Faculteit Wetenschappen

Departement Chemie

Mass spectrometry-based methods to explore higher-order protein structure:

Generating a fingerprint for biologics

Proefschrift voorgelegd tot het behalen van de graad van

doctor in de wetenschappen: chemie

aan de Universiteit Antwerpen te verdedigen door

Sneha Chatterjee

Promotor: Prof. dr. Frank Sobott
Co-promotor: Prof. dr. Geert Baggerman

Antwerpen, 28 Maart 2018

Acknowledgements

These last four years were filled with delightful experiences and there are quite a few people I would like to thank. My PhD adventure started in 2013, when after my master thesis I asked my promotor, Prof. Frank Sobott, to continue working in his lab as a PhD student. With his advice and support I was encouraged to successfully apply for an IWT grant. Thanks, Frank, for guiding me throughout these years and training me to become a better scientist. I do hope that one day you will consider writing a book about all the fun anecdotes you have told us during the lunch breaks. I wish you all the best for your new chapter at the University of Leeds.

If we turn the clock back a bit further to 2008, when I was visiting the Chemistry stall at the University of Antwerp, I met Prof. Filip Lemière, who was promoting this field. Thanks, Filip, for all the enthusiastic lectures through-out the years and for educating me in lab practices The mass spectrometry-related lectures of Frank and Filip during my studies were the reason I got interested in this field and I am happy I could end up in the BAMS group.
Furthermore, I will also remember the MS lectures of Prof. Luc Van Vaeck, who had given me the advice that I should not doubt myself and that can I achieve more than I think.

I would like to thank the other members of my jury as well. Prof. Geert Baggerman gave me valuable advice, especially during the mid-year PhD review, and showed me how to use the Orbitrap instrument. Thomas de Vijlder was very helpful these last years, from making our collaboration with J&J possible, to giving me valuable feedback on my data. I would like to thank Prof. Koen Augustyns, vice-dean of the faculty of pharmaceutical, biomedical and veterinary sciences, for his feedback on my thesis and his interesting questions during the preliminary defence. Finally, I would like to thank the two ladies in my jury, who are also two scientists whose experience I truly admire. Prof. Sarah Cianférani, an expert in MS of biotherapeutics, and Prof. Alison Ashcroft, an expert in biomolecular mass spectrometry, gave me valuable feedback to improve my thesis.

The foundation of my thesis was built on the different collaborations I had with universities and companies. Special thanks to my MedImmune collaborators; J.J., Nick, Dan and Christel for helping me during my research stays. Thank you, Machteld from KUL, for the interesting project on PAI-1 and nanobodies.

A big thank you to all the lovely people I worked with on the 4th floor of the V building! Jasper and Annika, who helped me during my Bachelor and Master Thesis projects. Jeroen and Albert were always there when I needed them, whether it was for advice or for when the instrument stopped working. A special thanks to Norbert and Glenn, without whom nothing in the lab would be working. I am also confident that Rani, Elise and Dietmar will run this lab perfectly. Jeroen, I look forward to attending your defence next!
I have a lot of love for my Chemistry ladies, Inne and Sanne!
I also want thank my Chemistry class mates, Ben, Carl, Jens and Wim, and look forward to the yearly class reunions.

All my life, I have been fortunate to have the support and advice from my family and friends. I thank my parents Dr. Nirjhar Chatterjee and Dr. Shyama Ghosh. When people say I am following in my parent's footsteps, I feel proud, as my mom and dad have already accomplished so much in their lives. They dedicated a lot to me and my brother, and we will always be grateful for their support, advice and love.

Thank you Nitish, my big brother, for all your support. You have been an inspiration to me as you balance your life with your lovely wife, Sofie, and your darling son, Arthur, while maintaining your career and MBA studies. Arthur, my little nephew, when I saw you for the first time in September '17 you made me forget all about my PhD stress. Your smiling face lights up our lives.

I would like to thank my Thakurdada, my Thakurma, my Dida and my Dadu for being the best grandparents. You have always supported me through everything and reminded me the importance of education and intelligence.

To my dear family-in-law, the Radmals, thanks for all your encouragement during these years and for taking care of me while I was locked up in my room, writing this thesis. I am happy I could end my PhD chapter now in order to fully enjoy our great party this summer!

To my fiancé, Abassin.
I will try to keep this brief, as I can write pages about how you have helped me these last years. While I could manage to keep it cool during the day, you saw the excitement, but also the frustrations, the restlessness and the over-thinking I encountered when I got home. Nevertheless, during every step you encouraged me and gave me confidence to go through every challenge. I am so lucky to have you as my partner. Thank you for everything. I look forward to moving into our new home and starting our married life this summer.

Science literacy is the artery through which the solutions of tomorrow's problems flow.
- Neil deGrasse Tyson

Summary

Therapeutic proteins have emerged as an important category of biopharmaceutical products, due to their use in the treatment of cancer, inflammation and infectious diseases. To characterise proteins, we require information on the primary to higher-order structure, as these distinct structural aspects control the general structure and basic function of the protein. Rather than maintaining their native state, proteins can undergo conformational changes, which can influence their functional aspects. Characterising the protein structure and the resulting protein function, will therefore enable further evaluation of the safety and quality of the protein as a biotherapeutic agent.

One limitation to do this is that, given their highly flexible and complex structures, there is no precise protocol yet for the complete and detailed characterisation of such large biomolecules. Mass spectrometry (MS) is a key technique for protein characterization, providing detailed information on primary to higher-order structure.

Our goal is to develop a set of MS based methods to thoroughly characterise proteins of biotherapeutic interest, especially the higher-order protein structure. This will decide the quality control of these therapeutic proteins whose biological efficacy, safety and immunogenicity are dependent on their structures.

In this thesis, a set of MS-based methods is used to characterise proteins of biotherapeutic interest such as antibodies, antibody fragments and Fc-fusion proteins. The native MS analysis of intact proteins and protein complexes is discussed. The conformational analysis of proteins was performed using ion mobility-mass spectrometry by monitoring the size and shape of proteins. Top-down electron transfer dissociation (ETD) of antibodies was discussed to analyse solvent-exposed regions of the protein. HDX-MS was employed to locate binding interfaces within protein complexes. The FPOP analysis was applied to intact antibodies and the oxidised samples were measured with native IM-MS.

The range of structural information provided by this set of MS-based methods proves how powerful these techniques can be and how they can aid in understanding the higher-order structure of proteins.

Samenvatting

Therapeutische eiwitten zijn een belangrijke categorie van biofarmaceutische producten. Ze worden gebruikt in de behandeling van kanker, ontstekingen en besmettelijke ziekten. Vergeleken met medicijnen gebaseerd op kleine moleculen, hebben therapeutische eiwitten betere biofysische, farmacokinetische en farmacodynamische eigenschappen.
Om eiwitten te karakteriseren willen we informatie verwerven over de primaire en hogere orde structuur, omdat de algemene eiwitstructuur en de eiwitfunctie afhangen van deze structurele aspecten. Eiwitten blijven niet in hun natieve toestand, maar kunnen conformationele veranderingen ondergaan en deze transities kunnen de functionele aspecten van het eiwit beïnvloeden. Door de karakterisatie van de eiwitstructuur en de resulterende eiwitfunctie, kan de veiligheid en kwaliteit van het eiwit geëvalueerd worden als biotherapeutische agens.

Het probleem is echter dat er nog geen exacte protocol bestaat voor de volledige en gedetailleerde karakterisatie van deze grote biomoleculen, omwille van hun zeer flexibele en complexe structuur. Massaspectrometrie (MS) is een van de belangrijkste technieken voor de karakterisatie van eiwitten, omdat het gedetailleerde informatie kan geven over de primaire tot hogere orde structuur.
Het doel is om een set van MS methoden te ontwikkelen om therapeutische eiwitten volledig te karakteriseren. Dit zal essentieel zijn voor de kwaliteitscontrole van deze eiwitten, omdat hun biologische efficiëntie, veiligheid en immunogeniciteit afhangen van hun structuur.
In deze thesis wordt een set van MS-gebaseerde methoden gebruikt om eiwitten van biotherapeutisch belang zoals antilichamen, antilichaamfragmenten en Fc-fusie eiwitten. De natieve MS analyse van intacte eiwitten en eiwitcomplexen wordt besproken.
De conformationele analyse van eiwitten met behulp van ionenmobiliteit-massaspectometrie (IM-MS) wordt uitgevoerd door de grootte en vorm van het eiwit op te volgen. Top-down elektronen transfer dissociatie (ETD) van antilichamen wordt besproken voor de analyse van de regio's van het eiwit die blootgesteld aan het solvent. HDX-MS was gebruikt om de bindingslocatie in eiwitcomplexen te lokaliseren. De FPOP techniek werd toegepast op intacte antilichamen en de geoxideerde stalen werden gemeten met natieve IM-MS.
Het breed aanbod aan informatie voorzien door deze set van MS-gebaseerde methoden bewijst hoe krachtig deze technieken zijn en hoe ze kunnen helpen in het beter begrijpen van het hogere-orde eiwitstructuur.

Table of Contents

List of abbreviatons

ADC	antibody-drug conjugate
ADCC	antibody-dependent cellular cytotoxcity
AtaT-AtaR	Aminoacyl tRNA acetyltransferase Toxin-Repressor
AUC	analytical ultracentrifugation
AURKC	aurora kinase C
BAMS	biomolecular and analytical mass spectrometry
BEH	ethylene-bridged hybrid
CCS	collision cross section
CDC	compliment-dependent cytotoxcity
CDR	complimentarity-determining region
CEM	chain ejection model
C_H	constant region of the heavy chain of an antibody
CH	variable domain of the heavy chain
CID	collision-induced dissociation
CIU	collision-induced unfolding
C_L	constant region of the light chain of an antibody
CRM	charged-residue model
CTDP	consortium for top-down proteomics
DNA	deoxyribonucleic acid
DT-IMS	drift-tube ion mobility mass spectrometry
DTT	dithiothreitol
e.g.	exempli gratia
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionisation
et al.	et alii
ETD	electron transfer dissociation
Fab	antigen-binding fragment of an antibody
Fc	crystallisible fragment of an antibody
FcRn	neonatal Fc receptors
FDA	food and drug administration
FPOP	fast photochemical oxidation of proteins
GLP-1	glucagon-like peptide 1
Hc	heavy chain of an antibody
HDX	hydrogen-deuterium exchange
HIV	human immunodeficiency virus
i.e.	id est
ICH	International conference on Harmonisation
IdeS	Immunoglobulin G-degrading enzyme of S. pyogenes
IEM	ion evaporation model
IgG	Immunoglobulin G
IM	ion mobility
IMS	ion mobility spectrometry
INCENP	inner centromere protein
Kd	dissociation constant

KGP	GingisKHAN protease
Lc	light chain of an antibody
LC-MS	liquid chromatography-mass spectrometry
m/z	mass-over-charge
mAbs	monoclonal antibodies
MALDI	matrix assisted laser desorption/ionisation
mRNA	messenger RNA
MS	mass spectrometry
MW	molecular weight
ncRNA	non-coding RNA
PAI-1	plasminogen activator inhibitor 1
PTM	post-translational modifications
Q-TOF	quadrupole-time-of-flight
RCL	reactive center loop
RNA	ribonucleic acid
RNA	RNA interference
rRNA	ribosomal RNA
scFv	single chain variable fragment
scpA/B	segregation and condensation proteins
SEC	size-exclusion chromaography
SEC-MALS	size-exclusion chromaography-multi-angle light scattering
SMC	structural maintenance of chromosomes
TAR-RNA	trans-activation response RNA
TAT	trans-activator of transcription
TCO	trans-cyclooctene
TOF	time-of-flight
t-PA	tissue-type plasminogen activator
tRNA	transfer RNA
ttz	tetrazine
TWIMS	travelling-wave ion mobility spectormetry
u-PA	urokinase-type plaminogen activator
UPLC	ultra performance liquid chromatography
VL	variable domain of the light chain

Chapter 1

Introduction

1.1 Biotherapeutic Proteins

Proteins are biological macromolecules that play many essential roles in living organisms. These roles include the transport and storage of molecules, catalysis or inhibition of chemical reactions, as well as immune protection. These specific characteristics allow us to design different proteins that will act as biotherapeutic agents and interact with particular targets, thereby affecting specific functions. Currently, proteins are the prime target for drugs, but these proteins or peptides can also be the drugs themselves. Proteins of biotherapeutic interest include antibodies, antibody fragments, antibody conjugates and other biomolecules which specifically interact with disease-related states of proteins.

Biologics, or biotherapeutics, are one of the fastest growing fields in the biotech and pharmaceutical industries. Their use in drug development relies on biopolymers, such as peptides, proteins and monoclonal antibodies. These therapeutic agents have been used in the treatment of cancer, inflammation and infectious diseases [1]. There has also been an increasing interest in the use of oligonucleotides in therapies. Until recently, the vast majority of drugs have been based on small molecules which are synthetically made and often derived from plant ingredients. In comparison, biotherapeutics have superior biophysical pharmacokinetic and pharmacodynamic properties, and the potential for the design of "tailor-made" drugs is vastly greater than that with chemical structures. The highly specific and complex set of functions is one of the advantages of biotherapeutics.
Monoclonal antibodies, antibody fragments and antibody-drug conjugates in particular are fast becoming an important class of biotherapeutics [2]. The worldwide annual revenue of biotherapeutics is more than 165 billion dollars [3]. **Table 1** shows the top 10 best-selling drugs in 2016. Major pharmaceutical companies are carrying out research on producing and targeting antibodies for diagnostic and treatment purposes.

Table 1: Top 10 best-selling pharmaceutical drugs in 2016 with the biotherapeutic drugs marked in grey. The remaining drugs in this list are based on small organic molecules. (www.genengnews.com)

Rank	Product	Active ingredient	Main indications	Manufacturer	Sales (USD)
1	Humira	Adalimumab	Arthritis	AbbVie	16.078 billion
2	Harvoni	Ledipasvir/ Sofosbuvir	Hepatitis C virus	Gilead Sciences	9.081 billion
3	Enbrel	Etanercept	Arthritis	Amgen and Pfizer	8.874 billion
4	Rituxan	Rituximab	Non-Hodghkin's lymphoma	Roche and Biogen	8.583 billion
5	Remicade	Infliximab	Crohn's disease	J&J and Merck	7.829 billion
6	Revlimid	Lenalidomide	Multiple myeloma	Celgene	6.974 billion
7	Avastin	Bevacizumab	Colorectal cancer	Roche	6.752 billion
8	Herceptin	Trastuzumab	Breast cancer	Roche	6.751 billion
9	Lantus	Insulin glargine	Diabetes	Sanofi	6.054 billion
10	Prevnar 13	CRM197 Protein	Pneumococcal pneumonia	Pfizer	5.718 billion

The first fully human monoclonal antibody, approved by the FDA, was Adalimumab (Humira) which is still one of the best selling drugs. Humira is a drug against the tumor necrosis factor (TNF) drug and it is used in the treatment of arthritis. It was a result of a collaboration between BASF Bioresearch Corporation and Cambridge Antibody Technology, U.K., manufactured and marketed by Abbott Laboratories, after Abbott's acquisition of BASF Pharma. Currently, Humira is sold by the pharmaceutical company Abbvie. Large pharmaceutical companies as well as biotechs and startups are developing novel technologies to design antibody-based, protein /peptide and oligonucleotide therapeutics that display superior properties including efficacy and selectivity. These novel drugs are essential as there is a need for better drugs for e.g., cancer, neurodegeneration, auto-immune diseases, etc. Next to creating a well-performing drug, they are also addressing issues concerning the stability and shelf-life of the therapeutics. These issues include the degradation and aggregation of therapeutic proteins during manufacture, storage or *in vivo* [4, 5]. Making drugs in an easy-to-make stable application form is currently a big challenge for many biopharmaceuticals. As a consequence, it is important to be able to thoroughly study these proteins of therapeutic interest. As these biotherapeutic compounds are large and complex structures, and can be sensitive to environmental conditions during production, they require very sophisticated quality control processes. This is why there is an urgent need for the structural characterisation of biotherapeutics, particularly aspects of higher-order structure (**Fig.1**), as the biological efficacy, safety and immunogenicity of these therapeutic agents need to be guaranteed.

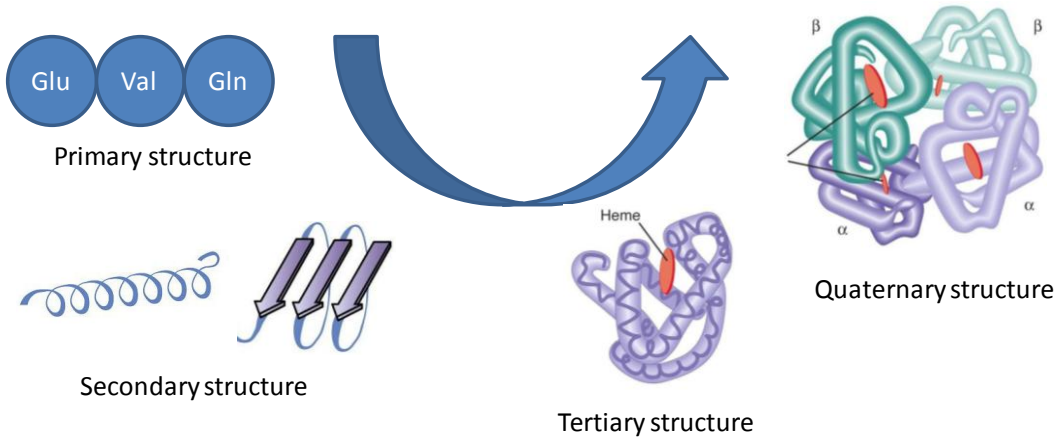


Figure 1: Schematic overview of primary to higher order protein structure.

In order to ensure the consistent quality of all protein therapeutics, the International Conference on Harmonisation (ICH), the Food and Drug Administration (FDA), and the European Medicine Agency (EMA) have established guidelines to regulate the manufacturing process of these biological compounds. These guidelines include descriptions of specific test procedures for biotechnological or biological products, and stability tests and comparability tests for changes in the manufacturing process. These analytical tests currently include electrophoresis, high resolution chromatography, peptide mapping, identity tests, and activity tests.

However, when these analytical methods are used to probe the higher-order structure (topology of disulfide linkages and PTMs (e.g., glycsoylations), conformation and aggregation) of therapeutic proteins, severe limitations exist due to the complexity of the molecules.

When a change occurs in the production process, manufacturers have to ascertain whether the higher-order structure is maintained in the product, or if a small change is significant for the efficacy and safety of the product.

1.2 Structural characterisation of biotherapeutics

The current and crucial limitation in this field is that there is no comprehensive analytical methodology to ensure the efficacy and safety of biotherapeutics. The cause of this issue lies in the difficulty to characterise the protein structure. We need to understand the structure-function relationship of these large, flexible and complex agents before we can safely deploy them in a therapeutic context. Currently, there is a lack of information on the link between sequence variations in the primary structure (e.g. mutations, truncations, disulfide variants) and how this affects protein folding and higher-order structure (e.g. conformational and aggregation behaviour).
A protein can naturally fold into a unique three-dimensional structure, i.e. its native conformation. However, proteins are generally not rigid biomolecules, as they are dynamic and can often change their conformation. These structural transitions influence the functional aspects of the protein.
Characterising the protein structure can enable us to link certain protein conformations to the resulting protein function. Once we know the specific function, the quality and the safety of the protein as a biotherapeutic agent can be further evaluated.

At this stage, different methods are used for protein characterisation:

Size exclusion chromatography (SEC), a size-based separation technique, is used for antibody aggregation analysis. SEC is often combined with light scattering techniques such as MALS (multi-angle light scattering). Protein aggregation is considered to be a critical attribute in quality control as it can induce undesirable consequences in protein function. Analytical ultracentrifugation (AUC) has also been used to analyse antibody aggregation by providing a quantitative size-distribution analysis.
Protein crystallography, which studies the three-dimensional structures of protein crystals at near-atomic resolution, has provided insight into biological processes. Small angle x-ray scattering can characterise antibodies in solution by measuring the scattering curve and determining the protein size on a nanoscale.
Nuclear magnetic resonance (NMR) spectroscopy has been used to look at protein binding interfaces by comparing the magnetic properties of specific atomic nuclei. It can also show protein degradation (e.g., oxidation, deamidation, etc.) and when the analysed protein crystallises in a different crystal form, it can indicate that the protein has undergone a conformational change.
Circular dichroism, which measures the difference in absorption between left- and right-polarised light, can investigate the secondary protein structure. It is a valuable tool for showing changes in conformation induced by e.g., temperature or denaturing agents. These currently-used biophysical techniques can provide important structural information, but in general a large amount of sample is needed and the experiments can be time-consuming to optimise. It is also not always evident to look at the heterogeneity (e.g., presence of different

modifications) of the sample. Furthermore most of these techniques do not allow simultaneous analysis of different protein conformations.

Mass spectrometry (MS) is already established as one of the key techniques for characterising proteins, as it can provide detailed information on their primary to higher order structure.

Novel MS methods can address issues of post-translational modifications (PTMs) and sequence variations in their entirety, as well as conformational and aggregation behavior. Comprehensive characterisation includes the determination of amino acid sequences, exact molecular weight, the full complement of post-translational modifications and the aggregation state of biotherapeutic compounds. Only the determination of molecular weight and bottom-up sequencing by MS are currently routine, while the other aspects of protein structure are much more difficult to define.
Developing a set of MS-based methods to characterise proteins of biotherapeutic interest will be essential for quality control, as their biological efficacy, safety and immunogenicity depend on their structure.

1.3 Developing an analytical toolbox for structural characterisation

Until now, there has been no comprehensive protocol formulated, and accepted by the industry as well as regulatory bodies, for the characterization of biotherapeutic agents. Developing such an analytical toolbox based on mass spectrometry methods can give us the means to provide detailed information on the structural aspects of biologicals with therapeutic capacity, which will be essential for defining their safety, efficacy and product consistency.

The goal of this research is to develop a set of high-throughput methods that will contribute to future regulatory standards for biotherapeutic proteins. These methods will characterise biotherapeutic agents and obtain unique profiles or fingerprints. In the long term these profiles can help to increase our understanding of structure-function relationships in the context of protein engineering.

The scientific objectives of this PhD project are:

- To determine the **global conformational properties** of proteins using native ion mobility mass spectrometry (**IM-MS**), and to understand how conformation affects protein function.
- To locate the **interactions** of antibodies or antibody fragments with targets using native mass spectrometry (**native MS**)
- To monitor **structural and dynamic aspects** of proteins by using hydrogen/deuterium exchange (**HDX**) in combination with mass spectrometry.
- To monitor **structural and dynamic aspects** of proteins by using fast photochemical oxidation of proteins (**FPOP**) in combination with mass spectrometry.
- To determine the **primary structure**, i.e. sequences, variations, disulfide bridges and post-translational modifications (PTMs), using top-down fragmentation methods based on Electron Transfer Dissociation (**ETD**).

Currently, these methods are not the routine techniques used for protein characterisation in the industry. By optimising these MS-based methods we aim to develop a more detailed and thorough characterization protocol. When we combine the proposed methods and integrate the obtained data, we could create a set of analytical tools that can effectively measure the quality, efficacy and safety of a product.

1.4 Characterising therapeutic agents and therapeutic targets

In the four year-journey for my PhD, I was lucky to participate in numerous interesting projects in collaboration with universities and companies. In this thesis the characterised biomolecules can be classified either as therapeutic agents, or as therapeutic targets.
The studied therapeutic agents include monoclonal antibodies, Fc-fusion proteins, nanobodies and single chain variable fragments (scFv). The studied therapeutic targets include RNA, kinase complexes and plasminogen activator inhibitor (PAI-1). For each project questions about the protein structure could be answered using mass spectrometry-based methods (**Table 2**).

Table 2: MS-based techniques used on therapeutic agents and targets.

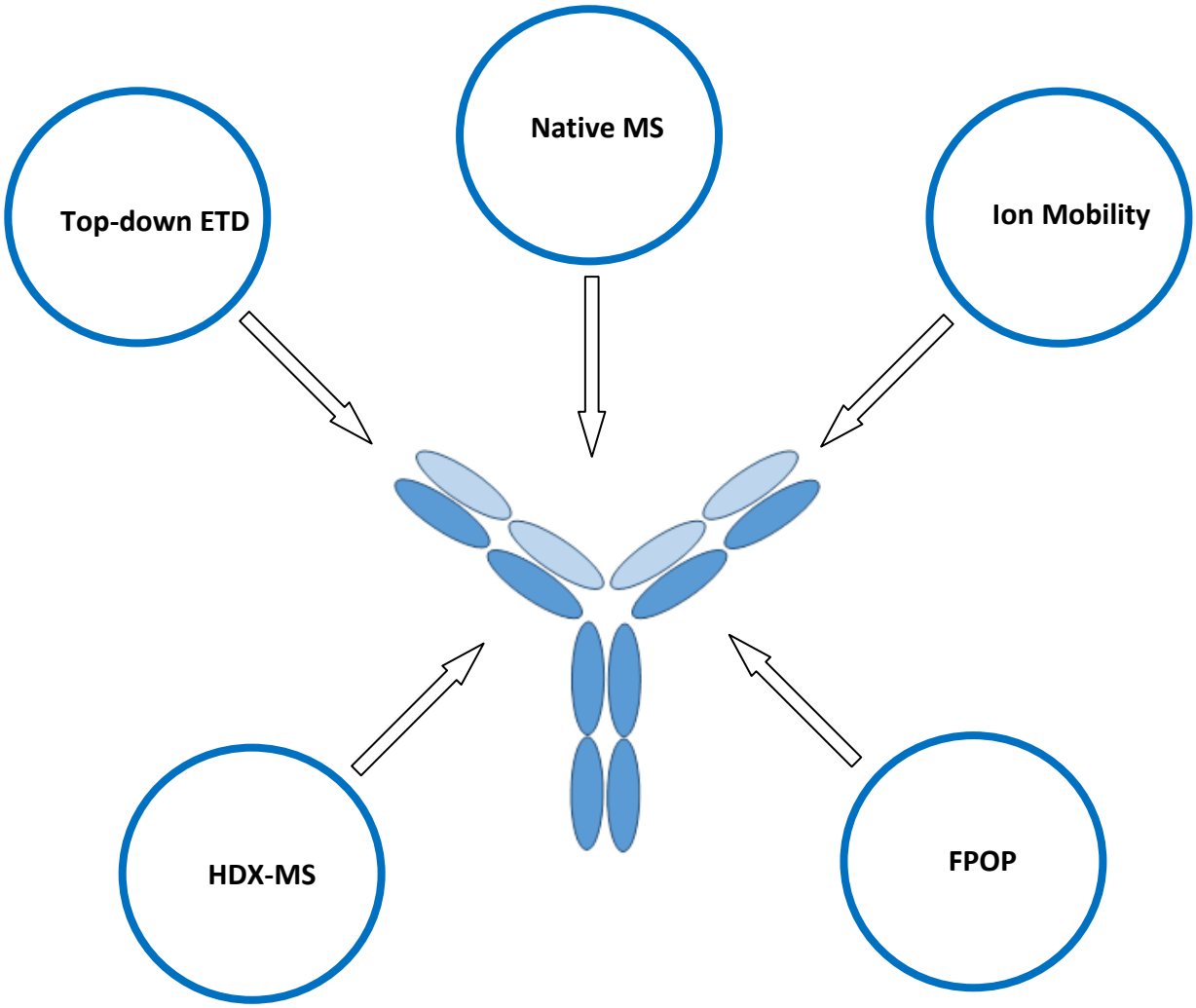
Technique	Antibody	Fc-Fusion protein	Nanobody ScFV	RNA	Kinase complex	PAI-1
Native MS	x	x	x	x	x	x
IM-MS	x	x	x	x	x	x
Top-down ETD	x					
HDX-MS		x				
FPOP	x					

1.5 Sources

[1] J. Mo, A. Tymiak, G. Chen, Structural mass-spectrometry in biologics discovery: advances and future trends, Drug Discovery Today, 17, Dec 2012
[2] A. Beck, S. Sanglier-Cianférani, A. Van Dorsselaer, Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry, Anal Chem. 5; 84(11):4637-46, Jun 2012
[3] R. R. Rader (Ed.), Biosimilars/Biobetters Pipeline Review, 2014. <www.biopharma.com>.
[4] K. Nandra, Therapeutic use of monoclonal antibodies and future trends in biotherapeutics, OBR review, 2013
[5] B. Kükrer, F. Vasco, E. van Duijn, P.T. Kasper, R.J. Vreeken, A.J.R. Heck, W. Jiskoot, Mass Spectrometric Analysis of Intact Human Monoclonal Antibody Aggregates Fractionated by Size-Exclusion Chromatography. Pharm Res.; 27(10):2197–2204, Oct 2010

Chapter 2

Mass spectrometry-based methods to characterise biotherapeutic proteins



2.1 The power of mass spectrometry

Mass spectrometry is one of the key techniques for characterising proteins, as it can provide detailed information on their primary to higher order structure. This technique has the ability to separate and identify components by differences in mass and charge.
A mass spectrometer is composed of three general components: an ion source, a mass analyzer and a detector. The molecular ions produced by the ion source are transferred to the mass analyzer by a voltage gradient. The mass analyzer separates the ions based on their mass-over-charge value (m/z). The sorted ions are then led to the detector, resulting in the formation of a mass spectrum.
There are two general approaches to analyze proteins with MS instrumentation (**Fig. 1**). The first is an intact or top-down approach that uses native electrospray ionization, ion mobility, and top-down fragmentation [1]. The second is a bottom-up approach that uses tandem mass spectrometry to provide peptide information after enzymatic digestion of the protein.

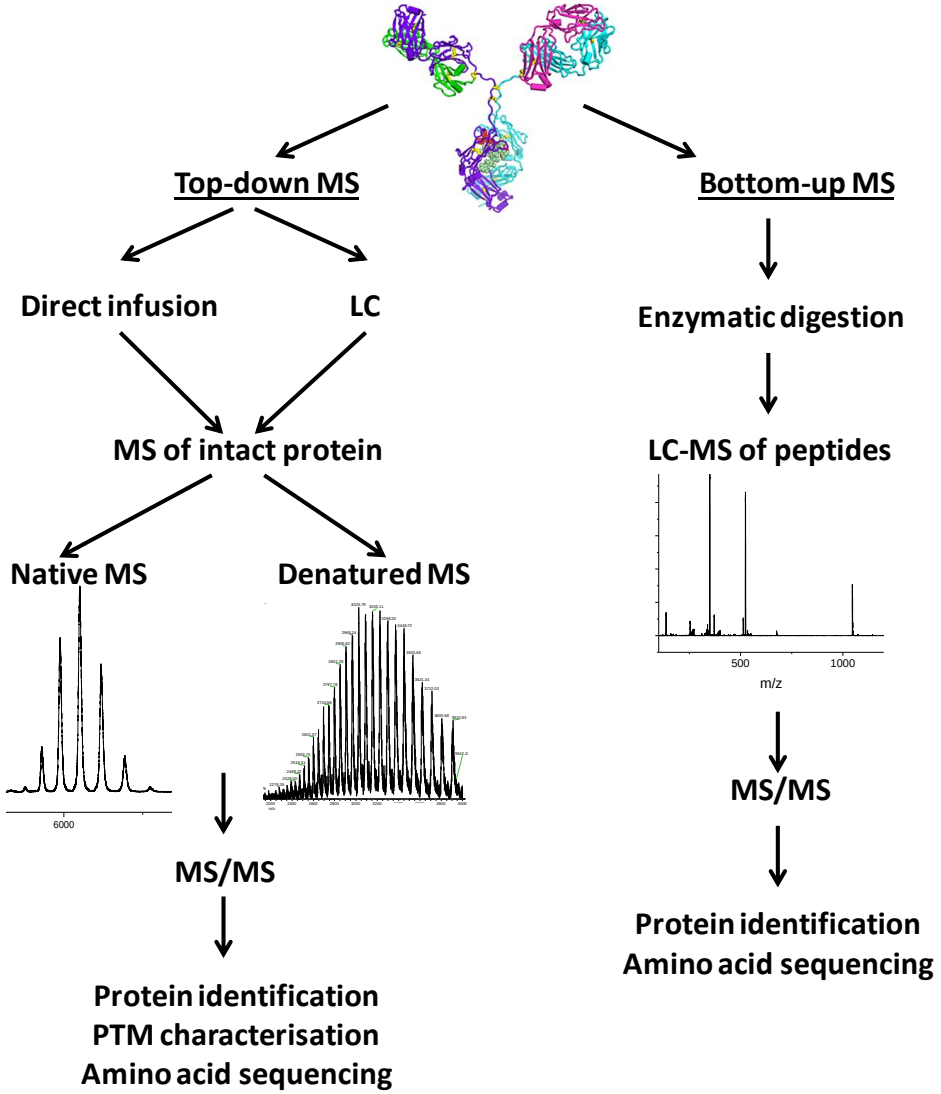


Figure 1: Overview of top-down vs. bottom-up mass spectrometry. The antibody structure shown is from PDB 1IGT and data were acquired on Synapt G2 MS.

In top-down MS the protein structure can be analysed in intact native or denatured form, while in bottom-up MS the digested protein is analysed (**Fig. 2**).
The advantage of bottom-up MS is that it can identify proteins in digests derived from complex mixtures. The bottom-up approach is the most widely used for protein identification and characterisation. The practical limitation of the bottom-up strategy is that usually only a fraction of the peptide population is identified. Furthermore, there is a loss of information about the post translational modifications (PTMs) due to the limited sequence coverage. Thus the full complement of PTMs and sequence variations (e.g., mutations and truncations) cannot be characterised easily.
The advantage of top-down MS is that there is more chance for complete protein sequence and it is also possible to locate and characterise the PTMS. Compared to bottom-up MS, top-down proteomics is a relatively young field and also has some limitations. The dissociation techniques used for sequencing require long ion scanning times. The generated complex spectra also limit the analysis to simple protein mixtures. The mechanisms of protein dissociation are currently less understood than those of peptide dissociation. With top-down MS there are more factors influencing the fragmentation such as the precursor ion charge state, the primary and higher order protein structure and the PTMS. As top-down MS is evolving, the understanding of the fragmentation mechanisms is important to study in order to improve the bioinformatic tools for high-throughput experiments.

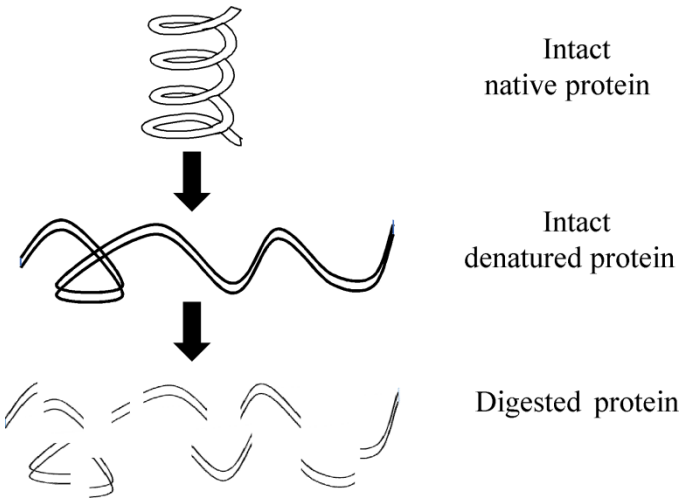


Figure 2: Overview of the different states encountered in protein MS analysis.
Figure published in [2]

In this thesis the focus lies on the methods that take an intact, native or top-down approach. The MS instrument used for the experiments was the Synapt G2 (**Fig. 3**), a quadrupole time-of-flight (Q-TOF) instrument with an ion mobility separation cell. This instrument combines the scanning capacity of a quadrupole mass filter with the high resolving power of a TOF detector. TOF instruments can have an approximate resolution of 20,000 to 80,000 (m/Δm) with 'm' the mass of the peak and 'Δm' the difference between two neighbouring peaks at full width at half maximum (FWHM) [3]. The technical details and the applications of the different methods are explained in this chapter.

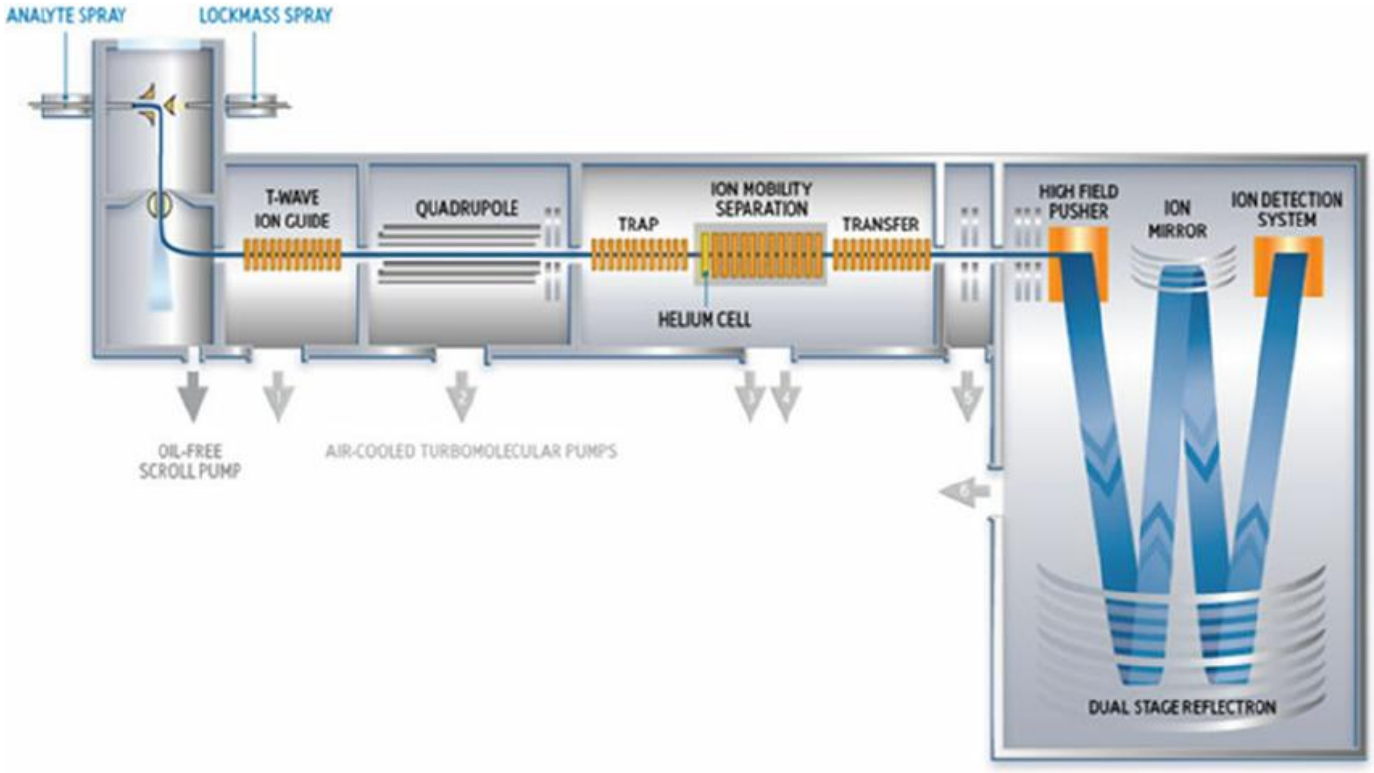


Figure 3: Schematic illustration of the quadrupole-IMS-TOF instrument Synapt G2 (Waters)

2.2 Electrospray ionisation (ESI)

Electrospray ionisation is a soft ionisation technique that transfers ions from solution into the gaseous phase using a metal or glass capillary that is held at a high electric potential (e.g., 3-5 kV for conventional ESI, <2 for nanoESI) (**Fig. 4**) [61]. The technique is commonly used in positive ion mode. Neutral compounds are converted to ions due to the presence of ions such as H^+, NH_4^+, Na^+ and K^+ in the analyte solution. The ESI process occurs at atmospheric pressure and involves three steps: (1) as the charged solution exits the needle a Taylor cone is formed where the electrostatic repulsion is high enough that a spray of charged droplets is dispersed at the capillary tip, then (2) these droplets undergo solvent evaporation and as the charge density on the shrinking droplets builds up it can reach the so-called Rayleigh limit, where the surface tension of the droplet is overcome by the increasing Coulombic repulsion, resulting in a Coulombic explosion and the formation of smaller droplets. (3) Finally, this process is repeated multiple times until the droplets are converted into gas phase ions. These emitted ions are then accelerated into the mass spectrometer.

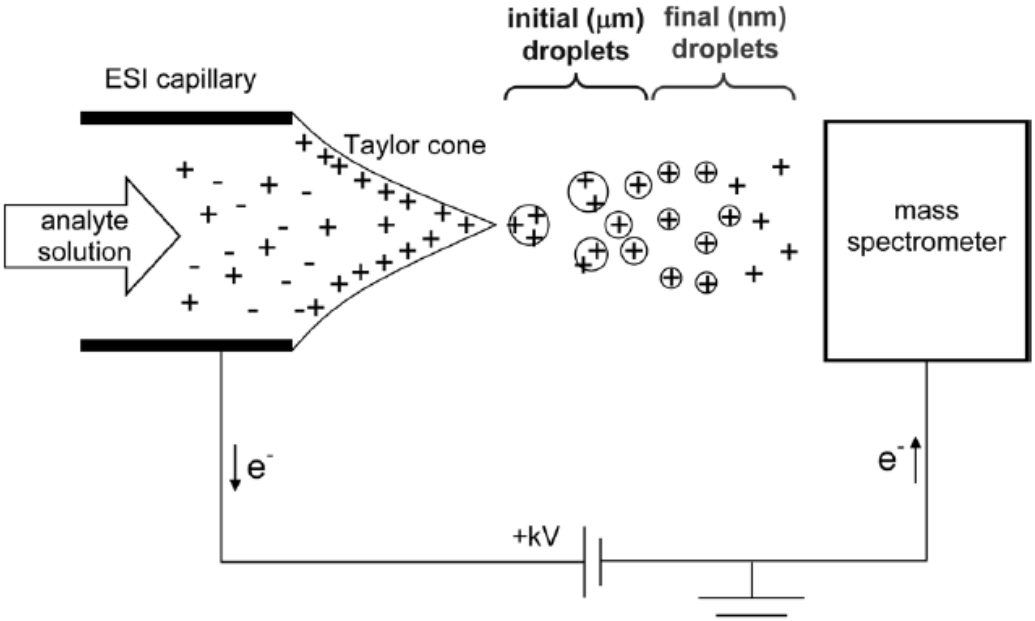


Figure 4: Schematic description of the ESI process in positive mode. Figure from [4]

The production of gas-phase ions from charged droplets can be described in three theories; the ion evaporation model (IEM), the charged-residue model (CRM) and the chain ejection model (CEM). Low molecular weight species are thought to be transferred in to the gas phase via the IEM (**Fig. 5a**). This model states that the electric field at the surface of the droplet is high enough for the ejection of small solvated gas phase ions. Large globular species are assumed to be released into the gas phase via CRM (**Fig.5b**). In this case the nanodroplets undergo extensive evaporation and as the last solvent shell disappears, the charge is transferred to the analyte which was in the vanished droplet. An additional model has been proposed for unfolded proteins. The CEM model describes the ESI mechanism for unfolded proteins or polymer chains that are disordered, partially hydrophobic, and capable of binding excess charge carriers (**Fig. 5c**) [5].

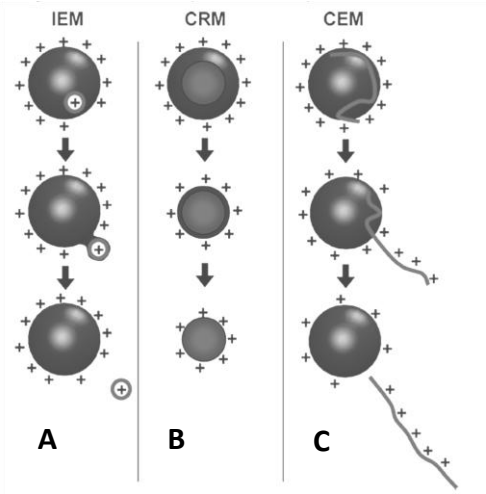


Figure 5: Schematic illustration of three suggested ESI mechanisms; (A) the ion evaporation model (IEM) for small molecules, (B) the charge residue model (CRM) for large globular analytes and (C) the chain ejection model (CEM) for disordered polymers. Fig. adapted from [4].

2.3 Native nano-ESI mass spectrometry

The introduction of nano-ESI has played an important role in the analysis of proteins. Due to the low flow rate of nanoliters per minute, small amounts of analyte can be analysed.
When working with nano-ESI-MS, typically 1 to 4 µL of sample (of e.g. 0.5-10 µM) is introduced into the mass spectrometer through a gold-coated glass capillary with an orifice of 1-5 µm, enabling flow rates of 1.2-30 µL/h (**Fig. 6**) [6]. The initially generated droplets are 100-1000 times smaller than those formed in conventional ESI, resulting in less need for solvent evaporation and lower amounts of non-volatile salts in the droplets [7].
In native mass spectrometry, a volatile buffer (e.g. ammonium acetate) at neutral pH is used with non-denaturing nano-ESI conditions to transfer the intact proteins into the gas phase [8]**.** In this manner the non-covalent interactions are largely preserved after ionisation and the natively folded protein can be studied. The use of volatile buffers is advised as non-volatile salts can adduct to protein ions and reduce sensitivity and mass measuring accuracy.
A typical mass spectrum of an intact protein displays a range of charge states in a Gaussian-like distribution. These observed charge states are formed due to a combination of factors such as the availability of ionisable sites, intramolecular interactions, Coulombic repulsion and the solvent accessible surface area [9][10][11][12]. Studies have shown that mass, charge and surface area are correlated for most proteins [12]. **Figure 7** shows the comparison of the mass spectra of native versus denatured myoglobin. Denatured myoglobin can carry more charges as its protein structure is more unfolded compared to the native myoglobin structure. It is believed that the lowest observed charge state corresponds to the most native-like structure in the gas phase.

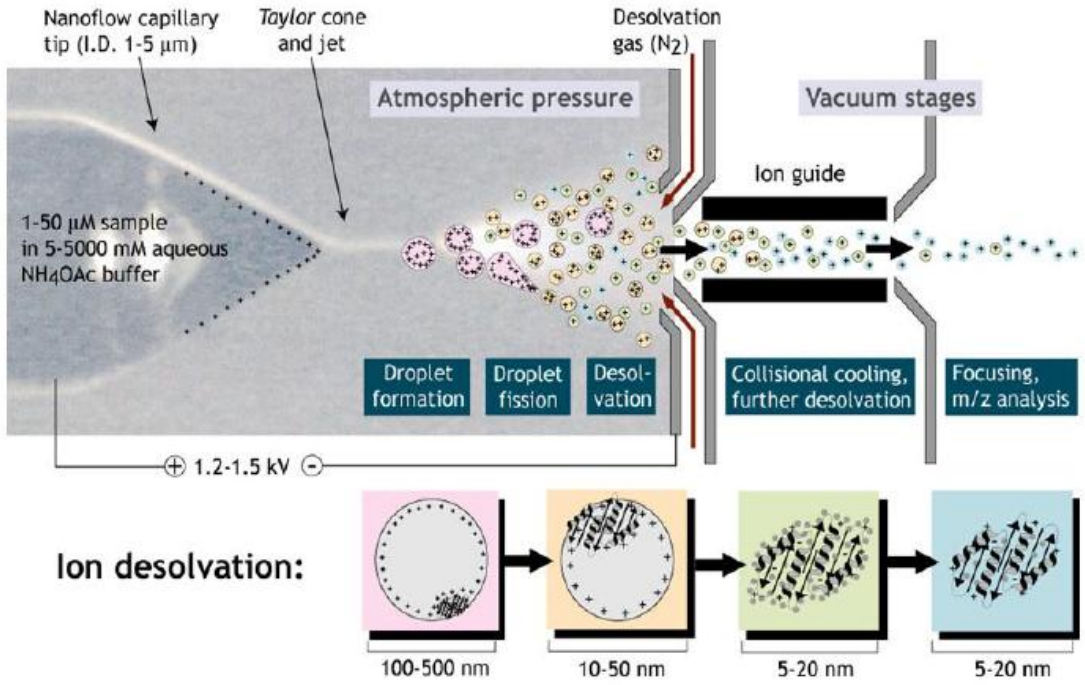


Figure 6: Schematic illustration of the nano-electrospray ionisation process. Figure from [6]

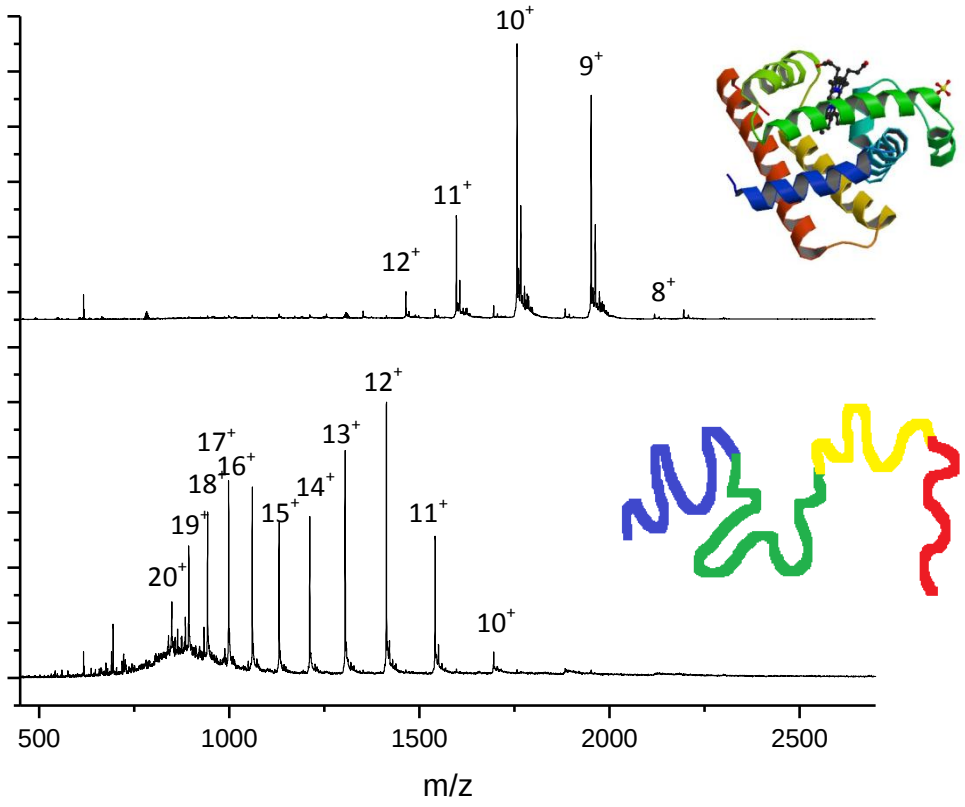


Figure 7: Mass spectra of (A) native myoglobin and (B) denatured myoglobin. Data acquired on a Q-TOF (Micromass, Waters).

The main advantages of native MS are that multiple proteins can be analysed simultaneously in their native-like state and it only requires a few microliters of sample at low micromolar concentration. The native mass spectra show relatively lower charged and fewer peaks compared to denaturing MS, resulting in less overlapping charge state distributions. Furthermore, native MS also allows the analysis of non-covalent complexes and it can study the dynamics of the higher-order structure.

However, when interpreting the results, it should be kept in mind that the MS experiments take place in the gas-phase, where the hydrophobic interactions become weaker and the electrostatic interactions become stronger than in the solution phase. Therefore, the relative abundance of the observed proteins and protein complexes can deviate from that in the solution phase. Yet, computational and experimental research has indicated that the transition from solution to the gas phase does not drastically modify biomolecules [13,14]. Several studies about native MS of proteins and protein complexes have been published, including the native MS analysis of monoclonal antibodies [15, 16]. For example, Marcoux *et al*.have used native MS to study an antibody-drug conjugate, trastuzumab emtanisine, to determine the average drug-antibody rato (DAR) and the drug load profiles [17]. This article also describes how native MS can provide clearly resolved spectra with fewer overlapping charge distributions compared to denaturing MS, allowing the calculation of the average DARs from the individual charge states. Furthermore, when using a high resolution mass analyser (Orbitrap) optimised for native MS, all the glycoforms were clearly resolved. Xuan *et al.* have used high resolution native MS to study antibody-antigen complexes and could determine to stoichiometry and the specific glycoforms observed [18].

2.4 Ion mobility-mass spectrometry (IM-MS)

Ion mobility-mass spectrometry (IM-MS) methods aim to provide information about the global protein structure (degree of compactness and flexibility) and conformational transitions [19,20,21]. With this information the degree of protein folding or misfolding can be addressed, and co-existing or dynamic structures simultaneously detected.

2.4.1 Linear drift-tube ion mobility spectrometry (DT-IMS)

Ion mobility is a technique that can separate ions based on their mobility through a gas. This ability is dependent on the charge, size and shape of the ion. In linear drift-tube ion mobility (DT-IM), the simplest setup of IM, a drift tube is used as an IM analyser and is filled with an inert drift gas, e.g. He [22]. An electric field is generated by, for example, a stack of rings and moves the ions through the gas. The time it takes for the ions to go through the drift tube is referred to as ion mobility drift time. The more compact and smaller analytes will undergo fewer collisions with the drift gas and will have shorter drift times than a more expanded molecule. This results in the separation based on size (**Fig.8**). The average area of an ion that collides with the drift gas as it tumbles through the drift tube is called the collision cross section (CCS). The CCS describes the rotationally averaged overall size and shape of an ion and is related to the ion mobility through **Equation 1** when using DT-IM [23].

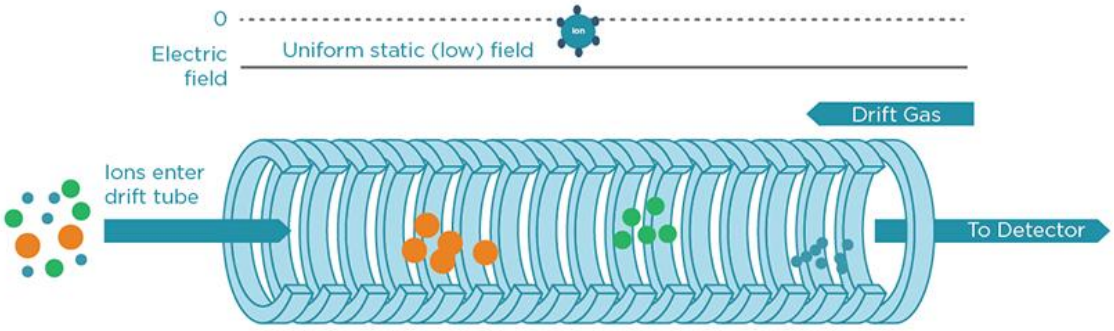


Figure 8: Schematic illustration of linear drift tube ion mobility. Figure adapted from [24].

The collision cross section or CCS (Ω) of an analyte is related to the charge of the analyte (z), the number density of the drift gas (N), ion mobility (K_0), the reduced mass of the ion (μ), the Boltzmann constant (k_b) and the gas temperature (T).

Equation 1: Collision cross section Ω in relation to mobility K_0

$$\Omega = \frac{3ze}{16N}\frac{1}{K_0}\left(\frac{2\pi}{\mu k_b T}\right)^{1/2}$$

When IM is coupled to native mass spectrometry, ions are simultaneously identified by their mass, charge and collision cross section. The advantage of DT-IM spectrometers, besides their potentially high resolving power, is that the analytical equations make it possible to directly calculate the CCS by solving **Equation 1** [23].

2.4.2 Travelling-wave ion mobility spectrometry (TWIMS)

As opposed to DT-IMS, where a constant electric field is used, Travelling-wave ion mobility spectrometry (TWIMS) uses a periodic electric field to guide ions through a stacked ring ion guide filled with a drift gas (e.g. N_2) (**Fig. 9**). The ions are moved by this wave-shaped electric gradient, while they are colliding with the drift gas, which can slow them down. More extended analytes will be slowed down more compared to smaller and more compact analytes. The compact analytes can follow the front of the travelling wave more easily and will have shorter drift times than the more extended analytes, which can be knocked over the travelling wave due to their lower mobility because of increased collisions with the drift gas, resulting in longer drift times. As this version of ion separation changes the relation between IM drift time and CCS, **Equation 1** cannot be used to calculate the CCS. Instead, a set of calibrants with known CCS values is used.

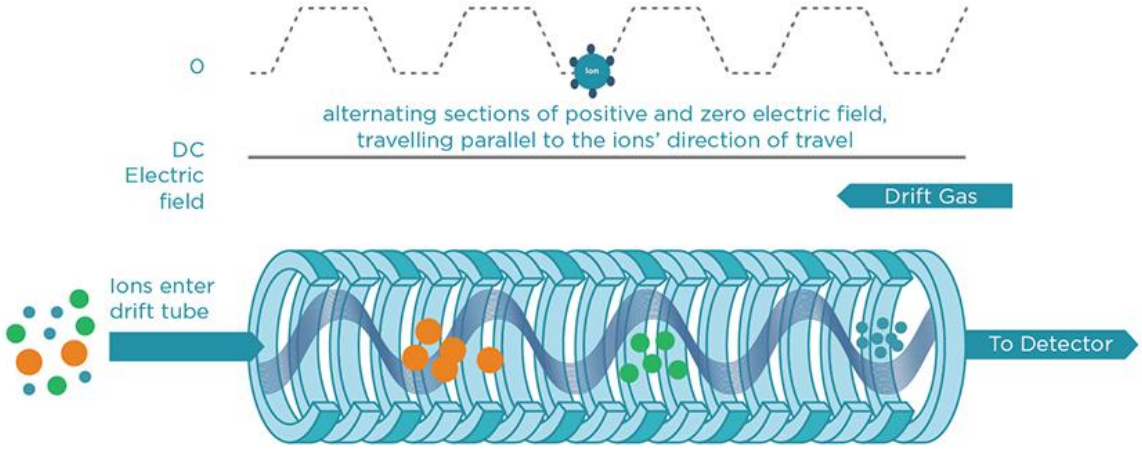


Figure 9: Schematic illustration of travelling-wave ion mobility spectrometry. Figure adapted from [24].

For globular proteins the experimental cross sections obtained by IM-MS are closely related to their molecular mass (**Fig.10**). Databases have been published for a large set of native-like proteins and can be used to calibrate the CCS values for other native-like protein structures [25]. Native IM-MS allows a simple shape and size measurement of monoclonal antibodies. Bagal *et al.* used native IM-MS to differentiate the conformation of IgG2 isomers caused by disulfide linkage variations [26]. Beck *et al.* have used IM-MS to compare therapeutic antibodies and biosimilar antibody versions [27]. Pacholarz *et al.* could observe subtle differences in the IM-MS date from the IgG1 and IgG2 subclasses. Furthermore, the experimental CCS data indicate that protein has collapsed in comparison to the X-ray structure. The molecular dynamics (MD) simulations indicate that the desolvation caused a contraction in the hinge region [28]. Pacholaz et al. have also shown that the IgGs display a wider range of conformations compared to other similarly sized protein complexes. This wider CCS range can be attributed to the dynamic and flexible structure of IgGs around the hinge region.

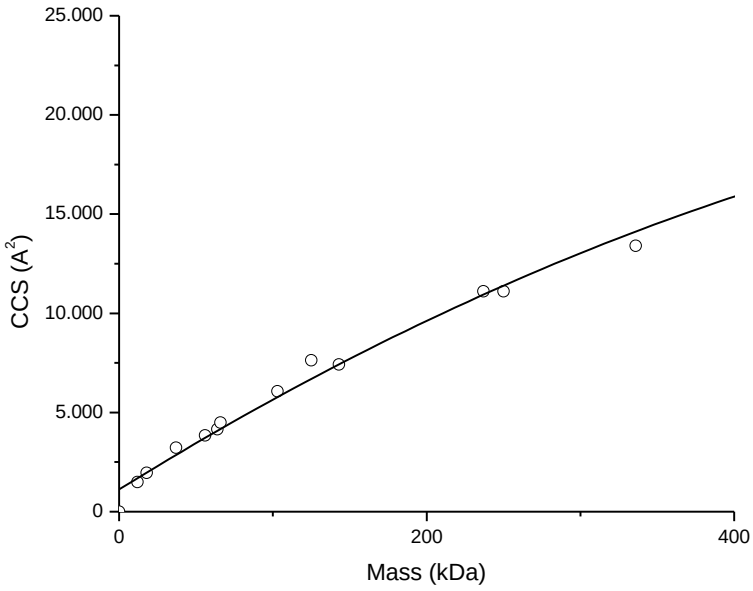


Figure 10: Experimental CCS values of calibrants (globular proteins) are correlated to their molecular weight. Data obtained from [25].

2.5 Collision-induced unfolding (CIU) and collision-induced dissociation (CID)

In ESI-MS, once the generated ions enter the mass spectrometer, their internal energy is determined by the combination of accelerating voltages applied in different sections of the instrument, and the pressure in the different compartments the ions pass through on the way to the detector. This pressure determines the frequency of collisions the ions undergo with the gas molecules in the instrument. Increasing the frequency and energy of these collisions can induce collisional heating which can lead to protein unfolding and dissociation. As the collision energy (CE) is increased the ion mobility is monitored and additional information can be provided on protein unfolding, prior and during dissociation.

Collision-induced unfolding (CIU; i.e. change in CCS detected by IM) can provide information on the protein architecture and (unfolding) stability, while collision-induced dissociation (CID; i.e. change in m/z detected by MS) can determine the (gas-phase dissociation) stability and the stoichiometry of non-covalent complexes. Ruotolo et al. have published a study where the collision unfolding of gas-phase proteins can be correlated to their folded domain structures in solution. For each step-wise transition the protein made from one conformation to another due to collisional unfolding, a correlation could be made to a domain unfolding in the protein structure [29].

Figure 11 shows the schematic workflow of the CIU method applied on the protein myoglobin. The protein is measured with native IM-MS to identify the mass and stoichiometries and derive the IM drift times (**Fig.11A-B**). The IM-drift time of a specific charge state is monitored at increasing collision energies in the trap cell noting the observed conformations and conformational changes (**Fig. 11C)**. A 3D plot is generated of the ion mobility drift times (or the derived collision cross sections) and the intensity of the observed conformations in function of the collision energy, resulting in a CIU plot. (**Fig.11D**)

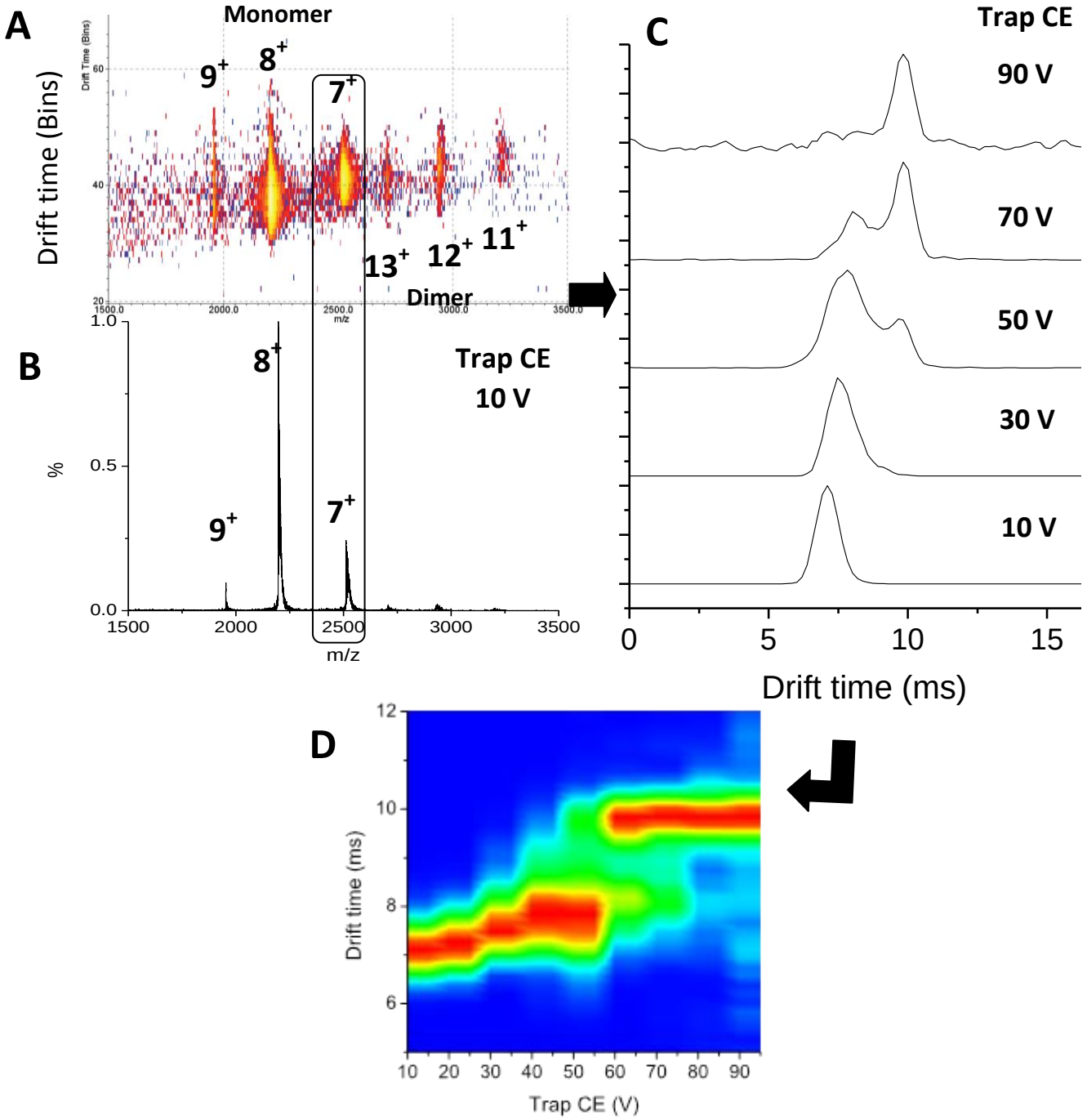


Figure 11: (A) Native Ion mobility spectrum of myoglobin (B) Native mass spectrum of myoglobin (C) Ion mobility drift times of the 7^+ charge state of myoglobin at increasing trap collision energy (D) collision induced unfolding (CIU) plot of 7^+ myoglobin.

Collision-induced dissociation (CID) is a fragmentation technique that breaks the weakest bonds (i.e. peptide bonds) as the analyte collides with the neutral gas (usually N_2 or Ar). When CID is applied on peptides the amide bonds in the peptide backbone are broken, resulting in the formation of b and y fragment ions (**Fig. 14**). Based on the detected masses of these fragment ions it is possible to identify functional groups or amino acid sequence. Such fragmentation patterns can be used to identify molecules of interest [30]. In protein complexes, the weakest bonds are usually the noncovalent bonds between subunits. The charge speeds up this process as the protein unfolding makes it easier to dissociate a subunit. Typically, the smallest subunit dissociates first, resulting in asymmetric dissociation. **Figure 12** displays a demonstration of CID applied on the protein myoglobin. A single charge state (7^+) is selected in the quadrupole and the MS/MS spectrum is monitored as the trap collision energy is increasing. At 50 V the heme group dissociates from myoglobin and at this point CID fragmentation is also observed. **Figure 13** shows the absolute (TIC) intensities of the observed MS peaks. When comparing the CID profile to CIU profile in **Figure 11D**, it is noticed that the collision-induced unfolding, dissociation starts to occur at 50 V.

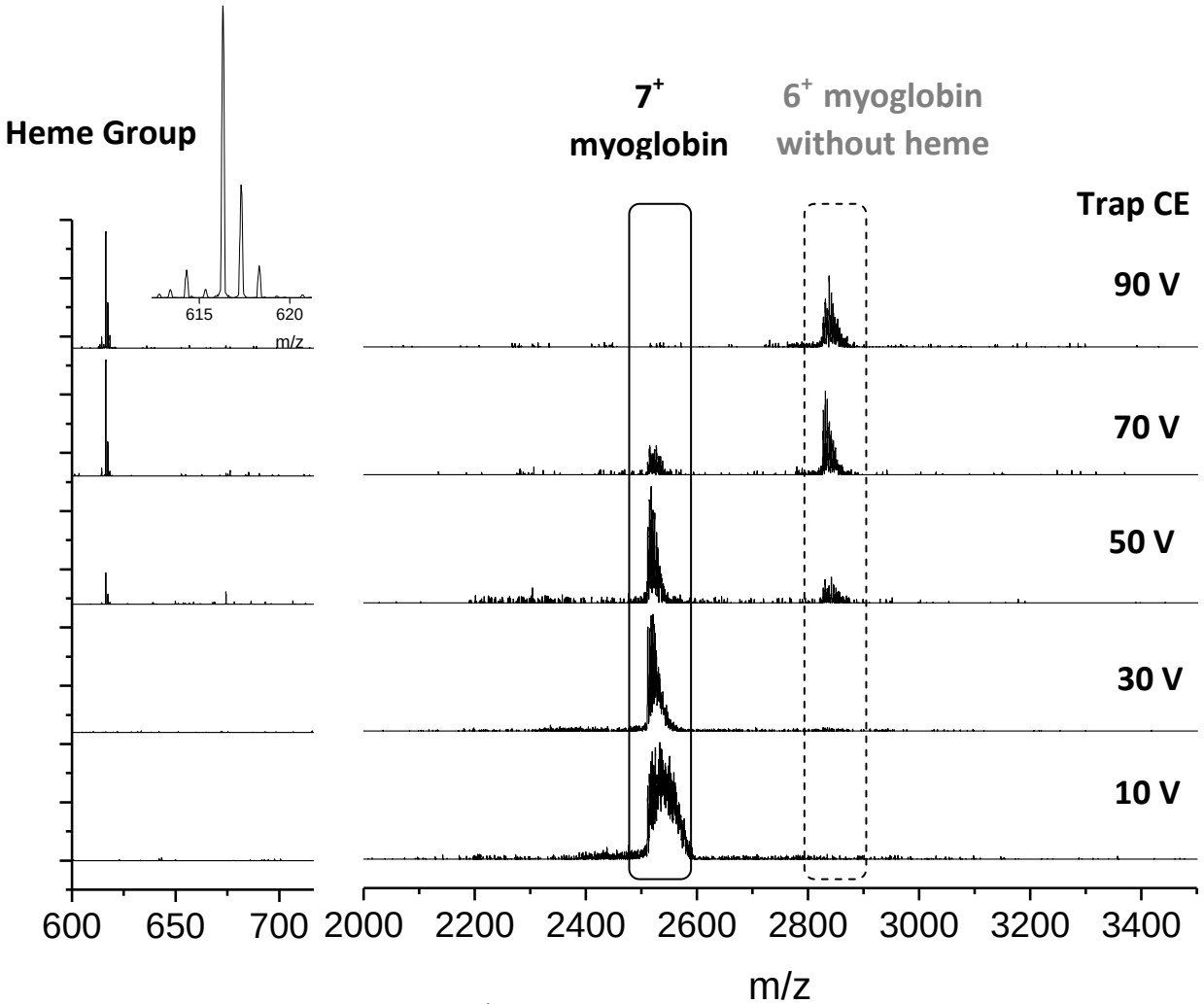


Figure 12: MS/MS spectra of the 7^+ charge state of myoglobin (17 kDa) at increasing trap voltages resulting in CID. The insert shows the zoomed in mass spectrum, displaying the dissociated heme group (616.3 Da)

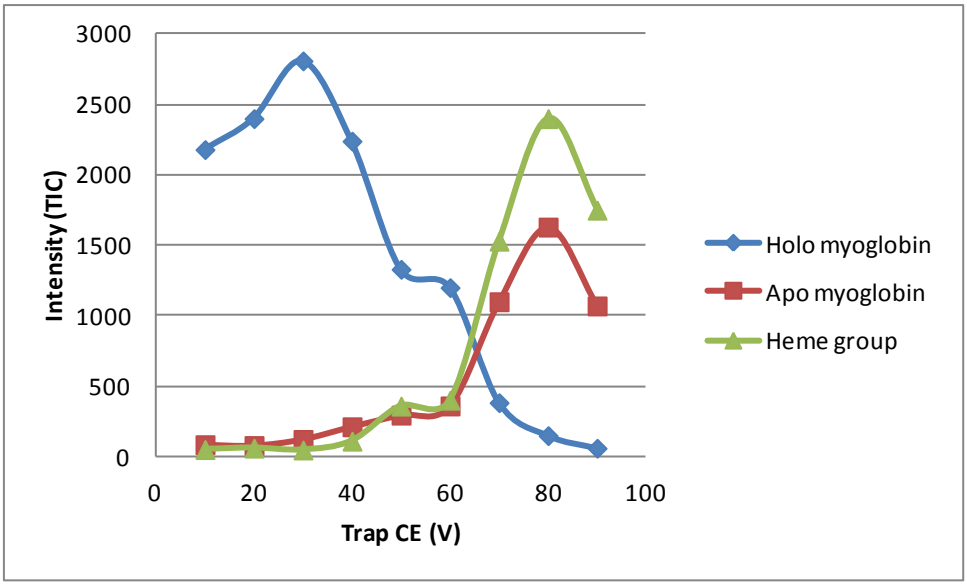


Figure 13: TIC Intensities (peak height) of the observed peaks in Fig. 12 at increasing trap voltages.

2.6 Top-down electron-transfer dissociation (ETD)

In mass spectrometry, electron-transfer dissociation (ETD) is a rapid fragmentation method which is now routinely used for sequencing of highly charged peptides. Conventional fragmentation methods such as collision-induced dissociation (CID) have the disadvantage that the PTMs are frequently lost during fragmentation, making it more difficult to determine the exact locations of the PTMs within the peptide. CID preferentially cleaves the peptide backbone at the weakly bound peptide bond resulting in b- and y-type fragment ions. On the other hand, ETD cleaves the NH-C_α bond resulting in c- and z-type ions. ETD leaves most PTMs intact and provides more extensive sequence coverage for proteins (**Fig. 14**). Furthermore, it leaves contacts made via side chains untouched, i.e., can be used to cut up a protein while the higher-order structure is preserved. ETD uses radical anions to cleave along the peptide backbone. In practice these anions are formed from compounds such as 1,4-dicyanobenzene or fluoranthene [31] by means of a glow discharge, i.e., a two-electrode device filled with the reagent gas across which voltage is applied to form, in this case, the radical anions. When these radical anions interact with the multiply charged protein sample c- and z- fragment ions can be formed, as displayed in **Figure 15,** and are analysed by mass spectrometry.

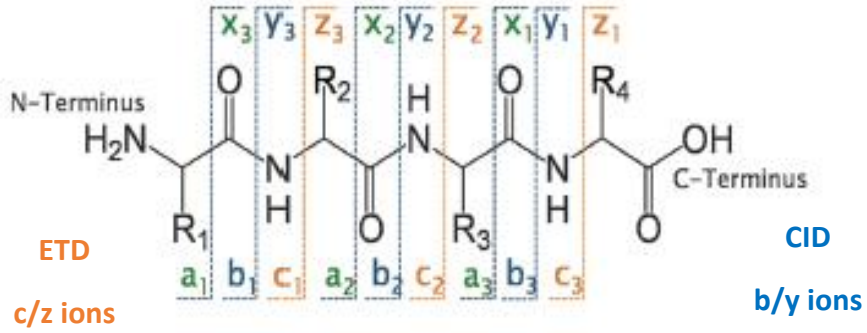


Figure 14: Illustration of b- and y-type ions produced by CID, compared with the c- and z-type ions produced by ETD. Figure adapted from [32].

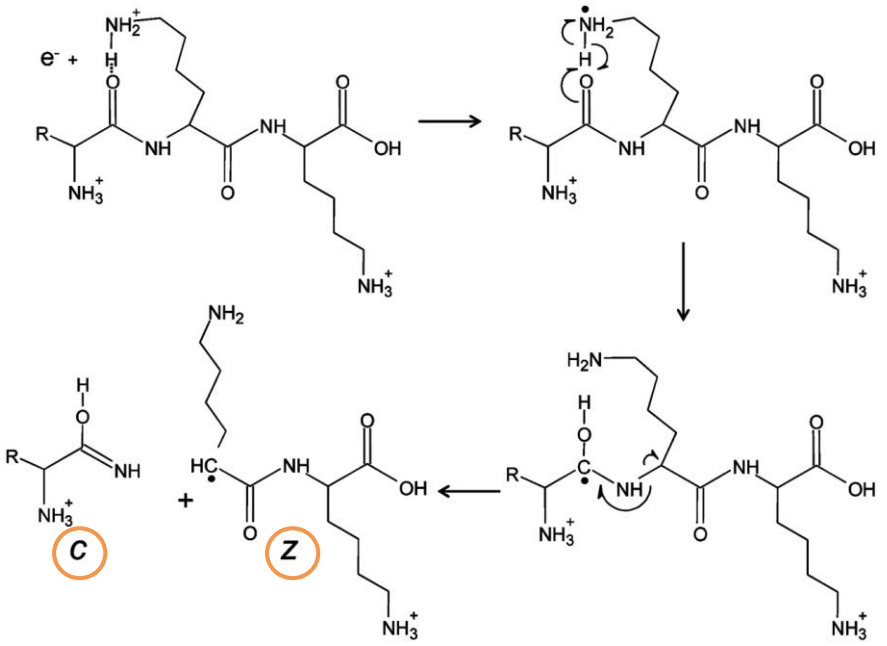


Figure 15: ETD fragmentation scheme, production of c and z fragment ions. Figure adapted from [33].

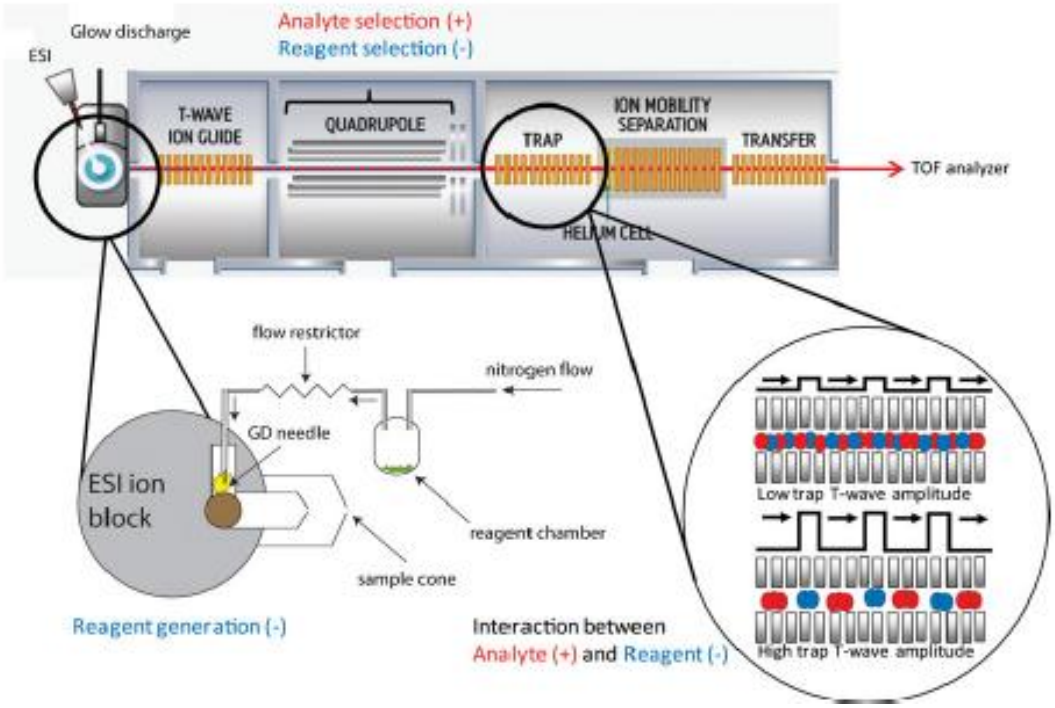


Figure 16: Schematic illustration of a Q-TWIMS-TOF instrument (Synapt G2) in which ETD has been implemented. The ETD reagent anions are generated by glow discharge. By controlling the traveling-wave parameters in the trap cell, the time and the degree of ion-ion interaction between analyte and reagent can be controlled. Figure from [34].

Figure 16 shows how the ETD process can be implemented on a Q-TWIMS-TOF instrument. Radical anions (i.e. ETD reagent ions) are generated in the source via a glow discharge between the sample cone and the extraction cone. The ETD reagent is carried by a nitrogen flow to the glow discharge needle. The polarity of the source up to and including the entrance of the trap cell are continuously switched. This way the trap cell is alternatively filled with ETD reagent anions and analyte cations. The travelling-wave height and velocity can determine the degree of ion-ion interaction as well as the residence time of the anions in the trap cell. The gas pressure in the trap also influences the ETD reaction. This fragmentation technique works best on long peptides or entire proteins, due to the influence of the higher charge state (z>2) on the ETD reaction. As a result, the ETD technique is ideal for top-down fragmentation, which implies the analysis of intact proteins in order to fragment the amino acids that are accessible. Usually this technique is performed on intact and denatured proteins to have better sequence coverage and analyse the PTMs which get lost during CID. However, top-down ETD can also be performed on native proteins. This top-down approach can give access to properties such as conformational states and protein interactions. Research has shown that top-down ETD can allow native surface mapping of proteins and non-covalent protein complexes. However as the charge states are much lower, the fragmentation efficiency and sequence coverage is reduced [21]. Lermyte *et al.* have applied top-down native ETD on alcoholdehydrogenase (ADH) (**Fig. 17**) and noted that only fragmentation of the surface-exposed amino acids was observed. This way they could correlate the data with the solvent–exposed surface area of the protein. However to release the resulting fragments, supplemental activation was necessary. This supplemental activation was controlled by tuning the travelling-wave trap and transfer cell.

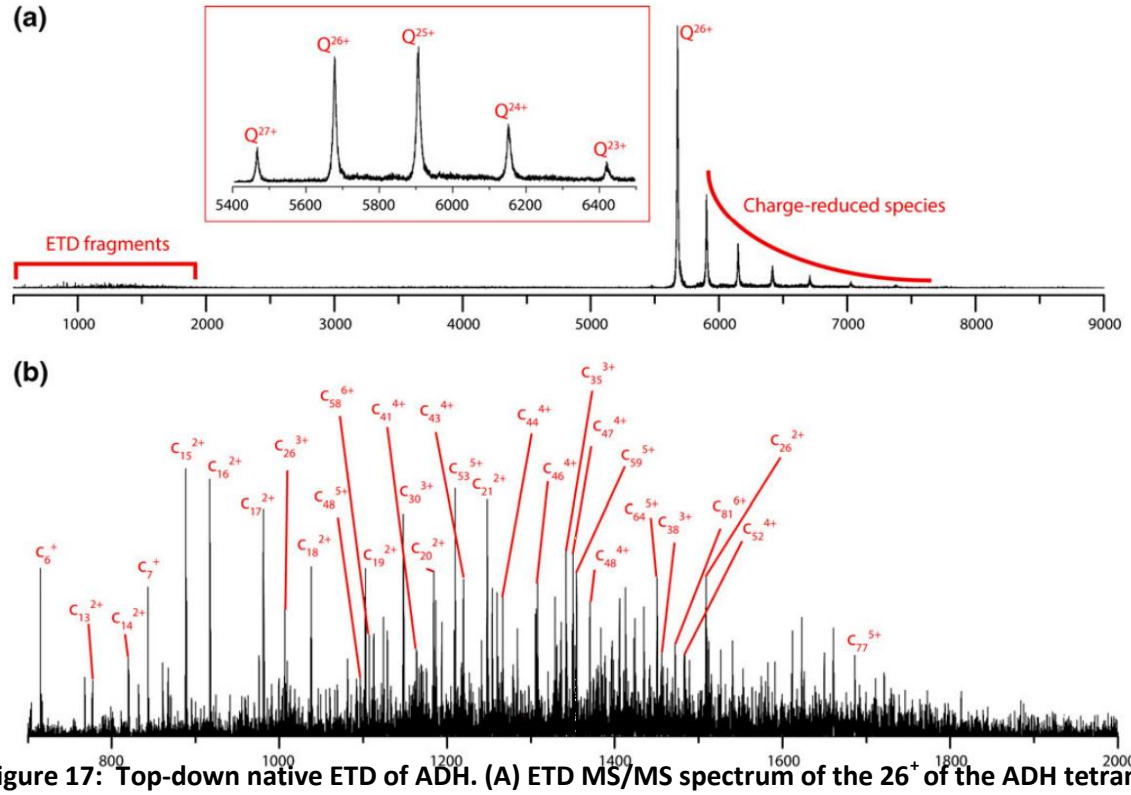


Figure 17: Top-down native ETD of ADH. (A) ETD MS/MS spectrum of the 26$^+$ of the ADH tetramer. The inset shows the full native spectrum of ADH prior to ETD. (B) Low m/z range of the MS/MS spectrum displayed in (A) where the ETD fragments can be observed. Figure from [34].

Top-down denaturing ETD has also been applied on intact denatured antibodies. Tsybin *et al.* have performed top-down ETD on a murine IgG1 using a high resolution Q-TOF mass spectrometer (**Fig. 18**), resulting in mostly z-ion fragments from C-terminal part of the antibody heavy chain [35]. However, when performing this technique on intact denatured antibodies only a limited coverage (~30 %) was detected. Therefore a middle-down approach has also recently been studied by first performing a limited digestion [36][37][38]. With this approach the glycosylation pattern of the antibody could be investigated [39].

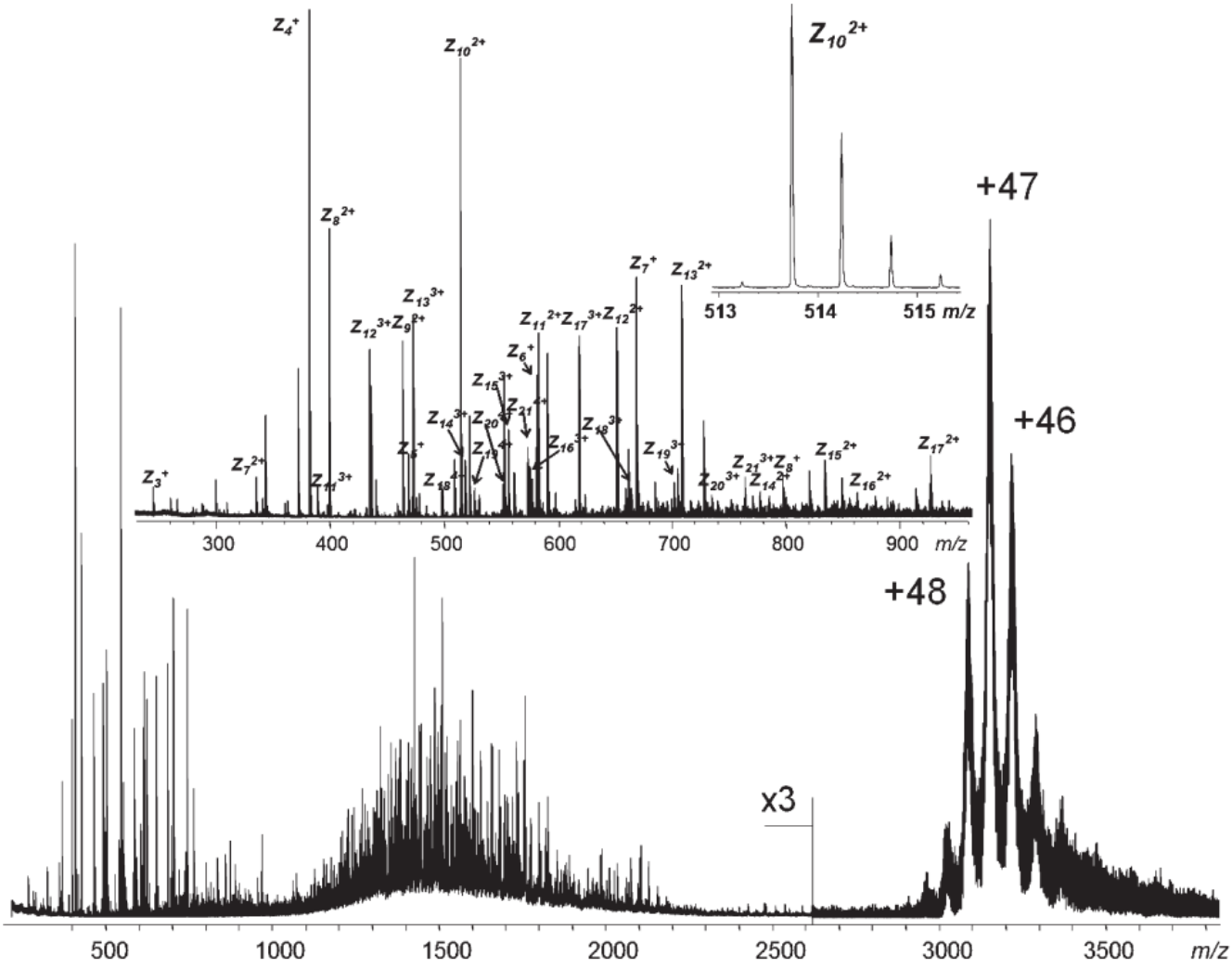


Figure 18: Tsybin *et al.* have performed top-down ETD of intact denatured murine IgG1 using LC-ESI-QTOF-MS/MS inset (250-950m/z shows the ETD fragment corresponding to mainly z-ion fragments of the C-teminal part of the heavy chain. The inset 513-515 m/z shows a typical isotopic distribution of doubly charged product ion, demonstrating the high resolving power achieved. The data was summed from 10 consecutives LC MS/MS runs Figure from [35]

The top-down native ETD protocol has not yet been optimised for intact antibodies. In this thesis it is shown that the sequence coverage is even more limited compared to intact denatured top-down MS. The native top-down ETD method does need more improvement. By optimising the combination of the top-down approach with native MS, information can be yielded on protein structure, interactions and identity in one single experiment.

2.7 Hydrogen-deuterium exchange (HDX)

The structural dynamics of a protein can influence its behavior. Whether the protein is more tightly folded or more unfolded due to the environmental conditions, the conformational change can alter the resulting function of the protein. There are a range of scientific questions about biotherapeutic protein dynamics, conformation and interactions. Hydrogen/deuterium exchange (HDX) MS can help answer these questions about the higher order structure of biotherapeutic proteins [40,41,42].

This technique relies on the fact that when a protein is exposed to deuterated water (D_2O), a rapid amide hydrogen/deuterium exchange is induced in disordered regions that lack stable hydrogen-bonding [42]. Proteins which are tightly folded would be much more protected from HDX, resulting in slow exchange. In other words, a higher exchange level indicates that the protein has a more unfolded or more flexible conformation than the protein with a lower exchange level. The HDX protection can come from the secondary protein structure or from interactions with ligands or other proteins.

Mass spectrometry is a key technique for measuring the mass differences before and after the isotope exchange. HDX-MS can thus be used to map conformational changes and it can also identify the flexible regions of a protein [43].

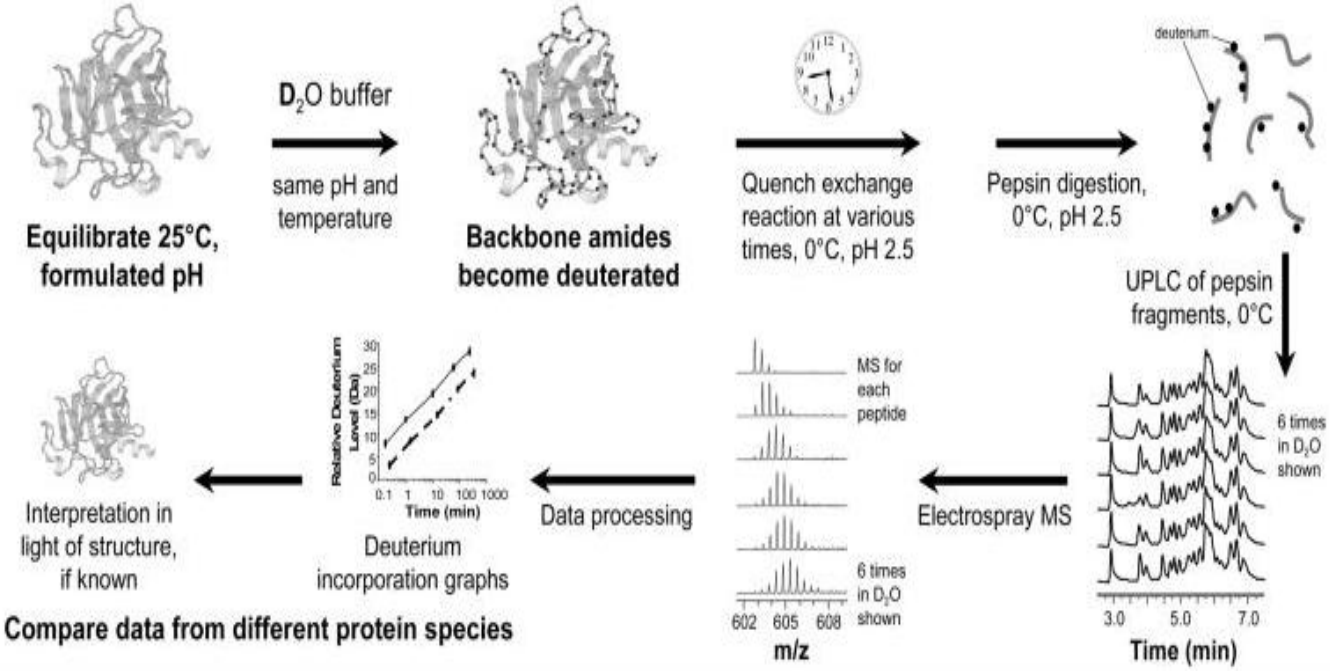


Figure 19: Schematic flow chart of the HDX-MS method. Figure adapted from [49].

Figure 19 shows a schematic representation of the HDX-MS process. The folded protein in solution is exposed to deuterium by diluting the sample with D_2O buffer. The accessible hydrogens exchange with the deuterium and this exchange reaction is quenched at various times by lowering the temperature and pH. The exchange rates are dependent on hydrogen-binding stability, solvent accessibility and protein dynamics. HDX can influence three types of hydrogens in proteins: those in carbon-hydrogen bonds, those in side-chain groups, and those

in the amide backbone bonds. As the exchange rates of hydrogens in carbon-hydrogen bonds are too slow to observe, and those of side-chain hydrogens are too fast (resulting in rapid back exchange when reaction is quenched), only the exchange of the backbone hydrogens can be registered. After the reaction is quenched, a peptide digestion is performed and the peptide fragments are analysed by liquid chromatography-mass spectrometry (LC-MS). The relative deuterium uptake can be monitored and can be processed into deuterium incorporation graphs to interpret the HDX-MS data.

HDX-MS has been applied to biologics to investigate epitope mapping (i.e. interaction between antibody and antigen) and to study the higher-order structure [44]. The higher-order structure of antibodies has been investigated using middle-down HDX-MS; Pan *et al.* have tried this approach by first performing a limited pepsin digestion and reducing the disulfide bonds [45]**.** Houde *et al.* have performed conformational analysis of intact monoclonal antibodies with HDX-MS and they investigated the changes before and after deglycosylation (**Fig. 20-21**) [46,47]. The changes in conformation as a result of modification could be detected based on the change in deuterium uptake [46]. By comparing the HDX data with and without glycosylations they noticed that specific regions in the hinge regions of the IgG displayed a significant change in deuterium uptake (**Fig. 20**).

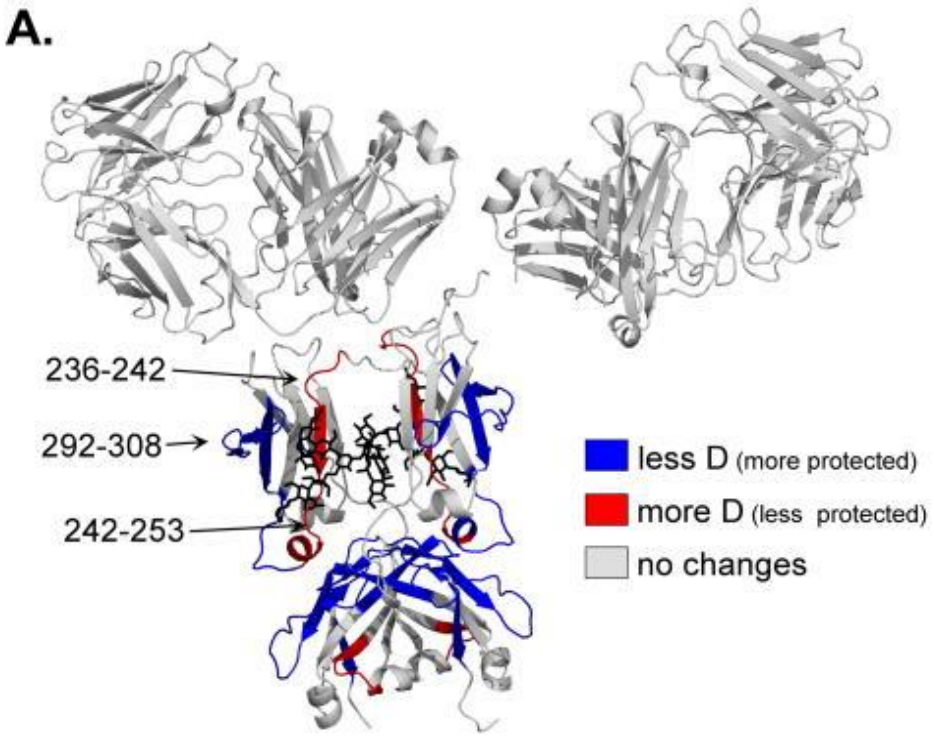


Figure 20: Houde *et al.* have compared the deuterium uptake of IgG1 with and without glycosylation. When the IgG was deglysocylated, the blue regions were protected from exchange (less deuterium) and the red regions were less protected from exchange (more deuterium uptake). Figure from [47].

Houde *et al.* studied the IgG1 deuterium incoporation at the peptide level in order to have a better understanding of the solution dynamics o f the intact IgG1 [47]. **Figure 21** displays the deuterium uptake profiles of IgG1 at 10 seconds, 1, 10 60 and 240 minutes exchange. After 10 seconds (**Fig.21 B**) about 10 % of the IgG (purple regions) was exchanged, leaving the protein mostly protected. After1 and 10 minutes the deuterium uptake increased (green and yellow regions) while some buried regions (black segments) were still protected. After 1 and 4 hours of exchange the deuterium uptake increased but the IgG1 remained about 50-60 % deuterated.

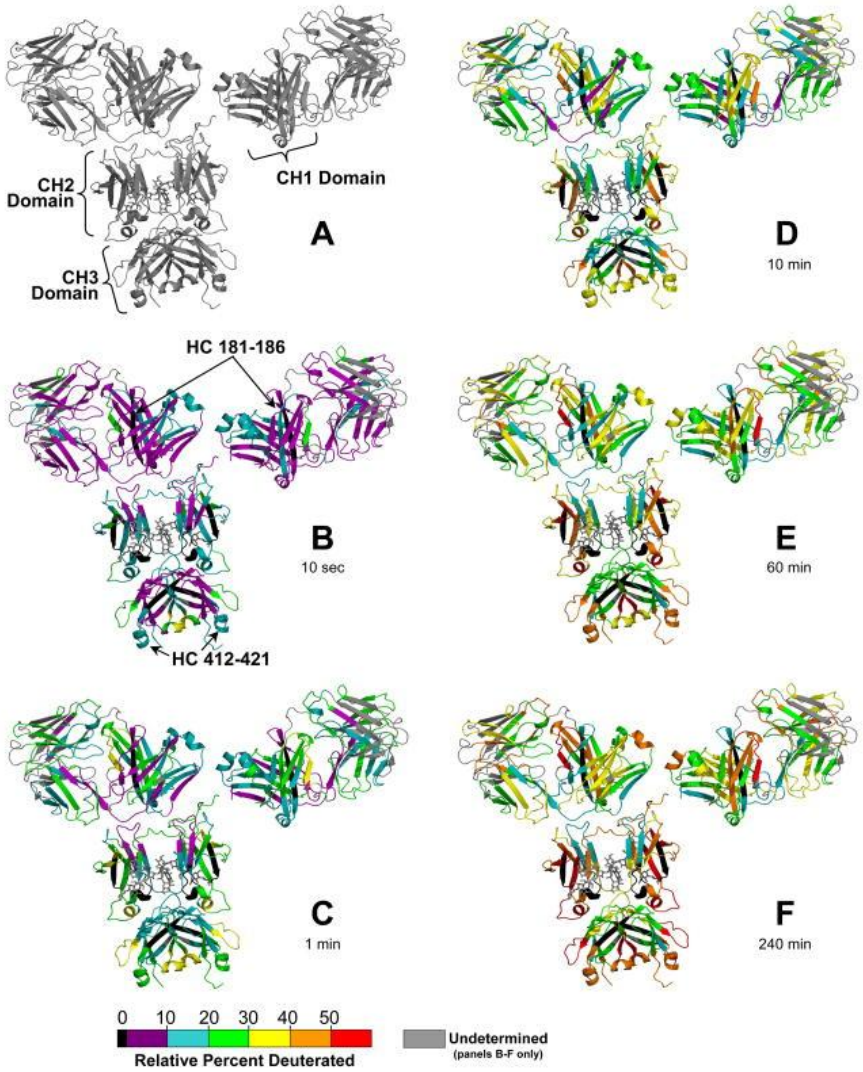


Figure 21: Houde et al. have monitored the global deuterium incorporation of IgG1. (A) The model structure of IgG1 is a model of the solved crystal structure of IgG1 Fab (PDB 3FZU) merged with the IgG1 B12 structure of the Fc region (PDB 1hzh) (100 % sequence homology) The relative percent deuterium incorporation is shown at 10 seconds, 1 minute, 10 minutes, 60 minutes and 240 minutes (B-F, respectively). Figure from [47].

HDX-MS has enabled the development of structure-activity relationships (SAR) by correlating the HDX-MS profiles of large numbers of ligands with their functional outputs [49]. HDX-MS will continue to play an important role in understanding protein structure-activity relationships.

2.8 Fast photochemical oxidation of proteins (FPOP)

Protein foot-printing is an approach to determine structural changes in proteins that occur due to binding, aggregation or other perturbations. When these perturbations occur, the solvent accessibility of certain regions of the protein (e.g., binding interfaces) can be modified. These structural changes can be probed by chemical reactions that happen more slowly when the region becomes less solvent-accessible [50].

Fast photochemical oxidation of proteins (FPOP) is a protein foot-printing MS-based method whereby solvent-accessible amino-acid residues are covalently labeled by oxidation with hydroxyl radicals [51]. As opposed to HDX-MS which is a foot-printing technique that labels the peptide backbone, the FPOP method labels the amino acid side chains of the protein and these labels are covalent, rather than subject to back-exchange.

In FPOP a pulsed laser is used to photolyse added hydrogen peroxide to generate OH-radicals in situ [52]. These OH-radicals interact with the protein sample and oxidise the solvent-accessible amino-acid residues. 14 of the 20 amino acids can be modified by hydroxyl radicals, potentially covering nearly 70% of the sequence of a typical protein [53].

The main advantages of this method are the irreversible nature of the labeling and the high reactivity of the radicals with the amino acid side chains [51]. In addition, the hydroxyl radicals have solvent properties similar to water (in terms of size and mobility), implicating the extent of oxidation depends directly on solvent accessibility [54].

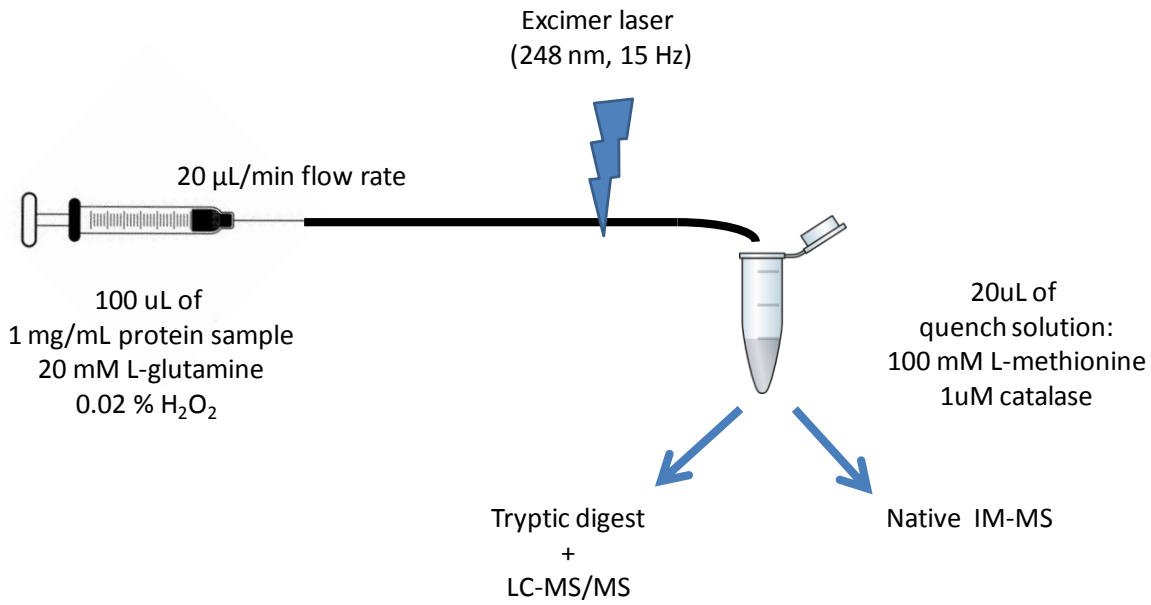
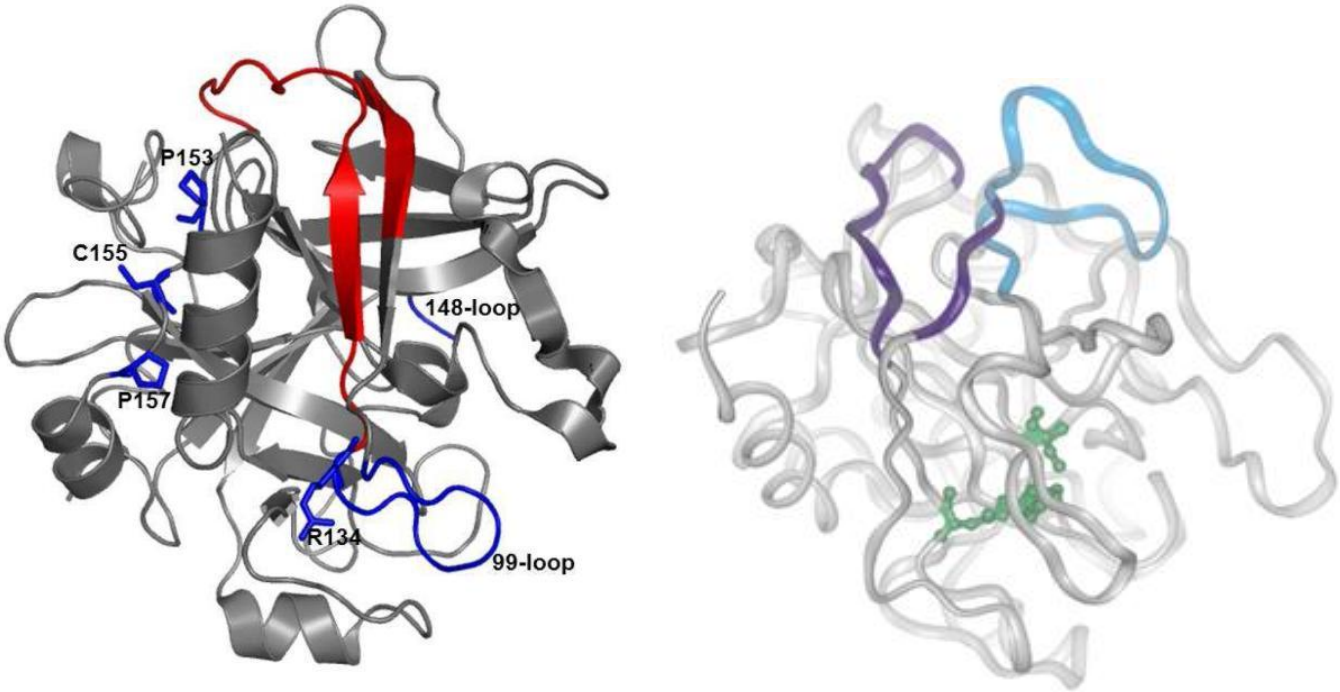


Figure 22: Schematic illustration of the FPOP method.

Figure 22 displays a schematic illustration of how the FPOP experiment works. Hydrogen peroxide is added to the protein sample immediately prior to the infusion through a capillary, after which the sample is irradiated with a pulsed laser. The generated hydroxyl radicals react

with the solvent accessible amino acids side chains and typically result in +16 mass additions [55]. Radical scavengers such as glutamine are added to reduce the hydroxyl radical lifetime to approximately 1 µs or low µs range [51,56]. This way the labeling time is faster than most protein (un)folding reactions. After the labeling reaction the sample is collected in a tube containing methionine and catalase to quench the excess hydrogen peroxide [56]. After digestion of the labeled protein, the sample can be analysed with LC-MS and the data can be compared to a control sample to investigate the differences in labeling. The labeled sample can also be studied without digestion in a top-down approach by using native MS.

FPOP has been applied in several studies to investigate epitope mapping and the data correlated well with data from crystal structures or HDX-MS (**Fig.23**) [57,58,59]. Jones *et al*. have used FPOP to map the epitope for thrombin, a protein involved in blood clotting. By comparing the FPOP data of thrombin with and without the bound antibody the interaction site could be elucidated [59]. Ortiz *et al.* have done a similar research but with HDX-MS and their research provided similar results [57]. Furthermore, FPOP has been used to investigate conformational states adopted during protein folding [55]. Zang *et al.* have investigated the FPOP approach on monoclonal antibodies to characterize local conformational changes and to reveal the solvent accessibility in the regions that would bind to the antigen [60]. Because of the fast and irreversible nature of this labeling technique, FPOP can be used to measure solvent accessibility and locate dynamic or flexible regions of a protein.



FPOP data for epitope mapping

HDX-MS data for epitope mapping

Figure 23: Epitope mapping of thrombin with FPOP and HDX-MS. (A) Jones *et al.* used FPOP to map the epitope region (i.e. where thrombin would bind to the anti-thrombin antibody) with the epitope in red and the loop regions that showed an increase in modification in blue. (B) Ortiz *et al.* mapped the epitope region of thrombin using HDX-MS (dark blue highly protected, light blue slightly protected). Figures from [57, 59].

2.9 Overview of the toolbox of MS-based methods

The set of methods discussed in this chapter can provide a wide range of information on the structural properties of proteins. **Table 1** shows an overview of the techniques and the main outcomes that can be derived from the experimental data.

Table 1: Overview of the MS-based methods used in this thesis

	Native IM-MS	**CIU & CID**	**Top-down ETD**	**HDX-MS**	**FPOP**
Techniques	m/z & IM drift time	Unfolding & dissociation pattern	Sequencing (native or denatured)	Labels backbone hydrogens	Labels amino acid side chains
Properties	Molecular weight and stoichiometry	Stability	Solvent accessibility	Solvent accessibility & secondary structure	Solvent accessibility
	Collision cross section	Stoichiometry	Structure	Dynamics	Dynamics
	Conformational properties	Conformational properties & Domain architecture	Interactions with target	Interactions with target	Interactions with target

2.9.1 Optimisation of methodology

Analytical techniques and methodologies need to be optimised in order to deliver accurate results. For each technique discussed in this thesis the optimisation is applied on different aspects of the experiment, e.g., sample preparation, instrument parameters and data analysis. Sample preparation is an important step for all analytical experiments as the type of buffer, the concentration and pH can influence the experimental results.

The instrument parameters also influence the proper analysis of the measured samples.

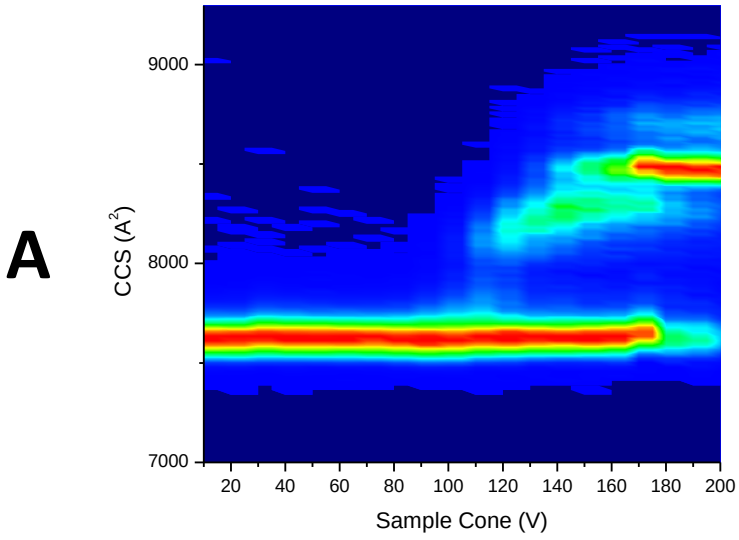
In native mass spectrometry in particular the key instrument parameters to tune are:

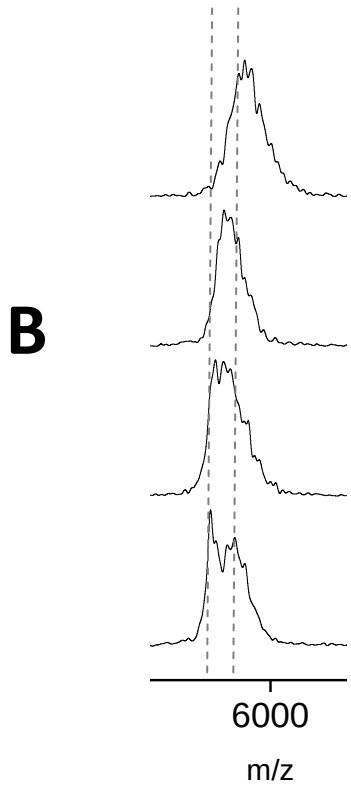
- The capillary voltage of the sample needle which can enhance or suppress ion density
- The sampling and extractor cone voltages which help draw the ions into the vacuum
- The nano-flow backing pressure which helps generate a spray mist of charged droplets
- The trap collision energy (additional declustering and ion transport)
- The transfer collision energy (additional declustering and ion transport)
- The trap bias (additional declustering and ion transport)
- The vacuum pressures inside the instrument (collisional heating or cooling)

When ion mobility is coupled to mass spectrometry extra key instrument parameters need to be properly tuned:

- The MS parameters as mentioned above
- The IMS gas flow
- The IMS wave velocity
- The IMS wave height

When using a high wave IMS wave height and low wave IMS velocity, internal energy can build up in the ions, therby activating (unfolding or dissociating) them, at a given IMS pressure. Thus these parameters need to be adjusted to avoid activation.





Sample cone (V)	Exp. Mass (Da)	Δ (Da)
30	149,030.00	
80	148,573.50	- 456
130	148,307.75	-723
180	148,157.25	-873

Figure 24: (A) The influence of the sampling cone voltage on the collision cross section of the antibody, trastuzumab (25^+ charge state). (B) Mass spectra of 25^+ at increasing sample cone voltages.

In **Figure 24** the CCS value of the 25^+ charge state of an antibody (trastuzumab) is monitored in function of increasing sampling cone voltage. **Fig. 24A** shows that the first transition starts at about 100 volts. Therefore, to keep the measurement as native as possible the sample cone value needs to stay under 100 V. In **Fig. 24B** the mass spectra of the 25^+ charge state is shown at different sampling cone voltages. As the voltage increases, the peak shifts to a lower m/z as adducts are removed. In this thesis we chose to keep the sample cone voltage at 30 V, to avoid unfolding and make sure that adducts and PTMs remain intact. With the settings used here, in general the sample cone needs to stay under 100 V for native MS, so it is possible to increase the sample cone voltage to e.g. 80 V to improve mass resolution by increased declustering, if needed.

For CIU and CID experiments IM-MS parameters also need to be properly tuned before increasing the trap collision energy in a step-wise manner.

The top-down ETD parameters that need to be tuned are:

- The discharge current
- The source temperature
- The make-up gas flow
- The desolvation temperature
- The MS parameters as mentioned above

The key parameters to optimise for HDX-MS:

- The incubation time after hydrogen-deuterium exchange (labeling time)
- The flow rate and gradient used when the samples are separated on the UPLC column
- The digestion performed on the pepsin column
- The full sequence coverage of the protein sample
- The MS parameters as mentioned above

The key parameters to optimise for FPOP:

- The frequency of the laser used to generate the hydroxyl radicals
- The capillary flow rate
- Incubation and quenching times (labeling times)
- The digestion performed after labeling
- LC parameters (flow rate, gradient, right column)
- MS parameters as mentioned above

The data analysis and interpretation is also an important factor when handing the outcome of these experimental techniques. Calibration of the instruments is also necessary to obtain accurate results.

2.10 Sources

[1] S.J. Hyung, B. T. Ruotolo, Integrating mass spectrometry of intact protein complexes into structural proteomics, Proteomics 12, issue 10 (2012), 1547-1564
[2] M. Tassi, J.De Vos, S. Chatterjee, F. Sobott, J. Bones,S.Eeltink, Advances in native high-performance liquid chromatography and intact mass spectrometry for the characterization of biopharmaceutical products, Journal of Separation Science, 2017,1-20
[3] R. Romero-Gonzalez, A. Garrido Frenich, Application in high resolution mass spectrometry, Elsevier, 2017, 23
[4] L. Konermann,E. Ahadi, A. D. Rodriguez, S. Vahidi,Unraveling the Mechanism of Electrospray Ionization, Analytical chemistry, 2013,85,2-9
[5] E. Ahadi, L. Konermann, Modeling the behavior of coarse-grained polymer chains in charged water droplets: implications for the mechanism of electrospray ionization JPhys Chem B, 2012, 116, 104-12
[6] A.Konijnenberg, A. Butterer, F. Sobott, Native Ion Mobility-Mass Spectrometry and Related Methods in Structural Biology, Biochimica et Biophysica Acta (BBA)- Proteins and Proteomics, Volume 1836, Issue 6, 2013, 1239-1256
[7] W. Jiskoot, D. Crommelin, Methods for structural analysis of protein pharmaceuticals, Springer science & business media, 2005, 437
[8] J. A. Loo, Electrospray ionization mass spectrometry: a technology for studying noncovalent macromolecular complexes, International Journal of Mass Spectrometry, Volume 200, Issues 1-3, 2000, 175-186
[9] I.A. Kaltashov, A. Mohimen, Estimates of protein surface areas in solution by electrospray ionization mass spectrometry, Anal. Chem. 2005, 77, 5370-9
[10] R. Grandori, Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry, 2003, Journal of mass spectrometry, 38, 11-15
[11] S. K. Chowdhury, V. Katta, B. T. Chait, Probing conformational changes in proteins by mass spectrometry, J. Am. Chem. Soc., 1990, 112, 9012-9013
[12] Z. Hall, C.V. Robinson, Do charge state signatures guarantee protein conformations? J Am Soc Mass Spectrom, 2012, 23,1161–1168
[13] T. Meyer, X. de la Cruz, M. Orozco, An atomistic view to the gas phase proteome, Structure, 2009, 17, 88-95
[14] K. Barylyuk, R. M. Balabin, D. Grunstein, R. Kikkeri, V. Frankevich, P. H. Seeberger, R. Zenobi, What Happens to Hydrophobic Interactions during Transfer from the Solution to the Gas Phase? The Case of Electrospray-Based Soft Ionization Methods, J Am Soc Mass Spectrom, 22, 1167-1177
[15] G. Terral, A. Beck, S. Cianferani, Insights from native mass spectrometry and ion mobility-mass spectrometry for antibody and antibody-based product characterization, Journal of Chromatography B, 2016, 1032, 79-90
[16] N. J. Thompson, S. Rosati, A. J .R. Heck, Performing native mass spectrometry analysis on therapeutic antibodies, Methods, 2014, 11-17
[17] J. Marcoux, T. Champion, O. Colas, E. Wagner-Rousset, N. Corvaïa, A. Van Dorsselaer, A. Beck, A. Cianferani, Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate
[18] Monoclonal Antibody and Related Product Characterization Under Native Conditions Using a Benchtop Mass Spectrometer, Y. Xuan, F. Debaene, J. Stojko, A. Beck, A. Van Dorselaer, S. Cianferani, M. Bromiski, Thermo Fisher application note 597
[19] E. Jurneczko, P.E. Barran, How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase, 2011, 136,20-8
[20] C. Uetrecht, R.J. Rose, E. Van Duijn, K. Lorenzen, A.J.R. Heck, Ion mobility mass spectrometry of proteins and protein assemblies, Chem Soc Rev, 39 (2010) 1633–1655.
[21] S. J. Allen, K. Gilles, T. Gilbert, M. Bush, Ion mobility mass spectrometry of peptide, protein, and protein complex ions using a radio-frequency confining drift cell, 2016, 141, 884-891

[22] K. Tang, A.A. Shvartsburg, H. Lee, D. C. Prior, M. A. Buschbach, F. Li, A. V. Tolmachev, G. A. Anderson, R. D. Smith, High-Sensitivity Ion Mobility Spectrometry/Mass Spectrometry Using Electrodynamic Ion Funnel Interfaces, Anal. Chem, 2005, 77, 3330-3339
[23] F. Lanucara, S. W. Holman, C. J Gray, C. E. Eyers, The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics, Nature Chemistry, 2014, 281-294
[24] 3 Types of IMS: An At-a-Glance Guide, Owlstone medical, 2017
[25] M.F. Bush, Z. Hall, K. Giles, J. Hoyes, C.V. Robinson, B.T. Ruotolo, Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology, Anal Chem, 2010, 82, 9557–9565
[26] D. Bagal, J.F. Valliere-Douglass, A. Balland, P.D. Schnier, Resolving disulfide structural isoforms of IgG2 monoclonal antibodies by ion mobility mass spectrometry, Anal. Chem., 2010 82, 6751-5
[27] A. Beck, F. Debaene, H. Diemer, E. Wagner-Rousset, O. Colas, A. Van Dorsselaer, S. Cianferani, Cutting-edge mass spectrometry characterization of originator, biosimilar and biobetter antibodies, J Mass Spectrom, 2015, 50 285-297
[28] K. J. Pacholarz, M. Porrini, R. A. Garlish, R. J. Burnley, R. J. Taylor, A. J. Henry, P. E . Barran, Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion-Mobility Mass Spectrometry and Molecular Modeling, Angew Chem Int Ed Engl, 2014, 7765-9
[29] Y. Zhong, L.Han, B. T. Ruotolo Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution
[30] Y. Yefremova, F. Teresa, J. Melder, B. D . Danquah, K. F. M Opuni, C. Koy, A. Ehrens, D. Frommholz, H. Illges, K. Koelbel, F. Sobott, M. O. Glocker, Apparent activation energies of protein–protein complex dissociation in the gas–phase determined by electrospray mass spectrometry, Analytical and Bioanalytical Chemistry, 2017, 409, 28, 6549-6558
[31] J. Martens, G. Berden, J. Oomens, Structures of Fluoranthene Reagent Anions Used in Electron Transfer Dissociation and Proton Transfer Reaction Tandem Mass Spectrometry, Anal. Chem, 2016, 88, 6126-9
[32] J. S. Brodbelt, Photodissociation mass spectrometry: new tools for characterization of biological molecules, Chem. Soc. Rev, 2014, 43, 2757-2783
[33] L. Elviri, ETD and ECD Mass Spectrometry Fragmentation for the Characterization of Protein Post Translational Modifications, Tandem Mass Spectrometry - Applications and Principles, 2012, 179-193
[34] F. Lermyte, A. Konijnenberg, J. P. Wiliams, J. M. Brown, D. Valkenborg, F. Sobott, ETD Allows for Native Surface Mapping of a 150 kDa Noncovalent Complex on a Commercial Q-TWIMS-TOF Instrument
[35] Structural Analysis of Intact Monoclonal Antibodies by Electron Transfer Dissociation Mass Spectrometry, Y. O. Tsbin, L. Fornelli, C. Stoermer, M. Luebeck, J. Parra, S. Nallet, F. M. Wurm, R. Hartmer, Analytical Chemistry, 2011, 83,8919-8927
[36] Y. Mao, S. G. Valejat, J. C. Rouse, C. L. Hendricksen, A G. Marshall, Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry, Anal. Chem., 2013, 85, 4423-4246
[37] L. M. Jones, H.Zang, W. Cui, M.L. Gross, Complementary MS Methods Assist Conformational Characterization of Antibodies with Altered S–S Bonding Networks, J Am Soc Mass Spectrom, 2013, 24, 835-45
[38] L. Fornelli, D.Ayoub, K. Aizikov, A.Beck, Y. O. Tsybin, Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry, Anal. Chem., 2014, 86, 3005-3012
[39] B.Q. Tran, C. Barton, J. Feng, A. Sandjong, S. H. Yoon, S. Awasthi, T.Liang, M.M. Khan, D.P.A. Kilgour, D. R. Goodlett, Y. Goo, Glycosylation characterization of therapeutic mAbs by top- and middle-down mass spectrometry , Journal of proteomics, 2016, 134, 93-101
[40] R. Lindner, U.Heintz, A.Winkler, Applications of hydrogen deuterium exchange (HDX) for the characterization of conformational dynamics in light-activated photoreceptors, Front Mol Biosc, 2015, 2-33
[41] B.Suchanova, R. Tuma, Folding and assembly of large macromolecular complexes monitored by hydrogen-deuterium exchange and mass spectrometry, Microb Cell Fact, 2008, 7, 12
[42] L. Konermann, J. Pan, Y. Liu, Hydrogen exchange mass spectrometry for studying protein structure and dynamics, 2011,3, 1224-34

[43] A. Tsirigotaki, R.V. Elzen, P. V. Veken, A.M. Lambeir, A. Economou , Dynamics and ligand-induced conformational changes in human prolyl oligopeptidase analyzed by hydrogen/deuterium exchange mass spectrometry, Sci Rep, 2017, 7, 2456
[44] H. Wei, J. Mo, L. Tao, R. J. Russel, A. A. Tymiak, G. Chen, R. Iacob, J. R. Engen Hydrogen/Deuterium Exchange Mass Spectrometry for Probing Higher Order Structure of Protein Therapeutics: Methodology and Applications
[45] J.Pan, S.Zhang, A. Chou, C.H. Borchers, Higher-order structural interrogation of antibodies using middle-down hydrogen/deuterium exchange mass spectrometry,Chem. Sci., 2016, 7 , 1480-1486
[46] D. Houde, J. R. Engen, Conformational Analysis of Recombinant Monoclonal Antibodies with Hydrogen/Deuterium Exchange Mass Spectrometry, Methods Mol Biol, 2013, 988, 269-289
[47] D. Houde, J. Arndt, W. Domeier, S. Berkowitz, J. R. Engen, Rapid characterization of IgG1 conformation and conformational dynamics by hydrogen/deuterium exchange mass spectrometry, Anal Chem, 2009, 81, 2644-2651
[48] HDX-MS guided drug discovery: small molecules and biopharmaceuticals, Chem. Soc. Rev. 2011, 40, 1224-1234
[49] D. Houde, S. A. Berkowitw, J. R. Engen, The Utility of Hydrogen/Deuterium Exchange Mass Spectrometry in Biopharmaceutical Comparability Studies, J Pharm Sci, 2011, 100, 2071-2086
[50] G. Xu, M.R. Chance, Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics, Chem. Rev. , 2007, 3514-3543
[51] B. C. Gau, J. S. Sharp, D. L. Rempel, M. L. Gross, Fast Photochemical Oxidation of Proteins Footprints Faster than Protein Unfolding, Anal. Chem, 2009; 81, 6563-6571
[52] L. Konermann, B.B. Stocks, T.Czarny, Laminar flow effects during laser-induced oxidative labeling for protein structural studies by mass spectrometry, Anal.Chem, 2010, 6667-6674
[53] J. Chen, W. Cui, D. Giblin, M. L. Gross, New Protein Footprinting: Fast Photochemical Iodination Combined with Top-down and Bottom-up Mass Spectrometry, J Am Soc Mass Spectrom , 2012, 23, 1306-1318
[54] J. G. Kiselar, M. R. Chance, Future Directions of Structural Mass Spectrometry using Hydroxyl Radical Footprinting, J. Mass Spectrom, 2010 45, 1373-1382
[55] A. N. Calabrese, J. R. Ault, S. E. Radford, A. E. Ashcroft, Using hydroxyl radical footprinting to explore the free energy landscape of protein folding. Methods, 2015, 89, 38-44.
[56] Y. Yan, Guodong C. Wei , R. Huang, J. Mo, D. L. Rempel, A. A. Tymiak, M. L. Gross, Fast Photochemical Oxidation of Proteins (FPOP) Maps the Epitope of EGFR Binding to Adnectin, J Am Soc Mass Spectrom, 25, 12, 2084-2092
[57] L. M. Jones, J. B. Sperry, J.A. Carroll, M.L. Gross, Fast Photochemical Oxidation of Proteins for Epitope Mapping, Anal. Chem, 2011, 83, 7657-7661
[58] Zhang, Y., et al., Mapping the Binding Interface of VEGF and a Monoclonal Antibody Fab-1 Fragment with Fast Photochemical Oxidation of Proteins (FPOP) and Mass Spectrometry. J Am Soc Mass Spectrom, 2017, 1-9
[59] A. Baerga-Ortiz, C. A. Hughes, J. G. Mandell, E. A. Komives, Epitope mapping of a monoclonal antibody against human thrombin by H/D-exchange mass spectrometry reveals selection of a diverse sequence in a highly conserved protein, Protein Sci, 2002, 11, 1300-8
[60] H. Zhang, W. Cui, M. L . Gross, Mass Spectrometry for the Biophysical Characterization of Therapeutic Monoclonal Antibodies, FEBS Lett, 2014, 588,308-317
[61] J.B. Fenn, M. Mann, CK. Meng, SF, Wong, CM Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, Science, 1989, 246, 64-71

Chapter 3

Exploration of the higher-order structure of monoclonal antibodies

3.1 Introduction

Monoclonal antibodies (mAbs) have been dominating the biopharmaceutical market ever since the introduction of therapeutic monoclonal antibodies in 1986, when the FDA approved the therapeutic antibody, muromonab, as a treatment for kidney transplant rejection [1]. The first therapeutic antibodies were made entirely from mouse cells with the hybridoma technique, where myeloma cells were fused with mouse spleen cells immunized with the targeted antigen.

However, as the human immune system would perceive these antibodies as foreign, there was a risk of an immune response against the murine antibodies. In order to reduce the immunogenicity of murine antibodies, chimeric antibodies were generated containing human constant domains and murine variable domains to maintain the specificity (**Fig. 1**). With genetic engineering, scientists could move on to the next step by placing the complementarity-determining regions (CDRs) of a murine antibody onto the variable region of a human antibody, creating humanized antibodies. In 1996 the FDA approved the first chimeric antibody, abciximab, which would reduce the risk of blood clots in patients with cardiovascular disease by binding to a specific receptor on platelets [3]. A year later the first humanized antibody, daclizumab, was approved for the prevention of organ transplant rejection [4].

The next step was the creation of fully human antibodies. One of the most used techniques for the production of human monoclonal antibodies is phage display, where a library of antibodies is expressed on the surface of a bacteriophage, followed by antigen-driven selection and phage propagation in E.coli. Another successful approach to produce human antibodies is transgenic mice technology, where human genes are expressed in mouse strains. In 2002, adalimumab became the first FDA-approved human monoclonal antibody and since then 23 more human mAbs followed with hundreds still in clinical trial. The creation of chimeric, humanized and human antibodies was a major breakthrough and showed impressive results in cancer therapy and auto-immune diseases [5]. However, the use of large cultures of mammalian cells and extensive purifications steps results in extremely high production costs and consequently limits the wide use of these drugs [5].

Another limitation pertaining to drug delivery is the large size of the antibody, since it has to penetrate tissues to reach the target cells. The binding affinity of the antibody to the antigen is also an important factor. The antibodies need to penetrate deep enough into the tumour to efficiently distribute over the target, instead of only tightly binding at the periphery of the tumour. The right balance of tumour targeting and tumor retention depends on factors such as association and dissociation rates [5]. Obtaining the ideal therapeutic antibody is therefore a challenge in antibody engineering.

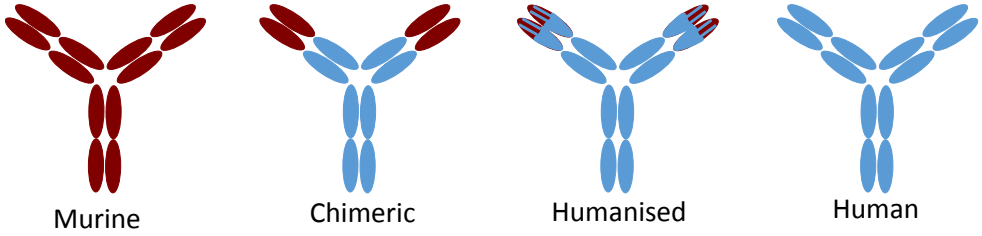


Figure 1: Schematic illustration of murine, chimeric and human antibodies.

3.1.1 Characteristics of a monoclonal antibody

Antibodies or immunoglobulins (Ig) are Y-shaped glycoproteins that are used by the immune system to help identify and destroy disease-causing pathogens (e.g., viruses, bacteria, parasites). An antibody consists of two identical heavy chains and two identical light chains joined together by disulfide bonds (**Fig. 2**). The heavy chains are connected by disulfide bonds in the flexible hinge region. The light chains can be classified based on their sequence as either kappa (κ) or lambda (λ). The Y-shaped protein can be divided into three sections: two antigen binding fragment (Fab) regions and one crystallisable fragment (Fc) region. The Fab regions consist of two variable (V) and two constant (C) domains. The two variable domains form the variable fragment (Fv) and they can bind to the epitope of a specific antigen via the complementarity-determining region (CDR). Since each variable domain consists of three CDRs, a single antibody has twelve CDRs to come in contact with a specific antigen. Most sequence variation is usually observed in this CDR region when comparing antibody sequences. The constant domain of the Fab region acts as a structural framework.

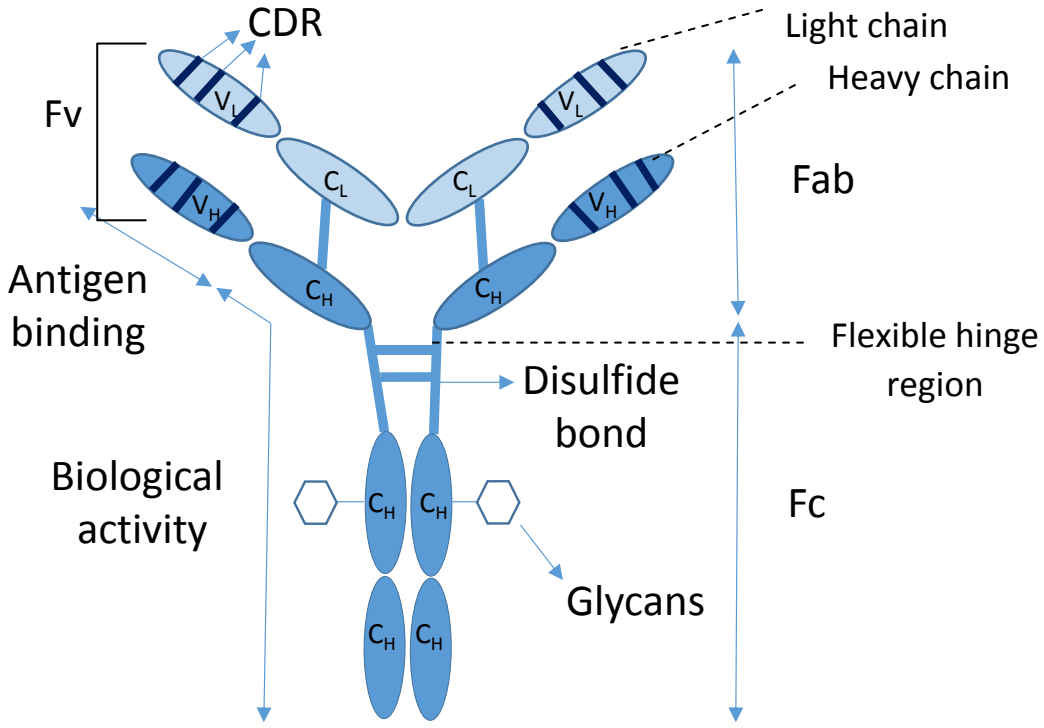


Figure 2: Schematic illustration of an antibody. Variable regions are indicated with V_L (light chain) and V_H (heavy chain). The constant regions are indicated with C_L (light chain) and C_H (heavy chain).

Post-translational glycosylation occurs in the constant region (C_H2) of the heavy chains of the antibody. The oligosaccharides are attached via an N-glycosidic bond to the asparagine residue Asn^{297} [6][7]. The N-glycans are mainly di-antennary complex structures and can influence the Fc region of the antibody [8]. An overview of commonly found antibody glycans is shown in **Fig.3**.

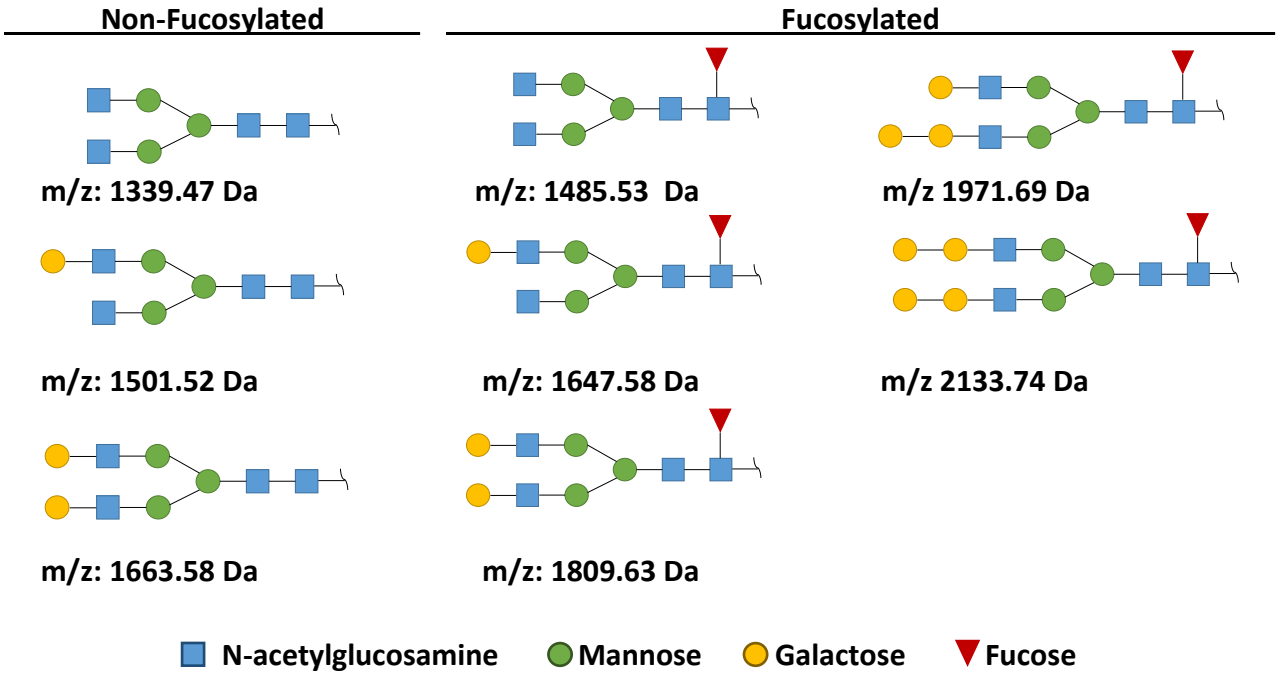


Figure 3: Schematic overview of N-linked glycan structures commonly found on IgG antibodies.

The constant region of the heavy chains determines the isotype or class of the antibody. In mammals, antibodies can be divided into five isotypes: IgA, IgD, IgE, IgG and IgM, based on the type of heavy chain (α, δ, ε, γ, or μ) (**Fig. 4**) [9].The heavy chains differ in amino acid sequence, number of constant domains and hinge structure. α, δ and γ have about 450 amino acids, while ε and μ contain about 550 amino acids. The average length of a light chain is 211-217 amino acids.

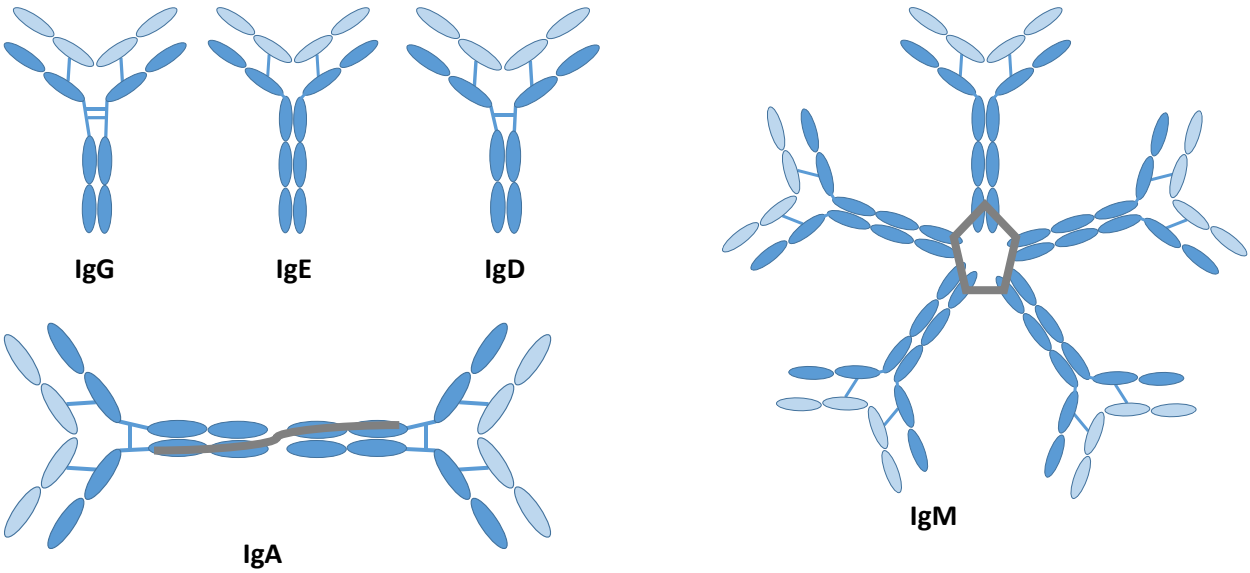


Figure 4: Schematic illustration of antibody isotypes: IgG, IgE, IgD, IgA and IgM.

IgG is the most abundant antibody isotype found in human serum (about 70-85 %). It also has the longest half-life (20-24 days) of the five isotypes [10]. IgGs are monomeric and have a

molecular weight of about 150 kDa. In humans, IgGs can be divided into four subclasses (IgG1, IgG2, IgG3 and IgG4) based on their different hinge regions (**Fig. 5**). IgG1 and IgG4 have 2 inter-chain disulfide bonds in the hinge region, while IgG2 has 4 and IgG3 has 11.

IgE is the least abundant serum antibody and found in the lungs, skin and mucous membranes. It has two additional constant domains in the hinge region compared to IgG. IgEs can bind to allergens and are associated with immediate hypersensitivity e.g., against pollen and peanuts.

IgD represent less than 1 % of the Igs in plasma, but exist in large quantities on the membrane of B-cells [10]. The basic structure of an IgD resembles that of an IgG, only with a more flexible hinge region. The exact function of IgD is still unknown.

IgA represents about 5-15 % of all antibodies in human serum and is predominantly found on epithelial cell surfaces [10]. IgA is mainly monomeric in serum, but in secretions (e.g., saliva, sweat and mucus) it is found as a dimer connected by a peptide chain, the secretory component, synthesised by epithelial cells of the gut.

IgM embodies 5-10 % of the immunoglobulins and the predominant antibody in the primary immune response [10]. It is predominantly found in the lymph fluid and the blood. Compared to IgGs it contains two additional constant domains and is predominantly found as a pentamer [9].

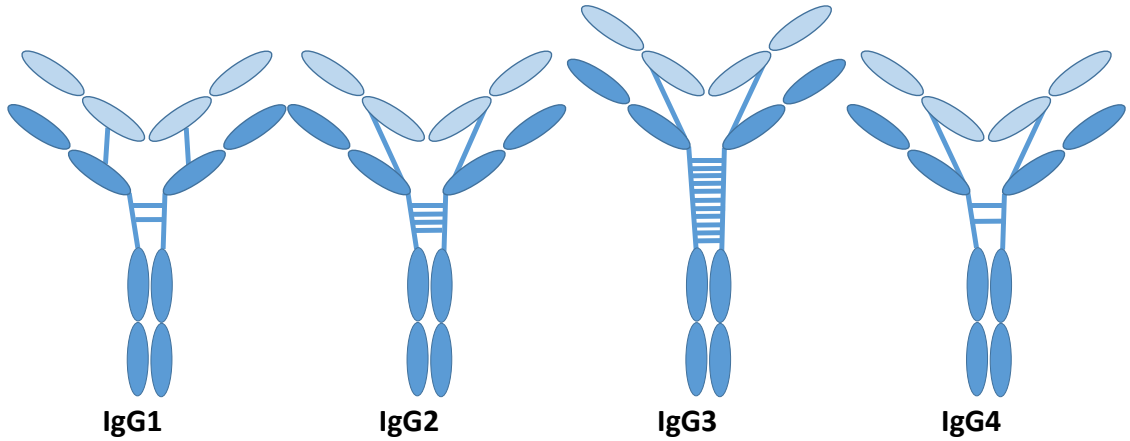


Figure 5: Schematic illustration of the four subclasses of human IgG: IgG1, IgG2, IgG3 and IgG4.

In this thesis, the main focus will be on the biggest class of antibodies, the IgGs. This class of antibodies can be further divided into four subclasses. Each IgG subclass has a unique profile due to their specific hinge region which would also determine the specific nature of their biological effects, i.e. effector functions. Examples of effector functions are antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC) [11]. In ADCC, the antibody binds to Fc receptors on the surface of effector cells, such as macrophages, leading to phagocytosis of the targeted cells. In CDC, the Fc region of the antibody triggers a cascade of reactions at the cell by interacting with Fc receptors, resulting in the killing of the targeted cells. Certain IgGs can also bind to receptors on placental trophoblasts, which results in the transfer of IgGs across the placenta, providing immunity to the fetus. Furthermore, the interaction of IgGs with neonatal Fc receptors (FcRn) also results in a longer serum half-life. The glycosylation patterns of IgGs can also influence the effector functions as the glycans have an effect on the stability of the IgGs [7].

A wide range of IgGs are being investigated for therapeutic applications. By choosing a specific IgG subclass, specific effector functions can also be selected [12]. In the current biopharmaceutical market of antibodies, IgG1 are the majority, due to their potent effector functions and being the most predominant subclass [13]. Subclasses IgG2 or IgG4 are considered when there is a need for a specific cellular activity. No therapeutic human IgG3s have been developed due to the short in vivo half-life of IgG3 compared to other subclasses [14]. There is also an increased likelihood for proteolysis of IgG3, due to their extensive hinge domain [13].

This part of the thesis will focus on full antibodies. Through our collaboration with Johnson & Johnson (BE & US), we received standard IgG1, IgG2 and IgG4.

During our collaboration with the Medicinal Chemistry research group of University of Antwerp (BE) we worked on the humanised IgG1, trastuzumab.

As part of an inter-laboratory study between 25 labs, we characterised three antibodies; NIST (IgG1), MSQ4 (IgG1) and trastzumab (IgG1).

3.1.2 Sources

[1] O. Leavy, Therapeutic antibodies: past, present and future, Nature Review Immunology 2010, 10, 297

[2] J.K.H. Liu, The history of monoclonal antibody development – Progress, remaining challenges and future innovations, Annals of Medicine and Surgery 2014, Vol. 3, Issue 4, 113-116

[3] J. Lefkovits, R.J. Ivanhoe, R.M. Califf, B.A. Bergelson, K.M. Anderson, G.L. Stoner, H.F. Weisman, E.J. Topol, Effects of platelet glycoprotein IIb/IIIa receptor blockade by a chimeric monoclonal antibody (abciximab). Am. J. Cardiol 1996, 77 (12), 1045-51

[4] D. Przepiorka, N.A. Kernan, C. Ippoliti, E.B. Papadopoulos, S. Giral, I Khouri, J.G. Lu, J. Gajewski, A. Durett, K. Cleary, R. Champlin, B.S. Andersson, S. Light, Daclizumab, a humanized anti-interleukin-2 receptor alpha chain antibody, for treatment of acute graft-versus-host disease, Blood, 2000, 95 (1), 83-89

[5] P. Chames, M. Van Regenmortel, E. Weiss, D. Baty, Therapeutic antibodies: successes, limitations and hopes for the future, Br. J. Pharmacol., 2009, 157(2), 220-223

[6] R. Hansen, A. J. Dicksen, R. Goodacre, G. M. Stephens, C. A. Sellick, Rapid Characterization of N-linked Glycans from Secreted and Gel-Purified Monoclonal Antibodies Using MALDI-ToF Mass Spectrometry, Biotechnology and Bioengineering,2010, 107(5),902-908

[7] K. Zheng, C. Bantog, R. Bayer, The impact of glycosylation on monoclonal antibody conformation and stability, MABS, 2011, 3(6), 568-576

[8] M. F. jennewein, G. Alter,The Immunoregulatory Roles of Antibody Glycosylation, Trends in Immunology, 38 (5), 358-372

[9] H.W. Schroeder, L. Cavacini, Structure and Function of Immunoglobulins, J. Allergy Clin. Immunol. ,2010, 125, 41-52

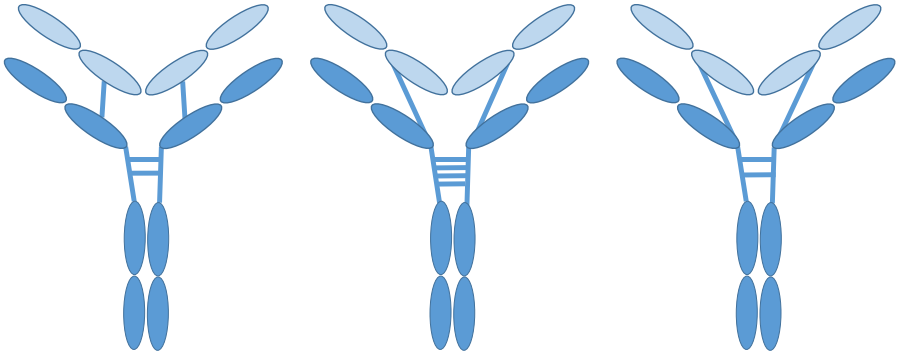
[10] J.M. Cruse, R.E. Lewis, Atlas of immunology, 2010, 3, 248-250

[12] S. Beers, M. J. Glennie, A. L. White, Influence of immunoglobulin isotype on therapeutic antibody function, Blood, 2016,127(9), 1097-1101

[13] V. Irani, A. J. Guy, D. Andrew, J. G. Beeson, P. A. Ramsland, J. S. Richards, Molecular properties of human IgG subclasses and their implications for designing therapeutic monoclonal antibodies against infectious diseases, Molecular Immunology, 2015, Vol. 67, Issue 2, 171-182

[14] R. Braster, S. Grewal, R. Visser, H. K. Einarsdottir, M. van Egmond, G. Vidarsson, M. Bogels, Human IgG3 with extended half-life does not improve Fc-gamma receptor-mediated cancer antibody therapies in mice, PLoS One, 2017, 12 (5)

3.2 Characterisation of standard antibodies



In collaboration with

Dr. Thomas de Vijlder[1]
Dr. Darryl Davis[2]

[1]Johnson & Johnson, Janssen Research and Development, Beerse, BE
[2]Johnson & Johnson, Research and Development, Pennsylvania, US

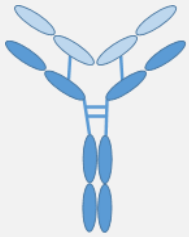
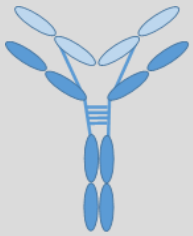
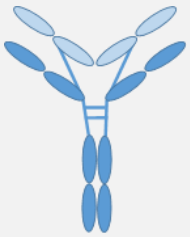
Part of this research was published in

Marco Tassi, Jelle De Vos, Sneha Chatterjee, Frank Sobott, Jonathan Bones, Sebastiaan Eeltink, Advances in native high-performance liquid chromatography and intact mass spectrometry for the characterization of biopharmaceutical products, *Journal of Separation Science*, 2017,1-20

3.2.1 Introduction

In this study we analysed four standard antibodies (two IgG1, an IgG2 and an IgG4) with MS-based methods in order to thoroughly characterise the structure and search for differences between the different IgG subclasses. **Table 1** shows an overview of these four studied antibodies and their respective general structures. The National Institute of Standard and Technology (NIST) has issued a monoclonal antibody reference material [1]. This material was donated to them by MedImmune and then characterised by NIST and collaborators. This reference material is a useful tool to evaluate a range of analytical methods. The NIST IgG1, IgG2 and IgG4 were characterised, together with CNTO 5825, a human monoclonal antibody (IgG1), which we obtained from Johnson & Johnson. Generally, IgGs are almost identical in sequence (90-95 %). This is also displayed in **Fig. 1** where the identical sequence is highlighted. I looked at these standard antibodies with native ion mobility mass spectrometry to compare their structural properties. In contrast, we also monitored the unfolding pattern of the three IgG subtypes. I performed a papain digest of the antibodies to study the resulting Fc and Fab regions of the IgG. The global structures of the antibodies were also studied with fast photochemical oxidation of proteins (FPOP) to find out the typical amount and sites of oxidations. Finally I applied size-exclusion chromatography-mass spectrometry (SEC-MS) to the antibodies to evaluate the analysis of an antibody mixture and to study them in complex formulations.

Table 2: Overview of the studied antibodies, IgG1, IgG2 and IgG4.

Antibody	Type	Isotype	Antigen	
NIST IgG1	Humanized	IgG1κ	RSV	
CNTO 5825	Human	IgG1κ	IL-13	
NIST IgG2	Human	IgG2	RSV	
NIST IgG4	Human	IgG4	RSV	

NIST:

NIST HC:

QVTLRESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIRQPPGKALEWLADIWWDDKKHYNPSLKDRLTISKDTSKNQVVLKVTNMDPADTATYYCARDMIFNFYF
DVWGQGTTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDK
RVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKC
KVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHE
ALHNHYTQKSLSLSPGK

NIST LC:

DIQMTQSPSTLSASVGDRVTITCSASSRVGYMHWYQQKPGKAPKLLIYDTSKLASGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCFQGSGYPFTFGGGTKVEIKRTVAA
PSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

CNTO:

CNTO5825 HC:

QVTLKESGPVLVKPTETLTLTCTVSGFSLSTYGVGVGWIRQPPGKALEWLAHIWWDDVKRYNPALKSRLTISKDTSKSQVVLTMTNMDPVDTATYYCARLGSDYDVW
FDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDK
KVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKC
KVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHE
ALHNHYTQKSLSLSPGK

CNTO5825 LC:

EIVLTQSPATLSLSPGERATLSCRASKSISKYLAWYQQKPGQAPRLLIYSGSTLQSGIPARFSGSGSGTDFTLTISSLEPEDFAVYYCQQHDYPYTFGQGTKLEIKRTVAAPSV
FIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

NIST IgG2

IgG2 HC

QVTLRESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIRQPPGKALEWLADIWWDDKKHYNPSLKDRLTISKDTSKNQVVLKVTNMDPADTATYYCARDMIFNFYF
DVWGQGTTVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVDK
TVERKCCVECPPCPAPPAAASSVFLFPPKPKDTLMISRTPEVTCVVVDVSAEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSVLTVLHQDWLNGKEYKCKVSN
KGLPSSIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPMLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALH
NHYTQKSLSLSPGK

IgG2 LC

DIQMTQSPSTLSASVGDRVTITCSASSRVGYMHWYQQKPGKAPKLLIYDTSKLASGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCFQGSGYPFTFGGGTKVEIKRTVAA
PSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

NIST IgG4

IgG4 HC:

QVTLRESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIRQPPGKALEWLADIWWDDKKHYNPSLKDRLTISKDTSKNQVVLKVTNMDPADTATYYCARDMIFNFYF
DVWGQGTTVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKR
VESKYGPPCPPCPAPEAAGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVS
NKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALH
NHYTQKSLSLSLGK

IgG4 LC:

DIQMTQSPSTLSASVGDRVTITCSASSRVGYMHWYQQKPGKAPKLLIYDTSKLASGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCFQGSGYPFTFGGGTKVEIKRTVAA
PSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Figure 1: Amino acid sequences of NIST IgG1, CNTO5825, NIST IgG2 and NIST IgG4. Identical parts of the sequences are highlighted in grey.

3.2.2 Experimental details

3.2.2.1 Native ion mobility mass spectrometry

The native IM-MS experiments were performed on a Synapt G2 HDMS Q-TOF instrument (Waters, UK). The samples were buffer exchanged into 250 mM ammonium acetate using Micro Bio-spin P6 columns (Bio-Rad) at final protein concentrations of ca. 1 mg/mL (6.7 uM). The sample was transferred into the mass spectrometer using gold-coated nanoESI needles (prepared in-house). The instrument was tuned to preserve native higher-order structure using the following parameters: spray capillary voltage 1.2 kV; nanoflow backing gas pressure 0.2 bar; sampling cone 30 V; extraction cone 1.0 V; trap collision energy 10 V; transfer collision energy 0 V; trap bias 45.0 V; IMS gas flow 90 mL/min.; IMS wave velocity 700 m/s; IMS wave height 40.0 V; backing pressure 3.0 mbar. The data were acquired and processed with Masslynx v4.1 software, and ion mobility drift times extracted using Driftscope v2.3 (both Waters). The collision cross sections (CCS) of the proteins were calibrated using known CCS values determined under native conditions as described previously [2].

3.2.2.2 Papain digestion

The papain was activated with 1 mM EDTA, 50 mM sodium phosphate and 10 mM cysteine at pH 7 for 15 minutes at 37 ^{o}C [3]. Activated papain was added to the antibody at a 60:1 (w/w) antibody to protease ratio. The mixture was incubated at 37 ^{o}C for 2 hours. The samples were buffer exchanged into 150 mM ammonium acetate using Micro Bio-spin P6 columns (Bio-Rad) before mass spectrometry analysis.

3.2.2.3 Fast photochemical oxidation of proteins

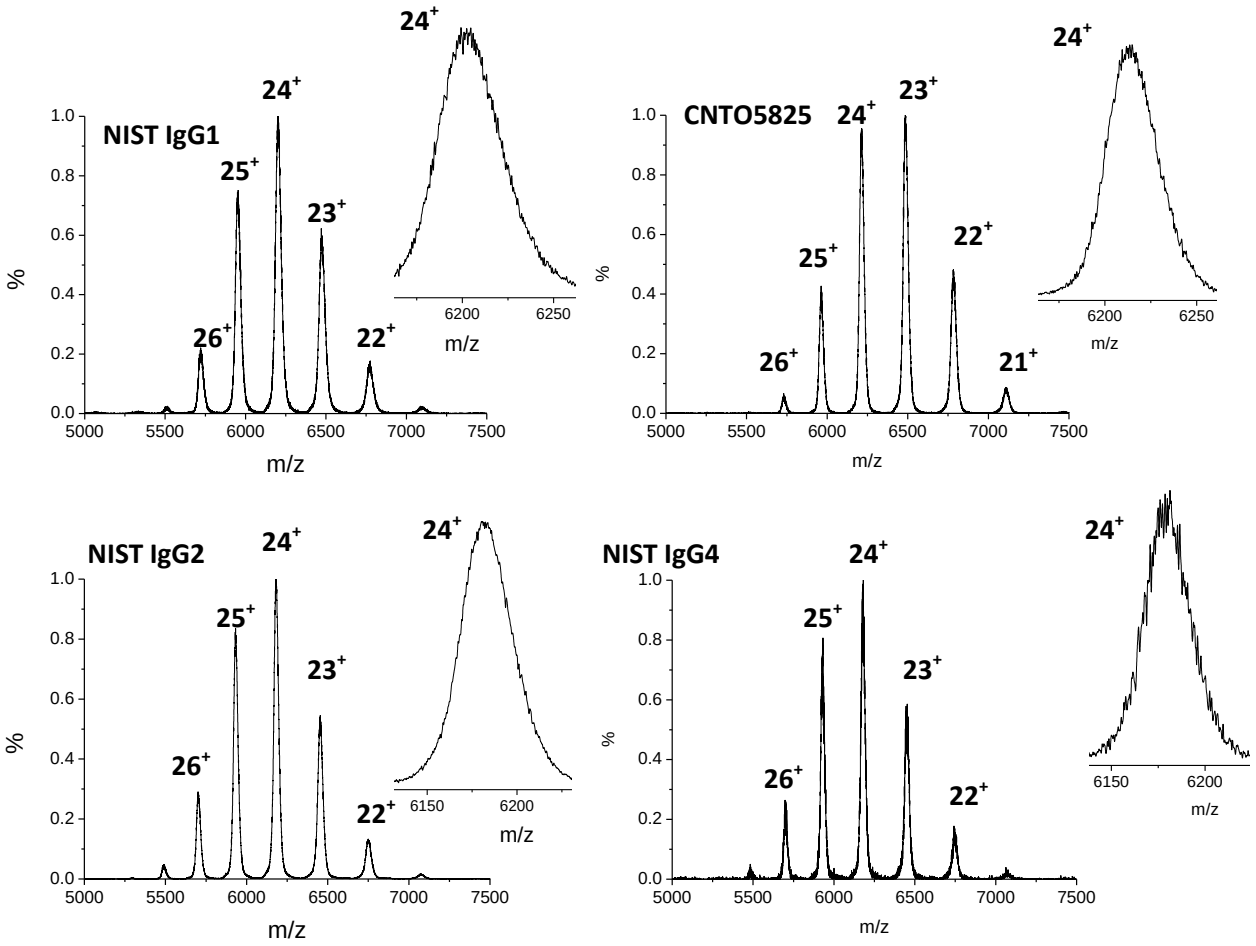
The IgG samples were buffer exchanged into 150 mM ammonium acetate using Amicon concentrator columns. 20 mM L-glutamine was added as a quenching reagent to the sample. 0.02% H_2O_2 was added immediately prior to laser irradiation. The sample was infused via a fused silica capillary (inner diameter of 100μM) at a 20 μL/min. flow rate. A Compex 50 Pro KrF excimer UV laser, operating at 248 nm (Coherent Inc., Ely, UK) with frequency of 15 Hz (laser beam width of <3 mm at the point of irradiation), was used to generate hydroxyl radicals. The sample was exposed to the laser through a window etched into the fused silica capillary. The capillary outflow was collected in a 1.5 mL tube that contained 20 μL of 100 mM L-methionine and 1 μM catalase in 150 mM ammonium acetate to quench the reaction. The control samples were handled in the same manner, without being subjected to laser irradiation. The samples were stored at -80 ^{o}C until analysis.

3.2.2.4 SEC-MS

An ACQUITY UPLC Protein BEH SEC column (200Å, 1.7 μm, 4.6 mm X 150 mm, 10- 500kDa) was coupled to standard in-line electrospray ionization (ESI) interface to detect the analytes with a Q-TOF-2^{TM} instrument (Micromass, Waters). For the chromatography experiment 100 mM ammonium acetate was used as mobile phase, with a flow rate of 0.1 mL/min and a column temperature of 30 ^{o}C. The Q-TOF settings were the following: spray capillary voltage 3.0 V, sampling cone 50 V, extractor cone 10 V, source temperature 150 ^{o}C, collision energy 50 V and backing pressure 10.0 mbar.

3.2.3 Native IM-MS of antibodies

When analyzing the four standard antibodies with native mass spectrometry, charge states 21^+ to 26^+ are typically observed (**Fig. 2**). The difference seen between the four spectra is the charge state distribution of CNTO5825 which is slightly shifted to a lower charge state. The experimental masses of the antibodies are shown in the table in **Fig.2**. The theoretical masses of the IgGs, based on their sequences are noted in this table as well. 3kDa was added to the calculated theoretical mass (most abundant glycoform 2xG1F) to estimate the respective masses of the glycosylated antibodies.



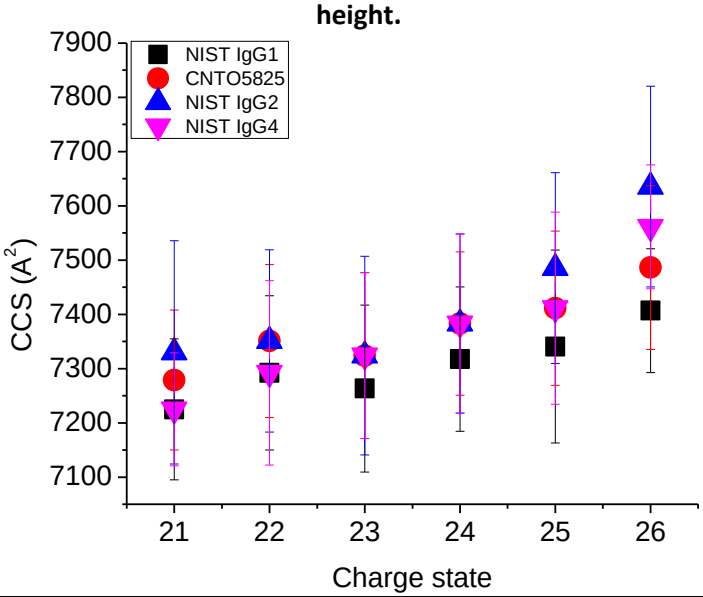
Antibody	Experimental MW (Da)	Theoretical* MW (Da)
NIST IgG1	148,889.4	148,436.88
CNTO5825	149,383.0	148,809.40
NIST IgG2	148,473.7	147,733.94
NIST IgG4	148,363.2	147,769.76

Figure 2: Native mass spectra of NIST IgG1, IgG2, IgG4, and CNTO5825. The insets show zoomed in mass spectra of the 24^+ charge state.* The mass of the most abundant glycoform (2xG1F= 3003 Da) was added to the calculated theoretical mass.

3.2.4 Collision cross sections determined by native ion mobility mass spectrometry

Ion mobility drift times were determined while using gentle non-denaturing instrument conditions. These drift time values can be converted into collision cross sections (CCS) by calibrating them with protein standards. **Table 2** displays the CCS values for each charge state of the four studied antibodies. The CCS values per charge state are similar for the four antibodies. The CCS values are similar to values previously obtained for IgGs, and it has been described that IgGs are more compact when measured with ion mobility, compared to molecular modeling [4]. The error bars are based on the width of the observed ion mobility drift time peak. This peak width can also be linked to the conformational heterogeneity, or "flexibility", of the analysed protein. The NIST IgG2 displays a slightly higher flexibility compared to the other three studied antibodies. This can be explained by the more flexible hinge region of IgG2 subtypes.

Table 2: Collision cross sections calculated for each charge state of antibodies NIST IgG1, IgG2 and IgG4 and CNTO5825. The error bars refer to the width of the ion mobility peak at half height.



NIST IgG1		CNTO 5825		NIST IgG2		NIST IgG4	
Charge state	**CCS (Å²)**	**Charge state**	**CCS (Å²)**	**Charge state**	**CCS (Å²)**	**Charge state**	**CCS (Å²)**
21⁺	7225 ± 130	**21⁺**	7279 ± 129	**21⁺**	7330 ± 206	**21⁺**	7225 ± 104
22⁺	7292 ± 142	**22⁺**	7350 ± 141	**22⁺**	7351 ± 168	**22⁺**	7292 ± 170
23⁺	7263 ± 154	**23⁺**	7324 ± 153	**23⁺**	7324 ± 183	**23⁺**	7324 ± 153
24⁺	7317 ± 133	**24⁺**	7382 ± 132	**24⁺**	7383 ± 165	**24⁺**	7383 ± 165
25⁺	7340 ± 178	**25⁺**	7411 ± 142	**25⁺**	7485 ± 176	**25⁺**	7411 ± 177
26⁺	7406 ± 114	**26⁺**	7486 ± 151	**26⁺**	7635 ± 185	**26⁺**	7561 ± 114
Average	7307		7372		7418		7366

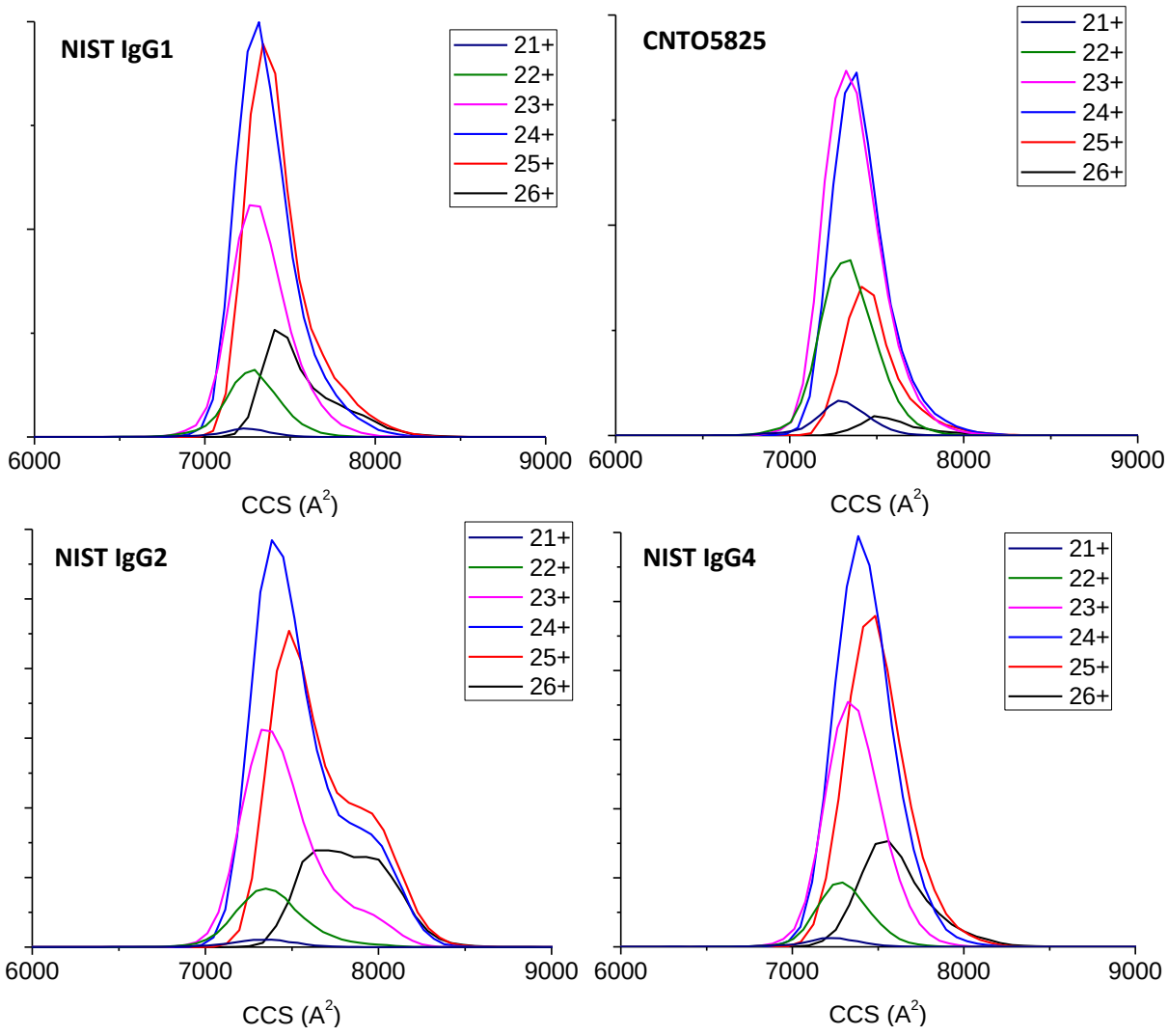


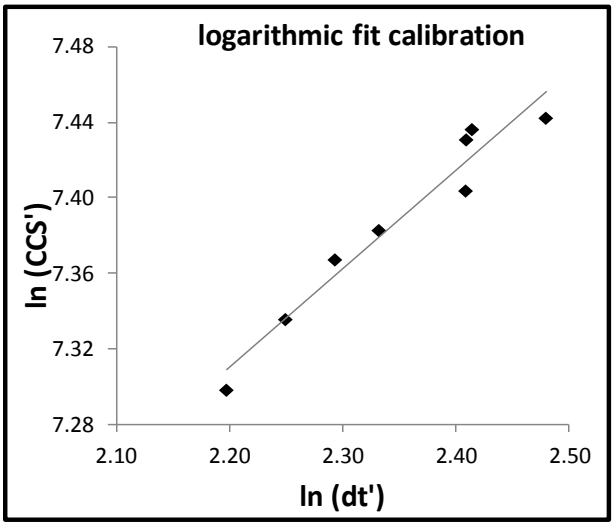
Figure 2.1: Collision cross section distributions per observed charge state for antibodies, NIST IgG1, IgG2, IgG4 and CNTO5825. The spectra are normalized to spectral intensity.

Figure 2.1 shows the CCS distribution (CCSD) for each observed charge state for the four measured antibodies. Again, here the flexibility of IgG2 is seen in the relatively broader CCS distributions, with a putative second, more extended state around 8000 Å^2 also visible This flexibility of antibodies has also been reported in literature, where the width of the CCSD peak was 1500 to 2000 Å^2 [4]. In our experiments the width of the CCSD peaks ranged from 750 to 1000 Å^2. This difference in CCSD distribution can be explained by the used instrument parameters in the literature.

The collision cross sections were calibrated using the known CCS values of alcohol dehydrogenase (ADH), concanavalin A (ConA) and bovine serum albumin (BSA). The CCS values found in literature are shown in **Table 3,** together with the measured drift time of each observed charge state. Based on these values, a logarithmic fit is generated with which the estimated CCS value of the sample can be calculated in this mass range.

Table 3: Overview of calibration of CCS values.

Calibrant	MW (kDa)	Charge state	Lit. N_2 CCS in Å^2	Drift time (ms)
ADH	150	23^+	7420	12.07
		24^+	7450	11.25
ConA	67	14^+	4490	11.30
		15^+	4490	10.02
		16^+	4470	9.11
BSA	103	19^+	6060	11.25
		20^+	6080	10.42
		21^+	6090	9.60



The average CCS values calculated for the antbodies ranged from 7307 to 7418 Å^2. In literature the CCS distribution ranged from 5000 to 10000 Å^2 depending on the measured charge state. Pacholarz *et al.* have shown the flexibility and the wide CCS distributions antibodies show per charge state [4]. For the 23^+ charge state they measured a CCS value of ca. 6100-6200 Å^2, which is a more compact value than the measured 7300 Å^2 in our results. An explanation could be the different IMS gas they used in their experiment (He vs N_2 in our experiments). The CCS values calculated using N_2 as drift gas have shown an increase in CCS value (compared to He) of about 1.10 fold. The measured CCS values still are more compact than the calculated CCS values resulting from molecular dynamics (MD) simulation found in literature, which estimate a CCs value of 8600 to 10600 Å^2 **[4]**. This was explained by the structural collapse around the hinge making the measured CCS value more compact.

Campuzano *et al.* have analysed a NIST monoclonal antibody using a Synapt G1 modified to a radio frequency (RF)-confining drift cell device with ion mobility measurements both in helium and nitrogen drift gases **[5]**. The derived CCS values for charge states 21^+ to 26^+ range from 6696 Å^2 to 6892 Å^2 in helium and 7223 Å^2 to 7403 Å^2 in nitrogen, respectively. These CCS values (in N_2) are closer to our experimental values of 7307 to 7418 Å^2.

3.2.5 Monitoring the unfolding of antibodies

By increasing the trap voltage in a step-wise manner while recording the ion mobility drift time, the unfolding of the protein can be monitored. Since the IgG subtypes have different hinge regions, a difference in unfolding pattern was expected. **Figure 3** shows the collision induced unfolding plots of the 24^+ charge state of the four antibodies. The first transition starts at 60 V. The second transition however starts at 100 V for NIST IgG1, CNTO5825 and NIST IgG4, while NIST IgG2 makes this transition at 130 V. After the second transition in the plot of NIST IgG2, two distinct conformations are observed, while the other three antibodies show one high intensity conformation. The same experiment was performed for the 23^+ charge state of the antibodies (**Fig.4**). For charge state 23^+ the first transition was seen at 70V. The second transition is at 140 V for CNTO5825 and IgG4. IgG2 shows a second transition at 180 V with two conformations. The three subtypes of IgG can thus be distinguished based on their collision induced unfolding plot.

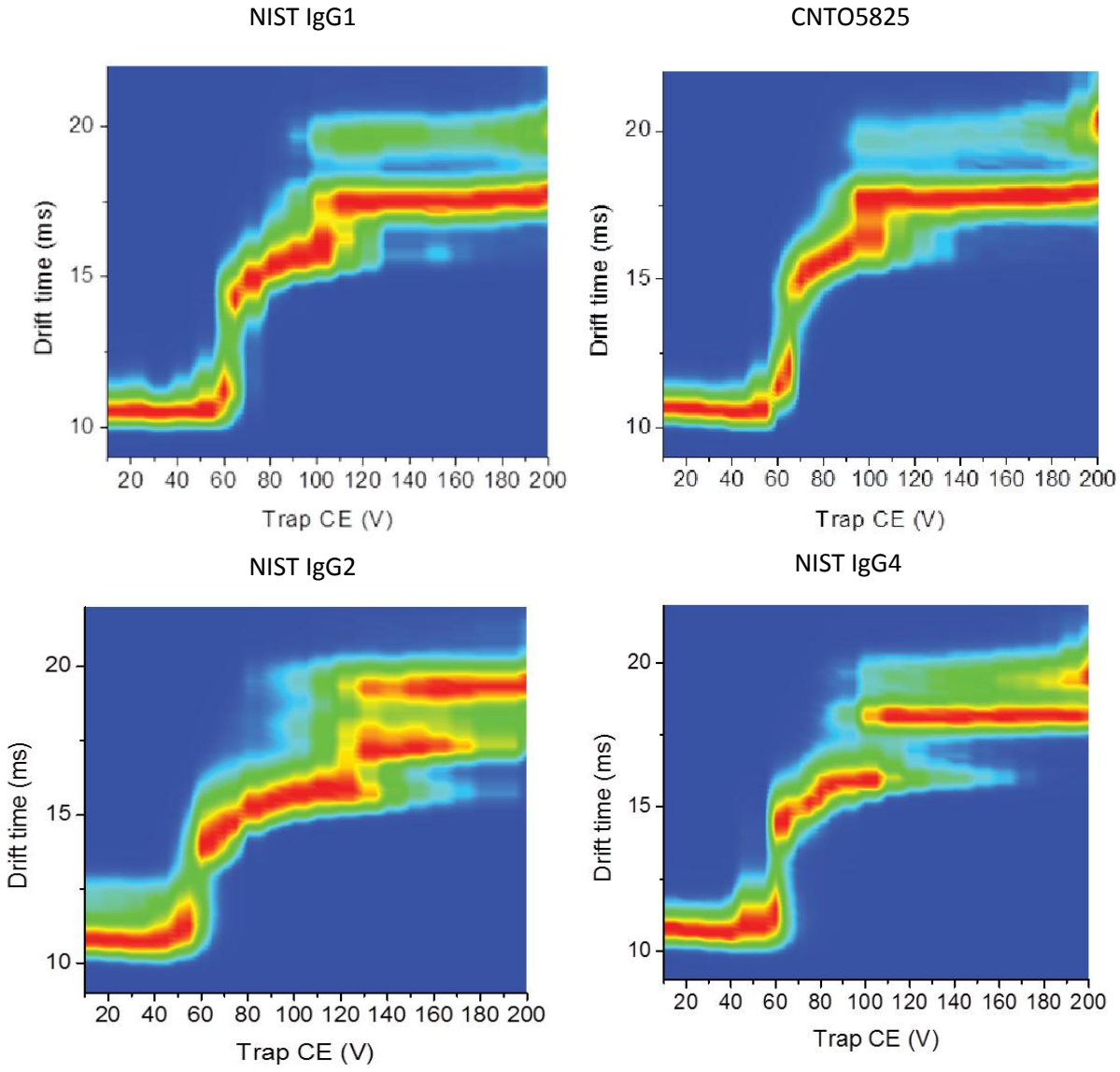


Figure 3: Collision induced unfolding plots for the 24^+ charge state of NIST IgG1, CNTO5825, NIST IgG2, and NIST IgG4.

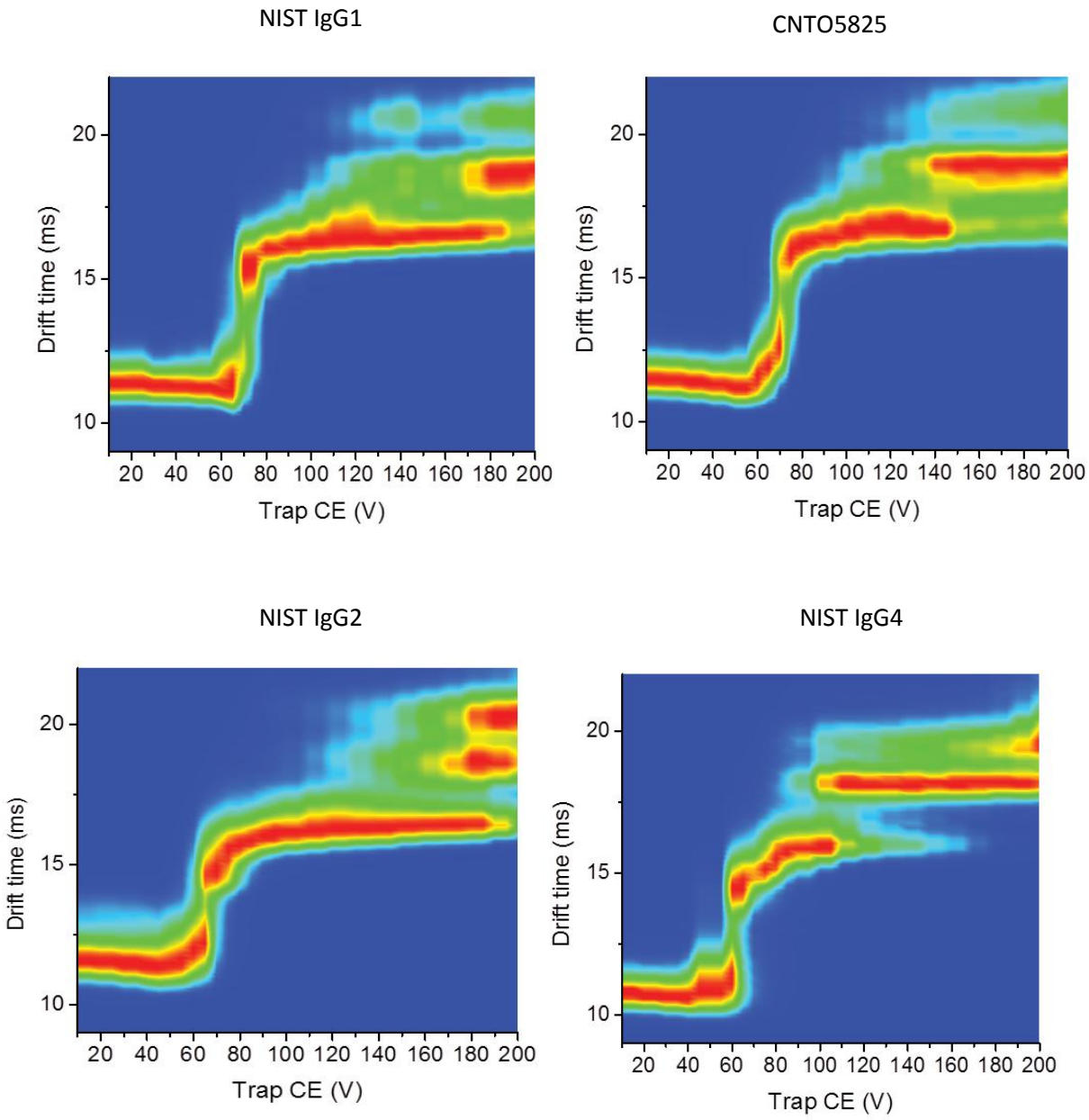


Figure 4: Collision induced unfolding plots for the 23^{+} charge state of NIST IgG1, CNTO5825, NIST IgG2, and NIST IgG4.

The CIU plots in Figures 3 and 4 were made in steps of 10 V, and peaks were processed by selecting the m/z range at full width at half maximum (i.e. the upper half of the peaks) to generate the drift time plots.

In Figures 4.1 and 4.2 the spectra are shown of the 23+ charge state of the NIST IgG1 at increasing trap collision energy. The m/z range of the left, center and right part of the peak was selected and the respective ion mobility drift times are shown on the left panel. It general, no big differences were observed, except for at the transition points, where the lower m/z range would transition slower (i.e. have a lower intensity drift time peak. This can be observed at 70 V and 170 V. From 10 V to 50 V there is no change in drift time and there is no difference when the left, center or right part of the m/z peak is selected. Assuming that the higher m/z part of the peak would have certain PTMs (glycosylations) that the left part of the peak would not have, this indifference in ion mobility drift time could mean that the glycosylation does not induce a large structural change that can be observed by ion mobility. (with the additional mass and volume of the glycans not directly detectable at the current mobility resolution). There is apparently however a small effect seen as a slight destabilization towards unfolding in the gas-phase CIU assays.

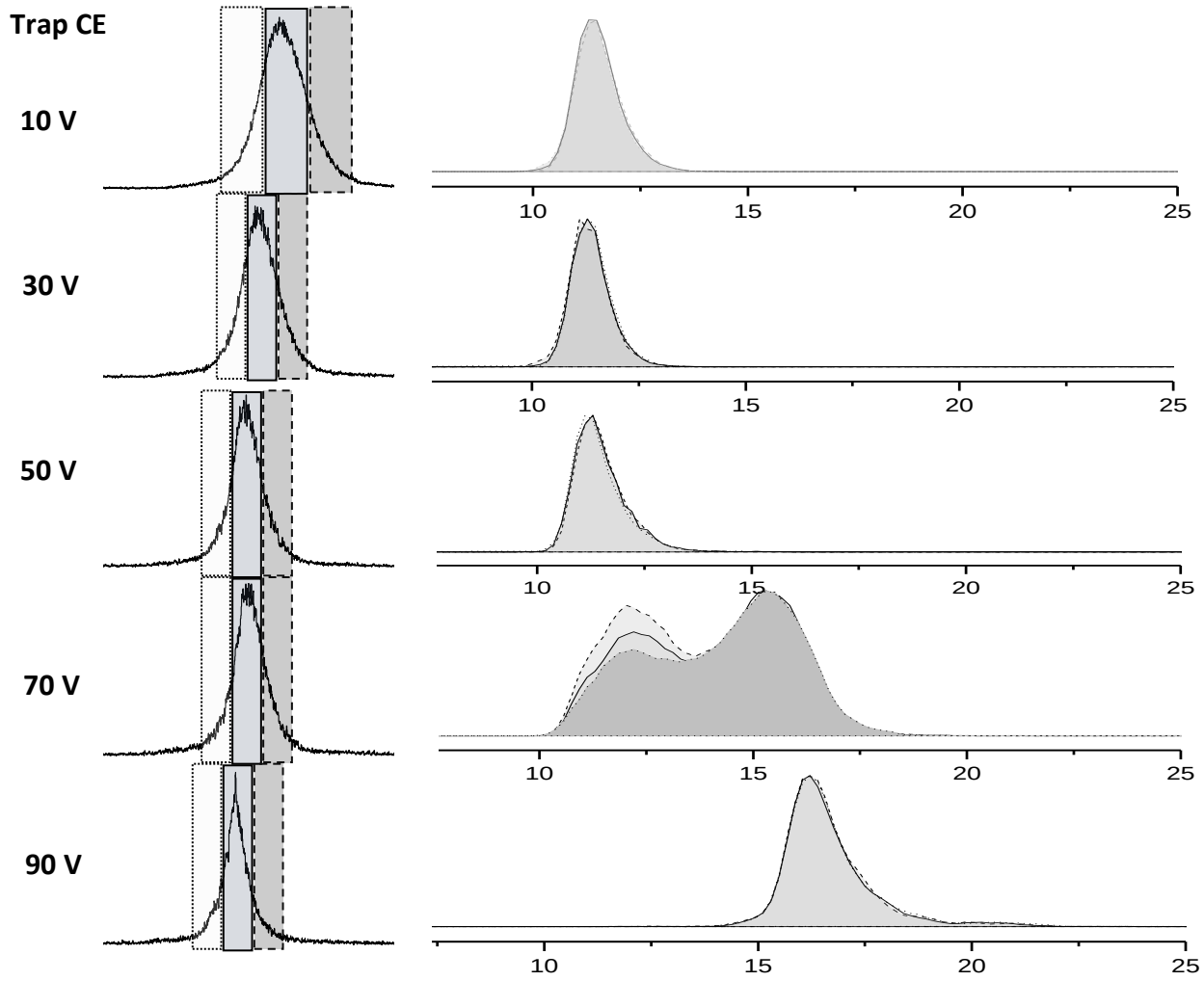


Figure 4.1: Mass spectra of the 23^{+} charge state of NIST IgG1 at increasing trap collision energies (10V to 90 V) with the respective ion mobility drift times selected at the left, center and right m/z range of the broad, heterogeneous antibody peak.

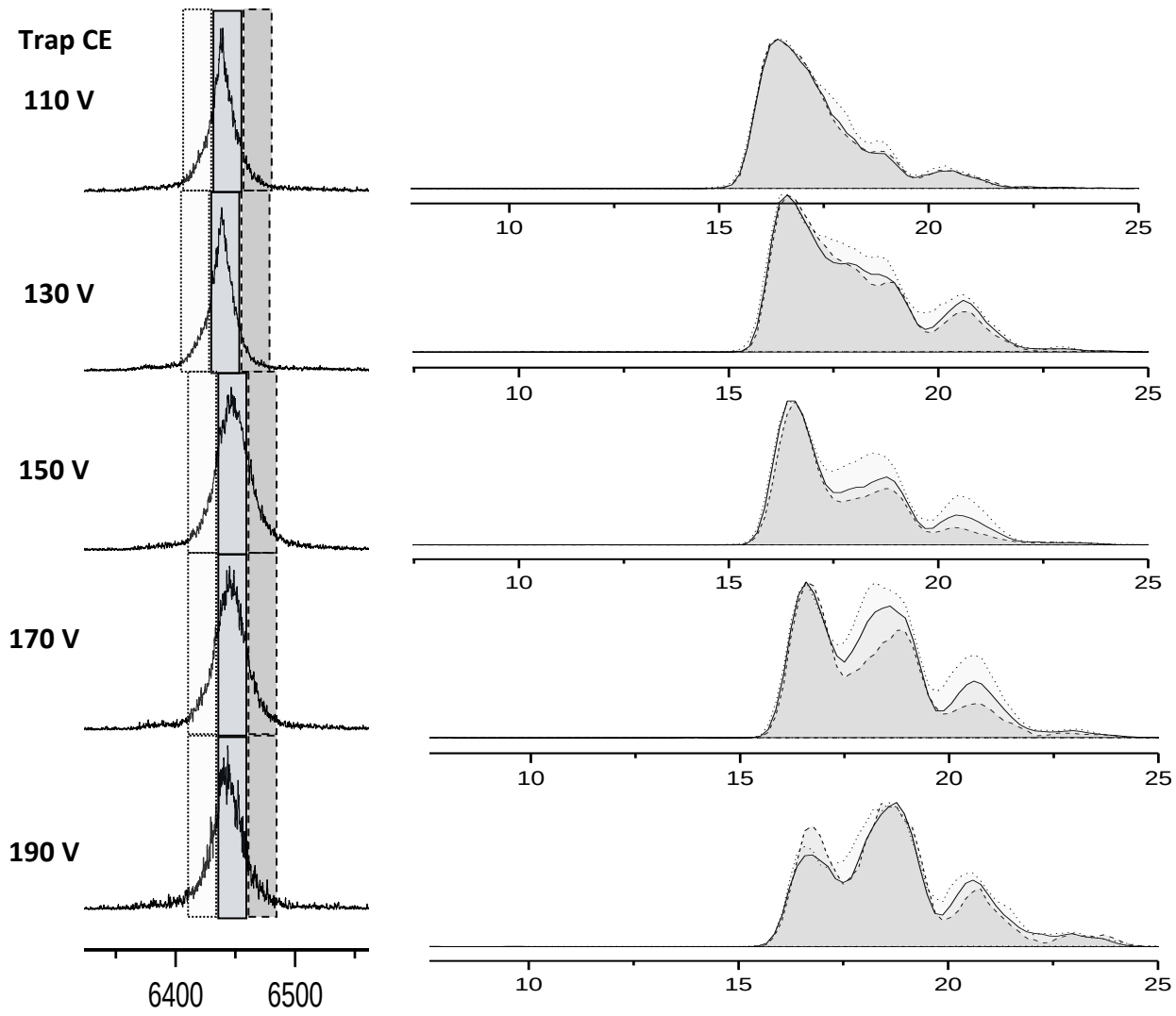


Figure 4.2: Mass spectra of the 23^{+} charge state of NIST IgG1 at increasing trap collision energies (110V to 190 V) with the respective ion mobility drift times selected at the left, center and right m/z range of the peak.

3.2.6 Freeze-thaw effect on antibodies

During research studies, samples are often stored at -80° C to preserve the protein structure. Furthermore, protein-based drugs also need cold storage. However freeze-thaw cycles can also affect the stability of protein samples. The stability of a protein can be linked to the primary and higher-order protein structure. Aggregation probably linked with unfolding is a particularly important property to assess. In practice, antibodies for therapeutic use are never stored frozen, but are kept at refrigerated conditions to avoid the freeze-thaw cycles. **Figure 5** shows a schematic overview of freeze-thaw experiment performed on NIST IgG1 and CNTO5825. The received IgG samples were put in the -80°C freezer when they arrived from J&J. Before the MS experiment they were thawed in the fridge (4°C). They were then diluted to 10 µM and desalted into 150 mM ammonium acetate using Bio-Spin P-6 columns (Bio-Rad) before the MS measurement.

Collision induced unfolding studies were conducted to assess the stability of IgG samples; the ion mobility drift time was plotted against the increasing trap collision energy (**Fig. 6A & 7A**). The samples were then frozen by keeping them at -80° C for one month and the CIU plot was created again, after thawing again in the fridge before measuring (**Fig. 6B & 7B**). Although there are no big differences in the transitions in the plot, there is a difference in the relative intensities between conformations.

After the first transition at 80 V, (**Fig. 6A & 7A)** the conformation at a drift time of 12 ms is still present at a relatively low intensity till 140 V. After the second freeze-thaw cycle (**Fig. 6B** & **7B**) this phenomenon remains till 100V. This could possibly indicate that this compact conformation is less stable after the freeze-thaw cycle. The effect of freeze-thaw cycles on antibodies was not very large in this study. A slight loss of stability can be indicated based on the relative intensities of conformations.

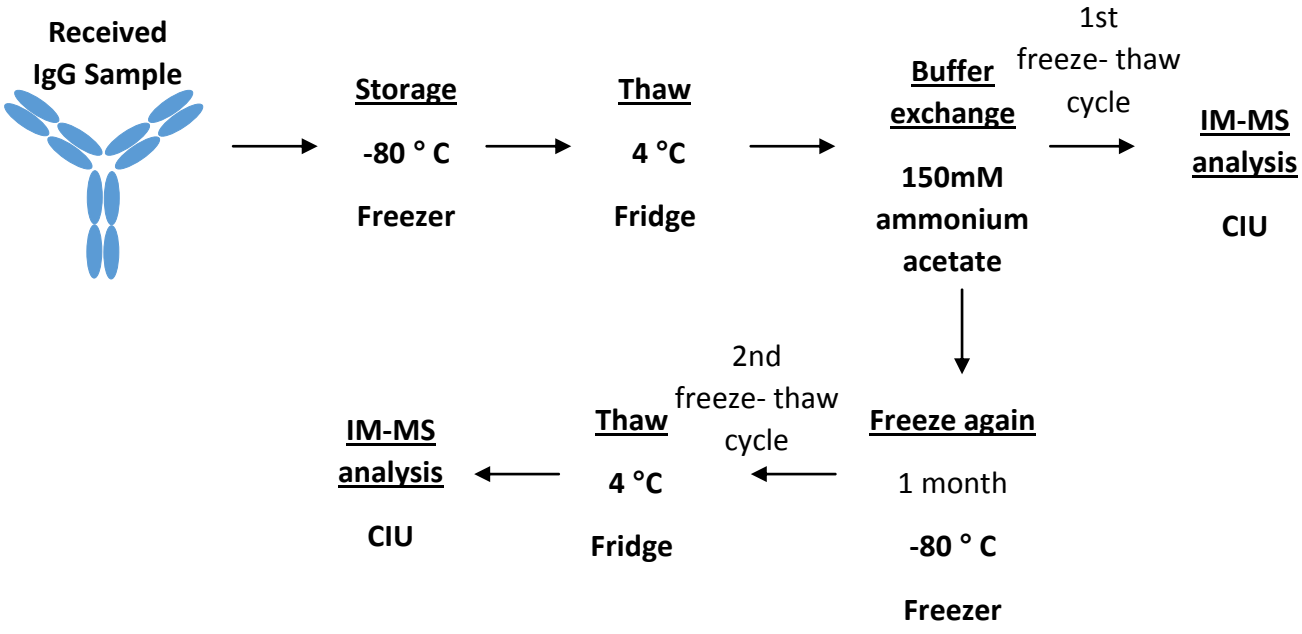


Figure 5: A schematic overview of the freeze-thaw experiment performed on NIST IgG1 and CNTO5825

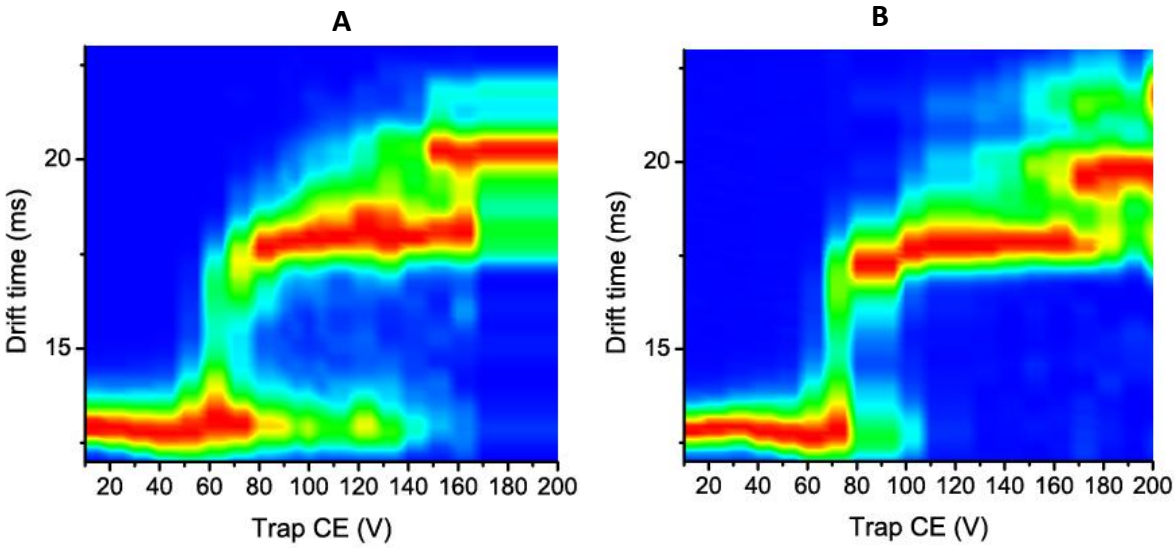


Figure 6: a) CIU plot of NIST after first freeze-thaw cycle. b) CIU plot of NIST after second freeze-thaw cycle.

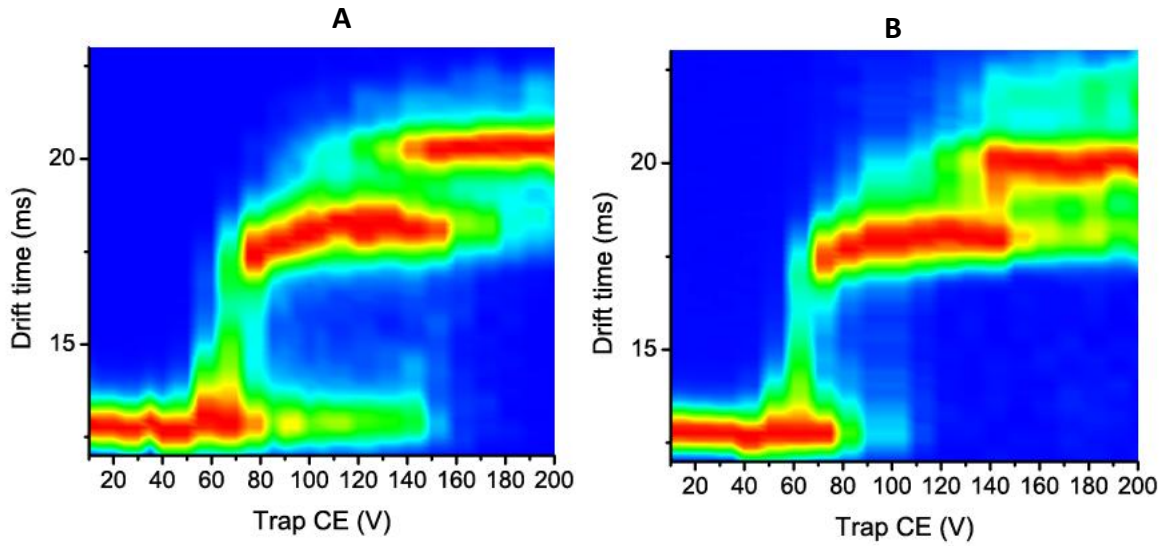


Figure 7: a) CIU plot of CNTO5825 after first freeze-thaw cycle. b) CIU plot of CNTO5825 after second freeze-thaw cycle.

3.2.7 Native IM-MS of mAbs digested with papain

Papain is a non-specific thiol-endopeptidase from *Carica papaya latex* that has a sulfhydryl group in the active site, which needs to be in reduced form for activity. EDTA and cysteine were added to respectively stabilize and activate papain. When activated papain is incubated with an antibody, peptide bonds in the hinge region are cleaved producing two Fab fragments and on one Fc fragment (**Fig.8**). In practice, this digest is formed when, for example, only the Fab fragment is needed for antigen-antibody binding studies without the interference of the Fc region. Cleaving a large antibody into its subunits also simplifies the characterisation of antibody variants. Since most of the IgGs differ in their sequence in their Fab or Fc region, an enzymatic approach was tested to cleave the antibody into these fragments. The three studied IgGs were incubated with papain and the resulting fragments were analysed with native IM-MS to monitor the differences between the IgG isotypes.

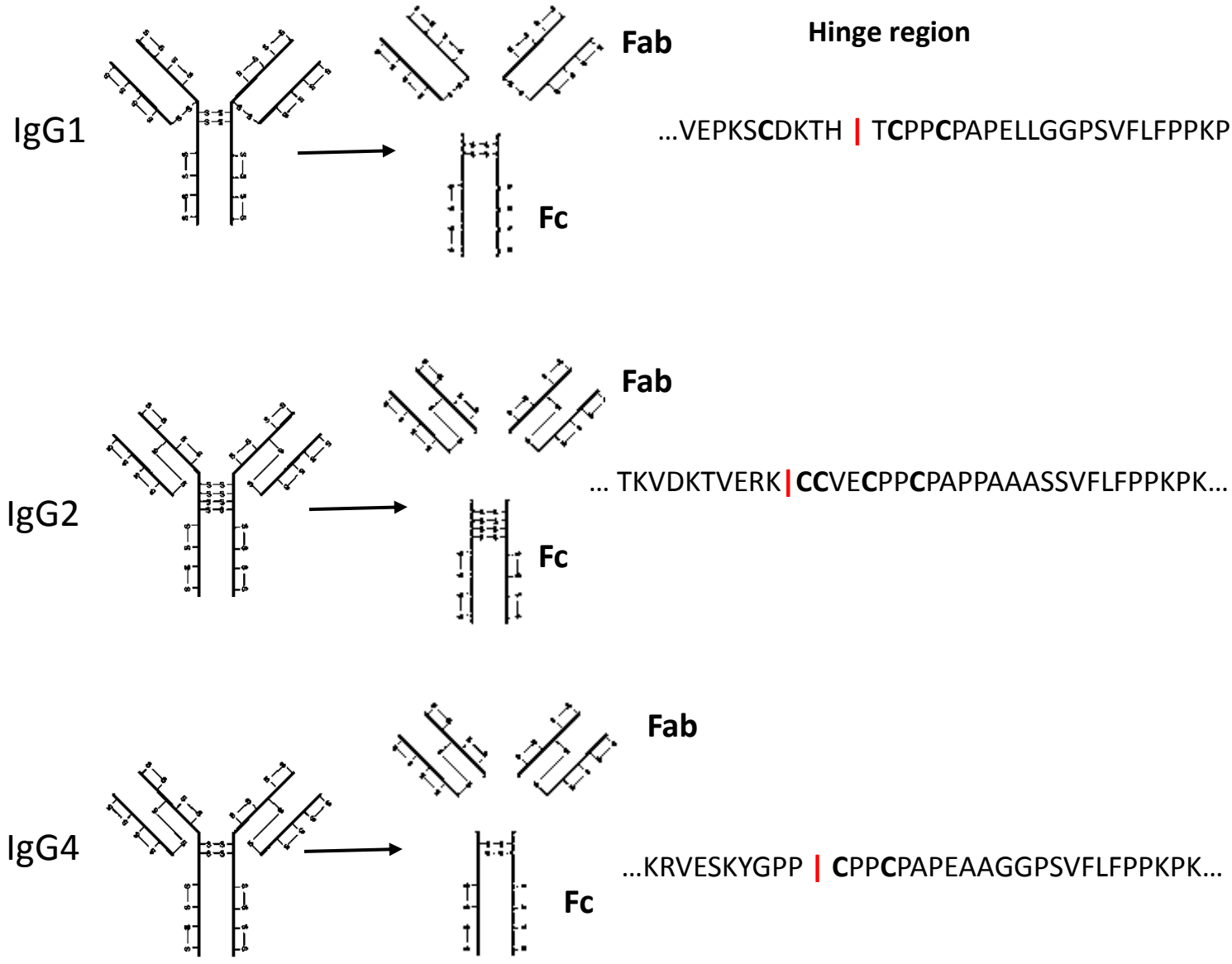
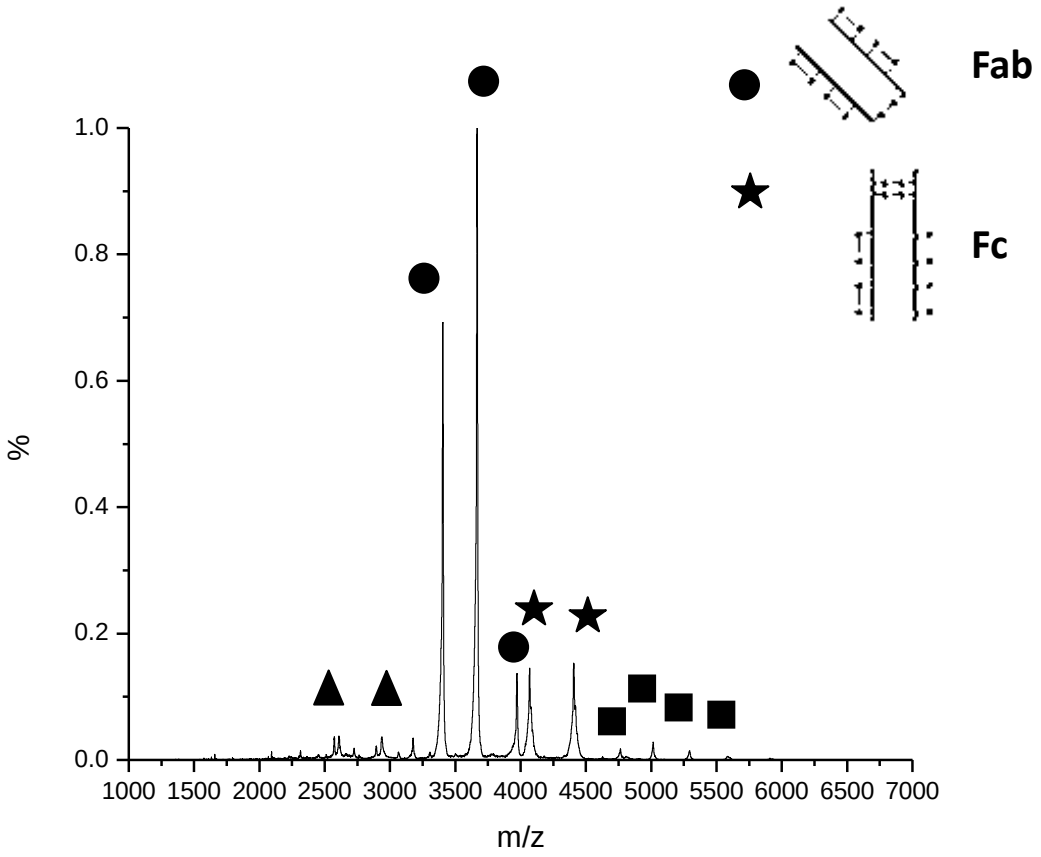


Figure 8: Schematic illustration of papain digestion of antibody into Fab and Fc fragments. The amino acid sequence of the hinge region is displayed with cleavage site marked in red.

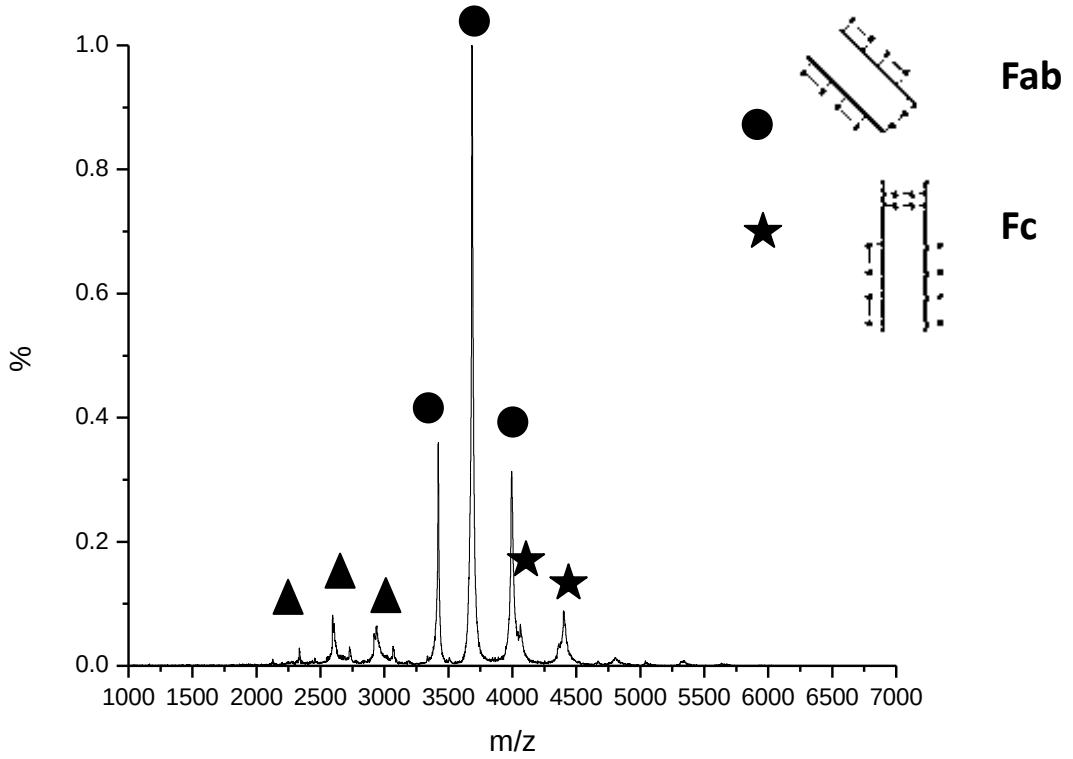
Figure 9 shows the native mass spectrum of NIST IgG1 after papain digestion. Next to the cleaved Fab domain (47651.0 Da) and Fc domain (52876.0 Da), the presence of light chain (23150.3 Da) and $(Fab)_2$ (95225.4 Da) is noticed. The table in **Figure 9** displays the experimental and theoretical masses which seem to match each component. The experimental mass of the Fc domain is 2.7 kDa higher than the theoretical mass due to post translational modifications (e.g., glycosylation).



Symbol	Experimental MW (Da)	Theoretical MW (Da)	ID
▲	23150.3	23127.74	LC
●	47651.0	47636.47	Fab
★	52876.0	50196.95	Fc
■	95225.4	95272.94	$(Fab)_2$

Figure 9: Native mass spectrum of NIST IgG1 after papain digestion. The table displays the experimental and theoretical molecular weight for each observed component of the antibody.

In **Figure 10** the native spectrum of CNTO5825 after papain digestion is shown. The spectrum displays the presence of the light chain (23428.5 Da), the Fab domain (47900.7 Da) and the Fc domain (52850.1 Da). The papain digestion worked efficiently as no $(Fab)_2$ or full IgG is seen in the spectrum.

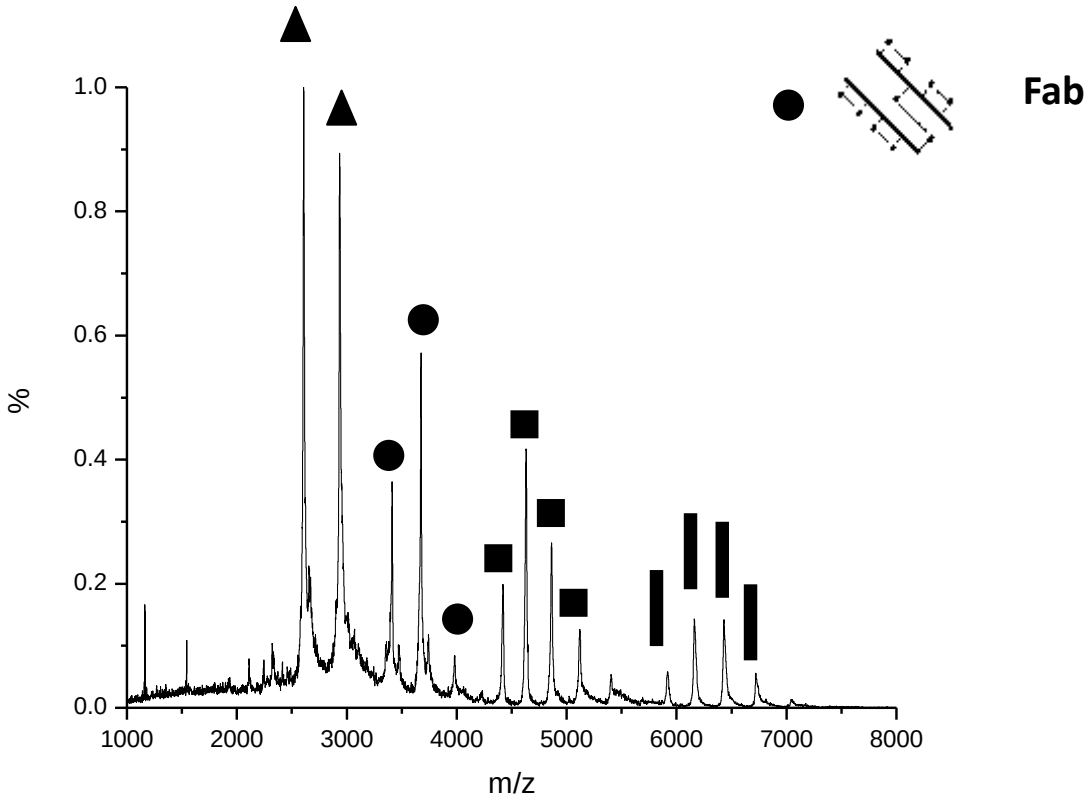


Symbol	Experimental MW (Da)	Theoretical MW (Da)	ID
▲	23428.5	23336.00	LC
●	47900.7	47854.79	Fab
★	52850.1	50609.31	Fc

Figure 10: Native mass spectrum of CNTO5825 after papain digestion. The table displays the experimental and theoretical molecular weight for each observed component of the antibody.

Tables 3 and **4** show the collision cross sections of the IgG fragments observed after papain digestion. Since NIST IgG1 and CNTO5825 are both in the IgG1 subclass, the collision cross sections are expected to be similar.

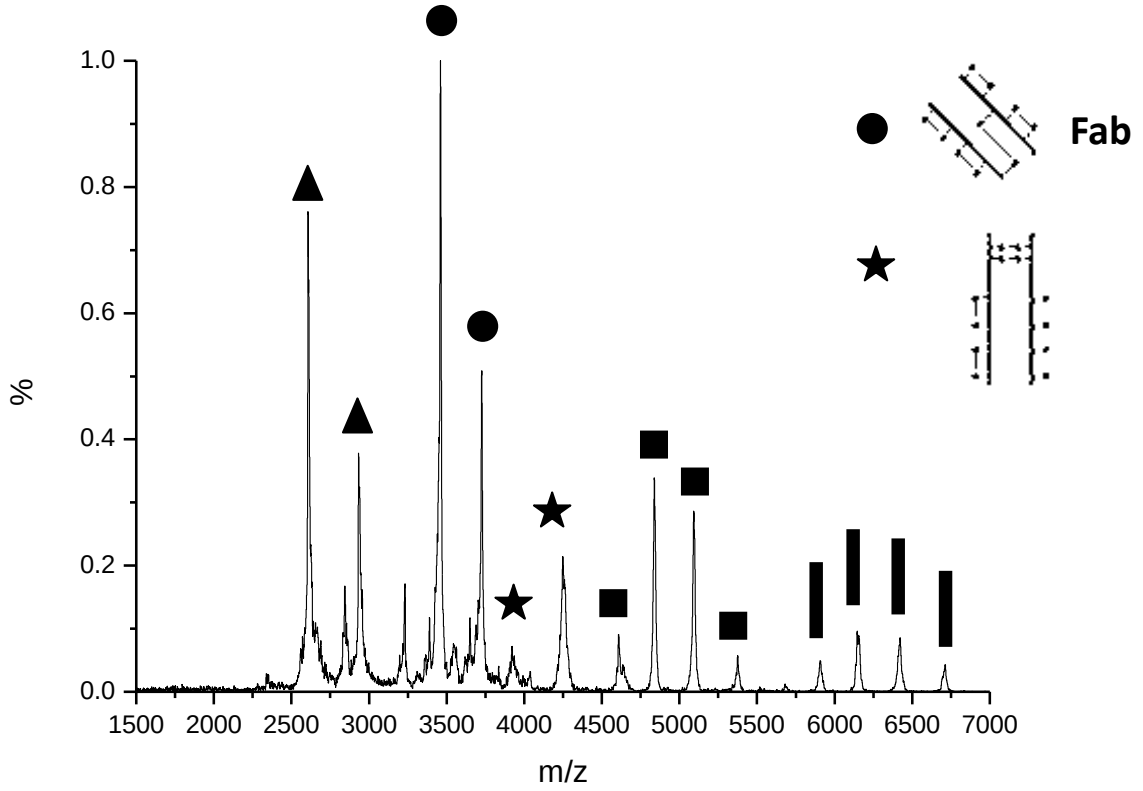
The papain digestion of NIST IgG2 was more difficult (**Figure 11**). This can be explained by the presence of more disulfide bonds in the hinge region compared to IgG1 and IgG4. After digestion the light chain (23457.61 Da) and the Fab domain (47725.34 Da) were observed. A subunit of 97200.39 Da was also observed which could be $(Fab)_2$ which was cleaved lower in the hinge. There was also full IgG present in the spectrum, but the Fc domain was not observed. Probably this region got over-digested or became unstable after cleavage and may have precipitated in solution.



Symbol	Experimental MW (Da)	Theoretical MW (Da)	ID
▲	23457.6	23127.74	LC
●	47725.3	47164.87	Fab
■	97200.4	94329.74	$(Fab)_2$
▌	147875.0	144730.94	IgG

Figure 11: Native mass spectrum of NIST IgG2 after papain digestion. The table displays the experimental and theoretical molecular weight for each observed component of the antibody.

The results for the papain digestion of NIST IgG4 are displayed in **Figure 12**. The spectrum displays the light chain (23453.92 Da), the Fab domain (48433.95 Da), the Fc domain (50926.06 Da), $(Fab)_2$ (95225.94 Da) and full IgG (148434.45 Da).



Symbol	Experimental MW (Da)	Theoretical MW (Da)	ID
▲	23453.9	23127.74	LC
●	48433.9	47504.30	Fab
★	50926.1	49794.15	Fc
■	95 225.9	95008.60	$(Fab)_2$
▌	148434.5	144766.76	IgG

Figure 12: Native mass spectrum of NIST IgG4 after papain digestion. The table displays the experimental and theoretical molecular weight for each observed component of the antibody.

Every IgG subtype showed a different result after papain digestion. The Fab domain was cleaved off each time. The Fc domain was observed for IgG1 and IgG4. **Tables 3** to **6** show the collision cross sections of the observed fragments after papain digestion. When comparing the subcomponent between the different subtypes, the Fab domains are similar in CCS value. The Fc domain of the two IgG1 are in the same CCS range and are slightly more compact than the Fc domain of the IgG4. The cleaved $(Fab)_2$ region also seems more compact for the IgG1 compared to IgG2 and IgG4 even with similar molecular weights. A possible explanation could be that the disulfide bond in the Fab domain in IgG1 is linked differently compared to those in the Fab domains of IgG2 and IgG4 (**Fig. 7**).

Table 3: Collision cross sections of the observed fragments of NIST IgG1 after papain digestion

NIST IgG1							
LC		Fab		Fc		$(Fab)_2$	
Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)
8^+	2156 ± 26	12^+	3567 ± 78	12^+	3636 ± 89	17^+	5486 ± 124
9^+	2276.97 ± 95	13^+	3584 ± 90	13^+	3706 ± 57	18^+	5607 ± 89
		14^+	3726 ± 94			19^+	5643 ± 161
						20^+	5759 ± 94

Table 4: Collision cross sections of the observed fragments of CNTO5825 after papain digestion

CNTO5825			
Fab		Fc	
Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)
12^+	3637 ± 79	12^+	3672 ± 70
13^+	3667 ± 54	13^+	3902 ± 144
14^+	3771 ± 61		

Table 5: Collision cross sections of the observed fragments of NIST IgG2 after papain digestion

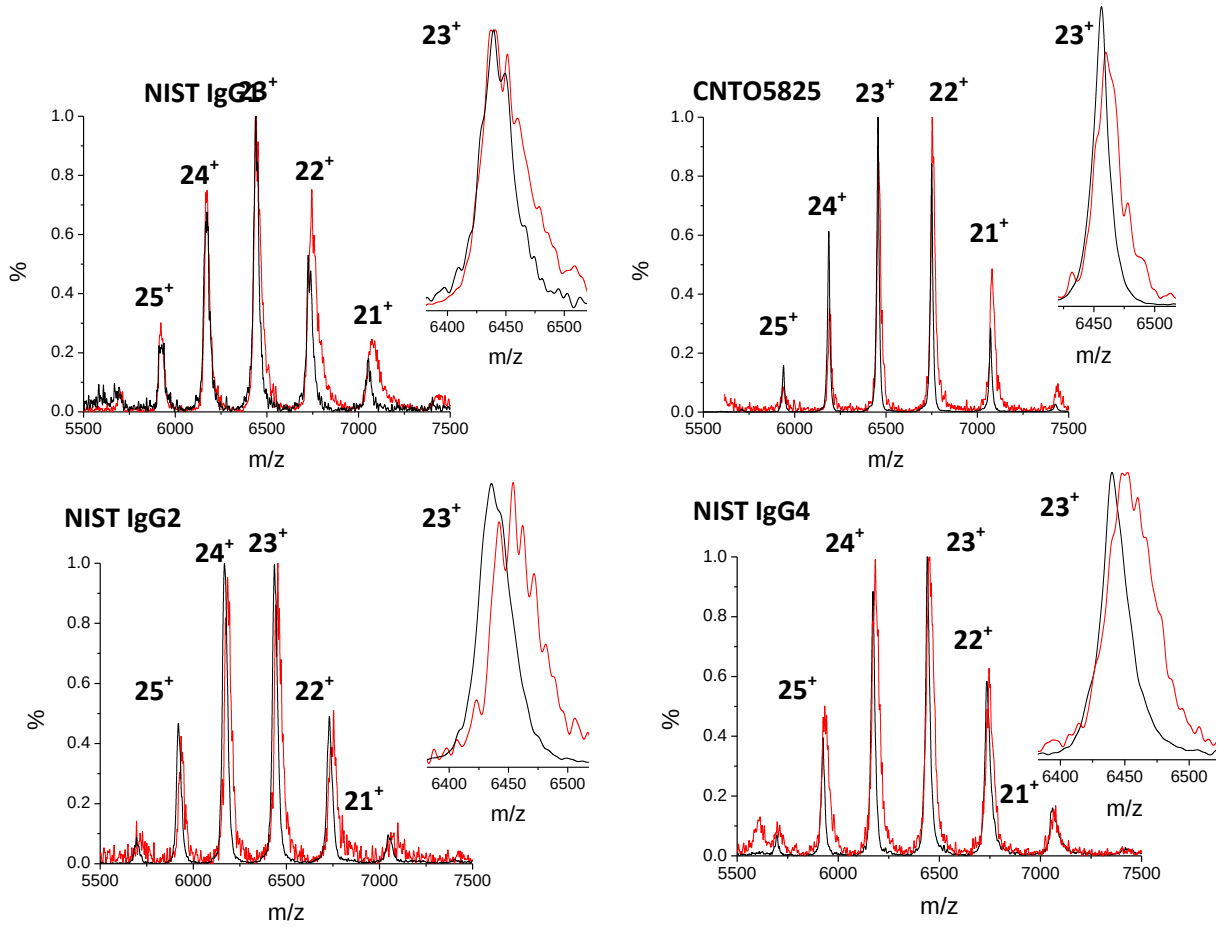
NIST IgG2					
LC		Fab		(Fab)2	
Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)
8+	2182 ± 54	12+	3567 ± 86	18+	5809 ± 79
9+	2276 ± 47	13+	3667 ± 170	19+	5866 ± 104
		14+	3726 ± 121	20+	6058 ± 124
				21+	6544 ± 189

Table 6: Collision cross sections of the observed fragments of NIST IgG4 after papain digestion

NIST IgG4							
LC		Fab		Fc		$(Fab)_2$	
Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)	Charge state	CCS ($Å^2$)
8^+	2207 ± 24	12^+	3673 ± 120	12^+	3940 ± 93	17^+	5907 ± 45
9^+	2276 ± 104	13^+	3707 ± 115	13^+	3978 ± 85	18^+	5973 ± 80
		14^+	3771 ± 123			19^+	6058 ± 170
						20^+	6177 ± 300

3.2.8 FPOP analysis of IgGs

Fast photochemical oxidation of proteins (FPOP) can be used to study protein-folding dynamics as it can oxidize regions of the domain which are solvent accessible. The FPOP technique was applied on four standard antibodies in order to map the structure by locating the oxidation sites. Unfortunately, the enzymatic digestion of the oxidised antibodies were not performed correctly or efficiently (i.e., nearly no peptides were detected during LS/-MS), so the oxidation sites could not be determined. However, the oxidised antibodies were analysed with native ion mobility mass spectrometry to monitor the affect of the oxidation on the protein structure (**Figure 13**). The oxidized antibodies showed a slight mass shift (187 to 270 Da) in the mass spectra and did not show any drastic change in charge state distribution.



Antibody	NIST IgG1	CNTO5825	NIST IgG2	NIST IgG4
Control	148148 Da	148451 Da	148041 Da	148131 Da
Oxidized	148418 Da	148638 Da	148267 Da	148365 Da
Δ mass	270 Da ≈ 17 O	187 Da ≈ 12 O	226 Da ≈ 14 O	234 Da ≈ 15 O

Figure 13: Native mass spectra of NIST IgG1, CNTO5825, NIST IgG2 and IgG4 (black line) overlaid with the spectra of their respective oxidized forms after FPOP (red line). The insets in the spectra show the zoomed in spectrum of the 23^+ charge state. The masses of the antibody before and after oxidation are displayed in the table. The difference in mass is linked to covalently bound oxygen.

The 24^+ charge state was selected to compare ion mobility drift times before and after oxidation (**Figure 14**). The slight shifts are due to the mass addition after oxidation. These drift times values were converted into CCS values and displayed in **Table 5**. The changes in CCS value are between 0.30 to 1.80 % which makes these differences relatively insignificant. This means that the oxidation of the antibodies did not influence the overall structure of the antibody.

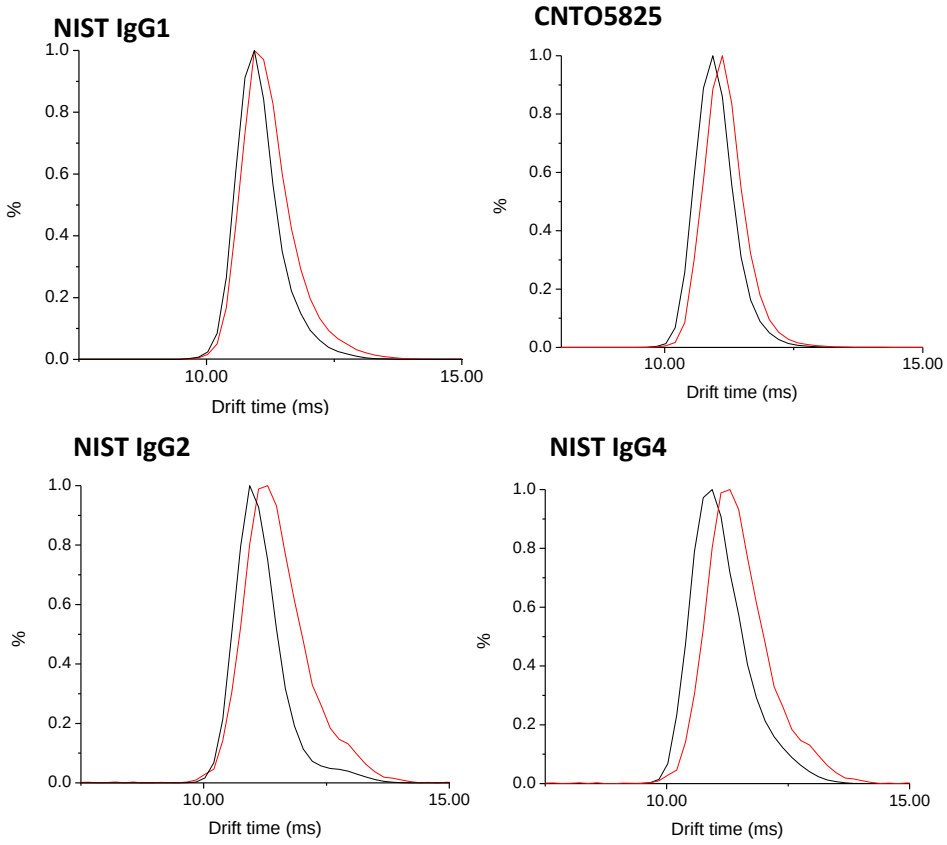


Figure 14: The ion mobility drift time of the 23^+ charge state of the four studied antibodies (full line) overlaid with the drift time of the oxidized form after FPOP (dotted line).

Table 5: Collision cross sections of the antibodies before and after oxidation with FPOP.

	NIST IgG1	CNTO5825	NIST IgG2	NIST IgG4
CCS Control ($Å^2$)	7466 ± 135	7466 ± 135	7466 ± 201	7466 ± 201
CCS Oxidized ($Å^2$)	7488 ± 245	7533 ± 134	7601 ± 202	7601 ± 256
Δ ($Å^2$)	22.41	67.53	134.63	134.52
Δ (%)	0.30	0.90	1.80	1.80

3.2.9 SEC-MS of IgGs

With a size exclusion chromatography (SEC) column coupled to the mass spectrometer the salts and contaminants in the sample can be separated from the protein in the column. **Figure 15** shows the chromatograms of NIST IgG1 and CNTO5825 and of the two antibodies mixed. The observed retention times were NIST IgG1 12.84 min and CNTO5825 12.56 min. Since these two antibodies differ slightly in amino acid sequence, this could be a reason why the retention times are slightly different. The slight difference in the sequence could cause a difference in shape and size. This slight difference was also detected in the ion mobility experiments.
The chromatogram shows a shifted and broader peak for CNTO5825, possibly through an error during this run such as an incorrect synchronization between the LC run and the MS detection, as this synchronization happened manually.

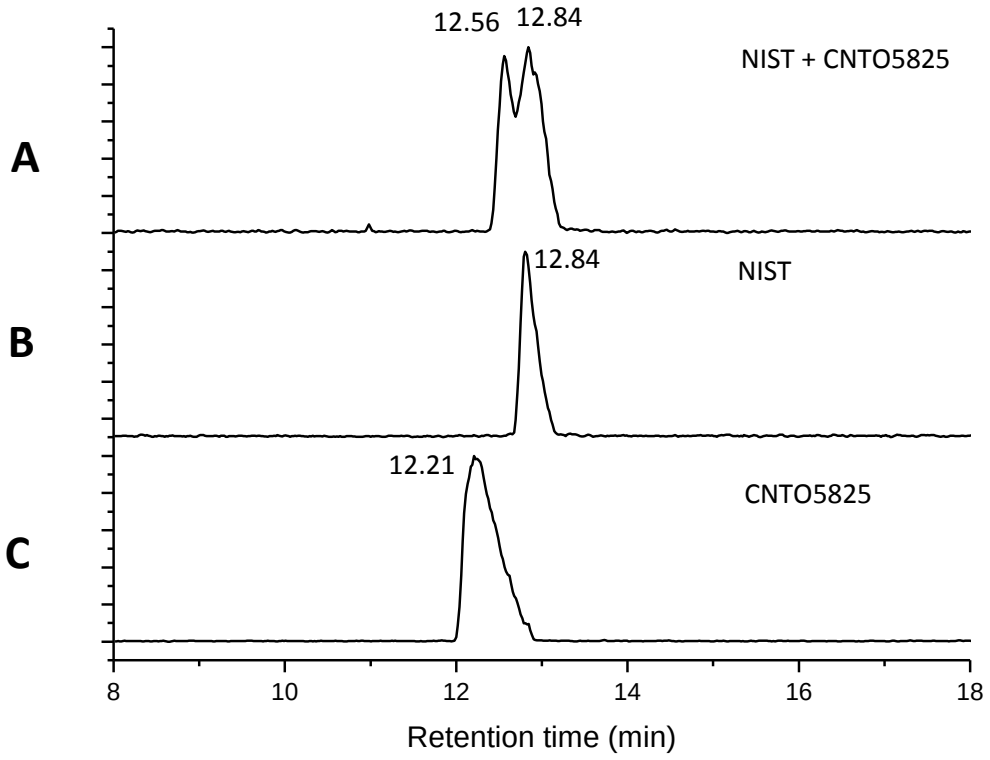


Figure 15: SEC chromatogram of A) a mixture of NIST and CNTO5825, and the two separate antibodies (B & C).

Figure 16 shows the mass spectra observed with SEC-MS. Charge states 21^+ to 29^+ are detected for CNTO5825 and NIST IgG1. This method thus shows a slightly more unfolded form of the antibody compared to the native MS with direct infusion nano-electrospray ionization (**Fig.2**).

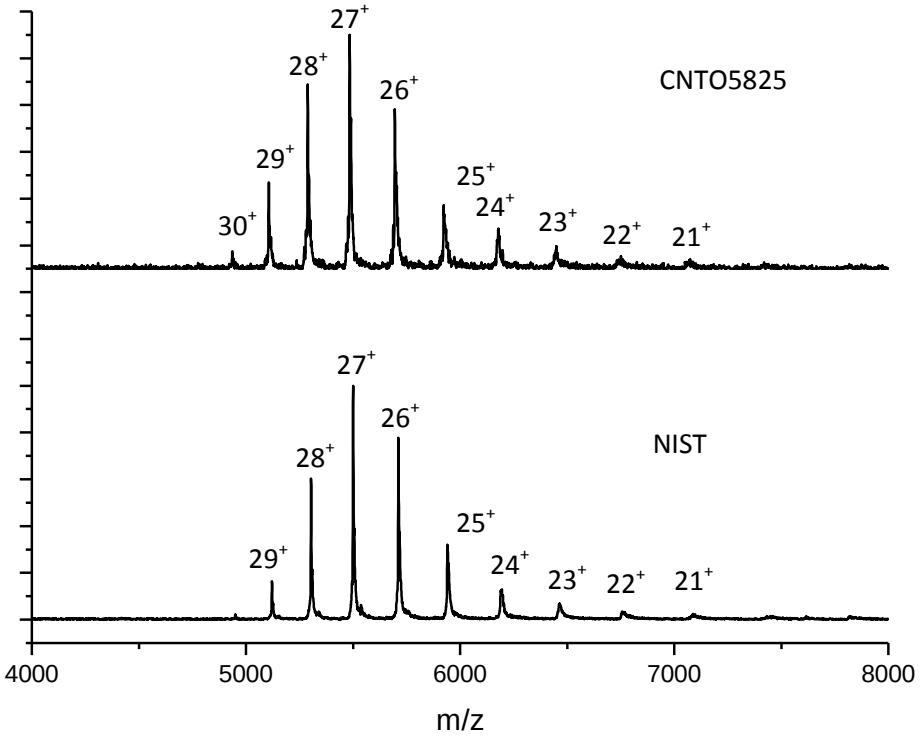


Figure 16: Mass spectra of A) CNTO5825 and B) NIST IgG1

The comparison between direct infusion nESI-MS and SEC-ESI-MS has been published in an article by Marco Tassi et al. in the Journal of Separation Science. This article is a review on native LC-MS of biopharmaceutical products and is attached at the end of this chapter.

3.2.10 Conclusion

This study on four standard antibodies showed us the versatility of MS-based methods by characterising different aspects of the protein structure. With native mass spectrometry the mass of the antibodies was determined. In combination with ion mobility the collision cross sections were calculated showing the similar size and shape of antibody but also the flexibility. The IgG2 subtype showed relative more flexibility compared to the other subtypes because of the larger hinge region. The subtypes IgG1, IgG2 and IgG4 could also be distinguished based on their collision induced unfolding plots. This difference can also be based on the stability of their hinge regions through which it would unfold differently. When the proteins went through a freeze-thaw cycle, similar CIU plots were produced, although a slight difference in relative intensities in conformations were observed.

To compare the Fc and Fab domain of each subtype, the two domains were produced via a limited papain digestion. The mass and collision cross sections were determined for each observed subunit. The Fc domains of NIST IgG1, CNTO5825 and NIST IgG4 had similar CCS values. The papain digestion of IgG2 was not properly optimized to produce the Fc domain. The $(Fab)_2$ domains of NIST IgG1, IgG2 and IgG4 had similar CCS values. It seems that papain cleaved lower in the hinge region of IgG1 compared to the other IgG subtypes as the observed $(Fab)_2$ has a relative higher molecular weight.

Fast photochemical oxidation of proteins (FPOP) was used to oxidize antibodies and to see the effect on its native structure. Based on the results of native IM-MS, the native structure was not significantly influenced by the oxidation via FPOP. Unfortunately as the digest of antibodies did not happen correctly, the sites of oxidation could not be determined.

With size exclusion mass spectrometry (SEC-MS) a mixture of the NIST IgG1 and CNTO5825 was analysed. The two antibodies could be separated on the chromatogram based on their slight difference in size. The mass spectra obtained after SEC-ESI-MS showed a more unfolded form of the antibodies compared to native nanoESI-MS with direct infusion.

3.2.11 Sources

[1] J. E. Schiel, D. Davis, O. V. Borisov, State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization, 2015, 1201

[2] Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Analytical Chemistry* 2010, *82* (22), 9557-9565.

[3] S. J. Berger, A. B. Chakraborty, J. C. Gebler, Waters, Development of a generic LC/MS methodology for protein-level analysis of IgG1 monoclonal antibodies and their related substructures.

[4] K. J. Pacholarz, M. Porrini, R. A. Garlish, R. J. Burnley, R. J. Taylor, A. J. Henry, P. E . Barran, Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion-Mobility Mass Spectrometry and Molecular Modeling, Angew Chem Int Ed Engl, 2014, 7765-9

[5] I. D. G. Campuzano, C. Larriba, D. Bagal, P. D. Schnier, Ion Mobility and Mass Spectrometry Measurements of the Humanized IgGk NIST Monoclonal Antibody, State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques, 2015, Chapter 4, pp 75–112

Received: 25 August 2017 | Revised: 29 September 2017 | Accepted: 29 September 2017

DOI: 10.1002/jssc.201700988

REVIEW ARTICLE

JOURNAL OF SEPARATION SCIENCE

Advances in native high-performance liquid chromatography and intact mass spectrometry for the characterization of biopharmaceutical products

Marco Tassi[1] | Jelle De Vos[1] | Sneha Chatterjee[2] | Frank Sobott[2,3,4]✉ | Jonathan Bones[5] | Sebastiaan Eeltink[1]

[1]Department of Chemical Engineering, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[2]Biomolecular & Analytical Mass Spectrometry, Antwerp University, Antwerp, Belgium

[3]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK

[4]School of Molecular and Cellular Biology, University of Leeds, Leeds, UK

[5]The National Institute for Bioprocessing Research and Training (NIBRT), Dublin, Ireland

Correspondence
Dr. Sebastiaan Eeltink, Vrije Universiteit Brussel, Pleinlaan 2, Belgium
Email: sebastiaan.eeltink@vub.be

✉Additional corresponding author
Professor Frank Sobott
Email: f.sobott@leeds.ac.uk

The characterization of biotherapeutics represents a major analytical challenge. This review discusses the current state-of-the-art in analytical technologies to profile biopharma products under native conditions, i.e., the protein three dimensional conformation is maintained during liquid chromatographic analysis. Native liquid-chromatographic modes that are discussed include aqueous size-exclusion chromatography, hydrophobic interaction chromatography, and ion-exchange chromatography. Infusion conditions and the possibilities and limitations to hyphenate native liquid chromatography to mass spectrometry are discussed. Furthermore, the applicability of native liquid-chromatography methods and intact mass spectrometry analysis for the characterization of monoclonal antibodies and antibody–drug conjugates is discussed.

KEYWORDS

bioprocessing, native chromatography methods, native mass spectrometry, protein therapeutics

1 | INTRODUCTION

The demand for biopharmaceuticals, defined as pharmaceutical products originating from modern molecular biology methods, is rapidly increasing due to their successful application in the treatment of various cancers and inflammatory diseases. Currently, there are more than 200 approved drugs available and the global market is expected to soon reach $278 billion [1,2]. Moreover, it is anticipated that >50% of new drug approvals will be biologics, rising to >70% by 2025 [3].

Innovations concerning the development of novel therapeutic proteins can be categorized into four groups depending on their pharmacological activity [4]. The first group involves protein therapeutics with enzymatic or regulatory activity that are prescribed to patients that exhibit protein-related deficiencies [5]. For example, a growth hormone deficiency due the lack of a specific protein that results in failure to grow at the expected rate. The second group concerns protein therapeutics with a special targeting activity. Protein therapeutics include peptides and protein derivatives [6], monoclonal antibodies (mAbs) that interact and interfere with a molecule or organism [6], and antibody–drug conjugates (ADCs) that act as a vehicle to deliver drugs to a specific biological site [7]. The third group involves protein vaccines that are used in the protection against deleterious infectious agents [8]. The fourth class regards protein diagnostic reagents that are used in clinical decision making [9].

Abbreviations: ADC, antibody–drug conjugate; AEX, anion exchange chromatography; CEX, cation exchange chromatography; CHO, Chinese hamster ovary; DAR, drug-to-antibody ratio; E. coli , Escherichia coli ; HIC, hydrophobic interaction chromatography; IgG, immunoglobulin G; mAb, monoclonal antibody; PTM, posttranslational modification

Conflict ofinterest: The authors have declared no conflict ofinterest.

mAbs and ADCs represent emerging classes of therapeutic agents. Over the last years more than 60 antibody derivatives have been approved by regulatory authorities for the treatment of various diseases including cancer [10], multiple sclerosis [11], rheumatoid arthritis [12], and asthma [13]. Recombinant mAbs (≈150 kDa) are composed of two identical heavy chains and two identical light chains linked by disulfide bridges, yielding a distinct Y-shape appearance. The part of the antibody which contains the antigen binding site is called the fragment of antibody binding. Large-scale production of mAbs mainly occurs in mammalians cell cultures using host cells such as Chinese hamster ovary (CHO) cells and mouse myeloma cells such as NS0 or SP2/0 [13]. Biopharmaceutical proteins of other classes, e.g., protein therapeutics with enzymatic or regulatory activity, are mainly produced by microorganisms such as bacteria and yeasts. Mammalian cells are used for the expression of glycosylated forms of these molecules such as enzyme replacement therapies due to the requirement for specific glycan based epitopes, e.g,. mannose-6-phosphate, needed for delivery to the lysosome upon administration. mAbs were traditionally developed and produced using hybridoma technology, i.e., methods in which hybrid cell lines are cultivated [14]. These cells combine the ability to produce large amounts of mAbs, derived from the B lymphocytes of an immunized animal, with the immortality and high rate of reproducibility of cancer cells, derived from immortalized myeloma cells [15]. More modern approaches for the development of monoclonal antibodies include the use of techniques such as phage display and humanized mouse models for target discovery followed by molecular optimization and expression of the developed mAb using industrial scale CHO cell culture. Antibodies provide the link between the innate and adaptive immune systems thereby requiring two specific features for optimal response: (1) high and specific antigen binding as determined by the complementarity determining regions encoded within the variable regions of the light and heavy chains and (2) the ability to interact with Fc receptors present on innate immune cells such as macrophages and monocytes to stimulate the immune response [16,17]. Glycosylation is the biological process in which the addition of glycans or polysaccharides to the antibody takes place. Advances in protein engineering, e.g., incorporation of non-natural amino acids, have facilitated the development of mAb-related products such as site-specific antibody–drug conjugates and biospecific antibodies [18,19]. Antibody–drug conjugates (ADCs) are biochemotherapeutical agents that combine the specificity of the mAb with the cytotoxic (anticancer) drug [20]. ADCs are produced by conjugation of the naked mAb with small drugs that exert cytotoxic activity. This class of therapeutics is extremely promising in cancer treatment, and whereas some are already commercially available, e.g., Brentuximab vedotin [21] and adotrastuzumab emtansine [22], many others are under development and investigation [23]. Bispecific antibodies (bs-mAbs) can interact with two different antigens at the same time, allowing highly efficient cancer treatment. Also, composite mixtures of mAbs are being exploited as novel biopharmaceutical products. bs-mAbs and composite mAb mixtures enlarge the molecular complexity of drug candidates, putting even greater demands on the analytical tools to characterize them [23].

Biopharmaceuticals are much more complicated to characterize than traditional small molecule active pharmaceutical ingredients. Regulatory guidelines require the characterization of the primary sequence, posttranslational modifications (PTMs), and higher order structures present on these molecules, using methods such as LC and MS [24]. These analyses are necessary to ensure that the quality of these biopharmaceuticals is maintained and to ensure the absence of unwanted PTMs such as nonhuman glycosylation epitopes, e.g. galactose alpha 1–3 galactose, or the presence of aggregated forms of the drug product and sub-visible particles that may be potentially immunogenic [25,26]. Chromatographic techniques such as SEC, cation exchange chromatography, and hydrophobic interaction chromatography (HIC) have, for many years, been the gold standard for the characterization of aggregates and higher order structures, charge variants and structural variants arising from PTMs such as oxidation, etc. [24]. The considerable advancements in stationary phase technology, combined with the advent of high-resolution MS under native conditions, represent key advances for the characterization of biopharmaceuticals [27]. As these recombinant proteins exist and exhibit their pharmacological functions as structured molecules, LC and MS methods that enable the characterization of these molecules in their native state are becoming more and more important as although still required and powerful, bottom-up approaches such as peptide mapping can often result in the loss of fine detail that exists on the molecule in its native form [28]. The ability to hyphenate native LC separation chemistries with high-resolution native MS represents an emerging and important tool that will provide information that will enable the linking of sequence to structure and potential functional implications [29].

The present review aims at providing a comprehensive overview on native LC workflows and native MS strategies applied for the characterization of biopharmaceutical products. Different native LC separation modes, including aqueous SEC, HIC, and IEC are discussed. Aspects of method optimization are discussed and major applications realized with the different native LC modes are highlighted. In addition, the application possibilities ofintact MS for the characterization of biopharmaceutical products are discussed and aspects of hyphenation to native LC are debated.

TABLE 1 Overview of SEC columns applied for the separation therapeutic proteins frequently reported in scientific literature

Column	Matrix	Chemistry	Particle size (µm)	Pore size (Å)	pH stability	Max Pressure (MPa)	Brand
Advanced Bio SEC	Silica	Silanol	2.7	300	2–8.5	10	Agilent Technologies
Yarra SEC-X300	Silica	Silanol	1.8	300	1.5–8.5	48	Phenomenex
Unix-C SEC 300	Silica	Diol	1.8	300	2–8.5	31	Sepax Technologies
Zenix SEC-300	Silica	Diol	3	300	2–8.5	7	Sepax Technologies
MAbPAc SEC 1	Silica	Diol	5	300	2–7.5	7	Thermo Fisher Scientific
TSKgel SuperSW mAb HR	Silica	Diol	4	250	2–7.5	8	Tosoh Bioscience
TSKgel UltraSW Aggregate	Silica	Diol	3	3000	2–7.5	12	Tosoh Bioscience
Protein-Pak SEC	Silica	Diol	10	300	2–8	30	Waters
XBridge Protein BEH SEC	Silica	Silanol	3.5	200	1–8	Not available	Waters

2 | NATIVE LC MODES

2.1 | Aqueous size-exclusion chromatography for the analysis of protein aggregates and fragments

In aqueous SEC, biomacromolecules are separated based on their difference in hydrodynamic volume, and hence on the difference in accessibility of proteins to the intraparticle pore volume of the resin (typically varying between 35 and 41%) in absence of solute interactions with the stationary-phase surface. The first size-based separation of biomolecules, i.e., peptides from amino acids, was reported by Lindqvist and Storgards using a column packed with starch [30]. The premier application area of aqueous SEC with respect to the characterization of therapeutic proteins is the quantitative determination of protein aggregation. Information on the molecular mass of monomeric proteins, possible aggregates, or protein fragments is typically obtained based on a calibration curve created using protein standards. The prediction error on molar–mass accuracy provided in this way is typically around 12% (when applying a flow rate matching the minimum of the Van Deemter curve) [31]. Table 1 provides an overview offrequently employed SEC columns, including particle and pore size of the resins and corresponding application area with respect to the characterization of biopharmaceutical products.

Whereas the selectivity provided by the SEC column is defined by the size of the intraparticle pore diameter, the efficiency in an SEC separation is (partially) governed by the particle diameter. SEC is considered a slow and low-resolution technique, especially compared to current-state-of-the-art RP–LC columns. Due to low diffusivity of macromolecules, the optimum flow rate (corresponding to the minimum plate height in the Van Deemter curve) is very low, and in practice modern SEC columns are operated in the C-term region of the Van Deemter curve. Often columns packed with rather large particles (5µm) are being employed, hence, relatively long interparticle diffusion distances compromise the separation efficiency due to resistance to mass transfer effects. De Vos et al. discussed the need to downscale particle size to maximize resolution while exploiting the current column-pressure limitations of 20 MPa [31]. Within this pressure range it was demonstrated that SEC separations could be conducted without compromising the selectivity or altering the protein conformation by shear effects. Furthermore, it was demonstrated that a factor of 2 in analysis time could be gained when using 3µm SEC resins instead of 5 µm particles, and optimizing the column length-to-particle-diameter ratio, such that the column efficiency was maintained [31]. The evaluation of SEC columns packed with sub-3 and sub-2µm particles for the analysis of mAbs and ADCs was described by Fekete et al. showing that an additional gain in time can be achieved without compromising analysis time [32,33]. The same group also reported the risk offorming on-column aggregates when applying small-particle columns under high-pressure conditions [34]. An alternative approach for method speedup was demonstrated by Diederich et al. who reported on a sub-2 min method for mAb aggregate analysis using a parallel interlaced SEC [35], following an approach described by Farnan et al. [36].

Derivatized porous silica has become the gold standard stationary-phase resin for SEC columns applied to biomacromolecule analysis. To reduce strong ionic interactions induced by acidic surface-silanol moieties, different surface procedures have been investigated. Diol-modified silica particles have emerged as current state-of-the-art, reducing ionic interactions and yielding minimal secondary hydrophobic

interactions. In 2010, SEC columns packed with porous hybrid organic/inorganic particulate material modified with diol chemistry became commercially available, apparently reducing residual surface silanol activity, improving pH stability, and increasing the mechanical strength and pressure rating of the columns [37,38]. It is important to note that the adjustment of the mobile-phase pH (between pH 5.5 and 8.5) and ionic strength (≤100 mM) is still required to counteract all interactions with residual silanol moieties. Kopaciewicz and Regnier reported on the effects of mobile phase pH and ionic strength on nonideal protein elution behavior [39]. Applying low ionic strength mobile phases (<100 mM phosphate buffer), electrostatic interactions may affect protein retention. At salt concentrations >500 mM, hydrophobic interaction effects may occur, see also discussion in 'Section 2.2'. Furthermore, the extent of these interactions was determined to be protein specific. Ricker and Sandoval validated these findings for the SEC analysis of mouse myeloma antibodies of similar molecular weight but of varying overall charge [40]. For weakly basic antibodies, good peak shapes and retention-time accuracies were observed applying mobile-phase ionic strengths between 50 and 400 mM phosphate buffer pH 7. At ionic strengths >600 mM peak broadening occurred and the retention time increased due to hydrophobic interactions. For a strongly basic antibody, an ionic strength of 200 mM was found to be optimal for the SEC analysis, with respect to retention time and peak shape. Reducing the ionic strength led to increased retention times due to electrostatic interactions. At a concentration of 400 mM and higher, peak broadening was observed and ultimately the peak profile shifted to higher retention-time values due to hydrophobic-interaction effects affecting the size-based separation. The effect of sodium and potassium additives on protein aggregation was investigated by Goyon et al. [41]. When comparing the ratios between high-molecular species and monomers for a large number of different mAbs and ADCs, no systematic trend in aggregation level was detected. Experiments showed that the addition of sodium or potassium to the mobile phase may, to a certain extent, affect the aggregation level, but this is likely a protein-specific effect [42].

Another factor found to critically affect the SEC performance is the pH of the mobile phase, since pH affects the equilibrium between charged and uncharged forms offunctional groups on both the column resin and the proteins, the latter determined by the pI, typically varying between 4 and 9 for antibodies [43]. Ricker et al. also conducted SEC experiments for mAbs applying a mobile phase with a pH range between 7.0 and 5.5 [40]. When lowering the mobile-phase pH, protein retention increased, hence application of higher ionic-strength mobile phases was required to mediate electrostatic interactions. When applying pH 5.5 and high ionic strength mobile phases, no peak broadening or shift in retention time were witnessed. At pH 5.5, the antibody became more positively charged (shielded by the higher salt content), which makes the antibody more polar, reducing its tendency for hydrophobic interactions.

During manufacturing and storage of biopharmaceuticals, size variants can arise that can alter the safety and efficacy of the product. Although SEC is known as the chromatographic mode with low efficiency and resolution, it is extremely powerful to assess aggregation and fragmentation. Figure 1 shows the SEC analysis of a Protein A purified mAb recombinantly expressed in CHO cells. This example perfectly illustrates the suitability of the technique in highlighting the presence of high and low molecular-weight variants. An important quality-control parameter that needs to be assessed during the production and storage of mAbs is the dissociation pattern of the hinge polypeptide connecting the fragment antigen binding part to the rest of the antibody. In a recent paper, Dada et al. correlated hinge fragments measured by SEC with a complementary CE-SDS electropherogram [44]. Another important SEC application is the determination of the molecular weights of the antibody light and heavy chains. Liu et al. compared the performance of an optimized SEC method with that of a gradient RP-LC method [45]. Whereas the retention time of the intact protein and heavy chain fragment coincided in the RP-LC methods, baseline resolution could only be achieved between intact antibody, the heavy chain, and light chain fragments with SEC. An SEC method to determine the ratio offree therapeutic mAbs and antidrug antibody complexed mAb in the serum of animals was described by Boysen et al. [46].

Hyphenation of LC, including aqueous SEC, to MS detection is desired to obtain accurate mass information. Kükrer et al. described an offline SEC–MS workflow for the analysis of dimer, trimer, and tetramer aggregates of stressed intact human mAb (IgG) [47]. A volatile ammonium acetate buffer system yielded poor chromatographic separation and MS performance. To overcome this problem, monomeric and aggregate IgG fractions were collected using SEC, applying a conventional 0.1 M phosphate buffer at pH 7.2, followed by dialysis of the biomacromolecule fractions and ESI–TOF-MS. Reanalysis of the dialyzed samples by SEC indicated that the oligomeric state of the different fractions was not measurable affected [47]. Shen et al. developed an online native SEC–MS workflow (which included a flow splitter reducing the solvent and salt intake before ESI) to study the effect of enzyme inhibitors on the protein quaternary structure [48]. Valliere-Douglas et al. presented a native SEC-based desalting method for analyzing cysteinyl-linked ADCs [49]. They also studied post-desalting dissociation of the denatured ADC during ESI-ionization by comparing with an orthogonal HIC separation of the mAbs conjugated with 0–8 drugs [49].

Different SEC–ESI–MS approaches have also been developed and applied to the characterization of biotherapeutics that include the application of organic solvents in their

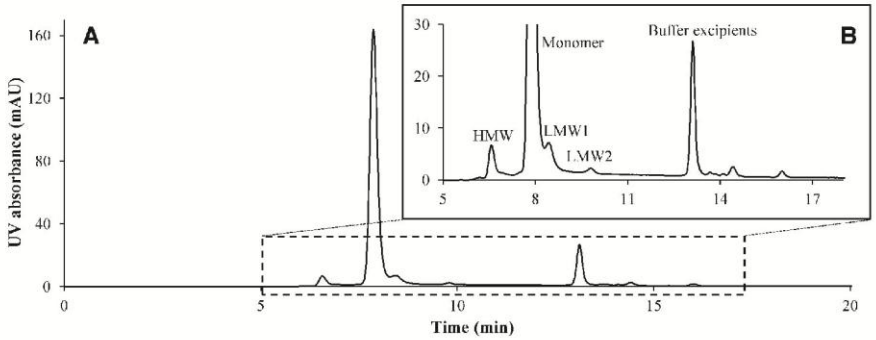


FIGURE 1 SEC analysis of a Protein A purified monoclonal antibody recombinantly expressed in Chinese hamster ovarian (CHO) cells performed on a 7.84.0 mm id× 300 mm long AdvanceBio SEC column packed with 2.7µm particles containing 300Å pores. Separation conducted applying a mobile phase of 150 mM sodium phosphate pH 7, a flow rate of 0.8 mL/min, and UV detection at 220 nm

workflow to advance the ESI spray stability [50–52]. Adding organic modifiers to the mobile phase is also frequently performed to suppress hydrophobic interactions and reduce peak tailing when analyzing highly hydrophobic biomacromolecules, such as ADCs [52,53]. It is highly probable that workflows that include organic solvents affect protein conformation, biological/enzymatic activity of biomacromolecules, protein–biomolecule interactions, and to certain extent also aggregation level. Although such workflows may be valuable, providing insights in the chemical structure, these workflows are not regarded as pure native LC. To further enhance the flow rate compatibility of SEC with MS detection and reduce the salt intake, it is mandatory to develop column technology with reduced column id. The number of SEC applications developed using sub-1 mm columns is limited. Rea et al. reported the use of 300µm id capillary SEC columns for mAb analysis purified from harvested cell culture fluid. After optimizing the fluidics to minimize system dispersion, picogram sensitivity was achieved in combination with UV detection [54]. Smoluch et al. applied a 300µm id column format for the online SEC–ESI–MS analysis of peptides in a mass range of 0.1–7 kDa [55].

To increase the performance of SEC, different aspects with respect to column technology and instrumentation need to be addressed. Whereas column-packing procedures for SEC columns with 5 µm particles have been fully optimized, and columns deliver reduced plate heights (h) of around 2, columns packed with small particles diameters do not yet reach their full expected kinetic performance [56]. Hence, column packing techniques to establish SEC columns need to be advanced. Also, column stability is deemed to be an issue. Recently Farrell et al. demonstrated the long-term stability for a current state-of-the-art SEC column packed with 5µm particles allowing for over 1500 consecutive runs, analyzing Bevacizumab aggregates, see Fig. 2 [57]. Similar experiments are required to demonstrate the robustness and the applicability of SEC columns packed with small particle diameters

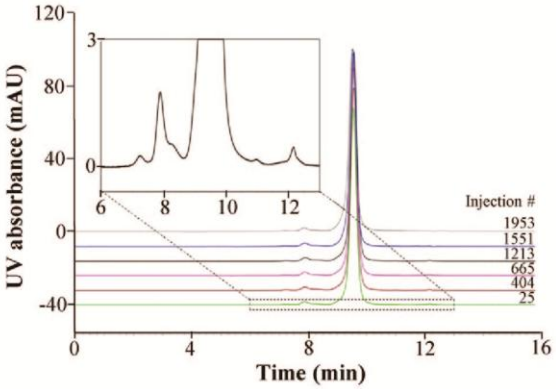


FIGURE 2 Overlay of selected SEC chromatograms extracted from over 1500 injections of bevacizumab performed on a 4.0 mm id × 300 mm long MAbPac SEC-1 column packed with 5µm macroporous particles applying 100 mM sodium phosphate pH 6.8 in 300 mM NaCl as the mobile phase. Adapted with permission from [57]

in a QC environment. To further enhance the kinetic performance, core–shell particles for SEC separations may represent a good alternative to columns packed with fully porous particles. Selectivity will be impaired, but the loss in selectivity will be small since more than 60–75% of the intraparticle pore volume is maintained. Similar to RP-LC, a gain of roughly 25% in efficiency can be expected due to improved A-, B-, and C-term characteristics [58,59]. Pirok et al. demonstrated the applicability of core–shell particle technology for SEC separations of polymers [60]. Columns packed with core–shell particles displayed outstanding resolution for specific (low molecular) weight polymer separations. Furthermore, a gain in analysis speed amounting up to one order of magnitude was demonstrated.

Peak volumes provided by columns packed with sub-3µm particles and small id columns are significantly lower than obtained using conventional SEC column technology. Hence, to preserve the high efficiencies provided by these columns,

it is important that the fluidic path is optimized with respect to extra-column dispersion. System-design requirements and aspect of tubing configurations influencing the separation performance have been described in a review by De Vos et al. [61]. The importance of system dispersion affecting high-resolution SEC separations has been addressed by Goyon et al. [41]. Moreover, when using small particle columns, thermal heating and possible shear-degradation effects need to be anticipated [62].

2.2 | Hydrophobic interaction chromatography for profiling differences in surface hydrophobicity

In 1948, Shepard and Tiselius reported on HIC using the term 'salting-out chromatography', observing that biomolecules bind to a hydrophobic surface material in the presence of salt [63]. Over the last years, HIC has gained significant importance for the characterization of biotherapeutics, allowing to obtain complementary information to RP-LC [64]. In contrast to RP-LC, in HIC mode, nondenaturing LC conditions are applied and hence, protein conformation and biological/enzymatic activity are maintained during the separation. When proteins are introduced in an aqueous environment, the protein surface will be shielded by ordered layers of water molecules, preventing hydrophobic interactions with the stationary phase [65]. During an HIC analysis, salt ions in the mobile phase lead to exclusion of water molecules from the surface, and the breakdown of the ordered layer is concomitant with an increase of entropy [66]. This favors the formation of hydrophobic noncovalent interactions between the hydrophobic patches situated at the proteins surface and the hydrophobic moieties on the stationary phase, decreasing the free energy. Protein elution based on difference in hydrophobic surface area is achieved by decreasing the salt concentration of the mobile phase in time.

The number of stationary phases available for HIC separations is relatively limited. This may be because effects of surface chemistry on protein conformation and hence HIC retention are still under debate. An overview offrequently used HIC columns and corresponding biopharma applications is provided in Table 2. Conventional columns are packed with 5 µm diameter particles. Typically, HIC resins are less hydrophobic as compared to their counterparts used in RP-LC. The most common column material used in HIC is either surface-modified silica or polymeric particles coated with short aliphatic groups, i.e., butyl-, hexyl-, or octyl-chains [67,68]. Whereas these columns are suitable for the analysis of highly hydrophobic biomacromolecules, particles functionalized with alkylamide functionalities, polyalkylimide chemistries, and alkyl ethers are applicable for the analysis of biomolecules with a wide range in hydrophobic surface area, including hydrophilic proteins [68]. It should be noted that different stationary-phase materials also induce protein specific retention effects [69].

HIC is typically performed applying an inverse ammonium sulfate gradient in 50–100 mM phosphate buffer pH 7. Protein retention is strongly affected by the salt concentration and the type of salt employed. The Hofmeister series, providing information on ions that stabilize the structure of proteins, has frequently been used to predict protein retention in HIC mode [70]. Soluble compounds that are well hydrated and form hydrogen bonds to water molecules will exclude water molecules from the protein and resin surface, hence promoting hydrophobic interactions. Salts that promote the formation of hydrophobic interactions are called kosmotropic, while salts that do not exhibit this property are called chaotropic [71,72]. However, Arakawa noticed that certain salts, including sodium phosphate and magnesium chloride, promote hydrophobic interactions regardless of their classification within the Hofmeister series [73]. Sodium chloride and ammonium acetate salts have been used to replace ammonium sulfate considering that the elution strength of 1 M ammonium sulfate is equivalent to ≅ 2.6 M sodium chloride and ≅ 3.3 M ammonium acetate [74]. Typically, kosmotropic salt systems are compatible with the analysis of hydrophilic biomacromolecules, whereas chaotropic salt systems are compatible with HIC analysis of hydrophobic proteins. To decrease retention of highly hydrophobic proteins, organic modifiers, including isopropanol and acetonitrile, are frequently added to the mobile phase [74]. The Eeltink research group recently demonstrated that the addition of only 2.5% ofisopropanol to the mobile phase may lead to protein conformational changes, significantly affecting the peak profile [69]. Complementary differential scanning calorimetry analysis demonstrated that the addition of a small amount of organic modifier leads to the denaturation of the protein investigated (α-lactalbumin) [69].

Two other parameters that influence the protein retention in HIC mode are the mobile-phase pH and the column temperature. The impact of pH depends on the pI of the protein. Good practice is to minimize the shift in pH between the pI of the protein and the pH of the mobile phase, to prevent possible 3D conformation changes, affecting the level of protein aggregation, or even induce protein denaturation. The effect of temperature on HIC separations is still under investigation [75,76]. Generally, it can be affirmed that an increase of temperature (i.e., column temperature) drives an increase in protein retention. The formation of hydrophobic interactions is an entropy driven process and the temperature increase favors a decrease in free energy. On the other hand, the increase of (column) temperature can induce undesired conformational changes of proteins, and possibly lead to a change in the strength of the hydrophobic interaction when the surface hydrophobicity is altered [77]. A safe range is retained to be in the temperature interval between 20 and 40 °C [78].

TABLE 2 Overview of HIC columns applied for the separation therapeutic proteins frequently reported in scientific literature

Column	Matrix	Chemistry	Particle size (µm)	Pore size (Å)	pH stability	Max Pressure (MPa)	Brand
Proteomix HIC 1.7	PS/DVB	Butyl/Ethyl	1.7	Non porous	2–12	50	Sepax Technologies
Proteomix HIC 5	PS/DVB	Phenyl/Butyl/ Propyl/Ethyl	5	Non porous	2–12	41	Sepax Technologies
MAbPAc HIC-10	Silica	Alkyl amide	5	1000	2–8	55	Thermo Fisher Scientific
MAbPAc HIC-20	Silica	Alkyl amide	5	1000	2–9	55	Thermo Fisher Scientific
MAbPAc HIC-Butyl	Polymer	Poly amide	5	Non porous	2–12	27	Thermo Fisher Scientific
TSKgel Butyl-NPR	Polymetha-crylate	Butyl	2.5	Non porous	2–12	20	Tosoh Bioscience
TSKgel Phenyl-5PW	Polymetha-crylate	Ether	13, 10	1000	2–12	2	Tosoh Bioscience
TSKgel Ether-5PW	Polymetha-crylate	Polyamine	10	1000	2–12	2	Tosoh Bioscience
Protein-Pak Hi Res HIC	Polymetha-crylate	Ether	10	Non porous	2–12	20	Waters

HIC has been successfully applied to characterize mAbs with respect to profiling PTMs, including monitoring of oxidation variants [79], aspartic acid isomerization [80], and domain misfolding [81]. In particular, oxidation of the amino acids exposed to the storage environment and micro-heterogeneities in the carboxy terminal chains are common PTMs that need to be monitored to guarantee the quality of mAbs products. Boyd et al. described the separation of native IgG1 from its oxidized Trp counterpart [79]. The authors also claimed that the HIC approach allows for profiling of oxidized methionine and isomerization/deamidation products. A comprehensive study to characterize mAbs variants resulting from variable N-and C-terminal processing and stress-induced modifications using HIC technology was performed by Valliere-Douglass et al. [80]. In this study, the authors also demonstrated the applicability of HIC to separate truncated antibodies from native species.

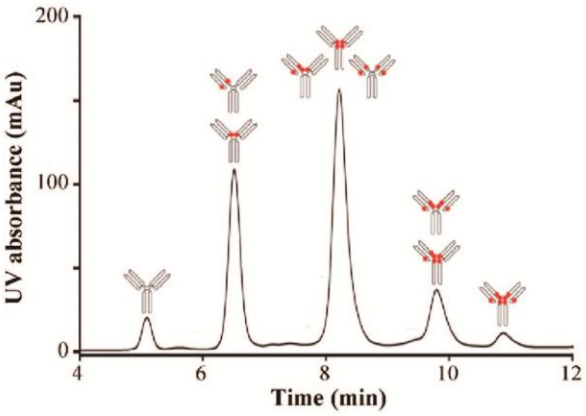


FIGURE 3 HIC separation of ADCs having different payload in which the retention time increases with increasing DAR. Adapted with permission from [81]

One of the key HIC applications is the determination of the average load of cytotoxic drug with respect to the antibody, i.e., the average drug-to-antibody ratio (DAR) of ADCs. Having information of the average DAR is essential, since this value determines the quantity of cytotoxic drug that will be transported to the targeted tumor cell, defining the efficacy of the chemotherapeutical distribution. Figure 3 shows the HIC separation of ADCs having different payloads [81]. The peaks were assigned using the unmodified antibody for the zero-drug peak and the absorbance ratio measured at 248 and 280 nm for the other peaks, since the drug and antibody have distinct absorbance maxima at these wavelengths. The cytotoxic drugs applied are typically hydrophobic, hence when the payload increases also the HIC retention time increases and the DAR ratio can be calculated by summation of the individual peak areas multiplied with their respective drug load divided by the total peaks area. Depending on the type of mAb (IgG1 or IgG2) used, the DAR varied between 2 and 8 for IgG1 and between DAR 2 and 12 for IgG2 [82]. DAR 0 refers to the mAb in which the conjugation with the cytotoxic drug did not occur, while odd DAR numbers (normally present in negligible amounts) refer to ADC in which the conjugation is incomplete. The latter two cases are considered as impurities in ADC analysis. In the case of ADCs derived from IgG1, different positional isomers can be present in the

DAR 2, DAR 4, and DAR 6 forms. Unfortunately, HIC has no sensitivity towards positional isomers whereas CE–SDS-PAGE [83] and also ion mobility may well have. The characterization of ADCs and their payloads using comprehensive LC modes has been described in an excellent review by Bobaly et al. [53].

2.3 | Ion-exchange chromatography for the analysis of charge variants

The relevance of IEC in biochemical studies was demonstrated already in 1949 by Cohn, who performed cation and anion-exchange separations for a trace-analysis study on the enzymatic formation and degradation of nucleic acids [84]. In recent years, IEC has been widely applied to monitor product quality and consistency of biotherapeutics. The separation is based on coulombic interactions between the stationary-phase surface, containing ionic functional groups, and the charges of the therapeutic protein. Since the disposition of charges at the protein surface depends on the native 3D protein conformation, proteins having structural diversities can be differentiated by means of IEC. The net charge of a therapeutic protein is not only determined by the amino acid residues on the protein backbone, but also charged glycans are accounting for a portion of the net charge of the protein. These charges not only affect the structure of the protein, and thus determine the stability and solubility of the therapeutic product, the charges also affect the binding affinity to receptors and functional groups of the stationary phase, influencing its biological activity [85,86]. The versatility of IEC in protein analysis is related to the fact that a wide range of separation conditions with respect to salt concentrations and pH are applicable. An overview offrequently used columns for the IEC characterization of biotherapeutics is provided in Table 3. The maximum pressure rating of the current commercially available material is currently 40 MPa, limiting the application of IEC under UHPLC conditions and thus also its possibility for method speedup. The majority of applications are performed using 4.6 or 2.1 mm id column formats. Rea and Farnan reported on the use of capillary columns formats, i.e., 400μm id columns packed with 5μm pellicular strong cation-exchange particles and 300μm id columns packed with 1.7μm nonporous weak cation exchange particles for the separation of mAb charge variants [87].

IEC separations can be performed using a salt gradient while keeping the mobile phase pH constant. This increase in ionic strength of the mobile phase promotes protein elution as the salt ions compete with the adsorbed protein molecules for the ion-exchange sites on the resin. Salt gradients provide good resolving power and robustness, but are product specific and time consuming to develop. Sodium chloride, usually dissolved in a <50 mM sodium phosphate buffer, is the most-employed eluent for separating proteins using salt gradients [88–90]. It is assumed that NaCl does not affect protein conformation. As the nature of the buffer cation and anion can affect protein retention and peak widths, the selection of the ideal salt buffer system is very important. [91,92]. The effects of eluent salts on the resolution of protein separation has been described by Gooding et al. and Regnier et al. [93,94]. Not all charge variants are generally resolved using a salt gradient in IEC mode, especially the acidic variants [95]. As the pH remained constant during the elution process, proteins with the same effective charge will be eluting with poor resolution. Nevertheless, the potential of cation exchange chromatography (CEX) for mAb characterization, applying a shallow gradient ofincreasing salt concentration (typically 200 mM NaCl) at constant pH, has been reported in several publications [88,96]. Flattening of the salt gradient only improves the resolution if the pH of the separation is operated near the pI of the proteins to be analyzed. As it is demanding in a high-throughput QC environment of biopharmaceutical industry to tailor salt systems for individual mAbs, alternative elution approaches are preferred.

Proteins can also be eluted in IEC mode by generating a pH gradient across the column. Irrespectively of how the pH gradient is formed, two modes of chromatofocusing can be distinguished, i.e., cation chromatofocusing where the stationary phase exhibits cation-exchange properties and a gradient running from low to high pH is generated, and anion chromatofocusing which employs an anion-exchange resin and proteins are eluted by applying pH gradient going from high to low pH. Generating pH gradients in IEC mode is generally called chromatofocusing, which is a pressure-driven chromatographic variant of IEF elution mechanism coined by Sluyterman and Elgersma [97–99]. Whereas conventional chromatofocusing uses an "internally generated" pH gradient, gradient chromatofocusing employs an "externally generated" pH gradient. In the former variant of this separation method, the buffer capacity of the stationary phase is used to convert a step change in pH after applying a mobile-phase of a given pH at the column inlet, while the IEC resin is preadjusted at a different initial pH. In this way, an internally generated pH gradient is generated as the packing material will buffer the pH step. This traveling pH wave allows to focus proteins, and releasing them once the pH gradient approaches the pI of the biomacromolecule. To generate an internal pH-wave, either an immobilized ampholytic buffer bound to a strong ion exchange resin, or noninteracting buffer species in conjunction with a weak ion-exchange resin being used [100–102]. The major challenge is to precisely generate the required pH gradient, while minimizing the ionic strength of the running buffer to reduce its effect on protein retention. The conventional elution buffers are polyampholytes. These molecules provide a high buffer capacity covering a broad pH range, but are poorly defined, and have been reported to interact with both the proteins as with the stationary phase resin [103]. Alternatively, a

TABLE 3 Overview of IEC columns applied for the separation therapeutic proteins frequently reported in scientific literature

Column	Matrix	Chemistry	Particle size (μm)	Pore size (Å)	pH stability	Max Pressure (MPa)	Brand
Agilent Bio SCX	PS/DVB	Sulfonic acid	10, 5, 3, 1.7	Non porous	2–12	68	Agilent Technologies
WP CBX	Silica	Sulfonic acid	5	300	2–8	45	Avantor Inc
Antibodix WCX	PS/DVB	Carboxylate	10, 5, 3, 1.7	Non porous	2–12	68	Sepax Technologies
Proteomix SCX	PS/DVB	Sulfonic acid	10, 5, 3, 1.7	Non porous	2–12	68	Sepax Technologies
BioBasic SCX LC	Silica	Sulfonic acid	5	300	2–8	40	Thermo Fisher Scientific
MabPac SCX	Polymer	Sulfonic acid	10	Non porous	2–12	20	Thermo Fisher Scientific
TSKgel Q-STAT	Polymer	Quaternary ammonium	10	Non porous	3–10	5	Tosoh Bioscience
TSKgel Bioassist Q	Polymetha-crylate	Polyamine	13, 10	4000	2–12	2	Tosoh Bioscience
Protein-Pak HiRes CM	Polymetha-crylate	Carboxymethyl	7	Non porous	3–10	15	Waters

combination of equally concentrated buffer species with equally spaced pK_a values in the chosen pH range can be employed. Kröner et al. provided an in silico optimization method of buffer compositions, resulting in well-controllable pH gradients with low ionic strength validated for characterization of more than 20 proteins [104].

Alternatively, chromatofocusing can also be performed by applying an externally generated pH gradient, i.e., by use of the gradient proportioning system of the LC pump. By gradually mixing the running buffer with successively greater proportions of an application buffer, while both buffers are set at different pH, a pH gradient is generated in time before entering the column. At the start of the pH gradient proteins are adsorbed on the column head and the proteins elute once the incoming pH gradient is slightly below the pI of the protein. The quality of the separation thus depends strongly on the solvent-proportioning capabilities of the LC equipment, as poorly controlled pH gradients can result in coelution of proteins with similar pI values. The formation of multistep or multivariable slope (non)-linear gradients over a wide pH range, and a buffer system compatible with both anion-and cation-exchange stationary phases that allows for an arbitrary start-and end-pH value and pH range are still not available for this separation mode [105]. Tsonev and Hirsch developed software that can precisely perform high-order polynomial fitting of titration curves for a dedicated buffer system, allowing controlled gradient formation of any desired shape and slope [105,106] for both cation-and anion-exchange separations. Furthermore, the algorithm also allows for software-driven control of pH gradients that can contain additives such as nonionic detergents, organic modifiers, salts, etc. Figure 4 shows the comparison between the optimization of an anion exchange separation of *Escherichia coli* acetone powders performed using a salt gradient (Fig. 4A), and using a pH gradient (Fig. 4B). The steepness of the salt gradients was decreased, at the expense of analysis time, whereas for the pH gradient separations only the slope of the gradient between pH 3.5 and 2.4 was varied, see Fig. 4C. The pH gradient separations are offering the best resolution, especially for the very acidic proteins present in the complex *E. coli* mixture.

IEC has emerged as the standard method for the determination of charge heterogeneity of monoclonal antibodies. It is important to measure product heterogeneity during the development and production process of mAbs, as many charge variants can arise due to PTM or product degradation processes. These modification processes of the parent protein include C-terminal lysine variants, N-terminal pyroglutamate formation, deamidation, glycation, and glycosylation, resulting in a modified isoelectric pH (pI) value of the mAb [107,108]. Vlasak et al. reviewed the analysis of charge-related heterogeneity in monoclonal antibodies [109]. IEC is less preferred to study ADC charge variants, as the linked cytotoxic drugs are changing the hydrophobic surface of the conjugated antibody resulting in unwanted secondary interactions with the stationary phase and consequently a poor resolution separation [110]. Some studies have been performed on retention time models for IEC separations using salt gradients [111,112] and pH gradients [113]. Fekete et al. applied a Drylab CEX model for the separation of mAb charge variants using both salt and pH gradients [114,115].

The contribution of various posttranslational modifications to monoclonal antibodies is diverse, with basic amino acids

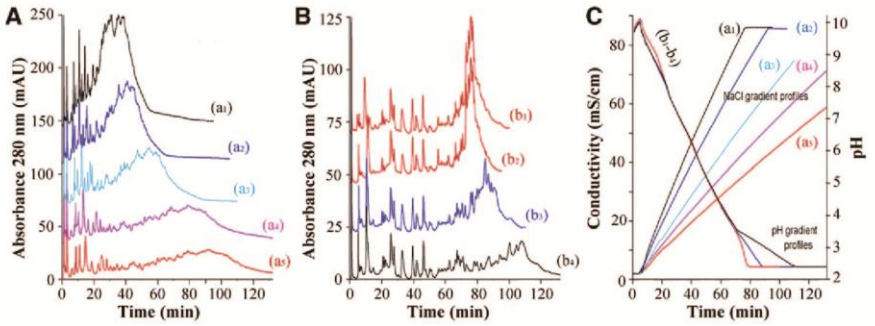


FIGURE 4 Optimization of anE. coli acetone powders separation in anion-exchange mode by (A) a salt gradient with decreasing the slope of the NaCl salt gradient in time, and by (B) application of a pH gradient; decreasing the slope of the pH-gradient in the range between pH 3.5–2.4. (C) shows the respective salt and pH gradient profiles. For the salt gradients, a 20 mM sodium carbonate buffer at pH 9.7 was used as mobile-phase A and 20 mM sodium carbonate buffer at pH 9.7 containing 1 mM NaCl was used as mobile-phase B. For the pH gradients, a proprietary pISep buffer (mixture of polyionic organic buffering molecules) at pH 2.4 was used as buffer A, and buffer B consisted of pISep buffer at pH 10.9. The column volume (CV) was approximately 2 mL. The applied gradient slopes are: a_1 13.6 mM NaCl/CV, a_2: 10.9 mM NaCl/CV, a_3: 8.0 mM NaCl/CV, a_4: 5.0 mM NaCl/CV, and a_5: 4.3 mM NaCl/CV for the salt-gradient profiles, and b_1 and b_2: 0.1 pH units/CV, b_3: 0.1 pH units/CV from pH 9.7–3.5 and 0.05 pH units/CV from pH 3.5–2.4, and b_4: 0.1 pH units/CV from pH 9.7–3.5 and 0.025 pH units/CV from pH 3.5–2.4 for the pH-gradient profiles. Adapted with permission from [105]

contributing to an increase in the mAbs pI whilst deamidation of asparagine residues and sialic acid present on N-glycans contribute to a decrease in the mAbs pI. These different contributors to the overall protein chemistry of the mAb make CEX the analysis method-of-choice to study mAb charge variants. In CEX mode, the separation of proteins is governed by the surface charge, charge distribution, and the geometry of the protein. CEX separations of mAbs are typically performed by applying a gradient with increasing salt concentration (i.e., 100–200 mM sodium chloride), while maintaining the pH of the buffer constant. The pH of the buffer depends on the pI of the mAbs under analysis but in general the pH range is between 7.5 and 9 [116]. Separation of mAbs in anion exchange chromatography (AEX) mode, is also being performed, mainly to separate oxidized variants of mAbs. Teshima et al. showed how AEX was effective in the analysis of three force-oxidized antibodies as compared to CEX. It was demonstrated that AEX revealed oxidized mAbs variants not monitored using CEX [117].

Jungbauer demonstrated the combined effect of a linear salt-and pH-gradient in IEC mode for the separation of protein isoforms of a human monoclonal antibody [118]. The IEC chromatogram was compared with IEF, and confirmation of elution order based on pI was shown. As the method relied on the reaction of mannitol with borate, the broad-scale applicative value was limited. Many publications have investigated internally generated pH-gradient CEX methods to separate mAb charge variants, however, they often employ cationic buffering agents which can lead to interactions with the stationary-phase chemistry. This deviates the shape of the applied pH-gradient from the ideal linear case, affecting protein retention and the resolution of the separation [119–121]. In an attempt to address this issue, research groups have investigated algorithms to correct for these deviations [105]; a simple mixture of buffering species that produce an internal linear gradient for neutral and acidic mAbs [122]; mixed-bed stationary phases consisting of small-pore weak IEC and large-pore strong IEC particles allowing for independent internal pH-gradient generation and protein binding [123,124]; external pH gradients in AEX using a mixture of amine buffering species as an application buffer and weakly acidic compounds as an elution buffer [125–127]; and shallow externally generated pH gradient of diethanolamine buffer on monolithic IEC stationary phases [128,129].

Another way of solving this issue is by using zwitterionic and acidic buffer substances with a pK_a range evenly distributed over the pH range and externally generate a pH and salt gradient. Typically, zwitterionic compounds tailored for biochemical research are used as buffering agents. It was shown that this allows for generating highly linear pH gradients, with even distribution of buffer capacity, for the analysis of charge heterogeneity of mAbs [130]. Recently, the ruggedness of a controlled gradient pH formation with a zwitterionic buffer system for the separation of mAb charge variants was demonstrated, showing good robustness of the method with <0.8% RSD for the retention times after more than 300 injections [131].

3 | PROTEIN MS

MS for intact protein analysis has proven to be essential in the field of biomolecule characterization. Fenn received the Nobel Prize in Chemistry in 2002 for the development

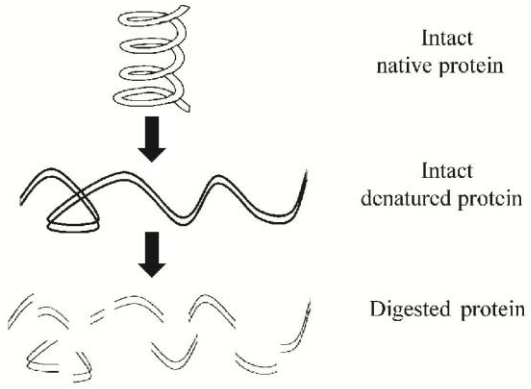


FIGURE 5 Schematic overview of the different state encountered in protein MS analysis

of ESI, allowing the transfer of biomacromolecules from an aqueous solution into the gas phase as molecular ions without fragmentation [132,133]. ESI can operate in the flow regime from 1 mL/min down to tens of nL/min, with the latter approach typically utilizing "static" (i.e., offline) spray from glass capillaries, also called nano-ESI [134]. Electrospray is now the dominant ionization method in many chemical, (bio)medical, and pharmaceutical MS laboratories, largely due to the ease with which it can be used to interface LC "inline" with different types of MS(/MS) instruments [135]. Denaturing MS-based strategies have been extensively applied to retrieve information on the intact mass of therapeutic proteins, including information on the amino acid sequence and PTMs, the DAR and drug load distribution, etc. [136].

A visualization of the different states of protein MS analysis is depicted in Fig. 5. Proteins encountered at physiological conditions remain their native 3D structure. In "intact denatured proteins" the 3D protein conformation is lost. The protein mass is particularly important when variations of the amino acid sequence, such as mutations or truncations, as well as post-translational or chemical modifications, e.g., deamidation, covalent linkers, need to be identified and quantified [136]. While the full set of modifications present as well as heterogeneity arising from the occupancy of the possible sites can be obtained from intact, denaturing MS, mapping the modification sites requires MS/MS approaches typically using bottom-up proteomics, but increasingly also middle-down and top-down MS/MS [137]. The often overlooked, but important shortcoming of these "standard" proteomics methods, however, is that they usually only give partial sequence information (i.e., not all expected peptides or MS/MS fragments are found back in the spectra), and therefore neither identify all modification sites nor typically characterize the full complement of modifications (i.e., their extent and heterogeneity). It has become increasingly obvious that the full knowledge of the primary sequence information, the "proteoform", i.e., the "chemical sum formula" of the protein and its sequence [138], can only be obtained by a combination ofintact protein MS with MS/MS approaches as they are used in proteomics, with or without prior digestion.

Extending the applicability of ESI–MS incorporating volatile buffer systems and physiological pH conditions, as well as modifications to the instruments to increase the mass range and the control over desolvation conditions, has led to the development of native MS [139,140]. Native MS has been extensively applied to study macromolecular assemblies, including stoichiometry and identity of binding partners [141,142], and in the last decade its applicability has been extended towards the MS analysis of biopharmaceutical products [24]. In native MS, it is believed that noncovalent weak interactions, i.e., van der Waals interactions, hydrogen bonds, and electrostatic interactions, are maintained, preserving the higher order, 3D protein structure during the MS analysis. It is generally recognized that changes in charge density in ESI–MS spectra correspond to conformational changes, i.e., the tertiary protein structure [143]. Figure 6A displays the charge–state distribution profile of an intact protein (antithrombin III) applying native MS conditions [144]. Due to the compact, folded state of the protein, the exposed surface that can be protonated is relatively small, therefore, yielding a relatively narrow charge envelope situated in the high m/z region (low z) compared to the same protein when applying denaturing ESI–MS conditions (Fig. 6C). Figure 6B shows that native and denatured protein states coexist at equilibrium applying mildly denaturing conditions. The intermediate charge-density ions correspond to proteins that contain domains that are unfolded, while other domains retain their native conformation. While unfolding proteins in denaturing MS usually allows accurate and precise mass determination (≤ 1 Da), desolvation conditions are more gentle in native MS and the folded protein often retains bound water or buffer ions, leading to a somewhat increased experimental mass compared to the expected value [145]. When determining the intact mass of proteins above 100 kDa, the native approach may become easier, as it produces fewer and lower charge states, whereas denaturing MS leads to a large number of closely spaced, highly charged peaks, which are difficult to resolve and correctly assign.

Ion mobility (IM) is now often coupled with native MS and several commercial platforms offer this option [140]. In IM, ions are separated by their collision cross-section, measured in nm^2 or $Å^2$, which depends on their charge but also their rotationally averaged size and shape—somewhat similar to gas-phase electrophoresis [146]. At each m/z, different coexisting conformers, isomers, or complex/aggregate topologies can be resolved as long as they differ in overall size by 2–3%. The measured mobilities ofions can be converted to collision cross-section values using a set of calibrants (e.g. protein

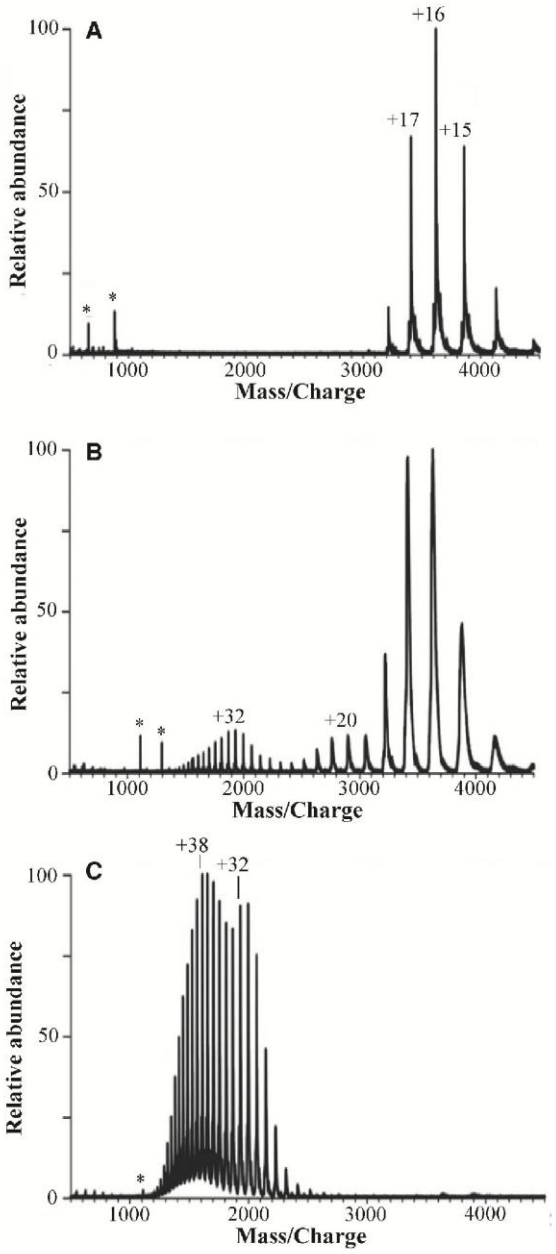


FIGURE 6 ESI mass spectra of anti-thrombin III (A) acquired under native MS conditions using 20 mM ammonium acetate, (B) using 20 mM ammonium acetate/methanol/formic acid 49:50:5 v/v/v%, and (C) denaturing conditions using 20 mM ammonium acetate/methanol/formic acid 45:50:5 v/v/v%. Adapted with permission from [144]

standards) with known structure. This technology has come to the fore in the last 10 years and recent examples include studies of protein folding/misfolding and aggregation, intrinsic disorder phenomena and the identification ofisomeric forms of metabolites, biomolecules, and complexes [147]. In the context of biopharmaceuticals characterization, IM has shown to be able to distinguish different glycoforms, even in cases where they cannot be resolved in LC, as well as disulfide isoforms [148].

3.1 | Conditions for direct infusion and hyphenation to LC

Gentle ionization, in which the noncovalent interactions involved in protein higher-order structure, i.e., folding and interactions, are maintained, is considered to be a critical step in native MS [143]. Most native MS is done in "static", offline nano-ESI using metallized glass capillaries, also called direct infusion, with a flow rate< 20 nL/min to minimize sample consumption, improve the tolerance of spraying aqueous buffer solutions, limiting the salt intake, and eliminating the need for desolvation gas and heating. Native MS can also be implemented at flow rates in the 200–300 nL/min range which are compatible with inline nano-LC, and in principle also at higher flow rates, although care has to be taken that ESI interface settings such as (hot) desolvation gas and source heating do not unfold the protein. Sample requirements for native MS and buffer conditions have been described by Hernandez and Robinson [149]. Typically, infusion of analyte at 1–20μM dissolved in 10 mM to 1 M aqueous ammonium acetate solution maintained near pH 7 or at the pH of choice, using an excess of ammonia or acetic acid, provides good MS spectra. Also, other ammonium salts and ammonium derivatives have been employed but acetates are found to perform better than bicarbonates [150]. Nonvolatile ions such as sodium and potassium are minimized using buffer exchange and other desalting methods, since these salts induce adduct formation, thereby lowering the mass resolution or suppressing signal entirely. Essential cofactors such as Mg or Zn ions can be added, but a large excess should be avoided. With respect to optimization of the MS settings, it is important that pressure in the transfer region between source and analyzer is optimized to ensure transmission of biomacromolecules. Modifications of MS instrumentation have been described in more detail by Rosati et al. [148].

Due to the stringent requirements with respect to infusion conditions, the number of reports describing the direct coupling between native LC and MS is limited. First of all, the flow rate compatibility with LC constitutes a problem. Conventional SEC, HIC, and IEC separations are still performed using either 4.6 mm id columns operated at a flow rate of 1 mL/min, or 2.1 mm id columns operated at 0.2 mL/min. Hence, postcolumn flow splitting is required to achieve direct coupling to MS by nano-ESI. Furthermore, the salts typically applied in SEC, HIC, and IEC modes are incompatible with MS analysis. In case of aqueous SEC, the phosphate buffer can be replaced by an acetate buffer. In gel filtration, typical buffers are fully native, but scaling down

is a major bottleneck. The sulfate ions typically used in HIC systems are also incompatible with MS, leading to significant signal suppression [151]. Volatile buffer acetate and tartrate and ammonium salt systems can be considered for HIC-MS analysis, but it should be noted that the choice affects protein retention and may limit the applicability. Xiu et al. reported a lack of retention for the HIC analysis of hydrophilic proteins using ammonium acetate as the mobile phase (as can be expected from the Hofmeister series) [151]. Ammonium tartrate dissolved in an ammonium acetate buffer provided similar elution strength compared to ammonium sulfate. The MS compatibility with respect to adduct formation was only demonstrated after desalting using ultra-centrifugation followed by RP-LC–MS analysis. It has been reported by Chen et al. that the desalting processes can induce variations in the conformation of the proteins, and hence the native conformation may be lost [152]. A viable approach to online HIC–ESI–MS was proposed by Chen et al., which involves the use oflow concentrations of ammonium acetate mobile phases (volatile and MS compatible) [152].

3.2 | MS characterization of mAbs and ADCs: Key examples

A protocol for conducting native MS analysis of mAbs and ADCs has been described by Thompson et al. [143]. Illustrative MS spectra of a 145 kDa purified mAb by direct infusion are demonstrated in Fig. 7A, yielding only seven charge states. A mass accuracy of 5 Da allows establishing the protein id with high confidence when the amino acid sequence is known and allows identifying modifications, including primary sequence mutations and C-terminal lysine clipping [143]. The natural isotopic peak width of the intact antibody was estimated to be 25 Da. A very accurate isotope pattern needs to be recorded to detect modifications such as deamidation, yielding a mass increase of 1 Da. Significant improvements in resolution and native MS technology have been reported over the years. For example, Rose et al. reported the use of Orbitrap MS yielding a resolution of 16 000 at m/z10 000 [153]. The applicability of native MS to probe the binding stoichiometries and affinities of mAb–antigen complexes was first demonstrated by Tito et al. [154]. Compared to SEC–UV or SPR spectroscopy yielding evidence for binding, or at best average-weight information, native MS provides accurate mass information. Tito et alalso performed control experiments to establish the specificity of the interactions [154].

The presence of microheterogeneities in the protein chains can derive from inconsistencies in the production process (differences in the cell lines) and therefore it is extremely important to perform batch to batch QC of the mAbs before their application as therapeutics. Most of the times these inconsistencies stem from heterogeneous glycosylation patterns. Figure 7B shows native MS spectra of a mAb with glycosylation and after deglycosylation using peptide-N-glycosidase F [143]. The presence of glycans increases both the mass and the heterogeneity of the MS signal, that in turns decreases the peak intensity. Intact MS analysis can be used to reveal the presence of different glycoforms or on the chains of the mAbs (mutations on the mAbs chain can dramatically alter the glycosylation) [148]. Rosati et al. performed both qualitative and quantitative analysis of glycosylation profiles on mAbs using high-resolution Orbitrap MS technology [148]. Figure 8 compares native MS spectra of IgGs obtained by direct-infusion experiment and after online SEC analysis, as performed by Chatterjee and Sobott. This experiment showed that the SEC analysis induced partial unfolding of the antibodies (without breaking disulfide bonds), as it appears with higher charge states.

With respect to the analysis of ADCs, MS enables the characterization of the drug load profile and distribution, and the DAR. Valliere-Douglas et al. and Sobott et al. reported on a method allowing to determine the intact mass of an ADC composed of noncovalently associated heavy and light chains,

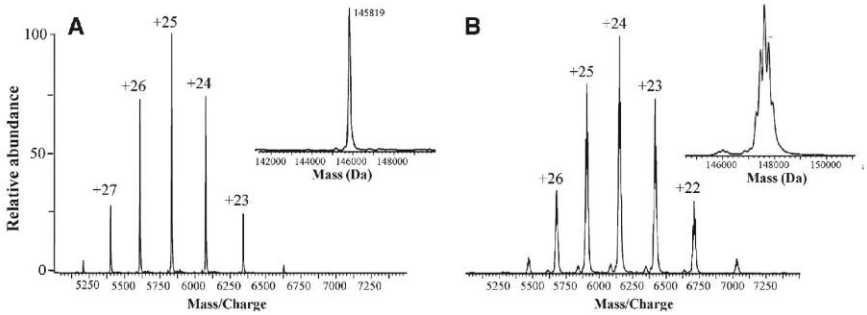


FIGURE 7 (A) Native MS spectrum of a deglycosylated mAb (IgG1) yielding a narrow charge envelope situated in the high m/z region and corresponding deconvoluted mass spectrum shown in the inset to determine the intact mass. (B) Subsection of a native MS spectrum from a glycosylated mAb displaying increased mass heterogeneity and corresponding deconvoluted mass spectrum in the inset revealing the presence of different glycoforms. Adapted with permission from [143]

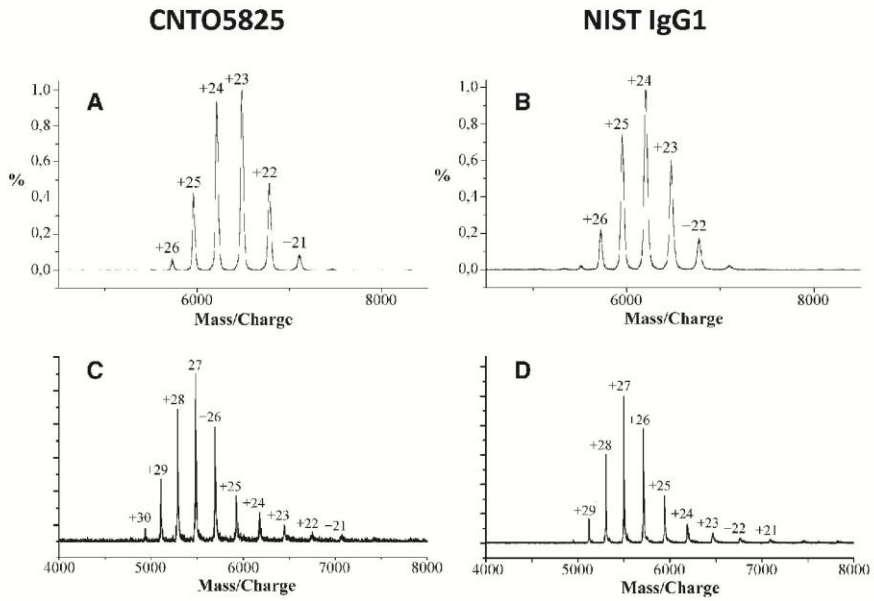


FIGURE 8 Native MS spectra of IgGs (CNTO5825 and NIST) obtained on a Q-TOF-2 instrument (Waters) after direct infusion (A and B) and after SEC analysis, indicating partial unfolding of the antibodies (without breaking disulfide bonds). LC conditions: Flow rate 0.1 mL/min; mobile phase= 100 mM ammonium acetate, pH 6.8; using a 4.6 mm id x 100 mm BEH SEC column packed with 1.7 µm particles (200 Å pores)

with a drug linked to interchain cysteine residues [49,155]. Debaene et al. conducted native MS experiments of Brentuximab vedotin (also an interchain cysteinyl-linked ADC) providing accurate mass measurements ofintact ADCs together with the average DAR and drug distribution [156]. The same group also characterized a lysine linked antibody drug conjugate (Tratuzumab emtansine) [156]. Extending the glycoprofiling experiments of mAbs, Rosati et al. also characterized the drug load and glycosylation patterns on IgG4 ADCs using high-resolution native MS [148].

4 | CONCLUDING REMARKS

Advances in the development of biotherapeutics are closely followed by innovations in the field of separation sciences and MS. The chemical heterogeneity of biopharmaceuticals in terms of polarity, size, and charge, require the use of complementary native LC techniques ideally hyphenated to MS to fully characterize (and quantify) the complex protein samples. This requires the use of separation technology with high resolving power to achieve the highest confidence in elucidating the biopharmaceutical product. Although columns packed with sub-2µm particles are being introduced and have become commercially available, the majority of LC experiments are still performed using conventional columns packed with 5 µm particles. However, to make a successful transition, the effects on protein unfolding induced by the mobile-phase composition applied, shear stress, and thermal effects need to be critically assessed. Furthermore, the pressure stability of currently available IEC, SEC, and HIC columns needs to be augmented to allow for operating pressures above 50 MPa. A promising (but currently underestimated) stationary phase type for biomacromolecule separation may be monolithic columns. The morphology can be optimized to achieve high efficiency separations by downscaling the globule size, while the macropore size can be tuned to minimize shear stress.

Conventional 4.6 mm id analytical columns for mAb analysis require several micrograms of mAbs to achieve adequate detection sensitivity oflow-abundant sample species. The yield of biopharmaceutical products coming from microwell-plate cell cultures is, however, limited and often insufficient for high-resolution LC analysis. This mandates the miniaturization of column formats allowing to increase detection sensitivity and to diminish sample consumption. It should be noted however, that extra-column band broadening needs to be minimized, imposing stringent requirements on instrumentation. An additional advantage of reducing the column format is that it decreases the salt-intake at the MS interface, effectively improving MS compatibility. The use of organic solvents that are conventionally added to the mobile phase, not only to improve spray drying but also to reduce the surface tension of the spray droplets leading to higher ionization yields, should be limited when performing bioanalysis. This is not only mandatory to maintain the protein conformation but also because buffers are known to lead to suppression ofion formation in the ion source and ion source contamination due to salt crust formation. Novel salt systems, buffering agents,

and the effects ofionic strength have to be further studied to improve MS compatibility. At the same time, it is mandatory to further study the effects ofionization conditions and MS conditions on protein conformation, to establish relevant biological conditions.

ACKNOWLEDGMENTS

JDV acknowledges a postdoctoral research grant of the Research Foundation Flanders (FWO, grant no. 12J6517N) and SC acknowledges a PhD research grant from the Institute for Innovation through Science and Technology (IWT). Dr. Koen Sandra (RIC, Kortrijk, Belgium) is acknowledged for providing Figure 1. Jonas-Frederik Jans (VUB) is acknowledged for his assistance in the preparation of the review manuscript.

ORCID

Sebastiaan Eeltink http://orcid.org/0000-0002-5465-0127

REFERENCES

1. http://www.persistencemarketresearch.com/mediarelease/biopharmaceutical-market.asp (last time accessed: May 12, 2017).
2. http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/rapid-growth-in-biopharma (last time accessed: July 7, 2017).
3. Walsh, G., Biopharmaceutical benchmarks 2014.Nat. Biotechnol. 2014, 32, 992–1000.
4. Leader, B., Baca, Q. J., Golan, D. E., Protein therapeutics: a summary and pharmacological classification.Nat. Rev. Drug Discov. 2008, 7, 21–39.
5. Bray, G. L., Gomperts, E. D., Courter, S., Gruppo, R., Gordon, E. M., Manco-Johnson, M., Shapiro, A., Scheibel, E., White, G. 3, Lee, M. A., Multicenter study of recombinant factor VIII (recombinate): safety, efficacy, and inhibitor risk in previously untreated patients with hemophilia A. The recombinate study group.Blood, 1994, 83, 2428–2435.
6. Weiner, L. M., Surana, R., Wang, S., Monoclonal antibodies: versatile platforms for cancer immunotherapy.Nat. Rev. Immunol. 2010, 10, 317–327.
7. Rostami, A. S., Qazi A.I., Sikorski, A.R. The clinical landscape of antibody drug conjugatedADC Rev. 2014, https://adcreview.com/articles/doi-10-14229jadc-2014-8-1-001/
8. Ohtake, S., Arakawa, T., Recombinant therapeutic protein vaccines. Protein Pept. Lett. 2013, 20, 1324–1344.
9. Ecker, D. M., Jones, S. D., Levine, H. L., The therapeutic monoclonal antibody marketMAbs 2015, 20, 9–14.
10. Mack, F., Ritchie, M., Sapra, P., The next generation of antibody drug conjugatesSemin. Oncol. 2014, 41, 637–652.
11. Helliwell, C. L., Coles, A. J., Monoclonal antibodies in multiple sclerosis treatment: current and future steps.Ther. Adv. Neurol. Disord. 2009, 2, 195–203.
12. Tanaka, T., Hishitani, Y., Ogata, A., Monoclonal antibodies in rheumatoid arthritis: comparative effectiveness of tocilizumab with tumor necrosis factor inhibitors.Biol. Targets Ther. 2014, 8, 141–153.
13. Li, F., Vijayasankaran, N., Shen, A. Y., Kiss, R., Amanullah, A., Cell culture processes for monoclonal antibody production.MAbs 2010, 2, 466–477.
14. Milstein, C., The hybridoma revolution: an offshoot of basic research.BioEssays 1999, 21, 966–973.
15. Carvalho, L. S., da Silva, O. B., Carneiro de Almeida, G., de Oliveira, J. D., Parachin, N., Carmo T. S., Fermentation Processes, InTech, Rijeka 2017.
16. Nelson, A. L., Antibody fragments. MAbs 2010, 2, 77–83.
17. Nimmerjahn, F., Ravetch, J. V., Fcgamma receptors as regulators ofimmune responses.Nat. Rev. Immunol. 2008, 8, 34–47.
18. Kontermann, R., Dual targeting strategies with bispecific antibodies. MAbs 2012, 4, 182–197.
19. Kontermann, R. E., Brinkmann, U., Bispecific antibodies. Drug Discov. Today, 2015, 15, 838–847.
20. Diamantis, N., Banerji, U., Antibody-drug conjugates—An emerging class of cancer treatment.Br. J. Cancer 2016, 114, 362–367.
21. van de Donk, N. W. C. J., Dhimolea, E., Brentuximab vedotin. MAbs 2012, 4, 458–465.
22. Lambert, J. M., Chari, R. V. J., Ado-trastuzumab emtansine (T-DM1): An antibody–drug conjugate (ADC) for HER2-positive breast cancer.J. Med. Chem. 2014, 57, 6949–6964.
23. Fan, G., Wang, Z., Hao, M., Li, J., Bispecific antibodies and their applications.J. Hematol. Oncol. 2015, 8, 130.
24. Berkowitz, S. A., Engen, J. R., Mazzeo, J. R., Jones, G. B., Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars.Nat. Rev. Drug Discov. 2012, 11, 527–540.
25. Bosques, C. J., Collins, B. E., Meador, J. W., Sarvaiya, H., Murphy, J. L., Dellorusso, G., Bulik, D. A., Hsu, I.-H., Washburn, N., Sipsey, S. F., Myette, J. R., Raman, R., Shriver, Z., Sasisekharan, R., Venkataraman, G., Chinese hamster ovary cells can produce galactose-α-1,3-galactose antigens on proteins.Nat. Biotechnol. 2010, 28, 1153–1156.
26. Narhi, L. O., Jiang, Y., Cao, S., Benedek, K., Shnek, D., A critical review of analytical methods for subvisible and visible particles. Curr. Pharm. Biotechnol. 2009, 10, 373–381.
27. Parr, M. K., Montacir, O., Montacir, H., Physicochemical characterization of biopharmaceuticalsJ. Pharm. Biomed. Anal. 2016, 130, 366–389.
28. Bush, D. R., Zang, L., Belov, A. M., Ivanov, A. R., Karger, B. L., High resolution CZE–MS quantitative characterization ofintact biopharmaceutical proteins: proteoforms ofinterferonβ-1. Anal. Chem. 2016, 88, 1138–1146.
29. Hayes, J. M., Frostell, A., Karlsson, R., Müller, S., Millan-Martin, S., Pauers, M., Reuss, F., Cosgrave, E., Anneren, C., Davey, G. P., Rudd, P. M., Identification of Fc gamma receptor glycoforms that produce differential binding kinetics for rituximab.Mol. Cell. Proteomics 2017, 16, 1770–1788.
30. Lindqvist, B., Storgards, T., Molecular-sieving properties of starch. Nature 1955, 175, 511–512.

3.3 IM-MS characterisation of a novel antibody-tetrazine conjugate

In collaboration with

Dr. Agnese Maggi[1]
Jens Fissers[1]
Prof. Dr. Koen Augustyns[1]

[1] Laboratory of Medicinal Chemistry, University of Antwerp, BE

Part of this research was published in

Agnese Maggi, Eduardo Ruivo, Jens Fissers, Christel Vangestel, Sneha Chatterjee, Jurgen Joossens, Frank Sobott, Steven Staelens, Sigrid Stroobants, Pieter Van Der Veken, Leonie Wyffels, Koen Augustyns,
Development of a novel antibody-tetrazine conjugate for bioorthogonal pretargeting,
Org. Biomol. Chem, 2016, 14, 7544

3.3.1 Antibody conjugates

Conjugated antibodies are antibodies linked to a small molecule and can have a range of applications. An example of conjugated antibodies used in the pharmaceutical industry is the antibody-drug-conjugate (ADC) which combines the selectivity of the antibody with the potency of the drug.
Antibody conjugates can also help in the detection of proteins by labelling the antibody with fluorescent dyes. This technique can be useful in molecular imaging. In this part of the chapter the antibody trastuzumab is studied which is first bound with a tetrazine conjugate in order to react with the fluorescent dye, trans-cyclooctene (TCO) (**Figure 1**). With mass spectrometry the amount of conjugates on the antibody was determined and with ion mobility mass spectrometry the collision cross section of the antibody and the antibody conjugate was compared.

Figure 1: Illustration of trastuzumab pretargeted with tetrazine conjugate and then labeled with TCO fluorescent dye via the IEDDA cycloaddition.

3.3.2 Experimental details

3.3.2.1 Determining exact mass with mass spectrometry

The MS experiments were performed on a high-mass modified Q-TOF-2^{TM} instrument (Micromass, Waters and MS Vision). The samples were buffer exchanged twice in 150 mM aqueous ammonium acetate at pH 7.0 using Micro Bio-Spin P-6 columns (Bio-Rad) at concentrations of 10 µM. Gold-coated nanoESI needles were prepared in-house in order to transfer 5 µL aliquots of sample.
The instrument was tuned using the following parameters: capillary voltage 1.6 kV; nanoflow 0.3 bar; sample cone 200 V; collision energy 100 V; backing pressure 4.0 mbar.

3.3.2.2 Determining collision cross sections with ion mobility mass spectrometry

The native IM-MS experiments were performed on a Synapt G2 HDMS Q-TOF instrument (Waters, UK). The instrument was tuned to preserve native higher-order structure using the following parameters: capillary voltage 1.2 kV; nanoflow backing gas pressure 0.2 bar; sampling cone 30 V; extraction cone 2.0 V; trap collision energy 10 V; transfer collision energy 2.0 V; trap bias 45 V; IMS gas flow 90.0 mL/min.; IMS wave velocity 700 m/s; IMS wave height 40.0 V; backing pressure 3.0 mbar.

The data were acquired and processed with Masslynx v4.1 software, and ion mobility drift times extracted using Driftscope v2.3 (both Waters). The collision cross sections (CCS) of the proteins were calibrated using known CCS values determined under native conditions.

3.3.3 Native IM-MS of trastuzumab and trastuzumab- tetrazine conjugate

Using native mass spectrometry the mass of trasuzumab was determined to be 148070 Da and shows charge states 22^+ to 27^+ (**Figure 2A**). When adding 10 and 100 equivalents of the tetrazine conjugate, on average 2 and 8 tetrazine adducts were observed respectively to bind covalently to the antibody (**Figure 2B**)

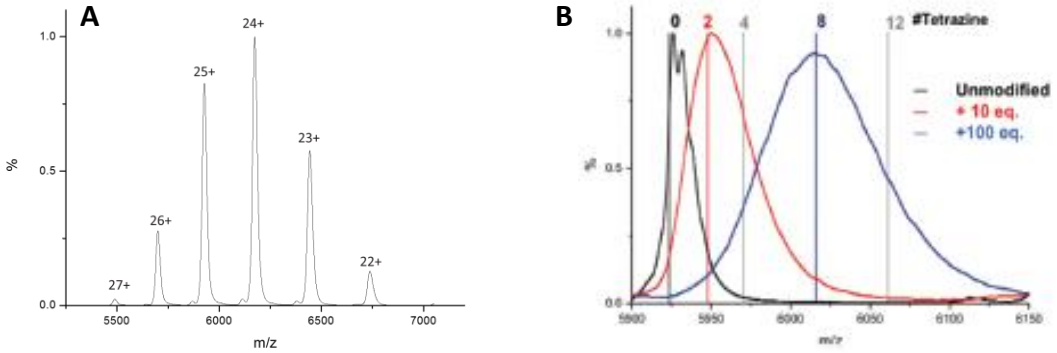


Figure 2: A) Mass spectrum of trastuzumab (148070 Da) with charge states from 22^+ to 27^+. B) 25+ charge state of trastuzumab (black) compared to the trastuzumab-tetrazine conjugates (10 and 100 equivalents of tetrazine)

Ion mobility mass spectrometry provided the collision cross sections for each observed charge state of trastuzumab and trastuzumab- tetrazine conjugate in **Table 1**. The difference in CCS value when the tetrazine was attached is between 0.24 to 1.62 %, which can be seen as an insignificant change. Thus, the binding of tetrazine to the antibody does not significantly change the collision cross section, and presumably the global structure, of the antibody.

Table 1: CCS values of trastuzumab and trastuzumab-tetrazine conjugate.

Charge state	CCS trastuzumab ($Å^2$)	CCS trastuzumab + ttz ($Å^2$)	Δ
21^+	7000.48	7114.15	1.62 %
22^+	7141.47	7242.55	1.41 %
23^+	7253.51	7340.26	1.19 %
24^+	7381.12	7455.77	1.01 %
25^+	7510.10	7543.85	0.44 %
26^+	7613.94	7632.59	0.24 %

Organic & Biomolecular Chemistry

PAPER

Cite this: Org. Biomol. Chem., 2016, 14, 7544

Development of a novel antibody – tetrazine conjugate for bioorthogonal pretargeting †

Agnese Maggi,[a] Eduardo Ruivo,[a] Jens Fissers,[a] Christel Vangestel,[b] Sneha Chatterjee,[c] Jurgen Joossens,[a] Frank Sobott,[c] Steven Staelens,[b] Sigrid Stroobants,[b,d] Pieter Van Der Veken,[a] Leonie wy ffels[b,d] and Koen Augustyns*[a]

Received 30th June 2016,
Accepted 12th July 2016

DOI: 10.1039/c6ob01411a

www.rsc.org/obc

Recently, bioorthogonal chemistry based on the Inverse Electron-Demand Diels – Alder (IEDDA) cycloaddition between 1,2,4,5-tetrazines and trans-cyclooctene (TCO) analogues added an interesting dimension to molecular imaging. Until now, antibodies (Abs) were tagged with TCO and after pretargeting they were reacted with tetrazines substituted with reporters. However, TCO tags have the tendency to degrade under physiological conditions, and due to their hydrophobic nature are buried within the protein. This results in loss of reactivity and a low Ab functional loading. To circumvent these problems, we report for the first time an approach in which tetrazines are used as tags for antibody (Ab) modification, and TCO as the imaging agent. We developed a new Ab – tetrazine conjugate, which displays a high functional loading, good stability and reactivity. We utilized this immunoconjugate for live-cell imaging together with novel TCO probes, resulting in selective and rapid labeling of SKOV-3 cells. Our approach may be useful for in vivo pretargeted imaging.

Introduction

Molecular imaging is a powerful tool, which enables the non-invasive visualization and quantification of biological processes occurring at the cellular level. Recently, the possibility to combine molecular imaging with bioorthogonal chemistry has strongly improved the applicability of this technique for in vitro and in vivo imaging under different modalities.

Generally, the term "bioorthogonal chemistry" refers to any chemical reaction that can occur inside a living system without interfering with native biochemical processes, and permits the covalent attachment of a probe molecule to a biomolecule of interest.[1–3] In order to achieve bioorthogonal ligation two successive steps are required: (1) introduction of a bioorthogonal tag onto the biological target (pretargeting step) and (2) reaction between the bioorthogonal tag and an externally introduced chemical probe (targeting step).

A well-established bioorthogonal reaction is the Inverse Electron-Demand Diels–Alder (IEDDA) cycloaddition between 1,2,4,5-tetrazines and trans-cyclooctene (TCO). This reaction was introduced in 2008[4–6] and is currently investigated for many applications.[7] In fact, its extremely fast kinetics[8,9] enables application for in vivo studies where reactions must occur rapidly at micro- to nanomolar concentrations.

It is noteworthy that tetrazine/TCO ligation has been used for pretargeted imaging of tumors both in living cells[10–13] and in mice.[14–16] In a typical bioorthogonal pretargeting approach the target (e.g. cell surface antigen) is first bound by a TCO-modified antibody (Ab), and then a tetrazine probe is introduced in order to achieve labeling through the cycloaddition reaction.

Different TCO–Ab conjugates have been developed in the last few years. These conjugates can be obtained either via the reaction between TCO–N-hydroxysuccinimide (TCO–NHS) esters and the Ab lysine residues,[17,18] or via a two-step procedure where the Ab is first decorated with azide functionalities and subsequently reacted with dibenzylcyclooctyne–TCO derivatives (DBCO–TCO).[19] Moreover, the TCO moieties can be attached to the Ab either directly or through short PEG spacers. Even though the type of attachment does typically not affect the Ab binding affinity, which is usually retained even at high TCO loadings, it can influence the effective number of reactive TCO moieties. Recently, Haun and co-workers have shown that when hydrophobic TCOs are directly attached to the antibody, up to 90% are not reactive due to their tendency to be buried within the protein. In contrast, the introduction of TCOs through hydrophilic PEG linkers results in a fully preserved reactivity.[19] Prior to this study, Robillard and

[a]Laboratory of Medicinal Chemistry, University of Antwerp, Belgium. E-mail: koen.augustyns@uantwerpen.be
[b]Molecular Imaging Center Antwerp, University of Antwerp, Antwerp, Belgium
[c]Biomolecular and Analytical Mass Spectrometry, University of Antwerp, Belgium
[d]Antwerp University Hospital, Department of Nuclear Medicine, Edegem, Belgium
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c6ob01411a

co-workers described an in vivo deactivation of TCO via isomerization to the unreactive cis-cyclooctene (CCO). Isomerization is probably catalyzed by copper bound to serum proteins, and its extent depends on the degree of exposure of the TCO tags on the Ab surface. Particularly, it has been demonstrated that the attachment of TCOs through PEG linkers results in a shorter half-life of the tags compared to direct attachment. [18]

Attaining a high Ab functional density is an important feature for the application of bioorthogonal pretargeting in animal models, where a high target-to-background ratio is desirable. The latter may be significantly reduced by both hydrophobic burying and tag exposure deactivation. Due to the slow clearance ofintravenously injected Abs, it can take several days to obtain an optimal biodistribution resulting in a prolonged exposure of the TCO tags to the physiological environment with consequent deactivation.

Replacement of TCO tags with more hydrophilic tetrazines (inversion offunctionalities) could o ffer a valid alternative to conventional Ab-pretargeting strategies, and may address hydrophobic burying. Moreover, with respect to future in vivo applications, the use of more hydrophobic labeled TCOs for targeted imaging might have an advantage over labeled tetrazines when cell membranes or the blood –brain barrier need to be crossed.

To date, the use of tetrazines as bioorthogonal tags for antibody-pretargeting has not been investigated. However, several studies demonstrated the utility of tetrazines as tags for pretargeted protein labeling, [20–23] and construction of labeled peptides. [24–26] Yi and co-workers reported the use of a Ab–tetrazine conjugate for protein capture. [27] Han et al. described Ab–quantum dot (QD) conjugates, which were assembled by reacting Ab –tetrazine conjugates with norbornene-coated QDs. [28] A challenge for the use of tetrazine –Ab conjugates for bioorthogonal-pretargeting is the development of tetrazines displaying long term stability under physiological conditions and, at the same time, high reactivity for the IEDDA reaction. Monoaryl-1,2,4,5-tetrazines and 3,6-di-(2-pyridyl)-1,2,4,5-tetrazines, commonly employed as bioorthogonal probes, exhibit fast kinetics but are susceptible to degradation in biological environments. [11,29]

Here we present the synthesis and characterization of a series of tetrazines bearing diverse substituents. We evaluated their stability and reactivity for the IEDDA reaction with (E)-cyclooct-4-enol and validated the use of these compounds for antibody labeling. Most importantly, we developed a novel Ab – tetrazine conjugate, which displays a good long-term physiological stability, high functional density, and fast kinetics with TCO modified-fluorophores. Furthermore, we demonstrate the effectiveness of the Ab –tetrazine conjugate using pretargeted cancer cell studies as the validation platform.

Results and discussion

Chemical synthesis of 1,2,4,5-tetrazine derivatives

The first step towards our goal was the development of a tetrazine derivative exhibiting an optimized compromise in terms of reactivity and stability. To this end, we investigated the effect of different substituents on the tetrazine core and developed a total of 8 compounds (Fig. 1). Furthermore, all the

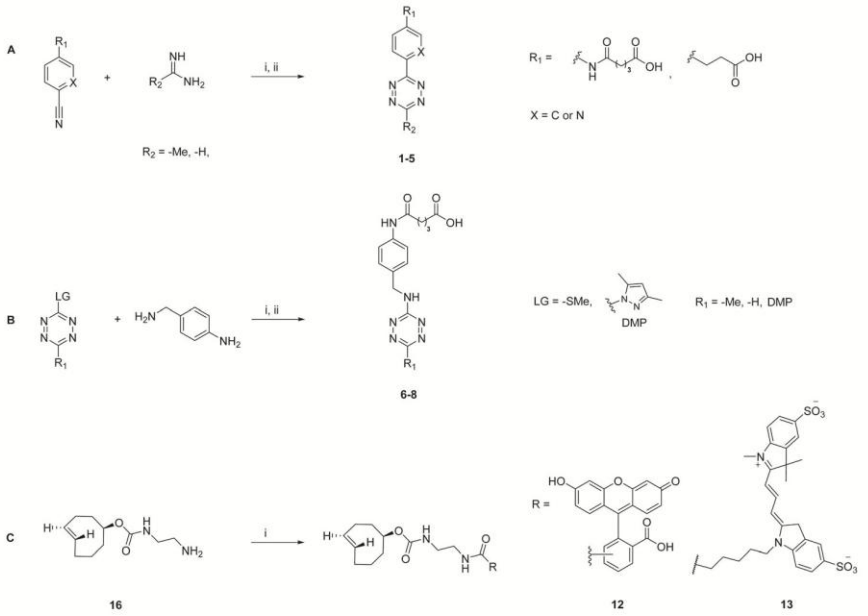


Fig. 1 Synthesis of tetrazine derivatives and TCO fluorescent dyes. (A) Synthesis of tetrazines 1– 5, reaction conditions: (i) Zn(OTf) $_2$ or S $_8$, NH $_2$NH $_2$·H $_2$O, EtOH, r.t.; (ii) NaNO $_2$, HCl 1 M, H $_2$O, 0 °C; (B) synthesis of tetrazines 6– 7, reaction conditions: (i) MeOH or toluene, r.t., Ar; (ii) glutaric anhydride, THF, reflux; (C) synthesis of TCO dyes 12 and 13, reaction conditions: (i) dye – NHS ester, DMF, TEA, r.t., N $_2$.

compounds were designed to possess a functional conjugation handle for antibody modification, such as a 5-oxopentanoic or a propionic acid chain.

Compounds 1–5 were prepared following a modification of a previously reported procedure, which entails the reaction between nitrile and amidine precursors with hydrazine monohydrate in the presence of a catalyst (sulfur or zinc triflate), followed by oxidation of the dihydrotetrazine intermediates (Fig. 1A).[30] After purification, tetrazines 1–5 were obtained in yields ranging from 11 to 40%. The comparatively moderate yields apparently are derived from the use of hydrazine monohydrate, which is more convenient and safe to use, instead of anhydrous hydrazine that was employed in the original procedure.[30]

The synthesis of compounds 6–8 was performed through a two-step protocol, where tetrazines bearing thiomethyl- or 3,5-dimethyl-1-pyrazol-1-yl (DMP) leaving groups were subjected to nucleophilic substitution with 4-aminobenzylamine. The resulting amino-derivatized tetrazine intermediates were subsequently reacted with glutaric anhydride to yield the final products in good overall yields (from 54 to 76% over two steps) (Fig. 1B).

Tetrazine stability and reaction kinetics with (E)-cyclooct-4-enol

The stability of the newly synthesized compounds was investigated next. Tetrazines 1–8 were incubated in either phosphate buffered-saline (PBS) or pure fetal bovine serum (FBS) at 37 °C, and the decrease of the tetrazine absorbance at selected wavelengths was monitored. The proportion ofintact tetrazine was determined after four days (see Table 1). The stability of tetrazines 9 and 10, which are commonly used for bioorthogonal studies and thus serve as a reference control, was also evaluated and compared to that of the newly synthesized compounds. As could be predicted the tetrazines containing electron donating groups (compounds 6–8) turned out to be very stable in either PBS or FBS, with up to 100% remaining intact after 4 days. On the other hand, for aryl substituted tetrazines (compounds 1–5) the stability was considerably reduced by replacing a methyl group with a hydrogen on the tetrazine core, and by the presence of electron withdrawing groups on the aryl substituent. The least stable compounds turned out to be tetrazines 2 and 10 with measured half-lives in PBS of 2 and 9 h respectively (Table 1).

The reactivity of tetrazines 1–9 towards the IEDDA reaction with (E)-cyclooct-4-enol was also evaluated (Table 1). Since the rate of the reaction between compounds 1–5 and the dienophile was too fast for reliable determination via UV-vis experiments, the reaction rates were measured by ^{1}H-NMR competition experiments with 3,6-di-2-pyridyl-1,2,4,5-tetrazine (DIPY).

In fact, the second order rate constant (k_2) of DIPY for the reaction with (E)-cyclooct-4-enol in MeOH is known to be 820 M^{-1} s^{-1}.[9] After mixing DIPY, the dienophile and the selected tetrazine in MeOH the ^{1}H NMR spectra of the mixture were analyzed, and the conversion rates of the two tetrazines were determined. From the relative conversion rates the final rate constants were calculated.[31]

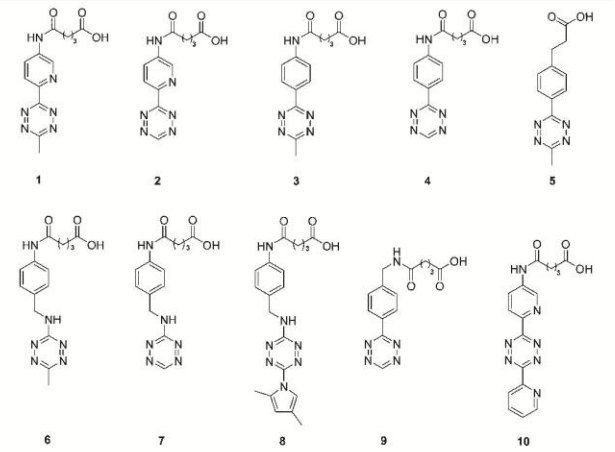
Due to their slower kinetics, the cycloaddition rates of tetrazines 6–8 could be directly measured by UV-Vis experiments. The rate constants were determined in PBS at 37 °C by reacting the different compounds with the dienophile under pseudo first order conditions, and measuring the disappearance of the tetrazine adsorption at selected wavelengths.

From this study we determined that the reactivity of aryl tetrazines 1–5 is inverse to their stability, whereas, for compounds 6–8 the presence of a hydrogen substituent on the tetrazine ring strongly affects the reactivity, with tetrazine 7 being 20 to 30 fold more reactive than tetrazines 6 and 8.

Overall, the range of reactivities and stabilities covered by compounds 1–8 make this library a valuable addition to the tetrazine compounds currently available for bioorthogonal pretargeting.

From these screenings we conditionally selected tetrazine 3 for the biological studies. This compound showed the best profile with regard to stability and reactivity with a half-life of

Table 1 Stability data and second order rate constants for tetrazines 1–10



Compound	Stability PBS % intact at day 4[a] ($t_{1/2}$, h)[b]	Stability FBS % intact at day 4[a] ($t_{1/2}$, h)[b]	k_2 (M^{-1} s^{-1}) reaction with (E)-cyclooct-4-enol	Reaction solvent[c]
1	48 ± 1	19 ± 4 (16)	210 ± 42	MeOH
2	0 (2)	—	902 ± 164	MeOH
3	96 ± 1	40 ± 20 (82)	136 ± 54	MeOH
4	41 ± 2	23 ± 15 (26)	875 ± 84	MeOH
5	100 ± 2	84 ± 6	95 ± 15	MeOH
6	100 ± 12	100 ± 6	0.41 ± 0.02	PBS
7	98 ± 2	95 ± 9	9.3 ± 0.6	PBS
8	89 ± 5	69 ± 30	0.31 ± 0.04	PBS
9	34 ± 1	19 ± 4 (23)	807 ± 145	MeOH
10[d]	0 (9)	—	—	—

[a] Data normalized to 100% at t = 0 h, and reported as the mean value of three measurements ± SD. [b] Halflife in brackets. [c] Temperature: 25 °C for reactions performed in MeOH, 37 °C for reactions performed in PBS. [d] The k_2 of tetrazine 10 could not be determined due to its poor solubility in MeOH.

82 h in serum at 37 °C, and a k_2 of 136 M^{-1} s^{-1} in MeOH at 25 °C (Table 1). It is noteworthy that the in vitro stability of tetrazine 3 is substantially higher compared to that of TCO derivatives commonly used for Ab modification, for which half-lives of circa 3 h were reported in serum. [18] The low rate constant of compound 3 in MeOH may rise in aqueous media. In fact, it has been previously demonstrated that the rate of the IEDDA reaction is accelerated by increasing solvent polarity and temperature. [32] This observation has been confirmed for tetrazine 9, which possesses a k_2 of 6000 M^{-1} s^{-1} as has been reported under physiological conditions. [10] For antibody modification tetrazine 3 was converted to the corresponding N-hydroxysuccinimide (NHS) ester derivative, tetrazine –NHS 11.

Development and characterization of the antibody –tetrazine conjugate

The next step in our study was the development of an Ab –tetrazine conjugate where tetrazine 3 is used as the bioorthogonal tag. To this end, anti-HER2/neu antibody trastuzumab was selected as the model antibody and tagged with tetrazine 3 via the standard amine-coupling procedure.

The tetrazine loading was modulated using two different molar equivalents (10 and 100) of the amino reactive tetrazine, and the number of incorporated tags was determined by quadrupole-time-of-flight (Q-TOF) mass spectrometry. To this end, trastuzumab was used as the reference and its mass spectrum was compared to that of the trastuzumab –tetrazine 3 conjugates. As shown in Fig. 2A the mass spectrum of trastuzumab reveals the presence of a 148 117 dalton (Da) Ab with charge states from 22+ to 27+. For comparison we selected the most intense peak (25+ charge state), and determined that treatment of trastuzumab with 10 equivalents of the tetrazine –NHS 11 resulted in the incorporation, on average, of 2 tetrazine tags per antibody, whereas, an average of 8 tags was incorporated after treatment with 100 equivalents (Fig. 2B).

The number of reactive tetrazines per antibody was also evaluated based on absorbance readings after the reaction with an excess of a TCO –fluorescein dye (12). We observed that there are 0.4 and 5 reactive tetrazines for the conjugates bearing 2 and 8 tags, respectively. Remarkably, 62% of tetrazine moieties resulted to be active for the conjugate bearing 8 tags. This result indicates that the direct attachment of tetrazine tags on the antibody does not influence their reactivity, with a high number of tags remaining functional towards the cycloaddition reaction.

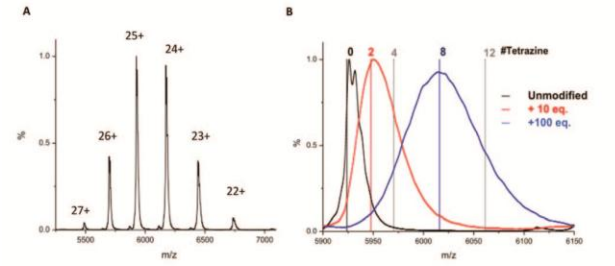


Fig. 2 Mass spectra of the trastuzumab – tetrazine 3 conjugates. (A) Mass spectrum of trastuzumab (148 117 Da) with charge states from 22+ to 27+; (B) 25+ charge state of trastuzumab (black) compared to trastuzumab – tetrazine 3 conjugates after treatment with 10 (red) and 100 (blue) equivalents of tetrazine – NHS 11.

The reaction rate of the trastuzumab –tetrazine 3 conjugate with TCO –fluorescein dye (12) was determined next. To this end, the immunoconjugate was reacted with different excesses of 12 in PBS at 37 °C. At selected time points aliquots were taken, and the reaction was quenched by addition of (E)-cyclooct-4-enol to the mixtures. After removal of the unreacted dienophiles the fluorescence signal was monitored over time and the pseudo-first-order rate constants (k_{obs}) were obtained (Fig. S6 †). From the different k_{obs} a k_2 of 3083 ± 352 M^{-1} s^{-1} was determined. As anticipated, the resulting rate constant was higher than that measured for tetrazine 3 in MeOH (k_2 = 136 M^{-1} s^{-1}).

Binding affinity of the trastuzumab –tetrazine 3 conjugate for HER2/neu receptors

We set out to determine whether the binding affinity of the immunoconjugate for the receptors is retained, and performed a saturation binding assay where triple negative human breast adenocarcinoma (MDA-MB-231, Her-2/neu negative) and human ovarian adenocarcinoma (SKOV-3, Her-2/neu positive) cells were used as the negative and the positive control respectively. Before proceeding with the assay, the cell lines were subjected to short tandem repeats (STR) and fluorescence in situ hybridization (FISH) analysis. The STR loci were concordant with the information given by the manufacturer (ATCC Rockville, Maryland) and in all the tumor cell lines the hybridization signal was uniform, whereas, FISH analysis demonstrated a high ratio of HER-2/neu over chromosome 17 in SKOV-3 (3.16), but no detectable expression in MDA-MB-231 cells.

In imaging studies, a high Ab functional loading is desirable, therefore we selected the trastuzumab –tetrazine 3 conjugate with the highest tetrazine loading (8 tags per Ab) for the binding assay. In order to introduce a fluorescent label and enable the tracking of the antibody on the cells, the immunoconjugate was reacted with TCO –fluorescein dye 12.

After incubation of the cells with this "pre-clicked" trastuzumab–tetrazine-3–TCO –fluorescein conjugate we determined a dissociation constant (K_D) of 3.9 nM which correlates with that of native trastuzumab (K_D = 5 nM as reported by the manufacturer), and that there is no unspecific binding to the cells (Fig. 3).

This result demonstrates that the binding affinity of the immunoconjugate is retained even with the higher tetrazine loading, and that this conjugate is suitable for cell imaging experiments.

Cell imaging experiments

To obtain a proof-of-concept for our approach we performed pretargeted cell imaging experiments using the newly develo-

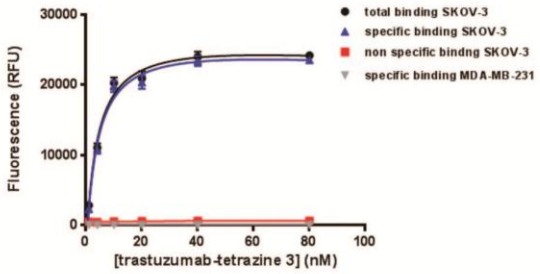


Fig. 3 Saturation plot. Binding of the trastuzumab –tetrazine- 3–TCO –fluorescein conjugate to SKOV-3 and MDA-MB-231 cells. Speci fic binding was de fined as total binding minus non-speci fic binding obtained by incubation of the immunoconjugate with 1 μM trastuzumab.

ped trastuzumab –tetrazine 3 conjugate in combination with the TCO fluorescent probes (Fig. 4 and 5).

To this end, SKOV-3 cells were incubated with 50 nM of the immunoconjugate bearing 8 tetrazine tags. After washing, Ab

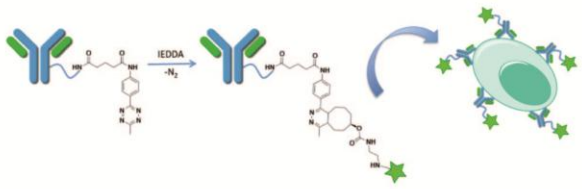


Fig. 4 Pretargeting scheme. Cells are pretargeted with trastuzumab –tetrazine 3 conjugates, and labeled with TCO fluorescent dyes via the IEDDA cycloaddition.

labeled cells were treated with two di fferent concentrations of TCO dye 12 (0.5 and 10 μM), washed again, and then imaged in the green fluorescent probe (GFP) channel. As shown in Fig. 5A, considerable membrane labeling was observed for cells treated with trastuzumab –tetrazine 3 even at low dye concentrations.

To track the antibody on the cells and at the same time verify the specificity of the bioorthogonal reaction, a dual labeled Ab-conjugate was also utilized.

The latter was prepared by treating trastuzumab with both tetrazine –NHS 11 and fluorescein –NHS ester. In this case, the Ab labeled cells were treated with Cy3 dye 13, and imaged using both GFP and red fluorescent probe (RFP) channels. The labeling of the cells was observed on both channels with good colocalization of the signal (Fig. 5Aiv –vi).

Negative controls were obtained by incubation of SKOV-3 cells with dyes 12 and 13, showing no unspecific labeling (Fig. S7iii and iv †). The cells were also labeled with control antibodies trastuzumab –fluorescein and the "pre-clicked" trastuzumab –tetrazine- 3–TCO –fluorescein (Fig. S7i and ii †).

From the perspective of applying our Ab –tetrazine conjugate for in vivo studies, we decided to assess its stability towards prolonged exposure to biological media. Particularly, we evaluated the stability of the trastuzumab –tetrazine 3 conjugate under cell labeling conditions by extending the Ab incubation time from 45 minutes to 48 hours. Membranous staining of the cells could still be observed after treatment with dye 12, indicating that the tetrazine moieties are still reactive (Fig. 5B).

These experiments clearly demonstrate the e ffectiveness of our Ab–tetrazine conjugates as pretargeting agents for bioorthogonal imaging in an in vitro set-up.

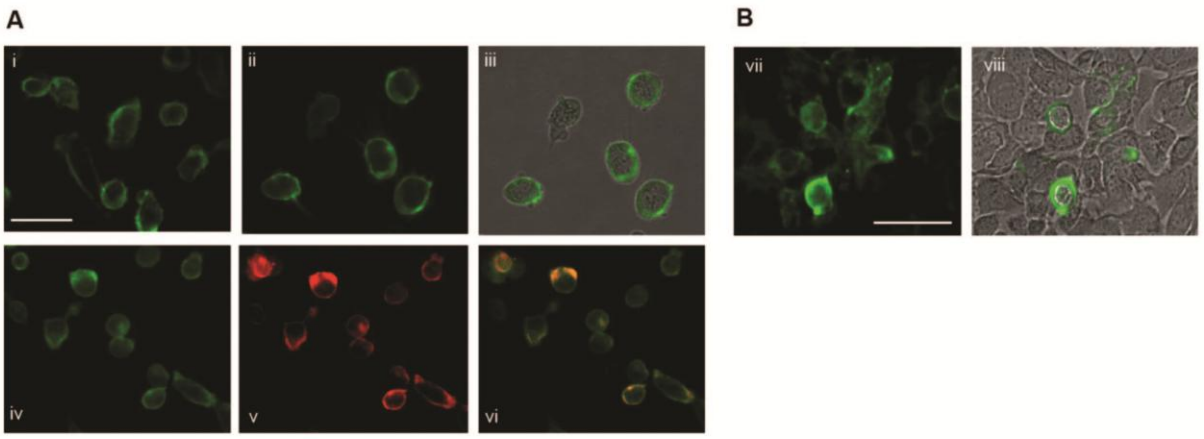


Fig. 5 Fluorescence microscopy images. (A) Images of SKOV-3 cells treated with trastuzumab –tetrazine 3 conjugates for 45 min. (i –iii) Cells were treated with trastuzumab –tetrazine 3 0.05 μM: (i) after treatment with dye 12 0.5 μM, (GFP channel), (ii) after treatment with dye 12 10 μM (GFP channel), (iii) example of overlay of the GFP channel and di fferential interference contrast (DIC). (iv –vi) Cells were treated with trastuzumab –tetrazine- 3–fluorescein 0.5 μM: (iv) image in the GFP channel, (v) after treatment with dye 13 10 μM (RFP channel), (vi) merged image of (iv) and (v). (B) Images of SKOV-3 cells treated with the trastuzumab –tetrazine 3 conjugate 0.05 μM for 48 h: (vii) after treatment with dye 12 10 μM (GFP channel), (viii) overlay of GFP and DIC. Positive and negative controls are given in the ESI (Fig. S7 †). Scale bar: 50 μm.

Conclusions

Due to their high target specificity, yet long circulation times within the body, Abs represent an excellent platform for bioorthogonal pretargeting both in cell and animal models. In fact, bioorthogonal pretargeting can significantly improve the target-to-background ratio resulting in a better target visualization and quantification compared to conventional immunotargeting techniques.

Since the introduction of the tetrazine/TCO ligation, many studies have reported the use of TCO modified Abs for pretargeted tumor imaging. However, the inverse approach has not been investigated yet, although inverting TCO with the tetrazine may address some issues related to the use of Ab–TCO conjugates. Particularly, the higher hydrophilicity of tetrazine compounds may prevent hydrophobic burying of the tags within the protein resulting in a higher e ffective functional loading.

In this study we developed a novel Ab –tetrazine conjugate and obtain a proof-of-concept for its applicability using real-time fluorescence cell imaging as a validation method. We synthesized a small library of novel 1,2,4,5-tetrazines all suitable for protein modification, and selected compound 3 for antibody labeling. This tetrazine displays a remarkable stability towards physiological media and still retains good cycloaddition reactivity with TCO. Next, we developed a novel antibody–tetrazine conjugate using tetrazine 3 as the bioorthogonal tag. The novel immunoconjugate was evaluated for pretargeted live cell imaging using two di fferent TCO fluorescent probes as imaging agents. This study demonstrates the viability of our approach, showing that the reaction between the Ab–tetrazine construct and TCO probes is fast even at micromolar concentrations and that the tetrazine tags are unaffected by prolonged exposure to physiological conditions. The high reactivity combined with the good functional density and stability of the Ab –tetrazine conjugate suggests that this strategy may be useful for in vivo positron emission tomography (PET) imaging. This will allow the in vivo translation of this approach.

Experimental

General procedures for the synthesis of 1,2,4,5-tetrazines

Method A (compounds 1 –5). Compounds 1–5 were synthesized by suspending 5-((4-cyanophenyl)amino)-5-oxopentenoic acid (14), 5-((6-cyano)-3-pyridinyl)-5-oxopentenoic acid (15) or 3-(4-cyanophenyl)-propanoic acid (1 eq.) in EtOH (2 mL). Acetamidine hydrochloride or formamidine acetate (4 to 10 eq.) was added followed by sulfur (1 eq.) or zinc triflate ($Zn(OTf)_2$, 0.5 eq.) and hydrazine hydrate (1.8 mL). The mixture was stirred at room temperature under argon for 16 to 48 h. Then it was diluted with water and cooled to 0 °C. $NaNO_2$ (10 eq.) was added, followed by dropwise addition of 1 M HCl. The acid was added until reaching pH 3. At this point gas evolution started and the solution turned pink or fuchsia. (Caution ! This step generates a large amount of toxic nitrogen oxide gas.) The mixture was stirred at 0 °C until gas evolution ceased, then EtOAc was added and the organic phase was separated. The aqueous phase was extracted three times with EtOAc. The combined organic phases were washed once with brine, and dried over $MgSO_4$. After removal of the solvent the crude residue was adsorbed onto Celite and subjected to reverse phase column chromatography (H_2O : MeOH gradient 95 : 5 to 20 : 80).

Method B (compounds 6 –8). Compounds 6–8 were synthesized by dissolving 3,6-bis-(DMP)-1,2,4,5-tetrazine (1 eq.), 3-methyl-6-thiomethyl-1,2,4,5-tetrazine (1 eq.) [33] or 6-thiomethyl-1,2,4,5-tetrazine (1 eq.) [33] in toluene (dry, 1 mL) or methanol (dry, 2 mL). 4-Aminobenzylamine (1 eq.) was added to the solution and the mixture was stirred at room temperature for 2 to 4 h under argon. After evaporation of the solvent the crude residue was adsorbed onto silica and subjected to column chromatography. The obtained amino derivatized tetrazines were subsequently refluxed with glutaric anhydride (5 eq.) in THF (dry, 2 mL) for 16 h. The final products were purified by reverse phase column chromatography (H_2O : MeOH, gradient 95 : 5 to 0 : 100).

General procedure for the synthesis of TCO dyes 12 and 13

(E)-Cyclooct-4-en-1-yl-(2-aminoethyl) carbamate (16) (1 eq.) was dissolved in anhydrous DMF (1 mL). Triethylamine (2 eq.) was added, followed by dye –NHS derivatives (1.5 eq.). The mixture was stirred overnight at room temperature under N_2. After evaporation of the solvent, the crude residue was purified by HPLC on a C-18 column (H_2O : ACN isocratic 70 : 30 + 0.1% FA) and freeze-dried. The ^{1}H-NMR spectrum of compound 13 (Fig. S3 †) revealed the presence of a 1 : 1 mixture of trans and cis isomers. Due to the very low reactivity of the cis isomer towards the IEDDA reaction the mixture was used as such for the cell imaging experiments. For simplicity, we refer only to the trans isomer in Fig. 1.

Antibody labeling

Trastuzumab (Herceptin®, Roche) stock in PBS pH 7.4, 144 μM (21 mg mL^{-1}) was diluted to a final concentration of 6.9 μM by addition of a PBS solution bu ffered at pH 9.1. The NHS ester derivative of tetrazine 3 (tetrazine –NHS 11) was dissolved in DMSO to make a 60 mM stock solution. For conjugation, the appropriate excess of tetrazine was aliquoted into the antibody solution, vortexed and reacted for 3 hours at 37 °C. The excess of tetrazine used in this experiment were 10 and 100 equivalents with respect to the antibody. The resulting Ab conjugates were purified via size exclusion chromatography (PD-10 column 5 K MWCO, GE Healthcare), concentrated to 1 mg mL^{-1} and stored in PBS at 4 °C. A dual labeled trastuzumab conjugate was obtained following a similar procedure where trastuzumab was simultaneously incubated with 100 equivalents of tetrazine –NHS 11, and 15 equivalents offluorescein –NHS ester.

Native mass spectrometry

The mass spectrometry experiments were performed on a Q-TOF-2TM instrument (Micromass, Waters). The samples

were buffer exchanged twice into 150 mM aqueous ammonium acetate at pH 7.0 using Micro Bio-Spin P-6 columns (Bio-Rad) at a concentration of 10 μM. Gold-coated nanoESI needles were prepared in-house in order to transfer 5 μL aliquots of the sample into the mass spectrometer. The instrument was tuned using the following parameters: capillary voltage 1.6 kV; nanoflow backing pressure 0.3 bar; sample cone 200 V; extractor 10 V; collision energy 100 V; backing pressure 4.0 mbar. The data were acquired and processed with Masslynx v4.1 software (Waters).

Measurement of tetrazine reactive tags on antibody

Trastuzumab –tetrazine 3 conjugates (18 μM) were treated with 20 fold excess of TCO –fluorescein dye 12 with respect to the Ab. The reaction was performed in PBS containing 1% of DMSO. After incubation for 1 h at 37 °C, the antibody conjugates were purified via size exclusion chromatography (PD-10 column 5 K MWCO, GE Healthcare), and subjected to UV-vis analysis using the following molar extinction coefficients: ε_{280} = 225 000 cm^{-1} M^{-1} for trastuzumab, [34] and ε_{500} = 70 000 cm^{-1} M^{-1}, CF = 0.3 (Thermo Scientific) for dye 12. Unmodified trastuzumab was subjected to the same treatment and used as the reference for these experiments. Each experiment was repeated in triplicate. The fluorescence spectra of the Ab conjugates were also recorded after treatment with dye 12 (Fig. S5 †).

Cell culture

Her-2/neu negative MDA-MB-231 and Her-2/neu overexpressing SKOV-3 cell lines were purchased from ATCC (Rockville, Maryland). Cells were routinely cultured in Dulbecco 's Modified Eagle Medium (DMEM) and DMEM/F12 medium respectively, supplemented with 10% (or 5% for SKOV-3 cells) heat inactivated FBS, 2 mM glutamine, 1% sodium pyruvate and 1% penicillin/streptomycin (Gibco, Life technologies). Cultures were maintained in exponential growth at 37 °C under a humidified atmosphere containing 5% CO_2. The HER-2 status was evaluated by FISH analysis. The identity of cell lines was confirmed with STR profiling. An 80 –90% confluent culture was detached with trypsin-EDTA or accutase (for SKOV-3) and washed three times with 1× PBS to remove any cell debris. Cells were counted with the Muse System (Merck Millipore, Count and Viability Assay kit) and resuspended in their respective medium at a concentration of 1 × 10^6 viable cells per 0.5 ml.

Cell saturation binding assay

The trastuzumab –tetrazine 3 conjugate with the highest tetrazine loading (8 tags per antibody) was reacted with dye 12 (7.4 eq. with respect to the antibody). After purification (PD-10 column 5 K MWCO, GE Healthcare), the resulting fluorescent conjugate, bearing 4 fluorophore moieties, was utilized for the saturation binding assay.

Briefly, increasing concentrations of the antibody conjugate (1 nM –80 nM, quadruplicate) were incubated with 1 × 10^6 SKOV-3 or MDA-MB-231 cells in a total volume of 1 mL cell medium for 2 h at 4 °C. To assess non-specific binding, unmodified trastuzumab (1 μM) was added (duplicates). Following incubation, the cells were pelleted via centrifugation (2000 rpm for 5 min at 4 °C) and, after removal of the supernatant, washed with cold 1× PBS + 0.2% tween 20 solution (repeated twice). After the last washing step the cells were resuspended in 500 μL 1× PBS and 200 μL aliquots from each tube were transferred to a 96-well plate (Nunclon Delta Surface, Thermo scientific) and the fluorescence intensity of the samples was measured (λ_{ex} = 500 nm, λ_{em} = 550 nm). The fluorescence intensity with the Ab concentration was fitted to a first-order exponential using GraphPad Prism 6, and the K_D was determined to be 3.9 nM.

Cell microscopy

Trastuzumab –tetrazine 3 (0.05 μM) and trastuzumab –tetrazine-3–fluorescein (0.5 μM) were incubated with 150 000 cells plated in 6-well plates, for 45 min in DMEM/F12 without phenol red, containing 5% FBS, 2 mM glutamine, 1% sodium pyruvate and 1% penicillin/streptomycin. The cells were then washed 3 times with cell medium before incubating with 0.5 and 10 μM of dyes 12 and 13 for 30 min in growth media. After washing 3 times with sterile 1× PBS, images were taken using an EVOS fluorescence digital inverted microscope in the GFP and RFP channels. Identical image acquisition settings were used for the experimental data. Images were analyzed using ImageJ applying identical leveling adjustments to all the experimental data across individual channels.

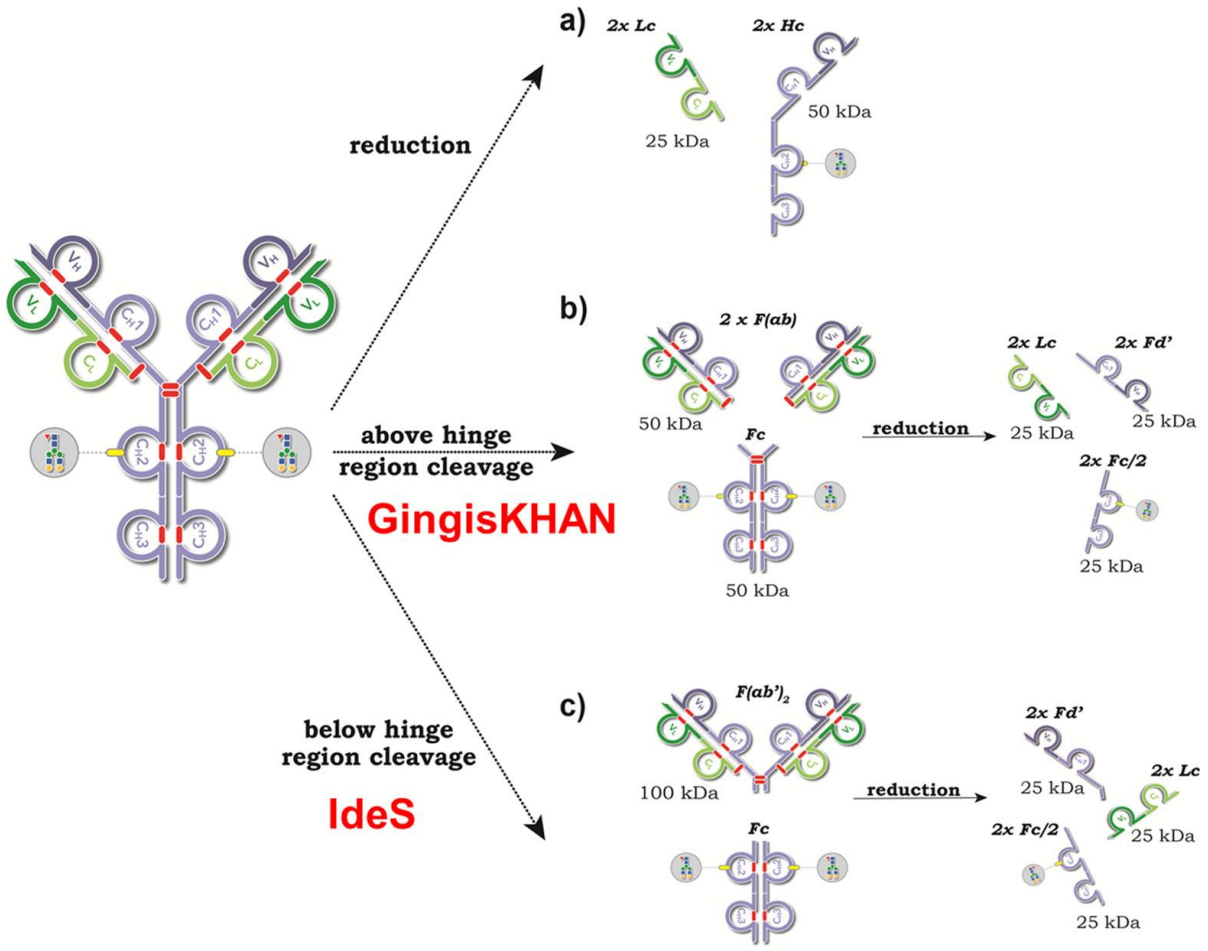
Acknowledgements

We thank Prof. Yves Guisez for his support in the mass analysis of the immunoconjugates; and Prof. Filip Lardon for access to the microscope. This work was supported by grants from the University of Antwerp (GOA-BOF) and FWO-IWT.

Notes and references

1 R. K. V. Lim and Q. Lin, Chem. Commun., 2010, 46, 1589.
2 C. S. McKay and M. G. Finn, Chem. Biol., 2014, 21, 1075.
3 E. M. Sletten and C. R. Bertozzi, Angew. Chem., Int. Ed., 2009, 48, 6974.
4 M. L. Blackman, M. Royzen and J. M. Fox, J. Am. Chem. Soc., 2008, 130, 13518.
5 N. K. Devaraj, R. Weissleder and S. A. Hilderbrand, Bioconjugate Chem, 2008, 19, 2297.
6 K. Braun, M. Wiessler, V. Ehemann, R. Pipkorn, H. Spring, J. Debus, B. Didinger, M. Koch, G. Muller and W. Waldeck, Drug. Des., Dev. Ther., 2008, 2, 289.
7 H. Wu and N. K. Devaraj, Top. Curr. Chem., 2015, 374, 3.
8 A. Darko, S. Wallace, O. Dimitrenko, M. M. Machovina, R. A. Mehl, J. W. Chin and J. M. Fox, Chem. Sci., 2014, 5, 3770.

3.4 Structural Analysis of Monoclonal Antibodies by Top-Down Mass Spectrometry: An Inter-Laboratory Study



In collaboration with

Prof. Dr. Yuri Tsybin[1]

Participants of the Consortium for Top-down Proteomics (25 labs)

[1]Biomolecular Mass Spectrometry Laboratory, EPFL ,Lausanne, Switzerland

Part of this research will be presented at the ASMS annual meeting in June 2018 in San Diego, and submitted for publication in 2018
in a joint project with the participants of the Consortium of Top-Down Proteomics

3.4.1 Introduction

The Consortium for Top-Down Proteomics (CTDP) is a non-profit organisation with members from academia and industry. For this study, a group of 25 labs (including myself representing the BAMS lab) worked together to characterise standard antibody drugs using Top-down MS methods. The goal of this study was to determine whether the current top-down MS technologies and protocols can add value to the current bottom-up MS approaches, such as ruling out artifacts, revealing new PTMS and their locations.

The CTDP would collect all the date from the different platforms and develop and validate the tools and protocols for top-down MS and data analysis, focusing on an efficient procedure for mAb characterisation. This way, the CTDP can build a database of mAb top-down data for further analysis and method development. This study will be an important contribution for the pharmaceutical and biotechnology industry to understand the potential value of top-down MS methods.

While most of the 25 labs measured the antibodies in denatured form, I used a native MS approach to analyse the samples (see also chapter 3.4.2). As the denatured form of a protein has more exposed regions compared to the native structure, the top-down MS data would be different. Rather than focusing on sequence coverage, which is assumed best when denatured and highly charged protein is used, I wanted to benchmark native approaches that preserve the higher-order structure. This approach would provide additional structural information.
The antibody samples, provided by the CTDP, were three standard IgG1s; Herceptin (trastuzumab), the NIST IgG1 standard, and the Sigma SiLu lite standard, MSQ4 (**Table 1**).

Trastuzumab is an antibody against the HER2 receptor, which is involved in breast cancer [1]. The NIST IgG is an antibody against the Respiratory Syncytial Virus (RSV). The Silu Lite antibody stands for Stable-isotope labeled universal antibody, and is produced by Sigma to use as an internal standard for quantitation of biotherapeutics.
The amino acid sequences of the antibodies are displayed in **Table 2** with the complementarity-determining regions (CDR) marked in red.

The antibodies were measured in intact state, but IgG subunits were also studied by reducing the disulfide bonds to separate the light and heavy chains. Furthermore, IgG subunits were analysed after enzymatically cleaving the hinge region with GingisKHAN or IdeS. **Figure 1** shows an overview of the analysed IgG structure and generated IgG subunits.

Table 1: Overview of analysed standard antibodies.

Name	Company	IgG subtype
MSQ4 (SiLu Lite)	Sigma	IgG1λ
NIST IgG1	NIST	HzIgG1κ
Trastuzumab (Herceptin)	Roche	HzIgG1κ

Table 2: The amino acid sequences of the heavy and light chains of the antibodies MSQ4, NIST and Trastuzumab. The CDR regions are marked in red.

SILu (MSQ4) heavy chain:
EVQLVESGGGLVQPGGSLRLSCVASGFTLNNYDMHWVRQGIGKGLEWVSKIGTAGDRYYAGSVKGRFTISRENAKDSLYLQMNSLRVGD AAVYYCARGAGRWAPLGAFDIWGQGTMVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQ SSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHE DPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTK NQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG
SILu (MSQ4) light chain
QSALTQPRSVSGSPGQSVTISCTGTSSDIGGYNFVSWYQQHPGKAPKLMIYDATKRPSGVPDRFSGSKSGNTASLTISGLQAEDEADYYCCS YAGDYTPGVVFGGGTKLTVLGQPKAAPSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYL SLTPEQWKSHRSYSCQVTHEGSTVEKTVAPTECS
NIST heavy chain:
QVTLRESGPALVKPTQTLTLTCTFSGFSLS**TAGMSVG**WIRQPPGKALEWLA**DIWWDDKKHYNPSLKD**RLTISKDTSKNQVVLKVTNMDP ADTATYYCAR**DMIFNFYFDV**WGQGTTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS GLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRV*EPKSCDKTHTCPPCP*APELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDP EVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKN QVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK
NIST light chain:
DIQMTQSPSTLSASVGDRVTITC**SASSRVGYMH**WYQQKPGKAPKLLIY**DTSKLASG**VPSRFSGSGSGTEFTLTISSLQPDDFATYYC**FQGSG YPFT**FGGGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADY EKHKVYACEVTHQGLSSPVTKSFNRGEC
Trastuzumab (HRC) heavy chain:
EVQLVESGGGLVQPGGSLRLSCAAS**GFNIKDTY**IHWVRQAPGKGLEWVAR**IYPTNGYT**RYADSVKGRFTISADTSKNTAYLQMNSLRAED TAVYYC**SRWGGDGFYAMDY**WGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQS SGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHED PEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKN QVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG
Trastuzumab (HRC) light chain:
DIQMTQSPSSLSASVGDRVTITCRAS**QDVNTAVA**WYQQKPGKAPKLLIY**SASFL**YSGVPSRFSGSRSGTDFTLTISSLQPEDFATYYC**QQHY TTPPT**FGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKAD YEKHKVYACEVTHQGLSSPVTKSFNRGEC

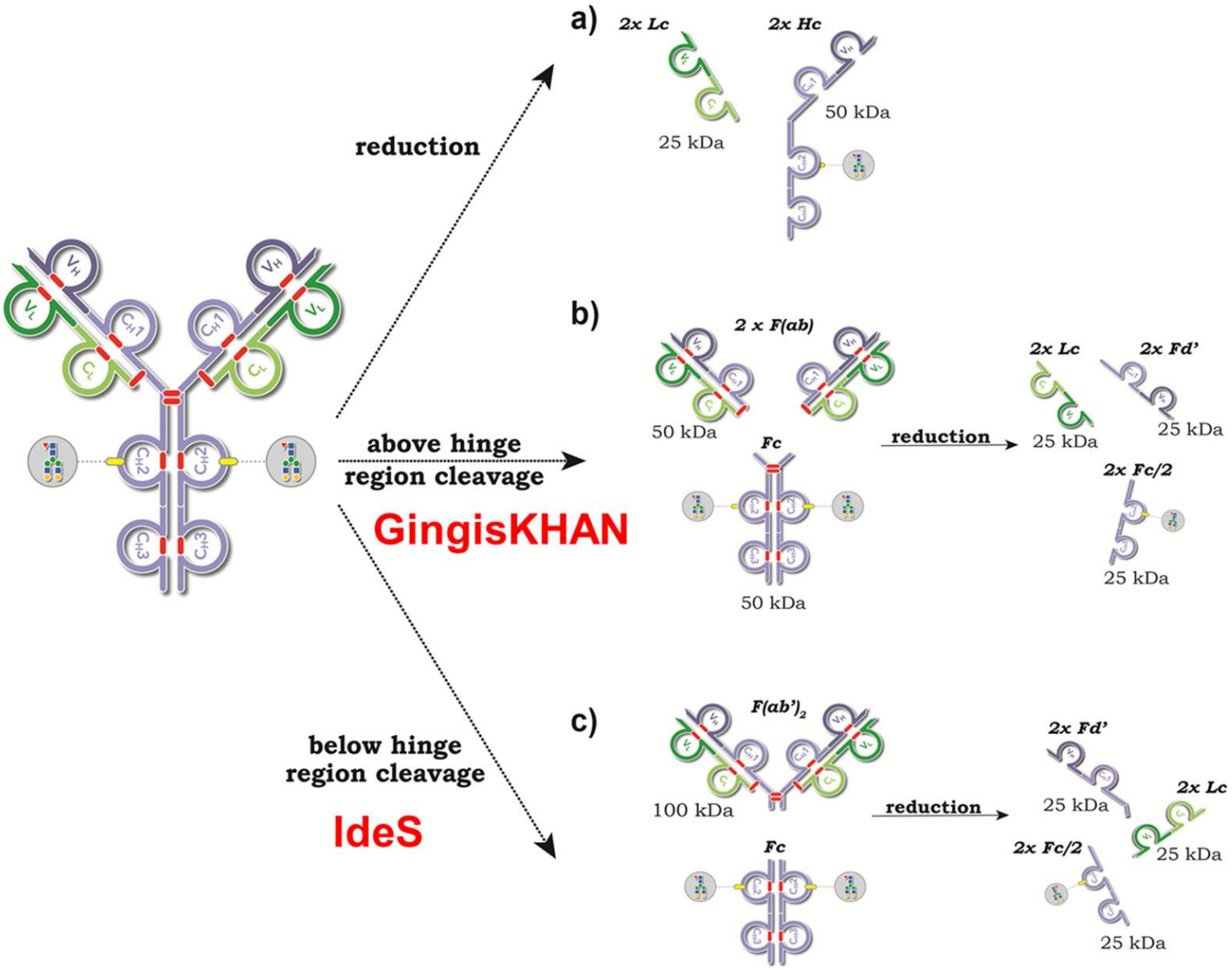


Figure 1: IgG structure and enzymatically-assisted structure-specific generation of IgG subunits (25, 50 and 100 kDa). Figure provided by CTDP.

3.4.2 Experimental details

Table 3 shows an overview of the types of analysis performed on the three standard antibodies. The same work packages (WP) were distributed to all 25 labs.

Table 3: Types of analysis performed

Work Program		Information	Comments	Performed (x)
WP1	Intact mAb (150 kDa) mass measurements	Intact MW	intact	x
WP2	Light and heavy chains	Chains MW	Reduced Mab	x
WP3	25 kDa subunit mass measurements	Subunits MW	IdeS digestion + reduced	x
WP4	50 kDa subunit mass measurements	Subunits MW	KGP digestion	x
WP5	100 kDa subunit mass measurements	Subunits MW	IdeS digestion	x
WP6	Top-down MS of 25 kDa mAb subunits	Subunit sequencing	IdeS digestion + reduced	
WP7	Top-down MS of 50 kDa mAb subunits	Subunit sequencing	KGP digestion	x
WP8	Top-down MS of 150 kDa intact mAb	mAb sequencing	Intact	x
WP9	Analysis of mAb mixtures			
WP10	Ion Mobility Separation			x

3.4.2.1 Sample preparation

The samples were buffer exchanged once with 150 mM ammonium acetate using Bio-spin size exclusion columns (Bio-rad). To obtain final concentrations of 1 mg/mL 150 mM ammonium acetate was used for dilution. For antibody reduction, the sample was incubated with 15 mM DTT at 60 °C for 1 hour.

3.4.2.2 Separation

The native IM-MS experiments were performed using direct infusion. Gold-coated nESI needles were made in-house.

3.4.2.3 Intact mass measurements

The experiments were performed on a Synapt G2 HDMS (Waters) in TOF mode. The following instrument parameters were used: capillary 1.5kV, sampling cone 70V, extraction cone 1.0V, nano flow gas pressure 0.1 bar, trap collision energy 150.0V, trap DC bias 45.0 V and backing pressure 5.0 bar.

3.4.2.4 Tandem mass spectrometry

The methods of fragmentation used were ETD and CID. The top-down fragmentation was applied on non-denatured samples. For the ETD measurements the instrument parameters were: capillary voltage 1.6kV, discharge current 15.0 mA, source temperature 30°C, sampling cone 170.0V, extraction cone 5.0 V, makeup gas flow 50 mL/min, desolvation temperature 150°C and backing pressure 5.0 mbar. For the CID measurements the instrument parameters were: capillary 1.5kV, sampling cone 70V, extraction cone 1.0V, nano flow gas pressure 0.1 bar, trap collision energy 170.0V, trap DC bias 45.0 V and backing pressure 5.0 mbar.

3.4.2.5 Data processing and bioinformatics

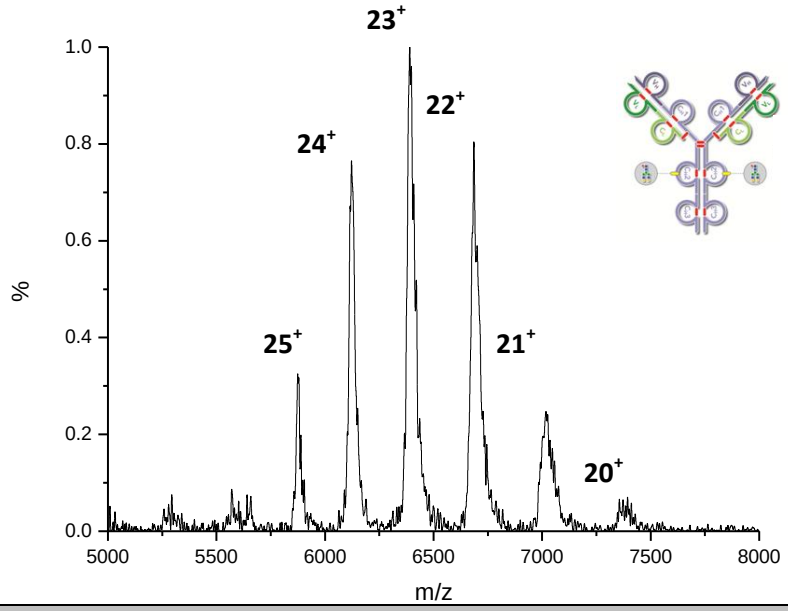
The data were acquired with Masslynx v4.1 (Waters). The spectra for native MS were smoothed 4 times over a smooth window of 10 channels. The sample mass was calculated using MaxEnt 1. The ETD spectra were a combination of 800 scans and were smoothed 5 times over a smooth window of 2 channels. The peak picking was performed manually in Masslynx, by selecting the peaks with high S/N. The CID spectra were a combination of 100 scans and were smoothed 5 times over a smooth window of 2 channels. The ion mobility drift times are displayed using Driftscope v2.3.

3.4.2.6 Complementary approaches

Native ion mobility mass spectrometry was performed on a Synapt G2 HDMS. The drift times were converted to collision cross sections using the drift times of standard protein calibrants as described previously [2]. The instrument parameters were the following; capillary 1.5kV, sampling cone 70V, extraction cone 1.0V, nano flow gas pressure 0.1 bar, trap collision energy 10.0V, trap DC bias 45.0 V and backing pressure 5.0 mbar.

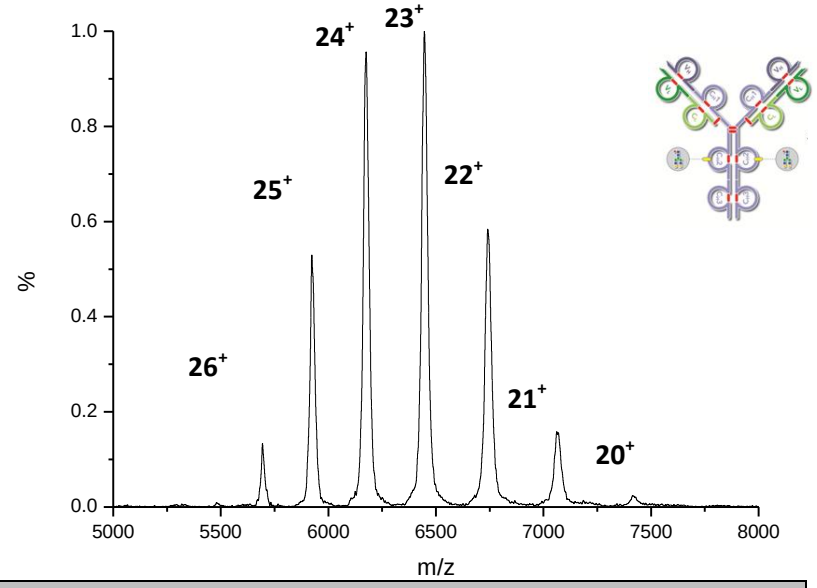
3.4.3 Intact Mass Measurement (WP1)

Figures 2 – 4 display the mass spectra of the intact antibodies MSQ4, NIST and Trastuzumab, respectively. The spectra of the IgG1s were similar in charge state distribution and mass, as expected.



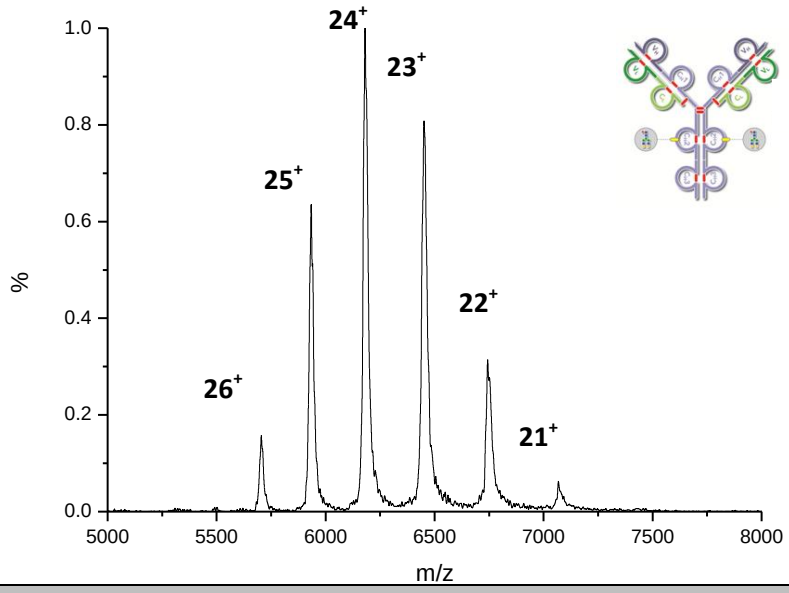
MSQ4 Intact		
IgG	**Experimental mass (Da)**	**Theoretical mass (Da)**
Intact	146 822	143 601

Figure 2: Mass spectrum of intact MSQ4.



NIST Intact		
IgG	**Experimental mass (Da)**	**Theoretical mass (Da)**
Intact	148 275	145 434

Figure 3: Mass spectrum of intact NIST IgG1.



Trastuzumab Intact		
IgG	**Experimental mass (Da)**	**Theoretical mass (Da)**
Intact	148 191	145 163

Figure 4: Mass spectrum of intact Trastuzumab.

3.4.4 Separated light and heavy chains mass measurements (WP2)

The reduction of the disulfide bonds of the IgG generates light and heavy chains(**Fig. 5**). **Figures 6-8** show the spectra of the reduced IgGs.

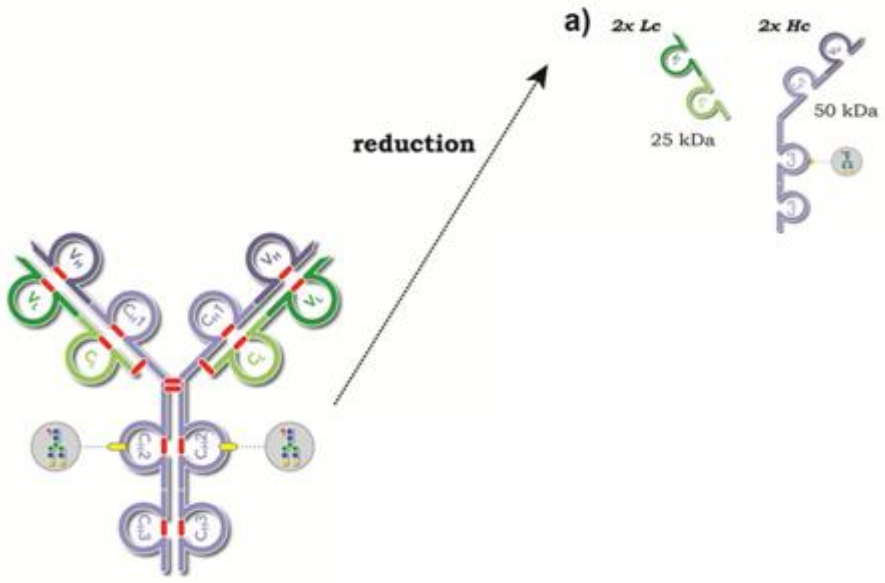
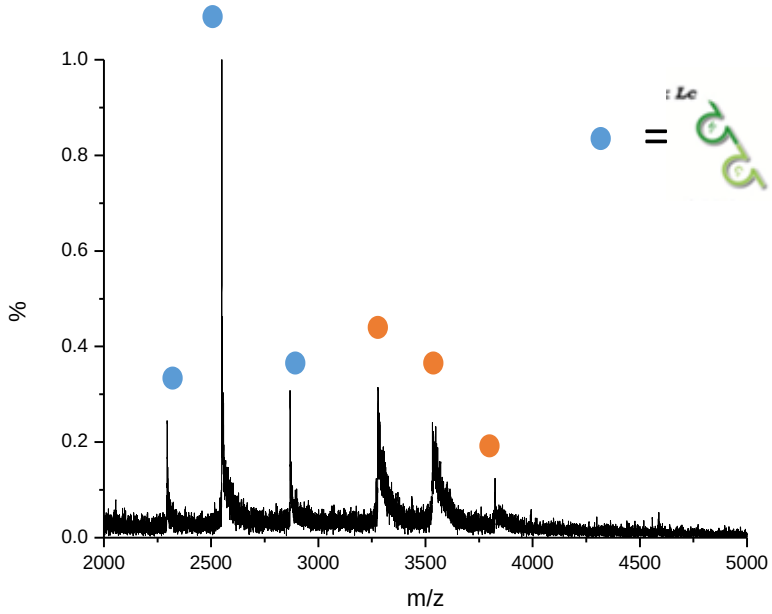
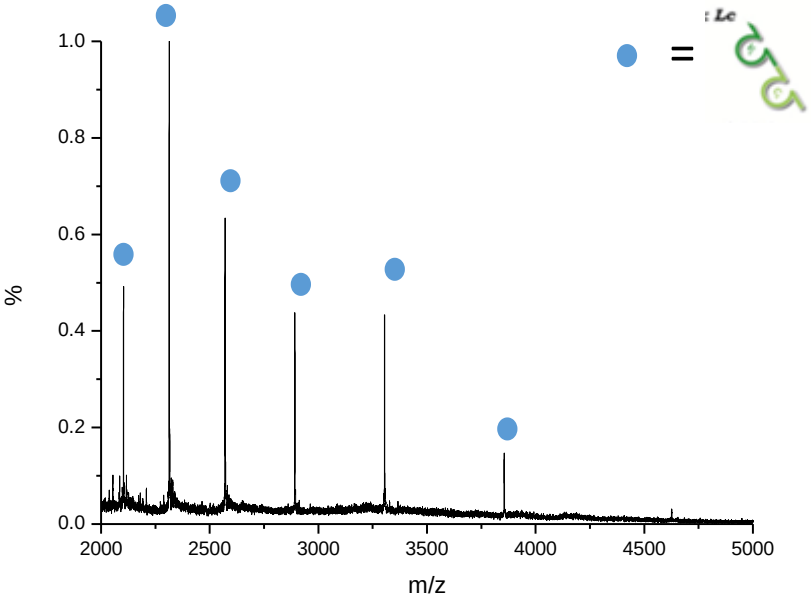


Figure 5: Schematic representation of the IgG subunits generated after reduction of the disulfide bridges.



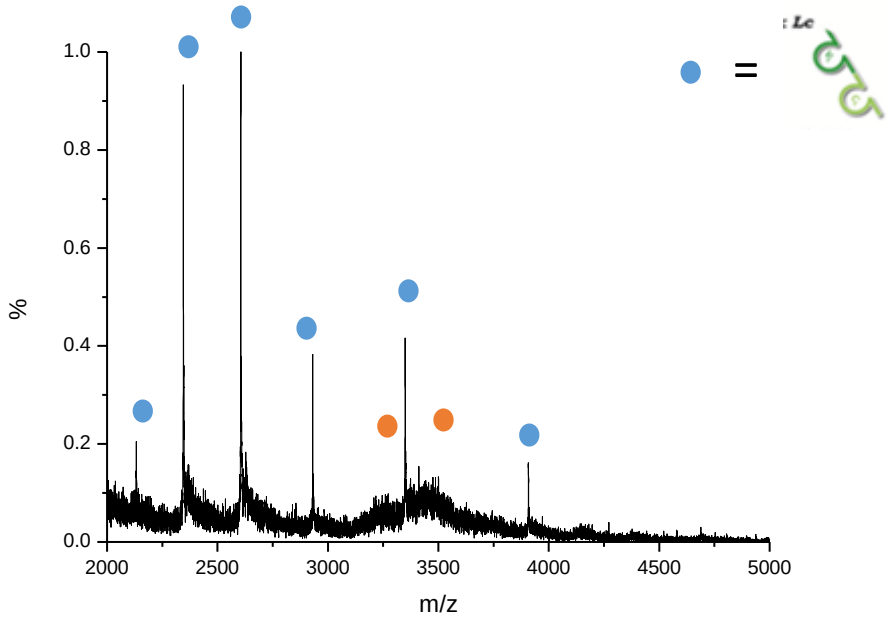
	Reduced MSQ4		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	22 940	22 959
●	**$(LC)_2$**	45 901	45918

Figure 6: Mass spectrum of reduced MSQ4



	Reduced NIST		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	23 126	23 127

Figure 7: Mass spectrum of reduced NIST



	Reduced Trastuzumab		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	23 441	23 443
●	**$(LC)_2$**	46 881	46 886

Figure 8: Mass spectrum of reduced Trastuzumab

3.4.5 25 kDa subunits mass measurements (WP3)

GinigsKhan cleaves just above the hinge region and generates two Fab subunits and one Fc subunit. When this digested sample is reduced six subunits are generated of ca. 25 kDa; two light chains (LC), two portions of the heavy chain which is included in the Fab fragment (Fd') and two halves of the Fc subunit (Fc/2) (**Fig. 9**).

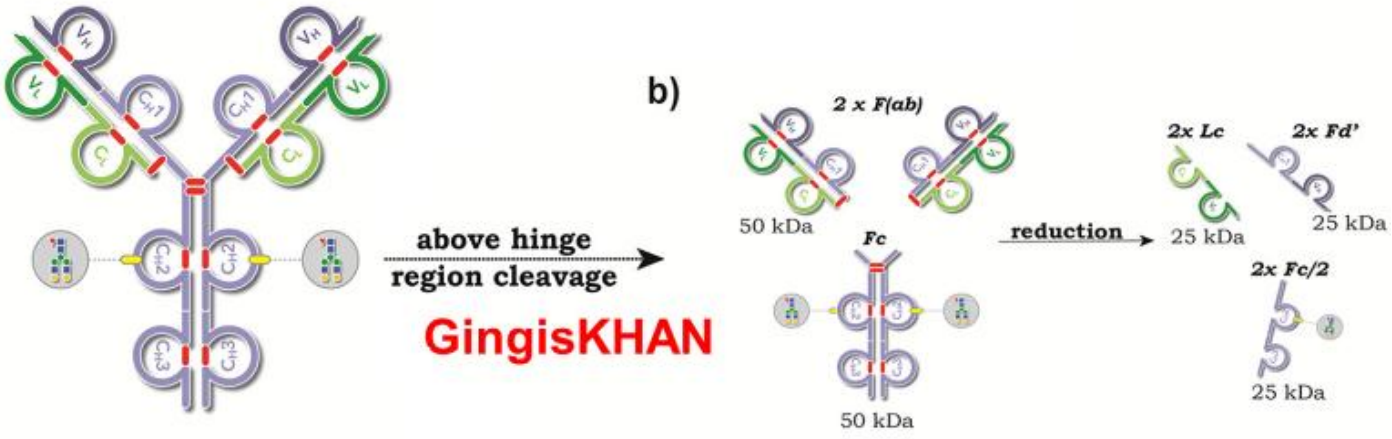
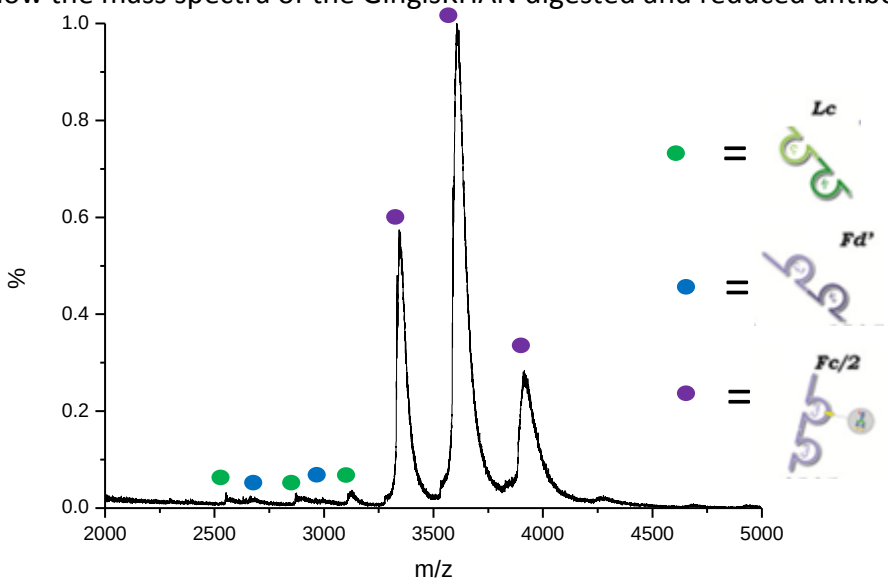


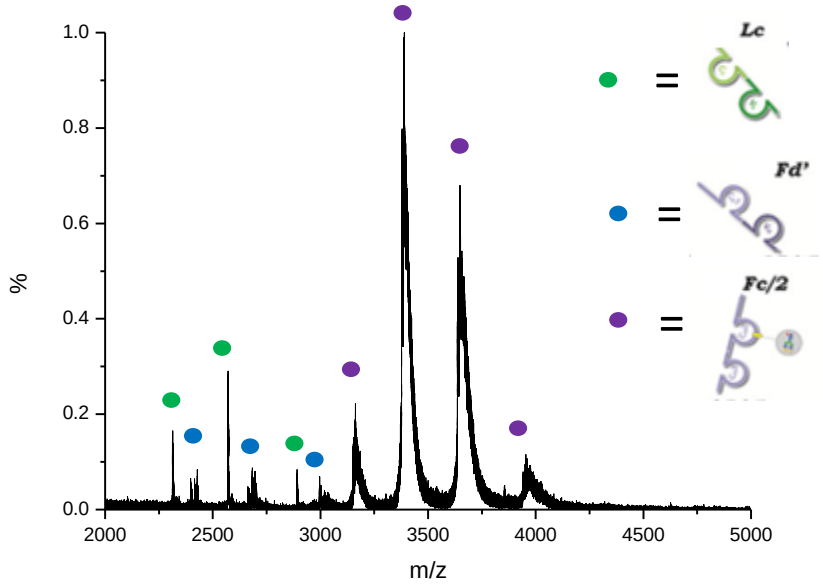
Figure 9: Schematic representation of the IgG subunits generated digestion with GingisKHAN and reduction.

Figures 10-12 show the mass spectra of the GingisKHAN digested and reduced antibodies.



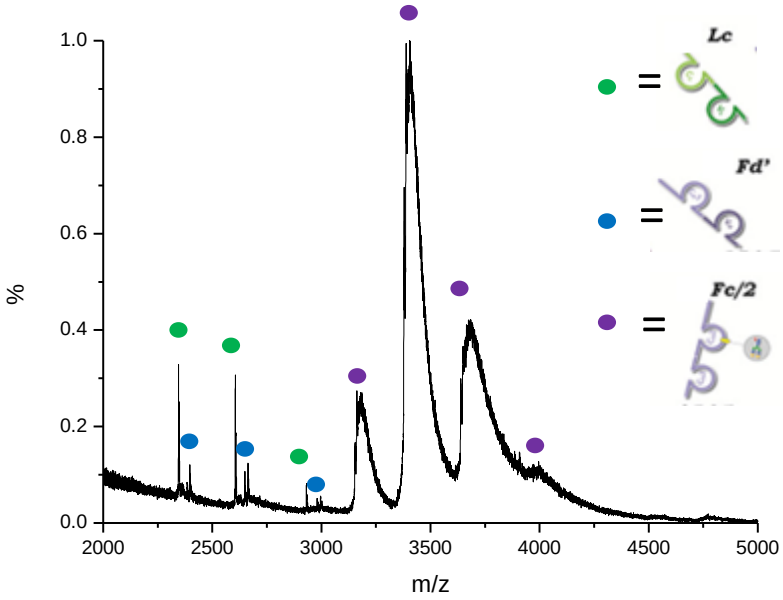
	MSQ4 + GINGISKHAN, Reduced		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	22 975	22 959
●	**Fd'**	23 945	23 799
●	**Fab**	46 814	46 741

Figure 10: Mass spectrum of MSQ4 cleaved with GingisKHAN and reduced.



	NIST + GINGISKHAN, Reduced		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	23 162	23 127
●	**Fd'**	24 260	24 288
●	**Fab**	47 469	47 398

Figure 11: Mass spectrum of NIST cleaved with GingisKHAN and reduced.



	Trastuzumab + GINGISKHAN, Reduced		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	23 460	23 443
●	**Fd'**	23 976	23 965
●	**Fab**	47 485	47 390

Figure 12: Mass spectrum of Trastuzumab cleaved with GingisKHAN and reduced.

The IdeS enzyme cleaves the antibody below the hinge region, generating a $(Fab)_2$ and a Fc subunit. After reducing the digested sample six subunits are generated; two light chains (LC), two heavy chain fragments (Fd') and two halves of the Fc subunit (Fc/2) (**Fig. 13**).

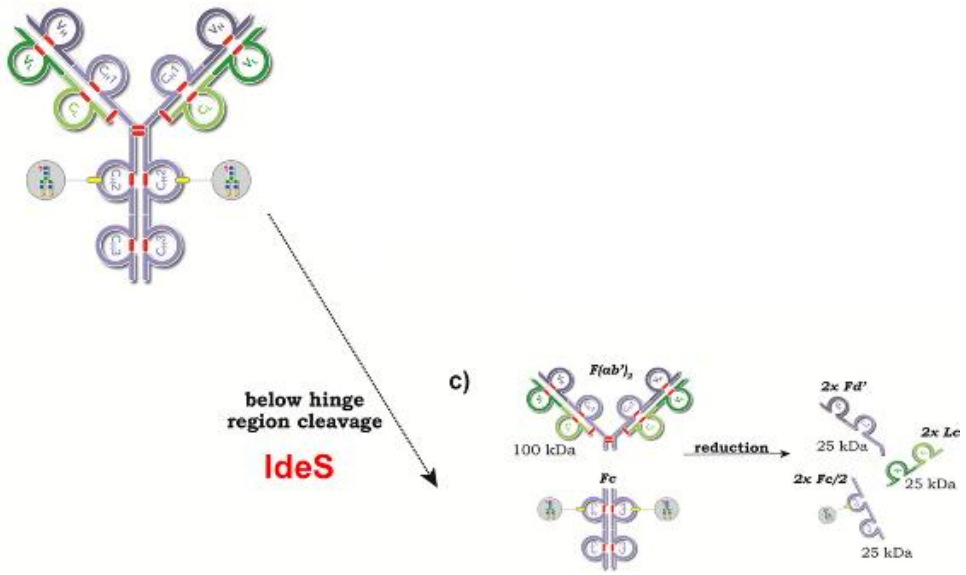
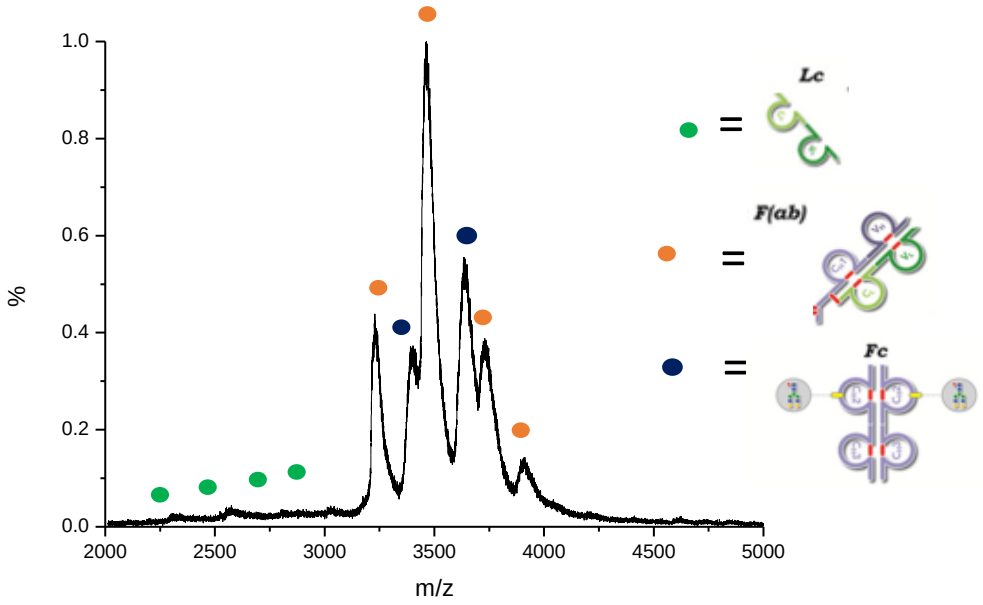


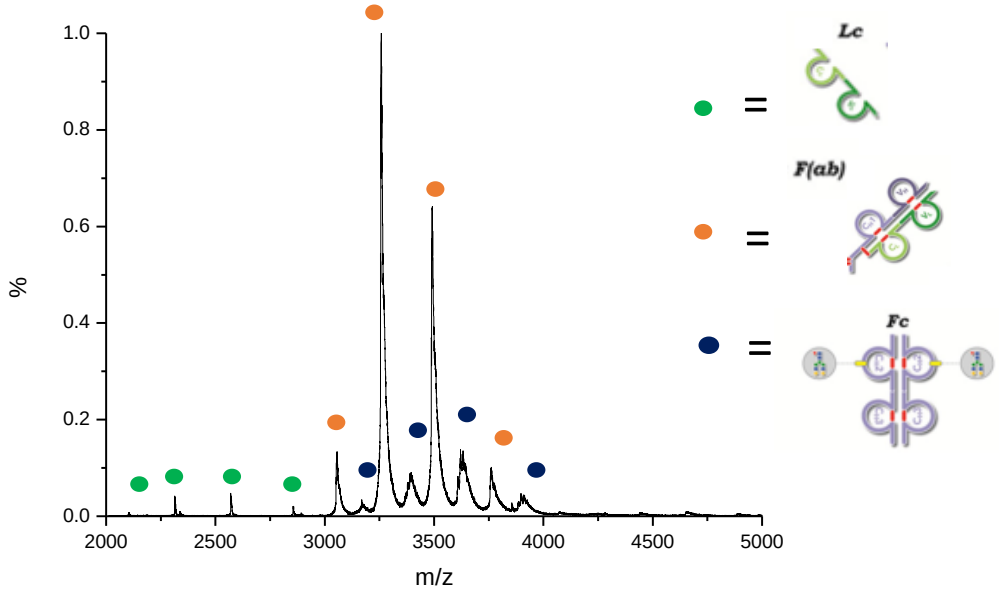
Figure 13: Schematic representation of the IgG subunits generated digestion with IdeS and reduction.

Figures 14-16 show the mass spectra of the IdeS digested and reduced antibodies.



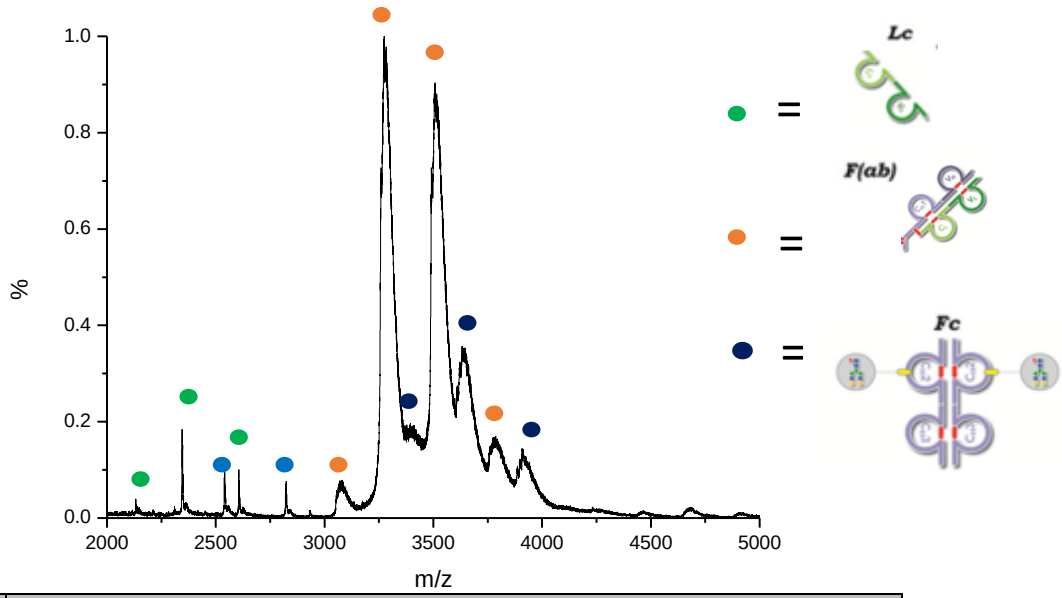
	MSQ4 + IdeS, Reduced		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	23 019	22 959
●	**Fab**	48 418	48 059
●	**Fc**	50 878	47 516

Figure 14: Mass spectrum of MSQ4 cleaved with IdeS and reduced.



	NIST + IdeS, Reduced		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	23 165	23 127
●	**Fab**	49 063	48 815
●	**Fc**	50 892	47 838

Figure 15: Mass spectrum of NIST cleaved with IdeS and reduced.



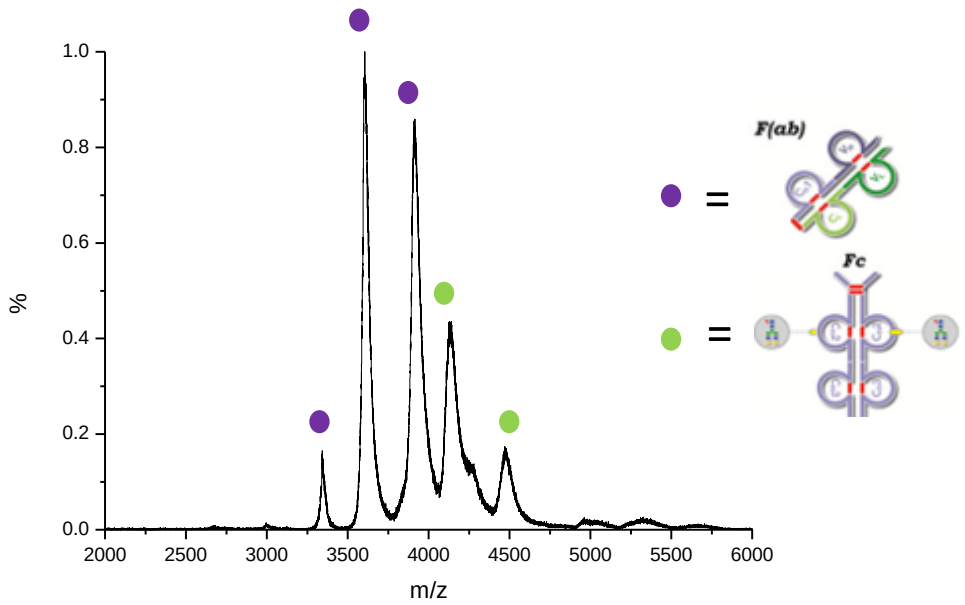
	Trastuzumab + IdeS, Reduced		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**LC**	23 456	23 443
●	**Fc/2**	25 407	23 790
●	**Fab**	49 095	48 808
●	**Fc**	50 859	47 580

Figure 16: Mass spectrum of NIST cleaved with IdeS and reduced.

3.4.6 50 kDa subunits mass measurements (WP4)

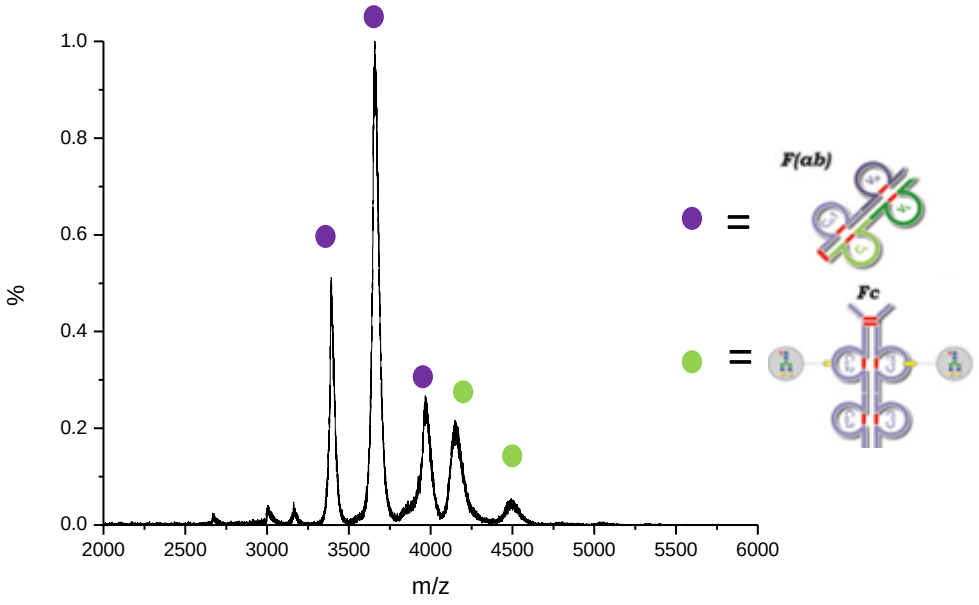
50 kDa IgG subunits can be generated via a GingisKHAN digest, which cleaves the antibody above the hinge region (**Fig. 17**). **Figures 18-20** show the mass spectra of the GingisKHAN digested antibodies.

Figure 17: Schematic representation of the IgG subunits generated digestion with GingsKhan.



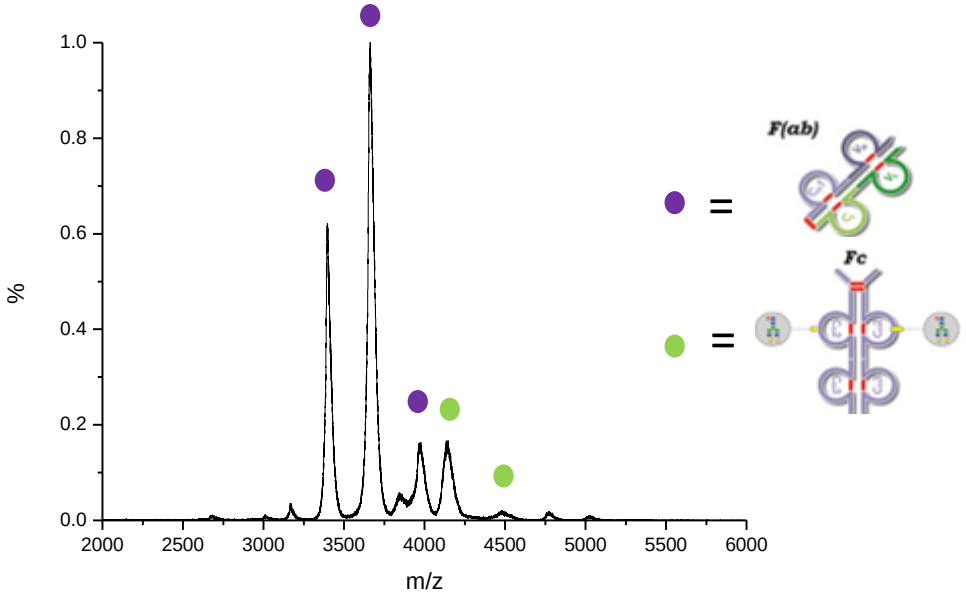
	MSQ4 + GINGISKHAN		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**Fab**	46 972	46 741
●	**Fc**	53 642	50 353

Figure 18: Mass spectrum of MSQ4 cleaved with GingisKHAN.



	NIST + GINGISKHAN		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**Fab**	47 553	47 398
●	**Fc**	53 949	50 673

Figure 19: Mass spectrum of NIST cleaved with GingisKHAN.



	Trastuzumab + GINGISKHAN		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**Fab**	47 676	47 390
●	**Fc**	53 855	50 416

Figure 20: Mass spectrum of Trastuzumab cleaved with GingisKHAN.

3.4.7 100 kDa subunits mass measurements(WP5)

100 kDa IgG subunits can be generated via an IdeS digest, which cleaves the antibody above the hinge region (**Fig. 21**).

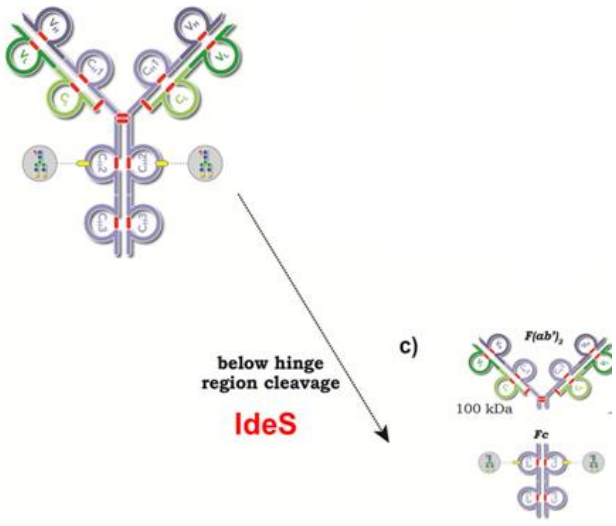
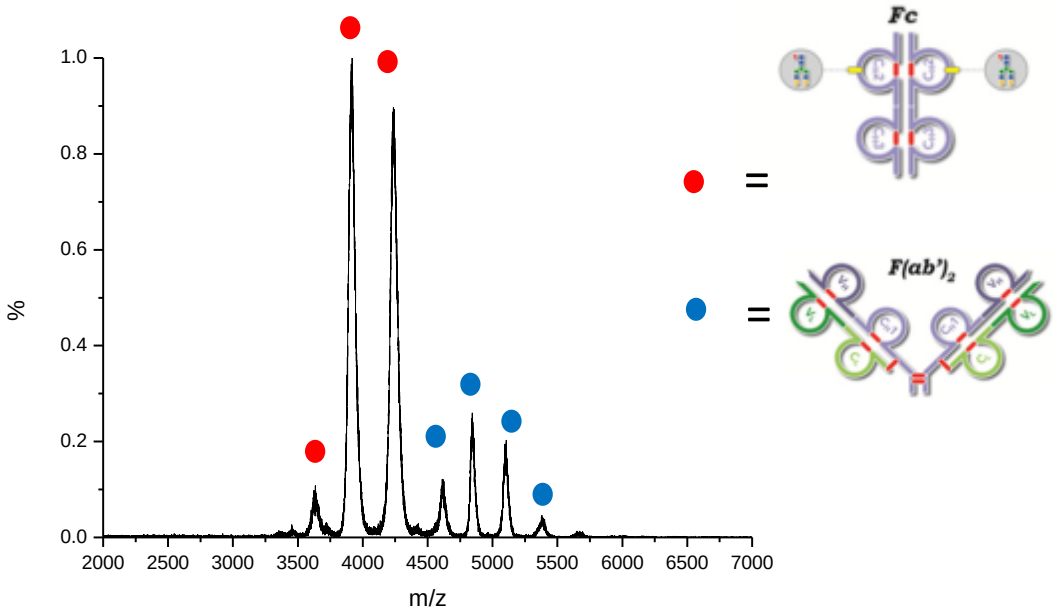
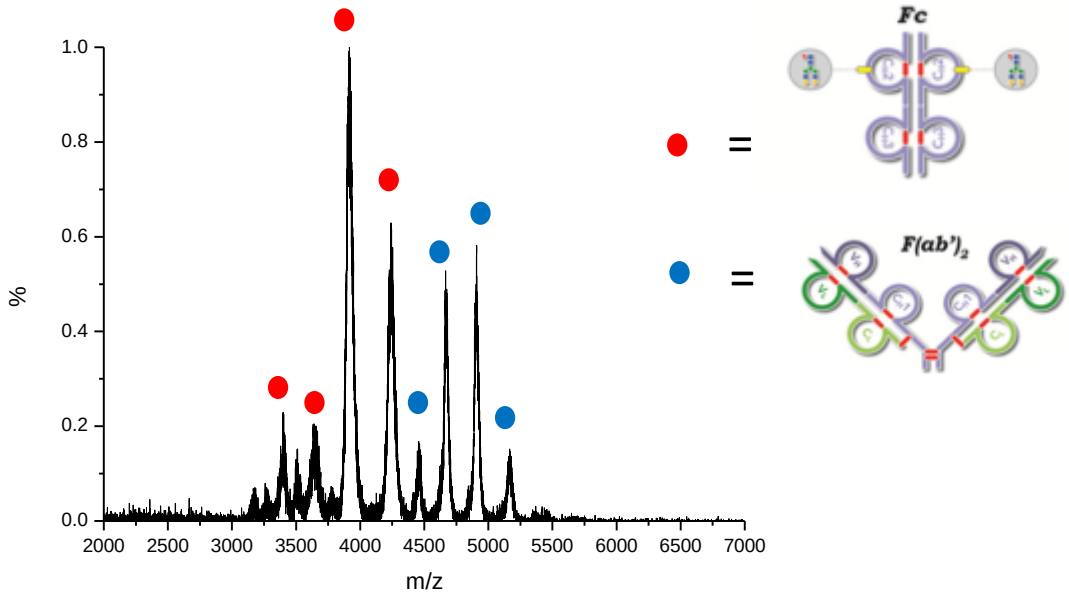


Figure 21: Schematic representation of the IgG subunits generated digestion with IdeS.



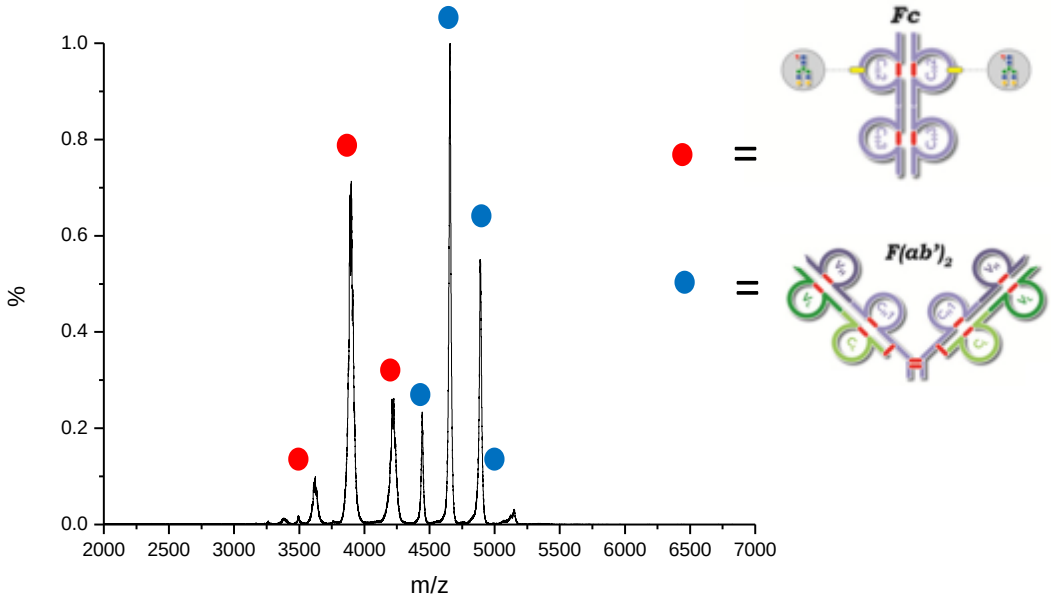
	MSQ4 + IdeS		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**Fc**	50 883	47 516
●	**$(Fab)_2$**	96 848	96 118

Figure 22: Mass spectrum of MSQ4 cleaved with IdeS.



	NIST + IDES		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**Fc**	50 881	47 838
●	**$(Fab)_2$**	98 139	97 630

Figure 23: Mass spectrum of NIST cleaved with IdeS.



	Trastuzumab + IdeS		
	IgG subunit	**Experimental mass (Da)**	**Theoretical mass (Da)**
●	**Fc**	50 545	47 580
●	**$(Fab)_2$**	97 789	97 616

Figure 24: Mass spectrum of Trastuzumab cleaved with IdeS.

3.4.8 Ion mobility separation (WP10)

With native ion mobility-mass spectrometry the collision cross-sections of the IgGs and IgG subunits were determined. Since the three studied IgGs are all of the same subtype IgG1 similar IM-MS results are expected. **Figure 25** shows the ion mobility spectra of the intact antibodies and **Table 4** shows the ion mobility drift times and CCS values for each observed charge state. **Figure 26** displays the ion mobility spectra of the IdeS digested antibodies with the drift times and CCS values of the observed charge states shown in **Table 5**. The ion mobility spectra of the GingisKHAN digested antibodies are displayed in **Figure 27** with the ion mobility drift times and CCS values shown in **Table 6**.

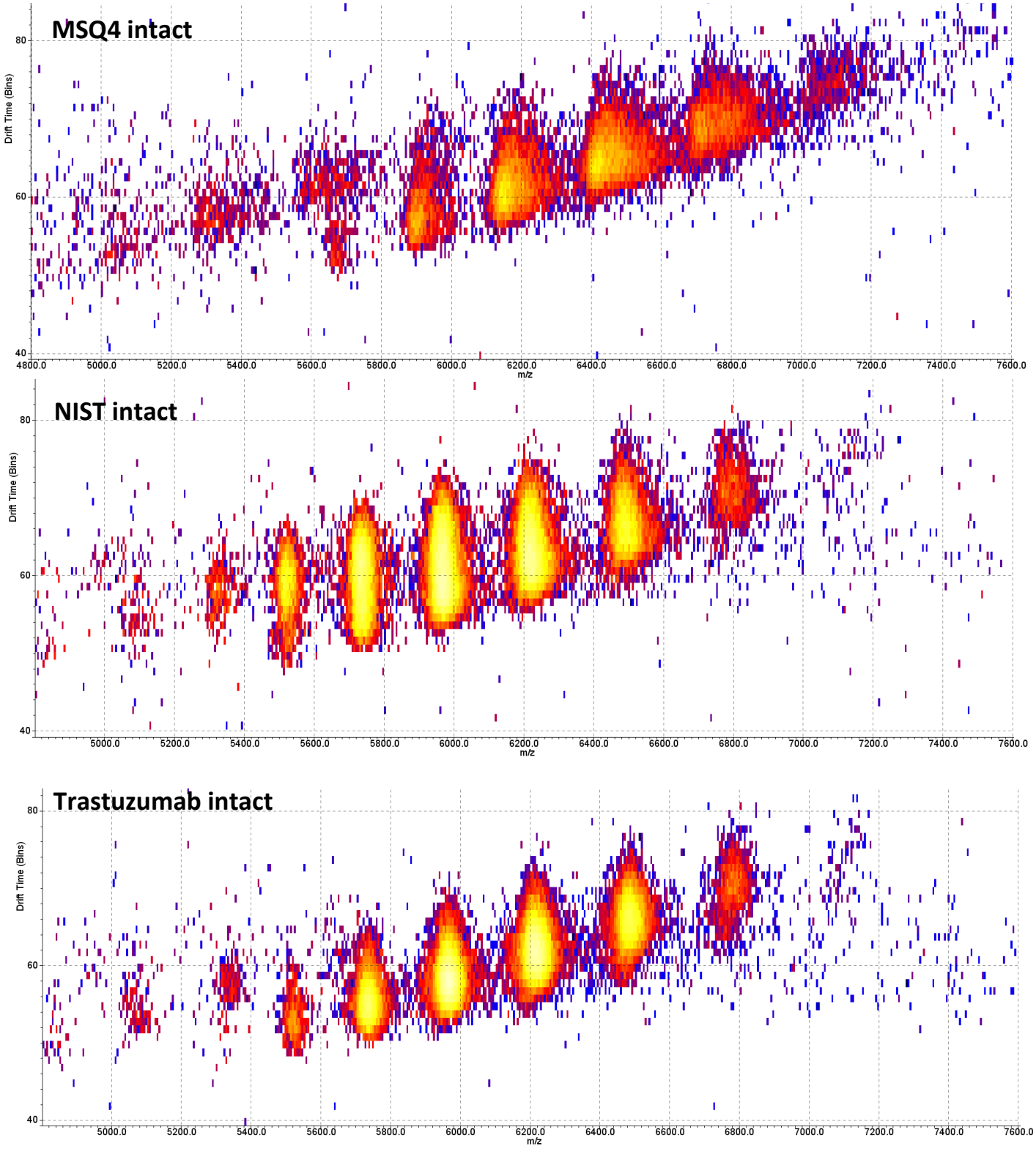


Figure 25: Ion mobility spectra of intact MSQ4, NIST and Trastuzumab.

Table 4: IM drift times and CCS values of intact MSQ4, NIST and Trastuzumab.

MSQ4 INTACT		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
21^+	13.30 ± 0.55	7385 ± 160
22^+	12.21 ± 0.54	7394 ± 172
23^+	11.30 ± 0.72	7420 ± 247
24^+	10.75 ± 0.55	7541 ± 202
25^+	10.21 ± 0.36	7644 ± 142
26^+	9.84 ± 0.16	7796 ± 67

NIST INTACT		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
22^+	12.39 ± 0.55	7452 ± 174
23^+	11.85 ± 0.17	7609 ± 58
24^+	11.12 ± 0.45	7677 ± 164
25^+	10.75 ± 0.25	7856 ± 96
26^+	10.94 ± 0.26	8247 ± 104
27^+	10.57 ± 0.43	8410 ± 180

Trastuzumab INTACT		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
22^+	12.39 ± 0.41	7452 ± 130
23^+	11.66 ± 0.54	7544 ± 183
24^+	10.94 ± 0.54	7611 ± 197
25^+	10.21 ± 0.49	7643 ± 193
26^+	9.66 ± 0.54	7720 ± 226
27^+	9.29 ± 0.31	7853 ± 138

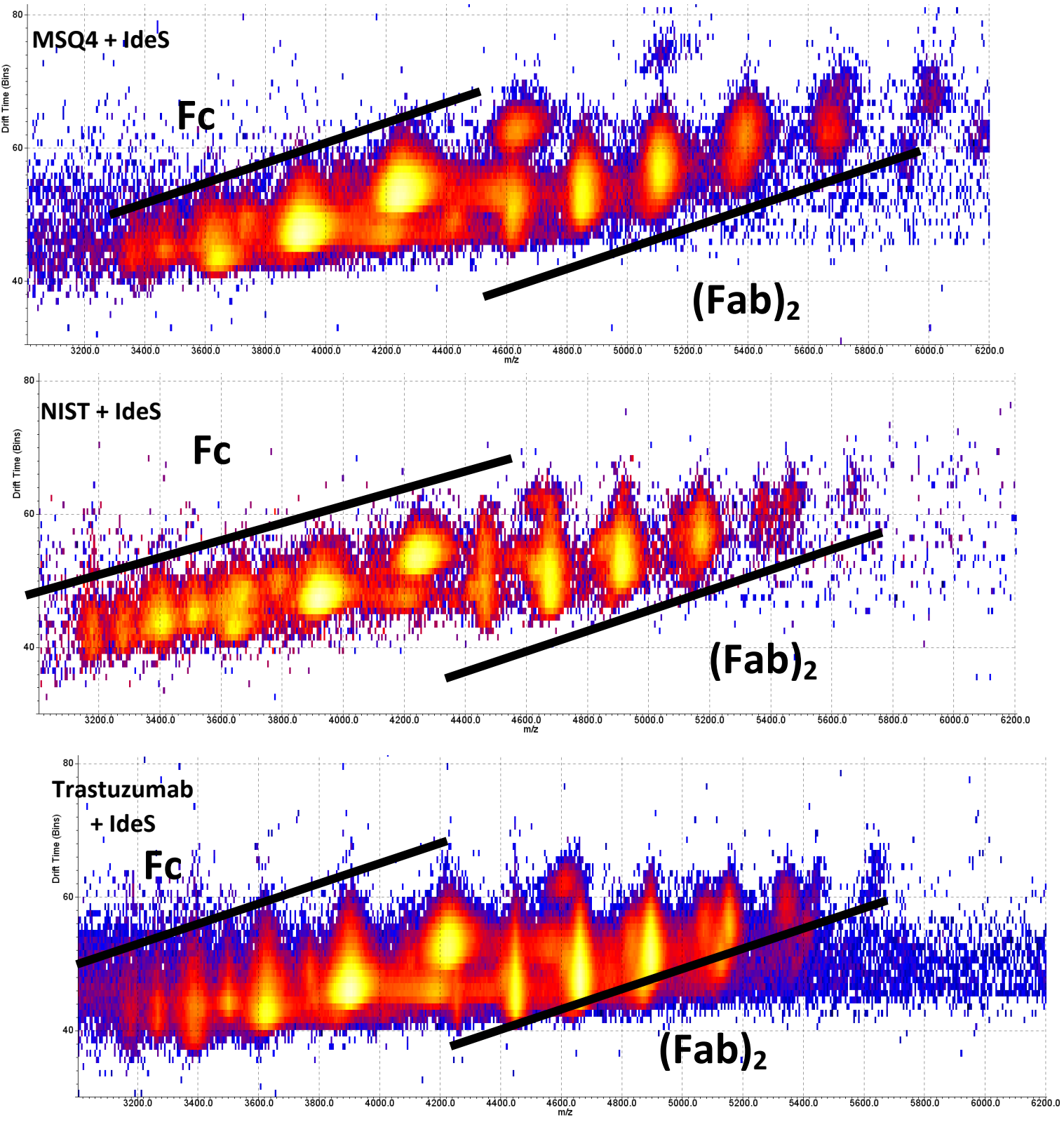


Figure 26: Ion mobility spectra of IdeS digested MSQ4, NIST and Trastuzumab.

Table 5: IM drift times and CCS values of IdeS digested MSQ4, NIST and Trastuzumab.

MSQ4 $(Fab)_2$		
Charge State	**Drift time (ms)**	**CCS ($Å^2$)**
18+	10.94 ± 0.36	5710 ± 99
19+	9.84 ± 0.46	5699 ± 140
20+	9.29 ± 0.55	5819 ± 181
21+	8.93 ± 0.47	5984 ± 165

NIST $(Fab)_2$		
Charge State	**Drift time (ms)**	**CCS ($Å^2$)**
18+	10.21 ± 0.59	5505 ± 138
19+	9.11 ± 0.61	5470 ± 192
20+	8.75 ± 0.55	5636 ± 186
21+	8.57 ± 0.50	5854 ± 180

Trastuzumab $(Fab)_2$		
Charge State	**Drift time (ms)**	**CCS ($Å^2$)**
19+	9.84 ± 0.54	5398 ± 164
20+	8.93 ± 0.57	5412 ± 190
21+	8.38 ± 0.62	5508 ± 224
22+	7.84 ± 0.90	5583 ± 348

MSQ4 Fc		
Charge State	**Drift time (ms)**	**CCS ($Å^2$)**
12+	9.29 ± 0.37	3493 ± 73
13+	8.20 ± 0.36	3542 ± 83
14+	7.47 ± 0.37	3631 ± 95

NIST Fc		
Charge State	**Drift time (ms)**	**CCS ($Å^2$)**
12+	9.48 ± 0.27	3531 ± 53
13+	8.38 ± 0.22	3583 ± 48
14+	7.65 ± 0.35	3677 ± 88

Trastuzumab Fc		
Charge State	**Drift time (ms)**	**CCS ($Å^2$)**
12+	9.29 ± 0.37	3493 ± 74
13+	8.02 ± 0.36	3501 ± 83
14+	7.47 ± 0.37	3631 ± 95

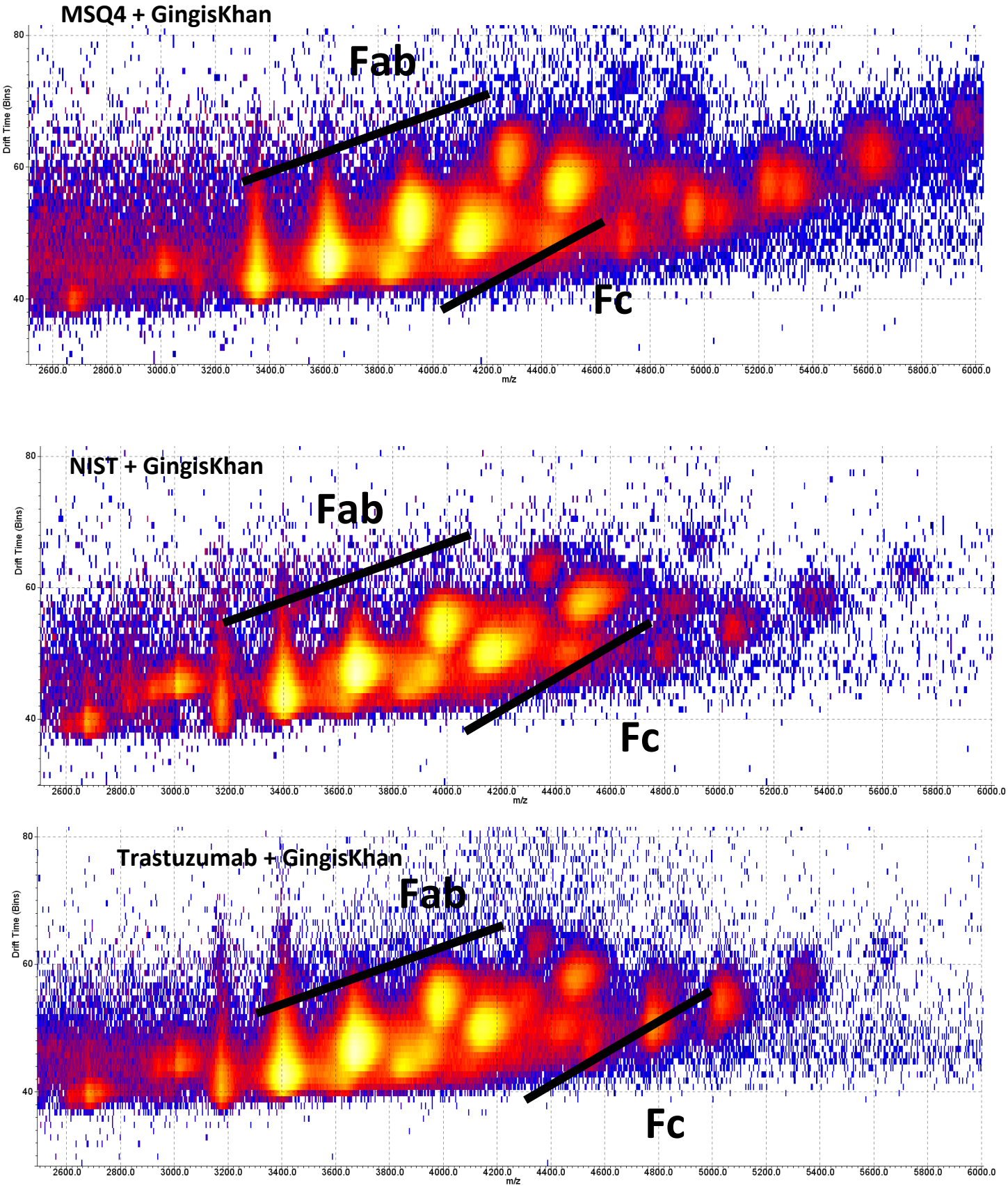


Figure 27: Ion mobility spectra of GingsKhan digested MSQ4, NIST and Trastuzumab.

Table 6: IM drift times and CCS values of GingisKHAN digested MSQ4, NIST and Trastuzumab

MSQ4 Fab		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
12+	9.11 ± 0.54	3458 ± 123
13+	8.02 ± 0.36	3501 ± 61
14+	7.47 ± 0.37	3632 ± 95

NIST Fab		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
12+	9.66 ± 0.18	3567 ± 35
13+	8.20 ± 0.36	3543 ± 82
14+	7.65 ± 0.19	3678 ± 48

Trastuzumab Fab		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
12+	9.48 ± 0.36	3531 ± 53
13+	8.20 ± 0.37	3543 ± 84
14+	7.47 ± 0.23	3632 ± 60

MSQ4 Fc		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
12+	10.02 ± 0.55	3636 ± 105
13+	8.57 ± 0.36	3626 ± 80

NIST Fc		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
12+	10.21 ± 0.36	3672 ± 51
13+	8.93 ± 0.28	3706 ± 61

Trastuzumab Fc		
Charge State	**Drift time (ms)**	**CCS (Å^2)**
12+	10.21 ± 0.36	3672 ± 68
13+	8.75 ± 0.36	3666 ± 80

3.4.9 Top-down analysis of 50kDa mAb subunits (WP7)

Using native top-down ETD, the aim is to identify the solvent-exposed regions of the antibody. To increase the solvent-accessibility the antibody can be cleaved into subunits, with IdeS or GingisKHAN. In this part of the chapter, the results of the NIST are discussed. The data for MSQ4 and Trastuzumab were analysed in a similar way and the data is summarized in the end of the chapter. In **Figure 28** the sequence of NIST is displayed and in the sequence coverage of IdeS digested NIST is highlighted In green. The detected ETD fragment ions are displayed in **Tables 7-8**. The acquisition time of this experiment was 800 scans. The LC coverage was 13/213 (6.10 %) and the HC coverage was 35/450 (7.78 %).

NIST heavy chain:
QVTLRESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIRQPPGKALEWLADIWWDDKKHYNPSLKDRLTISKDTSKNQVVLKVTNMDPADTAT
YYCARDMIFNFYFDVWGQGTTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTV
PSSSLGTQTYICNVNHKPSNTKVDKRV*EPKSCDKTHTCPPCP*APELLG
IdeS cleaves
GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPI
EKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVM
HEALHNHYTQKSLSLSPGK
NIST light chain:
DIQMTQSPSTLSASVGDRVTITCSASSRVGYMHWYQQKPGKAPKLLIYDTSKLASGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCFQGSGYPFTF
GGGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEV
THQGLSSPVTKSFNRGEC

Figure 28: Sequence of NIST IgG1 with the native ETD sequence coverage of IdeS digested NIST marked in green.

Table 7: List of detected ETD fragment ions of the $(Fab)_2$ subunit of IdeS digested NIST (MS/MS of 4903 m/z).

NIST HEAVY CHAIN (1-225)		
m/z	z	Fragment ion
575.52	2	C11(-8.5)
598.52	1	C5 (-17)
727.57	1	C6 (-17)
814.63	1	C7 (-17)
853.68	2	C16 (-8.5)
1039.80	1	C10 (-17)
1251.94	1	C12 (-17)

NIST LIGHT CHAIN		
m/z	z	Fragment ion
527.39	1	C4 (+22)
628.42	1	C5 (+22)
756.51	1	C6 (+22)
1027.70	1	C9 (+22)
1128.70	1	C10 (+22)
1399.94	1	C13 (+22)

Table 8: List of detected ETD fragment ions of the Fc subunit of IdeS digested NIST (MS/MS of 3920 m/z).

NIST HEAVY CHAIN (226-450)		
m/z	z	Fragment ion
380.30	1	C4 (+22)
527.39	1	C5 (+22)
640.49	1	C6 (+22)
1065.28	2	C19
1184.90	2	C11
1206.90	1	C11 (+22)
1313.04	1	C12
1335.01	1	C12 (+22)
1428.07	1	C13
1529.10	1	C14
1551.10	1	C14 (+22)
1886.38	1	C17

3.4.10 Top-down analysis after GingisKHAN digest of IgG

Native top-down ETD was performed on the GingisKHAN digest of the NIST IgG1. The sequence coverage is highlighted on the antibody sequence in **Figure 29** with the observed ETD fragment ions listed in **Tables 9-10**. The data was acquired for 1491 scans. The light chain sequence coverage was 13/213 (6.10 %) and the heavy chain sequence coverage was 58/450 (12.89 %).

NIST heavy chain:
QVTLRESGPALVKPTQTLTLTCTFSGFSLS**TAGMSVG**WIRQPPGKALEWLA**DIWWDDKKHYNPSLKD**RLTISKDTSKNQVVLKVTNMDPADTAT
YYCAR**DMIFNFYFDV**WGQGTTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTV
PSSSLGTQTYICNVNHKPSNTKVDKRV*EPKSCDK*
GingisKHAN cleaves
*THTCPPCP*APELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKE
YKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSR
WQQGNVFSCSVMHEALHNHYTQKSLSLSPGK
NIST light chain:
DIQMTQSPSTLSASVGDRVTITC**SASSRVGYMH**WYQQKPGKAPKLLIY**DTSKLASG**VPSRFSGSGSGTEFTLTISSLQPDDFATYYC**FQGSGYPFT**F
GGGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEV
THQGLSSPVTKSFNRGEC

Figure 29: Sequence of NIST IgG1 with the native ETD sequence coverage of GingisKHAN digested NIST marked in green(c-ions) and purple (z-ions).

Table 9: List of detected ETD fragment ions of the Fab subunit of GingisKHAN digested NIST (MS/MS of 3663 m/z).

NIST HEAVY CHAIN (1-225)		
m/z	z	Fragment ion
598.49	1	C5 (-17)
727.57	1	C6 (-17)
968.73	1	C9 (-17)
1039.77	1	C10 (-17)
1251.95	1	C12 (-17)
1476.99	1	C14 (-17)
1578.20	1	C15 (-17)
1706.18	1	C16 (-17)
1807.48	1	C17 (-17)
2020.90	1	C19 (-17)
2235.29	1	C20 (-17)
575.51	2	C11 (-17)
626.53	2	C12 (-17)
738.99	2	C14 (-17)
853.67	2	C16 (-17)
905.69	2	C17 (-17)
1067.39	2	C20 (-17)
1169.98	2	C22 (-17)
1220.86	2	C23 (-17)
1293.52	2	C24 (-17)
789.64	1	Z7
888.18	1	Z8
522.79	2	Z9
757.50	2	Z13
808.62	2	Z14
865.95	2	Z15

NIST LIGHT CHAIN		
m/z	z	Fragment ion
527.39	1	C4 (+22)
628.40	1	C5 (+22)
756.52	1	C6 (+22)
1027.70	1	C9 (+22)
1399.94	1	C13 (+22)

Table 10: List of detected ETD fragment ions of the Fc subunit of GingisKHAN digested NIST (MS/MS of 4101 m/z).

NIST HEAVY CHAIN (226-450)		
m/z	**z**	**Fragment ion**
1023.74	1	C10 (+1)
666.46	1	Z7 (-18)
772.60	1	Z8 (+1)
900.68	1	Z9 (+1)

3.4.11 Top-down analysis of 150kDa mAb (WP8)

The intact NIST antibody was also analysed with native top-down ETD and the sequence coverage is displayed in **Figure 29**.The observed ETD fragments are listed in **Table 11**. The acquisition time was 492 scans and the heavy chain sequence coverage was 38/450 (8.44 %).

NIST heavy chain:

QVTLRESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIRQPPGKALEWLADIWWDDKKHYNPSLKDRLTISKDTSKNQVVLKVTNMDPADTAT
YYCARDMIFNFYFDVWGQGTTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTV
PSSSLGTQTYICNVNHKPSNTKVDKRV*EPKSCDKTHTCPPCP*APELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVH
NAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWE
SNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK

NIST light chain:

DIQMTQSPSTLSASVGDRVTITCSASSRVGYMHWYQQKPGKAPKLLIYDTSKLASGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCFQGSGYPFTF
GGGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEV
THQGLSSPVTKSFNRGEC

Figure 29: Sequence of NIST IgG1 with the native ETD sequence coverage of intact NIST marked in green(c-ions) and purple (z-ions).

Table 11: List of detected ETD fragment ions of intact NIST (MS/MS of 6170 m/z).

NIST HEAVY CHAIN (1-450) 6170 m/z		
m/z	z	Fragment ion
484.77	2	C9 (-8.5)
520.46	2	C10 (-8.5)
576.19	2	C11 (-8.5)
598.52	1	C5 (-17)
684.80	1	Z7
727.57	1	C6 (-17)
739.06	2	C14 (-8.5)
772.60	1	Z8
789.63	2	C15 (-8.5)
814.64	1	C7 (-17)
853.66	2	C16 (-8.5)
899.58	1	Z9
904.22	2	C17 (-8.5)
960.77	2	C18 (-8.5)
1011	2	C19 (-8.5)
1027.68	1	Z10
1039.26	1	C10 (-17)
1067.22	2	C20 (-8.5)
1117.96	2	C21 (-8.5)
1128.76	1	Z11
1152.76	1	C11 (-17)
1220.71	2	C23
1251.95	1	C12 (-17)
1792.66	1	Z16
1856.67	2	C36 (-8.5)
1863.39	1	Z17

3.4.12 Native Top-down ETD summary

Table 12: Overview of the ETD sequence coverage for NIST IgG1, MSQ4 and Trastuzumab

		NIST intact	NIST +IdeS	NIST +GINGISKHAN
ETD coverage	HC	8.44 %	7.78 %	12.89 %
	LC		6.10 %	6.10 %
	mAb		13.88 %	18.99 %

		MSQ4 intact	MSQ4 +IdeS	MSQ4 +GINGISKHAN
ETD coverage	HC	3.11%	4.22 %	2.44 %
	LC			

		Trastuzumab intact	Trastuzumab +IdeS	Trastuzumab +GINGISKHAN
ETD coverage	HC	3.56 %	4.22 %	2.44 %
	LC			

3.4.12.1 Comparison with previous ETD data

Table 12 shows that the Native ETD sequence coverage was not optimal as it shows extremely low coverage. These ETD experiments were performed in begin 2017, but a similar native ETD experiment was performed in 2015 on an anti-actin antibody using the same instrument and settings showing better fragmentation performance. The sequence coverage of the light chain was 129/214 (60.28 %) and for the heavy chain 127/444 (28.60%) of the sequence was covered, resulting in a sequence coverage of 256/658 (38.90 %). Therefore, I expected a better sequence coverage when performing native ETD on the three standard antibodies of this project. Possibly the instrument settings have been changed during various instrument maintenances, and the ETD and MS/MS conditions were not adequately retuned for the ETD experiments.

Anti-Actin IgG Light chain

DIVLTQSPSSLSASLGDTITITCHASQNINVWLSWYQQKPGNIPKLLIYKASNLHTGVPSRFSGSGSGTGFTLTISSLQPEDIATYYCQQGQSYPLTFGG
GTKLEIKRADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSERQNGVLNSWTDQDSKDSTYSMSSTLTLTKDEYERHNSYTCEAT
HKTSTSPIVKSFNRNEC

Anti-Actin IgG Heavy chain

EVKLQESGGGLVQPGGSLKLSCATSGFTFSDYYMYWVRQTPEKRLEWVAYISNGGGSTYYPDTVKGRFTISRDNAKNTLYLQMSRLKSEDTAMYYC
ARHGGYYAMDYWGQGTTVTVSSAKTTAPSVYPLAPVCGDTTGSSVTLGCLVKGYFPEPVTLTWNSGSLSSGVHTFPAVLQSDLYTLSSSVTVTSST
WPSQSITCNVAHPASSTKVDKKIEPRGPTIKPCPPCKCPAPNLLGGPSVFIFPPKIKDVLMISLSPIVTCVVVDVSEDDPDVQISWFVNNVEVHTAQT
QTHREDYNSTLRVVSALPIQHQDWMSGKEFKCKVNNKDLPAPIERTISKPKGSVRAPQVYVLPPPEEEMTKKQVTLTCMVTDFMPEDIYVEWTNN
GKTELNYKNTEPVLDSDGSYFMYSKLRVEKKNWVERNSYSCSVVHEGLHNHHTTKSFSR

Figure 30: Sequence of anti-actin IgG with the native ETD sequence coverage marked in green (c-ions) and purple (z-ions).

Figure 31 shows the top-down ETD fragments previously obtained for the anti-actin antibody, modeled on a crystal structure of an IgG (PDB 1IGT). The surface-exposed Fab regions and the lower parts of the Fc region were mostly fragmented. When comparing these results with those of top-down denatured ETD of an antibody (**Fig. 32**), it can be shown more of the Fc domain could be fragmented and the heavy chain parts of the Fab-domains. By denaturing the protein these regions were more accessible for fragmentation. This is especially interesting when looking for the PTMs, which are located on the upper Fc domain.

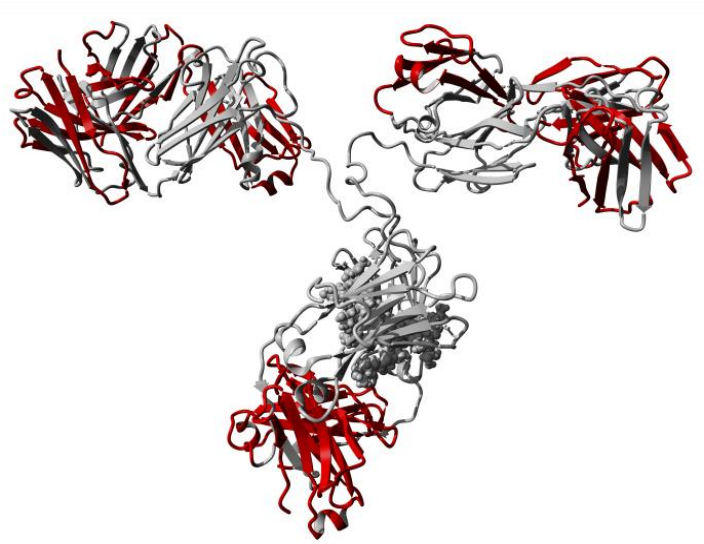


Figure 31: Top-down native ETD fragments (red) of anti-actin IgG modelled on IgG structure (PDB 1IGT). Data acquired on Synapt G2 HDMS

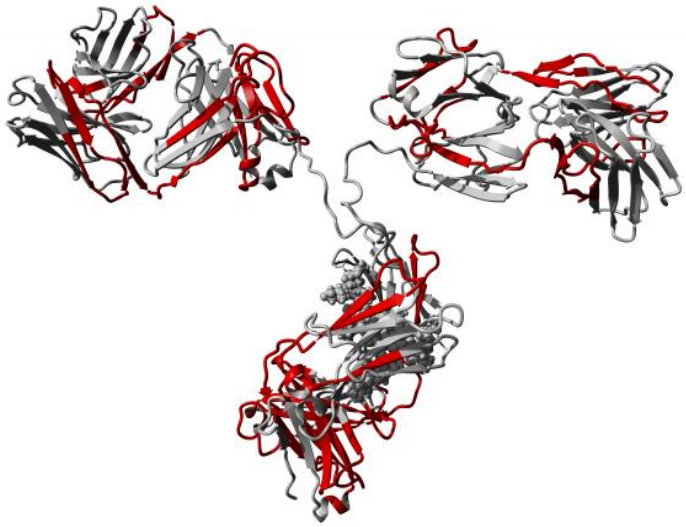


Figure 32: Top-down denatured ETD fragments (red) of a murine IgG modeled on IgG structure (PDB 1IGT). Data obtained from Tsybin *et al.* [3].

3.4.13 Conclusion

The antibodies MSQ4, NIST IgG1 and trastuzumab proved to be good standard antibodies to test these MS-based techniques.
With IM-MS the mass and collisions cross sections of the intact IgGs and IgG subunits were determined, providing consistent results for the three standard IgGs. This data can also be used as reference data sets when analysing intact antibodies or their subunits.
As IdeS cleaves the IgG under the hinge region, the generated Fc domain is also smaller in size and mass compared to the Fc generated by GingisKHAN.

Although the digested IgGs showed a slight increase in coverage, the general native top-down ETD experiments did not work out as planned. The instrument parameters need to be further optimised in order to solve this problem. It could also be that
the Synapt instrument was not sensitive enough to observe all the ETD fragments ions. However, previous ETD data showed that this method did work better, so there could have been a general problem with the instrument or a difference in settings which could not be resolved in order to obtain better sequence coverage. Native ETD could indicate the surface exposed regions of the antibody for an anti-actin IgG in a previous experiment. It could also be shown how the results of native ETD of IgG differ from those of denatured ETD, which shows more coverage of the inner or protected parts of the antibody.

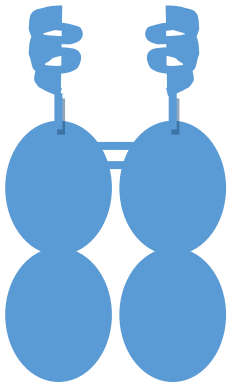
When the MS results of the 25 participating labs are summarised together, there will be also a broader data base based on different instruments and techniques, providing a complete overview of the relevant information this set of MS-based methods can offer.

3.4.14 Sources

[1] S.Maximiano, P. Magalhães, M.P. Guerreiro, M. Morgado, Trastuzumab in the Treatment of Breast Cancer, BioDrugs, 2016, 30, 75-86
[2] Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V., and Ruotolo, B. T. (2010) Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. Anal. Chem. 82, 9557–9565
[3] Structural Analysis of Intact Monoclonal Antibodies by Electron Transfer Dissociation Mass Spectrometry, Y. O. Tsbin, L. Fornelli, C. Stoermer, M. Luebeck, J. Parra, S. Nallet, F. M. Wurm, R. Hartmer, Analytical Chemistry, 2011, 83,8919-8927

Chapter 4

Characterisation of Fc-fusion proteins using an integrated structural proteomics approach



In collaboration with

Dr. Nick Bond[1]

Dr. Jonathan J. Phillips[2]

Dr. Christel Veyssier[3]

[1]Analytical Biotechnology, MedImmune, Cambridge, UK

[2]Living systems institute, University of Exeter, Exeter, UK

[3]Babraham research campus, F-star, Cambridge, UK

4.1 Introduction

Over the past three decades antibodies and antibody-derived products have become an important class of therapeutic agents due to their high specificity and stability. There have been 60 antibody drugs approved globally, covering a range of diseases, with sales of over 80 billion US dollars in 2015 [1]. Antibodies have shown results in the treatment of several conditions including autoimmune diseases such as arthritis and different forms of cancer. As protein-engineering is developing at a high rate, more sophisticated, engineered antibodies are produced with improved pharmacokinetics and efficiency.

One group of antibody derivatives that is growing in importance are Fc-fusion proteins. Fc-fusion proteins consist of an Fc domain of an immunoglobin that is linked to a peptide or protein of interest. This class of proteins is of biotherapeutic interest, because the presence of the Fc-domain increases the plasma half-life due to their interaction with Fc-receptors, and due to their slow renal clearance [3]. This means that the Fc-domain can improve the stability and the solubility of the bound peptide or protein.

Two glucagon-like peptide 1 (GLP-1) Fc-fusion proteins were analysed in this study; IP268 and IP118 (**Figure 1**). IP268 is a GLP-1 linked to the Fc-domain of human IgG1, and predominantly forms dimers. Dulaglutide (IP118) is a GLP-1 covalently linked to the Fc-domain of human IgG4, and does not form dimers. Dulaglutide is currently being used for treatment of type 2 diabetes [4]. The aim of this study is to optimise and apply a set of MS- and structural proteomics methods to structurally characterise these Fc-fusion proteins. Comparing the two constructs enhances our understanding of structural properties of fusion proteins, and specifically of the dimerisation interface of IP268.

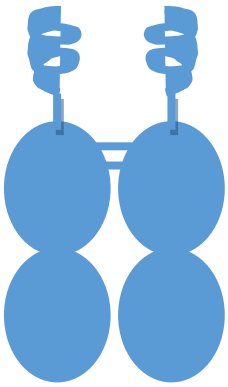


Figure 1: A schematic illustration of the studied Fc-Fusion protein monomer. The Fc-construct consists of an Fc-domain and two flexible linkers each attached to a GLP-1 peptide.

4.2 Materials and methods

Materials

The Fc-fusion proteins, IP268 and IP118, were donated by MedImmune (Cambridge, UK). The K100 stabilization kit was purchased from CovalX and ammonium acetate, urea and tris(2-carboxyethyl)phosphine (TCEP) were from Sigma Aldrich.

Native ion mobility-mass spectrometry (IM-MS)

A Synapt G2 HDMS Q-TOF instrument (Waters, UK) was used for the native IM-MS experiments. The samples were buffer exchanged into 250 mM ammonium acetate using Micro Bio-spin P6 columns (Bio-Rad) at final protein concentrations of ca. 5µM. 5µL aliquots were transferred into the mass spectrometer using gold-coated nanoESI needles (prepared in-house). The instrument was tuned to preserve native higher-order structure using the following parameters: capillary voltage 1.2 kV; nanoflow backing gas pressure 0.2 bar; sampling cone 30 V; extraction cone 1.0 V; trap collision energy 10 V; transfer collision energy 0 V; trap bias 45.0 V; IMS gas flow 90.0 ml/min; IMS wave velocity 700 m/s; IMS wave height 40.0 V; backing pressure 3.0 mbar.

The data were acquired and processed with Masslynx v4.1 software, and ion mobility drift times extracted using Driftscope v2.3 (both Waters). The collision cross sections (CCS) of the proteins were calibrated using known CCS values determined under native conditions as described previously [5].

Analytical ultracentrifugation (AUC)

Analytical ultracentrifugation (AUC) - Sedimentation velocity experiments were performed at the biophysics facility at the University of Cambridge. The experiments were conducted with an Optima XL-I (Beckman Coulter) centrifuge using an An60 Ti four-hole rotor. Standard double-sector Epon centrepieces equipped with sapphire windows contained 400 µL of IP268 at concentrations of 0.126 to 1.14 mg/mL. Interference data were acquired in the continuous mode at time intervals of ~200 s and rotor speed of 40,000 rpm, at a temperature of 20°C with systematic noise subtracted, without averaging and with radial increments of 0.003 cm. Absorbance data were also collected, at a wavelength of 230 nm. The density and viscosity of the buffer and the partial specific volume of the protein were calculated using Sednterp[6]. Multi-component sedimentation coefficient distributions were obtained from 100 scans by direct boundary modelling of the Lamm equation using Sedfit v.14.1 [7].

Size exclusion chromatography – multi –angle light scattering (SEC-MALS)

The SEC-MALS experiments were performed using a Zenix SEC-300 column coupled to an Agilent Diode-Array UV-detector, an Optilab T-reX refractive-index detector and a DAWN HELEOS quasi elastic light scattering detector. The latter allows measurement of the hydrodynamic radius in a Dynamic Light Scattering mode. 5 µL of 11.4 mg/mL protein stock solution was used for the molecular weight analysis and 10 µL of the same stock was used to determine the hydrodynamic radius. The data were acquired and processed with ASTRA (Wyatt).

Cross-linking

The cross-linking experiments were performed using a K100 CovalX kit. The stabiliser solution contains the cross-linkers 1,1'-(suberoyldioxy)bisazabenzotriazole (SBAT), 1,1'-(suberoyldioxy) bisbenzotriazole (SBBT,) and 1,1'-(glutaroyldioxy)bisazabenzotriazole (GBAT). SBAT and SBBT have a spacer arm length of 11.4 Å and GBAT has a length of 7.7 Å.
Protein concentrations of 0.14 to 4.56 mg/mL were prepared for the analysis. A final cross-linker concentration of 0.2 mg/mL was added to the protein samples. The non-covalent complexes were stabilised by the cross-linking agents and then directly analysed by MALDI mass spectrometry (Waters) with a CovalX HM3 Tuvo high-mass system. For each cross-linking experiment a control was made without reagents and analysed in parallel with MALDI MS.

Hydrogen deuterium exchange- mass spectrometry (HDX-MS)

The HDX-MS measurements were performed on an ACQUITY UPLC M-Class system with HDX technology (Waters) coupled to a Synapt G2 (Waters). The experiments were performed at MedImmune, UK.

Protein stock concentrations in the range of 0.5 mg/mL to 45 mg/mL were used to compare the predominantly dimeric state to the majority monomeric state of IP268. The samples were diluted 10 times with either 250 mM ammonium acetate buffer in 100% H_2O, pH 7.4, for the undeuterated samples, or 250 mM ammonium acetate in 99.9% D_2O, pD 7.4 for the deuterated samples. The final protein concentrations during the HDX experiments ranged from 0.05 mg/mL to 4.5 mg/mL. The samples were incubated at 20°C for 1 ,2, 5, 10 and 60 minutes after labeling, then quenched by reducing the pH with 250 mM ammonium acetate, 8 M urea and 0.5 M tris(2-carboxyethyl)phosphine (TCEP) at 0°C. The samples were injected into the ACQUITY UPLC system, where the digestion was performed at 0.5°C in the HDX manager using a pepsin column (2.0 x 30 mm). The peptides were separated on an ACQUITY UPLC BEH C18 column (Waters) at a flow rate of 40 µL/min with a 3% - 40% gradient of acetonitrile in water. The following MS conditions were used; ESI capillary voltage 3.0 kV, source temperature 80 °C, sampling cone 30 V, extraction cone 4.0 V, trap collision energy 4.0 V, backing pressure 2.08 mbar. The HDX data was processed using DynamX 3.0 (Waters)

4.3 Results and discussion

4.3.1 Determining mass and oligomeric state of Fc-fusion proteins with native MS

Native mass spectra of the Fc-fusion proteins, IP268 and IP118 were acquired at a 0.32 mg/mL concentration (**Figures 2 & 3**). For IP268 the experimental mass of the monomer was found to be 63,434.43 Da with charge states 12^+ to 14^+ (**Figure 2, triangles**). Based on the amino acid sequence, the theoretical mass of IP268 is 60,255.34 Da. The 3 kDa difference between the theoretical and experimental mass is due to post translational modifications such as N-glycosylations. The IP268 dimer was also detected with charge states 18^+ to 22^+ and an experimental mass of 126,603.91 Da (circles). In contrast IP118 only showed monomer (**Figure 3**). The IP118 monomer was observed with charge states 12^+ to 14^+ and an experimental mass of 62,603.05 Da. The theoretical mass of IP118 based on its amino acid sequence is 59,682.50 Da. The 3 kDa difference in mass is due to post-translation modifications (e.g. N-glycosylation).

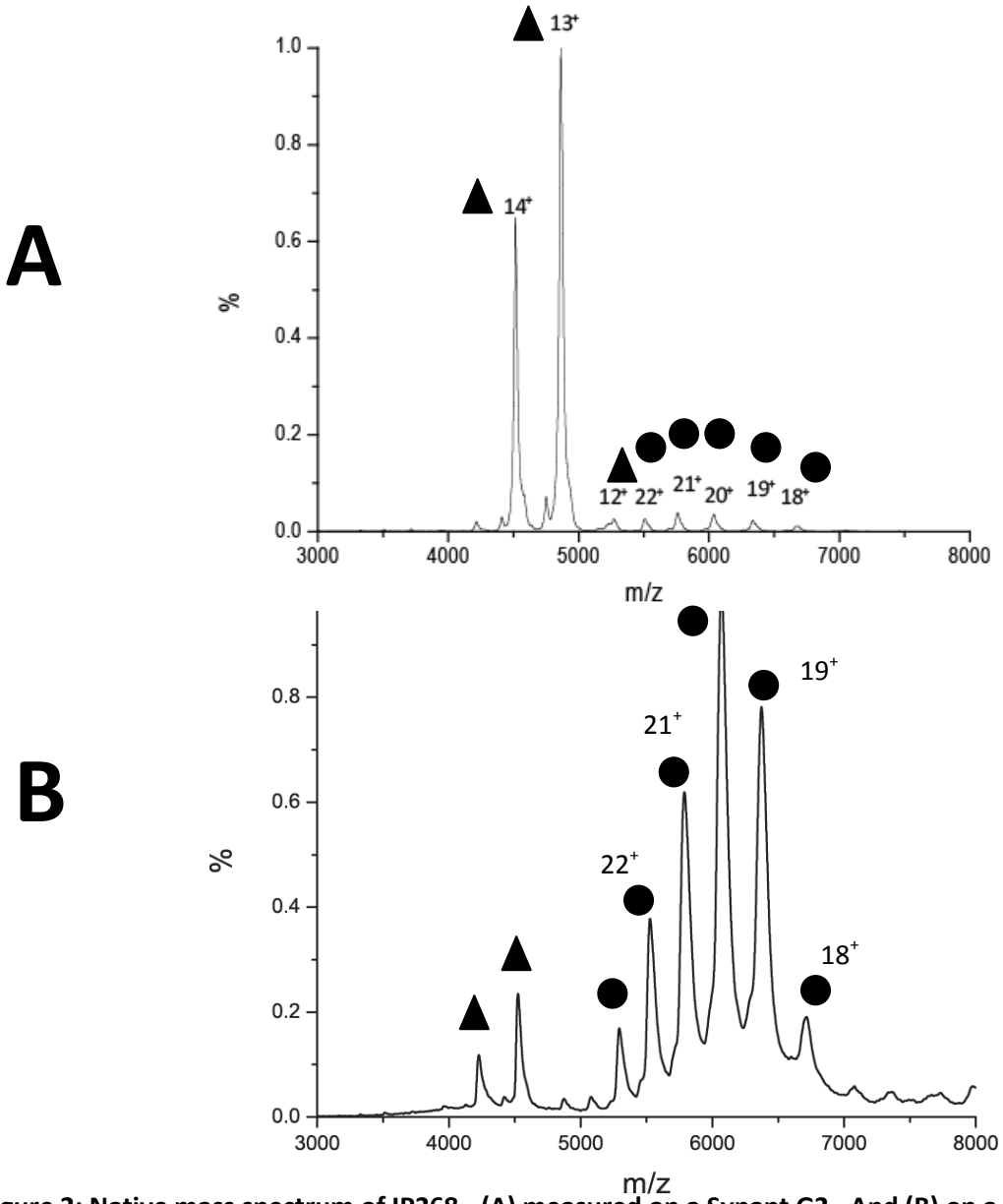


Figure 2: Native mass spectrum of IP268., (A) measured on a Synapt G2. And (B) on a Q-TOF modified for high mass. The monomer IP268 (▲) with charge states 12^+ to 14^+. The IP268 dimer (●) with charge states 18^+ to 22^+.

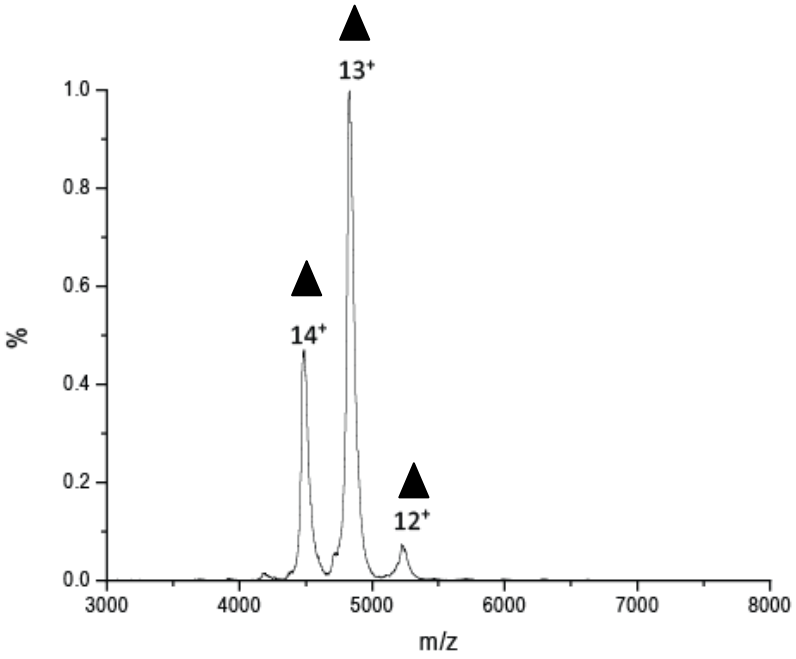


Figure 3: Native mass spectrum of IP118 monomer.

4.3.2 Determining collision cross section of Fc-fusion proteins with native IM-MS

Using native IM-MS, collision cross sections were obtained after calibration for each charge state of the Fc-fusion proteins (**Table 1**). The CCS values of the monomeric species range from 3716 to 3940 $Å^2$ for IP268 and from 3687 to 3899 $Å^2$ for IP118. The IP268 dimer shows CCS values from 6130 to 6796 $Å^2$. The dimerisation of IP268 showed an increase of 65% in CCS value when compared to the monomer (lowest charge states). This could mean that the overlapping surfaces of the two subunits are locked in such a way to form a compact ensemble.

Table 1: CCS values in $Å^2$ of IP268 monomer and dimer and IP118 monomer

	Charge state	CCS ($Å^2$)
IP268 monomer	12^+	3716 ± 26
	13^+	3824 ± 20
	14^+	3940 ± 24
IP268 dimer	19^+	6130 ± 55
	20^+	6289 ± 51
	21^+	6502 ± 67
	22^+	6645 ± 64
	23^+	6796 ± 69
IP118 monomer	12^+	3687 ± 56
	13^+	3807 ± 55
	14^+	3899 ± 75

4.3.3 Measuring molecular weight and radius with SEC-MALS

Next to the MS experiments, the Fc-fusion proteins were also studied in solution phase. IP268, IP118 and the calibrant, BSA were analysed with SEC-MALS to measure the approximate molecular weight and the hydrodynamic radius of the proteins. The elution times are shown in **Figure 4**. The BSA monomer eluted at 9.9 minutes and the dimer showed a small elution peak at 8.7 minutes. The IP118 monomer eluted at 9.7 minutes and no dimer was observed. IP268 is only observed in dimeric form at 8.8 minutes. This confirmed the oligomeric states of the Fc-fusion proteins. The difference observed in the MS experiments was that IP268 was present in both monomeric and dimeric state. In the MS experiments a sample concentration of 0.32 mg/mL, whereas the final concentration in the SEC-MALS experiments was 1.14 mg/mL. It could also be possible that the IP268 dissociated during the ESI process, despite the soft MS conditions, causing the detection of both monomeric and dimeric IP268 in the MS spectra.

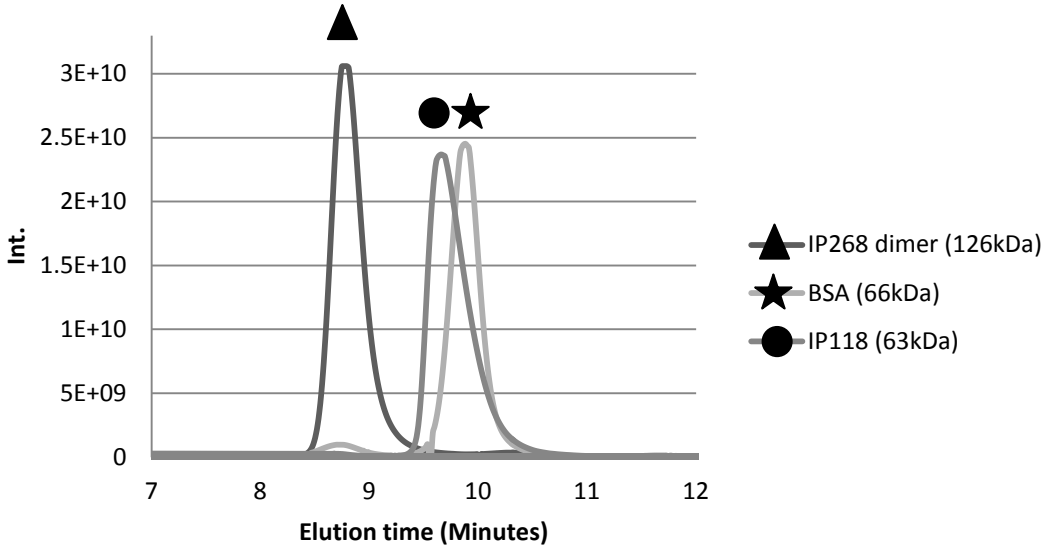


Figure 4: SEC-MALS elution time of IP268 dimer (blue), BSA (red) and IP118 (green). The BSA monomer eluted at 9.9 minutes and the BSA dimer showed a small elution peak at 8.7 minutes. The IP118 monomer eluted at 9.7 minutes and no Ip118 dimer was observed. IP268 is only observed in dimeric form at 8.8 minutes.

In **Table 2** the experimental molecular weights are shown of the Fc-peptide fusions determined by SEC-MALS. For the measurements marked high rep, a higher amount of protein was injected to allow sufficient signal fitting for the Rh data algorithm. IP118 was detected with a mass of 66828 to 69205 Da and hydrodynamic radius of 3.499 to 3.691 nm. IP268 was observed with a mass of 118478 to 117231 Da and a hydrodynamic radius of 5.007 to 5.064 nm.

Table 2: List of experimental molecular weight (rep 1, 2) and hydrodynamic radius (high rep1, 2) of the Fc peptide fusion determined by SEC-MALS. Rep1 and Rep2 are measurements with 60μg of protein injected for the determination of mass. IP118 was detected with a mass of 66828 to 69205 Da and IP268 with a mass of 118478 to 117231 Da. The high rep 1 and 2 are measurements with 120 μg of protein injected for the determination of hydrodynamic radius. IP118 was detected with a hydrodynamic radius of 3.499 to 3.691 nm and IP268 with a hydrodynamic radius of 5.007 to 5.064 nm.

Protein sample	Mw (kDa)	Rh(z) (nm)	Peak limits (min)
IP118 rep 1	66.828 ±0.386		9.605 - 10.182
IP118 rep 2	69.205 ±0.364		9.432 - 9.920
IP268 rep 1	118.478 ±0.529		8.644 - 8.969
IP268 rep 2	117.231 ±0.495		8.577 - 8.969
IP118 high rep 1		3.499 ±0.162	9.328 - 9.867
IP118 high rep 2		3.691 ±0.164	9.336 - 9.829
IP268 high rep 1		5.064 ±0.119	8.540 - 8.994
IP268 high rep 2]		5.007 ±0.123	8.540 - 8.994

The detected masses are in the same range as the masses detected with mass spectrometry. The hydrodynamic radius was also compared to the average CCS values determined with ion mobility. In **Table 3** the average CCS determined by ion mobility is shown and the average radius derived from this cross section value using the surface are formula of a circle, $\sqrt{((A)/\pi)}= r$. The average CCS radius of the IP268 dimer was 4.5 nm (11% smaller than the average hydrodynamic radius and the CCS radius of the IP118 monomer 3.5 nm (3% smaller than the average hydrodynamic radius)

Table 3: Comparison of average CCS determined by IM-MS and average hydrodynamic radius (Rh (z)) determined by SEC-MALS. The average CCS radius was derived from the IM-MS CCS value. The IP268 dimer CCS radius is 11% smaller than the hydrodynamic radius. The IP118 monomer CCS radius is 3% smaller than the average the hydrodynamic radius.

	Average CCS ($Å^2$)	Average CCS Radius (nm)	Average Rh(z) (nm)
IP268 dimer	6472	4.539	5.035
IP118 monomer	3798	3.477	3.595

4.3.4 Comparing CCS and collision induced unfolding plots of Fc-fusion proteins and Fc domains

IM-MS is a useful tool to detect structural changes when modifications are made to the sample. To compare the Fc-fusion protein, IP268, to its respective Fc domain, the full antibody, NIP109, was treated with LysC to separate the Fab and Fc domains. The native mass spectrum in **Figure 5** shows that the antibody digested into its respective Fab and Fc domains. The Fc domain has an experimental weight of 53639 Da with charge states 12^+ to13^+. After treatment with LysC there was also no dimer detected of the Fc-domain. This could indicate that the dimerization of the IP268 is triggered from the peptide-bound domain of the Fc-fusion protein.

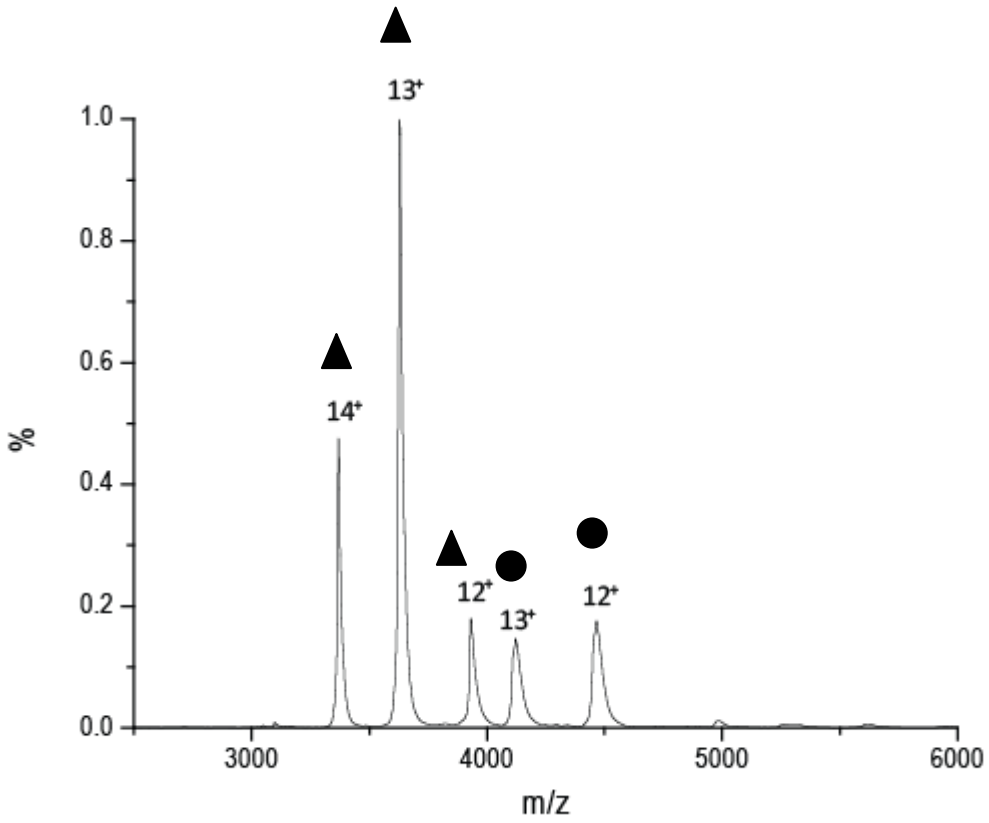


Figure 5: Native mass spectrum of antibody NIP109 treated with LysC. The Fab region (▲) has an experimental weight of 47206 Da with charge states 12^+ to 14^+. The Fc region (●) has an experimental weight of 53639 Da with charge states 12^+ to 13^+.

The full width at half maximum of the peaks 12^+ and 13^+ of the Fc-domain was selected to extract the ion mobility drift times. The same extraction was done for the 12^+ to 14^+ charge states of IP268. The drift times were converted to CCS values and are displayed in **Figure 6.** The collision cross section of the Fc domain of IP268 ranged from 3464 $Å^2$ to 3583 $Å^2$, while IP268 ranged from 3716 $Å^2$ to 3940 $Å^2$ (**Figure 6-a**). The fusion of the peptide to the Fc domain caused a 7% increase of the CCS value. A similar trend was observed for IP118 where the fusion of the peptide to the Fc domain caused an increase of 7-10%. For the Fc-domain of IP118 the CCS ranged from 3361 $Å^2$ to 3543 $Å^2$, while IP118 ranged from 3687 $Å^2$ to 4038 $Å^2$ (**Figure 6-b**). This increase could be explained by assuming that the two fused peptides are extended away from the Fc-domain, rather than adopting a more compact conformation with the peptides in proximity to the core.

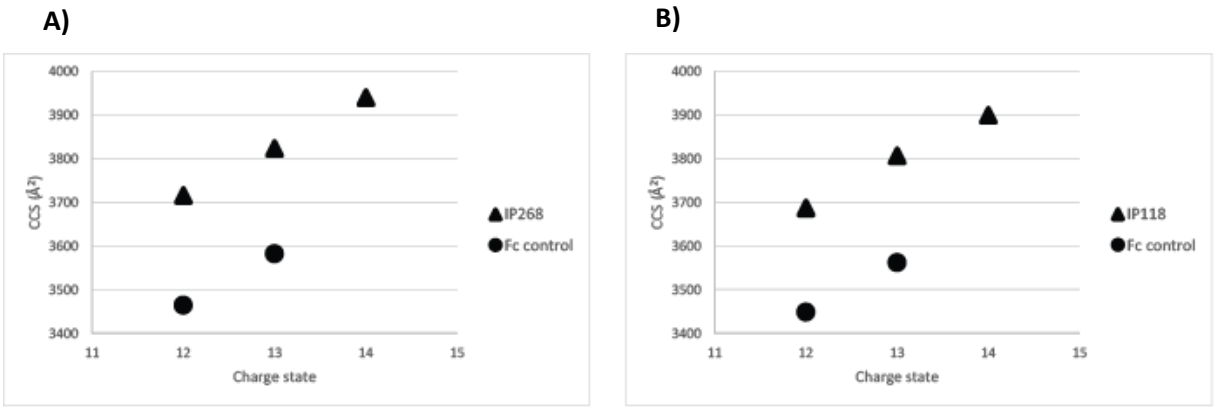


Figure 6: A) Collision cross sections of IP268 ranging from 3716 $Å^2$ to 3940 $Å^2$ and as a control the Fc-domain of NIP109 ranging from 3464 $Å^2$ to 3583 $Å^2$.
B) Collision cross sections of IP118 ranging from 3687 $Å^2$ to 3899 $Å^2$ and as a control the Fc-domain of IP118 ranging from 3448 $Å^2$ to 3561 $Å^2$.

By increasing the collision energy in the trap cell while recording the ion mobility drift time of a certain charge state (CIU, see methods chapter), the unfolding of the protein can be monitored. This experiment was performed for the Fc fusion proteins and their respective Fc-domain. The ion mobility drift times were converted to CCS values and plotted against the trap collision energy. **Figures 7** and **8** show the collision induced unfolding plots of the Fc-fusion proteins and corresponding Fc-domains. When comparing the Fc-domain of IP268 (**Figure 7-a**) to the full Fc-fusion protein IP268 (**Figure 7-b**), the first transition of the Fc domain at 60 V is at the same CCS value as the initial CCS value of IP268 from 10 to 70 V. This could indicate that the linking of the peptide to the Fc-domain partially unfolds the latter to this transition state. The same trend is seen for IP118 in **Figure 8 a-b** where the first transition state at 70V has the same CCS value as initial CCS value of IP118 from 10 V to 70 V. The Fc- fusion proteins are also showing more stability as they can tolerate higher voltages compared to their Fc-domains alone. This indicates that the linking of the peptides to the Fc-domain increases the size of the protein and increases its stability.

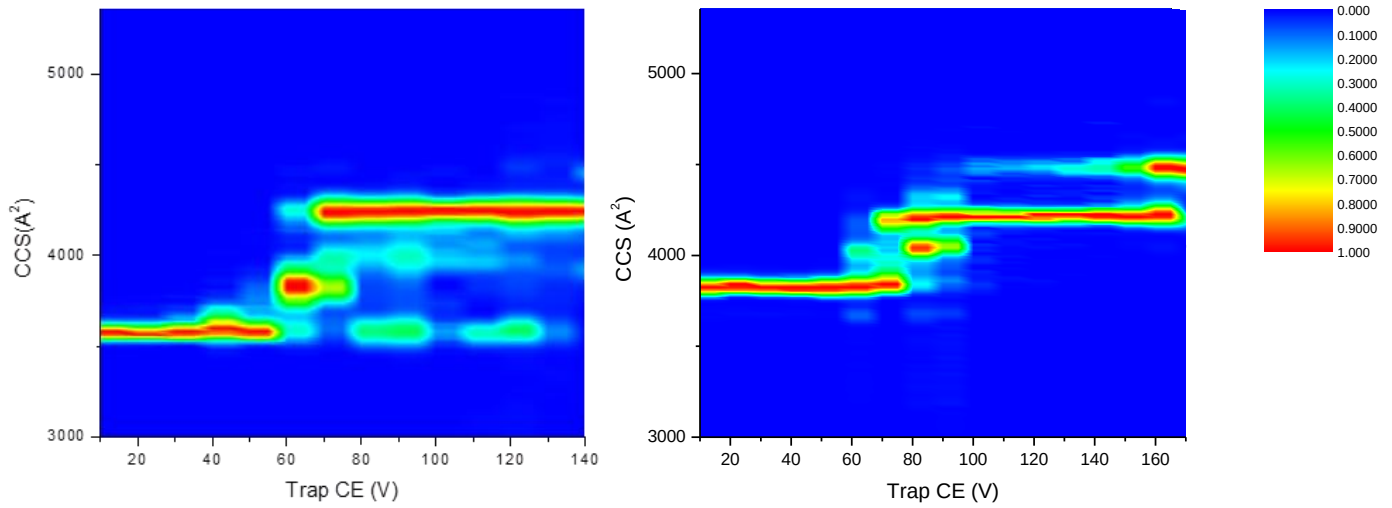


Figure 7: a) Collision induced unfolding plot of Fc-domain of NIP109 (13^{+}). b) Collision induced unfolding plot of Fc-peptide fusion, IP268 (13^{+}).

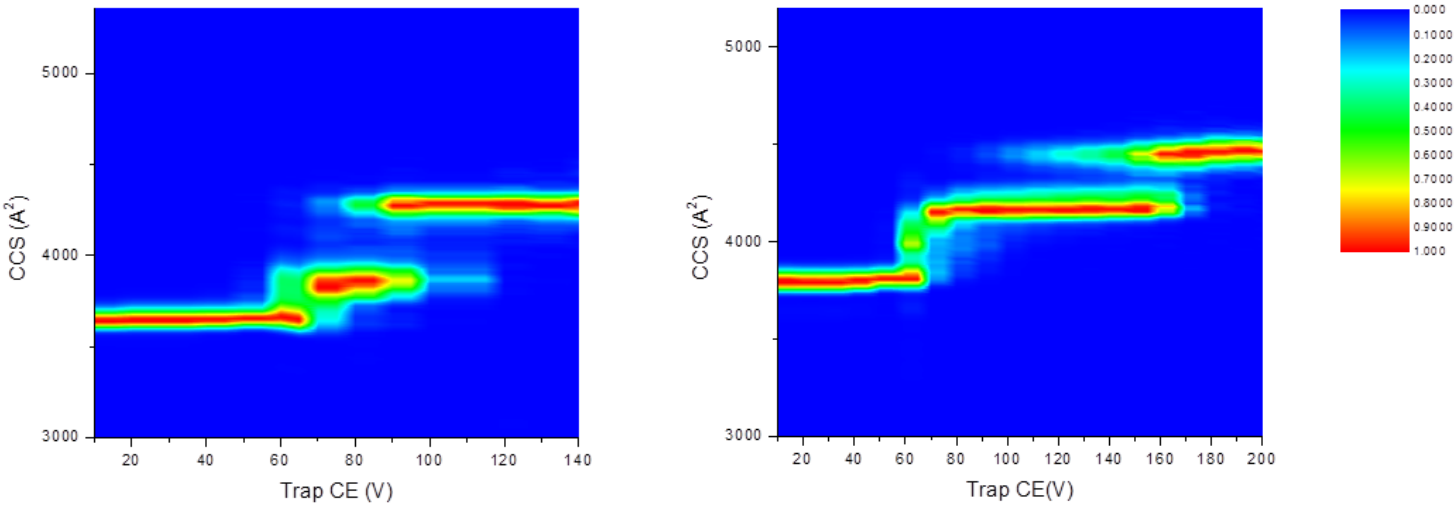
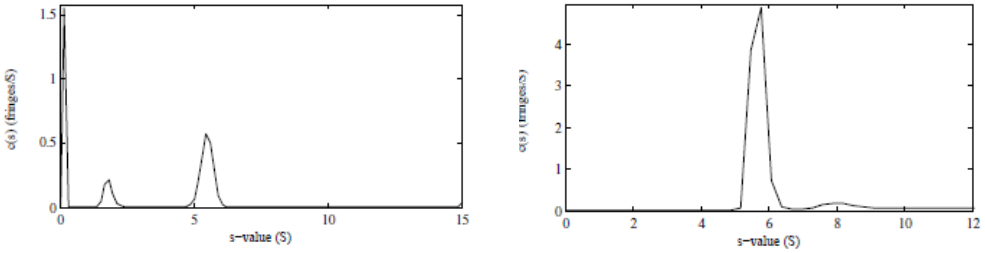


Figure 8: a) Collision induced unfolding plot of Fc-domain of IP118 (13^{+}). b) Collision induced unfolding plot of Fc-peptide fusion, IP118 (13^{+}).

4.3.5 Estimating the K_d with AUC

We attempted to determine the dissociation constant, K_d, of the IP268 dimer with analytical ultracentrifugation. We performed this experiment at final concentrations of 0.126 mg/mL and 1.14 mg/mL to find out at which concentration we would only see monomer or only dimer. In **Figure 9** we can see the detected species with their respective sedimentation values, assuming a uniform frictional ratio of $F_{k,w}$ = 1.46. The final root-mean-square deviation (r.m.s.d.) was 0.003 for interference and 0.007 for UV absorbance. For the higher concentration of 1.14 mg/mL only one species is observed with a sedimentation value of 6 S, corresponding with the IP268 dimer. However, for the low concentration of 0.126 mg/mL we still saw a mix of two species, corresponding to the monomer and dimer of IP268. Unfortunately, we could not lower the concentration sufficiently to only see the IP268 monomer.



Monomer +Dimer 0.126 mg/mL

Dimer 1.14 mg/mL

Figure 9: AUC- Sedimentation velocity results of IP268 at a final concentration of 0.126 mg/mL and at 1.14 mg/mL

To estimate the K_d based on the AUC measurements the molar concentrations of monomer and dimer need to be determined. The software SedFit can calculate the loading concentrations (i.e. peak area of every peak) for the observed peaks. For the IP268 dimer (126 kDa) this value was 0.316647, for the IP268 monomer (63 kDa) 0.025188 and for the single Fc chain of IP268 (31 kDa) it was 0.097218. The total of these SedFit units is 0.439053. This total equals the concentration that was loaded in the instrument multiplied by a constant (C). By using the absorbance (A), the extinction coefficient (ε) and the length of the cell (l), the concentration in the cell can be calculated using the formula of Lambert-Beer; c = A/(ε.L)= 0.916 mg/mL (**Eq. 1**). This makes the constant C equal to 0.439053 divided by 0.0916 (**Eq. 2**). The molar concentration for the monomer (monomer IP268 + Fc chain) and the dimer IP268 can be calculated using the loading concentration and the constant C (**Eq.3**). The dissociation constant (K_d) can then be calculated as in **Eq.4**, resulting in a K_D of 0.00987 mg/mL or 156 nM.

Eq. 1 **: c (IP268) = A/(ε.L) = 1.61/(1.47*1.2) = 0.91 mg/mL**

Eq. 2 **: Total loading conc. = 0.316647+0.025188+0.0972118=0.43905 = C x 0.0916 mg/mL**

→ C = 0.439053/0.0916 = 4.79 mg/mL

Eq. 3 **: [M] = 0.122406 x C & [D] = 0.316647 x C**

Eq. 4 **: K_d= $[M]^2/[D] = (0.0255)^2/ (0.0661)$= 0.00987 mg/mL**

K_d (Fc fusion) =(0.00987 mg/mL) /(63 kg/mol)= 156 nM

4.3.6 Determining K_d with covalX cross-linking

With CovalX cross-linking the oligomeric state of the Fc-fusion protein was investigated. By adding the cross-linking agents to a range of different protein concentrations, the relative intensity between monomer and dimer in the MALDI mass spectra was monitored. In **Figure 10-a** the spectra are shown of IP268 without cross-linker. With increasing concentration the relative intensity of IP268 also slightly increases. However this similar trend seems more like an artifact or ''plume effect'' (i.e., a series of peaks that attenuate with increasing mass) cause by the MALDI ionisation. When adding the cross-linker agents to the sample, the relative intensity of IP268 dimer increases in a more drastic manner with increasing protein concentration (**Figure 10-b**). The relative intensities of the IP268 dimer compared to the IP268 monomer are displayed in **Figure 11**. The cross-linker had a stabilising effect on the IP268 dimer and the relative intensity of the IP268 dimer increased with increasing protein concentration.

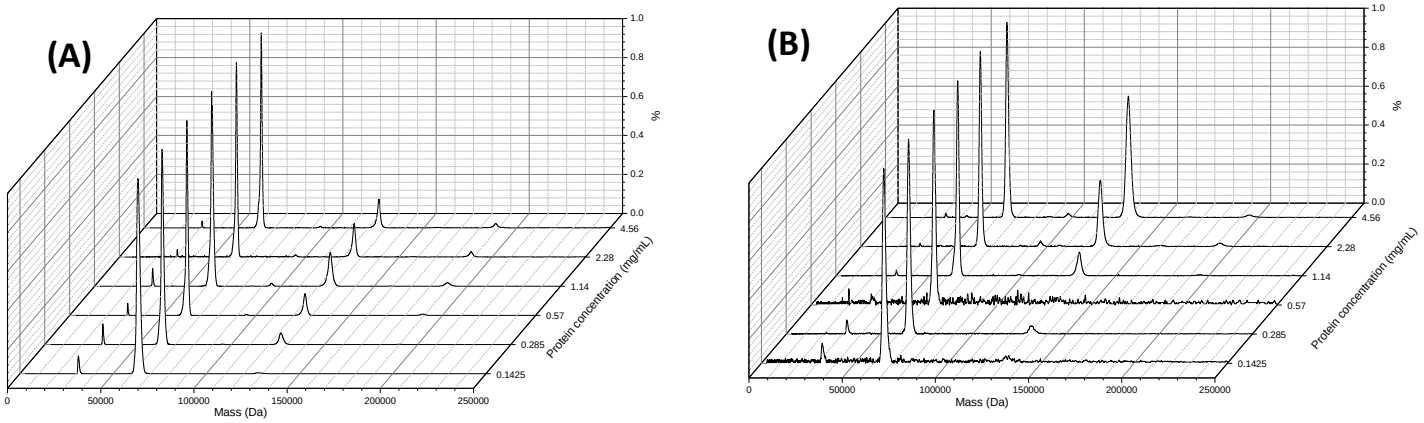


Figure 10: A) MALDI spectra of IP268 without cross-linker. B) MALDI spectra of IP268 with cross-linker.

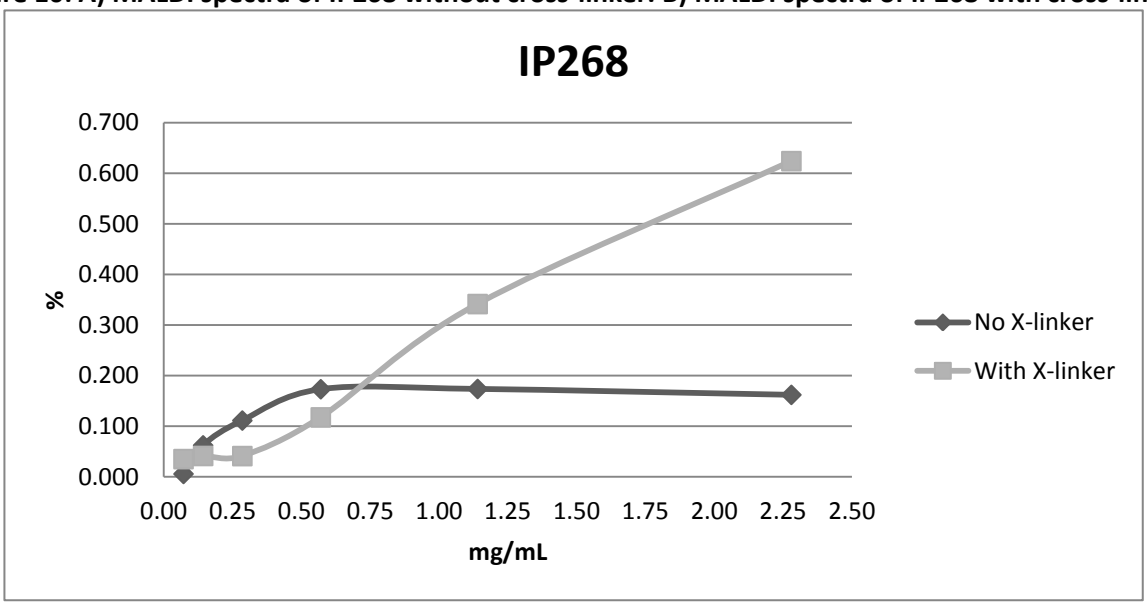


Figure 11: Relative intensity of IP268 dimer compared to monomer with cross-linker and without cross-linker.

As a control the same cross-linking experiment was performed with IP118. In **Figure 12-a** the spectra are shown of IP118 without cross-linker and shows the artifact again of dimers and trimers at increasing concentrations. In **Figure 12-b** the spectra are shown of IP118 with cross-linker and the relative intensity of the dimer is under 30% even at increasing protein concentrations. This confirms that the cross-linkers stabilized the monomer in this experiment and also indicates that the increasing IP268 dimer in **Figure 12-b** is not an artifact but due to the stabilization of the cross-linking agents. **Figure 13** shows the relative intensities of the detected IP118 dimer compared to IP118 monomer with and without cross-linker at increasing concentrations.

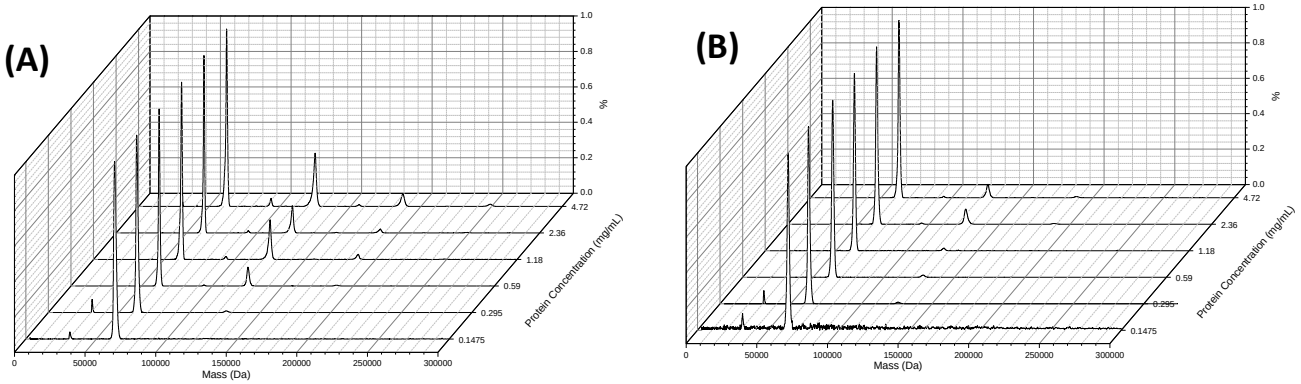


Figure 12: a) MALDI spectra of IP118 without cross-linker. b) MALDI spectra of IP118 with cross-linker.

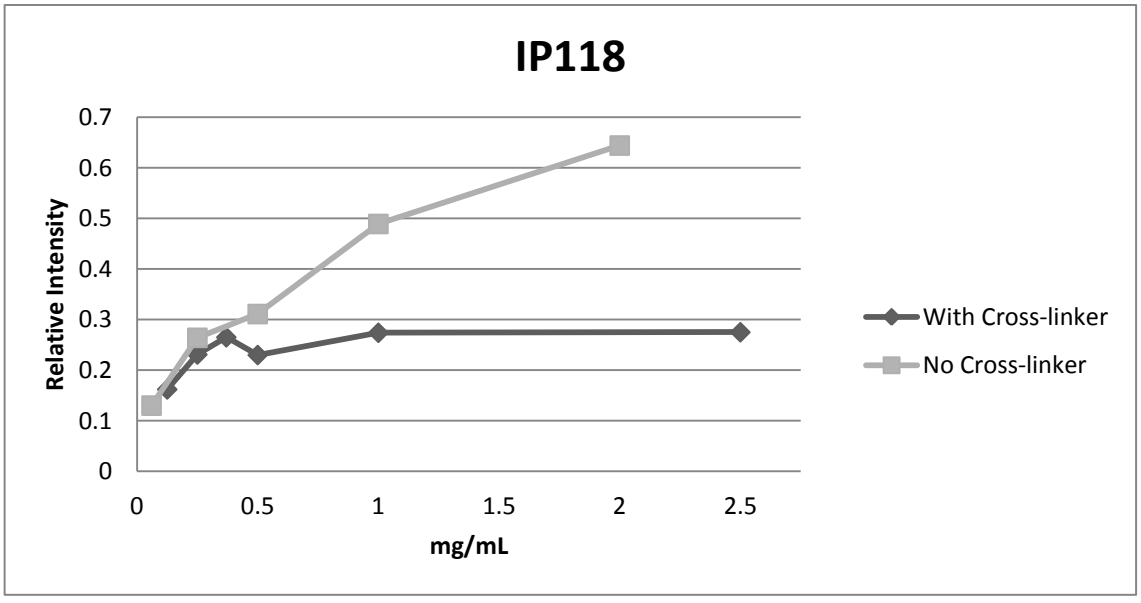


Figure 13: Relative intensity of IP118 dimer compared to monomer with and without cross-linker.

The cross-linking experiment of IP268 was repeated with more concentration intervals and the relative intensity of dimer compares to monomer is displayed in **Figure 14**. In this figure the concentrations used for each experiment are also indicated on the graph. This gives an indication at which concentration there would be a majority of the monomeric or dimeric form of IP268. To derive the K_D from this graph is difficult as even at low concentration of 0.1 mg/mL there is still both monomer and dimer present. **Table 4** displays the initial and final concentration used for the different techniques in this study.

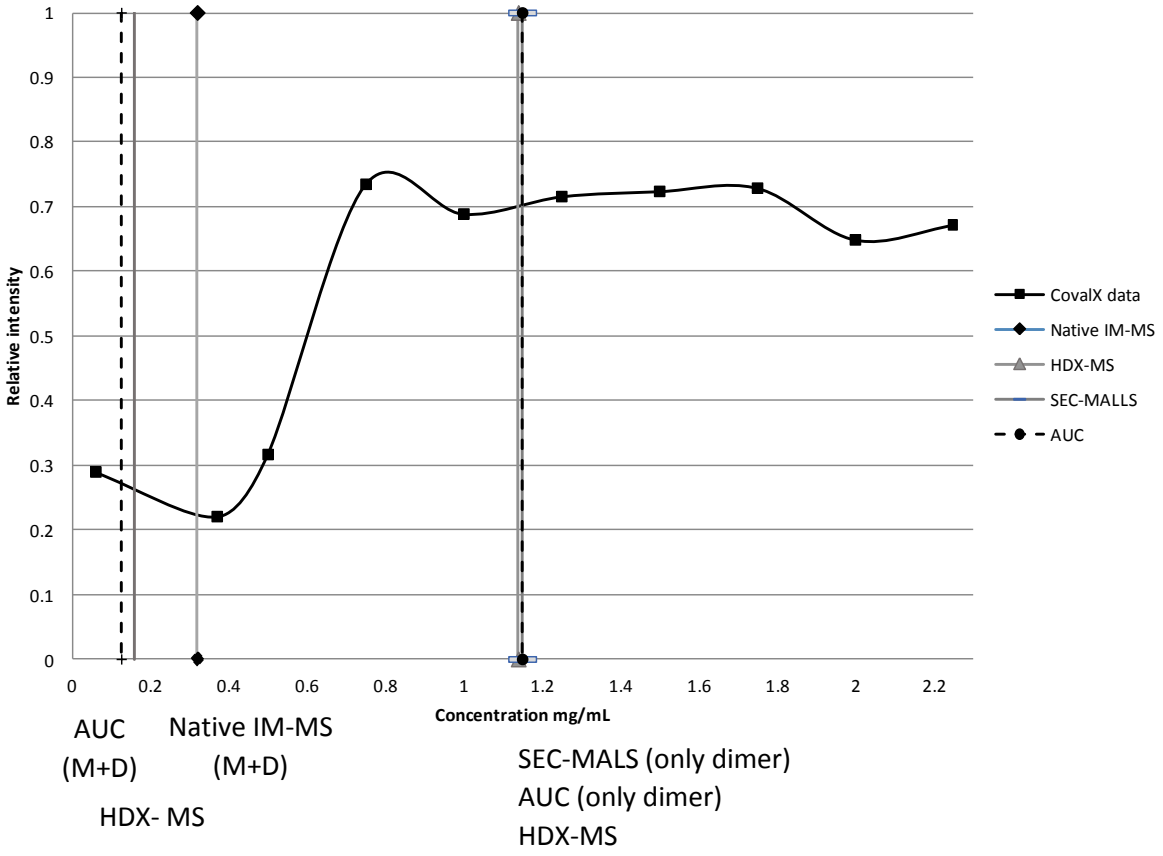


Figure 14: Overview of the collected data and the sample concentration used during the analysis of IP268.

Method	Initial concentration	Final concentration	Fc-Fusion Dimer (4xHC)	Fc-Fusion Monomer (2xHC)
Native IM-MS	0.32 mg/mL	0.32 mg/mL	2.5 uM	5 uM
HDX-MS	11.4 mg/mL 1.6 mg/mL	1.14 mg/mL 0.16 mg/mL	9 uM 1.25 uM	18 uM 2.5 uM
SEC-MALS	11.4 mg/mL	1.14 mg/mL	9 uM	18 uM
CovalX	0.07 mg/mL- 2.28 mg/mL	0.07 mg/mL- 2.28 mg/mL	0.56 uM- 18 uM	1.1uM- 36uM

Table 4: Overview of the collected data and the sample concentration used during the analysis of IP268.

4.3.7 Locating the dimerisation interface with HDX-MS

A series of IP268 concentrations (0.05 mg/mL, 0.1 mg/mL, 1 mg/mL/min, 4.5 mg/mL) was analysed by HDX-MS to locate the site of dimerisation. Based on the MS experiments of IP268 with and without peptide, it is assumed that the dimerisation happens in the region of the linked peptide. Therefore the residues 1-53 were analysed and compared when different concentrations of IP268 were used in the experiment.

The HDX experiments were run with a range of exposure times (i.e. 10 sec, 20 sec, 30 sec, 1 min, 5 min, 10 min and 60 min). An example of the deuterium uptake of a peptide (residue 1-9) of IP268 over the different exposure times is displayed in **Figure 15**. After less than 10 seconds of exposure time, there is an increase in mass, indicating that the peptide is becoming deuterated. This means that the corresponding region of the protein is surface exposed as it is easily accessible to the solvent. It has reached its maximum deuterium uptake under ten seconds. After one hour of exposure time it shows the same mass increase as after 10 seconds. The DynamX software calculated the relative deuterium uptake over all the peptides. This relative deuterium uptake can be plotted in function of the exposure time and also compared with other concentrations of IP268.
Figure 16 shows the difference in relative deuterium uptake for 4.5 mg/mL versus 0.05 mg/mL (**Fig. 16a**), for 4.5 mg/mL vs 0.1 mg/mL (**Fig .16b**) and 4.5 mg/mL vs 1.0 mg/mL of IP268 (**Fig. 16c**).

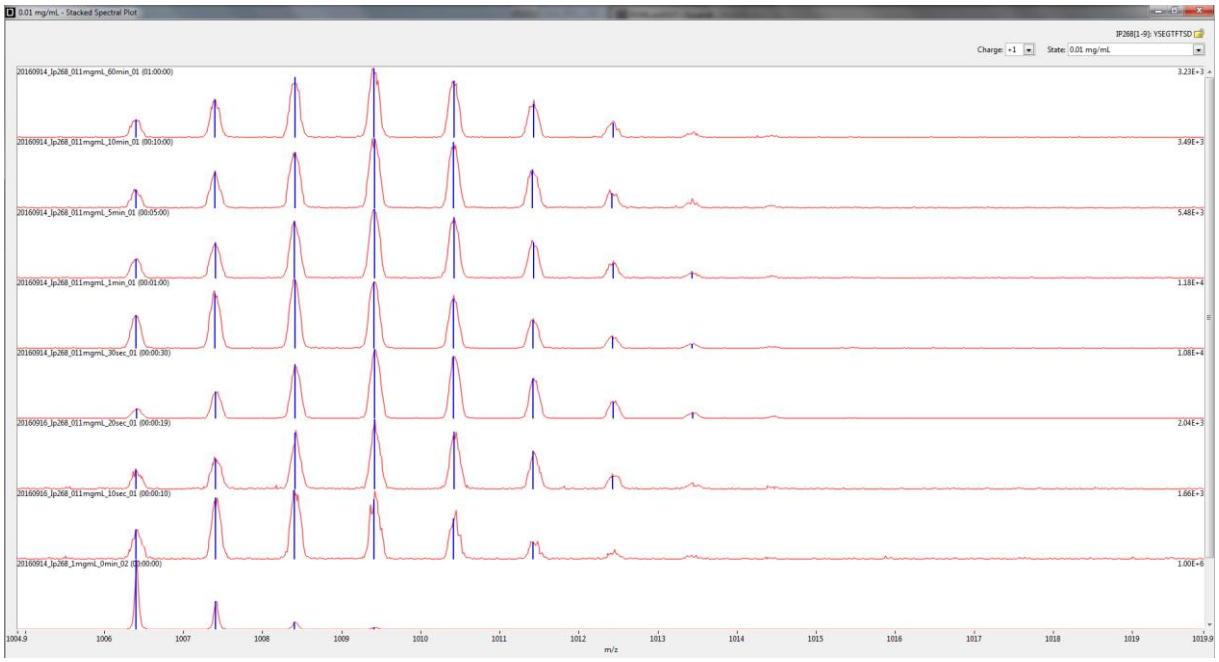


Figure 15: Stacked spectral plot of peptide 1-9 of IP268. At exposure time 0, 10 sec, 20 sec, 30 sec, 1 min, 5 min, 10 min and 60 min. After 10 seconds of exposure to D_2O, the peptide has increased in mass.

Three distinct areas in the selected peptides were displaying a significant difference when comparing different sample concentrations; peptide 1-9, 3-9 and 26-53. **Figure 17** shows the difference in deuterium intake for these peptides for each measured sample concentration. The observed trend for these peptides is that the difference index is decreasing as the concentration is increasing. This suggests that this area is slightly more protected at a higher concentration, although it is still flexible as it is susceptible to deuteration.

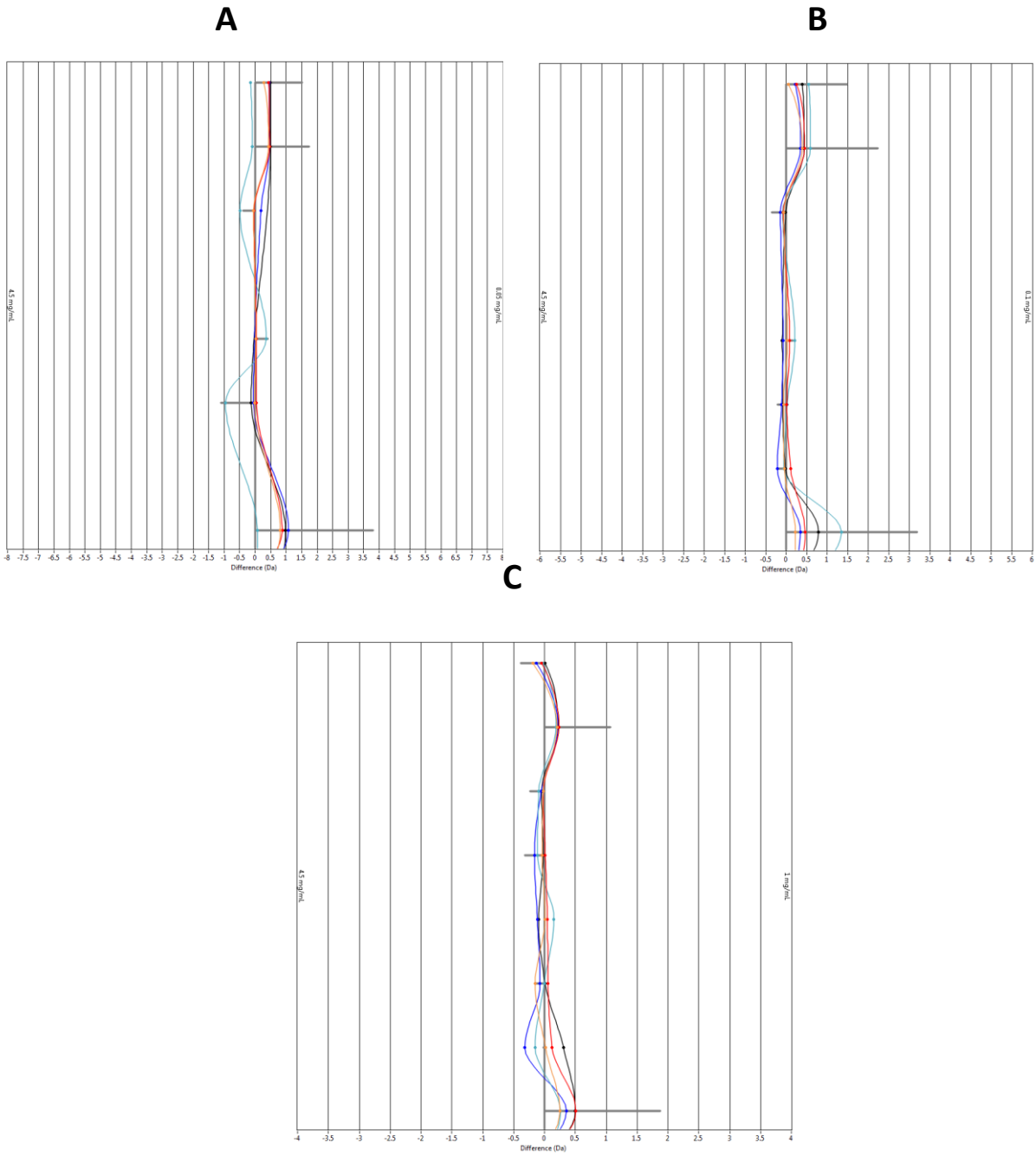


Figure 16: Difference in deuterium uptake for IP268 for (A) 4.5 mg/mL vs 0.05 mg/mL (B) 4.5 mg/mL vs 0.1 mg/mL and (C) 4.5 mg/mL vs 1 mg/mL

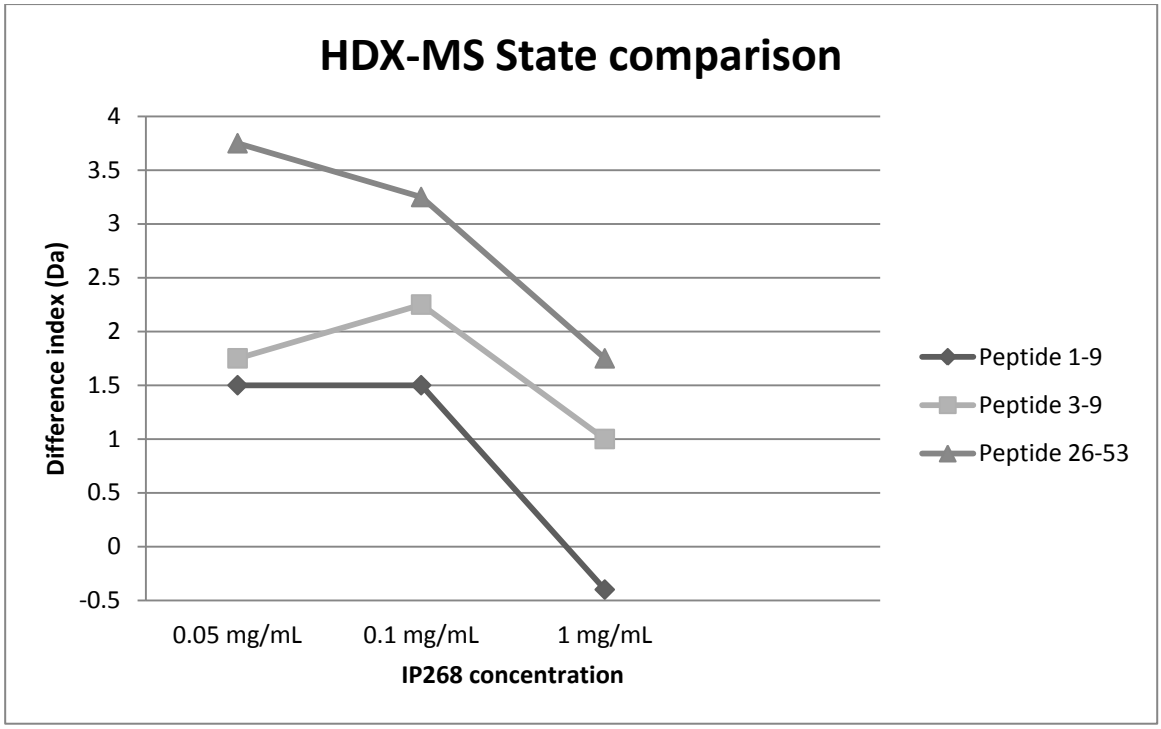


Figure 17: The difference index for three selected peptides for three different concentrations used in the HDX-MS experiment.

4.4 Conclusion

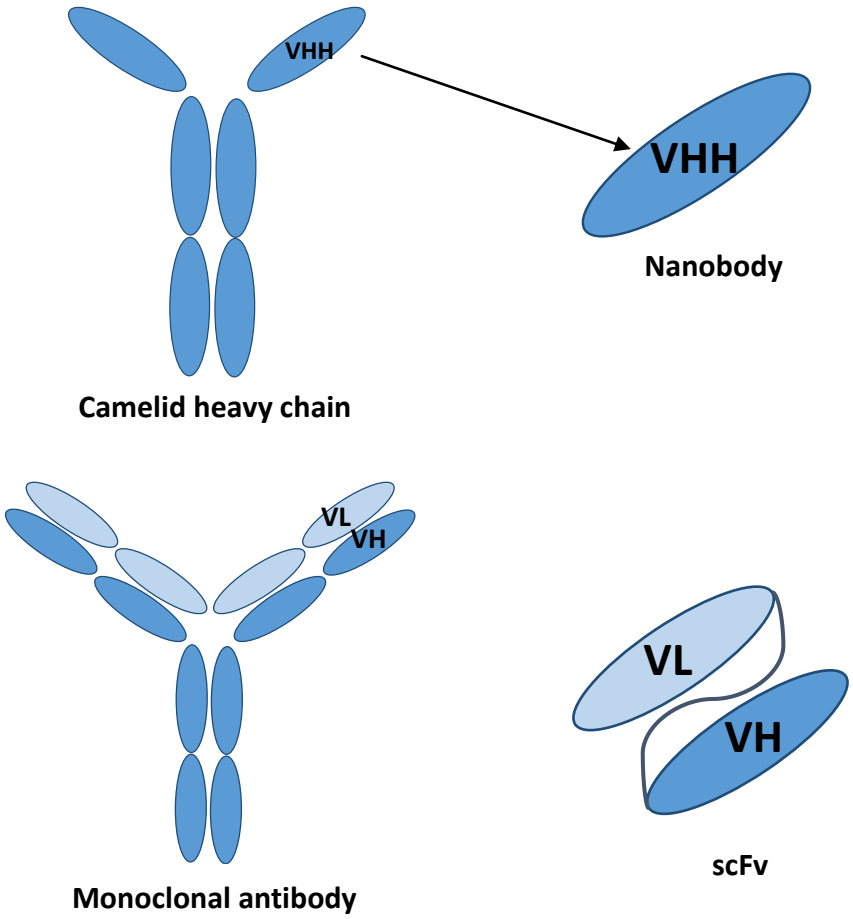
The broad range of analytical techniques applied on the Fc-fusion protein construct IP268, revealed a lot of information about the protein structure. With native IM-MS the molecular weight and CCS of the protein was determined. The Fc-fusion protein was compared to its Fc-analogue with native IM-MS. Based on the CIU plots, the Fc-fusion protein showed more stability compared to the Fc-fragment alone. The oligomeric state of IP268 was analysed by native MS, AUC, SEC-MALS and CovalX cross-linking, displaying the influence of the protein concentration on the ratio between monomer and dimer IP268. The hydrodynamic radius estimated by SEC-MALS was correlated with the CCS value obtained by ion mobility.
The AUC results could provide an estimate of the K_D for the dimerization of IP268. It was not feasible to go under this limit of 0.09 mg/mL to check if only monomer was observed. With CovalX cross-linking it was also not possible to decrease the concentration to the extent where only IP268 monomer was observed. The HDX-MS results suggest that the region where the linker is bound is more protected at higher protein concentrations, indicating the possible location of dimerisation.

4.5 Sources

[1] Martin Hall, Gregoire Pave, Company reports; Hardman & Co Life Sciences Research, 2016
[2] Daniel M. Czajkowsky, Jun Hu, Zhifeng Shao, Richard J. Pleass, Fc-fusion proteins: new developments and future perspectives, EMBO Mol Med (2012) 4, 1015-1028
[3] Timo R., Kristi Baker, Jennifer A. Dumont, Robert T. Peters, Haiyan Jiang, Shuo-Wang Qiao, Wayne I. Lencer, Glenn F. Pierce, and Richard S. Blumberg, Fc-fusion proteins and FcRn: structural insights for longer-lasting and more effective therapeutics, Crit Rev Biotechnol. 2015 Jun; 35(2): 235–254
[4] Jimenez-Solem E, Rasmussen MH, Christensen M, Knop FK.Dulaglutide, a long-acting GLP-1 analog fused with an Fc antibody fragment for the potential treatment of type 2 diabetes, Curr Opin Mol Ther. 2010;12(6):790–797
[5] Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Analytical Chemistry* **2010,** *82* (22), 9557-9565.
[6] Laue, T., Shah, B., Ridgeway, T., and Pelletier, S. Computer-aided interpretation of analytical sedimentation data for proteins, Royal Society of Chemistry Ed: Harding, S and Rowe, 1992, 90-125.
[7] Schuck, P. ,*Biophys J , 2000*, 78, 1606-1619.

Chapter 5

Study of nanobodies, single chain variable fragments and monoclonal antibodies against plasminogen activator inhibitor 1



In collaboration with
Machteld Sillen
Prof. Paul Declerck
Laboratory for Therapeutic and Diagnostic Antibodies
KU Leuven

5.1 Introduction

Antibody fragments such as Fab, scFv, diabodies and single-domain antibodies are engineered as alternatives to conventional monoclonal antibodies. These fragments maintain the targeting specificity of antibodies but can be produced more economically and possess other unique properties for a range of therapeutic applications [1].

Various strategic reasons for selecting an antibody fragment as a therapeutic agent instead of an intact IgG include:

- A smaller sized agent for the ability to penetrate tumors and for broader tissue distribution [2]
- Requirement for a short circulating half-life in serum [3]
- A molecule that lacks the Fc effector functionality (eliminating cellular response against the target and potential of dimerisation of receptors due to bivalency) [4]
- A molecule that can be manufactured in yeast or E. coli, resulting in cost reduction or the increase of the manufacturing scale [5]

In this study the focus lies on antibodies and antibody fragments against plasminogen activator inhibitor-1 (PAI-1), which is involved in cardiovascular diseases, cell migration and tumour development. [6]

5.1.1 Single-domain antibody (Nanobody)

A single-domain antibody fragment or Nanobody is a recombinant antibody fragment which has a molecular weight of about 15 kDa and is about 30% the size of a conventional Fab fragment (usually 50 kDa). The first nanobodies were produced from heavy-chain only antibodies found in camelids (**Fig. 1**).

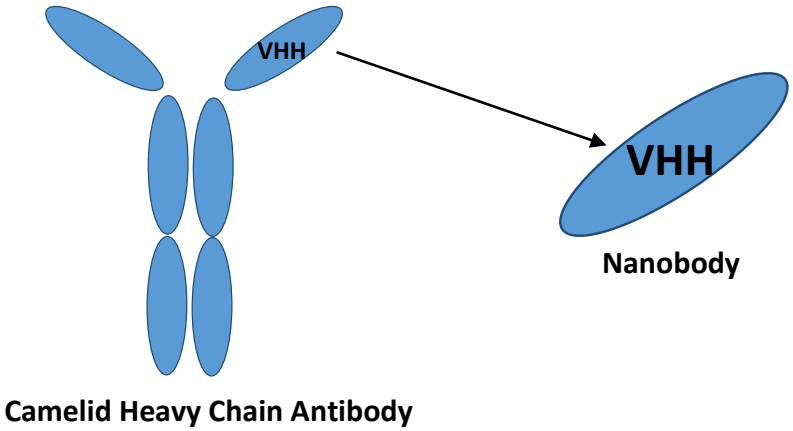


Figure 1: Schematic illustration of a single-domain antibody or nanobody.

While conventional antibodies bind to antigen through the variable domain of the heavy chain (VH) and of the light chain (VL), the camelid heavy chain antibodies bind to antigen through the VHH domain or nanobody.

Nanobodies have the inherited benefits of monoclonal antibodies such as high target affinity and selectivity and ease of discovery. In addition to these advantages, nanobodies are small proteins (only a tenth of the size of an antibody), so they can penetrate tissues more effectively

and can recognize hidden epitopes [7]. Nanobodies are also naturally soluble in aqueous solution and do not have the tendency to aggregate [8].
Their diagnostic and therapeutic use is being researched against a range of diseases such as cancer and cardiovascular diseases. In the field of antibody-based diagnostics, they can be used for immunoassays, as they can bind specifically with the target antigen. In antibody therapy this specific binding to target cells or proteins can stimulate the immune system of the patient to attack those targets.
Another important application is the use of these Nanobodies for structural determination of the target protein. The antibody fragments can bind to specific antigens and stabilize a particular conformational state of a protein, resulting in a conformationally uniform sample which simplifies the determination of the protein structure [9].

5.1.2 Single-chain variable fragment (scFv)

Single chain variable fragment (scFv) is an engineered variant in which the V_H and V_L domains are joined with a flexible polypeptide linker (**Fig. 2**). ScFvs are used as affinity reagents for diagnostics, therapeutics and proteomic analysis [10]. Target-specific scFvs are identified by various display technologies that utilise a genetic library of VH and VL gene segments. Besides retaining the specific antigen binding affinity of the parent IgG, these fragments also show improved pharmacokinetics for tissue penetration [11]. Since these fragments have a rapid clearance from blood, they can be coupled with drugs to result in lower exposure of the healthy tissue [12]. These are important properties, especially in cancer therapy. ScFvs also have a relatively lower commercial cost in large-scale production and it is possible to restructure them in order to improve their activity and production [13]. While scFvs are highly specific, they can be limited in their stability and by their aggregation tendency [14]. Therefore, the biophysical properties first have to be optimized before they are incorporated into therapeutic agents or clinical diagnostics. ScFvs are viewed as promising therapeutic agents, with several in clinical trials [15, 16].

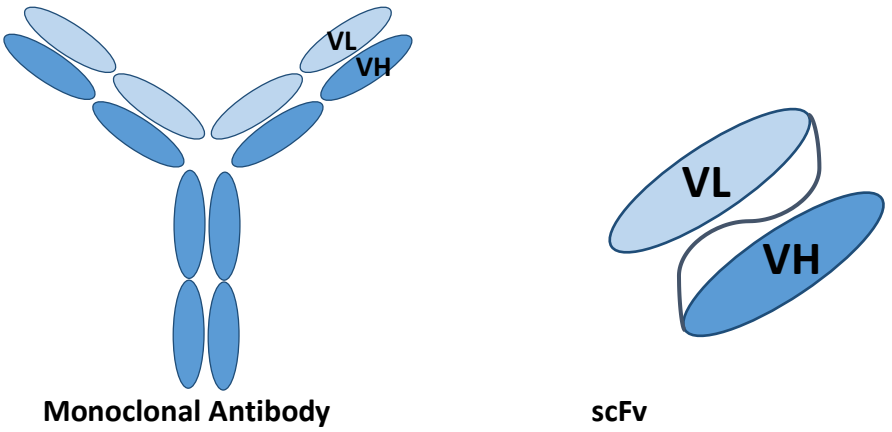


Figure 2: Schematic illustration of a single chain variable fragment or scFv.

5.1.3 Plasminogen activator inhibitor 1 (PAI-1)

Plasminogen activator inhibitor-1 (PAI-1) is a serine protease inhibitor (serpin) that regulates the fibrinolytic system by binding to and inhibiting plasminogen activators. The fibrinolytic system is involved in the degradation of fibrin, a fibrous protein involved in the clotting of blood. In this system, the inactive proenzyme, plasminogen, can be converted to the active enzyme, plasmin, which degrades fibrin into soluble degradation products. Two plasminogen activators have been identified: the tissue-type plasminogen activator (t-PA), and the urokinase-type plasminogen activator (u-PA). PAI-1 specifically inhibits t-PA and u-PA, which can result in the inhibition of fibrinolysis. PAI-1 has been identified in three different states; active, latent and cleaved (**Fig. 3**). In its active form (**Fig. 3B**), PAI-1 interacts with the target protease by a 20-residue surface-exposed loop called the reactive center loop (RCL), which includes the bait peptide bond Arg^{346}-Met^{347} (P1-P1')[17].

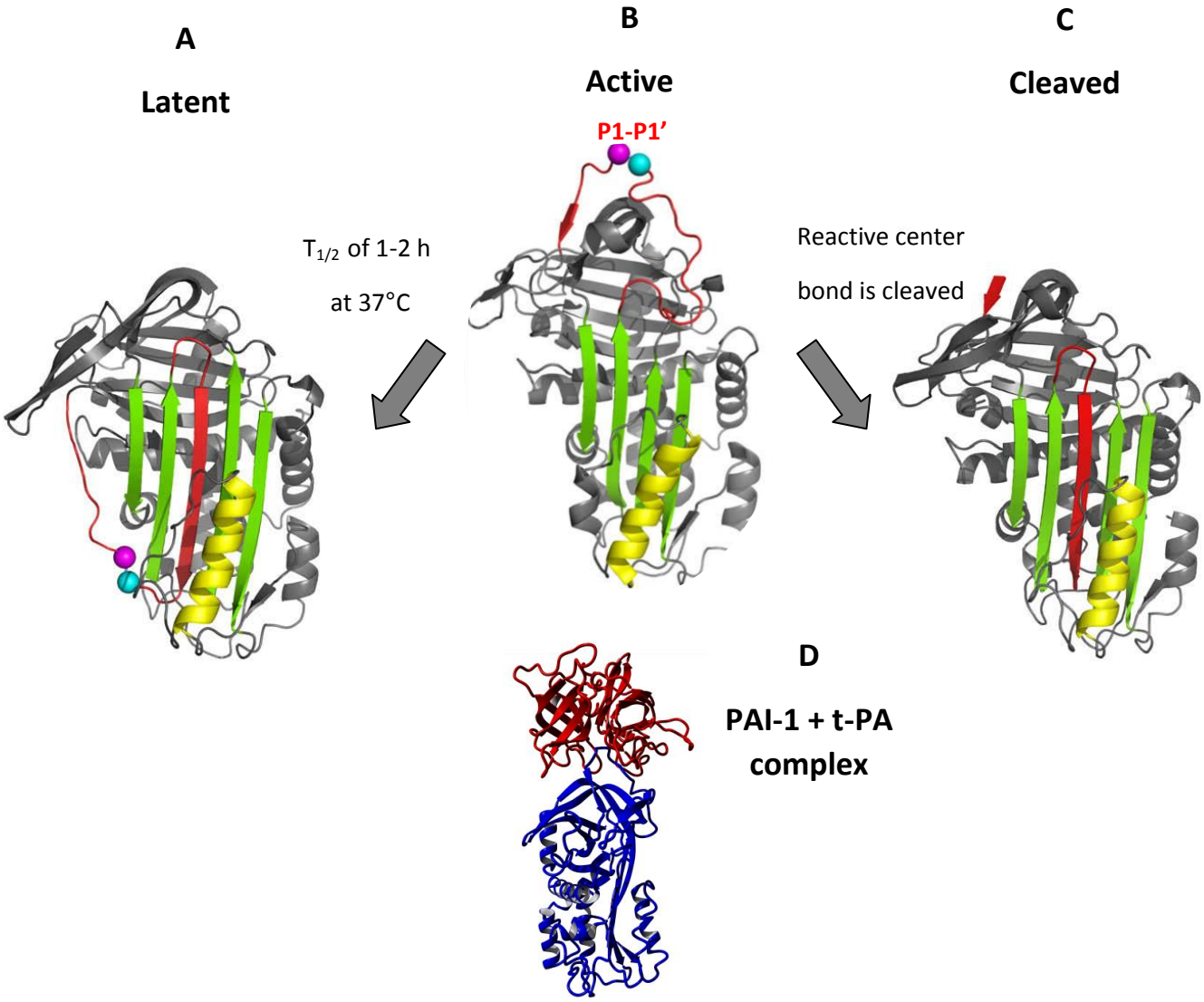


Figure 3: Crystal structures of latent (A), active (B) and cleaved (C) PAI-1. β-sheet A is displayed in green, the α-helix F in yellow and the reactive centre loop in red. [18] In Figure D the t-PA + PAI-1 complex is shown with PAI-1 in blue and t-PA in red (PDB 5brr).

In the active PAI-1 conformation, the RCL is exposed on the surface of PAI-1. T-PA can bind to this loop and cleave it at position P1-P1'. By cleaving this loop, an intermediate covalent complex is formed (**Fig. 3D**). This intermediate PAI-1/t-PA complex can however still be hydrolysed. When PAI-1 interacts with t-PA through the inhibitory pathway, the cleaved loop, which is at this time bound to t-PA, is inserted in the central beta-sheet and t-PA gets relocated to the opposite side of the PAI-1 molecule. This results in the distortion of the active site of t-PA (where it is bound to the RCL), making it impossible to hydrolyse the intermediate covalent bond so that t-PA gets trapped in a final covalent complex. To summarise, the protease is inhibited after interaction with the active form of PAI-1. In this study, a stabilised active mutant of PAI-1, PAI-1 W175F, was used in which the tryptophan residue 175 was replaced by a phenylalanine. Another stabilised active PAI-1 mutant is PAI-1 Stab which contains five mutations; N150H- K154T-Q301P-Q319L-M354I. The half-life of PAI-1 Stab is 150 hours at 37 °C and the crystal structure of this active form is represented by PDB 1DB2.

When PAI-1 interacts with t-PA through the substrate pathway, this intermediate covalent complex is hydrolysed, resulting in the release of t-PA (which remains active since the catalytic site is restored in its initial form by this hydrolysis). PAI-1 remains inactive since the loop was cleaved by t-PA **(Fig 3C).** It is observed that PAI-1-Stab reacts with t-PA via this substrate pathway in a slightly more efficient manner than wild-type PAI-1. This may be due to the stabilizing mutations in the region between alpha-helix F (the hinge region) and the central beta-sheet B. Movement of this hinge-region appears to be important for loop insertion (which is needed to trap tPA in the final covalent complex).

This substrate behaviour of PAI-1 can be induced in different ways: by mutations in the hinge region, by changes in external conditions or by addition of monoclonal antibodies [19,20,21,22]. The PAI-1 mutant, PAI-1 P14, in which the threonine residue 14 is replaced by a proline, shows substrate behavior when interacting with t-PA. This emphasises the importance of the reactive site loop. Although there has been a structure elucidated for a cleaved substrate variant of PAI-1 (PAI-1 A335P), there has not been any structure published yet for the intact substrate form of PAI-1 [18].

In the latent conformation of PAI-1, the N-terminal part of the RCL is inserted into β-sheet A, forming a new β-sheet (**Fig. 3A**). The reactive centre loop changes from solvent-exposed to mostly buried, which makes the reactive centre inaccessible to the target protease [23]. Active PAI-1 spontaneously converts to the latent conformation under physiological conditions with a half-life of 1-2 hours at 37°C [24]. This means that a sample of wild type PAI-1 can consist of a distribution of active, substrate and latent form of PAI-1. The rate of latency transition is reduced at a low pH (pH 5.5), high salt concentration (1 M NaCl) and low temperature (4 °C) [24]. Stablised active PAI-1 mutants can have half-lives of about 150 hours [25] W175F shows an increased functional half-life of the active conformation (≈7 hours) [26].

Table 1 shows an overview of the five variants of PAI-1 which were analysed in this study; PAI-1 wild type (PAI-1 WT), stabilised active PAI-1 (PAI-1 stab), stabilised active PAI-1 (PAI-1 W175F), PAI-1 with substrate behaviour (PAI-1 P14) and latent PAI-1 (PAI-1 Latent). The molecular weight, collision cross section and stability of these five variants were compared with native ion mobility-mass spectrometry (IM-MS) and by collision-induced unfolding (CIU).

Table 1: Overview of the analysed PAI-1 variants: PAI-1 wild type (PAI-1 WT), stabilised active PAI-1 (PAI-1 stab), stabilised active PAI-1 (PAI-1 W175F), PAI-1 with substrate behaviour (PAI-1 P14) and latent PAI-1 (PAI-1 Latent)

Name	Variant	Symbol	Structure
PAI-1 WT	Wild type		**Fig. 3A+B**
PAI-1 Stab	Active		**PDB 1DB2**
PAI-1 W175F	Active		**Fig 3B**
PAI-1 P14	Active		No structure available
PAI-1 Latent	Latent		**Fig 3A**

5.1.4 Nanobodies, ScFv and monoclonal antibodies against PAI-1

In this study the nanobodies VHH-2g-42, VHH-2w-64 and VHH-2-12 were analysed (**Table 2**). VHH2g-42 and VHH-2w-64 display moderate PAI-1 inhibition while VHH-2-12 inhibits PAI-1 activity up to only 10 % [27]. VHH-2g-42 may inhibit PAI-1 activity by preventing the initial PAI-1-t-PA complex formation [27]. VHH-2w-64 inhibits PAI-1 activity by inducing substrate behaviour of PAI-1 [27]. When VHH-2-12 binds to PAI-1 it causes a slight acceleration of the latency transition [27]. The molecular weight and the collision cross sections were determined for the nanobodies and their respective complexes with the two stabilised active forms of PAI-1 (PAI-1 Stab and PAI-1 W175F). PAI-1 Stab was also measured in complex with both VHH-2w-64 and VHH-2g-42.

Table 2: Overview of the analysed anti-PAI-1 nanobodies and the nanobodies in complex with PAI-1 Stab and with PAI-1 W175F

Name	Variant	Symbol
VHH-2g-42	Inhibits PAI-1	
VHH-2w-64	Induces substrate behaviour	
VHH-2-12	Induces latency	
PAI-1 Stab + VHH-2g-42	Inhibits PAI-1	
PAI-1 Stab + VHH-2w-64	Induces substrate behaviour	
PAI-1 Stab + VHH-2-12	Induces latency	
PAI-1 W175F + VHH-2g-42	Inhibits PAI-1	
PAI-1 W175F + VHH-2w-64	Induces substrate behaviour	
PAI-1 W175F + VHH-2-12	Induces latency	
PAI-1 Stab + VHH-2w-64 +VHH-2g-42	Inhibits PAI-1	

The scFv analysed in this study is scFv-33H1F7, which is derived from the corresponding PAI-1 neutralising monoclonal antibody, Mab 33H1F7. When scFv-33H1F7 binds to active PAI-1 it induces a non-inhibitory substrate behaviour of PAI-1 [28]. For comparison the anti-PAI-1 ScFv was measured in complex with the latent form PAI-1, and with the stabilized active form; PAI-1 Stab (**Table 3**).

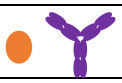
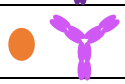
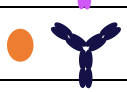
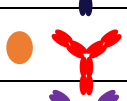
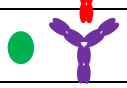
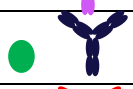
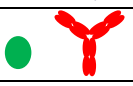
Table 3: Overview of the analysed anti-PAI-1 scFv-33H1F7 and PAI-1 complexes

Name	Variant	Symbol
ScFv-33H1F7	Induces substrate behaviour	
Latent PAI-1 +ScFv-33H1F7	Latent	
PAI-1 Stab +ScFv-33H1F7	Induces substrate behaviour	

5.1.5 Monoclonal antibodies (mAbs)

PAI-1 activity has also been regulated using monoclonal antibodies. In this study we looked at four anti-PAI-1 Mabs; Mab-33H1F7, Mab-55F4C12, Mab-8H9D4 and Mab-33B8. Mab-33H1F7, Mab-55F4C12 and Mab-8H9D4 are murine IgG1s that induce substrate behaviour of PAI-1 [29]. Mab-33B8 is a murine IgG1 that induces the acceleration of the active-to-latent conversion of Pai-1 [30]. The monoclonal antibodies were measured in complex with PAI-1 Stab and with PAI-1 W175F (stabilised active forms of PAI-1) (**Table 4**). The molecular weight and collision cross sections of the different proteins and protein complexes were determined with native IM-MS.

Table 4: Overview of the analysed anti-PAI-1 monoclonal antibodies and the monoclonal antibody in complex with PAI-1 Stab and PAI-1 W175F

Name	Variant	Symbol
Mab-33H1F7	Induces substrate behaviour	
Mab-55F4C12	Induces substrate behaviour	
Mab-8H9D4	Induces substrate behaviour	
Mab-33B8	Induces latency	
PAI-1 Stab + Mab-33H1F7	Induces substrate behaviour	
PAI-1 Stab + Mab-55F4C12	Induces substrate behaviour	
PAI-1 Stab + Mab-8H9D4	Induces substrate behaviour	
PAI-1 Stab + Mab-33B8	Induces latency	
PAI-1 W175F + Mab-33H1F7	Induces substrate behaviour	
PAI-1 W175F + Mab-55F4C12	Induces substrate behaviour	
PAI-1 W175F + Mab-8H9D4	Induces substrate behaviour	
PAI-1 W175F + Mab-33B8	Induces latency	

5.2 Experimental details

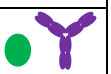
The PAI-1 samples, nanobodies, ScFv and monoclonal antibodies were provided by the Laboratory for Therapeutic and Diagnostic Antibodies, KU Leuven. The PAI-1 complexes were prepared at a 1:1 molar ratio between PAI-1 and the anti-PAI-1 protein. An overview of the measured samples is displayed in **Table 5**.

A Synapt G2 HDMS Q-TOF instrument (Waters, UK) was used for the native IM-MS experiments. The samples were in 150 mM ammonium acetate buffer at pH 5.5.
Gold-coated nanoESI needles (prepared in-house) were used to transfer the sample into the mass spectrometer. The instrument was tuned to preserve native higher-order structure using the following parameters: Nano-ESI capillary voltage 1.6 kV; nanoflow backing gas pressure 0.1 bar; sampling cone 30 V; extraction cone 1.0 V; trap collision energy 10 V; transfer collision energy 10 V; trap bias 45V; IMS gas flow 90 ml/min; IMS wave velocity 350 m/s; IMS wave height 25 V; backing pressure 5.0 mbar.

The data were acquired and processed with Masslynx v4.1 software, and ion mobility drift times extracted using Driftscope v2.3 (both Waters). The collision cross sections (CCS) of the proteins were calibrated using known CCS values determined under native conditions as described previously [31].

Collision-induced unfolding plots were made by plotting the drift time value in function of the trap collision energy value. The plots were made with OriginPro 8.5 software.

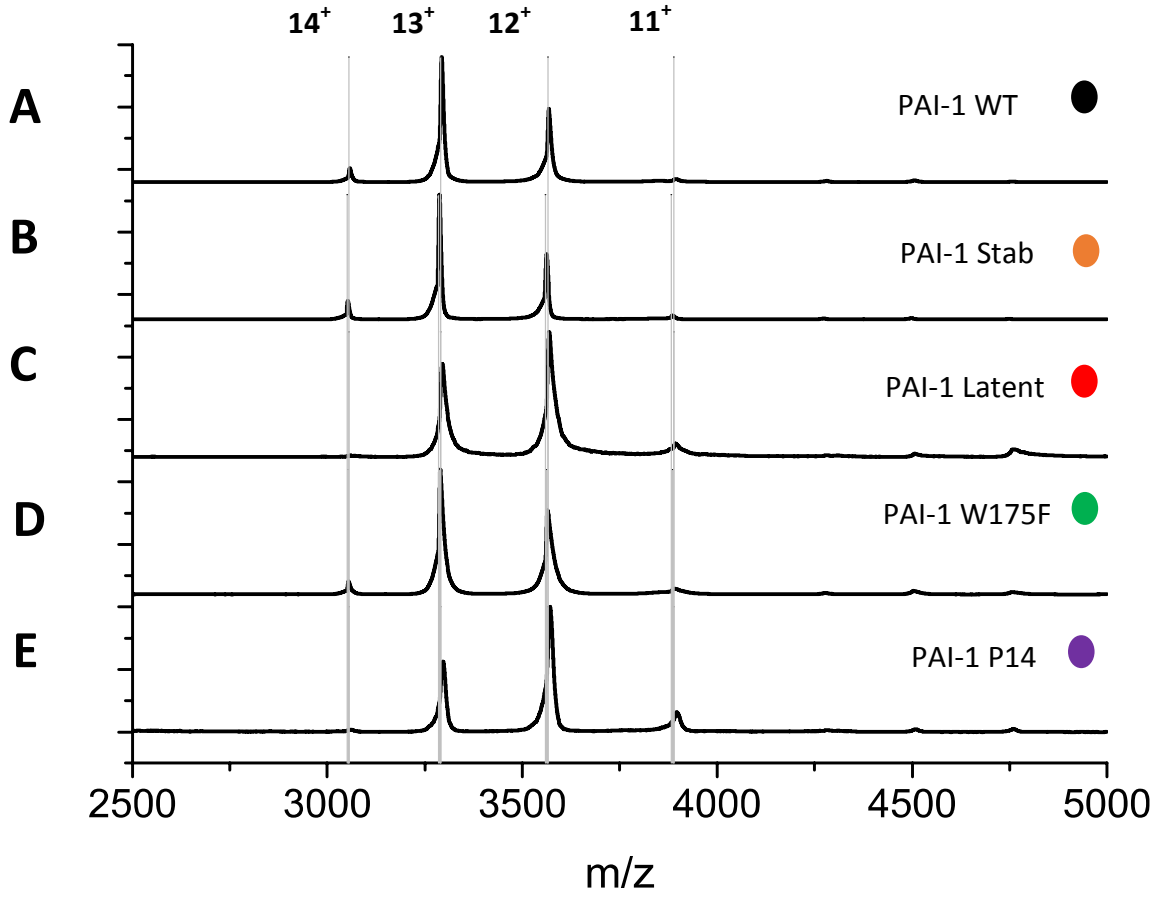
Table 5: Studied PAI-1 variants, nanobodies, scFV, antibodies and PAI-1 complexes.

Name	Variant	Type	Symbol
PAI-1 WT	Wild type	PAI-1	
PAI-1 Stab	Cleaved	PAI-1	
PAI-1 W175F	Active	PAI-1	
PAI-1 P14	Substrate	PAI-1	
PAI-1 Latent	Latent	PAI-1	
VHH-2g-42	Inhibits PAI-1	Nanobody	
VHH-2w-64	Induces substrate behaviour	Nanobody	
VHH-2-12	Induces latency	Nanobody	
PAI-1 Stab + VHH-2g-42	Inhibits PAI-1	PAI-1 + nanobody	
PAI-1 Stab + VHH-2w-64	Induces substrate behaviour	PAI-1 + nanobody	
PAI-1 Stab + VHH-2-12	Induces latency	PAI-1 + nanobody	
PAI-1 W175F + VHH-2g-42	Inhibits PAI-1	PAI-1 + nanobody	
PAI-1 W175F + VHH-2w-64	Induces substrate behaviour	PAI-1 + nanobody	
PAI-1 W175F + VHH-2-12	Induces latency	PAI-1 + nanobody	
PAI-1 Stab +VHH-2w-64 +VHH-2g-42	Inhibits PAI-1	PAI-1 + nanobody	
ScFv-33H1F7	Induces substrate behaviour	ScFv	
Latent PAI-1 +ScFv-33H1F7	Latent	PAI-1 + ScFv	
PAI-1 Stab +ScFv-33H1F7	Induces substrate behaviour	PAI-1 + ScFv	
Mab-33H1F7	Induces substrate behaviour	Mab	
Mab-55F4C12	Induces substrate behaviour	Mab	
Mab-8H9D4	Induces substrate behaviour	Mab	
Mab-33B8	Induces latency	Mab	
PAI-1 Stab + Mab-33H1F7	Induces substrate behaviour	PAI-1 + Mab	
PAI-1 Stab + Mab-55F4C12	Induces substrate behaviour	PAI-1 + Mab	
PAI-1 Stab + Mab-8H9D4	Induces substrate behaviour	PAI-1 + Mab	
PAI-1 Stab + Mab-33B8	Induces latency	PAI-1 + Mab	
PAI-1 W175F + Mab-33H1F7	Induces substrate behaviour	PAI-1 + Mab	
PAI-1 W175F + Mab-55F4C12	Induces substrate behaviour	PAI-1 + Mab	
PAI-1 W175F + Mab-8H9D4	Induces substrate behaviour	PAI-1 + Mab	
PAI-1 W175F + Mab-33B8	Induces latency	PAI-1 + Mab	

5.3 MS analysis of five PAI-1 variants

With native mass spectrometry the molecular weight was determined for the five studied PAI-1 variants (**Fig. 4**). The observed charge states were 11^+ to 14^+ with 12^+ and 13^+ forming the most intense peaks for each PAI-1 variant. The experimental and theoretical masses are listed in **Figure 4.**

The slight differences between experimental and theoretical mass can be explained by buffer and salt adducts that could not be removed through desalting of the sample. PAI-1 also has sites prone to chloride binding.



		Protein	Experimental mass (Da)	Theoretical mass (Da)	Δ (Da)
A	●	PAI- WT	42,839.54	42,769.20	70.34
B	●	PAI-1 Stab	42,785.64	42,701.10	84.54
C	●	PAI-1 Latent	42,779.52	42,769.20	10.32
D	●	PAI-1 W175F	42,839.53	42,730.10	109.43
E	●	PAI-1 P14	42,832.91	42,765.21	67.7

Figure 4: Native mass spectra of the PAI-1 wildtype (A) and the PAI-1 variants; (B) Stab, (C) Latent, (D) W175F and (E) P14.

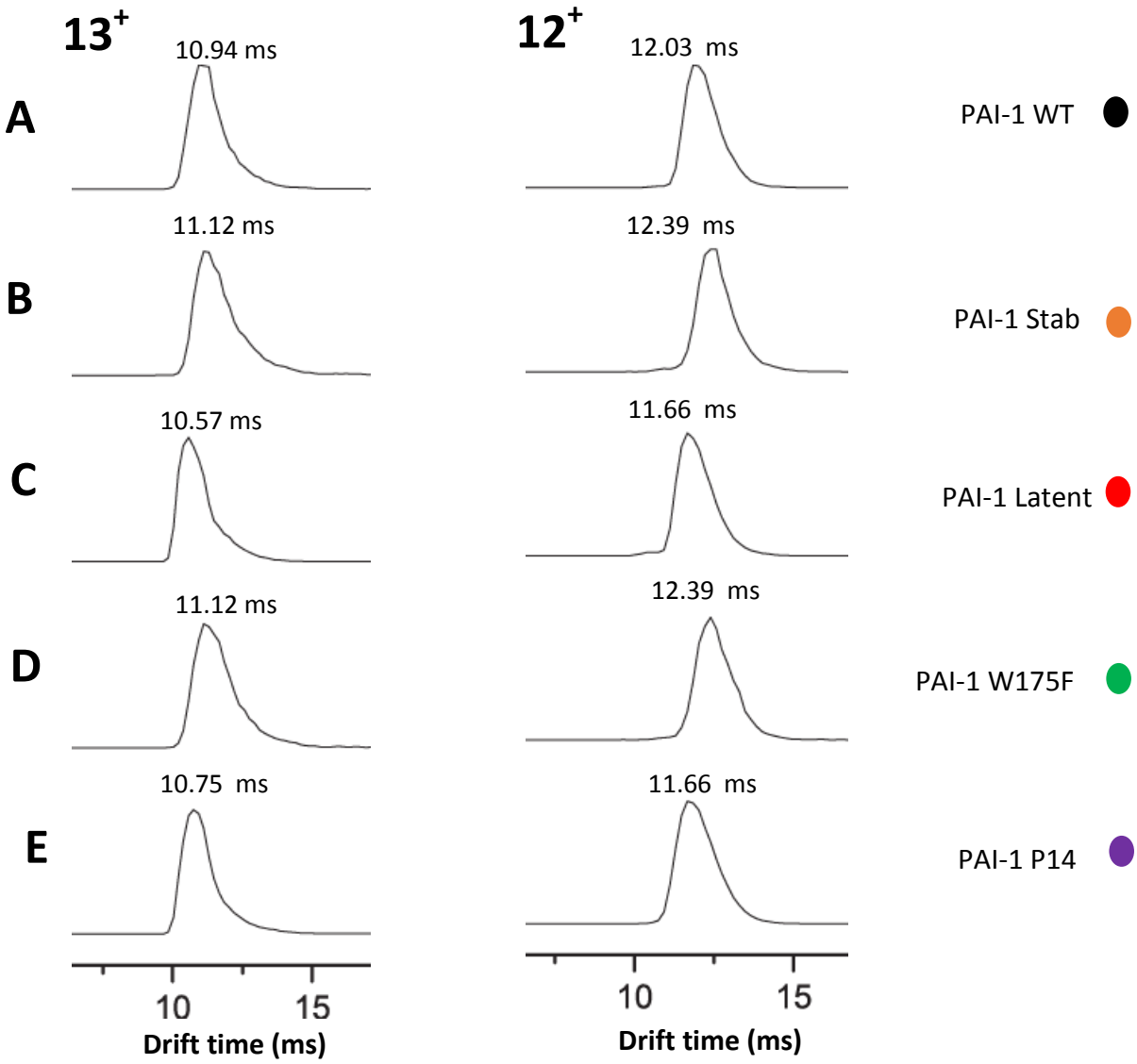
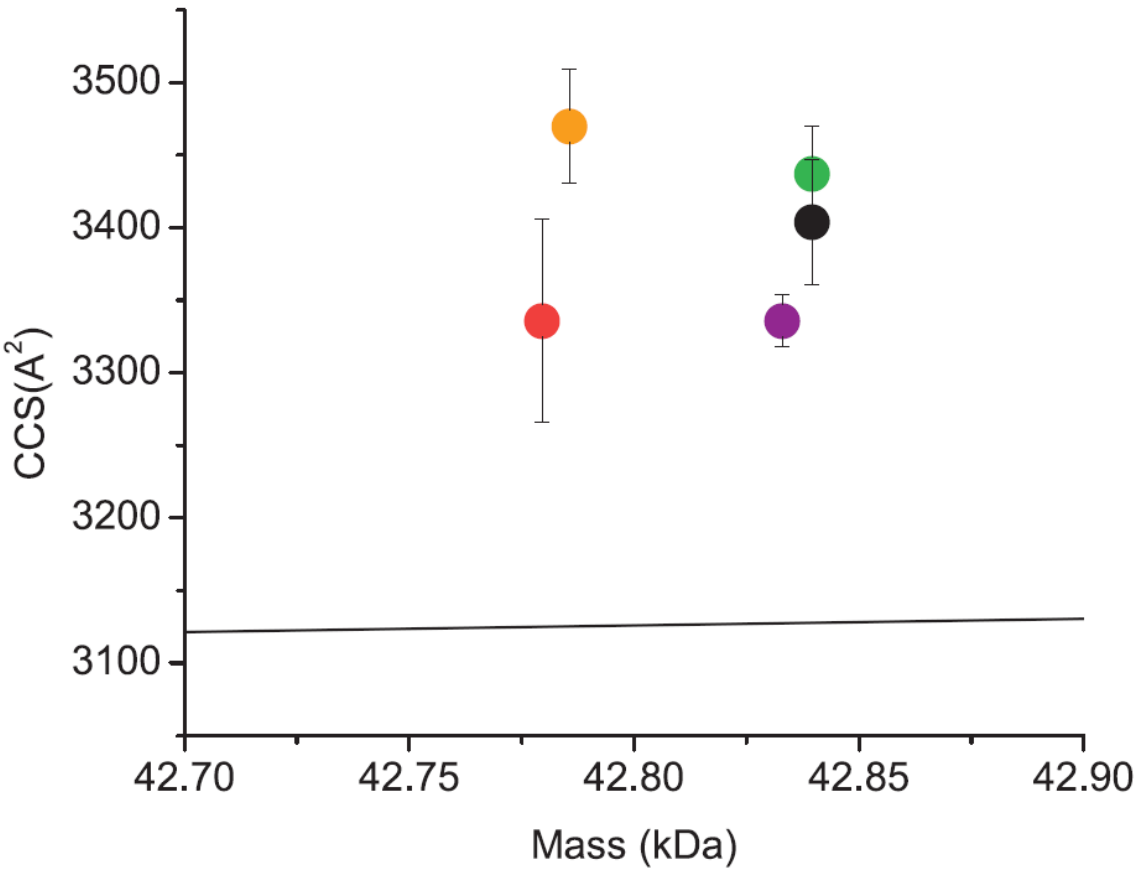


Figure 5: Ion mobility drift times of the 12+ and 13+ charge states of the PAI-1 wildtype (A) and the PAI-1 variants; (B) Stab, (C) Latent, (D) W175F and (E) P14.

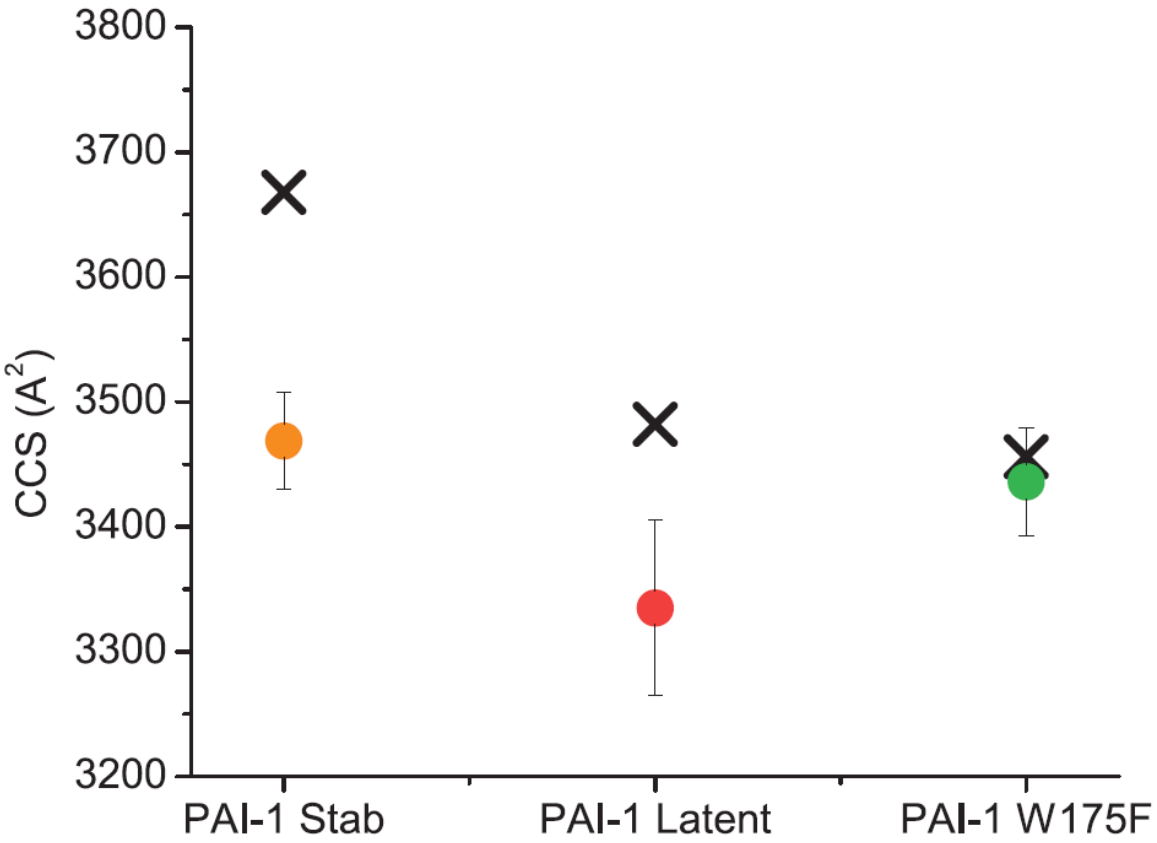
The ion mobility drift times of the 12^+ and 13^+ charge states of the PAI-1 variants are displayed in **Figure 5**. For each m/z peak there was one clear characterised conformation. These native ion mobility drift time values of the PAI-1 variants were converted into collision cross sections and are displayed in **Figure 6A**.

The most compact CCS values were shown for PAI-1 Latent and PAI-1 P14 (substrate). The largest CCS values were calculated for the stabilised active PAI-1 W175F and PAI-1 Stab. The difference between the largest and smallest CCS value is 4 % which makes the difference small but significant. The conformational change that is expected from active to latent form is the reorientation of the reactive central loop (RCL). Based on CCS alone it is difficult to distinguish the active and latent PAI-1 variant. However, the higher CCS values do correspond to the PAI-1 variants where the RCL is exposed (PAI-1 W175F and PAI-1 Stab), while the lower CCS values correspond to the PAI-1 variants where the RCL is inserted (PAI-1 Latent).



CCS	PAI-1 WT	PAI-1 W175F	PAI-1 Stab	PAI-1 Latent	PAI-1 P14
12+	3404 ± 33 Å^2	3436 ± 43 Å^2	3469 ± 39 Å^2	3335 ± 70 Å^2	3335 ± 18 Å^2
13+	3504 ± 27 Å^2	3541 ± 43 Å^2	3504 ± 51 Å^2	3391 ± 75 Å^2	3428 ± 18 Å^2
MW	42839.5Da	42839.5Da	42785.6 Da	42779.5 Da	42832.9 Da
		RCL exposed	*RCL exposed*	*RCL inserted*	*RCL inserted?*

Figure 6A: Collision cross sections of the PAI-1 wild type and the four studied PAI-1 variants. The black line represents the correlation between CCS and molecular weight based on globular calibrants.



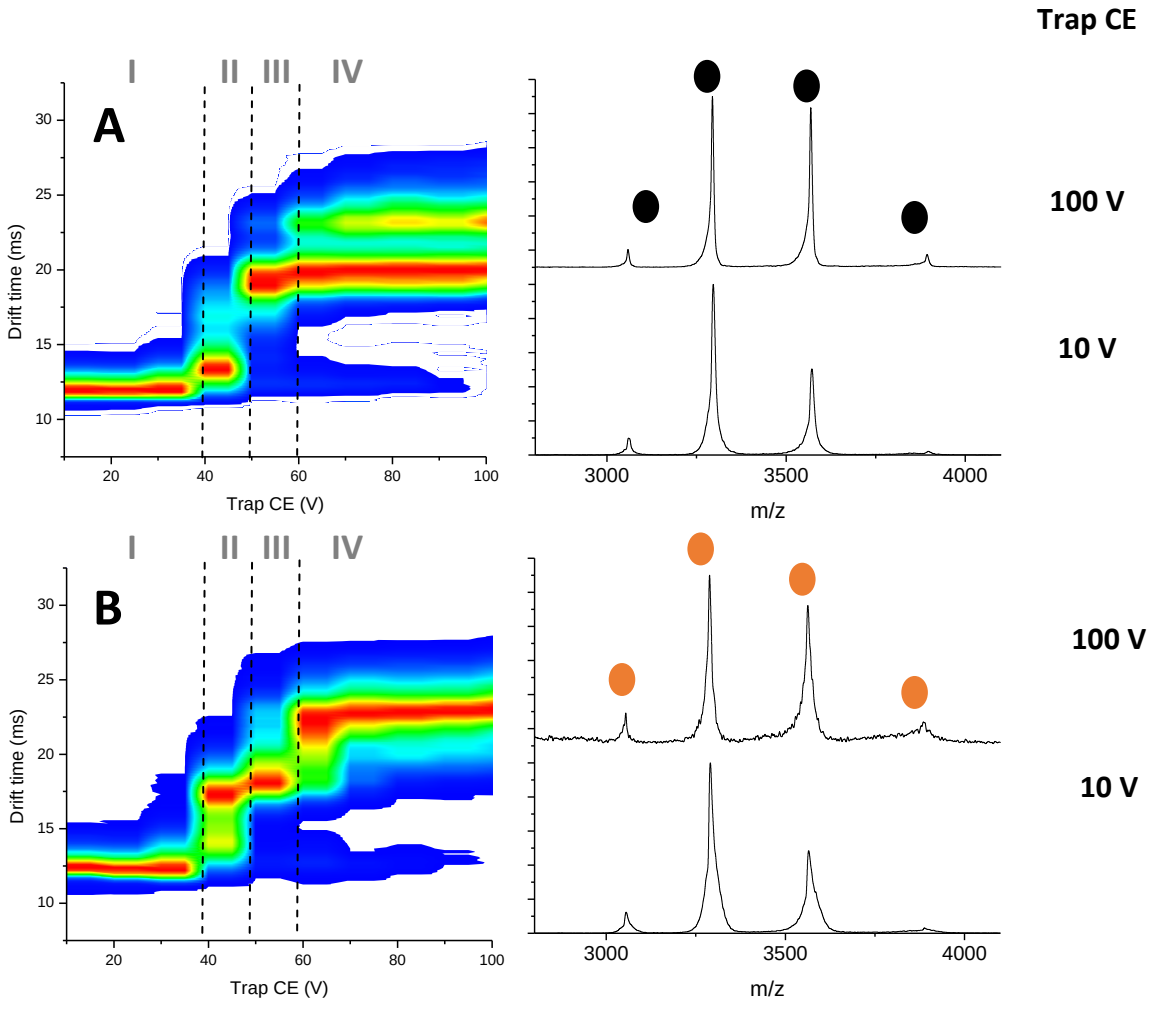
Protein	Experimental CCS (A^2)	Calculated CCS (A^2)	Δ (A^2)	Δ %
Active PAI-1 (1DVM chain A)	3469 ± 43	2697 x 1.36 = 3668	199	5.7 %
Latent PAI-1 (1 DVN)	3335 ± 70	2561 x 1.36 = 3483	148	4.4 %
PAI-1 W175F (3Q02)	3436 ± 43	2541 x 1.36 = 3456	20	0.6 %

Figure 6B: Comparison of the experimental CCS values of the 12^+ charge state of PAI-1 Stab, Pai-1 Latent and PAI-1 W175F with the calculated CCS values based on the crystal structures using IMoS software.

In **Figure 6B** the experimental CCS values of PAI-1 Stab, PAI-1 Latent and PAI-1 W175F are compared with the calculated CCS values based on the respective crystal structures (I.e. PDB 1DVM, 1DVN and 3Q02). These CCS values were calculated using the Ion Mobility Spectrometry Suite (IMoS) software in which the Projected Area (PA) method was applied as described previously [33]. In literature it is mentioned that the results of the PA method can underestimate the CCS value by a certain factor (i.e. from 1.14 to 1.36 fold) as it does not take in to account the gas scattering and ion-induced dipole interactions between the charged protein and the neutral gas molecules. [33,34]. When we multiply the calculated results with 1.36, they come even closer to the experimental values (**Fig. 6B**).

5.3.1 Unfolding patterns of PAI-1 variants

To investigate the stability and architecture of the folded protein, collision-induced unfolding plots were created for each PAI-1 variant (**Fig. 7**). The IM drift time for the five variants stays stable until a trap voltage of 40 V when there is a shift to an increased IM drift time (from 12.36 ± 0.51 ms to 14.0 ± 0.46 ms). This shift is larger for PAI-1 stab (**Fig. 7B**) and PAI-1 W175F (**Fig. 7C**) as there is also an intense signal at 17.31 ms which is not that intense in the other three plots. At 50 V another IM drift time shift occurs to 19.59 ± 0.64 ms for all the PAI-1 variants, except for PAI-1 Stab (**Fig. 7B**) where the shift is slightly smaller and goes to 18.04 ± 0.54 ms. At 60 V, PAI-1 Stab and PAI-1 W175F (**Fig. 7B & C**) show a different trend compared to the other three plots. PAI-1 Stab displays a drift time of 22.23 ± 0.80 ms and PAI-1 W175F a drift time of 24.24 ± 64 ms and maintains this until 100 V. For the wild type, latent and P14 variant of PAI-1 (**Fig. 7A, D, E**) there is a distribution of two conformations of which the lower one (20.66 ± 0.60 ms) is more intense. However, after 70 V the drift time of PAI-1 Latent is more intense at 24.33 ± 0.09 ms and for PAI-1 P14 this transition happens at 80 V.



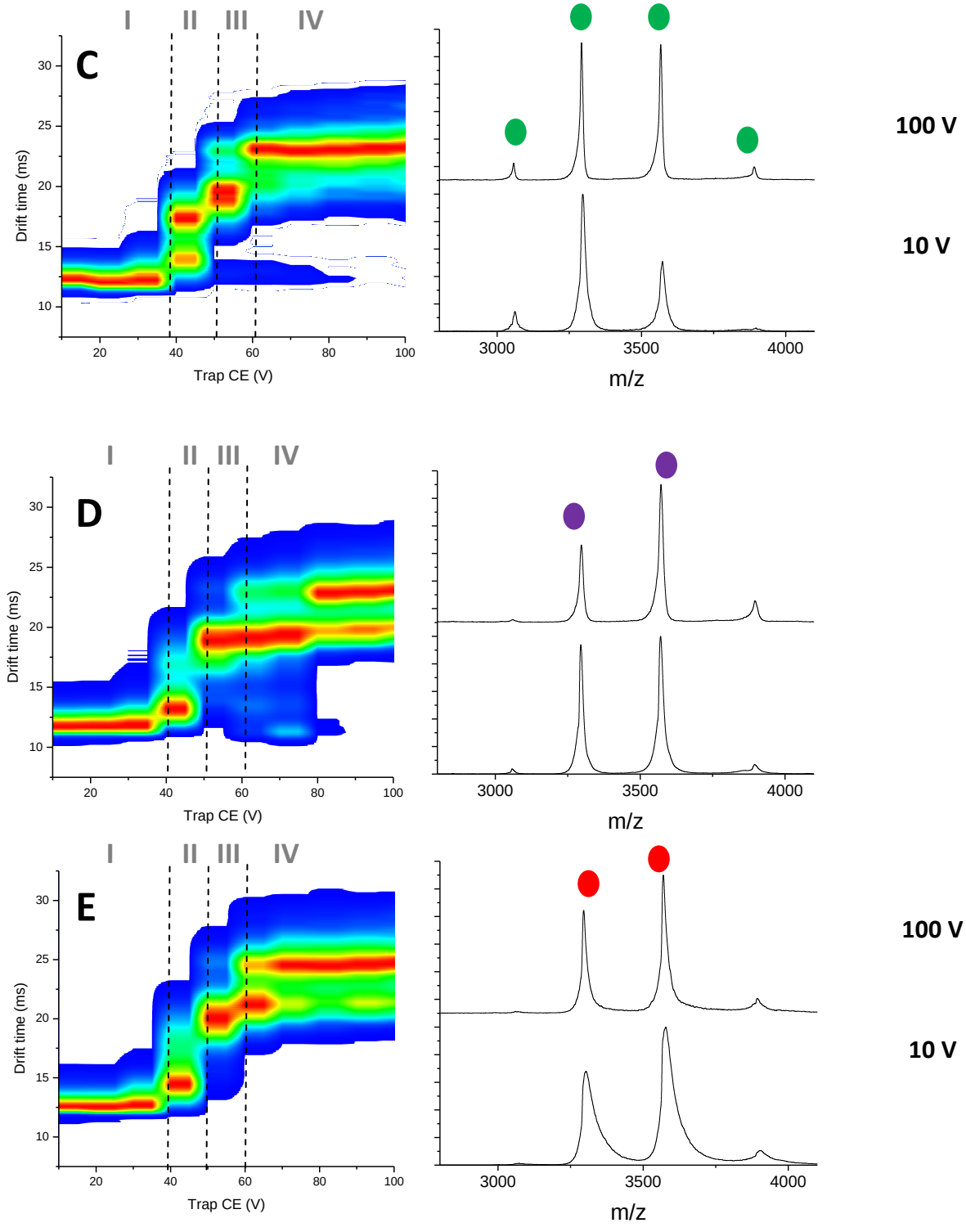
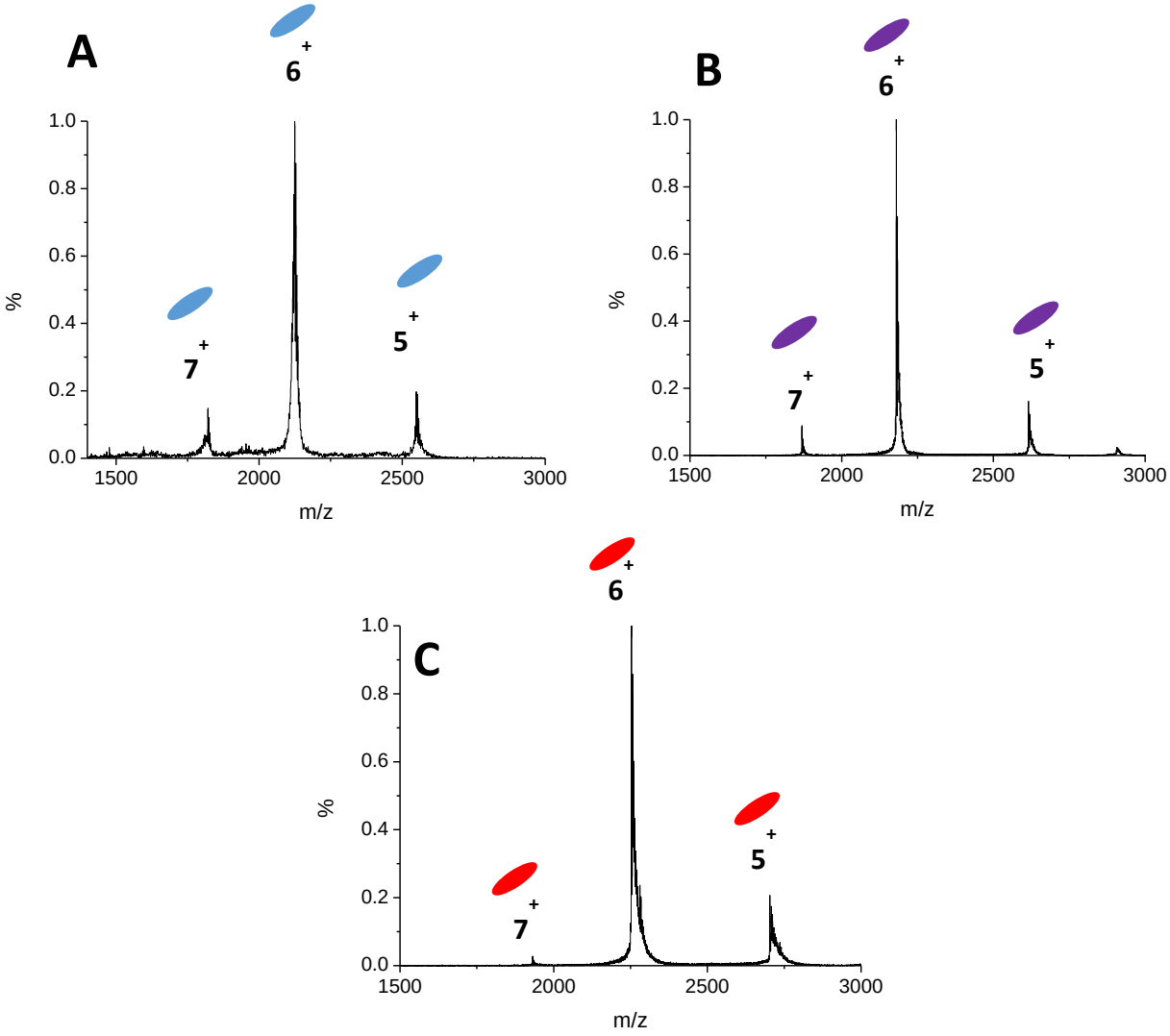


Figure 7: On the left panel are collision-induced unfolding plots of (A) the PAI-1 wildtype and the PAI-1 variants; (B) stab, (C) W175F, (D) P14 and (E) latent. On the right panel the corresponding mass spectra at trap 10 V and 100 V.

5.4 Characterising nanobodies against PAI-1

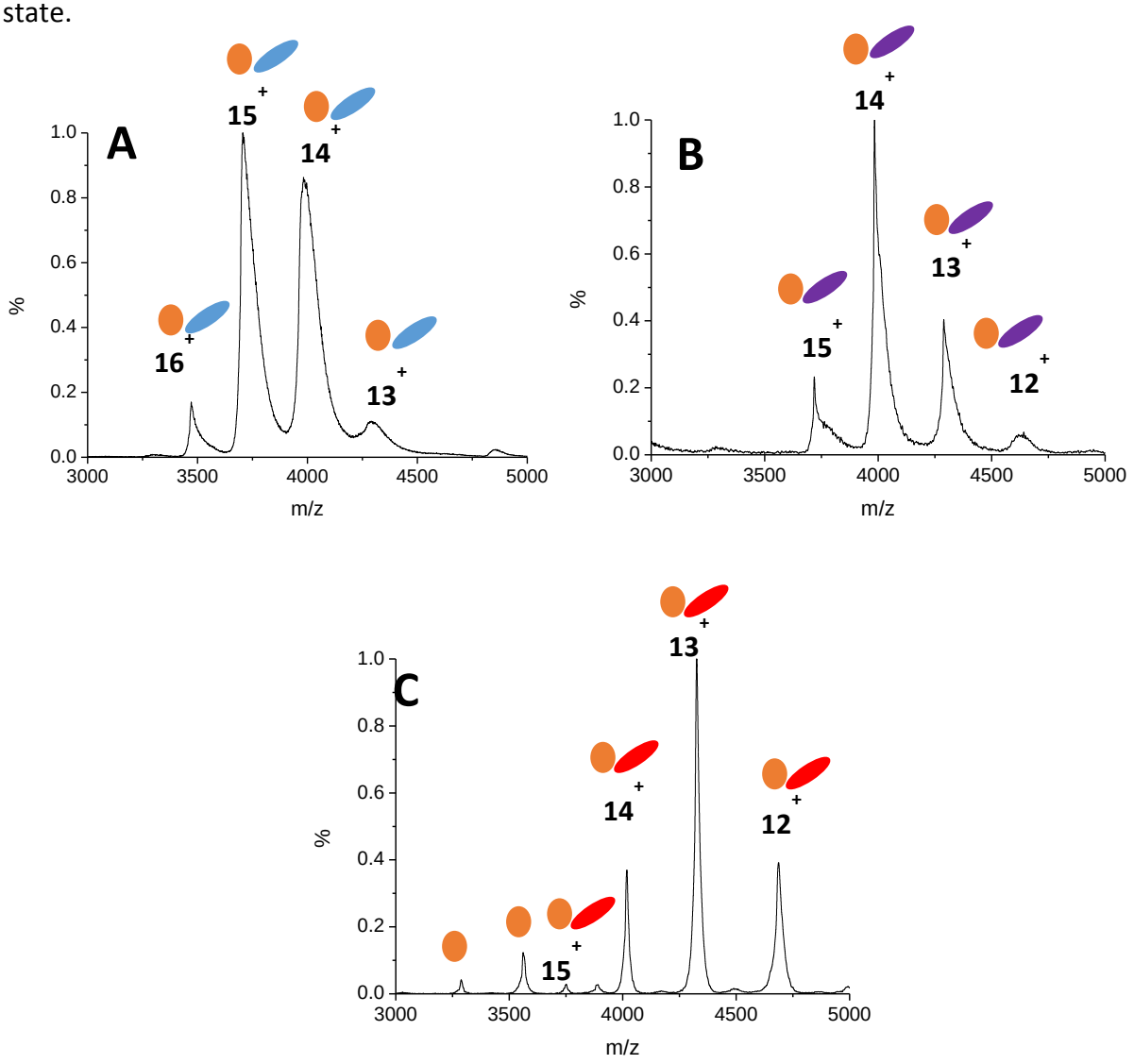
The three studied anti-PAI-1 nanobodies displayed similar native mass spectra with charge states ranging from 5^+ to 7^+ with the most intense peak corresponding to the 6^+ charge state (**Fig. 8**). The experimental and theoretical masses of the nanobodies are listed in **Figure 8**.



	Protein	Experimental mass (Da)	Theoretical mass (Da)
A	VHH-2g-42	12,742.6	12,739.29
B	VHH-2w-64	13,078.0	13,077.44
C	VHH-2-12	13,514.7	13,515.79

Figure 8: Native mass spectra of nanobodies (A) VHH-2g-42, (B) VHH-2w-64, (C) VHH-2-12.

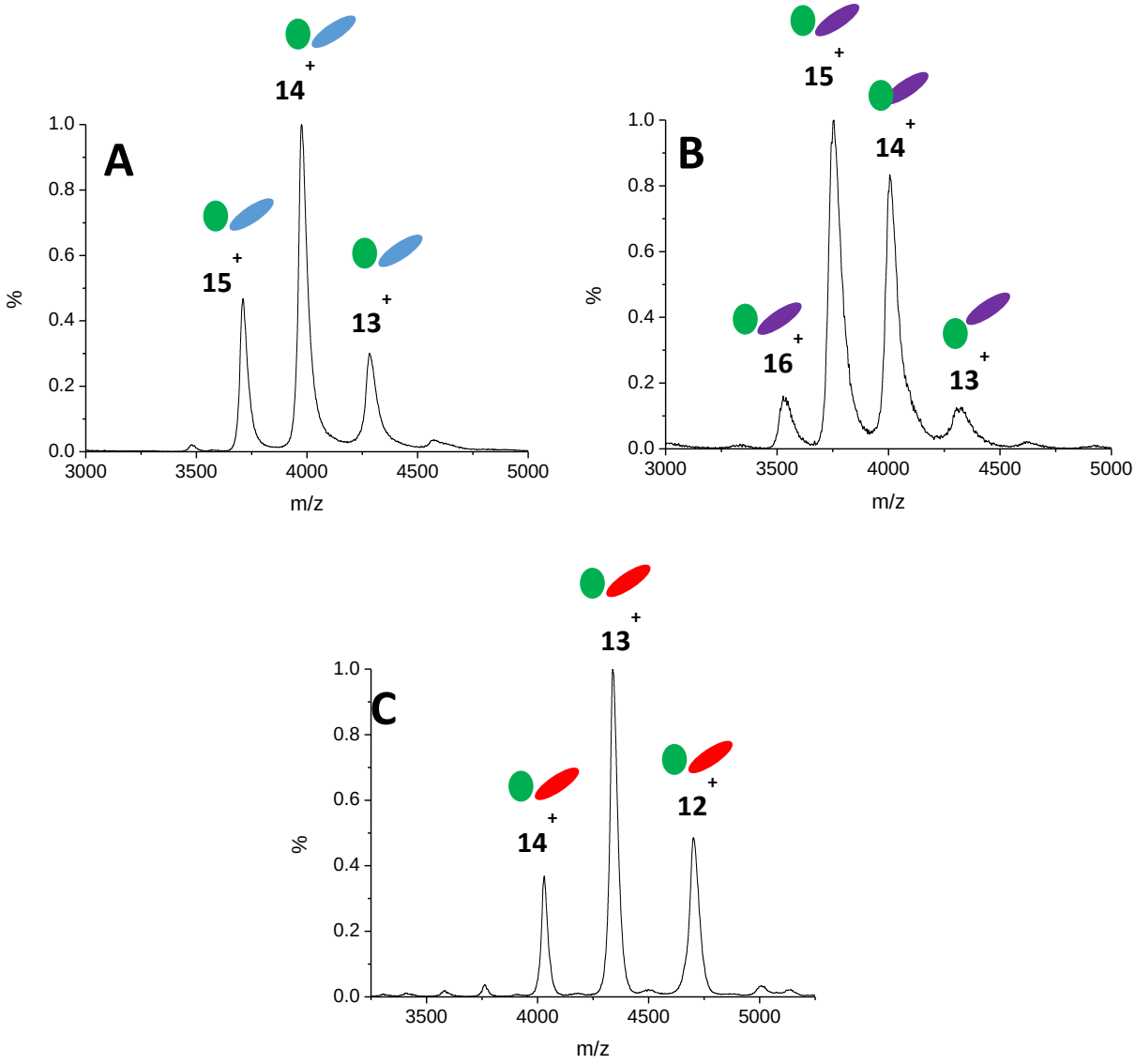
When the nanobodies are in a 1:1 complex with the stabilised active form of PAI-1 (PAI-1 Stab), their respective mass spectra show slight differences in their charge state distribution (**Fig.9**). VHH-2g-42 in complex with PAI-1 Stab has an experimental mass of 55,587.50 with a charge state distribution ranging from 13^+ to 16^+ with 14^+ and 15^+ corresponding with the most intense peaks. VHH-2w-64, which induces a substrate behaviour of PAI-1, has an experimental mass of 55,822.22 Da with a charge state distribution ranging from 12^+ to 15^+ with the most intense peak corresponding to the 14^+ charge state. VHH-2-12, which induces an acceleration of the active-to latent transition of PAI-1, shows an experimental mass of 56,243.50 Da with charge states ranging from 12^+ to 15^+ with the most intense peak corresponding to the 13^+ charge state.



	Protein	Experimental mass (Da)	Theoretical mass (Da)
A	PAI-1 Stab + VHH-2g-42	55,587.5	55,422.43
B	PAI-1 Stab + VHH-2w-64	55,822.2	55,760.58
C	PAI-1 Stab + VHH-2-12	56,243.5	56,198.93

Figure 9: Native mass spectra of PAI-1 Stab complexed with nanobody (A) VHH-2g-42, (B) VHH-2w-64, and (C) VHH-2-12.

Figure 10 shows the native mass spectra of nanobodies; VHH-2g-42, VHH-2w-64 and VHH-2-12, in their respective complex with the stabilized active form of PAI-1 (PAI-1 W175F). Here a slight difference in charge state distribution is also seen. VHH-2g-42 in complex with PAI-1 W175F has an experimental mass of 55,739.3 Da with charge states ranging from 13^+ to 15^+. VHH-2w-64 in complex with PAI-1 W175F has an experimental mass of 55,916.8 Da with charge states 13^+ to 16^+. VHH-2-12 in complex with PAI-1 W175F has an experimental mass of 56,419.5 Da with charge states 12^+ to 14^+.



	Protein	Experimental mass (Da)	Theoretical mass (Da)
A	PAI-1 W175F + VHH-2g-42	55,739.3	55,451.44
B	PAI-1 W175F + VHH-2w-64	55,916.8	55,789.59
C	PAI-1 W175F + VHH-2-12	56,419.5	56,227.94

Figure 10: Native mass spectra of (A) PAI-1 W175F complexed with nanobody VHH-2g-42. (B) PAI-1 W175F complexed with nanobody VHH-2w-64 (C) PAI-1 W175F complexed with nanobody VHH-2-12.

With ion mobility, the collision cross sections of the proteins can be determined. This gives another dimension to the observed data as the size and shape of the protein can be monitored.

Tables 6-8 show the CCS values of the three studied nanobodies and of the PAI-nanobody complexes. Assuming the most compact and native-like state would be the lowest significant charge state, the corresponding CCS values can be selected for comparison between different proteins and protein complexes.

When comparing the lowest charge state of each nanobody, which is 5^+, the CCS values are ranging from 1185 $Å^2$ for VHH-2g-42, 1569 $Å^2$ for VHH-2w-64 and 1595 $Å^2$ for VHH-2-12. The interesting aspect of this experiment is to compare the size of the complexes that these nanobodies form with PAI-1. For PAI-1 Stab complexed with the nanobody, the lowest common charge state is 13^+. For this charge state the CCS value for PAI-1 Stab + VHH-2g-42 is 4538 $Å^2$, for PAI-1 Stab + VHH-2w-64 it is 4666 $Å^2$ and for PAI-Stab + VHH-2-12 4441 $Å^2$. When comparing these CCS values to that of PAI-1 Stab, the complex increases in size by 29.49 % with VHH-2g-42 binding, 33.16 % with VHH-2w-64 and 26.74 % with VHH-2-12. As VHH-2-12 induces the latency transition of PAI-1, it makes sense that the PAI-1-nanobody complex would have a more compact CCS value as the reactive centre loop (RCL) of PAI-1 is then buried in the protein structure. VHH-2w-64 induces substrate-like behaviour of PAI-1, in which the RCL would also be buried in the protein. How VHH-2g-42 would prevent the interaction of PAI-1 with t-PA is still unknown. But the CCS value of the PAI-1 Stab + VHH-2g-42 complex shows a relative increase that lies between the PAI-1 P14 + nanobody complex and the latent PAI-1 + nanobody complex.

When the same CCS values are compared for the nanobody complexed with PAI-1 W175F, a stabilized active form of PAI-1, a slightly different trend was observed. The increase of the CCS values for the 13^+ PAI-1 + nanobody complex compared to PAI-1 W175F is 21.70 % for VHH-2g-42, 30.88 % for VHH-2w-64 and 23.54 % for VHH-2-12. In this case VHH-2-12 again shows a smaller change in CCS, as it induces a latency transition. VHH-2w-64 shows a relative larger change in CCS as it induces a substrate behaviour of PAI-1. VHH-2g-42 shows an even smaller change in CCS compared to VHH-2-12 when compared to the active PAI-1 W175F.

Even though PAI-1 Stab and PAI-1 W175F are both stabilized active forms of PAI-1, the relative increase in CCS between PAI-1 Stab and PAI-1 Stab + nanobody is slightly larger compared to the complex with PAI-1 W175F. This could suggest a different behavior of the mutant when it is bound to the nanobody, due to the difference in mutations.

Table 6: CCS values of the nanobody VHH-2g-42, different PAI-1 forms their nanobody complexes

CCS	VHH-2g-42	PAI-1 Stab	PAI-1 Stab +VHH-2g-42	PAI-1 W175F	PAI-1 W175F +VHH-2g-42
5^+	1185 ± 7 Å^2				
6^+	1479 ± 9 Å^2				
7^+	1871 ± 8 Å^2				
12^+		3469 ± 39 Å^2		3436 ± 43 Å^2	
13^+		3504 ± 51 Å^2	4538 ± 106 Å^2	3541 ± 43 Å^2	4309 ± 65 Å^2
14^+			4425 ± 109 Å^2		4276 ± 58 Å^2
15^+			4420 ± 95 Å^2		4379 ± 59 Å^2
16^+			4821 ± 110 Å^2		
MW	12,742.6 Da	42,785.6 Da	55,587.5 Da	42,839.5 Da	55,739.3 Da

Table 7: CCS values of the nanobody VHH-2w-64, different PAI-1 forms their nanobody complexes

CCS	VHH-2w-64	PAI-1 Stab	PAI-1 Stab +VHH-2w-64	PAI-1 W175F	PAI-1 W175F +VHH-2w-64
5^+	1569 ± 7 Å^2				
6^+	1803 ± 7 Å^2				
7^+	2067 ± 7 Å^2				
12^+		3469 ± 39 Å^2	4680 ± 80 Å^2	3436 ± 43 Å^2	
13^+		3504 ± 51 Å^2	4666 ± 40 Å^2	3541 ± 43 Å^2	4635 ± 98 Å^2
14^+			4677 ± 54 Å^2		4569 ± 88 Å^2
15^+			4700 ± 48 Å^2		4622 ± 107 Å^2
16^+					4672 ± 114 Å^2
MW	13,078.0 Da	42,785.6 Da	55,822.2 Da	42,839.5 Da	55,916.8 Da

Table 8: CCS values of the nanobody VHH-2-12, PAI-1 and the PAI-1-nanobody complex

CCS	VHH-2-12	PAI-1 Stab	PAI-1 Stab +VHH-2-12	PAI-1 W175F	PAI-1 W175F +VHH-2-12
5^+	1596 ± 9 Å^2				
6^+	1603 ± 8 Å^2				
12^+		3469 ± 39 Å^2	4482 ± 19 Å^2	3436 ± 43 Å^2	4395 ± 93 Å^2
13^+		3504 ± 51 Å^2	4441 ± 17 Å^2	3541 ± 43 Å^2	4375 ± 30 Å^2
14^+			4497 ± 8 Å^2		4424 ± 77 Å^2
15^+			4500 ± 8 Å^2		
MW	13,514.7 Da	42,785.6 Da	56,243.5 Da	42,839.5 Da	56,419.5 Da

Figures 11 to **13** show the experimental CCS values plotted in function of the corresponding molecular weight. The plotted trendline is based on the CCS values of globular calibrants (open circles). The CCS values of the nanobodies and PAI-1 are following this predicted trend. The PAI-1 + nanobody complex rejects this trend with a 12-21 % increase in CCS.

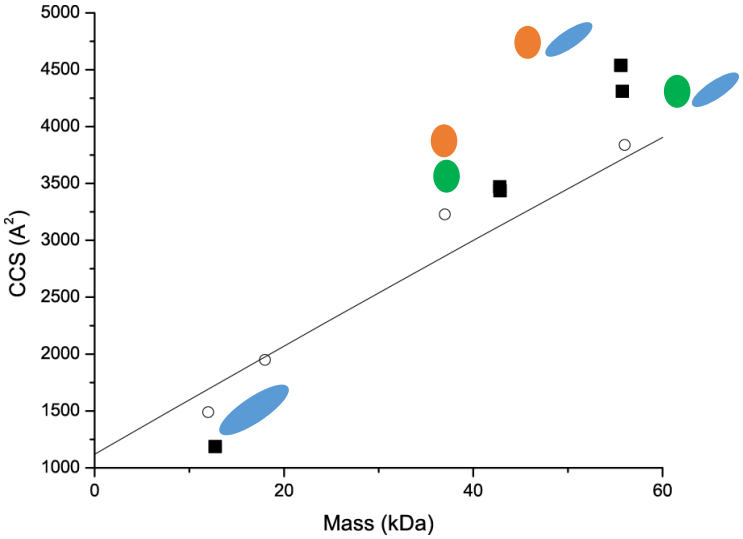


Figure 11: Experimental CCS values (■) of the low-charge states of nanobody VHH-2g-42, PAI-1 and PAI-1-nanobody complex

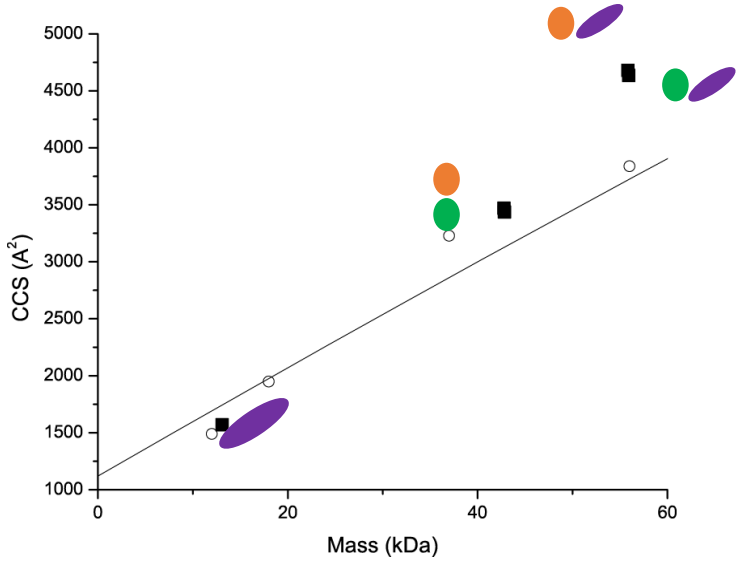


Figure 12: Experimental CCS values (■) of the low-charge states of nanobody VHH-2w-64, PAI-1 and PAI-1-nanobody complex

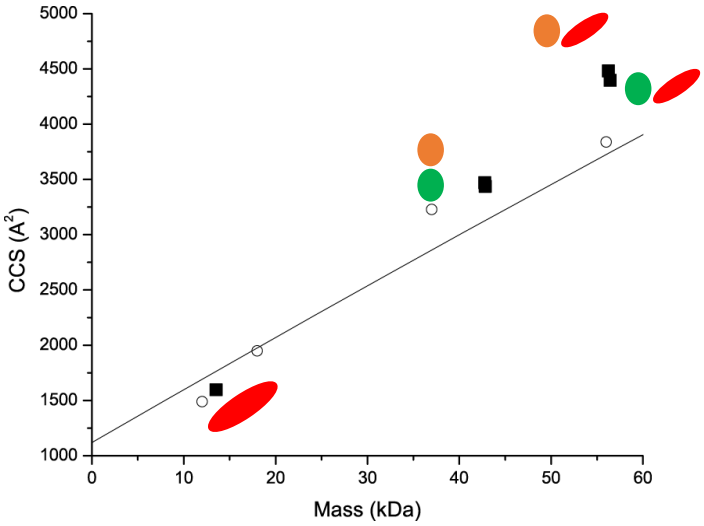
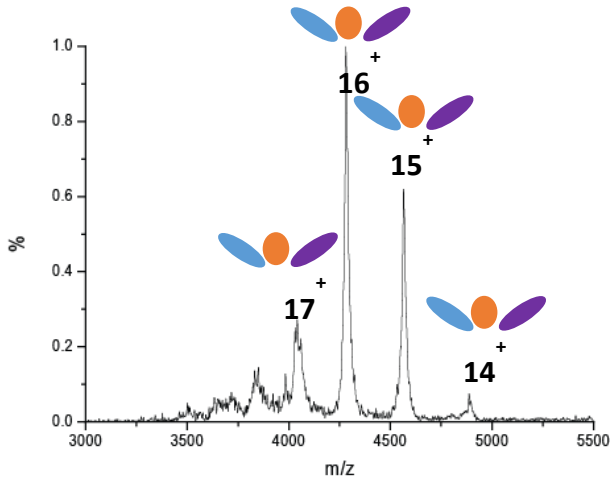


Figure 13: Experimental CCS values (■) of the low-charge states of nanobody VHH-2-12, PAI-1 and PAI-1-nanobody complex

In a separate experiment, PAI-1 Stab was complexed with VHH-2g-42 and VHH-2w-64 at the same time. **Figure 14** shows the mass spectrum of this complex, which has a subunit ratio of 1:1:1 and an experimental mass of 68,483.01 Da with charge states 14^+ to 17^+. The CCS values of this PAI-1 + nanobody complex are displayed in **Table 10**. The CCS value of the 14^+ charge state of this complex can be compared to those of the respective 1:1 PAI-1 + nanobody complexes. The CCS value of PAI-1 Stab + VHH-2g-42 + VHH-2w-64 is 5,160.71 Å^2, while that of PAI-1 Stab + VHH-2g-42 is 4,425.09 Å^2 and that of PAI-1 Stab + VHH-2w-64 is 4,666.93 Å^2. Compared to the 14^+ charge state of PAI-1 Stab + VHH-2g-42 there is an increase in CCS value of 16.62 %. Compared to the CCS value of PAI-1 Stab + VHH-2w-64 there is an increase of 10.58 % in CCS value. **Figure 15** shows that the measured CCS value of the $(VHH)_2$ + PAI-1 complex evades the predicted CCS value for a globular complex with 15 %.



Protein	Experimental mass (Da)	Theoretical mass (Da)
PAI-1 Stab + VHH-2g-42 + VHH-2w-64	68,483.0	68,481.86

Figure 14: Native mass spectrum of PAI-1 stab complexed with both nanobodies VHH-2g-42 and VHH-2w-64 in a 1:1:1 stoichiometry.

Table 9: CCS values of the nanobodies VHH-2-12, VHH-2w-64, PAI-1 Stab and the PAI-1-nanobody complex

CCS	VHH-2-12	VHH-2w-64	PAI-1 Stab	PAI-1 Stab +VHH-2g-42 +VHH-2w-64
5^+	1596 ± 9 Å^2	1569 ± 7 Å^2		
6^+	1603 ± 8 Å^2	1803 ± 7 Å^2		
7^+		2067 ± 7 Å^2		
12^+			3469 ± 39 Å^2	
13^+			3504 ± 51 Å^2	
14^+				5160 ± 21 Å^2
15^+				5161 ± 18 Å^2
16^+				5179 ± 21 Å^2
17^+				5192 ± 46 Å^2
MW	13,514.6 Da	13,078.0 Da	42,785.6 Da	68,483.0 Da

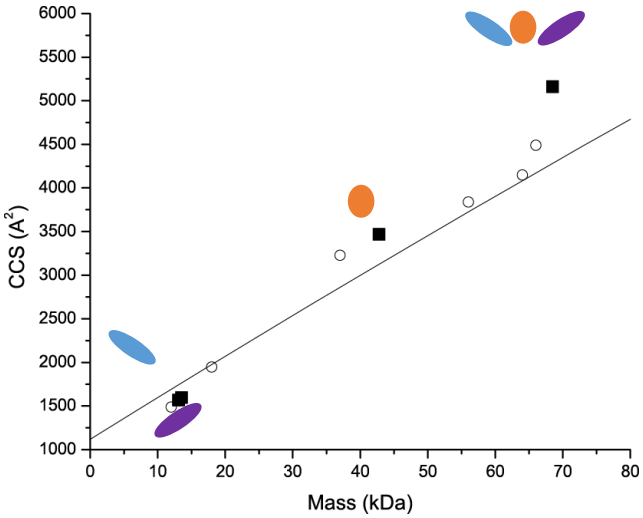


Figure 15: Experimental CCS values of the low-charge states of nanobodies VHH-2-12 and VHH-2w-64, PAI-1 and the PAI-1-nanobody complex

5.4.1 Unfolding pattern and stability of nanobodies

The three studied anti-PAI-1 antibodies displayed slightly different CIU behaviour (**Fig. 16**). For VHH-2g-42, the drift time at 10 V is at 9.66 ± 0.27 ms and has a less intense peak at 13.30 ± 0.55 ms (**Fig. 16A**). At 20 V VHH-2g-42 shows a transition to 10.21 ± 0.46 ms and 13.67 ± 0.73 ms. After 30 V the conformation at 13.67 ms stays stable till 100 V.
VHH-2w-64 displays a drift time of 14.76 ± 0.55 ms at 10 V with a less intense peak at 13.12 ± 0.55 ms which disappears at 30 V (**Fig 16B**). The peak at 14.76 ms remains stable till 100 V. VHH-2-12 has a drift time of 10.94 ± 0.20 ms with a less intense peak at 14.40 ± 0.37 ms at 10 V (**Fig 16C**). However, this distribution of intensity changes at 30 V, where the peak at 14.40 ms is slightly more intense than the peak at 10.94 ms. The three studied nanobodies show slightly different transitions but are overall flexible as their ion mobility peaks are broad and for VHH-2-12 there seem to be two conformations present.

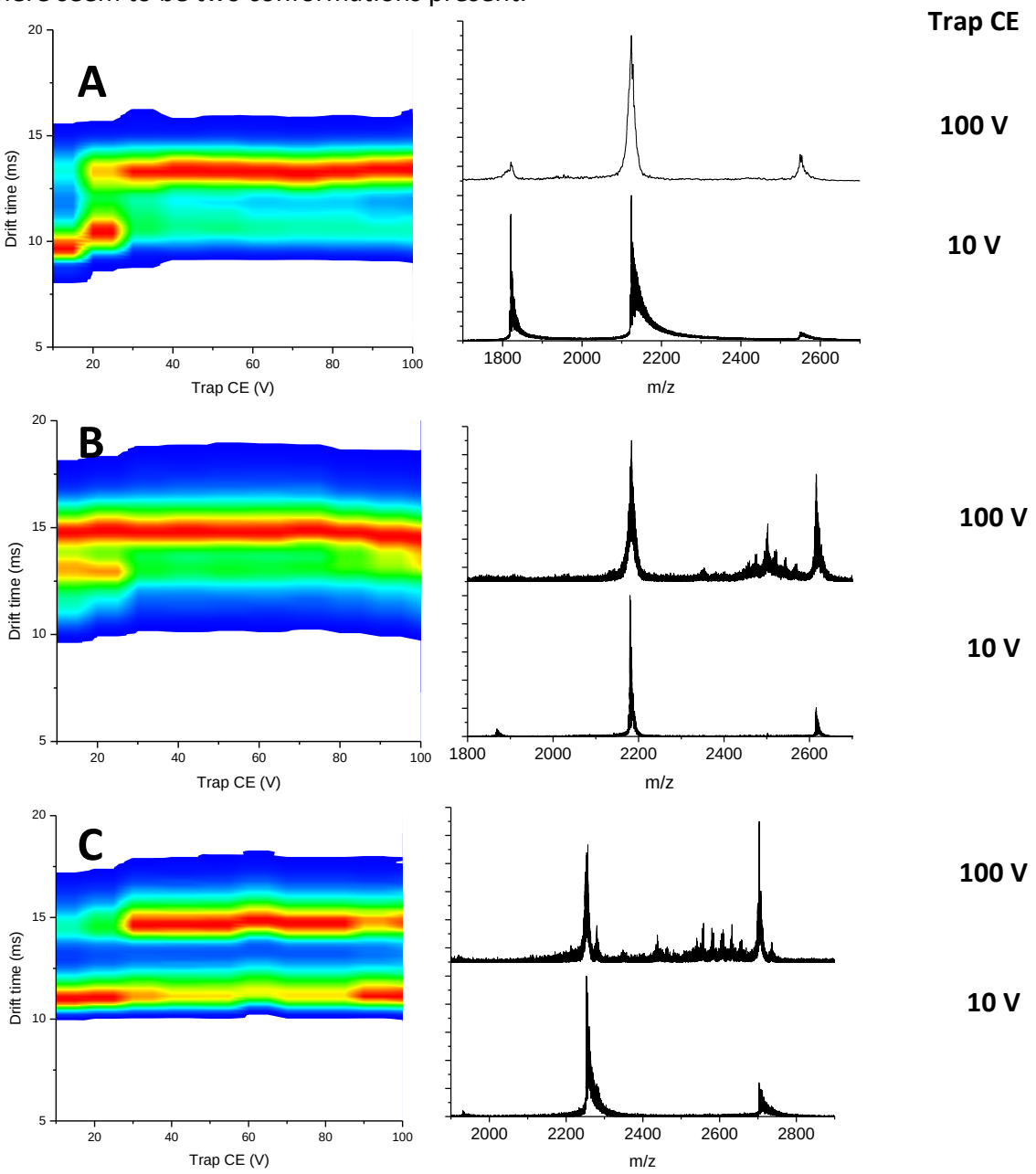


Figure 16: Collision-induced unfolding plots of nanobodies, (A) VHH-2g-42, (B) VHH-2w-64, (C) VHH2-12

5.4.2 Influence of nanobodies on stability of PAI-1 complex

When the studied nanobodies were bound to PAI-1, each complex showed a different unfolding pattern as displayed in the CIU plots in **Figure 17**. The VHH-2w-64 + PAI-1 Stab complex shows a first transition at 30 V, whereas VHH-2g-42 + PAI-1 Stab and VHH-2-12 + PAI-1 Stab show their first transition at 40 V. The VHH-2g-42 + PAI-1 Stab then displays a transition every 10 V until it reaches 90 V where it stays constant. VHH-2w-64 + PAI-1 Stab and VHH-2-12 + PAI-1 Stab show transitions every 10 V from 40 V to 60 V after which it stays stable. VHH-2-12 induces the latent form of PAI-1 and VHH-2w-64 induces the substrate form of PAI-1. How exactly VHH-2g-42 inhibits PAI-1 is not known, but the unfolding pattern of this complex is different compared to the other two and shows more stability as it reaches the most unfolded observed state at a higher voltage. The least stable would be the VHH-2w-64 + PAI-1 Stab complex which starts to transition at relative lower voltages.

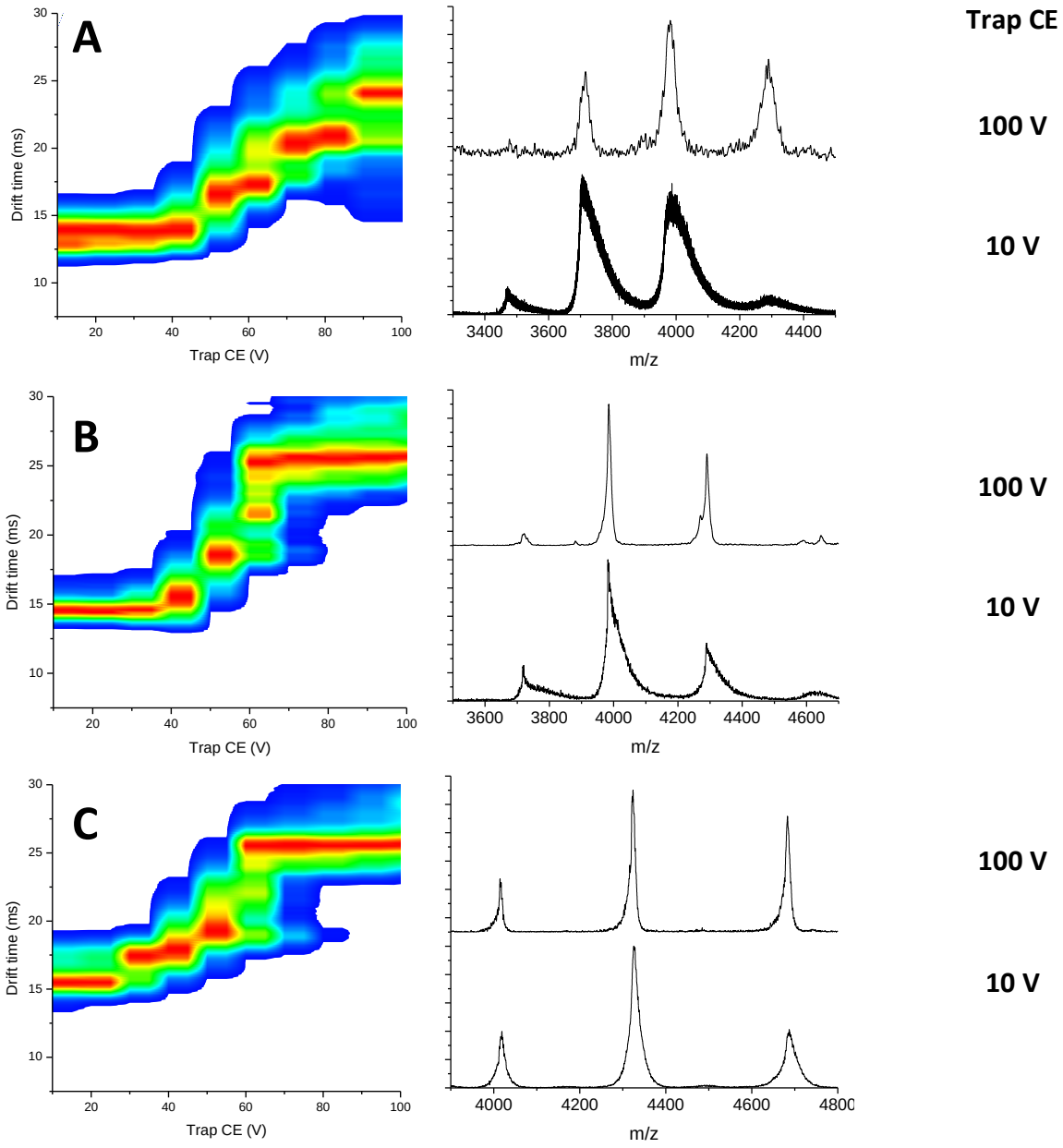
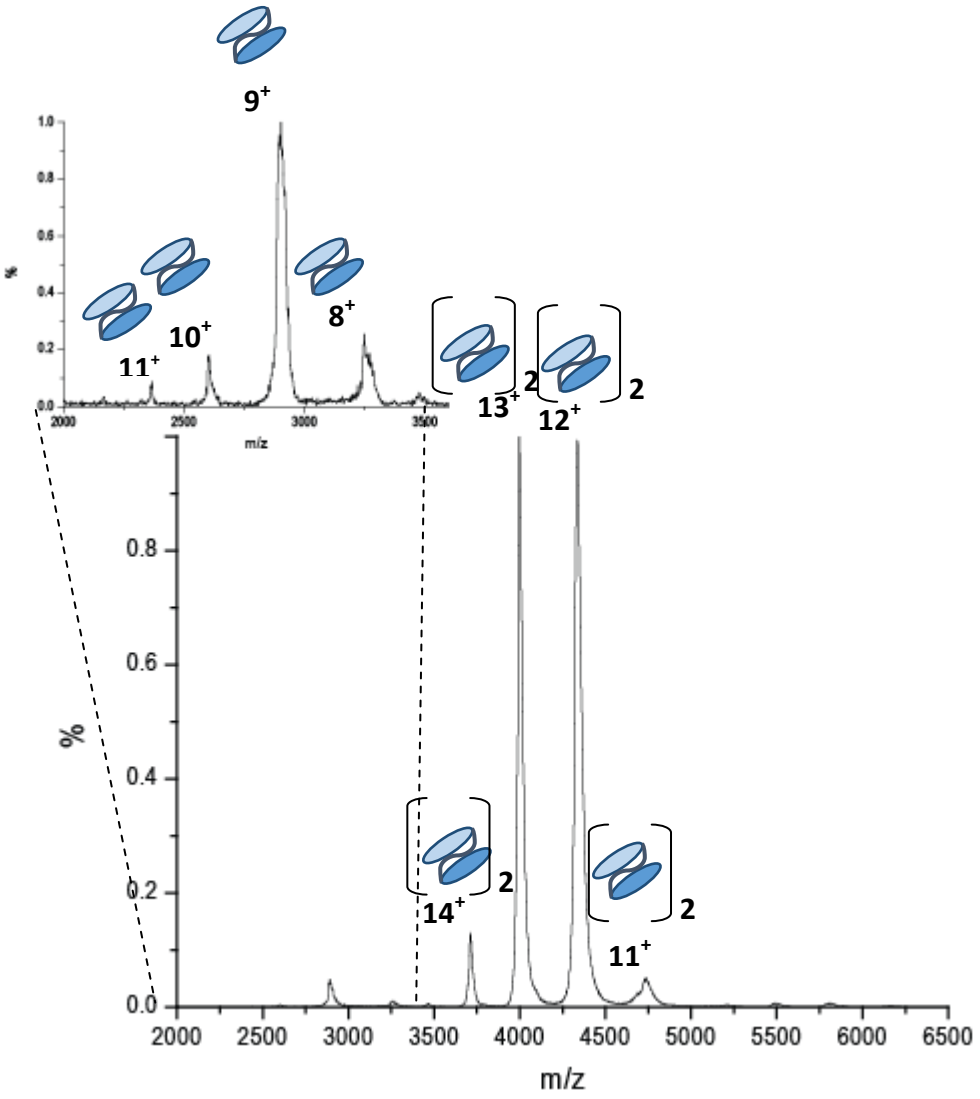


Figure 17: Collision-induced unfolding plots (A) VHH-2g-42 + stab, (B) VHH-2w-64 + stab, (C) VHH-2-12 + stab

5.5 Characterising an scFv against PAI-1

ScFV-33H1f7 is a single-chain variable fragment that inhibits PAI-1 by inducing substrate behaviour. **Figure 18** shows the mass spectrum of scFV-33H1F7 indicating that it is observed as a dimer of 51,947.00 Da with charge states 11^+ to 14^+. Small amounts of monomer were also observed with an experimental mass of 25,996.3 Da. It is believed that this dimerisation of scFv is due to the solution conditions of the experiment; pH 5.5 and 150 mM ammonium acetate. Dimers can be trapped by factors which stabilise the VH-VL interface, such as high ionic strength and a pH below 7.5 [32]. The presence of the antigen is also a factor that can stabilise the scFv dimers [32].

To confirm that the dimer is not due to a possibly too high protein concentration used during this experiment, the scFv was measured at lower concentration up to 350 nM (**Fig. 19**) Even at a protein concentration of 350 nM the scFv was observed as a dimer.



Protein	Experimental mass (Da)	Theoretical mass (Da)
ScFv-33H1F7 monomer	25,996.3	25,927.40
ScFv-33H1F7 dimer	51,947.0	51,854.80

Figure 18: Native mass spectrum of scFv-33H1F7. The inset shows the mass spectrum zoomed in at 2000-3500 m/z.

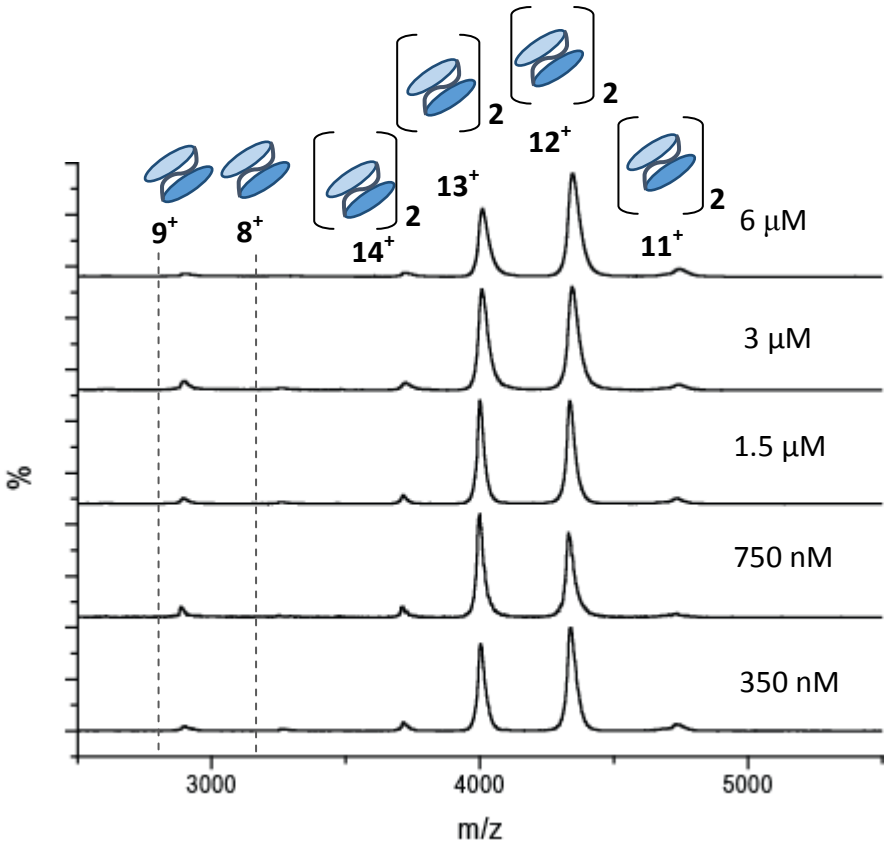
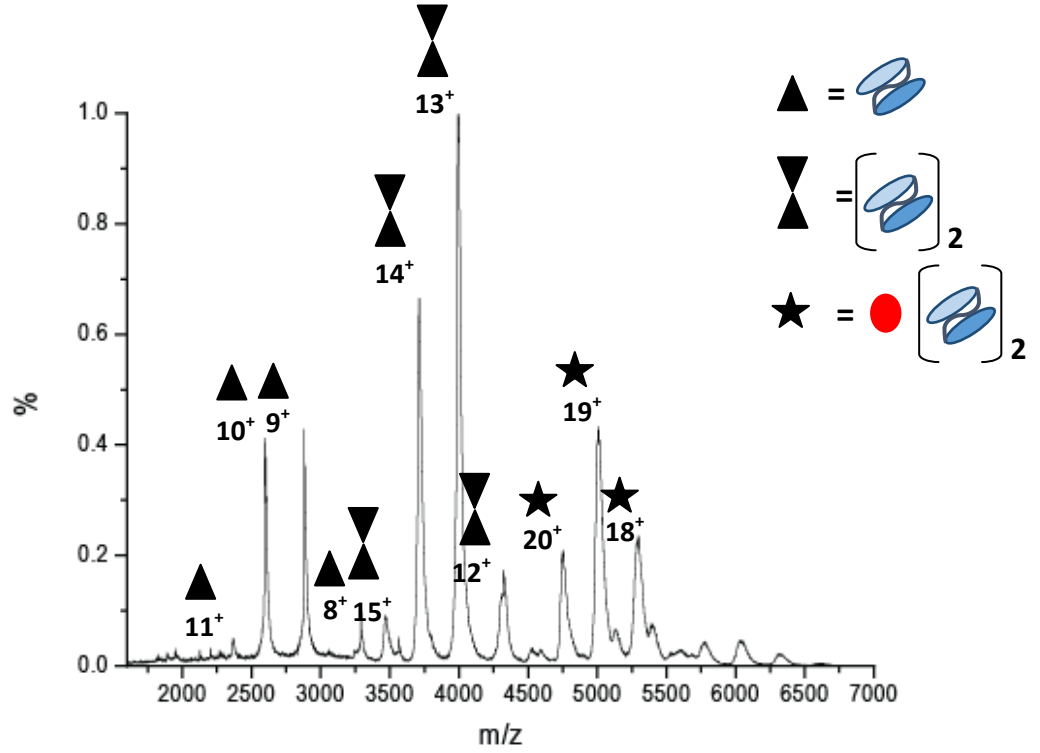


Figure 19: Native mass spectra of scFv-33H1F7 at a range of protein concentrations.

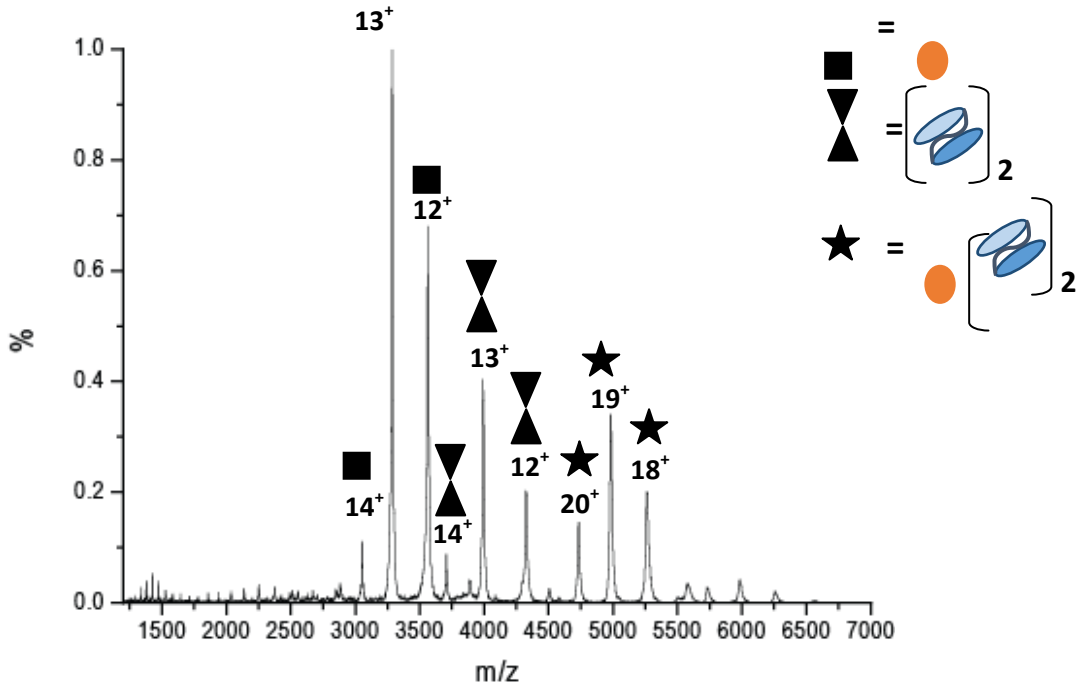
ScFv-33H1F7 was measured in complex with PAI-1 Stab and with PAI-1 Latent to check the complex stoichiometry with native mass spectrometry and the collision cross sections with native ion mobility. When scFv-33H1F7 interacts with latent PAI-1 (**Fig. 20**), the complex consists of a scFv dimer and a PAI-1 monomer and has an experimental mass of 94,787.7 Da. Next to this complex there is still some scFv monomer and scFv dimer observed but no free PAI-1 Latent. There is also a small presence of the complex with scFv dimer and PAI-1 dimer detected in the range of 5000-6500 m/z.
When scFv interacts with PAI-1 Stab (**Fig. 21**), the stoichiometry of the complex also indicates a scFv dimer binding a PAI-1 stab monomer, with an experimental mass of 94,787.1 Da. In this case there was still free PAI-1 stab and free scFv dimer observed in the mass spectrum, as well as the presence of a complex consisting of scFv dimer and PAI-1 dimer in the range of 5500-6500 m/z.

Table 11 displays an overview of the CCS values calculated for the scFV monomer, scFv dimer, and the scFv-PAI-1 complex with PAI-1 Stab and with PAI-1 Latent. Selecting the CCS values of the lowest charge states of the proteins; the scFv monomer has a CCS value of 2337.2 $Å^2$; the scFv dimer 3813 $Å^2$; scFV dimer + PAI-1 Stab 5919 $Å^2$ and scFv dimer + PAI Latent a value of 6192 $Å^2$. The CCS values for the PAI-1 Stab complex and the PAI-1 Latent complex are similar with a 4 % difference in size. Although scFv-33H1F7 induces substrate behaviour of active PAI-1, the latent PAI-1 would possibly not make this transition.



	Protein	Experimental mass (Da)	Theoretical mass (Da)
▲	ScFv-33H1F7 monomer	25,998.2	25,927.40
⧗	ScFv-33H1F7 dimer	51,897.0	51,854.80
★	ScFv-33H1F7 dimer + PAI-1 Latent	94,787.7	94,624.00

Figure 20: Native mass spectrum of latent PAI-1 complexed with scFv-33H1F7.



	Protein	Experimental mass (Da)	Theoretical mass (Da)
■	Pai-1 Stab	42,716.8	42,701.10
●	ScFv-33H1F7 dimer	51,994.5	51,854.80
★	ScFv-33H1F7 dimer + Pai-1 Stab	94,787.1	94,538.00

Figure 21: Native mass spectrum of PAI-1-Stab complexed with scFv-33H1F7.

Table 10: CCS values of ScFv-33H1F7, PAI-1 Stab, PAI-1 Latent and the PAI-1-scFv complexes

CCS	ScFv monomer	ScFv dimer	PAI-1 Stab	PAI-1 Stab +ScFv dimer	PAI-1 Latent	PAI-1 Latent +ScFv dimer
8^+	2337 ± 47 $Å^2$					
9^+	2374 ± 48 $Å^2$					
10^+	2400 ± 53 $Å^2$					
11^+		3812 ± 81 $Å^2$				
12^+		3696 ± 49 $Å^2$	3469 ± 39 $Å^2$		3335 ± 70 $Å^2$	
13^+		3650 ± 60 $Å^2$	3504 ± 51 $Å^2$		3391 ± 75 $Å^2$	
14^+		3690 ±44 $Å^2$				
18^+				5918 ± 65 $Å^2$		6191 ± 85 $Å^2$
19^+				5899 ±52 $Å^2$		6148 ± 119 $Å^2$
20^+				5996 ± 54 $Å^2$		6835 ± 118 $Å^2$
MW	25,998.2 Da	51,994.5 Da	42,716.8 Da	94,787.1 Da	42,779.5 Da	94,787.7 Da

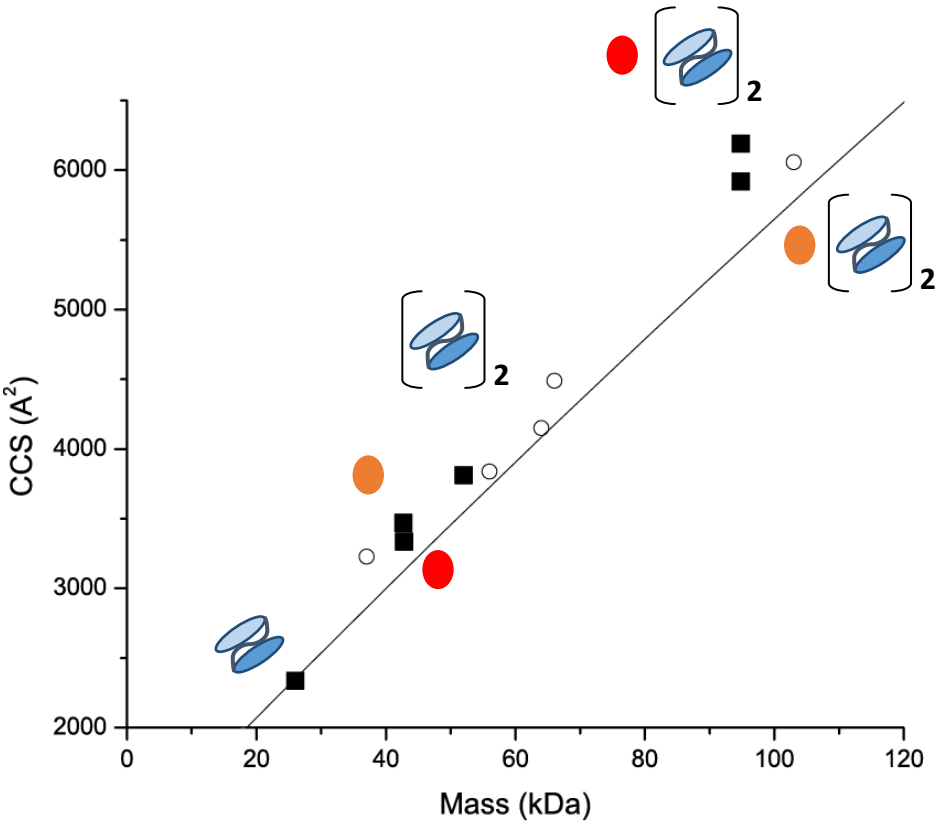


Figure 22: Experimental CCS values of the low-charge states of ScFv-33H1F7, PAI-1 Stab, PAI-1 Latent and the PAI-1-scFv complexes

5.5.1 The influence of scFv-33H1F7 on PAI-1 complex stability

The unfolding plot of the scFv-33H1F7 dimer shows a transition at 40 V and another one at 50 V (**Fig. 23A**). When the scFv binds to PAI-1 the unfolding plot differs for PAI-1 Stab and PAI-1 Latent (**Fig 23B &C**). The scFv+PAI-1 Stab complex displays clear transitions at 50 V, 60V and at 100 V and stays stable till 150 V. On the other hand, the scFv +PAI-1 Latent complex shows a slight transition at 60 V and a transition at 100 V after which the complex is lost, which could mean that this complex is relatively more unstable. The scFv-33H1F7 would induce substrate behaviour of PAI-1 stab. However, the binding of scFv-33H1F7 to latent PAI-1 results in a relatively unstable complex.

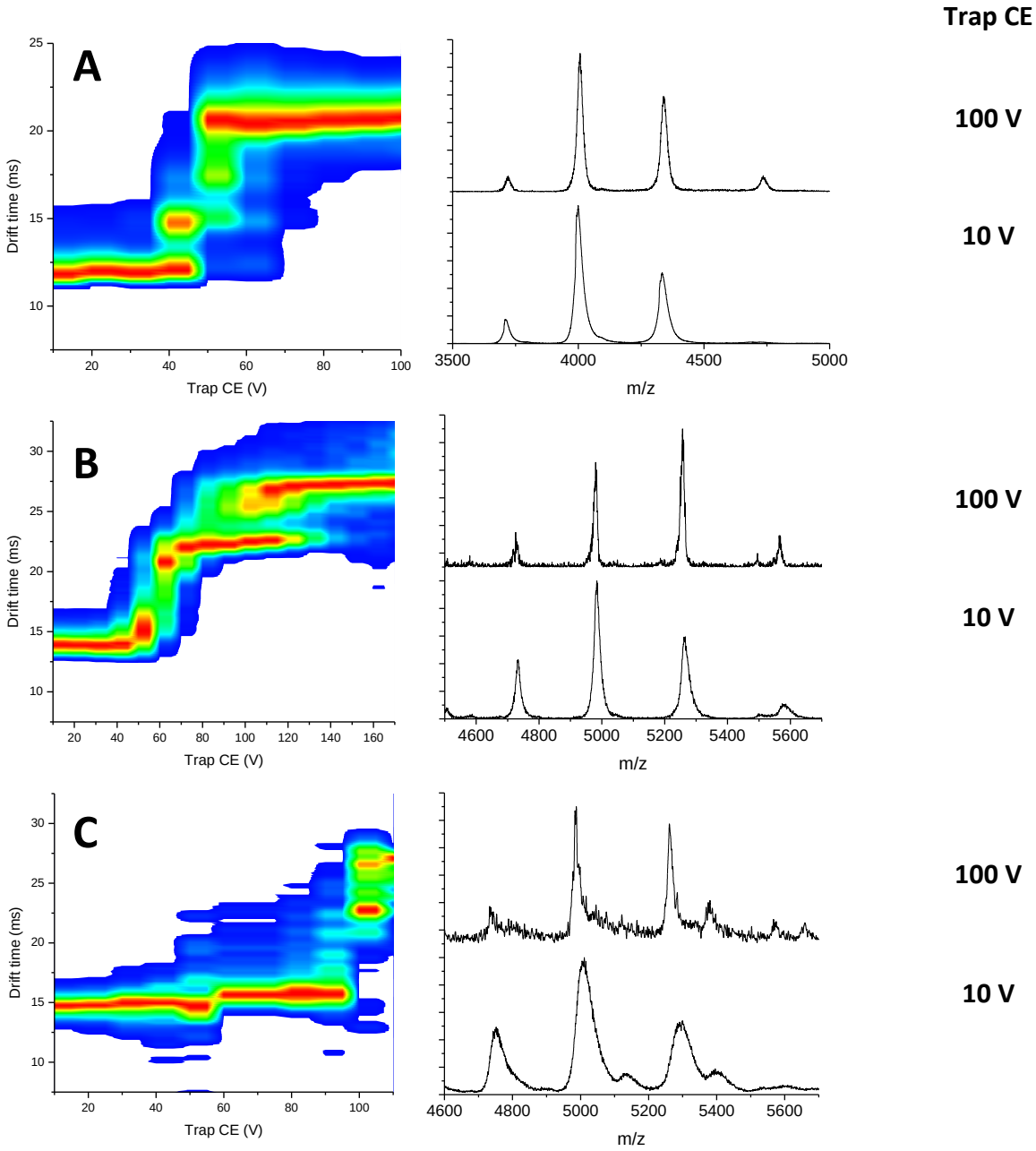


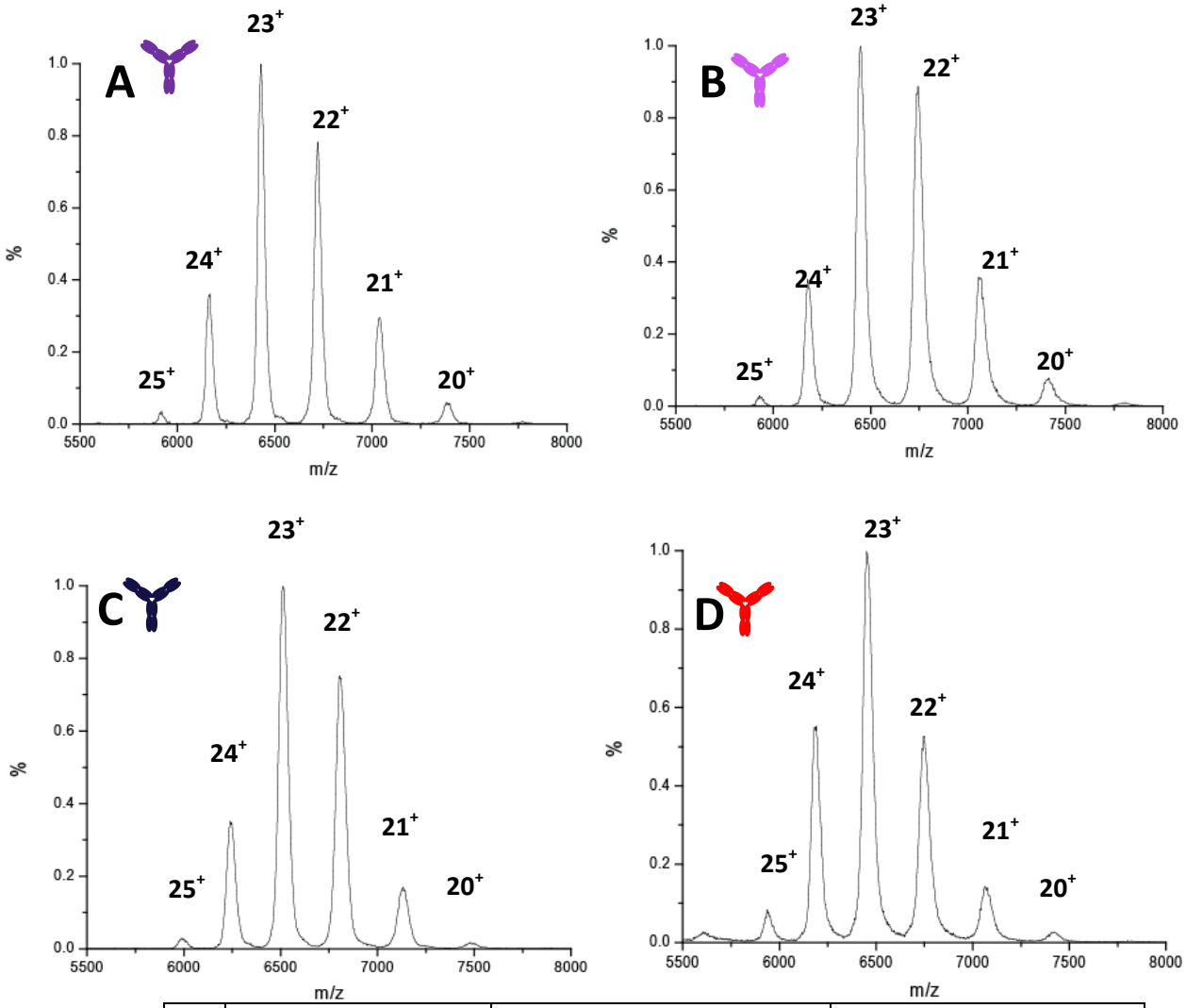
Figure 23: Collision-induced unfolding plots (A) ScFv-33H1F7, (B) ScFv-33H1F7 + PAI-1 Stab, (C) ScFv-33H1F7 + PAI-1 Latent

5.6 Monoclonal antibodies against PAI-1

Four anti-PAI-1 monoclonal antibodies were analysed with native IM-MS. Mab-33H1F7, Mab-55F4C12 and Mab-8H9D4 induce substrate behaviour of PAI-1 [20] and Mab-33B8 induces the acceleration of the active-to-latent conversion of PAI-1 [21].

The Mabs were measured separately and in complex with PAI-1 Stab and with PAI-1 W175F.

Figure 24 displays the native mass spectra of the four studied anti-PAI-1 antibodies. The same charge state distribution was observed for the four antibodies (20^+ to 25^+), with the most intense peak corresponding to the 23^+ charge state.

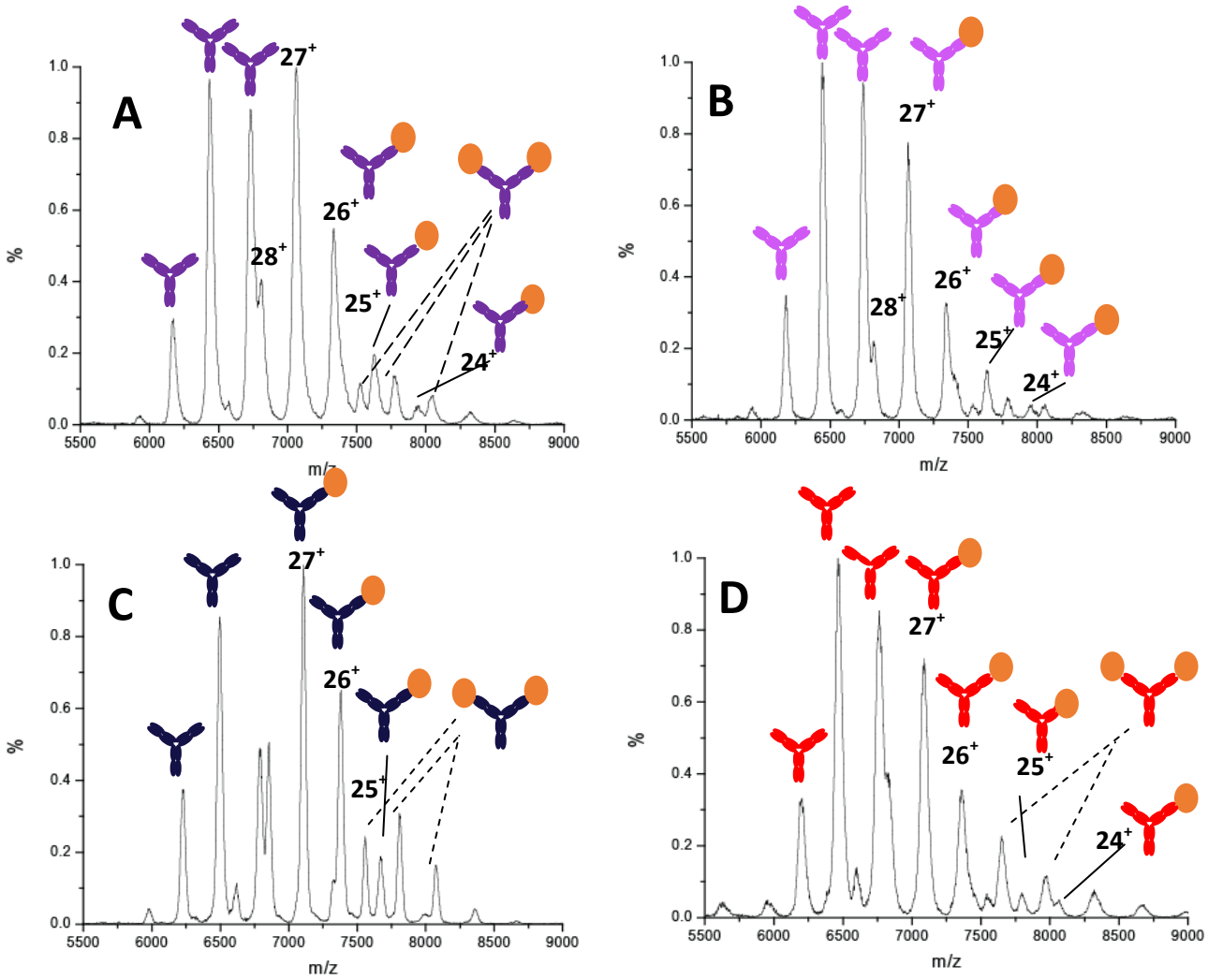


	Protein	Experimental mass (Da)	Theoretical mass
A	**Mab-33H1F7**	147,869.2	148 kDa
B	**Mab-55F4C12**	148,267.4	148 kDa
C	**Mab-8H9D4**	148,309.2	148 kDa
D	**Mab-33B8**	148,981.2	148 kDa

Figure 24: Native mass spectra of anti PAI-1 antibodies (A) Mab-33H1F7, (B) Mab-55F4C12, (C) Mab-33B8 and (D) Mab-8H9D4.

When PAI-1 Stab was added to the antibody in a 1:1 ratio, there was a similar trend observed in the mass spectra of PAI-1 Stab + Mab (**Fig. 25**). The observed complex consisted of monomeric PAI-1 Stab and monomeric antibody and the PAI-1 Stab + Mab complexes had a similar charge state distribution ranging from 24^+ to 28^+ with the most intense peak corresponding to the 27^+ charge state. The experimental mass ranged from 189,962.4 Da for Mab-33H17F7+PAI-1 Stab to

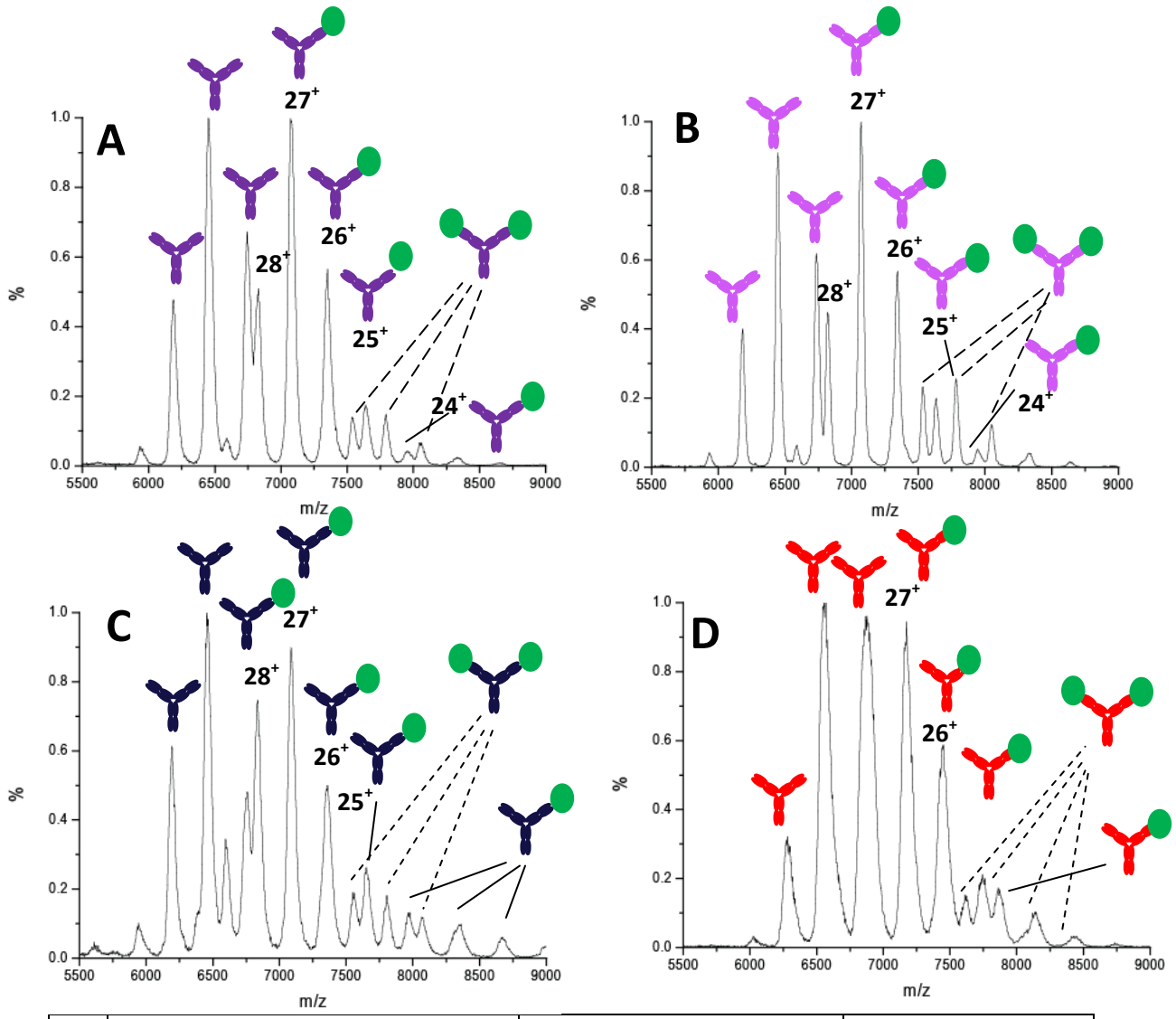
191,553.3 Da for Mab 33B8+PAI-1 Stab. The free antibody is still observed in the mass spectra in the 6000-7000 m/z range.



	Protein	Experimental mass (Da)	Theoretical mass
A	**Mab-33H1F7 + Pai-1 Stab**	189,962.4	191 kDa
	Mab-33H1F7 + (Pai-1 Stab)$_2$	232,955.5	233 kDa
B	**Mab-55F4C12 + Pai-1 Stab**	190,754.6	191 kDa
	Mab-55F4C12 + (Pai-1 Stab)$_2$	233,457.8	233 kDa
C	**Mab-8H9D4 + Pai-1 Stab**	190,582.9	191 kDa
	Mab-8H9D4 + (Pai-1 Stab)$_2$	233,054.3	233 kDa
D	**Mab-33BB + Pai-1 Stab**	191,553.3	191 kDa
	Mab-33BB + (Pai-1 Stab)$_2$	233681.4	233 kDa

Figure 25: Native mass spectra of anti PAI-1 antibodies (A) Mab-33H1F7 + Pai-1 Stab, (B) Mab-55F4C12 + Pai-1 Stab, (C) Mab-8H9D4 + Pai-1 Stab and (D) Mab-33B8 + Pai-1 Stab.

The PAI-1 W175F + Mab complexes showed similar MS results as the PAI-1 Stab + Mab complexes (**Fig. 26**). The complex consists of monomeric antibody and monomeric PAI-1 W175F with a charge state distribution of 24^+ to 28^+, and 27^+ charge state as the most intense peak. The free antibody is observed in the 6000-7000 m/z range. The experimental mass of the Mab+ PAI-1 W175F complex ranges from 190,064.34 Da for Mab-8H9D4 + PAI-1 W175F to 191,227.53 Da for Mab-338B + PAI-1 W175F.



	Protein	Experimental mass (Da)	Theoretical mass
A	**Mab-33H1F7 + Pai-1 W175F**	190,982.8	191 kDa
	Mab-33H1F7 + (Pai-1 W175F)$_2$	233659.4	233 kDa
B	**Mab-55F4C12 + Pai-1 W175F**	190,907.5	191 kDa
	Mab-55F4C12 + (Pai-1 W175F)$_2$	233,537.2	233 kDa
C	**Mab-8H9D4 + Pai-1 W175F**	190,064.3	191 kDa
	Mab-8H9D4 + (Pai-1 W175F)$_2$	233,521.9	233 kDa
D	**Mab-33BB + Pai-1 W175F**	191,227.5	191 kDa
	Mab-33BB + (Pai-1 W175F)$_2$	233,884.3	233 kDa

Figure 26: Native mass spectra of A) Mab-33H1F7 + Pai-1 W175F B) Mab-55F4C12 + Pai-1 W175F C) Mab-8H9D4 +Pai-1 W175F D) Mab-33B8 + Pai-1 W175F.

Tables 12-15 display the CCS values of the anti-PAI-1 antibodies and the PAI-1 + antibody complexes. The anti-PAI-1 antibodies have CCS values of 7500 $Å^2$ to 7548 $Å^2$ when comparing their 20^+ charge states. When comparing the complex with PAI-1 Stab to the complex with PAI-1 W175F, no significant differences are seen at the lowest charge state (24^+) for the four antibodies.

The CCS value of the 24^+ charge state of the monoclonal antibody can be compared to the CCS value of the Mab + PAI-1 Stab complex. For Mab-33H1F7 the complex shows an increase of 17.8 %, for Mab-554C12 an increase of 19.5 %, for Mab-8H9D4 an increase of 18.0 % and for Mab 33B8 an increase of 18.5 %. As the first three mentioned antibody induce the same substrate behaviour of PAI-1 these similar CCS values were expected. Mab-33B8 induces an active-to latent transition of PAI-1, but this form is not distinguishable when comparing to the other CCS values.

Figures 27-30 display the experimental CCS values plotted in function of the molecular weight. The trend line is based on the CCs values of globular calibrants. The CCS values of the antibody, PAI-1 and the PAI-1 + antibody complex follow the predicted trend, suggesting that the overall globular shape remains when forming the complex.

Table 11: CCS values of the antibody Mab-33H1F7, PAI-1 and the PAI-1-antibody complex.

CCS	PAI-1 Stab	PAI-1 W175F	Mab-33H1F7	PAI-1 Stab + Mab-33H1F7	PAI-1 W175F + Mab-33H1F7	$(PAI\text{-}1\ Stab)_2$ + Mab-33H1F7	$(PAI\text{-}1\ W175F)_2$ +Mab-33H1F7
12^+	3469 ± 39 Å^2	3436 ± 43 Å^2					
13^+	3504 ± 51 Å^2	3541 ± 43 Å^2					
20^+			7500 ± 58 Å^2				
21^+			7534 ± 47 Å^2				
22^+			7570 ± 61 Å^2				
23^+			7625 ± 57 Å^2				
24^+			7644 ± 46 Å^2	8999 ± 68 Å^2	8999 ± 86 Å^2		
25^+			7692 ± 46 Å^2	9026 ± 57 Å^2	8967 ± 62 Å^2		
26^+				9011 ± 55 Å^2	9011 ± 69 Å^2		
27^+				9020 ± 78 Å^2	9020 ± 88 Å^2		
28^+				9065 ± 71 Å^2	9065 ± 54 Å^2		
29^+						10468 ± 64 Å^2	10468 ± 49 Å^2
30^+						10541 ± 58 Å^2	10468 ± 55 Å^2
31^+						10587 ± 75 Å^2	10510 ± 66 Å^2
MW	42716.9 Da	42839.5 Da	147869.2 Da	189962.4 Da	190982.8 Da	232955.5 Da	233659.4 Da

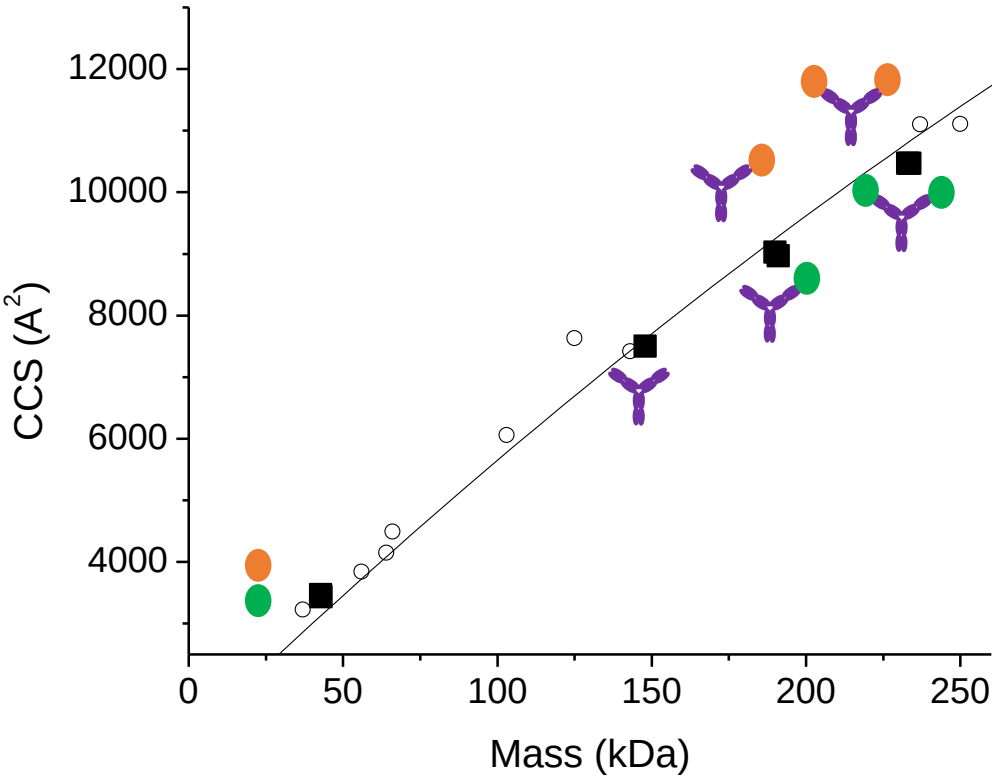


Figure 27: Experimental CCS values of the low-charge states of antibody Mab-33H1F7, PAI-1 and the PAI-1-antibody complex

Table 12: CCS values of the antibody Mab-55F4C12, PAI-1 and the PAI-1-antibody complex.

CCS	PAI-1 Stab	PAI-1 W175F	Mab-55F4C12	PAI-1 Stab + Mab-55F4C12	PAI-1 W175F + Mab-55F4C12	$(PAI\text{-}1\ Stab)_2$ + Mab-55F4C12	$(PAI\text{-}1\ W175F)_2$ + Mab-55F4C12
12^+	3469 ± 39 Å^2	3436 ± 43 Å^2					
13^+	3504 ± 51 Å^2	3541 ± 43 Å^2					
20^+			7500 ± 57 Å^2				
21^+			7484 ± 64 Å^2				
22^+			7461 ± 62 Å^2				
23^+			7507 ± 62 Å^2				
24^+			7580 ± 47 Å^2	9056 ± 63 Å^2	8999 ± 61 Å^2		
25^+			7692 ± 54 Å^2	9084 ± 55 Å^2	8967 ± 46 Å^2		
26^+				9074 ± 60 Å^2	9011 ± 50 Å^2		
27^+				9088 ± 77 Å^2	9020 ± 52 Å^2		
28^+				9089 ± 31 Å^2	9137 ± 59 Å^2		
29^+						10331 ± 59 Å^2	10400 ± 50 Å^2
30^+						10396 ± 67 Å^2	10468 ± 62 Å^2
31^+						10587 ± 52 Å^2	10432 ± 54 Å^2
MW	42,716.9 Da	42,839.5 Da	148,267.4 Da	190,754.6 Da	190,907.6 Da	233,457.8 Da	233,537.2 Da

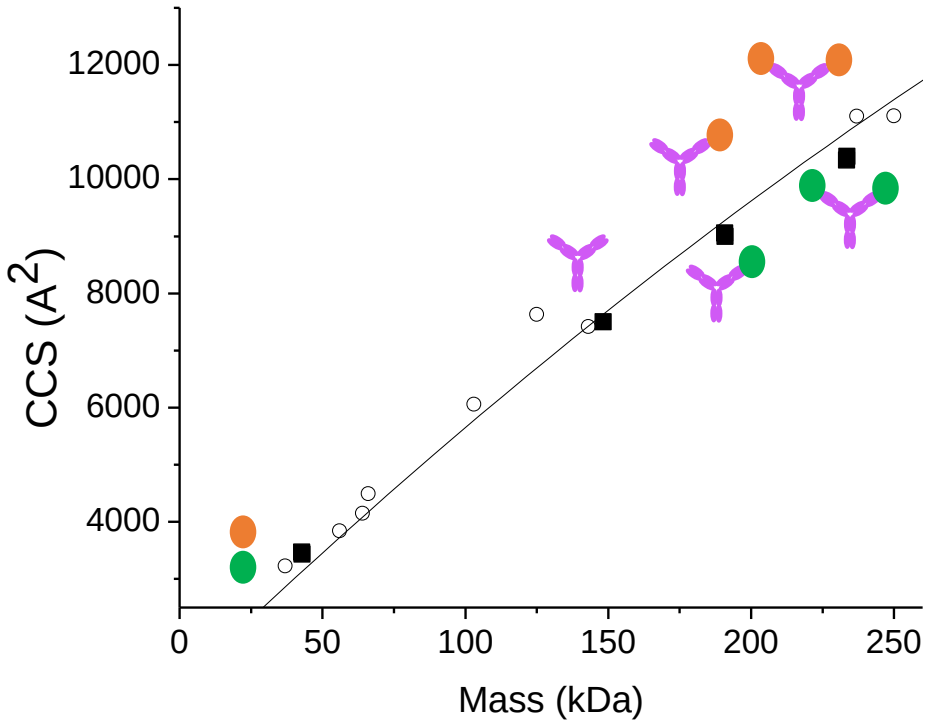


Figure 28: Experimental CCS values of the low-charge states of antibody Mab-55F4C12, PAI-1 and the PAI-1-antibody complex

Table 13: CCS values of the antibody Mab-8H9D4, PAI-1 and the PAI-1-antibody complex.

CCS	PAI-1 Stab	PAI-1 W175F	Mab-8H9D4	PAI-1 Stab + Mab-8H9D4	PAI-1 W175F + Mab-8H9D4	$(PAI\text{-}1\ Stab)_2$ + Mab-8H9D4	$(PAI\text{-}1\ W175F)_2$ + Mab-8H9D4
12^+	3469 ± 39 Å^2	3436 ± 43 Å^2					
13^+	3504 ± 51 Å^2	3541 ± 43 Å^2					
20^+			7548 ± 87 Å^2				
21^+			7483 ± 85 Å^2				
22^+			7515 ± 99 Å^2				
23^+			7507 ± 104 Å^2				
24^+			7580 ± 75 Å^2	8944 ± 74 Å^2	8944 ± 72 Å^2		
25^+			7691 ± 97 Å^2	8907 ± 60 Å^2	8966 ± 80 Å^2		
26^+				9011 ± 66 Å^2	9011 ± 84 Å^2		
27^+				9020 ± 78 Å^2	9088 ± 98 Å^2		
28^+				9004 ± 46 Å^2	9137 ± 70 Å^2		
29^+						10674 ± 53 Å^2	10672 ± 73 Å^2
30^+						10688 ± 78 Å^2	10541 ± 64 Å^2
31^+						10967 ± 79 Å^2	11046 ± 89 Å^2
MW	42,716.9 Da	42,839.5 Da	148,309.2 Da	190,582.9 Da	190,064.3 Da	233,054.2 Da	233,521.9 Da

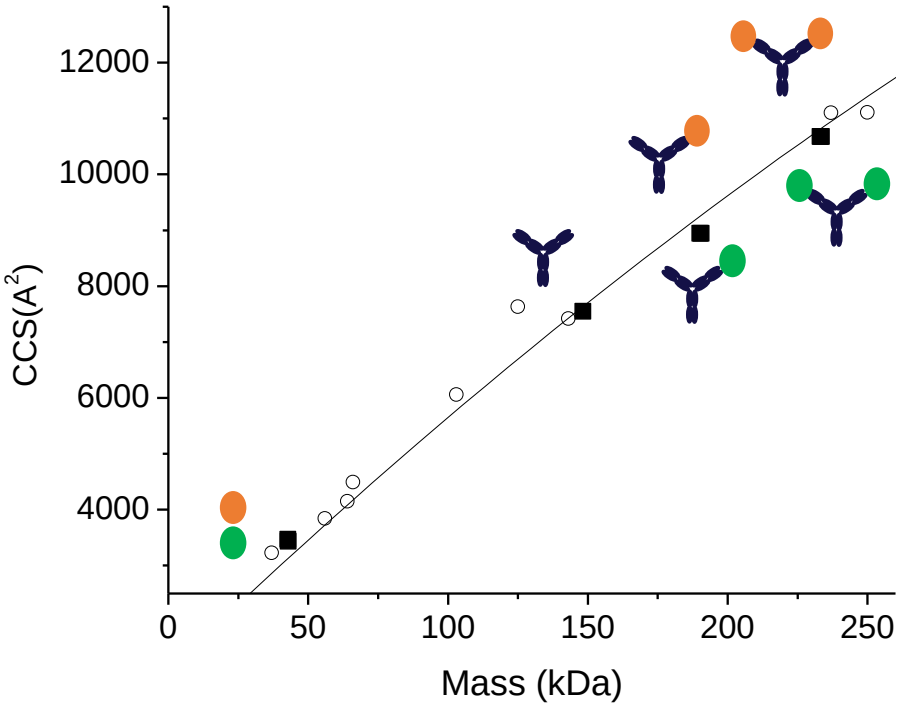


Figure 29: Experimental CCS values of the low-charge states of antibody Mab-8H9D4, PAI-1 and the PAI-1-antibody complex

Table 14: CCS values of the antibody Mab-33B8, PAI-1 and the PAI-1-antibody complex.

CCS	PAI-1 Stab	PAI-1 W175F	Mab-33B8	PAI-1 Stab + Mab-33B8	PAI-1 W175F + Mab-33B8	$(PAI\text{-}1\ Stab)_2$ + Mab-33B8	$(PAI\text{-}1\ W175F)_2$ + Mab-33B8
12^+	3469 ± 39 Å^2	3436 ± 43 Å^2					
13^+	3504 ± 51 Å^2	3541 ± 43 Å^2					
20^+			7548 ± 73 Å^2				
21^+			7583 ± 65 Å^2				
22^+			7570 ± 78 Å^2				
23^+			7566 ± 68 Å^2				
24^+			7643 ± 71 Å^2	9056 ± 71 Å^2	8999 ± 93 Å^2		
25^+			7763 ± 79 Å^2	9142 ± 45 Å^2	9203 ± 78 Å^2		
26^+				9199 ± 45 Å^2	9199 ± 100 Å^2		
27^+				9222 ± 47 Å^2	9424 ± 110 Å^2		
28^+				9495 ± 42 Å^2			
29^+						10807 ± 46 Å^2	10674 ± 106 Å^2
30^+						10830 ± 50 Å^2	10830 ± 93 Å^2
31^+						10893 ± 45 Å^2	10893 ± 87 Å^2
MW	42,716.9 Da	42,839.5 Da	148,981.2 Da	191,553.3 Da	191,227.5 Da	233,681.4 Da	233,884.3 Da

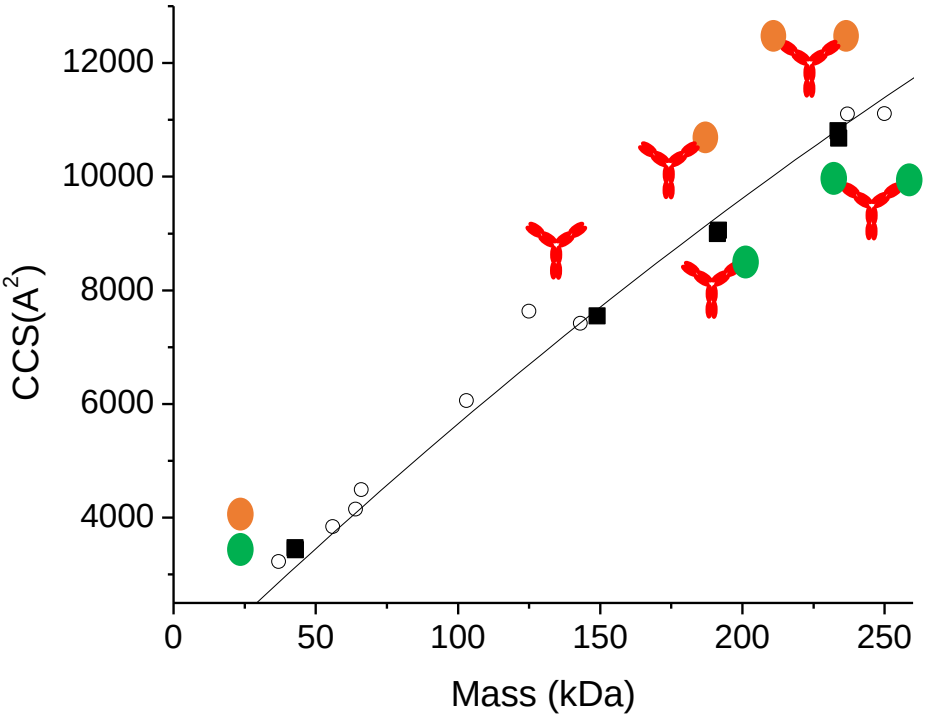


Figure 30: Experimental CCS values of the low-charge states of antibody Mab-33B8, PAI-1 and the PAI-1-antibody complex

5.6.1 Unfolding of anti-PAI-1 antibodies complexed with PAI-1

The collision-induced unfolding of anti-PAI-1 antibodies in complex with PAI-1 Stab was monitored and the results are displayed in **Figure 20**. The overall plots are similar for the four studied antibody-PAI-1 complexes. The transitions occur at 80 V, 90 V and 100 V, after which the signal stays stable until 160 V for Mab-331F7 + PAI-1 Stab (**Fig. 20A**) and until 180 V for the other three Mab + PAI-1 Stab complexes (**Fig. 20B, C &D**). Mab-33B8 displays the clearest transitions and an additional small transition at 110 V.

A similar trend was observed when the anti-PAI-1 antibodies were measured in complex with active PAI-1 W175F (**Fig. 21**). A first slight transition was observed at 60 V for the complexes with the substrate inducing Mabs; 33H1F7, 55F4C12 and 8H9D4 (**Fig. 21A, B &C**). The following transitions occur at 80 V, 90 V and 100 V, after which it stays stable until 170 V.

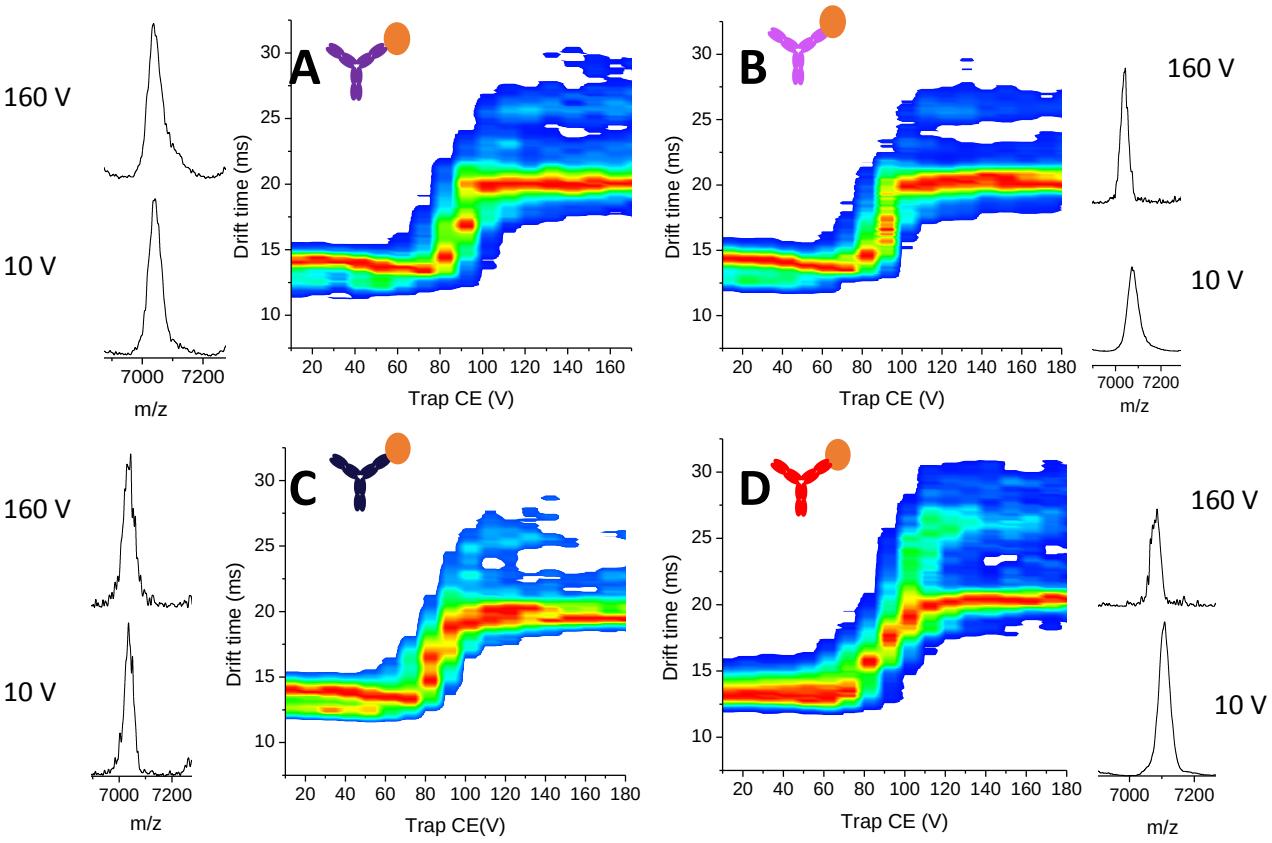


Figure 31: Collision-induced unfolding plots of anti PAI-1 antibodies (A) Mab-33H1F7 + Pai-1 Stab, (B) Mab-55F4C12 + Pai-1 Stab, (C) Mab-8H9D4 +Pai-1 Stab and (D) Mab-33B8 + Pai-1 Stab. Insets show the m/z peak (at different trap voltages) selected to extract the IM drift time.

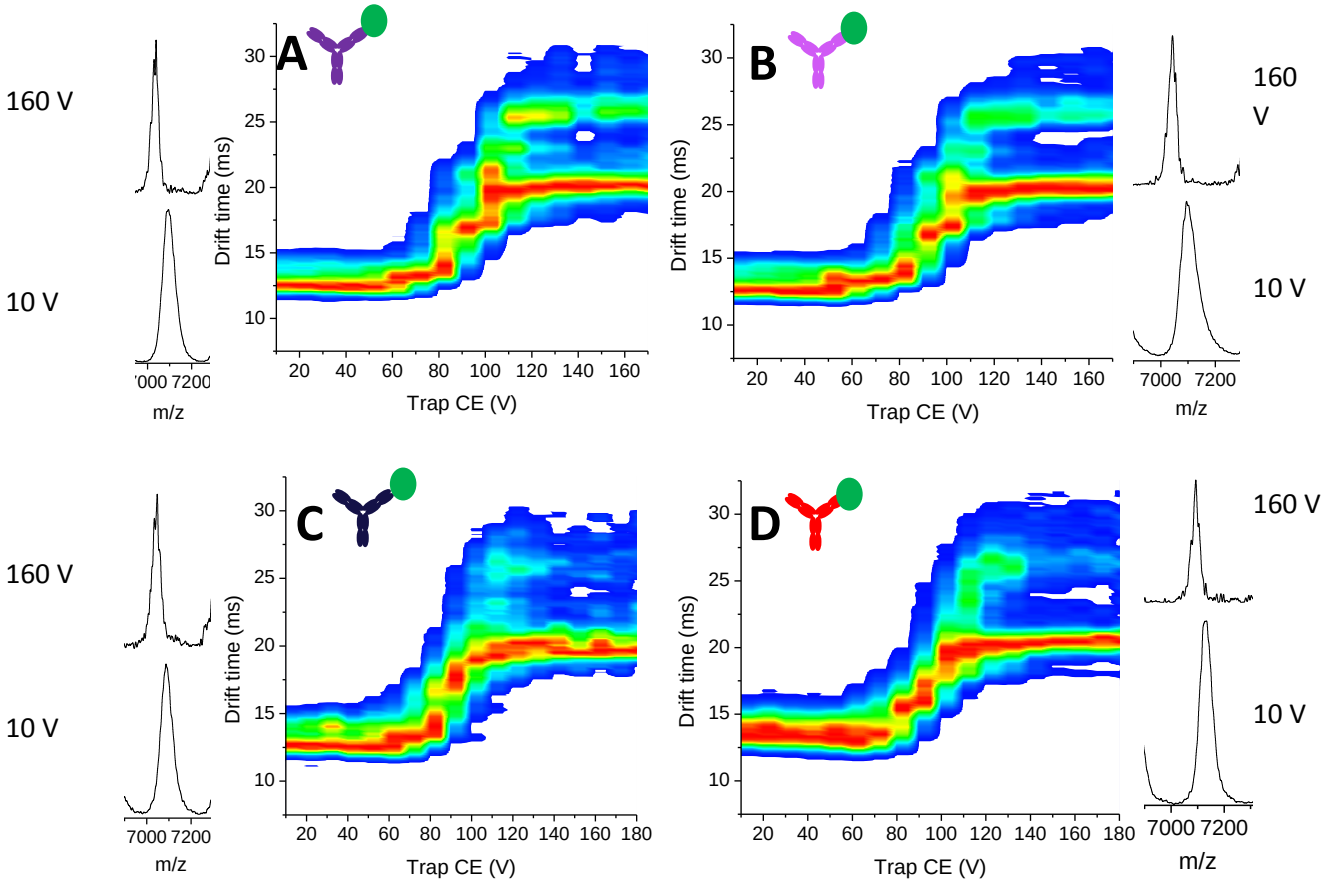


Figure 32: Collision-induced unfolding plots of (A) Mab-33H1F7 + PAI- W175F, (B) Mab-55F4C12 + PAI-1 W175F, (C) Mab-8H9D4 +PAI-1 W175F and (D) Mab-33B8 + PAI-1 W175F. Insets show the m/z peak (at different trap voltages) selected to extract the IM drift time.

5.7 Conclusions

The combination of native mass spectrometry, ion mobility and collision-induced unfolding provided a set of information that could aid in characterising and distinguishing variants. Next to structural analysis this information can also be used to investigate the reproducibility of a pattern.

For the studied PAI-1 variants, which are similar in mass and CCS values, the collision-induced unfolding plots could distinguish an active PAI-1 from a latent PAI-1.

The comparison of nanobodies, which inhibit PAI-1 in different ways, resulted in unique data sets for each nanobody and nanobody + PAI-1 complex.

ScFv-33H1F7 in complex with PAI-1 Stab could be distinguished from its complex with PAI-1 Latent, based on their collision-induced unfolding plot.

The monoclonal antibodies bound to PAI-1 displayed similar results in native MS and CCS values. When comparing their CIU plots the complex with latent-form inducing Mab-33B8 has a slightly different trend compared to the other three substrate inducing mAbs.

5.8 Sources

[1] P. Holliger, PJ. Hudson, Engineering antibody fragments and the rise of single domains. Nat. Biotechnol (2005). 23: 1126-1136

[2] M.M. Schmidt, K0D. Wittrup, A modeling analysis of the effects of molecular size and binding affinity on tumor targeting Mol. Cancer. Ther, 2009, 8,2861-71

[3] M. Valgimigli, G. Campo, M. Tebaldi, R. Carletti, C. Arcozzi, R.Ferrari, and G. Percoco, Abciximab: a reappraisal of its use in coronary care, Biologics, 2008,2, 29-39

[4] W. Strohl, L. Strohl, Therapeutic antibody engineering,: current and future advances driving the strongest growth area in the pharmaceutical industry, 12;265-274

[5] D.C. Andersen, D.E. Reilly, Production technologies for monoclonal antibodies and their fragments, Curr. Opin Biotechnol 2005, 15,456-62

[6] P.J. Declerck, A. Gils, Three Decades of Research on Plasminogen Activator Inhibitor-1: A Multifaceted Serpin, Semin thromn Hemost (2013), 39. 356-364

[7] Revets H, De Baetselier P, Muyldermans S, Nanobodies as novel agents for cancer therapy. Expert. opin. Biol. Theor (2005), 5: 111-124.

[8] S. Muyldermans , Single domain camel antibodies: current status. J. Biotechnol (2001). 74: 277-302

[9] J. Steyaert , B.K. Kobilka, Nanobody stabilization of G protein-coupled receptor conformational states, Curr Opin Strucural Biology, 21, 567-72 Aug 2011

[10] K.D. Miller, J. Weaver-Feldhaus, S.A. Gray, R.W Siegel, M.J. Feldhaus, Production, purification, and characaterization of human scFV antibodies expressed in Saccharomyces cerevisiae, Pichia pastoris, and Escherichia coli, Protein Expr Purif (2005), 42, 255-67

[11] M.M Harmsen, H.J. De Haard, Properties, production, and applications of camelid single-domain antibody fragments, Appl. Microbiol. Biotechnol (2007), 77:13-22

[12] Z. A. Ahmad, S.K. Yeap, A.M. aLi, W.Y. Ho, N. B.M. Alitheen, H. Hamid, scFV Antibody: Principles and Clinical Application, Clinical and Developmental Immunology, Volume 2012 (2012, Article IF 980250

[13] M. Whitlow, D. Filpula, single-chain Fv proteins and their fusion proteins, Methods,2, 1991, 97-105

[14] A. Lehmann, J.H. Wixted, M.V. Shapovalov, H.Roder, R.L. Dunbrack,M.K. Robinson, Stability engineering of anti-EGFR scFv antibodies by rational design of a lambda-to-kappa swap of the VL framework using a structure-guided approach, Mabs, 2015, 7,1058-71

[15] J. G. Elvin, R.G. Couston, C.F. van der Walle, Therapeutic antibodies: Market considerations, disease targets and bioprocessing

[16] A.P. Herrington-Symes, M. Farys, H. Khalili, S. Brocchini, Antibody fragments: Prolonging circulation half-life. Adv. Biosci. Biotech. 2013, 4, 689–698.

[17] J.K. Jensen, L. C. Thompson, J. C. Bucci, P. Nissen, P.G.W. Gettins , C. B. Peterson, P.A.Andreasen, J. Preben Morth, Crystal structure of plasminogen activator inhibitor-1 in an active conformation with normal thermodynamic stability. Journal of biological chemistry, 285, 34, 29709-29717, 2011

[18] M. Dewilde, S.V. Strelkob, A. Rabijna, P. J. Declerck, Journal of structural biology, High quality structure of cleaved PAI-1-stab, 165, 126-132 ,2009

[19] Declerck, P.J., De Mol, M., Vaughan, D.E., Collen, D., 1992. Identification of a conformationally distinct form of plasminogen activator inhibitor-1, acting as a noninhibitory substrate for tissue-type plasminogen activator. J. Biol. Chem. 267, 11693–11696.

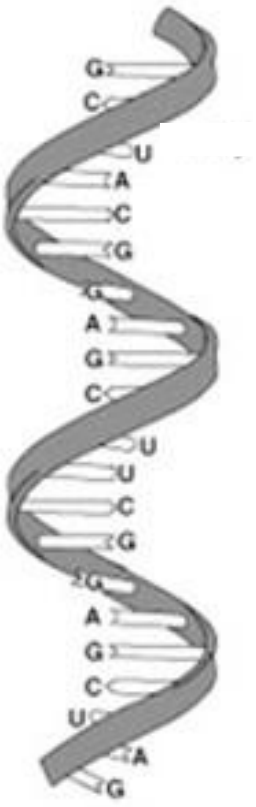
[20] Lawrence, D.A., Olson, S.T., Palaniappan, S., Ginsburg, D., 1994. Serpin reactive center loop mobility is required for inhibitor function but not for enzyme recognition. J. Biol. Chem. 269, 27657–27662.

[21] Andreasen, P.A., Egelund, R., Jensen, S., Rodenburg, K.W., 1999. Solvent effects on activity and conformation of plasminogen activator inhibitor-1. Thromb. Haemost. 81, 407–414.

[22] Wind, T., Jensen, M.A., Andreasen, P.A., 2001. Epitope mapping for four monoclonal antibodies against human plasminogen activator inhibitor type-1 – Implications for antibody-mediated PAI-1-neutralization and vitronectin-binding. Eur. J. Biochem. 268, 1095–1106.
[23] Mottonen, J., Strand, A., Symersk y, J., Sweet, R.M., Danley, D.E., Geoghegan, K.F., Gerard, R.D., Goldsmith, E.J., Structural basis of latency in plasminogen activator inhibitor-1. Nature,1992, 355, 270–273.
[24] Sancho, E., Tonge, D.W., Hockney, R.C., Booth, N.A., Purification and characterization of active and stable recombinant plasminogen-activator inhibitor accumulated at high levels in Escherichia coli. Eur. J. Biochem.1994, 224,125–134.
[25] Stoop, A.A., Eldering, E., Dafforn, T.R., Read, R.J., Pannekoek, H., 2001. Different structural requirements for plasminogen activator inhibitor 1 (PAI-1) during latency transition and proteinase inhibition as evidenced by phage-displayed hypermutated PAI-1 libraries. J. Mol. Biol. 305, 773–783.
[26] Verheyden S., Sillen A., Gils A. Declerck P.J., 2003 Engelborghs Y., Tryptophan Properties in Fluorescence and Functional Stability of Plasminogen Activator Inhibitor 1, Biophysical Journal, 85,501-510
[27] Zhou X, Hendrickx ML, Hassanzadeh-Ghassabeh G, Muyldermans S, Declerck PJ, Generation and in vitro characterisation of inhibitory nanobodies towards plasminogen activator inhibitor 1, Thromb Haemost. 2016 Nov 30;116(6):1032-1040.
[28] Deveiter J., Booth N.A., Declerck P.J., Gils A., Bispecific targeting of thrombin activatable fibrinolysis inhibitor and plasminogen activator inhibitor-1 by a heterodimer diabody, J. Thromb Haemost, 11, 1884-91, 2008
[29] Komissarov A.A. Declerck P.J. Shore J.D., Mechanisms of conversion of plasminogen activator inhibitor 1 from a suicide inhibitor to a substrate by monoclonal antibodies, J. Biol chem., 2002, 46, 43858-65
[30] Verhamme I., Kvassman J.O., Day D., Debrock S., Vleugels N., Declerck P.J., Shore JD, Accelerated conversion of human plasminogen activator inhibitor-1 to its latent form by antibody binding, J. Biol. Chem., 1999, 275, 17511-7
[31] Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Analytical Chemistry* **2010,** *82* (22), 9557-9565.
[32] K. M. Arndt, K.M. Müller, A. Plückthun, Factors Influencing the Dimer to Monomer Transition of an Antibody Single-Chain Fv Fragment, Biochemistry 1998, 37, 12918-12926
[33] C. Larriba, C. J. Hogan, Ion mobilitites in diatomic gases: measurement versus prediction with non-specular scattering models, J. Phys. Chem. A, 2013, 117, 3887-3901
[34] J. L. Benesch, B. T. Ruotolo, Mass Spectrometry: an Approach Come-of-Age for Structural and Dynamical BiologyCurr. Opin. Struct. Biol. 2011, 21, 641–649

Chapter 6

Analysis of RNA therapeutics



6.1 Introduction

Ribonucleic acid (RNA) is emerging as a new category of molecule with therapeutic interest. RNA therapeutics can be defined as the use of RNAs as therapeutic agents or the modulation of RNA structure and function with more traditional drugs (e.g. small molecules or peptides). However, issues such as the instability of RNA, potential immunogenicity, and the requirement of a delivery vehicle for efficient transport to the targeted cells, have hindered the clinical progress of some RNA-based drugs. Nonetheless, solutions to these problems are provided by improved synthetic delivery carriers and chemical modifications of the RNA [1]. The focus on research and development of RNA therapies is growing as more than 50 RNA or RNA-derived therapeutics have reached clinical testing. To date, four oligonucleotide therapeutics have been approved by the FDA and the expected approvals will launch this class of molecules as the third major drug category after biologics and small molecules [2]. In this chapter, the different types of RNA as well as the range of RNA therapies will be described.

6.1.1 **Ribonucleic acid (RNA)**

RNA usually consists of a single-stranded chain of nucleotides. Each nucleotide consists of ribose, a phosphate group and a nitrogenous base (adenine, guanine, cytosine or uracil) (**Fig. 1**). RNA primary structure resembles DNA as it is composed of nitrogenous bases linked to sugar structures. Yet, the higher-order structure of RNA mimics proteins to some extent, as it can fold into complex structures with unique binding sites suitable for small molecules, peptides or proteins [3]. These characteristics make RNA an attractive therapeutic target or agent. Since there are important diseases caused by RNA viruses, such as the human immunodeficiency virus (HIV), hepatitis C virus (HCV) and respiratory syncytial virus (RSV), this emphasizes the importance of developing compounds that can target RNA [3].

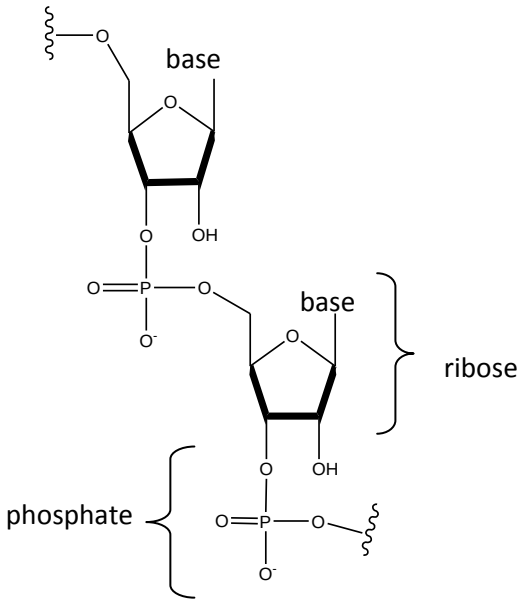


Figure 1: Chemical structure of RNA

There are several types of RNA, which can be classified based on their functions (**Fig.2**). Messenger RNA (mRNA) is involved in protein-coding, as it transcribes the genetic information from DNA into a series of three-base code words [4]. However, there are RNA molecules that are not translated into proteins, called non-coding RNA (ncRNA) [5]**.** The different types of non-coding RNA include the translation-related RNAs, ribosomal RNA (rRNA) and transfer RNA (tRNA). Ribosomal RNA directs the translation of mRNA into proteins. Transfer RNA (tRNA) transfers amino acids, which correspond to each three-nucleotide codon of tRNA, to the ribosome. These amino acids are then joined together to make peptides and proteins. Research continues to uncover the important role of non-coding RNA in epigenetic gene regulation (i.e., the contribution of an organism's environment and experience to its biology) [6]**.** These non-coding RNAs can be short (i.e., shorter than 200 nucleotides, or long (i.e., larger than 200 nucleotides). There are several subtypes of long and short ncRNAs and these can be grouped according to their functions [6]. RNA interference (RNAi) technology uses short RNA molecules to regulate the expression of a gene of interest [7]. Small interfering RNAs (siRNA), which are short pieces of double stranded RNA, can initiate the specific degradation of a targeted cellular mRNA. Long non-coding RNA can be involved in epigenetic gene silencing and in tumor suppression [8].

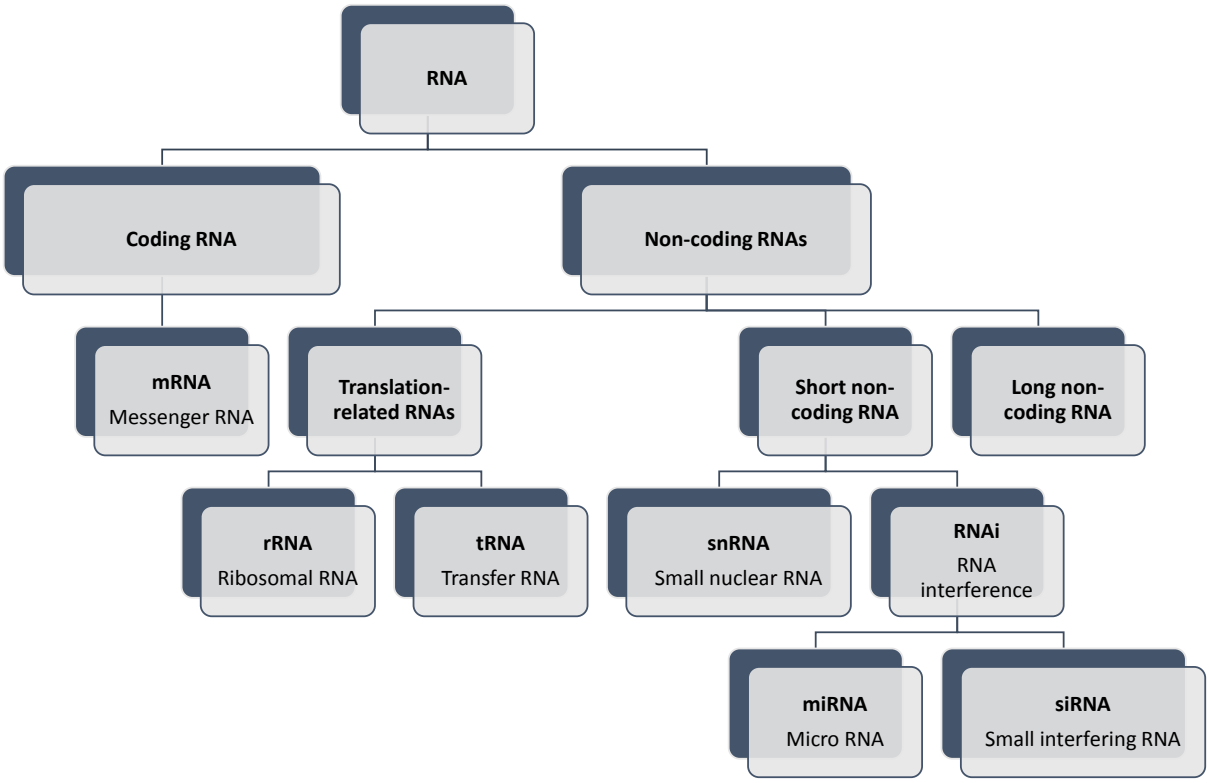


Figure 2: General overview of different types of RNA.

6.1.2 **RNA-based therapeutics**

RNA-based therapeutics have vast potential to treat a variety of diseases. Despite issues such as delivery, specificity, stability and immune activation, the development of RNA therapeutics is advancing by focusing on these hurdles. By improving synthetic delivery carriers and implementing chemical modifications of RNA therapeutics, emerging drugs are showing promising results in clinical trials [1]. RNA-based therapeutics include: inhibitors of mRNA translation (antisense), RNA interference (RNAi), catalytically active RNA (ribozymes), and RNA that binds proteins or other molecular ligands (aptamers). **Figure 3** gives an overview of the current RNA-based therapies with a brief description of their mechanism and the related approved drugs.

THERAPY	DESCRIPTION	TARGET	MODE OF ACTION
Antisense oligonucleotide	Single-stranded DNA or RNA	Messenger RNA (mRNA) or pre-mRNA	Blocks protein translation or affects splicing
Aptamer	Single-stranded DNA or RNA	Proteins, small molecules, toxins, or even whole cells	Inactivates or modifies target. (See "Antibody Alternatives," *The Scientist*, February 2016.)
Small interfering RNA (siRNA)	Double-stranded RNA	mRNA	Hijacks natural gene-silencing pathway in cells that uses interference (RNAi) to trigger target degradation
Anti-microRNA	Single-stranded RNA	microRNA	Inactivates microRNA, leading to changes in gene expression
MicroRNA mimic	Double-stranded RNA	mRNA	Augments activity of endogenous microRNA to block translation

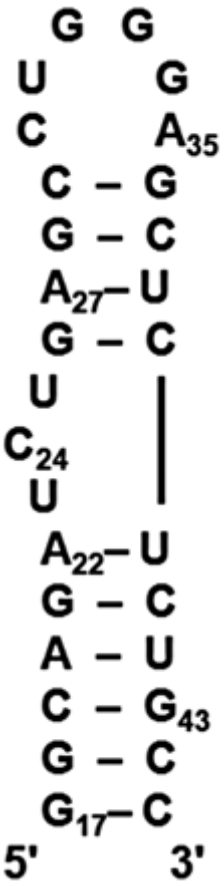
THERAPY	MAIN COMPANIES CURRENTLY INVOLVED IN CLINICAL OR PRECLINICAL RESEARCH	DRUGS THAT HAVE BEEN APPROVED OR ARE IN LATE-STAGE CLINICAL TRIALS
Antisense oligonucleotide	• Ionis Pharmaceuticals • Sarepta Therapeutics	• Fomivirsen (approved 1998) • Mipomersen (approved 2013) • Eteplirsen (approved 2016) • Nusinersen (Phase 3 trial halted in August due to positive results) • Alicaforsen (currently in Phase 3 trials) • Aganersin (currently in Phase 3 trials)
Aptamer	• SomaLogic • RIBOMIC • Aptagen	• Pegaptanib (approved 2004) • Pegpleranib (currently in Phase 3 trials)
Small interfering RNA (siRNA)	• Alnylam Pharmaceuticals • Dicerna Pharmaceuticals • Arrowhead Pharmaceuticals • Arbutus Biopharma • Quark Pharmaceuticals	• Patisiran (results from Phase 3 trial expected mid-2017) • QPI-1002 (currently in Phase 3 trials)
Anti-microRNA	• Santaris (owned by Roche) • Regulus Therapeutics	None
MicroRNA mimic	• Mirna Therapeutics • Regulus Therapeutics	None

Figure 3: General overview of oligonucleotide-based therapeutics. Figure taken from [2]

6.1.3 Sources

[1] J.C. Burnett, J.J. Rossi, RNA-based Therapeutics-Current Progress and Future Prospects, Chem. Biol. 2012, 19(1), 60-71
[2] C. Offord, Oligonucleotide Therapeutics Near Approval, The scientist, 2016, Dec
[3]K.L. McKnight, B.A. Heinz, RNA as a target for developing antivirals, Antiviral Chemistry 1 Chemotherapy 2003, 14, 61-73
[4] H. Lodish, A. Berk, S.L. Zipusky, The Three Roles of RNA in Protein Synthesis, Molecular Cell Biology 4th Ediition, 2000
[5] A.F. Palazzo, E. S. Lee, Non-coding RNA: what is functional and what is junk? Front. Genet., 2015, 6, 2
[6] V.J. Peschansky, C. Wahlestedt, Non-coding RNAs as direct and indirect modulators of epigenetic regulation, Epigenetics, 2014, 9, 3-12
[7] N. Agrawal, P. V. N. Dasaradhi, A.Mohmmed, P. Malhotra, R.K. Bhatnagar, S.K. Mukherjee, RNA Interference: Biology, Mechanism, and Applications, Microbiol. Mol. Biol. Rev., 2003, 67, 657-685
[8] B. Santosh, A. Varshney, P. K. Yadava, Non-coding RNAs: Biological functions and applications, Cell Biochem. Funct., 2015, 33, 14-22

6.2 TAR-RNA peptide binding



In collaboration with

Dr. Matthew D. Shortridge
University of Washington, US
Department of Chemistry

6.2.1 Introduction

Trans-activation response (TAR) is a RNA-element required to stimulate viral gene expression and virus replication [1]. The TAR sequence is located at the 5'-noncoding region of HIV-1 mRNA's, and forms a stable stem-loop structure in vitro (**Figure 1**). The TAT (trans-activator of transcription) protein interacts with the TAR sequence in order to regulate viral gene expression [1]. The HIV TAR-RNA has been considered a target for inhibiting HIV replication due to its complex role in facilitating transcription of viral DNA [2]. Many of the known TAR-binding small molecules often have poor affinity, inadequate physicochemical properties or poor selectivity [2]. Hence, finding the right small molecules or peptides that bind to TAR remains an important challenge. The goal of this project was to evaluate native IM-MS for RNA-peptide binding and to provide structural restraints, derived from the collision cross sections, for further structural modeling.

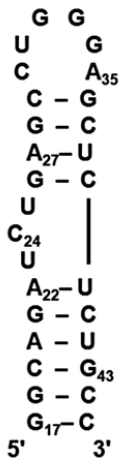


Figure 1: TAR-RNA structure

6.2.2 Experimental details

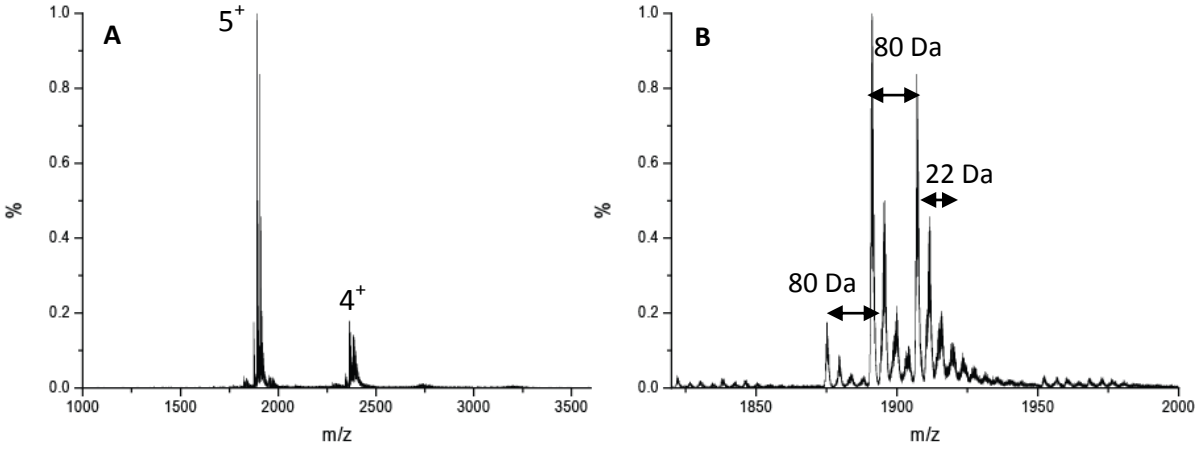
The RNA samples were provided by Dr. Matthew D. Shortridge of the Chemistry department of the University of Washington. Stock RNA sample (77uM in water) was heated for 4 minutes at 90°C, and then cooled at -20°C until sample was frozen. This last step helps to ensure that the stem loop structure is formed. Before the MS experiment the samples were buffer-exchanged into 150 mM ammonium acetate using Micro Bio-spin columns (Bio-Rad) and diluted to a concentration of 10 μM. The instrument was tuned to preserve native higher-order structure using the following parameters: spray capillary voltage 1.2 kV; nanoflow gas pressure 0.1 bar; sampling cone 20 V; extraction cone 1.0 V; trap collision energy 4.0 V; transfer collision energy 0 V; IMS wave velocity 300 m/s; IMS wave height 35.0 V ; backing pressure 3.0 mbar. The data were acquired and processed with Masslynx v4.1 software, and ion mobility drift times extracted using Driftscope v2.3 (both Waters). As calibrants, an ESI- Low concentration tuning mix (Agilent) was used (Table 1). The CCS values of six compounds from this tuning mix have been previously calibrated by Gabelica et al. in negative mode on a linear drift tube ion mobility spectrometer (DT-IMS; Gabelica, personal communication). These calibrated CCS values were used to calculate the CCS values of RNA molecules using the CCS calibration method as described previously [3]. In this experiment negative calibrants were used, because no equivalent compound calibrations were available.

Table 1: List of low mass ESI calibrants (Agilent) used for the calibration of TAR-RNA and complexes. The CCS values were obtained in negative mode.

Calibrant	Mass (Da) Negative mode	CCS value (Å^2)
Hexakis(2,2-difluoroethoxy)phosphazine	601.98	118.6
Hexakis(1H, 1H, 3H-tetrafluoropropoxy)phosphazine	1033.99	190.0
Hexakis(1H, 1H ,4H-hexafluorobutyloxy)phosphazine	1333.97	215.5
Hexakis(1H, 1H, 5H-octafluoropentoxy)phosphazine	1633.95	244.3
Hexakis(1H, 1H, 6H-decafluorohexyloxy)phosphazine	1933.93	272.2
Hexakis(1H, 1H, 7H-dodecafluoroheptoxy)phosphazine	2233.91	295.8

6.2.3 Native IM-MS of TAR-RNA

With native mass spectrometry, the 5^+ and 4^+ charge state of TAR-RNA were observed, with an experimentally determined mass of 9451.4 Da (**Figure 2**). The theoretical mass of the TAR sequence is 9446.2 Da. When zooming into the 5+ charge state additional peaks were noted, with an additional mass of 80 Da and 22 Da, which can indicate phosphate (instead of a hydroxyl group) and sodium adducts, respectively. The low charge states indicated that the TAR-RNA was still in its compact native state.



	Experimental Mass (Da)	Theoretical Mass (Da)
TAR-RNA	9451.4	9446.18

Figure 2: Native mass spectrum of TAR-RNA (A) and TAR-RNA zoomed in on 5^+ charge state (B).

Figure 3 shows the ion mobility drift times of the 3^+, 4^+ and 5^+ charge states of TAR-RNA , as well as the presence of the TAR-RNA dimer (6^+ and 7^+). When observing the narrowness of the drift time peaks in the y-axis, they appear to be compact structures that did not show multiple conformations per charge state.

The collision cross sections of TAR-RNA were derived from the ion mobility drift times and are displayed in **Table 2**.

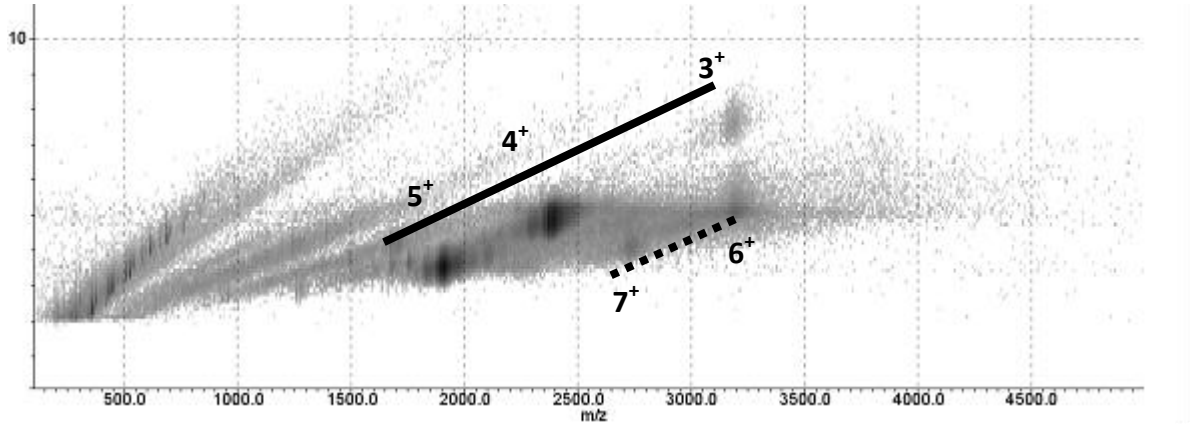


Figure 3: Native ESI-MS ion mobility drift times in function of m/z values of TAR-RNA.

Table 2: Collision Cross Sections of the observed charge states of TAR-RNA

Experimental Mass (Da)	TAR-RNA	CCS (Å^2)
9451.4	5^+	**575.0 ± 0.5**
9451.4	4^+	**569.9 ± 0.5**
9451.4	3^+	**567.8 ± 0.2**

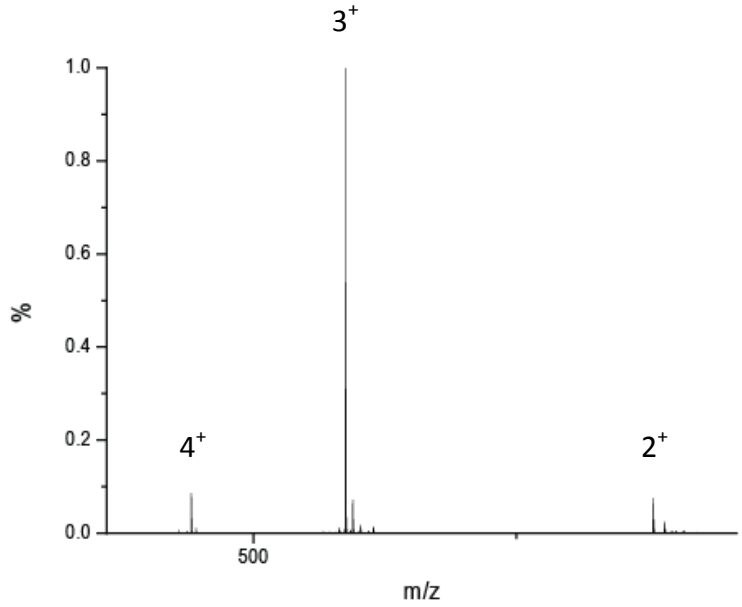
Experimental Mass (Da)	TAR-RNA dimer	CCS (Å^2)
18965.3	7^+	**881.9 ± 22.8**
18965.3	6^+	**867.3 ± 21.7**

6.2.4 Interaction of TAR-RNA with a cyclic peptide

In this study, the interaction of TAR-RNA with a cyclic peptide, JB64, was reported (**Figure 4**). This peptide has a theoretical mass of 1735.08 Da. With native mass spectrometry, a mass of 1758.12 Da was observed (**Figure 5**), which can be assigned as the cyclic peptide with one sodium adduct (+23 Da).

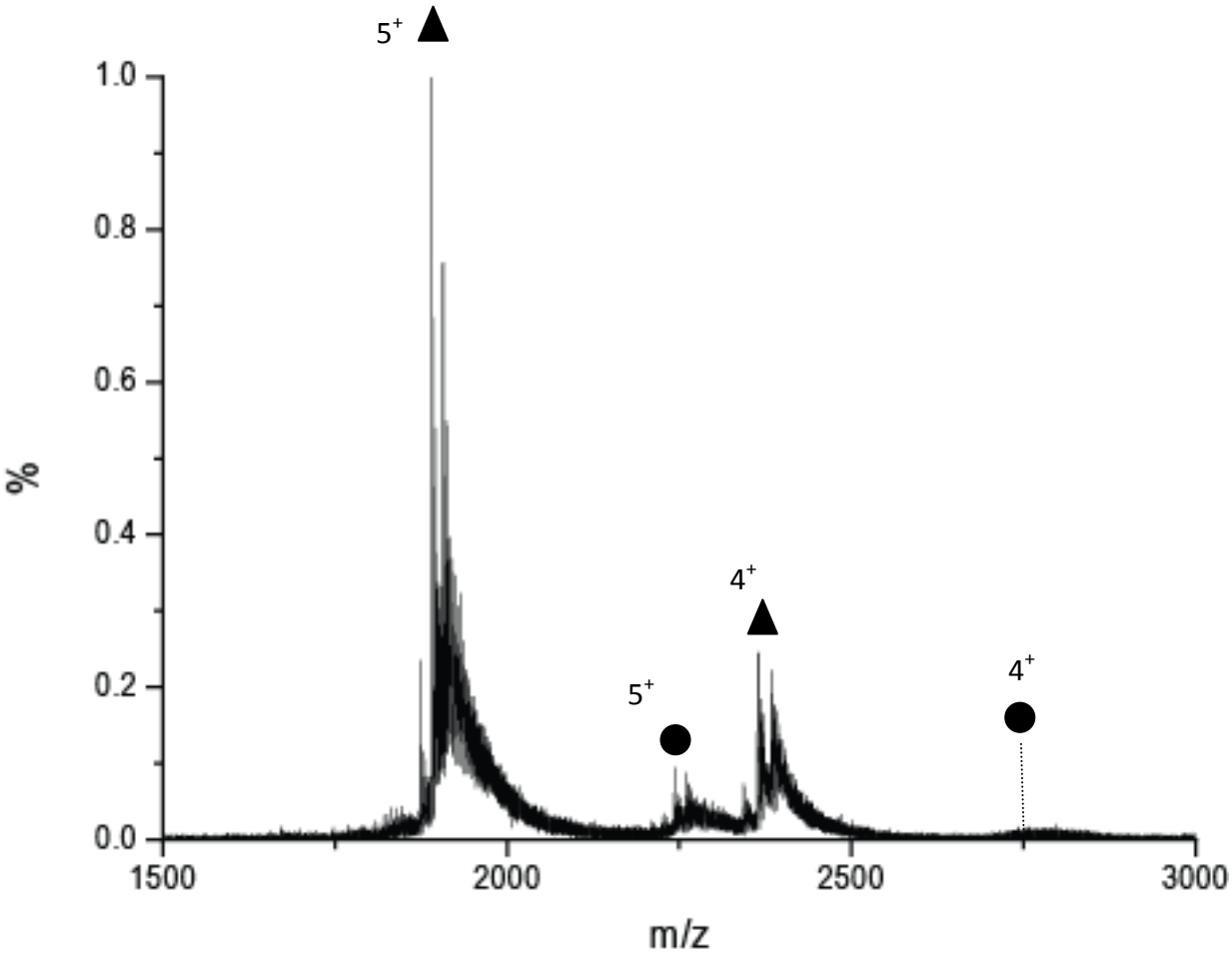
cyclo(RVRTQKGRRIRIpP)

Figure 4: Sequence of cyclopeptide JB64.



	Experimental Mass (Da)	Theoretical Mass (Da)
Cyclopeptide JB64	1758.1	1735.08

Figure 5: Native mass spectrum of cyclopeptide JB64.



	Experimental Mass (Da)	Theoretical Mass (Da)
TAR-RNA	9451.4	9446.18
TAR-RNA + JB64	11215.2	11181.26

Figure 6: Native mass spectrum of TAR-RNA with added cyclopeptide JB64. The peaks corresponding to TAR-RNA are labeled with a triangle and the complex with a circle.

When adding the cyclic peptide, JB64 to TAR-RNA at a 1:1 molar ratio, the presence of the TAR-RNA complexed with one JB64 unit was noted (**Figure 6**).

Since the relative intensity of the peak at 2243 m/z was low, an MS/MS experiment was performed on this peak to confirm that it corresponds to the RNA-peptide 1:1 complex. In Figure 7 the MS/MS spectra are shown with increasing trap collision energy. When a trap collision energy of 80 volts was applied, the RNA-peptide complex dissociated into its subunits.

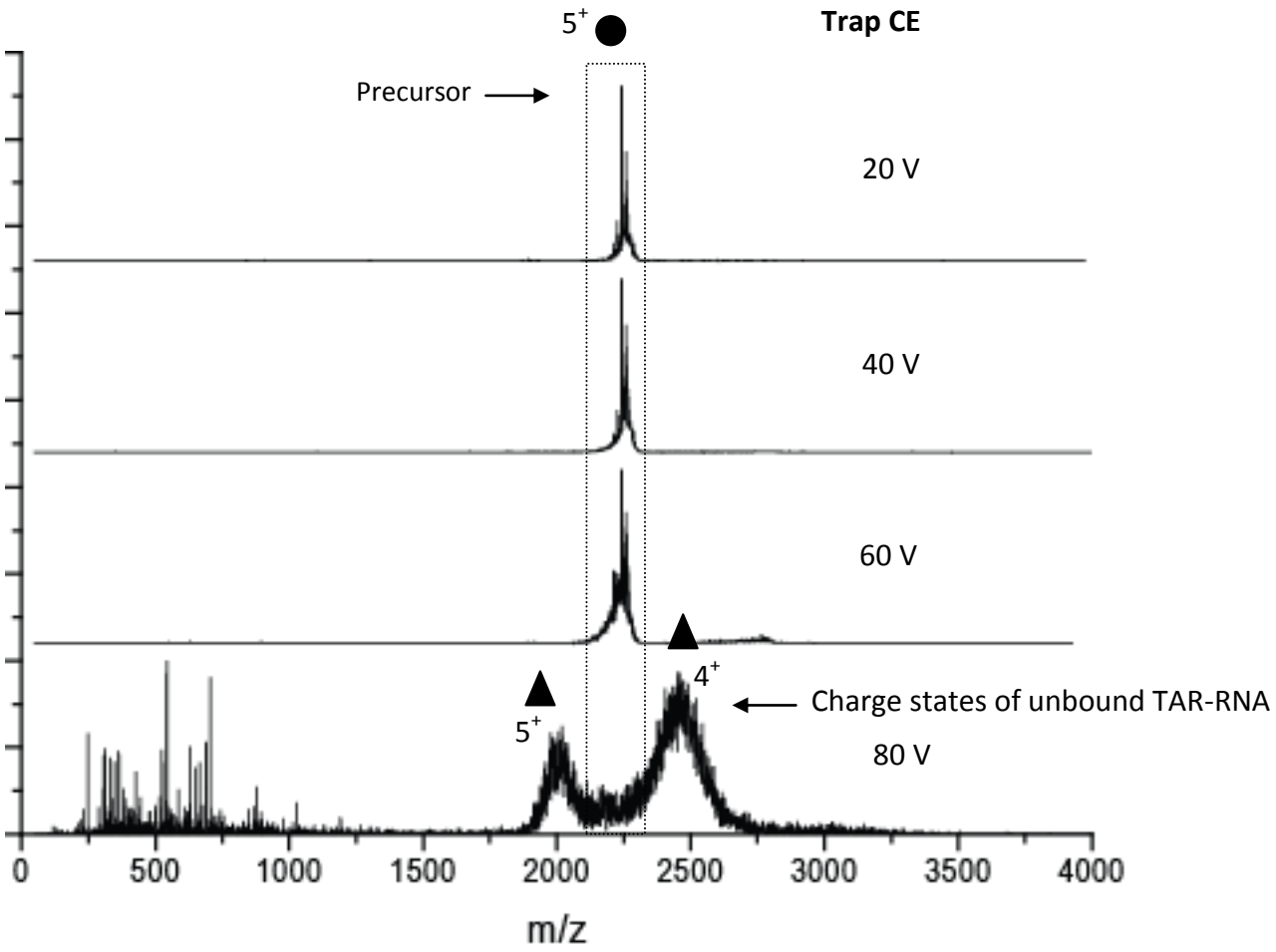


Figure 7: MS/MS spectra of the TAR-RNA + JB64 complex (●)peak at 2243 m/z with increasing trap collision energy. At 80 V the complex is dissociated as unbound TAR-RNA (Δ) appears

Table 3 shows the collision cross sections derived from the ion mobility drift times of TAR-RNA and the TAR-RNA-peptide complex. The cyclic peptide, JB64, has a collision cross section from 466.4 to 521.4 Å^2 and the TAR-RNA-peptide complex a collision cross section of 800.0 Å^2. When looking at the effect of the increased trap collision energy on the ion mobility drift time, no difference was seen. Thus, the RNA complex structure is very stable and is tolerant to the voltage changes in the MS instrument.

Table 3: Collision cross sections of cyclopeptide JB64 and the TAR-RNA+JB64 complex.

JB64	CCS (Å^2)
2^+	**466.4**
3^+	**479.5**
4^+	**521.4**

TAR-RNA + JB64	CCS (Å^2)
5^+	**800.0**

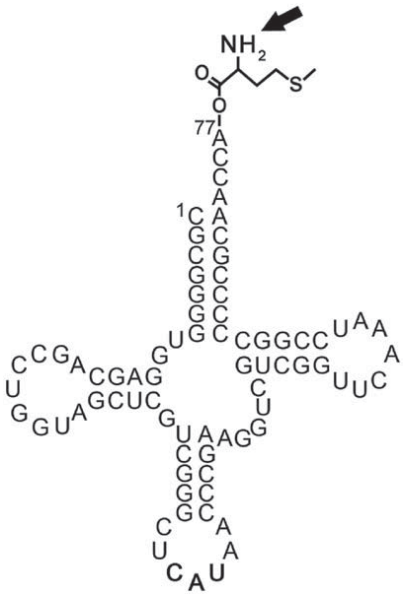
6.2.5 Conclusion

Native IM-MS was used to determine the mass and collision cross section of TAR-RNA and of the cyclopeptide, JB64. With ion mobility, the TAR-RNA dimer peaks could be more clearly distinguished from the TAR-RNA monomer.
When adding the cyclic peptide to TAR-RNA, a 1:1 complex was observed. The presence of this complex was confirmed by performing an MS/MS experiment of the complex peak. The native IM-MS data provides a reference for mass and size, which can be used for further comparison with calculated structures.

6.2.6 Sources

[1] S. Roy, U. Delling, C.H. Chen, C.A. Rosen, N. Sonenberg, A bulge structure in HIV-1 TAR RNA is required for Tat binding an Tat-mediated trans-activation, Genes & Development 1990, 4, 1365-1373
[2] J. Sztuba-Solinska, S.R. Shenoy, P.Gareiss, L.R.H. Krumpe, S.F.J.LeGrice, B.R. O'Keefe, J.S.Schneekloth, Jr, Identification of Biologically Active, HIV TAR RNA-Binding small Molecules using Small Molecule Microarrays
[3] Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Analytical Chemistry* **2010,** *82* (22), 9557-9565.

6.3 AtaT blocks translation initiation by N-acetylation of the initiator tRNA^{fmet}



In collaboration with

Dukas Jurenas[1,2]
Abel Garcia-Pino[1]

[1] Laboratoire de Génétique et Physiologie Bactérienne, Université Libre de Bruxelles, Gosselies, Belgium.
[2]Department of Biochemistry and Molecular Biology, Vilnius University Joint Life Sciences Center, Vilnius, Lithuania

This research was published in

Dukas Jurėnas, Sneha Chatterjee, Albert Konijnenberg, Frank Sobott, Louis Droogmans, Abel Garcia-Pino, Laurence Van Melderen,
AtaT blocks translation initiation by N-acetylation of the initiator tRNA^{fMet},
Nature Chemical Biology, 2017, 13,640-646

6.3.1 Introduction

Toxin-antitoxin systems (TA) are found in bacterial genetic elements and chromosomes. They are involved in the regulation of processes such as plasmid stabilization and protection against phages. When the toxin and the antitoxin components are proteins, the modules are called type II TA modules. AtaT-AtaR is a novel type II TA operon found in *Escherichia coli* and stands for **A**minoacyl **t**RNA **a**cetyltransferase **T**oxin-**R**epressor. The toxin, AtaT, contains an N-acetyltransferase GNAT domain (Gcn5-related N-acetyltransferase) which can modify substrates from small molecules, such as antibiotics, to macromolecules. The goal of this project was to demonstrate that AtAT inhibits translation initiation by acetylating the free amine group of methionine on the $tRNA^{fMet}$, using acetyl coenzyme A (AcCoA) as acetyl donor (**Fig. 1**). The acetylation of the initiator Met-tRNAfMet prevents the formation of an initiation complex compatible with translation initiation. This would result in the inhibition of proteins synthesis and subsequent growth arrest.

This chapter focussed on how mass spectrometry assisted in confirming the formation of the AtaT-AtaR complexes and the acetylation of $tRNA^{fMet}$.

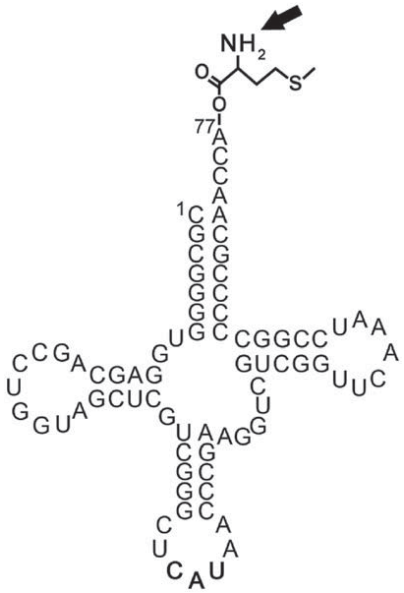


Figure 1: Illustration of initiator $tRNA^{fMmet}$, the arrow indicates the amine group modified by AtaT.

6.3.2 Experimental details

Samples of the AtaT-AtaR complex were prepared at 20 µM protein concentration in 100 mM ammonium acetate buffer, pH 6.9. The treated and non-treated Met-$tRNA^{fMet}$ samples (10 µM) were also prepared in 100 mM ammonium acetate buffer, pH 6.9. All the samples were introduced into the vacuum of the mass spectrometer using nanoelectrospray ionization with in-house-prepared gold-coated borosilicate glass capillaries with a voltage of approximately +1.6 kV. Spectra were recorded on a quadrupole TOF instrument (Q-TOF2, Waters, Manchester, UK) modified for transmission of native, high *m*/*z* protein assemblies as described previously [1].

6.3.3 Native MS of AtaR-AtaT complex

AtaR-AtaT is a heterogeneous mixture of complexes with different toxin:antitoxin ratios. The predominant species observed at native MS conditions consist of AtaT:AtaR at 1:1 and 2:2 ratios, which highlights the fact that the interaction between AtaT and AtaR might be stronger than the expected AtaR dimer that is unstable even under gentle desolvation conditions (**Fig.2**).

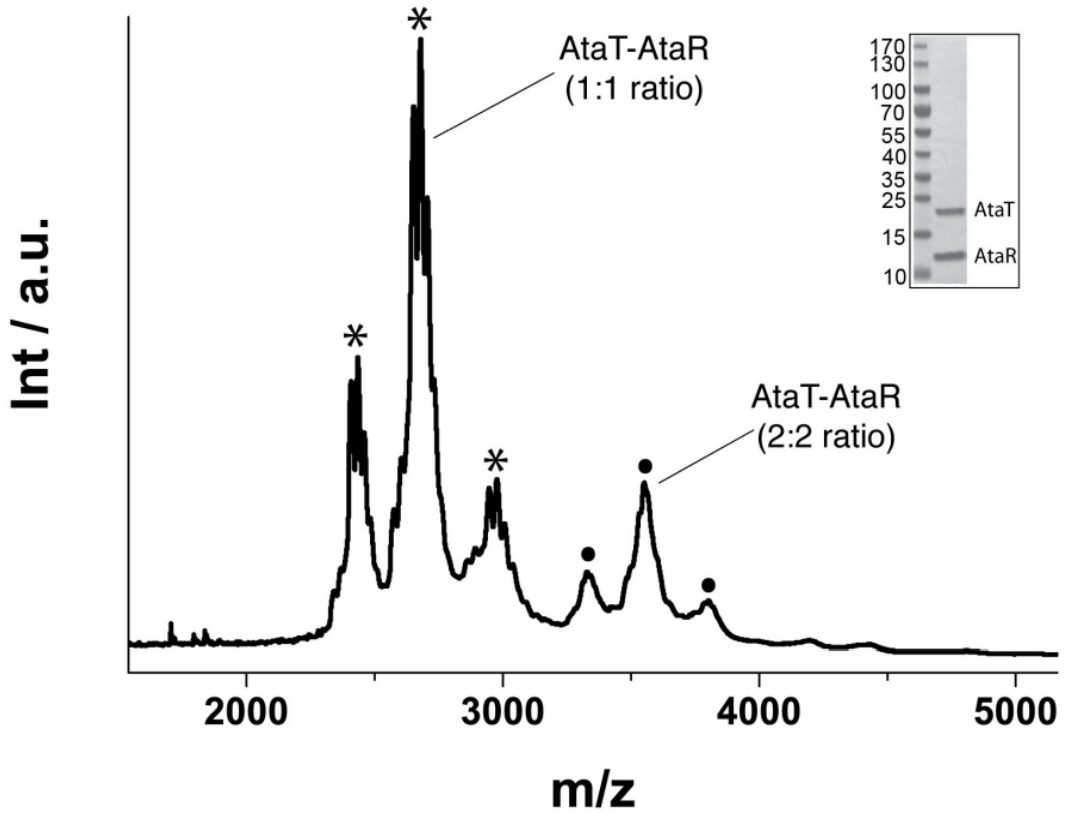


Figure 2: Native mass spectrum of the AtaT-AtaR complex. The inset shows the SDS-PAGE gel showing the presence of AtAT and AtaR

6.3.4 Native MS of treated $tRNA^{fMet}$

Figure 3 shows that the mass difference between the peaks corresponding to the $tRNA^{fMet}$ modified with the methionyl-cargo (treated and non-treated) is 41.2 Da, which is in the range of the expected average increase in mass corresponding to the introduction of an acetyl group (42.01 Da). The remaining peaks (labelled in the figure) correspond to non-charged $tRNA^{fMet}$ and $tRNA^{fMet}$ species containing an additional base (a common by-product from the in vitro synthesis of tRNAs)

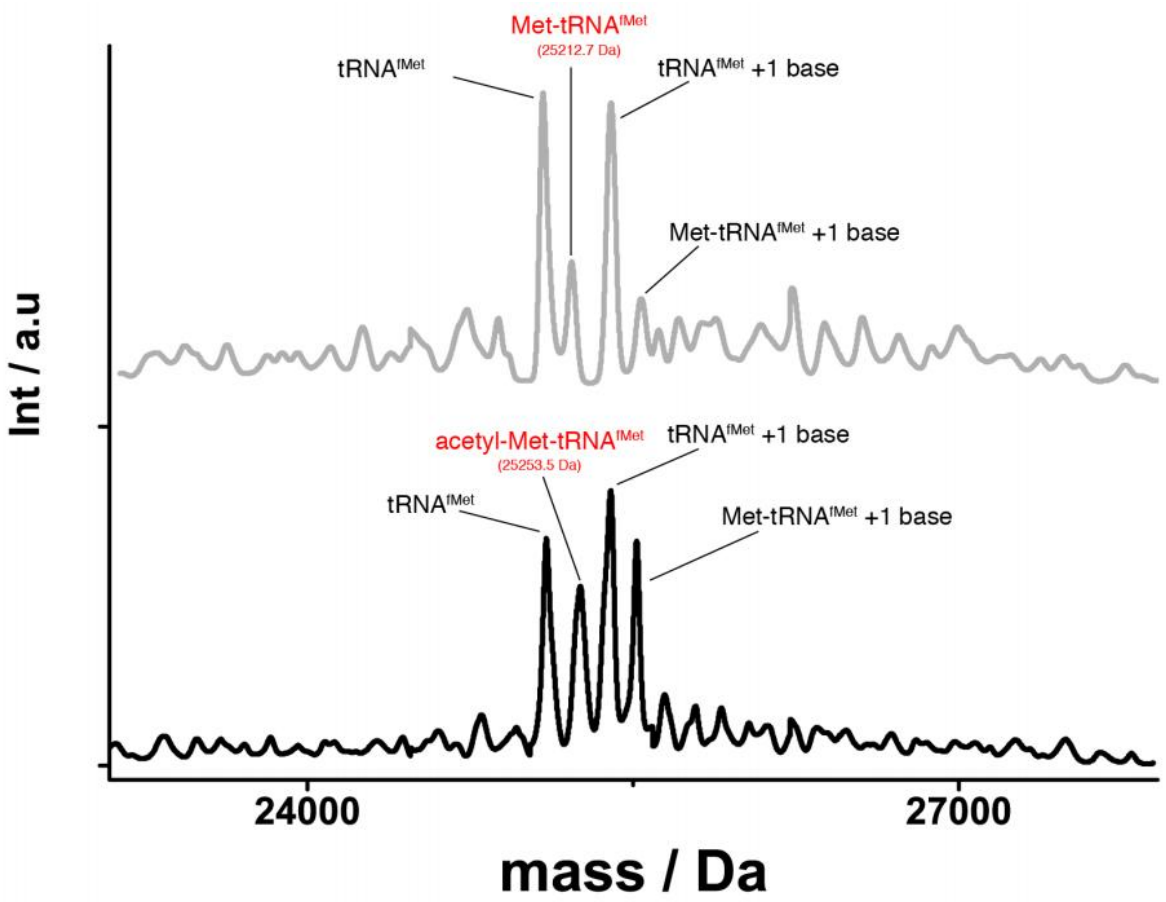


Figure 3: Native mass spectrum of Met-tRNA^{fMe}t in grey (25 212.1 Da) and Met-tRNA^{fMe}t treated with AtaT and acetyl-CoA in black (25 253.5 Da). The relevant peaks are assigned in red.

6.3.5 Conclusion

With mass spectrometry, the AtaT-AtaR complex could be detected in a 1:1 ratio, in addition the presence of a 2:2 complex was detected. Mass spectrometry of Met-tRNA^{fMet} before and after acetylation by AtaT confirmed that treatment with the enzyme results in an increase of mass of approximately 41 Da that corresponds to the approximate mass of an acetyl group. The full study can be found in the following article.

6.3.6 Sources

[1] Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-1407 (2002).

ARTICLE
PUBLISHED ONLINE: 3 APRIL 2017 | DOI: 10.1038/N CHEMBIO .2346

AtaT blocks translation initiation by N-acetylation of the initiator t RNA fMet

Dukas Jurėnas [1,2], Sneha Chatterjee [3], Albert Konijnenberg [3], Frank Sobott [3–5], Louis Droogmans [6], Abel Garcia- Pino [7]* & Laurence Van Melderen [1]*

Toxin–antitoxin (TA) loci are prevalent in bacterial genomes. They are suggested to play a central role in dormancy and persister states. Under normal growth conditions, TA toxins are neutralized by their cognate antitoxins, and under stress conditions, toxins are freed and inhibit essential cellular processes using a variety of mechanisms. Here we characterize ataR–ataT , a novel TA system, from enterohemorrhagic Escherichia coli. We show that the toxin AtaT is a GNAT family enzyme that transfers an acetyl group from acetyl coenzyme A to the amine group of the methionyl aminoacyl moiety of initiator tRNA. AtaT specifically modifies Met-tRNA ^{fMet}, but no other aminoacyl-tRNAs, including the elongator Met-tRNA ^{Met}. We demonstrate that once acetylated, AcMet-tRNA ^{fMet} fails to interact with initiation factor-2 (IF2), resulting in disruption of the translation initiation complex. This work reveals a new mechanism of translation inhibition and confirms Met-tRNA ^{fMet} as a prime target to efficiently block cell growth.

Toxin–antitoxin (TA) systems are widespread in bacterial mobile genetic elements and chromosomes[1–3]. They take part in the regulation of important processes such as plasmid stabilization and protection against phages[4,5]. Accumulating evidence indicates that TA systems are involved in the switch to a persister state (highly tolerant to antibiotics) and modulate cell growth under stress conditions[6–7].

TA modules are typically classified according to the nature and mode of action of their antitoxins[8,9]. These ubiquitous modules typically consist of a toxic protein and its cognate unstable antitoxin whose presence prevents toxin activity or synthesis. In type II TA modules, both components are proteins. During normal growth, antitoxins form a tight complex with their cognate toxins that neutralizes their activity. However, once the toxins are released, they target essential cellular processes, resulting in transient cell growth arrest.

Type II toxins present a variety of molecular mechanisms to achieve their various functions, such as targeting peptidoglycan synthesis[10], replication[11–13] and translation. For example, toxins from the RelE and MazF families cleave mRNAs with little specificity, with or without the assistance of ribosomes, respectively[14–16]. Toxins from the VapC family are PIN-domain endonucleases that specifically cleave tRNA and rRNA. Enteric VapCs cleave tRNA ^{fMet} at the anticodon stem–loop[17], whereas Mycobacterium tuberculosis VapC20 cleaves the sarcin–ricin loop of 23S rRNA[18]. HipA inhibits glutamyl-tRNA synthetase by phosphorylation, resulting in the inhibition of the production of Glu-tRNA^{Glu} (refs. 19,20), and Doc phosphorylates the elongation factor EF-Tu, thereby interfering with the formation of the ternary complex EF-Tu–GTP–aminoacyl-tRNA [21,22].

Here, we identify aminoacyl tRNA acetyltransferase toxin–repressor complex (AtaT–AtaR), a novel type II TA operon found in Escherichia coliO157:H7. The AtaT toxin contains a Gcn5-related N-acetyltransferase (GNAT) domain. GNAT enzymes modify a myriad of substrates, from small molecules, such as antibiotics and metabolites, to macromolecules[23]. We demonstrate that AtaT inhibits translation initiation by specifically acetylating the free amine group of methionine charged on tRNA^{fMet}, using acetyl coenzyme A (AcCoA) as an acetyl group donor. Acetylation of the initiator Met-tRNA ^{fMet} prevents interaction with IF2 and formation of an initiation complex compatible with translation initiation. This results in the efficient inhibition of protein synthesis and cell growth arrest.

RESULTS

ataRT is a TA system with a putative acetyltransferase toxin

ataT (locus Z4832) of E. coli O157:H7 is predicted to encode AtaT, a 19.7-kDa acetyltransferase from the GNAT family of proteins. The ataR gene (locus Z4833), located 6 bp upstream from ataT, encodes a putative 9.9-kDa RHH-domain protein, AtaR (Supplementary Results, Supplementary Fig. 1). GNAT domain proteins with similar genetic organization have been demonstrated to act as TA pairs[24,25]. On the basis of these observations, we hypothesized that this gene pair encoded a TA module.

In order to validate this putative operon as a bona fide TA module, we cloned the open reading frames (ORFs) in compatible vectors carrying different inducible promoters, and tested the effect on E. coli DJ624 viability. Expression of ataT, encoding the putative GNAT toxin AtaT, alone caused cell growth inhibition, while coexpression ofataR, encoding the putative antitoxin, restored cell viability (Fig. 1 a).

Next, we asked if AtaR–AtaT is a type II TA system, with AtaR (antitoxin) and AtaT (toxin) forming a tight complex resulting in the neutralization of AtaT activity. To test this hypothesis, we labeled the proteins from the ataR–ataT operon with two different affinity tags (His-antitoxin and toxin-strepII) and performed Ni-Sepharose affinity chromatography purification. The AtaR and AtaT proteins co-purified (Fig. 1 b), and a complex was separated from excess antitoxin after size exclusion chromatography (Fig. 1 c). Based on analytical gel filtration, we hypothesized that two toxin

[1]Laboratoire de Génétique et Physiologie Bactérienne, Université Libre de Bruxelles, Gosselies, Belgium. [2]Department of Biochemistry and Molecular Biology, Vilnius University Joint Life Sciences Center, Vilnius, Lithuania. [3]Biomolecular and Analytical Mass Spectrometry Group, Department of Chemistry, University of Antwerp, Antwerp, Belgium. [4]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK. [5]School of Molecular and Cellular Biology, University of Leeds, Leeds, UK. [6]Laboratoire de Microbiologie, Université Libre de Bruxelles, Gosselies, Belgium. [7]Laboratoire de Biologie Structurale et Biophysique, Université Libre de Bruxelles, Gosselies, Belgium. *e-mail: agarcia@ulb.ac.be or lvmelder@ulb.ac.be

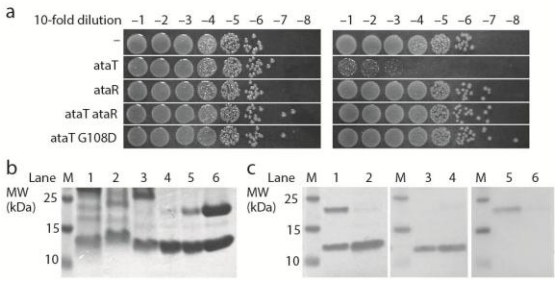


Figure 1 | The ataT–ataR gene pair constitutes a type II TA system. (a) Overnight cultures of *E. coli* strains transformed with pBAD33 and pKK223-3 vectors or derivatives expressing ataT, the ataT G108D mutant, ataR, or both ataT and ataR were serially diluted (10^{-1}- to 10^{-8}-fold). Dilutions were spotted on LB medium supplemented with appropriate antibiotics and 0.2% glucose (repression conditions, left) or 0.2% arabinose and 1 mM IPTG (induction conditions, right). (b) Ni-affinity purification of the His-AtaR–AtaT-strepII complex. Elution fractions were resolved by SDS–PAGE and stained with Coomassie blue. Lanes 1–6: fractions corresponding to 0, 20, 50, 80, 150 and 500 mM imidazole, respectively. (c) Subjecting fraction 6 to size exclusion chromatography resulted in two peaks corresponding to the His-AtaR–AtaT-strepII complex and His-AtaR antitoxin as seen on SDS–PAGE, stained with Coomassie blue (lanes 1 and 2), anti-His (lanes 3 and 4) or anti-strepII (lanes 5 and 6) western blots. M, molecular weight marker. Images of all of the full gels are shown in Supplementary Figure 12.

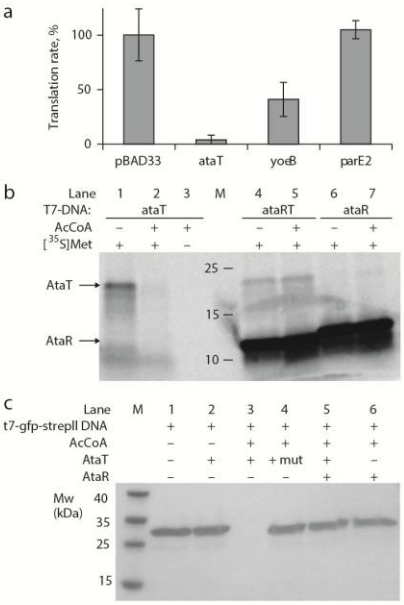


Figure 2 | AtaT inhibits translation in an AcCoA-dependent manner. (a) In vivo translation rate measured by incorporation of [^{35}S]methionine after 1 h expression either of ataT or of yoeB or parE2 (encoding type II toxins that inhibit translation and replication, respectively) as controls. Translation rate in the different strains is normalized to that in the strain containing the empty pBAD33 vector. Error bars indicate s.d. of three independent experiments. (b) Synthesis of AtaT, AtaR or both expressed from the T7 promoter in the in vitro transcription–translation system supplemented with [^{35}S]methionine. Reactions were carried out with AcCoA (lanes 2, 5 and 7) or without AcCoA (lanes 1, 4 and 6). Lane 3 is a control without [^{35}S]methionine. Samples were resolved by SDS–PAGE and exposed to phosphor storage screen. (c) Synthesis of GFP-strepII reporter protein expressed from the T7 promoter in in vitro transcription–translation system. Products of translation reactions were resolved by SDS–PAGE followed by western blot with anti-strepII tag antibodies. "Mut" indicates the use of the AtaT G108D mutant. Images of the full gels are shown in Supplementary Figure 12.

units and two antitoxin units associated to form a 60-kDa complex (Supplementary Fig. 2). Native MS showed that AtaR–AtaT is a heterogeneous mixture of complexes with different toxin:antitoxin ratios (Supplementary Fig. 3). The predominant species observed at low collision energy consisted of AtaT:AtaR at 1:1 and 2:2 ratios. TA complexes with variable stoichiometries have been described for other systems[26–29]. This is a crucial feature in the regulation of transcription of these operons, allowing a link between toxin neutralization and operon repression[30].

Taken together, these data showed that the ataR–ataT gene pair encoded a functional type II TA system similar to the recently described TacTA module from *Salmonella typhimurium*, which also contains a GNAT-domain toxin[24].

AtaT inhibits translation in the presence of acetyl-CoA

To determine which cellular process AtaT inhibits, we measured the incorporation of radiolabeled precursors for replication, transcription and translation upon toxin expression in vivo. The incorporation of [^{35}S]methionine, shown in Figure 2a, was severely affected, indicating that AtaT inhibits translation in vivo, without interfering with transcription or replication (Supplementary Fig. 4). As previously mentioned, AtaT is predicted to be an acetyltransferase, and it possesses the conserved GNAT family topology (Supplementary Fig. 5). We therefore reasoned that translation should be functional in an in vitro transcription–translation system in the absence of its potential substrate, AcCoA. Indeed, AtaT was produced to a detectable amount in the absence of AcCoA. Upon addition of AcCoA, the protein was no longer synthesized (Fig. 2b), showing that AcCoA was essential for catalysis. Addition of AcCoA did not interfere with the in vitro production of the AtaT–AtaR complex or the antitoxin AtaR (Fig. 2b). Moreover, in vitro translation of a reporter protein (GFP-strepII) was tested in different conditions. In the presence of purified AtaT toxin and AcCoA, GFP-strepII was not produced, while addition of AtaR antitoxin restored GFP-strepII synthesis (Fig. 2c). To validate the enzymatic activity of AtaT, a G108D mutation was introduced in the conserved AcCoA-binding pocket consensus sequence (Q/RxxGxG/A) (Supplementary Fig. 5)[31]. The G108D mutation inactivated AtaT both in vivo and in vitro (Figs. 1a and 2c, lane 4). Together, these data showed that in the presence of AcCoA, AtaT blocked translation, and that the antitoxin AtaR restored translation completely, even in the presence of AcCoA (Fig. 2c). Moreover, these data indicated that the target of AtaT was present in the in vitro transcription–translation system.

AtaT acetylates tRNAs

We used isotope-labeled [^{14}C]AcCoA to monitor the acetylation reaction catalyzed by AtaT. The product of the in vitro translation reaction, resolved on SDS–PAGE gels, revealed no isotope signal, suggesting that either the target was not a protein or the modification was not stable under SDS–PAGE conditions. Alternatively, we performed a size fractionation of the reaction product. The ^{14}C signal dotted on a nitrocellulose membrane is retained in fractions corresponding to a size between 30 and 50 kDa (Supplementary Fig. 6). This was confirmed by native PAGE of the in vitro translation reactions treated with AtaT and [^{14}C]AcCoA (Fig. 3a). Gels were

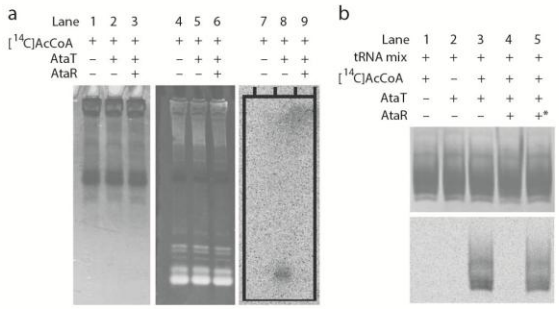


Figure 3 | AtaT acetylates tRNAs. (a) Acetylation reaction in the in vitro transcription–translation system supplemented with [^{14}C]Ac CoA. Reaction products resolved by native PAGE in three replicas and stained with Coomassie blue (lanes 1–3) or with ethidium bromide (lanes 4–6). Gel comprising lanes 7–9 was dried and exposed to phosphor storage screen. (b) In vitro acetylation reactions of tRNA mixture purified from E. coli supplemented with [^{14}C]Ac CoA. Reaction products were resolved by native PAGE, stained with methylene blue (upper panel) and exposed to phosphor storage screen (lower panel). Lane 1, control tRNAs with [^{14}C] Ac CoA; lane 2, control tRNAs with Ata T toxin; lane 3, tRNAs with Ata T and [^{14}C]Ac CoA; lane 4, AtaR was premixed with Ata T before acetylation reaction; lane 5, AtaR was added after acetylation reaction (indicated by *). The reaction was then allowed to continue for 30 min. Images of the full gels are shown in Supplementary Figure 12.

stained with either Coomassie blue or ethidium bromide or exposed to obtain an autoradiography image. Interestingly, the ^{14}C radioactive signal did not match the band pattern from the Coomassie-stained proteins, but did match that the bands corresponding to ethidium bromide–stained nucleic acids (Fig. 3a). Based on this migration pattern, it stands to reason that substrates of AtaT are tRNAs. To test this hypothesis, a purified mixture of tRNAs from E. coli was treated with AtaT and [^{14}C]AcCoA and resolved on a native PAGE gel. Autoradiography confirmed that tRNAs were acetylated by AtaT, and that AtaR specifically inhibited this reaction, but was unable to reverse it (Fig. 3b).

AtaT modifies the aminoacyl moiety of tRNA

tRNAs are extensively modified as part of their post-transcriptional maturation. N^4-Acetylcytidine is the only known acetylated nucleoside of bacterial tRNAs[32]. To test whether AtaT performed a similar type of modification, an E. coli tRNA mixture [^{14}C]acetylated by AtaT was subsequently digested with RNase P1, and the nucleotide products were resolved by 2D thin layer chromatography (TLC)[33]. Intriguingly, the radioactive signal did not correspond to the N^4-acetylcytidine position, and the migration pattern differed significantly from that observed from cognate nucleotides (Supplementary Fig. 7). Notably, the aminoacyl moiety at the CCA tail of aa-tRNAs constitutes an alternative acetylation site to the tRNA nucleotides. To examine this, we removed the amino acid moiety from the tRNA mixture either by alkaline treatment, known to disrupt weak ester bonds and de-acylate tRNAs[34], or by using $CuSO_4$. The latter treatment is known to specifically uncharge aminoacyl-tRNA, but not peptidyl-tRNA[35]. A purified E. coli tRNA mixture was subjected to either treatment before and after the acetylation reaction catalyzed by AtaT using [^{14}C]AcCoA as a substrate (Fig. 4). In both cases, no ^{14}C signal was detected when the aa-tRNA mixture was uncharged prior to acetylation, indicating that acetylation occurred only on aminoacylated tRNAs (Fig. 4a, lanes 3 and 5). Moreover, the signal was also lost when the aminoacyl moieties were removed from the tRNAs after the acetylation by alkaline treatment. This indicated that acetylation occurred on the aminoacyl moiety of aa-tRNAs (Fig. 4a, lane 4). By contrast, the $CuSO_4$ treatment failed to remove the amino acids from AtaT-acetylated aminoacyl-tRNAs (Fig. 4a, lane 6). This was a strong indication that upon AtaT treatment, the acetylated-aa–tRNA species no longer resembled translation-compatible aminoacyl-tRNA (sensitive to $CuSO_4$) and, rather, were closer to species where the amine group is not free but blocked, as in peptidyl-tRNAs. In addition, aa-tRNAs that had been chemically acetylated (by acetic anhydride treatment) were no longer acetylated by AtaT (Fig. 4a, lane 7), confirming that AtaT acetylated aa-tRNAs at the free amine group.

AtaT activity is specific to the methionine on Met-tRNA^{fMet}

To identify the target of AtaT in vitro, we produced and tested aminoacylated-tRNA species for each amino acid (Fig. 4b). We found that AtaT was highly specific to initiator tRNA^{fMet} charged with methionine (Met-tRNA^{fMet}) (Fig. 4b). Only weak acetylation was detected for other tRNAs, including the elongator Met-tRNA^{Met} (Fig. 4b). This suggested that AtaT not only recognized the CCA tail of aa-tRNA, but was also able to discriminate between the aa-tRNA species based on the acceptor stem and the aminoacyl moiety.

The treatment with AtaT and [^{14}C]AcCoA of all possible variants of tRNA^{fMet} (not aminoacylated, charged with methionine, charged with methionine and formylated, charged with methionine and chemically acetylated), either synthetized in vitro or purified in vivo, showed that the ^{14}C-labeled acetyl group was transferred only to the free amine group of tRNA^{fMet} charged with methionine (Fig. 4c,d and Supplementary Fig. 8).

MS of Met-tRNA^{fMet} before and after acetylation by AtaT confirmed that the enzyme treatment resulted in an increase in mass of approximately 40 Da, which corresponds to the approximate mass of an acetyl group (Supplementary Fig. 9).

To directly confirm these results, we performed alkaline treatment on AtaT-treated Met-tRNA^{fMet}. AtaT-modified or acetic anhydride–modified methionine (used as reference) removed from tRNA^{fMet} was analyzed by TLC and MS. The TLC results showed that after AtaT treatment, the recovered methionine was acetylated (Supplementary Fig. 10), and MS showed that AtaT treatment resulted in a mass approximately matching that of acetyl-methionine (Supplementary Fig. 11). Additionally, no acetyl-methionine was detectable from untreated Met- tRNA^{Met} (Supplementary Figs. 10 and 11).

Acetylation of Met-tRNA^{fMet} precludes translation initiation

The interaction between initiator tRNA and IF2 is crucial for correct assembly of the 30S initiation complex (30S IC). The recent structure of a 70S ribosome in complex with IF2 and fMet-tRNA^{fMet} showed that IF2 recognizes the aminoacyl moiety of fMet-tRNA^{fMet} (ref. 36). Therefore, we hypothesized that acetylation of Met-tRNA^{Met} could disrupt this interaction and interfere with translation initiation.

To test this, we used isothermal titration calorimetry (ITC) to measure the interaction of IF2 with fMet-tRNA^{fMet} and acMet-tRNA^{fMet}. In the absence of the other initiation complex components, IF2 bound fMet-tRNA^{fMet} with a K_d of ~1 μM (similar to previously reported values[36,37]). By contrast, IF2 affinity for acMet-tRNA^{fMet} was below detection level (Fig. 5a). Next, we tested the effects of initiator tRNA acetylation on the assembly of 30S IC. The 30S IC was reconstituted in vitro by incubating 30S ribosomes, mRNA, IF1 and IF2 with formylated or acetylated [^{35}S]Met-tRNA^{fMet}. The efficiency of complex formation was measured by the incorporation of ^{35}S-labeled initiator tRNA. Our results showed that there was a ten-fold decrease of 30S IC formation in the presence of acetylated Met-tRNA^{fMet} (acetylated enzymatically with AtaT or chemically with acetic anhydride) compared to formylated Met-tRNA^{Met} (Fig. 5b). Together, these data suggested that AtaT acetylation of initiator tRNA^{fMet} precluded the interaction of IF2 and acMet-tRNA^{Met},

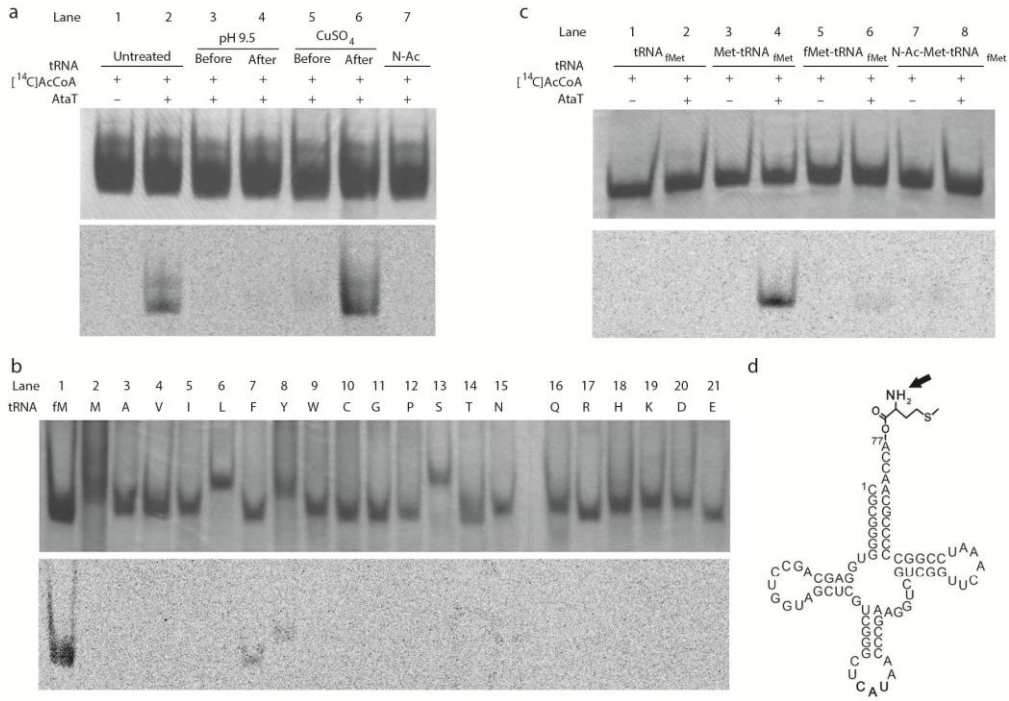


Figure 4 | AtaT acetylates the amine group of the methionine charged on the initiator tRNA. (a) In vitro acetylation reactions of aminoacyl moiety of E. coli tRNA mixture supplemented with [^{14}C]Ac CoA. Products were resolved by native PAGE, stained with methylene blue (upper panel) and exposed to phosphor storage screen (lower panel). Lane 1, control (E. coli tRNA mixture incubated with [^{14}C]Ac CoA); lane 2, addition of Ata T; lane 3, alkaline treatment of tRNA mixture using Tris-H Cl (pH 9.5) before acetylation; lane 4, same treatment after acetylation; lane 5, $CuSO_4$ treatment of tRNA mixture before acetylation; lane 6, same treatment after acetylation; lane 7, acetic anhydride treatment of tRNA mixture before acetylation. (b) Ata T acetylation reactions with individual aa-tRNA species using [^{14}C]Ac CoA. Products were resolved by native PAGE, stained with methylene blue (upper panel) and exposed to phosphor storage screen (lower panel). Lane 1, charged initiator tRNA (Met-tR NA^{fMet}); lanes 2–21, tRNAs charged with their respective amino acids. tRNA species are indicated in Online Methods. (c) Acetylation of synthetic tR NA^{fMet} species with [^{14}C]Ac CoA and with or without AtaT (controls). Reaction products were resolved by native PAGE, stained with methylene blue (upper panel) and exposed to phosphor storage screen (lower panel). Reactions were carried out using uncharged tRNA^{fMet} (lanes 1, 2), tR NA^{fMet} charged with methionine (lanes 3, 4), tR NA^{fMet} charged with methionine and formylated (lanes 5, 6), tR NA^{fMet} charged with methionine and chemically N-acetylated (lanes 7, 8). (d) Representation of initiator tRNA^{fMet}. The arrow indicates the amine group modified by AtaT. Images of the full gels are shown in Supplementary Figure 12.

resulting in the inhibition of translation initiation. To assess the validity of this model in vivo, we characterized ribosomal fractions from E. coli overexpressing ataT or both ataT and ataR. Our results showed that the expression of ataT alone led to extensive accumulation of ribosome assembly intermediates compared to control or ataR and ataT coexpression (Fig. 5c,d), which strongly supports our model.

DISCUSSION

Bacterial type II TA modules have been shown to activate under stress conditions. They are proposed to assist the stress survival machinery based on their ability to modulate key cellular processes and reversibly arrest cell growth. Translation is a preferred target of type II toxins[9]. They hijack translation at almost every step from mRNA, tRNA and rRNA cleavage to inactivation of translation factors[17–19,21,38]. We unravelled the mechanism of toxicity of the novel type II TA toxin AtaT. Here, we showed that AtaT inhibited translation by acetylating initiator Met-tRNA^{fMet} at the amine group of the methionine moiety.

AtaT belongs to a new class of bacterial N-acetyltransferases of the GNAT family. GNAT enzymes, which are found in all domains of life[23,39], acetylate a myriad of targets using AcCoA as a donor group. Knowledge on bacterial GNAT enzymes is scarce, with the best-studied cases being aminoglycoside-N-acetyltransferases and the three protein acetyltransferases RimI, RimJ and RimL that are known to acetylate the N-termini of ribosomal proteins S18, S5 and L12, respectively[23,40,41]. Notably, the GNAT enzyme TmcA is implicated in bacterial translation by modifying the anticodon (CUA) wobble base of elongator tRNA^{Met}, to prevent misreading of the similar AUA codon[32]. We showed that unlike TmcA, AtaT recognizes both the aminoacyl-CCA moiety and the double-stranded stem of initiator tRNA^{fMet}, and acetylates the amino acid moiety rather than the tRNA itself (Fig. 4b).

During the formation of the 30S IC, IF2 and fMet-tRNA^{fMet} must bind to the 30S subunit to prime it for subunit joining. IF2 recognizes the CCA-fMet end of fMet-tRNA^{fMet} via its β-barrel C2 domain[42,43] (Fig. 6a–c). The structure of the fMet-tRNA^{fMet}–70S ribosome complex shows the terminal A-fMet docks in the cavity formed between the β1-β2 and β4-β5 loops (Fig. 6c). It is worth noting that the fMet moiety is surrounded by a network of interactions involving the hydrocarbon region of the side chain of R847, and the π-electrons of the F848 ring of IF2. As shown in Figure 6c, the formyl group is in close proximity to the phenyl group of F848 (ref. 42); therefore, it is not surprising that when an acetyl group is

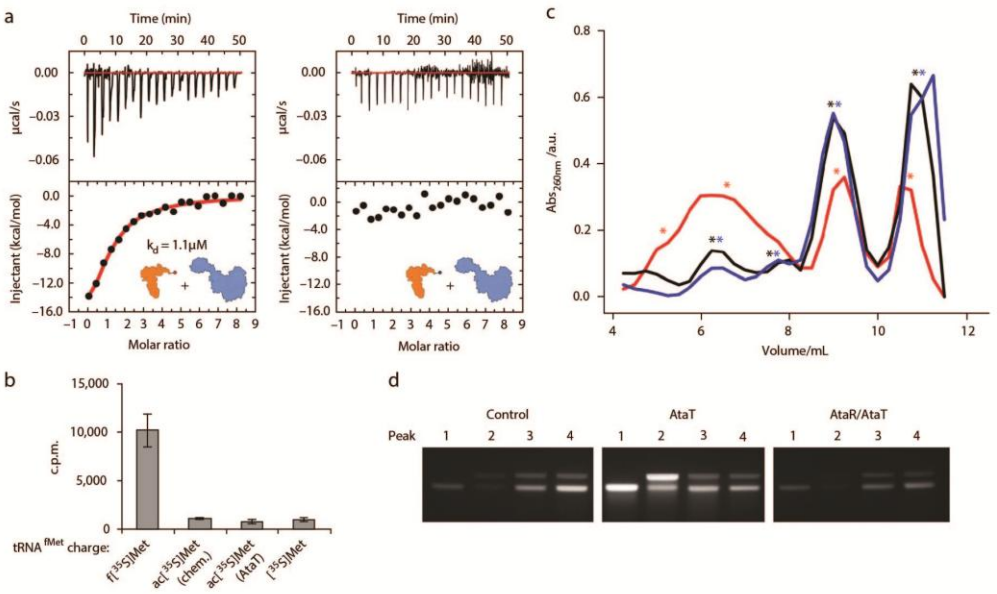


Figure 5 | Translation initiation inhibition in vitro and in vivo by AtaT. (a) Interaction of IF2 with fMet-tRNA^{fMet} (left) and acMet-tRNA^{fMet} (right) monitored by ITC. Top panels show corrected heat rate and bottom panels normalized heat. (b) In vitro formation of 30S IC using 30S ribosomes, IF1, IF2, mRNA and tRNA^{fMet} charged with [^{35}S]Met and modified as indicated on the x axis. The complex formation was allowed for 10 min, protein complexes were trapped on nitrocellulose filters and incorporation of isotope labeled tRNA^{fMet} was measured in scintillation counter (the data represent mean values ± s.d., each measurement was repeated at least three times). (c) Ribosome profiles. Cultures of E. coli strains transformed with pBAD33 and pKK223-3 vectors (black curve) or derivatives expressing the ataT gene (red curve) or the ataT and ataR genes (blue curve) were grown to an OD_{600nm} of 0.2. Arabinose was added at 0.2% for 1 h and cultures were treated with 0.5 mg/ml of chloramphenicol for 3 min. Cell extracts were centrifuged on sucrose gradient, fractions were collected top down and OD_{260nm} was measured. a.u., arbitrary units. (d) Fractions from c indicated with an asterisk (*) were used for rRNA extraction to confirm the identity of peaks 1–4 (left to right); rRNA was analyzed by agarose gel electrophoresis followed by staining with ethidium bromide. Images of the full gels are shown in Supplementary Figure 12.

modeled instead of a formyl group (Fig. 6d), the additional methyl moiety introduces clashes likely leading to structural rearrangements (Fig. 6d). Considering that the simple addition of a formyl group to Met-tRNA^{fMet} strongly enhances affinity and selectivity for IF2 (ref. 44), it is to be expected that a disruptive modification, such as acetylation, would have a catastrophic impact on the assembly of the initiation complex. The latter is particularly relevant because evidence suggests that simultaneous arrival of IF2 and fMet-tRNA^{fMet} to the 30S IC may dominate in vivo[37]. Moreover, since acMet-tRNA^{fMet} is a dead-end product, whereas fMet-tRNA^{fMet} is continuously used, the activity of AtaT will irrevocably lead to translation inhibition by the accumulation of acMet-tRNA^{fMet} (Fig. 6a). Indeed, we show that AtaT-dependent acetylation of Met-tRNA^{fMet} precludes the binding of IF2, formation of 30S IC in vitro and in vivo, and assembly of 70S ribosomes. On the basis of these data, we propose that in vivo AtaT efficiently competes with methionyl-tRNA formyltransferase to modify the methionine moiety of Met-tRNA^{fMet}.

In a recent work, TacT, a distant homolog GNAT toxin from Salmonella typhimurium (24% sequence identity with AtaT), was shown to acetylate multiple elongation tRNAs, thereby inhibiting translation at the elongation step[25]. Although the bases of this inhibition and its impact on ternary complex formation require further investigation, this work suggests that TacT has a broader specificity compared to AtaT, which is highly specific for the initiator Met-tRNA^{fMet}. This constitutes a remarkable functional divergence within this class of acetyltransferase type II toxins: one subfamily of toxins targets translation at the initiation step while another, with relaxed specificity, targets elongation.

This fuzzy, or relaxed, specificity seems to be a common functional feature within families of type II toxins. The members of the RelE family show differences in mRNA cleavage specificity[14] and even variation in their dependence on the ribosome for catalysis. Different VapC toxins cleave different tRNAs as well as 23S rRNA[17,18,45]. More strikingly, toxins from the Doc/Fic family show versatile molecular mechanisms and targets. While Doc phosphorylates the translation elongation factor EF-Tu, FicT AMPylates the GyrB and ParE subunits of DNA gyrase and topoisomerase IV, respectively[13,21,46]. It should be noted that sequence similarity within toxin families is usually low, something that is most likely a contributing factor to the observed broad range of specificity and activity. However, the selective pressure driving the divergence in specificity remains to be investigated.

In terms of physiological function, the TacTA system was shown to promote persister cells formation in *S. typhimurium*[24]. Further work on the impact of AtaT-dependent translation-initiation inhibition in the context of stress response and persistence is needed to unravel the biological roles of the *ataRT* system. The work presented here represents a crucial step forward in this challenging endeavor.

Received 13 May 2016; accepted 12 January 2017;
published online 3 April 2017;

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

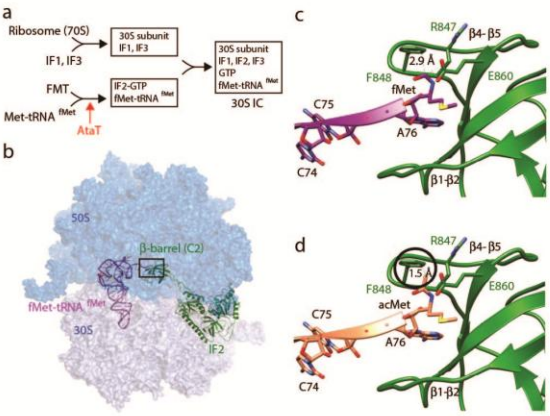


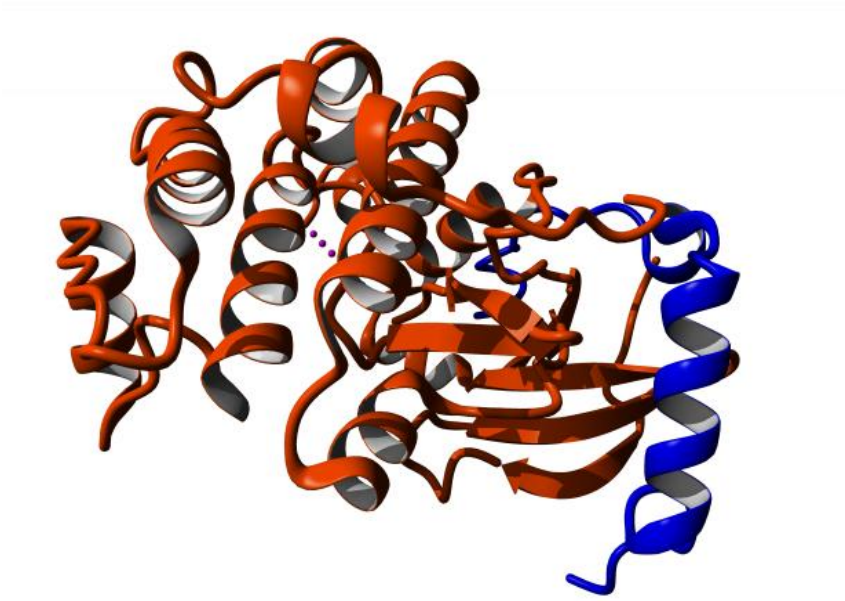
Figure 6 | Proposed mode of action of AtaT. (a) Scheme of translation initiation in bacteria. During this step of translation, the initiation factors IF1, IF2 and IF3 are tasked with ribosome subunit dissociation and anti-association, selection of initiator aa-tRNA, selection of the correct translation start site, and subunit joining at the start codon. AtaT (labeled in red) interferes with the initiation process by acetylating the initiator Met-tRNA^{fMet}. (b) Cryo electron microscopy structure of fMet-tRNA^{fMet} and IF2 bound to the E. coli ribosome (PDB ID 3JCJ; ref. 42). The 50S and 30S subunits are colored in blue, IF2 is shown in green and fMet-tRNA^{fMet} in magenta. The A-formyl-Met end is recognized by the C2 β-barrel domain of IF2. (c) Detailed view on the interaction between IF2 and fMet-tRNA^{fMet} at the CCA end. The aliphatic side chains of R847 and E860 together with the aromatic ring of F848 enclose the formylmethionine. Notably, the phenyl group of F848 is in close contact with the formyl group. (d) If an acetyl group is modeled on the amine moiety of methionine, the distances between the extra methyl group and the ring of F848 become less than 1.5 Å (black circle).

References

1. Leplae, R. et al. Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* 39, 5513–5525 (2011).
2. Makarova, K.S., Wolf, Y.I. & Koonin, E.V. Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. Biol. Direct 4, 19 (2009).
3. Pandey, D.P. & Gerdes, K. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* 33, 966–976 (2005).
4. Magnuson, R.D. Hypothetical functions of toxin-antitoxin systems. J. Bacteriol. 189, 6089–6092 (2007).
5. Van Melderen, L. Toxin-antitoxin systems: why so many, what for? Curr. Opin. Microbiol. 13, 781–785 (2010).
6. Maisonneuve, E. & Gerdes, K. Molecular mechanisms underlying bacterial persisters. Cell 157, 539–548 (2014).
7. Brauner, A., Fridman, O., Gefen, O. & Balaban, N.Q. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat. Rev. Microbiol.* 14, 320–330 (2016).
8. Yamaguchi, Y. & Inouye, M. Regulation of growth and death in Escherichia coli by toxin-antitoxin systems. *Nat. Rev. Microbiol.* 9, 779–790 (2011).
9. Hayes, F. & Van Melderen, L. Toxins-antitoxins: diversity, evolution and function. *Crit. Rev. Biochem. Mol. Biol.* 46, 386–408 (2011).
10. Mutschler, H., Gebhardt, M., Shoeman, R.L. & Meinhart, A. A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. PLoS Biol. 9, e1001033 (2011).
11. Bernard, P. & Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J. Mol. Biol. 226, 735–745 (1992).
12. Jiang, Y., Pogliano, J., Helinski, D.R. & Konieczny, I. ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase. Mol. Microbiol. 44, 971–979 (2002).
13. Harms, A. et al. Adenylylation of gyrase and topo IV by FicT toxins disrupts bacterial DNA topology. *Cell Rep.* 12, 1497–1507 (2015).
14. Goeders, N., Drèze, P.L. & Van Melderen, L. Relaxed cleavage specificity within the RelE toxin family. J. Bacteriol. 195, 2541–2549 (2013).
15. Pedersen, K. et al. The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. Cell 112, 131–140 (2003).
16. Zhang, Y. et al. MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. Mol. Cell 12, 913–923 (2003).
17. Winther, K.S. & Gerdes, K. Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. Proc. Natl. Acad. Sci. USA 108, 7403–7407 (2011).
18. Winther, K.S., Brodersen, D.E., Brown, A.K. & Gerdes, K. VapC20 of Mycobacterium tuberculosis cleaves the sarcin-ricin loop of 23S rRNA. Nat. Commun. 4, 2796 (2013).
19. Germain, E., Castro-Roa, D., Zenkin, N. & Gerdes, K. Molecular mechanism of bacterial persistence by HipA. Mol. Cell 52, 248–254 (2013).
20. Kaspy, I. et al. HipA-mediated antibiotic persistence via phosphorylation of the glutamyl-tRNA-synthetase. Nat. Commun. 4, 3001 (2013).
21. Castro-Roa, D. et al. The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. Nat. Chem. Biol. 9, 811–817 (2013).
22. Garcia-Pino, A. et al. Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. J. Biol. Chem. 283, 30821–30827 (2008).
23. Vetting, M.W. et al. Structure and functions of the GNAT superfamily of acetyltransferases. Arch. Biochem. Biophys. 433, 212–226 (2005).
24. Cheverton, A.M. et al. A Salmonella toxin promotes persister formation through acetylation of tRNA. Mol. Cell 63, 86–96 (2016).
25. Iqbal, N., Guérout, A.M., Krin, E., Le Roux, F. & Mazel, D. Comprehensive functional analysis of the 18 Vibrio cholerae N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron. J. Bacteriol. 197, 2150–2159 (2015).
26. Dao-Thi, M.H. et al. Intricate interactions within the ccd plasmid addiction system. J. Biol. Chem. 277, 3733–3742 (2002).
27. Overgaard, M., Borch, J., Jørgensen, M.G. & Gerdes, K. Messenger RNA interferase RelE controls relBE transcription by conditional cooperativity. Mol. Microbiol. 69, 841–857 (2008).
28. Garcia-Pino, A. et al. Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. Cell 142, 101–111 (2010).
29. Afif, H., Allali, N., Couturier, M. & Van Melderen, L. The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. Mol. Microbiol. 41, 73–82 (2001).
30. Loris, R. & Garcia-Pino, A. Disorder- and dynamics-based regulatory mechanisms in toxin-antitoxin modules. *Chem. Rev.* 114, 6933–6947 (2014).
31. Neuwald, A.F. & Landsman, D. GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. Trends Biochem. Sci. 22, 154–155 (1997).
32. Ikeuchi, Y., Kitahara, K. & Suzuki, T. The RNA acetyltransferase driven by ATP hydrolysis synthesizes N4-acetylcytidine of tRNA anticodon. EMBO J. 27, 2194–2203 (2008).
33. Grosjean, H., Keith, G. & Droogmans, L. Detection and quantification of modified nucleotides in RNA using thin-layer chromatography. Methods Mol. Biol. 265, 357–391 (2004).
34. Schuber, F. & Pinck, M. On the chemical reactivity of aminoacyl-tRNA ester bond. I. Influence of pH and nature of the acyl group on the rate of hydrolysis. Biochimie 56, 383–390 (1974).
35. Janssen, B.D., Diner, E.J. & Hayes, C.S. Analysis of aminoacyl- and peptidyl-tRNAs by gel electrophoresis. Methods Mol. Biol. 905, 291–309 (2012).
36. Mitkevich, V.A. et al. Thermodynamic characterization of ppGpp binding to EF-G or IF2 and of initiator tRNA binding to free IF2 in the presence of GDP, GTP, or ppGpp. J. Mol. Biol. 402, 838–846 (2010).
37. Tsai, A. et al. Heterogeneous pathways and timing of factor departure during translation initiation. Nature 487, 390–393 (2012).
38. Neubauer, C. et al. The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE. Cell 139, 1084–1095 (2009).
39. Vetting, M.W., Bareich, D.C., Yu, M. & Blanchard, J.S. Crystal structure of RimI from Salmonella typhimurium LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18. Protein Sci. 17, 1781–1790 (2008).
40. Tanaka, S., Matsushita, Y., Yoshikawa, A. & Isono, K. Cloning and molecular characterization of the gene rimL which encodes an enzyme acetylating ribosomal protein L12 of Escherichia coli K12. Mol. Gen. Genet. 217, 289–293 (1989).
41. Yoshikawa, A., Isono, S., Sheback, A. & Isono, K. Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12. Mol. Gen. Genet. 209, 481–488 (1987).
42. Sprink, T. et al. Structures of ribosome-bound initiation factor 2 reveal the mechanism of subunit association. Sci. Adv. 2, e1501502 (2016).
43. Guenneugues, M. et al. Mapping the fMet-tRNA(f)(Met) binding site of initiation factor IF2. EMBO J. 19, 5233–5240 (2000).
44. Milon, P. et al. The ribosome-bound initiation factor 2 recruits initiator tRNA to the 30S initiation complex. *EMBO Rep.* 11, 312–316 (2010).

Chapter 7

Characterisation of kinases and their complexes



In collaboration with
Dr. Jon Elkins
Dr. Kamal Abdul Azeez

Structural Genomics Consortium, University of Oxford, UK

7.1 Introduction

Kinases are enzymes that add phosphates to specific substrates and can influence the function of the modified substrate. As kinases are involved in various cell processes, any disruption in their activity can have consequences. Abnormal kinase activity has been associated with diseases including cancer, arthritis and neurologic disorders [1]. In order to develop therapeutics to target kinase, these kinases need to be fully characterized.

Most pharmaceutical companies have programs in the field of kinase modulation [2]. However, they have become cautious due to the often severe side effects and off-target effects, e.g. toxicity. Understanding allosteric effects is the main focus at the moment. Until now the most used type of drug to target kinases are small-molecule kinase inhibitors, as they can be delivered orally, are amenable to large-scale production and their chemical structure can be easily fine-tuned. Besides small molecules, monoclonal antibodies and RNA are also being targeted against kinases [3]. However, not all patients respond to kinase inhibitors in the same way. For instance when treating cancer, there can be different routes within the kinase network by which cells can proliferate. Therefore, drug development can benefit from a better understanding of the kinase activity expressed in each cancer cell population [4].

In this study we investigated the higher-order structure of Aurora kinase C – INCENPA complexes. There are three Aurora kinases in mammals, Aurora kinase A, Aurora kinase B and Aurora kinase C (AURKA, AURKB and AURKC). Although the kinase domains of the Aurora kinases have significant homology (AURKC has 84% and 68% sequence identity to AURKB and AURKA respectively), they have different regulatory mechanisms and perform different functions.
Aurora kinase C (AURKC) is a part of the Chromosomal Passenger Complex (CPC) which is involved in mitosis, but the exact function of Aurora C is still unknown. INCENPA or inner centromere protein A is a centromere-interacting protein that is also a part in of the chromosomal passenger complex and is involved in the regulation of Aurora C. AURKC is bound to the C-terminal 'IN-box' region of INCENP which contains a conserved C-terminal TSS motif. It has been shown that the phosphorylation of the TSS motif increases AURKB activity. This points to the essential and fundamental role of the TSS motif in INCENP. In this chapter AURKC-INCENP A complexes were characterised with various INCENP mutants in order to obtain insight into the structural basis of their interaction.

Due to their role in mitotic regulation, aurora kinases are seen as targets for inhibition for treatment of cancers [5]. Understanding the structure, flexibility and conformational variability of these kinase complexes will help us to get a better view of their dynamic behavior. The goal of this study is to characterize the higher-order structure of three variants of the Aurora C – INCENPA complex and the influence of a kinase inhibitor on the kinase complex structure.

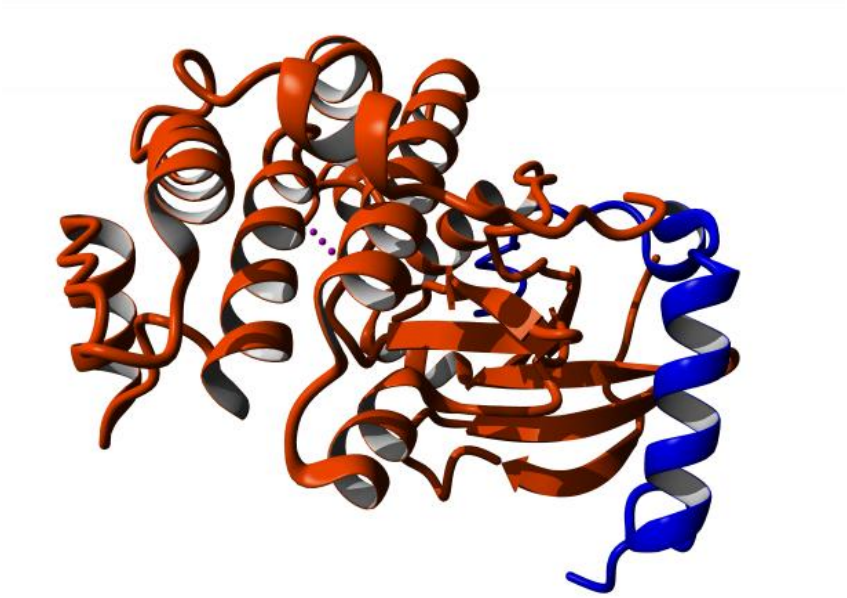


Figure 3: Illustration of Aurora C (red)-INCENPA (blue) kinase complex

7.2 Experimental details

The wild type and mutant variants of Aurora C-INCENPA and the kinase inhibitor, VX-680, were provided by the Structural Genomics Consortium (SGC), University of Oxford. Ammonium acetate buffer was purchased from Sigma Aldrich.
The native IM-MS measurements were performed on a Synapt G2 HDMS (Waters). The samples were buffer exchanged into 250 mM ammonium acetate using Micro Bio-spin P6 columns (Bio-Rad), and sprayed at gentle instrument tuning conditions; capillary voltage 1.2 kV, sampling cone 30 V, trap CE 5.0 V, transfer CE 0.0 V, trap bias 45 V, nanoflow gas pressure 0.1 bar, backing pressure 3 mbar, source temperature 30 ^{o}C, IMS wave velocity 300 m/s, IMS wave height 35 V.

7.3 Native mass spectrometry of Aurora C:INCENPA complex

We looked at the intact mass of the wild type AuroraC-INCENPA (**Fig. 1**) and two variants of the complex with mutated INCENP. The first variant has one mutation (S894A) in INCENPA, the second variant has three mutations (S893A, S894A, W897A) in INCENPA. S894A prevents the majority of the hydrogen bonding around the phosphorylated TSS motif while the triple-mutant complex was prepared to give the maximum chance of distinguishing the effect of the TSS motif. **Figure 2** shows the native mass spectra of the three variants under identical conditions. The three complexes have the same charge state distribution at mass range 2500 to 4000 m/z, and the mutated variants also show dissociated INCENP at 900-2000 m/z. The dissociated aurora kinase is also observed between peaks of the complex at 2500-3500 m/z. This could imply that the complexes with mutated INCENP are less stable than the wild type Aurora C-INCENPA complex. The experimental masses of the three complexes are 40,747.2 Da for the wild type complex, 40,224.3 Da for the single mutant S894A and 40,090.8 Da for the triple mutant S893A/ S894A/W897A. Their corresponding theoretical masses are 40500.5 Da,

40059.0 Da and 39927.9 Da respectively. The differences between the experimental and theoretical masses of ca. 247, 165 and 163 Da respectively correspond to 3 (240 Da) or 2 (160 Da) phosphorylations.

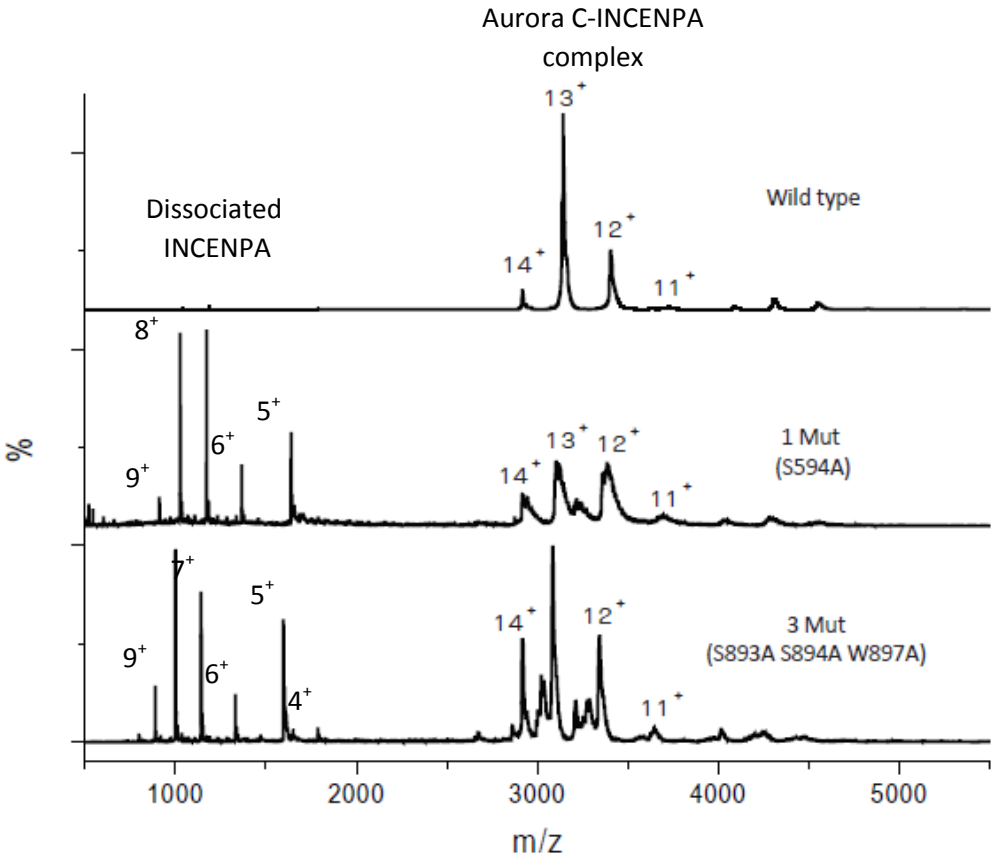


Figure 2: Native MS of the three variants show charge states 11^+ to 14^+ of the complex.

7.4 Collision cross-section of Aurora C:INCENPA complex

The rotationally-averaged collision cross-section (CCS) of a protein is related to its structure and conformational states. The collision cross sections of the three variants of Aurora C-INCENPA complex are displayed in **Figure 3** for each charge state. We hypothesised that the mutant complexes would have greater structural flexibility in the region of the TSS motif, which would be less strongly bound to Aurora C. The mutant complexes would therefore have a different average shape and cross-sectional area compared to wild-type.The mutated variants show a slightly more compacted CCS value than the wild type complex. The collision cross sections of the kinase complexes range from 3084 to 3529 $Å^2$ for the wild type complex, from 3065 to 3502 $Å^2$ for the single mutant and from 3046 to 3476 $Å^2$ for the triple mutant. This could imply that these specific mutations in the INCENPA subunit cause a conformational change, resulting in a smaller CCS value.

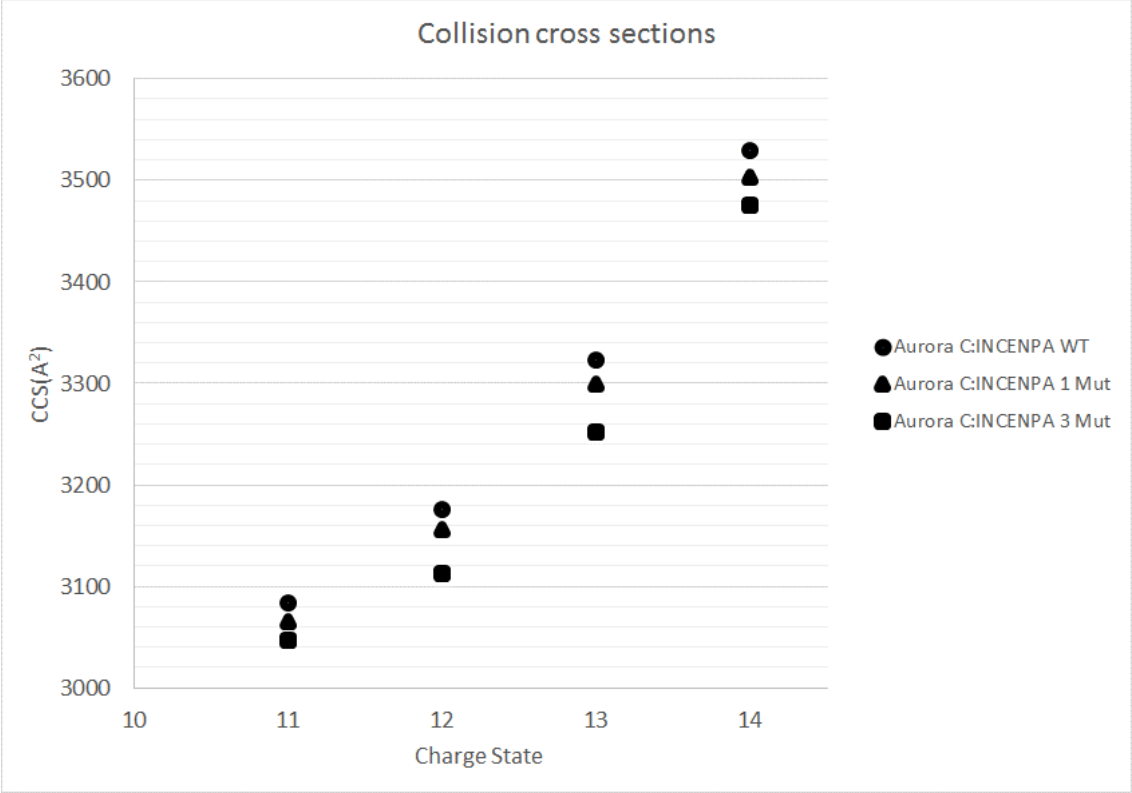


Figure 3: Collision cross sections of Aurora C-INCENPA Wild type (●), single mutant (▲) and triple mutant (■).

7.5 Collision induced unfolding of Aurora C:INCENPA complex

We can induce protein unfolding by increasing the collision energy in the trap cell of the Synapt G2. The trap collision energy was increased in a stepwise manner and the CCS monitored. These collision-induced unfolding plots describe the conformational stability of the proteins (**Figure 4**). The three variants show slight differences in transition states. In **Figure 4A**, the wild type complex shows a first transition at 30 V, then at 40 V and at 50 V it remains constant till 100 V. The single mutant complex shows the same transitions (**Figure 4B**), with a slight transition starting already at 20 V, which could imply that the mutated complex is less stable than the wild type complex. The triple mutant complex shows transitions at 30 V and 40 V, after which it stays constant till 90 V (**Figure 4C**). The triple mutant complex also is less stable as it the CCS value jumps more quickly to around 4000 $Å^2$ compared to the other two variants.

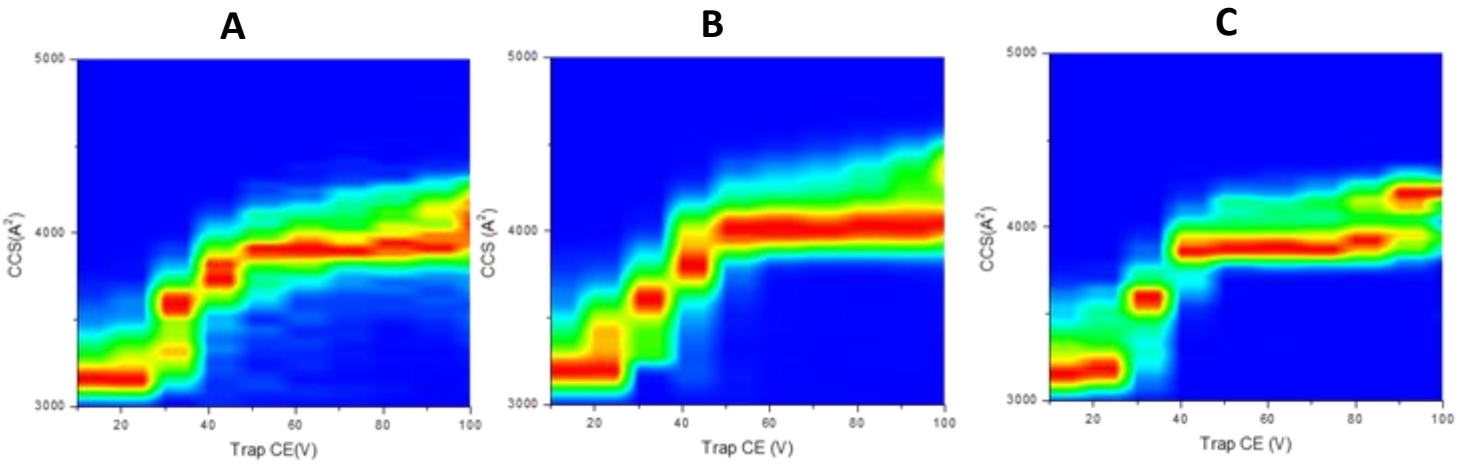


Figure 4: Collision induced unfolding plots of the 13^{+} charge state of Aurora C-INCENPA Wild type (A), single mutant (B) and triple mutant (C).

7.6 Binding of kinase inhibitor, VX-680, to the Aurora C-INCENPA complex

VX-680 is a small molecule with a mass of 464 Da **(Figure 5)**, known to inhibit Aurora C. With native mass spectrometry the binding of the kinase inhibitor to the Aurora C – INCENPA complex was detected at a 2:1 kinase to inhibitor ratio **(Figure 6**). We could also observe that the charge state distribution shifted to lower values, when the inhibitor was bound to the complex. This could suggest a change in structure due to the binding of the inhibitor.

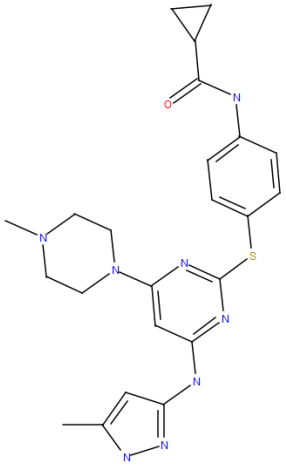


Figure 5: Chemical structure of VX-680. 464 Da small molecule that functions as a kinase inhibitor to Aurora C-INCENPA.

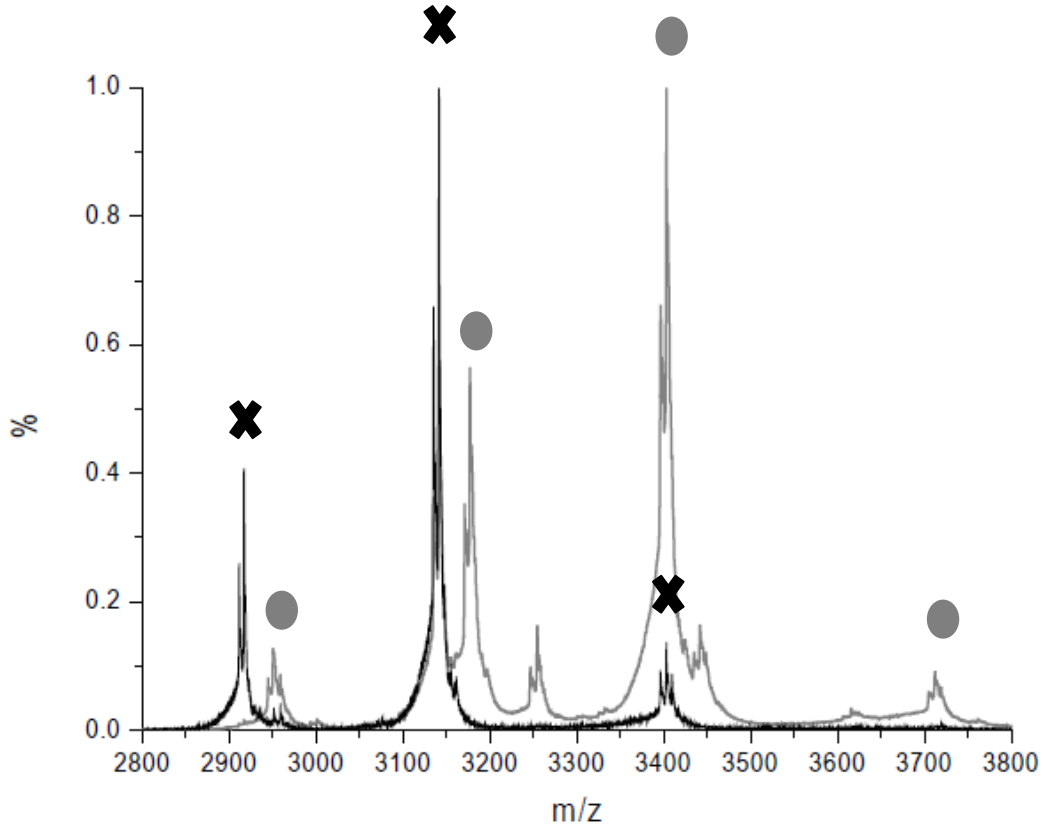


Figure 6: Native mass spectra of Aurora C- INCENPA complex (X) and of the complex bound with 1 inhibitor (●) when the inhibitor is added to a 2:1 complex to inhibitor ratio.

Different molar ratios between kinase complex and inhibitor were analysed with mass spectrometry in order to control the binding affinity of the inhibitor. **Figure 6** shows that binding was observed at a wild type to inhibitor ratio of 2:1.

Figure 7 displays the effect on the observed mass spectra when more inhibitor is added to the kinase complex the relative intensity of the inhibited complex increases. Furthermore, at a wild type to inhibitor ratio of 1:5, the peaks shift to a higher m/z range, i.e. charge state reduction occurs

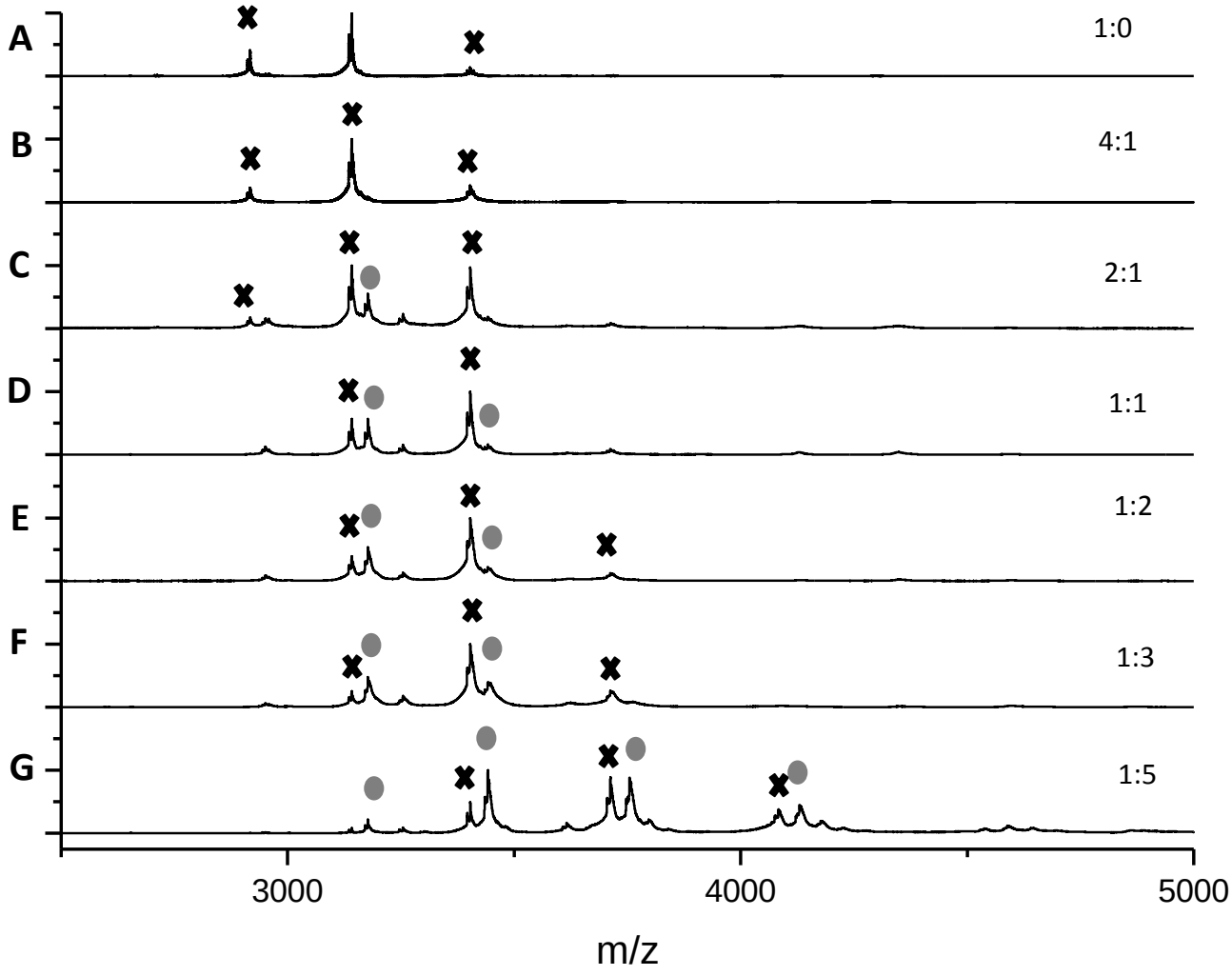


Figure 7: Native mass spectra of Aurora C- INCENPA complex with different complex to inhibitor ratios; 1:0 (A); 4:1 (B); 2:1 (C); 1:1 (D)); 1:2 (E); 1:3 (F); 1:5 (G). The peaks labelled with (X) represent the unbound wild type complex and the peaks labelled with (●) represent the complex bound with 1 inhibitor.

7.7 Conformational changes due to binding of inhibitor to the aurora kinase complex

With ion mobility-mass spectrometry, we looked for any conformational topological changes in the protein complex structure after binding with the inhibitor, VX-680. By selecting one charge state (13^+) and converting its corresponding ion mobility drift time to collision cross section, a distribution of the CCS values for this charge state (i.e. arrival time distribution) can be displayed (**Figure 8**). While the wild type complex shows one abundant peak at around 3300 $Å^2$, the inhibitor-bound complex shows a wider range of conformations at reduced CCS values. This can also be linked to the charge distribution shift of the inhibited complex seen in the native mass spectra (**Figure 6**).

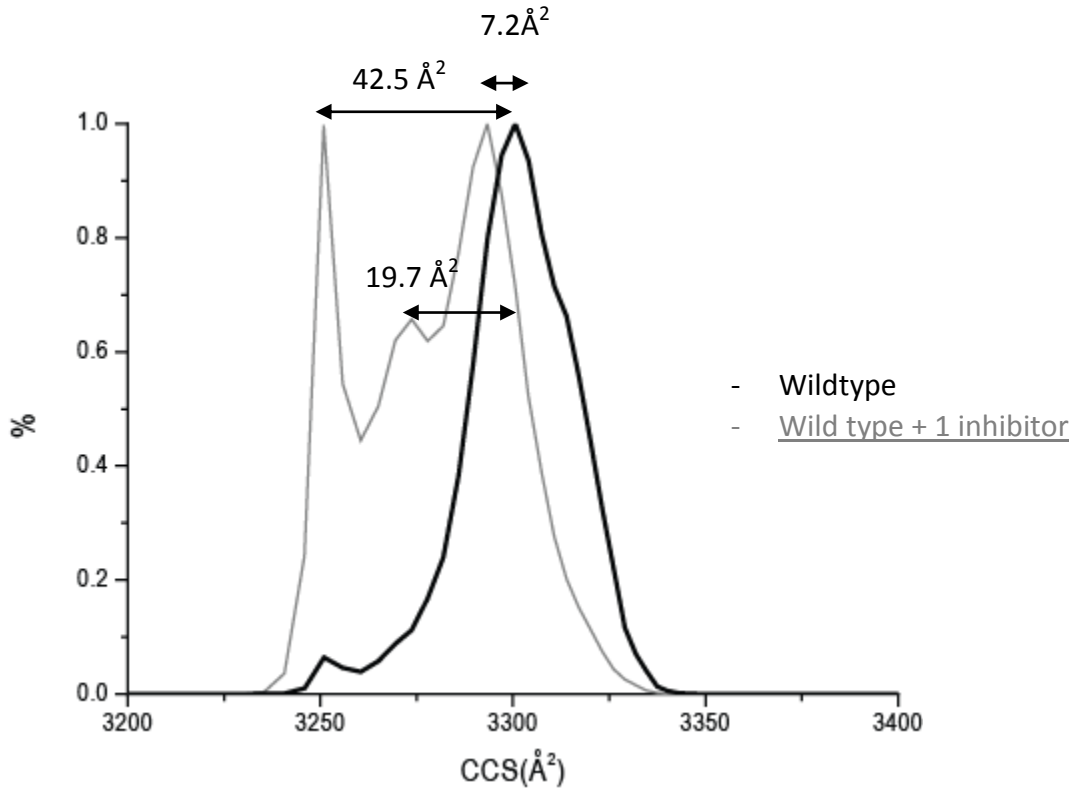


Figure 8: Collsion cross section of the 13^+ charge state of the Aurora C – INCENPA wild type complex (black) and of the complex bound with 1 inhibitor (grey). The differences in CCS values are noted above the CCS distributions.

Figure 9 shows how the collision cross section of the kinase complex is affected when adding increasing amounts of inhibitor. With ion mobility the inhibited complex is already observed at 1:0.25 wild type to inhibitor ratio, proving the ion mobility method improves the sensitivity. By adding the inhibitor a more compact form of the protein complex appears stabilised, however the same CCS values are observed when increasing the amount of inhibitor added.

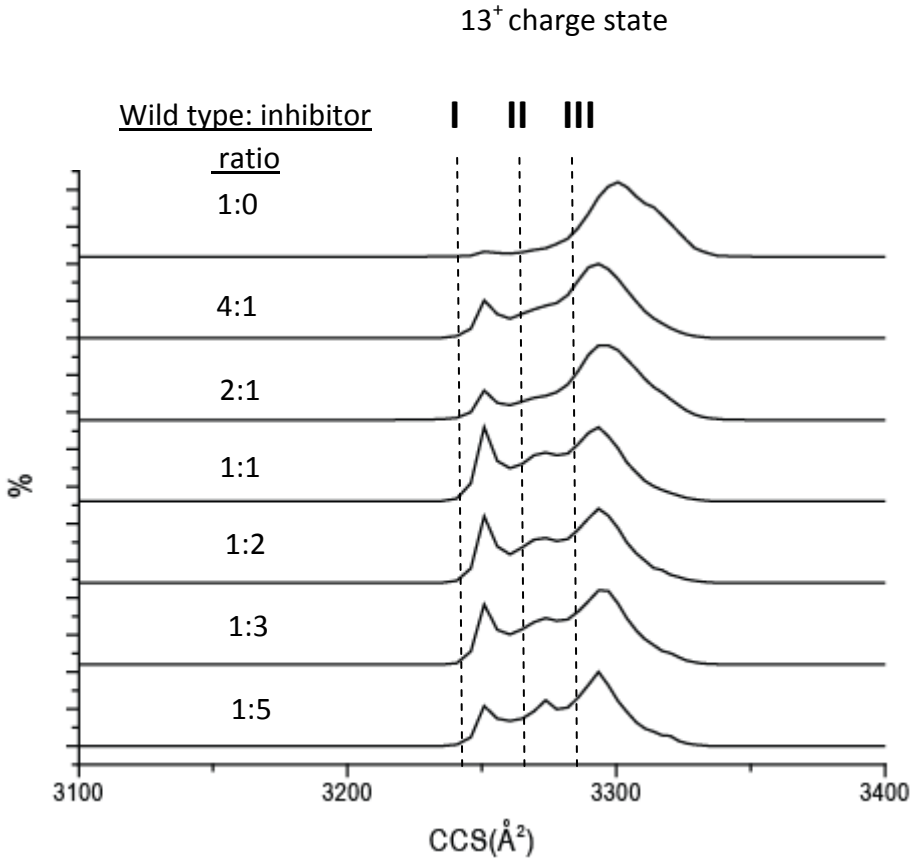


Figure 9: Collision cross section of the 13^+ charge state Aurora C - INCENPA complex with different complex to inhibitor ratios. Three distinct conformations could be observed (labeled I, II and III).

7.8 Conclusion

Native ion mobility-mass spectrometry allowed us to distinguish three variants of the Aurora C-INCENPA complex. With native mass spectrometry alone we could show that the mutated complexes were slightly less stable than the wild type complex, as the mutated INCENPA would more easily dissociate from the Aurora C kinase.
When comparing the ion mobility drift times of the three variants, the mutated complexes showed a more compact CCS value compared to the wild type complex. We looked at the stability of the three complexes by comparing their collision induced unfolding plots. The wild type complex needed more energy to unfold compared to the mutated complexes.
We looked at the binding affinity of the kinase inhibitor, VX-680, to the wild type kinase complex. With native mass spectrometry the binding of one inhibitor to the complex was observed at a wild type kinase complex to inhibitor ratio of 1:0.5. The inhibited kinase complex showed a shift in charge state distribution to higher m/z range, suggesting a change in conformation. This was confirmed with the performed ion mobility mass spectrometry results, where the addition of the kinase inhibitor induced a conformational change with lower CCS values and a wider range of conformations.
Together with the corresponding crystal structures, which have been generated at the SGC at the University of Oxford, this allows us to paint a more dynamic picture of the structural changes associated with these kinase interactions (an article is being drafted at the moment).

7.9 Sources

[1]L. Beltran, P. Casado, J.C. Rodriguez-Prados, P.R. Cutillas, Global profiling of protein kinase activities in cancer cells by mass spectrometry, Journal of Proteomics (2012), 77, 492-503
[2] C.J. Wenthur, P. R. Gentry, T. Mathews, C. W. Lindsley., Drugs for allosteric sites on receptors, Annu Rev Pharmacol Toxicol (2014), 54, 164-184
[3] M.K. Paul, A.K. Mukhopadhyay, Tyrosine kinase – role and significance in cancer, Int J Med Sci (2004), 1(2), 101-115
[4] C.L. Sawyers, The cancer biomarker problem, Nature (2008), 452, 548-552
[5] F. Girdler, K.E; Gascoigne, P.A. Eyers, S. Hartmuth, C. Crafter, K. M. Foote, N. J. Keen, S. S. Taylor, Validating Aurora B as an anti-cancer drug target, Journal of Cell Science (2006) 119, 3664-3675

Chapter 8

Studies of proteins of biopharmaceutical interest

In collaboration with:

SMC study

Dr. Gemma Fisher
Prof. Mark Dillingham
University of Bristol

CRL5^{SOCS2} Study

Dr. Esther M. Martin[1]
Dr. Emil Bulatov[2]
Prof. Alessio Ciulli[2]
[1]University of Antwerp, now MedImmune
[2]University of Dundee

Part of this research was published in

Emil Bulatov, Esther M. Martin, Sneha Chatterjee, Axel Knebel, Satoko Shimamura, Albert Konijnenberg, Clare Johnson, Nico Zinn, Paola Grandi, Frank Sobott and Alessio Ciulli, Journal of Biochemical Chemistry (2014) 290, 4178-4191

8.1 Introduction

Native nano-electrospray ionization-mass spectrometry can analyse large intact proteins and protein complexes up to the mega Dalton range. Snijder *et al.* have used native MS to measure 18 MDa virus assemblies [1]. They modified commercial MS instrument (QTOF) to improve the transmission of large ions, resulting in accurate mass analysis of intact protein complexes. Native MS studies have also been performed on 800 kDa non-covalent protein complex, GroEL. This tetradecameric chaperone complex is a popular high-mass standard for TOF MS systems and has been used to study intact mass studies, dissociation studies and conformational studies using MS-based methods [2,3,4].

To optimise the selection of ions up to an m/z value of 32,000, the instrument is usually calibrated over the range of 1,000 to 32,000 m/z using a solution of caesium iodide as this compound can form charged salt clusters, $(CsI)_nCs^+$ that extend over a wide mass range.

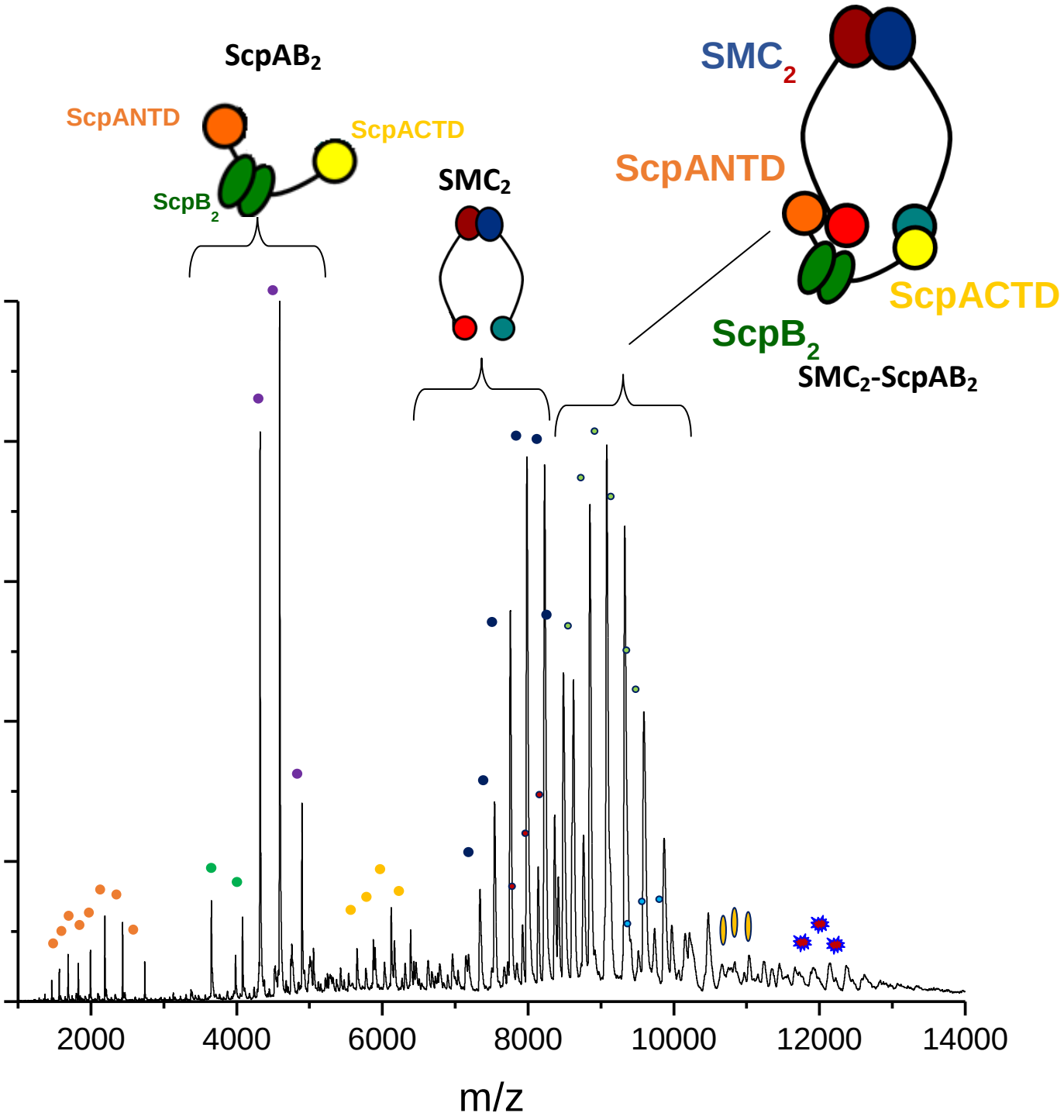
During my PhD, native MS was used to analyse intact protein complexes (e.g., 40 kDa Aurora kinase complexes (Chapter 7)), large intact proteins (e.g., 150 kDa antibodies (Chapter 3)) and large intact protein complexes (e.g., 190 kDa antibody-antigen complexes (Chapter 5)).

8.2 Structural maintenance of chromosomes

Another large protein complex I studied was the Structural Maintenance of Chromosome protein (SMC) in complex with segregation and condensation proteins, ScpA and ScpB. The SMC protein is a 270 kDa dimer that regulates chromosomes for proper replication and distribution of DNA [5,6]. This SMC protein can form complexes with ScpA (29 kDa) and ScpB (22 kDa), which assist in the binding of the protein complex to the chromosomes [7].

Understanding the structure of the SMC protein is important as it is involved in not only in mitosis but also DNA organization, e.g., condensation. Understanding the structure and architecture of these complexes can aid in uncovering mechanisms of how they behave during the cell cycle and in disorders such as cancers.

Figure 1 shows the mass spectrum of SMC protein interacting with ScpA and ScpB mixed together at a (1:1:1) ratio. The most prominent peaks are those representing the SMC dimer (270 kDa), $ScpAB_2$ (73 kDa) and the SMC_2-$ScpAB_2$ complex (344 kDa). This shows how the stoichiometry of a large protein complex can be elucidated with native mass spectrometry.



	Protein	Experimental mass (Da)	Theoretical mass (Da)
●	ScpB	21,915	22,038.2
●	$(ScpB)_2$	43,855	44,076.4
●	$ScpAB_2$	73,466	73,536.4
●	$(ScpAB_2)_2$	146,956	147,072.8
●	SMC_2	271,568	270,780.0
●	SMC_2-ScpA	301,198	300,240.0
●	SMC_2-$ScpAB_2$	345,214	344,316.4
●	SMC_2-$(ScpAB_2)_2$	418,706	417,852.8
●	$(SMC_2$-$ScpAB_2)_2$	688,941	688,632.8
●	$(SMC_2$-$ScpAB_2)_2$-ScpA	717,559	718,092.8

Fig. 1 Mass spectrum of SMC interacting with ScpA and ScpB.

8.3 Sources

[1] J.Snijder, R. J. Rose, D. Veesler, J. E. Johnson, A. J.R. Heck, Studying 18 Mega Dalton Virus Assemblies with Native Mass Spectrometry, Angew. Chem. Int Ed Engl., 2013, 52, 4020-4023
[2] I. Campuzano, K. Giles, T. McKenna, C. J. Hughes, J. I. Langridge, The application of synapt high definition mass spectrometry for the conformational studies of protein complexes,2007, Waters application note 720002343en
[**3]** K. Giles, J. P. Williams, I.Campuzano, Enhancements in travelling wave ion mobility resolution, Rapid Commun. Mass Spectrom, 2011, 1559-1566
[4] M. Zhou, C.M.Jones, V.H. Wysocki, Dissecting the large noncovalent protein complex GroEL with surface-induced dissociation and ion mobility-mass spectrometry, Anal Chem, 2013, 85,8262-7
[5] A. V. Strunnikov, SMC proteins and chromosome structure, Trends Cell Biol. 1998, 8, 11, 454-459
[6] T. E. Melby, C. N. Ciampaglio, G. Briscoe, H. P. Erickson, The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge, J Cell Biol., 1998, 142,1595-1604

8.4 Biophysical studies on interactions and assembly of $CRL5^{SOCS2}$

In this study, native IM-MS was used to investigate components of the $CRL5^{SOCS2}$ complex. CRLs target post-translationally modified substrates for ubiquitination and proteasomal degradation. SOCS proteins are substrate receptors of CRL5 complexes. The $CRL5^{SOCS2}$ can bind to a growth hormone receptor and negatively regulates growth hormone signalling for ubiquitination and proteasomal degradation. Understanding the assembly of the complex can aid in development of small molecules targeting CRL complexes.
It was confirmed that the components of the complex exist in monomeric state. The IM-MS data also support the proposed structural models based on crystal models. The experimental methods and the discussed results are displayed in the following article.

⌘ Author's Choice

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 290, NO. 7, pp. 4178–4191, February 13, 2015

Biophysical Studies on Interactions and Assembly of Full-size E3 Ubiquitin Ligase

SUPPRESSOR OF CYTOKINE SIGNALING 2 (SOCS2)-ELONGIN BC-CULLIN 5-RING BOX PROTEIN 2 (RBX2) *

Received for publication, October 29, 2014, and in revised form, November 28, 2014 Published, JBC Papers in Press, December 11, 2014, DOI 10.1074/jbc.M114.616664

Emil Bulatov [‡§], Esther M. Martin [¶], Sneha Chatterjee [¶], Axel Knebel [‖], Satoko Shimamura**, Albert Konijnenberg [¶], Clare Johnson [‖], Nico Zinn**, Paola Grandi**, Frank Sobott [¶1], and Alessio Ciulli [‡§2]

From the [‡]Division of Biological Chemistry and Drug Discovery, College of Life Sciences, and the [‖]Medical Research Council Phosphorylation and Ubiquitylation Unit, College of Life Sciences, Sir James Black Center, University of Dundee, Dundee DD1 5EH Scotland, United Kingdom, the [§]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom, the [¶]Department of Chemistry, University of Antwerp, 2020 Antwerp, Belgium, and **Cellzome GmbH, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Background: The component subunits of CRL E3 ligases assemble into specific complexes.
Results: Components of CRL5 ^{SOCS2} were identified from human cell lysate, the full-size complex was reconstituted in vitro, and protein-protein interactions were biophysically characterized.
Conclusion: CRL5 ^{SOCS2} components exist in a monomeric state, and proposed structural models are supported by ion mobility mass spectrometry.
Significance: We provide structural insights into the assembly of full-size CRL5 ^{SOCS2} that can aid development of small molecules targeting CRL complexes.

The multisubunit cullin RING E3 ubiquitin ligases (CRLs) target post-translationally modified substrates for ubiquitination and proteasomal degradation. The suppressors of cytokine signaling (SOCS) proteins play important roles in inflammatory processes, diabetes, and cancer and therefore represent attractive targets for therapeutic intervention. The SOCS proteins, among their other functions, serve as substrate receptors of CRL5 complexes. A member of the CRL family, SOCS2-EloBC-Cul5-Rbx2 (CRL5 ^{SOCS2}), binds phosphorylated growth hormone receptor as its main substrate. Here, we demonstrate that the components of CRL5 ^{SOCS2} can be specifically pulled from K562 human cell lysates using beads decorated with phosphorylated growth hormone receptor peptides. Subsequently, SOCS2-EloBC and full-length Cul5-Rbx2, recombinantly expressed in Escherichia coli and in Sf21 insect cells, respectively, were used to reconstitute neddylated and unneddylated CRL5 ^{SOCS2} complexes in vitro. Finally, diverse biophysical methods were employed to study the assembly and interactions within the complexes. Unlike other E3 ligases, CRL5 ^{SOCS2} was found to exist in a monomeric state as confirmed by size exclusion chromatography with inline multiangle static light scattering and native MS. Affinities of the protein-protein interactions within the multisubunit complex were measured by isothermal titration calorimetry. A structural model for full-size neddylated and unneddylated CRL5 ^{SOCS2} complexes is supported by traveling wave ion mobility mass spectrometry data.

Cullin RING E3 ubiquitin ligases (CRLs) [3] represent the largest known family of E3 enzymes in the ubiquitin-proteasome system (>200 CRLs of a total of >600 E3 enzymes) and play a significant role in cancer and other diseases (1). In some types of cells, up to 20% of the proteasome-dependent degradation is mediated by CRLs (2). Assembly of the multisubunit CRLs was initially reported for the archetypal Skp1-Cul1-Rbx1 complex (3). CRLs use modular subunit organization consisting of interchangeable adaptors (Skp1, elongin B, elongin C, DDB1, and BTB), substrate receptors (F box, SOCS box, DCAF, and BTB), cullin scaffolds (cullin 1–7), and RING domain proteins (Rbx1 and Rbx2) to enable the assembly of a large number of functionally diverse E3 ligase complexes (1, 4, 5).

* This work was supported by United Kingdom Biotechnology and Biological Sciences Research Council David Phillips Fellowship BB/G023123/1 (to A. C.), European Research Council Starting Grant ERC-2012-StG-311460 DrugE3CRLs (to A. C.), and the Government of the Republic of Tatarstan (Ph.D. studentship to E. B.). Molecular biophysics was supported by Wellcome Trust strategic award 100476/Z/12/Z (to the University of Dundee). The Synapt G2 mass spectrometer was funded by a grant from the Hercules Foundation-Flanders.
⌘ Author's Choice—Final version full access.
[1] A Francqui Research Professor at the University of Antwerp.
[2] To whom correspondence should be addressed: College of Life Sciences, University of Dundee, Dow Street, Dundee DD15EH, Scotland, United Kingdom. Tel.: 44-1382-386230; E-mail: a.ciulli@dundee.ac.uk.

[3] The abbreviations used are: CRL, cullin-RING E3 ubiquitin ligase; BTB, bric-a-brac·tram-track·broad complex; CCS, collision cross-section(s); CRL5^{SOCS2}, SOCS2-EloBC-Cul5-Rbx2 complex; CTD, C-terminal domain; Cul5, cullin 5; DCAF, DDB1 and Cul4-associated factor; DDB1, DNA damage-binding protein 1; EloB, elongin B; EloC, elongin C; EloBC, elongin B-elongin C complex; EHSS, exact hard sphere scattering; FBXO31, F box-only protein 31; GHR, growth hormone receptor; TWIM-MS, travelling wave ion mobility-mass spectrometry; ITC, isothermal titration calorimetry; NEDD8, neural precursor cell expressed developmentally down-regulated protein 8; NTD, N-terminal domain; PA, projection approximation; PDB, Protein Data Bank; Rbx, RING box protein; Skp, S-phase kinase-associated protein; SEC-MALS, size exclusion chromatography and multiangle light scattering; SOCS2, suppressor of cytokine signaling 2; TCEP, tris(2-carboxyethyl) phosphine hydrochloride; TEV, tobacco etch virus; SH2, Src homology 2.

The N-terminal domain (NTD) of cullin proteins consists of three five-helix bundles ("cullin repeats") that form a long stalk architecture, and a globular C-terminal domain (CTD, or "cullin homology domain"). Cullin NTD recruits variable substrate receptors either directly or via an adaptor protein, whereas cullin CTD serves as a docking site for RING domain proteins that in turn recruit a cognate ubiquitin-loaded E2 (6). RING domain proteins contain a distinct Zn^{2+}-binding domain characterized by a canonical RING motif. Selection of the substrate receptors for a particular CRL occurs through a specific receptor LP XP motif that forms a minor yet crucial supplementary interaction with cullin NTD (7, 8).

SOCS2 is a member of the SOCS box protein family that, in association with the adaptor elongin B-elongin C complex (EloBC), cullin 5 scaffold, and Rbx2, constitutes a CRL5 ^{SOCS2} E3 ligase. SOCS2 contains three structural domains: a conserved C-terminal SOCS box domain that binds to adaptor EloBC; a central SH2 domain mediating recruitment of phosphorylated tyrosine-containing sequence of the substrate; and a variable N-terminal region that facilitates interaction with the substrate. CRL5 ^{SOCS2} negatively regulates growth hormone signaling by targeting growth hormone receptor (GHR) for ubiquitination and proteasomal degradation (9). Phosphorylated tyrosine 595 serves as the key structural determinant of GHR recognition by the SH2 domain of SOCS2 (10). Crystal structures of SOCS2-EloBC (PDB code 2C9W) (11) and SOCS2-EloBC-Cul5 $_{NTD}$ (PDB code 4JGH) (12) describe structural features of the protein-protein interfaces within these left-arm complexes. However, the details of assembly of the full-size CRL5 ^{SOCS2} E3 complex both in vivo and in vitro remain missing.

In recent years, interest in studying the structure, function, and assembly of CRLs has been growing, notably driven in part by their potential role as drug targets in a number of human diseases (13–16). However, only a few studies have investigated the full-size CRL complexes biophysically, primarily due to difficulties in obtaining some of the protein components recombinantly, in particular full-length cullins. Furthermore, large heteromeric protein complexes such as CRLs are notoriously difficult to crystallize into diffraction quality crystals. Therefore, it seems promising to engage the strengths of diverse biophysical methods in order to facilitate characterization of both the individual subunits and the full-size complexes as well as to provide a means for examining their association and interactions.

Here, we show that all components of the CRL5^{SOCS2} could be pulled down from a cell lysate via SOCS2-mediated recognition of the phosphorylated GHR_pY595 peptide immobilized on beads. The full-length E3 ligase complex was then reconstituted in vitro using purified recombinant proteins and characterized biophysically. Investigations of assembly and interactions within the complex were carried out using size exclusion chromatography and multiangle light scattering (SEC-MALS), isothermal titration calorimetry (ITC), and nanoelectrospray traveling wave-ion mobility mass spectrometry (TWIM-MS).

EXPERIMENTAL PROCEDURES

Pull-down Experiments—Pull-down experiments were performed using biotinylated GHR-derived 11-mer peptides phosphorylated (GHR_pY595) or not (GHR_Y595) on tyrosine 595, harboring an aminohexanoic acid as spacer after the biotin (Biotin-aminohexanoic acid-PVPDpYTSIHIV-amide), and immobilized on high capacity Neutravidin beads. Competition experiments were performed by incubating human K562 total cell lysate with 100 μM non-biotinylated phosphorylated peptide (GHR_pY595) and the immobilized (Biotin-aminohexanoic acid-PVPDpYTSIHIV-amide) beads fo r 2 h at 4 °C. After washing, bound proteins were eluted with SDS-sample buffer and prepared for tandem mass tags labeling and MS analysis as described previously (17).

Protein Expression and Purification—Recombinant human SOCS2 (amino acids 32–198), elongin C (amino acids 17–112), and elongin B (amino acids 1–118) were co-expressed in Escherichia coli BL21(DE3) from the pLIC (SOCS2) and pCDF_ Duet (EloBC) plasmids (gifts from A. Bullock, Structural Genomics Consortium, Oxford, UK). A starter culture was grown overnight from a single transformant colony using 50 ml of LB medium containing 100 μg/ml ampicillin and 50 μg/ml streptomycin. Starter culture then was used to inoculate 7 liters of LB medium containing 100 μg/ml ampicillin and 50 μg/ml streptomycin. The cells were grown at 37 °C until A_{600} ~0.7 and cooled to 18 °C, and then protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 12 h.

Recombinant human Cul5$_{NTD}$ (N-terminal domain, residues 1–386) was expressed in E. coli BL21(DE3) from a pNIC plasmid encoding sequence for Cul5$_{NTD}$, containing His_6 and FLAG tags at the C-terminal end and a tobacco etch virus (TEV) cleavage site, as described previously (18). Briefly, 50 ml of starter culture was grown overnight using a single transformant colony in LB medium containing 50 μg/ml kanamycin and used to inoculate 2 liters of LB medium supplemented with 50 μg/ml kanamycin. The cells were grown at 37 °C until A_{600} ~0.7 and cooled to 18 °C, and protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 12 h.

Recombinant SOCS2-EloBC and Cul5 $_{NTD}$ were independently purified using the following protocol. The cell pellets were harvested by centrifugation at 5,000 rpm and 4 °C for 30 min and resuspended in binding buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP). The supernatant was treated with 10 μg/ml DNase I, 10 m M $MgCl_2$ for 30 min and then filtered through a 0.22-μm filter. The sample was applied on a HisTrap column (GE Healthcare), and the resin was washed with wash buffer (50 mM HEPES, pH 7.5, 20 m M imidazole, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP) and then the bound proteins were eluted with an incremental gradient of elution buffer (50 mM HEPES, pH 7.5, 500 m M imidazole, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP). Fractions containing protein were pooled, and the His_6 tag was cleaved off by overnight dialysis in the presence of TEV protease at 4 °C in binding buffer. The protein was applied to a HisTrap column for a second time, collecting the flow-through, and then concentrated and purified on a HiLoad 16/60 Superdex 75 column with running buffer 25 mM HEPES, pH 7.5, 250 m M NaCl, 0.5 m M TCEP.

Recombinant human Cul5 (amino acids 1–780) and Rbx2 (amino acids 1–113) were co-expressed in Sf21 insect cells using pFastBac™ Dual vector in the Bac-to-Bac® baculovirus expression system. In this vector, Cul5 is N-terminally tagged with a fragment of bacterial PBP5 (Dac tag) (19), which can be removed with TEV protease, as described previously (20).

Bacmids for Dac-TEV-Cul5/Rbx2 were generated in DH10BAC cells and transfected into Sf21 cells, using Cellfectin II® reagent (Invitrogen). The transfected cells were kept for 7 days at 27 °C in Insect Express® medium (Lonza), supplemented with ANTI-ANTI® (Invitrogen). The cells were sedimented, and the virus-containing supernatant was used to infect 150 ml of Sf21 culture at a density of 1.5×10^6 cells/ml. For the expression of RING E3 ligases, we supplemented the Insect Express medium with 5 μM $ZnCl_2$. After 5 days the cells, were collected under sterile conditions, and the supernatant was used to infect 2 liters of Sf21 cell culture for 3 days. The Cul5-Rbx2 protein complex was purified using the following procedure. Cells were harvested by centrifugation at 3,500 rpm and 4 °C for 15 min and then resuspended in 25 ml of 50 mM HEPES, pH 7.4, 0.1 mM EGTA, 1 μM $ZnCl_2$, 1 mM TCEP, 1 mM Pefabloc®, and 20 μg/ml leupeptin (both from Apollo Scientific) and incubated for 15 min at 4 °C. Cells were sheared using a 50-ml tight fit Dounce homogenizer, and insoluble material was removed by centrifugation at 40,000 rpm and 10 °C for 20 min. To perform Dac affinity purification, the supernatant was gently mixed with ampicillin-Sepharose at room temperature for 50 min. The Sepharose was collected by centrifugation and washed six times in 10 volumes of 50 mM HEPES, pH 7.4, 150 mM NaCl, 0.1 mM EGTA, 1 μM $ZnCl_2$, 1 mM TCEP. To recover untagged Cul5-Rbx2, the Sepharose was incubated with TEV protease (50 μg of TEV, 1 ml of Sepharose) overnight at room temperature and drained and washed through Econopac® filter units (Bio-Rad). The Cul5-Rbx2 protein was concentrated and further purified by preparative size exclusion chromatography using HiLoad 16/600 Superdex 200 column (GE Healthcare) with 50 mM HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 0.5 mM TCEP.

The identity and purity of all obtained proteins was confirmed using denaturing electrospray ionization-MS and SDS-PAGE. All proteins were flash-frozen using liquid nitrogen and stored at −80 °C.

Reconstitution of SOCS2-EloBC-Cul5$_{NTD}$ and SOCS2-EloBC-Cul5-Rbx2 Protein Complexes—To form quaternary SOCS2-EloBC-Cul5$_{NTD}$ complex, Cul5$_{NTD}$ and SOCS2-EloBC were mixed together at a 1:1.1 molar ratio and incubated at room temperature for 30 min, following purification of the complex using a HiLoad 16/600 Superdex 75 column in 25 mM HEPES, pH 7.5, 250 mM NaCl, 0.5 mM TCEP.

To form the pentameric complex, Cul5-Rbx2 and SOCS2-EloBC were mixed at a 1:1.1 molar ratio and incubated at room temperature for 30 min. The protein complex was purified using a HiLoad 16/600 Superdex 200 column in 50 mM HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 0.5 mM TCEP. SDS-PAGE analysis of SOCS2-EloBC, Cul5$_{NTD}$, SOCS2-EloBC-Cul5$_{NTD}$, Cul5-Rbx2, and SOCS2-EloBC-Cul5-Rbx2 is shown in Fig. 3*B*.

Formation of NEDD8 ~Cul5-Rbx2 Covalent Conjugate—To obtain the conjugate, E1 activating enzyme APP-BP1/UBA3 (1 μM), E2 conjugating enzyme UBE2F (5 μM), NEDD8 (40 μM), and Cul5-Rbx2 (5 μM) were incubated at 37 °C for 1 h in 50 mM HEPES, pH 7.4, 150 mM NaCl, 5 mM DTT, 10 mM $MgCl_2$, 0.2 mM ATP. A negative control experiment was performed in the same solution containing 0.1 mM EDTA, with no $MgCl_2$ or ATP added. The completion of the neddylation on the Cul5 subunit was confirmed by SDS-PAGE (Fig. 8).

SEC-MALS— SEC-MALS experiments were performed using Dionex Ultimate 3000 UHPLC system (Thermo Scientific) with an inline Wyatt miniDAWN TREOS MALS detector and Optilab T-rEXTM refractive index detector. Molar masses spanning elution peaks were calculated using ASTRA version 6.0.0.108 (Wyatt). SEC-MALS data were collected for the following samples: 1) SOCS2-EloBC at 130 μM; 2) Cul5$_{NTD}$ at 110 μM; 3) SOCS2-EloBC-Cul5$_{NTD}$ at 90 μM; 4) Cul5-Rbx2 at 30 μM; 5) SOCS2-EloBC-Cul5-Rbx2 at 35 μM. The experiments were performed using a Superdex 200 10/300 GL column (GE Healthcare) with running buffer 50 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM TCEP. The scattering signal was collected at 44, 90, and 136° using $\lambda = 658.5$ nm incident light. Resulting data were processed in Microsoft Excel, and peaks were normalized.

TWIM-MS— The native TWIM-MS experiments were conducted on a Synapt HDMS G2 instrument (Waters, Milford, MA), which has been described previously (21). Samples following gel filtration were buffer-exchanged into 500 mM aqueous ammonium acetate at pH 7.0, using Micro Bio-Spin P-6 columns (Bio-Rad), at concentrations in the range of 5–10 μM. Aliquots of 3–5 μl were transferred to gold-coated nano-electrospray ionization needles prepared in house. The instrument was tuned to ensure the preservation of non-covalent interactions (22), using the following parameters: capillary, 1.2 kV; sample cone, 40 V; extraction cone, 0.5 V; nanoflow gas pressure, 0.3 bar; trap collision energy, 4.0 V; transfer collision energy, 3.5 V; backing pressure, 4 millibars, trap pressure, 3.4 millibars. For the measurement of the full 148-kDa complex, SOCS2-EloBC-Cul5-Rbx2, the backing pressure was increased to 5 millibars to facilitate the transmission of high m/z signal. Gas pressure in the ion mobility cell was 3.0 millibars, and helium and N_2 gas flows were 180 and 90 ml/min, respectively, with a trap bias of 50 V. The traveling wave velocity was 800 m/s with a traveling wave height of 40 V. The data were acquired and processed with MassLynx version 4.1 software (Waters), and drift times were extracted using Driftscope version 2.3 (Waters). The experimental collision cross-sections (CCS) of the protein complexes were determined by calibration with known protein cross-sections determined under native conditions as described previously (23).

Calculation of Theoretical CCS— Theoretical CCS values of the protein complexes were calculated from model structures, obtained by docking individual protein subunits together, using the program MOBCAL with both the projection approximation (PA) and the exact hard sphere scattering (EHSS) methods (24, 25). The PDB files were cleaned (i.e. by resolving dihedral conflicts and adding missing side chains and removing crystal water molecules) prior to the PA or EHSS calculation. The theoretical CCS was compared with the experimental CCS of the lowest available charge state for that species in the mass spectra,

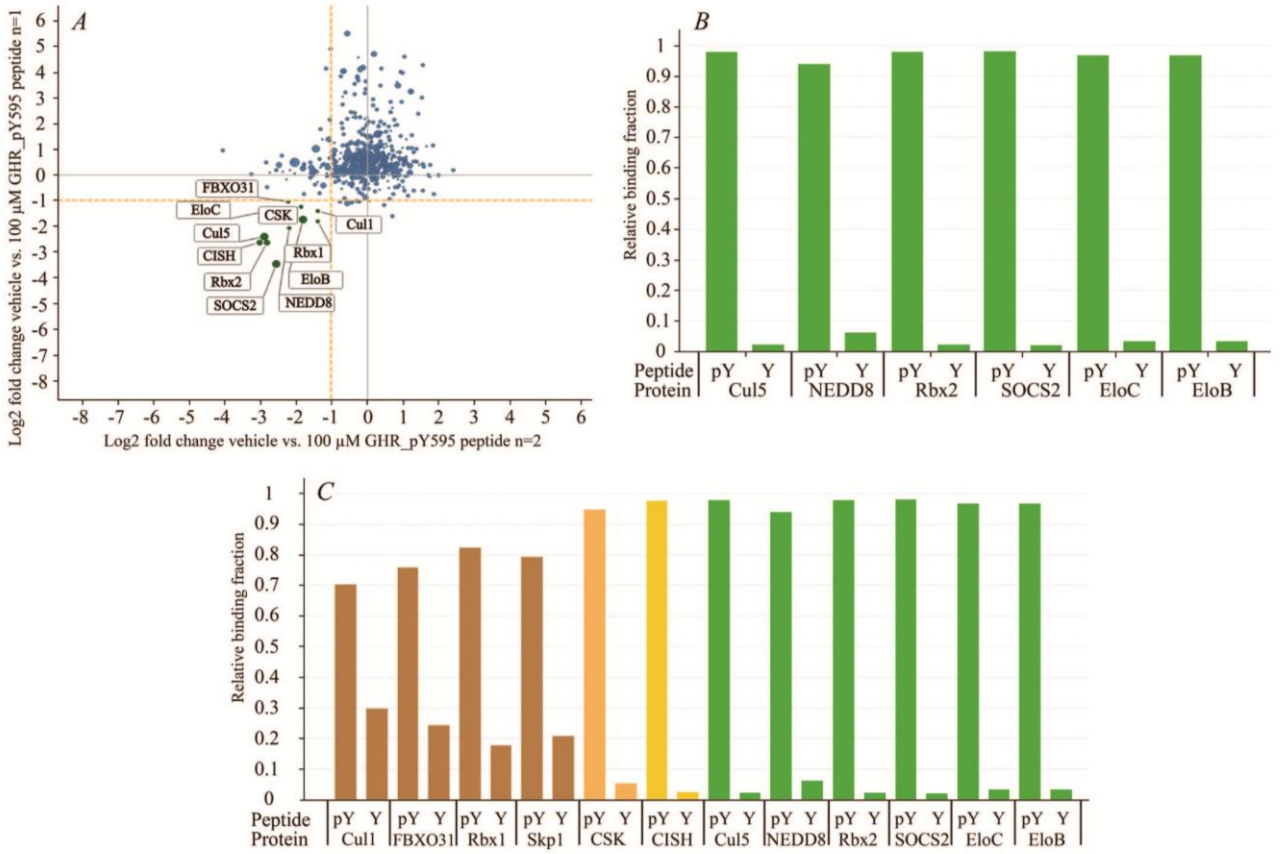


FIGURE 1. Components of CRL5 ^{SOCS2} E3 ligase can be captured from cell lysate using phosphorylated substrate peptide attached to the beads. A, scattered plot of the peptide pull-down experiments after MS analysis: log2 -fold changes of all proteins captured on the phosphorylated GHR_pY595 peptide beads after competition with a 100 μM concentration of the same peptide in the lysate. Each protein is indicated by a circle, and the size of the circle is proportional to the MS1 value. The labeled proteins are at least 50% displaced by the addition of the peptide in both experiments (log2 -fold change $\leq$1, orange dotted lines). B, relative binding profile of the components of CRL5^{SOCS2} E3 ligase complex captured on the beads decorated with phosphorylated (pY) versus non-phosphorylated (Y) GHR_(p)Y595 peptide. C, binding profiles for other proteins specifically recruited by the beads. Components of the CRL5^{SOCS2} E3 ligase are colored in green.

which corresponds to the most native-like structure of the protein complex (Fig. 4, A–C, bottom panels).

ITC— Experiments were conducted using an iTC200 microcalorimeter instrument (GE Healthcare). GHR_pY595 peptide (350 μM, PVPDpYTSIHIV-amide), GHR_Y595 (350 μM, PVPDYTSIHIV-amide), and phosphotyrosine (2 m M, Tyr(P)) were titrated into SOCS2-EloBC (30 μM) at 298 K. Temperature-dependent experiments to study interaction between SOCS2-EloBC and Cul5 $_{NTD}$ were performed by titrating Cul5$_{NTD}$ (450 μM) into SOCS2-EloBC (60 μM) at 298, 303, and 308 K. Prior to all titration experiments, sample proteins were dialyzed into 50 mM HEPES, pH 7.5, 250 m M NaCl, 0.5 m M TCEP. Peptides and Tyr(P) were dissolved in the same buffer. Obtained data were analyzed and fitted using the Microcal Origin version 7.0 software package. Binding enthalpy, dissociation constants, and stoichiometry were determined by fitting the data using a one-set-of-site binding model.

Molecular Modeling of Protein Complexes—Due to the absence of an Rbx2 crystal structure, its closest homolog, Rbx1, was used for the model construction. The structural model of the SOCS2-EloBC-Cul5-Rbx1 complex was prepared in PyMOL, using the crystal structure of Cul1-Rbx1-Skp1-Skp2 (PDB code 1LDK) (26) as the initial template. To construct the model, SOCS2-EloBC-Cul5 $_{NTD}$ (PDB code 4JGH) (12) was superimposed on the template by aligning its Cul5$_{NTD}$ subunit with the Cul1$_{NTD}$ of the template. After that, Cul5$_{CTD}$ -Rbx1 (PDB code 3DPL) (27) was aligned with Cul1 $_{CTD}$ subunit of the template to generate a model of the full-length E3 ligase. The resulting model of the CRL5 ^{SOCS2} complex was used to obtain the model for Cul5-Rbx1. To generate a model of the "open" neddylated complexes, NEDD8~Cul5$_{CTD}$ -Rbx1 (PDB code 3DQV) (27) was aligned with the Cul1 $_{CTD}$ subunit of the template. Alternatively, to prepare a model of the closed neddylated complexes, NEDD8 was simply added from aligned 3DQV onto the non-neddylated Cul5$_{CTD}$ -Rbx1 and SOCS2-EloBC-Cul5-Rbx1.

RESULTS

Components of CRL5^{SOCS2} E3 Ligase Can Be Pulled Down from Human Cell Lysates Using Phosphopeptide-modified Beads—Specific subunits of E3 ligase SOCS2, EloB, EloC, Cul5, and Rbx2 are known to function as a CRL5 ^{SOCS2} complex (28). The SH2 domain of SOCS2 recognizes and specifically binds a GHR sequence containing the phosphorylated tyrosine

TABLE 1
Proteins enriched by the phosphorylated GHR_pY595-modified beads
The proteins were specifically captured by the beads and displaced by the phosphorylated GHR peptide.

Protein	UniProt ID	Comments
SOCS2	O14508	Suppressor of cytokine signaling 2. Substrate recognition domain of Cul5^{SOCS2} E3 ligase. Contains SH2 domain that recognizes substrate phosphotyrosine residues.
EloB	Q15370	Transcription elongation factor B.
EloC	Q15369	Transcription elongation factor C. Complex of EloB and EloC (EloBC) serves as adaptor domain of Cul5^{SOCS2} E3 ligase.
Cul5	Q93034	Cullin 5. Scaffold domain of the Cul5 ^{SOCS2} E3 ligase.
Rbx2	Q9UBF6	RING box protein 2. Contains RING-type zinc finger, recruits E2 conjugating enzyme.
NEDD8	Q15843	Neural precursor cell expressed developmentally down-regulated protein 8. Ubiquitin-like protein, can form covalent conjugate with Cullin that enhances the E3 ligase activity.
CISH	Q9NSE2	Cytokine-inducible SH2-containing protein. Component of SCF E3 ligase, can recognize phosphotyrosine.
Rbx1	P62877	RING-box protein 1. Contains RING-type zinc finger, recruits E2 conjugating enzyme.
CSK	P41240	c-Src kinase. Contains SH2 domain that can recognize phosphotyrosine residues.
Cul1	Q13616	Cullin 1. Scaffold component of SCF E3 ligase.
FBXO31	Q5XUX0	F box-only protein. Substrate recognition component of SCF E3 ligase. Can recognize certain phosphorylated substrates.
Skp1	P63208	S-phase kinase-associated protein 1. Adaptor component of SCF E3 ligase.

Tyr(P) ^{595}. We envisaged that SOCS2 and components offull-length CRL5 ^{SOCS2} E3 ligase should be amenable for capturing from cell lysate using substrate peptides immobilized on beads. With this aim, we performed pull-down experiments from human K562 cell lysate using beads decorated with both phosphorylated GHR_pY595 (PVPDpYTSIHIV-amide, positive control) and non-phosphorylated GHR_Y595 (PVPDYTSIHIV-amide, negative control) peptides. Mass spectrometry analysis revealed a reproducible and limited set of proteins captured and subsequently displaced by the phosphorylated peptide (Fig. 1A, bottom left corner). All components of the CRL5 ^{SOCS2} (SOCS2, EloB, EloC, Cul5, and Rbx2) were among this protein set.

Binding profile (Fig. 1 B) shows that endogenous components of CRL5 ^{SOCS2} E3 ligase were only captured by GHR_pY595-beads. In contrast, no significant capturing from cell lysates was observed with non-phosphorylated GHR_Y595-containing beads (Fig. 1B). This observation shows that the Tyr(P) residue plays a key role in recognition of the substrate by the CRL^{SOCS2} complex in the cell and that phosphorylation of the peptide is essential for specific interaction with the E3 ligase. Interestingly, we also detected NEDD8 as a protein specifically pulled down by the phosphorylated peptide. NEDD8 is a ubiquitin-like protein that is known to be covalently attached to cullins and acts as a CRL activator by inducing scaffold dynamics and increasing conformational flexibility of the E3 enzyme (27, 29). Identification of NEDD8 suggests that the active neddylated complex is also being pulled down in the assay.

Surprisingly, in addition to the expected CRL5 ^{SOCS2} complex subunits, we detected subunits of the CRL1^{FBXO31} E3 ligase, namely FBXO31, Rbx1, Cul1, and Skp1 proteins, as being captured by the beads and displaced by the phosphorylated peptide (Fig. 1A and Table 1), indicating specific binding. FBXO31 is an F box protein that binds phosphorylated substrates; therefore, it could have been recruited by the GHR_pY595 peptidedirectly. However, significant recruitment of the four CRL1 ^{FBXO31} subunits was also observed by the non-phosphorylated GHR peptide (Fig. 1C). This would imply a degree of phosphorylation-independent interaction, either directly with the beads or indirectly via binding to the components of the CRL5^{SOCS2} E3 ligase complex.

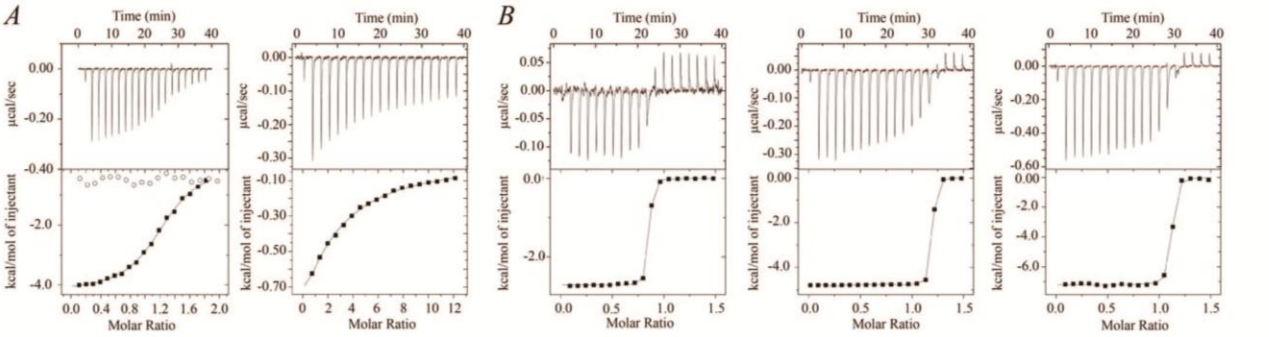
Moreover, CSK (C-terminal Src kinase) and CISH (cytokine-inducible SH2-containing) proteins were also recruited (Fig. 1 A and Table 1). Both CSK and CISH contain the SH2 domain; therefore, both proteins were probably directly recruited by the phosphorylated peptide.

SOCS2-EloBC Forms a Weak Interaction with GHR and a Tight Interaction with Cul5 $_{NTD}$—To determine the affinity of interaction and the thermodynamic parameters of binding between SOCS2-EloBC and substrate GHR or scaffold Cul5 $_{NTD}$, we performed isothermal titration calorimetry experiments (Fig. 2). SOCS2-EloBC binds GHR_pY595 peptide with $K_d = 1.8\ \mu M$, which is consistent with the previously reported value (11), and binds Tyr(P) with $K_d = 191\ \mu M$, both at 298 K (Fig. 2A). The binding affinity for phosphotyrosine is ~100-fold weaker than for the phosphorylated peptide, suggesting that other peptide residues make some contribution to interaction with the protein. However, negative control titration using non-phosphorylated GHR_Y595 peptide showed no binding (Fig. 2A), reinforcing the key contribution of the phosphate group to substrate binding.

We next determined the affinity of SOCS2-EloBC for Cul5 $_{NTD}$ by measuring a $K_d = 11$ nM for the interaction (Fig. 2C). These data are in good agreement with the previously reported K_d of 28 nM (by ITC) (12). In addition, two groups independently reported $K_d = 7$ nM (by ITC) (30), $K_d = 10$ nM (by ITC), and $K_d = 47$ nM (by surface plasmon resonance) (31) for this interaction, albeit using SOCS box domain instead of the whole SOCS2 protein in complex with EloBC.

To test the potential cooperativity ofinteractions at the GHR/SOCS2-EloBC/Cul5 $_{NTD}$ interfaces, we performed titration of GHR_pY595 peptide into SOCS2-EloBC-Cul5 $_{NTD}$ and titration of Cul5 $_{NTD}$ into GHR_pY595-SOCS2-EloBC complex. No change in the K_d or ΔH values was observed in either case, suggesting no cooperativity or cross-talk between these interactions.

The interaction between SOCS2-EloBC and the Cul5 scaffold is high affinity and crucial to the assembly of CRL complex. To provide further insights into the nature of this interaction, we performed temperature-dependent ITC titrations and determined a change in heat capacity $\Delta C_p = -450$ cal/mol/K (titration curves shown in Fig. 2B). Fig. 2D demonstrates a plot with a temperature-dependent change of thermodynamic parameters of SOCS2-EloBC/Cul5 $_{NTD}$ interaction. The experimental ΔC_p value is calculated from the slope of theΔH linear regression. As a comparison, previously reported ΔC_p values for ASB9-EloBC/Cul5 $_{NTD}$ and Vif-EloBC/Cul5 $_{NTD}$ interactions were found to be −350 cal/mol/K (18) and −300 cal/mol/K (30), respectively.



C

Material in Syringe	Material in Cell	T (K)	N	Kd (µM)	ΔH (cal/mol)	-TΔS (cal/mol)	ΔG (cal/mol)
GHR_pY595	SOCS2-EloBC	298	1.2	1.8 ± 0.1	-4,227 ± 32	-3603 ± 66	-7,839 ± 34
pY	SOCS2-EloBC	298	1.0	191 ± 4.0	-4,856 ± 54	-214 ± 66	-5070 ± 12
$Cul5_{NTD}$	SOCS2-EloBC	298	0.8	11.3 ± 1.6	-2,714 ± 6	-8,127 ± 91	-10,841 ± 84
$Cul5_{NTD}$	SOCS2-EloBC	303	1.2	5.6 ± 1.0	-4,788 ± 8	-6,661 ± 116	-11,449 ± 109
$Cul5_{NTD}$	SOCS2-EloBC	308	1.1	19.7 ± 4.3	-7,218 ± 27	-3,655 ± 163	-10,873 ± 136

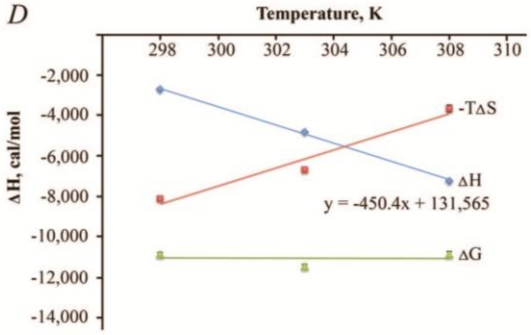


FIGURE 2. ITC data demonstrate weak interaction of SOCS2-EloBC with phosphorylated substrate GHR and tight interaction with scaffold $Cul5_{NTD}$. A, ITC titration curves of 350 µM GHR peptides (left) and 2 mM Tyr(P) (right) into 30 µM SOCS2-EloBC protein at 298 K. Phosphorylated GHR_pY595 peptide (black squares) and non-phosphorylated GHR_Y595 (white circles) are shown on the left. B, titration curves for temperature-dependent ITC of 450 µM $Cul5_{NTD}$ into 60 µM SOCS2-EloBC at 298 (left), 303 (middle), and 308 K (right). C, table summarizing data obtained from ITC experiments. D, plot demonstrating temperature dependence of thermodynamic parameters and calculation of heat capacity (ΔC_p) for SOCS2-EloBC/$Cul5_{NTD}$ interaction.

TABLE 2

Large negative ΔC_p value for SOCS2-EloBC/$Cul5_{NTD}$ interaction indicates a highly hydrophobic interface between the proteins; theoretical ΔSASA values for SOCS2-EloBC/$Cul5_{NTD}$ interaction calculated using GetArea and NACCESS programs

		SASA (Å^2)			ΔSASA (Å^2)		
		Polar	Apolar	Total	Polar	Apolar	Total
GetArea	SOCS2-EloBC	7,270.0	10,908.7	18,178.7			
	$Cul5_{NTD}$	7,709.6	11,708.7	19,418.4			
	SOCS2-EloBC-$Cul5_{NTD}$	14,361.0	21,291.4	35,652.5	-618.6	-1,326.0	-1,944.6
NACCESS	SOCS2-EloBC	7,992.5	10,375.8	18,368.3			
	$Cul5_{NTD}$	8,393.4	11,216.6	19,610.0			
	SOCS2-EloBC-$Cul5_{NTD}$	15,688.9	20,378.7	36,067.6	-697.0	-1,213.7	-1,910.7

We next calculated the theoretical solvent-accessible surface area values in GetArea (32) and NACCESS (33) software using the crystal structure of SOCS2-EloBC-$Cul5_{NTD}$ complex (PDB code 4JGH) as a model (12) (Table 2). Theoretical ΔC_p values were calculated using the following equation (34),

$$\Delta C_p = \Delta c_{ap}\Delta ASA_{ap} + \Delta c_p\Delta ASA_p \quad \text{(Eq. 1)}$$

where ΔASA is the apolar (ap) and polar (p) surface buried upon interaction of the proteins, and Δc is the area coefficient, representing per Å^2 contribution of residues in heat capacity change. The polar and non-polar area coefficients represent values empirically determined from a range of protein data sets by different groups (35–39) (reviewed in Ref. 34). We observe good agreement between theoretical and experimental data when using area coefficients according to Refs. 39 and 37 (Table 3).

SOCS2-EloBC Forms Stable Monomeric Complexes with $Cul5_{NTD}$ and Cul5-Rbx2—To validate formation of CRL5^{SOCS2} and determine the stoichiometry of subunits in the complex, we demonstrated assembly of the full-length E3 ligase in vitro using recombinantly expressed and purified protein components (schematic representation in Fig. 3A). SOCS2 and EloBC were co-expressed in E. coli to obtain the SOCS2-EloBC ternary complex, and

Cul5$_{NTD}$ was independently expressed in E. coli. The Cul5-Rbx2 protein complex was co-expressed in Sf21 insect cells.

To characterize the purified protein complexes, biophysical analyses were carried out using SEC-MALS (Fig. 3), native MS, and TWIM-MS techniques (Figs. 4 and 5). SOCS2-EloBC-Cul5$_{NTD}$ and SOCS2-EloBC-Cul5-Rbx2 protein complexes were formed by mixing SOCS2-EloBC and either Cul5$_{NTD}$ or Cul5-Rbx2 components in equimolar amounts and then purified using size exclusion chromatography (Fig. 6).

TABLE 3

Large negative ΔC_P value for SOCS2-EloBC/Cul5$_{NTD}$ interaction indicates a highly hydrophobic interface between the proteins; comparison between theoretical and experimental ΔC_P values shows good agreement

		ΔC_p (cal/mol/K)	
	Source	GetArea	NACCESS
experimental		-450.0	
theoretical	Spolar et. al.	-337.5	-290.4
	Murphy and Friere	-434.7	-363.8
	Myers	-314.6	-276.1
	Makharadze and Privalov	-548.1	-474.2
	Robertson and Murphy	-286.0	-278.1

The SEC-MALS elution profiles of the different protein components and complexes show that they all exist as monomeric and monodisperse entities (Fig. 3D). The molar mass over elution peaks is shown in corresponding colors. Molecular weight values of eluted proteins are summarized Fig. 3C. The results of SEC-MALS analysis confirm the formation of expected protein complexes with experimentally determined molecular weights that correlate well with theoretical values.

Experimental Collision Cross-sections for Protein Complexes Are in Good Agreement with Theoretical Values—To validate the structural model of SOCS2-EloBC-Cul5-Rbx2, TWIM-MS was used to examine the molecular weight and stoichiometry of the intact protein complexes as well as to confirm their topology by CCS measurements. The protein components SOCS2-EloBC and Cul5$_{NTD}$ alone were first analyzed using native MS, and the resulting spectra are shown in Fig. 7, A and B. The masses were confirmed as ~43 and 45 kDa, respectively. Theoretical and experimental masses for each complex are shown in Table 4.

Combining SOCS2-EloBC with Cul5$_{NTD}$ produced an 89-kDa complex, which could be detected with charge states ranging from 16+ to 20+ (Fig. 4A). Some free SOCS2-EloBC was also observed in this spectrum, with the same charge states as the native mass spectrum of SOCS2-EloBC alone, indicating that some dissocia-

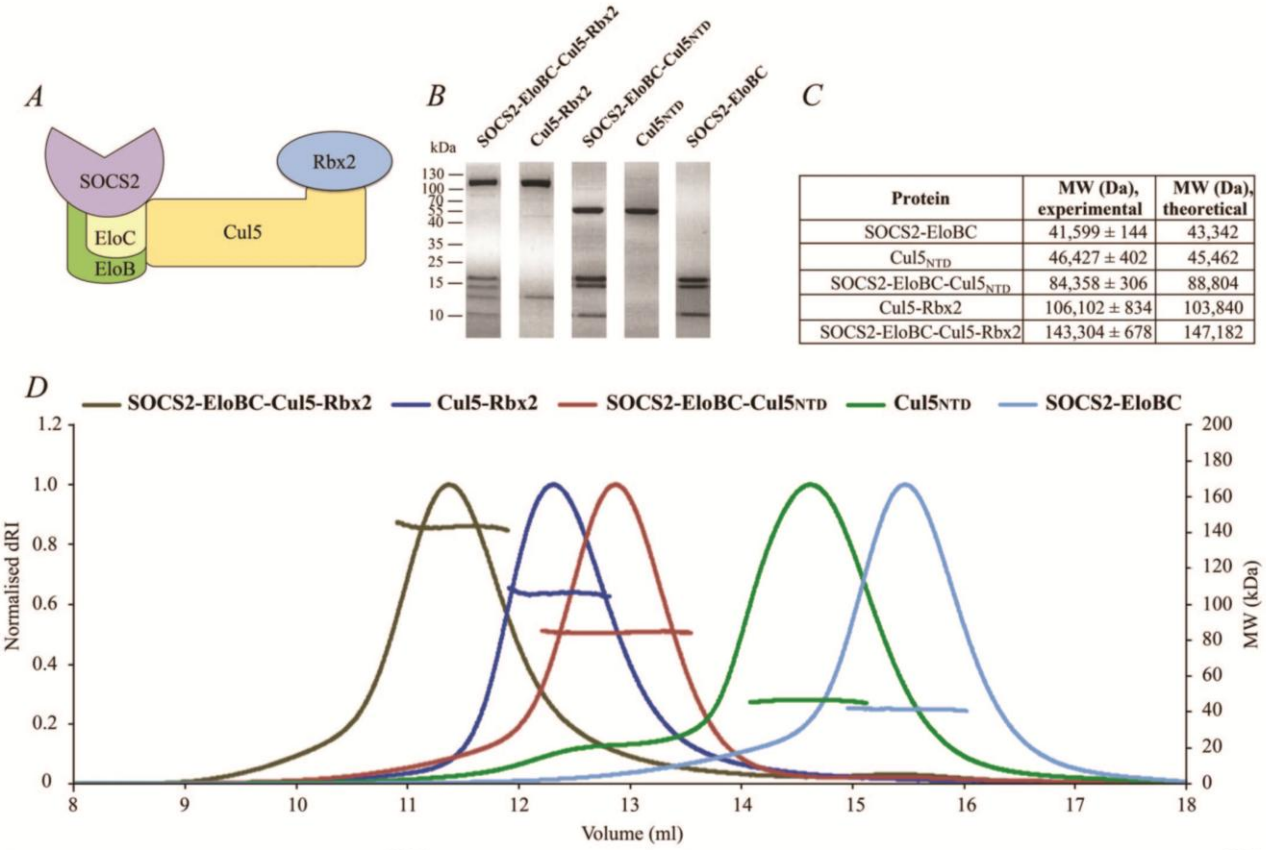


FIGURE 3. Recombinant components of CRL5^{SOCS2} assemble into the monomeric full-size protein complex. A, schematic representation of CRL5^{SOCS2}, which includes substrate receptor SOCS2, adaptor EloBC complex, scaffold Cul5, and RING domain protein Rbx2. B, SDS-polyacrylamide gel images of the purified protein complexes. C, table showing a comparison of theoretical protein molecular weights against values experimentally determined by SEC-MALS. D, SEC-MALS elution profiles for the individual components of CRL5^{SOCS2} and their complexes, including full-size SOCS2-EloBC-Cul5-Rbx2 complex.

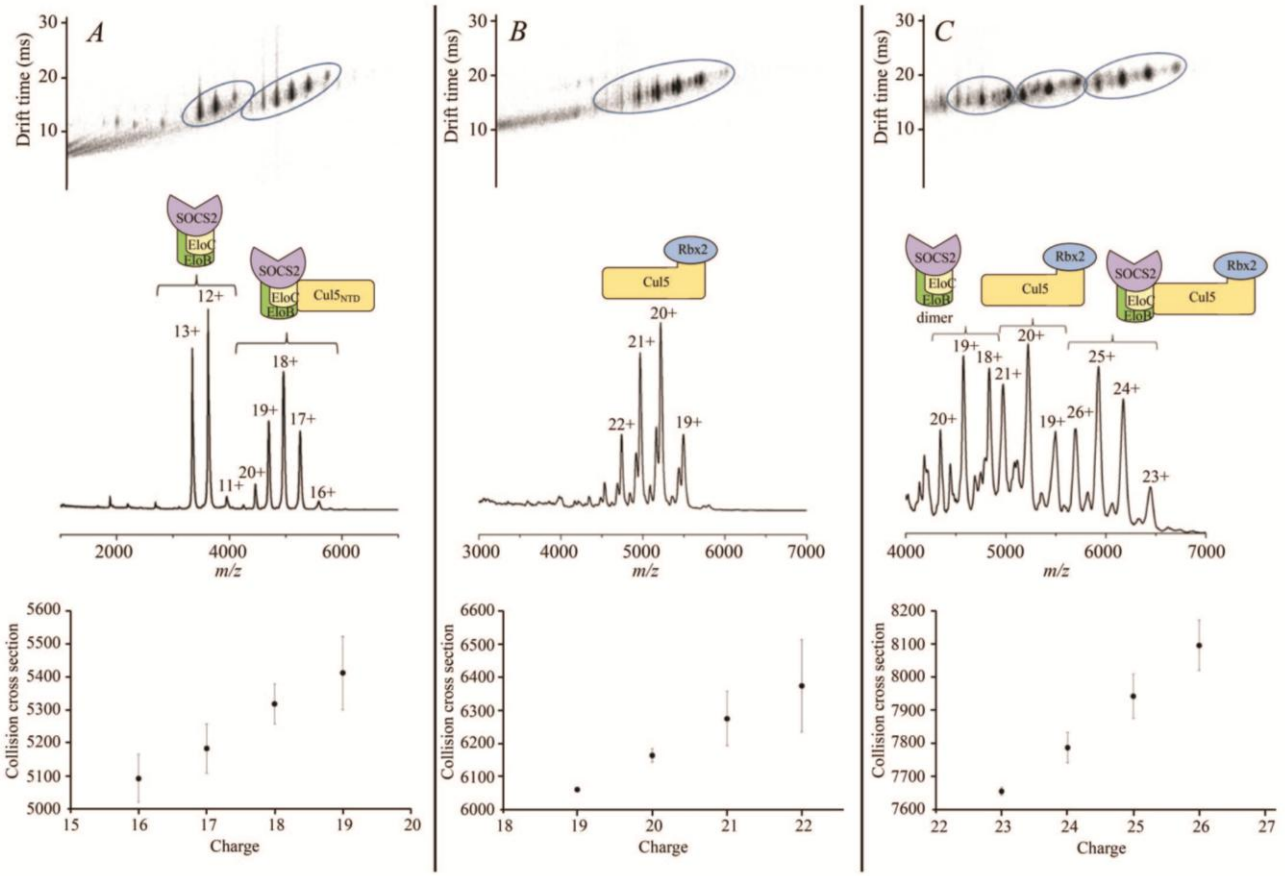


FIGURE 4. Ion mobility drift time plot (top), corresponding native mass spectra (middle), and collision cross-sections (bottom) for the CRL5 SOCS2 complexes and their components. A, SOCS2-EloBC-Cul5$_{NTD}$; B, Cul5-Rbx2; C, full-size complex SOCS2-EloBC-Cul5-Rbx2. Error bars, S.D.

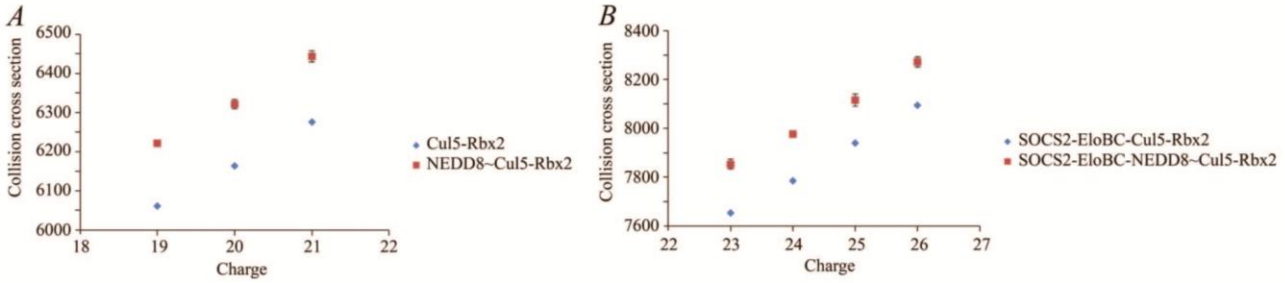


FIGURE 5. Experimental ion mobility data for a range of charge states suggests an increase in CCS values upon neddylation of the protein complexes. Shown are collision cross-sections for Cul5-Rbx2 and NEDD8~Cul5-Rbx2 (A) and SOCS2-EloBC-Cul5-Rbx2 and SOCS2-EloBC-NEDD8~Cul5-Rbx2 (B). Error bars, S.D.

tion in solution occurs (Fig. 7A). Table 4 shows the experimental CCS values compared with the theoretical ones calculated with the PA and the EHSS methods. It is normally expected that the experimental values would be smaller than the EHSS results and larger than the PA results (24, 25). The collision cross-section determined using ion mobility for SOCS2-EloBC-Cul5 $_{NTD}$ is 5,092 Å^2 for the most native charge state (16+; Fig. 4A, bottom), which is reasonably close to the theoretical value calculated for the model (Table 4, PA value 5,306 Å ^{2}).

A typical spectrum of the Cul5-Rbx2 complex is shown in Fig. 7C with a predominant 6+ charge state for Rbx2 and a series of charge states from 19+ to 22+ representing the binary complex (104 kDa). Fig. 4 B shows the Cul5-Rbx2 complex in more detail in addition to the drift time plot. Moreover, there are less intense peaks to the left-hand side of the predominant peak corresponding to a loss of ~1 kDa from the complex that may represent a truncation in the Cul5 subunit. These species are clearly separated, however, by their ion mobility (Fig. 4B), so it is possible to calculate a collisional cross-section for the intact complex. The CCS value from these data for the most native 19+ charge state was found to be 6,061 Å^2 (Fig. 4B, bottom), which compared well with the theoretical value (Table 4, PA value 5,988 Å ^{2}).

The native mass spectrum of the SOCS2-EloBC-Cul5-Rbx2 showed intense peaks at 3,000–4,000m/z, indicating a relative abundance offree SOCS2-EloBC, with charge states 11 + to 13+, as described previously (Fig. 7A). It is possible that there is

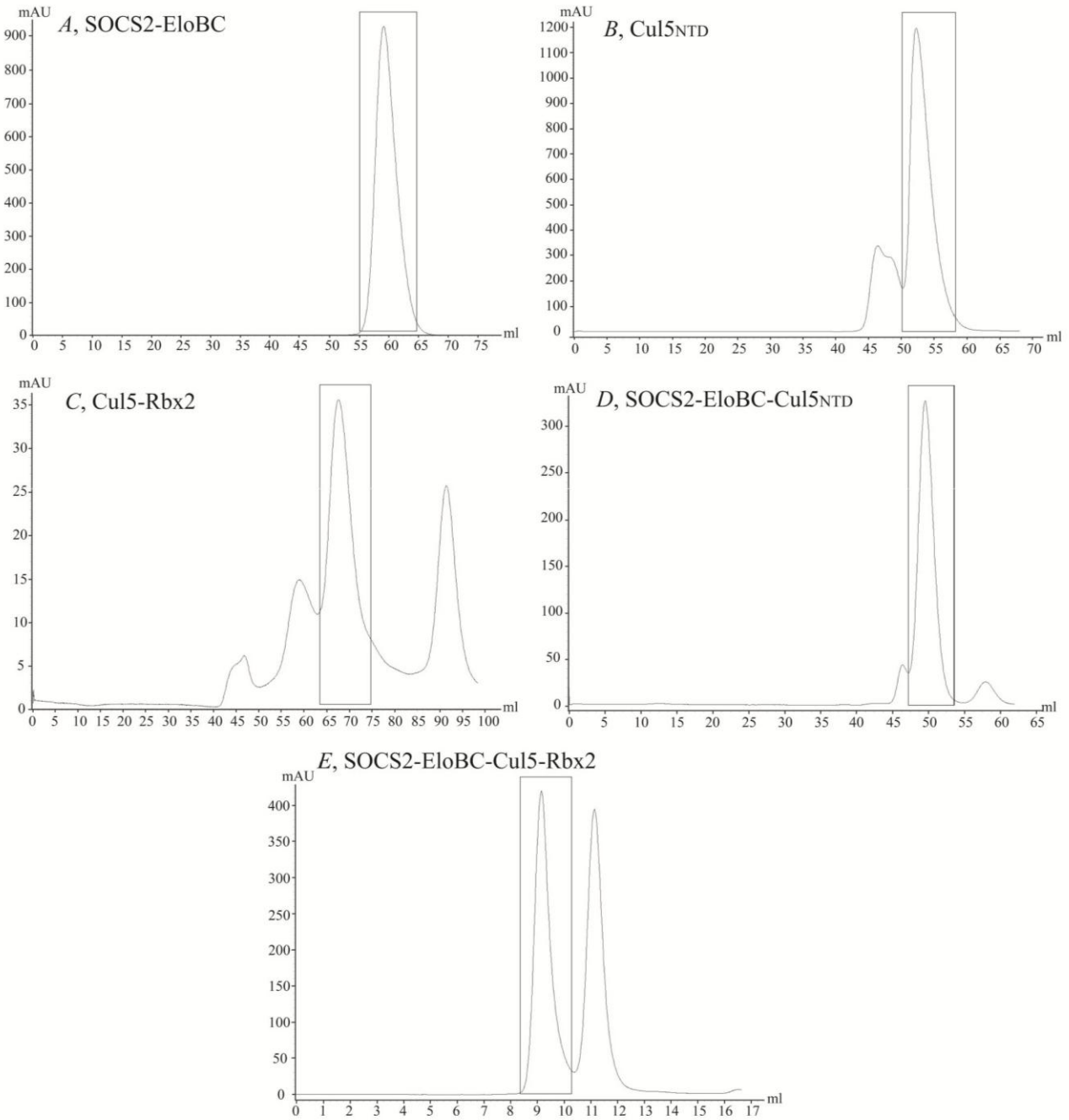


FIGURE 6. Gel filtration UV traces. A, SOCS2-EloBC; B, Cul5$_{NTD}$; C, Cul5-Rbx2; D, SOCS2-EloBC-Cul5$_{NTD}$; E, SOCS2-EloBC-Cul5-Rbx2. Proteins were purified using multistep purification with size exclusion chromatography as the last step. Peak fractions corresponding to the appropriate proteins were pooled together. The purity and identity of each protein were confirmed using denaturing electrospray ionization-MS (data not shown) and SDS-PAGE.

an excess of SOCS2-EloBC in these samples or that this subunit has a greater ionization efficiency compared with the other protein components. Fig. 4C (bottom) depicts the 4,000–7,000 m/z range of the SOCS2-EloBC-Cul5-Rbx2 spectrum that shows peaks ranging from 4,300 to 4,900 m/z representing a small amount of SOCS2-EloBC dimer. Second, at 4,700–5,600 m/z, the Cul5-Rbx2 complex is detected with charge states from 19+ to 22+. Finally, the peaks representing the full 148-kDa complex, SOCS2-EloBC-Cul5-Rbx2, are in the range of 5,600–6,600 m/z, with charge states 23+ to 26+.

The CCS values measured for each charge state of the full 148-kDa complex are displayed in Fig. 4C (bottom). For the lowest charge state of SOCS2-EloBC-Cul5-Rbx2, an experimental cross-section of 7,653 Å^2 was determined compared with a theoretical CCS of 7,918 Å^2 (Table 4), confirming the structural model as shown in Fig. 10D. In this case, the experi-

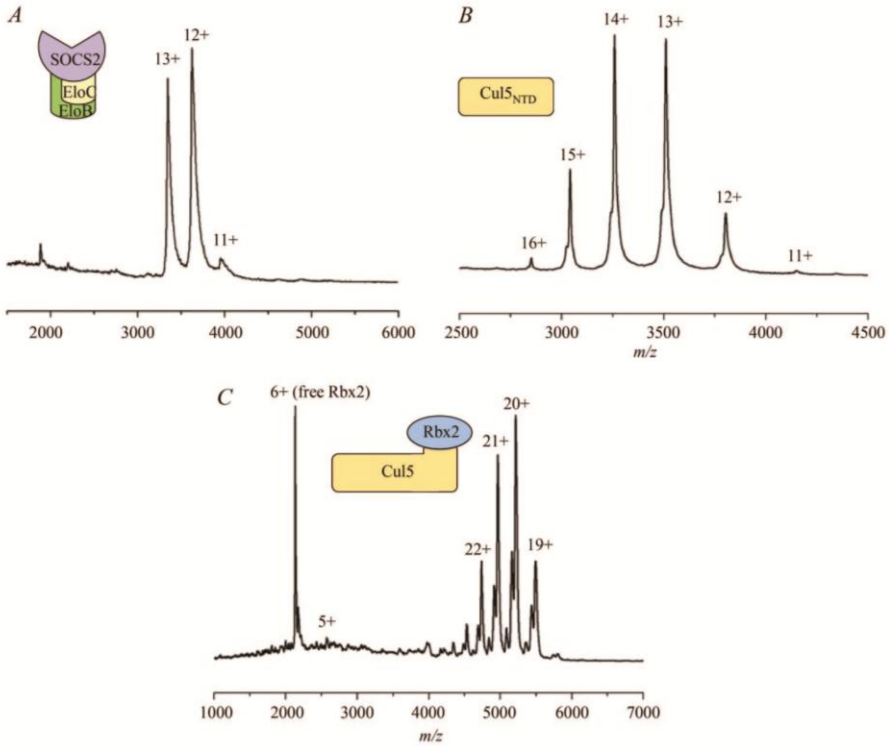


FIGURE 7. Native MS spectra. A, SOCS2-EloBC; B, Cul5$_{NTD}$; C, Cul5-Rbx2.

TABLE 4

Summary of the experimental and theorectical CCS data

Shown are masses measured for the protein complexes and observed charge state ranges. Shown is a comparison of experimental CCS values of the protein complexes versus their theoretical values calculated using the PA and the EHSS methods. Presented data also include calculated theoretical CCS values for "open" and "closed" NEDD ~Cul5-Rbx1 and SOCS2-EloBC-NEDD ~Cul5-Rbx1 complexes.

Protein	Theoretical mass (Da)	Experimental mass (Da)	Charge states	Theoretical CCS (Å^2)		Experimental CCS (Å^2)
				PA	EHSS	
SOCS2-EloBC	43,341.6	43,566.5	11^+ - 13^+			
Cul5$_{NTD}$	45,462.0	45,612.8	12^+ - 15^+			
SOCS2-EloBC-Cul5$_{NTD}$	88,803.6	89,275.1	16^+ - 20^+	5,306	6,844	5,092
Cul5-Rbx1/2	103,840.0	104,218.8	19^+ - 23^+	5,988	7,822	6,061
SOCS2-EloBC-Cul5-Rbx1/2	147,181.6	148,121.5	23^+ - 26^+	7,918	10,298	7,653
				"Open" / "Closed"	"Open" / "Closed"	
NEDD8~Cul5-Rbx1/2	112,911.9	112,981.3	19^+ - 22^+	6,511 / 6,338	8,491 / 8,288	6,222
SOCS2-EloBC-NEDD8~Cul5-Rbx1/2	156,253.5	157,000.0	23^+ - 26^+	8,333 / 8,230	10,843 / 10,701	7,853

mentally determined value is slightly smaller than the theoretical value, which could indicate that the structure is slightly more compact than the model suggests.

To investigate the effect of neddylation on the complex assembly, we performedin vitro neddylation assays on the purified Cul5-Rbx2 complex (Fig. 8) and used the reaction product NEDD8 ~Cul5-Rbx2 to reconstitute neddylated full complex SOCS2-EloBC-Cul5-Rbx2. The masses for neddylated complexes (Table 4, 113 and 157 kDa, respectively) are in agreement with the theoretical mass for the addition of NEDD8. First, the same range of charge states was observed for the non-neddylated and neddylated complexes in the native mass spectra (Fig. 9). This would indicate that no significant conformational rearrangement had occurred. Second, the CCS values measured by TWIM-MS for each charge state of the neddylated complexes were compared with those of the non-neddylated ones (Fig. 5A and B), showing an increase in CCS of 150–200 Å ^{2} in each case (Table 4). We compared the experimental data with two alternative models: one "open" model that assumes a conformational change upon neddylation, as observed crystallographically for NEDD8 ~Cul5$_{CTD}$ -Rbx1 (27), and a second "closed" model that simply has NEDD8 added onto the non-neddylated complex without any conformational rearrangement, with the aim to distinguish them based on the TWIM-MS data. The

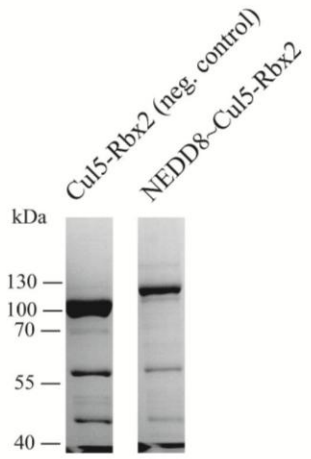


FIGURE 8. SDS-polyacrylamide gel images of the neddylation reaction products. No neddylated product was observed in the negative control reaction (left). Right, NEDD8~Cul5-Rbx2.

calculated CCS increase for the two alternative neddylated models (NEDD8 ~Cul5-Rbx1) versus the non-neddylated one (Fig. 10, A–C), is 350 Å^2 for the closed model and 523 Å^2 for the open model (PA method; Table 4), whereas for SOCS2-EloBC-NEDD8 ~Cul5-Rbx1, the neddylation accounts for an extra 312 Å^2 (closed) and 415 Å^2 (open). It would therefore appear that the increase in size is predominantly due to the addition of NEDD8 rather than a significant conformational change.

DISCUSSION

Here, we show that all CRL5^{SOCS2} components SOCS2, EloBC, Cul5, and Rbx2 can be specifically pulled down from the human cell lysates with subsequent validation of their identity by MS analysis. These components were recombinantly expressed and purified and then assembledn vitro into different sized complexes up to the full-size E3 ligase.

Biophysical studies of the full-size CRLs are important for better understanding the principles of assembly and to gain insight into their structural architecture. This is particularly relevant for the cases where crystal structures are not available, as for CRL5^{SOCS2}. We addressed this by presenting the first report of in vitro assembly offull-size human CRL5^{SOCS2} reconstituted from recombinant components and provide a biophysical analysis of the obtained complexes.

The structural model of the SOCS2-EloBC-Cul5-Rbx2 complex was validated by TWIM-MS studies. The experimentally measured CCS values are in agreement with the theoretically calculated ones, although the molecular architecture ofSOCS2-EloBC-Cul5$_{NTD}$ and SOCS2-EloBC-Cul5-Rbx2 appears to be slightly more compact than predicted.

Modification of the cullin scaffold with NEDD8 protein is crucial for the activation of CRLs (40). Previous structural studies showed that NEDD8 promotes a conformational rearrangement of the Cul5-Rbx1 component of CRL5 (27). Such a structural alteration enables Rbx1-E2~ubiquitin to extend toward the substrate receptor subunit, thereby promoting substrate polyubiquitination. One of the proteins identified in the pull-down experiments was NEDD8, supporting the presence of the active neddylated complex inside cells. Comparison of TWIM-MS data between neddylated and non-neddylated Cul5-Rbx2 and SOCS2-EloBC-Cul5-Rbx2 complexes showed an increase of 150–200 Å^2 in CCS values (Table 4). Interestingly, the increase in calculated CCS values defined by the addition of NEDD8 is ~2–3 times larger than the conformational change of the Rbx1 subunit in the open models (Fig. 10, C and F). Therefore, the difference between open and closed neddylated models is not significant enough, and we cannot distinguish them based on the experimental data. The observed change in CCS is therefore largely due to the addition of the extra NEDD8 subunit.

One of the main limitations of biophysical studies of the whole multisubunit CRLs is the difficulty of obtaining all of the components in appropriate amount and quality, particularly

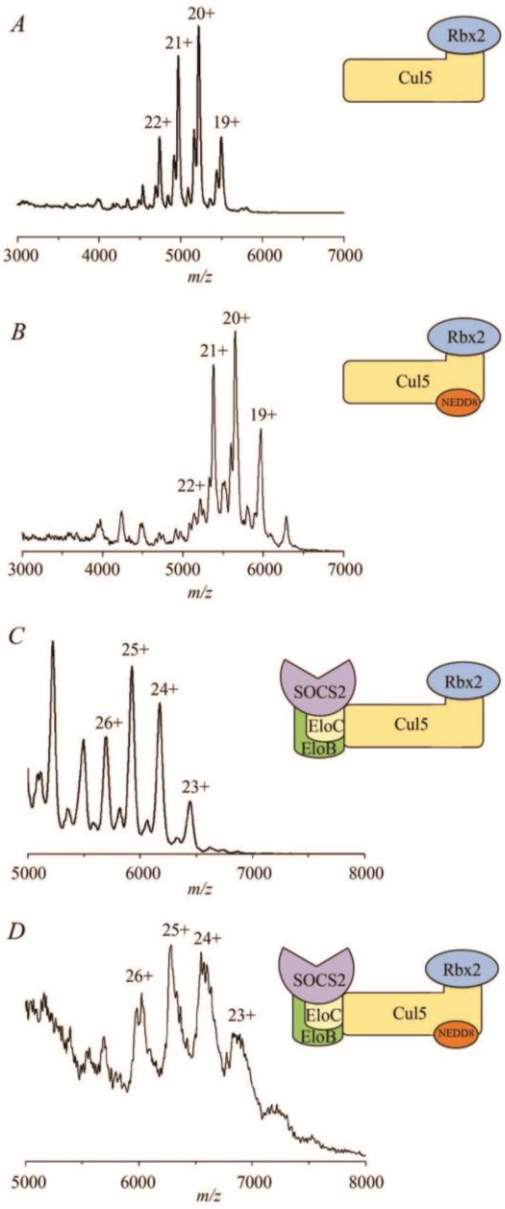


FIGURE 9. Comparison of native MS spectra. A, Cul5-Rbx2; B, NEDD8~Cul5-Rbx2; C, SOCS2-EloBC-Cul5-Rbx2; D, SOCS2-EloBC-NEDD8~Cul5-Rbx2.

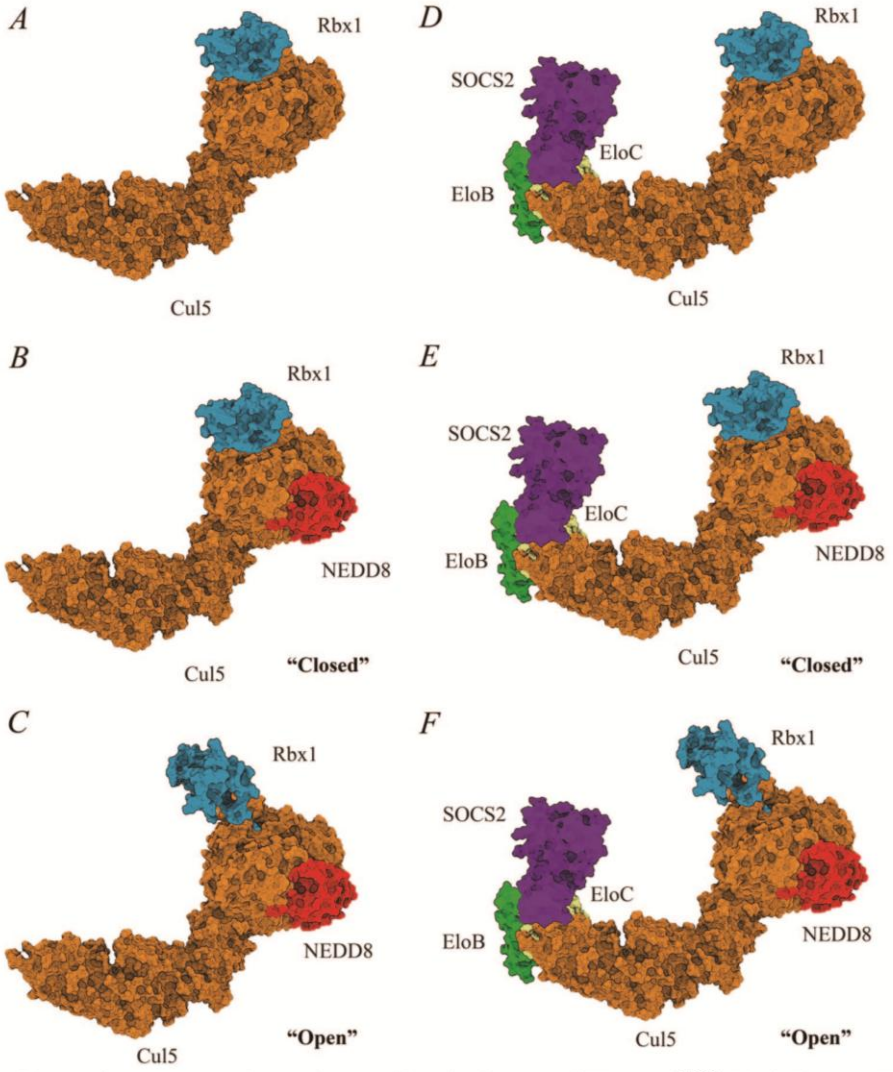


FIGURE 10. Structural models provide important insights into the assembly and architecture of CRL5^{SOCS2}. A, Cul5-Rbx1; B, NEDD8~Cul5-Rbx1 (closed model); C, NEDD8~Cul5-Rbx1 (open model); D, SOCS2-EloBC-Cul5-Rbx1; E, SOCS2-EloBC-NEDD8~Cul5-Rbx1 (closed); F, SOCS2-EloBC-NEDD8~Cul5-Rbx1 (open). The models were assembled using available crystal structures. Due to the lack of Rbx2 crystal structure, its closest homolog, Rbx1, was used instead. These models were used for calculation of the theoretical CCS values.

full-length cullins. Expression and purification of stable full-length cullin scaffolds in complex with RING domain proteins is not trivial and has been previously reported only for Cul1-Rbx1 (26), Cul4A-Rbx1 (41), Cul4B-Rbx1 (42), and Cul5-Rbx2 (20). As a result, there were only a few cases in the literature describing characterization of the full-size CRLs assembled from recombinant subunits (26, 42). To purify the Cul5-Rbx2 complex in this study, we used the Dac tag technology, which provides additional stability and solubility to the protein complex and additionally improves the yield of recombinant proteins (19). This approach has also proven to be successful for purification of Cul2-Rbx1 complex [4] and could be further extended to other cullins and large multisubunit complexes.

In certain cases, CRLs exist and function in homo- or hetero-oligomeric states. The biological implications of CRL oligomerization are postulated to include activity regulation, enhancement of substrate ubiquitination, and alternative mechanistic aspects of ubiquitin transfer (6). For example, several studies have shown that CRL3 can dimerize via an adaptor BTB domain (43) or through NEDD8-mediated interaction between two Cul3 scaffolds (44). CRL1 was also demonstrated to be able to dimerize via the receptor Cdc4 (cell division control protein 4), resulting in enhanced ubiquitination of substrate Sic1 (45). Additional examples include other BTB receptor/adaptor subunits of CRL3 (46–49), F box receptors of CRL1 (45, 50–53), and the DCAF receptor of CRL4 (54). More recently, a two-site model for substrate recognition was proposed for CRL3^{KLHL11} based on the crystal structure of KLHL11-Cul3$_{NTD}$ (49). However, no evidence for dimerization of elongins, cullin 2 or cullin 5, or SOCS subunits has been reported to date. In this work, using SEC-MALS and native MS techniques, we have established that CRL5^{SOCS2} exists in a monomeric state. We provide

[4] A. Knebel, unpublished results.

a structural model validated by TWIM-MS studies that suggests a similar mechanism of ubiquitin transfer to a previously reported monomeric CRL1^{Skp2} complex (26). Therefore, according to the proposed model, CRL5^{SOCS2} oligomerization does not seem to be necessary for enzyme activity.

Our measured K_d values for the interaction of SOCS2-EloBC with GHR_pY595 peptide or Cul5$_{NTD}$ are in good agreement with previously reported data (11, 31). Weak SOCS2-EloBC/GHR_pY595 interaction ($K_d = 1.8\ \mu M$) could suggest low selee tivity toward a particular substrate and instead the ability to target a variety of phosphorylated proteins. In contrast, the interaction of SOCS2-EloBC with scaffold Cul5$_{NTD}$ is very tight (K_d = 11 nM at 298 K). The large negative ΔC_p value (−450 cal/mol/K) for the SOCS2-EloBC/Cul5$_{NTD}$ interaction indicates a major contribution of the hydrophobic interface and further reflects the high affinity (34, 38) (e.g. when compared with other related interactions, such as ASB9-EloBC/Cul5$_{NTD}$ and Vif-EloBC/Cul5$_{NTD}$) (18, 30). Overall, these results indicate the structural importance of the SOCS2-EloBC/Cul5$_{NTD}$ interface for assembly and stability of the CRL5^{SOCS2}.

As the next logical step following the current study, we believe it would be important to develop an assay to measure activity of the recombinant CRL5^{SOCS2} against the substrate GHR protein resulting in ubiquitination and the subsequent proteasomal degradation of the latter. Such an assay could be useful for testing the potency of small molecule modulators of CRL5^{SOCS2} activity. In accordance with this, a recent example demonstrates in vitro reconstitution of murine CRL5^{SOCS3}, containing SOCS3, a close homolog of SOCS2, as a substrate receptor subunit (55). The authors used co-expressed Cul5$_{NTD}$, Cul5$_{CTD}$, and Rbx2 proteins to form a complex with SOCS3-EloBC, and the assembled E3 ligase then demonstrated activity in the ubiquitination assay against substrates JAK2 and gp130.

In addition, it would be important to obtain the crystal structure of the receptor SOCS2 bound to the substrate GHR depicting the details of the interface between these two proteins. This could substantially advance the development of inhibitors of this interaction (i.e. structural phosphotyrosine analogs or isosteres). The biophysical insights into the interactions and assembly of the full-size CRL5^{SOCS2} E3 ligase reported in our study will aid future developments in this direction.

Acknowledgments—We thank Colin Hammond and Prof. Tom Owen-Hughes (College of Life Sciences, University of Dundee) for assistance with SEC-MALS experiments.

REFERENCES

1. Sarikas, A., Hartmann, T., and Pan, Z.-Q. (2011) The cullin protein family. Genome Biol. 12, 220
2. Soucy, T. A., Smith, P. G., Milhollen, M. A., Berger, A. J., Gavin, J. M., Adhikari, S., Brownell, J. E., Burke, K. E., Cardin, D. P., Critchley, S., Cullis, C. A., Doucette, A., Garnsey, J. J., Gaulin, J. L., Gershman, R. E., Lublinsky, A. R., McDonald, A., Mizutani, H., Narayanan, U., Olhava, E. J., Peluso, S., Rezaei, M., Sintchak, M. D., Talreja, T., Thomas, M. P., Traore, T., Vyskocil, S., Weatherhead, G. S., Yu, J., Zhang, J., Dick, L. R., Claiborne, C. F., Rolfe, M., Bolen, J. B., and Langston, S. P. (2009) An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. Nature 458, 732–736
3. Feldman, R. M., Correll, C. C., Kaplan, K. B., and Deshaies, R. J. (1997) A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. Cell 91, 221–230
4. Bosu, D. R., and Kipreos, E. T. (2008) Cullin-RING ubiquitin ligases: global regulation and activation cycles. Cell Div. 3, 7
5. Lydeard, J. R., Schulman, B. A., and Harper, J. W. (2013) Building and remodelling Cullin-RING E3 ubiquitin ligases. EMBO Rep. 14, 1050–1061
6. Zimmerman, E. S., Schulman, B. A., and Zheng, N. (2010) Structural assembly of cullin-RING ubiquitin ligase complexes. Curr. Opin. Struct. Biol. 20, 714–721
7. Kamura, T., Maenaka, K., Kotoshiba, S., Matsumoto, M., Kohda, D., Conaway, R. C., Conaway, J. W., and Nakayama, K. I. (2004) VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. Genes Dev. 18, 3055–3065
8. Mahrour, N., Redwine, W. B., Florens, L., Swanson, S. K., Martin-Brown, S., Bradford, W. D., Staehling-Hampton, K., Washburn, M. P., Conaway, R. C., and Conaway, J. W. (2008) Characterization of Cullin-box sequences that direct recruitment of Cul2-Rbx1 and Cul5-Rbx2 modules to elongin BC-based ubiquitin ligases. J. Biol. Chem. 283, 8005–8013
9. Greenhalgh, C. J., Rico-Bautista, E., Lorentzon, M., Thaus, A. L., Morgan, P. O., Willson, T. A., Zervoudakis, P., Metcalf, D., Street, I., Nicola, N. A., Nash, A. D., Fabri, L. J., Norstedt, G., Ohlsson, C., Flores-Morales, A., Alexander, W. S., and Hilton, D. J. (2005) SOCS2 negatively regulates growth hormone action in vitro and in vivo. J. Clin. Invest. 115, 397–406
10. Greenhalgh, C. J., Metcalf, D., Thaus, A. L., Corbin, J. E., Uren, R., Morgan, P. O., Fabri, L. J., Zhang, J.-G., Martin, H. M., Willson, T. A., Billestrup, N., Nicola, N. A., Baca, M., Alexander, W. S., and Hilton, D. J. (2002) Biological evidence that SOCS-2 can act either as an enhancer or suppressor of growth hormone signaling. J. Biol. Chem. 277, 40181–40184
11. Bullock, A. N., Debreczeni, J. E., Edwards, A. M., Sundstro¨m, M., and Knapp, S. (2006) Crystal structure of the SOCS2-elongin C-elongin B complex defines a prototypical SOCS box ubiquitin ligase. Proc. Natl. Acad. Sci. U.S.A. 103, 7637–7642
12. Kim, Y. K., Kwak, M.-J., Ku, B., Suh, H.-Y., Joo, K., Lee, J., Jung, J. U., and Oh, B.-H. (2013) Structural basis of intersubunit recognition in elongin BC-cullin 5-SOCS box ubiquitin-protein ligase complexes. Acta Crystallogr. D 69, 1587–1597
13. Cohen, P., and Tcherpakov, M. (2010) Will the ubiquitin system furnish as many drug targets as protein kinases? Cell 143, 686–693
14. Petroski, M. D. (2008) The ubiquitin system, disease, and drug discovery. BMC Biochem. 10.1186/1471-2091-9-S1-S7
15. Andérica-Romero, A. C., Gonza´lez-Herrera, I. G., Santamaría, A., and Pedraza-Chaverri, J. (2013) Cullin 3 as a novel target in diverse pathologies. Redox Biol. 1, 366–372
16. Zhao, Y., and Sun, Y. (2013) Cullin-RING Ligases as attractive anti-cancer targets. Curr. Pharm. Des. 19, 3215–3225
17. Bantscheff, M., Hopf, C., Savitski, M. M., Dittmann, A., Grandi, P., Michon, A.-M., Schlegl, J., Abraham, Y., Becher, I., Bergamini, G., Boesche, M., Delling, M., Du¨mpelfeld, B., Eberhard, D., Huthmacher, C., Mathieson, T., Poeckel, D., Reader, V., Strunk, K., Sweetman, G., Kruse, U., Neubauer, G., Ramsden, N. G., and Drewes, G. (2011) Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. Nat. Biotechnol. 29, 255–265
18. Thomas, J. C., Matak-Vinkovic, D., Van Molle, I., and Ciulli, A. (2013) Multimeric complexes among ankyrin-repeat and SOCS-box protein 9 (ASB9), ElonginBC, and Cullin 5: insights into the structure and assembly of ECS-type Cullin-RING E3 ubiquitin ligases. Biochemistry 52, 5236–5246
19. Lee, D. W., Peggie, M., Deak, M., Toth, R., Gage, Z. O., Wood, N., Schilde, C., Kurz, T., and Knebel, A. (2012) The Dac-tag, an affinity tag based on penicillin-binding protein 5. Anal. Biochem. 428, 64–72
20. Kelsall, I. R., Duda, D. M., Olszewski, J. L., Hofmann, K., Knebel, A., Langevin, F., Wood, N., Wightman, M., Schulman, B. A., and Alpi, A. F. (2013) TRIAD1 and HHARI bind to and are activated by distinct neddylated Cullin-RING ligase complexes. EMBO J. 32, 2848–2860
21. Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T., and Scrivens, J. H. (2007) An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument.

Chapter 9

Conclusions and Outlook

9.1 General conclusions

The set of mass spectrometry-based methods used in this thesis, provides an insight regarding structural information concerning several scientific projects. A variety of samples could be analysed that ranged from monoclonal antibodies, nanobodies and Fc-fusion proteins to RNA, kinase complexes and other protein complexes. This highlights the diverse applications this set of methods can address. However, proper data analysis and interpretation remains very important.

Native IM-MS

With native IM-MS, the molecular weight and collision cross section of various antibodies (IgG1, IgG2 and IgG4) and antibody fragments (Fab and Fc generated by lysine digest, GingisKHAN digest or IdeS digest) could be determined. Although the antibody subtypes differ in their hinge region a similar size and shape could be calculated for each antibody. Antibody conjugates were studied with native IM-MS to check the amount of ligands bound to the antibody and to confirm that no significant conformational changes occurred due to the covalent modification, which mean that the overall structure is preserved.

Using native IM-MS, the size and shape of an Fc-fusion protein was compared to its Fc-analogue. It was also shown how the presence of the peptide could induce dimerisation, as the mass spectrum of the Fc-analogue alone showed monomer Fc only.

Nanobodies and ScFv were also studied with native IM-MS. Similar studies were conducted for the complexes with their antigen, which in this case was the plasminogen activator inhibitor 1 (PAI-1). The size and shape of five variants of PAI-1 were compared. Some of these variants could be distinguished based on their slight difference in CCS value.

The native IM-MS data for TAR-RNA was determined and was compared to when TAR-RNA interacted with a cyclic peptide.

Native IM-MS could distinguish three variants of the Aurora kinase C–INCENPA complex. When this complex interacted with a small molecule kinase inhibitor, native IM-MS could detect a conformational change caused by the inhibition.

Large protein complexes can be analysed with native IM-MS to determine the stoichiometry of the complex.

Advantages: A widely applicable and sensitive technique that can provide the mass, size and shape of proteins and non-covalent protein complexes using only a few microliters of sample at micromolar concentration. The calculated collision cross sections derived from the IM data give a quick estimate of the size and of the possible conformations. Using IM-MS the effect of covalent modifications on the protein structure can also be investigated.

Disadvantages: As the measurement occurs in the gas phase, during the interpretation of the results the possible effect of the solution to gas transition needs to be considered. Furthermore, a thorough purification and a volatile buffer are required for resolved spectra. The calculated CCS values have proven to be a good indication of the size of the analyte, but in most cases are still different from the results of modeling studies (gas-phase vs. solution?) Thus, this result should be a complementary estimate of the size (together with e.g., X-ray, NMR structures, or computational modeling). Very small structural differences (as little as ΔCCS ~ 2%) can be

detected with the IM-MS technology used here, but need to be carefully verified and are difficult to interpret.

Collision induced unfolding

With collision-induced unfolding, a unique unfolding pattern could be plotted for every protein and protein complex, which aids in distinguishing protein variants. The IgG subtypes IgG1, IgG2 and IgG4 were distinguished based on their unique CIU plot. Slight differences were observed in CIU plots of monoclonal antibodies after freeze-thaw cycles. The stability of an antibody conjugate could also be monitored via CIU. By comparing the CIU plot of an Fc-fusion protein with that of its Fc-analogue, it could be shown that the Fc-fusion is more stable as it could tolerate higher voltages. The CIU plots of antibody-antigen complexes were also investigated to monitor their stability and unfolding patterns.

Advantages: The CIU technique can distinguish analytes based on their unique unfolding pattern, allowing a classification of the studied molecules. Furthermore, the stability of a protein or protein complex can be monitored in one simple experiment. CIU plots can be used to compare the effect of one mutation, bound ligand, etc. with others, i.e. they allow to classify such structure perturbing factors fast and efficiently, without always easily explaining what changes are occurring.

Disadvantages: The interpretation of a CIU plot is not always straightforward. It is important that these plots are reproducible. In practice, there is also a stable and intense MS signal necessary for a reliable CIU plot.

Native top-down ETD

Native top-down ETD was attempted on intact antibodies and antibody fragments and showed limited sequence coverage. This could have been a result of the Synapt G2 instrument not being sensitive enough to detect all fragments, or the instrument may have needed more fine-tuning to get better sequence coverage. However, previous experiments have shown better performance of the instrument. Thus, the instrument parameters must have been not sufficiently optimised for the experiments.

*Advantages: Using native top-down ETD, instead of denatured, allows the protein (complex) to maintain it*s non-covalent bonds during the measurement. This results in additional information on the surface accessible area of the protein or protein complex.

Disadvantages: The instrument needs to be properly tuned to detect the ETD fragments, which is not always straight forward. The precursor signal intensity and the intensity of the radical anion need to be sufficiently high to generate intense fragments. Nonetheless long acquisition times (>1000 scans) are needed to obtain intense fragment ion spectra. Furthermore, improved software should be available to analyse and interpret the raw data. Thus, the data analysis and interpretation can be tedious.

HDX-MS

HDX-MS was applied on an Fc-fusion protein at various concentrations to track changes in deuterium uptake, in order to locate the dimer interface. The deuterium uptake of the fused peptide dropped when increasing the protein concentration, which indicated that the Fc-fusion protein dimerises via the linked peptide. This was also suggested from the native MS data, as

the presence of the linked peptide was needed in order to see Fc-fusion protein dimer in the mass spectra.

Advantages: Information can be obtained about interaction sites, flexible protein regions and the solvent accessible surface area. The instrumentation and data analysis software have been improved to keep up with the updated technique and high-throughput measurements.

Disadvantages: Despite a range of useful HDX-MS software, data analysis can still be time-consuming. Furthermore, the HDX labeling is a reversible process, so there is chance of back-exchange during the experiment. It is also necessary to obtain good sequence coverage of the protein for proper data analysis.

FPOP

Monoclonal antibodies were oxidised using FPOP and the oxidised antibodies were analysed with native IM-MS to check the effect of FPOP on the native structure. Based on the results, the native structure was not significantly influenced by oxidation. The oxidation sites, however, could not be determined during this FPOP experiment, as something went wrong during the enzymatic digestion.

Advantages: Highly reactive and irreversible labeling technique, which can be used to investigate protein interactions and solvent accessible surface area. Unlike HDX-MS, it can label amino acid side chains.

Disadvantages: A relatively new technique, so instrumentation and data analysis software need to be further optimised for high-throughput experiments and reproducible results.

9.2 Creating a fingerprint for biologics

The set of MS-based methods discussed in this thesis can offer a range of structural information on proteins. This set of methods can be looked at as a toolbox to solve scientific questions by approaching the analyte from different perspectives.

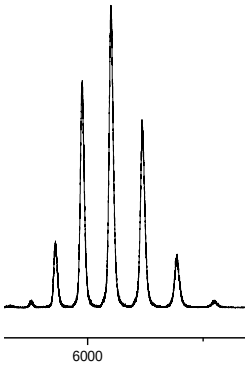
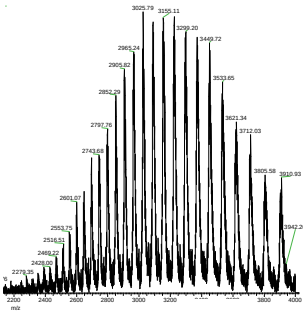
To thoroughly understand the structure and structure-function relationships of a protein it is important have an overview or a summary of its structural characteristics.

For biotherapeutics, or biologics, it is equally important to have a profile or fingerprint of the biomolecule of interest (antibody, antibody-drug conjugate, peptide, RNA, etc.) as it can be useful for protein engineering or for the quality control of the biotherapeutic drug.

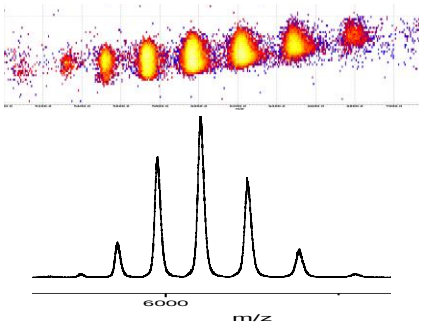
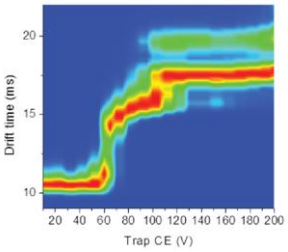
Table 1 (I-IV) displays an overview of how the discussed set of MS-based methods (focused on higher-order structure) can be used to generate a unique fingerprint for a biologic.

Table 1: An overview of the different techniques to generate a unique profile for a biologic.

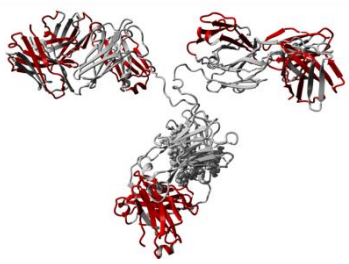
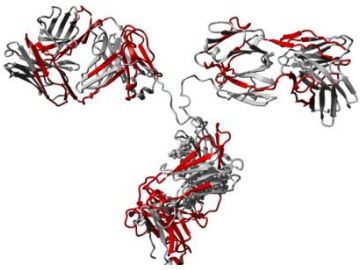
I

	Native MS	Denatured MS
Techniques	m/z intact native IgG	m/z intact denatured IgG
		
Properties	Mass of non-covalent complex	Exact mass of protein subunits
	Conformational properties	PTMs

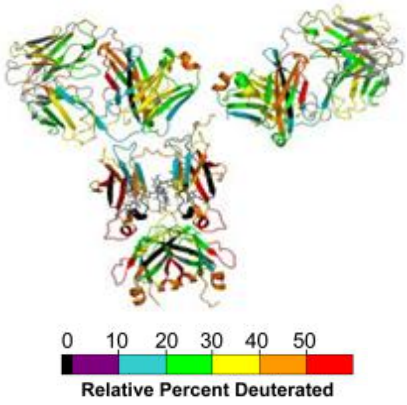
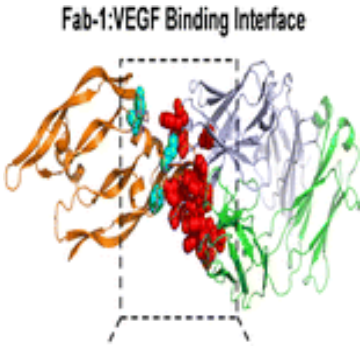
II

	Native IM-MS	CIU & CID
Techniques	m/z & IM drift time	Unfolding & dissociation pattern
		
Properties	Molecular weight and stoichiometry	Stability
	Collision cross section	Stoichiometry
	Conformational properties	Conformational properties & Domain architecture

III

	Top-down ETD	
Techniques	Sequencing Native IgG	Sequencing Denatured IgG
		
Properties	Solvent accessibility	Solvent accessibility
	Dynamics	Dynamics
	Interactions with target	PTMs

IV

	HDX-MS	FPOP
Techniques	Labels backbone hydrogens	Labels amino acid side chains
	 Data from Houde *et al.*[3]	 Data from Zhang *et al.*[4]
Properties	Solvent accessibility & secondary structure	Solvent accessibility
	Dynamics	Dynamics
	Interactions with target	Interactions with target

9.3 Outlook of MS techniques for higher-order structure determination

Mass spectrometry has played an important role in determining the primary structure of proteins. In recent years, MS-based methods that probe higher-order structure and conformational dynamics of proteins are getting more attention in the biopharmaceutical industry. With these higher-order MS techniques, links can be suggested between protein conformation and biological activity, stability and safety of biologics [1]. Top-down and bottom-up MS based techniques can be combined to generate a complete dataset in order to create a unique profile or fingerprint for the studied protein.
The MS technology is continuously improving by advances such as high-resolution mass spectrometry, e.g., quadrupole-time of flight (Q-TOF) and Orbitrap MS. The development of hybrid MS instruments also allow the ability to combine the best features from different technologies. Furthermore, sample purification is still an important step in MS experiments, making techniques such as liquid chromatography (LC) and size-exclusion chromatography (SEC) indispensable. Additionally, these techniques allow the analysis of complex formulations.

One of the limitations to overcome is to simplify data-analysis when transforming these MS-based methods to high-throughput techniques, for rapid protein characterisation.
A significant amount of manual labour is required to analyse complex sets of data (e.g. for HDX-MS, FPOP, top-down ETD). This analysis time can be shortened by developing relevant bioinformatics and reporting tools. At the same time, it is also crucial that the techniques employed offer robust detection of small changes, e.g., 1-2 % unfolding, which may affect the quality of the biologic.
To do this, detailed methodology studies are essential to comprehend the impact of changes in protein structure. It has to be ascertained whether the differences seen in the experimental data are actual changes in the sample, or a result of the variability in the samples or protocols. Furthermore, the large sets of data generated need to be scientifically and biologically relevant. While currently MS analysis occurs in-vitro, future goals are to study the same samples in vivo. Imaging spectrometers are being developed and optimised for this approach [2].
Ultimately, I predict that there will be continous technical development in the MS-instrumentation in order to provide novel capabilities with higher sensitivity, mass accuracy and resolution.

MS-based methods have shown spectacular advance in the last two decades and based on the published studies by practitioners in academia and industry this field has a promising future. The interest in the biopharmaceutical industry in these techniques is also growing rapidly. When the hurdles are overcome to make this MS-based toolbox a collection of high-throughput methods, standard protocols can be made to elucidate protein structures of primary to higher-order structure. It is very exciting to witness how this field will evolve over the next decades and how it will contribute to the unique profile of a biologic.
Until then, it will be the hard work of research scientists, PhD students and their supervisors that will keep proving and improving the power of MS-based methods.

9.4 Sources

[1] I.A. Kaltashov, C.E. Bobst, R. R. Abzalimov, S. A. Berkowitz, D. Houde, Conformation and dynamics of biopharmaceuticals: transition of mass spectrometry-based tools from academe to industry, J Am Soc Mass Spectrom, 2010, 21, 323-337

[2] C. A. Challener, Mass Spectrometry Measures Up to Analytical Challenges, BioPharm International, 2017, 30, 14-17

[3] D. Houde, J. Arndt, W. Domeier, S. Berkowitz, J. R. Engen, Rapid characterization of IgG1 conformation and conformational dynamics by hydrogen/deuterium exchange mass spectrometry, Anal Chem, 2009, 81, 2644-2651

[4] Zhang, Y., et al., Mapping the Binding Interface of VEGF and a Monoclonal Antibody Fab-1 Fragment with Fast Photochemical Oxidation of Proteins (FPOP) and Mass Spectrometry. J Am Soc Mass Spectrom, 2017, 1-9

Curriculum Vitae

Curriculum Vitae

Sneha Chatterjee

CONTACT:
Hoevelei 80
2630 Aartselaar, Belgium

GSM:
+32472/69.34.19
E-mail:
sneha1302@yahoo.com

PERSONALITY:
Responsible
Independent
Patient
Punctual
Creative
Motivated

LANGUAGES:
English (Fluent),
Dutch (Fluent),
French & Bengali

INTERESTS:
Saxophone, piano, science and travel.

WORK EXPERIENCE

2018-*Present* Senior Consultant Gi^3 team Deloitte
R&D grants, tax credits and incentives
Deloitte Belgium, Brussels, BE

EDUCATION

2014 – 2017 PhD in Chemistry, IWT scholarship
Thesis: The characterisation of proteins of biotherapeutic interest by novel mass spectrometry-based methods.
BAMS research group, University of Antwerp, BE

2014-2017 Ombuds Bachelor Students Chemistry, UA

2011-2013 Master in Chemistry
Thesis: Cross-linking of prokaryotic SMC protein complexes.
BAMS research group, University of Antwerp, BE

2008-2011 Bachelor in Chemistry
Thesis: Development of an internal standard for the LC-MS analysis of Melphalan-DNA adducts.
BAMS research group, University of Antwerp, BE
Thesis: Calculation of the dispersion interaction energy by DFT-D methods.
Structural Chemistry research group, University of Antwerp, BE

2002-2008 High school diploma
Major: Latin & Sciences
Sint-Lievens college, Antwerp, BE

SKILLS & EXPERIENCE

Technical skills:
Native mass spectrometry, ion mobility, top-down proteomics, cross-linking, liquid chromatography-mass spectrometry, analytical chemistry and proteomics.
Professional skills:
Planning and organizing multiple projects, team-work, supervising students, time management, and problem-solving.

Collaborations with universities and industries:
MedImmune (UK): Research stay at the analytical biotechnology department, protein analysis.

J&J (BE & US): MS analysis of monoclonal antibodies.

CTDP (Global): Consortium of top-down proteomics of monoclonal antibodies, a 25 laboratory study.

KULeuven (BE): MS analysis of nanobodies.

Oxford University (UK): MS analysis of kinase complexes.

TALKS & POSTER PRESENTATIONS

21 March 2017 COST Research conference on native MS and structural proteomics, Heraklion, Crete, GR
Talk: Ion mobility methods to characterize kinase complexes
21-24 Jan 2016 Sanibel conference, Characterization of Protein Therapeutics by Mass Spectrometry, Clearwater, Florida, USA
Poster: IM-MS characterization of Fc-Fusion proteins
19-20 Nov 2015 Symposium on Structural Proteomics, Halle, DE
Poster: IM-MS characterization of Fc-Fusion proteins
24-25 Nov 2014 Affinity-MS in Life Sciences, Opel-Forum, Rüsselsheim, DE
Talk: Ion mobility methods for the characterization of dynamic protein complexes
8-9 April 2014 Strasbourg Workshop on Native IM-MS, Institut Pluridisciplinaire Hubert Curien, Strasbourg, FR
Poster: Novel cross-linking and detection strategies for recombinant proteins
28 March 2014 GOA 4D Protein structure meeting, University of Antwerp, BE
Talk: Characterization of biotherapeutic proteins by novel MS-based methods
3 May 2013 Belgian Society for Mass Spectrometry meeting in Antwerp, BE
Poster presentation: Cross-linking of SMC proteins

AWARDS

21 Jan 2016 ASMS Student Stipend at Clearwater Beach, FL, USA
Travel grant to attend and present at the ASMS Sanibel conference in Clearwater, FL, USA
5 Nov 2015 ASMS Student Stipend at Seattle, WA, USA
Travel grant to attend the ASMS fall workshop in Seattle, WA, USA
1 Jan 2014 IWT scholarship at Brussels, BE
Research grant to start PhD on 'The characterization of Proteins of biotherapeutic interest by novel MS-based methods'.

ATTENDED CONFERENCES

30 June–02 July 2015 Waters HDX & Advanced MS Characterization, Manchester, UK

15-16 June 2015 Next generation antibodies and protein analysis, Ghent, BE

27-28 Nov 2014 Symposium on Structural Proteomics, Antwerp, BE

22 Oct 2014 Waters Biopharm. Seminar, Zellik, BE

17-19 Sep 2014 VIB International PhD Symposium, Antwerp, BE

27 May 2014 MS in Structural Biology, Ghent, BE

17 March 2014 Structural Biology Symposium, KU Leuven, BE

PUBLICATIONS

09 Oct 2017 Marco Tassi, Jelle De Vos, Sneha Chatterjee, Frank Sobott, Jonathan Bones, Sebastiaan Eeltink, Advances in native high-performance liquid chromatography and intact mass spectrometry for the characterization of biopharmaceutical products, *Journal of Separation Science*, 2017,1-20

12 Jan 2017 Dukas Jurenas, Sneha Chatterjee, Albert Konijnenberg, Frank Sobott, Louis Droogmans, Abel Garcia-Pino, Laurence Van Melderen, AtaT blocks translation initiation by N-acetylation of the initiator tRNA(fMet), *Nature Chemical Biology* 13,640-646 (2017)

19 July 2016 Agnese Maggi, Eduardo Ruivo, Jens Fissers, Christel Vangestel, Sneha Chatterjee, Jurgen Joossens, Frank Sobott, Steven Staelens, Sigrid Stroobants, Pieter Van Der Veken, Leonie Wyffels, Koen Augustyns, Development of a novel antibody-tetrazine conjugate for bio-orthogonal pretargeting, *Organic Biomolecular Chemistry* 14 (31), 7544-7551 (2016)

11 Dec 2014: Emil Bulatov, Esther M. Martin, Sneha Chatterjee, Axel Knebel, Satoko Shimamura, Albert Konijnenberg, Clare Johnson, Nico Zinn, Paola Grandi, Frank Sobott and Alessio Ciulli, Biophysical studies on interactions and assembly of full-size E3 ubiquitin ligase: suppressor of cytokine signaling 2 SOCS2):ElonginBC:Cullin5:RING-box protein 2 (Rbx2),
Journal of Biological Chemistry 290,4178-4191 (2015)

MANUSCRIPT IN PREPARATION

- Top-down MS of monoclonal antibodies (Chapter 3.4)
- MS-based characterisation of a Fc-fusion protein (Chapter 4)
- Analysis of Nanobodies and ScFv against PAI-1 (Chapter 5)
- Conformational analysis of Aurora C Kinase complex (Chapter 7)