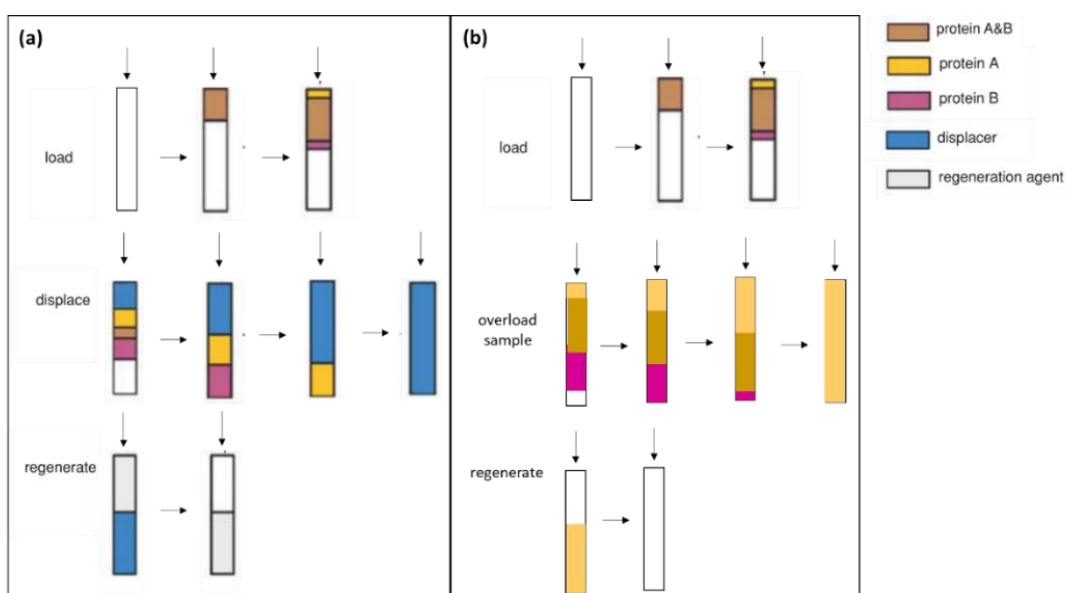


**Figure 3.** Different elution mode in chromatography: isocratic elution (a), gradient elution (b), and displacement elution (c) (49).



A study from Zhang, T., et al. (39) has successfully achieved the separation of acidic, main, and basic proteoforms of monoclonal antibody using cation exchange displacement chromatography. They have demonstrated that DC allows significant enrichment of proteoforms with high purity and recovery. It has shown the possibility of using DC for the separation and purification of proteoforms.

However, despite the success of DC, the use of displacer is raising a concern that it may co-elute with the target molecules. Sample displacement chromatography (SDC) is an alternative approach which allows DC without displacer. It takes advantages of the fact that during sample application onto a column, samples components already arrange themselves according to their affinity toward the stationary phase (50). The competition of the sample components for the binding site of stationary phase can be used for their separation (**figure 4**). This phenomenon present in any type of liquid chromatography such as reserved-phase, ion exchange, hydrophilic interaction, or affinity chromatography (50). The components with higher affinity compete more successfully for the binding sites than the lower affinity components, thus the latter ones are displaced from the column. The sample displacement can be optimally performed under overloading condition, the separation happens during the sample loading, thus the column capacity is optimally used.



**Figure 4.** Illustration of displacement chromatography (DC) principle (a) and sample displacement chromatography (SDC) (b). Modified from McAtee, C.P. (51).

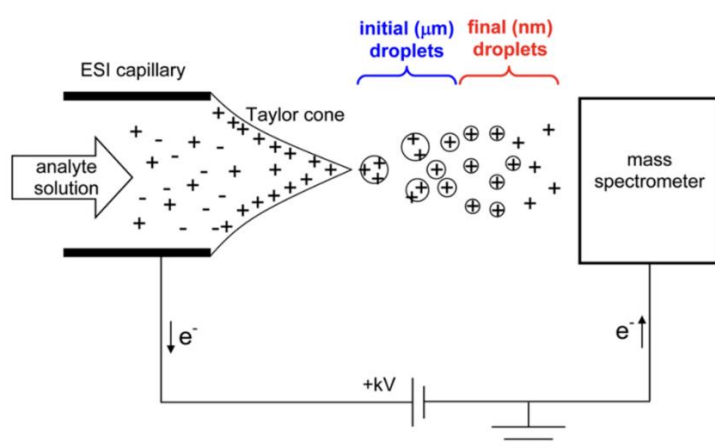
SDC was used for the first time for peptides separation (50,52) and since then has been used for separation and purification of proteins (53), separation of human plasma protein fraction Cohn IV-4 (54), and therapeutic antibody charge variants (40). The sample displacement effect can be utilized for the separation of molecules by using segmented columns connected in series. After sample loading, the column segments will be disconnected, and the components bind to the stationary phase of each individual column are eluted separately. Therefore, sample components are distributed and separated throughout the system. Since proteoforms hold subtle differences in their physical and chemical properties that will also affect their chromatography behaviours, in principle SDC should also work for enrichment and purification of proteoforms.

### **Mass spectrometry for proteoforms**

To support the identification of proteoforms, LC is coupled with detection technique such as UV and mass spectrometry (MS). MS is preferred than UV as it shows higher sensitivity (55) and is now the most widely used technique for proteins and proteoforms identification (56). There are two main approaches that are most used, bottom-up or shotgun approach and top-down approach (57). Bottom-up approach relies on peptides generated by chemical cleavage or enzymatic. The peptide mixture is subjected to LC/MS analysis and later identification is done by comparing the peptide fragments against an in-silico digested database of protein sequences (58). However, as it identifies proteins based on peptides derived from enzymatic digestion of samples, peptides from all proteoforms are mixed. It cannot be assumed that a peptide derived from a tryptic digest is unique to a single protein or a proteoform as it is highly possible that a peptide is shared by more than one protein. Top-down mass spectrometry approach on the other hand, offers ease in correlating modification to a specific source giving unambiguous information of each proteoform as it utilizes the intact proteins for analysis instead of peptides. With the use of intact proteins, all the information related to the protein and the proteoforms accompanied with it is retained. The problem of shared peptide in a heterogenous proteoform mix evident in bottom-up approach is avoided (59–61).

To analyse the intact proteins or proteoforms, sample can be introduced in the MS as direct injections/infusion or coupled to front end fractionation techniques such as liquid chromatography or capillary electrophoresis (62,63). Direct injection is preferable as it allows faster analysis, avoid possible protein precipitation due to high organic solvents such as acetonitrile, and avoid the loss of proteoforms in a chromatography column (64,65). The experiment can be carried out either in denatured or native stage of protein. On the denatured stage, protein is unfolded, exposing the sites for protonation, while in the native stage the noncovalent interaction is preserved, and protein stays folded.

To produce ions in mass spectrometry, 'soft ionisation' methods which are less destructive to sample molecules and can retain the integrity of the molecules as they induce little to no fragmentation is used. Both matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) (66–70) can be used. ESI (**figure 5**) is preferred over MALDI because ESI imparts more charges per protein (71,72). This enables the mass determination of large biomolecules using mass analysers with moderate mass-to-charge ratio upper limits (for example,  $m/z \leq 4,000$ ), which happens to offer the highest resolving power (73). ESI generates  $[M + zH]^{z+}$  with multiple charges ( $z \gg 1$ ), allowing the detection of large molecules on mass spectrometry with limited  $m/z$  range.



**Figure 5.** Schematic depiction of ESI source operated in positive mode (73).

## Aim

In therapeutic proteins, the presence of some proteoforms may reduce efficacy and may even induce harmful side effects. Thus, identification of these undesired proteoforms and their removal is an important issue. For both actions, identification and removal, liquid chromatography (LC) is an important method. LC of proteoforms is one of the most challenging problems in analytical chemistry and in down-stream-processing because proteoforms are coded by the same gene that makes them physically and chemically very similar. Moreover, undesired proteoforms are often present in a low abundance. The displacement mode in LC is advantageous with respect to its separation efficacy compared to the gradient mode. Furthermore, displacement chromatography in the batch mode (SDBC) offers a very simple enrichment and fractionation of proteins without expensive equipment, as was demonstrated by Kotasinska et al. (53).

It is hypothesized that SDBC should also work for the separation of proteoforms. Therefore, this thesis aims to answer the question whether SDBC is suitable for separating proteoforms of the same protein and to develop a strategy for the purification of proteoforms using SDBC. Protein purification parameter screening (PPS) will be used to determine the most optimal condition for the purification of proteoforms using SDBC. Top-down mass spectrometry analysis of the intact proteoforms will be used to analyse the samples from PPS and SDBC.

## Results

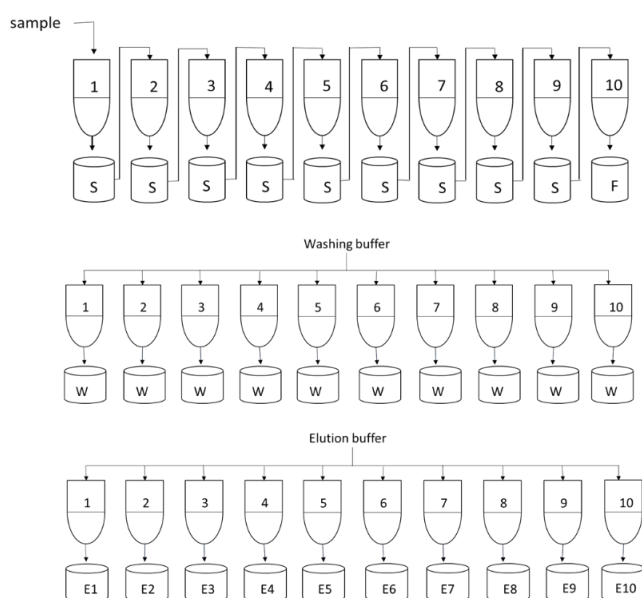
### Method Optimisation for Protein Purification Parameter Screening (PPS) and Sample Displacement Batch Chromatography (SDBC)

#### *The choice of device for sample displacement chromatography*

Considering that proteoforms are often present in low abundance and the differences between proteoforms are very subtle, it is necessary to develop methods with certain conditions in which the purification of proteoforms is maximal. To determine the most optimal conditions for enrichment and purification of proteoforms with sample displacement chromatography (SDC), several protein purification parameter screening (PPS) experiments were carried out to ensure the efficiency of the separation and purification. A cation exchange chromatography (CEX) column (Sepax Proteomix SCX-NP3, 4.6 x 50 mm, 3 $\mu$ m) was tested for sample displacement chromatography. To obtain optimal sample displacement effect, an overloading condition was necessary. Since most commercially available chromatography columns have big dimensions with high binding capacity, high amount of sample is required. Sample displacement chromatography in a batch is an alternative for column chromatography (74,75). Instead of packing chromatography material (resin) into a column, a suspension of resin was used in batch chromatography, giving a flexibility to customise the amount of resin used and therefore reducing the amount of sample needed.

Sample displacement batch chromatography (SDBC), developed by Kotasinska et al. (53,75), was tested. First, the appropriate vial format for SDBC was tested. 384-wells microtiter plate, 96-wells plate, and 1.5 mL reaction vials were tested for batch chromatography. Limitation of the total volume that could be used in 384-wells microtiter plate resulted in problems with liquid handling because of surface tension resulting in air bubbles trapped in the cavities. This problem was avoided by using 96-well plates which have larger cavities. However, the resin easily settled on the bottom of the plate since the plate can be shaken only in horizontally way, thus limiting the contact between sample and resin. Reaction vials (RVs, 1.5 mL size) were then chosen, allowing the use of buffer volume up to 1 mL, and giving possibility to be shaken in 360 degree which maximised the contact between sample and resin.

In SDBC, 10 RVs were used as 10 segments as shown in **figure 6** in which the resin was distributed in equal amount. Sample was diluted in sample application buffer, loaded in the first segment, incubated, and continuously shaken. After sedimentation of the resin, the supernatant from the first segment was transferred to the second segment, incubated, and shaken. The process was repeated until the last segment. With this system, proteoforms were distributed throughout the segments according to their affinity to the resin. Washing steps were performed to remove unbound proteoforms from the segment. Proteoforms bound to the resin in each segment were then eluted with 1 Molar sodium chloride (NaCl), buffer exchanged to water, and each fraction was analysed with mass spectrometry in the top-down modus (for details see sub chapter: optimisation of sample preparation for top-down mass spectrometry).



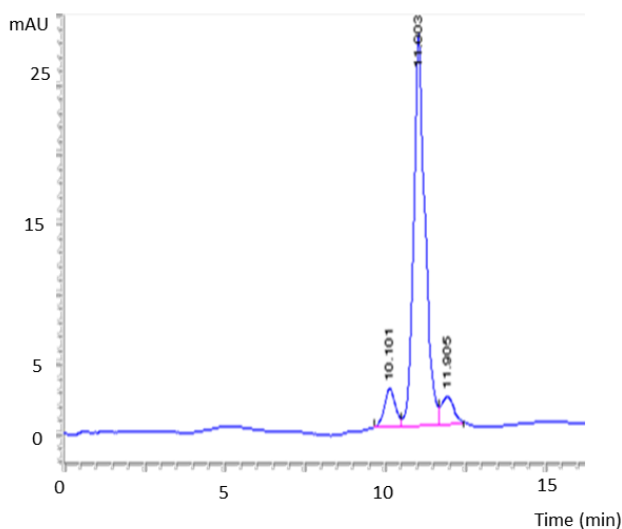
**Figure 6.** Scheme of the steps of sample displacement batch chromatography (SDBC) using 10 segments. **S**, supernatant; **F**, flow-through; **W**, wash; **E**, eluate. Modified from Heikaus, L., et al., (74).

### *Sample selection and calculation of binding capacity*

Ovalbumin from chicken egg white was chosen as a model sample for PPS and SDBC. Ovalbumin is not a therapeutic protein however its complexity represents the complexity of therapeutic proteins which make it a good model protein to study proteoforms. Ovalbumin is

the most prominent protein in egg white, expressed in the magnum of the oviduct and is responsible for egg white formation. Up to 60-65 % of the total protein mass in egg white is ovalbumin. The exact physiological function of ovalbumin is still remain unknow, it is hypothesised to be a storage protein (76). Chicken egg white ovalbumin consists of 386 amino acids with molecular weight around 44 kDa. The initial methionine is cleaved from mature protein (77). Ovalbumin possesses high diverse amount of PTMs (78,79). It is embellished with various PTMs, including N-terminal acetylation in the amino acid glycine position 2 (77), phosphorylation in the amino acid serine position 69 and 345, one disulfide bond between cysteine position 74 and cysteine position 121 (80), and extensive glycosylation in amino acid asparagine position 293.

An initial experiment using a CEX column (Sepax Proteomix SCX-NP3) with a NaCl gradient elution showed two small peaks eluting before and after the main peak which correspond to acidic and basic proteoforms of ovalbumin (**figure 7**). It was estimated that the total amount of low abundance ovalbumin proteoforms was approximately 10 % of the total amount of ovalbumin.

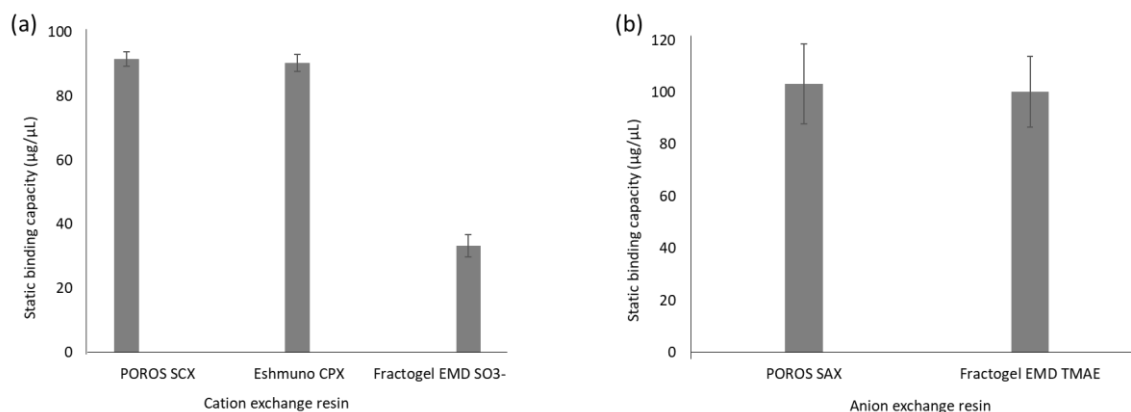


**Figure 7.** Chromatogram of a cation exchange gradient chromatography (CEX) of commercial ovalbumin. Detector: UV 280 nm; column: Sepac Proteomix SCX NP-3; buffer A: 20 mM formate buffer pH 3; buffer B: 1M NaCl in buffer A. Gradient elution 0-3 min (0% B); 3-8 min (0-100% B); 8-13 min (100% B); flowrate 0.2 mL/min.

The presence of different modifications in different proteoforms contribute to their difference in net surface charge, resulting in slightly different isoelectric point (pI), making ion exchange chromatography as a good choice for proteoforms purification. In phosphorylation, negatively charged phosphate group substitutes the neutral hydroxyl groups on serine, tyrosine, or threonine residues. It induces an acidic shift on the pI. Deamidation, acetylation, and glycosylation induced by sialic acid also change the pI to a lower point. Uncyclized N-terminal glutamine, C-terminal lysine, C-terminal proline amidation, and succinimide formation of an aspartic acid contribute to more basic protein and increase the pI. Basic glycoproteins show a decrease of 0.22-0.3 pI unit whereas acidic glycoproteins show a decrease of only 0.1-0.2 pI unit (81). Therefore, ion exchange chromatography resin was chosen for enrichment and separation of proteoforms.

Three different CEX resins, named POROS SCX, Eshmuno CPX, Fractogel EMD SO<sub>3</sub><sup>-</sup> and two anion exchange chromatography (AEX) resins, named POROS SAX and Fractogel EMD TMAE were used for PPS. Since sample displacement work best in an overloading condition, the experimental static binding capacity of each resin was determined. First, the binding capacity of each resin was estimated from the binding capacity provided by the manufacture which is 100 µg/µL for POROS SCX, 120 µg/µL for Eshmuno CPX, 130 µg/µL for Fractogel EMD SO<sub>3</sub><sup>-</sup>, 140 µg/µL and 100 µg/µL for POROS SAX and Fractogel EMD TMAE, respectively. Each resin was diluted in the respected sample application buffer (buffer A) to get binding capacity of 2 µg/µL resin suspension. 50 µL of resin suspension was then used for every PPS experiment and SDBC to theoretically bind ±100 µg of total protein. **Figure 8** shows experimental static binding capacity of each resin for the ovalbumin sample. All resin showed binding capacity close to what was estimated except for Fractogel EMD SO<sub>3</sub><sup>-</sup> which had a low binding capacity (less than 40 µg/µL). Therefore, the amount of resin used in PPS and SDBC was adjusted in a way that 100 µg of total protein will be bound to the resins, allowing enough amount for further analysis. In the end, 1 µL resin was used for all the resins except for Fractogel EMD SO<sub>3</sub><sup>-</sup> in which 2 µL resin was used. 1 mg of sample (10 times overload of the static binding capacity) was used for the PPS and SDBC experiments.





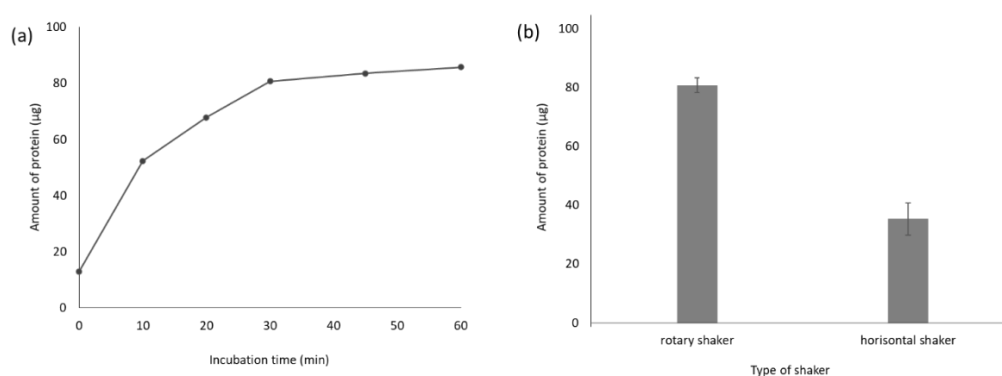
**Figure 8.** Static binding capacities of different types of resin, cation exchange resin (CEX) (a) and anion exchange resin (AEX) (b) for ovalbumin sample. Buffer composition, buffer A: 25 mM formate buffer pH 4 for CEX and 25 mM phosphate buffer pH 7 for AEX; buffer B: 1M NaCl in respected buffer A.

### *Optimisation of incubation time & incubation method*

To perform purification parameter screening and sample displacement chromatography in a batch mode, sample was loaded in a reaction vial, mixed with resin suspension, and incubated to give enough contact between the proteoforms and the resin. Different incubation times and incubation methods were tested to get the optimal time and method for sample incubation and to maximise the binding of proteoforms to the resin. **Figure 9a** shows the total amount of ovalbumin proteoforms eluted from CEX resin after different incubation times. When the protein was loaded to the reaction vial, mixed with the resin, and immediately rested until the resin sedimented (indicated with 0 incubation time in the **figure 9a**) and then the protein was eluted, the protein did not have a chance to contact and bind with the resin. A significant increase in the total amount of bound protein was observed with increase incubation time. From 10 minutes to 30 minutes, the total amount of protein bound to the resin was two times higher. Despite further increase in incubation time to 45 and 60 minutes, only a slight increase in the total amount of bound protein was observed which were 81 µg total protein when incubated for 30 minutes, 84 µg and 86 µg when incubated for 45 and 60 minutes, respectively. Therefore, 30 minutes incubation was decided for the PPS and SDBC experiment.

In addition to the incubation time, the way the batch chromatography (in reaction vial) was shaken played an important role in maximising contact between proteoforms and resin.

When a horizontal shaker was used, sedimentation of the resin in the bottom of the tube was observed, limiting the overall contact with the proteoforms in the sample solution, resulting in lower amount of proteoforms bound and thus eluted from the resin (**figure 9b**). Rotational shaker in 360 degree allowed continues movement of the resin suspension, giving more chances for contact. In the end, the method optimisation showed the optimal way to perform sample displacement chromatography in 1.5 mL reaction vials with sample incubation of 30 minutes in rotational shaker. This method was then used for the PPS and SDBC experiments.

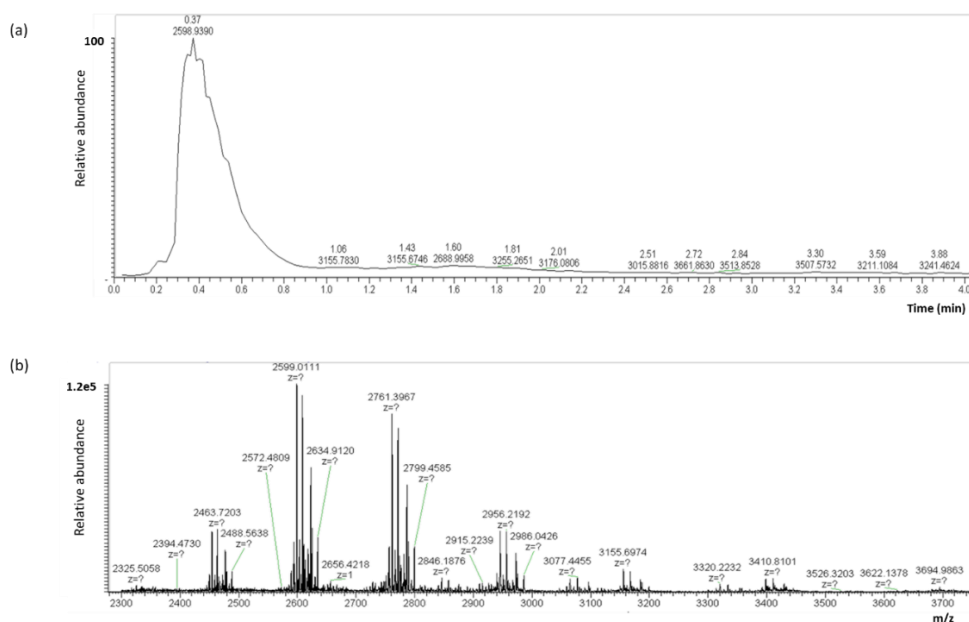


**Figure 9.** Total amount of eluted ovalbumin protein from cation exchange (CEX) POROS SCX resin after different incubation times (**a**) and applying different types of shakers for 30 minutes (**b**). Buffer compositions: see **figure 8**.

### *Optimisation of sample preparation for top-down mass spectrometry*

Apart from the method optimisation for SDBC, attention is given to the sample preparation prior to mass spectrometry analysis. Flow injection analysis coupled with mass spectrometry analysis (FIA-MS) was used to analyse all samples from PPS and SDBC fractions. Top-down mass spectrometry approach offers ease in correlating modification to a specific source giving unambiguous information of each proteoform as it utilizes the intact proteins for analysis instead of peptides. With the use of intact proteins, all the information related to the protein and the proteoforms accompanied with it is retained. The problem of shared peptide in a heterogenous proteoform mix evident in bottom-up approach is avoided (59–61). Direct injection is preferable as it allows faster analysis, The method for the quantification of proteoforms with FIA-MS was developed by Gaikwad, M., (82). In FIA-MS, the sample of intact protein is directly injected to the mass spectrometer allowing faster analysis and avoiding

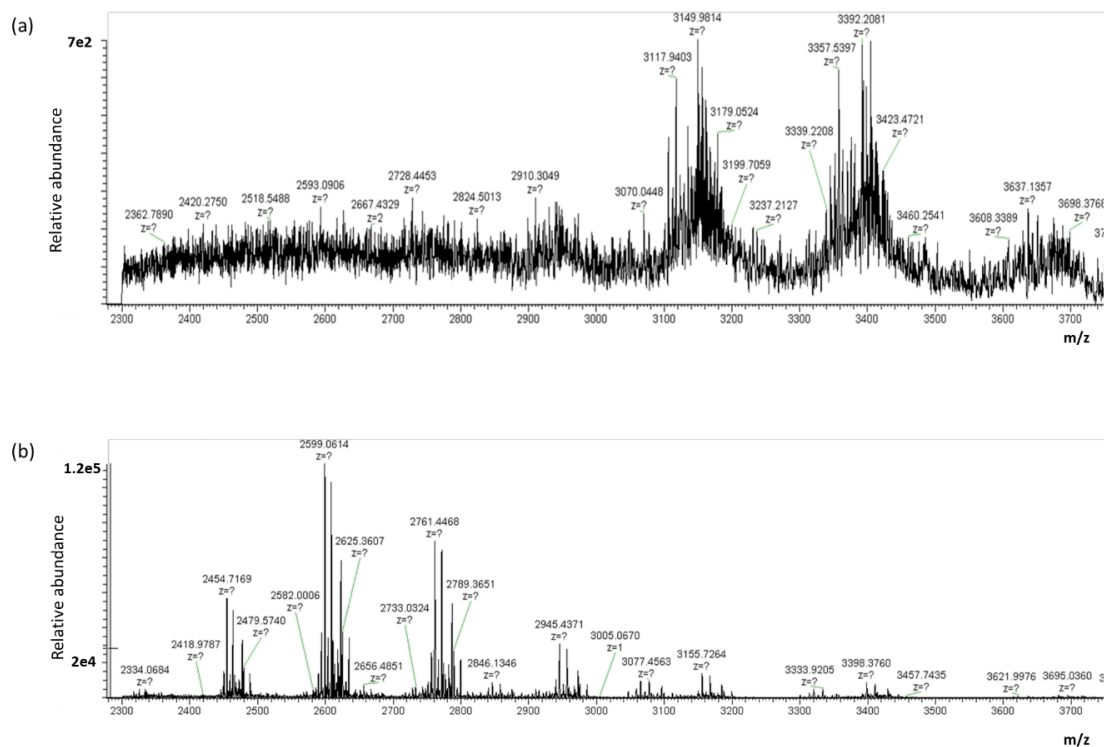
possible protein precipitation due to high organic solvents such as acetonitrile, as well as avoiding the loss of proteoforms in a chromatography column (64,65).. A FIA-diagram of ovalbumin sample and its corresponding full-scan spectra are represented in **figure 10**.



**Figure 10.** FIA diagram of ovalbumin sample from the FIA-MS analysis using hybrid quadrupole-orbitrap mass spectrometer **(a)** and corresponding full-scan spectrum **(b)**. For details see method section: FIA-MS.

In the PPS and SDBC experiments, 1 M NaCl was used as elution buffer (buffer B). The presence of salt can hinder the mass spectrometry analysis as it is known to give ion suppression effect or peak splitting due to the formation of adduct. The presence of Na<sup>+</sup> is particularly problematic, especially when a direct sample injection into an ESI source is used. The study from Donnelly, D., P., et al. (83), showed that as low as 1.5 mM sodium chloride (NaCl) caused a 50% signal loss. . A heterogenous composition such as  $[M + zH + n(Na - H) + m(Cl + H)]z^+$  where both n and m can cover a wide range, tend to happen in the present of non-volatile salt such as NaCl (84). Therefore, a buffer exchange to MS compatible solution was done. All samples were buffer exchanged to 100% water via ultrafiltration. **Figure 11** shows how the mass spectra of ovalbumin were significantly improved after a proper buffer exchange to water. Three filtration steps were inadequate to remove most of the salt and to

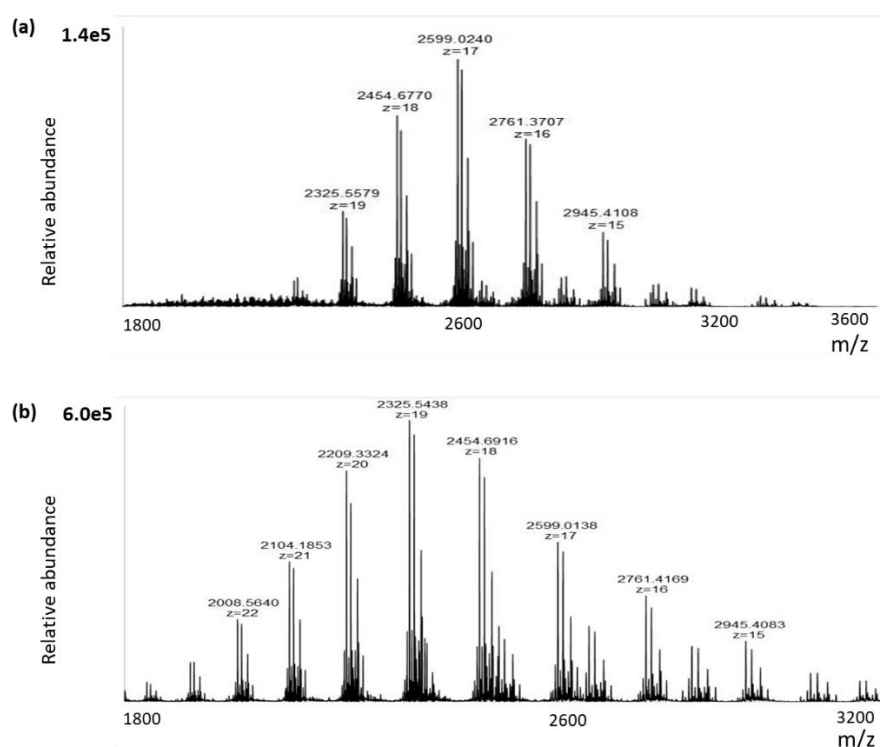
improve the MS signal (**figure 11a**). After six filtration steps, the signal intensity increased from  $7e2$  to  $1.2e5$  and the background noise was significantly reduced, hence better spectra.



**Figure 11.** Full-scan spectra of ovalbumin sample eluted from CEX resin after three times filtration (a) and after six times filtration (b) to buffer exchanged the sample to water. FIA-MS analysis used hybrid quadrupole-orbitrap mass spectrometer. For details see method section: FIA-MS.

Several studies have reported the use of supercharging reagent to decrease the number of sodium adducts of the proteoforms formation and give additional desalting effect to further improve the spectra quality (85–87). One of the most used supercharging reagents is an organosulfur compound sulfolane ( $(\text{CH}_2)_4\text{SO}_2$ ). Sulfolane modulates protein charge values during the ESI which results in large shift in the mass spectral charge stage distributions to the higher charge stage. Sulfolane also helps for further desalting by binding to the sodium ions in solution. It reduces the sodium available to adduct to protein ions (87). A study conducted by Cassou and William (85) has showed that the addition of < 3% of sulfolane was effectively reduce sodium adduction to protein ions by an average of 80%. Another study found that the addition of 5% sulfolane to ovalbumin sample has been proven to help improving the spectra quality of ovalbumin (82). Sulfolane has been proven to be more

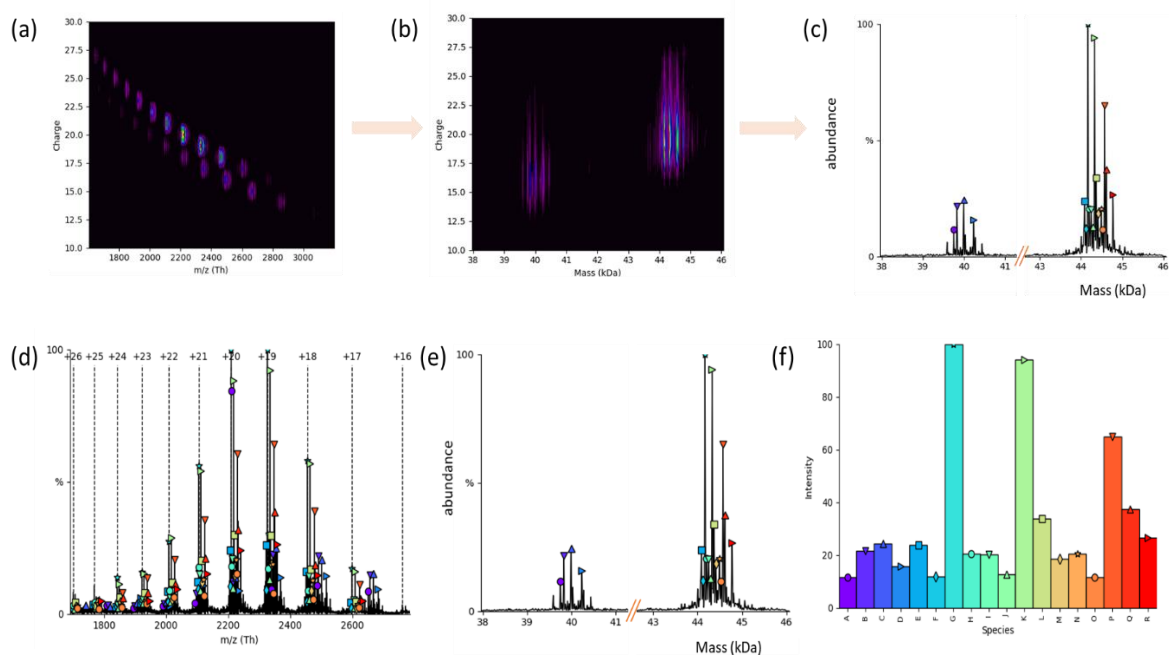
effective at protein ion desalting compared to other supercharging agents such as *m*-nitrobenzyl alcohol (*m*-NBA) (85). Therefore, in this work, 5% sulfolane was added to the ovalbumin sample prior to mass spectrometry analysis. **Figure 12** shows the full-scan spectra of ovalbumin with and without the addition of 5% sulfolane. With the addition of sulfolane the average charge state of ovalbumin proteoforms ions increased by an average of +3 from the sample containing no supercharging agent. Moreover, the addition of sulfolane reduced the noise and improved the intensity of the proteoforms.



**Figure 12.** Full-scan spectra of ovalbumin sample without additional sulfolane (a) and with additional 5% sulfolane as supercharging (b) from the FIA-MS analysis using hybrid quadrupole-orbitrap mass spectrometer. For details see method section: FIA-MS.

The mass spectrum of ovalbumin was then deconvoluted using UniDec deconvolution software from Marty, M. T., et al., (88). UniDec uses Bayesian deconvolution algorithm framework to separate the mass and charge dimensions. **Figure 13** shows the output of UniDec in which two-dimensional delta matrix containing m/z in one dimension and charge in the other dimension is implemented (**figure 13a**). UniDec initially assumes that all charge states have equal probability. The algorithm then proceeds in three steps. Firstly, the charge state distributions are smoothed. Secondly, the m/z vs charge matrix is summed along the

charge axis and convolved with the peak shape. Thirdly, the measured spectrum is compared with the simulated spectrum and adjusted. Finally, the  $m/z$  vs charge matrix is transformed into a matrix of mass vs charge as shown in **Figure 13b**. The sum of the zero-charge mass spectra is used for peak detection (**figure 13c**). UniDec then quantifies the relative abundance of the proteoforms from the deconvoluted spectra by extracting intensities of each proteoform from all charge states.



**Figure 13.** Deconvolution process of ovalbumin sample by UniDec deconvolution software; start with implementation of the charge dimensions into a  $m/z$  vs charge matrix (a), then transformed into a mass vs charge matrix (b), and final deconvoluted spectra in zero charge state (c), annotation of ovalbumin proteoforms and corresponding charge states in the MS spectra (d), deconvoluted of the MS spectra into zero charge state (e), and output of quantification in relative abundance (f).

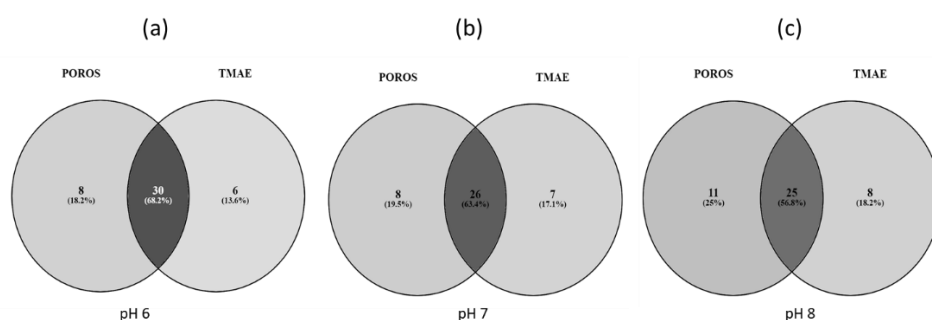
### Purification Parameter Screening (PPS) for Ovalbumin Proteoforms

The method optimisation showed the optimal method to perform sample displacement chromatography in 1.5 mL reaction vials with sample incubation of 30 minutes in rotational shaker and the analysis of the sample with FIA-MS in top-down modus. Next, a series of protein purification parameter screenings (PPS) were performed. Parameters, including different pH values, different resins, and different NaCl concentrations added in sample application buffer were screened to find the optimal conditions for sample displacement

batch chromatography (SDBC) of ovalbumin sample (the detail about parameter screening conditions can be found in the method section: protein purification parameter screening).

#### *Screening of pH and resin for ovalbumin proteoforms purification*

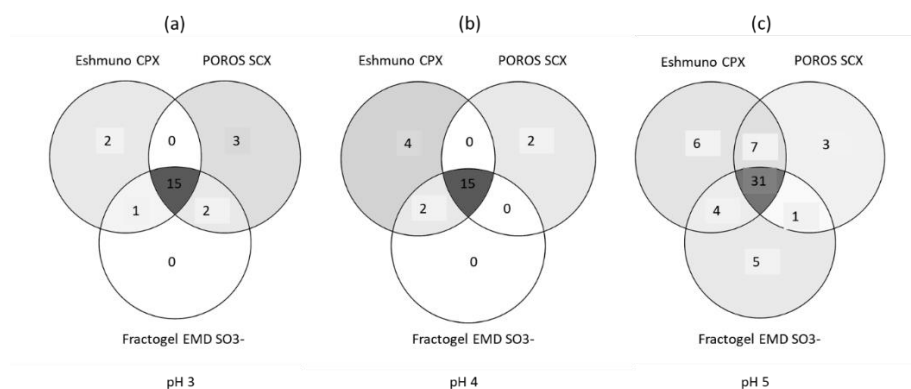
To maintained ion exchange capacity across different pH values, strong ion exchangers, implying that these exchangers do not lose or take up protons and remain fully charged over broad pH range were chosen. Two AEX resins, named POROS SAX and Fractogel EMD TMAE and three CEX resins, named POROS SCX, Eshmuno CPX, Fractogel EMD SO<sub>3</sub><sup>-</sup> were tested. Those resins are different in term of exchanger group and matrix which may have different selectivity for different proteoforms. For the AEX resins, Fractogel EMD TMAE media consist of synthetic methacrylate based polymeric beads while the POROS SAX has polystyrene-divinylbenzene backbone. **Figure 14** shows the overlaid of detected ovalbumin proteoforms in the PPS of both AEX resins in different pH values, in which around 18% to 25% of the proteoforms were detected uniquely in POROS SAX, while 13% to 18% of the proteoforms were detected only in Fractogel EMD TMAE.



**Figure 14.** Number of ovalbumin proteoforms detected with FIA-MS and deconvoluted with UniDec software in the PPS in different anion exchange resins. Buffer composition, buffer A: 25 mM bis-tris buffer pH 6 (a); 25 mM phosphate buffer pH 7 (b); 25 mM tris buffer pH 8 (c), buffer B: 1M NaCl in respected buffer A.

For the CEX resins, the ion exchanger groups of the Fractogel EMD SO<sub>3</sub><sup>-</sup> and the Eshmuno CPX are sulfoisobutyl groups and bonded to linear polymer chains. The POROS SCX resin has rigid polymeric beads coated with a hydrophilic polymer onto which the sulfopropyl groups are

covalently attached. In every pH value, Eshmuno CPX showed better performance with regards to the number of detected proteoforms, followed by POROS SCX and Fractogel EMD SO<sub>3</sub><sup>-</sup> (**Figure 15**).



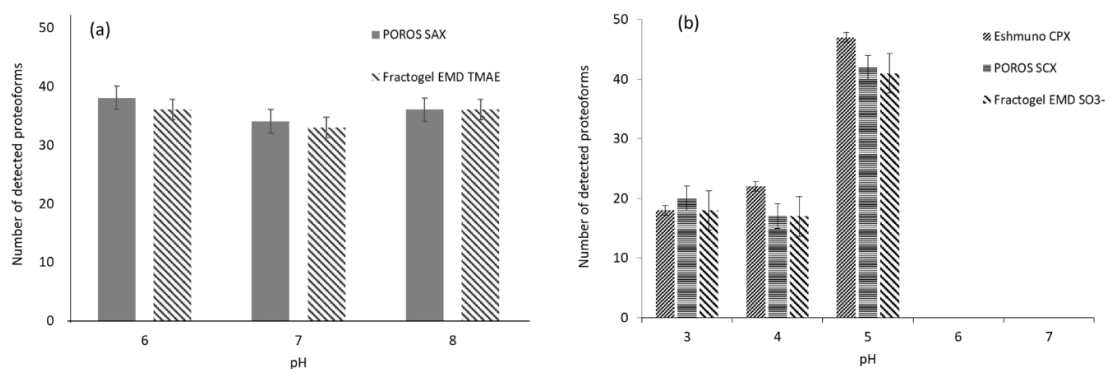
**Figure 15.** Number of proteoforms detected with FIA-MS and deconvoluted with UniDec software in the PPS in different cation exchange resins. Buffer composition, buffer A: 25 mM formate buffer pH 3 (a); 25 mM formate buffer pH 4 (b); 25 mM acetate buffer pH 5 (c), buffer B: 1M NaCl in respected buffer A.

Ovalbumin has an isoelectric point (pI) value range from 4.8 to 5.2 and possibly has a broader pI range since it is such a complex protein with many different modifications, but the pI variation among the ovalbumin proteoforms is very subtle. The pH of sample application buffer (buffer A) was chosen to maximise the difference in total net surface charge of the proteoforms. Three different pH values for AEX (pH 6, 7, and 8) and five different pH values for CEX (pH 3, 4, 5, 6, and 7) were tested.

AEX is beneficial to enrich and purify more acidic proteoforms. In this study, using sample application buffer pH 6, 7, and 8 in AEX, in average of 36 ovalbumin proteoforms were detected in each pH (**figure 16a**). Since the theoretical pI range of ovalbumin proteoforms is lower than 5.5, in all AEX conditions the proteoforms were theoretically negatively charged and all competed for binding sites. The choice of performing AEX in pH lower than 5 is limited by the availability of ideal counter ions. In CEX experiments, no proteoform bound to the resin and detected when the sample was applied in pH 6 and 7 (**figure 16b**). It can be assumed that no ovalbumin (or very less) proteoform has pI higher than 6, therefore, in pH 6 and 7 most proteoforms were negatively charged and did not bind to the resin. In all different CEX resins,

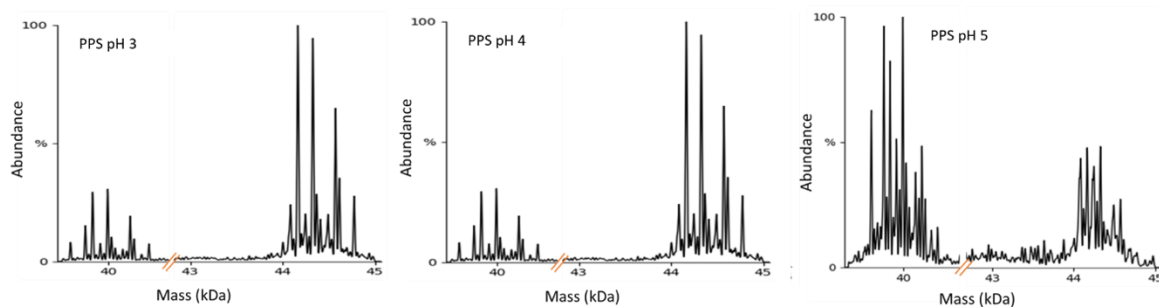


number of proteoforms detected in pH 3 and 4 was considerably low (approximately 20 proteoforms) compared to the number of proteoforms detected in pH 5 which was more than two times higher.



**Figure 16.** Number of ovalbumin proteoforms detected with FIA-MS and deconvoluted with UniDec software in PPS with different pH and different anion exchange chromatography resins (a) and cation exchange chromatography resins (b). Buffer composition: see **figure 14** and **figure 15**.

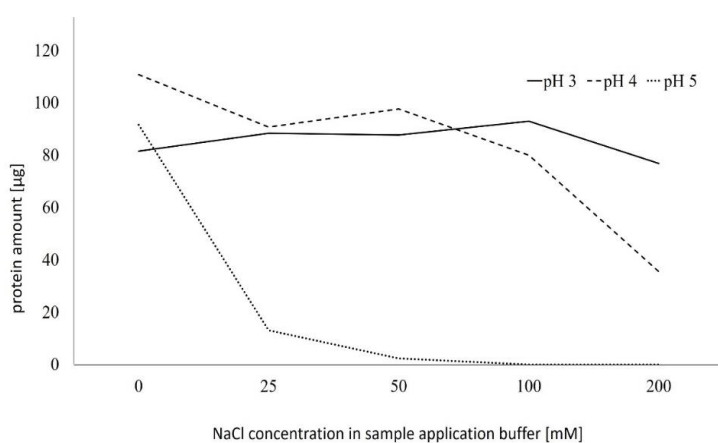
From the deconvoluted spectra of CEX PPS in different pH as shown in **figure 17**, similar patterns were seen in both pH 3 and 4. A significant difference was observed when the pH of the sample application buffer was risen to 5, in which more proteoforms were detected. In buffer pH 5, most proteoforms were negatively charged and only proteoforms with pI higher than 5 were theoretically positively charged. Most of the proteoforms were thus eliminated and more low-abundant proteoforms were enriched and detected. In contrast to sample in pH 3 and 4, most (if not all) of the ovalbumin proteoforms were positively charged. Although, different proteoforms will have slightly different affinity toward the resin in which higher pI proteoforms having higher affinity to the CEX resin, in pH 3 and 4 most proteoforms have the possibility to compete for the binding site. The high-abundant proteoforms are still dominant and the low-abundant proteoforms are not detected. In the end, best parameters and conditions from PPS experiments yielding the best results with respect to the number of detected proteoforms significantly different from other conditions were chosen for SDBC. Therefore, CEX resin Eshmuno CPX and sample application buffer pH 5 were chosen for SDBC of ovalbumin.



**Figure 17.** Deconvoluted spectra of ovalbumin PPSs from cation exchange Eshmuno CPX resin in different pH(s). Buffer composition, buffer A: 25 mM formate buffer pH 3 and 4, 25 mM acetate buffer pH 5, buffer B: 1M NaCl in respected buffer A. For deconvoluted parameters, see method section: data analysis.

### *The effect of additional NaCl in sample application buffer*

It has been reported that the addition of small amount NaCl in sample application buffer may be beneficial for improving selectivity (53). In this study, the addition of different NaCl concentrations ranging from 25mM to 200 mM in sample application buffer was tested. The total amount of ovalbumin eluted from the Eshmuno CPX resin was consistently decreasing as the concentration of additional NaCl increased (**figure 18**). Mild effect was seen in the sample application buffer pH 3 and 4 even when using as high as 100mM NaCl but very strong effect was observed in pH 5 even with the addition of as little as 25 mM NaCl where the amount of ovalbumin bound and eluted form the column decreased significantly. It was concluded that no NaCl in sample application buffer was necessary for SDBC of ovalbumin.

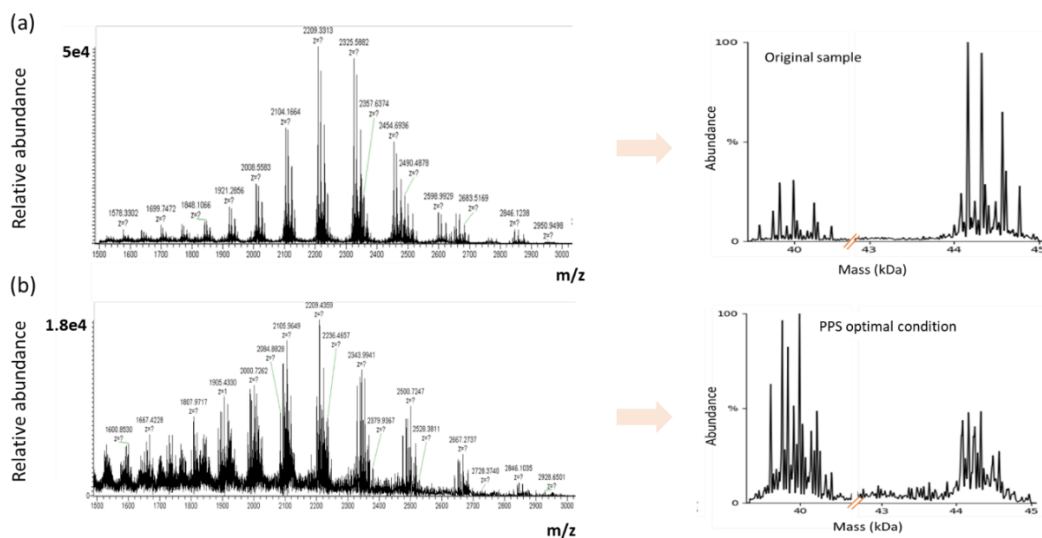


**Figure 18.** Total amount of ovalbumin protein eluted from the cation exchange Eshmuno CPX resin with different NaCl concentration in buffer A. Protein amount was measured with BCA test. Buffer composition: see **figure 17**.

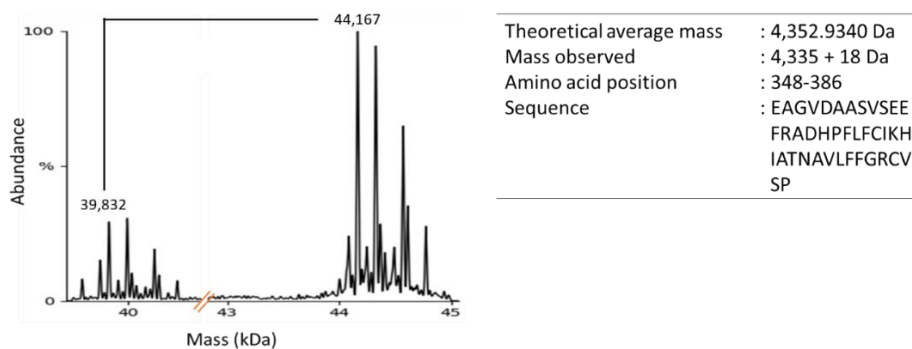
### *Analysis of ovalbumin proteoforms in PPS optimal condition and original sample*

From the method development and PPS experiments, optimal condition for purification of ovalbumin proteoforms was found in using strong CEX resin Eshmuno CPX in batch chromatography using 1.5 mL reaction vials, acetate buffer pH 5 as sample application buffer, 1 M NaCl as the elution buffer, and 30 minutes samples incubation in continuous rotary shaker. FIA-MS in top-down approach was used for proteoforms analysis and UniDec deconvolution software was chosen to analyse the data. Proteoforms detected in ovalbumin original sample and proteoforms detected after PPS optimal condition were then compared.

In PPS optimal condition, a significant enrichment of ovalbumin proteoforms were observed. **Figure 19** shows the MS spectra and corresponding deconvoluted spectra of ovalbumin original sample and ovalbumin eluted from PPS optimal condition. Two clusters of proteoforms were detected in masses of  $\pm 40$ kDa (lower molecular weight/LMW) and  $\pm 44$ kDa area (higher molecular weight/HMW). The cluster of LMW ovalbumin proteoforms were significantly enriched in the PPS optimal condition, while the cluster of HMW proteoforms were considerably reduced. There is a mass difference of 4352 Da between the proteoform with mass of 44167 Da (which belongs to the group of HMW proteoforms) and the proteoform with mass of 39832 Da (which belongs to the group of LMW proteoforms) as shown in **figure 20**. Further MS2 fragmentation clarified that 44167 Da in HMW is ovalbumin proteoform with complete amino acid sequence and additional acetylation, N-glycosylation, and two phosphorylation, while the proteoform of 39832 Da (from the LMW cluster) was most probably C-terminal truncated form of the proteoforms 44167 Da on amino acid position 348 to 386.



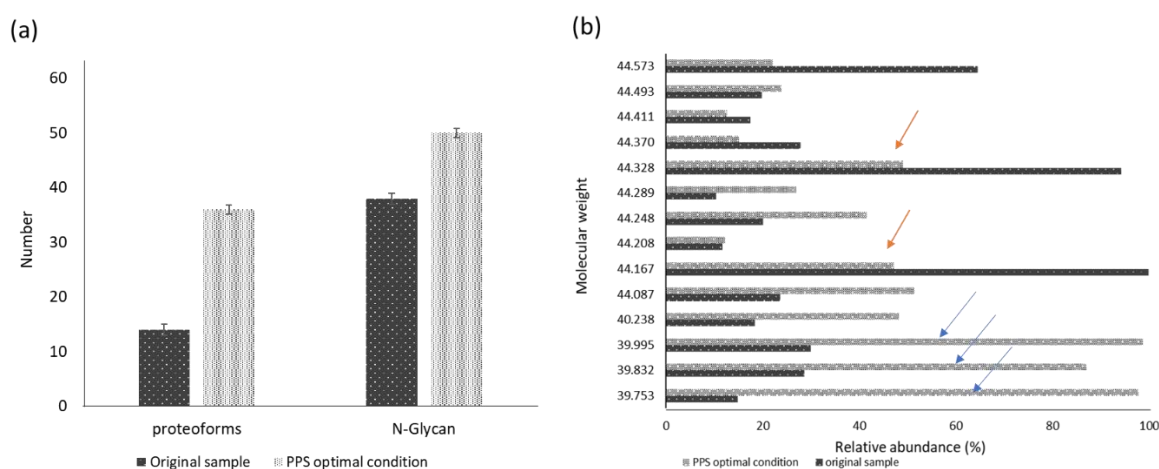
**Figure 19.** Full-scan spectrum and corresponding deconvoluted spectra of ovalbumin original sample (a) and ovalbumin eluted from PPS optimal condition (b). PPS condition, buffer A: 25 mM acetate buffer pH 5, buffer B: 1M NaCl in buffer A, resin: cation exchange Eshmuno CPX. For deconvoluted parameters, see method section: data analysis.



**Figure 20.** Deconvoluted spectra of ovalbumin original sample showing 2 clusters of proteoforms, focus on two proteoforms in lower molecular weight and higher molecular weight which differed by  $\sim 4,352$  Da. For deconvoluted parameters, see method section: data analysis.

Furthermore, **Figure 21a** shows that in PPS optimal condition more than twice the number of proteoforms were detected compared to the original sample. Then, from both eluates of PPS optimal condition and ovalbumin original sample, the attached N-glycan was released by PNGase and analysed separately. It showed an inline result with the number of detected proteoforms in which more N-glycans compositions were identified in PPS optimal condition. In addition to the detection of more proteoforms and their corresponding N-glycans, a closer look at the common proteoforms that were detected in both original sample and PPS eluate

showed a significant enrichment of some proteoforms as seen in **figure 21b**. In the original sample, proteoforms with mass of 44167, 44328, and 44573 Da were three most abundant proteoforms. Then, the relative abundance of those three proteoforms were considerably reduced in PPS optimal condition. On the contrary, proteoforms with mass of 39995, 39753, and 39832 Da, previously less abundant in the original sample, were significantly enriched after the PPS.

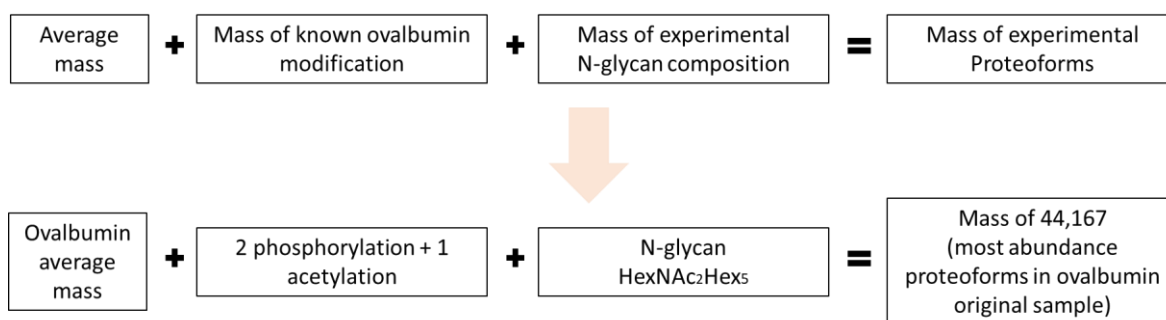


**Figure 21.** Number of ovalbumin proteoforms and corresponding N-glycans composition detected in ovalbumin original sample and ovalbumin eluted from PPS optimal condition **(a)**; relative abundance of common proteoforms detected in both ovalbumin original sample and elution of PPS optimal condition **(b)**. Orange and blue arrows represent examples of decreasing and increasing abundance of some proteoforms. PPS conditions: see **figure 19**. List of all detected proteoforms and identified N-glycan can be found in the **supplement 1 and 2**.

### *Identification of ovalbumin proteoforms with mass-matching*

From the observed experimental masses, identification of ovalbumin proteoforms was performed by mass-matching. **Figure 22** schemes the strategy used for mass-matching identification in which theoretical masses of proteoforms were obtained by adding the mass of known ovalbumin post translational modifications (PTMs) and the mass of experimental N-glycan composition to the ovalbumin average mass. The initial methionine of ovalbumin was cleaved from mature protein thus the average mass of ovalbumin was derived from the amino acid sequences position number 2 to 386 that is 42750 Da (89). Then, the mass of known PTMs

in ovalbumin named N-acetylation of amino acid (delta mass 42 Da) and one or two phosphorylation (delta mass 80 Da) were added. The mass of all N-glycan compositions was obtained from experimental released N-glycan using PNGase was also added. The possibility of ovalbumin proteoforms lacking N-acetylation and lacking phosphorylation were also considered for mass-matching identification. In addition, potential succinimide formation of an aspartic acid (mass decrease 18 Da) was also considered. An example was given for the proteoforms with mass of 44167 that was identified as ovalbumin with two phosphorylation, one acetylation, and one N-glycan which composed of two N-acetylhexosamine and one hexose. **Table 1** listed all ovalbumin proteoforms identified by mass-matching.



**Figure 22.** Scheme of identification of ovalbumin proteoforms by mass-matching strategy giving example of identification of ovalbumin proteoform with molecular weight 44,167 Da. Complete possible identification of other ovalbumin proteoforms is listed in **table 1**.

**Table 1.** List of ovalbumin proteoforms identified by mass-matching.

No.	Mass (Da)	*PTMs	**N-glycan composition
1	39593	GPPS	HexNAc <sub>2</sub> Hex <sub>4</sub>
2	39674	GPS	HexNAc <sub>2</sub> Hex <sub>5</sub>
3	39753	GPA	HexNAc <sub>2</sub> Hex <sub>5</sub>
4a	39832	GPPA	HexNAc <sub>2</sub> Hex <sub>5</sub>
4b		GA	HexNAc <sub>2</sub> Hex <sub>6</sub>
5	39915	GS	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>
6	39954	GAS	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>
7	39995	GPPA	HexNAc <sub>2</sub> Hex <sub>6</sub>
8	40036	GPAS	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>
9a	40076	GA	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>
9b		GPPS	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>
10	40158	GP	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>
11	40200	GPA	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>
12	40238	GPP	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>

13	40279	GPPA	Neu5Ac1HexNAc3Hex3Fuc1
14	40441	GS	Neu5Ac1HexNAc4Hex5Fuc1
15a	44070	GP	Neu5Ac1HexNAc3Hex3Fuc1
15b		GPAS	HexNAc2Hex5
16a	44086	GPA	HexNAc2Hex5
16b		GP	HexNAc3Hex4
17a	44123	GPP	HexNAc2Hex5
17b		GAS	Neu5Ac1HexNAc2Hex4
18a	44167	G	HexNAc3Hex5
18b		GPPA	HexNAc2Hex5
18c		GPP	HexNAc3Hex4
19a	44208	G	HexNAc4Hex4
19b		GPPA	HexNAc3Hex4
19c		GPP	HexNAc4Hex3
19d		GP	HexNAc2Hex6
20a	44231	GPS	HexNAc3Hex5
20b		GPAS	HexNAc2Hex6
20c		GPPAS	HexNAc4Hex3
21a	44248	GPA	HexNAc2Hex6
21b		GPPA	HexNAc4Hex3
21c		GP	HexNAc3Hex5
22a	44289	G	Neu5Ac1HexNAc3Hex4
22b		GPA	HexNAc3Hex5
22c		GPP	HexNAc2Hex6
22d		GP	HexNAc4Hex4
23a	44328	GA	Neu5Ac1HexNAc3Hex3Fuc1
23b		GPPA	HexNAc2Hex6
23c		GPP	HexNAc3Hex5
24a	44370	G	HexNAc4Hex5
24b		GPPA	HexNAc3Hex5
24c		GPP	HexNAc4Hex4
24d		GP	Neu5Ac1HexNAc3Hex4
25a	44394	GPPA	HexNAc4Hex3Fuc1
25b		GPPS	HexNAc5Hex3
25c		GPPAS	HexNAc4Hex4
26a	44411	G	HexNAc4Hex4Red-HexNAc1
26b		GPA	Neu5Ac1HexNAc3Hex4
26c		GPPA	HexNAc4Hex4
26d		GPP	HexNAc5Hex3
26e		GP	HexNAc3Hex6
27a	44476	GPA	HexNAc4Hex4Fuc1
27b		GPS	HexNAc5Hex4
27c		GPAS	HexNAc4Hex5
27d		GPPAS	Neu5Ac1HexNAc3Hex4
28a	44493	GPA	HexNAc4Hex5
28b		GPPA	Neu5Ac1HexNAc3Hex4

28c		GPP	HexNAc3Hex6
28d		GA	HexNAc5Hex4
29a	44531	GPPA	HexNAc3Hex6
29b		GPP	HexNAc4Hex5
30a		G	HexNAc5Hex5
30b	44573	GPPA	HexNAc4Hex5
30c		GPP	HexNAc5Hex4
31a		GPPA	HexNAc5Hex4
31b	44613	GPP	HexNAc6Hex3
31c		GP	HexNAc4Hex6
32a		GA	HexNAc7Hex3
32b	44697	GPA	HexNAc5Hex5
33c		GPP	HexNAc4Hex6
33d		GP	HexNAc6Hex4
34a		G	HexNAc6Hex5
34b	44776	GPPA	HexNAc5Hex5
34c		GPP	HexNAc6Hex4
35	33639		
36	39630		
37	39711		
38	39791		
39	39872		
40	40361		
41	43638		
42	43713		
43	43878		

\*PTMs, N-glycosylation (N), phosphorylation (P), acetylation (A), and succinimide (S). \*\*Proposed N-glycan compositions, Hex (hexose, galactose/mannose), HexNAc (N-acetylhexosamine), Neu5Ac (N-acetylneuraminic acid), and Fuc (fucose). Preferred N-glycan structure for each composition can be found in the supplement 3.

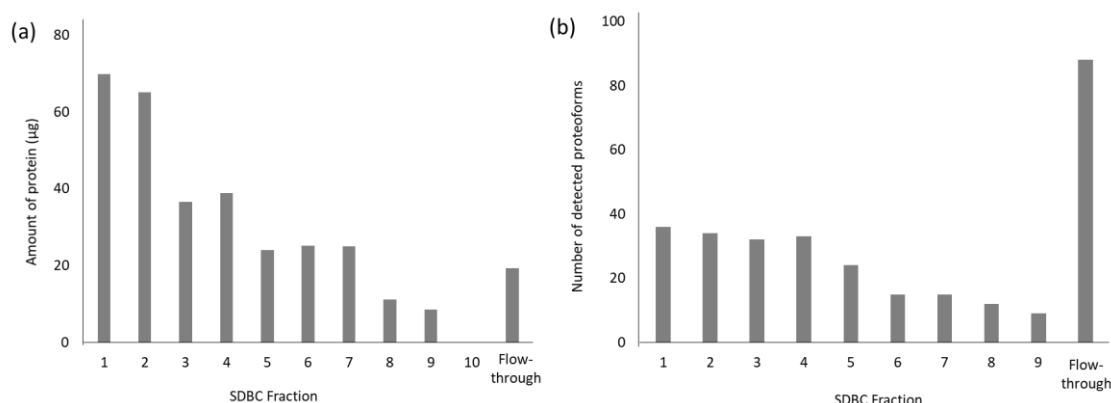
### Sample Displacement Batch Chromatography (SDBC) for Ovalbumin Proteoforms

The optimal conditions for ovalbumin proteoforms purification obtained from PPS experiment were used for the sample displacement batch chromatography (SDBC). The SDBC experiment was carried out in ten segments using ten reaction vials in the size of 1.5 mL as described in method optimisation. The parameters and conditions from the PPS optimal condition were Eshmuno CPX resin, acetate buffer pH 5 for sample application buffer (buffer A), and 1M NaCl in buffer A for the elution buffer (buffer B). With this system, proteoforms were distributed and separated throughout the segments according to their affinity to the



resin. In total, ten fractions from ten segments, one original sample and one flow-through were analysed by FIA-MS.

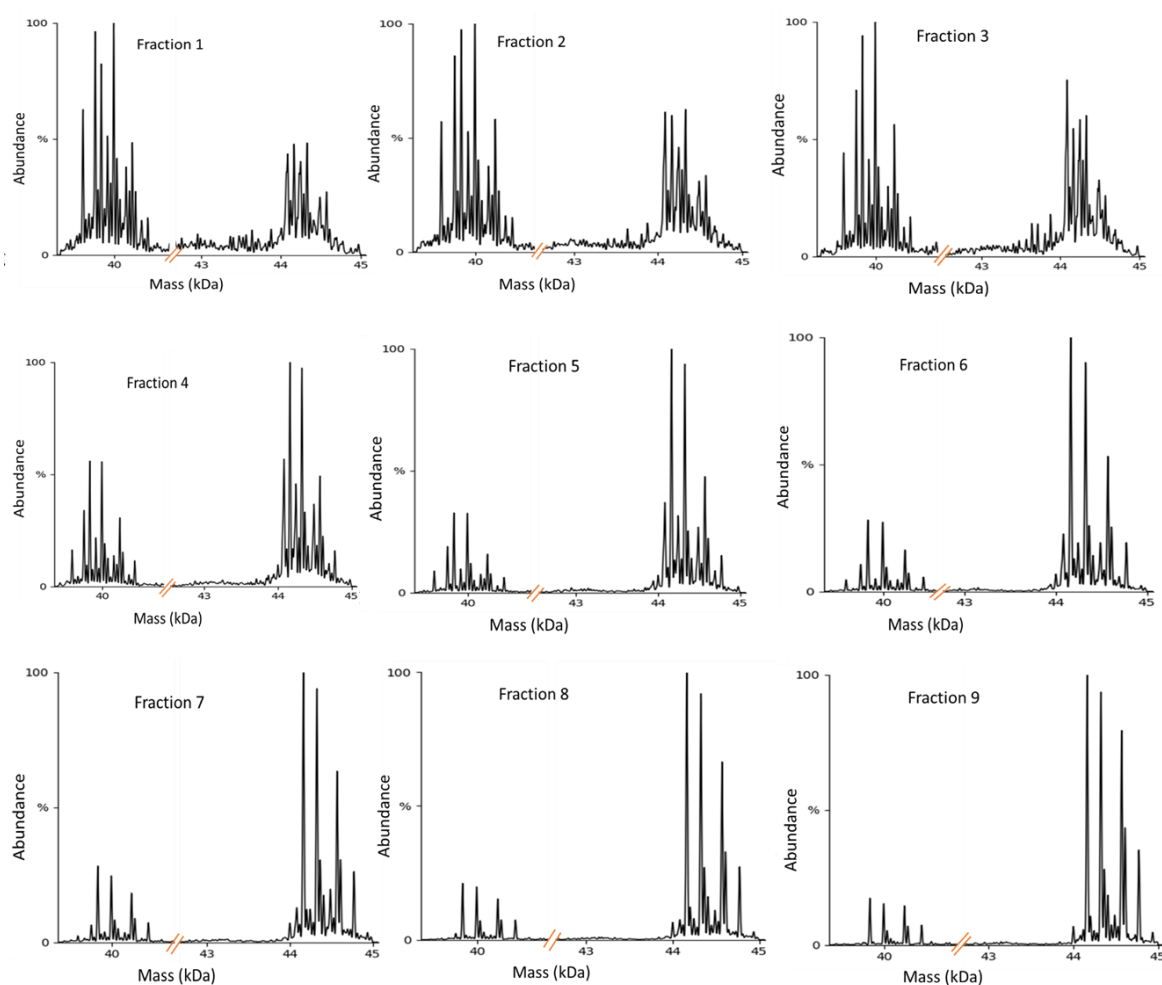
**Figure 23a** shows the total amount of ovalbumin proteoforms eluted from all SDBC fractions, demonstrating a decrease of total amount from the first segment to the last segment as it typically happens in displacement chromatography. At fraction number ten, no more ovalbumin proteoforms bound to the resin. The MS spectrum of each fraction was deconvoluted with UniDec deconvolution software and the number of detected proteoforms in each fraction was given in **figure 23b**. The number of detected proteoforms was decreasing in the later fractions (fraction 5 to fraction 9), hinting purification and enrichment of some proteoforms in different SDBC fractions. High number of proteoforms were detected in the SDBC flow-through hinting good purification of ovalbumin proteoforms from other contaminant or other proteins from the commercial sample.



**Figure 23.** Total amount of ovalbumin proteoforms eluted from each SDBC fraction measured with BCA test (a) and number of detected ovalbumin proteoforms in each SDBC fraction detection with FIA-MS and deconvolution with UniDec software (b). For detail, see method section: SDBC.

The deconvoluted spectra of all SDBC fractions can be seen in **figure 24**, giving an overview of how SDBC worked for the enrichment and purification of ovalbumin proteoforms. In the first two fractions, LMW proteoforms in the cluster of  $\pm 40$ kDa were highest in abundance. The pattern of proteoforms abundance slightly changed in the fraction 3 and 4 in which an increase in abundance of HMW proteoforms (around  $\pm 44$ kDa) was observed. From fraction 5, the abundance of LMW proteoforms was significantly decreased and the HMW

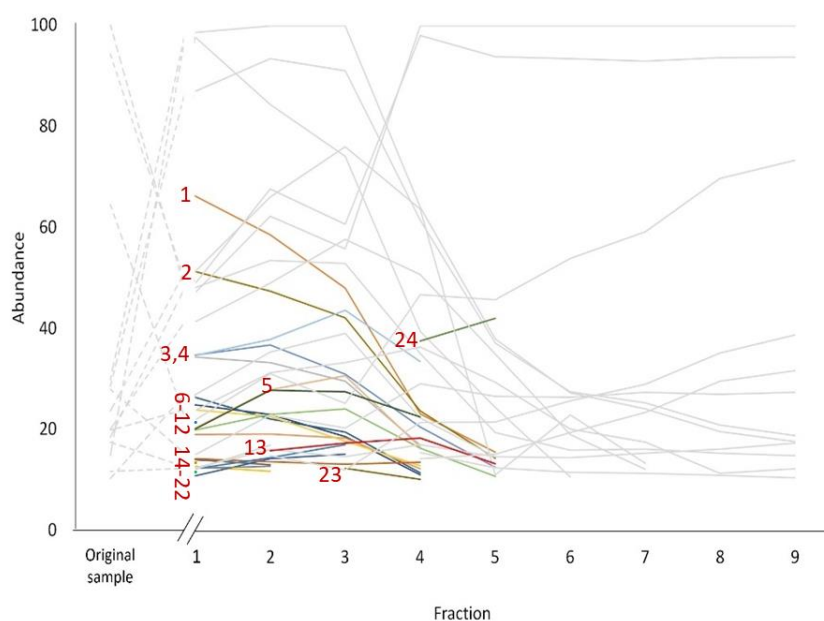
proteoforms became the most abundant. To the later fractions, the abundance of LMW proteoforms continuously decreased until the fraction 9. It can be seen clearly that the abundance of proteoforms had significantly changed from fraction 1 which indicating some degree of purification of proteoforms. All proteoforms detected in SDBC fraction was listed in **supplement 1**.



**Figure 24.** Deconvoluted spectra of each ovalbumin SDBC fraction. List of all proteoforms detected in each fraction can be found in the **supplement 1**. For deconvoluted parameters, see method section: data analysis.

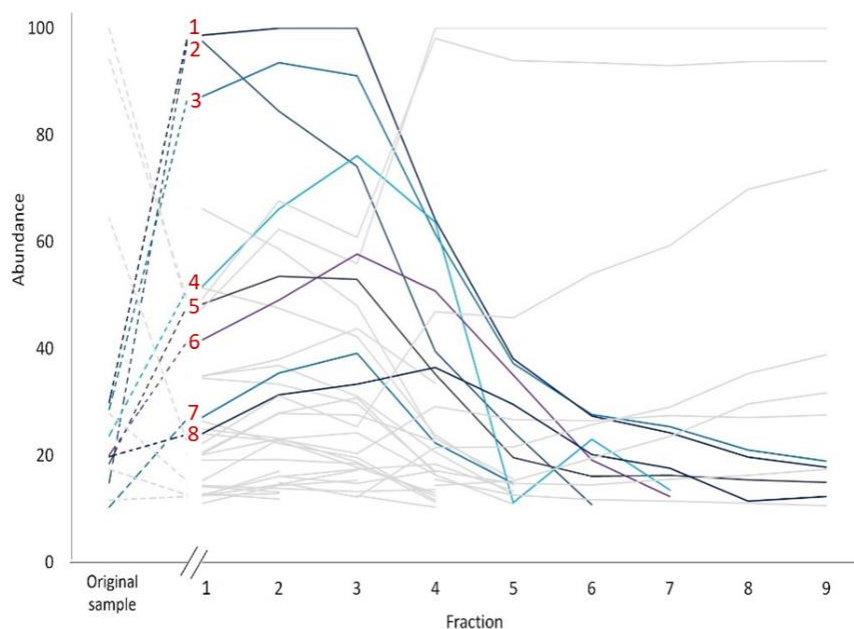
The relative abundance of each proteoforms in all SDBC fractions was then quantified using UniDec deconvolution software as previously described in the result section '*sample analysis with mass spectrometry*'. The relative abundance of all proteoforms detected in ovalbumin original sample and in the SDBC fractions was plotted as chromatograms in **figure 25** which

focuses on 24 proteoforms enriched and detected only in SDBC early fractions. These 24 ‘new’ proteoforms were previously undetected in the original sample due to their very low abundance, but then were enriched and detected in SDBC early fractions (fraction 1 to 5). Most of those proteoforms had bound to the resin in early fraction and were no longer detected in the later fractions. In other words, in the fractions 6 to 9, other ovalbumin proteoforms have been purified from these low abundant proteoforms. As the SDBC experiment was performed using CEX chromatography resin, those low abundant proteoforms were predicted as more basic ovalbumin proteoforms. Identification of these proteoforms by mass-matching was conducted and it was observed that 12 out of 24 proteoforms shows a mass decrease of ~18 Da which potentially corresponds to succinimide formation of an aspartic acid (**table 1**).

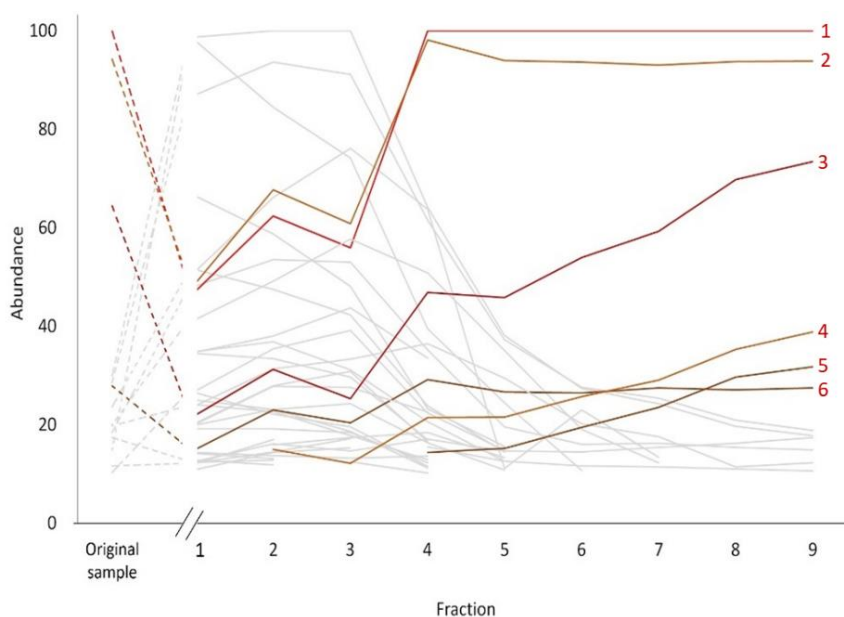


**Figure 25.** Chromatograms of all proteoforms detected in ovalbumin original sample and in the eluates of SDBC fractions, focus on proteoforms enriched and detected only in SDBC early fractions (represented in colour-lines) while other proteoforms were presented in grey-lines. Mass of proteoforms in **1-24**: 39593; 39630; 39674; 39711; 39791; 39872; 39915; 39954; 40036; 40076; 40158; 40200; 40279; 40361; 40441; 43638; 43713; 43878; 44070; 44123; 44231; 44394; 44476; 44531 Da. Relative abundance of the proteoforms was obtained from summed intensity-based quantification using UniDec deconvolution software in which the highest abundance proteoform in each fraction was given a value of 100 and the abundance of other proteoforms was calculated relatively to the most abundance proteoform.

Furthermore, SDBC was also able to separate ovalbumin proteoforms into two categories: highly enriched proteoforms in early fractions and highly enriched proteoforms in later fractions. **Figure 26** shows the chromatograms of proteoforms in SDBC fractions which focus on highly enriched proteoforms in early fractions. These proteoforms were previously detected in low abundance in the original sample and had been highly enriched in SDBC early fractions. **Figure 27** shows the chromatograms of proteoforms in SDBC fractions which focus on highly enriched proteoforms in later fractions. These proteoforms were previously detected in low abundance in the early fractions and had been highly enriched in the later fractions. The identification of the proteoforms in both categories are listed in **table 1** in which highly enriched proteoforms in early fraction are most probably ovalbumin with one N-glycan, one acetyl, and one phosphate attached on it, while the highly enriched proteoforms in the later fractions are most probably ovalbumin with one N-glycan, one acetyl, and two phosphates attached on it. It clarified that an additional one phosphorylation made the proteoforms less basic and thus displaced by more basic proteoforms and moved to the later fractions.



**Figure 26.** Chromatograms of all proteoforms detected in ovalbumin original sample and in the eluates of SDBC fractions, focus on highly enriched proteoforms in early SDBC fractions (represented in colour-lines) while other proteoforms were presented in grey-lines. Mass of proteoform in 1-8: 39995; 39753; 39832; 44086; 44248; 40237; 44288; 44493 Da. For detail about relative abundance, see **figure 25**.



**Figure 27.** Chromatograms of all proteoforms detected in ovalbumin original sample and in the eluates of SDBC fractions, focus on proteoforms detected in higher abundance in the later SDBC fractions (represented in colour-lines) while other proteoforms were presented in grey-lines. Mass of proteoform in 1-6: 44167; 44328; 44573; 44613; 44775; 44369 Da. For detail about relative abundance, see **figure 25**.

Identification of ovalbumin proteoforms in SDBC fractions by mass-matching also shows a high probability that one mass may correspond to several proteoforms with various combination of PTMs as shown in **table 2**. For the example, a detected mass of 44328 Da may corresponds to at least three different proteoforms either ovalbumin with one N-glycan two phosphorylation one N-terminal acetylation (GPPA), or one N-glycan two phosphorylation without N-terminal acetylation (GPP), or one N-glycan one N-terminal acetylation without phosphorylation (GA). When ovalbumin is more phosphorylated, negative charge is added, while when N-terminal of ovalbumin is acetylated, a positive charge is lost. These subtle differences in chemical properties contribute to the different pI and to the overall surface charge of proteoforms in the given pH buffer solution. They lead to different chromatography behaviour and affinity of the proteoforms toward CEX resin in which proteoforms with GPPA are more acidic and were displaced by more basic proteoforms (GPP and GA). This phenomenon rises the possibility that a mass of 44328 Da detected in SDBC early fractions

and later fractions may belong to completely different proteoforms. The same principle is also applied for all other detected masses.

In addition to the intact protein analysis of all SDBC fractions, a release N-glycan was performed for each SDBC fraction. **Table 3** shows that in SDBC experiment there were 20 different N-glycan compositions that were identified in two distinct clusters, N-glycans identified only in SDBC early fractions and N-glycans identified only in later fractions. However, further analysis of each N-glycan showed that the clustering of N-glycan does not directly correlate to the chromatography behaviour of the proteoforms in CEX. Nevertheless, it demonstrated that there was a separation of ovalbumin proteoforms in SDBC early and later fractions.

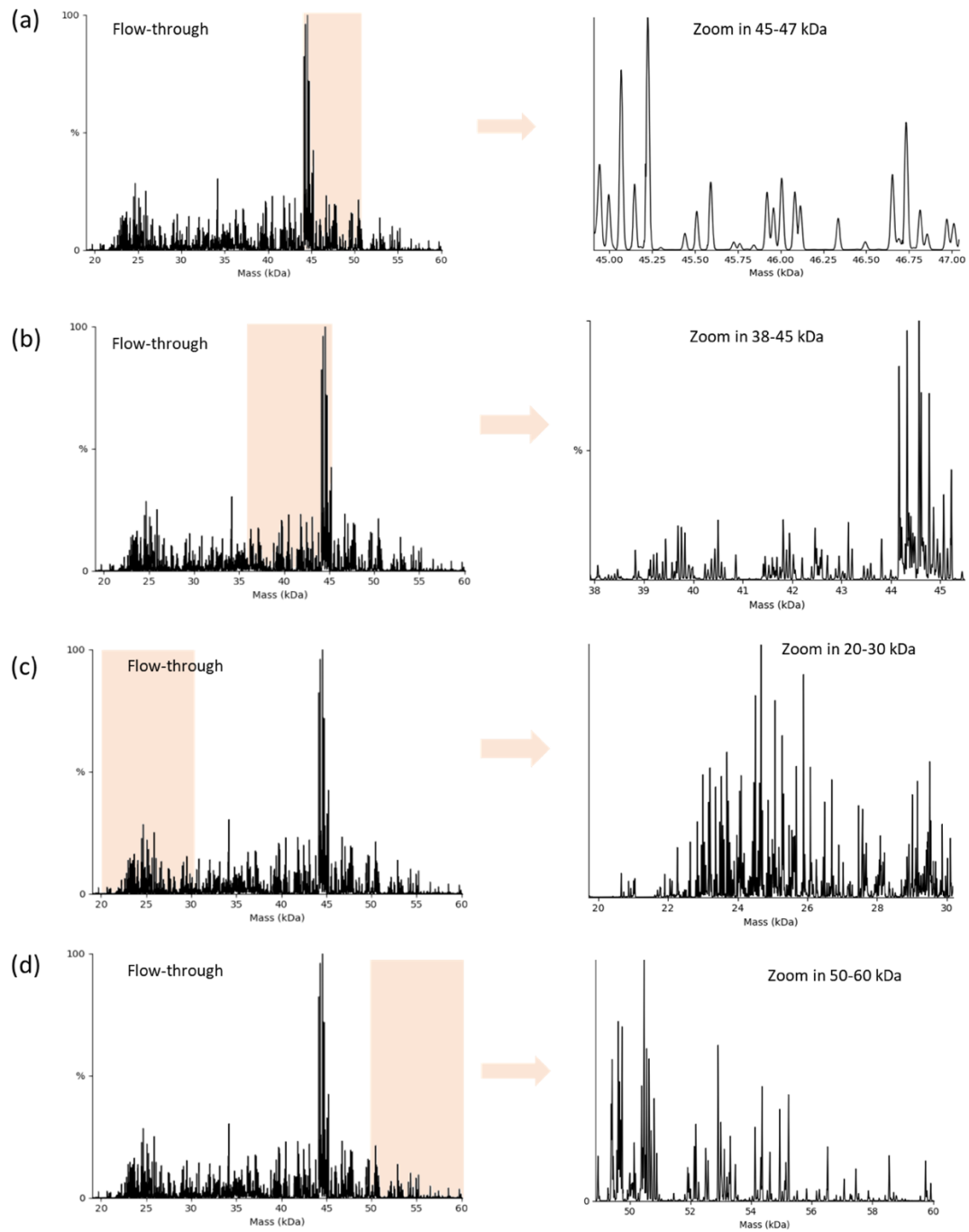
**Table 2.** Example of ovalbumin proteoforms enriched in the later SDBC fractions, identified by mass-matching, showed possibility of different proteoforms with the same mass.

Proteoform molecular weight (Da)	Possible identifications	Possible N-Glycan compositions
44167	<ul style="list-style-type: none"> <li>- Ovalbumin + G + 2P + A</li> <li>- Ovalbumin + G + 2P</li> <li>- Ovalbumin + G</li> </ul>	<ul style="list-style-type: none"> <li>- HexNAc2Hex5</li> <li>- HexNAc3Hex4</li> <li>- HexNAc3Hex5</li> </ul>
44328	<ul style="list-style-type: none"> <li>- Ovalbumin + G + 2P + A</li> <li>- Ovalbumin + G + 2P</li> <li>- Ovalbumin + G + A</li> </ul>	<ul style="list-style-type: none"> <li>- HexNAc2Hex6</li> <li>- HexNAc3Hex5</li> <li>- Neu5Ac1HexNAc3Hex3Fuc1</li> </ul>
44573	<ul style="list-style-type: none"> <li>- Ovalbumin + G + 2P + A</li> <li>- Ovalbumin + G + 2P</li> <li>- Ovalbumin + G</li> </ul>	<ul style="list-style-type: none"> <li>- HexNAc4Hex5</li> <li>- HexNAc5Hex4</li> <li>- HexNAc5Hex5</li> </ul>
44613	<ul style="list-style-type: none"> <li>- Ovalbumin + G + 2P + A</li> <li>- Ovalbumin + G + 2P</li> <li>- Ovalbumin + G + P</li> </ul>	<ul style="list-style-type: none"> <li>- HexNAc5Hex4</li> <li>- HexNAc6Hex3</li> <li>- HexNAc4Hex6</li> </ul>

**Table 3.** N-glycans identified in ovalbumin SDBC, showing two distinct clusters in early and later fractions.

No.	N-glycan composition	Fraction								
		1	2	3	4	5	6	7	8	9
1	HexNAc2Hex <sub>3</sub> Fuc <sub>1</sub>									
2	HexNAc3Hex <sub>4</sub> Fuc <sub>1</sub>									
3	Neu5Ac <sub>1</sub> HexNAc3Hex <sub>5</sub>									
4	HexNAc3Hex <sub>7</sub>									
5	HexNAc5Hex <sub>5</sub>									
6	Neu5Ac <sub>1</sub> HexNAc3Hex <sub>7</sub>									
7	HexNAc7Hex <sub>5</sub>									
8	Neu5Ac <sub>1</sub> HexNAc8Hex <sub>6</sub>									
9	HexNAc3Hex <sub>4</sub> Fuc <sub>1</sub>									
10	Neu5Ac <sub>1</sub> HexNAc5Hex <sub>3</sub>									
11	HexNAc6Hex <sub>3</sub> Fuc <sub>1</sub>									
12	HexNAc1Hex <sub>2</sub>									
13	HexNAc2Fuc <sub>1</sub>									
14	HexNAc2Hex <sub>1</sub> Fuc <sub>1</sub>									
15	HexNAc3Hex <sub>3</sub> Fuc <sub>1</sub>									
16	HexNAc5Hex <sub>3</sub> Fuc <sub>1</sub>									
17	Neu5Ac <sub>1</sub> HexNAc2Hex <sub>3</sub>									
18	HexNAc1Hex <sub>3</sub>									
19	HexNAc7Hex <sub>3</sub> Fuc <sub>1</sub>									
20	HexNAc2									

In addition, the flow-through from the SDBC experiment was analysed. A very crowded deconvoluted spectra was observed as shown in **figure 28**. Zooming in on the area of mass around 20-30 kDa and mass around 50 to 60 kDa showed many masses in which a further bottom-up method identified the present of other proteins or proteoforms which do not belong to ovalbumin, such as ovomucoid, ovotransferin, and other ovalbumin related proteins. Those proteins usually present in the extract of chicken egg. This signify that SDBC was able to further purify ovalbumin from the contaminating proteins in the commercial sample. Another zoom in the deconvoluted spectra of the SDBC flow-through showed masses in the molecular weight around 38-45 kDa and 45-47 kDa which also potentially correspond to either other proteoforms from other proteins in the commercial sample or a very acidic ovalbumin proteoforms that were not binding to the CEX resin.

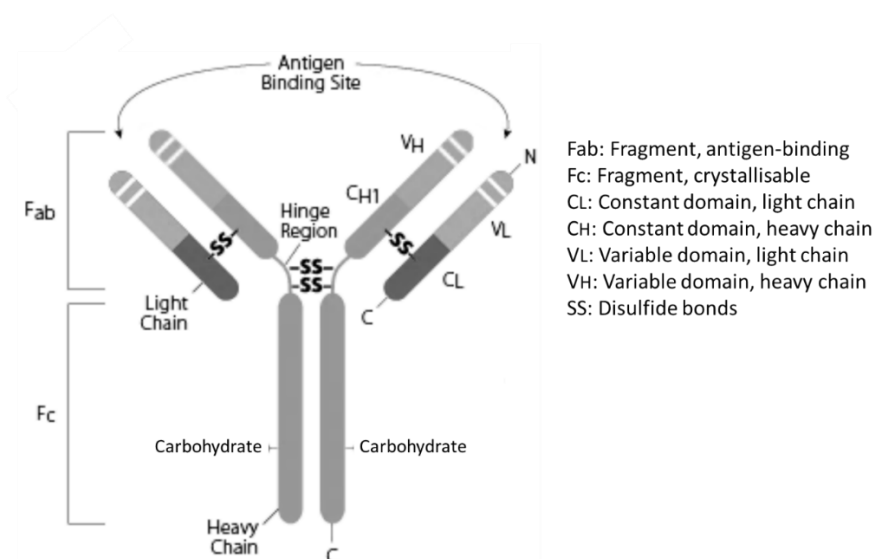


**Figure 28.** Deconvoluted spectra of the flow-through of ovalbumin SDBC fraction. For details, see method section: data analysis.



### Application of PPS and SDBC for Monoclonal Antibody (mAb)

The study of enrichment and purification of ovalbumin proteoforms with SDBC demonstrated that sample displacement batch chromatography can be used for the enrichment and purification of proteoforms. Therefore, further application of SDBC for one of the therapeutic proteins monoclonal antibody adalimumab was conducted. Adalimumab, a tumour necrosis factor (TNF) inhibitor, is a recombinant human monoclonal antibody consists of two kappa light chains and two IgG1 heavy chains (**figure 29**) with a total molecular weight of approximately 148 kDa. Each light and heavy chain consists of 214 amino acid residues and 451 amino acid residues, respectively. Adalimumab has specific binding to TNF both soluble and transmembrane forms, thus blocking TNF's binding to the cell surface TNF receptors (TNFR) p55 (TNFRI) and p75 (TNFRII) (90). It is indicated for the treatment of rheumatoid arthritis, Crohn's disease, ulcerative colitis, hidradenitis suppurativa, ankylosing spondylitis, intestinal Behcet's disease, psoriatic arthritis, and psoriasis (91).



**Figure 29.** Annotated diagram of immunoglobulin (IgG) structure (92).

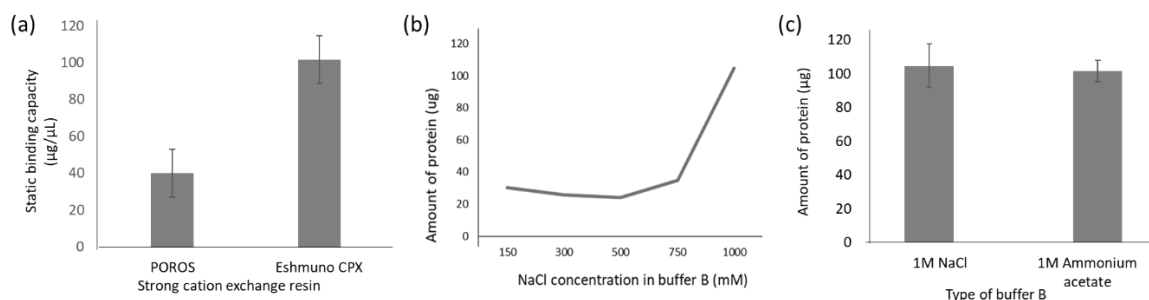
Proteoforms in adalimumab includes C-terminal lysine truncation, deamidation, succinimide aspartic acid (Asp) formation, and glycation. Three main glycoforms in adalimumab are G0F/G0F, G0F/G1F, and G1F/G1F or G0F/G02F. Adalimumab is highly abundance in basic species which are dominated by three forms differing in the number of C-terminal lysine

residues. C-terminal lysine variants underwent succinimide formation of an Asp residue. Hydrolysis C-terminally of Asp is one of the most commonly occurring mAb degradation pathways. Deamidation of adalimumab generates acidic species. Single and double deamidated variants were found in adalimumab. The total of asparagine Asn329 deamidation in adalimumab is around ~25% (93).

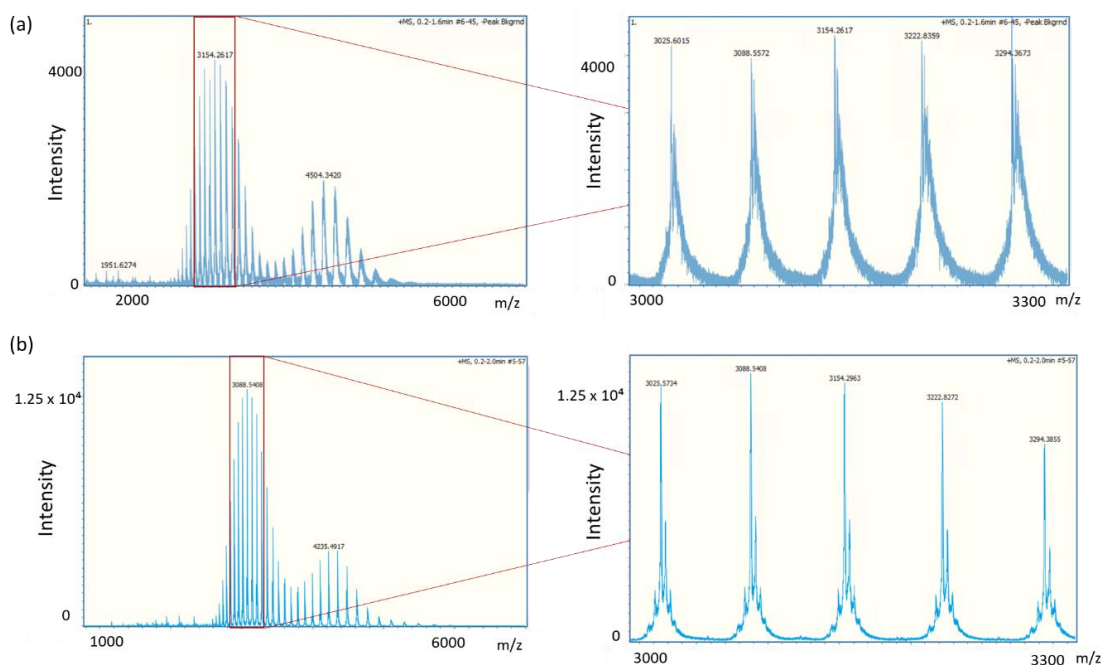
The method used for the PPS and SDBC of adalimumab is the same as the PPS and SDBC for ovalbumin as described in the previous chapter method optimisation for PPS and SDBC. However, the purification of adalimumab proteoforms requires different conditions with the purification of ovalbumin proteoforms. Therefore, to ensure the use of optimal condition for adalimumab proteoforms purification, a series of PPS experiments were performed prior to SDBC. Two strong CEX chromatography resin named Eshmuno CPX and POROS SCX, different pH values of buffer A, and different NaCl concentrations in elution buffer (buffer B) were tested. Cation exchange resins were chosen as the pI of adalimumab proteoforms is high (around 9). Eshmuno CPX resin showed higher static binding capacity (101  $\mu\text{g}/\mu\text{L}$ ) for adalimumab compared to POROS SCX resin in which the static binding capacity is less than half of the theoretical value (**figure 30a**). The theoretical binding capacity according to the manufacture is 120 $\mu\text{g}/\mu\text{L}$  and 100 $\mu\text{g}/\mu\text{L}$  for Eshmuno CPX resin and POROS SCX resin, respectively.

To elute adalimumab proteoforms from the resin, NaCl was used as the elution buffer. The present of salt can hinder the mass spectrometry analysis as it is known to give ion suppression effect or peak splitting due to the formation of adduct. In attempt to reduce the sodium adduct and the effect of ion suppression, different concentrations of NaCl in the elution buffer (buffer B) were tested. However, **Figure 30b** shows that NaCl concentration less than 1M was not enough to fully elute the adalimumab proteoforms from the resin. The samples were then buffer exchanged to water and analysed by FIA-MS. Despite the attempt to buffer exchange the adalimumab samples to water using ultrafiltration for six times, the quality of adalimumab mass spectra was not significantly improved. Therefore, 1M ammonium acetate was then tested to replace NaCl as an elution buffer. **Figure 30c** shows that the use of 1M Ammonium acetate gave comparable result to 1M NaCl regarding the total

amount of adalimumab proteoforms eluted from the Eshmuno CPX resin. After the elution of adalimumab proteoforms with 1M ammonium acetate, the sample was also buffer exchanged to water six times. **Figure 31** shows that the quality of mass spectra was considerably improved after the elution buffer was changed from NaCl to ammonium acetate.

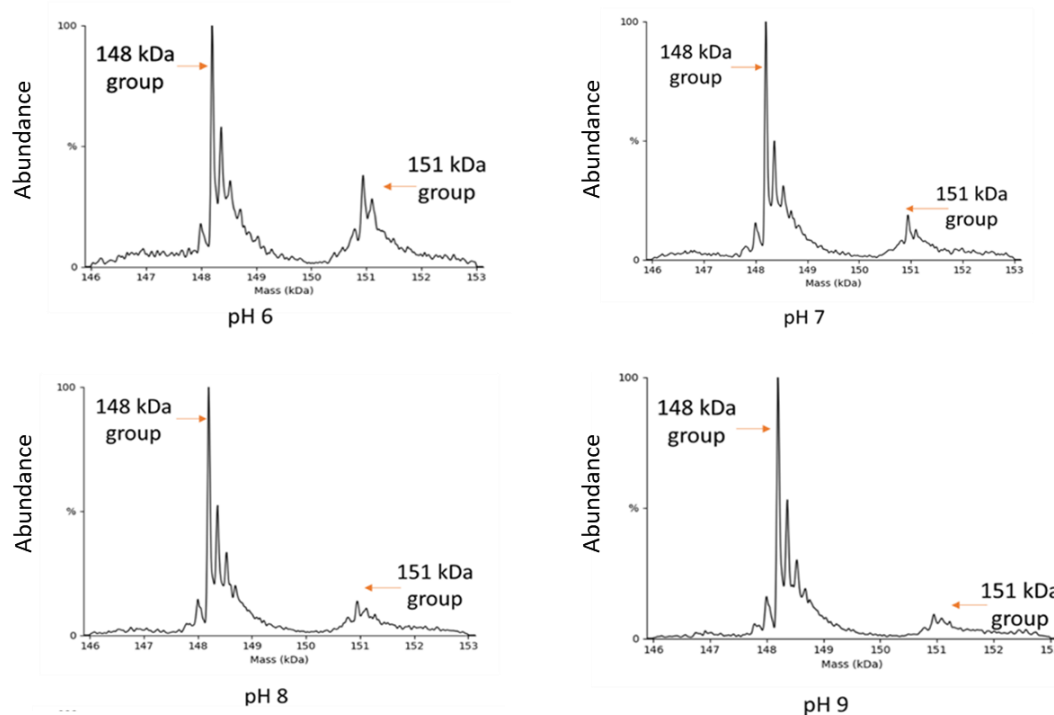


**Figure 30.** Static binding capacity of adalimumab in different resin (a), total amount of adalimumab protein eluted with different NaCl concentration in buffer B (b) and eluted with different type of buffer B (c). Buffer A: ammonium acetate buffer pH 9. Protein amount was measured with BCA test.



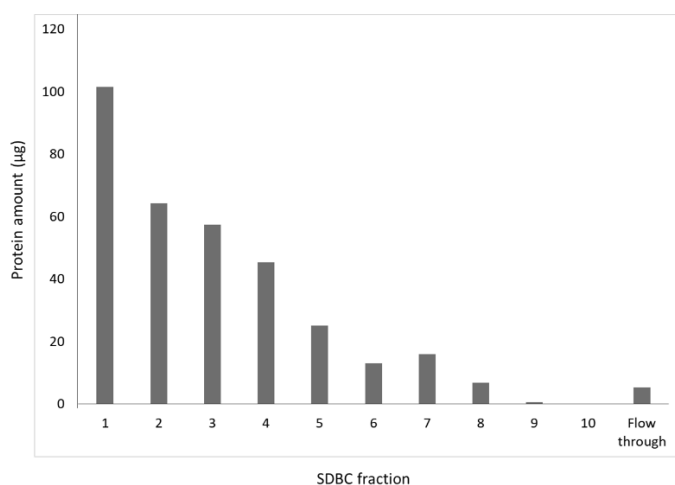
**Figure 31.** Full scan spectrum of adalimumab from CEX Eshmuno CPX resin eluted with 1M NaCl (a) and with 1M Ammonium acetate (b) zoomed in 3000-3300 m/z from the FIA-MS analysis using quadrupole-time of flight (qTOF) mass spectrometer. For details see method section: FIA-MS.

For the screening of different pH, sample application buffer in pH 6 to 9 were tested (For detail composition of buffer, see the method section: PPS for Adalimumab). In those pH values, most of the adalimumab proteoforms will theoretically be positively charged and compete for the binding site in the cation exchange resin but their total net surface charges are different thus different patterns of proteoforms profile is expected. **Figure 32** shows the deconvoluted spectra of adalimumab PPS in different pH values. Two clusters (groups) of adalimumab proteoforms were detected in the PPS experiments, a group of proteoforms with mass around 148 kDa and 151 kDa. In the PPS using buffer A pH 6 the group of 151 kDa proteoforms were enriched. In the contrary, with pH 9 of buffer A the abundance of 151 kDa proteoforms were reduced. It can be assumed that proteoforms in the group of 151 kDa are adalimumab proteoforms with lower pI values or more acidic. Therefore, SDBC in pH 6 can be used to further enrich and analyse the group of 151 kDa proteoforms, while the SDBC in pH 9 can be used to further remove the 151 kDa proteoforms and purify the main adalimumab proteoforms in the group of 148kDa.



**Figure 32.** Deconvoluted spectra of adalimumab PPS eluates from cation exchange Eshmuno CPX resin in different pH(s). Buffer composition, buffer A: 25 mM MES buffer pH 6, 25 mM phosphate buffer pH 7, 25 mM HEPES buffer pH 8, 25 mM ammonium acetate buffer pH 9, buffer B: 1M NaCl in respected buffer A. For deconvoluted parameters, see method section: data analysis.

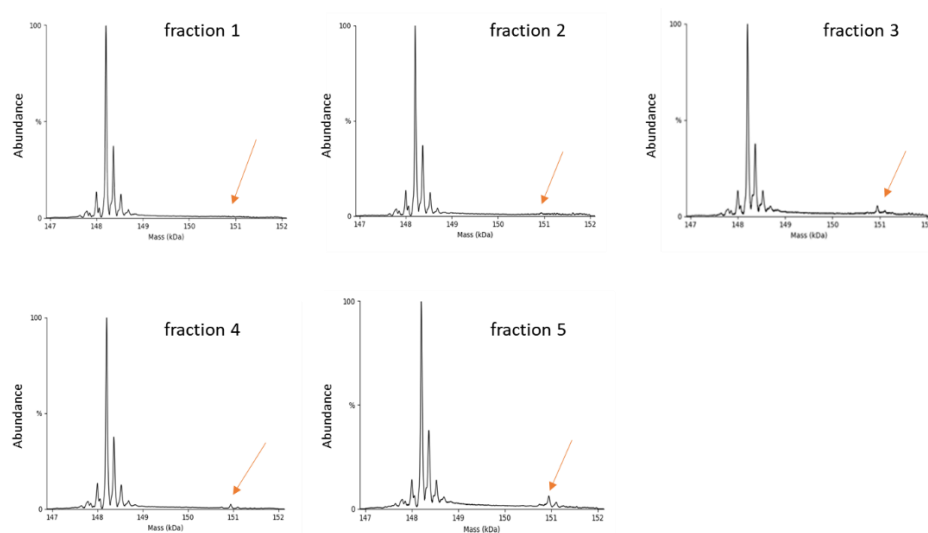
Based on the results from PPS experiments, SDBC for purification and enrichment of adalimumab proteoforms was performed using cation exchange resin Eshmuno CPX with 25 mM buffer A pH 9. pH 9 was chosen for the purification of main adalimumab proteoforms (group of 148 kDa) from the proteoforms in the group 151 kDa. Ammonium acetate 1M was chosen as buffer B. SDBC was carried out in ten segments using 1.5 mL reaction vials as described in method optimisation. In this system, proteoforms were distributed throughout the segments according to their affinity to the resin. **Figure 33** shows the total amount of adalimumab eluted from all SDBC fractions, demonstrating a decrease of amount from the first segment to the last segment as it typically happens in displacement chromatography. At fraction number nine, the amount of adalimumab bound and eluted was very low and in fraction number ten no more adalimumab proteoforms bound to the resin. In total, 8 SDBC fractions, one original sample, and one flow-through were analysed with FIA-MS.



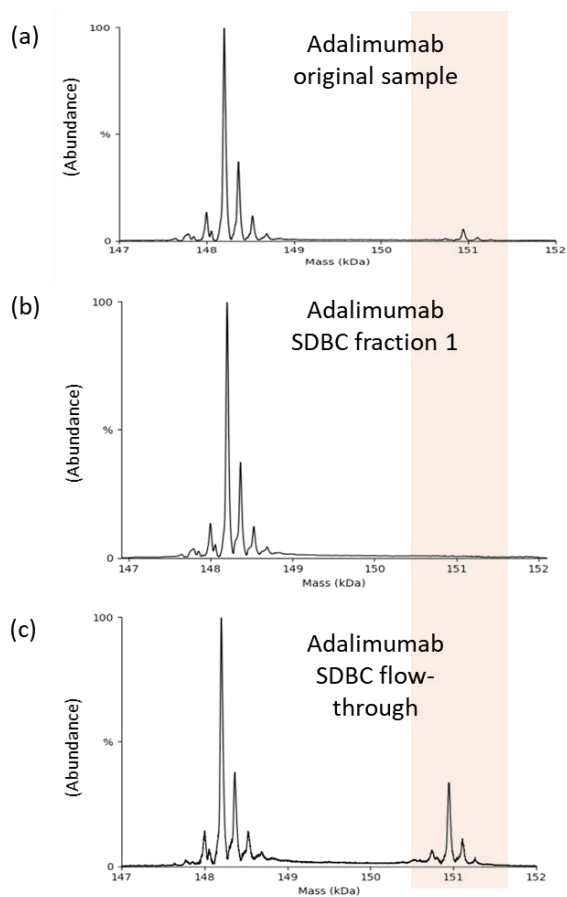
**Figure 33.** Total amount of adalimumab eluted from each SDBC fraction measured with BCA test. Details of SDBC conditions, see method section: SDBC.

Due to the continues problem with contamination, the FIA-MS used for the SDBC of adalimumab was performed in quadrupole time of flight (qTOF) instrument instead of hybrid quadrupole-orbitrap mass spectrometer, optimised by Gaikwad, M., (82). **Figure 34** shows the deconvoluted spectra of SDBC adalimumab of five early fractions which demonstrated the subtle increase of abundance of group 151 kDa proteoforms from the first to fifth fraction.

From this pattern, it can be assumed that there is more increase of abundance of 151 kDa proteoforms occur in the later fractions. However, from fraction 6 to 8 the mass spectra got very complicated, the deconvolution spectra were distorted, and results analysis could not be done accurately. Nevertheless, from the general overview of deconvoluted spectra of adalimumab in SDBC early fractions showed a high possibility that SDBC may work for the purification and enrichment of adalimumab proteoforms. **Figure 35** shows that with SDBC high abundance of the group 151 kDa adalimumab proteoforms were detected in the flow-through of SDBC which demonstrated a significant purification of main adalimumab proteoforms (group of 148 kDa) from the low abundant proteoforms (group of 151kDa).



**Figure 34.** Deconvoluted spectra of adalimumab SDBC 5 early fractions. Details of SDBC conditions and deconvoluted parameters, see method section: SDBC and data analysis.



**Figure 35.** Deconvoluted spectra of adalimumab original sample (a), adalimumab SDBC fraction 1 (b) and adalimumab SDBC flow-through (c). Details of SDBC conditions and deconvoluted parameters, see method section: SDBC and data analysis.

## Discussion

In contrast to small molecule drugs, therapeutic proteins are characterised by their composition of many very similar but not identical species (proteoforms) regarding their composition of atoms. This heterogeneity may arise due to alternatively spliced RNA transcripts, genetic variations, or post translational modifications (PTMs). Therefore, there are many proteoforms encoded by a single gene (5,94). The analysis of proteoforms is highly challenging because proteoforms are physically and chemically very similar and often present in a very low abundance.

Looking at current published studies, a well-established chromatography for separation and purification techniques still mainly focusses on the separation and purification of proteins mixture. There are not many studies aiming to do the separation and purification of proteoforms. Among those studies, ion exchange chromatography (IEX) is mainly used for the analysis of proteoforms (charge variants) in monoclonal antibody either in salt gradient elution (93) or pH gradient elution (35). Very few have utilised displacement chromatography (DC) or sample displacement chromatography (SDC) despite several benefits associated with utilising DC or SDC. In DC and SDC, the quality of separation of DC may be better than gradient chromatography mode, and the enrichment effect on trace components and the concentrating effect yielding the highest concentration of analytes eluting from the stationary phase compared to any other elution modes (43,51,95).

For many years, sample displacement chromatography (SDC) has been used for the separation and purification of proteomics, mixture of peptides and proteins, and separation of human plasma protein fraction Cohn IV-4 (50,52–54). SDC takes advantage of the fact that during sample application, sample components arrange themselves according to the affinity toward the stationary phase. The competition of the sample components for the binding site of the stationary phase can be used for their separation. Further fractionation can be achieved by connecting several columns together in which the components with higher affinity will be retained in the early columns and the lower affinity components will be displaced and bound to the later columns. Therefore, sample components are distributed and separated



throughout the system. Since proteoforms hold subtle differences in their physical and chemical properties that will also affect their chromatography behaviours, in principle SDC should also work for enrichment and purification of proteoforms.

*This study aims to answer the question whether SDC is suitable for separating proteoforms of the same protein and to develop a strategy for the purification of proteoforms using sample displacement batch chromatography (SDBC).*

First, method optimisation for the batch chromatography was performed. From the result obtained in this study, the most benefit was obtained when performing the SDC in batch mode using reaction vials compared to cation exchange chromatography column (Sepax Proteomix SCX-NP3, 4.6 x 50 mm, 3 $\mu$ m). In SDC, the main separation occurs during sample loading and thus it optimally happens in an overloading condition. However, it is not always easy to find appropriate column dimension since protein purification column with volume smaller than 0.1 mL is rare. Therefore, to perform SDC in a column, a high amount of sample is required. A study by Zhang, T., et al. (39) using displacement chromatography in a cation exchange (CEX) column has shown that 200 mg of monoclonal antibody was needed, while another study by Khanal, O., et al. (96) has used around 10 mg of monoclonal antibody for one experiment of SDC using in-house-packed columns. This need of high amount of sample may hinder the experiment when dealing with limited amount of sample. In batch mode, the amount of chromatography resin can be customised, thus the amount of sample used can also be adjusted to what is needed. In this study, only 1  $\mu$ L of resin and 1 mg of ovalbumin sample were used in sample displacement batch chromatography (SDBC). SDBC is a better choice especially when dealing with precious samples with limited availability.

In addition, SDBC does not need a pump system, remove high backpressure problem, and scaling-up or scaling-down is very simple. Batch chromatography can be easily performed in several devices, such as microtiter 384-wells plate, 96 deep well-plates or any type of reaction vials. In this study, the most optimal device is reaction vial in the size of 1.5 mL in which higher buffer volume can be used to avoid high sample viscosity and give possibility to be shaken in 360 degree to allow maximal contact between sample and resin. Ten segments using 10

reaction vials are used for SDBC. Using this batch system, the transfer of supernatant from one segment to the next segment is simpler and easier compared to connecting 10 columns segments which requires more complex instrument setup. Moreover, as preparative column gradient LC is one of the most expensive purification steps, SDBC offers an economical benefit as it is less expensive since the binding capacity of the stationary phase is used more efficiently (54,97).

In the end, the method optimisation showed the optimal way to perform SDBC in 1.5 mL reaction vials with sample incubation of 30 minutes in rotational shaker. This method was then used for the PPS and SDBC experiments. Parameters, including different pH values of sample application buffer, different resins, and different additional NaCl in sample application buffer were screened to find the optimal parameters for SDBC of ovalbumin sample.

Since proteoforms hold only very subtle differences regarding their pI, pH of the sample application buffer plays a major role for the success of enrichment and purification. In this study, it is found that sample application buffer in pH 5 in CEX resin works best to enrich more basic ovalbumin proteoforms. In pH 5, ovalbumin proteoforms have maximum difference in the total net surface charge, thus maximise the purification process. Moreover, to change the interaction between proteoforms and resin, small concentration of NaCl can be added to the sample application buffer. A study of sample displacement chromatography for mixture of proteins (lysozyme, cytochrome C, ribonuclease A and myoglobin) has demonstrated that in the presence of NaCl in sample application buffer, lysozyme was purified better from other proteins (53). However, it is important to note that in the case of protein purification, different proteins have relatively large difference in terms of physical and chemical properties and NaCl might help to prevent the binding of weaker affinity proteins and be beneficial for improving selectivity and maximising purification process. Results from this thesis show that the addition of NaCl in sample application buffer was not beneficial for ovalbumin proteoforms purification especially when using sample application buffer in pH 5. It can be explained that the proteoforms whose pI is only slightly higher than 5 are too weak to bind to the resin with the presence of even a small concentration of NaCl.

In addition to the importance of pH, matrix, bead composition and functional group of the resin also contribute to the success of enrichment and purification. Results obtained from this study shows that strong CEX resin, Eshmuno CPX performs best for ovalbumin regarding the highest number of detected proteoforms compared to Fractogel EMD SO<sub>3</sub><sup>-</sup> and POROS SCX. This result can be explained by looking at the structure Eshmuno CPX. Both Eshmuno CPX and Fractogel EMD SO<sub>3</sub><sup>-</sup> resin have sulfoisobutyl as their functional group and have tentacle structure which allows multi-point interaction and reduces sterical hindrance between the functional group and the proteoforms. However, they differ in term of their polymeric bead. Compared to Fractogel beads which are composed of hydrophilic methacrylate copolymers and have medium rigidity, the beads in Eshmuno resin are hydrophilic polyvinylether polymers which are characterised by their high rigidity due to high cross-linking level (98). These rigid beads in Eshmuno CPX cause proteoforms to bind stronger. POROS SCX resin is known to have robust salt tolerance and consistent binding capacity across broad range of salt concentration for the purification of protein mixtures by minimising competitive binding between the protein of interest and counter ions. It also has rigid polymeric beads coated with a hydrophilic polymer onto which the sulfopropyl groups are covalently attached. The rigid beads in POROS SCX are also known to cause protein to bind stronger. However, for the study of ovalbumin proteoforms, Eshmuno CPX resin performs slightly better than POROS SCX, possibly due to the availability of tentacle structure in Eshmuno CPX resin which allows multi-point interaction and reduces sterical hindrance between the functional group and the proteoforms (99).

This thesis shows that strong CEX resins work well for the enrichment of ovalbumin proteoforms. CEX resins or IEX resins in general are available in wide range varieties regarding the type of matrix, bead composition, and functional group which may offer different selectivity for different proteoforms. Nevertheless, there are also many other chromatography resins which may also have high potential for proteoforms enrichment and purification such as mixed-mode chromatography resins and proteomimer. Those resins have been used successfully for the enrichment and purification of protein from a mixture (100–

104). Therefore, in the future, their potential can be further explored for the enrichment and purification of proteoform.

Considering the results obtained in PPS experiments, it is concluded that the best parameters and conditions for purification of ovalbumin proteoforms are using Eshmuno CPX resin, acetate buffer pH 5 for sample application buffer (buffer A) and 1M NaCl in buffer A as the elution buffer (buffer B). These parameters and conditions are used for SDBC of ovalbumin.

Results obtained in this study show the enrichment of C-terminal truncated form of ovalbumin on amino acid position 348 to 386. This identification of C-terminal truncated forms of ovalbumin is in agreement with the study of complexity of ovalbumin by Füssl, F., et al. (79), in which they have identified ~4300 Da mass difference of LMW species that is suggested to be truncated of the peptide bond in the C-terminal of ovalbumin. Moreover, by using SDBC, 24 'new' low-abundant basic ovalbumin proteoforms were detected and LMW proteoforms were significantly enriched in SDBC early fractions. This gives a proof that SDBC is effective even for the separation of closely related proteoforms. Furthermore, in this study, SDBC was also able to separate ovalbumin proteoforms into two big categories: highly enriched proteoforms in early fractions and highly enriched proteoforms in later fractions (**figure 26** and **27**). In addition to the viability of SDBC for the enrichment of ovalbumin proteoforms, this study also shows that with SDBC, ovalbumin was further purified from other proteins in the commercial sample such as ovomucoid, ovatransferin, and other ovalbumin related proteins (**figure 28**). This shows SDBC as an effective method for the enrichment and purification of proteoforms from the same protein and for the removal of unwanted contaminant proteins.

The caveat in this study is the high heterogeneity and complexity in ovalbumin sample make the proteoforms separation and purification look less distinct. The highly enriched proteoforms in the later SDBC fractions were seemed to be detected in very less abundance in the early SDBC fractions. However, in the analysis of intact proteoforms by mass spectrometry, only the m/z values of proteoforms were detected and then deconvoluted to obtain the mass or molecular weight of each proteoforms. Therefore, there is high possibility

that one detected mass may refer to several different proteoforms. The mass-matching identification shows high possibility that proteoforms with different N-glycan compositions or different combination of PTMs have the same molecular weight (**table 1 and 2**).

Moreover, the heterogeneity and complexity are brought to the next level by the nature of commercial ovalbumin sample used in this study which may be extracted from hundreds of different hens. Therefore, genetic variations and possible amino acid substitutions combined with PTMs will only increase the complexity of the proteoforms. In this study, a total of 42 ovalbumin proteoforms were detected. Then, identification by mass-matching considering the possibility that different proteoforms may have the same mass, 86 proteoforms were detected (77 of them were identified). In this identification, only 4 types of most reported PTMs in ovalbumin named N-glycosylation, phosphorylation, N-terminal acetylation, and succinamate formation were included. One N-glycan site in amino acid asparagine position 293 contributes the most to the heterogeneity of ovalbumin proteoforms as its biosynthesis is template free-driven process which results in various forms of branching, glycan composition and other types of isomerism. In this study, 67 different N-glycan structures were identified from one N-glycosylation site of ovalbumin. Moreover, the possibility of amino acid substitution from cysteine to alanine in position 74 and 121 and substitution of arginine by threonine in position 340 have been previously reported (105,106). Therefore, the real possible number of ovalbumin proteoforms is probably far more than what has been detected and identified in this study. Nevertheless, SDBC performed in this study was able to enrich and noticeably separate them in early and later fractions.

Following the positive outcome from SDBC for the enrichment and purification of proteoforms in ovalbumin. SDBC was applied for the enrichment and purification of proteoforms in a monoclonal antibody (mAb) adalimumab.

From the obtained PPS results, CEX resin Eshmuno CPX gave better performance compared to POROS SCX resin in term of static binding capacity for adalimumab. This once again highlights the benefit of tentacle structure in Eshmuno CPX which allows multi-point interaction and reduces sterical hindrance between the functional group and the proteoforms

combined with rigid bead that makes proteoforms bind stronger (107,108). PPS results also shows that compared to ovalbumin, adalimumab suffers more from the effect of nonvolatile salt. For ovalbumin sample, the problem with nonvolatile salt was solved by simply buffer exchanging the NaCl solution to water. For adalimumab, the spectra quality was still highly effected and could not be rescued even after the NaCl solution was buffer exchanged to water for six times (**figure 31**). The spectra quality was finally improved by changing the NaCl to ammonium acetate.

The use of direct injection in this study also comes with limitation regarding the effect of nonvolatile salt. Although it is beneficial to avoid the loss of proteoforms in chromatography column, to keep the proteoforms from denatured stage, and to give faster analysis time, not having up-front column before mass spectrometer comes with a caveat of more noticeable salt effect. This limitation can be overcome by using nanospray and to increase the in-source CID in order to support the desolvation and declustering as has been suggested by Rosaty, S., et al., (109).

Furthermore, the results from SDBC of adalimumab in this study shows different patterns of deconvoluted spectra of SDBC early fractions and SDBC flow-through (**figure 34** and **35**). The patterns indicate changes in proteoforms composition across different fractions and purification of the main adalimumab proteoforms in the group 148 kDa from the low-abundant proteoforms in the group 151 kDa. However, the complexity of the spectra and the limitation of MS software for intact protein have caused the detection, identification, and relative quantification of each adalimumab proteoform in SDBC fraction still very much challenging particularly when aiming to address the low abundance proteoforms.

As the spectra turned to be very complicated, the deconvolution was distorted. Consequently, analysis could not be properly conducted. Major improvement and optimisation of the MS analytical software need to be addressed first. The data analysis in top-down approach until now is still in the phase of continues development. Some initial attempts have been trying to improve the deconvolution of intact proteins containing multiple charge states (88,110–115). There are several approaches to do deconvolution of MS spectra such as peak assignment

algorithm which extracts a list of peaks from the spectrum and assign a charge to each. Isotopic algorithms require high-resolution data with isotope-resolved spectra. The mass difference between peaks is known from the atomic masses of the isotopes, the charge can be inferred directly from the difference in  $m/z$ . There is another study which has used simulation algorithm in which multiple hypothetical mass and charge distributions are generated from which a spectrum is simulated. The simulated spectrum that fits the data best is then taken to be the most correct based on a metric of minimum chi-squared or maximum entropy (110–113). However, none of them has fully addressed the deconvolution of isotopically unresolved mass spectra data and the complicity of proteoforms in large molecules particularly regarding the low abundance proteoforms.

Moreover, to this date, there is little consensus to judge the quality of the deconvolution. A study by Marty, M. T. (111), has proposed a universal scoring function for ESI deconvolution to judge its quality based on the uniqueness and fit of the deconvolution to the data, peak shape consistency across different charge states, charge state distribution smoothness, and separation and symmetry of the peak. However, high-quality deconvolution does need high-quality intact MS spectra which is one thing that also still needs to be improved for adalimumab samples in this study.

Despite several challenges regarding the complexity of proteoform and limitation regarding the availability of well-established analytical system, the results obtained in this study show the utility and possibility of sample displacement batch chromatography for simple and effective separation of proteoforms. This in the future can be used for removal of harmful proteoforms and tool for enrichment of low abundant proteoforms for further analysis or to support and enhance molecular characterisation.

## Materials and Reagents

Chemicals and reagents used in this work are listed in **table 4**. The used of instruments are specified in the **table 5**.

**Table 4.** Materials, chemicals, and reagents used in this work and their corresponding distributor.

<b>Material/chemical/reagent</b>	<b>Distributor</b>
POROS SCX cation exchange resin	Thermo Fisher scientific, Bremen, Germany
POROS SAX anion exchange resin	Thermo Fisher scientific, Bremen, Germany
Eshmuno CPX cation exchange resin	Merck KGaA, Darmstadt, Germany
Fractogel EMD SO <sup>3-</sup> cation exchange resin	Merck KGaA, Darmstadt, Germany
Fractogel EMD TMAE anion exchange resin	Merck KGaA, Darmstadt, Germany
Ovalbumin from chicken egg white	Sigma-Aldrich, St. Louis, USA
Adalimumab	Institute of Bioprocess Science and Engineering, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria
Formic acid (FA) for MS	Fluka, Honeywell, Charlotte, USA
Acetic acid	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Sodium formate	Merck KGaA, Darmstadt, Germany
Sodium acetate	Merck KGaA, Darmstadt, Germany
2-(N-morpholino) ethanesulfonic acid (MES)	Sigma-Aldrich, St. Louis, USA
Disodium phosphate (Na <sub>2</sub> HPO <sub>4</sub> )	Sigma-Aldrich, St. Louis, USA
Sodium phosphate (NaH <sub>2</sub> PO <sub>4</sub> )	Sigma-Aldrich, St. Louis, USA
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma-Aldrich, St. Louis, USA
Ammonium acetate	Merck KGaA, Darmstadt, Germany
Trisma-base	Sigma-Aldrich, St. Louis, USA
Bis-Tris	Sigma-Aldrich, St. Louis, USA
Sodium chloride	Sigma-Aldrich, St. Louis, USA
Sulfolane	Sigma-Aldrich, St. Louis, USA
Water LC-MS grade	Merck KGaA, Darmstadt, Germany
Acetonitrile (ACN) LC-MS grade	Merck KGaA, Darmstadt, Germany



Ammonium bicarbonate (ABC)	Fluka, Honeywell, Charlotte, USA
Dithiothreitol (DTT)	Sigma-Aldrich, St. Louis, USA
Iodoacetamide (IAA)	Sigma-Aldrich, St. Louis, USA
Trypsin	Promega, Mannheim, Germany
PNGase F	Promega, Mannheim, Germany
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, St. Louis, USA

**Table 5.** List of instrumentation used

<b>Instrument</b>	<b>Distributor</b>
Sepax Protomix SCX NP-3, 4.6 x 50 mm, 3µm	Sepax Technologies, Inc., Newark, Germany
Acclaim PepMap® RSLC, 75µmx500µm, C18, 2µm, 100Å	Thermo Fisher scientific, Bremen, Germany
Acclaim PepMap® µ-precolumn, C18, 300 µm × 5 mm, 5 µm, 100 Å,	Thermo Fisher scientific, Bremen, Germany
Sep-Pak RP-SPE C18 cartridge	Waters, Manchester, UK
1.5 mL tube	Eppendorf, Hamburg, Germany
Agilent 1200 HPLC system	Agilent Technologies, Santa Clara, USA
Waters ACQUITY UPLC system	Waters, Manchester, UK
Dionex ultimate 3000 nano UPLC	Thermo Fisher scientific, Bremen, Germany
Elute UHPLC system	Bruker Daltonik GmbH, Bremen, Germany
Orbitrap QExactive	Thermo Fisher scientific, Bremen, Germany
MaXis II ETD qToF	Bruker Daltonik GmbH, Bremen, Germany
Orbitrap Fusion	Thermo Fisher scientific, Bremen, Germany

## Methods

### Cation exchange chromatography of ovalbumin

In this experiment the following buffers were used

Sample application buffer (buffer A): 20 mM formate buffer pH 3

Elution buffer (buffer B): 1M NaCl in buffer A

The cation exchange HPLC used a Sepac Protemix SCX NP-3. The chromatography was performed on Agilent 1200 HPLC system at 30°C with the mobile phase flow rate at 0.2 mL/min. The separation was obtained with gradient elution of 0-3 min (0% B); 3-8 min (0-100% B); 8-13 min (100% B). UV 280 nm was used as detector.

### Chromatography resin preparation

The binding capacity for each chromatography material (resin) was calculated based on the information provided by the manufacture. The resin then diluted to get binding capacity of 2µg/µL resin and equilibrated in respected buffer A. 50 µL resin was then used for every protein purification parameter screening (PPS) experiments and sample displacement batch chromatography (SDBC) to theoretically bind ±100 µg total protein.

### Protein purification parameter screening (PPS)

In the PPS experiments for ovalbumin the following buffers were used

Sample application buffer (buffer A) for PPS with cation exchange resin: 25 mM formate buffer (pH 3 and pH 4), 25 mM acetate buffer (pH 5), 25 mM MES buffer (pH 6), and 25 mM phosphate buffer (pH 7)

Sample application buffer (buffer A) for PPS with anion exchange resin: 25 mM bis-tris buffer (pH 6), 25 mM phosphate buffer (pH 7), and 25 mM tris buffer (pH 8)

Elution buffer (buffer B): 1M NaCl in respected buffer A

In the PPS experiments for adalimumab the following buffer were used

Sample application buffer (buffer A) for PPS with cation exchange resin: 25 mM MES buffer (pH 6), 25 mM phosphate buffer (pH 7), 25 mM HEPES buffer (pH 8), and 25 mM ammonium acetate buffer (pH 9).

Elution buffer (buffer B): 1M NaCl in respected buffer B

All different PPS conditions were listed in the **table 6** below:

**Table 6.** list of experiment conditions for protein purification parameter screening (PPS)

	Resin		pH	Concentration of buffer A (mM)	NaCl concentration in buffer B (mM)	Additional NaCl in buffer A (mM)
PPS of ovalbumin	Anion exchange	POROS SAX	6, 7, 8	25	1000	0
		Fractogel EMD TMAE	6, 7, 8	25	1000	0
	Cation exchange	POROS SCX	3, 4, 5, 6, 7	25	1000	0
		Eshmuno CPX	3, 4, 5, 6, 7	25	1000	0, 25, 50, 100, 200
		Fractogel EMD SO <sup>3-</sup>	3, 4, 5, 6, 7	25	1000	0
	PPS of adalimumab	Cation exchange	POROS SCX	6, 7, 8, 9	10, 15, 20, 25	150, 300, 500, 750, 1000
Eshmuno CPX			6, 7, 8, 9	10, 15, 20, 25	150, 300, 500, 750, 1000	0

One milligram sample was diluted in each respected buffer A for each condition. 1.5mL Eppendorf tubes were used as batch in which one tube represented one parameter condition. For each condition, 50µL of resin suspension was put in the tube. Sample was loaded into the tube, incubated, and continuously shaken for 30 minutes in rotational shaker, except for the experiment of the effect of incubation time, 0, 10-, 20-, 30-, 45-, and 60-minutes incubation time and for the experiment of the type of shaker rotational and horizontal shakers were used. After sedimentation of the resin, the supernatant was discarded from the tube. A washing step was then performed 3 times using the same buffer A and the supernatant was discarded from the tube. After that, to elute the proteoforms from the resin, 200µL of buffer

B was added into the tube, incubated, and shaken for 30 minutes. After sedimentation of the resin, the supernatant was collected. The total amount of protein was then measured using a BCA test.

### **Sample displacement batch chromatography (SDBC)**

For SDBC experiment of ovalbumin the following resin and buffers were used

Resin: cation exchange Eshmuno CPX

Sample application buffer (buffer A): 25 mM acetate buffer (pH 5)

Elution buffer (buffer B): 1M NaCl in buffer A

For SDBC experiment of adalimumab the following resin and buffers were used

Resin: cation exchange Eshmuno CPX

Sample application buffer (buffer A): 25 mM ammonium acetate buffer (pH 9)

Elution buffer (buffer B): 1M ammonium acetate

Ten Eppendorf tubes were used as ten segments. Eshmuno CPX resin suspension was put into ten 1.5 mL Eppendorf tubes in equal amounts (50 $\mu$ L of resin in each tube). One mg original sample was dissolved in 1 mL buffer A, loaded into the first tube, incubated, and continuously shaken for 30 min in rotational shaker. After sedimentation of the resin, the supernatant from the first batch was transferred to the second tube, incubated, and continuously shaken for 30 min in rotational shaker. The process was repeated until the last tube. A washing step was performed 3 times using the same buffer A and the supernatant was discarded from the tube. After that, to elute the proteoforms from the resin, 200  $\mu$ L of buffer B was added into the tube, incubated, and continuously shaken for 30 minutes. After sedimentation of the resin, the supernatant was collected from each segment. The total amount of protein was then measured using a BCA test.

### **Sample preparation for mass spectrometry**

All samples collected from ovalbumin and adalimumab PPS and SDBC experiments were buffer exchanged to 100% water using MWCO ultrafiltration for 6 times. For ovalbumin 5% sulfolane was added to each sample as a supercharging agent. All samples were then analysed by mass spectrometry.

### **Flow injection analysis – mass spectrometry (FIA-MS)**

#### *Ovalbumin*

FIA-MS was used for the analysis of all ovalbumin PPS and SDBC fractions. LC autosampler from Waters ACQUITY UPLC system (Waters) was used to directly inject the sample into the MS. Q-Exactive Hybrid-Quadrupole-Orbitrap™ mass spectrometry (Thermo Scientific) was used. 1.5µL of each sample was injected. Mobile phase of 100% water was used for 4 minutes run in a flow rate of 0.075mL/min for 1.5 minutes, then 0.1mL/min for 1.5 minutes and back to 0.075mL/min for 1 minute. MS parameters were set at scan range 1200 to 3600, resolution was set to 17500, AGC target was  $5 \times 10^6$ , micro scan count was 4, capillary temperature was set 275°C and in-source CID was 30eV.

#### *Adalimumab*

FIA-MS was used for the analysis of all adalimumab SDBC fractions. LC autosampler, Elute UHPLC (Bruker Daltonics Inc.) was used to directly inject the sample into the MS coupled to quadrupole-time of flight (qToF) mass spectrometry, MaXis II™ (Bruker Daltonic Inc.) as detector. Mobile phase of 100% water was used for 4 minutes run in a flow rate of 0.075mL/min. ESI in positive mode was used to acquire intact adalimumab only in full scan mode. The ion source parameters were 4500V capillary voltage, dry temperature 200°C, dry gas flow rate of 8.0mL/min and end plate offset of 500 V. The MS parameters were 2500 to 6100 m/z scan range, 150eV of in-source CID, ion energy 6eV, collision energy 12eV, pre pulse storage of 40 µs (transfer time was maintained at 200µs and the collision RF to 3800 Vpp).

## Data analysis

### *Ovalbumin*

The MS1 spectra was deconvoluted using UniDec GUI deconvoluted software (ver. 4.0.2 from Martys et. al.) with parameters were set at data processing m/z range 1500 to 3000, charge range from 10 to 35, sample mass every 0.5Da, peak detection range was 15Da, and peak detection threshold was set to 0.15. Results were exported and processed further in Excel spreadsheet.

### *Adalimumab*

The MS1 spectra was then deconvoluted using UniDec GUI deconvoluted software (ver. 4.0.2 from Martys et. al.) with parameters were set at data processing m/z range 2500 to 5000, charge range from 30 to 60, mass range of 147 to 152 kDa, sample mass every 1Da, peak detection range was 10Da, and peak detection threshold was 0.03. Results were exported and processed further in Excel spreadsheet.

## **N-glycan release, purification, and permethylation**

To obtain the free N-glycans, the extracted proteins were reduced with DTT and alkylated with IAA for then desalting and buffer exchange were performed using 3k centrifugal filters to an ABC solution. 1:30 (v/w) PNGase F was added and incubated at 37°C for 24 h to release N-glycans, followed by trypsin digestion for another 20 h. N-glycans and tryptic peptides were separated using a RP-SPE C18 cartridge. The RP cartridge was conditioned with 5 mL methanol and equilibrated with 10 mL 5% (v/v) acetic acid, respectively. Then each digested sample was loaded into the cartridge and N-glycans were eluted with 5 mL 5% (v/v) acetic acid. The N-glycan solvent was evaporated using a SpeedVac<sup>TM</sup> vacuum concentrator. Then the optimized solid phase permethylation was performed. Briefly, 10 µg/µL borane-ammonia were added to dried N-glycans and incubated at 60°C for 1 h. After evaporation, 110 µL water/DMSO (10/100, v/v) and 100 µL iodomethane was added and transferred into sodium hydroxide beads (200 mg) in glass vial by rotation for 10 min. The solution was transferred into a new

vial with one more addition of 150  $\mu$ L DMSO into the sodium hydroxide beads. After addition of 200  $\mu$ L 5% (v/v) acetic acid, permethylated N-glycans were purified using chloroform-water extraction and dried by a SpeedVac<sup>TM</sup> vacuum concentrator, followed by MS analysis.

### **Nano LC-MS/MS for permethylated N-glycans**

Permethylated N-glycans were first dissolved into 0.1% (v/v) FA, transferred to an autosampler, and injected into a Dionex Ultimate 3000 UPLC system. N-glycans were purified and desalted using an RP C<sub>18</sub> trapping column, Acclaim PepMap<sup>TM</sup>, 100  $\mu$ m $\times$ 2 cm, 5  $\mu$ m, 100Å at a flow rate of 3  $\mu$ L/min with 2% solvent B (0.1% (v/v) FA in ACN) and 98% solvent A (0.1% (v/v) FA) and transferred to an analytical RP C<sub>18</sub> column, Acclaim PepMap<sup>TM</sup> RSLC, 75  $\mu$ m $\times$ 50 cm, 2  $\mu$ m, 100Å, at a flow rate of 0.2  $\mu$ L/min, for chromatographic separation. For permethylated glycans, a 90 min gradient was used, starting with 10% solvent B. Solvent B increased to 30% in 5 min followed by a linear gradient elevating the concentration of 75% in 70 min and finally increased to 95% in 80 min. Eluted, derivated N-glycans were ionized using a nano spray ion source for electrospray ionization at a capillary voltage of 1.8 kV. Derivative N-glycan ions were transferred to a tribrid quadrupole-orbitrap-ion trap Orbitrap Fusion mass. For MS<sub>1</sub> scanning, an orbitrap mass analyzer was used with an orbitrap resolution of 120,000 FWHM at m/z 200; the maximum injection time was 120 ms, to an AGC target of  $2\times 10^5$ ; m/z scan range was from 450 to 2,000. Data dependent acquisition was used in the top speed mode. For CID-MS/MS, the most intense precursor ions were selected for fragmentation and isolated using an isolation window of 3; the normalized collision energy of CID was set to 35%; fragment ions were injected to an ion trap with maximum injection time was 20 ms at an AGC target of  $1\times 10^5$ . The data were visualized and analysed using Xcalibur software.

### **Identification for differential N-glycomics**

MaxQuant version 1.6.2.3 (<http://www.maxquant.org>) was used to extract masses of all tentative glycan precursors in an "allPeptides.txt" file from the MS raw data. In this

“allPeptides.txt” file, all the masses from the “Mass” column are extracted as “Mass.csv” and matched to the possible monosaccharide compositions using designed R-scripts (<https://github.com/guan181992/Glyco-informatics>), outputting the result file named as “Monosaccharide composition.csv”. Three tissue groups derived N-glycan MS raw data were processed by mouse derived R-script. After matching to monoisotopic m/z, bundled sequencing algorithm is used to illustrate the N-glycan structure with the MS2 data using GlycoWorkbench in the version 2.1 stable build 146.



## References

1. Schlüter H, Apweiler R, Holzhütter HG, Jungblut PR. Finding one's way in proteomics: A protein species nomenclature. *Chem Cent J*. 2009;3(1):1–10.
2. Jungblut PR, Holzhütter HG, Apweiler R, Schlüter H. The speciation of the proteome. *Chem Cent J*. 2008;2(1):1–10.
3. Smith LM, Kelleher NL. Proteoform: A single term describing protein complexity. *Nat Methods*. 2013;10(3):186–7.
4. Mathur M, Kim CM, Munro SA, Rudina SS, Sawyer EM, Smolke CD. Programmable mutually exclusive alternative splicing for generating RNA and protein diversity. *Nat Commun* [Internet]. 2019;10(1). Available from: <http://dx.doi.org/10.1038/s41467-019-10403-w>
5. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, et al. How many human proteoforms are there? *Nat Chem Biol* [Internet]. 2018;14(3):206–14. Available from: <http://dx.doi.org/10.1038/nchembio.2576>
6. Khoury GA, Baliban RC, Floudas CA. Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Sci Rep*. 2011;1:1–5.
7. Ramazi S, Zahiri J. Post-translational modifications in proteins: resources, tools and prediction methods. *Database*. 2021;2021(7):1–20.
8. Virág D, Dalmadi-Kiss B, Vékey K, Drahos L, Klebovich I, Antal I, et al. Current Trends in the Analysis of Post-translational Modifications. *Chromatographia* [Internet]. 2020;83(1):1–10. Available from: <https://doi.org/10.1007/s10337-019-03796-9>
9. Kumar A, Narayanan V, Sekhar A. Characterizing Post-Translational Modifications and Their Effects on Protein Conformation Using NMR Spectroscopy. *Biochemistry*. 2020;59(1):57–73.
10. Lee HHL, Heo CE, Seo N, Yun SG, An HJ, Kim HI. Accurate Quantification of N-Glycolylneuraminic Acid in Therapeutic Proteins Using Supramolecular Mass Spectrometry. *J Am Chem Soc*. 2018;140(48):16528–34.
11. Muthana SM, Campbell CT, Gildersleeve JC. Modifications of glycans: Biological significance and therapeutic opportunities. *ACS Chem Biol*. 2012;7(1):31–43.
12. Ducry L. Challenges in the Development and Manufacturing of ADC. *Ther Proteins Methods Protoc Methods Mol Biol* [Internet]. 2012;899:489–97. Available from: <http://link.springer.com/10.1007/978-1-61779-921-1>
13. Rege NK, Wickramasinghe NP, Tustan AN, Phillips NFB, Yee VC, Ismail-Beigi F, et al. Structure-based stabilization of insulin as a therapeutic protein assembly via enhanced aromatic–aromatic interactions. *J Biol Chem*. 2018;293(28):10895–910.
14. Hassiki R, Behi J, Said N Ben, Boujbel L, Bouhaouala-Zahar B. Biochemical and Biological Quality Control of Two Recombinant Human Erythropoietin Biosimilar Products. *Curr Pharm Anal*. 2017;13(6):566–73.
15. Discovery D, Haselberg R, Jong GJ De, Somsen GW. Sensitive Glycoform Profiling of b - Interferon-1a ( Avonex ) and recombinant human Erythropoietin by CESI-TOF-MS. 2017;(August 2016):1–6.
16. Akash MSH, Rehman K, Tariq M, Chen S. Development of therapeutic proteins: Advances and challenges. *Turkish J Biol*. 2015;39(3):343–58.
17. Beck A, Liu H. Macro- and Micro-Heterogeneity of Natural and Recombinant IgG Antibodies. *Antibodies*. 2019;8(1):18.
18. Goh JB, Ng SK. Impact of host cell line choice on glycan profile. *Crit Rev Biotechnol* [Internet]. 2018;38(6):851–67. Available from: <https://doi.org/10.1080/07388551.2017.1416577>
19. Gervais D. Protein deamidation in biopharmaceutical manufacture: Understanding, control and impact. *J Chem Technol Biotechnol*. 2016;91(3):569–75.
20. Moss CX, Matthews SP, Lamont DJ, Watts C. Asparagine deamidation perturbs antigen presentation on class II major histocompatibility complex molecules. *J Biol Chem* [Internet]. 2005;280(18):18498–503. Available from: <http://dx.doi.org/10.1074/jbc.M501241200>

21. Liu YD, van Enk JZ, Flynn GC. Human antibody Fc deamidation in vivo. *Biologicals* [Internet]. 2009;37(5):313–22. Available from: <http://dx.doi.org/10.1016/j.biologicals.2009.06.001>
22. Hermeling S, Crommelin DJA, Schellekens H, Jiskoot W. Structure-immunogenicity relationships of therapeutic proteins. *Pharm Res*. 2004;21(6):897–903.
23. Delorme E, Lorenzini T, Giffin J, Martin F, Jacobsen F, Boone T, et al. Role of Glycosylation on the Secretion and Biological Activity of Erythropoietin. *Biochemistry*. 1992;31(41):9871–6.
24. Erbayraktar S, Grasso G, Sfacteria A, Xie Q wen, Coleman T, Kreilgaard M, et al. Asialoerythropoietin is a nonerythropoietic cytokine with broad neuroprotective activity in vivo. *Proc Natl Acad Sci U S A*. 2003;100(11):6741–6.
25. Satzer P, Jungbauer A. High-capacity protein A affinity chromatography for the fast quantification of antibodies: Two-wavelength detection expands linear range. *J Sep Sci*. 2018;41(8):1791–7.
26. Liu H, Gaza-Bulsecu G, Chumsae C. Analysis of Reduced Monoclonal Antibodies Using Size Exclusion Chromatography Coupled with Mass Spectrometry. *J Am Soc Mass Spectrom*. 2009;20(12):2258–64.
27. Aguilar MI. Reversed-phase high-performance liquid chromatography. *Methods Mol Biol* [Internet]. 2004 [cited 2021 May 21];251:9–22. Available from: <https://link.springer.com/protocol/10.1385/1-59259-742-4:9>
28. Buszewski B, Noga S. Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique [Internet]. Vol. 402, *Analytical and Bioanalytical Chemistry*. Springer; 2012 [cited 2021 May 21]. p. 231–47. Available from: </pmc/articles/PMC3249561/>
29. Muneeruddin K, Bobst CE, Frenkel R, Houde D, Turyan I, Sosic Z, et al. Characterization of a PEGylated protein therapeutic by ion exchange chromatography with on-line detection by native ESI MS and MS/MS. *Analyst*. 2017;142(2):336–44.
30. Muneeruddin K, Nazzaro M, Kaltashov IA. Characterization of Intact Protein Conjugates and Biopharmaceuticals Using Ion-Exchange Chromatography with Online Detection by Native Electrospray Ionization Mass Spectrometry and Top-Down Tandem Mass Spectrometry. *Anal Chem*. 2015;87(19):10138–45.
31. Lee YF, Jöhnck M, Frech C. Evaluation of differences between dual salt-pH gradient elution and mono gradient elution using a thermodynamic model: Simultaneous separation of six monoclonal antibody charge and size variants on preparative-scale ion exchange chromatographic resin. *Biotechnol Prog*. 2018 Jul 1;34(4):973–86.
32. Zhu K, Zhao J, Lubman DM, Miller FR, Barder TJ. Protein pI shifts due to posttranslational modifications in the separation and characterization of proteins. *Anal Chem*. 2005;77(9):2745–55.
33. Healthcare G. *Ion Exchange Chromatography & Chromatofocusing: Principles and Methods*. GE Heal Handbooks. 2016;170.
34. Leblanc Y, Ramon C, Bihoreau N, Chevreux G. Charge variants characterization of a monoclonal antibody by ion exchange chromatography coupled on-line to native mass spectrometry: Case study after a long-term storage at +5 °C. *J Chromatogr B Anal Technol Biomed Life Sci* [Internet]. 2017;1048:130–9. Available from: <http://dx.doi.org/10.1016/j.jchromb.2017.02.017>
35. Lingg N, Berndtsson M, Hintersteiner B, Schuster M, Bardor M, Jungbauer A. Highly linear pH gradients for analyzing monoclonal antibody charge heterogeneity in the alkaline range: Validation of the method parameters. *J Chromatogr A* [Internet]. 2014;1373:124–30. Available from: <http://dx.doi.org/10.1016/j.chroma.2014.11.021>
36. Bailey AO, Han G, Phung W, Gazis P, Sutton J, Josephs JL, et al. Charge variant native mass spectrometry benefits mass precision and dynamic range of monoclonal antibody intact mass analysis. *MAbs* [Internet]. 2018;10(8):1214–25. Available from: <https://doi.org/10.1080/19420862.2018.1521131>
37. Bailey AO, Han G, Phung W, Gazis P, Sutton J, Josephs JL, et al. Charge variant native mass

- spectrometry benefits mass precision and dynamic range of monoclonal antibody intact mass analysis. *MAbs*. 2018 Nov 17;10(8):1214–25.
38. Alekseychik L, Su C, Becker GW, Treuheit MJ, Razinkov VI. High-Throughput Screening and Analysis of Charge Variants of Monoclonal Antibodies in Multiple Formulations. *SLAS Discov*. 2017;22(8):1044–52.
  39. Zhang T, Bourret J, Cano T. Isolation and characterization of therapeutic antibody charge variants using cation exchange displacement chromatography. *J Chromatogr A* [Internet]. 2011;1218(31):5079–86. Available from: <http://dx.doi.org/10.1016/j.chroma.2011.05.061>
  40. Khanal O, Kumar V, Westerberg K, Schlegel F, Lenhoff AM. Multi-column displacement chromatography for separation of charge variants of monoclonal antibodies. *J Chromatogr A* [Internet]. 2019;1586:40–51. Available from: <https://doi.org/10.1016/j.chroma.2018.11.074>
  41. Ralph Himmelhoch S. Chromatography of Proteins on Types of Adsorbent Available-Matrix. *Anal Biochem* [Internet]. 1966;17:273–86. Available from: <http://linkinghub.elsevier.com/retrieve/pii/0076687971220280>
  42. Tiselius. Displacement development in adsorption analysis. *Ark Kemi Miner Geol*. 1943;16A:1–18.
  43. Horváth C, Nahum A, Frenz JH. High-performance displacement chromatography. *J Chromatogr A*. 1981;218(C):365–93.
  44. Trusch M, Tillack K, Kwiatkowski M, Bertsch A, Ahrends R, Kohlbacher O, et al. Displacement chromatography as first separating step in online two-dimensional liquid chromatography coupled to mass spectrometry analysis of a complex protein sample-The proteome of neutrophils. *J Chromatogr A* [Internet]. 2012;1232:288–94. Available from: <http://dx.doi.org/10.1016/j.chroma.2012.02.029>
  45. Kwiatkowski M, Krösser D, Wurlitzer M, Steffen P, Barcaru A, Krisp C, et al. Application of Displacement Chromatography to Online Two-Dimensional Liquid Chromatography Coupled to Tandem Mass Spectrometry Improves Peptide Separation Efficiency and Detectability for the Analysis of Complex Proteomes. *Anal Chem*. 2018;90(16):9951–8.
  46. Newburger J, Guiochon G. Utility of the displacement effect in the routine optimization of separations by preparative liquid chromatography. *J Chromatogr A*. 1990;523(C):63–80.
  47. Ren J, Yao P, Chen J, Jia L. Salt-independent hydrophobic displacement chromatography for antibody purification using cyclodextrin as supermolecular displacer. *J Chromatogr A* [Internet]. 2014;1369:98–104. Available from: <http://dx.doi.org/10.1016/j.chroma.2014.10.009>
  48. Pinto NDS, Frey DD. Displacement chromatography of proteins using a retained pH front in a hydrophobic charge induction chromatography column. *J Chromatogr A* [Internet]. 2015;1387:53–9. Available from: <http://dx.doi.org/10.1016/j.chroma.2015.01.087>
  49. Heikaus L, Schlüter H. Sample displacement chromatography for protein purification. *Am Pharm Rev*. 2014;17(2).
  50. Srajer Gajdosik M, Clifton J, Josic D. Sample displacement chromatography as a method for purification of proteins and peptides from complex mixtures. *J Chromatogr A* [Internet]. 2012;1239:1–9. Available from: <http://dx.doi.org/10.1016/j.chroma.2012.03.046>
  51. McAtee CP. Displacement Chromatography of Proteins. *Curr Protoc Protein Sci*. 2010;(SUPPL. 59):1–14.
  52. Badalà F, Nouri-mahdavi K, Raoof DA. Sample displacement chromatography as a method for purification of proteins and peptides from complex mixtures. *J Chromatogr*. 2008;144(5):724–32.
  53. Kotasińska M, Richter V, Thiemann J, Schlüter H. Cation exchange displacement batch chromatography of proteins guided by screening of protein purification parameters. *J Sep Sci*. 2012;35(22):3170–6.
  54. Heikaus L, Hidayah S, Gaikwad M, Kotasinska M, Richter V, Kwiatkowski M, et al. Sample displacement batch chromatography of proteins. *Methods Mol Biol*. 2021;2178:285–99.

55. Baldrey SF, Brodie RR, Morris GR, Jenkins EH, Brookes ST. Comparison of LC-UV and LC-MS-MS for the determination of taxol. *Chromatographia*. 2002;55(SUPPL.).
56. Aebersold R, Mann M. Mass spectrometry-based proteomics: Abstract: Nature. *Nature* [Internet]. 2003;422(6928):198–207. Available from: <https://www.nature.com/articles/nature01511><http://www.nature.com/articles/nature01511><http://www.nature.com/nature/journal/v422/n6928/abs/nature01511.html>
57. Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. Top down versus bottom up protein characterization by tandem high- resolution mass spectrometry. *J Am Chem Soc*. 1999;121(4):806–12.
58. Gundry RL, White MY, Murray CI, Kane LA, Fu Q, Stanley BA, et al. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr Protoc Mol Biol*. 2009;(SUPPL. 88).
59. Schaffer L V., Shortreed MR, Cesnik AJ, Frey BL, Solntsev SK, Scalf M, et al. Expanding Proteoform Identifications in Top-Down Proteomic Analyses by Constructing Proteoform Families. *Anal Chem*. 2018;90(2):1325–33.
60. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. 2011;480(7376):254–8.
61. Kou Q, Wu S, Tolić N, Paša-Tolić L, Liu Y, Liu X. A mass graph-based approach for the identification of modified proteoforms using top-down tandem mass spectra. *Bioinformatics*. 2017 May 1;33(9):1309–16.
62. Catherman AD, Skinner OS, Kelleher NL. Top Down proteomics: Facts and perspectives. *Biochem Biophys Res Commun* [Internet]. 2014;445(4):683–93. Available from: <http://dx.doi.org/10.1016/j.bbrc.2014.02.041>
63. Shen X, Yang Z, McCool EN, Lubeckyj RA, Chen D, Sun L. Capillary zone electrophoresis-mass spectrometry for top-down proteomics. *TrAC - Trends Anal Chem* [Internet]. 2019;120:115644. Available from: <https://doi.org/10.1016/j.trac.2019.115644>
64. King ACF, Giorio C, Wolff E, Thomas E, Roverso M, Schwikowski M, et al. Direct Injection Liquid Chromatography High-Resolution Mass Spectrometry for Determination of Primary and Secondary Terrestrial and Marine Biomarkers in Ice Cores. *Anal Chem*. 2019;91(8):5051–7.
65. Scheffler K. Top-down proteomics by means of orbitrap mass spectrometry. *Methods Mol Biol*. 2014;1156:465–87.
66. Rappsilber J, Moniatte M, Nielsen ML, Podtelejnikov A V., Mann M. Experiences and perspectives of MALDI MS and MS/MS in proteomic research. *Int J Mass Spectrom*. 2003;226(1):223–37.
67. Richards G. Physical chemistry. *Nature*. 1978;272(5648):101–2.
68. Rebane R, Kruve A, Liigand P, Liigand J, Herodes K, Leito I. Establishing Atmospheric Pressure Chemical Ionization Efficiency Scale. *Anal Chem*. 2016;88(7):3435–9.
69. Cohen SL, Chait BT. Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal Chem*. 1996;68(1):31–7.
70. Lotz W. An empirical formula for the electron-impact ionization cross-section. *Zeitschrift für Phys*. 1967;206(2):205–11.
71. Nadler WM, Waidelich D, Kerner A, Hanke S, Berg R, Trumpp A, et al. MALDI versus ESI: The Impact of the Ion Source on Peptide Identification. *J Proteome Res*. 2017;16(3):1207–15.
72. Shin E, Cha S. Comparison of MALDI and ESI for relative quantification of citrullination via skewed isotopic distribution patterns. *Bull Korean Chem Soc*. 2015;36(5):1373–7.
73. Konermann L, Ahadi E, Rodriguez AD, Vahidi S. Unraveling the mechanism of electrospray ionization. *Anal Chem*. 2013 Jan 2;85(1):2–9.
74. Heikaus L, Hidayah S, Gaikwad M, Kotasinska M, Richter V, Kwiatkowski M, et al. Sample displacement batch chromatography of proteins. In: *Methods in Molecular Biology* [Internet]. Humana Press Inc.; 2021 [cited 2021 Apr 12]. p. 285–99. Available from:

- <https://pubmed.ncbi.nlm.nih.gov/33128756/>
75. Kotasinska M, Richter V, Kwiatkowski M, Schlüter H. Sample displacement batch chromatography of proteins. *Methods Mol Biol.* 2014;1129:325–38.
  76. da Silva M, Beauclercq S, Harichaux G, Labas V, Guyot N, Gautron J, et al. The family secrets of avian egg-specific ovalbumin and its related proteins Y and X. *Biol Reprod.* 2015;93(3):1–7.
  77. Thompson EOP, Fisher WK. A Correction and Extension of the Acetylated Amino Terminal Sequence of Ovalbumin. *Aust J Biol Sci.* 1978;31(5):443–6.
  78. Yang Y, Barendregt A, Kamerling JP, Heck AJR. Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. *Anal Chem.* 2013;85(24):12037–45.
  79. Füssl F, Criscuolo A, Cook K, Scheffler K, Bones J. Cracking Proteoform Complexity of Ovalbumin with Anion-Exchange Chromatography-High-Resolution Mass Spectrometry under Native Conditions. *J Proteome Res.* 2019;18(10):3689–702.
  80. Takahashi N, Tatsumi E, Orita T, Hirose M. Effects of hrole of the intrachain disulfide bond of ovalbumin during conversion into S-ovalbumin. *Biosci Biotechnol Biochem.* 1996;60(9):1464–8.
  81. Barrabés S, Sarrats A, Fort E, De Llorens R, Rudd PM, Peracaula R. Effect of sialic acid content on glycoprotein pl analyzed by two-dimensional electrophoresis. *Electrophoresis.* 2010;31(17):2903–12.
  82. Gaikwad M. Fast quantification of intact therapeutic protein species. Doctoral Thesis. University of Hamburg; 2021.
  83. Donnelly DP, Rawlins CM, DeHart CJ, Fornelli L, Schachner LF, Lin Z, et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods* [Internet]. 2019;16(7):587–94. Available from: <http://dx.doi.org/10.1038/s41592-019-0457-0>
  84. Metwally H, McAllister RG, Konermann L. Exploring the mechanism of salt-induced signal suppression in protein electrospray mass spectrometry using experiments and molecular dynamics simulations. *Anal Chem.* 2015;87(4):2434–42.
  85. Cassou CA, Williams ER. Desalting protein ions in native mass spectrometry using supercharging reagents. *Analyst* [Internet]. 2014;139(19):4810–9. Available from: <http://dx.doi.org/10.1039/C4AN01085J>
  86. Iavarone AT, Williams ER. Supercharging in electrospray ionization: Effects on signal and charge. *Int J Mass Spectrom.* 2002;219(1):63–72.
  87. Metwally H, McAllister RG, Popa V, Konermann L. Mechanism of Protein Supercharging by Sulfolane and m-Nitrobenzyl Alcohol: Molecular Dynamics Simulations of the Electrospray Process. *Anal Chem.* 2016;88(10):5345–54.
  88. Marty MT, Baldwin AJ, Marklund EG, Hochberg GKA, Benesch JLP, Robinson C V. Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal Chem.* 2015;87(8):4370–6.
  89. Thompson EOP, Fisher WK. A Correction and Extension of the Acetylated Amino Terminal Sequence of Ovalbumin. Vol. 31, *J. Bioi. Sci.* 1978.
  90. Tracey D, Klareskog L, Sasso EH, Salfeld JG, Tak PP. Tumor necrosis factor antagonist mechanisms of action: A comprehensive review. *Pharmacol Ther.* 2008;117(2):244–79.
  91. Lee N, Lee JAJ, Yang H, Baek S, Kim S, Kim S, et al. Evaluation of similar quality attribute characteristics in SB5 and reference product of adalimumab. *MAbs* [Internet]. 2019;11(1):129–44. Available from: <https://doi.org/10.1080/19420862.2018.1530920>
  92. Immunoglobulin Structure and Classes | Thermo Fisher Scientific - DE [Internet]. [cited 2021 Apr 22]. Available from: <https://www.thermofisher.com/de/de/home/life-science/antibodies/antibodies-learning-center/antibodies-resource-library/antibody-methods/immunoglobulin-structure-classes.html>
  93. Füssl F, Trappe A, Cook K, Scheffler K, Fitzgerald O, Bones J. Comprehensive characterisation of the heterogeneity of adalimumab via charge variant analysis hyphenated on-line to native

- high resolution Orbitrap mass spectrometry. *MAbs* [Internet]. 2019;11(1):116–28. Available from: <https://doi.org/10.1080/19420862.2018.1531664>
94. Leduc RD, Schwämmle V, Shortreed MR, Cesnik AJ, Solntsev SK, Shaw JB, et al. ProForma: A Standard Proteoform Notation. *J Proteome Res*. 2018;17(3):1321–5.
  95. Ahrends R, Lichtner B, Buck F, Hildebrand D, Kotasinska M, Kohlbacher O, et al. Comparison of displacement versus gradient mode for separation of a complex protein mixture by anion-exchange chromatography. *J Chromatogr B Anal Technol Biomed Life Sci* [Internet]. 2012;901:34–40. Available from: <http://dx.doi.org/10.1016/j.jchromb.2012.05.037>
  96. Khanal O, Kumar V, Westerberg K, Schlegel F, Lenhoff AM. Multi-column displacement chromatography for separation of charge variants of monoclonal antibodies. *J Chromatogr A*. 2019 Feb 8;1586:40–51.
  97. Kotasinska M, Richter V, Kwiatkowski M, Schlüter H. Sample displacement batch chromatography of proteins. *Methods Mol Biol*. 2014;1129:325–38.
  98. Fractogel® EMD CEX + Eshmuno® CPX media: Performance comparison in mAbs purification. Merck; 2020.
  99. Eshmuno® CPX Chromatography Resin | Biopharmaceutical Manufacturing | Merck [Internet]. [cited 2021 Apr 26]. Available from: <https://www.merckmillipore.com/DE/de/products/biopharmaceutical-manufacturing/downstream-processing/chromatography/ion-exchange-chromatography/eshmuno-resin/eshmuno-cpx-resin/ylyb.qB.kYkAAAFAWgBkiQpy,nav?ReferrerURL=https%3A%2F%2Fwww.google.com%2F>
  100. Zhang K, Liu X. Mixed-mode chromatography in pharmaceutical and biopharmaceutical applications. *J Pharm Biomed Anal*. 2016 Sep 5;128:73–88.
  101. Arakawa T, Ponce S, Young G. Isoform separation of proteins by mixed-mode chromatography. *Protein Expr Purif*. 2015 Jul 28;116:144–51.
  102. Gilar M, Yu YQ, Ahn J, Fournier J, Gebler JC. Mixed-mode chromatography for fractionation of peptides, phosphopeptides, and sialylated glycopeptides. *J Chromatogr A*. 2008 May 16;1191(1–2):162–70.
  103. Altomare A, Fasoli E, Colzani M, Paredes Parra XM, Ferrari M, Cilurzo F, et al. An in depth proteomic analysis based on ProteoMiner, affinity chromatography and nano-HPLC-MS/MS to explain the potential health benefits of bovine colostrum. *J Pharm Biomed Anal* [Internet]. 2016;121:297–306. Available from: <http://dx.doi.org/10.1016/j.jpba.2016.01.013>
  104. Hartwig S, Czibere A, Kotzka J, Paßlack W, Haas R, Eckel J, et al. Combinatorial hexapeptide ligand libraries (ProteoMiner™): An innovative fractionation tool for differential quantitative clinical proteomics. *Arch Physiol Biochem*. 2009;115(3):155–60.
  105. Ishimaru T, Ito K, Tanaka M, Tanaka S, Matsudomi N. The role of the disulfide bridge in the stability and structural integrity of ovalbumin evaluated by site-directed mutagenesis. *Biosci Biotechnol Biochem* [Internet]. 2011 [cited 2021 Apr 12];75(3):544–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/21389617/>
  106. Arii Y, Hirose M. Probing the serpin structural-transition mechanism in ovalbumin mutant r339t by proteolytic-cleavage kinetics of the reactive-centre loop. *Biochem J* [Internet]. 2002 Apr 15 [cited 2021 Apr 12];363(2):403–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/11931671/>
  107. Liu HF, Ma J, Winter C, Bayer R. Recovery and purification process development for monoclonal antibody production [Internet]. Vol. 2, mAbs. *MAbs*; 2010 [cited 2021 Apr 12]. p. 480–99. Available from: <https://pubmed.ncbi.nlm.nih.gov/20647768/>
  108. Connell-Crowley L, Nguyen T, Bach J, Chinniah S, Bashiri H, Gillespie R, et al. Cation exchange chromatography provides effective retrovirus clearance for antibody purification processes. *Biotechnol Bioeng* [Internet]. 2012 Jan [cited 2021 Apr 12];109(1):157–65. Available from: <https://pubmed.ncbi.nlm.nih.gov/21837666/>
  109. Rosati S, Van Den Bremer ET, Schuurman J, Parren PW, Kamerling JP, Heck AJ. In-depth

- qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *MAbs*. 2013;5(6):917–24.
110. Kou Q, Wu S, Liu X. A new scoring function for top-down spectral deconvolution [Internet]. Vol. 15, *BMC Genomics*. 2014. Available from: <http://www.biomedcentral.com/1471-2164/15/1140>
111. Marty MT. A Universal Score for Deconvolution of Intact Protein and Native Electrospray Mass Spectra. *Anal Chem*. 2020;92(6):4395–401.
112. Reid DJ, Diesing JM, Miller MA, Perry SM, Wales JA, Montfort WR, et al. MetaUniDec: High-Throughput Deconvolution of Native Mass Spectra. *J Am Soc Mass Spectrom*. 2019;30(1):118–27.
113. Millán-Martín S, Carillo S, Füssl F, Sutton J, Gazis P, Cook K, et al. Optimisation of the use of sliding window deconvolution for comprehensive characterisation of trastuzumab and adalimumab charge variants by native high resolution mass spectrometry. Vol. 158, *European Journal of Pharmaceutics and Biopharmaceutics*. 2021. p. 83–95.
114. Jeong K, Kim J, Gaikwad M, Hidayah SN, Heikaus L, Schlüter H, et al. FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-Down Proteomics. *Cell Syst*. 2020;10(2):213–218.e6.
115. Basharat AR, Ning X, Liu X. EnvCNN: A Convolutional Neural Network Model for Evaluating Isotopic Envelopes in Top-Down Mass-Spectral Deconvolution. *Anal Chem*. 2020 Jun 2;92(11):7778–85.

## Supplements

### Supplement 1. List of all detected proteoforms in ovalbumin

No.	Proteoforms mass (Da)	Original sample	PPS optimal condition	SDBC fractions									
				1	2	3	4	5	6	7	8	9	
1	33639		v	v									
2	39592.5		v	v	v	v	v	v					
3	39630		v	v									
4	39673.5		v	v	v								
5	39710.5		v	v									
6	39753	v	v	v	v	v	v	v	v				
7	39790.5		v	v	v	v	v						
8	39832	v	v	v	v	v	v	v	v	v	v	v	v
9	39871.5		v	v	v								
10	39914.5		v	v	v	v	v	v					
11	39953.5		v	v	v	v	v						
12	39995	v	v	v	v	v	v	v	v	v	v	v	v
13	40036		v	v	v	v	v	v					
14	40076		v	v	v	v	v						
15	40158		v	v	v	v	v						
16	40199.5		v	v	v	v	v						
17	40237.5	v	v	v	v	v	v	v	v	v	v	v	v
18	40278.5		v	v	v	v	v	v					
19	40361		v	v	v	v							
20	40441		v	v	v	v	v						
21	43637.5		v	v	v								
22	43713					v							



23	43877.5		v	v	v	v							
24	44069.5		v	v			v	v					
25	44086	v	v	v	v	v	v	v	v	v	v		
26	44123		v	v	v	v	v						
27	44167	v	v	v	v	v	v	v	v	v	v	v	v
28	44207.5	v	v	v	v		v	v	v	v	v	v	v
29	44231		v	v	v	v	v						
30	44248	v	v	v	v	v	v	v	v	v	v		
31	44288.5	v	v	v	v	v	v	v					
32	44328	v	v	v	v	v	v	v	v	v	v	v	v
33	44369.5	v	v	v	v	v	v	v	v	v	v	v	v
34	44394					v	v						
35	44411	v	v	v	v	v	v	v	v	v	v	v	v
36	44475.5		v	v	v	v	v						
37	44493	v	v	v	v	v	v	v	v	v	v	v	v
38	44531				v	v	v	v					
39	44573	v	v	v	v	v	v	v	v	v	v	v	v
40	44613	v			v	v	v	v	v	v	v	v	v
41	44696.5						v						
42	44775.5	v					v	v	v	v	v	v	v

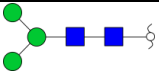
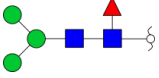
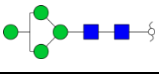
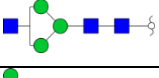
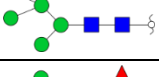
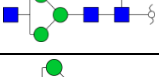
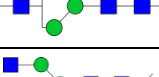
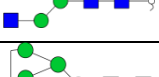

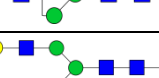

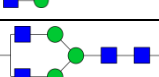

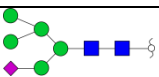
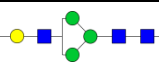
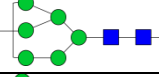
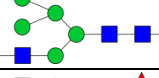

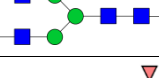
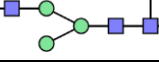



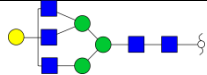
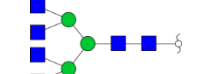
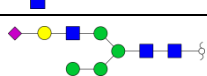
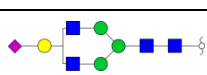
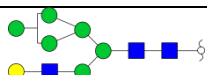
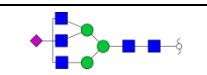
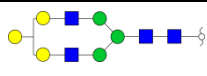
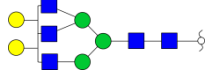
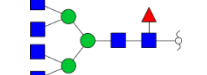



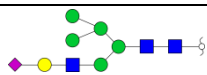
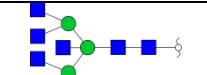
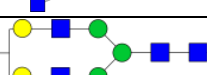


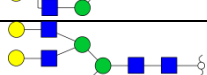




61	HexNAC <sub>1</sub> Hex <sub>2</sub>	v	v	v								
62	HexNAC <sub>2</sub> Fuc <sub>1</sub>		v	v								
63	HexNAC <sub>2</sub> Hex <sub>1</sub>	v	v	v	v	v	v	v		v	v	v
64	HexNAC <sub>1</sub> Hex <sub>3</sub>	v	v	v	v			v		v	v	v
65	HexNAC <sub>2</sub> Hex <sub>1</sub> Fuc <sub>1</sub>		v	v								
66	HexNAC <sub>2</sub> Hex <sub>2</sub>	v	v	v	v	v	v			v	v	v
67	HexNAC <sub>2</sub> Hex <sub>2</sub> Fuc <sub>1</sub>	v										

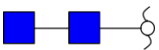

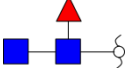
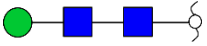
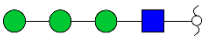
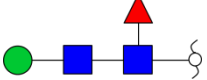

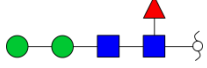
N-glycan compositions: Hex (hexose, galactose/mannose), HexNAC (N-acetylhexosamine), Neu5Ac (N-acetylneuraminic acid), and Fuc (fucose). Preferred N-glycan structure for each composition can be found in the **supplement 3**. Details of N-glycan released by PNGase can be found in the method section.

## Supplement 3. Preferred N-glycan structure of ovalbumin

No.	N-Glycan compositions	Preferred N-glycan structures	Permethylated monoisotopic Mass (Da)
1	HexNAc <sub>2</sub> Hex <sub>3</sub>		1164.6251
2	HexNAc <sub>2</sub> Hex <sub>3</sub> Fuc <sub>1</sub>		1338.7166
3	HexNAc <sub>2</sub> Hex <sub>4</sub>		1368.7249
4	HexNAc <sub>3</sub> Hex <sub>3</sub>		1409.7515
5	HexNAc <sub>2</sub> Hex <sub>5</sub>		1572.8247
6	HexNAc <sub>3</sub> Hex <sub>3</sub> Fuc <sub>1</sub>		1583.8407
7	HexNAc <sub>3</sub> Hex <sub>4</sub>		1613.8512
8	HexNAc <sub>4</sub> Hex <sub>3</sub>		1654.8778
9	HexNAc <sub>2</sub> Hex <sub>6</sub>		1776.9245
10	HexNAc <sub>3</sub> Hex <sub>4</sub> Fuc <sub>1</sub>		1787.9438
11	HexNAc <sub>3</sub> Hex <sub>5</sub>		1817.9510
12	HexNAc <sub>4</sub> Hex <sub>3</sub> Fuc <sub>1</sub>		1828.9699
13	HexNAc <sub>4</sub> Hex <sub>4</sub>		1858.9776
14	HexNAc <sub>5</sub> Hex <sub>3</sub>		1900.0041
15	Neu5Ac <sub>1</sub> HexNAc <sub>2</sub> Hex <sub>5</sub>		1933.9983
16	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>4</sub>		1975.0249
17	HexNAc <sub>2</sub> Hex <sub>7</sub>		1981.0242
18	HexNAc <sub>3</sub> Hex <sub>6</sub>		2022.0508
19	HexNAc <sub>4</sub> Hex <sub>4</sub> Fuc <sub>1</sub>		2033.0639
20	HexNAc <sub>4</sub> Hex <sub>5</sub>		2063.0773
21	HexNAc <sub>3</sub> Hex <sub>4</sub> Fuc <sub>1</sub>		2074.0946

22	HexNAc <sub>5</sub> Hex <sub>4</sub>		2104.1039
23	HexNAc <sub>6</sub> Hex <sub>3</sub>		2145.1304
24	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>5</sub>		2179.1247
25	Neu5Ac <sub>1</sub> HexNAc <sub>4</sub> Hex <sub>4</sub>		2220.1512
26	HexNAc <sub>3</sub> Hex <sub>7</sub>		2226.1539
27	Neu5Ac <sub>1</sub> HexNAc <sub>5</sub> Hex <sub>3</sub>		2261.1789
28	HexNAc <sub>4</sub> Hex <sub>6</sub>		2267.1771
29	HexNAc <sub>5</sub> Hex <sub>5</sub>		2308.2036
30	HexNAc <sub>6</sub> Hex <sub>3</sub> Fuc <sub>1</sub>		2319.2228
31	Neu5Ac <sub>1</sub> HexNAc <sub>2</sub> Hex <sub>7</sub>		2342.2003
32	HexNAc <sub>6</sub> Hex <sub>4</sub>		2349.2302
33	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>6</sub>		2383.2258
34	HexNAc <sub>7</sub> Hex <sub>3</sub>		2390.2567
35	Neu5Ac <sub>1</sub> HexNAc <sub>4</sub> Hex <sub>5</sub>		2424.2510
36	Neu5Ac <sub>1</sub> HexNAc <sub>5</sub> Hex <sub>4</sub>		2465.2775
37	HexNAc <sub>4</sub> Hex <sub>6</sub> Fuc <sub>1</sub>		2482.2929
38	HexNAc <sub>5</sub> Hex <sub>6</sub>		2512.3034
39	Neu5Ac <sub>1</sub> HexNAc <sub>2</sub> Hex <sub>8</sub>		2546.3122
40	HexNAc <sub>6</sub> Hex <sub>5</sub>		2553.3300
41	HexNAc <sub>7</sub> Hex <sub>3</sub> Fuc <sub>1</sub>		2564.3679

42	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>7</sub>		2587.3267
43	HexNAc <sub>7</sub> Hex <sub>4</sub>		2594.3565
44	Neu5Ac <sub>1</sub> HexNAc <sub>4</sub> Hex <sub>6</sub>		2628.3569
45	HexNAc <sub>8</sub> Hex <sub>3</sub>		2635.3831
46	Neu5Ac <sub>1</sub> HexNAc <sub>5</sub> Hex <sub>5</sub>		2669.3773
47	HexNAc <sub>5</sub> Hex <sub>7</sub>		2716.4039
48	Neu5Ac <sub>1</sub> HexNAc <sub>3</sub> Hex <sub>8</sub>		2791.4264
49	HexNAc <sub>7</sub> Hex <sub>5</sub>		2798.4599
50	HexNAc <sub>8</sub> Hex <sub>4</sub>		2839.4828
51	Neu5Ac <sub>1</sub> HexNAc <sub>5</sub> Hex <sub>6</sub>		2873.4763
52	HexNAc <sub>9</sub> Hex <sub>3</sub>		2880.51
53	Neu5Ac <sub>1</sub> HexNAc <sub>6</sub> Hex <sub>5</sub>		2914.5036
54	Neu5Ac <sub>1</sub> HexNAc <sub>7</sub> Hex <sub>4</sub>		2955.534
55	HexNAc <sub>8</sub> Hex <sub>5</sub>		3043.5826
56	Neu5Ac <sub>1</sub> HexNAc <sub>8</sub> Hex <sub>4</sub>		3200.658
57	HexNAc <sub>8</sub> Hex <sub>6</sub>		3247.6824
58	Neu5Ac <sub>1</sub> HexNAc <sub>8</sub> Hex <sub>5</sub>		3404.7563
59	Neu5Ac <sub>1</sub> HexNAc <sub>8</sub> Hex <sub>6</sub>		3608.8545











60	HexNAC <sub>2</sub>		552.3258
61	HexNAC <sub>1</sub> Hex <sub>2</sub>		715.40033
62	HexNAC <sub>2</sub> Fuc <sub>1</sub>		726.41606
63	HexNAC <sub>2</sub> Hex <sub>1</sub>		756.4256
64	HexNAC <sub>1</sub> Hex <sub>3</sub>		919.5003
65	HexNAC <sub>2</sub> Hex <sub>1</sub> Fuc <sub>1</sub>		930.52682
66	HexNAC <sub>2</sub> Hex <sub>2</sub>		960.5254
67	HexNAC <sub>2</sub> Hex <sub>2</sub> Fuc <sub>1</sub>		1134.6149









N-glycan compositions: Hex (hexose: ● galactose/ ● mannose), ■ HexNAC (N-acetylhexosamine), ◆ Neu5Ac (N-acetylneuraminic acid), and ▲ Fuc (fucose). Details of N-glycan released by PNGase can be found in the method section.



## Risk and Safety

Risk and safety pictograms of potentially hazardous chemicals used throughout this study, based on the Globally Harmonized System of Classification and Labelling of Chemicals (GHS), GHS hazard and precautionary statements.

Chemicals	GHS symbol	GHS hazard statement	GHS precautionary statements
Formic acid (FA)	GHS02  GHS05  GHS06 	H226 H302 H314 H331 EUH071	P210 P280 P301 + P330 + P331 P304+P340 P305+P351+P338 P308 + P310
Acetic acid	GHS02  GHS05 	H226 H314	P210 P280 P301+P330+P331 P305+P351+P338 P308+P310
Sodium acetate		H319	P264 P280
2-(N-morpholino)ethanesulfonic acid (MES)		H315 H319	P264 P280 P302+P352 P305+P351+P338 P321 P332+P313 P337+P313 P362+P364
Disodium phosphate (Na <sub>2</sub> HPO <sub>4</sub> )	 	H318 H319	P264, P280, P310 P305+P351+P338 P337+P313
Ammonium acetate		H315 H319 H335 H402	P261, P264, P271, P273, P280, P312, P362 P302+P352 P305+P351+P338 P332+P313 P337+P313

Acetonitrile (ACN)	GHS02  GHS07 	H226 H314	P210 P280 P301+P330+P331 P305+P351+P338 P308+P310
Ammonium bicarbonate (ABC)	GHS07 	H302	P301+P312 P330
Dithiothreitol (DTT)	GHS07 	H302 H312 H332	P261 P280 P301+P310 P304+P340 P361 P501
Iodoacetamide (IAA)	GHS06  GHS08 	H301 H317 H334	P261 P280 P301 + P310 P342 + P311
Trypsin	GHS07  GHS08 	H315 H319 H334 H335	P302 + P352 P304+P340 P305+P351 P342+P311

## Acknowledgements

Alhamdulillahirobbilalamin, all praise be to Allah. Surely due to His blessing I came to this point in life.

I would like to thank everyone who supported me academically and psychologically,

I sincerely thank my supervisor, Prof. Dr. Hartmut Schlüter for welcoming me to his group, mass spectrometry proteomics at the University Medical Centre Hamburg (UKE). Thank you for supervising, teaching, guiding, and giving me never ending support.

I thank the Horizon 2020 Marie Skłodowska-Curie Action ITN 2017 of the European Commission (H2020-MSCA-ITN-2017) for the funding. I also thank everyone in the analytics for biologics (A4B) consortium for the amazing collaborations, networking, and opportunities to always learn. To Manasi Gaikward and Kim Jihyung, thank you for helping in developing method for my sample analysis and data analysis.

I thank everyone in the mass spectrometry proteomics group and in the institute of clinical chemistry and laboratory medicines, UKE, for the endless help and fantastic working environment.

I thank my parents, siblings, extended families, and friends for their unconditional supports.

I thank myself for trying the best searching positivity in everything. Thank you for coming this far and not giving up.

## Declaration

### Declaration on oath

I hereby declare that this thesis has been generated by me from my own research and experiments. I have not used or written other than the acknowledged resources and aids. The submitted written version corresponds to the version on the electronic storage medium. I hereby declare that I have not previously applied or pursued for a doctorate (Ph.D. studies).

A handwritten signature in blue ink, appearing to be 'Siti Nurul Hidayah', written in a cursive style.

Hamburg, 21 May 2021

Siti Nurul Hidayah

---

Place and Date

Signature