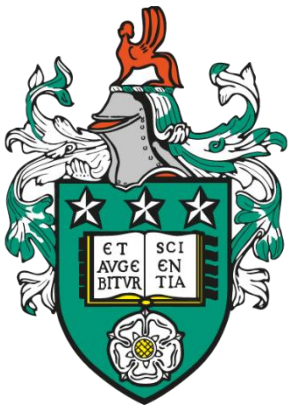
Detection of flow-induced perturbations of antibody structure using mass spectrometry



Samantha Marie Lawrence

University of Leeds

Astbury Centre for Structural Molecular Biology

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

February 2023

Declaration

The candidate confirms that the submitted work is her own, where appropriate credit has been given within the thesis to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from this thesis may be published without proper acknowledgement.

Throughout this thesis, the work directly attributable to the candidate is as follows:

i) Literature search and compilation of the manuscript.

ii) Candidate has performed all the experimental work and data analysis unless otherwise stated in the text.

Acknowledgements

Firstly, many thanks go to my supervisors, Professors David Brockwell, Sheena Radford and Frank Sobott. Your continued support and countless meetings over the last four and a half years have been imperative to completing this thesis. Frank, thank you for jumping into my supervisory team early on in the project – having a trio of Professors has been a brilliant adventure.

I wish to thank the BBSRC Collaborative Training Partnership and my industrial partner AstraZeneca for the funding for this project. Specifically I would like to thank Dr Nick Bond for his support, I feel honoured to have worked with you both in my undergraduate degree and throughout my postgraduate degree. I want to thank Dr Janet Saunders for her help with mAb expression and purification; Dr Piera Marchetti for the HDX results and for answering all of my many questions; Dr Romina Hofele for generating the PyMol structures; Dr Paul Devine for encouraging me to apply to Leeds for a PhD in the first place; and a big thank you to Drs Christopher Lloyd, Nicholas Darton and David Lowe for your discussions over the years in our AZ-Leeds meetings.

To everyone who has been a friendly face over the last four and a half years, in the lab, on level 10, and throughout FBS, I am so grateful for the great memories throughout my PhD. It has been fantastic to have been a part of the Squishers team – Dr Leon Willis, Dr Ioanna Pagani, and Alex Page, thank you for all the science chats and laughs over the years. I will always remember our after-meeting debriefs in the small office! Thank you Prof Nik Kapur for your input during the Squishing meetings. To Sophie Cussons, thank you for letting me distract you from work for lengthy catch ups about houses, pets and weddings. To Dr Lorna Kelly, thank you for your friendship in my first year and all the adventures between then and now. Huge thanks go to Nasir Khan, for being a fantastic lab manager and always being up for a cuppa and a natter; and of course your generosity over the years supplying a never-ending source of biscuits, yummy curries, and beautiful plants! To my cohort, Dr James Whitehouse, Dr Samuel Haysom and Dr Romany Horne, I am so grateful to have laughed and worked alongside you! Romany, thank you for being such a wonderful friend throughout the years of PhD. Your enthusiasm and drive have kept me going, and I am constantly in awe of your – and Dr Jim Horne's – passion for science. I was honoured to be your Maid of Honour, and I am honoured to call you my friend.

Thank you to everyone who is, and has been, part of the Mass Spec team. Specifically Dr James Ault and Dr Rachel George, thank for the huge amounts of support during my times in MS. Dr Anton Calabrese, thank you for being a voice of motivation when things have been tough. To the MS crew, I'm so grateful for all the trips to Old Bar, and for welcoming me into the catacombs/office for work- and non-work catch ups. To my USA gals – you are amazing! Mel, my east-Leeds buddy, thank you

for being an integral part in my Chapter 4; working alongside you – and taking long Café Nero breaks – are amongst my favourite times during this PhD.

To my wonderful friends who have been instrumental in supporting me since before (and throughout) this PhD. Many thanks go to Dr Matthew Edgeworth for the laughs and lessons at the beginning of my Mass Spec journey at AZ. Dr Sarah Harbach, thank you for all of the many catch ups, PhD rants, and for your support throughout our 8+ years of Biochemistry. To Georgie Milburn, Sarah Quickfall, Rebecca Horton, and Claire Lancaster, thank you for always being available for a motivational call and for fabulous weekend adventures.

(Un-acknowledgements go to COVID-19, for disrupting years 2 and 3. Doing a lab-based PhD in a global pandemic is interesting!)

The biggest thank you goes to my family, Mum, Dad and Matt, for always being ready for a family catch up and a hike in the Peaks. Thank you for your unwavering support when times were stressful, your help with ferrying belongings about when moving houses, your motivating positivity and the best hugs.

And finally, to my incredible fiancé Mike. Without your unwavering love, endless patience, hilarious silliness, and ability to keep me constantly supplied with tea, writing this thesis would have been a million times harder. I'm eternally grateful you decided to humour me at the start of my PhD by coming to dance classes when no-one else would, and for inviting me to help you 'focus' during your write-up by chain-drinking Miro cappuccinos and talking about anything but science. Thank you for sharing your dreams with me, and for letting me turn your house into *our* house, with a bit of organisation and a lot of love, including three wonderful work-from-home distra-cat-ions. I love you. Here's to becoming the Drs Davies!

Abstract

Monoclonal antibody biopharmaceuticals have emerged over the last few decades as a powerful class of therapies, with exquisite specificity and safety. However, these protein-based therapies are prone to environmental stresses which can trigger unfolding and aggregation, causing major roadblocks to their manufacture at many stages. Hydrodynamic flow, which describes the flow of fluid, can cause stress to proteinaceous molecules in solution, as they encounter forces which apply stress (including shear and extensional stress) to the proteins as they move with velocity in a particular environment. Hydrodynamic flow-induced can cause aggregation of proteins, and there is an urgent need to characterise the mechanisms involved in unfolding and aggregation to inform the manufacture of therapies more resistant to these issues in the future.

This thesis presents a set of three monoclonal antibodies (mAbs) with similar sequences but strikingly different physiochemical properties. Previously, an anti-nerve growth factor mAb, MEDI-1912 (WFL herein), was developed with picomolar affinity for its target but exhibited poor biophysical properties (including low solubility, a long retention time in a high performance size exclusion chromatography (HP-SEC) column, and self-association to form higher order species analysed by analytical ultracentrifugation (AUC)). The region of this mAb responsible for the poor properties was identified, leading to the development of a triple mutant, STT, with improved properties. Aggregation-prone WFL and its less-aggregation-prone counterpart STT were studied, alongside an additional variant 114 identified through a directed evolution screen of WFL. 114 scored more highly than WFL and STT in an assay previously used to rank aggregation-prone species, indicating its reduced aggregation potential. A combination of chromatographic, spectroscopic, and mass spectrometric techniques were employed, with the intention to contribute to an understanding of what may cause the difference in properties of these three therapeutically-relevant highly homologous mAbs. Even through 114's retention of aggregation hotspot residues W, F and L, its 4 additional mutations rescue the beneficial physiochemical properties seen for STT. The susceptibility of these mAbs to flow-induced aggregation were ranked, and formed the basis of the method built upon in the second and third results chapters.

To build a method capable of fingerprinting the stages of unfolding with a covalent labelling approach, various digestion methods and liquid chromatography (LC) methods were compared, iteratively building a robust liquid chromatography-tandem mass spectrometry (LC-MS/MS) methodology for use in the final chapter. Tandem mass spectrometry refers to the use of the mass spectrometer for fragmentation of introduced peptides into individual amino acids,

allowing for the sequencing of the introduced sample and detection of any modification to each individual amino acid by piecing together the resulting spectra. Thus, the fast photochemical oxidation of proteins (FPOP) procedure for conformation-sensitive labelling was applied to STT and successfully combined with the optimised LC-MS/MS protocol, for the generation of modified peptides which could be compared with un-modified peptides.

Finally, the FPOP-LC-MS/MS analysis of the three mAbs in parallel revealed strikingly different oxidation patterns between the complementarity determining regions (CDRs). The peptide-level analysis of the oxidation patterns of these mAbs revealed that regions in the heavy chain complementarity determining regions (CDRs) showed the most variation in labelling pattern between the three mAbs. Additionally, the mAbs were also subjected to hydrodynamic flow, and this 'stressed' sample was compared to the native sample using the optimized FPOP-LC-MS/MS protocol. WFL demonstrated the most protection from oxidation in CDRs, whereas 114 showed slight increases in oxidation for most peptides, with STT remaining the most constantly labelled before and after flow stress. These experiments demonstrate the applicability of using peptide-level FPOP analysis to begin to unpick the mechanisms of flow-induced unfolding.

Overall, the data presented here provides a springboard for future exploration of the flow-FPOP-LC-MS/MS methodology. Understanding the flow-induced structural perturbations of proteins will ultimately aid the design of more unfolding-resistant variants, and contribute to the economic production of current and next generation high-value biopharmaceuticals.

Table of contents

List of figures

List of tables

List of equations

List of abbreviations

AC	Alternative current
AC-SINS	Affinity-capture self-interaction nanoparticle spectroscopy
ADH	Alcohol dehydrogenase
APR	Aggregation-prone region
AUC	Analytical ultracentrifugation
BSA	Bovine serum albumin
CDR	Complementarity-determining region
CEM	Chain ejection model
C_H	Constant domains in the heavy chain (1, 2 and 3)
CHO	Chinese hamster ovary
C_L	Constant domains in the light chain
CID	Collision induced dissociation
CRM	Charged residue model
dAbs	Single domain antibodies
DC	Direct current
DDA	Data dependent acquisition
DIA	Data independent acquisition
DLS	Dynamic light scattering
DSF	Differential scanning fluorimetry
DTT	Dithiotheitol
ECD	Electron capture dissociation
EFD	Extensional flow device
EM	Electron microscopy

ESI	Electrospray ionization
Fab	Antigen binding fragment
Fc	Fragment crystallizable
FDA	Food and Drug Association
FPOP	Fast photochemical oxidation of proteins
Fv	Variable fragment
FWHM	Full width at half maximum
HDX	Hydrogen-deuterium exchange
HEK	Human embryonic kidney
HIC	Hydrophobic interaction chromatography
HOS	Higher order species
HPLC	High performance liquid chromatography
IAM	Iodoacetamide
IEM	Ion evaporation model
IEX	Ion exchange chromatography
Ig	Immunoglobulin
IMS	Ion mobility spectrometry
LC	Liquid chromatography
LIT	Linear ion trap
mAb	Monoclonal antibody
MeCN	Acetonitrile
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MWCO	Molecular weight cut off
m/z	Mass to charge ratio

NGF	Nerve growth factor
NMR	Nuclear magnetic resonance spectroscopy
ppm	Parts per million
PTM	Post translational modification
RF	Radio frequency
RMSD	Root mean square deviation
RPLC	Reverse phase liquid chromatography
RT	Retention time
SAP	Spatial aggregation propensity
SASA	Solvent accessible surface area
scFv	Single-chain variable fragments
SEC	Size exclusion chromatography
SMAC	Standup monolayer adsorption chromatography
TAP	Therapeutic Antibody Profiler
TFA	Trifluoroacetic acid
ToF	Time of flight
UV	ultraviolet
V_H	Variable domain on the heavy chain
V_L	Variable domain on the heavy chain
vWF	von Willebrand Factor
XIC	Extracted ion chromatogram

Chapter 1: Introduction Part 1

Protein structure, misfolding and aggregation in biopharmaceuticals

---

1 Introduction

1.1 Part 1 – Protein structure, misfolding and aggregation in biopharmaceuticals

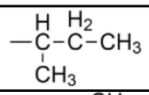
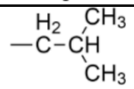
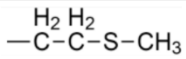
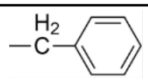
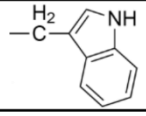
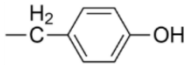
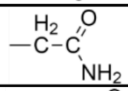
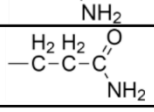
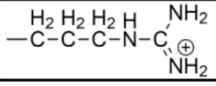
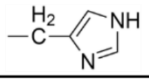
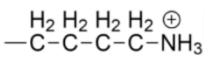
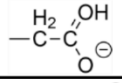
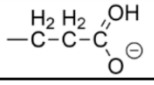
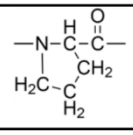
1.1.1 Protein folding

Proteins are essential for maintaining almost all biochemical processes. Their three-dimensional conformation, dictated by the sequence of amino acid building blocks, provide each protein with specific functions. A protein's conformation is created by the folding of one or more polypeptide chains, where these chains are composed of a mix of twenty different types of amino acids, all with differing chemical and structural arrangements of their side chains.

In the 1960s, key scientific advancements led the exponential interest in the fields of structural biology and protein folding. Anfinsen's thermodynamic experiments on ribonuclease A demonstrated its ability to refold after denaturation without the help of any biological machinery; he postulated that the ability for the protein to find its lowest free-energy state, both thermodynamically stable and functional, must arise from the primary amino acid sequence alone (Anfinsen *et al.*, 1961). However, Levinthal's work on the protein folding problem demonstrated that a random search for the correct state of a folded protein is not an effective method of finding the stable energy-minimised structure (Zwanzig *et al.*, 1992), and random sampling all of the possible conformations of folding would be impossible on a biologically relevant timescale (Levinthal, 1968). As proteins can find and adopt their native fold in as little as a few microseconds, it is clear that there must be other forces at work in protein folding, other than just the primary sequence randomly sampling folds. In fact, the folding equilibrium and kinetics of a certain sequence are influenced by physical chemistry (Dill *et al.*, 2012), where there is a hierarchy of folding patterns, locally (secondary) and more global (tertiary) folding influences.

The variety of shape, size, charge and hydrophobicity of the amino acid side chains (Table 1.1) all contribute to the precise fold a protein. The functionalities of amino acid side chains can also be expanded through post-translational modifications (mostly occurring post-folding), which further increase the diversity of a protein's capabilities, for example adjustment of its surface properties or tuning the specificity to perform its function.

Table 1.1 Summary of the twenty naturally occurring amino acids, including their abbreviations, side chain structures, and the monoisotopic masses of the condensed residue (average mass can be found in Appendices Table 7.1).

Name	Abbreviation	Side chain (R group)	Classification	Condensed residue monoisotopic mass (Da)
Alanine	Ala, A		Hydrophobic, small	71.04
Isoleucine	Ile, I		Hydrophobic	113.08
Leucine	Leu, L		Hydrophobic	113.08
Methionine	Met, M		Hydrophobic, sulfur-containing	131.04
Phenylalanine	Phe, F		Hydrophobic	147.07
Tryptophan	Trp, W		Hydrophobic	186.08
Tyrosine	Tyr, Y		Hydrophobic	163.06
Valine	Val, V		Hydrophobic	99.07
Asparagine	Asn, N		Polar uncharged	114.04
Glutamine	Gln, Q		Polar uncharged	128.06
Serine	Ser, S		Polar uncharged	87.03
Threonine	Thr, T		Polar uncharged	101.05
Arginine	Arg, R		Positive charge	156.10
Histidine	His, H		Positive charge	137.06
Lysine	Lys, K		Positive charge	128.09
Aspartic acid	Asp, D		Negative charge	115.03
Glutamic acid	Glu, E		Negative charge	129.04
Cysteine	Cys, C		Sulfur-containing	103.01
Glycine	Gly, G		Small	57.02
Proline	Pro, P		Uncharged	97.05

Addressing the Levinthal's paradox, further research suggested that there must be a sequence of events and local interactions that take place in order for a molecule to be guided to its stable state via a folding pathway (Levinthal, 1968). Today, protein folding is often represented by energy landscapes or folding funnels (Dill & Chan, 1997) (Figure 1.1). This visual demonstrates that the pathway from an unfolded chain to the folded native state progresses downhill towards low-energy and low-entropy states, where the native state is reached at the energy minimum (Dill *et al.*, 2012; Dill & Chan, 1997) Sampling the funnel is faster than sampling all of the hypothetical combinations of dihedral angles in a protein backbone.

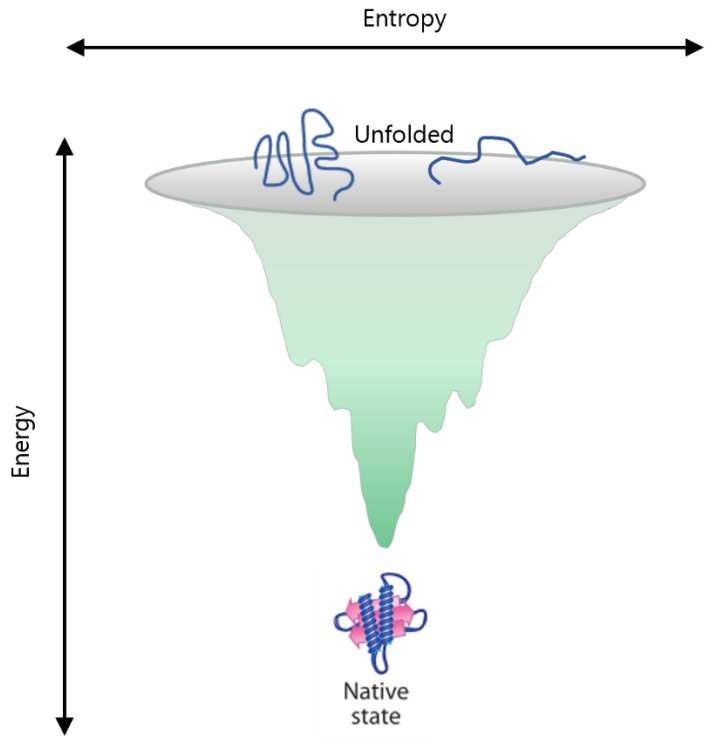


Figure 1.1 Idealised energy landscape of protein folding.

Internal free energy of the system is represented by the vertical axis, and conformation entropy in the polypeptide chain is represented across the width of the funnel, where the green area stands for the possible molecular conformations. Protein folding begins by sampling conformations in an energy landscape. As the number of intramolecular contacts increase, the internal free energy decreases leading to reduced conformational freedom. This conformation sampling 'funnels' down into the native state. Adapted from (Y. E. Kim et al., 2013).

During the progression of folding, there can be multiple routes through conformational space, sampling a wide range of partially folded structural intermediates (Jahn & Radford, 2005;

Vendruscolo *et al.*, 2003). These heterogeneous intermediates may be 'on-pathway', which form low energy kinetic traps (Brockwell & Radford, 2007; Jahn & Radford, 2008), adding roughness to the folding funnel but ultimately do not lead to aberrant folding behaviour (Figure 1.1, ripples at the edges of the funnel). The folding polypeptide chain can be driven into more native-like states through hydrophobic and electrostatic interactions, which contribute to a negative change in free energy (ΔG) for folding. Hydrogen bonds may contribute to the creation of local secondary structures, such as alpha helices and beta strands, which in turn may lead to the association of partially-folded intermediate structures with lower internal energy. Arrangements of these secondary structure elements may lead to the burial of hydrophobic regions into a core, creating tertiary structure which further stabilise the chain. Overall, the interactions the chain makes along its folding pathway usually directs the peptide into environments closer to its native state (Jahn & Radford, 2008). Also, importantly, some proteins are natively unfolded, where the lowest ΔG state under given conditions is disordered. These proteins are not considered as intermediates, partially folded or misfolded kinetically trapped states; they are natively intrinsically disordered, depending on solvent environment and the presence or absence of interaction partners.

In the context of protein production in a cell, a polypeptide chain is assembled by the ribosome, a molecular multimeric protein responsible for translating messenger RNA. As the chain of condensed amino acids – connected through peptide bonds – emerges from the ribosome, folding may begin to occur. One mechanism which has evolved to manage the risk of protein misfolding and work to assist proteins to fold efficiently, is the use of chaperones. These proteins stabilise or assist protein folding in a catalytic manner, by providing a 'safe', often hydrophobic environment in which the protein can rearrange itself, either passively or actively (requiring adenosine triphosphate), thereby reducing the population of proteins in partially unfolded states (Y. E. Kim *et al.*, 2013). Emerging nascent polypeptide chains can be co-translationally folded through the action of chaperones during the translation process – one example is the heat shock protein (Hsp) 70 which contribute to co-translational de novo folding of nascent polypeptide chains, and are estimated to assist the folding of 10-20% of all bacterial proteins (Mayer & Bukau, 2005).

1.1.1.1 Misfolding and aggregation of proteins

Due to the immense number of possible conformations a polypeptide chain can adopt, and the complexity of the protein folding funnel, it would be extremely unlikely for protein misfolding

to never occur. Therefore, the realistic folding funnel contains 'off-pathway' folding routes and the process of aggregation (depicted in Figure 1.2).

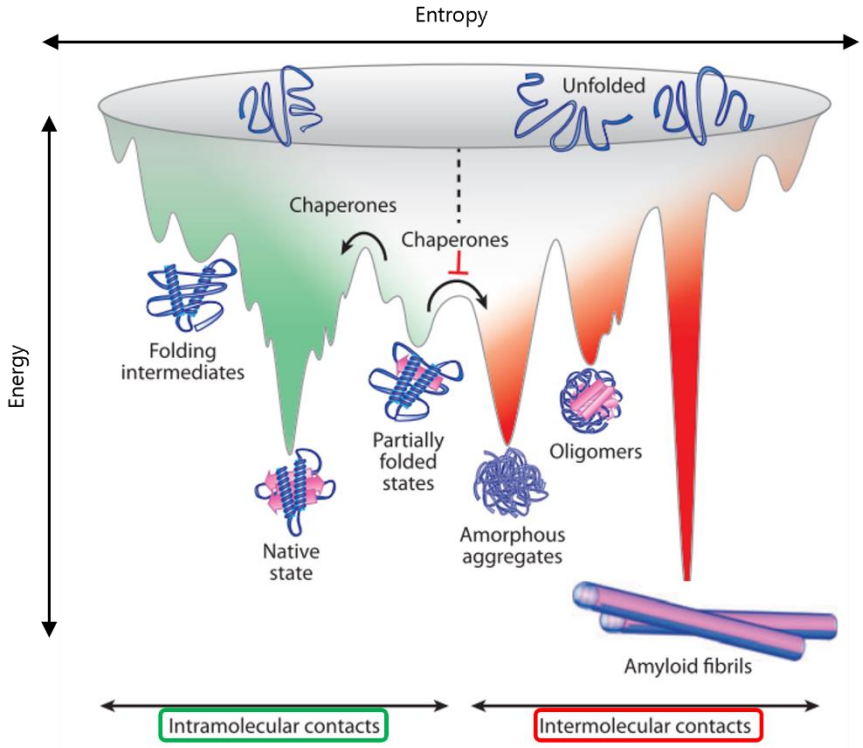


Figure 1.2 Realistic energy landscape of protein folding, misfolding and aggregation.

Proteins are inherently aggregation prone, and can become kinetically trapped into 'on-pathway' or 'off-pathway' intermediates. For example, exposure of aggregation-prone regions could promote the formation of amorphous aggregates. Higher order species such as amyloid fibrils can form as a consequence of oligomerisation from a nucleus. Figure adapted from (Y. E. Kim et al., 2013).

'Off-pathway' protein folding intermediates hinder the formation of the native state, forming misfolded states that slow the rate of folding, or even species that that can be associated with human disease (Brockwell & Radford, 2007). This can occur when the misfolded states exist in a more stable state than the native, or are kinetically trapped in a folding well, where significant energy-demanding reorganisation is required (Jahn & Radford, 2008) (Figure 1.2). As discussed above, chaperones can lower the energy barrier for re-folding of peptide chains into a more native-like fold. However, should it not be possible to recover the misfolded protein to a native state, a cell can employ protein degradation pathways to remove the misfolded protein. One example is by utilising the chaperone heat shock cognate (Hsc) 70 to target misfolded proteins

to the lysosome for degradation by lysosome (Ulrich Hartl *et al.*, 2011). Crucially, if purified proteins (such as those used in biopharmaceuticals) become misfolded and aggregated *in vitro*, the chaperone machinery is not intrinsically present and the species are not removed, which may lead to the propagation of aberrant species which may be highly undesirable.

The potential of a protein to become trapped in non-native folds depends on a variety of properties, including destabilising mutations in the amino acid sequence, or changes to cellular environments such as changes in temperature, pH, the presence or absence of ligands, hydrophobicity (such as in a membrane environment) or through strong electrostatic forces (such as in condensates) (Chi *et al.*, 2003). The influence of these destabilising conditions may also trigger the native protein to behave non-natively. Aberrant behaviour of the native state of proteins, or of misfolded, partially stable proteins, can trigger self-assembly in a process called aggregation. Aggregation refers to the self-association of protein monomers into assemblies other than the native structure (Ratanji *et al.*, 2014), induced by either conformational or chemical alterations (Esfandiary *et al.*, 2015). As native proteins are expressed in cells on the cusp of their solubility and stability, and must retain inherent flexibility for function approval (W. Wang & Roberts, 2018), they are inherently unstable in aqueous solution, and favourable intramolecular contacts within proteins can be replaced with intermolecular interactions. Indeed, some proteins can aggregate natively, forming functional amyloid fibres which are stable protein polymers (for example curli, formed by enteric bacteria in biofilm formation (Chiti & Dobson, 2006). Overall, proteins have the potential to aggregate through several pathways, both reversible and irreversible (Figure 1.3).

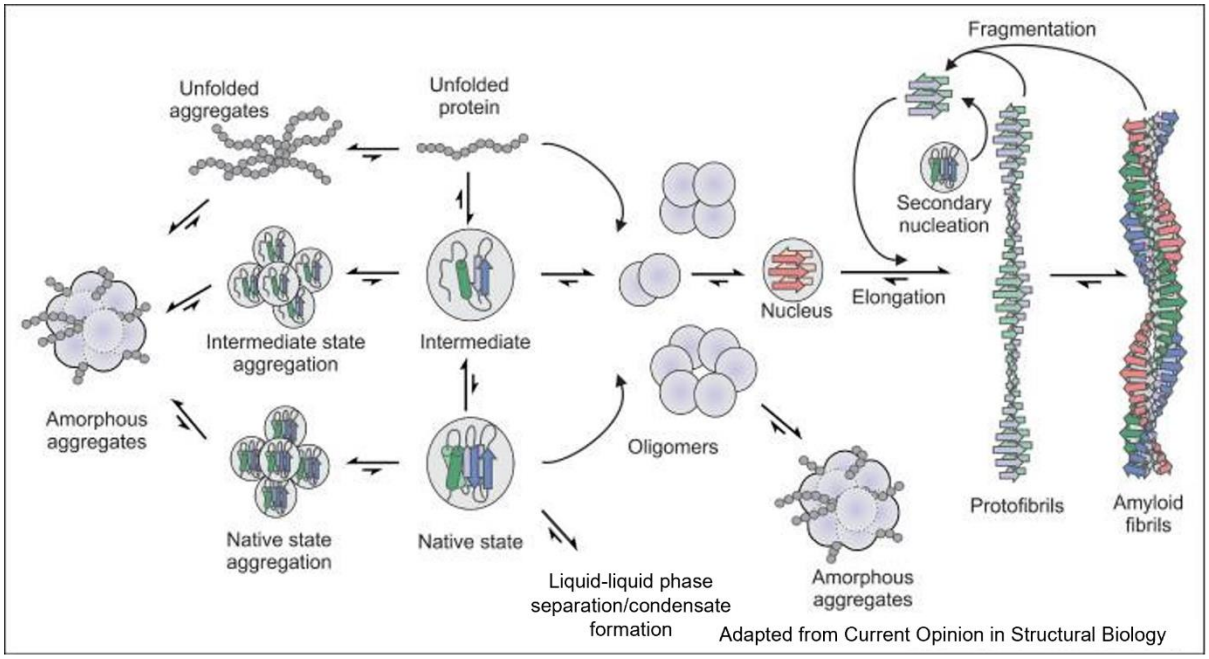


Figure 1.3 Schematic illustration of aggregation pathways, adapted from (Ebo, Guthertz, et al., 2020).

Aggregation precursors may be the native state of a protein, or ones that are partially- or fully-unfolded. Amorphous aggregation (left side of the figure) can occur through several aggregation-prone species activated through environmental stressors, which associate into larger species which are disordered. Conversely, amyloid formation often progresses through ordered arrangements of monomers into oligomeric nuclei. This nucleus provides a nucleation point for assembly into protofibrils and amyloid. Fibrils also can fragment and elongation can nucleate from these ends.

Aberrant behaviour of partially stable proteins can trigger the formation of several types of aggregates, amongst these there are amorphous and ordered aggregates (Figure 1.3). The nucleation of molecules to form higher-order species may be irreversible or reversible, and aggregates may be soluble or insoluble, depending on the protein involved and stimuli (Chi *et al.*, 2003). Ordered aggregate structures are found in misfolding diseases, including amyloid disorders (Chiti & Dobson, 2017), where the amyloid fibrils are irreversibly formed. Amyloids are characterised by their cross-β structure which makes them the most thermally-stable and low-energy aggregates known (Knowles *et al.*, 2014). Conversely, amorphous aggregates may form by unstructured organisation of proteins through hydrophobic patches or colloidal interactions (Brummitt *et al.*, 2011; Esfandiary *et al.*, 2015).

In most natively folded proteins, the majority of hydrophobic residues are buried from solvent as this is energetically favourable (Jahn & Radford, 2008). In some instances, hydrophobic residues can be solvent-exposed for the protein to function, for example there are high

incidences of aromatic tyrosines and tryptophans in the complementarity-determining regions (CDRs) of antibodies (Ausserwöger *et al.*, 2022; X. Wang *et al.*, 2010).

It is common for aggregation-prone regions (APRs) on proteins to contain hydrophobic motifs, and APRs are often protected from forming intramolecular interactions when they are buried in the hydrophobic core. APRs can become exposed due to environmental stresses such as chemical and mechanical changes to the protein's environment (Esfandiary *et al.*, 2015; Meric *et al.*, 2017), and these exposed regions can trigger an aggregation cascade, producing protein oligomers of varying size from dimers to higher-order aggregates (Figure 1.3).

Additionally, native monomers have been shown to reversibly self-associate through hydrophobic surface patches known as 'hotspots', or electrostatic colloidal interactions (C. L. Dobson *et al.*, 2016). These collections of proteins can be triggered to become irreversible at high concentration through the formation of covalent bonds such as disulfide bonds (Chi *et al.*, 2003).

Determining which aggregation mechanism is dominant for different protein sequences remains a challenge. It is important to develop techniques able to identify small transient fluctuations in a protein's conformation, to further understand the process of aggregation.

1.1.2 Biopharmaceuticals

The term 'biopharmaceutical' refers to a therapeutic wholly or partially synthesised by a living system and utilised in biotechnology, including recombinant antibodies, nucleic acid- and genetically engineered cell-based products (Walsh & Walsh, 2022). As this thesis focusses on therapeutic antibodies, their structure, development and manufacture will be covered in this section.

Monoclonal antibodies (mAbs) and antibody derivatives dominate the approved biopharmaceutics in the clinic, representing 53.5% of all approvals within the last four years (Walsh & Walsh, 2022). Therapeutic mAbs are a result of tremendous research into the adaptation of antibodies for specific purposes.

1.1.2.1 Antibody structure

Antibodies originate from the Immunoglobulin (Ig) superfamily. Of the five types of Ab in higher vertebrate species – IgM, IgA, IgD, IgG and IgE – IgG is the dominant isotype found in the body,

and therapeutic mAbs are typically IgGs (Buss *et al.*, 2012; Schroeder & Cavacini, 2010). IgGs consist of two identical heavy (approximately 50 kDa) and two identical light (approximately 25 kDa) chains, arranged in a characteristic 'Y' shape (Figure 1.4 a). The chains are folded into structural Ig domains that consist of two β-sheets arranged in an Ig-fold (each oval in Figure 1.4). The heavy chain is comprised of four Ig domains, whereas the light chain is comprised of two such domains. These heavy and light chains associate into the Y-shaped molecule, which is commonly divided into two regions: the crystallisable fragment (Fc) and two antigen binding fragments (Fab). In each Fab domain, there are two Ig domains from the heavy chain (V_H and CH_1) and two domains from the light chain (V_L and C_L) (Figure 1.4). Each variable fragment Fv (V_H and V_L, Figure 1.4 b) contains six CDR loops (three per variable domain) which are responsible for specific interactions of antibodies with their antigen. The diversity of the amino acid sequences in these loops gives rise to a vast diversity in antigen recognition. A variety of modalities can be used for drug development additionally to the mAb, including single domain antibodies (dAbs), single-chain variable fragments (scFvs), antibody-drug conjugates, Fc-fusions (Figure 1.4 b-e) and more, including bispecifics and Fab fragments (Kaplon *et al.*, 2022; Mullard, 2022).

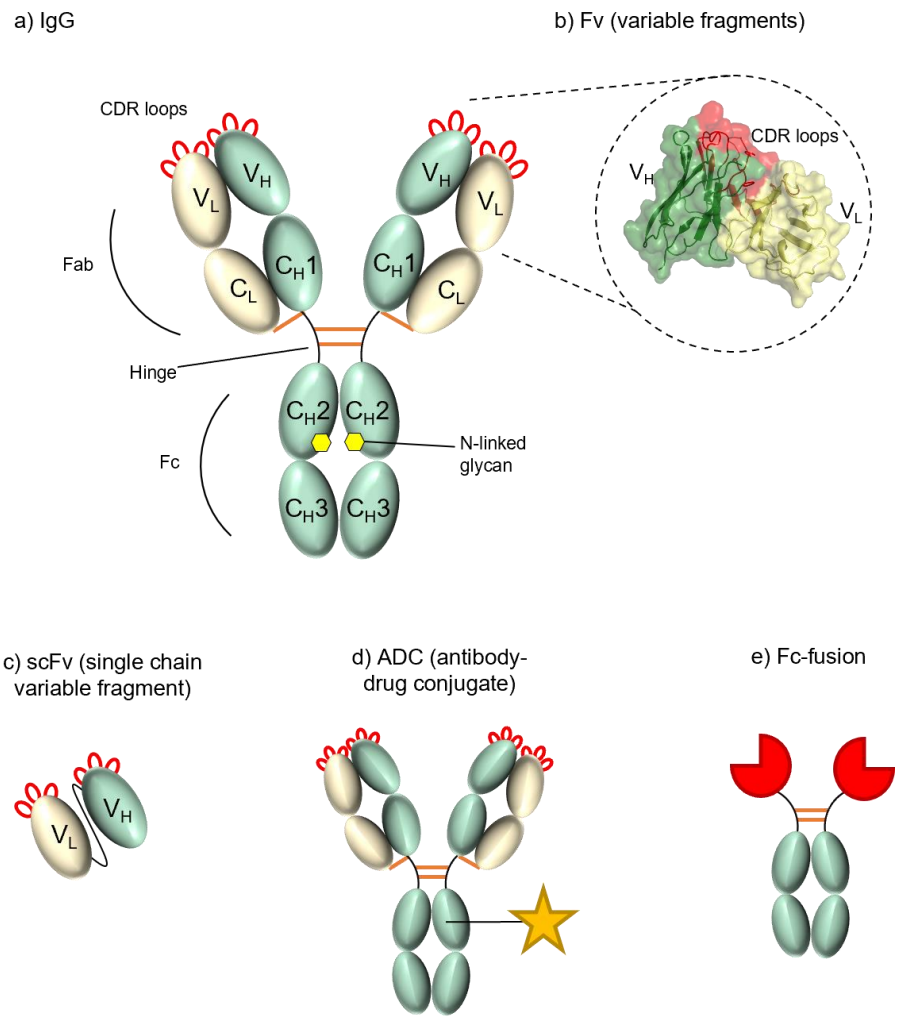


Figure 1.4 Schematic representation of a monoclonal antibody and some mAb-based therapeutic scaffolds.

a) IgG molecules are Y shaped and have a mass of approximately 150 kDa. Their heavy chains (green) and light chains (yellow) are connected by intermolecular disulfide bonds (orange). Their Fab and Fc (fragment crystallisable) sections are labelled along with the complementarity determining regions (CDRs) and the hinge (in IgGs, the heavy chains are connected by two disulfide bonds here). b) Variable fragments from the heavy chain (V_H, green) and light chain (V_L, yellow), with their CDRs in red. c) A single chain variable fragment (scFv) is comprised of the V_H and V_L domains connected by a flexible glycine-serine linker (GS repeats). d) Antibody-drug conjugate (ADC), an IgG scaffold with an added drug payload (star) often conjugated to cysteines engineered into the sequence. e) Fc-fusion protein, an Fc fused to two pharmaceutical molecules (red).

IgGs are held in their quaternary arrangement by disulfide bonds. Disulfide bonds increase the stability of protein folds (Anfinsen, 1973; Feige *et al.*, 2010). The disulfide bonding of IgGs varies between subclasses (IgG1-4) (H. Liu & May, 2012). The bonds are highly conserved within each subclass and are responsible for the distinct subclasses (IgG1-4), of which the classical disulfide bond networks were defined in the 1960s (H. Liu & May, 2012). Additionally, the disulfide bonds

within an IgG-like biomolecular structure vary in solvent exposure depending on the location between chains (inter-chain) or being buried within the anti-parallel beta-sheet Ig domain structures (intra-chain). Within an Ig fold, there is a disulfide bond connecting the two beta sheets (between strand B and strand F, (Feige *et al.*, 2010)) which is widely maintained in the Ig superfamily (Schroeder & Cavacini, 2010).

1.1.2.2 The history of the development of mAbs as a therapeutic

One of the historical challenges for industrial mAb production was the generation of a cell line that secretes a single type of antibody. Hybridoma technology was developed by Köhler and Milstein in 1975 (Köhler & Milstein, 1975) and has been used in academia and industry for mAb production for at least the last 40 years. In this technique, murine antibody-producing B cells are harvested from antigen-inoculated mice and fused with immortalised B cell myeloma cells, producing a hybridoma cell line that secretes a single type of antibody. This was a breakthrough in biopharmaceutical therapeutic production, resulting in the first US Food and Drug Administration-approved mAb called Orthoclone OKT3 (Muromonab-CD3) in 1986 for the treatment of transplant rejection (Leavy, 2010). However, significant disadvantages of the approach are that murine mAbs in humans are associated with allergic reactions, and being poor recruiters of effector functions in human systems (Buss *et al.*, 2012). Advances in protein engineering technology supported the creation of humanised recombinant mAbs (Winter & Harris, 1993): a chimera antibody is formed from human and non-human species to modify the sequence to be more human-like, and therefore be less immunogenic.

Subsequently, alternative approaches that avoid humanisation arose from the development of *in vitro* display technologies (Bradbury *et al.*, 2011), including phage and yeast display and optimisation of antibody engineering (Buss *et al.*, 2012). Display technologies give control over screening for highly specific and high-affinity antibodies (Buss *et al.*, 2012; Elgundi *et al.*, 2017). In 2018, Winter and Smith were awarded part of The Nobel Prize in Chemistry for their development of phage display, and using this technique humanised scFvs are created in bacteria. The display technologies are used as selection techniques for evolving proteins, and has been successful in generating antibodies currently in use in the clinic (Bradbury *et al.*, 2011).

The first approved human recombinant mAb developed using phage display was Humira (Adalimumab, from AbbVie), specific against human tumour necrosis factor. It has been demonstrated to have long-term safety as a treatment for rheumatoid arthiritis (Buss *et al.*, 2012) as well as plaque psoriasis, Crohn's disease and other disease indications (Gordon *et al.*,

2019). Humira is worth approximately US$21 billion in the global mAb market currently (Mullard, 2022).

Today, the market value of biopharmaceuticals continues to rise. In light of the COVID-19 pandemic, COVID-19 vaccines have had the largest impact upon the biopharmaceutical landscape in commercial, technological and medicinal terms. The long-time best-selling biopharmaceutical mAb Humira was displaced in 2021 by drug sales of the mRNA vaccines Comirnaty (from Pfizer/BioNTech) and Spikevax (from Moderna), cumulatively generating revenues of US$54.5 billion (Walsh & Walsh, 2022). The speed of Comirnaty's development from discovery to approval was 1.5 years, setting new precedents for the drug development pipeline which averages at 10.7 years (Mullard, 2022). Even despite these record-breaking mRNA-based therapies, mAbs still represented 80.2% of total protein-based global biopharmaceutical sales in 2021, up from 76.9% in 2020 (Walsh & Walsh, 2022). The global predicted market value of therapeutic antibodies for 2025 is US$300 billion (Lu *et al.*, 2020).

In 2022, the Food and Drug Association (FDA) approved 37 new drugs, where biologics accounted for 41% of the total approvals, and 30% of the total approvals were for antibody-based therapeutics, which includes mAbs, bispecifics (engineered antibodies that bind two different epitopes) and antibody-drug conjugates (ADCs, where small-molecule drugs are chemically linked to mAbs) (Mullard, 2023). The use of engineered antibodies to treat disease is expanding rapidly where the late-stage clinical pipeline of antibody therapeutics grew by over 30 % between 2020 and 2021 (Kaplon *et al.*, 2022). Antibodies in late-stage trials or approval provide therapy for various conditions, including and not limited to cancers, autoimmunity diseases and infectious diseases. In November 2021, 45 % of the primary indications for antibody therapeutics approved or in review were for oncology (Kaplon *et al.*, 2022). However, the variety of scaffolds for mAb and mAb-like molecules (Figure 1.4) take years to develop and require rigorous testing to obtain regulatory approval before market entry (Elgundi *et al.*, 2017).

1.1.2.3 Biopharmaceutical discovery pipeline

A candidate biopharmaceutical must pass through many stages of discovery and development to become a marketed product. As mentioned before, the drug development pipeline averages at 10.7 years (Mullard, 2022). Many stages of drug discovery are shared between small molecule therapies and biopharmaceuticals, such as the discovery, screening, optimisation and clinical trials, but for proteinaceous therapies such as mAbs, their structures tend to be more 'engineered' than constructed through screening processes. All drug compounds must be

carefully chosen to avoid immunogenic effects, and their development is highly rigorous, and can cost several billions of dollars (Dimasi *et al.*, 2016). The main stages of a biopharmaceutical pipeline can be summarised into discovery, development, and trial phases, before approval and large-scale manufacture (Figure 1.5).

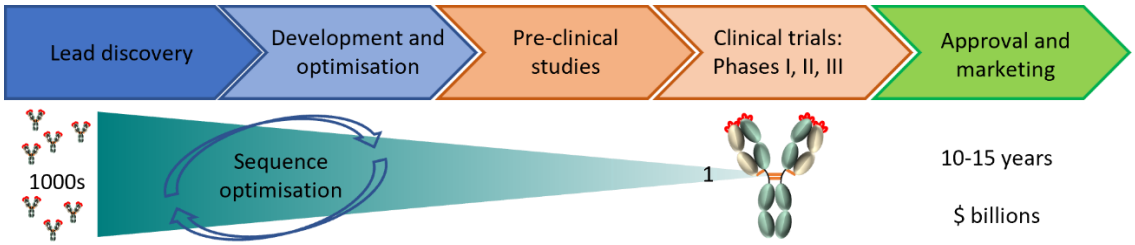


Figure 1.5 Overview of the drug discovery pipeline.

Once a target is identified, thousands of lead compounds are screened. If developability guidelines are met, they progress to optimisation and careful assessment of developability, iteratively improving the desired sequence. Less than 10 molecules are typically taken forward for pre-clinical trials to assess safety and pharmacokinetics before clinical trials in human patients.

In the early stages, the discovery of lead compounds to target specific pathways or diseases lay the foundations for drug development (Lead Discovery, Figure 1.5). Here, the mode of action and biological activity can be determined. At this early stage, it is important to assess the developability of the molecules – their feasibility to progress from discovery to development – by evaluating physiochemical properties (Bailly *et al.*, 2020). As there can be thousands of candidates, efficient and high-throughput screening of candidates is crucial, ranking their properties such as improved stability, solubility, binding affinities and reduced aggregation. *In silico* predictive tools can also be used to evaluate sequence liabilities and predict aggregation risk before expression (Bailly *et al.*, 2020), such as the CamSol solubility score (Sormanni *et al.*, 2017), Aggrescan (Conchillo-solé *et al.*, 2007; Kuriata *et al.*, 2019) and Spatial Aggregation Propensity (SAP) score (Chennamsetty *et al.*, 2009).

Once there are candidates which meet binding and activity requirements, they are characterised using biophysical methods. The developability of each candidate is closely assessed throughout the processes, screening for optimal stability, formulation, and low aggregation (Razinkov *et al.*, 2013). This is an iterative process to identify the prime sequence for the candidate. Eventually, less than 10 molecules would be chosen for pre-clinical trials in animals, before the commencement of human trials. This step is essential for assessing the pharmacokinetics and

efficacy of the candidate compound, and to assess the safety of administering the therapeutic (Walsh, 2003).

Eventually, successful compounds enter into clinical trials in humans. These trials are based around four phases (Walsh, 2003). Phase I involves safety testing in healthy volunteers, and successful molecules progress to phase II for safety and efficacy testing on a small number of patients, around 100-300. Phase III involves larger-scale trials with comparisons to alternatives or placebos. Phase IV (post-marketing surveillance) occurs to monitor the approved medicine over time, potentially assisting with extending the patent lifetime of the drug.

Once a molecule has been chosen for market, manufacturing must be scaled up. The manufacturing of mAbs is a 'platform process', where similar molecules are produced in a closed workflow. The process from a bioreactor to a formulated product can be divided into upstream and downstream processes. A typical mAb production process in found in Figure 1.6.

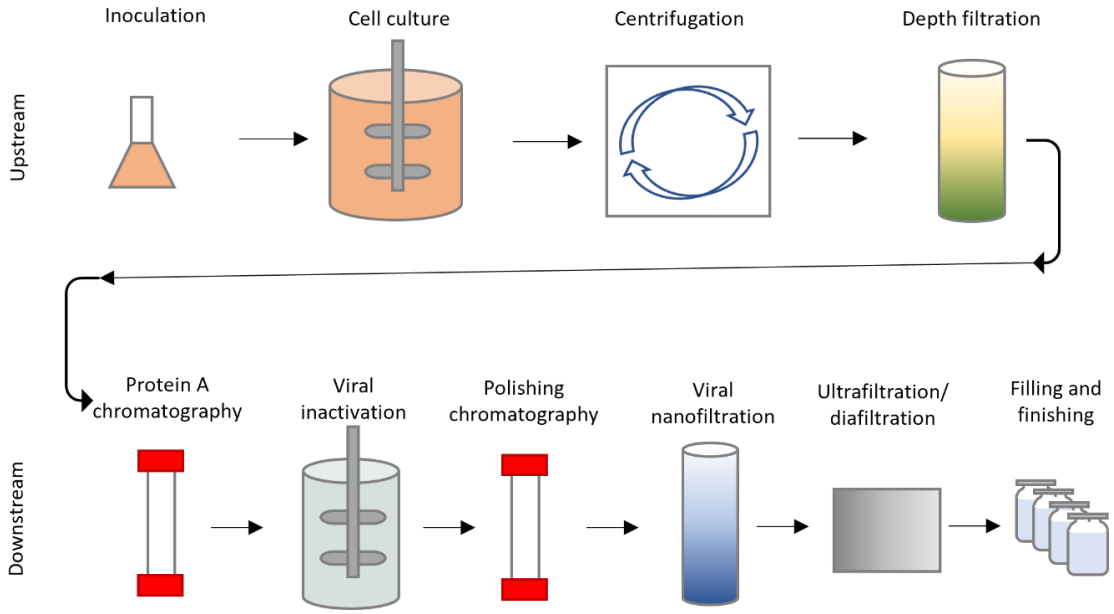


Figure 1.6 The biopharmaceutical manufacturing flowchart summarising the upstream and downstream processes.

Cell cultures are inoculated to overexpress the protein of interest. When the optimum cell density is reached, this is used to initiate large scale fermentation. Cells are separated from expressed protein through centrifugation and depth filtration. Clarified mAb is purified through protein A chromatography, and elution at low pH is beneficial for viral inactivation. Polishing chromatography steps such as ion exchange chromatography are performed to further remove any impurities. Ultrafiltration/diafiltration concentrates the product and buffer exchanges it into an formulation buffer for storage and administration. The product is often lyophilised (dried) in vials, or stored at high concentrations, for dilution before administration to patients. Figure adapted from (Faustino Jozala et al., 2016; Shukla & Thömmes, 2010).

Mammalian cells, such as Chinese hamster ovary (CHO) (J. Y. Kim *et al.*, 2012) or human embryonic kidney (HEK) (P. Thomas & Smart, 2005) cells are the most common expression systems for mAbs due to their cellular machinery which guides correct folding and post-translational modification, compared to prokaryotic cells (Tripathi & Shrivastava, 2019). The cells are transfected with an expression plasmid containing the protein of interest, and these are used for the inoculation of growth medium in a large production-scale bioreactor (Walsh, 2003) (Figure 1.6). Crude product can be harvested after the fermentation process using centrifugation and filtration, which removes cell debris before downstream manufacturing (Shukla & Thömmes, 2010).

The downstream production focuses on the removal of any contaminants so that the final purified product is able to satisfy the quality and purity requirements set by the approving body (such as the FDA) (Walsh, 2003). Protein-A affinity chromatography is used to capture mAbs; as the matrix consists of bound Protein-A from *Staphylococcus aureus* (42 kDa protein in the cell wall composed of five Fc binding domains and a cell-wall binding domain (Murphy & Kennedy, 2016)), the Fc regions of mAbs bind and impurities wash through (Murphy & Kennedy, 2016). Often >98% purity is obtained from this first purification step (Shukla & Thömmes, 2010). MAbs are eluted in low pH conditions as this weakens the interaction with Protein-A, which aids the inactivation of any virus particles not cleared in the wash steps (Shukla & Thömmes, 2010).

After this, polishing chromatography steps are employed to further remove impurities, such as ion-exchange chromatography (IEX) or hydrophobic interaction chromatography (HIC) (Marichal-Gallardo & Álvarez, 2012). Finally, the mAbs are concentrated and buffer exchanged into the final formulation buffer in the ultra-/dia-filtration stage (H. F. Liu *et al.*, 2010), before the sample is stored in vials, lyophilised, or pre-filled syringes before transportation (Rathore & Rajan, 2008) (Figure 1.6). Overall, the multi-step manufacturing process is highly complex, and throughout the whole process there are factors which adversely affect the molecules efficacy and safety, discussed in the next section.

1.1.2.4 Monoclonal antibody aggregation models

One of the major obstacles which all biopharmaceuticals face during the stages of the manufacturing pipeline is the inherent aggregation potential of proteins (Elgundi *et al.*, 2017). Protein stability is highly dependent on the environmental conditions that the protein is exposed to. In this thesis, the focus is on biopharmaceutical mAbs, and how potential changes to the environment of a mAb through all stages of expression, purification and formulation may 'stress'

the molecule, and induce aggregation through partially unfolded aggregation-prone species (Chi *et al.*, 2003).

Firstly, it is important to consider the current literature surrounding the pathways of antibody aggregation. Multiple research groups demonstrate that there is a variety of proposed steps leading up to amorphous aggregated species from monomer. Lumry and Eyring proposed a mechanistic model of protein aggregation, where native structures would reversibly form partially-unfolded species, which then would irreversibly aggregate (Y. Li & Roberts, 2009; Lumry & Eyeing, 1958). Classically, ordered aggregation pathways, such as those performed by amyloid proteins, proceed through a sigmoidal representation of aggregate formation over time, where ordered, insoluble amyloid aggregate structures form (Chiti & Dobson, 2017). Conversely, large globular mAbs can follow alternative aggregation kinetics; there is increasing evidence that many aggregation pathways initiate from activated, partially unfolded metastates (Brader *et al.*, 2015; Svilenov & Winter, 2019). Proteins undergo inherent fluctuations in conformation which may briefly expose aggregation-prone sequences or hotspots, and promote associations (W. Wang & Roberts, 2018). These non-native oligomers can be considered as the initial nucleus of aggregation on which several polymerisation steps may be followed. Following initial dimerisation, aggregate growth can occur through a variety of means: monomer addition, aggregate-aggregate cluster interactions, condensation or a combination of these. No universal model has yet been proposed that fully satisfies the prediction of mAb aggregation rates and route (W. Wang & Roberts, 2018). In fact, a two-step Finke-Watzky model (initial slow nucleation, followed by rapid autocatalytic surface growth) was shown to fit the aggregation kinetics of a commercially available mAb (bevacizumab) better than the Lumry-Eyring model (Oliva *et al.*, 2015). To understand more about the mechanisms of aggregation of mAbs, researchers have varied the kinetics of mAb aggregation through external factors, including temperature and pH changes, in order to propose more detailed mechanisms.

When considering time-dependent mAb aggregation over a range of concentrations and temperatures, kinetic models of aggregation have been proposed, where aggregation proceeds via dimer formation (Bunc *et al.*, 2022). In this example from Bunc *et al.*, a branched mechanism seemed to model the data best (obtained from Size Exclusion Chromatography (SEC) elution profiles); at low temperatures (< 40 ˚C) native monomers ‘N’ transitioned into a kinetic intermediate ‘I’, which further aggregates through forming a dimer of ‘NI’. Conversely, the authors speculate that at higher temperatures (≥ 40 ˚C) a separate high-temperature intermediate ‘D’ is formed from N, which further aggregates through D_2 dimers (Bunc *et al.*, 2022). Additionally, the authors link mAb thermodynamic stability (the overall mAb free energy

of denaturation ΔG_d), calculated from denaturation using chaotropic agent urea, to aggregation propensity. The authors show a significant correlation between the rate constant k_I (the formation of the intermediate 'I' in the LT pathway) and ΔG_d: when mAb stability is low ($\Delta G_d < 3$ kcal/mol) there is an increased amount of non-native protein concentration, and therefore is a lower threshold of switching aggregation pathways to that triggered through dimer formation from partially unfolded species (Bunc *et al.*, 2022).

Additionally, differences in pH were shown to affect the aggregate species formed when therapeutically-relevant humanised antibody fragment A33Fab was exposed to wide ranges of pH, temperature and ionic strength (Chakroun *et al.*, 2016). Aggregation was observed to occur from the native state, partially unfolded state, and unfolded state, dependent on the pH and ionic strength. At low pH there was a stronger destabilising effect of ionic strength, and aggregation proceeded through partially and fully unfolded mechanisms. Additionally, the debated relationship between conformational stability and aggregation rate was only significant for the Fab when incubated at temperatures close to the stability limit (the melting temperature) and aggregation from the unfolded state dominated (Chakroun *et al.*, 2016). Interestingly, the rank order of aggregation rates between low (4 ˚C), medium (23 ˚C) and high (45 ˚C) temperatures did not correlate well, which indicated that the commonly-used 'forced degradation' studies performed in industrial development of biopharmaceuticals may not be a representative screen for predicting aggregation behaviour at low temperatures (i.e. through storage).

Finally, researchers investigating protein aggregation under extensional flow (Willis *et al.* 2020) have developed a device which is being used to investigate the unfolding of mAbs in hydrodynamic stress, and are using this to begin to unpick the kinetics of mAb unfolding and aggregation. From the studies mentioned above, and the wealth of additional studies where aggregation mechanisms have been proposed for mAbs, there is clearly a need to further unpick the mechanisms of mAb aggregation, triggered by a wide range of influencing factors.

1.1.2.5 The undesirability of monoclonal antibody aggregation

Overall, mAb aggregation can be highly undesirable. A summary of potential stress inducers a mAb may be exposed through in industrial processes can be found in Figure 1.7.

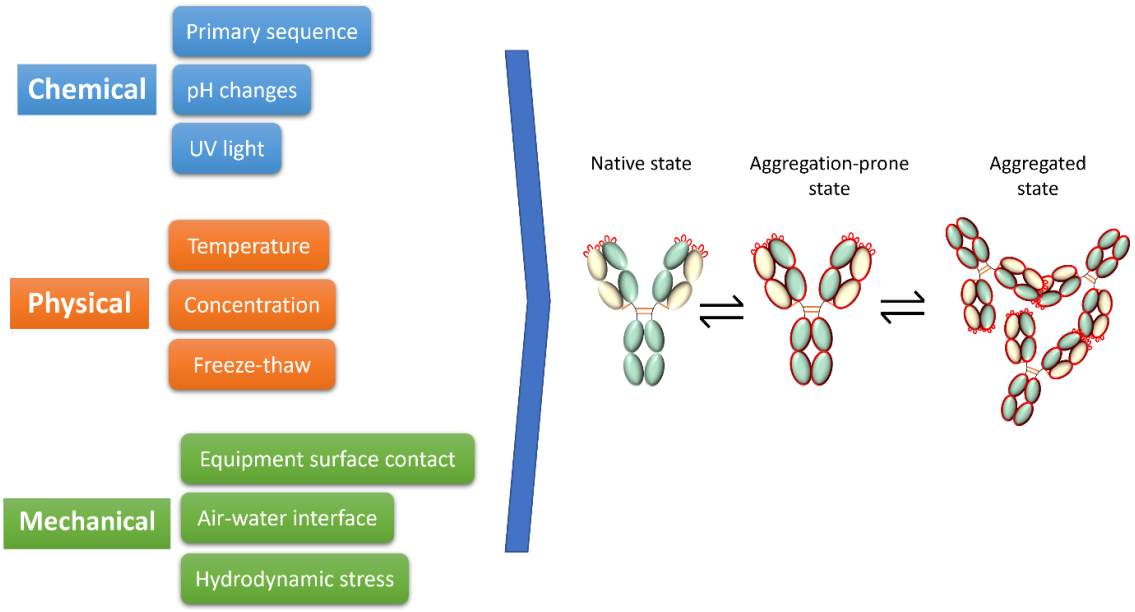


Figure 1.7 Potential inducers of mAb aggregation.

The figure lists a few examples of stresses that mAbs may encounter, both in the manufacturing pipeline and generally as a protein expressed on the cusp of its stability. These stresses are broadly chemical, physical and mechanical. The way that mAbs respond is dependent on the primary sequence. The stresses may trigger the perturbation of the native state to one that is aggregation-prone, which may lead to aggregation. Note the mechanism is not defined although there are a variety of mechanisms proposed in literature – these pathways are dependent on a range of conditions.

During the large-scale upstream and downstream manufacturing of mAbs (Figure 1.6), there are an array of different environments and stresses which can trigger unfolding and aggregation (Figure 1.7) (Esfandiary *et al.*, 2015; Rathore & Rajan, 2008). A biopharmaceutical mAb is likely expressed in HEK or CHO cells, and even at this early stage of development, some mAbs show signs of aggregation based on their primary sequence when expressed at low levels (1 – 10 mg) (C. L. Dobson *et al.*, 2016). It is apparent that identifying aggregation-prone candidates in these early stages of drug discovery will decrease the likelihood of these problems being detected later down the pipeline, to prevent the loss of time and money (Jain *et al.*, 2017).

Downstream, in the chromatography polishing and viral inactivation stages (Figure 1.6), exposure to pH and ionic strength changes can cause loss of yield of the mAbs through aggregation (Mazzer *et al.*, 2015; Shukla & Thömmes, 2010). Additionally, mechanical stress (such as stirring and shaking, pumping) can be imparted on molecules (Mahler *et al.*, 2009). Degradation has been shown to occur at interfaces between the drug product and solids (such as steel, glass) (Bee *et al.*, 2011) and gases (air) (Ghazvini *et al.*, 2016). The passage of mAbs through tubing, filling devices and pumps can be potential triggers of aggregation in the

manufacturing pipeline, at shear positions such as flow through tubing or connectors (Kalonia *et al.*, 2018). Exposure to stainless steel has been shown to induce mAb aggregation by exacerbating shear stress (friction between the solution and the material wall initiates horizonal planes of frictional force which can impart mechanical stress onto a molecule) (Bee *et al.*, 2010; Kalonia *et al.*, 2018). Aggregation triggered through shear stress at steel tubing interfaces is particularly undesirable due to these being present after the final filtration steps during fill/finish (Figure 1.6), meaning the aggregates would remain in the drug product. Additionally, passage of mAbs through tubing and pumps mean that the products are exposed to hydrodynamic stress throughout the manufacture. The effects of encountering hydrodynamic flow on proteins and biopharmaceuticals, and the stress this may cause, is further explored in the Section 1.1.4: Fluid flows and their effects on molecular structure.

Higher order association of protein structures is highly undesirable for the performance, purity and quality of product and threatens their approval (W. Wang & Roberts, 2018). The effect of aggregates on the immune system is not fully understood. It has been shown that aggregates can trigger an immunogenic reaction, varying from mild symptoms (such as a rash) to anaphylaxis and even death (Ratanji *et al.*, 2014; X. Wang *et al.*, 2010). Product aggregation would also prevent the precise dosage of the active ingredient. This highlights the importance of preventing aggregate formation in biopharmaceutical products.

One factor that is important to help prevent aggregation and to prolong shelf life is by formulating the product in favourable conditions (W. Wang & Roberts, 2018). Common formulation excipients include polysorbates (80 and 20) and free amino acids. Arginine (Table 1.1) is a commonly used excipient in mAb formulations as it has a protective effect; the aliphatic side chain is thought to interact with hydrophobic groups on the mAb, shielding it from hydrophobic protein-protein interactions; and the charged termini and guanidinium R-group may also interact electrostatically with other amino acid side chains on the mAb (Baynes *et al.*, 2005; N. A. Kim *et al.*, 2016). Arginine is also added to formulations as a way of decreasing viscosity of the solutions (Inoue *et al.*, 2014). MAbs are formulated in high concentrations (often >50 mg mL^{-1}) (Elgundi *et al.*, 2017) because the required doses are often 100s of milligrams, and the concentration must be high as it would be uncomfortable and slow for patients to have large volumes injected subcutaneously (Mitragotri *et al.*, 2014). In addition to the unpleasantness of the viscous solutions, the 'syringability' of the solution (the force required to eject) would be high, which could also deleteriously affect protein structure (see Section 1.1.4: Fluid flows and their effects on molecular structure) (Baek & Zydney, 2018). Subcutaneously delivered mAbs

are often provided in pre-filled syringes with silicone oil lubricated barrels (Bee *et al.*, 2011), and there is evidence to suggest the oil influences mAb aggregation and therefore immunogenicity effects in patients (Chisholm *et al.*, 2015).

Overall, understanding where the issues arise in manufacture is an important step towards characterising and eventually avoiding future problems in biopharmaceutical production. It is apparent that there is a strong need to predict, observe and understand the aggregation of these molecules, which will be the topic of the next section.

1.1.3 Techniques to study and overcome biopharmaceutical aggregation

Proteinaceous therapeutics need to be extensively characterised and monitored throughout and after their manufacture. This is important to report on their inherent properties, alongside structural and biological integrity, and to detect and monitor impurities induced by the process (Mahler *et al.*, 2009). However, one of the main challenges for studying aggregation in biopharmaceuticals is that there is no single analytical method which covers the entire range of aggregate sizes which could be generated. Several methods are often employed and compared to understand each product.

There are a variety of techniques used to generally study protein structure. Widely used techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) have, for many years, provided scientists with detailed insights into protein structural features (Alberts *et al.*, 2002). Recent advancements in high resolution methods such as cryo-electron microscopy (cryo-EM) are rapidly gaining popularity (Callaway, 2020) by providing the means to visualise the folded structure of proteins through microscope images. Even though these powerful techniques hold the ability to provide detailed structural understanding, these techniques have drawbacks based on protein size, or the complexity of protein preparation. The large datasets often need a lot of time and resources for thorough analysis which is not ideal for fast-paced biopharmaceutical research. For the quality control of a biopharmaceutical product, it is also required to characterise impurities which may make up under 1% of the total, which structural biology techniques may not detect.

Routinely, chromatography and light-scattering-based techniques are employed to characterise biopharmaceutical aggregation (Den Engelsman *et al.*, 2011). These techniques are lower resolution in comparison to NMR, crystallography and cryo-EM, and do not provide specific information on the organisation of side chains but rather valuable information about protein size, shape, surface properties and oligomeric interactions. A subset of these are summarised in

this section. Additionally, advancements in spectroscopic-based techniques such as mass spectrometry (MS) have increased its popularity in routine assessment of aggregation. MS provides crucial site specific information to support the analysis of protein structure, which will be elaborated on further in Introduction Part 2 – Mass Spectrometry in structural biology: understanding protein conformation.

1.1.3.1 In silico predictors of aggregation-prone proteins

The power of computational predictions can be applied to enhance protein engineering approaches, to aid the prediction of the sequences involved in aggregation (Ebo, Guthertz, *et al.*, 2020). *In silico* computational techniques are useful for predicting aggregation-prone regions and informing initial product design; or the rational design of improved biopharmaceuticals. There are a menagerie of protein aggregation prediction algorithms, including SAP (Chennamsetty *et al.*, 2009) and Aggrescan (Conchillo-solé *et al.*, 2007; Kuriata *et al.*, 2019; Zambrano *et al.*, 2015), which can be used for predictions of protein aggregation based on their sequence and how this relates to structure.

Aggrescan calculates aggregation propensity for each amino acid, based on an aggregation-propensity scale for natural amino acids derived from in vivo experiments, where green fluorescent protein (GFP) was used as a folding reporter for amyloid-beta 42 (Aβ42) mutants (Conchillo-solé *et al.*, 2007). In the initial calculation experiments, the Aβ42 mutants were fused to GFP and the in vivo fluorescent levels measured; if the substitution triggered a lower overall fluorescence, that amino acid was proposed to have caused increased aggregation of the Aβ42. From this screening process, all amino acids were provided with a score. Therefore, the output Aggrescan score of a protein of interest is calculated through a sliding window average calculation for a region of amino acids in the sequence, which is centered on every residue's alpha carbon in the sequence. This enables the putative identification of aggregation 'hot spots', or APRs. Advancements to the Aggrescan server lead to the development of Aggrescan 3D (Kuriata *et al.*, 2019; Zambrano *et al.*, 2015) which corrects the aggregation score of APRs by projecting onto a 3D protein structure, collecting structurally corrected aggregation values for each amino acid based on its structural context. The server also projects a 'dynamic mode' where protein flexibility is simulated, enhancing the predictions of APR exposure (Kuriata *et al.*, 2019). This feature can also suggest protein variants with optimised solubility.

It is well known that an input of many generic factors must be considered when predicting aggregation. Recently, a program called Therapeutic Antibody Profiler (TAP) (Raybould *et al.*,

2019) was developed, by analysing clinical stage therapeutics and picking out the key developability guidelines important for biopharmaceutical proteins (Jain *et al.*, 2017). This program raises 'flags' if the input molecule contains properties that are non-conforming to the developability guidelines. The factors calculated include CDR length, hydrophobicity patches in the CDR regions, positive and negative patches on protein surface, and the structural Fv charge (the net charge of V_H and V_L, Fv labelled in Figure 1.4).

1.1.3.2 Separation-based methods for understanding biopharmaceutical aggregation

It is highly important to be able to detect and categorise polydispersity of mAb solutions during the biopharmaceutical development process, in order to identify potentially problematic aggregate formation. There is no clear threshold of when a soluble aggregate is considered a 'particle', but generally anything greater than 100 µm in diameter (assuming a spherical shape) is classified as a visible particle (Roesch *et al.*, 2022), and these are commonly identified through visual inspection. Below 100 µm these particles are considered aggregates; protein oligomers are generally 10-100 nm, where sub-micrometer particles/nanometer aggregates range between 0.1-1 µm, and subvisible particles/micrometer aggregates are between 1-100 µm (Roesch *et al.*, 2022). A summary of applicable particle analysis techniques to these size ranges can be found in (Roesch *et al.*, 2022).

Soluble aggregates below 100 nm in size are commonly analysed using size-exclusion chromatography (SEC) (Mahler *et al.*, 2009), due to its reproducibility and the speed of analysis, ~15 minutes per run (Hong *et al.*, 2012). This technique separates proteins based on size (incorporating their molecular weight and volume) and shape, and these factors affect their permeation through porous matrices. The TSKgel column is the industry standard for SEC, and is a silica based column. SEC coupled to high performance liquid chromatography (HPLC) can be used to separate mixtures containing associated species with a size range of 5-1000 kDa. Because of this range, SEC is used in industry to separate mAb mixtures to quantify the percentage of soluble higher order species (HOS, aggregates) or lower order species (such as fragments or unpaired chains) (Telikepalli *et al.*, 2014). However, aggregation-prone mAbs have been demonstrated to associate with silica SEC matrices (C. L. Dobson *et al.*, 2016). Therefore, there are an arsenal of chromatography techniques for further understanding mAb surface biophysical properties.

Hydrophobic interaction chromatography (HIC) is a technique performed at high concentrations of salt in the running buffer, in order to enhance any hydrophobic interactions to the

hydrophobic matrix, and desorption is initiated by reducing the salt content (Murphy & Kennedy, 2016). It is a well-established technique for post-protein A polishing, but the high salt content required for mAb binding may trigger denaturation leading to aggregation (Murphy & Kennedy, 2016).

Standup monolayer adsorption chromatography (SMAC) has been used as a screening assay for mAb self-association and aggregation (Kohli & Geddie, 2017). The SMAC HPLC columns have a hydrophobic monolayer with terminal hydrophilic groups which mimic the exterior of a protein. Antibodies with colloidal instability are more prone to interact non-specifically to the column and elute later from the column (Kohli & Geddie, 2017).

Additionally, industry commonly use analytical ultra-centrifugation (AUC) to separate macromolecular species with different densities and to quantitate aggregates in solution, as an orthogonal method to SEC (Bou-Assaf *et al.*, 2022; Gabrielson *et al.*, 2010). In a sedimentation velocity experiment, samples are centrifuged, the separation and sedimentation of different molecular weight species in the sample are monitored in real time by absorbance or interference detectors, providing a size distribution (Bou-Assaf *et al.*, 2022). This process is not restricted by oligomeric size, and there is no potential for column interaction, unlike SEC. Samples can be analysed in their formulation buffer, and common practice is to have the sample concentration corresponding to an absorbance of 1.0 to ensure a good signal-to-noise ratio. However, this may require mAb concentration to be between 0.6-2.4 mg mL^{-1} depending on the pathlength of the instrument (Bou-Assaf *et al.*, 2022), and low concentrations should be avoided so that the ability to detect minor species is not impaired, meaning the technique is not best suited to early stage investigations of aggregation where one may be sample limited.

1.1.3.3 Light scattering methods for sizing particles

Sub-micrometer and sub-visible particles that are too small to be analysed through visual inspection, but too big to be separated from monomeric proteins and oligomers by SEC, are important to observe as they form the most concern in the manufacturing and regulating process (Ratanji *et al.*, 2014; W. Wang *et al.*, 2012). Light scattering techniques are used to detect, size and count aggregates in solution.

The most widely used light scattering method for characterising mAb candidates during early discovery stages is dynamic light scattering (DLS). This technique estimates the size distribution of a sample by measuring the scattering of light through the random motion of particles in a

suspension. The rate of diffusion of the scattered species is relative to the rate of decay of the fluctuations in the scattered light. The resulting correlogram is used to create size distribution histograms. DLS is often performed in a plate format allowing fast screening of different samples and conditions (Geng *et al.*, 2014). Connolly *et al.* demonstrated the power of DLS in improving mAb candidate selection and characterisation, where the authors measured the diffusion interaction parameter (a measure of pairwise intermolecular interactions) for 29 unique mAbs in several solution conditions and at concentrations between 1-20 mg mL^{-1} (Connolly *et al.*, 2012). These were compared with viscosity measurements at concentrations of up to 175 mg mL^{-1}, and strong correlations were found between the data, showing that these measurements can be made with low volumes of sample, for example using sample from small scale expression trials, negating the need to use resources to screen samples (Geng *et al.*, 2014). The authors used this study to demonstrate the power of the technique to identify concerns relating to viscosity at low concentrations to predict their viscosity at high concentrations before formulation and delivery (Connolly *et al.*, 2012).

Additionally, light scattering techniques have been used to generate thermally induced unfolding profiles to monitor mAb aggregation behaviour when stressed using changes in temperature and pH (Bhambhani *et al.*, 2012). The combination of static right angle light scattering, circular dichroism, intrinsic- and extrinsic- fluorescence measurements allowed the development of an assay to screen for suitable excipients (additives) for the mAb formulation (Bhambhani *et al.*, 2012). The investigation into how excipients may stabilise mAb products is highly useful for biopharmaceutical development.

1.1.3.4 Spectroscopic techniques for understanding unfolding

Spectroscopy is the interaction of electromagnetic radiation with matter. Fluorescence spectroscopy uses photons to excite electrons in molecules, followed by a relaxation step, which initiates the emission of photons at a longer wavelength. Fluorescence techniques are used to analyse the extent of biotherapeutic aggregation, often utilising the intrinsic fluorescence property of the aromatic amino acid tryptophan. Tryptophans are often monitored in intrinsic fluorescence measurements of mAbs, as excitation at 280 nm leads to emission at 285-450 nm (Chen & Barkley, 1998), and the changes to the excitation maxima have been linked to aggregation in mAbs (Joubert *et al.*, 2011). As tryptophans are exposed during unfolding, they emit photons of longer wavelengths than when buried in the hydrophobic core of mAb folds, because of their more polar environment. Similarly, extrinsic fluorophores added into the

chemical environment surrounding proteins can be used to report on their structure, such as Nile red (Demeule *et al.*, 2007), that has been shown to interact with hydrophobic patches on stressed IgGs. However, the formulation of biopharmaceuticals can interfere with the analysis by altering the fluorescence characteristics of the dyes used (Mahler *et al.*, 2009).

Differential scanning fluorimetry (DSF) is used in an industrial setting to detect aggregation in low concentrations (Semisotnov *et al.*, 1991). It is a fluorescence emission spectroscopy measurement that calculates an apparent melting temperature T_m from the midpoint of a fluorescence emission intensity versus temperature plot. Extrinsic dyes such as SYPRO orange can be used as reporters of unfolding as it binds to exposed hydrophobic regions (S. Shi *et al.*, 2013). More recently, advances in DSF technology have seen the development of machines which can detect intrinsic tryptophan fluorescence through epifluorescence of a very small sample volume in a plate format, such as the SUPR-DSF from Protein Stable.

A well-used approach for commenting on mAb colloidal self-interaction propensity is affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS). This technique uses gold nanoparticles coated with capture and non-capture antibodies (goat anti-human, and goat, non-specific, respectively) to immobilise mAbs in dilute solutions (~0.001-0.05 mg mL^{-1}) (Geng *et al.*, 2014). Monoclonal antibodies that have a strong propensity to associate will bind both the capture antibody and each other, bringing the nanoparticles closer together, which triggers a red-shift in the wavelength of the maximum absorbance of the mixture compared to that of the coated nanoparticles alone. This technique has been successfully implemented in the discrimination of mAbs with low and high solubility differing by single mutations in the CDRs (Sule *et al.*, 2013). Additionally, this technique has also been shown to correlate with HP-SEC retention time shifts of a panel of mAbs with increasing aggregation propensities (Ebo, Saunders, *et al.*, 2020).

1.1.3.5 Spectrometric techniques

Mass spectrometry can be used to monitor the stability of biopharmaceutical molecules at many stages throughout a manufacturing process. MS techniques report accurate masses of biopharmaceuticals, can identify aggregation or degradation, and determine sites of chemical modifications (Den Engelsman *et al.*, 2011; Jiskoot *et al.*, 2012). Examples of these techniques, their applications to structural biology, how they can be used to interrogate biopharmaceutical mAb aggregation, and their use in this thesis, are explored in Introduction Part 2 – Mass Spectrometry in structural biology: understanding protein conformation.

Overall, a powerful arsenal of structural biology techniques are being applied to one of the greatest challenges in the biopharmaceutical research field: to understand the mechanism and the principles behind protein folding, misfolding and aggregation (Dill *et al.*, 2008, 2012). There is a need for screening tools that are compatible with large numbers of samples, but cope with low sample volume and are easily reproducible. As there is not a 'one-size-fits-all' method for predicting the development of a successful biopharmaceutical product, there needs to be further research into better ways to understand structure and stability. Specifically how hydrodynamic stress affects biopharmaceutical structure and leads to aggregation is a controversial topic in the field, and forms the topic of the next section.

1.1.4 Fluid flows and their effects on molecular structure

During a protein's manufacture it can be exposed to a variety of different stresses, such as centrifugation, filtration, pumping and filling, as discussed in Sections 1.1.2: Biopharmaceuticals and 1.1.3: Techniques to study and overcome biopharmaceutical aggregation. Amongst the least well understood stresses to biopharmaceuticals is how mechanical stress, particularly hydrodynamic forces, can influence unfolding and aggregation. Proteins are subjected to a variety of flow forces in large-scale bioreactors, between containers, and throughout their manufacture to product packaging and transport. Understanding the impact of fluid forces on aggregation may ultimately improve our ability to engineer proteins with better structural stability, increasing the probability of the future development of successful therapeutics (Jiskoot *et al.*, 2012).

1.1.4.1 Shear and extensional flow

Two of the most common fluid flow mechanics encountered during biopharmaceutical manufacture are shear and extensional flows (Figure 1.8). In shear flow, proteins are exposed to shear stress when they encounter parallel layers of fluid flow moving at different velocities in the flow direction. These layers create a velocity gradient, where the velocity is lower toward the outer layers of the fluid closer to the edge of the vessel, meaning the gradient of velocity decreases perpendicularly to the flow direction. A parabolic curve of velocity gradient is formed because friction at the edges of the pipe wall cause the velocity of the central fluid to be higher than the peripheral fluid (C. R. Thomas & Geer, 2011). Thus, shear rate is maximal at the pipe

wall (Figure 1.8). Friction between the layers of solution and the material wall initiates horizonal planes of frictional force which can impart mechanical stress onto a molecule.

An extensional flow is created when a fluid is forced to accelerate through a constriction point, rapidly increasing the velocity of the fluid in the direction of flow. A protein passing through this constriction experiences a strain rate which could cause protein elongation or extension. The rates of shear or strain applied to molecules is directly related to the velocity of the fluid in the case of laminar flow (Figure 1.8) (C. R. Thomas & Geer, 2011).

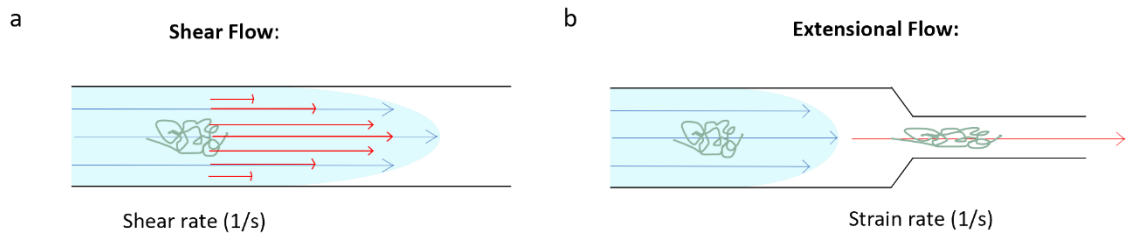


Figure 1.8 Schematic examples of shear flow and extensional flow in a laminar flow system.

a) Shear rate is maximal at the wall where there is the most friction, creating the parabolic profile of flow, in a laminar flow system. b) A constriction point causes a rapid increase in velocity in the direction of flow, in a laminar flow system.

Damage to the proteins through these hydrodynamic forces may be caused by a multitude of different effects to the protein. For example, disruption of tertiary structures; changes to secondary or tertiary structures through unfolding; promotion of aggregation to give insoluble aggregates; or a combination of above. Many devices have been designed to mimic flow forces encountered in industry, and have been used to subject elongated and globular proteins to flow forces to investigate unfolding and aggregation. A device used in this thesis will be explained in Section 1.1.4.4: Extensional Flow Device (EFD) to mimic aggregation under flow, which, through a computational fluid designed system, subjects proteins to hydrodynamic force (J. Dobson *et al.*, 2017; Willis *et al.*, 2018, 2020), and is used to investigate the effects on biopharmaceuticals.

1.1.4.2 The effects of flow on model proteins

A protein that undergoes functionally-relevant changes to its structure by hydrodynamic forces is von Willebrand Factor (vWF). vWF is an elongated multimeric glycoprotein, and monomers of vWF can join end-on-end into structures of up to 15 µm in length (Bergal *et al.*, 2022).

Hydrodynamic forces in the bloodstream are known to regulate vWF's molecular mechanisms of tension-dependent binding to proteinaceous partners such as collagen, for immobilising damaged blood vessels and proteins involved in platelet recruitment by the unfolding of the platelet binding site when the force is sufficient (Aponte-Santamaría *et al.*, 2015). Recruitment of platelets to the protein mesh formed by vWF rapidly increases repair to the affected blood vessels.

Most globular proteins form insoluble aggregates under shear and extensional flow stress. When lysozyme, bovine serum albumin (BSA) and alcohol dehydrogenase (ADH) were exposed to these stresses in a four-roll mill apparatus (mimicking finish and fill steps in biopharmaceutical manufacture), Simon *et al.* found that extensional flows triggered more aggregation than shear flows (Simon *et al.*, 2011). The larger proteins BSA and ADH were more susceptible to flow effects than lysozyme and formed larger visible aggregates and a greater amount.

1.1.4.3 The effects of flow on biopharmaceuticals

Many bioprocessing steps subject biopharmaceuticals to mechanical stress (Figure 1.6 and Figure 1.7), and may trigger unfolding of proteins in solution (J. Dobson *et al.*, 2017; Willis *et al.*, 2018, 2020). Despite many research efforts over the last few decades there is still no clear consensus to the magnitude and type of hydrodynamic stress needed to induce mAb aggregation (Bee *et al.*, 2009; Brückl *et al.*, 2016; Duerkop *et al.*, 2018; Grigolato & Arosio, 2020).

Devices mimicking flow forces executed by biopharmaceutical manufacturing procedures provide an insight into the behaviour of mAbs under flow. Brückl *et al.* subjected a mAb to a shear rate of 3840 s^{-1} in laminar flow conditions and used circular dichroism to make inline biophysical measurements (Brückl *et al.*, 2016). The results show that shear was not able to alter the mAb fold under the experimental conditions. The authors suggest interactions at liquid-air interfaces during biopharmaceutical processing steps could trigger aggregation from the introduction of bubbles in the device. Similarly, Bee *et al.* subjected a concentrated mAb solution (150 mg mL^{-1}) to hydrodynamic stress using high shear rates in excess of those expected during normal processing operations (>2.5 x 10^5 s^{-1}), but observed no aggregation, suggesting that air-liquid interfaces, contamination by particles, or pump cavitation stress could be more important causes of aggregation than shear (Bee, Stevenson, *et al.*, 2010). These examples, however, do not comment on the effect of extensional force on mAbs.

1.1.4.4 Extensional Flow Device (EFD) to mimic aggregation under flow

In the University of Leeds, colleagues from the School of Mechanical Engineering and the School of Cellular and Molecular Biology designed and built a flow device to subject proteins to defined shear- and extensional- flow fields (J. Dobson *et al.*, 2017). This device, referred to as the Extensional Flow Device (EFD), produces laminar, non-turbulent flow in two syringes connected by a capillary (Figure 1.9). Molecules entering the capillary are subjected to strain at an extensional flow region at the point of constriction, and also high shear rates along the length of the capillary, as illustrated by Computational Fluid Dynamics modelling (J. Dobson *et al.*, 2017) (Figure 1.9).

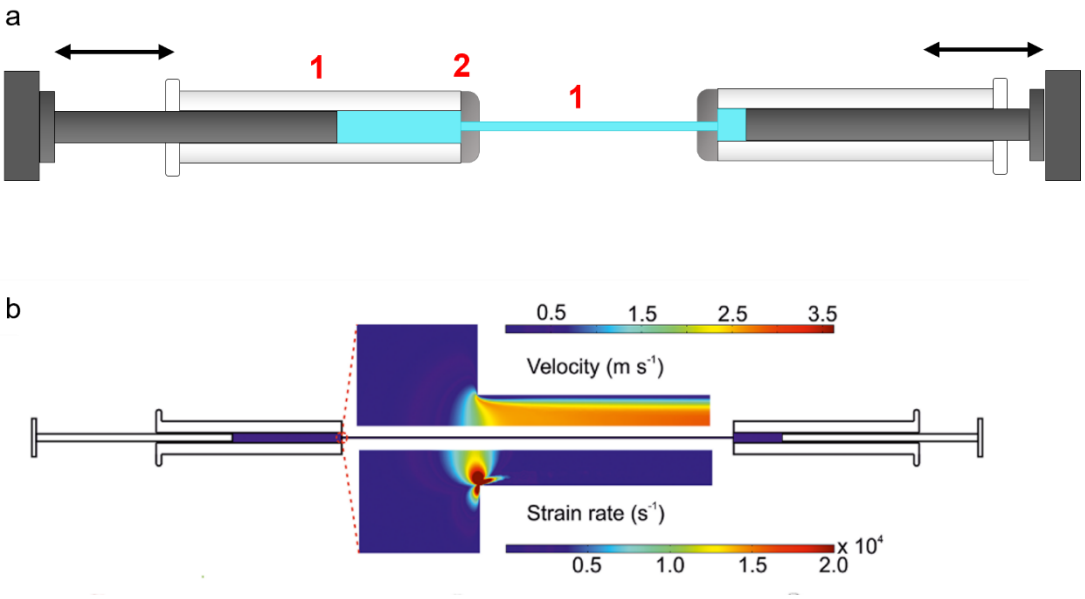


Figure 1.9 Schematic representation of the extensional flow device (EFD) used in this thesis.

a) Two syringes are connected by a capillary, and a stepper motor is used to push the baffles horizontally (black arrows), shuttling protein solution (turquoise) between the two syringe barrels. The proteins in the solution experience shear forces (1) in the syringe barrels and the capillary, and extensional force (2) as the constriction point between the syringe barrel and capillary. b) Computational fluid dynamic model (J. Dobson *et al.*, 2017) *of the fluid velocity (upper) and strain rate (lower). Fluid velocity is slow and laminar in the syringe. The contraction point triggers a rapid increase in fluid velocity and the proteins experience high strain at this position. b) taken from* (Willis *et al.*, 2018).

Using model globular proteins, the EFD has been shown to trigger aggregation at a magnitude dependent on the velocity and concentration of sample. The device induced aggregation of BSA by exposing new hydrophobic surfaces with greater self-affinity (J. Dobson *et al.*, 2017). The EFD has also been used for studying mAbs and their response to extensional flow. Willis *et al.*

compared three IgG1 molecules for their propensity to aggregate under flow conditions (Willis, 2018; Willis *et al.*, 2018). The aggregation induced in the extensional flow device was measured by analysing the percentage of protein aggregate in pellet compared to protein remaining in solution after centrifugation. Two out of three of these mAbs studied form the basis of this thesis and will be explained more thoroughly in Section 1.3: Basis of the study.

In 2020, Willis *et al.* demonstrated the power of the EFD as a uniquely sensitive developability tool, by comparing EFD-induced aggregation data from 33 IgG1 mAbs to aggregation data obtained previously from 12 other biophysical assays (Willis *et al.*, 2020). The authors could create a hierarchical clustering family tree to represent the EFD as a unique biophysical assay compared to those in the reference dataset in (Jain *et al.*, 2017). The power of the technique to trigger aggregation is apparent, and the use of this device to probe mAb unfolding will be explored in this thesis (see Section 1.4: The aims of the thesis, after Introduction Part 2 – Mass Spectrometry in structural biology: understanding protein conformation).

Chapter 1: Introduction Part 2

Mass Spectrometry in structural biology: understanding protein conformation

---

1.2 Part 2 – Mass Spectrometry in structural biology: understanding protein conformation

High resolution structures of proteins and complexes have been determined using X-ray crystallography and NMR spectroscopy for many years. As discussed in Part 1, there is no perfect technique to capture protein native-state, aggregation-prone-state, and oligomeric-state conformation, but rather a wide breadth of techniques should be used in parallel to gather important information to inform the output as a whole. Likewise, the limitations of each technique need to be considered so that other techniques which are complementary can be applied to the example at hand.

Mass Spectrometry is an analytical technique that determines masses through measuring the mass to charge ratio (*m/z*) of gas-phase ions. There are a diverse range of mass analysers and detector systems to date, all built on the foundations laid out by Thomson and Aston's work on elemental isotopes in 1913 (Thomson, 1913). For a sample to be analysed through MS, it must be introduced into the gas phase through ionisation (or in the case of MALDI, sample is desorbed into the gas phase via matrix molecules, and then ionised), and entered into a vacuum system for separation and detection. Resulting data is analysed and mass information is gleaned. This process is summarised in Figure 1.10.

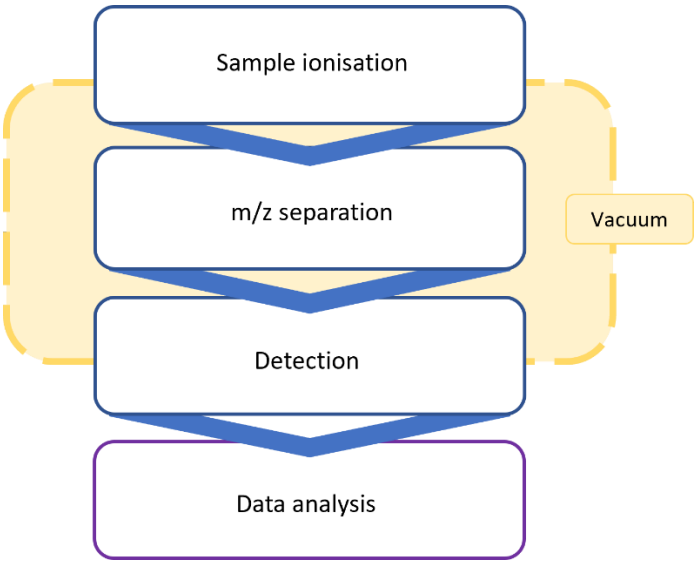


Figure 1.10 Overview of the main stages in a mass spectrometer.

Samples are ionised and introduced into the gas phase (e.g. in ESI, ions are introduced into the gas phase and then vacuum; in MALDI the ions are made in vacuum). The ions are accelerated before separation by m/z in mass analyser(s). The ions are detected, and then the data is analysed.

This thesis focuses on the use of Mass Spectrometry as part of the structural biology technological toolbox to further investigate protein structure. A complete review of the ionisation methods, and the different types of mass analysers and detectors, is out of the scope of this thesis, however the main concepts behind commonly used mass spectrometers will be outlined. Additionally, this part of the introduction will describe some common uses for MS to understand protein structure, conformation, unfolding and aggregation, with a focus on mAbs.

1.2.1 Ionisation

Ions can be generated through 'hard' or 'soft' ionisation techniques. Generally, for studying biomolecules such as proteins or peptides, soft ionisation techniques are used (Fenn *et al.*, 1989), as the process imparts lower amounts of energy onto a sample compared to hard ionisation techniques, such as electron ionisation (Bleakney, 1929). Lower energy ionisation has a lower risk of fragmenting peptide bonds, and better retains any non-covalent interactions, especially important for Native Mass Spectrometry techniques. Typically large non-volatile biomolecules are placed in volatile buffers to aid ionisation, and adopt charges such as $[M+nH]^{n+}$, where M is the molecular weight and n is an integer (greater than or equal to 1).

Electrospray ionisation (ESI) is a soft ionisation technique commonly used in protein analysis. ESI adds charges to biomolecules at atmospheric pressure by passing a high potential through a sample-filled capillary with an opening at the end. An electrode is placed at the front of the mass spectrometer entrance, charged oppositely to the ions formed, to ensure ions are attracted to the front of the system (the ions are positive when working in positive mode, most common for protein analysis, so the entrance to the MS would be negative). As the voltage is applied on the metal-coated glass capillary, a Taylor cone of liquid is formed at the front of the sample-containing volatile solvent, generating a cone shape where an aerosol of charged droplets detatch (Taylor, 1964) (Figure 1.11). The highly-charged fine spray droplets decrease in size due to desolvation. In standard ESI, desolvation can be helped with a stream of nebulising gas such as nitrogen at the front of the instrument, and a drying gas perpendicular to the Taylor cone, which is not used in nano-ESI. The droplets gradually reduce in size as they are desolvated, until Coulombic repulsion from charges on the droplet surface overcomes the force of surface tension – the Rayleigh limit (Rayleigh, 1882). When this point is reached, droplet fission occurs and droplets get smaller, and this process repeats until charge is transferred onto the sample. The gas-phase ions are attracted to the oppositely-charged source (Figure 1.11).

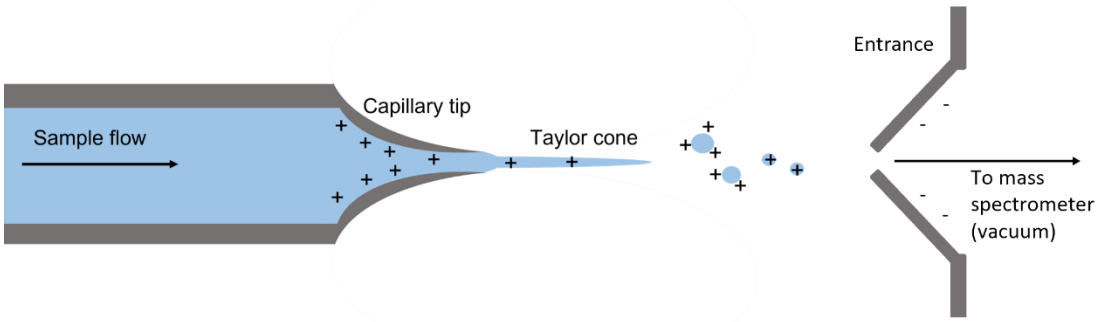


Figure 1.11 Overview of an electrospray ionisation source and the formation of a Taylor cone.

Voltage applied over a metal-coated capillary containing the analyte in a volatile buffer causes the solution to form a Taylor cone as it is sprayed from the capillary. Charged droplets containing the analyte are ejected, until ions enter the gas phase.

There are several models in the field to try to explain the process behind the transfer of H+/positive charge to ions. These include the charged residue model (CRM) (Dole *et al.*, 1968), the ion evaporation model (IEM) (Iribarne & Thomson, 1976), the chain ejection model (CEM) (Konermann *et al.*, 2013) and, recently suggested bead ejection mechanism (Khristenko *et al.*, 2023). The CRM suggests that continued desolvation and Coulombic fission eventually leads to completely desolvated ions, and is proposed to be what occurs for large globular samples. In comparison, the IEM suggests that ions are ejected from the droplet surface once the radii of the droplets have been sufficiently reduced, and is the most likely mechanism for small (in)organic ions (Konermann *et al.*, 2013). Thirdly, the CEM could be the process for the ionisation of unfolded proteins, where the exposed hydrophobic regions on the unfolded chain cause them to rapidly reach the droplet surface and are ejected starting with the terminus (Konermann *et al.*, 2013). Finally the bead ejection mechanism is proposed to be a hybrid of CRM and CEM, where globular domains tethered by disordered linkers are ejected a domain at a time from a parent droplet (Khristenko *et al.*, 2023).

ESI-MS has improved with the development of nano-ESI (nESI). This method creates smaller initial droplets through the use of smaller capillaries (<5 µm orifices), increasing the efficiency of desolvation. The flow rates for in-line nESI are much lower than standard ESI (around 20-100 nL min^{-1} (M. Wilm & Mann, 1996) compared to conventional ESI flow rates of 1-20 µL min^{-1} (M. S. Wilm & Mann, 1994), which means less sample is consumed. Static nESI is a variant of nESI where no pumping of the liquid is required and flow is initiated by capillary action. Additionally,

with nESI there is no need for high temperatures or a nebulising gas. Overall, this technique is one of the most widely used ionisation sources in biological MS, and has been applied to study a wide range of protein sizes, from hundreds of Da up to several-MDa complexes (Loo, 2000; Snijder, 2013).

1.2.2 *m/z* separation using mass analysers

After introduction into the vacuum of the system, the ions must be separated according to their *m/z*. The vacuum provides a low pressure environment which is essential to prevent unwanted ion-molecule collisions, improving transmission and resolution of the ions. There are a variety of different mass analysers, including quadrupole, time of flight (ToF), ion trap and orbitrap. There are benefits and limitations to each technique which dictate the preferred technique for particular uses. These analysers are normally compared for characteristics such as resolving power, sensitivity, accuracy, scan speed (the rate at which mass spectra can be acquired) and mass range limit (the *m/z* range over which an analyser can measure ions).

Mass resolution can be defined as the ability of the analyser to separate two peaks with a close *m/z* value. There are two general definitions for 'resolution', both defined by the equation below (Equation 1.1). When there is a single peak, resolution can be defined as the width of the peak on the *m/z* scale at its half-height (*Δm/z*), called the full width at half maximum (FWHM) (Hoffmann & Stroobant, 2007). Additionally, where there are two peaks, these are said to be resolved when the valley between adjacent peaks of the same intensity is below a defined percentile threshold (different for each mass analyser).

Equation 1.1 Definition of resolution.

Resolution (R) is equal to the mass to charge ratio of the ion divided by the change in mass to charge ratio, either the width of the peak at half height (FWHM) or the distance between two peaks with a defined % overlap.

$$R = \frac{m/z}{\Delta m/z}$$

The resolving power of an instrument refers to its ability to distinguish two ions of similar mass with equal intensity.

The mass accuracy of an analyser is typically measured in parts per million (ppm). This is defined in the equation below (Equation 1.2). It is a measure of how close the calculated mass is to the theoretical mass of the molecule.

Equation 1.2 Definition of mass accuracy.

The divergence of the observed mass from the theoretical mass gives a value for the accuracy of the observed mass in parts per million.

$$Mass\ accuracy = \left(\frac{Theoretical\ mass - Observed\ mass}{Theoretical\ mass}\right) x\ 10^6$$

1.2.2.1 Quadrupole analysers

A quadrupole mass analyser consists of four parallel metal rods arranged in a bundle, where the opposite two rods are in pairs where an opposite polarity direct current (DC) is applied to each pair. A radio frequency (RF) alternating current (AC) is also applied to the paired rods, with a maximum amplitude which is greater than the applied DC voltage. Ions enter into the space between the four rods and are driven through the quadrupole in a spiral trajectory, driven by dynamic potential changes between the rods. The waveforms of AC current through adjacent rods are maintained at 180 ˚ out of phase which causes repulsion and attraction of ions to maintain their path through the analyser.

The different planes of motion created by the potential across the rods can create high pass and low pass *m/z* filters which remove ions that have unstable trajectories in the analyser (Hoffmann & Stroobant, 2007). The stability of trajectories of ions mainly depend on the AC change frequency and the magnitude of both the AC and DC voltages. This tuneability means that the quadrupole can be used as a scanning mass analyser to sequentially allow ions of different *m/z* to the detector by ramping the AC and DC voltages. Additionally, the quadrupole provides fast scan speeds, and can cope with continuous ion infusion which is beneficial for coupling with ESI. However, quadrupoles have inherently low resolving power (Hoffmann & Stroobant, 2007) and are limited in the *m/z* range they can detect, which is typically capped at 4000 *m/z* (Collings & Douglas, 1997; Hoffmann & Stroobant, 2007).

Often a quadrupole is used as an ion guide rather than an analyser, by removing the DC potential in RF-only mode. This allows the passage of a wide range of *m/z* ions simultaneously and this is ideal for coupling to other mass analyser systems, such as TOFs, orbitraps and linear ion traps.

Mass spectrometers used in this thesis with quadrupole analysers are the Orbitrap Fusion Lumos Tribrid, Orbitrap Exploris 240, Velos Linear Trap Quadrupole (LTQ) Orbitrap and Q-Exactive UHMR (see Section 2.2: Methods).

1.2.2.2 Linear ion trapping analysers

Linear ion traps (LITs) are similar to quadrupole mass analysers, but have three distinct segments, where an RF field is applied to the centre section and the two end sections are controlled with DC potentials. This allows the repulsion of ions from the ends back and forth axially, confining the ions to the centre of the device in three dimensions. This allows the analyser to select or scan through *m/z* ions, and eject ions sequentially, to generate a spectrum.

LITs can be used as the sole mass analysers in an MS as their scan speed and achievable resolution are similar to quadrupoles (Douglas *et al.*, 2005). In some commercial instruments, LITs are also used as storage devices and collision cells before other mass analysers. The LIT can be filled with collision gas to excite ions and induce fragmentation.

Mass spectrometers used in this thesis with LITs are Orbitrap Fusion Lumos Tribrid and the Velos Linear Trap Quadrupole (LTQ) Orbitrap (see Section 2.2: Methods).

1.2.2.3 Time of Flight analysers

Time of Flight analysers are a pulsed analysis technique. The amount of time it takes ions to pass from an acceleration stage through the flight tube (a field free region) and to the detector, is proportional to the square root of an ion's *m/z* value. Because of this, higher *m/z* ions take longer to traverse the flight tube. ToF analysers inherently have a theoretically limitless detectable mass range, making this analyser ideally suited to study large biomolecules.

However, because a pusher voltage is used when ions enter the analyser to provide kinetic energy needed to accelerate the ions, a difference in the distribution of ions in the pusher region can mean ions have an uneven exposure to the acceleration voltage. This difference in voltage applied results in different arrival times for the ions, which can lower the resolution. Methods employed to equalise flight times include the use of the use of orthogonal acceleration ToF analysers (Guilhaus *et al.*, 2000) and/or reflectron ToF analysers (Mamyrin *et al.*, 1973).

1.2.2.4 Orbitrap analysers

An orbitrap mass analyser is an orbital ion trap device. It consists of a spindle-shaped electrode inside of a trapping unit, where a current applied to the central electrode allows for the stable orbit of ions. Ions are injected into the orbitrap in packets, and they begin to oscillate around the central spindle due to the applied electrical field creating a pull towards the centre of the spindle. This packet of ions rapidly goes out of phase such that ions with more initial kinetic energy orbit with a higher frequency than those with lower initial kinetic energy. The frequency of oscillation along the central axis is inversely proportional to their *m/z* values. The different oscillation frequencies are then used as a read-out for generating a mass spectrum. Orbit occurs without the need for an applied field on the outer electrodes (which are held at virtual ground potential (Makarov, 2000)), and the outer electrodes are used for image current detection (see Section 1.2.3: Detectors).

The scanning speeds for orbitrap analysers are slower than quadrupole, ion trap, and ToF mass analysers, in part due to the requirement for the packet of ions to find their oscillation frequency before mass analysis for reliable measurements. The frequency of the oscillations are typically maintained around 10 Hz (Olsen *et al.*, 2009) but can be decreased if a higher resolution is required. Their resolution also decreases as a function of $(m/z)^{0.5}$ (Olsen *et al.*, 2009) which is a disadvantage compared to quadrupoles, where the resolution is largely independent of *m/z*. However, orbitraps are inherently high resolution, and advancements in technology has allowed orbitraps to achieve up to 1,000,000 resolution (FWHM) at 200 *m/z* (Schmidt *et al.*, 2018). They are often coupled to continuous ionisation sources such as curved linear ion traps for ion accumulation and packed control, which then end up being the limiting factor in the detectable mass range limit of the instrument (Hoffmann & Stroobant, 2007). Recent advancements have enabled orbitraps to study protein complexes of up to several MDa (Snijder, 2013).

Mass spectrometers used in this thesis with orbitrap analysers are the Orbitrap Fusion Lumos Tribrid, Orbitrap Exploris 240, Velos Linear Trap Quadrupole (LTQ) Orbitrap and Q-Exactive UHMR (see Section 2.2: Methods).

1.2.3 Detectors

The purpose of a detector is to generate a signal from the separated ions, which are interpreted by a computer to generate the mass spectrum output (Bracewell, 1986). This is the most common detector for ToF instruments is an electron multiplier. Here, electrons collide with a dynode plate, which creates an electron cascade to amplify the signal in the detector (Hoffmann & Stroobant, 2007). It has a fast response time which is important in ToF instruments for recording precise time measurements. However the path of the ions are disrupted when they strike the detector.

In orbitrap mass analysers, where the outer electrodes are used as a detector (image current detection), the path of an ion (the oscillation frequency) in the mass analyser is proportional to its *m/z* and so this path must be undisturbed for accurate *m/z* measurements. Image current detection is a non-disruptive method of ion detection where ions moving between conductive metal surfaces induce a current. Overlapping image current signals from multiple different *m/z* ions in the trap are processed through Fourier transform algorithms to generate a mass spectrum. The outer electrodes in an orbitrap serve as the detector surfaces. During the ion separation, there is a delay between the introduction of the ions to the orbit of the electrodes, and this time allows for ions with differing *m/z* to radially de-phase, which reduces any interfering signals from in-phase orbiting (Makarov, 2000).

1.2.4 Analysis of MS data

For positively charged ions, the following formula is used to calculate mass (Equation 1.3). This is dependent on knowing the charge state of the ion (n).

Equation 1.3 Determination of mass from the mass spectrum.

Mass (for positive ions) is equal to the mass to charge multiplied by the charge state (n), where the mass of the additional protons is subtracted. Mass of proton = 1.0072 Da.

$$Mass = \frac{m}{z}\, n \, - \, 1.0072n$$

For the analysis of biomolecules ionised by ESI, the charge state can be determined in two ways using the distribution of different charge states. Where the resolution is sufficient to resolve

isotope peaks, the biggest contribution usually is from ^{13}C isotopes, and so knowing the difference between these peaks will be 1 Da, the charge state would be equal to the reciprocal of the difference in *m/z* of two peaks side by side in an isotope distribution (Equation 1.4). For larger species, the ^{13}C is less easily resolved, so the charge state is determined from the adjacent peaks in a charge state distribution, described in Equation 1.5, as n must be an integer and adjacent peaks are n+1 (lower *m/z*) or n-1 (higher *m/z*). This distribution shows that the adjacent peaks have the same mass and ionised from the same molecule, but a different number of acquired protons. This calculation gives the average mass rather than the monoisotopic mass.

Equation 1.4 Charge state determined from isotopic resolution.

$$n = (\Delta m/z)^{-1}$$

Equation 1.5 Charge state determined by charge state distribution.

$m/z_{(1)}$ and $m/z_{(2)}$ are adjacent peaks in a charge state distribution such that $m/z_{(1)}$ is of higher value on the m/z axis and $m/z_{(2)}$ has a lower value on the m/z scale.

$$n = \frac{m/z_{(2)} - 1.0072}{m/z_{(1)} - m/z_{(2)}}$$

1.2.5 Liquid chromatography – mass spectrometry (LC-MS)

Liquid chromatography (LC) before MS is highly useful for separating complex mixtures of samples before the ionisation process. In LC, an aqueous sample is passed through an analytical column, and the sample is resolved by its physical properties (such as size and charge) depending on the method of LC (summarised in Section 1.1.3.2: Separation-based methods for understanding biopharmaceutical aggregation). Reversed-phase LC (RP-LC) is the most commonly used LC technique to separate biological proteins or peptides. This technique separates sample due to polarity (Molnár & Horváth, 1976) using a column coated with alkyl chains; a C18 column has 18 carbon length alkyl chains commonly used for peptide analyses, and typically for proteins a C8 with shorter length chains is used. The hydrophobic, non-polar chains cause retention of hydrophobic, non-polar sample, which will elute later than sample that is more polar and hydrophilic. The elution of sample is performed by gradually increasing the

concentration of organic (less polar) solvent (acetonitrile, MeCN) in the buffer (Molnár & Horváth, 1976).

Technological advancements in MS instrumentation have allowed the development of systems which link to MS by direct ionisation. These include high-performance LC (HPLC) and ultra-performance LC (UPLC), which offer improved resolution than standard LC, a lower sample volume and higher pressures, along with flow rates compatible with the nESI source.

1.2.6 Tandem mass spectrometry (MS/MS)

Establishing an LC-MS technology suitable for investigating a sample mixture is important. Multiple mass analysers in a MS can be used in tandem, first for fragmentation and then analysis, to begin to interrogate sequences from the mass to charge data. When two mass analysers are used in succession for fragmentation followed by analysis, the process is referred to as MS/MS or tandem MS. A commonly used MS/MS method for biomolecules is the product ion scan (Figure 1.12), where ions (proteins or peptides) are selected in a mass analyser, collide with gas in a collision cell, then scanned in another mass analyser to generate a product ion spectrum. The output spectrum is often used for determining the initial sequence of the selected precursor through sequencing. In 'top-down' MS/MS, proteins are introduced into the mass spectrometer intact and fragmented within the fragmentation cell; 'bottom-up' MS/MS involves proteolytic digestion followed by the introduction of peptides into the MS1 region (further discussed in Section 1.2.6.1: Peptide sequencing and bottom-up proteomics).

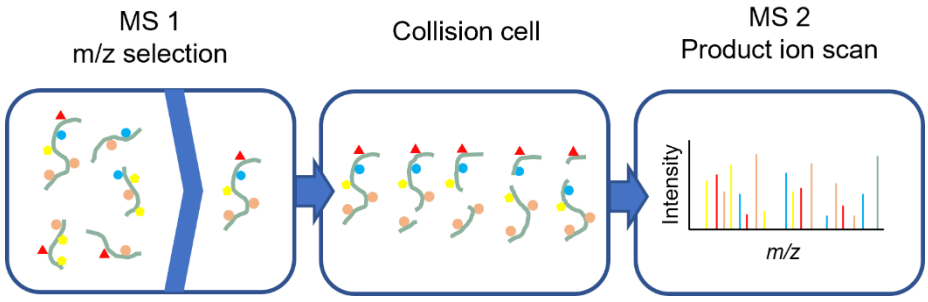


Figure 1.12 The workflow of a product ion scan.

Ions are selected in the first mass analyser (usually a quadrupole). The ions are fragmented in a collision cell and then passed into a second mass analyser which separates the product ions.

The most common fragmentation method used in sequencing biomolecules is collision induced dissociation (CID) (Jennings, 1968). A collision gas (usually argon, nitrogen or helium) in the collision cell collides with the ions, 'heating' them by transferring kinetic energy into internal energy. This energy is distributed through the ion which breaks the kinetically weakest bonds first, usually non-covalent interactions, then weak covalent ones. CID of proteins and peptides results in the breakage of the peptide bond to yield b- and y- ions, the N-terminal and the C-terminal fragments respectively (Figure 1.13). Other fragmentation methods may favour fragmentation of different bonds along the amino acid chain, generating a- and x- or c- and z-ions. CID of each individual peptide typically is only carried out once. As the masses of the full amino acid sequence, the resulting product ions, and each amino acid is known, the data can be analysed and a peptide sequence can be built up from the resulting peaks (bottom-up proteomics). When the initial sample introduced into the MS from LC is peptides, tandem MS/MS using CID can generate an understanding of peptide sequence; sequence tags (partial sequences of many/most peptides) are generated, and compared with the masses of theoretical fragments in typical database searches. This allows the identification of any modifications (and their positions on the peptide). Applications are discussed in Section 1.2.6.1: Peptide sequencing and bottom-up proteomics.

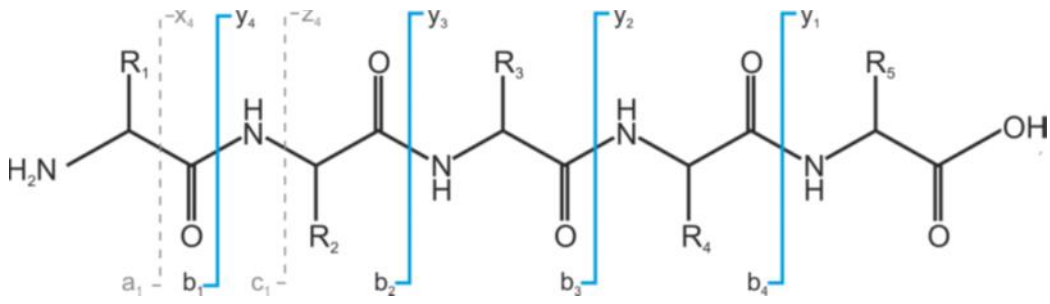


Figure 1.13 Schematic of a pentapeptide where the patterns of fragmentation possible are labelled.

The b- and y- ions are labelled in blue, where the b- ions are the N-terminal fragments and the y- ions are the C-terminal fragments. Figure adapted from (Roepstorff & Fohlman, 1984).

Alternative fragmentation methods include higher energy C-trap dissociation (HCD; where collisions with buffer gas induce fragmentation), and electron capture dissociation and electron transfer dissociation (ECD and ETD respectively; the use of low energy electrons to fragment). ECD and ETD are useful to get additional fragmentation data and for applications where non-covalent interactions need to be maintained.

1.2.6.1 Peptide sequencing and bottom-up proteomics

Standard 'bottom-up' MS/MS proteomics approaches involve proteolytic digestion before LC separation, followed by the introduction of peptides into the MS1 region, fragmentation and then sequencing. This process is summarised in Figure 1.14.

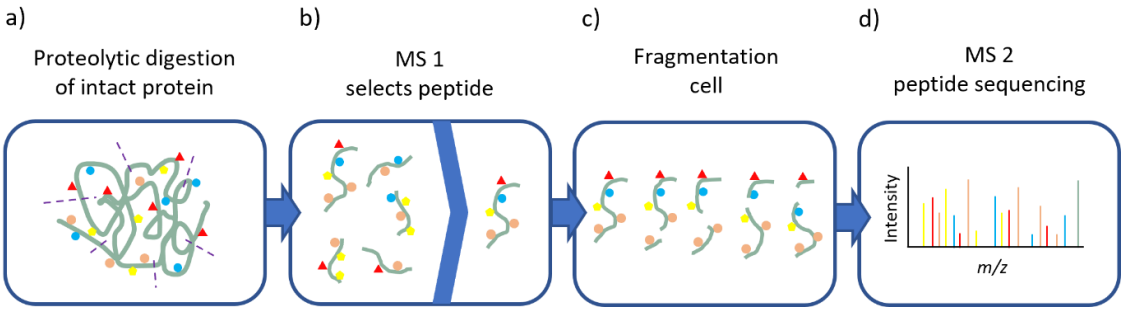


Figure 1.14 Proteolytic digestion before product ion scan MS/MS.

a) Sample is digested in solution into peptides which are separated by LC before ESI. b) In the MS, ions are selected in the first mass analyser (usually a quadrupole). c) The ions are fragmented in a collision cell and then d) passed into a second mass analyser which separates the product ions for sequencing.

Firstly, before digestion, proteins containing disulfide bonds are treated with a reducing agent and then an alkylating agent, which reduce the bond into free cysteines and then cap the free cysteines, respectively. This modification (carbamidomethylation) results in a mass increase of 57.02 Da and is presumed to be a fixed modification of cysteine residues in the tryptic digest protocol. The resulting capped amino acid does not reform the disulfide bond, to prevent reformation of the native state of the protein or another aberrant intra- or intermolecular disulfide bond at the protein or peptide level. Often, if a disulfide linkage is buried in the hydrophobic core of a folded protein, a protein can be exposed to heat to increase the internal energy of the system aiding unfolding, at a temperature at which the enzyme chosen for the digestion procedure remains active. Alternatively a sample can be exposed to chaotropic conditions before reduction (such as guanidine hydrochloride) which acts to stabilise a more unfolded and open structure, increasing the efficiency of reduction.

A reduced and alkylated sample is then digested with an enzyme to generate fragments of primary sequence (proteolytic peptides such as trypsin, in the case of this thesis, Figure 1.14 a). The resulting peptides are separated by LC in-line with the mass spectrometer, where elution peaks correspond to particular peptides. Then 'bottom-up' MS/MS is used to determine the

amino acid sequence of each peptide, and any covalent modifications they may carry, which is typically performed by product ion scan and collision induced dissociation (CID) (Figure 1.14 b-d).

1.2.6.2 LC-MS/MS data acquisition

As discussed above, each LC elution peak corresponds to peptides that are in turn ionised for MS/MS. Peptides with specific *m/z* are ideally selected one after another in the quadrupole (or ion trap) mass analyser in the mass spectrometer, and are fragmented, resulting in multiple fragment ions with different masses and also some unfragmented parent peptides. There needs to be a fast acquisition of the intact precursor and fragment ion spectra for each ion, which can be achieved through rapid automated switching between high (fragmentation) and low (precursor acquisition) energies in the collision cell. There are generally two acquisition modes: data dependent, and data independent, acquisition.

Data-dependent acquisition (DDA) modes first acquire a full *m/z* intact spectra scan of a short LC elution period by deactivating fragmentation. The user will have chosen a number of the most intense *m/z* ions (TopN) from this precursor ion scan (MS1, see Figure 1.12) to be selected by the first mass analyser for fragmentation, and spectral analysis by the second mass analyser. This cycle repeats, taking full scans and then subsequent top ion fragmentation spectra, until the LC elution finishes. A dynamic exclusion parameter can be introduced so that the same *m/z* value is not repeatedly selected in a particular time frame, to utilise the time for other less-intense ions to be chosen for fragmentation. The DDA cycles must be kept short in order to not miss any closely-eluting ions, but this needs to be balanced with a broad enough time needed to generate a thorough analysis of the eluting ions. The ability to pre-select *m/z* ions before fragmentation is popular as it makes the analysis simpler.

Data-independent acquisition (DIA) does not use any pre-selection of *m/z* ions before rapid switching between low and high energy modes for intact and fragmentation scans, respectively. This results in quick acquisition of MS1 and MS2 spectra for eluting ions, but these are complex mixtures of multiple ion precursors, making the assignments more challenging. MS^E can be employed to simplify the analysis process through matching precursor ions to their fragment ions by retention time analysis, or alternatively by ion mobility correlation analysis.

1.2.7 Understanding conformation and detecting aggregation using Mass Spectrometry

A wide selection of techniques have been employed for protein structural analysis, briefly discussed in Section 1.1.3: Techniques to study and overcome biopharmaceutical aggregation. Mass Spectrometry has proven to be a crucial addition to the method toolbox. Specifically in the study of protein aggregation, there are key analytical features in many MS techniques which provide valuable insights (Pukala, 2023). MS is sensitive, and only requires small sample amounts. Often, samples require less meticulous preparation than X-ray crystallography and NMR, and the timescales of analysis are often quick in comparison. ESI (discussed in Section 1.2.1: Ionisation) can be gentle enough to preserve noncovalent binding interactions and stoichiometry. Additionally, non-covalent and covalent labels can be utilised to footprint protein structure. The plethora of techniques used to investigate protein conformation are summarised in Figure 1.15. This section will focus on the benefits of using a selection of these MS techniques as biophysical characterisation tools for understanding mAb structure and aggregation.

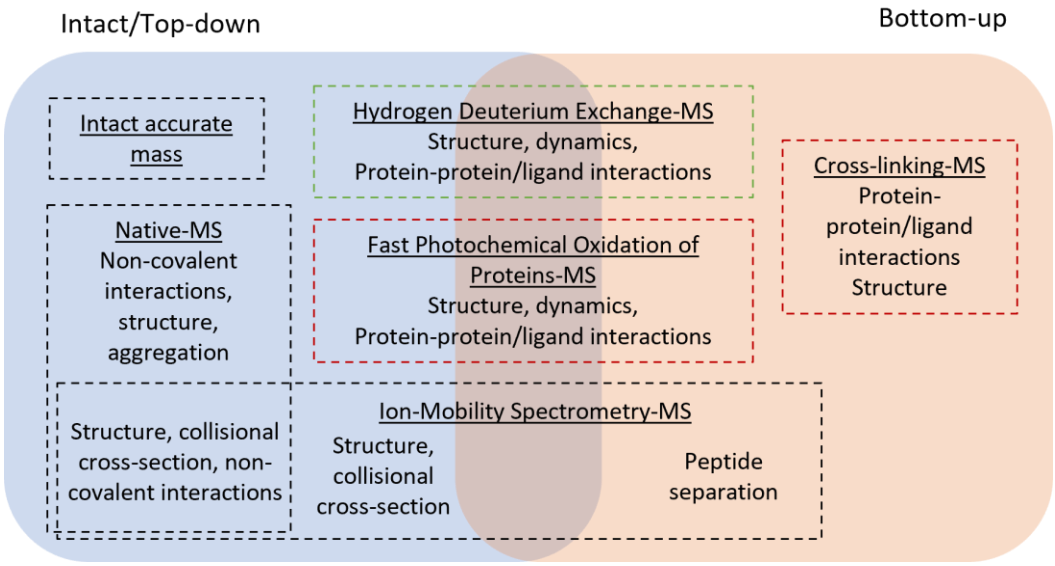


Figure 1.15 Schematic showing the different applications of structural MS methods for the analysis of protein primary to tertiary structure.

Broadly these techniques split into two categories, intact (the polypeptide chain is not disrupted) or proteomic (enzymatic digestion and/or MS fragmentation is used to generate proteins from the polypeptide chain). Black dash = no labelling; green dash = reversible labelling; red dash = irreversible labelling. Adapted from (H. Zhang et al., 2014).

1.2.7.1 Probing gas-phase conformation

1.2.7.1.1 Native Mass Spectrometry

Native ESI-MS is a 'top-level' technique, where native molecule structure is preserved as natively as possible, including non-covalent interactions within the native fold, to study biomolecular structure and stoichiometry (Figure 1.15). Experimental optimisation can be extensive, including using volatile and non-denaturing solvent conditions prior to ionisation, and optimising instrument parameters such as the use of trapping after the source and using higher pressure ion guides to improve collisional-cooling of ions, important for improved ion transmission (Chernushevich & Thomson, 2004). Native MS is becoming essential in biopharmaceutical development and control (Allison *et al.*, 2020), therefore pushing the requirements for improved standardisation of the analysis process. It is used for characterising recombinant proteins and analysing mAb structure and interactions (Deslignière *et al.*, 2021, 2022; H. Zhang *et al.*, 2014). Native MS has been used to provide information on higher-order species (HOS) present in mAb formulations (Terral *et al.*, 2016), where a direct readout of dimer/HOS intensity is visible. Additionally, success has been made in coupling online Size-Exclusion Chromatography (see Section 1.1.3.2: Separation-based methods for understanding biopharmaceutical aggregation) to native MS, where species of different size (mainly higher-order aggregates) triggered due to high temperature changes and storage exposure are isolated for simultaneous identification and quantification (Haberger *et al.*, 2016; Terral *et al.*, 2016). The SEC process is also beneficial for desalting during the process, adding to the rapid analysis without the need for pre-buffer exchange (Deslignière *et al.*, 2021).

1.2.7.1.2 Ion-mobility Mass Spectrometry (IM-MS)

The capabilities of native MS can be expanded by coupling to ion-mobility (IM) (Figure 1.15), where molecules can be separated by size and shape in the MS by colliding with inert gas molecules. This gas-phase separation technique can separate co-populated conformations of ions with the same *m/z*. This has successfully been used to study small aggregation-prone amyloid proteins (Bernstein *et al.*, 2010) and larger aggregation-prone mAbs (Vallejo, Jeon, *et al.*, 2022). The IM drift time of each ion can be converted into collision cross-section (CCS) values, which assumes that the CCS corresponds to the averaged rotational 2D projection of the biomolecule's 3D structure. Conformations identified by native IM-MS can be compared to the predicted values *in silico* (for example with IMPACT (Ion Mobility Projection Approximation Calculation Tool) (Marklund *et al.*, 2015) which have been calculated from X-ray crystallography

or NMR experiments. Proteins are considered to remain in a 'native-like' conformation in the timescales of an IM-MS experiment. However, there is debate about the extent that proteins can retain their conformations – for example, mAbs have been shown to undergo gas phase collapse when the lack of solvent (and therefore the hydrophobic effect) leads to hinge movement (Devine *et al.*, 2017).

Furthermore, collisional heating of mAbs (collision-induced unfolding, CIU), followed by IM separation and MS analysis can provide information on gas phase stability. It has been shown to distinguish mAb isotypes (IgG1/2/3/4) with different patterns of disulfide bonding and glycosylation (Pacholarz *et al.*, 2016; Terral *et al.*, 2016), useful quality control for mAb analysis in industrial manufacture. Recently, Vallejo and colleagues demonstrated the power of IM-MS to investigate mAb structures that have been stressed to aggregate with temperature and pH changes, and found suggestions that differences in mAb CDRs drive differential responses to degradation, influencing HOS (Vallejo, Jeon, *et al.*, 2022).

1.2.7.2 Probing solution-phase conformation with footprinting techniques

Over the past two decades, advances in liquid chromatography (LC) coupled to tandem MS (LC-MS/MS) technology means that MS/MS sequencing of individual peptides can be accurately and reliably assigned to spectra with increasing ease (Graf *et al.*, 2020). MS/MS sequencing is a routine method applied to the development and quality control of biotherapeutic molecules, in order to identify changes to the sequence that could impact safety and efficacy by affecting surface charge and solubility. These include sequence heterogeneity through post-translational modifications (PTMs) arising from recombinant mAb production (Beck & Liu, 2019), or any markers of degradation, for example deamidation of asparagine residues in accelerated degradation studies or oxidation of methionine during storage (Jefferis, 2016).

Sequence modification using non-covalent and covalent probes has been harnessed as a tool to probe for solvent-accessible parts of structures, which create a 'footprint' of protein structure in solution. Footprinting refers to the use of a label which attaches to solvent-exposed sections of a protein's tertiary structure. The position of these labels on the primary sequence can then be used to map onto sections of tertiary structure, for applications such as native state conformation, protein-protein/ligand interactions, sites of unfolding and aggregation and more (Figure 1.15). This thesis uses reversible labelling by Hydrogen deuterium exchange (HDX-LC-MS/MS: Chapter 3.5) and irreversible labelling by Fast Photochemical Oxidation of Proteins (FPOP-LC-MS/MS: Chapters 4.3 and 5) to report on solution-phase structural conformation.

1.2.7.2.1 Hydrogen-deuterium exchange-Mass Spectrometry (HDX-MS)

Hydrogen deuterium exchange (HDX) is a non-covalent modification strategy where solvent-accessible hydrogen is exchanged with deuterium, a heavy isotope of hydrogen with one neutron, which is detectable on the peptide backbone (see explanation in the next paragraph). Due to this +1 Da mass difference, HDX can be monitored by shifts in a spectrum over the time of exchange. The deuterated protein is typically proteolytically cleaved into peptides for LC-MS analysis. The differences in deuterium uptake are localised at a peptide level resolution (Kan *et al.*, 2013).

HDX for structural MS applications is typically performed as a continuous labelling technique, where sample is diluted into a deuterated buffer and left to equilibrate for fixed periods of time, before the reaction is quenched at discreet time points at low temperature and pH to limit further exchange. However, the exchange of H and D at the position of covalent N-H bonds are reversible, and back exchange can occur even when quenched (Englander, 2006), but this exchange is more rapid on the side chains of amino acids compared to the backbone. Exchange of the backbone labile protons (hydrogens involved in N-H, O-H and S-H bonds) occurs on a real-time measurable time scale (Konermann *et al.*, 2011). Therefore, HDX comments on changes to peptide backbone solvent accessibility over time, with near-complete coverage of every amino acid (but proline) (X. R. Liu *et al.*, 2020).

HDX is the most common footprinting technique example for assessing protein dynamics in solution, and can be used for characterisation of protein-protein/ligand interactions (Chalmers *et al.*, 2011; J. Li *et al.*, 2017), epitope mapping to lead therapeutic design (Zhu *et al.*, 2021), and identifying regions prone to aggregate (Benhaim *et al.*, 2020; Knight *et al.*, 2022; H. Zhang *et al.*, 2014; Y. Zhang *et al.*, 2013). The development of pulse-labelling HDX-MS has enabled the monitoring of rapid events such as the early stages of aggregation, and has been used to understand the mechanism of aggregation of intrinsically disordered amyloid proteins (Y. Zhang *et al.*, 2013). Additionally, HDX-MS is becoming increasingly incorporated into biotherapeutic development (Benhaim *et al.*, 2020), as a tool for ensuring biocomparability (Houde *et al.*, 2012), and for structurally evaluating mAb native conformation (Pan *et al.*, 2015). Recently, HDX has been used to study an IgG4 under temperature stress, which elucidated residues in the variable and constant domains responsible for stress-induced dimerisation (Knight *et al.*, 2022).

Because of the drawback of reversible exchange, it is becoming more popular in the field to use a complement of footprinting techniques alongside HDX, such as Fast Photochemical Oxidation of Proteins (FPOP) and crosslinking – for investigating protein structure, conformation and

aggregation, further adding confidence to the results of each technique (Vallejo, Rojas, *et al.*, 2022; Wagner & Gross, 2022). Covalent labelling with FPOP will be discussed in the next section.

1.2.7.2.2 Fast photochemical oxidation of proteins (FPOP)-mass spectrometry

Hydroxyl radical footprinting (HRFP) is one of the most widely used covalent labelling methods to characterise protein structure and dynamics (L. Wang & Chance, 2017). HRFP utilises hydroxyl radicals (·OH) to oxidise amino acid side chains on a protein's surface. This oxidation label will persist in its position on a protein's primary sequence, so that following digestion the label can be identified, and used to assign which areas of primary sequence were initially surface-exposed in the protein fold at the time of labelling. Hydroxyl radicals can be generated in aqueous solution by a variety of methods including radiolysis of water by synchrotron X-ray, or UV photolysis of hydrogen peroxide (H_2O_2) (B. Zhang *et al.*, 2018; M. M. Zhang *et al.*, 2019). A common goal of all HRFP methodologies is to limit the level of exposure and minimise overall reaction time, in order to avoid triggering modification-induced unfolding, whilst capturing fast phenomena at the point of irradiation. The tuning of the level of exposure to radical chemistry (by adjusting the starting H_2O_2 concentration and amount of scavenging molecules in the solution) is to avoid additional labelling occurring at a later stage, compromising the output with observations which do not reflect the protein structure at the precise timepoint of footprinting. Sensitive tuning of the average lifetime of radicals in an FPOP experiment is around 1 microsecond, in order for a sample to only be exposed to radicals in a time period that is arguably faster than the time taken for protein structural rearrangement (Hambly & Gross, 2005; X. R. Liu *et al.*, 2020). Therefore, rapid generation of radicals created by laser photolysis in a flow system would be preferable for exposing samples over a nanosecond time scale (B. Zhang *et al.*, 2018). This method of capturing the system in a nanosecond time range becomes important in systems which are not in equilibrium, therefore capturing several instances before, during, and after an event would be an exciting application of the technique.

Fast Photochemical Oxidation of Proteins (FPOP) is a method that is incorporated under the blanket of HRFP techniques, and was developed in the early 2000's (Hambly & Gross, 2005). The technique is summarised below in Figure 1.16. A krypton-fluoride (KrF) excimer laser is used in the FPOP procedure to liberate ·OH from H_2O_2. Inert gases Kr and F in the laser housing are excited under the influence of a high voltage electric discharge, briefly form an unstable KrF complex through gaseous collisions, but decay instantly, generating nanosecond pulses of UV laser light at 248 nm (Figure 1.16 a). The sample is prepared in buffer to the desired

concentration, along with a scavenger amino acid such as L-histidine to tune radical exposure (the lifetime of primary ·OH radicals is sub-microsecond in the FPOP protocol) (B. Zhang *et al.*, 2018). Immediately before flowing sample through a capillary at a fixed flow rate in the path of the laser, H_2O_2 is added to the sample as the source of the oxygen radicals. In the laser path, oxygen radicals are generated from H_2O_2 and react on sub-microsecond time scales with solvent-exposed amino acid side chains in the analyte (Figure 1.16 a). The sample then flows immediately into a quench mixture containing L-methionine and catalase (Figure 1.16 b). L-methionine (highly reactive amino acid) further scavenges any residual radicals, and catalase decomposes any remaining H_2O_2 to water and oxygen (Calabrese *et al.*, 2015). A control sample which contains H_2O_2 but not exposed to the laser is also quenched at the end of the reaction. Another control without H_2O_2 might also be included, but is not essential. The end mixture is taken for reduction, alkylation, digestion (Figure 1.16 c) and LC-MS/MS analysis (Figure 1.16 d).

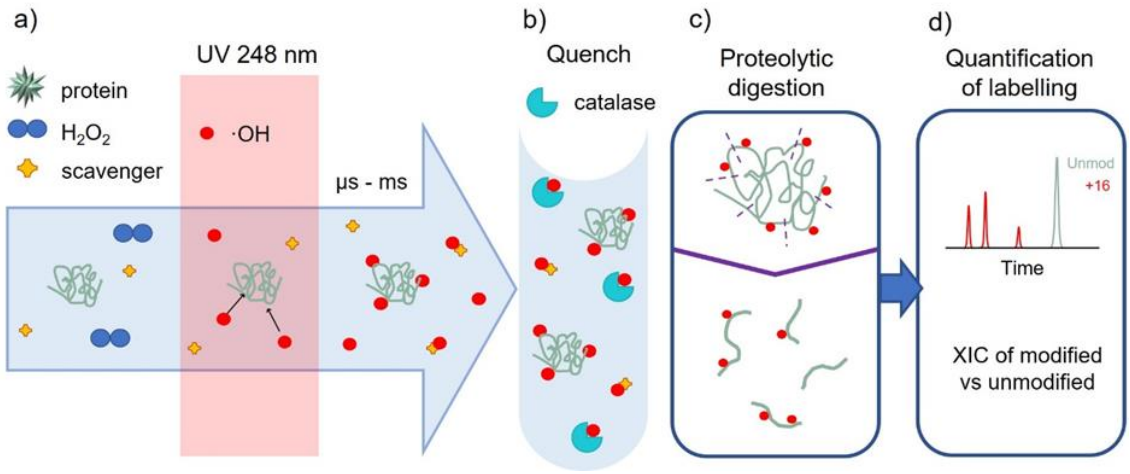


Figure 1.16 Simplified FPOP-LC-MS/MS workflow.

a) Protein sample is mixed with a scavenger amino acid and H_2O_2, and flowed through the path of a pulsed KrF laser (248 nm UV) (blue arrow demonstrates constant flow). The H_2O_2 is split into hydroxyl radicals in the path of the laser (red circles). The labelling occurs at sub-µs (assumed nanosecond) timescales. b) Irradiated sample is collected in the presence of quench solution containing additional scavenger amino acids and catalase. c) Resulting intact sample is proteolytically digested and resulting peptides are analysed by LC-MS/MS. d) The labelling is quantified (% modified) by comparing the normalised relative intensity of the area in the extracted ion chromatograms (XICs) of each peptide in the sequence. The unmodified peak is compared to that of the modified, for example +16 Da to unmodified (see Table 1.2 for common modifications).

As ·OH radicals can attack almost any region of a protein, attack of the α-carbon can (less commonly) result in fragmentation of the polypeptide backbone (Xu & Chance, 2007), however the α-carbon position is usually sterically shielded by the amino acid side chains which become modified.

There are a range of modifications that each amino acid side chain can get. By far the most common modification is +16 Da hydroxylation, although many others are possible, and some of the common modifications in FPOP are summarised in Table 1.2. This is not an exhaustive list, but an illustration of common modifications. It is important to note that carbamidomethylation (+57.02 Da) is presumed to be a fixed modification after (reduction and) alkylation of cysteine residues in the tryptic digest protocol of proteins in this thesis.

Table 1.2 A list of common FPOP modifications.

Modification (+ abbreviation)	**Mass difference (Da)**	**Common amino acid reaction mechanism**
One hydroxylation (+16)	+15.99	Addition of OH and abstraction of H
Two hydroxylations (+32)	+31.99	Addition of OH and abstraction of H, x2
Three hydroxylations (+48)	+47.99	Addition of OH and abstraction of H, x3
Carbonyl formation (+14)	+13.98	Addition of O and abstraction of 2H
Single oxidation following carbamidomethylation (+73)	+73.01	Cysteine side chains modified by iodoacetamide, then singly oxidised
Double oxidation following carbamidomethylation (+89)	+89.00	Cysteine side chains modified by iodoacetamide, then doubly oxidised
Decarboxylation (-30)	-29.99	Loss of CO_2 from carboxylic acid side chains
Deguanidination (-43)	-43.00	Loss of guanidine from arginine

Hydroxyl radicals preferentially attack large hydrophobic residues or sulfur-containing side chains, in a hierarchical manner over other amino acid side chains (Xu & Chance, 2007) (the amino acid reactivity hierarchy can be found in Table 1.3). All 20 naturally occurring amino acids can be labelled with hydroxyl radicals, however due to the several orders of magnitude difference between the most reactive (Cysteine) and the least (Glycine), it is only common to see 14 out of the 20 routinely labelled in standard FPOP experiments of intact proteins (L. Wang & Chance, 2017), due to competing local reactivity of neighbouring amino acids and differences in amino acid sequences. The small van der Waals radius of the radical ensures it has a greater surface accessibility than other larger covalent labelling reagents, such as diethylpyrocarbonate (Limpikirati *et al.*, 2020) or diazirine-based carbene probes (Manzi *et al.*, 2016).

Table 1.3 Rate constants for reaction of amino acid side chains with hydroxyl radicals and common mass additions. Taken from (Xu & Chance, 2007).

Side chain	**Abbreviation**	**Reactivity rate (M^{-1} s^{-1})**	**Common Δmasses (Da)**
Cys (most reactive)	C	3.5×10^{10}	+48, +32, -16
Trp	W	1.3×10^{10}	+16, +32, +48 etc
Tyr	Y	1.3×10^{10}	+16, +32
Met	M	8.5×10^{9}	+16, +32, -32
Phe	F	6.9×10^{9}	+16, +32, +48
His	H	4.8×10^{9}	+16, -22, -10, +5
Arg	R	3.5×10^{9}	-43, +16, +14
Ile	I	1.8×10^{9}	+16, +14
Leu	L	1.7×10^{9}	+16, +14
Val	V	8.5×10^{8}	+16, +14
Pro	P	6.5×10^{8}	+16, +14
Gln	Q	5.4×10^{8}	+16, +14
Thr	T	5.1×10^{8}	+16
Lys	K	3.5×10^{8}	+16, +14
Ser	S	3.2×10^{8}	+16
Glu	E	2.3×10^{8}	-30, +16, +14
Ala	A	7.7×10^{7}	+16
Asp	D	7.5×10^{7}	-30, +16
Asn	N	4.9×10^{7}	+16
Gly	G	1.7×10^{7}	-

Since its development, FPOP has been applied to many structurally diverse and biopharmaceutically relevant systems (Cornwell & Ault, 2022). FPOP has been used to highlight the areas of structure occluded by binding partners, known as epitope mapping, in examples such as transmembrane proteins (K. S. Li *et al.*, 2017), and monoclonal antibodies (Y. Zhang *et al.*, 2017). As well as *in vitro* labelling, FPOP can successfully be used to label cells to probe protein structure *in vivo* (Espino *et al.*, 2015, 2020).

FPOP is suited for investigating aggregation-prone proteins (Cornwell & Ault, 2022) including amyloid systems (Cornwell *et al.*, 2021; K. S. Li *et al.*, 2016), and amorphous aggregate systems thought to be triggered through partial unfolding and conformational change. Indeed, FPOP has been used to highlight areas of long-range conformational changes in mAbs, where in an IgG1's Fab domain had a region of protection and exposure on the Fab domain triggered by Fc binding (L. Shi *et al.*, 2019). MAbs with differing amounts of reversible self-association have been compared in detail using FPOP, which showed protection in the C_L-C_H1 and C_H1-C_H2 interfaces in the mAb more aggregation-prone than its counterpart (Cornwell *et al.*, 2019).

FPOP can probe similar changes in structure as HDX. The rationale behind FPOP is that side chains with greater solvent accessibility will undergo more oxidative labelling than those buried from the surface. It can comment on conformation, dynamics, and protein-protein/ligand interactions, because these trigger changes the accessibility of side chains to the solvent, therefore changing degree of modification. Additionally, overall levels of FPOP modification can be detected by introducing the sample into the MS intact, and quantifying the observed changes in mass distributions, which can also be done with HDX samples. However, unlike HDX, the covalent label is not susceptible to back exchange and therefore the reaction does not need to be quenched in low temperatures and pH. This also means the proteolytic enzyme used for digestion is not restricted to acid proteases.

Despite FPOP being an attractive technique, there are still shortcomings in the current understanding of the technique. The effect of local microenvironment on the extent of modification, in addition to solvent accessible surface area, complicates the analysis and might affect the interpretation of any structural changes (Cornwell *et al.*, 2019; Xie *et al.*, 2017). The analysis to amino-acid-level resolution is arduous, time intensive and manual. Nevertheless, the potential of FPOP as part of the structural MS toolbox is clear. How this technique might be used to capture protein conformation has been explored, and now its application to studying the stages involved in protein unfolding and aggregation forms the basis and aims of this thesis.

1.3 Basis of the study

1.3.1 WFL and STT: model monoclonal antibodies

In this thesis, the biophysical characteristics of a pair of highly homologous monoclonal antibodies (mAbs) with different aggregation properties will be studied. MEDI-1912 (which will be referred to as WFL) was a product of *in vitro* affinity maturation of the parent mAb MEDI-578, developed against nerve growth factor for the treatment of chronic pain (C. L. Dobson *et al.*, 2016). WFL had enhanced picomolar affinity (Kd = 69 pM) for its target, nerve growth factor (NGF), but displayed poor biophysical characteristics, reversible self-association character demonstrated by Analytical Ultracentrifugation, non-specific binding to SEC column matrices, and oligomer formation by dynamic light scattering (DLS). This mAb interacted with column matrices and adsorbed to filter membranes, resulting in poor yields during purification, along with displaying colloidal instability (low solubility and a propensity to precipitate) (C. L. Dobson *et al.*, 2016).

Hydrogen-deuterium exchange experiments identifying the aggregation interface through protection, was analysed alongside *in vitro* analysis of the variable region of the protein using the computational modelling Spatial Aggregation Propensity software (SAP) (Chennamsetty *et al.*, 2009). Together this work identified a hydrophobic patch on the surface of WFL's variable regions (V_H and V_L) as being prone to aggregation (C. L. Dobson *et al.*, 2016). This surface hydrophobicity aligns with the areas predicted to be aggregation-prone using Aggrescan3D (Zambrano *et al.*, 2015) (Figure 1.17 a). Consequently, these data aided the rational design of the variant STT, where the triple amino acid substitution W30S, F31T and L57T reverted these amino acids back to the parent MEDI-578 sequence, and in turn reduced the hydrophobic patch surface area (Figure 1.17 b).

WFL and STT share 99.6% sequence similarity, and the latter displays reduced interaction with the column matrix compared to WFL (Figure 1.17 c). Analytical ultracentrifugation (AUC) at 1 mg mL^{-1} demonstrated that WFL formed higher-order oligomers (Figure 1.17 d) but STT remained monomeric in solution (Figure 1.17 e). At 0.1 mg mL^{-1} however, both WFL and STT are monomeric (C. L. Dobson *et al.*, 2016). Therefore, the pair of mAbs developed by AstraZeneca provide an ideal model system for studying self-association in biopharmaceutically-relevant IgGs.

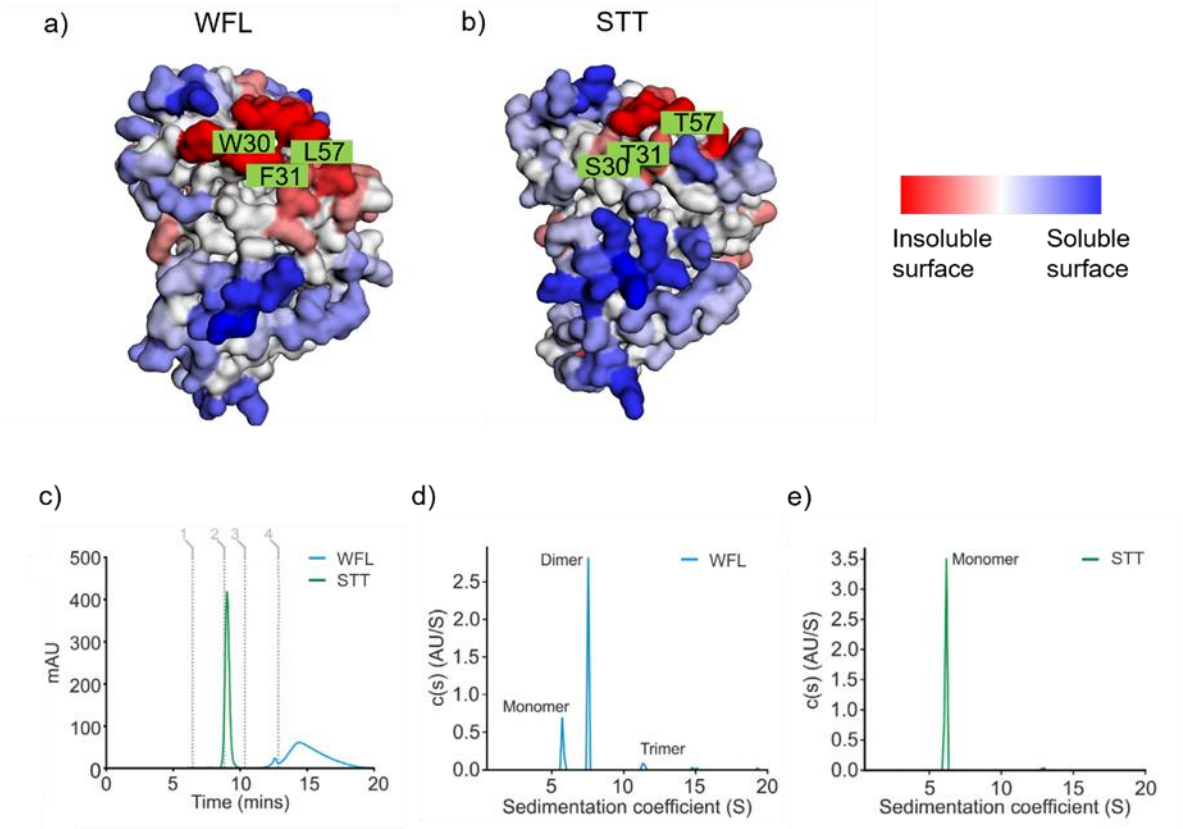


Figure 1.17 WFL and STT are structurally similar but demonstrate different biophysical properties.

a) Aggrescan3D analysis of WFL_scFv reveals a surface exposed hydrophobic patch (red). Residues W30, F31 and L57 are highlighted (green = mutations in the heavy chain). b) Aggrescan3D analysis of STT, with the mutated residues W30S, F31T and L57T highlighted (green = mutations in the heavy chain). c) High performance size exclusion chromatography (HP-SEC) elution profiles of WFL (green) and STT (blue) at 280 nm (mAU), 1 mg mL^{-1} in 0.1 M sodium phosphate, 0.1 M sodium sulphate, pH 6.8. Grey lines indicate elution times of calibrant proteins: 1, Thyroglobulin (670 kDa); 2, IgG (158 kDa); 3, Ovalbumin (44 kDa) and 4, Vitamin B12 (1.35 kDa). d) AUC of WFL at 1 mg mL^{-1}. e) AUC of STT at 1 mg mL^{-1}. HP-SEC and AUC data from (C. L. Dobson et al., 2016).

WFL and STT have different flow-induced aggregation behaviour (see Section 1.1.4: Fluid flows and their effects on molecular structure) (Willis *et al.*, 2018), however the exact mechanism of aggregation and the conformational changes associated with these amino acid substitutions remain poorly investigated. WFL and STT have been previously compared by FPOP-MS (Cornwell *et al.*, 2019), revealing protection in WFL's C_L-C_H1 and C_H1-C_H2 interfaces distal to the substitution sites in the V_H, indicating long-range effects of the substitutions on mAb conformation that differ between the variants. How far the substitutions influence stability and unfolding in response to flow stress remains unstudied. Additionally, as the key method used for comparing the mAbs is FPOP, the difference in reactivity of the residue side chains to

hydroxyl radicals (hydrophobic tryptophan and phenylalanine residues are more reactive than serine and threonine, for example (Table 1.3, (Xu & Chance, 2007)) makes it difficult to directly compare the hydrophobic V_H region in WFL to the homologous V_H region in STT. Therefore to help with this, a third variant, 114, will be studied, and is described in the next section.

1.3.2 Variant 114

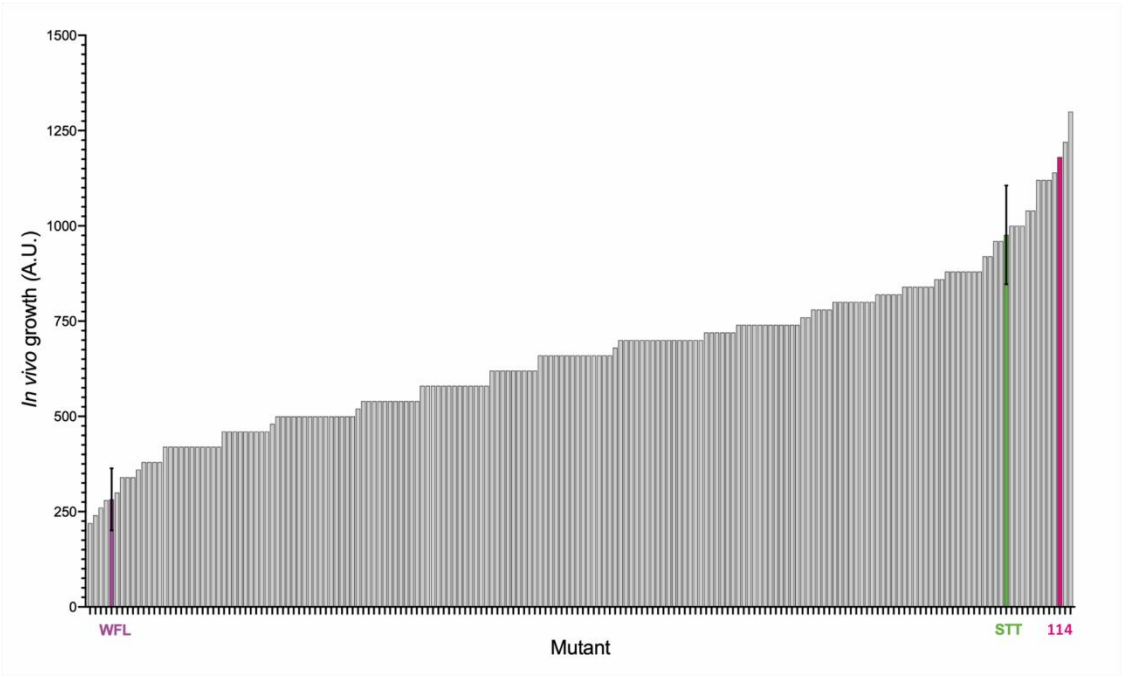
A tripartite β-lactamase enzyme assay (TPBLA) has previously been used as a directed evolution screen to identify and rank aggregation-prone proteins and peptides (Ebo, Saunders, *et al.*, 2020; Saunders *et al.*, 2016). Here, a test protein or peptide sequence is inserted between two halves of the β-lactamase enzyme sequence, connected by a glycine-serine flexible linker. The plasmid containing this DNA is transformed into *E. coli* cells, which are cultivated, allowing the expression of the proteins *in vivo*. Cells are grown in media containing increasing amounts of ampicillin (a β-lactam antibiotic). If the test sequence can properly fold – indicating a low aggregation potential – then the two halves of the β-lactamase enzyme will form as one, and cells will be able to breakdown the antibiotic and survive. However, if the test sequence aggregates, or misfolds and is degraded, then the β-lactamase enzyme will also be pulled into this aggregation or degradation in the cell, and therefore not correctly function, resulting in cell death. This cell death assay can be used to screen a huge variety of different sequences and antibiotic conditions.

Recently, the assay has been applied to the variable regions of mAbs, to select for sequences that are more aggregation-resistant than the parent sequence. An error-prone PCR library of the variable domain of WFL in scFv format was subjected to the assay to screen for variants with improved developability (Ebo, Saunders, *et al.*, 2020). The mutational frequency profile from 315 evolved WFL_scFv colonies revealed 12 hotspot residues, 9 in the V_H (most of which are hydrophobic, solvent exposed and clustered in the CDR regions) and 3 in the V_L: W30 and F31 are hotspot residues residing in a hydrophobic surface patch in the V_HCDR1.

185 variants were randomly selected and ranked by *in vivo* growth score using the TPBLA. 181 of the 185 variants identified in the selection had higher scores in the TPBLA assay than WFL, and 12 had higher scores than STT (Figure 1.18 a). Of these, variant number 114 was chosen to study further in this thesis as it contains four mutations in the heavy and light chain collectively, yet retains the aggregation-prone hydrophobic patch characterized by the W and F residues (Figure 1.18 b). The Aggrescan3D *in silico* prediction visualises a hydrophobic patch surface area

similar to WFL (Figure 1.18 c). A more detailed sequence alignment can be found in Appendices: Figure 7.1.

a)



b)

Variant	Mutations
WFL	-
STT	W30S, F31T, L57T
114	F29S (V_H), I52N (V_H), S2G (V_L), T77A (V_L)

c)

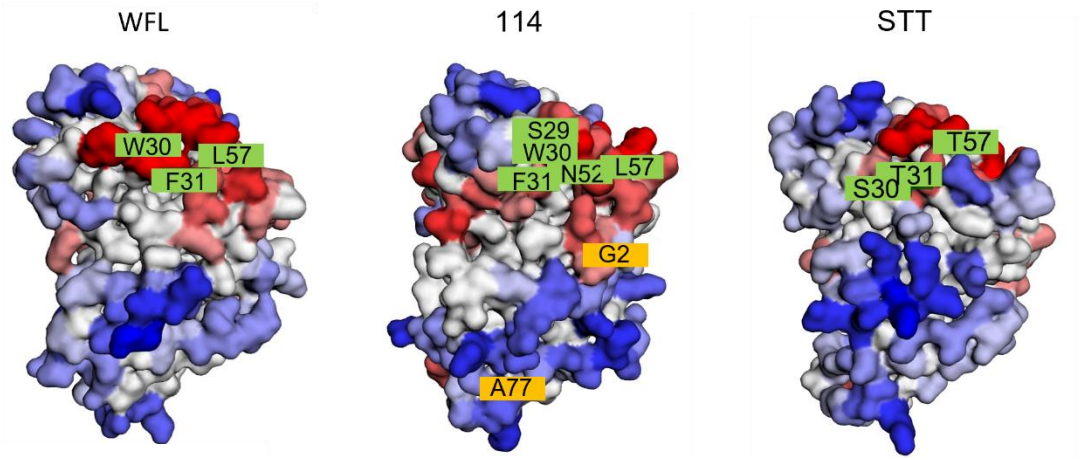


Figure 1.18 114 is structurally similar to WFL and STT.

a) Ranked in vivo growth score of 185 evolved scFv-WFL variants from the TPBLA. Grey bars = mean score of each mutant. ScFv-WFL (purple, mean in vivo *growth score 280 A.U.) and scFv-STT (green, mean* in vivo *growth score 975 A.U.) error bars represent n = 16 biological repeats. ScFv-114 (pink,* in vivo *growth score 1180 A.U.). Recreated from (Ebo, Saunders, et al., 2020). b) Mutations in 114 and STT compared to WFL. c) Aggrescan3D analysis of 114 (center),*

alongside WFL (left) and STT (right), revealing the surface exposed hydrophobic patches on each variant (indicated by red surface area). Residues W30, F31 and L57, and additional mutations are highlighted (green = mutations in the heavy chain, yellow = mutations in the light chain).

1.4 The aims of the thesis

In this thesis, the applicability of an FPOP-proteomics workflow for capturing the solution-phase structures of mAbs will be demonstrated. FPOP will be used to capture any perturbations to a protein structure, and specifically in this work, how a protein structure may be changed after a hydrodynamic force event. After flow stressing, the protein solution will be exposed to the FPOP laser as quickly as feasibly possible after the event. This means there is potential to capture any persistent partial unfolded states or conformations triggered in response to flow. Overall, this workflow holds the potential to form the basis of understanding the process of flow-induced unfolding leading to aggregation of therapeutic mAbs.

The aims of this thesis are to:

1. Characterise the biophysical properties of WFL, STT and 114, and to compare their solution-based propensities for associations.
2. Build a robust methodology for performing LC-MS/MS on FPOP-modified peptides, with a view to observing the highest coverage possible.
3. Test the effectiveness of a flow-FPOP-MS workflow to characterise any structural changes influenced by flow stress, and to determine influences that amino acid substitutions may make to their susceptibility to unfold, and aggregate.

Chapter 2

Materials and methods

2 Materials and methods

2.1 Materials

2.1.1 Technical equipment

Table 2.1 List of technical equipment used in this thesis

Analytical instruments and equipment	
20 cm capillary emitter column C18 bead column (inner diameter 75 µm, packed with 3 µm Reprosil-Pur 120 C18 media)	Dr. Maisch
EASY-nLC 1000 LC	Thermo Scientific
HPLC polypropylene insert vials (0.3 mL)	VWR
HPLC vial caps	ThermoFisher
Immobilised ethylene bridged hybrid (BEH) pepsin Enzymate column	Waters Ltd.
LC40 HPLC system	Shimadzu
M-Class Acquity LC system with HDX manager	Waters Ltd
Proteomix HIC Butyl-NP5 non-porous column	Sepax Technologies
RF20A fluorescence detector	Shimadzu
TSKgel GSK3000XL HPLC column	TOSOH Bioscience
Ultimate 3000 RSLCnano LC	ThermoFisher
VanGuard pre-column Acquity UPLC BEH C18 trap column (1.7 µm, 2.1 µm x 5 µm)	Waters Ltd.
Zenix SEC-300 column	Sepax Technologies
Centrifuges	
Optima TLX Ultracentrifuge	Beckman Coulter
Thickwall Polycarbonate ultracentrifuge tubes	Beckman Coulter
TLA100 rotor	Beckman Coulter
Spectrophotometer	

UV-1800 UV/Visible Scanning Spectrophotometer	Shimadzu
1 cm path length fluorescence cuvettes	Hellma Analytics

Extensional Flow equipment

Arduino Microcontroller	Arduino
Breadboard base	Thor Labs
Borosilicate Glass capillaries (0.3 mm internal diameter, 75 mm length)	Sutter Instruments
Ferrule compression fittings	Hamilton
Gas-tight 1 mL syringes 1001 RN model	Hamilton
Gilson P10 O-ring	Gilson
Stepper motor	Haydon Switch and Waterbury CT Instrument Co.
Syringe clamps	Thor Labs

Mass Spectrometers

Orbitrap Fusion Lumos Tribrid	Thermo Scientific
Orbitrap Exploris 240	Thermo Scientific
Velos Linear Trap Quadrupole (LTQ) Orbitrap	Thermo Scientific
Q-Exactive UHMR	Thermo Scientific

Software

Aggrescan 3D 2.0 webserver	http://biocomp.chem.uw.edu.pl/A3D2/
BioPharma Finder	ThermoFisher
HDExaminer	Sierra Analytics
MassLynx V4.1	Waters
Origin Pro 2020	OriginLab Corp.
PEAKS Studio v10	Bioinformatics Solutions Inc.
PyMOL (Version 4.4)	Schrödinger
UniDec v5.1.1	(Marty *et al.*, 2015); University of Oxford 2017-2019, University of Arizona

XCalibur v4.0.27.19 - FreeStyle	ThermoFisher

Other Equipment

Coherent COMPexPRO 50F Krypton Fluoride Excimer Laser	Coherent Inc., Ely, UK
Electrospray borosilicate capillaries (1.5 mm OD x 0.78 mm ID, filamented)	World Precision Instruments
EZ-2 Personal Evaporator	GeneVac, Ipswich, UK
Jenway 3020 Bench pH Meter	Bibby Scientific, Stone, UK
LEAP HDX platform	LEAP Technologies, FL, USA
Protein Low-Bind tubes	Eppendorf
P-97 Micropipette Puller	Sutter Instrument Company
Midi GeBaFlex-tube Dialysis cassettes (8 kDa MWCO)	Generon
Rapigest SF surfactant	Waters Corp., Manchester, UK
Sep-Pak Vac 1cc (50 mg) tC18 cartridges	Waters Corp., Manchester, UK
Polaron SC7620 Sputter coater	Quorum Technologies
S-Trap cartridges	Protifi
SUPR-DSF fluorescence plate reader	Protein Stable
Syringe-driven 0.22 and 0.45 µm filters	Merck Millipore and Jet Biofil
Vivaspin columns, 5 kDa MWCO, 20 mL	Sartorius UK Ltd, Epsom, UK
Zeba Spin Desalting columns, 7 kDa MWCO	Thermo Scientific, UK

2.1.2 Chemicals table

Table 2.2 List of chemicals used in this thesis (all reagent grade unless stated)

A	
Glacial acetic acid	Fisher Scientific, Loughborough, UK
Acetonitrile (MeCN)	BioSolve Chimie, France
Acrylamide, 30% (*v/v*) Bis-Acrylamide stock solution	Severn Biotech, Kidderminster, UK
Ammonium acetate stock solution (7.5 M) (HPLC-grade)	Sigma-Aldrich, St. Louis, MO, USA
Ammonium persulfate	Sigma-Aldrich, St. Louis, MO, USA
L-Arginine	Sigma-Aldrich, St. Louis, MO, USA
C	
Chloroform	Sigma-Aldrich, St. Louis, MO, USA
D	
1,2-Dithiothreitol (DTT)	Formedium, Norfolk, UK
E	
Ethanol	Fisher Scientific, Loughborough, UK
F	
Formic acid (FA) 99% (*w/v*), MS grade	BioSolve Chimie, France
G	
Guanidine hydrochloride (GdnHCl)	Sigma-Aldrich, St. Louis, MO, USA
Guanidine hydrochloride (GdnHCl) 8M solution	BioChemica, UK
H	
HPLC-grade H_2O	Merck, MA, USA

Hellmanex III	Hellma Analytics, Essex, UK
L-Histidine	Sigma-Aldrich, St. Louis, MO, USA
Hydrochloric acid (37% *w/v*)	Fisher Scientific, Loughborough, UK
Hydrogen Peroxide, 30% (*w/v*) solution	Fluka (Sigma-Aldrich, St. Louis, MO, USA)

I

Iodoacetamide	Sigma-Aldrich, St. Louis, MO, USA

M

Methanol	Sigma-Aldrich, St. Louis, MO, USA
L-Methionine	Sigma-Aldrich, St. Louis, MO, USA

P

Phosphoric acid	Sigma-Aldrich
Potassium phosphate (KH_2PO_4)	Sigma-Aldrich, St. Louis, MO, USA

S

Sodium acetate	Sigma-Aldrich, St. Louis, MO, USA
Sodium chloride (NaCl)	Fisher Scientific, Loughborough, UK
Sodium dodecyl sulfate (SDS) 10% (*w/v*)	Sigma-Aldrich, St. Louis, MO, USA
Sodium hydroxide (NaOH)	Fisher Scientific, Loughborough, UK
Sodium succinate dibasic hexahydrate	Sigma-Aldrich, St. Louis, MO, USA
Sodium sulfate (Na_2SO_4)	Sigma-Aldrich, St. Louis, MO, USA

T

1M Triethylammonium bicarbonate (TEAB)	Thermo Fisher Scientific, Loughborough, UK
Trifluoroacetic acid (TFA) 1mL aliquots	Thermo Fisher Scientific, Loughborough, UK
Pierce™ Trypsin Protease, 1 mg, MS Grade	Thermo Fisher Scientific, Loughborough, UK

U

Urea	Sigma-Aldrich, St. Louis, MO, USA

2.1.3 Buffer table

All buffers were filtered with 0.22 µm filter before use. Buffers for chromatography were also degassed before use.

Table 2.3 Buffers used in this study

Formulation buffer	20 mM sodium succinate dibasic hexahydrate 125 mM L-arginine, pH 6.0
Non-protective solution/Native MS solution	150 mM ammonium acetate, pH 6.0
AZ HPLC running buffer	100 mM sodium phosphate, 100 mM sodium sulfate pH 6.8
HIC dilution and running buffer	1.8 M ammonium sulfate, 100 mM sodium phosphate, pH 6.5
HIC elution buffer	100 mM sodium phosphate pH 6.5
SMAC running buffer	150 mM sodium phosphate, pH 7.0
PBS buffer	pH 7.4
HSA buffer	20 mM L-histidine, 120 mM sucrose, 80 mM L-arginine, pH 6
S-Trap Binding buffer	100 mM TEAB, pH 7.1 (phosphoric acid) (from 1M stock) 9:1 methanol:100 mM TEAB solution
S-Trap Elution buffer 1	50 mM TEAB (from 1M stock)
S-Trap Elution buffer 2	0.2% *v/v* formic acid
S-Trap Elution buffer 3	50% acetonitrile, 0.2% *v/v* formic acid
Peptides mobile buffer A	HPLC-grade H_2O, 0.1% *v/v* formic acid
Peptides mobile buffer B	Acetonitrile, 0.1% *v/v* formic acid
AZ Denaturing Buffer	7.2M GdnHCl, 90mM Tris, 0.1mM EDTA, 45.5 mM DTT, pH 7.6
HDX labelling buffer	20 mM sodium succinate dibasic hexahydrate 125 mM L-arginine, pD 6.0
HDX quench buffer	100 mM potassium phosphate 8 M Guanidine HCl, pH 2.3
FPOP Scavenger solution	100 mM L-Histidine in formulation buffer
FPOP Quench solution	1 µM catalase 100 mM L-methionine in formulation buffer

2.2 Methods

2.2.1 Industrial mAbs used in this thesis

2.2.1.1 IgG purification

The IgG1 WFL_mAb and STT_mAb were provided by the Biologics expression team at AstraZeneca. 114 expression was performed by Dr Janet Saunders (AstraZeneca). IgG expression vectors for 114_mAb were created by cloning GeneArt strings of the mutated V_H and V_L domains into expression vectors (V_H domain into IgG V_H IgG1 TM YTE expression vector (pEU1.6); V_L domain into the IgG V_L lambda expression vector (pEU4.4)). The plasmids were co-transfected into HEK293 mammalian cells and IgG expressed. IgG proteins were purified from the culture medium using Protein A chromatography.

WFL and STT mAbs were provided on ice in formulation buffer (Section 2.1.3: Buffer table). 114 mAbs were provided in PBS and dialysed (see Section 2.2.1.2: Dialysis) into formulation buffer within 24 hours of shipment to Leeds. Aliquots were snap-frozen in liquid nitrogen and stored at -80 ˚C until required. For any experiment, aliquots were defrosted on bench and then stored on ice until use.

2.2.1.2 Dialysis

Aliquots of 114 were received on ice in PBS buffer. The solution was transferred to Midi GeBaFlex-tube Dialysis cassettes (8 kDa MWCO, Generon) and dialysed against the desired buffer or solution (1:200 (*v:v*) ratio) for 3 rounds of 4 hours, changing into fresh buffer or solution between rounds.

2.2.1.3 Concentration measurement

Sample was diluted to appropriate approximate concentration using its buffer (normally 1:20), so that the absorbance readings would land within the accurate range of the UV-1800 UV/Visible Scanning Spectrophotometer (Shimadzu). UV-transparent cuvettes (STARSTEDT) were used for absorbance measurements at 280 nm were made relative to the blank buffer. Concentration was determined using the pre-determined molar extinction coefficients and the molecular weights of the mAbs (Section 2.2.1.4, Table 2.4) using the equations below:

Equation 2.1 Calculating concentration using absorbance.

$$C\ (M) = \frac{A_{280}}{\epsilon_{280}\ l}$$

Where:

Equation 2.2 Calculation of concentration in g L^{-1} from concentration in molarity and the molecular weight.

$$C\ (g\ L^{-1}) = C\ (M)\ x\ MW\ (g\ mol^{-1})$$

2.2.1.4 Table of molecular masses and molar extinction coefficients

Table 2.4. Molecular mass and molar extinction coefficients (based on primary sequence) of proteins used in this study.

Protein	**Molecular weight (MW) (Da)**	**Molar extinction coefficient, ε_{280} (M^{-1} cm^{-1})**
WFL	148,422	239,440
STT	148,107	228,440
114	148,303	239,440

2.2.2 Bioinformatics methods

2.2.2.1 In silico *modelling*

The models of WFL_mAb, STT_mAb and 114_mAb were created kindly by Romina Hofele (AstraZeneca) using Schrodinger, where scFvs were grafted onto an IgG1 template. The scFv predictions of WFL, STT and 114 were generated by mutating PDB 5JZ7, the Fab structure of the WFL precursor MEDI-578 (Dobson 2016) in PyMol 2.1.0. All the predicted structures were minimised at pH 6 to correspond to the pH of the formulation buffer and ammonium acetate solution.

2.2.2.2 Aggrescan3D

Aggrescan3D 2.0 (Kuriata *et al.*, 2019) server was used to predict the aggregation propensity of the scFv sequences in dynamic mode. A 10 Å radius was applied and stability calculation option was selected using FoldX to optimise the input structure. Web server address: http://biocomp.chem.uw.edu.pl/A3D2/

2.2.3 Chromatography

2.2.3.1 High performance size-exclusion chromatography

The HP-SEC measurements were performed using a Shimadzu LC40 HPLC system coupled with a RF20A fluorescence detector, and a TOSOH TSKgel GSK3000XL HPLC column (Phase diol, L x I.D. 30 cm x 7.8 mm, 5 µm particle size; Sigma Aldrich) running at 0.5 mL min^{-1} with formulation buffer in a column oven set to 20 ˚C. 20 µL of 1 mg mL^{-1} sample was loaded and absorbance measurements were taken at 280 nm. The area under the intensity curves were extracted for monomer loss comparison. A calibrant mix was run for quality control and calibration purposes (blue dextran (void volume); alcohol dehydrogenase tetramer (145 kDa); alcohol dehydrogenase dimer (82 kDa); ovalbumin (43 kDa); cytochrome C (12 kDa); Vitamin B12 (1.35 kDa).

2.2.3.2 High performance size-exclusion chromatography for monomer loss assay

The HP-SEC % monomer measurements were performed as per the method described in 2.2.3.1: High performance size-exclusion chromatography, except for the eluent buffer was AZ running buffer (0.1 M sodium phosphate, 0.1 M sodium sulfate pH 6.8) to remain consistent with previous data in collaboration with Dr Leon Willis.

2.2.3.3 Hydrophobic Interaction Chromatography

Hydrophobic interaction chromatography-HPLC was performed on the Shimadzu LC40 HPLC system using a Proteomix HIC Butyl-NP5 non-porous column (4.6 x 35 mm) (Sepax Technologies). Samples were diluted 1:1 with 1.8 M ammonium sulfate, 0.1 M sodium phosphate, pH 6.5. 5 µg were loaded onto the column and a gradient run was performed at 1 mL min^{-1} for 26 minutes, eluting with 0 - 100% water containing 0.1 M sodium phosphate pH 6.5. Absorbance measurements were taken at 280 nm.

2.2.3.4 Stand-up Monolayer Adsorption Chromatography

Stand-up Monolayer Adsoption Chromatography was performed on the Shimadzu LC40 HPLC system and a Zenix SEC-300 column (L x I.D. 4.6 cm x 300 mm, pore size 300 Å, 3 µm particle size; Sepax Technologies). Eluent was 150 mM sodium phosphate, pH 7.0. 2 µg of protein was loaded and the flow rate was 0.35 mL min^{-1}. Absorbance measurements were taken at 280 nm.

2.2.4 Affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS)

AffiniPure goat anti-human IgG Fcγ Fragment specific (IgGα-Fc) and ChromePure Goat IgG, whole molecule (IgG_{WHOLE}) (Jackson ImmunoResearch) were buffer exchanged into 20 mM potassium acetate, pH 4.3 and diluted to 0.4 mg mL^{-1}. 600 µL IgGα-Fc and 400 µL IgG_{WHOLE} were incubated with 9 mL citrate-stabilised 20 nm gold nanoparticles (Expedeon) for 2 hours at room temperature. Blocking of the nanoparticles was done with 0.1 µM 2000 MW thiolated PEG (Sigma-Aldrich) for 2 hours at room temperature. Nanoparticles were concentrated to 800 µL in siliconised Eppendorf tubes (VWR) and stored at 4 ˚C. Antibody sample was prepared to 50 µg mL^{-1} in HSA buffer (20 mM L-histidine, 120 mM sucrose, 80 mM L-arginine, pH 6). 45 µL sample was mixed with 5 µL nanoparticle solution and incubated for 30 minutes at room temperature, before loading onto a 384-well polystyrene UV transparent plate (Thermo Scientific). Absorbance from 400 to 700 nm was read in 1 nm increments. The maximum absorbance was determined (the plasmon wavelength) and the redshift in plasmon wavelength was calculated by comparison with the plasmon wavelength of nanoparticles alone (the values were subtracted). Measurements were made in duplicate.

2.2.5 Ultra-low-volume dynamic light scattering (DLS)

Sample was filtered with 0.22 µm filter prior to being prepared to 1 mg mL^{-1} in either formulation buffer or 150 mM ammonium acetate pH 6.0. A buffer baseline was recorded for normalisation of the data collected. Sample was loaded by capillary action into glass capillary cuvettes (Malvern) and capped with sealing compound. Capillaries were held in a sample holder and placed into the light path (633 nm) in the Zetasizer Ultra DLS (Malvern). Particulates were measured using side scatter (90 ˚) and plotted with Debye. Correlogram data was used to extract size (diameter (nm)) versus volume of solution (percent) plots. Measurements collected in technical triplicate.

The correlogram data was fitted with a single-exponential decay function (Equation 2.3 below) to extract R^2 values (goodness of fit).

Equation 2.3 Single Exponential Decay equation fit to correlation function data.

$$y = y_0 + Ae^{\frac{-(x-x_0)}{\tau}}$$

Where A is the amplitude, y_0 is the y axis intercept, x_0 is the x-axis intercept and τ is the delay time, i.e. the amount the intensity trace shifts from the original prior to averaging.

2.2.6 Differential scanning fluorimetry

Differential scanning fluorimetry was performed with a SUPR-DSF fluorescence plate reader (Protein Stable) reading intrinsic tryptophan fluorescence by epi-fluorescence. Intrinsic tryptophan fluorescence refers to the excitation at 280 nm of fluorescent amino acid tryptophan residues found naturally in the protein of interest's sequence. As a protein gradually unfolds during the temperature ramping from 30 ˚C to 100 ˚C, the emission spectrum changes. Here, intrinsic tryptophan fluorescence was monitored, by excitation at 280 nm and measuring the fluorescence emission at 310 – 420 nm. The change in tryptophan emission spectra is monitored: there is a shift in fluorescence emission from 350 nm (more folded) to 330 nm (more unfolded) through previously internally buried tryptophans becoming more exposed to the solvent. By plotting the change in this ratio (first derivative) versus temperature, the apparent temperatures of unfolding ($T_{m,app}$) can be extracted. The higher the T_m indicates higher thermal stability. It is important to note that extrinsic fluorescence can also be measured using an external (reporter) fluorophore such as SYPRO orange but was not performed here.

The SUPR-DSF machine relies on epi-fluorescence where the excitation light and emission light travel through the same objective lens; here, the source is above the sample well and excitation light is transmitted down into the sample well, and then the light emitted from the sample travels back up toward the detector (near the source).

All mAbs were measured at a concentration of 1 mg mL^{-1} in either formulation buffer or 150 mM Ammonium Acetate pH 6.0. 20 µL of 1 mg mL^{-1} samples were loaded (in technical triplicate) into a 394-well black PCR plate and sealed with adhesive sealing film. Plates were loaded into the sample tray in the SUPR-DSF and equilibrated to 30 ˚C. Then, samples were heated from 30 ˚C to 100 ˚C in 1 ˚C increments and epi-fluorescence was measured at each temperature by exciting at 280 nm and measuring emission at 310 – 420 nm.

The ratio of the fluorescence intensity emission at 350 nm and 330 nm as a function of temperature was normalised to buffer signal, then converted into a moving average of 5 points to smooth the data. This was used to calculate the transition temperatures $T_{m,app}1$ and $T_{m,app}2$, by fitting the data to a double Boltzmann three-state model (with an orthogonal distance regression iteration algorithm) using the following equation (Equation 2.4). The fit was then used to calculate the first derivative where the peak apexes corresponded to the $T_{m,app}$ (data processed using Origin).

Equation 2.4 Double Boltzmann fit.

$$y = y_0 + A\left[\frac{p}{1+e^{\frac{x-x1}{k1}}} + \frac{1-p}{1+e^{\frac{x-x2}{k2}}}\right]$$

Where *y* is the observed signal (fluorescence intensity), y_0 is the offset, *A* is the span/height, *p* is the initial signal value, *x1* is the centre of the first transition, *x2* is the centre of the second transition, *k1* is the slope of the first transition, *k2* is the slope of the second transition.

Equation was chosen with reference to (Chakroun *et al.*, 2016; Nashine *et al.*, 2013). First 10 points and last 10 points were excluded from the fitting analysis due to high signal-to-noise.

$T_{m,app}$ data for identified transitions from the three replicates of each protein in each buffer were averaged and reported as $T_{m,app}1$ and $T_{m,app}2$.

2.2.7 Flow Device

2.2.7.1 Apparatus set up

The extensional flow device (J. Dobson *et al.*, 2017; Willis *et al.*, 2018) consists of two modified 1 mL Hamilton gas-tight glass syringes (inner bore diameter 4.61 mm) connected via a 75 mm glass capillary (0.3 mm inner diameter, 75 mm length). The compression fitting was designed to produce a contraction in the fluid path, leading to a 238-fold increase in fluid velocity. Before and after each experiment, the syringes and connectors were cleaned with 2 % (*v/v*) Hellmanex-III solution followed by MilliQ-grade H_2O and filtered formulation buffer (Table 2.3).

For the extensional flow experiments, one syringe was filled with the 500 µL of 0.22 µm-filtered protein solution then secured with claps onto an optics board (Thor Labs). The second empty syringe was then connected to the clamped syringe via the capillary and also clamped, ensuring there was no air in the syringes. A stepper motor, controlled by a microcontroller, was used to shuttle the protein solution between the syringes, maintaining constant velocity and flow rate (values used in this study are listed in Table 2.5). The protein solution was removed after the desired number of passes. A protein solution, kept at an ambient temperature for the duration of a given stress experiment was used as a control.

Table 2.5 Plunger velocities in the device and the corresponding center-line strain rates.

Plunger Velocity (mm s^{-1})	Center-line strain rate (s^{-1})	Capillary wall shear rate (s^{-1})
8	11750	50375
16	23421	100751

2.2.7.2 Hydrodynamic force experimental workflow

Sample was made to the desired concentration for the experiments (0.25 mg mL^{-1} or 1 mg mL^{-1}) and kept on ice until use. 500 µL protein sample at the desired concentration were loaded into the syringes and shuttled for 10, 20, 50, 100, 200, 300 or 500 passes at 8 mm s^{-1} or 16 mm s^{-1}. Non-stressed on-bench control (500 µL sample in an Eppendorf) was placed at room temperature for the duration of the most amount of passes in the experiment (i.e. the length of time it takes for 500 passes). Resulting stressed sample was collected in a fresh Eppendorf. All samples were kept on ice before centrifugation and the set up was cleaned and re-run as in Section 2.2.7.1: Apparatus set up.

2.2.7.3 Monomer loss quantification

To determine monomer loss, three x 150 µL of quiescent and stressed protein samples (three tubes per sample of 0.5 mL) were loaded into Ultracentrifuge Tubes (UC tubes) (Beckmann Coulter). The samples were centrifuged at 30,000 rpm, 30 min, 4 ˚C using a TLA100 rotor (Beckmann Coulter) in a Optima TLX Ultracentrifuge (Beckmann Coulter). Following this, 100 µL supernatant was removed from each tube and pooled.

The A_{280} of WFL samples were measured by UV-Visible Spectroscopy, whereas the STT and 114 samples were sealed into HPLC polypropylene insert vials (0.3 mL) (VWR) and the HP-SEC trace at A_{280} over a 25 minute run at 0.5 mL min^{-1} was measured following the procedure in Section 2.2.3.2: High performance size-exclusion chromatography for monomer loss assay. Loss of monomer from solution as a result of squishing was quantified by the loss in absorbance (compared to the zero passes control) using UV-Visible Spectroscopy (for WFL) or HP-SEC for STT and 114 (see Section 2.2.3.2).

Exponential decay fitting was calculated in Origin Pro 2020 using the single-exponential decay function using the equation below:

Equation 2.5 Single exponential decay function for monomer loss

$$y = A_1 exp^{(-\frac{x}{t_1})} + y_0$$

Where t_1 = time constant (t). This in turn was used to calculate the half-time:

Equation 2.6 Calculation of the half-time from a single-exponential decay function

$$t_{1/2} = t_1 \, ln_2$$

2.2.8 Mass Spectrometry

2.2.8.1 Native MS

Samples were prepared by buffer exchange into 0.15 M Ammonium acetate buffer pH 6.0, using the dialysis procedure (Section 2.2.1.2: Dialysis). Final samples were prepared to a working concentration of 0.1 mg mL^{-1} for experiments, calculated based on the absorbance at 280 nm and the Beer-Lambert law, using respective extinction coefficients.

Electrospray borosilicate capillaries (1.5 mm OD x 0.78 mm ID, filamented) were pulled in-house using P-97 Micropipette Puller (Sutter Instrument Company) and coated with palladium using a sputter coater (Polaron SC7620). Samples were loaded into tips and ESI was initiated on the Q-Exactive UHMR (Thermo Scientific) by applying the following instrument parameters (further tune file parameters listed in Appendices 7.2 Related information for Materials and Methods).

Table 2.6 Instrument parameters for Native MS experiments. Extended example tune file and instrument parameters can be found in Appendices 7.2.

Instrument parameter	**Native MS conditions**
Cone voltage (kV)	1.10 – 1.50
Capillary Temperature (˚ C)	250.00
Sheath gas flow rate	0.00
Aux gas flow rate	0.05
Sweep gas flow rate	0.00
Aux temperature (˚ C)	0.85
Max spray current (A)	50.00
Probe heater temperature (˚ C)	350.00
S-lens RF level	200.00
Ion Source	NSI

Spectra were collected for 2 minutes and summed before being analysed using UniDec v5.1.1 (Marty *et al.*, 2015). All samples were analysed in positive ionisation mode.

2.2.8.2 LC-MS/MS using S-Traps

Sample at 0.08 mg mL^{-1} in formulation buffer was diluted 1:1 with 10% SDS (*w/v*). Sample was then reduced with DTT (final concentration 20 mM) for 10 minutes, 95 ˚C, shaking at 600 rpm. After this, sample was cooled to room temperature for 5 minutes and then alkylated with iodoacetamide (final concentration 40 mM) for 30 minutes, 20 ˚C in the dark, no shaking. Following this, sample was acidified with 12% phosphoric acid (*w/v*) (acid to sample ratio 1:9) and diluted 7:1 (*v/v*) with S-Trap binding buffer (detailed in Table 2.3).

Trypsin vials (Thermo Scientific) were reconstituted in 50 mM ammomium bicarbonate pH 7.4 buffer to 0.02 µg $µL^{-1}$ and then 1 µg was added to the sample. The following steps up to the addition of more trypsin were completed as rapidly as possible. Immediately, sample was added to S-Traps (200 µL at a time) and spun at 4000 x g, 30 seconds until excess liquid had flown through (flow through was discarded to waste), and sample addition was repeated until all sample had passed through the column. Quickly following this, the columns were washed with S-Trap binding buffer (130 µL at a time) and spun at 4000 x g, 30 seconds, for three repeats. Finally, 30 µL of 0.02 µg $µL^{-1}$ trypsin was added to the top of the S-Trap and pushed gently to soak the full column (taking care to not push air into the column), and immediately the S-Traps were loosely capped and incubated 1 hour, 47 ˚C, no shaking.

Peptides were eluted by spinning 4000 x g, 1 minute with 40 µL of Elution buffer 1 (Table 2.3), followed by 40 µL of Elution buffer 2 (Table 2.3), and finally by 35 µL of Elution buffer 3 (Table 2.3). The resulting volume of peptide suspension (approximately 80 µL) were split into two Low-Bind tubes (Eppendorf) and then evaporated in an EZ-2 Personal Evaporator (GeneVac) to approximately 5 µL to remove volatile acetonitrile. Peptides were reconstituted with 0.1% formic acid (*v/v*) to a final concentration of 0.5 µM by the addition of the solution to the low-bind tube and gentle shaking, 30 minutes. Following this, sample was transferred to LC-MS vials and stored in the Ultimate 3000 RSLCnano LC autosampler.

Tryptic peptides (3 µL at 0.5 µM) were injected onto the Ultimate 3000 RSLCnano LC system equipped with a 20 cm capillary emitter column C18 bead column (inner diameter 75 µm, packed with 3-µm Reprosil-Pur 120 C18 media, Dr. Maisch) prepared in-house, where peptides were separated using reverse-phase chromatography with a two-section linear gradient of 2 - 30% and 30 - 60% (*v/v*) MeCN in H_2O (both containing 0.1% (*v/v*) formic acid) over a 90 minute range at 0.5 µL min^{-1} (total run time 120 minutes). Peptides were analysed by ESI-MS/MS through direct infusion onto an Orbitrap Exploris 240 (ThermoFisher Scientific) operating in Orbitrap mode (60,000 resolution, rapid mode). Data dependent acquisition (DDA) was

performed in the Orbitrap (MS/MS mode was Orbitrap) where HCD fragmentation was employed to sequence the peptides. TopN = 20 scans, maximum injection time = 200 ms, apex detection = on, dynamic exclusion = exclude after 1 times, for 3 seconds.

2.2.8.3 Fast Photochemical Oxidation of Proteins – LC-MS/MS

Samples were prepared to a final concentration of 0.1 mg mL^{-1} (6.7 µM) in formulation buffer plus scavenger solution. L-Histidine in formulation buffer is used as a scavenger amino acid and is prepared to a final concentration of 10 mM in the sample mixture. A control sample was prepared identically and placed on the bench for the duration of the reaction, but was not exposed to the laser. Immediately before irradiation, 3 µL of 5% (*v/v*) hydrogen peroxide is added to 100 µL protein solution (and the control solution) before loading into a Hamilton syringe, to give a final H_2O_2 concentration of 0.15% (*v/v*). Sample was passed at a flow rate of 20 µL min^{-1} through a fused silica capillary (internal diameter 100 µm). The protein mixture is passed through the UV irradiation path (beam width 3mm) of a COMPexPRO 50F Krypton Fluoride Excimer Laser (Coherent Inc., Ely, UK) operating at 248 nm to trigger decay of H_2O_2 into ·OH radicals. The firing frequency was 15 Hz and the pulse duration was 20 nanoseconds. Laser power was kept constant at 110 mJ. Labelled protein sample was collected into an Eppendorf containing 20 µL quench solution of 100 mM L-methionine and 1 µM catalase in formulation buffer, mixed by flicking, then placed immediately on ice. 20 µL quench solution was also added to the control sample, mixed by flicking, and also placed immediately on ice.

Following UV irradiation and quenching, final concentration of the sample was 0.08 mg mL^{-1} and sample was reduced, alkylated, digested and ran on the LC-MS/MS procedure as described in Section 2.2.8.2: LC-MS/MS using S-Traps.

2.2.8.4 Flow – Fast Photochemical Oxidation of Proteins – LC-MS/MS

Samples were stressed at 1 mg mL^{-1} as in Section 2.2.7.2: Hydrodynamic force experimental workflow. Immediately after the hydrodynamic force experiment (or for the corresponding amount of time for the control sample without hydrodynamic force), samples were diluted. The final solution was 1:1:8 of sample:scavenger:formulation buffer, so that the final concentration of protein was 0.1 mg mL^{-1} (6.7 µM) and final L-histidine concentration was 10 mM. This solution was stored on ice and immediately irradiated as in Section 2.2.8.3: Fast Photochemical Oxidation

of Proteins – LC-MS/MS (the dead time between the final hydrodynamic force event and irradiation was consistently 5 minutes).

2.2.8.5 FPOP data processing and analysis

Peptide raw data was searched with PEAKS Studio v10.6 against the FASTAs for the HC and LC of the corresponding mAb. Trypsin cleavage was specified as semi-selective (allowing for up to a maximum of 3 missed cleavages). The peptides with good MS/MS data identified in PEAKS was used to search for peptides in the raw Thermo data files using FreeStyle.

Variable mass additions of +16 Da, +32 Da and +14 Da were searched to identify FPOP oxidations (carbamidomethylation modifications were searched where applicable (+73 Da and +89 Da being singly and doubly oxidised cys side chains)). Data were quantified at the peptide level using FreeStyle (part of Xcalibur software (v4.0.27.19)) by integrating peaks in the extracted ion chromatograms (XICs) of each peptide ion for each identifiable charge state, for the modified and unmodified versions of each peptide. Additionally, some MS/MS data were manually curated at a residue level in order to identify and assign FPOP modifications with confidence. The equation used to quantify FPOP modifications is below (Equation 2.7):

Equation 2.7 Quantifying FPOP oxidations.

The modified peak of interest was generated from XIC is quantified against the ion counts of all other identifiable versions of the peptide, both unmodified and modified.

$$\%\, Modified = \frac{\sum Modified_{Peak}}{Unmodified + \sum Modified_{All\ peaks}}$$

2.2.8.6 Hydrogen-Deuterium Exchange – LC-MS/MS

HDX-MS experiments were carried out by Dr Piera Marchetti (AstraZeneca) using an automated sample handling robot (LEAP technologies) coupled to an M-Class Acquity LC system and HDX manager (Waters Ltd). WFL, STT and 114 samples were prepared to 1 mg mL^{-1} in formulation buffer, and placed in vials in the LEAP HDX platform robot (Trajan Scientific and Medical) at 1 ˚C. Samples were diluted 1:9 with deuterated formulation buffer (20 mM sodium succinate, 125 mM L-arginine, pD 6.0) to a concentration of 0.1 mg mL^{-1} and incubated for 30 seconds, 60 seconds, 300 seconds and 1800 seconds, after which the labelled solution was immediately quenched by dilution 1:1 into quench buffer (100 mM potassium phosphate, 8 M Guanidine HCl, pH 2.3) at 1 ˚C. This was immediately diluted 1:1 with 100 mM potassium phosphate, pH 2.3,

giving a final quench pH ~2.5 and a final guanidine hydrochloride concentration of 2 M. 50 µl of quenched sample was passed through an immobilised ethylene bridged hybrid (BEH) pepsin Enzymate column (Waters Ltd.) at 20 ˚C at a flow rate of 400 µL min^{-1}. Peptides were trapped using a VanGuard Acquity UPLC BEH C18 pre-column trap column (Waters Ltd.) and then transferred to a C18 column (Waters Ltd.). Peptides were separated by a gradient elution of 0-35% (*v/v*) MeCN, 0.1% (*v/v*) formic acid in H_2O over 6 minutes at 35 µL min^{-1}, then 35-40% (*v/v*) MeCN, 0.1% (*v/v*) formic acid in H_2O over 1 minute at 35 µL min^{-1}. Peptides were analysed using a Orbitrap Fusion Lumos Tribrid (Thermo Scientific) operating in DDA mode for the peptide map, and CID fragmentation for HDX data.

2.2.8.7 HDX data processing

HDX data were processed using BioPharma Finder (Thermo Scientific) and HDExaminer (Sierra Analytics). HDX data were visualised by exporting PyMOL scripts from HDExaminer and mapped onto the mAb structures (Section 2.2.2.1: In silico modelling).

Chapter 3

Biophysical characterisation of monoclonal antibodies

3 Biophysical characterisation of model monoclonal antibodies

3.1 Objectives

The aim of this thesis was to investigate how the structure of biotherapeutic monoclonal antibodies (mAbs) are perturbed by hydrodynamic forces. Previously, two closely related mAbs were compared by a variety of characterisation techniques including HP-SEC and AUC, and one mAb (WFL) was shown to be highly aggregation prone and one was resistant to aggregation (STT) (Section 1.3.1: WFL and STT: model monoclonal antibodies). In this chapter, a wealth of complementary biophysical characterisation techniques were applied to this pair of mAbs, alongside a third variant (114) which was isolated by screening a variant library derived from WFL (Section 1.3.2: Variant 114). This third variant, which retained WFL residues but had four other mutations in the variable domains, demonstrated the potential to have improved biophysical properties through its identification in the TPBLA screen by scoring more highly than both WFL and STT. From this, it was hypothesised that variant 114 could provide insights into how its mutations might mask the negative influence of a hydrophobic surface patch in WFL. It was also hypothesised that hydrodynamic flow conditions could be chosen to perturb the native state of the three mAbs to a flow-induced aggregation state, to allow the use of labelling methods to visualise any meta-stable state. This may provide a crucial structural snapshot for yielding insights into an unfolding pathway leading to aggregation.

This chapter will explore how the variant 114 compares to WFL and STT in a range of biochemical techniques. Techniques commonly used in industry to assess developability (the selection of candidate molecules with improved biophysical characteristics) are used. Observations of how the peripheral mutations in 114 might influence the impact of the WFL residues are discussed in this chapter, and will aim to inform the interpretation of any protein structural changes that may be identified by labelling-mass spectrometry in response to mechanical flow stresses in the subsequent chapters.

3.2 Characterising monoclonal antibody monomers

There are a variety of techniques employed by industry during biopharmaceutical development and manufacture to monitor what species may be present in a sample. Comprehensive analysis of molecular properties of variants is important for prioritising the most promising candidates for development, without exhibiting undesirable characteristics, such as low solubility, bad colloidal stability or aggregation (Jain *et al.*, 2017). A biophysical understanding of monomer

behaviour using chromatographic techniques is invaluable for comparing variants and will form the basis of this section.

3.2.1 High Performance-Size Exclusion Chromatography (HP-SEC)

Previously, HP-SEC was carried out on WFL (MEDI-1912), compared its parent mAb, MEDI-578 (containing S, T and T residues), in 0.1 M sodium phosphate, 0.1 M sodium sulphate, pH 6.8 (C. L. Dobson *et al.*, 2016). In these experiments, the samples were analysed at 1 mg mL^{-1}, and the absorbance at 280 nm (mAu) was measured. WFL exhibited a late retention time with a broad asymmetric peak shape, indicative of interactions with the gel matrix in the column, whereas MEDI-578 demonstrated an elution time and profile expected for a typical monoclonal antibody (C. L. Dobson *et al.*, 2016). Therefore, design of the triple mutant MEDI-1912_STT (STT herein), with sequence identical to WFL but with W30S, F31T and L57T mutations, allowed the investigation of the influence of the three substitutions on the mAb's biophysical behaviour. The Aggrescan 3D soluble surface area predictions, and the original data from Dobson *et al.* found in Section 1.3.1: Figure 1.17, is displayed in the figure below again for the reader's ease (Figure 3.1).

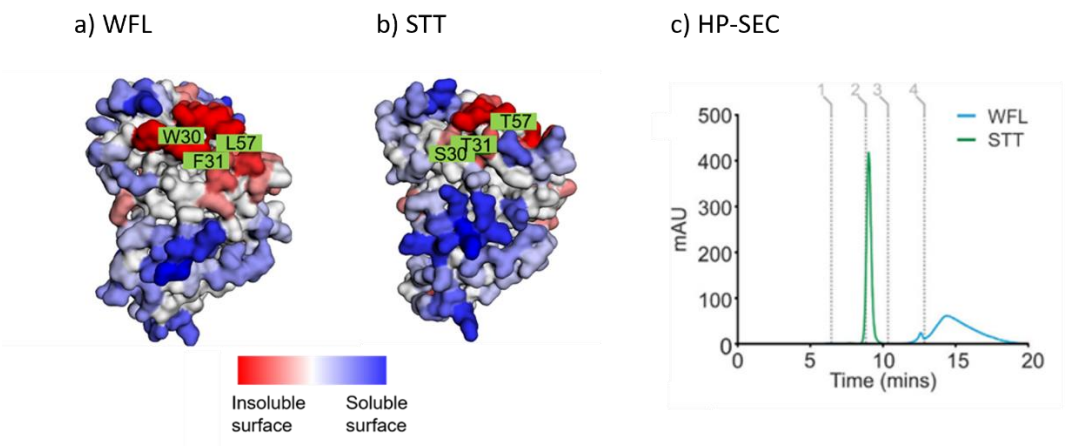


Figure 3.1 WFL and STT are structurally similar but demonstrate different HP-SEC elution profiles.

a) Aggrescan3D analysis of WFL_scFv reveals a surface exposed hydrophobic patch (red). Residues W30, F31 and L57 are highlighted (green = mutations in the heavy chain). b) Aggrescan3D analysis of STT, with the mutated residues W30S, F31T and L57T highlighted (green = mutations in the heavy chain). c) High performance size exclusion chromatography (HP-SEC) elution profiles of WFL (green) and STT (blue) at 280 nm (mAU), 1 mg mL^{-1} in 0.1 M sodium phosphate, 0.1 M sodium sulphate, pH 6.8. Grey lines indicate elution times of calibrant proteins: 1, Thyroglobulin (670 kDa); 2, IgG (158 kDa); 3, Ovalbumin (44 kDa) and 4, Vitamin B12 (1.35 kDa).

Subsequent purifications of WFL and STT protein were formulated in a storage buffer containing L-arginine, an amino acid known to mitigate hydrophobic interactions (Baynes *et al.*, 2005; N. A. Kim *et al.*, 2016). The storage buffer, known as formulation buffer herein, is 20 mM sodium succinate, 125 mM L-arginine, pH 6.0 (Table 2.3 Buffers used in this study).

In this thesis, WFL, STT and 114 were expressed, then dialysed into and stored in formulation buffer. 114 and STT were analysed using HP-SEC in formulation buffer using TSKgel GSK3000XL HPLC column (TOSOH Bioscience) (the same HP-SEC column matrix used previously (C. L. Dobson *et al.*, 2016)) and the intensity of absorbance at 280 nm (mAu) were compared (Figure 3.2). The column was calibrated with a mixture of globular proteins to verify the elution of the mAbs at a time corresponding to an approximate mass of 150 kDa (shown by the grey dashed lines overlayed in Figure 3.2). As WFL demonstrated an atypical elution in previous experiments (Figure 3.1 c), it was not repeated for this thesis.

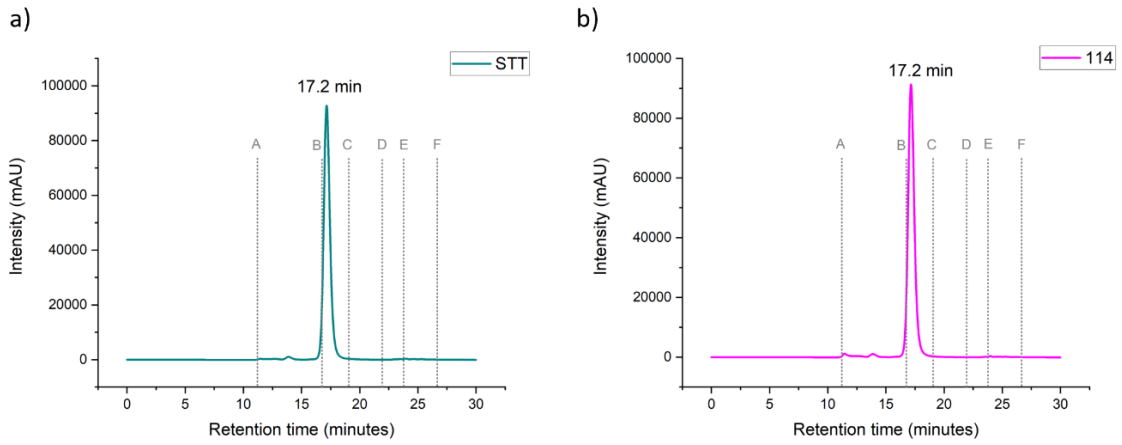


Figure 3.2 HP-SEC traces for STT and 114.

20 µL of 1 mg mL^{-1} sample was injected onto a TSKgel GSK3000XL HPLC column at a flow rate of 0.5 mL min^{-1} for 30 minutes. a) STT (teal) elutes at a maximum at 17.2 minutes. b) 114 (pink) elutes at a maximum at 17.2 minutes. Data representative of n=3 technical repeats. Grey lines = column calibration protein elution times: A, blue dextran (void volume); B, alcohol dehydrogenase tetramer (145 kDa); C, alcohol dehydrogenase dimer (82 kDa); D, ovalbumin (43 kDa); E, cytochrome C (12 kDa); F, Vitamin B12 (1.35 kDa).

The results of the HP-SEC experiments demonstrated ideal elution profiles at 17.2 min for both STT and 114 (Figure 3.2), which line up with an expected mass of approximately 150 kDa for a

mAb. The elution profiles are symmetrical demonstrating no interaction with the column matrix. Overall, 114 appears to not interact with the column under these conditions. The presence of the mutations (two variable heavy chain and two variable light chain) appear to mitigate the influence that the WFL residues have to cause mAb interaction to the column.

To further understand WFL's aberrant behaviour in comparison to STT and 114, orthogonal chromatography techniques are employed to comment on any hydrophobic interaction character that the proteins may have.

3.2.2 Hydrophobic Interaction Chromatography (HIC)

As WFL was previously shown to interact with the TSKgel GSK3000XL HPLC column to a much higher degree than STT, it was postulated that the hydrophobic group of residues surrounding and including WFL in the variable regions of the mAb influence the migration of the molecule, due to the sample interacting with the column matrix. To investigate this, hydrophobic interaction chromatography was carried out on the samples where sample was diluted 1:1 with 1.8 M ammonium sulfate, 100 mM sodium phosphate, pH 6.5, and 5 µg was loaded onto the Proteomix HIC Butyl-NP5 non-porous column (Sepax). Ammonium sulfate is used in the running buffer to promote stationary phase-protein interactions (Fekete, Veuthey, *et al.*, 2016). The sample was eluted with an inverse salt gradient of 0-100% with 100 mM sodium phosphate pH 6.5 (Figure 3.3). Absorbance at 280 nm (mAU) is measured. The results have not been normalised to the highest elution peak to compare retention to the column. Elution later in the run, at a lower salt concentration, indicates the protein is more hydrophobic and has a stronger interaction with the matrix.

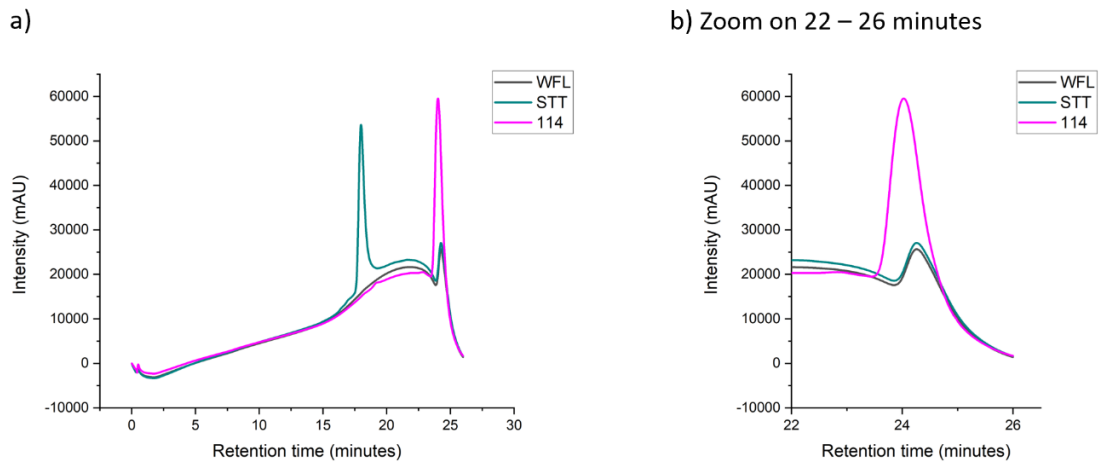


Figure 3.3 HIC trace of WFL, STT and 114.

5 µg of sample was eluted from a HIC Butyl-NP5 non-porous column (Sepax) over 26 minutes. WFL (black) is compared to STT (teal) and 114 (pink). a) Trace from 0 – 26 minutes. b) Trace from 22 – 26 minutes.

STT (teal) eluted at approximately 17 minutes, whereas 114 (pink) eluted at 24 minutes. This may be due to the presence of the W and F residues in the variable regions increasing the degree of interaction of 114 compared to STT. WFL (black) does not appear to elute from the column. There is a peak for each sample at 24.5 minutes (in 114, this is obscured by the mAb monomer elution peak), which may correspond to highly hydrophobic impurity. The peaks for STT and 114 remained broadly symmetrical.

This chromatography method demonstrates a much greater discrepancy between the retention time of STT and 114 compared to HP-SEC (Figure 3.2). Here, 114's global hydrophobicity is greater than STT. It is worth noting this chromatography technique cannot comment on which residues or domains are interacting with the column. However, the elution of 114 with lower salt % in the running buffer indicates a higher surface hydrophobicity than STT. WFL appears to not elute at low salt conditions.

For insights into whether the hydrophobic surface patches on the proteins contribute to their colloidal stability and self-association, stand-up monolayer chromatography (SMAC) was employed.

3.2.3 Stand-up Monolayer Adsorption Chromatography

SMAC was used to address self-interaction between the sample. A Zenix SEC-300 column was used for the experiments, where the resin consists of a hydrophobic monolayer with terminal hydrophilic groups to mimic the exterior of a protein. Retention times of antibodies have been found to be inversely related to colloidal stability, which result in longer retention for samples more prone to aggregation or precipitation (Kohli 2015). Here, 2 µg sample was loaded onto the column using a mobile phase of 150 mM sodium phosphate pH 7.0 (Figure 3.4).

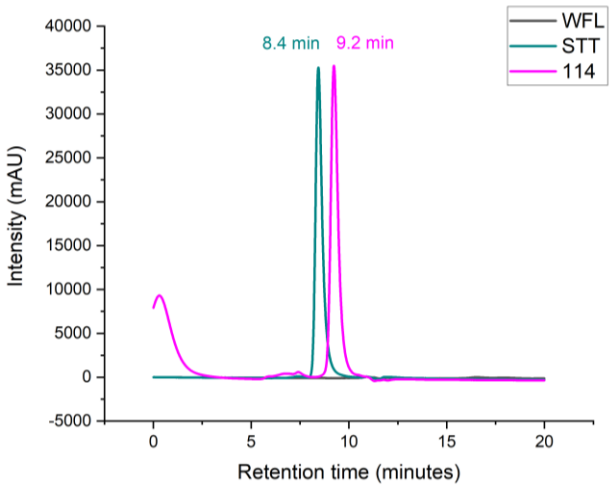


Figure 3.4 SMAC traces for WFL, STT and 114.

STT (teal) elutes at 8.4 minutes and 114 (pink) elutes at 9.2 minutes. WFL (black) is featureless, indicating the sample does not elute in this time scale.

WFL does not elute from the SMAC column in this time range as there is no peak visible in the trace (Figure 3.4). STT elutes at the expected retention time for monoclonal antibodies, whereas 114 has a slightly longer retention time with a slightly broader trace at the base of the peak, indicating interaction on the column. The protein-mimicking resin retains WFL to a much higher extent than 114 and STT, indicating WFL is sticky to its counterparts through hydrophobic and hydrophilic interactions; 114 is marginally influenced by the hydrophilic character of the column which competes for interaction with the protein over its self-interaction. Additionally, there is less discrepancy in the retention time between STT and 114 in SMAC, compared to the retention time observed using HIC (comparing Figure 3.3 and Figure 3.4). The stronger interaction with a hydrophobic surface in HIC, in comparison to one which is mixed with hydrophobic and

hydrophilic patches, further supports 114 being less self-association- and aggregation-prone than WFL, and closer in properties to STT.

To further understand colloidal self-interaction propensity of the proteins, techniques to capture oligomeric state were performed.

3.3 Characterising monoclonal antibody oligomers

Antibodies are successful as biopharmaceuticals as they can be engineered to bind selectively to targets through their CDRs, which is important for low polyspecificity and low immunogenicity (Jain *et al.*, 2017). Binding of the mAb through its CDRs or any other part of the structure to itself or other non-specific targets is undesirable for efficacy, immunogenicity and manufacture. This next section focusses on how WFL, STT and 114 perform in some biophysical techniques that can be employed to monitor oligomeric state.

3.3.1 Affinity-Capture Self-Interaction Nanoparticle Spectroscopy

The affinity-capture nanoparticle assay measures colloidal self-interactions, and has been widely used in biopharmaceutical development for screening antibody self-association. The process uses capture antibody-coated gold nanoparticles, which remain dispersed in solution or are pulled closer together in solution, based on the amount of interaction between test IgGs (Figure 3.5). The absorbance from 400 nm to 700 nm is measured and the plasmon wavelength plotted, where a higher nm shift in wavelength indicates the molecules have a higher propensity for each other.

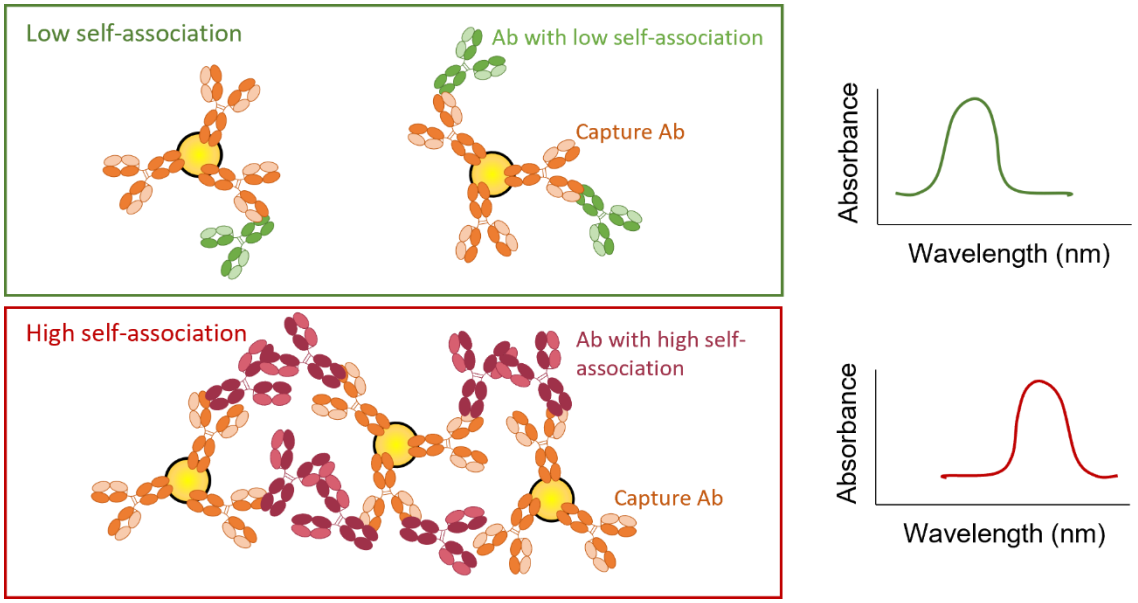


Figure 3.5 Schematic of AC-SINS.

Gold nanoparticles coated with the capture antibody (orange) bind test IgGs (green and pink). If the test IgGs have low self-interaction, the particles remain dispersed in solution resulting in a lower measured plasmon wavelength. When test IgGs have higher propensity to aggregate, particles are pulled closer together resulting in a higher plasmon wavelength (red shifted).

Here, WFL, STT and 114 were incubated with gold nanoparticles in histidine-sucrose-arginine buffer (20 mM L-histidine, 120 mM sucrose, 80 mM L-arginine, pH 6), which is widely used in industry as a model buffer for AC-SINS experimentation. The mAbs were assessed for self-association by observing the red shift compared to free nanoparticles in buffer (Figure 3.6). The summarised red shift is presented in bar chart format (Figure 3.7). Experiments were performed by Dr Janet Saunders (AstraZeneca).

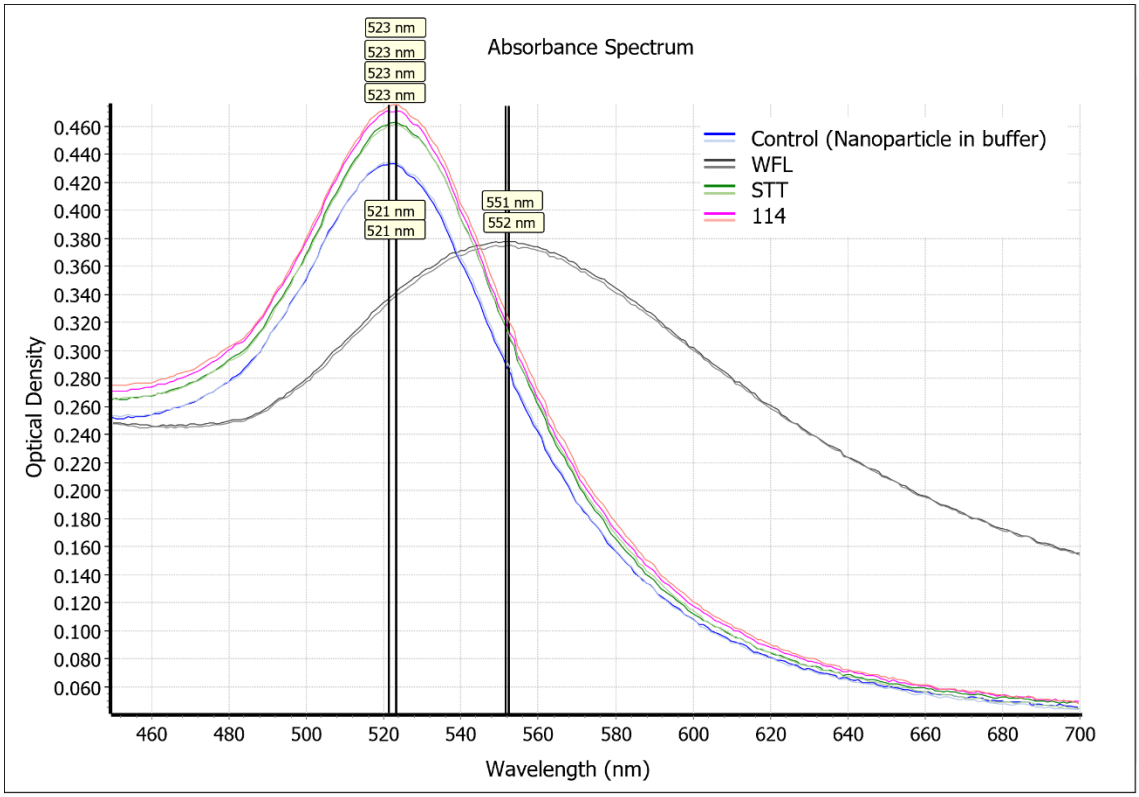


Figure 3.6 AC-SINS plasmon wavelength red shift absorbance spectrum data for WFL, STT and 114.

Optical density is plotted against wavelength (nm). All data collected in 20 mM L-histidine, 120 mM sucrose, 80 mM L-arginine, pH 6 buffer in duplicate. Data in blue = control (free nanoparticles in buffer, n=2); grey = WFL (n=2); green = STT (n=2); pink = 114 (n=2). Data collected by Dr Janet Saunders, AZ.

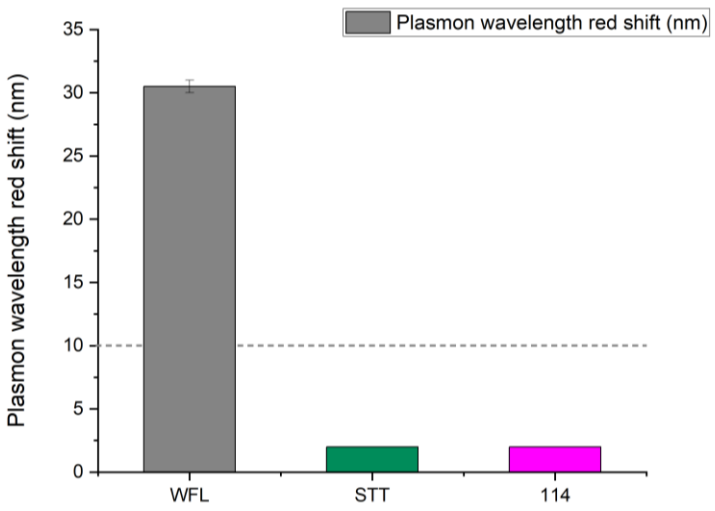


Figure 3.7 AC-SINS plasmon wavelength red shift for WFL, STT and 114.

Data in grey = WFL; teal = STT; pink = 114. Error bars show the standard deviation from n = 2. The maximum red shift limit for developability specified by AstraZeneca is 10 nm (grey dashed line). Data collected by Dr Janet Saunders, AZ.

WFL had the largest red shift (30.5 nm ± 0.5 nm) compared to STT (2.0 nm ± 0.0 nm) and 114 (2.0 nm ± 0.0 nm). WFL's wavelength shift is over the guideline threshold for classifying a mAb as low risk in development (grey line at 10 nm), supporting the poor biophysical properties of the WFL patch which may contribute to self-association and aggregation. STT and 114 did not demonstrate any self-associating character, where both mAbs had a similar red shift pattern. These data support the improved properties of 114 over WFL, even with the presence of W, F and L residues, indicating that the four other mutations in 114 must contribute to shielding the mAb from self-association. The results confirm that 114 has a reduced self-association character compared to WFL, like STT, despite 114 retaining the W, F and L residues.

As the formulation buffer of these mAbs contains arginine, it was important to observe the size distribution of these proteins in their formulation buffer natively. Previously, AUC was performed with WFL and STT at 1 mg mL^{-1} (Figure 1.17, (C. L. Dobson *et al.*, 2016)), where WFL is predominantly dimeric (>75% dimer) and STT is monomeric. Orthogonal techniques to observe the distribution of particle size – DLS and AF4 – were used in this thesis to study the oligomeric state of the mAbs in solution. The behaviour of the mAbs in a solution other than the formulation buffer was measured for comparison.

3.3.2 Ultra-low-volume dynamic light scattering (DLS)

Sample was prepared to 1 mg mL^{-1} in formulation buffer and in 150 mM ammonium acetate, pH 6.0. Approximately 2 µL of sample was drawn into glass capillary cuvettes through capillary action and light scattering was measured (Figure 3.8 (formulation buffer) and Figure 3.9 (150 mM ammonium acetate)). Work was performed in collaboration with Dr Juhi Patel (AstraZeneca).

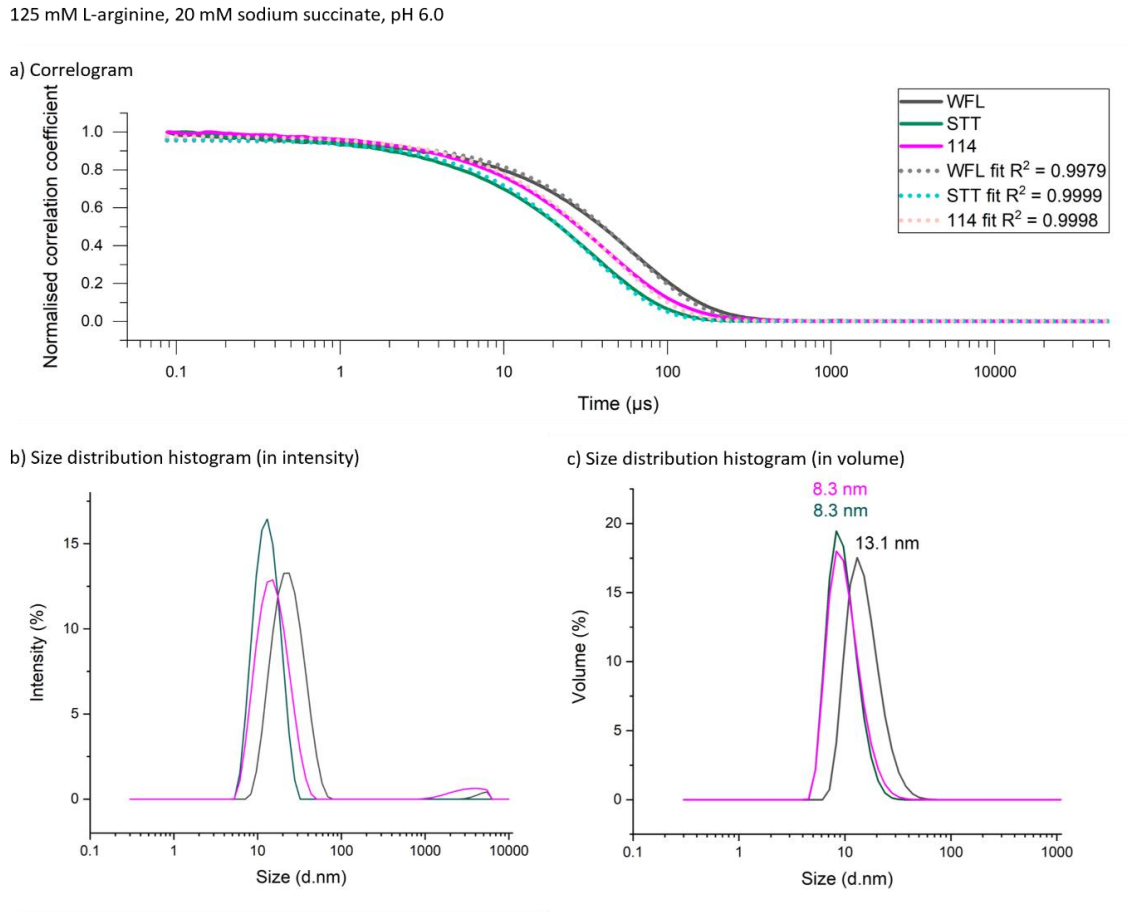


Figure 3.8 Ultra-low-volume DLS for WFL, STT and 114 in formulation buffer.

Correlograms (a) and the corresponding size distribution histograms in intensity (b) and volume (c). The R^2 values shown are for the goodness of fit for a single-exponential decay to the data. The R^2 values shown are for the goodness of fit for a single-exponential decay to the data. Data for each of the results are representative of n = 3 technical replicates. Data collected by the PhD candidate.

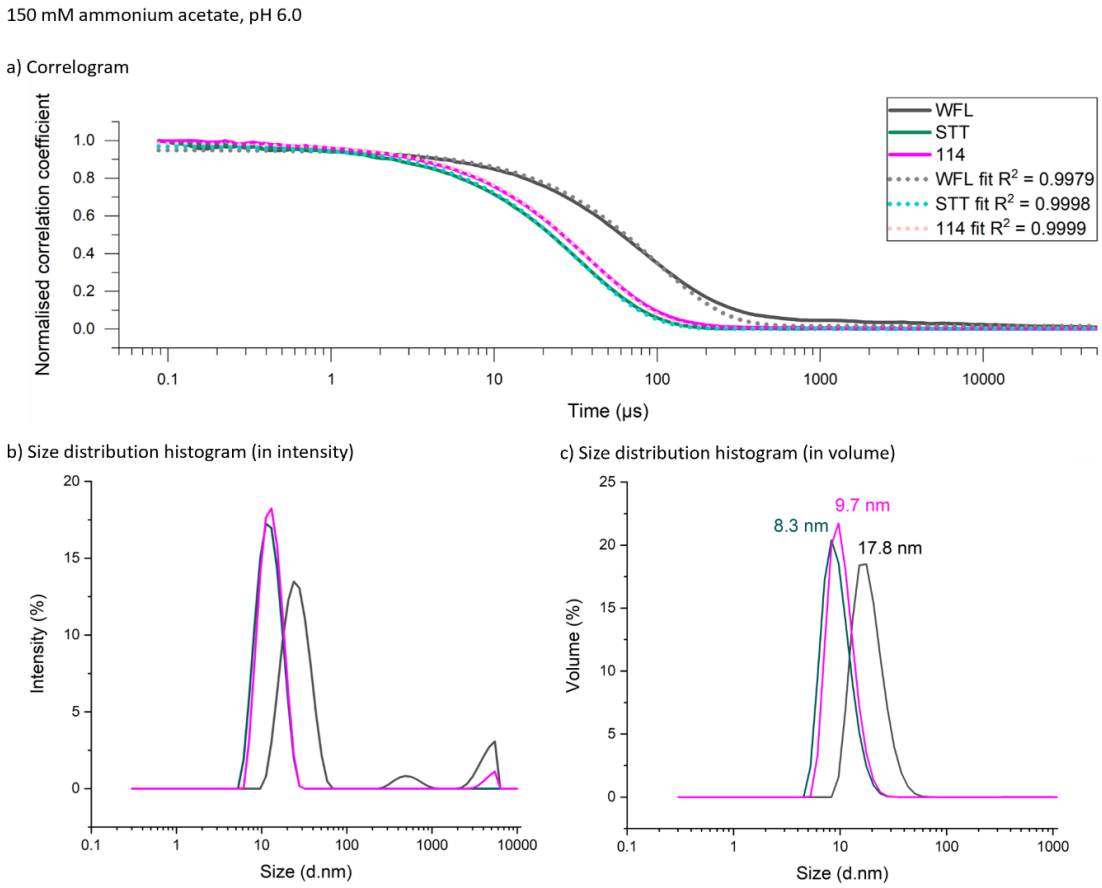


Figure 3.9 Ultra-low-volume DLS for WFL, STT and 114 in 150 mM ammonium acetate pH 6.0 solution.

Correlograms (a) and the corresponding size distribution histograms in intensity (b) and volume (c). The R^2 values shown are for the goodness of fit for a single-exponential decay to the data. The R^2 values shown are for the goodness of fit for a single-exponential decay to the data. Data for each of the results are representative of n = 3 technical replicates. Data collected by Dr Juhi Patel (AstraZeneca).

The correlograms (Figure 3.8 a and Figure 3.9 a) comment on the decay of signal over time relative to the original signal at time 0. The correlogram is statistical measure of the degree of non-randomness in an apparently random data set, when applied to a time dependent intensity trace. The larger the particles are in the system, the more slowly they diffuse, meaning it will take longer for the complete decay of the signal. This is supported with WFL being more prone than 114 and STT to associating into larger structures in both formulations, as both correlograms for WFL decay after a greater time than STT and 114.

The intensity output represents the proportion of molecules in the solution based on their volume assuming the volume of a sphere. From this representation in both buffers, STT does not have any secondary histogram distributions with a diameter in the 1000-10000 nm range,

whereas both WFL and 114 do (Figure 3.8 b and Figure 3.9 b). In ammonium acetate solution, WFL has three histogram distributions indicating it has a greater propensity to self-associate in this solution.

The intensity output is used to calculate a volume (%) histogram, based on the knowledge that larger particles scatter more light than smaller particles proportional to the sixth power of its diameter i.e. Intensity I α r^6 (from Rayleigh's approximation) (Filipe *et al.*, 2010; Minton, 2016). WFL exhibits a broader volume histogram than STT and 114 in both formulations (Figure 3.8 c and Figure 3.9 c). The average diameter (nm) of WFL is larger than STT and 114, where the difference to STT in formulation buffer is 13.1 nm compared to 8.3 nm (Figure 3.8 c, but in 150 mM ammonium acetate pH 6.0, WFL's average particle diameter increases to 17.8 nm (Figure 3.9 c). This indicates that in both buffers, WFL exists as a heterogeneous mixture at 1 mg mL^{-1}, where the mAb associates into oligomers, contributing to the observed average diameter being greater than monomer. In ammonium acetate, WFL has a higher propensity to exist in larger self-associations than in formulation buffer, indicating that the formulation buffer does act to protect WFL self-association somewhat. STT appears to have the same average nm diameter independent of the formulation, whereas 114 exhibits a shift to greater average diameter in ammonium acetate (9.7 nm average compared to 8.3 nm average for STT). This indicates that 114 is sensitive to some self-association depending on its solution components at a concentration of 1 mg mL^{-1}. Given more time in this thesis, it would have been valuable to repeat at a range of concentrations to separate insoluble dimer formation from any self-association.

3.4 Interrogating monoclonal antibody stability

The mAbs in this study have so far been characterised in monomeric form and how these mAbs may exist as oligomers. This thesis aims to shed light on how monomers are predisposed to form higher-order species through subtle sequence-based differences. Therefore, this next section focuses on ways that the monomeric state may be stressed, and how their intrinsic stabilities could be compared through thermal and mechanical stress.

3.4.1 Assessing thermodynamic stability using differential scanning fluorimetry

The three mAbs were examined for changes in thermal stability based on their subtle sequence changes. Differential scanning fluorimetry (DSF) was performed using the SUPR-DSF (Protein Stable), which observed changes in absorbance by epi-fluorescence. WFL, STT and 114 were subjected to temperature unfolding by ramping from 30 ˚C to 100 ˚C, and the thermal shifts of intrinsic tryptophan fluorescence in formulation buffer and 150 mM ammonium acetate solution were measured, by excitation at 280 nm and measuring the fluorescence emission at 310 – 420 nm, using SUPR-DSF (Protein Stable) (Figure 3.10).

The change in tryptophan emission spectra was monitored, from 350 nm (more folded) to 330 nm (more unfolded) through previously internally buried tryptophans becoming more exposed to the solvent. By plotting the change in this ratio (first derivative) versus temperature, the apparent temperatures of unfolding ($T_{m,app}$) were be extracted. The higher the T_m indicates higher thermal stability. This technique was employed to measure the transition temperatures corresponding to the unfolding of the C_H2 domain (T_m1) and the Fab (T_m2) (Ionescu *et al.*, 2008; Menzen & Friess, 2013). The melting temperatures ($T_{m,app}1$ and $T_{m,app}2$) were extracted from the first derivative of the 350/330 nm ratio plots (Figure 3.11). The experiments were performed at a concentration of 1 mg mL^{-1}, where previous studies with AUC (Figure 1.17, (C. L. Dobson *et al.*, 2016)) and DLS (Figure 3.8 and Figure 3.9) proved STT and to be mainly monomeric and WFL was mainly dimeric.

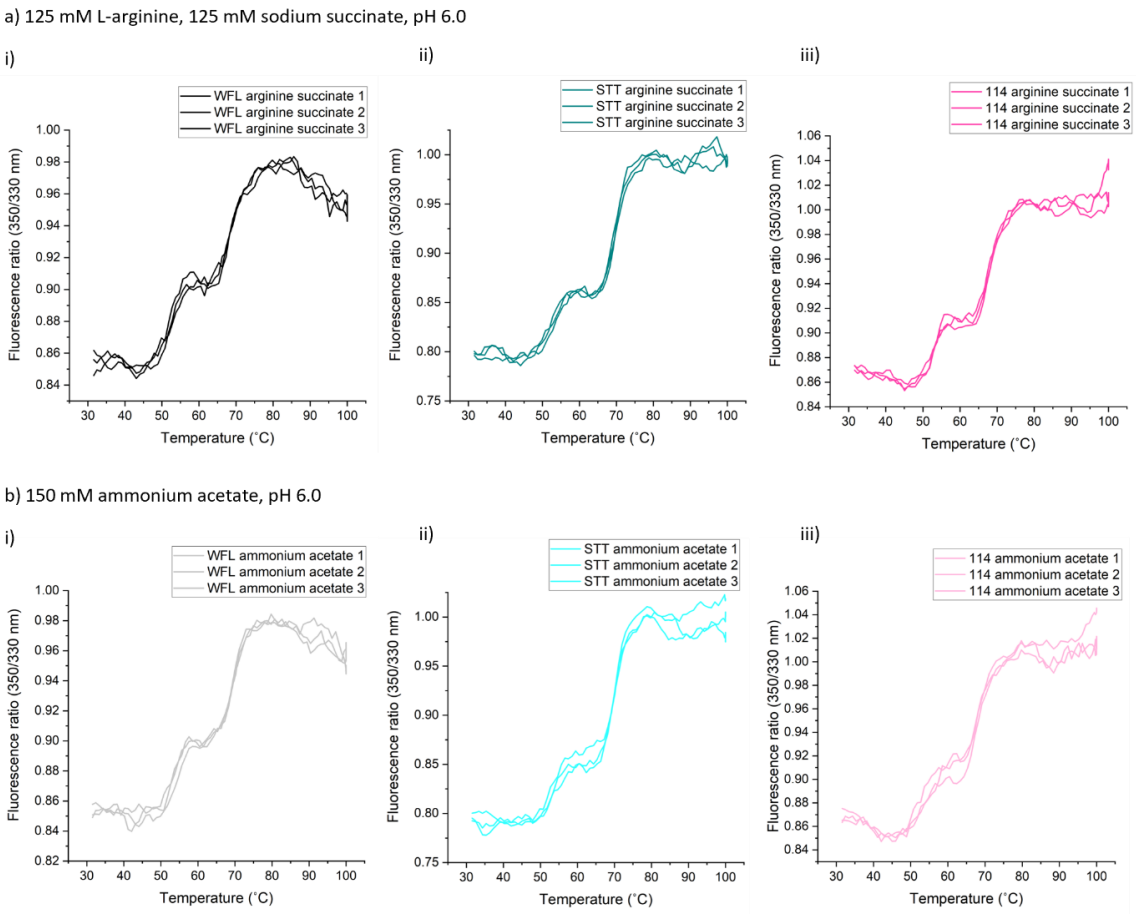
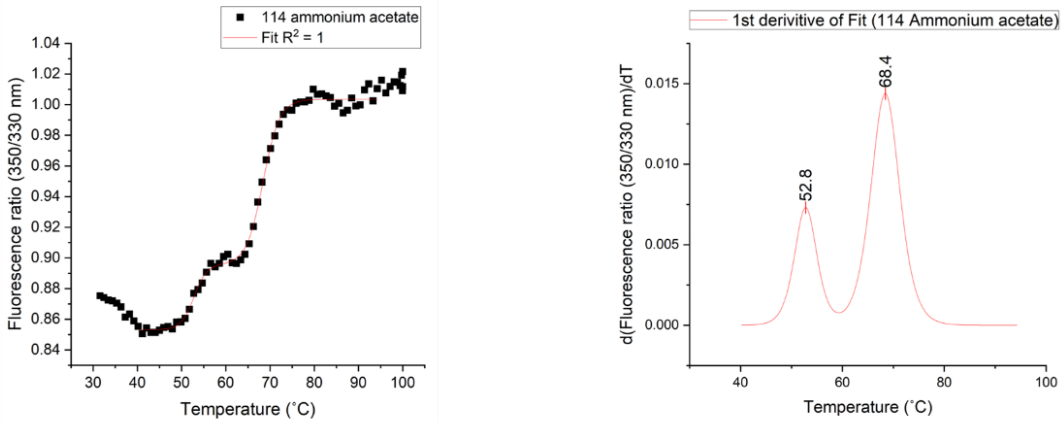


Figure 3.10 Differential scanning fluorimetry thermal melt ratio (350/330 nm) measurements for WFL, STT and 114 in two solution conditions measured by epifluorescence.

a) Thermal unfolding of WFL (i), STT (ii) and 114 (iii) at 1 mg mL^{-1} in formulation buffer. b) Thermal unfolding of WFL (i), STT (ii) and 114 (iii) at 1 mg mL^{-1} in ammonium acetate solution. Samples were heated between 30 and 100 ˚C and excited with 280 nm light. Emission spectra was measured between 310 and 420 nm. Displayed fluorescence ratio has been normalised to buffer-only control and smoothed by calculating a rolling average of ratio over a range of 5 ˚C.

a) Typical fit to the thermal unfolding data and the corresponding 1st derivative graph



b)

i)

Sample (in **arginine succinate**)	$T_{m,app}1$ (˚C)	$T_{m,app}2$ (˚C)
WFL	51.8 ± 0.4	69.1 ± 0.2
STT	52.3 ± 0.9	70.0 ± 0.2
114	52.5 ± 0.5	68.4 ± 0.2

ii)

Sample (in **ammonium acetate**)	$T_{m,app}1$ (˚C)	$T_{m,app}2$ (˚C)
WFL	53.6 ± 0.6	69.5 ± 0.1
STT	53.5 ± 0.4	70.2 ± 0.3
114	53.0 ± 0.9	68.7 ± 0.4

Figure 3.11 Extracted apparent thermal stability values for WFL, STT and 114.

a) A representative thermal unfolding shift showing the fit calculated with the double Boltzmann equation and iterated using the orthogonal distance regression analysis function in Origin Pro 2020, and the corresponding first derivative graph where the peak corresponds to the temperature of unfolding (in ˚C). b) $T_{m,app}$ values for samples in formulation buffer arginine succinate (i) and ammonium acetate solution (ii).

The transition temperatures of the three antibodies are comparable to themselves in both buffers, with the $T_{m,app}1$ in formulation buffer 1 ˚C lower than in ammonium acetate, whilst the $T_{m,app}2$ is unaffected by buffer. Overall, compared to each other, the results suggest the mAbs have similar stabilities of their C_H2 and Fab domains. The third transition (of the C_H3 domain) is either not visible in this temperature range or it is happening concurrently with the Fab transition (Menzen & Friess, 2013). The variant STT is more stable than WFL in both buffers, indicated by the higher melting temperature needed. 114 has the least stable Fab in both buffers (lowest $T_{m,app}2$ in Figure 3.11 b).

Samples in the arginine-containing buffer appear to be more prone to unfolding at lower temperatures. This is seen by a 1 ˚C drop in melting temperature observed, from three technical repeats. Previously it has been shown that arginine does not lead to stabilisation of globular proteins under thermal stress, but can decrease the melting temperature of RNase A and lysozyme (Arakawa & Tsumoto, 2003). Additionally, Arakawa and Tsumoto suggest that arginine can suppress aggregation by solubilising the unfolded state.

To probe the mechanical stability of the mAbs, and how this may be affected by the formulation the proteins are in, the samples were exposed to hydrodynamic stress in both formulation buffer and ammonium acetate solution.

3.4.2 Assessing mechanical stability under hydrodynamic force

Previously, the use of the extensional flow device (detailed in Section 1.1.4.4: Extensional Flow Device (EFD) to mimic aggregation under flow) has been proven to trigger aggregation of proteins (J. Dobson *et al.*, 2017; Willis, 2018). Additionally, the presence of arginine in the formulation buffer has been demonstrated to reduce aggregation of mAbs when stressed using the device, compared to in ammonium acetate solution (Willis *et al.*, 2018). In these experiments, the concentration of protein remaining in solution after clarification using ultracentrifugation is compared, calculated from either: the drop in the absorbance (at 280 nm) peak area measured by HPLC; or by the drop in the absorbance (at 280 nm) measured by spectrophotometric absorbance. This results are displayed as a percentage of monomer remaining, compared to the concentration of an un-stressed sample (which is stored at room temperature for the duration of the longest stress experiment). The results in this section were performed in collaboration with Dr Leon Willis where his contributions to the experiments have been provided.

3.4.2.1 Experimental considerations

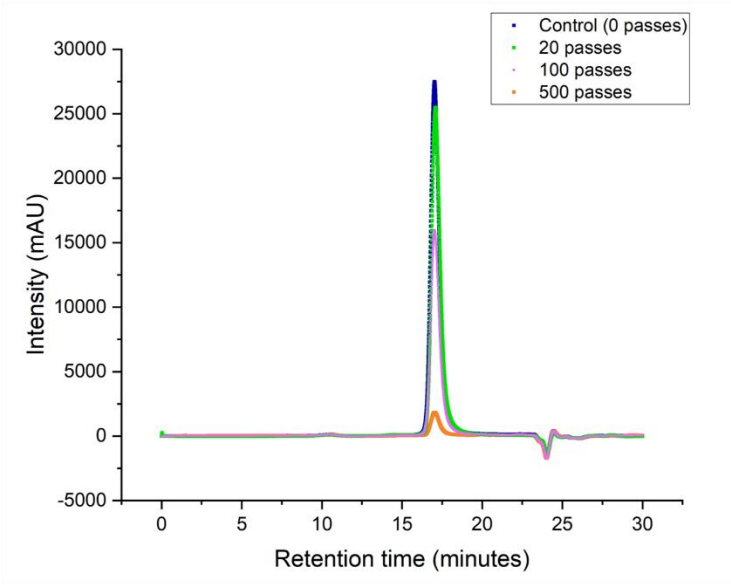
3.4.2.1.1 Soluble material is monomeric

In order to investigate the decrease of soluble monomer due to exposure to hydrodynamic flow, an initial mAb concentration of 0.25 mg mL^{-1} was chosen for the following experiments. This concentration eliminates the results being influenced by any higher order self-oligomerisation, which is not triggered by hydrodynamic flow in the measurements. At 1 mg mL^{-1}, WFL is mainly

dimeric (C. L. Dobson *et al.*, 2016) but reversibly self-associates, and at 0.1 mg mL^{-1}, both WFL and STT are monomeric (C. L. Dobson *et al.*, 2016). Therefore, at 0.25 mg mL^{-1} WFL is assumed to be mainly monomeric, and STT and 114 are assumed monomeric at this concentration. This assumption can be verified by observing the HP-SEC raw traces used for the calculation of monomer loss in this section.

Here, 114 was stressed for 0 passes (the control was held at room temperature for the same time as the 500 passes sample took for flow exposure), 20 passes, 100 passes and 500 passes at 16 mm s^{-1} in 150 mM ammonium acetate pH 6.0. The protein were clarified by ultracentrifugation (Section Methods: 2.2.7.2, 2.2.7.3) to remove pellet-able material, and the soluble material was subjected to the HP-SEC absorbance quantification process. Monomer loss is calculated by taking the area under the peak at 17 minutes known to correspond to mAb monomer (Figure 3.2). The absolute intensity traces were displayed (Figure 3.12).

a) Raw HP-SEC trace for a typical Monomer Loss experiment



b) Zoom to 10% intensity

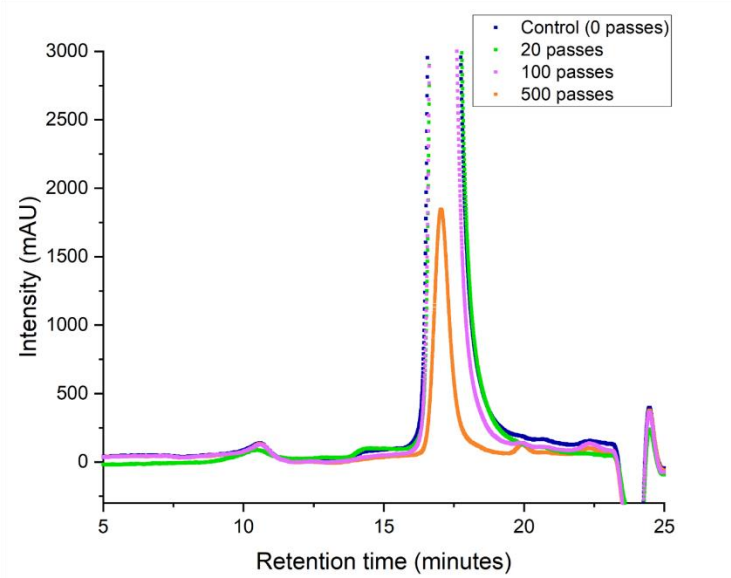


Figure 3.12 Raw HP-SEC traces from the monomer loss experiments to illustrate the calculation of percentage protein loss from flow exposure.

a) Raw intensity traces of 114 after 0 passes, 20 passes, 100 passes and 500 passes, clarification by ultracentrifugation and the soluble material (20 µg) loaded onto a TOSOH HP-SEC column. b) Zoom into the trace in a) where intensity is displayed at 10% of the overall most intense signal.

The area under the peak at 17 minutes (Figure 3.12) decreases as a function of pass number and is used to calculate percentage monomer loss by comparison to the control (normalised to 100%). In Figure 3.12 b, the peak at 11 minutes can be seen for all traces, indicating that there

is a small percentage of the sample that exists in a higher order structure than monomer. However, this peak corresponds to less than 1.5% of the total area of each trace in this experiment (control = 0.14%; 20 passes = 0.12%; 100 passes = 0.23%; 500 passes = 1.46%) and so the potential to influence the total intensity of the monomer peak is negligible. Additionally, as this is present in every sample, including the control, the presence of these higher order species observed appear to have appeared independently of the flow-exposure process and therefore their influence on the amount of monomer loss is not valid when considering the influence of flow on the aggregation of these mAbs. Therefore, for the following data, even though the monomer peak is the only peak taken into account for the calculation of loss of monomer, the conclusions in this section should be taken alongside the small caveat that the total remaining mAb product after ultracentrifugation contains trace amounts of soluble higher order species.

3.4.2.1.2 At low concentration, loss of monomer is strain-rate independent

Two different plunger velocities were utilised in this section. Experiments on WFL and STT were performed by Dr Leon Willis, where the plunger velocity was 8 mm s^{-1} for consistency to his previous work. Experiments on 114 were performed by the PhD candidate, where the plunger velocity was 16 mm s^{-1}, for consistency to the conditions chosen for the following chapters of this thesis. The plunger velocity of 16 mm s^{-1} provides a higher strain rate for a shorter amount of time than at 8 mm s^{-1}, which is important for rapidly triggering any changes in mAb structure before immediate labelling by Fast Photochemical Oxidation of Proteins (FPOP). The flow experiments on 114 in this section were performed at 16 mm s^{-1} and not repeated at 8 mm s^{-1} because the amount of 114 mAb sample received from collaborators AstraZeneca was limited, therefore only a limited number of experiments could be performed. Due to these differing experimental conditions, it was essential to establish that the data is comparable, by showing that at low concentration (0.25 mg mL^{-1}), loss of monomer is strain rate-independent (Willis (manuscript in preparation)).

Therefore, STT monomer loss at 8 mm s^{-1} and 16 mm s^{-1} were compared to 114 monomer loss at 16 mm s^{-1}, where the concentration of the mAbs was 0.25 mg mL^{-1} in 150 mM ammonium acetate pH 6.0 solution (Figure 3.13). Sample were stressed in the flow device at an initial concentration of 0.25 mg mL^{-1} for 10, 20, 50, 100, 200, 300 and 500 passes (where a 0 passes control constituted 100% monomer remaining). Aggregation was monitored by quantifying the

concentration of monomer remaining in a clarified sample by HPLC. This work was performed in collaboration with Dr Leon Willis.

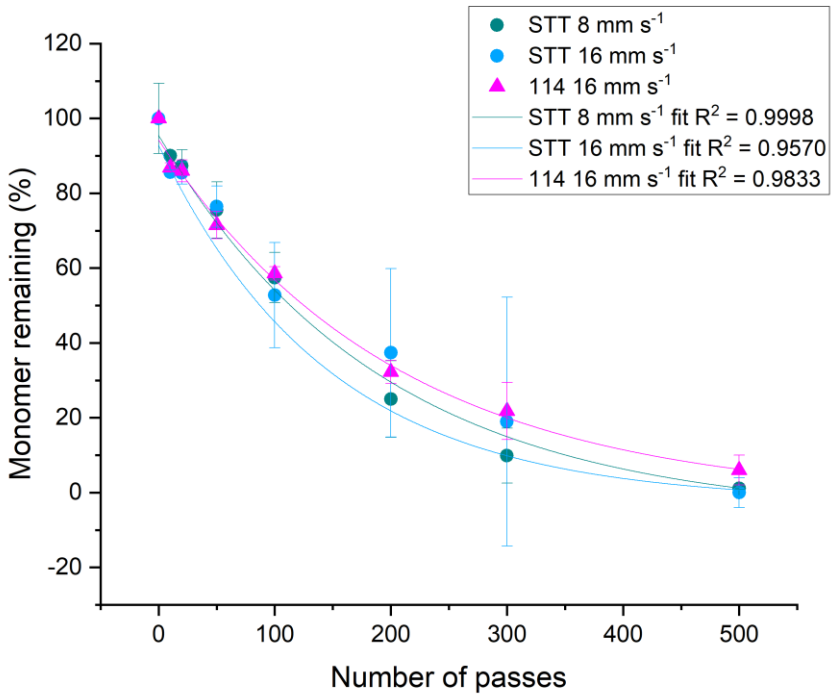


Figure 3.13 Monomer remaining of STT and 114 is strain rate independent at 8 mm s^{-1} and 16 mm s^{-1}.

Protein was prepared at an initial concentration of 0.25 mg mL^{-1} in 150 mM ammonium acetate pH 6.0. STT was stressed at 8 mm s^{-1} (teal circles) and 16 mm s^{-1} (light blue circles), for 0, 10, 20, 50, 100, 200, 300 and 500 passes. 114 was stressed at 16 mm s^{-1} (pink triangles), for 0, 10, 20, 50, 100, 200, 300 and 500 passes. Exponential decay fits were calculated for the data points and weighted to the error (standard deviation of n = 3 biological repeats). STT data from Dr Leon Willis. 114 data from the PhD candidate.

The results in Figure 3.13 demonstrate that the overall trend of monomer loss for STT and 114 are similar for both 8 mm s^{-1} and 16 mm s^{-1}. The decay fit is similar (within error) and the end point of monomer remaining is below 6% for all data sets, where STT (8 mm s^{-1}) = 1.2 ± 0.3%; STT (16 mm s^{-1}) = 0.0 ± 4.0%; 114 (16 mm s^{-1}) = 6.0 ± 4.0% (Figure 3.13).

3.4.2.2 Monomer loss experiments in acetate solution

The effect of hydrodynamic force on the mAbs in 150 mM ammonium acetate pH 6.0 will be explored in this section. This solution removes any protective effects that L-arginine has on the aggregation of the mAbs, and therefore a large effect on the percentage monomer remaining can be compared. At the initial concentration 0.25 mg mL^{-1}, the mAbs are assumed monomeric.

WFL, STT and 114 were prepared to an initial concentration of 0.25 mg mL^{-1} in 150 mM ammonium acetate pH 6.0, and were stressed in the flow device for 10, 20, 50, 100, 200, 300 and 500 passes (where a 0 passes control constituted 100% monomer remaining). Aggregation was monitored by quantifying the concentration of monomer remaining in a clarified sample by HPLC (for STT and 114) and by spectrophotometric absorbance (for WFL, due to its propensity to interact with the HP-SEC column matrix) (Figure 3.14). This work was performed in collaboration with Dr Leon Willis.

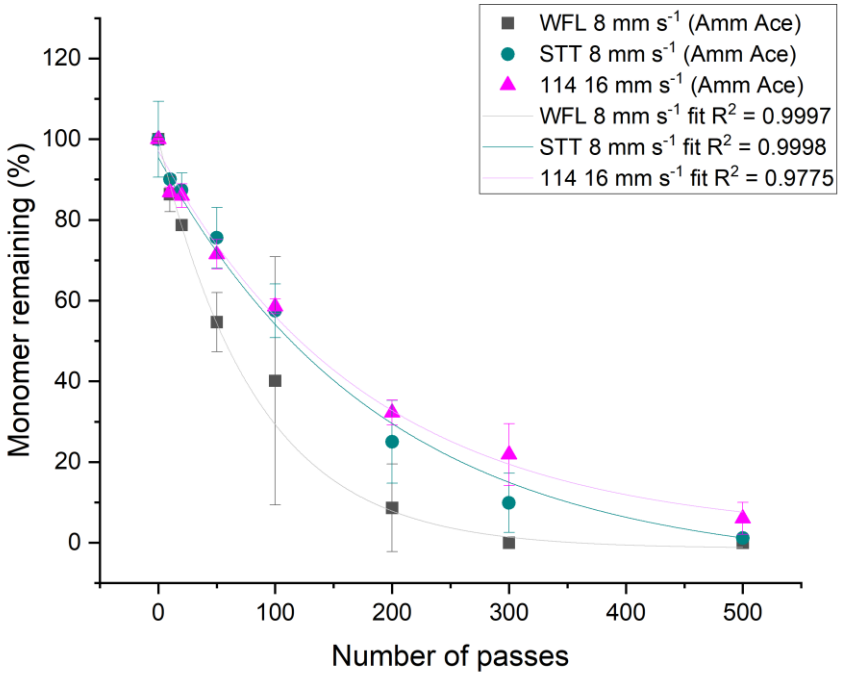


Figure 3.14 Monomer remaining of WFL, STT and 114 over 0-500 passes in 150 mM ammonium acetate, pH 6.0.

0.25 mg mL^{-1} WFL (black squares) was stressed at 8 mm s^{-1}, 0.25 mg mL^{-1} STT (teal circles) was stressed at 8 mm s^{-1}, and 0.25 mg mL^{-1} 114 (pink triangles) was stressed at 16 mm s^{-1}, for 0, 10, 20, 50, 100, 200, 300 and 500 passes. Exponential decay fits were calculated for the data points and weighted to the error (standard deviation of n = 3 biological repeats). WFL and STT data from Dr Leon Willis. 114 data from the PhD candidate.

WFL showed a higher rate of monomer loss in fewer passes compared to the other mAbs. The WFL data has a steeper exponential fit and reaches a baseline plateau after 300 passes. This trend for WFL supports WFL being more aggregation-prone than STT and 114 under hydrodynamic flow conditions in ammonium acetate solution.

STT and 114 have similar initial decays under 100 passes with a shallower gradient, compared to WFL. At 500 passes, the percentage of 114 monomer remaining is higher than the baseline of STT and WFL, suggesting the 114 monomer is more resistant to aggregation than both STT and WFL under these conditions after 500 passes. This finding supports the suggestion that 114 has reduced aggregation propensity than STT, as it was identified in the initial evolution screen of WFL_scFv as it had a higher TPBLA score than both WFL and STT (Figure 1.18).

3.4.2.3 Monomer loss experiments in formulation buffer

Next, monomer loss of STT and 114 was compared in formulation buffer, which is protective from aggregation (Willis *et al.*, 2018) and so the loss of monomer is minimal. STT and 114 were stressed at 0.25 mg mL^{-1} for 10, 20, 50, 100, 200, 300 and 500 passes (Figure 3.15).

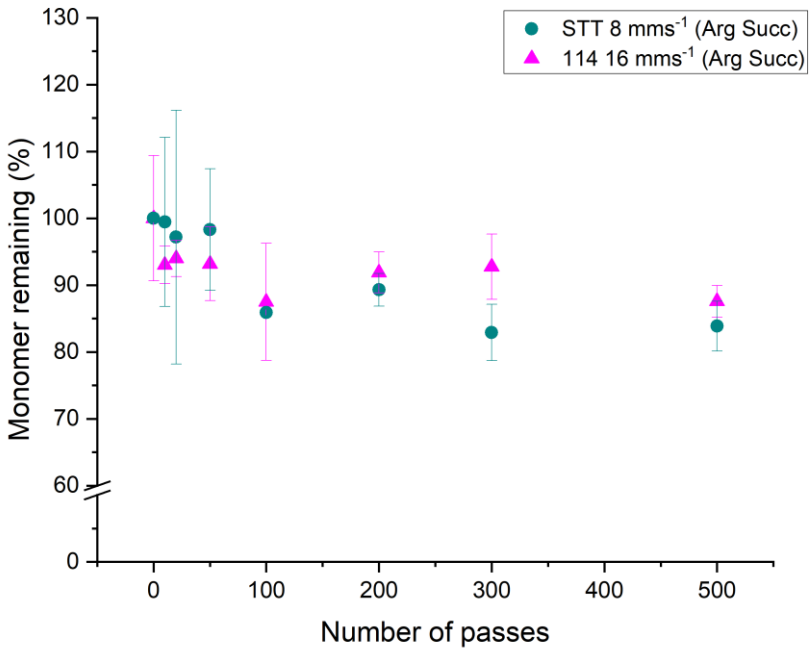


Figure 3.15 Monomer remaining of STT and 114 over 0-500 passes in formulation buffer.

0.25 mg mL^{-1} STT (teal circles) was stressed at 8 mm s^{-1}, and 0.25 mg mL^{-1} 114 (pink triangles) was stressed at 16 mm s^{-1}, for 0, 10, 20, 50, 100, 200, 300 and 500 passes. Error bars show standard deviation of n = 3 biological repeats. STT data from Dr Leon Willis. 114 data from the PhD candidate.

The results in Figure 3.15 demonstrate that both 114 and STT do not aggregate (less than 15% monomer loss) after 500 passes in formulation buffer. The results support the knowledge that arginine-containing buffer protects protein from aggregation under hydrodynamic stress in the flow device (Willis, 2018; Willis *et al.*, 2018). 114 acts similar to STT where the loss of monomer over 500 passes reaches a plateau at around 90% monomer remaining. Under conditions that lead to aggregation of STT and 114 (200 passes in 150 mM ammonium acetate pH 6.0), these mAbs lose 75.0% and 67.7% monomer respectively (Figure 3.14), however in protective conditions (200 passes in formulation buffer), STT and 114 lose 10.7% and 8.1% monomer, respectively (Figure 3.15).

3.4.2.4 Validating the flow conditions for the FPOP-LC-MS/MS section of this thesis

Overall, using the flow device to stress mAbs demonstrates that the device can trigger loss of monomer through insoluble pelletable aggregate. However, the process and mechanism by

which monomers transition from a native state, through an aggregation-prone state, to loss of monomer is still unknown, hence the need for a labelling technique which could take a snapshot, or fingerprint, of the species present. It is clear from these experiments that 114 acts similarly or is even more resistant to flow than to STT, and that WFL is the most prone to loss of monomer by hydrodynamic stress, making these proteins an exciting model system to study. The crux of this thesis (in following chapters) will aim to elucidate the difference between the activated, aggregation-prone state of these mAbs triggered through hydrodynamic stress.

Therefore, the conditions chosen for the pioneering flow-coupled-to-FPOP experiments needed to be a carefully considered balance of conditions. On one hand, there needed to be enough passes trigger a perturbation to 114 and STT monomer structure (enough consecutive passes) – but over a short time period – to be able to capture any structural changes to the aggregation-resistant proteins. On the other hand, as WFL is aggregation-prone and the number of passes needed to trigger a high extent of monomer loss is much lower than that of STT and 114, the pass number chosen needed to be low to limit the influence of aggregated material in the flow-FPOP output as much as possible. Additionally, the flow experiments needed to be performed at 1 mg mL^{-1} to allow a 10 times dilution into formulation buffer and scavenger.

Considering these aspects, the following conditions were tested in preparation for the FPOP-LC-MS/MS protocol later in this thesis: WFL STT and 114 were stressed at 1 mg mL^{-1} for 200 passes at 16 mm s^{-1} in formulation buffer. 200 passes at 16 mm s^{-1} takes approximately 10 minutes in the flow device. Aggregation was monitored by quantifying the concentration of monomer remaining in a clarified sample by HPLC (for STT and 114) and by spectrophotometric absorbance (for WFL, due to its propensity to interact with the HP-SEC column matrix) (Figure 3.16).

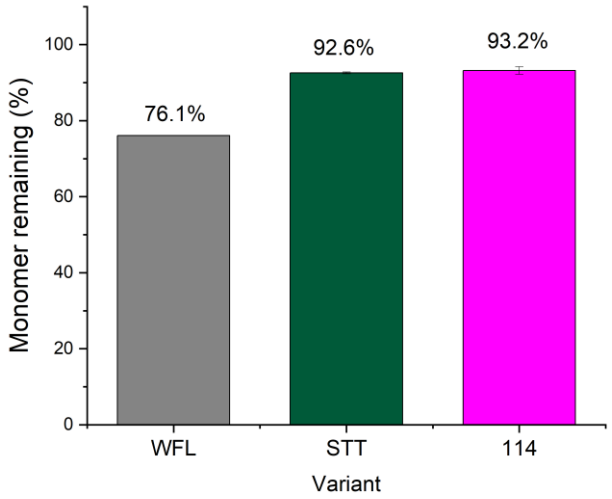


Figure 3.16 WFL, STT and 114 monomer remaining under flow-FPOP-LC-MS/MS conditions.

Sample in formulation buffer at 1 mg mL^{-1} were stressed for 200 passes at 16 mm s^{-1}. Monomer loss was quantified using the loss of absorbance intensity HPLC (for STT and 114) and by spectrophotometric absorbance (for WFL). Error bars show standard deviation of n = 2 biological repeats. All data from the PhD candidate.

The loss in monomer over 200 passes is minimal for the three variants. As WFL is more aggregation-prone, as explored in various techniques throughout this chapter, the monomer loss in these conditions were to be expected. The minimal loss of STT and 114 add confidence to the hypothesis that choosing these conditions for investigation with labelling-MS, we could capture partial unfolding while inducing minimal insoluble aggregation. Using Fast Photochemical Oxidation of Proteins (FPOP) after hydrodynamic stress with the device in these conditions will allow the capture of any states of mAb perturbed from the native state. This could allow us to understand which regions of the three mAbs are more prone to perturbation, and may add weight to identifying specific regions responsible for the pathway to aggregation.

It was important to verify that WFL, STT and 114 do not perturb from their native state into a partially unfolded or aggregated state over time, without applying hydrodynamic force or temperature-induced stress. For this, Hydrogen-Deuterium Exchange coupled to Mass Spectrometry (HDX-MS) was performed.

3.5 Probing differences in intrinsic dynamics over time with HDX-MS

HDX-MS was used as a labelling technique to monitor any local dynamics the mAbs may exhibit over a 30 minute period in formulation buffer. This time scale covers – and exceeds – the time scale of the flow coupled to FPOP experiment, where 200 passes at 16 mm s^{-1} took approximately 10 minutes. This experiment in formulation buffer at 1 mg mL^{-1} was important for probing any local dynamics changes that may have arisen due to the substitutions of amino acids in the three mutants. HDX-MS has been used for tracking mutation-induced differences in exchange kinetics in mAbs before (Majumdar *et al.*, 2015), where mutations in an IgG1's C_H2 domain triggered distal flexibility changes in V_H, C_H1, and V_L domains. The aim here was to see whether the substitutions of amino acids between WFL, STT and 114 significantly altered local or distant modification accessibility from the sites of mutation. The experiments in this section were designed by the PhD candidate, the lab work performed by Dr Piera Marchetti (AstraZeneca), and the analysis performed by the PhD candidate.

Native WFL, STT and 114 at 10 mg/mL were diluted 1:10 with deuterated formulation buffer (125 mM L-arginine, 20 mM sodium succinate dibasic hexahydrate, pD 6.0) and equilibrated for 30 seconds, 60 seconds, 300 seconds or 1800 seconds before immediate quenching with guanidinium-containing quench buffer at pH 2.4 at 0 ˚C. Then sample was loaded onto a pepsin column and peptides were immediately analysed using LC-MS/MS to identify deuterium uptake over time. The FASTA sequences of the mAbs were entered into BioPharma Finder, a peptide mapping analysis software which, from the MS/MS data, generates a list of identified peptides from a proteomics study (giving a coverage map of overlapping peptides). Then the software HDExaminer (Sierra Analytics) matches up spectra at several time points of deuteration, to map the overall behaviour of exchange of each protein by looking at the shift in the spectra for each peptide over time to map deuterium uptake.

From this matching process, the overall deuteration signature is overlayed onto coverage maps of the mAbs (Figure 3.17). The coverage map is a compiled based on the heat map of deuterium uptake, generated from all of the results for each peptide over the time course (0.5 minutes to 30 minutes). Therefore the colours (green and yellow in Figure 3.17) of each peptide in the map represent the confidence value that the software has assigned for the assignment of the deuterium occupancy to the expected retention time of the peptide, indicating that the data matches the expected deuterium uptake curve over time for the peptide. Green = good agreement in the solved deuterium occupancy assignment for the peptide at all four time points (unified data trend for the overlapping peptides); yellow = semi-good agreement in the solved

deuterium occupancy assignment for the peptide at all four time points (a less unified data trend for the overlapping peptides).

The overall coverage was calculated from all of the peptides in the coverage map, and is found in Table 3.1. Data for these peptides were then used for differential heat maps and assignment to the protein structure.

Table 3.1 The peptide coverage for WFL, STT and 114 in the pepsin digestion of the HDX experiments (both green = confident and yellow = semi-confident deuterium occupancy resolution assignments contributed to the overall % coverage).

	Overall sequence coverage (%)	
	Overall	Per chain
WFL	79.9	HC 74.2 LC 85.6
STT	78.2	HC 70.7 LC 85.6
114	81.1	HC 78.8 LC 83.3

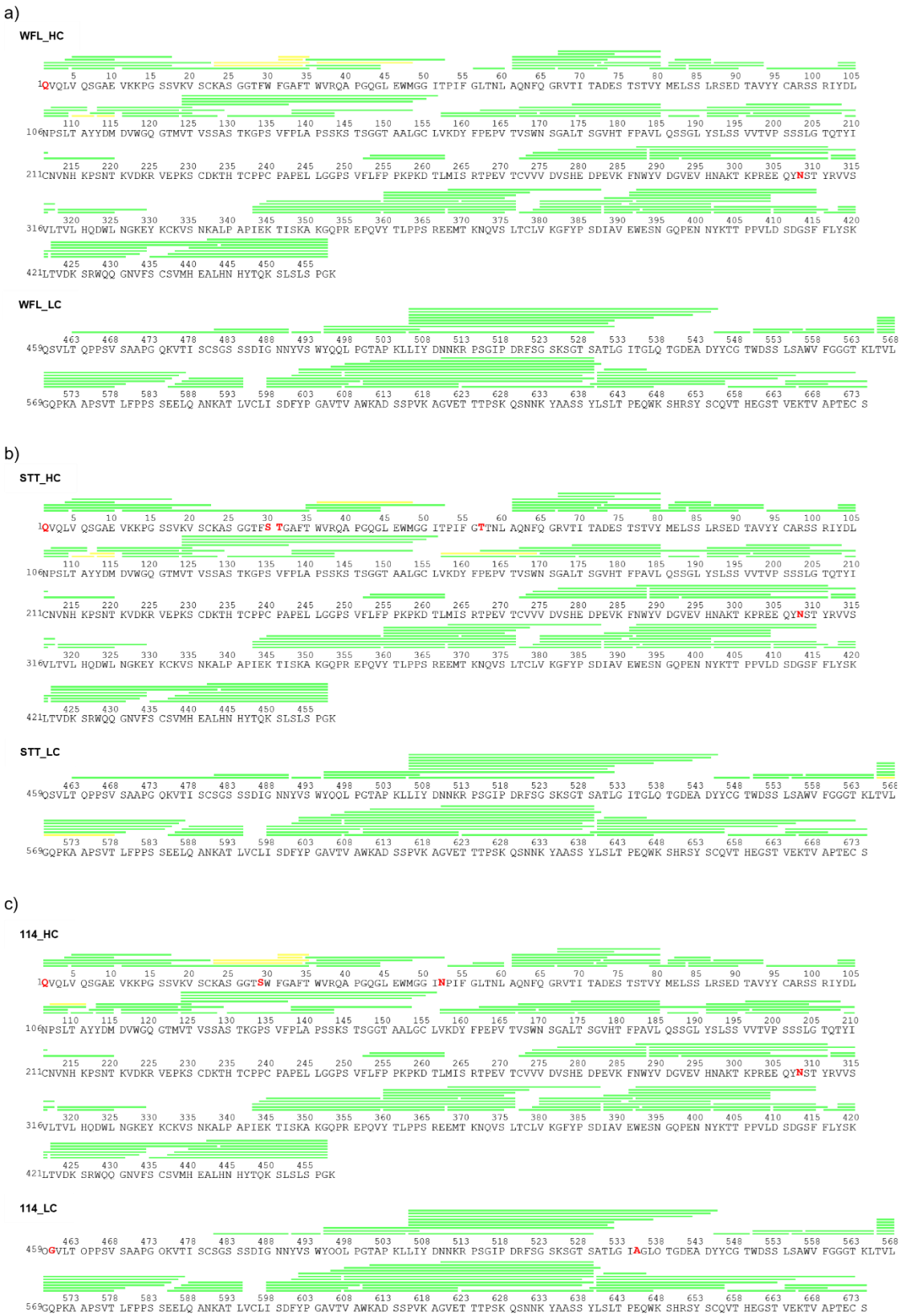


Figure 3.17 Peptide coverage maps for WFL, STT and 114.

Sequences for WFL (a), STT (b) and 114 (c) are displayed with their corresponding peptides above (green and yellow bars). Green = high confidence in the solved deuterium occupancy assignment for all four time points, yellow = medium

confidence in the solved deuterium occupancy assignment for all four time points. Mutations in comparison to 'wildtype' sequence WFL are shown in red lettering. Data incorporating n = 3 technical repeats.

The coverage of the mAbs after pepsin digestion was greater than 78% for all of the sequences (Table 3.1). The regions of coverage of the peptides were mostly common, indicating the mAbs were unfolded similarly in the guanidine hydrochloride- and TCEP-containing quench leading to similar patterns of digestion, important for comparability between the variants. From these data, heat maps were generated corresponding to deuterium uptake over time. These maps were compared to each other resulting in residual plots which compare the differences in percentage deuterium uptake on each peptide between the different mAbs over the time points measured (Figure 3.18, Figure 3.19, Figure 3.20). The residual plots visualise the differential change in deuterium as a percentage (%D) between a mAb pair, and represent the comparison of two butterfly plots, making it easier for the reader to compare differences in uptake. This representation is powerful for identifying any regions of sequence that differ in deuterium uptake over time between the variants, highlighting partial unstructuring or protection triggered by the differences in sequence. A significance level of 10% is a standard cut-off percentage specified by the analysis software used, HDExaminer. Below this value, the deuterium uptakes in the mAbs being compared are considered similar and statistically insignificant. Where one mAb peptide has a higher level of exchange than another at a particular time point and the %D values differ by more than 10%, this is marked as a significant change in deuterium, indicating a region of less or more solvent exposure.

Comparing WFL to STT (Figure 3.18), WFL to 114 (Figure 3.19), and STT to 114 (Figure 3.20), we can identify regions of sequence with different propensities to exchange hydrogen with deuterium depending on the variant.

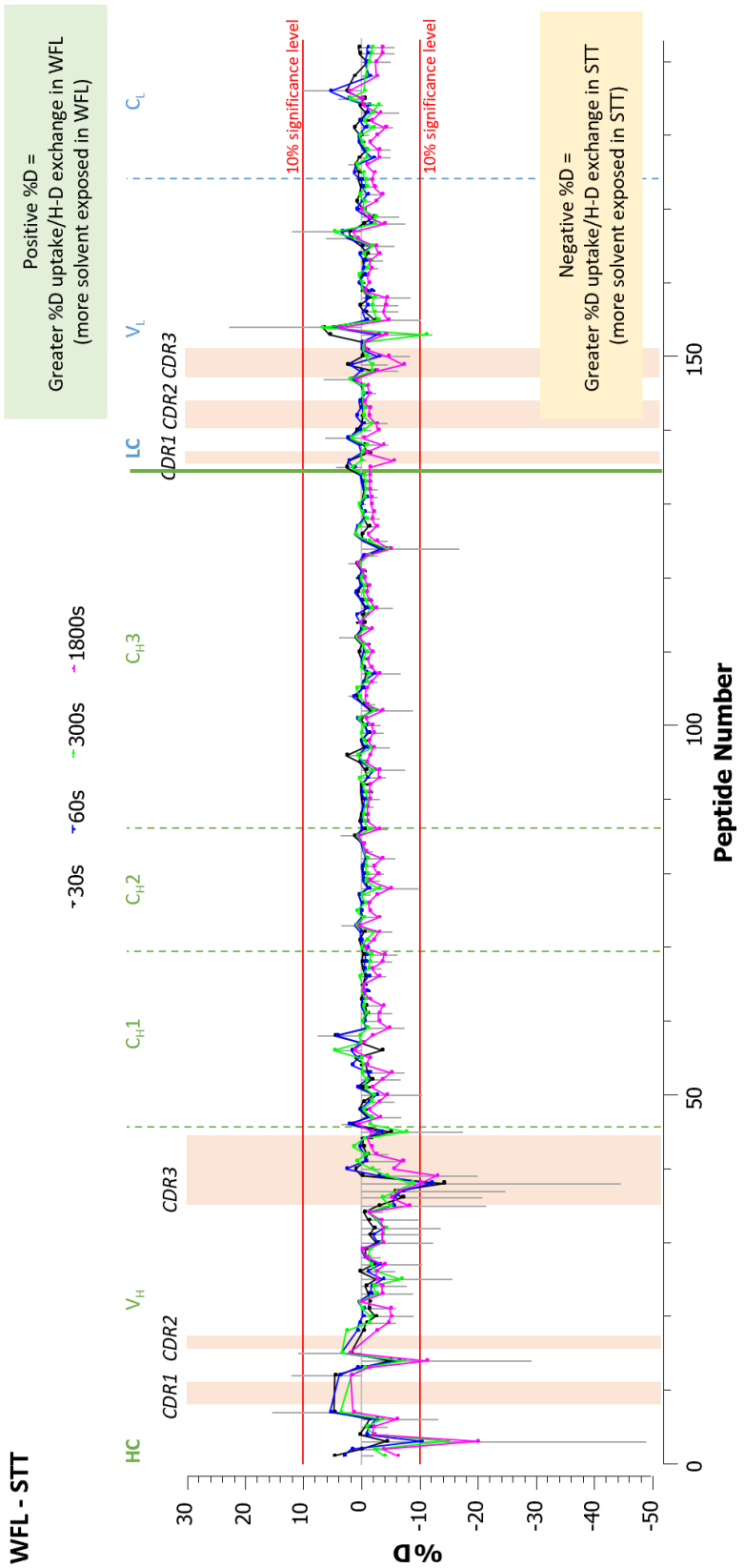


Figure 3.18 Residual WFL – STT percentage deuterium differential uptake plot comparing uptake in WFL to uptake in STT over 30 seconds, 1 minute, 5 minutes and 30 minutes. The difference in uptake is expressed in + or - %D for each peptide identified. The %D for each peptide over the time points (30 s = black, 1 min/60 s = blue, 5 min/300 s = green, 30 min/1800 s = pink) is plotted against the peptides as they are numbered in order from the N-terminus of the heavy chain (HC = green labels and dashed dividers) to its C-terminus, followed by the N-terminal of the light chain (LC = blue labels and dashed dividers) to its C-terminus. CDRs are highlighted in orange (V_HCDR1 = #7-11; V_HCDR2 = #16-17; V_HCDR3 = #35-44; V_LCDR1 = #136-138; V_LCDR2 = #140-144; V_LCDR3 = #148-151). Grey vertical bars = cumulative %D for all of the time points for that peptide. Red horizontal bars = 10% significance level specified by HDExaminer software.

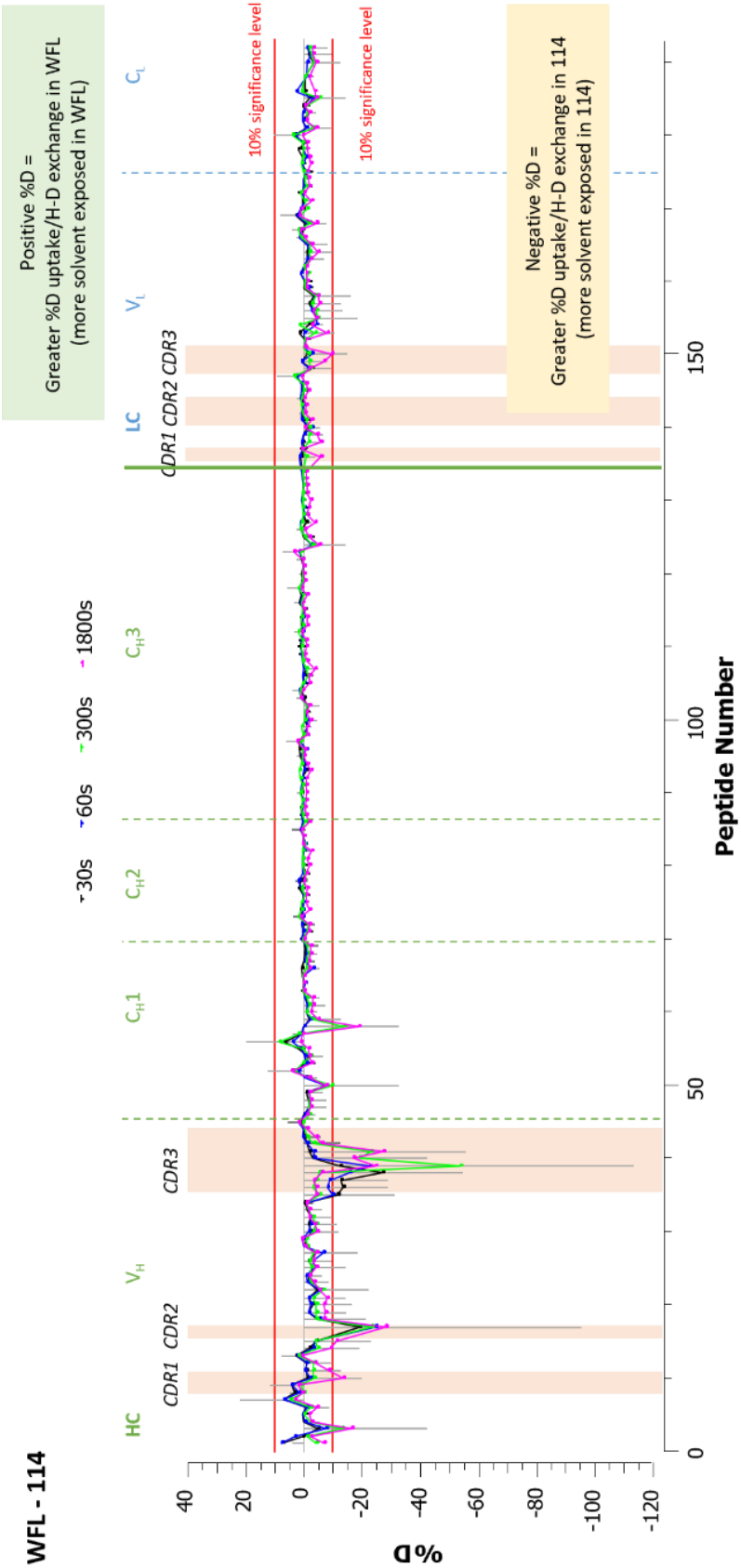


Figure 3.19 Residual WFL – 114 percentage deuterium differential uptake plot comparing uptake in WFL to uptake in 114 over 30 seconds, 1 minute, 5 minutes and 30 minutes. The difference in uptake is expressed in + or - %D for each peptide identified. The %D for each peptide over the time points (30 s = black, 1 min/60 s = blue, 5 min/300 s = green, 30 min/1800 s = pink) is plotted against the peptides as they are numbered in order from the N-terminus of the heavy chain (HC = green labels and dashed dividers) to its C-terminus, followed by the N-terminal of the light chain (LC = blue labels and dashed dividers) to its C-terminus. CDRs are highlighted in orange (V_HCDR1 = #7-11; V_HCDR2 = #16-17; V_HCDR3 = #35-44; V_LCDR1 = #136-138; V_LCDR2 = #140-144; V_LCDR3 = #148-151). Grey vertical bars = cumulative %D for all of the time points for that peptide. Red horizontal bars = 10% significance level specified by HDExaminer software.

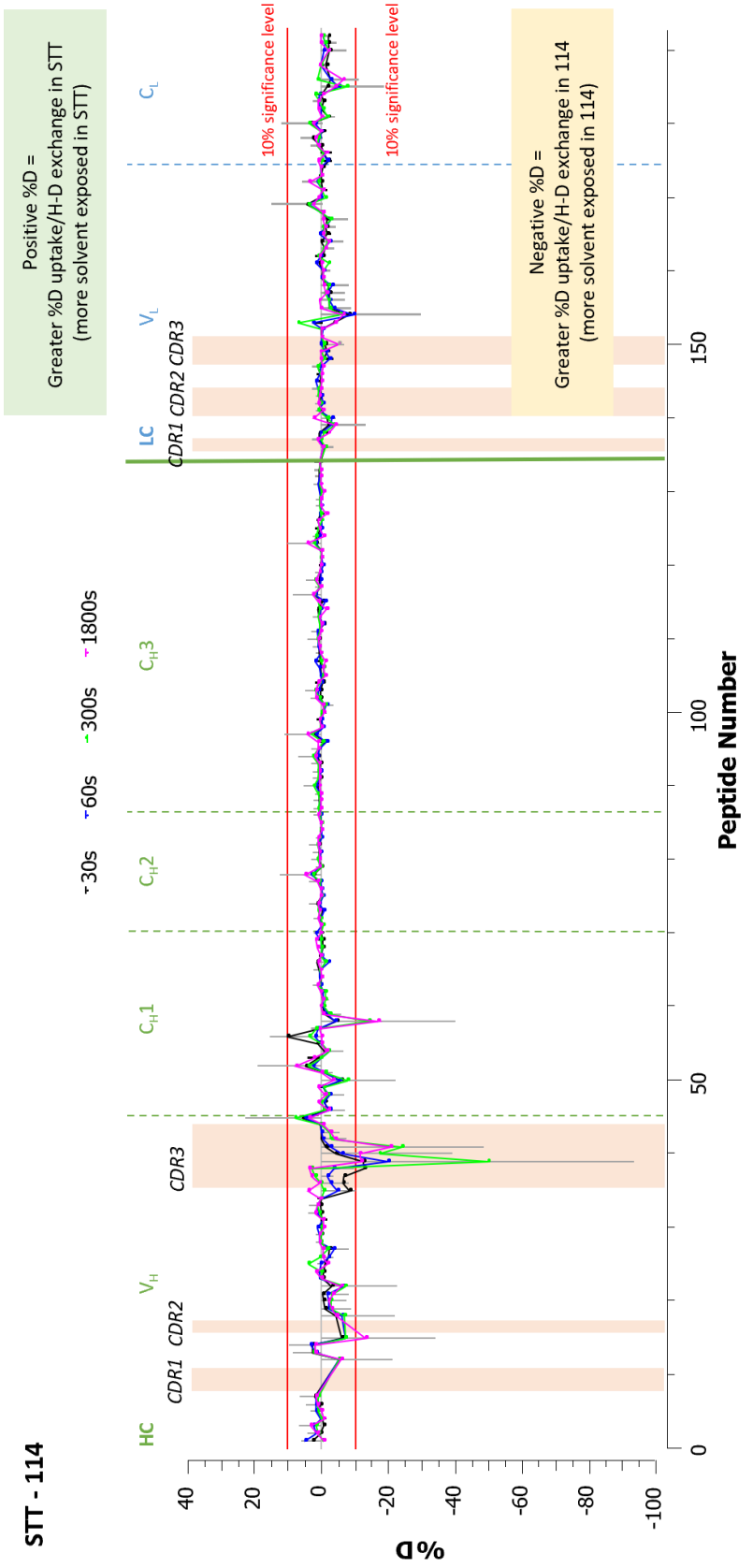


Figure 3.20 Residual STT – 114 percentage deuterium differential uptake plot comparing uptake in STT to uptake in 114 over 30 seconds, 1 minute, 5 minutes and 30 minutes. The difference in uptake is expressed in + or - %D for each peptide identified. The %D for each peptide over the time points (30 s = black, 1 min/60 s = blue, 5 min/300 s = green, 30 min/1800 s = pink) is plotted against the peptides as they are numbered in order from the N-terminus of the heavy chain (HC = green labels and dashed dividers) to its C-terminus, followed by the N-terminal of the light chain (LC = blue labels and dashed dividers) to its C-terminus. CDRs are highlighted in orange (V_HCDR1 = #7-11; V_HCDR2 = #16-17; V_HCDR3 = #35-44; V_LCDR1 = #136-138; V_LCDR2 = #140-144; V_LCDR3 = #148-151). Grey vertical bars = cumulative %D for all of the time points for that peptide. Red horizontal bars = 10% significance level specified by HDExaminer software.

In the time course of the experiment, the differential deuterium uptake for the majority of peptides in the constant domains C_H2, C_H3 and C_L showed a gradual increase, indicative of similar degrees of protection. This is demonstrated by the black, blue, green and pink bars corresponding to different time points, with the difference in uptake generally remaining under the 10% significance level set by HDExaminer (Figure 3.18, Figure 3.19, Figure 3.20). The exceptions for this are mainly focussed around the V_H domains, and the constant domain C_H1 when comparing 114 to WFL or STT (Figure 3.19 and Figure 3.20) – the constant domain C_H1 in WFL and STT had uptakes over the time course that did not surpass the 10% significance level (Figure 3.18). Overall, the results suggest that, globally, the mAbs share similar solution-based propensities to exchange deuterium over time, in turn suggesting similar global structures; with the observation that their V_H regions have differences in their ability to exchange deuterium over time, indicating a difference in protection.

The grey vertical bars in the figures represent the difference in the cumulative %D over the four time points. This occurs if a particular peptide in one mAb had a large %D uptake at each time point, compared to the other mAb paired to it in the differential plot. For example, the largest differences in uptake were in the first half of the HC peptides (corresponding to the V_H and C_H1) and parts of the LC (mainly V_L) to a lesser extent, when 114 was compared to both WFL (Figure 3.19) and STT (Figure 3.20).

The V_H region for all three mAbs had the most significant differences in differential %D uptake when comparing the mAbs to each other. Most specifically, this was significant in the peptides between the N-terminus and V_HCDR1, the peptides spanning the V_HCDR2, and the peptides spanning the V_HCDR3. Focussing on the V_HCDR3, STT gained significantly more deuterium than WFL over the 30 second, 60 second and 1800 second time points, to an extent of 14% (Figure 3.18). When comparing 114 to WFL (Figure 3.19) however, every time point demonstrated a 30% or higher difference in uptake, and at the 300 second time point this reached a difference of 50%, indicating the CDR peptides in 114 are much more accessible to solvent and therefore exchange than WFL. 114 compared to STT further indicates 114 is more solvent accessible in this region, compared to STT, where the differences at 30 second, 60 second and 1800 second time points reached 20%, and the 300 second time point reached a difference of 50% for some peptides. The V_HCDR3 is known to be involved in the recognition and interaction of WFL and STT with its binding partner NGF (C. L. Dobson *et al.*, 2016) and so is solvent exposed, allowing deuterium exchange. Perhaps the fact that 114 showed the highest %D uptake in this region

comments on the accessibility of the V_HCDR3, which would in turn aid the recognition of NGF, however this is speculation and has not been tested.

These striking differences in %D uptake over time for each pair of mAbs are more clearly observed when comparing the deuterium uptake plots for key peptides in the V_HCDR3 region (Figure 3.21). The uptake plots over the four time points for peptides #39, #40 and #41 for the three variants are shown, where the error bars indicate the variation between three technical replicates. For comparison, peptide #33 situated between V_HCDR2 and V_HCDR3 is also shown as there is no significant difference between the variants or any significant increase in uptake over time (the peptide is buried/protected from exchange).

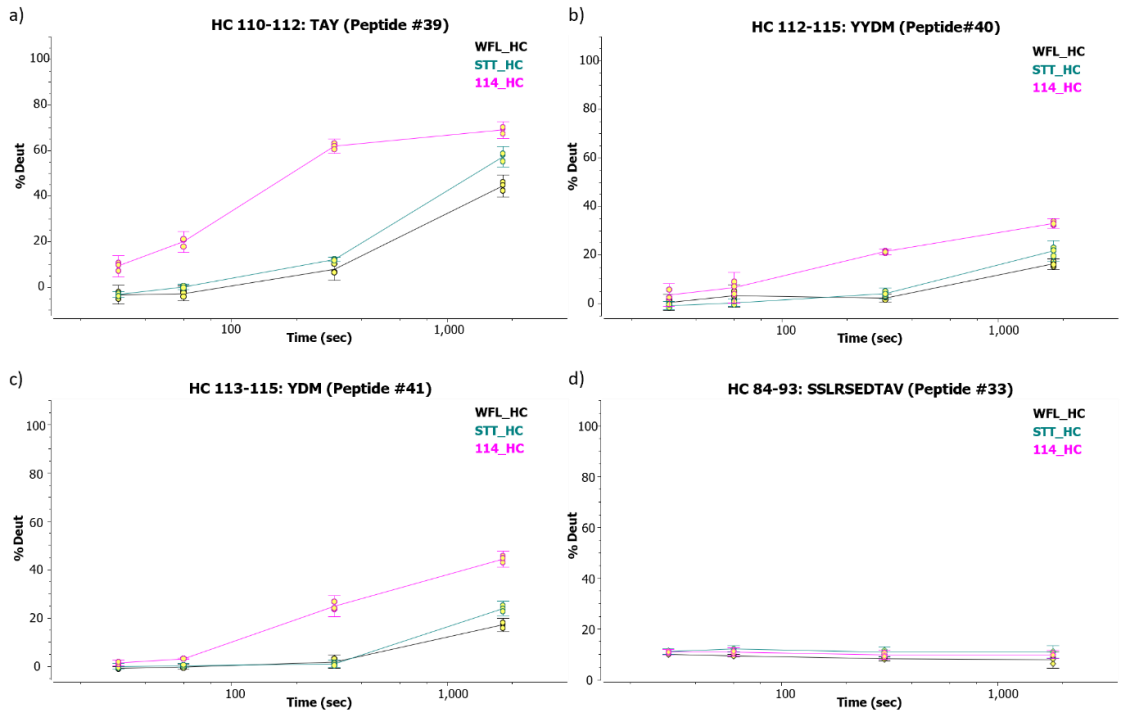


Figure 3.21 Deuterium fractional uptake plots for example peptides with greater than 10% significant difference in uptake between 114 and the other mAbs.

Peptides in the CDR3 region (a, b, c) show significant increase in uptake at 60 seconds, 300 seconds and 1800 seconds time points for 114, compared to STT and WFL. A peptide with no significant change in relative uptake (d) is included for comparison. Data error bars represent SD from n = 3 experiments.

The %D uptake plots demonstrate a significantly different uptake for 114 for the 300 seconds and 1800 seconds time points (Figure 3.21). The data demonstrate time-resolved differences between 114 and its comparative mAbs WFL and STT in the V_HCDR3 region. This indicates WFL and (to a lesser extent) STT experience more protection from exchange than 114, meaning the fractional uptake of deuterium is slower than for 114. This is seen clearly in Figure 3.21 a, where 114 reaches 60% deuterium after 300 seconds, which remained at a similar percentage at 1800 seconds. For both STT and WFL, deuterium uptake was low for the first three time points but increased toward 55% (STT) and 45% (WFL) after 1800 seconds of exchange, showing a significant difference where STT exchanged to a higher percentage over this time period for this peptide. In fact, for the three V_HCDR3 peptides studied here, the 300 seconds time point showed the greatest change in %D for 114 (Figure 3.21). This is prevalent to a much lesser extent in the comparison of STT with WFL. This as a whole indicates that the kinetics of exchange for 114 are faster than WFL and STT in this region.

It is important to note that the peptides in Figure 3.21 have been assigned different spectral confidence scores between the variants, seen by the green and yellow representations on the coverage map (Figure 3.17) – medium confidence assignment in WFL and STT, but high confidence in 114. This represents how confident the software was at assigning the actual mass spectra for the peptide at the observed charge state, to the theoretical isotope cluster for that peptide. The software assigns the confidence based on several factors: how well the observed spectra aligns with the retention time window that was predicted from the theoretical isotope cluster for the peptide; the ion intensity; and the background signal noise. The raw spectra data for one repeat of the peptide #41 YDM (Figure 3.21 c) has been displayed in Figure 3.22, as an example of how the spectra is classified as high or medium confidence by the HDExaminer software. Figure 3.22 shows one example of non-deuterated control spectra and the 30 minute exposure spectra for each variant, alongside the retention time window used to extract the data over the theoretical extracted ion chromatogram of the peptide calculated by the software.

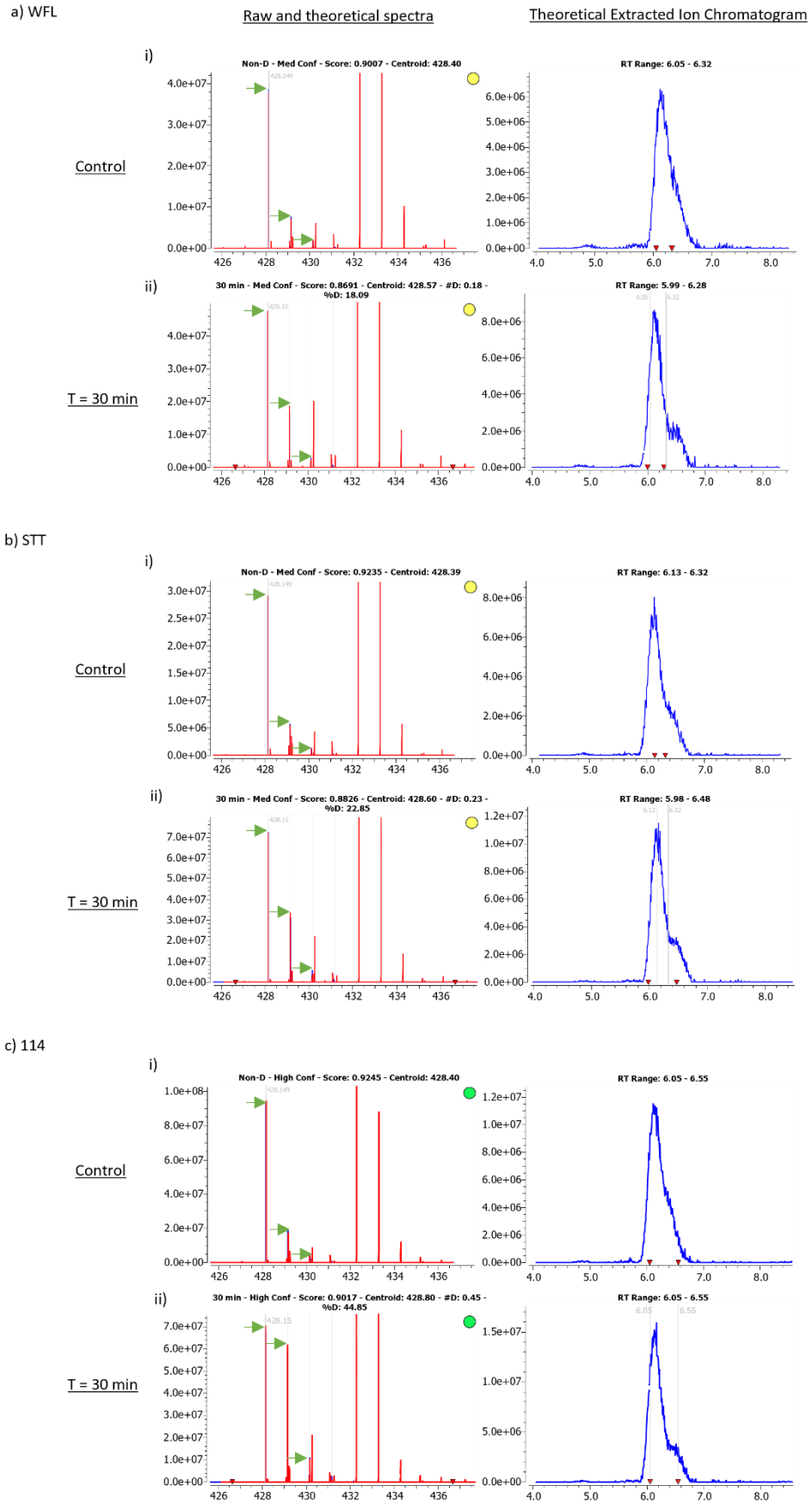


Figure 3.22 HDX spectra for peptide YDM for WFL, STT and 114.

Data for peptide #41 YDM in WFL (a), STT (b) and 114 (c) are shown for the control (i) (T = 0, no deuterium exposure) and for the 30 minutes time point of exchange (ii) (T = 30 min). Left graph: the raw MS data for the peptide (red) for the chosen m/z window is overlayed with the theoretical isotope cluster (blue trace) for the peptide at charge state z = 1. Green arrows indicate the peaks that match the theoretical isotope distribution. Right graph: the theoretical extracted ion chromatogram from the theoretical isotope cluster. The XIC provides the retention time window for the peptide data. The red triangles indicate the RT window used to extract the raw spectra from. This window for the control is represented by grey vertical lines in the theoretical extracted ion chromatogram graph for the T = 30 minute timepoint.

By manually looking at the raw data in Figure 3.22, the user can determine that the data support the significance of the differences mentioned in Figure 3.21. Even though the peptide YDM was single charged, predisposing the ion intensity to being low, the intensity of the spectra were enough above the noise to be able match the data to the theoretical isotope distribution (the red spectra overlays well with the blue predicted trace, indicated by the arrows). The data for 114 was a higher abundance (over 2x) and this contrast in intensity compared to the same peptide in WFL and STT would have contributed to the increased confidence in the score given to this peptide. The spectra for each peptide in each sample were, however, close to the noise, indicated by the other peaks present which do not correspond to the YDM peptide, and this is reflected by the broad theoretical retention time peak (blue graphs, right hand side). Despite this, the data classified by HDExaminer as medium confidence for WFL and STT lines up to the theoretical distribution similarly to the data for 114, which was high confidence. Therefore, this lends confidence to the significant differences in deuterium uptake detailed in the deuterium uptake plots in Figure 3.21.

To display the wealth of data collected in these experiments, the global %D values from Figure 3.18, Figure 3.19 and Figure 3.20 were translated into a visual output mapped onto the 3D predicted structure of mAbs WFL and STT (Figure 3.23). The differences in exchange are colour-coded based on the peptide level differences in exchange between the two variants being compared. The colouration gradient was prepared in 10% increments, where blue = -100%D (more protection in one variant compared to the other) and red = 100%D (more exchange in one variant compared to the other). These data were mapped onto the 3D PyMOL structure of the STT mAb. The energy-minimised models were generated from a prediction of the scFv sequence grafted onto an IgG1 template. From this representation of the data it is clear to see that the majority of the structures have globally similar exchange (grey colouration) over time, indicating that any structural differences in formulation buffer natively (without induced stress such as flow) are minimal. The area around the V_HCDR3 region where STT and 114 have greater

%D than WFL over time are seen in light pink (Figure 3.23 aii, bii and cii), corresponding to between 30 and 50% more deuterium than the 'wildtype' state. This is most pronounced comparing 114 to WFL over 1800 seconds.

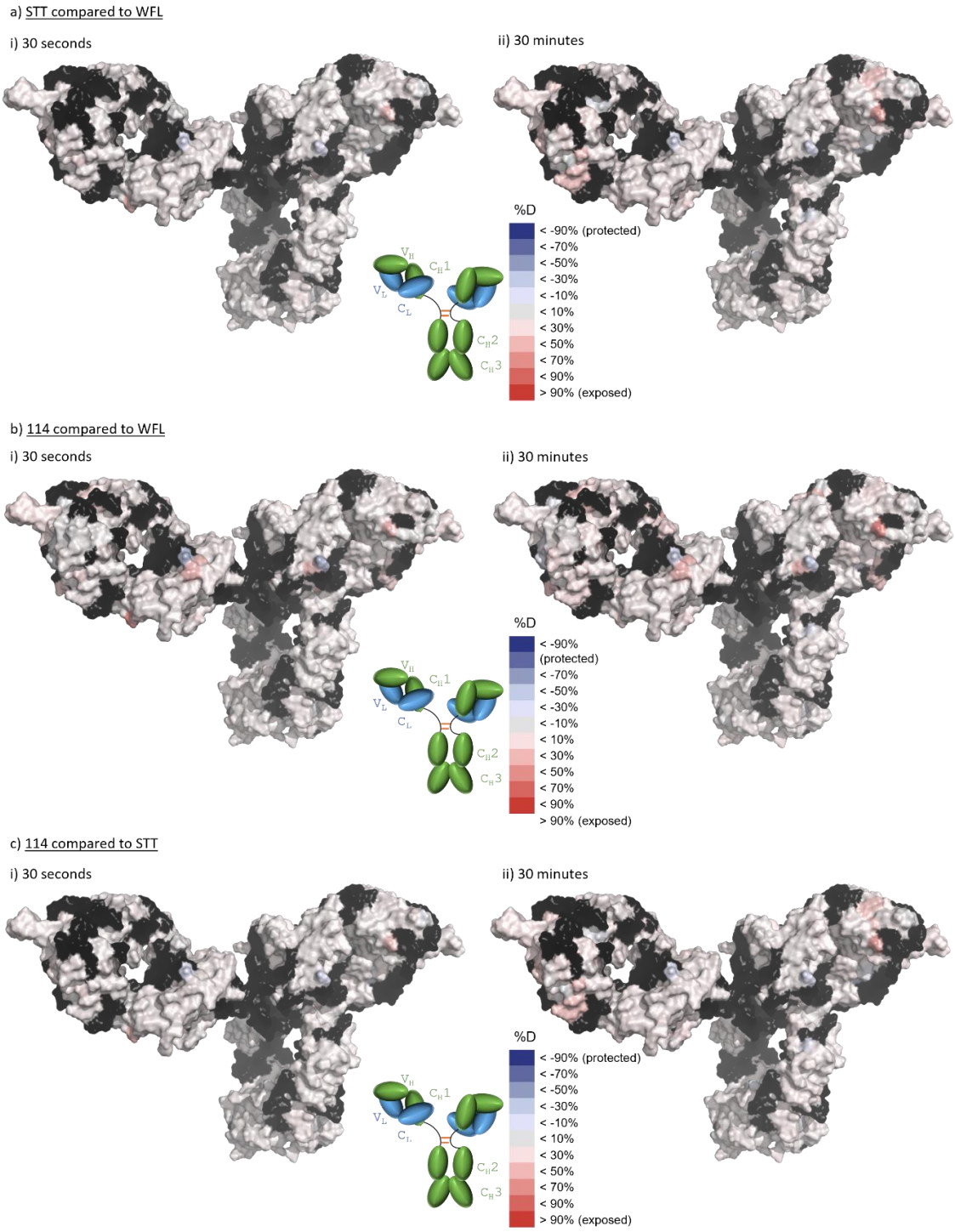


Figure 3.23 Deuterium uptake differences represented in coloured PyMol structures for WFL, STT and 114.

Differences in deuterium uptake at amino acid resolution between two mAbs (a) WFL – STT; b) WFL – 114; c) STT – 114), at two time points (i) 30 seconds; ii) 30 minutes). Each amino acid is coloured to represent differences in deuterium uptake at each time point, where more blue = less uptake of deuterium in the second mAb of the pair compared to the first, and red = more uptake of deuterium in the second mAb of the pair compared to the first. Black = sequence with no coverage. Energy-minimised structural prediction model of STT generated by Dr Romina Hofele (AstraZeneca) with Schrodinger, pH 6.0.

Overall, the HDX data support the hypothesis that the structure of the mAbs are globally similar when incubated in deuterated formulation buffer. The application of flow to these proteins, and how their structure is subsequently changed, will be captured by FPOP and will be the subject of the final results chapter. The next chapter (Development of an in-house method for FPOP label fingerprinting of monoclonal antibodies) follows the method development to enable the use of an in-house FPOP-LC-MS/MS method.

3.6 Discussion

In this chapter, several approaches to probe the biochemical characteristics of WFL, STT and 114 were employed to investigate the impact of mutations on their biochemical characteristics and stability. The interaction of WFL with industry-standard silica-based SEC columns in comparison to STT which does not, is in agreement with previous studies (C. L. Dobson *et al.*, 2016). The additional four mutations in 114 have been shown to mitigate the impact of the WFL patch, through HP-SEC and SMAC, but to a lesser extent in HIC, indicating the hydrophobic character of the WFL in fact plays a role in native mAb surface chemistry, and that colloidal stability may stem from the impact of other residues in the mAb surface.

The two heavy chain mutations in 114 (F29S and I52N) highlight two positions of sequence which have previously been identified as part of the mutational hotspot of WFL's V_H and V_L chains in scFv format (Ebo, Saunders, *et al.*, 2020). The authors identified 12 residues which had a high level of mutational frequency when the scFv sequence was subjected to directed evolution. To isolate the influence of the positions further, each of the 12 hotspot residues were mutated individually to the most common amino acid substitution at that position, and the TPBLA score was ranked (Figure 3.24). 11 out of 12 of the mutations increase the score to values above that of WFL. STT's score is also provided for comparison.

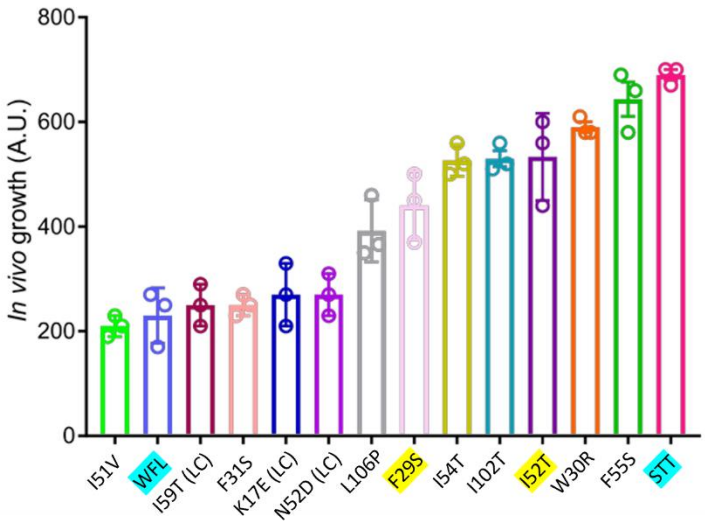


Figure 3.24 Ranked in vivo growth scores from the TPBLA survival curves of 14 scFv sequences.

WFL and STT are highlighted in blue; substitution positions of interest (V_H 29 and 52) are highlighted in yellow. Error bars indicate the standard error of the mean (n = 3 biologically independent experiments). Figure adapted from (Ebo, Saunders, et al., 2020).

The phenylalanine at F29 was most frequently substituted to a serine, and this mutation also arose in the variant 114. This supports the suggestion that a more hydrophilic residue in this position is beneficial for the improved TPBLA score. The point mutation at position I52 was most frequently mutated to a threonine in the screen, and although in 114 this position mutated to an asparagine, both of these amino acids are polar and uncharged, again reducing the overall hydrophobicity at these positions. There is a clear correlation of pressure to mutate hydrophobic amino acids in the insoluble patch on WFL, reducing hydrophobicity in the hotspot region on the V_H in 114, which may explain the drastic difference in retention on the HPLC column. The precise substituted residue does not seem to matter, rather the residue substituted should have a polar uncharged side chain replacing a hydrophobic one.

Overall, the data are in agreement that aggregation-prone WFL tends to stick to hydrophobic surfaces, such as silica HPLC column and protein-like (including itself) through HP-SEC, AC-SINS and DLS. Its aggregation behaviour due to protein surface chemistry is supported by the mechanical stress data where the apparent half-time of WFL monomer loss is over two times faster than STT and 114 (Figure 3.14). Flow data in this chapter does not reveal the flux of protein aggregation through unfolded intermediates, and whether the interactions are triggered by surface-mediated unfolding or bulk-mediated unfolding (Willis, manuscript in preparation). However, as the flow-induced aggregation of WFL is more pronounced than STT or 114 in non-

protective conditions, its structure may more likely be perturbed than STT or 114 in a protective buffer. Following from this, it could be postulated that WFL would be more likely than STT or 114 to show flow-induced changes in the FPOP labelling procedure, due to its aggregation-prone nature in biophysical studies in this chapter.

To fully determine the transitions the mAbs make in response to temperature, it would be useful to perform DSC to observe the change in specific heat for each transition (C_P). This has been previously used to identify the order of transitions of unfolding of C_H2, Fab and C_H3 regions when mutations in the Fab regions change the stability of the unfolding of the Fc (Garber & Demarest, 2007; Ionescu *et al.*, 2008). The cooperative unfolding of the Fab Ig domains generally gives a maximal C_P value about three times greater than the C_H2 or C_H3 domains due to its size (Garber & Demarest, 2007), which cannot be extracted from epi-fluorescence of tryptophan exposure. From the DSF experiments above, it is unclear whether the mutations in 114 compared to WFL or STT have changed the thermal stability of the Fab in respect to each other as the DSF here was performed on the full mAb (Figure 3.10 and Figure 3.11).

Arginine is commonly used in protein formulations to suppress aggregation during protein storage, by neutralising opposite charges and masking hydrophobic regions (Jain *et al.*, 2017). This observation was demonstrated by a reduction in red shift (lower nm) for 114 compared to the highly aggregation-prone WFL in arginine-containing buffer in AC-SINS (Figure 3.7). The AC-SINS of 114 displayed a red shift below the developability limit, with a similar value to STT, even with the W30 and F31 residues in the V_H hotspot, supporting the hypothesis that the peripheral hydrophilic mutations in 114's HC can counteract the molecule's overall propensity to stick to itself. Further experiments with non-arginine-containing solution 150 mM ammonium acetate pH 6.0 saw all three mAbs having a higher propensity to aggregate under extensional flow stress (Figure 3.14) than in formulation buffer (Figure 3.15 and Figure 3.16), supporting the role of arginine as a protector from aggregation. Proteins stored in the formulation buffer for up to 1800 seconds showed little to no disruption to fractional uptake during the time scale of the hydrogen-deuterium exchange experiments, supporting the buffer as an aggregation-suppressor over time (Figure 3.18, Figure 3.19, Figure 3.20, Figure 3.23).

The HDX analysis was performed to probe structural stability corresponding to the timescale of the proposed FPOP experiment and longer – 200 passes at 16 mm s^{-1} takes 10 minutes in the flow device (Willis *et al.*, 2018), whereas the mAbs were observed up to 30 minutes. The results demonstrate that under the timescale of the experiment, the difference in the %D uptakes between the pairs of mAbs are globally similar (Figure 3.18, Figure 3.19, Figure 3.20, Figure 3.23),

demonstrating that the structure of the mAbs are globally similar over the observation interval. This supports the hypothesis that, under ambient conditions without flow stress, the majority of the mAb structure remains native. Additionally, the HDX experiments highlighted a short section of amino acids in the V_HCDR3 of STT and 114 which exchange to a higher extent than WFL after 300 seconds in deuterated formulation buffer (Figure 3.21). This is consistent with previous HDX and chemical cross-linking data which show HC CDR3 as the site of Fab-Fab interaction in WFL (J. Dobson *et al.*, 2017). In the timescale of this experiment, the HC CDR3 of WFL could be transiently self-associating with other WFL molecules and shielding increases in deuterium uptake. As 114 and STT are less prone to self-association, their HC CDR3 are more freely accessible to the solvent, and show greater %D uptake over the experimental timescale. 114 has the greatest change in comparative %D uptake when compared to WFL at earlier time points (60 seconds and 300 seconds) in the experiment, which is seen to a lesser extent in STT (Figure 3.21). This may correlate to STT and 114 having improved biophysical properties over WFL. How the time-resolved dynamics elucidated by HDX in native conditions might compare to the perturbation of structure captured by flow followed by FPOP will be the subject of the next few chapters.

In conclusion, the three mAbs in this chapter display differing biophysical properties in the variety of experiments performed, and also exhibited varied aggregation behaviour in response to hydrodynamic force. It is therefore essential to develop a sensitive labelling method which can be applied to mAbs, either native or have been exposed to flow force, to capture any subtle structural perturbations or rearrangements. The photo-inducible reaction of the label in FPOP will be suited for capturing any transient changes to structure. Additionally, the use of mass spectrometry as a tool to identify the positions of labels will be highly valuable for pinpointing the peptides and residues responsible. The design of an in-house FPOP-LC-MS/MS method for this purpose will form the subject of the next chapter.

Chapter 4

Development of an in-house method for FPOP label fingerprinting of monoclonal antibodies

---

4 Development of an in-house method for FPOP label fingerprinting of monoclonal antibodies

4.1 Objectives

The preliminary objective of this thesis was to develop a method to identify areas of biopharmaceutical molecules which have a propensity to unfold due to flow forces. The covalent labelling technique FPOP was chosen as an approach to fix oxygen labels to solvent-exposed amino acid side chains, creating a footprint of the accessible surface area identifiable by MS sequencing. The identification of residues at a peptide- and residue-level resolution stems from the development of a robust, accurate and reproducible bottom-up protein sequencing method (LC-MS/MS) that obtains the highest possible sequence coverage with a high quality dataset. This method will then be applied to the biotherapeutic mAbs studied in the thesis (and indeed other proteins of interest) to investigate, and even correlate, subtleties in primary sequences which confer differing aggregation properties to the resulting tertiary structure.

The suitability of FPOP for capturing any differences in surface exposed regions of related proteins, for example between WFL, STT and 114, before or after exposure to hydrodynamic force in the flow device (detailed in Section 1.1.4.4: Extensional Flow Device (EFD) to mimic aggregation under flow) is apparent. However, the subsequent protein digestion protocol and the intricacies of how to analyse and interpret the output data still need careful optimisation. This chapter aims to address the following:

1. Optimise the sequence coverage obtained in our laboratory of unlabelled mAb STT, to identify digestion and LC-MS/MS conditions ideal for peptide mapping the mAbs in the study.
2. Develop FPOP-LC-MS/MS methodology for studying biopharmaceutical protein structure.

4.2 Optimising protein digestion and LC-MS/MS methodology

4.2.1 Standard in-solution digestion method validation

Irrespective of any FPOP modification on a protein surface, it is necessary to obtain a high coverage of the overall protein sequence in a robust, reproducible LC-MS/MS methodology. STT was provided by AstraZeneca for the optimisation process in this chapter.

A standard protocol for protein digestion was investigated for its suitability to digest STT (summary of the protocol is in Figure 4.1). This protocol was used for peptide mapping of any protein submitted to the Mass Spectrometry facility in the University of Leeds and was a starting point for my optimisation of the digestion process.

Briefly, unmodified STT at 0.1 mg mL^{-1} in formulation buffer was exposed to dithiothreitol (DTT) for 1 hour, 56 ˚C, shaking at 600 rpm (final concentration of DTT was 10 mM). The sample was allowed to cool to room temperature and the resulting free thiols were alkylated with iodoacetamide for 45 minutes in the dark, 20 ˚C, shaking at 600 rpm (final concentration of iodoacetamide was 55 mM). A 1:50 w/w ratio of trypsin:protein was then added and the sample was incubated for 16 hours, 37 ˚C, shaking at 600 rpm. The resulting tryptic peptides were subjected to a clean-up step using Waters Sep-pak SPE C18 cartridge, and the eluted peptides were dried in a EZ-2 Personal Evaporator. Peptides were reconstituted with 0.1% trifluoroacetic acid (TFA) to 1 µM final concentration and 1 µL was injected onto an Easy-nLC 1000 LC system equipped with a custom made 30-cm capillary emitter column, inner diameter 75 µm, packed with 3 µm Reprosil-Pur 120 C18 medium, and coupled to a LTQ Velos Orbitrap. Peptides were separated by an elution gradient of 1-30% v/v MeCN in H_2O over 45 minutes at 0.3 µL min^{-1}, where both mobile phases contained 0.1% v/v formic acid. Using these conditions, I tested the ability of the un-optimised digestion procedure to create peptides identifiable in the peptide mapping analysis software PEAKS v10.6, where trypsin was specified to be semi-selective, allowing up to 4 missed cleavages. The inherent mass accuracy of the LTQ Velos Orbitrap when sequencing parent ions by MS/MS was specified as 15 ppm during the search (Figure 4.1).

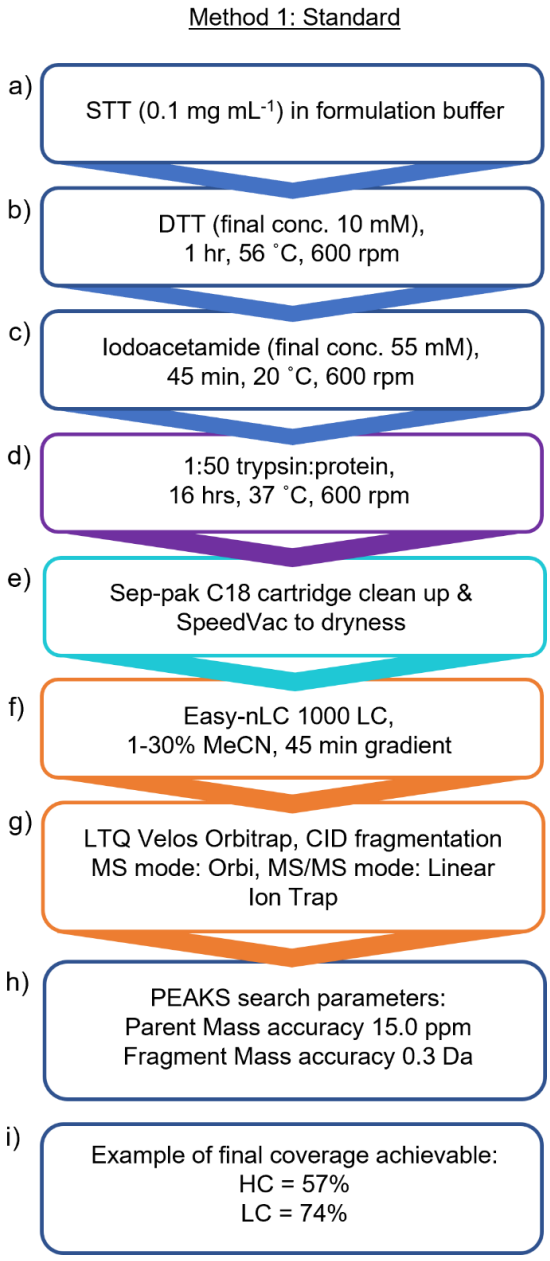


Figure 4.1 Method 1: Summarised standard in-solution digestion method conditions.

a-c) STT was subjected to reduction and oxidation (solution-based protein = blue steps). d) Protein was subjected to digestion (solution-based peptides = purple step). e) The resulting peptides were cleaned, dried, and reconstituted in LC-MS compatible solution (clean-up of peptides = aquamarine). f-i) 1 µM was injected on the Easy-nLC 1000 LC system coupled to the LTQ Velos Orbitrap and analysed using PEAKS v10.6 (LC-MS = orange; analysis = deep blue).

The peptide map of STT was generated using PEAKS v10.6, and the percentage coverage by detected peptides was HC = 57% and LC = 74% (Figure 4.2). Carbamidomethylation (the addition of +57.02 Da due to free cysteine side chain reacting with iodoacetamide, preventing disulfide reformation) is assumed to be a fixed modification in all data analysis. The detected peptides

cover ten of the twelve hotspot residues (residues are marked with an *, Figure 4.2). These hotspot residues have been previously identified as contributing to hydrophobic patches on the surface of this family of mAbs (Figure 1.18 114 is structurally similar to WFL and STT.), and are important for the aggregation properties of the mAbs (Ebo, Saunders, *et al.*, 2020). Additionally, the peptides generated from this method cover the C_L-C_H1 interface, a section of mAb surface area which has previously shown significantly reduced modification in WFL compared to STT by FPOP-LC-MS/MS (Cornwell *et al.*, 2019). The modified residues were V136 and P138 in C_H1; W152 and an unidentified position between P186-K190 in C_L, and all of these residues were covered by peptides here.

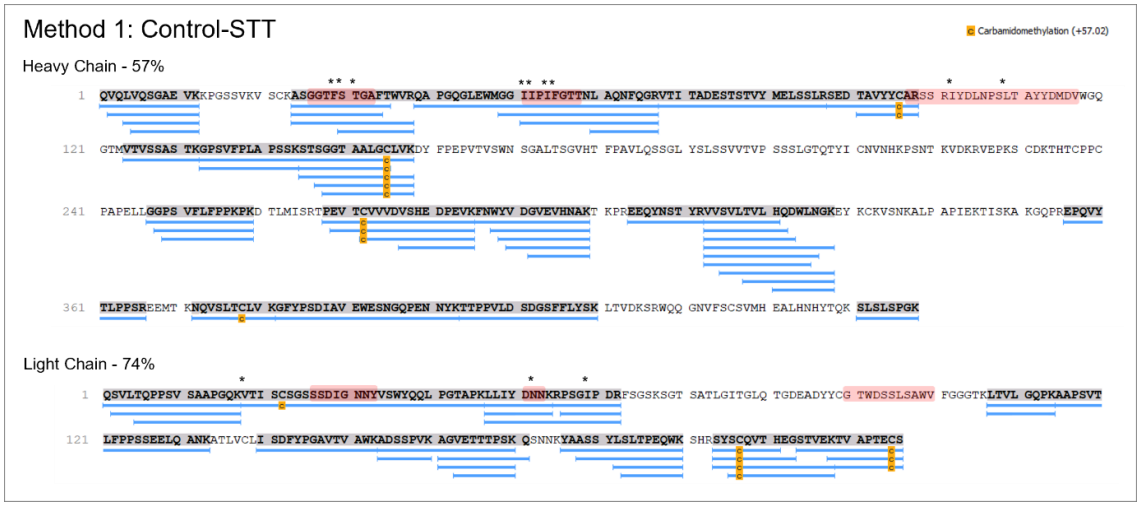


Figure 4.2 Method 1 example peptide coverage map for Control-STT following in-solution tryptic digestion and analysis using the LTQ Velos Orbitrap.

*Identified peptides for the full Heavy Chain (upper) and the full Light Chain (lower) are shown as blue lines below the sequence. CDR regions 1, 2 and 3 for each chain are highlighted in red. Hotspot residues are marked with an *. Data were obtained on the Easy-nLC 1000 LC system coupled to the LTQ Velos Orbitrap.*

However, the CDR3 regions in HC and LC (highlighted in red in Figure 4.2) were not covered by peptides, which means there are hotspot residues in the HC which cannot be investigated if this digestion method is taken forward. The lack of sequence coverage here may be due to these regions not being efficiently digested in the procedure because of their hydrophobic character which may cause self-association. The lack of coverage could also arise from the un-optimised capabilities of the Velos Orbitrap mass spectrometer using this protocol; as the peptides elute from a short home-packed LC column, their elution times may overlap and therefore the whole

packet of ions may not be fragmented efficiently before the next packed is injected. Therefore the MS/MS fragmentation may be incomplete, leading to the software being unable to assign peptide sequences with high enough confidence.

It was apparent that the method of LC-MS/MS needed to be optimised for higher coverage, as a greater coverage of the mAbs will allow a more complete picture when localising any oxidation modifications made to amino acids during the FPOP-LC-MS/MS experiments in the final results chapter of this thesis. Ballpark figures of at least 80% heavy chain and 95% light chain were personally set as targets for overall coverage of the peptides. I set the prediction for the heavy chain coverage to be lower than that of the light chain. This was because there is a low density of lysines and arginines over the hinge region (between the C_H1 and C_H2 of the heavy chain), and therefore when observing the peptides expected from the digestion protocol using online *in vitro* digestion prediction software ExPASy Peptide Mass, it was anticipated that I would lose around 20% of the sequence here, as the peptide lengths would be too long for efficient ionisation and subsequent detection in the MS procedure.

A Thermo Scientific Orbitrap Exploris 240, with inherently higher ppm ion mass accuracy, was obtained by the Mass Spectrometry facility in the University of Leeds and then was next used for the FPOP-LC-MS/MS experiments. The method in Figure 4.1 was used as an initial starting point when optimising STT digestion and the peptides were analysed in the Exploris 240.

4.2.2 Adapting the standard in-solution digestion method for new LC-MS instrumentation

The assumed accuracy of the Thermo Orbitrap Exploris 240 was set to 10 ppm (where previously the inherent ppm accuracy of the LTQ Velos Orbitrap was 15 ppm) when searching for peptide matches in PEAKS v10.6. These ppm values are limits which are manually set to avoid false positive identifications in the analysed data set, but are more conservative than the maximum achievable accuracy of each instrument to avoid missing data. Additionally, the MS/MS mass accuracy for Exploris was set to 0.05 Da (where previously the inherent accuracy of the LTQ Velos Orbitrap was 0.3 Da). Overall, the improved accuracy with the new instrument adds confidence to the assigned MS/MS sequencing when using PEAKS.

STT (at 0.1 mg mL^{-1} in formulation buffer) was exposed to the reduction, alkylation, digestion and clean up procedure as in Figure 4.1 a-e) but analysed using an Ultimate 3000 RSLCnano LC system coupled to a Thermo Orbitrap Exploris 240. The peptides were separated by an elution

gradient of 1-30% v/v MeCN in H_2O over 60 minutes (previously 45 minutes) at 0.3 µL min^{-1}, where both mobile phases contained 0.1% v/v formic acid. This method is summarised in Figure 4.3. The resulting peptide coverage map (Figure 4.4) illustrates a HC coverage of 39% and LC coverage of 62%. Again, the coverage map succeeds in covering the C_L-C_H1 interface, but does not cover the CDR regions (highlighted in red), or the twelve hotspot residues (residues are marked with an *).

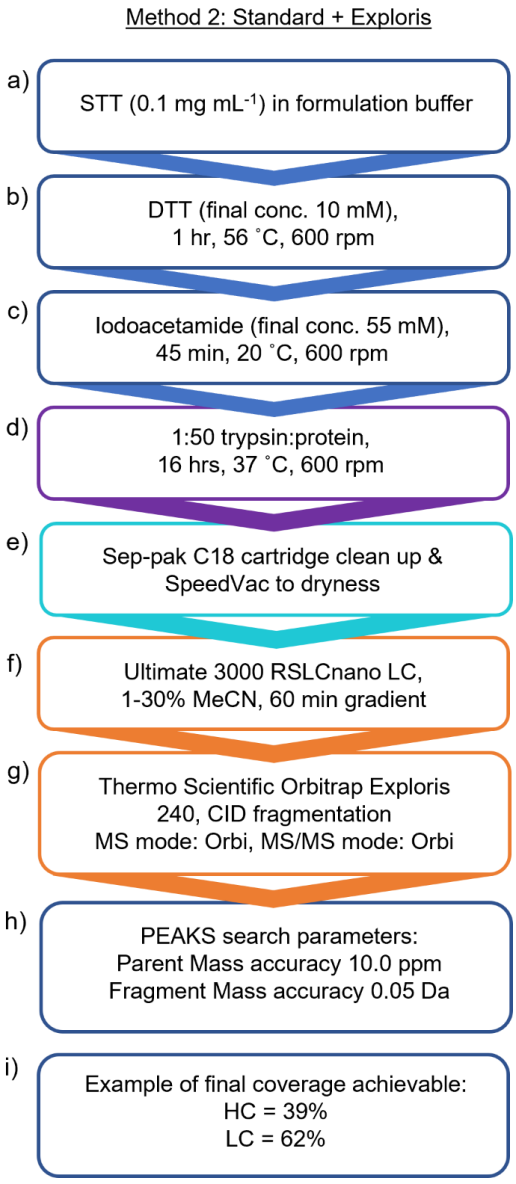


Figure 4.3 Method 2: Summarised standard in-solution digestion method conditions, using the Exploris 240 mass spectrometer.

a-d) STT was subjected to reduction, oxidation and digestion (solution-based protein = blue steps; solution-based peptides = purple step) e) The resulting peptides were cleaned, dried, and reconstituted in LC-MS compatible solution (clean-up of peptides = aquamarine). f-i) 1 µM was injected on the Ultimate 3000 RSLCnano LC system coupled to the Exploris 240 Orbitrap and analysed using PEAKS v10.6 (LC-MS = orange; analysis = deep blue).

Figure 4.4 Method 2 example peptide coverage map for Control-STT following in-solution tryptic digestion and analysis using the Exploris 240 Orbitrap.

*Identified peptides for the full Heavy Chain (upper) and the full Light Chain (lower) are shown as blue lines below the sequence. CDR regions 1, 2 and 3 for each chain are highlighted in red. Hotspot residues are marked with an *. Data were obtained on the Ultimate 3000 RSLCnano LC system coupled to the Exploris 240 Orbitrap and analysed using PEAKS v10.6.*

It was postulated that the low coverage of STT was due to incomplete digestion of the antibody into tryptic peptides. It is important that the IgG is denatured, unfolded and reduced sufficiently prior to digestion in order to permit trypsin access to every region of the protein. Previous attempts to improve digestion of an industry-standard antibody NISTmAb involved investigating the use of chaotropic denaturing reagents in the buffer before reduction with DTT (Mouchahoir & Schiel, 2018). Therefore, we next investigated using industry-standard digestion protocols as a means of improving coverage. It was important to investigate whether the use of chaotropic agents in the reduction buffer would increase reduction efficiency, and additionally alkylation and digestion efficiency.

4.2.3 Applying an industry protocol to improve digestion efficiency

Biopharmaceutical products are often formulated at high concentrations for storage and transport (50 – 150 mg mL^{-1}) (Baek & Zydney, 2018), and diluted before administration. However during transport, these highly concentrated samples may be exposed to mechanical stress and aggregation may occur. Aggregation can cause issues with insolubility of sample or loss of yield. Therefore, industry can use a variety of methods to monitor sample aggregation (e.g. SEC, DLS). Additionally, an important quality control procedure to identify any post-translational

modifications that could appear during storage (such as spontaneous oxidation) is to subject sample to tryptic digestion and LC-MS/MS analysis. It is crucial to monitor and avoid unwanted post-translational modifications in order to comply with regulations and for the reliability of repeat assessments of the molecule's potency (Grassi & Cabrele, 2019).

For routine testing of biotherapeutics in industrial development, a peptide mapping protocol needs to be reproducible and robust between all varieties of protein platforms including mAbs, bispecifics and antibody-drug conjugates (Fekete, Guillarme, *et al.*, 2016). Often proteins are placed in denaturing buffer conditions to increase the likelihood of offering good sequence coverage (above 80% HC and 95% LC) of any format, without any need to deviate from a standard protocol between different platforms and their formulations. The chemical denaturing solution conditions are designed to be robust to work with a wide variety of protein structures in a wide variety of formulations, even if this means the protocol is not optimised for particular systems.

AstraZeneca perform their routine protein digestion protocol using a sample concentration of 5 mg mL^{-1}. This works well for industrial procedures with material being formulated at high concentrations as there is lots of available sample. However, this starting concentration was not feasible for my experiments, due to low sample availability, and the propensity for WFL to undergo concentration-induced self-association (see Figure 1.17, (C. L. Dobson *et al.*, 2016)). At 5 mg mL^{-1}, the concentration-dependent self-association would cause shielding/protection of regions of protein ordinarily surface-accessible in monomeric conditions in the FPOP workflow. Additionally the high protein concentration would serve as a self-scavenger, reducing the reactivity of the radicals. As the FPOP method relies on the hydrogen peroxide being distributed over the surface space for labelling all surface-exposed areas of the samples, a high concentration would cause WFL, STT and 114 to be in different dispersity conditions, making the experiment unreliable. Therefore it was imperative to develop an in-house LC-MS/MS method which digested a low concentration (0.1 mg mL^{-1}) of protein efficiently.

STT (at 0.1 mg mL^{-1} in formulation buffer) was subjected to the standard protocol kindly provided by AstraZeneca (summarised in Figure 4.5, Method 3). The original standard procedure used by AstraZeneca was designed to work with sample at 5 mg mL^{-1}, a concentration which is 50 times higher than is used for an in-house FPOP protocol. It was decided to apply the AstraZeneca method to a sample at 0.1 mg mL^{-1} to align with the concentration that will be used in future FPOP protocols in order to assess its effectiveness (Figure 4.5, Method 3). Additionally, a modified version of the protocol was designed (summarised in Figure 4.5, Method 4) with two

variations which aimed to decrease the dilution of the sample further during the protocol: sample was initially dried in order to be reconstituted in chemical denaturant and not diluted into it; and dialysed into 2M urea and not 6M which would require dilution afterwards. These improvements were successful in increasing the coverage of STT (HC = 31% and LC = 62% for Method 3; HC = 62% and LC = 75% for Method 4) (Figure 4.6). However the coverage of the HC and LC still ideally would need to be improved.

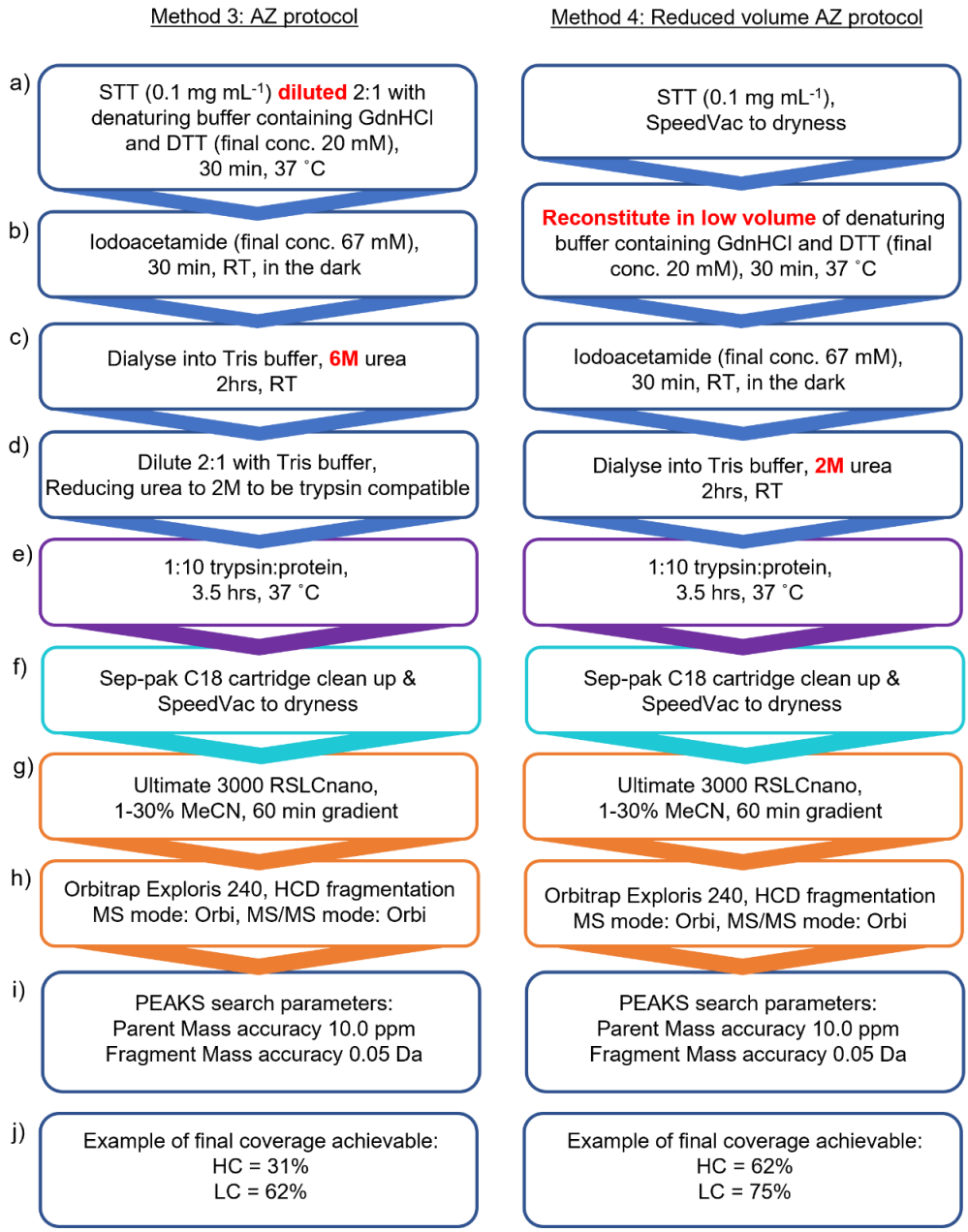


Figure 4.5 Methods 3 and 4: AZ protocols.

Method 3 was the original standard procedure provided by AstraZeneca, however with starting material at 0.1 mg mL^{-1}. Method 4 was the procedure in Method 3 with two points adjusted to reduce sample dilution. a-e) STT was subjected to reduction, oxidation and digestion. f) The resulting peptides were cleaned, dried, and reconstituted in LC-MS compatible solution. g-j) 1 µM was injected and analysed on the Ultimate 3000 RSLCnano LC system coupled to the Exploris 240 LC-MS/MS system and analysed using PEAKS v10.6. Key differences between Method 3 and 4 are in bold red. Solution-based protein = blue; solution-based peptides = purple; clean-up of peptides = aquamarine; LC-MS = orange; analysis = deep blue.

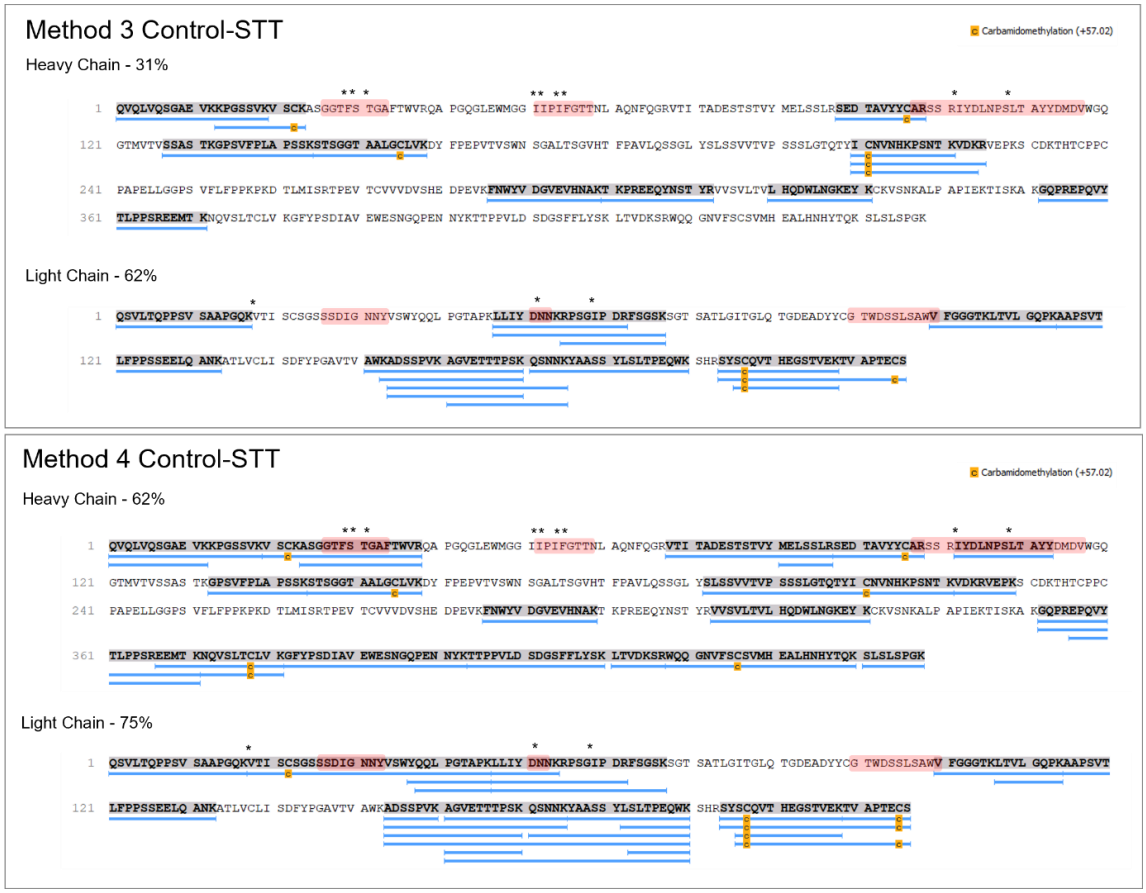


Figure 4.6 Methods 3 and 4 example peptide coverage map for Control-STT following industry-derived in-solution tryptic digestion protocols.

*Method 3 and Method 4 are described in Figure 4.5. Identified peptides for the heavy chain and the light chain are shown as blue lines below the sequence. CDR regions 1, 2 and 3 for each chain are highlighted in red. Hotspot residues are marked with an *. Data were obtained on the Ultimate 3000 RSLCnano LC system coupled to the Exploris 240 Orbitrap.*

The resulting coverage from Methods 3 and 4 did not improve over the standard in-solution digestion methods in Figure 4.1 and Figure 4.3. The main difference in the AstraZeneca protocols Method 3 and 4 was the use of chaotropic agents; guanidinium hydrochloride was used to aid unfolding of protein before reduction, and urea was to keep the reduced and alkylated protein from refolding before digestion. Therefore it was important to test the necessity of these reagents to prevent refolding using a control protein Bovine Serum Albumin, which contains 17 disulfide bonds. Control protein Bovine Serum Albumin (at 0.1 mg mL^{-1} in 150 mM ammonium acetate pH 7.4) was subjected to Method 3. Its sequence coverage was 52% with the AstraZeneca protocol. This result suggested that, even with a standard protein known to aid troubleshooting of digestion methods due to its general good digestion efficiency, the use of

low protein concentration with this method resulted in low coverage. It was postulated that peptides were being lost in the process after the digestion, perhaps in clean-up steps.

The LC-MS set up with the Ultimate 3000 RSLCnano LC system coupled to a Thermo Orbitrap Exploris 240 differs slightly to that in industry – our LC systems have not used guard columns before chromatography using a homemade column. In most configurations, a guard column precedes commercial columns to remove any salts or impurities before chromatography. Pre-columns also can be used to focus the peptide fraction of the loaded sample. As my system did not have a guard column, there needed to be an additional clean-up step using Sep-pak columns (seen in Figure 4.1 e; Figure 4.3 e; and Figure 4.5 f). It was postulated this could be contributing to the lower than expected sequence coverage when trying an industry-standard protocol in-house, as low abundance peptides from starting concentration 0.1 mg mL^{-1} could be susceptible to being lost in the clean-up step. Therefore overall, the resulting sequence coverages for both BSA and STT demonstrate that the industry protocol is not ideal for achieving high coverage under our laboratory conditions, and when the starting concentration is compatible with the FPOP procedure, 0.1 mg mL^{-1}. Thus, the industry protocol was no pursued further, but a variety of other methods which aid solubility of the mAbs with the aim of increasing digestion efficiency were screened.

4.2.4 Screening digestion protocols

A variety of protocols were screened for increased digestion efficiency. A summary of the work is listed in Table 4.1. The conditions covered three categories: in-solution digestion method with a pre-treatment of precipitation and subsequent re-solubilisation in Rapigest SF surfactant (Waters), which is a mild denaturant that solubilises and unfolds substrates and so could aid solubility of the IgG (Method 5); in-gel digestion and extraction of peptides (Method 6); and digestion on a column cartridge 'S-Trap (Protifi)' and subsequent elution (Method 7). Methods 5 and 6 require clean up with Sep-pak cartridges (such as in Methods 1-4 in the previous sections), whereas Method 7 does not, as the S-Trap cartridge itself is washed with the peptides bound, eluting impurities and salts, before subsequently collecting the cleaned peptides. This in-built clean up method provided the best coverage out of the methods screened (HC = 90% and LC = 69%, Table 4.1). The coverage maps are compared in Figure 4.7, where the Method 7 S-Trap protocol (Table 4.1 c) excelled in providing high coverage of STT.

Table 4.1 Sequence coverage comparisons for screened digestion protocols for STT at 0.1 mg mL^{-1}.

Method category	**Method description**	**Coverage (%)**
Method 5) In-solution with precipitation and surfactant pre-treatment	• Chloroform and methanol precipitation of STT (0.1 mg mL^{-1}). • Pellet resuspended with Rapigest surfactant to resolubilise STT. • Mixture was subjected to Method 2 (Figure 4.3), including Sep-pak clean up.	HC = 18 LC = 44
Method 6) In-gel digestion	• STT (0.1 mg mL^{-1}, 10 µL) ran through a denaturing SDS-PAGE gel. • Excised gel piece was subjected to Method 2 (Figure 4.3), including Sep-pak clean up.	HC = 66 LC = 68
Method 7) S-Trap digestion	• Sample was reduced and alkylated before trypsin was added. • S-Trap column was primed with trypsin. • Sample is added to column and incubated for 2 hours. • Sample was eluted from the column in MS-compatible buffer.	HC = 80 LC = 69

Figure 4.7 Methods Methods 5, 6 and 7: example peptide coverage maps for Control-STT following in-solution, in-gel, or on column tryptic digestion protocols.

*Method 5) Map of Control-STT which was precipitated and reconstituted in Rapigest surfactant before reduction, alkylation, digestion and clean-up. Method 6) Map after Control-STT was excised from SDS-PAGE gel and reduced, alkylated, digested and extracted, then cleaned up. Method 7) Map of Control-STT after in-solution reduction and alkylation, then digested on an S-Trap cartridge. Identified peptides for the heavy chain and the light chain are shown as blue lines below the sequence. CDR regions 1, 2 and 3 for each chain are highlighted in red. Hotspot residues are marked with an *. Data obtained on the Ultimate 3000 RSLCnano LC system coupled to the Orbitrap Exploris 240.*

I postulated that the sequences not covered in the peptide maps in Figure 4.7 a and b could be due to binding to the C18 matrix in the Sep-pak clean up columns and as a result the peptides could not be efficiently eluted from the column during the clean-up process. In accordance with this idea, the digestion using the S-Trap column method does not contain the additional clean up step and so the coverage obtained was higher. Following this, the S-Trap column digestion protocol was chosen as the base method of digestion of STT.

4.2.5 Optimising the LC-MS/MS methodology results in high sequence coverage of STT

After choosing S-Traps as the method of digestion and clean-up, the LC-MS/MS methodology was next targeted to improve the sequence coverage found by LC-MS/MS.

Early changes to the LC-MS/MS method from the initial starting point in Method 1 (Figure 4.1) to Method 2 involved the increase of the gradient elution from 45 minutes to 60 minutes. The shallower gradient of acetonitrile acts as an improved separation technique for resolving lowly abundant peptides. Additionally, 3 µL of peptides were injected (previously 1 µL) as the longer gradient causes dilution of peptides, and so the greater volume of sample loaded helped with signal increase and quality. After manipulating the protocol up to the digestion of the sample in to peptides, Method 7 using S-Traps yielded the best coverage so far.

Subsequently, further optimisation of the LC-MS/MS method coupled to the S-Trap digestion was set up to systematically test Method 8: a method with an elution gradient of 90 minutes (with a total run time of 120 minutes). Method 8 was used as a basis for systematically testing 3 parameters: a two-stage ramped elution gradient (Method 9), a modification to the apex detection parameter (Method 10), and a modification to the dynamic exclusion parameter (Method 11).

The adjusted methods were applied separately to analyse the same Control-STT S-Trap peptide sample. The experiment was designed this way in order to attribute improvements to the MS/MS to a particular change in condition. Each run on the LC was separated by a 30 minute column wash method (3 µL buffer load). Each mobile phase for LC contained 0.1% v/v formic acid. Overall, the key differences in Methods 1-11 are explained in Table 4.2.

Table 4.2 Methods 1-11: Optimising LC-MS/MS methodology

Method	**Volume loaded (µL)**	**Total run time (min)**	**LC elution gradient**	**Additional method adjustment**
1)	1	60 (0.3 µL min^{-1})	• 1% MeCN, 5 min (equilibration) • 1-30% v/v MeCN, 45 min • 80% MeCN, 10 min (column wash)	
2-7)	3	90 (0.3 µL min^{-1})	• 1% MeCN, 5 min (equilibration) • 1-30% v/v MeCN, 60 min • 80% MeCN, 10 min (column wash)	
8)	3	120 (0.5 µL min^{-1})	• 2% MeCN, 10 min (equilibration) • 2-30% v/v MeCN between 11 – 102 min • 80% v/v MeCN between 102 – 111 min (column wash) • 2% MeCN, 9 min (equilibration)	Dynamic exclusion: exclude after n = 1, for a duration of 3 seconds Apex detection: 10 seconds expected peak width
9)	3	120 (0.5 µL min^{-1})	• 2% MeCN, 10 min (equilibration) • 2-30% v/v MeCN between 11 – 72 min, • 30-60% v/v MeCN between 72 – 102 min • 80% v/v MeCN between 102 – 111 min (column wash) • 2% MeCN, 9 min (equilibration)	Ramped elution gradient in two stages
10)	3	120 (0.5 µL min^{-1})	Base method: 8	Apex detection set to 20 seconds expected peak width
11)	3	120 (0.5 µL min^{-1})	Base method: 8	Dynamic exclusion: exclude after n = 3 within 30 sec, for a duration of 5 sec.

The 120 minute run time (with 90 minute gradient) used in all four methods conferred a higher overall coverage (Table 4.3) than the previous 90 minute run time (with 60 minute gradient) (Table 4.1, Method 7). Methods 8 and 9 showed the highest coverage (Table 4.3).

When investigating the PEAKS output, the total number of MS/MS scans gathered in the Orbitrap per sample can be compared (#MS/MS scans). Additionally, it is possible to see the number of these MS/MS scans which the software has successfully assigned to part of the protein sequence (number of peptide precursor sequence matches, #PSM), and from this the number of peptides which are taken forward as being good enough quality in sequence to include in the coverage map (#Peptides and #Sequences). These are restricted from the initial #MS/MS scans due to the manually confined parameter of false discovery rate (FDR; for all PEAKS analyses, the FDR was set to 0.1%, as the lower the FDR percentage, the less chance that part of the sequence coverage has been assigned using a false positive MS/MS signal).

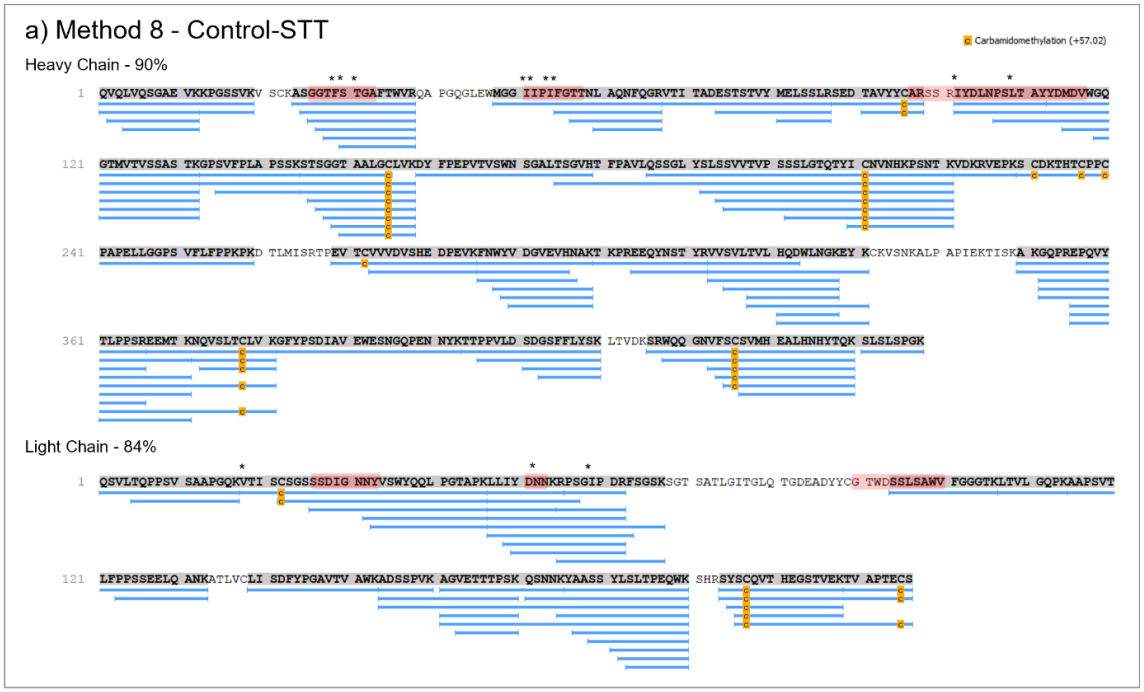
Comparing the overall LC-MS/MS outputs of the methods (Table 4.3), Method 11 had the largest number of MS/MS scans (column 3), but this did not confer the highest amount of confidently identified peptides used in the creation of the final coverage map (column 5; blue bars in Method 11, Figure 4.8). This could be due to the dynamic exclusion parameter being set to n = 3, meaning peptides with the same *m/z* value would need to appear 3 times during the period of 30 seconds in the LC elution profile to be placed on an MS/MS exclusion list for 5 seconds. This could be contributing to the same peptide being sequenced too many times unnecessarily, contributing to the high value of MS/MS scans (column 3) but no proportional increase in the resulting number of peptides at the end (column 5). Therefore, the dynamic exclusion should remain at n = 1, with a time duration of 3 seconds to exclude the highly abundant peptides from being sequenced too many times, allowing the sequencing of lowly abundant peptides.

The apex detection parameter is a prediction of the elution peak width based on the gaussian distribution of a peak over the baseline. By increasing this value, the time anticipated for a peak apex to be reached is longer, and so the peptides sampled for MS/MS are taken later in the elution profile. The longer apex detection duration in method 10 reduced the number of peptides scanned and subsequently the coverage dropped compared to Methods 8 and 9, supporting the use of a low apex detection value in the final method used.

The number of MS/MS scans for Method 9 was higher than Methods 8 and 10 (column 3), and also resulted in the greatest number of peptides taken forward for the coverage map (columns 4-6; blue bars in Method 9, Figure 4.8). Therefore, Method 9 was chosen for future application to FPOP-LC-MS/MS procedures.

Table 4.3 Comparison of sequence coverage between four differing LC-MS/MS methodologies.

Method	Coverage	#MS/MS scans	#PSMs	#Unique peptides	#Sequences
8	HC = 90 LC = 84	9662	610	175	133
9	HC = 81 LC = 99	10307	636	204	156
10	HC = 81 LC = 84	8620	556	154	121
11	HC = 84 LC = 84	33464	5408	179	138


a) Method 8 - Control-STT
Carbamidomethylation (+57.02)
Heavy Chain - 90%
1 QVQLVQSGAE VKKPGSSVKV SCKASGGTFS TGAFTWVRQA PGQGLEWMGG IIPIFGTTNL AQNFQGRVTI TADESTSTVY MELSSLRSED TAVYYCARSS RIYDLNPSLT AYYDMDVWGQ
121 GTMVTVSSAS TKGPSVFPLA PSSKSTSGGT AALGCLVKDY FPEPVTVSWN SGALTSGVHT FPAVLQSSGL YSLSSVVTVP SSSLGTQTYI CNVNHKPSNT KVDKRVEPKS CDKTHTCPPC
241 PAPELLGGPS VFLFPPKPKD TLMISRTPEV TCVVVDVSHE DPEVKFNWYV DGVEVHNAKT KPREEQYNST YRVVSVLTVL HQDWLNGKEY KCKVSNKALP APIEKTISKA KGQPREPQVY
361 TLPPSREEMT KNQVSLTCLV KGFYPSDIAV EWESNGQPEN NYKTTPPVLD SDGSFFLYSK LTVDKSRWQQ GNVFSCSVMH EALHNHYTQK SLSLSPGK
Light Chain - 84%
1 QSVLTQPPSV SAAPGQKVTI SCSGSSSDIG NNYVSWYQQL PGTAPKLLIY DNNKRPSGIP DRFSGSKSGT SATLGITGLQ TGDEADYYCG TWDSSLSAWV FGGGTKLTVL GQPKAAPSVT
121 LFPPSSEELQ ANKATLVCLI SDFYPGAVTV AWKADSSPVK AGVETTTPSK QSNNKYAASS YLSLTPEQWK SHRSYSCQVT HEGSTVEKTV APTECS
b) Method 9 - Control-STT
Carbamidomethylation (+57.02)
Heavy Chain - 81%
1 QVQLVQSGAE VKKPGSSVKV SCKASGGTFS TGAFTWVRQA PGQGLEWMGG IIPIFGTTNL AQNFQGRVTI TADESTSTVY MELSSLRSED TAVYYCARSS RIYDLNPSLT AYYDMDVWGQ
121 GTMVTVSSAS TKGPSVFPLA PSSKSTSGGT AALGCLVKDY FPEPVTVSWN SGALTSGVHT FPAVLQSSGL YSLSSVVTVP SSSLGTQTYI CNVNHKPSNT KVDKRVEPKS CDKTHTCPPC
241 PAPELLGGPS VFLFPPKPKD TLMISRTPEV TCVVVDVSHE DPEVKFNWYV DGVEVHNAKT KPREEQYNST YRVVSVLTVL HQDWLNGKEY KCKVSNKALP APIEKTISKA KGQPREPQVY
361 TLPPSREEMT KNQVSLTCLV KGFYPSDIAV EWESNGQPEN NYKTTPPVLD SDGSFFLYSK LTVDKSRWQQ GNVFSCSVMH EALHNHYTQK SLSLSPGK
Light Chain – 99%
1 QSVLTQPPSV SAAPGQKVTI SCSGSSSDIG NNYVSWYQQL PGTAPKLLIY DNNKRPSGIP DRFSGSKSGT SATLGITGLQ TGDEADYYCG TWDSSLSAWV FGGGTKLTVL GQPKAAPSVT
121 LFPPSSEELQ ANKATLVCLI SDFYPGAVTV AWKADSSPVK AGVETTTPSK QSNNKYAASS YLSLTPEQWK SHRSYSCQVT HEGSTVEKTV APTECS

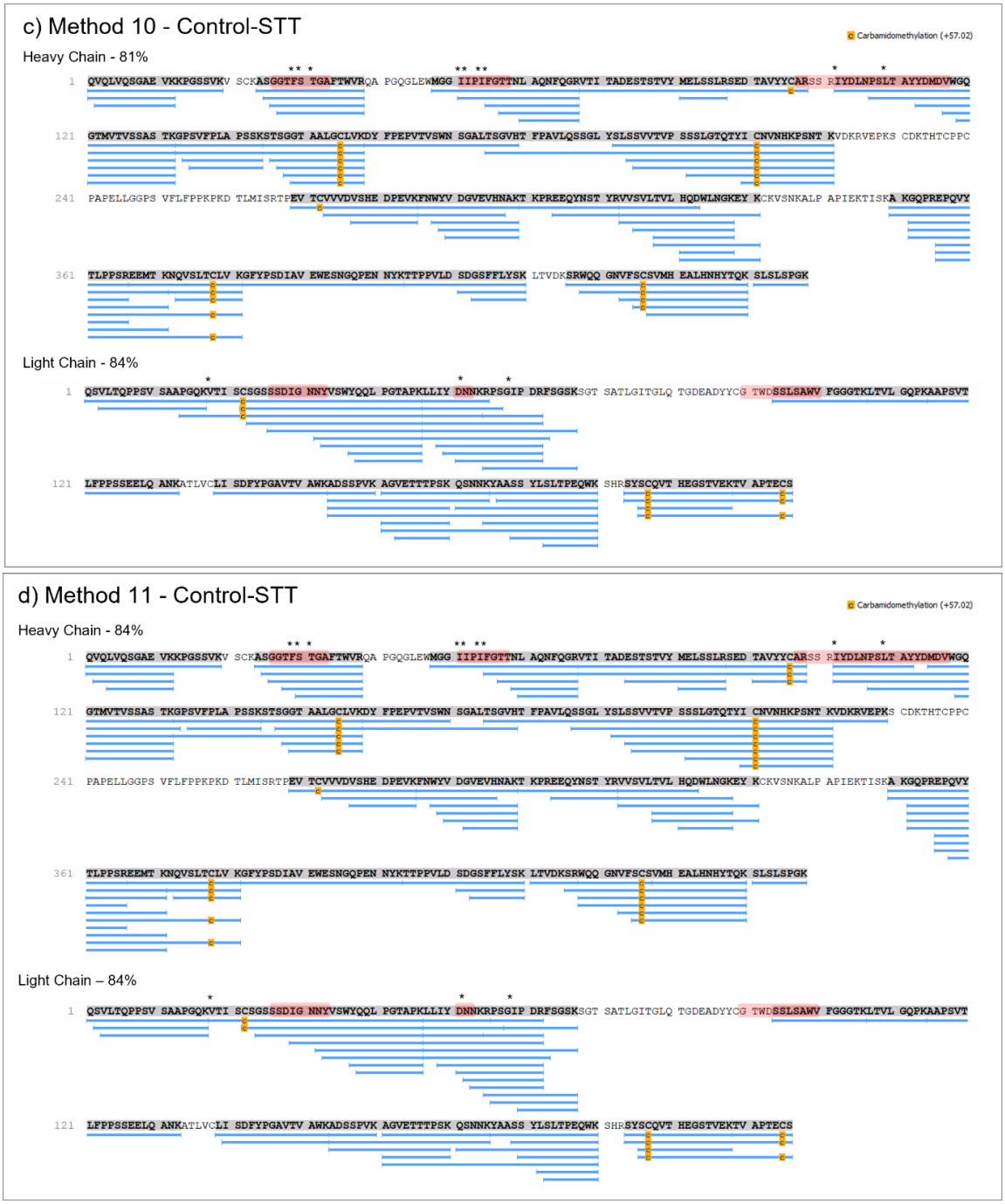


Figure 4.8 Methods 8, 9, 10 and 11: example peptide coverage maps demonstrating the difference in coverage between four LC-MS/MS method iterations.

*a) Method 8. b) Method 9. c) Method 10. d) Method 11. Identified peptides for the heavy chain and the light chain are shown as blue lines below the sequence. CDR regions 1, 2 and 3 for each chain are highlighted in red. Hotspot residues are marked with an *. Data were obtained on the Ultimate 3000 RSLCnano LC system coupled to the Exploris 240 Orbitrap.*

Overall, this chapter has explored a wide range of parameter space, digestion methodology and LC-MS/MS methodology. The differences in the resulting method compared to the initial un-optimised method used with the LTQ Velos are shown in Figure 4.9. The optimised method demonstrates that I can detect almost all of the sequence, which gives the highest possible chance for the detection of unique FPOP modifications in an FPOP-LC-MS/MS protocol.

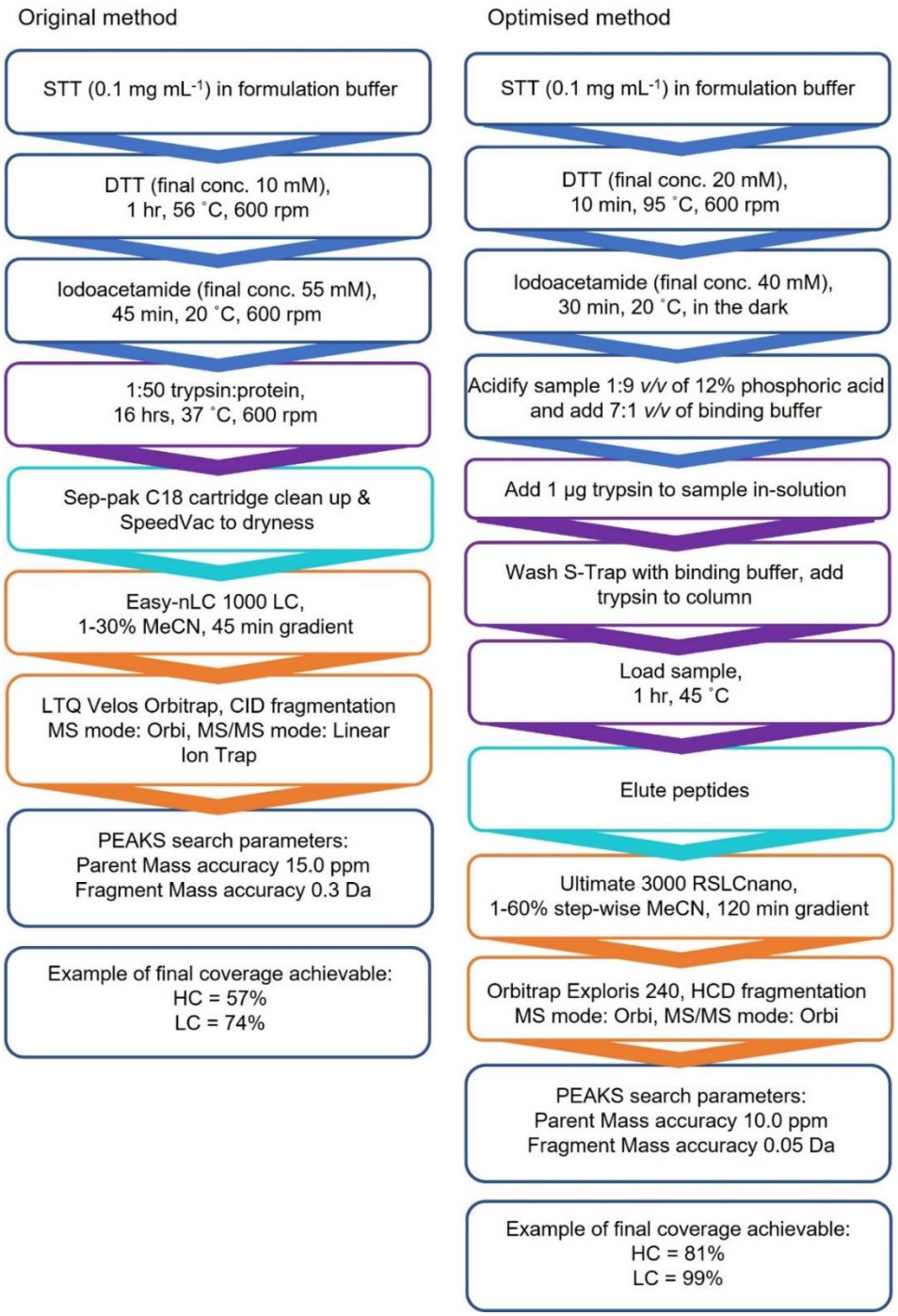


Figure 4.9 Summarised original standard in-solution digestion method conditions compared to the optimised conditions, using S-Trap digestion.

Solution-based protein = blue; solution-based peptides = purple; clean-up of peptides = aquamarine; LC-MS = orange; analysis = deep blue.

4.3 Oxidative labelling of STT coupled to the optimised LC-MS/MS method

The first aim of this work was to develop a robust in-house method for sequence mapping of proteins. With this methodology finalised in Figure 4.9, the next step was to subject STT to oxidative labelling by FPOP, and analyse with the optimised LC-MS/MS procedure. It was important to confirm that the labelled sample yields similar sequence coverage to unlabelled sample.

STT was subjected to oxidation by FPOP in its formulation buffer. The FPOP procedure is detailed in Method Section 2.2.8.3: Fast Photochemical Oxidation of Proteins – LC-MS/MS. Briefly, STT was prepared to 0.1 mg mL^{-1} in formulation buffer containing a final concentration of 10 mM L-histidine as a scavenger. Immediately before irradiation, 3 µL of 5% H_2O_2 was added before loading into a syringe connected to the capillary in the path of the laser. The sample was irradiated by the laser by flowing through a capillary at 20 µL min^{-1} and collected in a Eppendorf containing 20 µL quench solution (100 mM L-methionine and 1 µM catalase in formulation buffer). The amount of peroxide used was necessary to overcome the radical scavenging effect of the arginine excipient in the formulation buffer (Cornwell, 2019). This quenched sample has a final mAb concentration of 0.08 mg mL^{-1} (0.54 µM). Control-STT sample was prepared at the same time as FPOP-STT in the same way containing H_2O_2 for the time of the laser-exposed experiment, but not exposed to the laser itself, and also quenched with 20 µL quench solution. The laser power is kept constant at 110 mJ, with a firing frequency of 15 Hz and a pulse duration of 20 ns. The beam width was approximately 3 mm and unchanged between experiments.

4.3.1 Native Mass Spectrometry confirms global FPOP modifications

20 µL of irradiated sample was exchanged into 150 mM ammonium acetate pH 6.0 buffer overnight by dialysis (Methods Section 2.2.1.2: Dialysis). The control sample without being exposed to the FPOP laser was also dialysed into ammonium acetate. The sample was then directly infused into a Q-Exactive UHMR mass spectrometer (Thermo Scientific) and native spectra were obtained (Figure 4.10). The overall charge state distribution remains similar suggesting no change in overall structure. The glycoform pattern is seen in both control and modified sample. The addition of FPOP modifications in Figure 4.10 b has resulted in a higher baseline and less clear distinction between glycoform masses, and resulted in trailing edges to each charge state mass, showing there is heterogeneity to the amount of oxidation each mAb has experienced. Additionally the peak position of each charge state in the distribution has shifted to a higher mass, indicating additions of FPOP modifications.

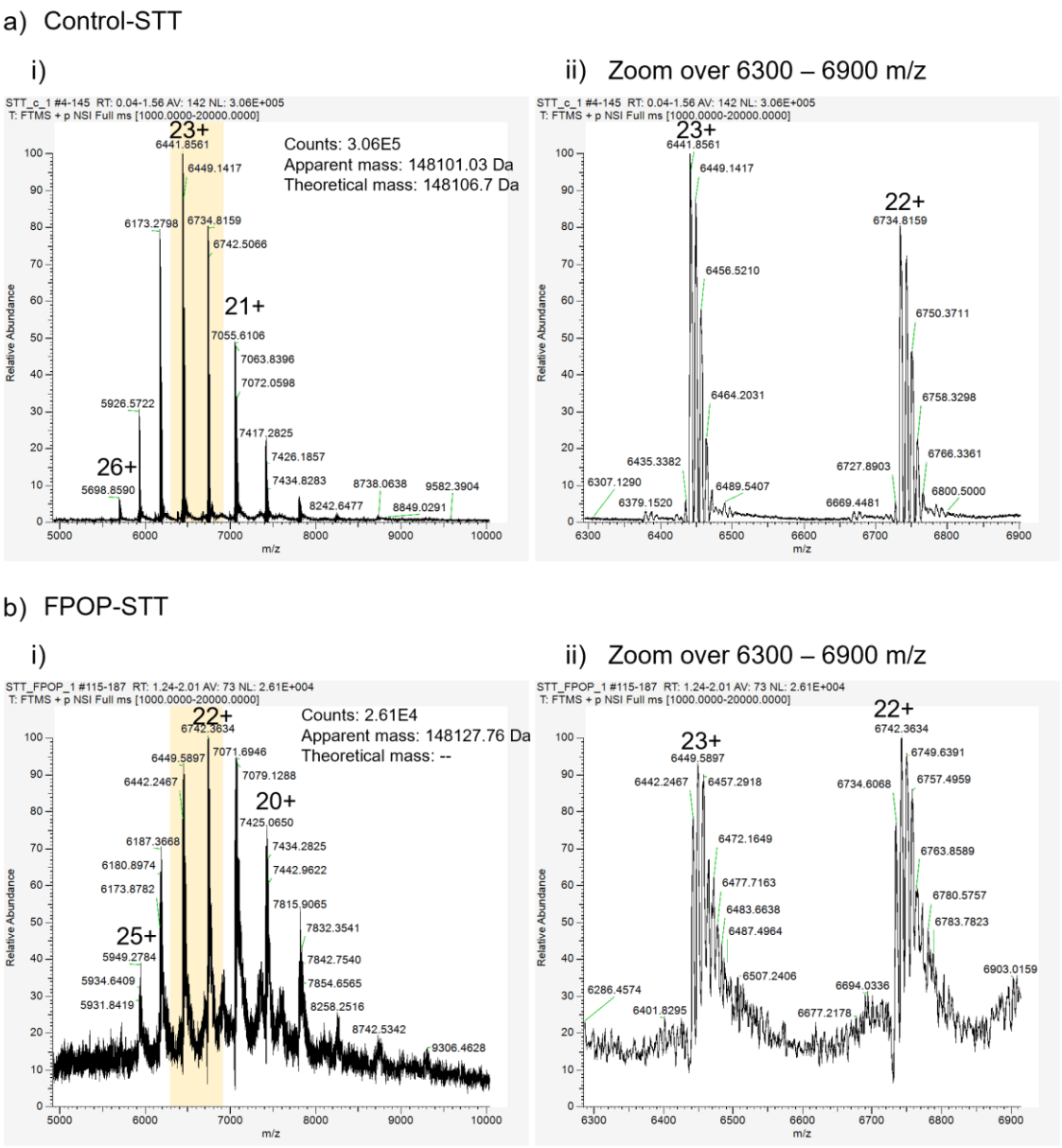


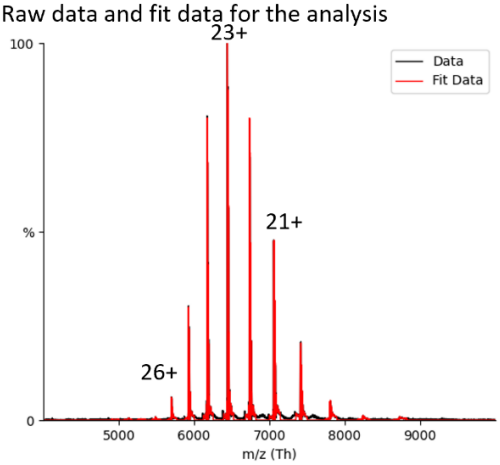
Figure 4.10 Native raw spectra for STT, control and FPOP modified.

a) Control-STT (without laser exposure) raw data, shown in i) m/z range of 5000 – 10000 (yellow block indicates the window expanded in ii), and ii) m/z range of 6300 – 6900. b) FPOP-STT raw data, shown in i) m/z range of 5000 – 10000 (yellow block indicates the window expanded in ii), and ii) m/z range of 6600 – 7200. Data generated on a Q-Exactive UHMR mass spectrometer. Tune file and instrument parameters for a) listed in Appendices 7.2.

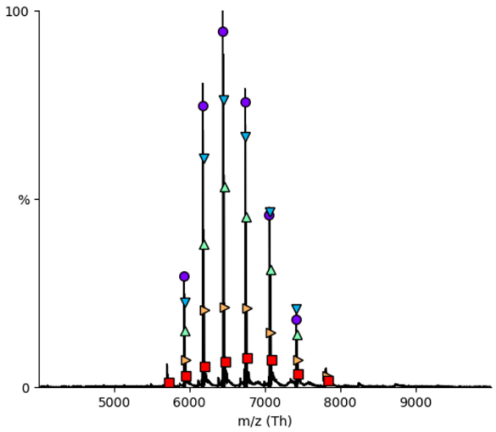
The raw data in Figure 4.10 was processed using UniDec v5.1.1 (Marty *et al.*, 2015) to estimate apparent masses for the FPOPed sample compared to the control. The output for the control is in Figure 4.11 and FPOPed STT is shown in Figure 4.12.

Control-STT

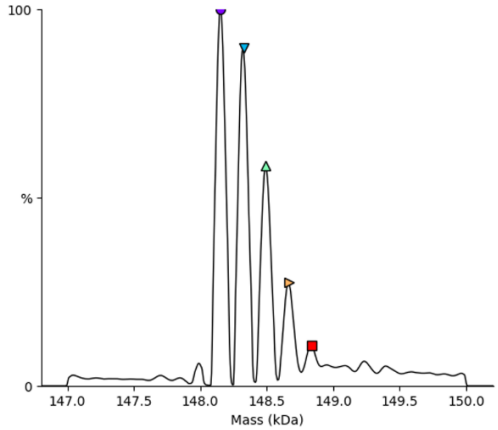
a) Raw data and fit data for the analysis



b) Peak selection



c) Mass distribution calculation

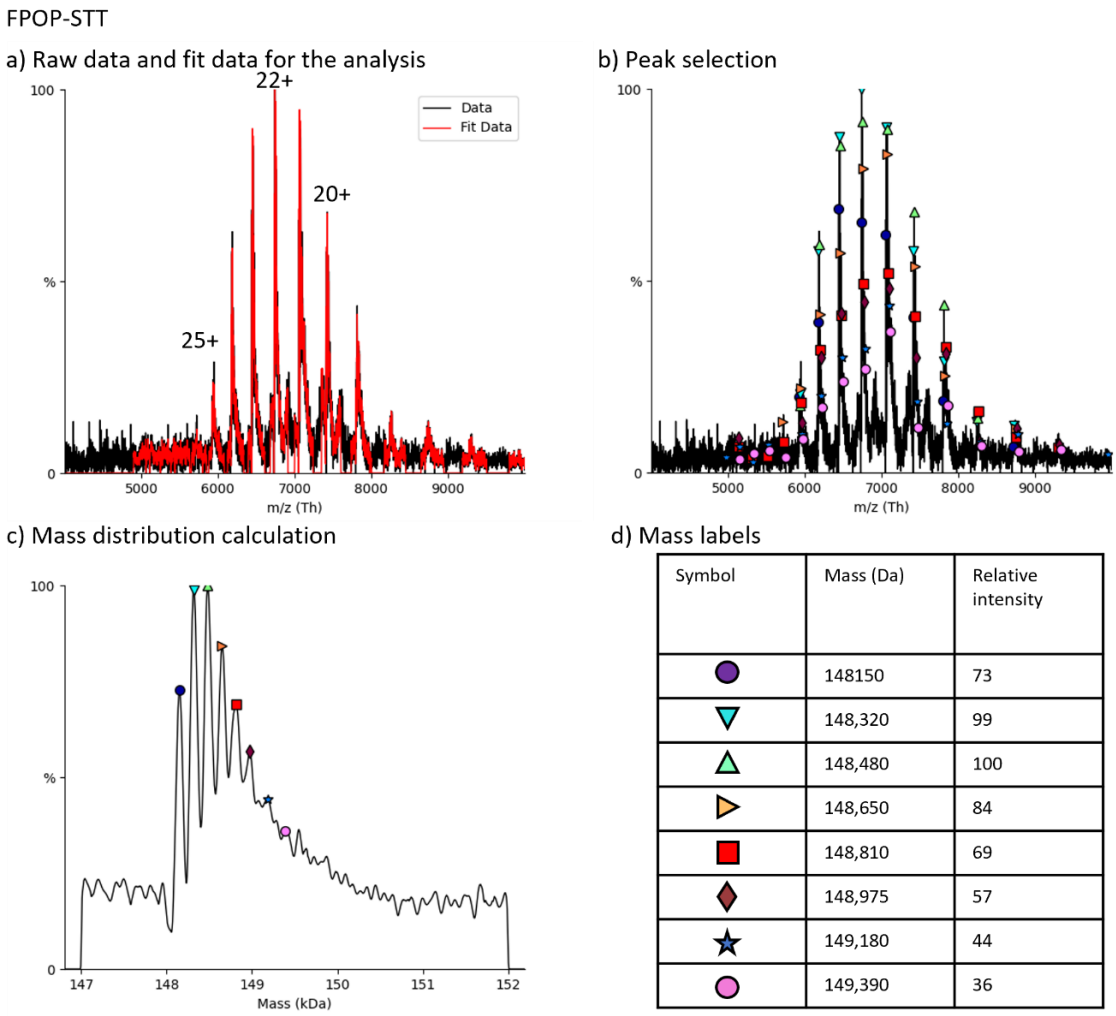


d) Mass labels

Symbol	Mass (Da)	Relative intensity
●	148,150	100
▼	148,325	90
▲	148,490	58
▶	148,665	28
■	148,835	11

Figure 4.11 Native MS data for control STT (non-irradiated control sample).

a) Raw native MS spectra annotated with charge states. The fit of the data = red, used for peak selection and mass distribution calculations. b) Selected peaks annotated with symbols. c) Mass distribution against intensity, labelled with the corresponding symbols from b). d) Mass label values in Da and their relative intensity. Data generated on a Q-Exactive UHMR mass spectrometer and analysed using UniDec software v5.1.1.



Symbol	Mass (Da)	Relative intensity
●	148150	73
▼	148,320	99
▲	148,480	100
▶	148,650	84
■	148,810	69
◆	148,975	57
★	149,180	44
●	149,390	36

Figure 4.12 Native MS data for irradiated STT (FPOP sample).

a) Raw native MS spectra annotated with charge states. The fit of the data = red, used for peak selection and mass distribution calculations. b) Selected peaks annotated with symbols. c) Mass distribution against intensity, labelled with the corresponding symbols from b). d) Mass label values in Da and their relative intensity. Data generated on a Q-Exactive UHMR mass spectrometer and analysed using UniDec software v5.1.1.

The extracted masses are approximated to the nearest 50 Da, which is apparent by the differences between the expected masses (control STT mass = 148,106.7 Da, apparent mass = 148,150 Da). The extraction of the apparent FPOP masses demonstrate that there is a high heterogeneity to the FPOPed species, along with the high baseline and the increased difficulty for the software to resolve peaks at higher mass values. However, it is important to note that the FPOP modifications do not cause the apparent global molecular weight of STT to become over-heterogeneous (which would indicate over-modification) and the base mass is still within 50 Da of the control. From the shift in the most abundant mass in the control being 148,150

(purple circle in Figure 4.11) to 148,480 in the FPOPed sample (green triangle in Figure 4.12), the average number of oxidations per mAb can be approximated to 21 modifications of +16.

These results clearly demonstrate that the FPOP procedure works to modify sample to a moderate amount. Next, control STT and FPOPed STT were subjected to the optimised LC-MS/MS procedure outlined in Figure 4.9.

4.3.2 Peptide coverage maps are comparable regardless of oxidation percentages

Non-irradiated control STT and irradiated FPOP STT at approximately 0.1 mg mL^{-1} were digested by the procedure in Figure 4.9. A coverage map was then generated to compare the ability of sample to be digested when oxidised.

The coverage map for control STT and FPOP STT are shown in Figure 4.13. It is key to note that the searches in PEAKS were made with trypsin specified as semi-specific, with a maximum of 3 miscleavages allowed. Additionally, the PEAKS program was used to search for FPOP modifications (+16 Da, +32 Da, etc.) but these were not shown on the coverage map for simplicity of comparing the resulting peptides.

Overall, the digestion patterns are similar (comparing Figure 4.13 a and b) with a very close overall percentage coverage, showing that the sample exposed to the FPOP laser is digested as efficiently as that without laser exposure.

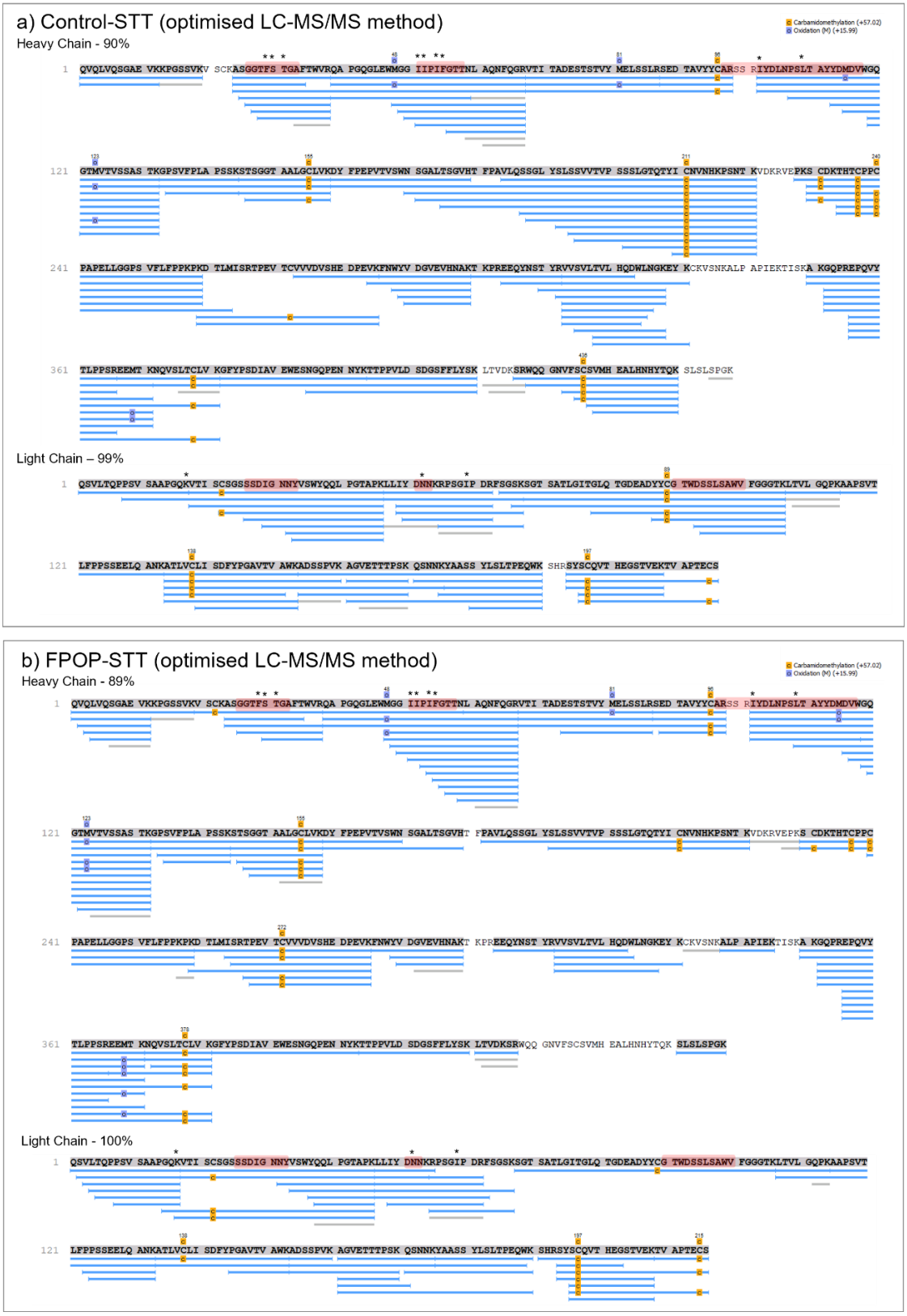


Figure 4.13 Example peptide coverage maps of control-STT and FPOP-STT, following the optimised LC-MS/MS procedure.

a) Control STT (0.1 mg mL^{-1}) prepared with hydrogen peroxide but not irradiated, b) FPOP-STT (0.1 mg mL^{-1}) prepared with hydrogen peroxide and exposed to the KrF laser. Identified peptides for the full Heavy Chain (upper) and the full

*Light Chain (lower) are shown as blue lines below the sequence. CDR regions 1, 2 and 3 for each chain are highlighted in red. Hotspot residues are marked with an *. Data were obtained on the Ultimate 3000 RSLCnano LC system coupled to the Orbitrap Exploris 240.*

4.3.3 Observing peptide-level oxidation modification differences between non-irradiated control and irradiated FPOP sample

Before commencing with the analysis of all the mAbs explored in this thesis in the final results chapter, it was important to use PEAKS as a modification searching tool, to give a map of possible oxidation sites over the sequence. This visual map will be used to make a list of searchable peptides for peptide-level oxidation analysis in the final chapter of this thesis.

The FPOP STT data in Figure 4.13 b was analysed using PEAKS to detect hydroxylations of the peptides. This proof of principle experiment demonstrates the PEAKS software assigns hydroxylations to amino acids through MS/MS sequencing. This output map is displayed in Figure 4.14. From this map, regions of STT that are most prone to oxidation natively in formulation buffer at 0.1 mg mL^{-1} are clearly seen. For simplicity, the modifications shown in Figure 4.14 refer to hydroxylations of +16 Da and +32 Da, the most common modifications typically found in an FPOP dataset (see Table 1.2 A list of common FPOP modifications.). However during thorough analysis, additional modifications are searched. These other modifications include +48 Da (three oxidations per peptide); the formation of carbonyl groups +14 Da; and +73 Da or +89 Da, singly and doubly oxidised cysteine side chains following carbamidomethylation). Additionally, example MS/MS spectra collected for the peptide 24-ASGGTFSTGAFTWVR-38 is shown in Figure 4.15, illustrating one variation of the position of labelling of the peptide by +16 and +32 (Figure 4.15).

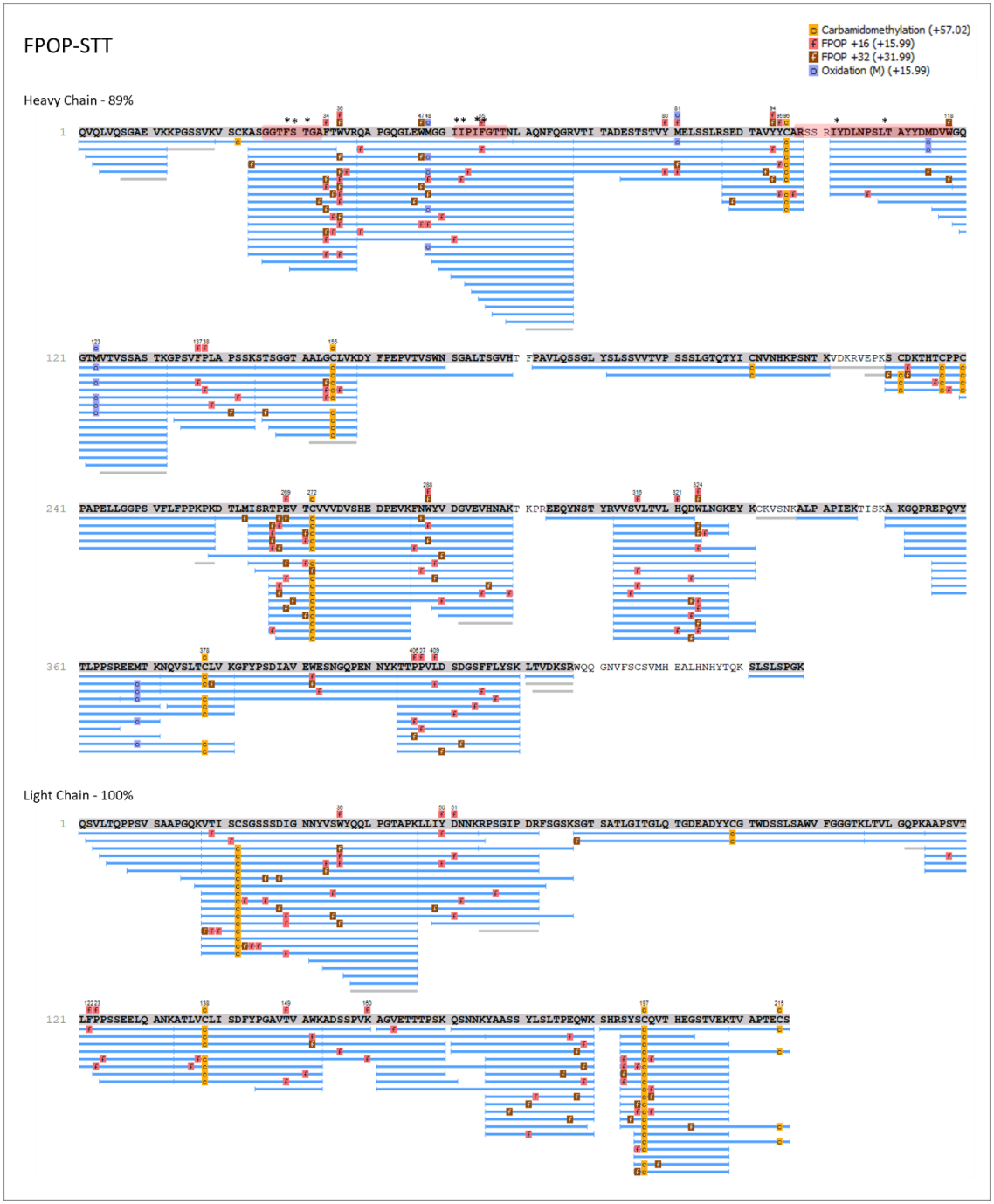
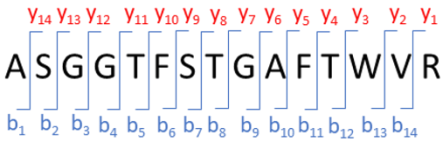


Figure 4.14 Peptide map for an example FPOP-STT sample, showing several modification types.

*Identified peptides for the full Heavy Chain (upper) and the full Light Chain (lower) are shown as blue lines below the sequence. CDR regions 1, 2 and 3 for each chain are highlighted in red. Hotspot residues are marked with an *. Data were obtained on the Orbitrap Exploris 240.*



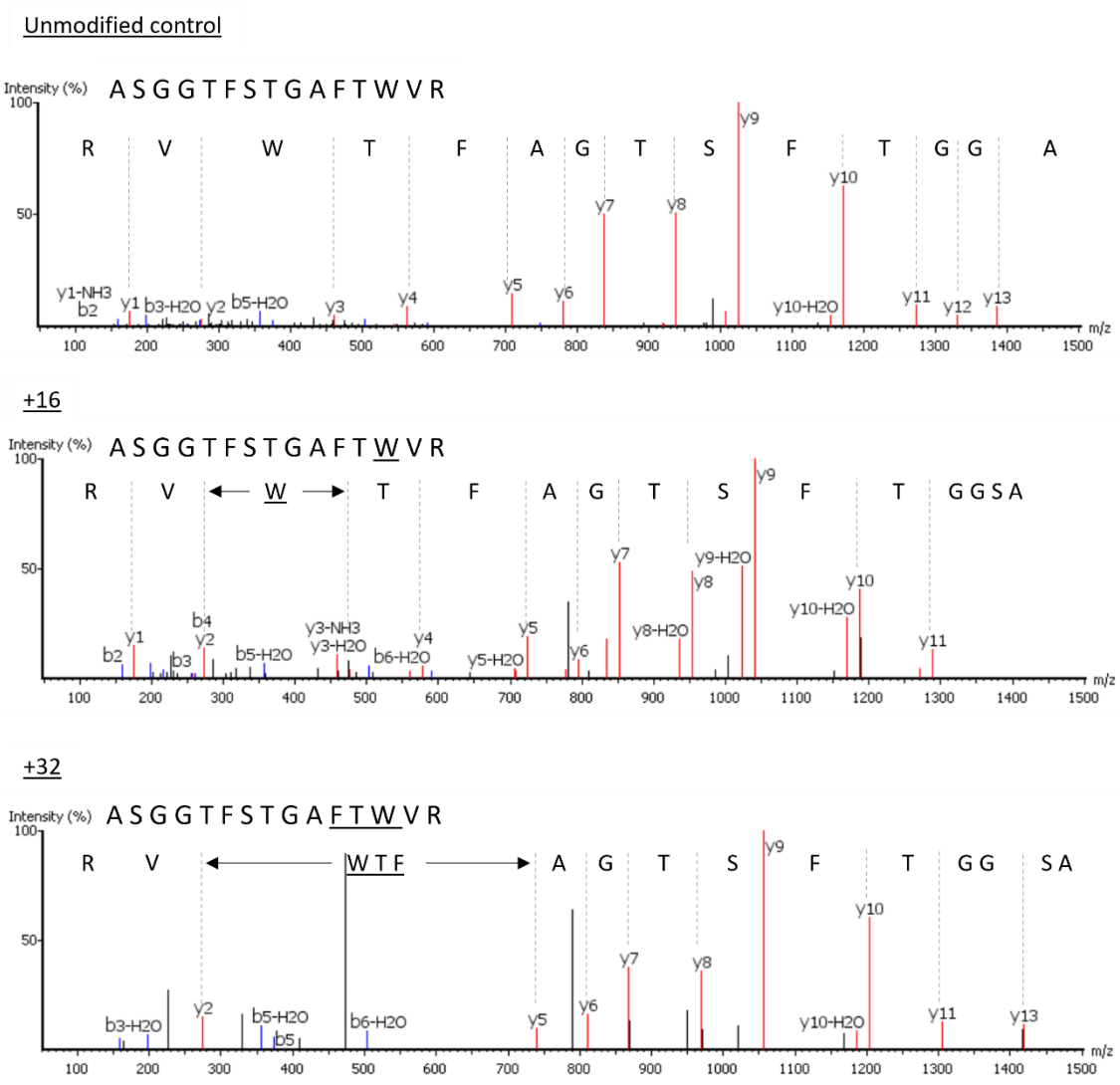


Figure 4.15 Example MS/MS spectra illustrating the +16 and +32 modifications assigned.

MS/MS sequencing data for peptide 24-ASGGTFSTGAFTWVR-38 from the dataset in Figure 4.14 when unmodified (top panel, m/z = 772.8787), with an example of spectra for the assignment of one oxidation +16 (middle panel, m/z = 780.8757) and an example of spectra for the assignment of two oxidations +32 (bottom panel, m/z = 788.8737). The amino acid sequence is annotated in above the spectra, where underlined letters correspond to where the modification site assignment was made.

In Figure 4.15, the unmodified spectra shows successful sequencing throughout the peptide, with signal above the noise for each fragmented peptide in the y-ion series (top panel). The +16 modified example (middle) shows the modification localised to the tryptophan W36; this is the

most abundant position of modification, but there is also observable MS/MS signal for the modification at phenylalanine F34 (not shown here). Additionally, in the +32 example (bottom panel), the sequencing is not good enough quality to localise the two modifications to one/two amino acids; there is likely to be a combination of MS/MS traces where there is; +32 on W36, +16 on each of W36 and F34; and even +32 on F34 (however the inherent reactivity rate of W is 1.88 times higher than F (Introduction Part 2, Table 1.3; W = 1.3×10^{10} M^{-1} s^{-1}; F = 6.9×10^{9} M^{-1} s^{-1}).

As previously mentioned, the format in Figure 4.14, which displays all of the PEAKS-identified oxidations on a peptide map is useful for identifying the regions of STT that are most prone to oxidation. Of course, this is a top-level display where it is necessary to visit the MS/MS data corresponding to each peptide in order to confirm the location of oxidation and even discover lowly-abundant isotope distributions that PEAKS may not identify which show further oxidation. The results here in Figure 4.14 have been pre-filtered in PEAKS analysis software, to remove any false assignments of MS/MS which could have falsely identified a peptide. This method of pre-filtering the MS/MS assignments allows the user to specify a false discovery rate (FDR), where the software can remove identification based on a calculation of its certainty of being a true assignment. Typically, a threshold of 0.1% is recommended. The output of the pre-filtering performed for each protein sequence input in PEAKS is a significance score (a P score in the form of $-10_{log}P$) which can be used for comparison between proteins of different sequence. The $-10_{log}P$ score is the threshold score to indicate the likelihood of a true identification of the modification position on any peptide. The higher the score is, the more confident the software is at its MS/MS assignment.

In Figure 4.14, the FDR has been manually set to 0.1% threshold, which for this data gives a $-10_{log}P$ score of 45.5. One may wish to make this threshold higher (a lower % threshold), which triggers PEAKS to re-calculate the assignments, and this aids to exclude any promiscuous assignments on the edge of the threshold value. For analysis of FPOP data to amino-acid level resolution, the quality of the MS/MS sequencing needs to be highly reliable for confidently assigning oxidation to each residue, and therefore it would be important to manually check all of the suggested sequencing data per peptide, to decide whether to categorise the assignment as credible or not; this could be aided by creating a higher % threshold of the assigned data in PEAKS. However, for the majority of the FPOP data analysis in the final chapter of this thesis, all peptides were investigated for total percentage oxidation at peptide-level resolution, which meant that the PEAKS' assignment of the FPOP position was just used as a possible site of

oxidation to explore, and the oxidation modification percentages were calculated from the extracted ion chromatogram (XIC) data without the need for MS/MS sequencing data.

The experiments in this thesis run on the assumption that background oxidation is comparable between every sample analysed. Therefore, the background oxidation has not been calculated, considering the total level for this will be the same between every sample, and the 'control' refers to the unmodified form of each peptide contained within the same sample dataset as the modified material (the same datafile). In this sense, the oxidised control is an internal reporter of background oxidation. Quantifying oxidation in this way will mean that, a) background oxidation 'cancels out' when considering the calculation of % modification (Equation 2.7 Quantifying FPOP oxidations.) – in this equation, the % modified will be based on the sum of the modified peak area(s), divided by the total oxidation in the unmodified control and modified peptides; and b) there is no introduction of error which could be the case when comparing oxidation between different samples due to errors in volumes of peroxide/scavenger in the sample preparation process. Therefore, the level of background oxidation in all experiments is assumed to be low and comparable due to the same laser power settings and the samples being prepared in the same way with the same exposure times to peroxide. The level of in-solution oxidation from peroxide exposure is assumed to be vastly lower than when the peroxide is induced to form radicals by the FPOP laser.

An exception to the low level of background oxidation is that of the amino acid methionine, which is assumed to be oxidised to a high level in any MS experiment (Cornwell *et al.*, 2019; Johnson *et al.*, 2019; H. Zhang *et al.*, 2014). Additionally, all samples are assumed to have the same low level of background oxidation in the ESI source due to spontaneous radical formation in the ESI plume. These oxidations can be identified through having the exact same retention time as the unmodified peptide in the extracted ion chromatograms (XICs) for the peptide. Following this, I will assume that the background oxidation in all of my subsequent data will be comparable and therefore I will not individually correct data for background oxidation to simplify data analysis. Future experiments to quantify background oxidation would be to mix the sample with hydrogen peroxide in the same ratios as experimented on here, but do not expose the sample to the KrF laser; then quench the reaction after 5 minutes just as designed in the Methods. This would indicate how much labelling may occur spontaneously in solution.

Overall, this chapter has built a narrative of the optimisation of an FPOP-LC-MS/MS protocol to study mAbs. The technique will be applied to WFL, 114 and STT in the next chapter of this thesis, where samples native in solution will be compared to each other. Samples will also be stressed

using hydrodynamic force, and compared with the native state, to begin to unpick the beginnings of mAb unfolding and aggregation in response to flow force.

4.4 Discussion

This work demonstrated an improvement of sequence coverage from HC = 57%, LC = 74%, to HC = 81% and LC = 99%. During the method development, the volume of sample loaded onto the LC column was increased from 1 µL to 3 µL. It has been demonstrated that the amount of FPOP sample loaded onto the LC column influences the variability in signal intensity of peptides, such that low sample loading (for example 1 µL of 0.5 µM, as was used in all of the Sections in this chapter) can cause significant variation in results between separate repeats (Abolhasani Khaje *et al.*, 2018). In the context of FPOP modification, the abundance of unmodified XIC peak area compared to modified for the same peptide is used for quantifying the percentage modification, and so it is important to load a larger amount of peptides in order to attempt to keep variability to a minimum. This is because the difference of abundance of FPOP modification observed between 'biological' repeats (in this case this refers to separate instances of preparing protein from stock aliquots of the same batch) needs to be kept to a minimum, for confident assignment of oxidation at a particular site, if it occurs in each repeat.

During the PEAKS searches, the false discovery rate is manually set by the user (in this chapter, I chose to set this to 0.1%) to filter out any low scoring peptides before analysing the data in too much detail, simplifying the analysis somewhat. The lower the false discovery rate, the more stringently peptides are discarded, if the MS/MS sequencing data for that instance the software matched the data to a peptide is poor enough to fall below the threshold. This manually-set threshold limits the amount of peptides identified by the software to form part of the coverage maps (blue lines), reducing the apparent coverage. However, during manual quality control of the sequencing process, some peptides that might satisfy the 0.1% false discovery rate could be found to be too insufficiently sequenced for the user to confirm the data corresponds to the correct peptide. Therefore, manually observing the MS/MS spectra is a highly important step to further reject peptides which the software has assigned but do not satisfy the user's requirements, such as the b and y ion intensities being 3 times higher than the noise. This added complexity to the analysis further confirms the sensitive nature of FPOP data analysis and highlights further the need for a robust and reproducible methodology (e.g. a bespoke software to automate this process).

Obtaining high sequence coverage is important in the context of investigating how different amino acids on a protein surface may influence a distal structural change in that protein. The observation of how the triple mutation STT conferred drastic changes to the reversible self-association properties of WFL is still not well understood. Previously, FPOP-MS and HDX-MS experiments elucidated differences in the conformation and structure of WFL and STT (Cornwell *et al.*, 2019). The two parallel techniques were in agreement at many solvent accessible regions, but a key finding was that using FPOP to probe solvent accessible surface area specifically, the mutations WFL to STT in the CDR regions resulted in significantly different labelling patterns at the interfaces of the C_L−C_H1 and C_H1−C_H2 domains. This long-range effect of the substitutions in the V_H could be attributed to a change in structure and orientation of the C_L−C_H1 interface, with significant ($p < 0.01$) protection from modification and exchange in WFL compared to STT. The authors postulate that, although the significance of these conformational differences with regards to the poor pharmacological properties of WFL cannot be established from these data alone, the differences observed between WFL and STT may contribute to the propensity of WFL to form aggregates and undergo self-association. Additionally, changes in modification in the C_L−C_H1 interface are thought to be related to antigen binding at the CDR regions (Sela-Culang *et al.*, 2012). Therefore a conformational link between the two regions was suggested, and obtaining good coverage over this region with the improved in-house methodology developed in this chapter will allow further investigation into the long range effects using FPOP-LC-MS/MS. This work will also be important in further elucidating any influence the WFL mutations may have on the unfolding and aggregation of the mAbs, providing insight into sequence effects on the propensity to unfold due to hydrodynamic stress. Furthermore, the addition of 114 to the WFL/STT story will add to the understanding of how mutations in the V_H and V_L might contribute to changes in conformation of the mAb in other parts of the structure.

The potential of using FPOP-MS to detect subtle structural differences between proteins is not well understood. Applying the optimised method in this chapter for screening of novel biotherapeutic candidates for their surface-exposed residues in quiescent conditions, compared to how their structures change after being stressed (such as during extensional flow), is an important application of the time-sensitive labelling technique. FPOP is tuned to limit the exposure of proteins to reactive oxygen species to the sub-millisecond time scale (B. Zhang *et al.*, 2018). Using FPOP to report on transient unfolding has not been explored to date, so will be a powerful application of the protocol in the next chapter of this thesis. The overall methodology shows high applicability for aiding the selection of promising robust candidates. Investigating ways of mitigating protein aggregation in biopharmaceutical development is of critical

importance for patient safety, product efficacy, and the future development of biopharmaceutical products.

In summary, the data presented in this chapter has demonstrated the rational improvements made to an LC-MS/MS protocol for antibody peptide mapping. I can achieve high coverage on sample at 50 times lower concentration than used in industry. The ability of the protein to be modified by FPOP has been validated and the next stages for applying the technique to observe and characterise the position of lowly abundant modifications on a structure have been outlined.

Chapter 5

FPOP-LC-MS/MS comparing three mAb variants

5 FPOP-LC-MS/MS comparing three mAb variants

5.1 Objectives

Identifying and mitigating aberrant protein aggregation in biopharmaceuticals is of critical importance for both their manufacture and for patient safety (see Section 1.1.2.4: Monoclonal antibody aggregation). The development of biopharmaceuticals would greatly benefit from ways to identify aggregation-prone sequences in the drug candidates early in the development pipeline, before time- and labour-expensive experiments. The pharmaceutically-relevant family of model mAbs WFL, STT and 114 used in this thesis have identical sequences apart from three or four amino acid substitutions in the V_H or/and V_L regions; these confer drastic changes to their biophysical properties, making them ideal candidates to study aggregation in mAbs. As seen in previous chapters, STT differs from WFL at three positions in the CDR region (W30S, F31T, L57T) but has highly improved physiochemical properties (Figure 1.17) and improved clearance *in vivo* (C. L. Dobson *et al.*, 2016). The third mAb studied, 114, retains WFL residues but has two substitutions located in the V_H mutational hotspot region (F29S and I52N) and two other substitutions in the V_L domain (S2G and T77A). These mutations, identified using directed evolution, confer improved biophysical properties over WFL, including a reduced hydrodynamic flow-induced aggregation propensity. 114 displays similar biophysical properties, and tendency to aggregate in response to flow, to STT (Section 3.4.2: Assessing mechanical stability under hydrodynamic force).

This triad of proteins with contrasting properties resulting from such minor sequence changes provides a unique opportunity for further understanding what the key factors of flow-induced aggregation might be, using the in-house FPOP protocol optimised in Chapter 4. A similar methodology has previously been applied to the study of these large, complex proteins at residue-level resolution (Cornwell *et al.*, 2019), and the previous chapter demonstrated the successful adaptation of a protocol for use in this thesis. This final chapter will therefore probe the limits of the optimised FPOP-LC-MS/MS protocol for assessing conformational changes associated with aggregation propensity of the three mAbs at the peptide-level resolution. Of particular interest is to further understand the phenomenon of flow-induced unfolding and aggregation, and the initial stages of protein manipulation and deformation due to hydrodynamic flow. With this in mind, WFL, STT and 114 were subjected to FPOP-LC-MS/MS with the following aims:

1. To identify any differences in structure between the three closely-related mAbs prior to force exposure, which may add to the understanding of the mAbs's propensity to aggregate;
2. To identify any flow-induced changes to the structure of the mAbs unstressed and stressed, relative to each variant, highlighting regions proximal or distal to the substitution sites which have more of a propensity to unfold under flow.

5.2 Initial characterisation and overview

WFL, STT and 114 were labelled by the FPOP-LC-MS/MS procedure optimised in Chapter 4 (the methodology is listed in Section 2.2.8.3: Fast Photochemical Oxidation of Proteins – LC-MS/MS. The mAbs were analysed in their formulation buffer (20 mM sodium succinate, 125 mM L-arginine, pH 6.0) at 0.1 mg mL^{-1}, conditions at which both WFL and STT are known to be predominantly monomeric (C. L. Dobson *et al.*, 2016) and 114 was shown to be monomeric (at 1 mg mL^{-1}) through ultra-low-volume DLS (Section 3.3.2: Ultra-low-volume dynamic light scattering (DLS)). Briefly, 10 mM L-histidine was used in the final solution as a scavenger and 3 µL of 5% *v/v* H_2O_2 was added as a source of oxygen radicals immediately before flowing through the laser path at 20 µL min^{-1}. The solution was collected in an Eppendorf containing quench solution (20 µL of 100 mM L-methionine and 1 µM catalase in formulation buffer). The final mixture at approximately 0.08 mg mL^{-1} was subjected to the reduction, alkylation and digestion with trypsin outlined in Section 4.3: Oxidative labelling of STT coupled to the optimised LC-MS/MS method.

This procedure yielded good sequence coverage for the heavy and light chains of the three mAbs. An example peptide map is shown in Figure 5.1, displaying the peptides which were identified through MS/MS in blue, and the peptides common in all samples and therefore used for FPOP analysis highlighted in green. A total of 22 peptides were used, as these were found to be present in all replicates of the three mAbs (Figure 5.1, Table 5.1). These chosen peptides also had unmodified and modified forms present in the same data file, and were found reproducibly present between repeats (n=2 for all sample types). Of these 22 peptides, the MS/MS identified by PEAKS was consulted to provide confidence in the identification of the unmodified and modified peptides, using the minimal criteria that there were four or more amino acids in a series present with intensity of the b or y ion series three times the noise level. Peptide-level resolution of modification was mainly explored in this chapter (i.e. mass of unmodified peptide, mass +16 Da, and mass +32 Da; indicating no oxidation, +one oxidation and +two oxidations per

peptide respectively). This is a subsection of modifications that could have also been quantified but to explore all possible modifications is out of the scope of this thesis. Additionally, this chapter explores some examples of residue-level resolution and how this might add to the peptide-level analysis.

In the interpretation of all FPOP results in this thesis, it is important to consider both inherent reactivity of individual amino acids to ·OH radicals, and solvent-accessible surface area (SASA) effects, when interpreting the level of oxidation of peptides in a peptide-level analysis. Inherent reactivity is dependent on the chemistry of the amino acid, and so is independent of the surroundings the amino acids have as part of a sequence/fold (the local structural influences). The value of inherent reactivity is discreet for each amino acid (see Introduction Table 1.3). Therefore, changes to reactivity of a position in a sequence would be due to an amino acid substitution and therefore purely on its chemistry. The SASA of each amino acid refers to the influence of its environment on the accessibility that any ·OH radicals could have for a reaction with the amino acid side chain. This is influenced by protein fold – including other reactive amino acids brought close which may scavenge the reaction and the orientation of the side chain, any bound ligands or solution excipients that may act to shield the amino acid side chains. The interplay of SASA effects on amino acid- and peptide-level oxidation analysis still fairly unknown and can be aided through mapping results onto a structure.

Figure 5.1 Example peptide coverage map of STT following FPOP and tryptic digestion.

a) Heavy chain. b) Light chain. Black letters = protein sequence. Bold letters and grey background = the amino acids which make up the peptides identifiable by PEAKS analysis software (all peptides identified are the grey lines below the corresponding sequences). Amino acids highlighted in red = CDR regions. Asterisks = mutational hotspot residues as identified in (Ebo, Saunders, et al., 2020). Pink arrows = amino acid substitution positions. Grey lines = identified peptides not used for FPOP analysis; green lines = heavy chain peptides chosen for FPOP analysis; blue lines = light chain peptides chosen for FPOP analysis (all green and blue peptides were identified in all replicates, and had both modified and unmodified versions, see Table 5.1). Orange cysteine residues = sites of carbamidomethylation (carbamidomethylation (+57.02 Da) is presumed to be a fixed modification after reduction and alkylation of cysteine residues in the FPOP-LC-MS/MS procedure).

Table 5.1 Summary of modifications observed following FPOP of the mAbs WFL, STT and 114.

The peptide 'code' (shortenings of each peptide's sequence used to refer to each individual peptide) were constructed to help the reader refer to each peptide. 'Codes' take the first four amino acids (AA) of each peptide, and bracketing them with the position of the start amino acid of the full peptide, and the position of the last amino acid of the full peptide. In the peptide sequence, the CDRs are underlined (V_LCDR3 was not identified). Residues that differ between WFL and STT and WFL/STT and 114 are shown in red. The oxidations searched for in the data were +16 and +32 for every identifiable charge state. Further details including m/z values searched can be found in Appendices: Table 7.3.

Peptide 'code' (start-first four AAs-end)	Peptide sequence	Position	Oxidations identified
24-ASGG-38	ASGGTF**WF**GAFTWVR (WFL)	HC, V_H (CDR1)	+16, +32
	ASGGTF**ST**GAFTWVR (STT)	HC, V_H (CDR1)	+16, +32
	ASGGT**SWF**GAFTWVR (114)	HC, V_H (CDR1)	+16, +32
39-QAPG-67	QAPGQGLEWMGGI**I**PIFG**L**TNLAQNFQGR (WFL)	HC, V_H (CDR2)	+16, +32
	QAPGQGLEWMGGI**I**PIFG**T**TNLAQNFQGR (STT)	HC, V_H (CDR2)	+16, +32
	QAPGQGLEWMGGI**N**PIFG**L**TNLAQNFQGR (114)	HC, V_H (CDR2)	+16, +32
68-VTIT-87	VTITADESTSTVYMELSSLR	HC, V_H	+16, +32
88-SEDT-98	SEDTAVYYCAR	HC, V_H	+16, +32
102-IYDL-132	IYDLNPSLTAYYDMDVWGQGTMVTVSSASTK	HC, V_H (CDR3)	+16, +32
133-GPSV-144	GPSVFPLAPSSK	HC, C_H1	+16
145-STSG-158	STSGGTAALGCLVK	HC, C_H1	+16, +32
286-FNWY-299	FNWYVDGVEVHNAK	HC, C_H2	+16, +32
313-VVSV-328	VVSVLTVLHQDWLNGK	HC, C_H2	+16, +32
356-EPQV-371	EPQVYTLPPSREEMTK	HC, C_H3	+16
372-NQVS-381	NQVSLTCLVK	HC, C_H3	+16, +32
382-GFYP-403	GFYPSDIAVEWESNGQPENNYK	HC, C_H3	+16, +32
404-TTPP-420	TTPPVLDSDGSFFLYSK	HC, C_H3	+16, +32
1-Q(S/G)VL-17	Q**S**VLTQPPSVSAAPGQK (WFL & STT)	LC, V_L	+16, +32
	Q**G**VLTQPPSVSAAPGQK (114)	LC, V_L	+16
18-VTIS-46	VTISCSGSSSDIGNNYVSWYQQLPGTAPK	LC, V_L (CDR1)	+16, +32
47-LLIY-62	LLIYDNNKRPSGIPDR	LC, V_L (CDR2)	+16
107-LTVL-114	LTVLGQPK	LC, V_L/C_L	+16
115-AAPS-133	AAPSVTLFPPSSEELQANK	LC, C_L	+16
134-ATLV-153	ATLVCLISDFYPGAVTVAWK	LC, C_L	+16, +32
154-ADSS-170	ADSSPVKAGVETTTPSK	LC, C_L	+16
176-YAAS-190	YAASSYLSLTPEQWK	LC, C_L	+16, +32
194-SYSC-208	SYSCQVTHEGSTVEK	LC, C_L	+16, +32

The peptides used for quantification of FPOP modifications (Figure 5.1, Table 5.1) were mapped onto the energy-minimised structure prediction of STT generated using computational modelling platform Schrödinger (structure PDB file created by Romina Hofele, AstraZeneca), to visualise the overall coverage of the peptides used for FPOP quantification of the mAbs (Figure 5.2).

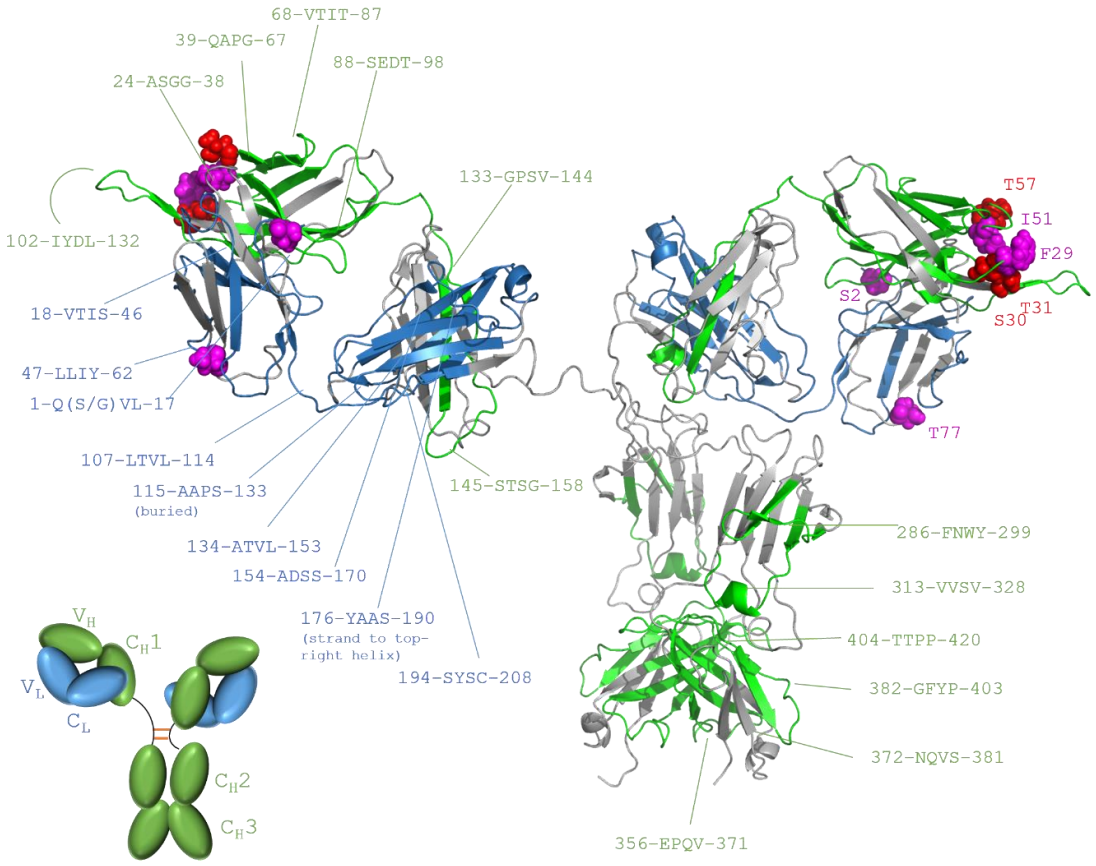


Figure 5.2 Modified peptides identified following FPOP-LC-MS/MS of WFL, STT and 114, mapped onto a model of STT.

Energy minimised structure prediction (created using Schrodinger) of STT in pH 6.0. Peptides used for FPOP quantification are shown in green (heavy chain) and blue (light chain) (all green and blue peptides were identified in all replicates, and had both modified and unmodified versions). Position of each peptide is labelled (one HC and one LC for simplicity) with their four letter code; line points to the midpoint of each peptide. Unidentified peptide sequence = grey. Sites of substitution = spheres (red spheres = positions of STT/WFL substitutions, magenta spheres = positions of 114's four substitutions. Inset (bottom-left) = simplified mAb structure. Structural prediction created using the primary sequence in Schrodinger simulation software, and visualised and annotated using PyMOL.

In the structure in Figure 5.2, the peptides were coloured green (heavy chain (HC)) and blue (light chain (LC)), contrasted with the rest of the sequence in grey. The mutation positions were depicted as spheres in red (WFL/STT) and pink (unique to 114) for ease of visualising their relative positions on the mAb. The positions of each peptide have been labelled with the four letter code, which gives the first four amino acids in the peptide, bracketed by the start and end positions of the full peptide (for example, 88-SEDT-98 is HC peptide SEDTAVYYCAR which correspond to residues 88 to 98 in the sequence. The lines on the diagram connect the four letter label to approximately the middle amino acid in the corresponding peptide.

FPOP modification was explored through the comparison of LC peak areas for particular peptides – using their extracted ion chromatograms (XICs). Difference in the peak area between unmodified and all of the peaks for the modified was added for all observable charge states (if present in every sample), giving a percentage modification value for every peptide. The results are the mean of n=2, where the error bars indicate the range. The heavy chain data are in Figure 5.3 and the light chain data are in Figure 5.4.

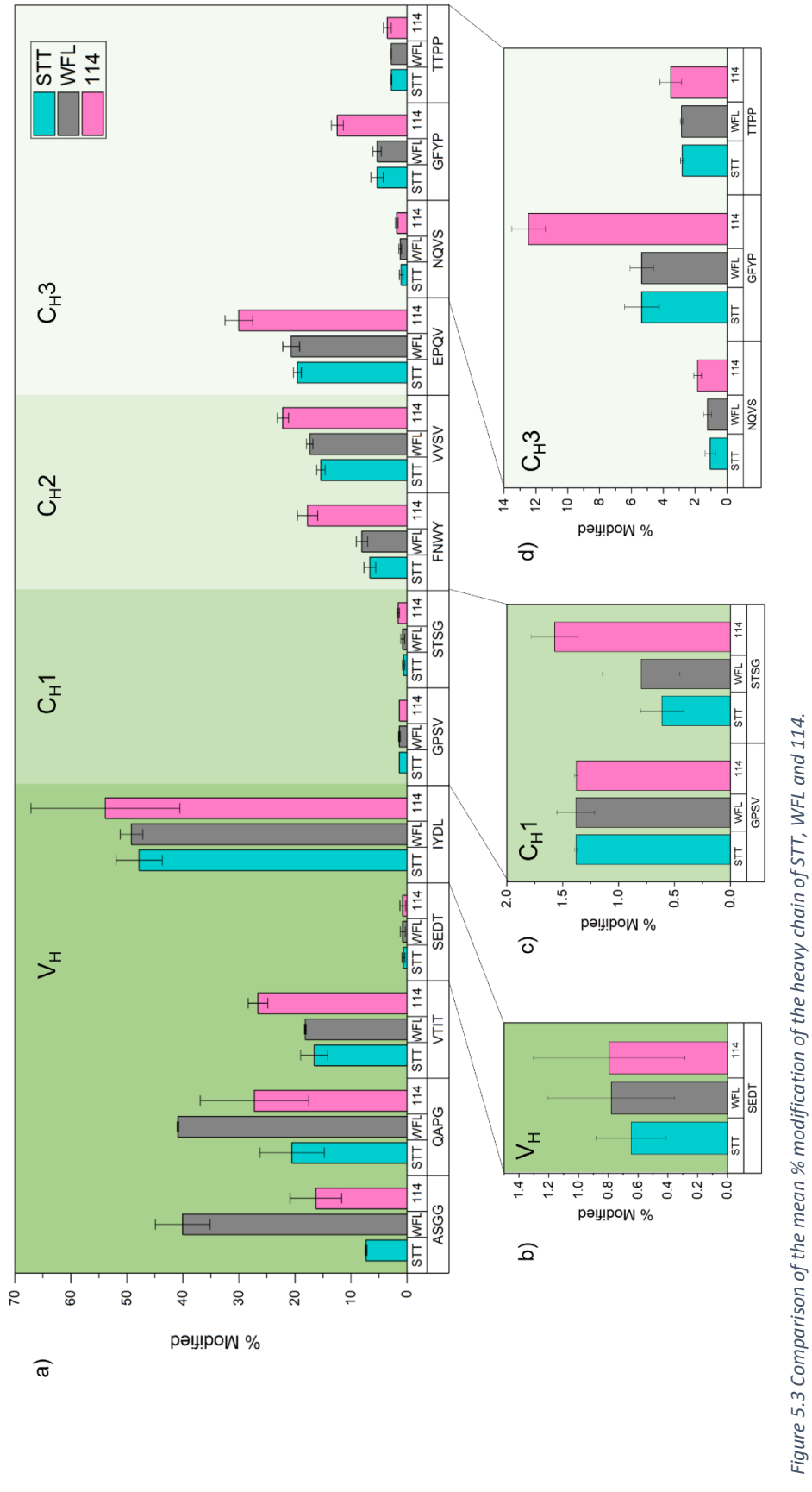


Figure 5.3 Comparison of the mean % modification of the heavy chain of STT, WFL and 114.

a) Summary of the mean % modification (and range error bars) for 13 peptides that were identified in the heavy chain of each mAb, where each label below the bar charts corresponds to the first 4 amino acids in the peptide sequence. Teal = STT; grey = WFL; pink = 114. Values shown are the mean (n=2) and the error bars indicate the range between duplicate datasets. Zoomed in quantification of low abundance modification on peptides are in b) SEDT, c) GPSV and STSG, d) NQVS, GFYP and TTPP. Values are listed in Appendices: Table 7.4

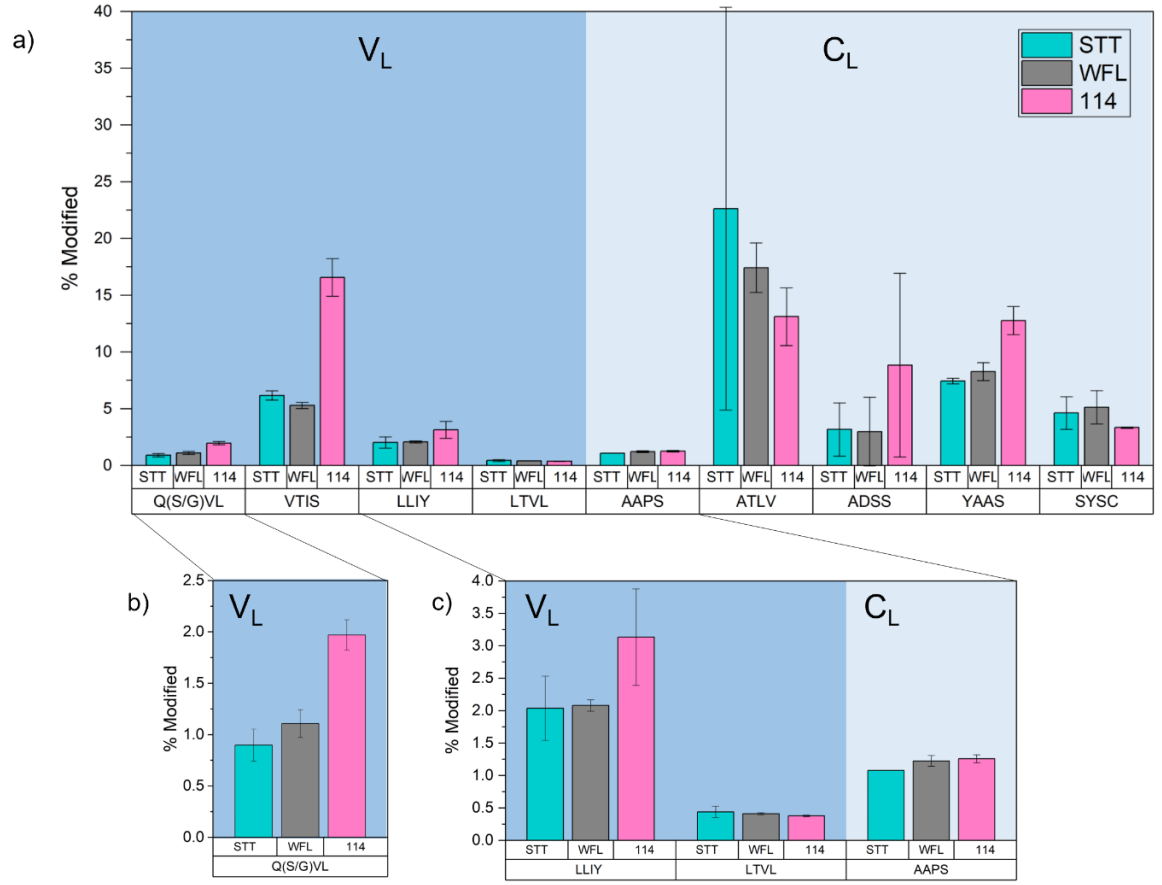


Figure 5.4 Comparison of the mean % modification of the light chain of STT, WFL and 114.

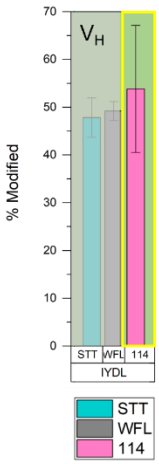
a) Summary of the mean % modification (and range error bars) for 9 peptides that were identified in the light chain of each mAb, where each label below the bar charts corresponds to the first 4 amino acids in the peptide sequence. Teal = STT; grey = WFL; pink = 114. Values shown are the mean (n=2) and the error bars indicate the range between duplicate datasets. Zoomed in quantification of low abundance modification on peptides are in b) Q(S/G)VL, c) LLIY, LTVL and AAPS. Values are listed in Appendices: Table 7.4

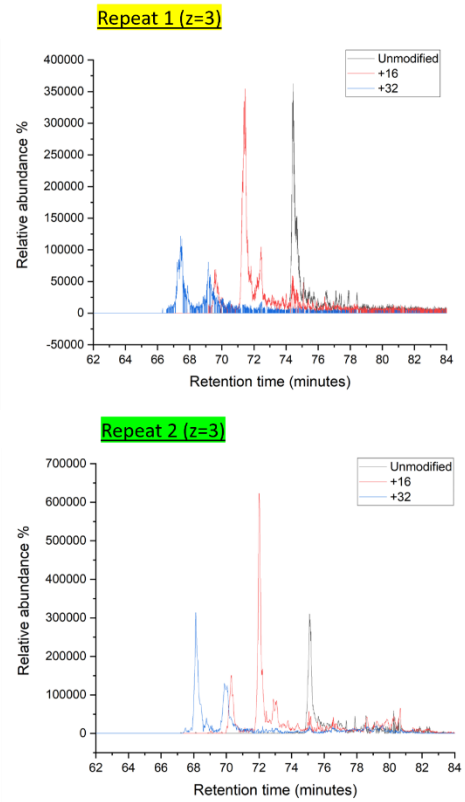
5.2.1 Replicate range bar differences are likely ion intensity issues rather than capturing conformational differences

In Figure 5.3 and Figure 5.4, the error bars indicate the range between two repeats. This average range for the control data is ± 1.69%, however there are a few examples where the range is high, and specifically in one variant compared to the other two. One example from the heavy chain (102-IYDL-132), where 114 showed a range of 13.31% between the two repeats, and one example from the light chain (134-ATLV-153), where STT showed a range of 17.75% between the two repeats, were investigated further. The XICs for all observable charge states were compared and the relative abundance – not normalised, in order to visualised absolute intensities – displayed, along with the calculation of percent labelling from the peak areas of the peaks corresponding to each modification (Figure 5.5 and Figure 5.6, respectively).

a) XIC traces demonstrating abundance variation between repeats

102-IYDLNPSLTAYYDMDVWGQGTMVTVSSASTK-132





b) Overall percentage modification calculation demonstrating the range variation between repeats

$$\% \, Modified = \frac{\sum Modified_{Peak}}{Unmodified + \sum Modified_{All\ peaks}}$$

Sample Repeat 1 of 2	Peak area (z=3)		
	Absolute	Percentage modified (of total modifications identified: z=3) (%)	
STT unmodified	8.71 e^6	-	Overall % modification
STT +16	1.09 e^7	33.04	47.18
STT +32	4.67 e^6	14.14	

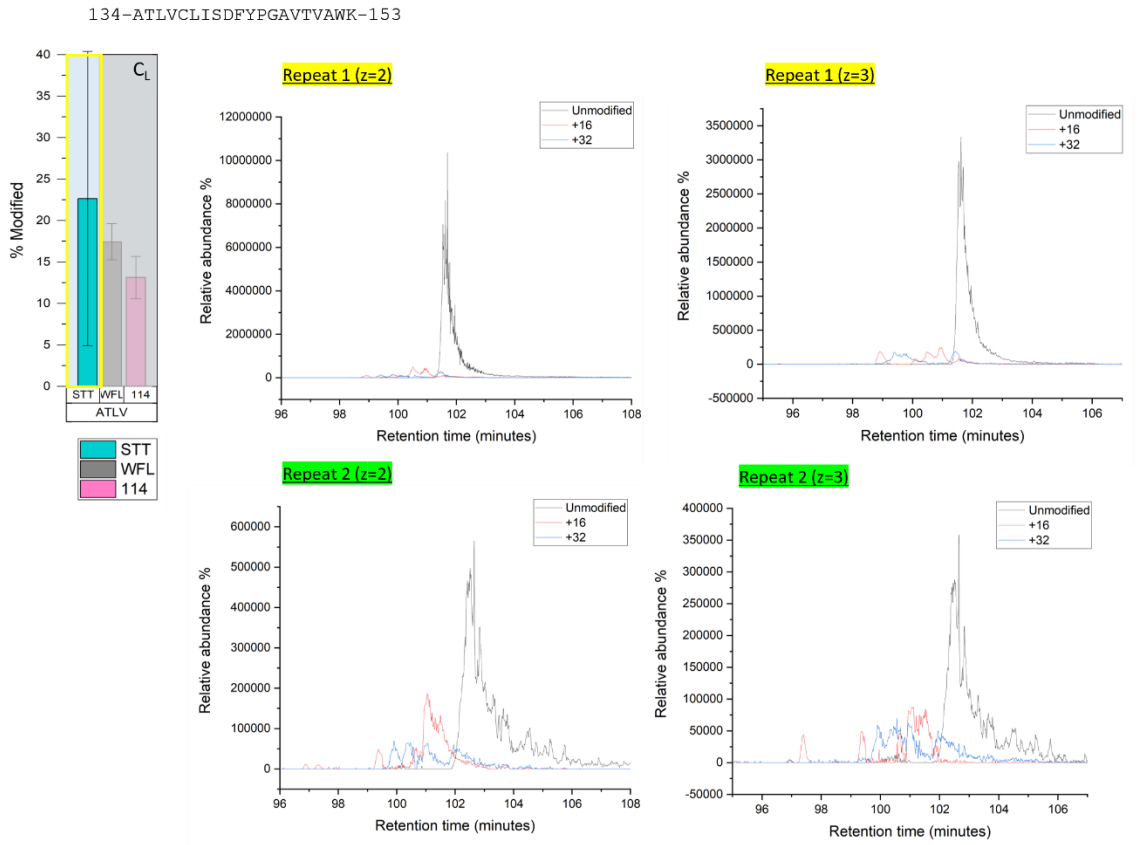
Sample Repeat 2 of 2	Peak area (z=3)		
	Absolute	Percentage modified (of total modifications identified: z=3) (%)	
STT unmodified	7.62 e^6	-	Overall % modification
STT +16	1.44 e^7	37.35	60.50
STT +32	1.65 e^7	23.15	

Range = 13.31 %

Figure 5.5 XICs and percentage modification calculations for peptide 102-IYDL-132 in 114.

a) XIC traces for the z=3 charge state (only observable charge state) for the unmodified peptide (black trace), the +16 (red trace) and the +32 (blue trace). b) The breakdown of the calculation of overall modified percentages of each repeat (n = 2). The absolute intensity value corresponds to the area under the peak(s) in the corresponding XICs that relate to the peptide, summed. This is then used in context with the total area of all peaks observed (every unmodified and modified peak) to calculate percentage modified (see equation in the graph and Methods: Equation 2.7). The difference in the overall % modification of both repeats is stated as the range.

a) XIC traces demonstrating abundance variation between repeats



b) Overall percentage modification calculation demonstrating the range variation between repeats

$$\% \, Modified = \frac{\sum Modified_{Peak}}{Unmodified + \sum Modified_{All\ peaks}}$$

Sample Repeat 1 of 2	Peak area (z=2)		Peak area (z=3)			
	Absolute	Percentage modified (of total modifications identified: z=2 & 3) (%)	Absolute	Percentage modified (of total modifications identified: z=2 & 3) (%)		
STT unmodified	1.76 e^8	-	8.25 e^7	-	Total (%)	Overall % modification
STT +16	1.68 e^7	5.62	1.08 e^7	3.60	9.22	13.75
STT +32	6.28 e^6	2.09	7.30 e^6	2.43	4.52	

Sample Repeat 2 of 2	Peak area (z=2)		Peak area (z=3)			
	Absolute	Percentage modified (of total modifications identified: z=2 & 3) (%)	Absolute	Percentage modified (of total modifications identified: z=2 & 3) (%)		
STT unmodified	2.8 e^7	-	1.65 e^7	-	Total (%)	Overall % modification
STT +16	8.09 e^6	12.44	4.27 e^7	6.56	19.00	31.50
STT +32	4.01 e^6	6.16	4.12 e^6	6.34	12.50	

Range = 17.75 %

Figure 5.6 XICs and percentage modification calculations for peptide 134-ATLV-153 in STT.

a) XIC traces for the z=2 and z=3 charge states for the unmodified peptide (black trace), the +16 (red trace) and the +32 (blue trace). b) The breakdown of the calculation of overall modified percentages of each repeat (n = 2). The absolute intensity value corresponds to the area under the peak(s) in the corresponding XICs that relate to the peptide,

summed. This is then used in context with the total area of all peaks observed (every unmodified and modified peak) to calculate percentage modified (see equation in the graph and Methods: Equation 2.7). The difference in the overall % modification of both repeats is stated as the range.

The XICs for all observable charge states in the two peptides show highly similar oxidation patterns, meaning the traces appear to line up well at the observed retention times between repeats. In Figure 5.5, the unmodified V_H 102-IYDL-132 peptide intensities were similar between the repeats, whereas +16 peak area was a larger percentage of the total area observed in Repeat 2 (a), red trace in bottom graph, and b) bottom table) than it was in Repeat 1. Additionally, the peak area for +32 was also high in Repeat 2, and both of these have contributed to the high overall % modification compared to Repeat 1. Here, these differences might be suggesting that an alternative conformation of 114 might be present, due to the unmodified intensities being similar between the repeats but the modified intensities differ. However, looking at the V_L 134-ATLV-153 peptide in Figure 5.6, where both z=2 and z=3 charge states were observable for each repeat, the abundances are clearly halved in the second repeat, and the peak shape for the unmodified is poorer, indicating a signal intensity difference is the main cause of the overall difference in observed % modified.

Therefore, overall, the large differences seen between replicates of all three mAbs in this dataset in this chapter are likely to be due to fluctuations in signal intensity, and are less likely to be capturing any evidence for differences in conformation of the protein where the peptides have been derived from, and consequently, it would be highly valuable to generate more repeats of the datasets to tease this difference apart.

5.2.2 Overlapping peptides would add confidence to modification assignments

Investigating whether peptide-level FPOP analysis has captured different dominant states of the mAb structures in solution is part of the aims for this chapter, and a method to improve the confidence on the assigned percentage modification on each region of a peptide would be to use overlapping peptide information over a region of interest. This method of assigning confidence through overlapping peptides is used in HDX-MS experiments (see Section 3.5 Probing differences in intrinsic dynamics over time with HDX-MS), to build a confidence map of percentage exchange of hydrogen with deuterium (labelling). HDX experimentation uses pepsin, a broad specificity enzyme, which is used to reliably generate overlapping fragments of the same sample in the same conditions. Here in the FPOP experiments, trypsin is used to digest

specifically after lysines and arginines (if not before a proline), and this reliability is used to ensure like-for-like peptides are examined in the procedure; however, miscleavages can occur, resulting in fragments which do not end or/and start where the predicted digestion pattern would be. In theory, this natural abundance of miscleavages could be utilised for assigning oxidation more confidently to shorter stretches (or indeed individual amino acids) of the full length peptides. In practice, however, from the results displayed in Figure 5.3 and Figure 5.4, the intensity of the unmodified miscleaved peptides were generally 2-3 orders of magnitude lower in intensity than the full-length peptides, meaning their XICs were close to the noise level of the data acquired and accurate peak areas could not be attained. Furthermore, rarely were the modified forms of these shorter peptides observable. Unfortunately, no patterns could be drawn from the data from miscleaved peptides in this dataset.

Therefore, an improvement for future methods of peptide-level FPOP analysis would be to incorporate single digestion (trypsin alone) and also tandem digestion from enzymes (for example, chymotrypsin in combination with trypsin), to create shorter predictable peptides which are overlapping when the datasets are analysed side-by-side. This would allow one to build a % modification heat map, and confidence could be given to the assignment of modification to narrower sections of peptides, still without the need for amino-acid level analysis. Because of time constraints, this was not performed in this thesis. However, from the results generated here in this chapter, the interpretation of the current data was made with the basis that the conclusions from the n = 2 dataset should not be overstated, and that more repeats of this data should allow statistical significance to be generated from the results.

5.3 Global observations of native WFL, STT and 114 oxidation

Overall, the highest percentage oxidations observed are in the peptides spanning the HC CDRs; 24-ASGG-38 (V_HCDR1), 39-QAPG-67 (V_HCDR2) and 102-IYDL-132 (V_HCDR3) (Figure 5.3 a, V_H). The LC CDR peptides 18-VTIS-46 (V_LCDR1) and 47-LLIY-62 (V_LCDR2) oxidised to a lower extent, and V_LCDR3 was not identified in any variant. The highest amount of oxidation for all three variants was in peptide 102-IYDL-132, which incorporates the V_HCDR3, supporting the observations made by the HDX experiments in Section 3.5: Probing differences in intrinsic dynamics over time with HDX-MS, where the V_HCDR3 had a high % deuterium uptake relative to most of the rest of the sequence. It is known that, generally, the V_HCDR3 of mAbs is inherently flexible for function, and makes a large conformational change upon antigen binding (Sela-Culang *et al.*, 2012). The HDX labelling experiments measured deuterium uptake for WFL, STT and 114 after incubation

for up to 20 minutes. Residues 110-TAYYDM-115 appeared to have high deuterium uptake over this time period. This stretch of residues is contained in the 102-IYDL-132 peptide quantified by FPOP. By HDX, 114 had the highest propensity to be labelled in this region compared to the other mAbs, shown in the differential deuterium uptake plots. The difference in deuterium uptake at 5 minutes in the V_HCDR3 region was -50% for WFL compared to 114, and -40% for STT compared to 114, indicating a much greater uptake of deuterium in 114 than the other variants in this region at this time point (Figure 3.19 and Figure 3.20). At the 30 minute time point, the differences between WFL and STT compared to 114 were -20% deuterium uptake, still indicating a higher uptake in 114 over time and an overall flexibility in the V_HCDR3 region. Conversely, when labelled by FPOP, the three mAbs showed similar percentage modification levels over the 102-IYDL-132 peptide containing the V_HCDR3: 47-54% for 102-IYDL-132 (STT: 47.8 ± 4.1%, WFL: 49.1 ± 2.0% and 114: 53.8 ± 13.3%, Figure 5.3 a). The mean percentage oxidation of 114 was higher (53.8 ± 13.3%) than WFL and STT, but because had the largest error (range of the data) between the experimental results, more work would be needed to understand if this difference could be significant.

As discussed above in Section 5.2.1, the large range difference between the 114 sample are likely to be resulting from differences in ion intensity between the two repeats rather than commenting on conformation. However, considering the comparison of the trends of mean deuterium uptake (in HDX) compared to oxidative labelling (in FPOP) between the variants themselves, it is unclear why the FPOP labelling would show a similar level of FPOP labelling when comparing the variants, but the HDX over the 5 and 30 minute timepoints showed 114 having greater levels of deuterium exchange. One suggestion may be that, due to the FPOP peptide 102-IYDL-132 being long, it is showing general solvent accessibility and this has several regions which are solvent exposed and therefore ‘dilute’ the observed effect of increased solvent accessibility of the CDR3 region, compared to HDX which has much shorter peptides localised to the region of interest. This was further explored in Section 5.2.1 above. This explanation of the difference between the two MS methods is particularly likely due to the fact that peptide 102-IYDL-132 (identical in all variants) contains two methionine residues, which are highly reactive to oxidation and would cause the total oxidation of the peptide to be high, swamping the effects of differences in oxidation elsewhere on the peptide. Therefore, to test this further (and not relying on peptide miscleavages), it would be useful to implement a sequential digestion step in the procedure of this project to localise oxidation to regions of interest more clearly. An additional interpretation could be that the large difference between the two repeats in 114 indicates that the FPOP may be capturing the existence of two folding

states of 114 inherently present in solution, similar to the large difference between the states captured in HDX at 5 minutes and 30 minutes time points. However this is unlikely to be the case between the n = 2 samples here. More repeats of the datasets would begin to tease this information apart, or additionally, incubation of sample before FPOP for comparable time periods would allow the investigation of this point further. Overall, the agreement in global high levels of labelling between these two distinct experiments supports the combined use of FPOP and HDX to comment on differences in solvent accessible surface area (SASA) of the mAbs in the CDR3 region, where 114 has a slightly greater inherent flexibility in its CDR3 than WFL and STT.

For the majority of the peptides in both HC (9 out of 13 peptides, Figure 5.3) and LC (5 out of 9 peptides, Figure 5.4), 114 has a globally higher mean percentage oxidation than STT and WFL; the exceptions being 24-ASGG-38 (V_HCDR1), 39-QAPG-67 (V_HCDR2), 88-SEDT-98 and 133-GPSV-144 in the HC, and 107-LTVL-114, 115-AAPS-133, 134-ATVL-153 and 194-SYSC-208 in the LC. The reason for this is unclear. As FPOP probes SASA, a tentative explanation for this overall higher level of oxidation could be that native 114 is globally more open in structure than both STT and WFL, and exists in a slightly more relaxed folded state (higher SASA) than its parent mAbs at equilibrium in formulation buffer.

In terms of the heavy chain CDRs 1 and 2 (24-ASGG-38 (V_HCDR1) and 39-QAPG-67 (V_HCDR2)), 114 has oxidation percentages that are closer to STT than WFL; the mean percentage modified (and ranges) of WFL for these two peptides far exceed 114 and STT; for 24-ASGG-38, the percentage modifications are WFL = 40.1 ± 4.9%, which is much greater than 114 (= 16.3 ± 4.6%), which in turn is greater than STT (= 7.3 ± 0.2%). For 39-QAPG-67, the percentage modifications are WFL = 40.9 ± 0.2%, which is much greater than 114 (= 27.3 ± 9.7%), which in turn is greater than STT (= 20.5 ± 5.7%). This is explored in more detail in the next section.

5.3.1 Oxidation in the regions surrounding F29, W30 and F31

CDR1 contains two (WFL/STT) or three (114) amino acids of interest. V_HCDR1 is sampled by peptide 24-ASGG-38 – here, W30 and F31 have been rationally mutated to S30 and T31 in STT, and F29 was substituted to S29 in 114. The relevant results from Figure 5.3 have been displayed in Figure 5.7 for clarity.

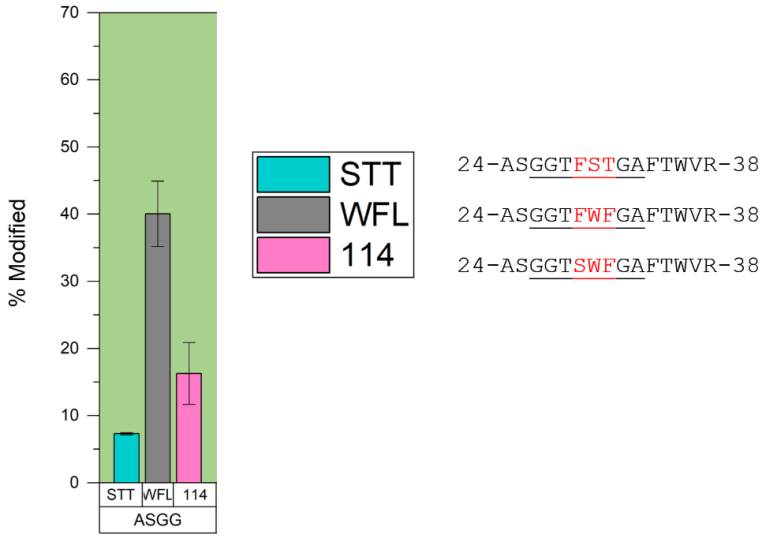


Figure 5.7 Percentage modified in FPOP for the V_HCDR1 peptide 24-ASGG-38.

The abbreviated label below the bar charts correspond to the first 4 amino acids in the peptide sequence. Teal = STT; grey = WFL; pink = 114. Summary of the mean % modification (n=2) and the error bars indicate the range between duplicate datasets. Peptide sequences are shown, with CDR1 underlined and positions of mutation in red.

Hydrophobic residues W and F are highly reactive to hydroxyl radicals compared to S and T (expected fold difference in reactivity is 40.6 times and 13.5 times, respectively (see Introduction Table 1.3 (Xu & Chance, 2007)) which is reflected by the lower oxidation in STT (Figure 5.3, Figure 5.7). The oxidation in STT would likely be on F29 as the scavenging effect of the W residue in the centre of the CDR region of the peptide is not there. All three variants have other phenylalanine (F34) and tryptophan (W36) residues in the peptides, however these do not make up part of the CDR loop and are usually buried from solvent, so would scavenge the oxidation from the CDR peptides to a lesser extent. Interestingly, 114's oxidation in this peptide is more similar to STT than WFL, where only 1 out of 3 highly reactive residues have been mutated. 114 has W and F residues, but the mutation F29S appears to have a greater effect at lowering the oxidation level, compared to replacing W and F with S and T. For example for STT, the change from W to S decreases the expected reactivity at this position by ~41 times, and F to S by ~14 times. However, for 114 the change from F to S decreases the expected reactivity by ~14 times, but the overall modification percentage is closer to the value for STT. This could indicate that the sequences SWF and FST might have similar solvent accessibilities and therefore overall reactivity levels, based on the similar oxidation results of STT and 114 on the 24-ASGG-38 peptide (7.3 ± 0.2% and 16.3 ± 4.6%, respectively, Figure 5.3 and Figure 5.7). The solvent-exposed hydrophobic and bulky phenylalanine F29 residue has been replaced with a polar small serine, which highlights

the potential of a structural rearrangement which shields the hydrophobic W and F residues more from radicals in 114.

Previous studies on the oxidation of reactive residues in the V_HCDR1 and V_HCDR2 regions of WFL compared to STT suggest a structural rearrangement of the V_HCDR2 loop in the V_H domain (Cornwell *et al.*, 2019). Cornwell *et al.* investigated the residue-level modification in the two peptides spanning 24-38 and 39-67 (24-ASGG-38 and 39-QAPG-67). F29 and F34, located either side of the W30S and F31T substitutions in the 24-ASGG-38 peptide, were only modified in STT, but oxidised to a low percentage (less than 0.5% oxidation (Cornwell *et al.*, 2019)). Additionally, W36 (found in all three variants in the 24-ASGG-38 peptide) was found to be almost six times more oxidised in WFL than STT (7.5% compared to 1.3% respectively), which was surprising as it is usually buried from solvent, therefore the results indicated a change in conformation around this residue where it was more exposed to solvent. In line with these previous results, the peptide data for 24-ASGG-38 and 39-QAPG-67 in Figure 5.3 shows a greater level of oxidation in the CDRs 1 and 2 of WFL over STT and 114. This difference may be coming from the exposure of W36 in the WFL variant, but as this can only be investigated through residue-level information, my WFL XIC results for the 24-ASGG-38 peptide in this study were not of a sufficient spatial resolution to assign oxidation to particular amino acids. If one considers 114 alongside these results, one can see an apparent STT<114<WFL oxidation hierarchy in these CDR regions, potentially indicating that the W36 remains more 'STT-like' and buried, compared to the structural rearrangement which may have occurred for WFL.

Previously, Cornwell *et al.* reported that F55 was two times more oxidised in STT than WFL (Cornwell *et al.*, 2019). This amino acid is proximal to the L57T substitution site in the V_HCDR2 loop, and is spatially proximal to the W30S and F31T substitutions. On first observation, my results seem to contradict this observation; the mean % oxidised for the peptide 39-QAPG-67 (spanning Phe 55) was two times higher in WFL than STT, at 40.9 ± 0.2% modified rather than 20.5 ± 5.7% respectively, Figure 5.3 a, expanded in Figure 5.8). The published results showed one residue oxidised in the sequence 53-PIFGLTNLAQNFQGR-67; however the peptide that was used for quantification was peptide 39-QAPGQGLEWMGGIIPIFGLTNLAQNFQGR-67 (identified in every repeat of my mAb data, whereas 53-PIFGLTNLAQNFQGR-67 was only found in unoxidised form in the STT sample). This longer peptide contains more residues (with reactive W47 and F55 as the most potentially reactive sites of oxidation), and additionally a methionine at position 48, which is highly prone to oxidisation even at a background level in the ESI process (Cornwell *et al.*, 2019; Johnson *et al.*, 2019; B. Zhang *et al.*, 2018). Therefore, this peptide has the opportunity to be oxidised to a greater extent than the shorter peptide, which could account for the

increased oxidation overall in the WFL peptide, due to the first half of the peptide potentially being more solvent exposed in WFL's orientation compared to STT which would overshadow the change in F55 oxidation alone.

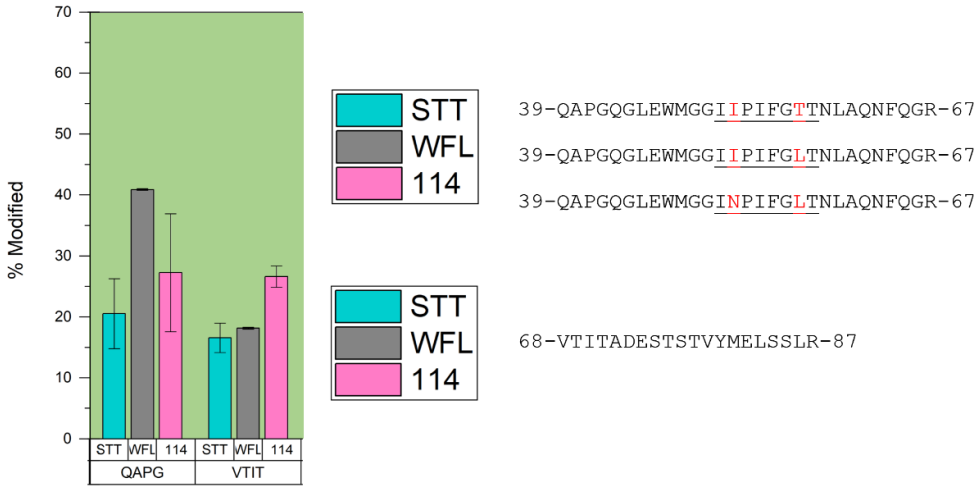


Figure 5.8 Percentage modified in FPOP for the V_HCDR2 peptide 39-QAPG-67, and the following peptide 68-VTIT-87.

The abbreviated label below the bar charts correspond to the first 4 amino acids in the peptide sequence. Teal = STT; grey = WFL; pink = 114. Summary of the mean % modification (n=2) and the error bars indicate the range between duplicate datasets. Peptide sequences for QAPG are shown, with CDR2 underlined and positions of mutation in red.

Furthermore, peptides 39-QAPG-67 and 68-VTIT-87 are close in sequence and proximal to each other in the three-dimensional fold of the mAbs (based on the predicted structure in Figure 5.2). Their differences in oxidation patterns between the variants (Figure 5.8) might provide information on how the substitutions might have affected the local structure and solvent accessibility. In 114, the mutation of I52N in the V_HCDR2 (sampled in peptide 39-QAPG-67) decreases the inherent reactivity rate of the residue at this site, from 1.8 x 10^9 M^{-1} s^{-1} to 4.9 x 10^7 M^{-1} s^{-1} (~37-times reduction in reactivity). However, the oxidation of the peptide 39-QAPG-67 is 27.3 ± 9.7% for 114 compared to 40.9 ± 0.2% for WFL, indicating that there are additional influences (such as changes in solvent accessibility) which have affected the oxidation of this peptide, not just the change in reactivity of the residue. The I51N substitution in 114 may have influenced the native fold of the V_HCDR2 region as it is proximal to peptide 68-VTIT-87. 68-VTIT-87 has an identical sequence in all three mAbs, and is modified between 16-27% (Figure 5.8), where WFL = 18.2 ± 0.2%, STT = 16.6 ± 4.9%, and 114 = 26.6 ± 1.8%. This peptide is situated close to the V_HCDR2, where I70 (in 68-VTIT-87) is packed close to the hydrophobic isoleucine run

(51-IIPI-54) in the V_HCDR2 of peptide 39-QAPG-67. The substitution of hydrophobic I52 with polar residue N52 in 114 reduces the hydrophobicity of the buried core area, which potentially has decreased the packing ability of the loop over the hydrophobic regions. This may be having the effect of increasing the effective solvent accessible area in this peptide loop, therefore increasing the accessibility for hydrogen peroxide to oxidise the 68-VTIT-87 peptide in 114 compared to both WFL and STT.

5.3.2 Observing oxidation around the C_L-C_H1 and C_H1-C_H2 interfaces

In addition to differences in FPOP labelling in V_HCDR2, Cornwell *et al.* also observed significant decreases in WFL labelling around the C_L-C_H1 interface compared to STT (Cornwell *et al.*, 2019). They reported decreases in labelling at V136 (C_H1), F138 (C_H1), W152 (C_L) and 186-PEQWK-190 (C_L) in WFL, all sections proximal to the C_L-C_H1 interface (Cornwell *et al.*, 2019). These positions have been labelled in Figure 5.9 on the energy-minimised model of STT. The position of these protected areas may suggest WFL and STT have different C_L-C_H1 orientations. Visualising the amino acid positions on this structure, an alternative interpretation could be proposed, where WFL is more likely to populate a structure where the Fab arms are closer together over the hinge region in the native structure (where the red arrow points in Figure 5.9), or that intermolecular interactions were being detected (mAbs shielding each other from oxidation in this region during the labelling process).

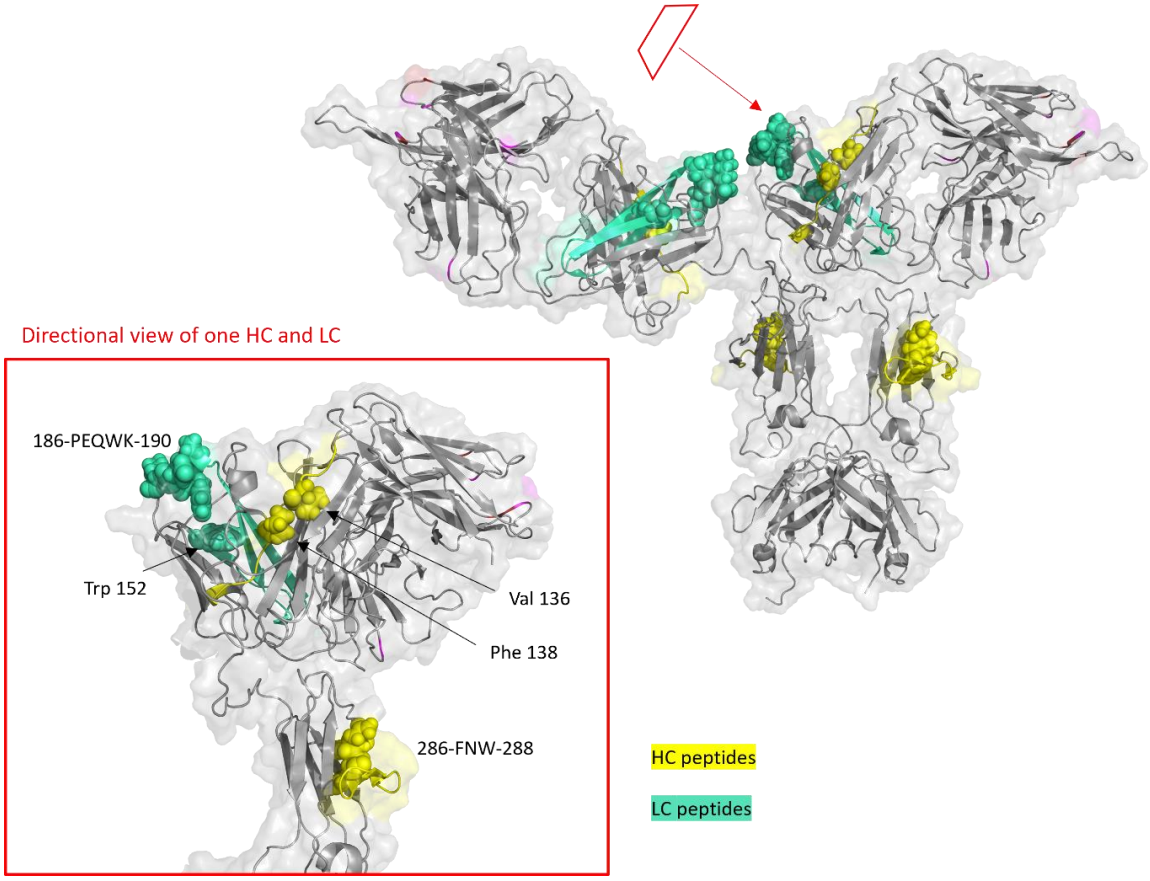


Figure 5.9 Amino acid positions in the C_L-C_H1 and C_H1-C_H2 interfaces that showed differences in labelling between WFL and STT in previous work.

Oxidation positions identified in (Cornwell et al., 2019) are represented in spheres on the ribbon diagram. The full peptides these amino acids reside in are coloured accordingly on the ribbon diagram; HC peptides = yellow; LC peptides = teal. Red trapezoid represents the directional field of view of the insert image and red arrow gives the direction the insert image was taken from, to visualise the positions of the residues from above the Fab arm.

In the published work, W152 (C_L) had a 5-fold decrease in oxidation compared to STT (Cornwell *et al.*, 2019). In comparison, at peptide-level resolution in this thesis, peptide 134-ATLV-153 (which contains W152) showed a similar level of labelling between the variants (although the observed range of modification for the measurements was high), where STT = 22.6 ± 17.8%, WFL = 17.4 ± 2.2, 114 = 13.1 ± 2.5 (Figure 5.4). For this peptide, 114 appears to have a slightly lower % oxidation than WFL (13.1% compared to 17.4%), suggesting more protection from oxidation in 114 than WFL, however there is no large fold decrease in labelling between STT and WFL at the peptide-level analysis compared to the published data. This might be due to the effects of oxidation on other amino acids in this peptide that counteracts any decreases in oxidation in the WFL variant. Additionally, my results show no difference in oxidation of the HC 133-GPSV-144

peptide (which spans the amino acids V136 (C_H1) and P138 (C_H1) amino acids mentioned in Figure 5.9) for all three variants (STT = 1.4 ± 0.01%, WFL = 1.4 ± 0.2%, 114 = 1.4 ± 0.01% (Figure 5.3). Also, my results show almost no difference in oxidation of peptide 176-YAAS-190 (containing 186-PEQWK-190 (C_L)) between STT (7.4 ± 0.2%) and WFL (8.3 ± 0.8%), but a slight increase in oxidation for 114 (12.8 ± 1.2%) (Figure 5.4).

The differences between my data and the published oxidation changes in these C_L-C_H1 interface regions may indicate that the peptide-level analysis method has been – in this case – insensitive to any small changes in oxidation between individual amino acids, and therefore may have missed subtle information which detects local structural rearrangements that affect amino acid orientation. Any large structural rearrangements are likely to affect more than one residue, particularly ones adjacent to the residue in 3D space. Therefore a peptide-level analysis may be more inclusive of all changes to SASA in a section of sequence.

Changes to the C_L-C_H1 interface are thought to be related to antigen binding, suggesting a conformational link between the variable regions and the constant domains (Sela-Culang *et al.*, 2012). My results do indicate a slight increase in solvent accessibility of 114 in the proximity of 186-PEQWK-190 (C_H1) residues, based on the higher oxidation level for the 176-YAAS-190 peptide (114 (12.8 ± 1.2%) compared to STT (7.4 ± 0.2%) and WFL (8.3 ± 0.8%)), therefore the HC mutations in 114, and their proximity to the hydrophobic CDRs, could be distally affecting the C_L-C_H1 interface to a greater extent than WFL or STT's amino acid sequence in the variable regions.

Additionally, Cornwell *et al.* report a slight decrease in oxidation (<2-fold) for WFL compared to STT on the upper section of the C_H2 domain: residues 286-FNW-288 (Cornwell *et al.*, 2019), which form part of the 286-FNWY-299 peptide (identical between the variants). Again, my peptide-level oxidation results display little to no difference between WFL and STT (8.1 ± 1.0% compared to 6.7 ± 1.0% respectively), however 114 had increased oxidation (17.8 ± 1.8%). This might support the suggestion that 114's C_H1 and C_H2 domains tend to be in a different orientation than WFL's or STT's, allowing a greater SASA for labelling in the upper C_H2 region, or that there is a more significant change to the C_L-C_H1 interface in 114 which affects the Fab arm local protection.

5.3.3 Light chain mutations in 114 do not appear to affect local oxidation

One out of two LC mutations in 114 are observable in the peptides identified in this study: 1-Q(S/G)VL-17 spans the S2G mutation in the V_L (the distal T77A (V_L) mutation was not observed). In WFL and STT, which have identical sequences for the 1-Q(S/G)VL-17 peptide, there was no difference in the low levels of modification identified (WFL = 1.1 ± 0.1%, STT = 0.9 ± 0.2%, Figure 5.4). 114 was modified to 2.0 ± 0.2%. Serine residues have a low propensity to oxidise (serine's expected reactivity rate with ·OH = $3.2 \times 10^8\ M^{-1}\ s^{-1}$) and glycine has an even lower propensity (G = $1.7 \times 10^7\ M^{-1}\ s^{-1}$) (Xu & Chance, 2007), a decrease in expected reactivity of 18-times, however the full peptide has an increased modification (two times) compared to the peptide in WFL and STT with a serine residue. Whether this doubling of oxidation level is significant or not would need to be tested with more repeats.

It is interesting to note that the neighbouring peptide 18-VTIS-46 (containing the V_LCDR1) has a large increase in oxidation of almost 3-times in 114 (16.6 ± 1.7%) compared to both WFL and STT (5.2 ± 0.3% and 6.2 ± 0.4%, respectively (Figure 5.4)). The V_LCDR1 loop is situated close to the site of the S2G mutation on peptide 1-Q(S/G)VL-17, so perhaps the mutation has triggered a change in orientation of the two peptides in relation to each other, so that both peptides in 114 show an increased oxidation percentage compared to WFL and STT.

Additionally, the S2G substitution is proximal to the V_HCDR2 peptide 39-QAPG-67, specifically the QAPG amino acids which create an unstructured turn. How the mutation might have influenced localised changes to oxidation propensity (due to scavenging or differences in structural rearrangement) cannot be attributed here using peptide-level analysis, as the length of the 39-QAPG-67 peptide is long and the effects in the hydrophobic CDR2 region might overshadow any oxidation changes localised in the N-terminal section of the peptide.

5.4 Under hydrodynamic flow, the propensity to be oxidised changes

WFL, STT and 114 were prepared to 1 mg mL^{-1} and subjected to 200 passes at 16 mm s^{-1} in the hydrodynamic flow device detailed in Section 1.1.4.4: Extensional Flow Device (EFD) to mimic aggregation under flow (Methods Section 2.2.7: Flow Device). The flow conditions finalised in Chapter 4: Development of an in-house method for FPOP label fingerprinting of monoclonal antibodies were followed for WFL, STT and 114, where the key differences in the pre-treatment of the proteins are detailed in Figure 5.10.

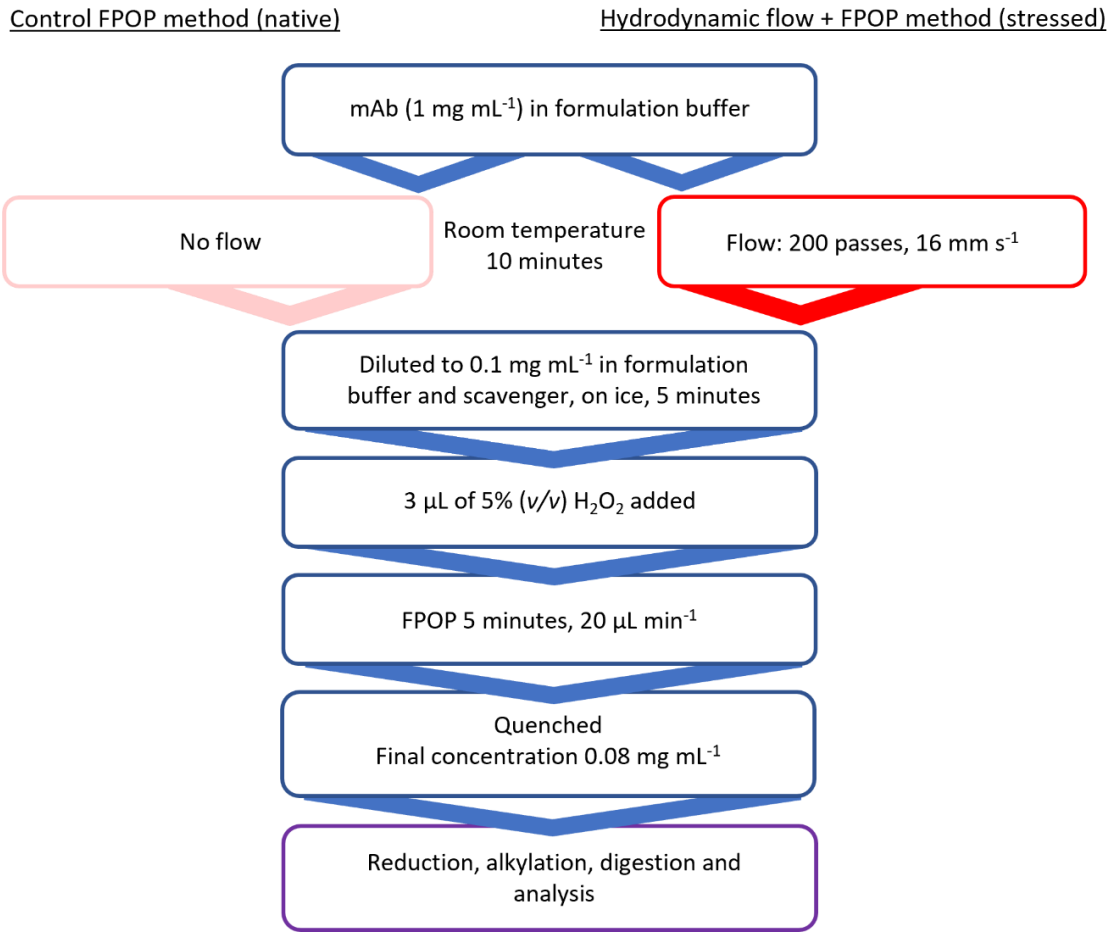


Figure 5.10 Summary of the FPOP workflow without or with flow stress.

The length of time between the end of the force experiment and the commencement of FPOP laser irradiation was 5 minutes, and samples were stored on ice during the dilution and transport to the laser (Figure 5.10). The control (native) sample was matched as closely to the stressed sample experiment as possible; the control sample was stored on the bench (not on ice) for the

duration of the flow experiment; the dilution was made at the same time as the stressed sample; the peroxide was mixed immediately before exposure to the FPOP laser, and the sample was quenched after the FPOP procedure. This carefully curated procedure allowed me to attribute the changes seen entirely to the effects of hydrodynamic force. Therefore, FPOP modification differences between control and stressed sample was explored through the comparison of XIC peak areas. This area difference between unmodified and modified was taken for all observable charge states (if present in every sample) and compared, giving a % modification value for every peptide. The heavy chain data is in Figure 5.11 and the light chain data is in Figure 5.12.

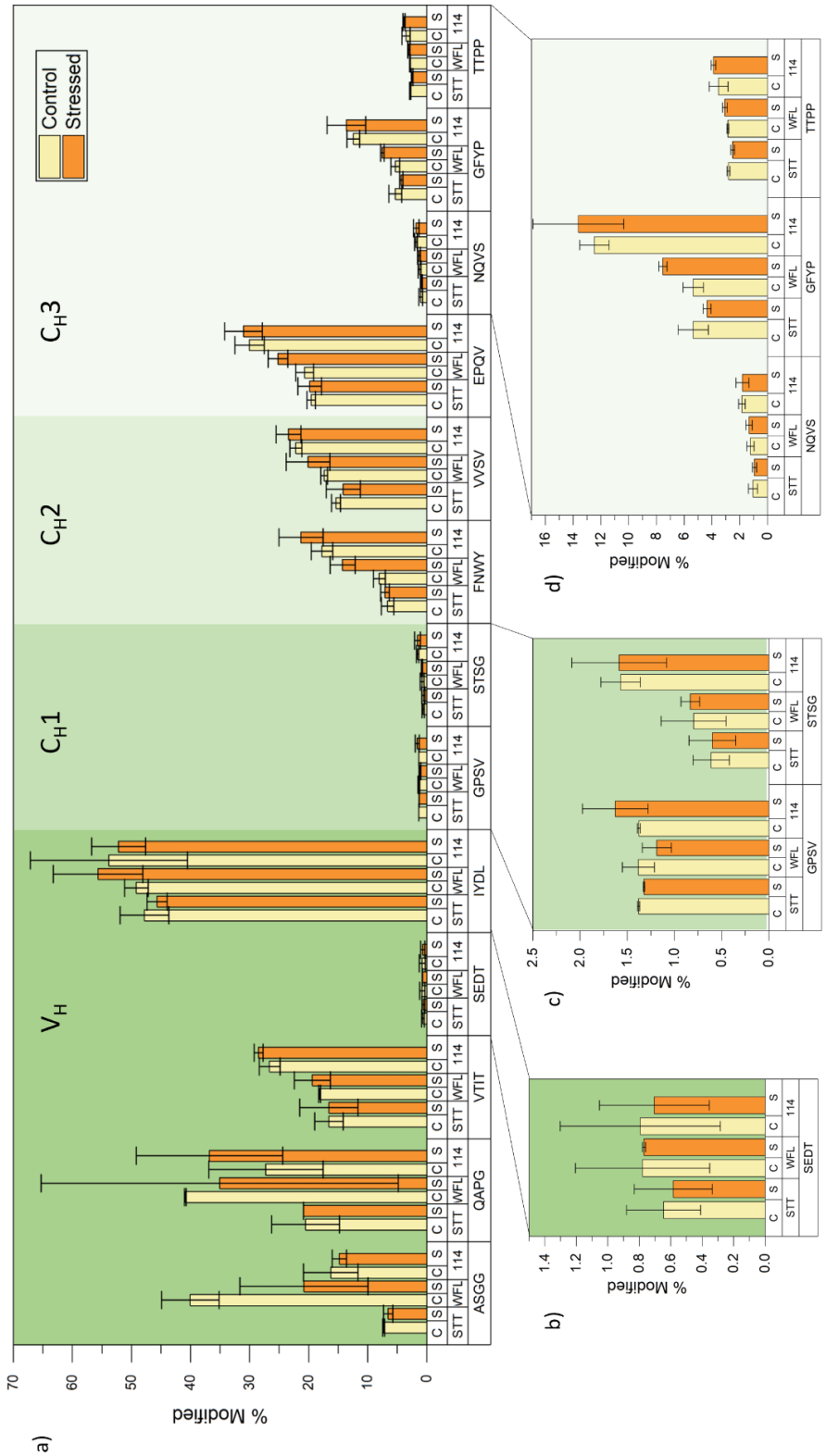


Figure 5.11 Comparison of the mean % modification of the heavy chain of control (un-stressed) and hydrodynamically stressed STT, WFL and 114.

a) Summary of the mean % modification (and range error bars) for 13 peptides that were identified in the heavy chain of each mAb, where each label below the bar charts corresponds to the first 4 amino acids in the peptide sequence. Yellow = control, orange = stressed (200 passes, 16 mm s^{-1}). Values shown are the mean (n=2) and the error bars indicate the range between duplicate datasets. Zoomed in quantification of low abundance modification on peptides are in b) SEDT, c) GPSV and STSG, d) NQVS, GFYP and TTPP. Values are listed in Appendices: Table 7.4.

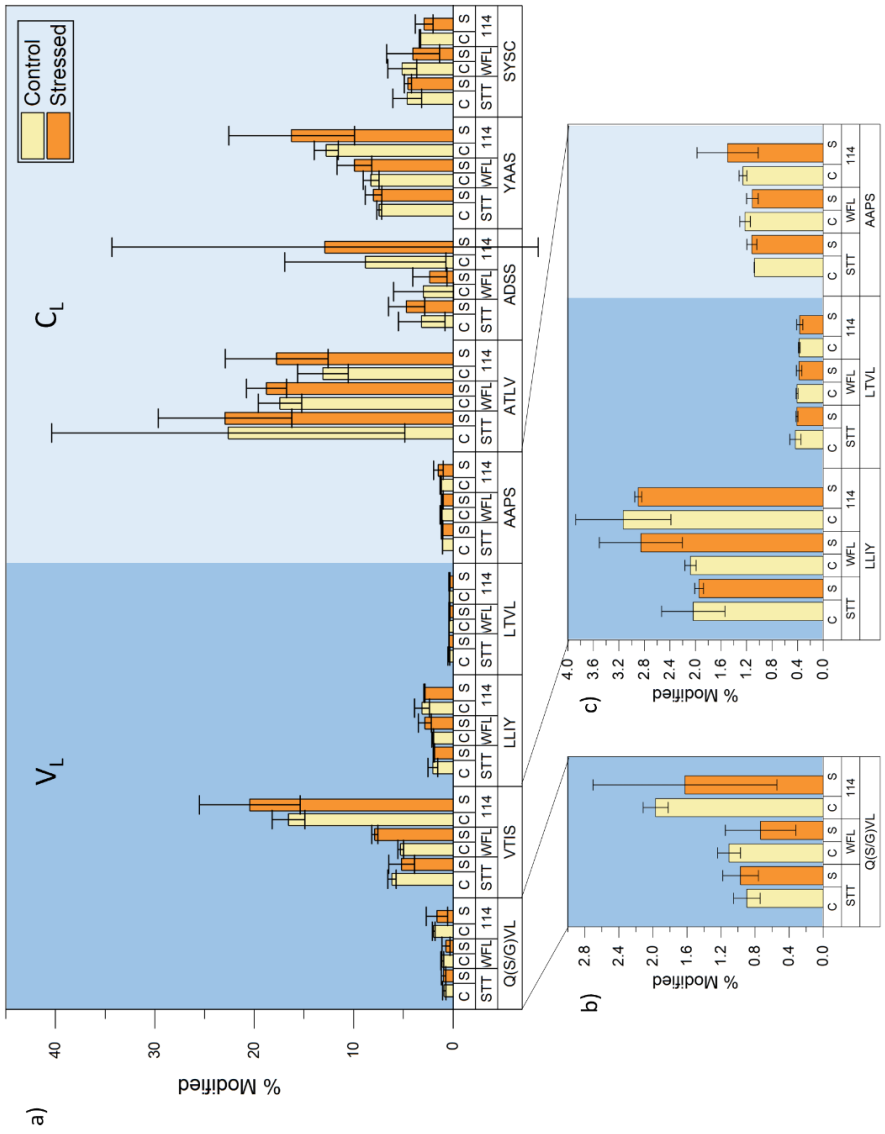


Figure 5.12 Comparison of the mean % modification of the light chain of control (un-stressed) and hydrodynamically stressed STT, WFL and 114.

a) Summary of the mean % modification (and range error bars) for 9 peptides that were identified in the light chain of each mAb, where each label below the bar charts corresponds to the first 4 amino acids in the peptide sequence. Yellow = control, orange = stressed (200 passes, 16 mm s^{-1}). Values shown are the mean (n=2) and the error bars indicate the range between duplicate datasets. Zoomed in quantification of low abundance modification on peptides are in b) Q(S/G)VL, c) LLIY, LTVL and AAPS. Values are listed in Appendices: Table 7.4.

5.4.1 Changes to oxidation levels in response to hydrodynamic stress

Overall, the differences between the mean % modification of control sample compared to stressed are subtle. The majority of heavy and light chain peptides in each protein tend to show slight increases to the mean % modified under flow (however the range between the measurements (n=2 experiments) increases, indicating a greater variability in the observed oxidation after exposure to flow). The most striking exceptions to the increases to the mean percentage oxidation triggered by flow exposure are in the HC peptides 24-ASGG-38 and 39-QAPG-67 (containing the V_HCDR1 and V_HCDR2 respectively) (Figure 5.11) and the LC peptides 1-Q(S/G)VL-17, 18-VTIS-46 (containing the V_LCDR1) and 154-ADSS-170 (Figure 5.12).

The overall low level of difference in oxidation in the majority of the peptides is an expected outcome; the flow conditions chosen in Chapter 4 of this thesis were using a high strain (16 mm s^{-1}), for a reasonable, but not high, number of passes (200). This amounted to 76.1% soluble WFL remaining, and 92.6% STT and 93.2% 114 remaining, measured by the monomer loss assay (Figure 3.16). These flow conditions were chosen for study to perturb native structures but to limit the amount of protein lost into insoluble aggregations. These conditions aimed to allow the study of any subtle structural perturbations and early unfolding, not involved in late stages of aggregation. It was hypothesised the changes to FPOP modification amounts will be subtle in these preliminary experiments. It would be fascinating to delve into this more deeply, probing different times after the force event, different concentrations, different buffer conditions, and so on. However, in this thesis, the novel application of FPOP to detect changes in oxidation before and after extensional force at one discrete time point has been demonstrated, and the key differences will be discussed in this section.

5.4.2 Exploring oxidation differences in the HC CDRs

The most clear differences before and after flow stress are in the peptides spanning V_HCDR1 and V_HCDR2. In the control data for the peptide 24-ASGG-38, which spans V_HCDR1 and W and F residues (Figure 5.3 and Figure 5.7), WFL has the highest percentage modification over 114 and STT (WFL = 40.1 ± 4.9%, 114 = 16.3 ± 4.6%, STT = 7.3 ± 0.2%). After exposure to the hydrodynamic force, the mean oxidation level for this peptide reduces by a half for WFL (WFL_S = 20.8 ± 10.9%), where there is virtually no reduction in labelling for 114 and STT (114_S = 14.8 ± 1.2, STT_S = 6.6 ± 0.8% (Figure 5.11)). An explanation of this could be that there is a structural rearrangement of

WFL's V_HCDR1 region in response to flow, where the oxidation-prone residues have been placed in a more shielded environment and were less prone to oxidise compared to the control, and STT and 114 do not have a structural change in response to flow. As discussed in the sections above, the mutation F29S in peptide 24-ASGG-38 in 114 appears to counteract the presence of W and F somewhat, where the percentage oxidation of the control sample is more similar to STT than WFL.

An alternative suggestion to the reduction in oxidation of WFL's peptide 24-ASGG-38 after flow exposure is that the CDR1s are becoming more involved in intermolecular associations (and that STT and 114 are both resistant to unfolding and associating in this region). The CDRs of antibodies are often composed of hydrophobic residues (Ausserwöger *et al.*, 2022), and indeed the presence of the hydrophobic 29-FWF-31 motif provides an ideal association interface. This interface has previously been mapped using HDX and chemical cross-linking analysis of WFL (C. L. Dobson *et al.*, 2016). The drop in WFL oxidation after flow exposure in this experiment might be demonstrating that the mAb is forming flow-induced associations. However, this is unlikely to be the sole reason for the decrease in oxidation from 40.1 ± 4.9% to 20.8 ± 10.9% after flow exposure; in the flow experiments in Chapter 3 where flow conditions were chosen, WFL loses 23.9% monomer (Figure 3.16), and although this material can be pelleted out, is remains in the total sample in the FPOP experiments due to the fact that it would take a lot of time to pellet out and therefore there is less chance of seeing any transient structures influenced by the flow experiments. This material could be influencing the drop in global oxidation however this is only 23.9% of the total signal and may not be associating in a way which shields the CDR1. Additionally, there could be soluble oligomers which form during the flow process (which may or may not associate via the CDR1) that influence the total oxidation observed. These oligomers could form transiently (reversible self-interactions) or be irreversible, and shield the CDR1. Therefore, to decrease the oxidation at the 24-ASGG-38 peptide by a half, there must be additional in-solution affects which decrease the reactivity to radicals at this peptide, such as the presence of soluble oligomers (which are able to be digested and then observed in the MS procedure). Also, the drop in oxidation after flow could be due to local rearrangements of the CDR1 which place the FWF residues into a more buried position in the mAb V_H-V_L interfaces, or perhaps the residues are rearranged to be closer to other hydrophobic aromatic residues which may further scavenge the ·OH and prevent oxidation at the 24-ASGG-38 peptide. Further work would need to be made to validate these suggestions.

The peptide 39-QAPG-67, which spans the V_HCDR2, also appears to have changes to the mean percentage modified, however the large range of modification (the error bars) of the data make

it difficult to conclude any significant changes to peptide exposure or protection (Figure 5.11). This peptide in 114 tentatively appears more modified, however whether this is significant or not will need to be investigated through repeating the experiments. Additionally, for after flow exposure, the light chain peptides 134-ATLV-153, 154-ADSS-170, and 176-YAAS-190 have increased mean percentage modifications (Figure 5.12), but again the large range of modification (the error bars) of the data make it difficult to interpret whether these apparent increases in labelling after flow are real. As these three light chain peptides have identical sequences in all three variants, it would be interesting to investigate whether the substitutions elsewhere in the sequences have differing effects on structural rearrangements to the C_L region containing these peptides.

5.4.3 Residue level information may inform oxidation patterns after flow in the V_HCDR1

Changes to retention time for oxidised peptides, relative to the unmodified version, is a well-established phenomenon (Y. Zhang *et al.*, 2017). The relationship between changes in retention time of peptides in an FPOP experiment, and the nature of the modified peptide – such as which residue is modified – has been investigated recently (Cornwell, 2019; Cornwell *et al.*, 2019). This means that, when a peptide has a modification in the sequence, the residue that is modified (and the location of the modification on the side chain, in cases such as tryptophan and phenylalanine) affects its retention time. The MS/MS analysis can identify the sites of amino-acid resolution, which can allow the comparison of very localised changes in structure, through side-chain rearrangements and changes to solvent accessibility.

Therefore, to try to further understand the flow-induced change in oxidation in the peptide 24-ASGG-38, amino-acid resolution analysis was employed for this peptide from the three mAb variants (Table 5.2, Figure 5.13). Modifications for +16, +32 and +48 were searched, and MS/MS was used to identify which residue was most commonly modified for each peptide with a resolvable retention time. If the resulting fragment ion spectra allowed the assignment of modification position, the XICs were plotted and compared. The results are summarised in Table 5.2, and the XICs for ASGG peptide for each variant are shown in Figure 5.13.

Table 5.2 Common modifications positions identified for the 24-ASGG-38 peptide for the three mAb variants.

*Positions of interest in the sequence are highlighted in red type. Asterisks (*) mark the residues where there is MS/MS evidence for modification localisation at amino-acid level accuracy, where red * = +16 Da, blue * = +32 Da, and green * = +48 Da. Green bars indicate where the modification could not be localised to a particular amino acid, but rather somewhere within the stretch of amino acids the green bar spans.*

	Modification position - FLOW	Modification position + FLOW
WFL * = +32 * = +16	ASGGTFWFGAFTWVR	ASGGTFWFGAFTWVR
STT * = +48 * = +32 * = +16	ASGGTFSTGAFTWVR	ASGGTFSTGAFTWVR
114 * = +32 * = +16	ASGGTSWFGAFTWVR	ASGGTSWFGAFTWVR

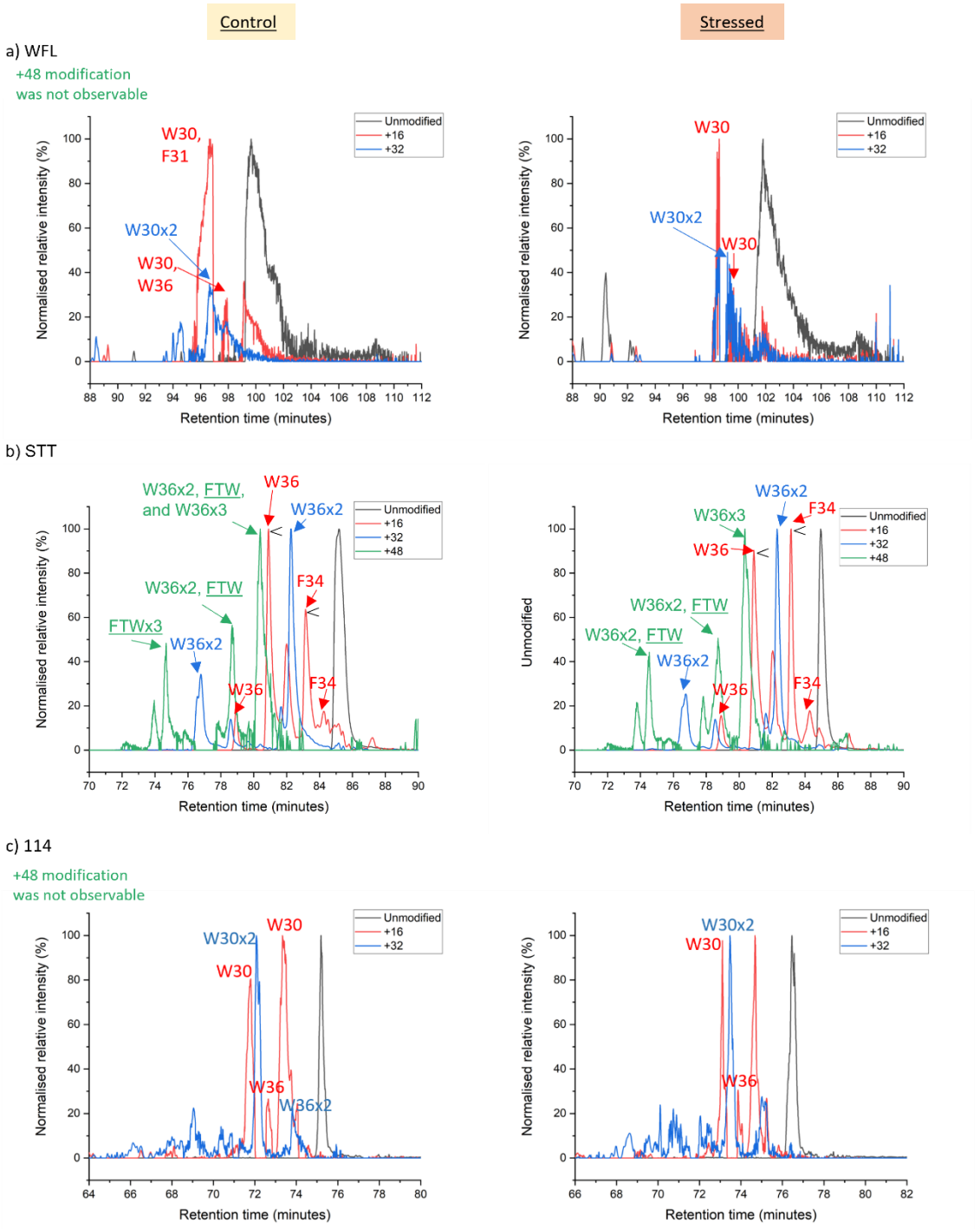


Figure 5.13 Representative XICs for peptide 24-ASGG-38 from WFL, STT and 114, without (control) and after (stressed) flow exposure.

Unmodified peptide XIC = black; +16 modified peptide XIC = red; +32 modified peptide XIC = blue; +48 modified peptide = green. +48 modification was only found at a level above the noise in STT. Control XICs = left hand column; stressed XICs = right hand column. a) WFL; b) STT; c) 114. The modification positions which were confirmed from the MS/MS (using PEAKS) are annotated. < indicates a change in intensity discussed in the text.

The WFL XICs are poorly resolved with broad elution peaks, indicating that the hydrophobic-rich peptide interacts strongly with the C18 column. This is disadvantageous as the broad peaks can mask any discrete elution times which result from different positional isomers. To improve this, the concentration of organic buffer used to elute the peptides would need to be higher and at a steeper gradient to improve the XIC resolution. A control peptide's representative XIC is shown in Figure 5.14 to demonstrate that other peptides in WFL act similarly to the corresponding peptide in the other variants. The peptide 102-IDYL-132 XICs for all three variants are similar and this peptide does not have solubility problems in WFL. This control demonstrates that the peptide 24-ASGG-38 has a solubility issue contributing to the broad peak shapes in Figure 5.13, but this is an exceptional case as the other peptides in the sample have a more resolved peak shape for FPOP analysis. The peptide 102-IDYL-132 was chosen as this example due to its similar levels of FPOP between the three variants, both in the control and the stressed sample (Figure 5.14).

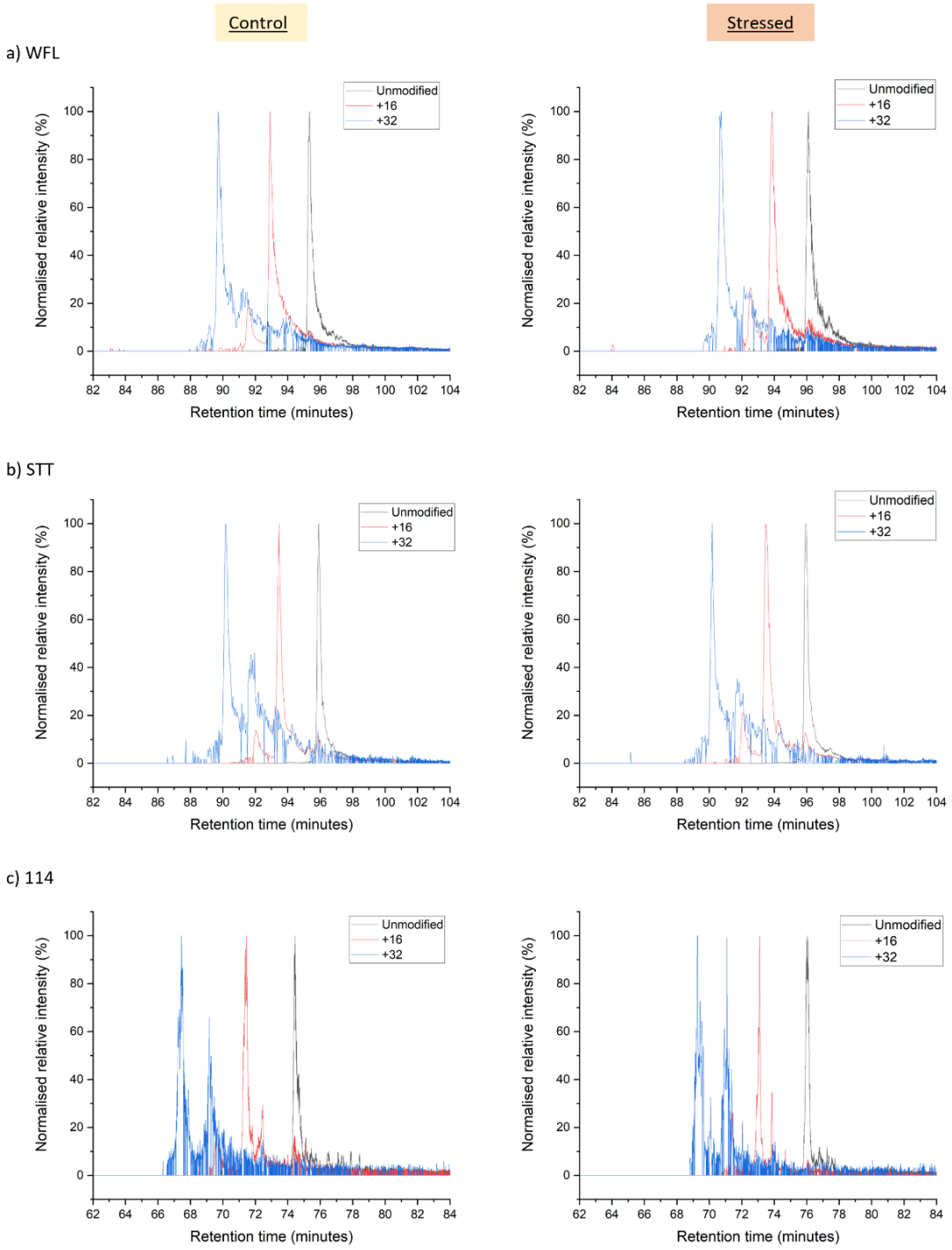


Figure 5.14 Representative XICs for peptide 102-IDYL-132 from WFL, STT and 114, without (control) and after (stressed) flow exposure.

Unmodified peptide XIC = black; +16 modified peptide XIC = red; +32 modified peptide XIC = blue. Control XICs = left hand column; stressed XICs = right hand column. a) WFL; b) STT; c) 114.

Despite the poorly resolved trace for WFL in Figure 5.13, it is clear that shifts in the XIC traces occur upon +16 and +32 modification. When stressed, the apparent position for +16 modification shifts from W30, F31 and W36 to only W30. This accords with the drop in mean % modification identified on this peptide after flow (20.8 ± 10.8%, compare to 40.1 ± 4.9% without flow (Figure 5.11). For the +32 trace, the only resolvable peak demonstrated that both oxidations were found in the tryptophan residue, and it is known that multiple oxidations on the W30 residue would have attached in several positions on the W30 ring (positional isomers of W oxidation, also identified in (Cornwell *et al.*, 2019)). The multiple other peaks in the XIC indicates that there could be a combination of positions (meta-, ortho- and para-) over the aromatic ring where these oxidations would attach, which can affect the migration of the peptide chain in the column before MS. The +48 trace was unidentifiable, which may be to do with the fact that the peptide in general is poorly resolved, and the extra oxidation would make the peptide even more hydrophilic.

The XICs extracted for STT are clearly resolved. Here, +16, +32 and +48 XICs are highly similar before and after flow exposure respectively. Interestingly, there are however subtle changes to the preference of F or W for some +16 modifications before and after stress (indicated by < on b) in Figure 5.13. The preference for W36 appears to switch to F34 – this indicates that a subtle structural arrangement may have placed F34 into a more exposed position after flow, such that the hydroxyl radical may attack the phenylalanine more preferably. A structural rearrangement is more likely to be the reason for this switch here, than intensity issues (discussed in Section 5.2.1 above), due to the STT traces being well resolved and there is a wealth of MS/MS data for the assignment of the modification positions.

114's XIC trace is similar to WFL, due to the main positions of oxidation being the reactive residues W30 and F31. From the MS/MS data, there is a preference for two oxidations to be positioned on W30. When stressed, the XIC intensities for both +16 and +32 do not appear to change drastically, which is supported by the similar mean percentage modification reported in Figure 5.11. Additionally, the +48 XIC was not identifiable in 114, similar to WFL.

Due to the peptide 24-ASGG-38 being rich in hydrophobic, highly reactive residues (most F and W residues in WFL>114>STT), more information could be extracted by looking at the +48 Da modification positions. Indeed, the investigation of +48, +64, +14 and other traces may provide rich information for the majority of the applicable peptides in the datasets in this chapter. Considering these future improvements, weight is added to the observation that FPOP is a highly

information-rich technique, and lots can be done to extract the wealth of information captured in each experiment.

Overall, the results in this chapter demonstrate that peptide-level modification analysis is highly valuable for a general understanding of how a protein may be FPOP modified. There is a clear advantage for the peptide-level analysis for several reasons, including that it is less time-consuming than residue-level analysis, and that it can be employed when overall significant oxidation occurs between conditions to gain an inclusive capture of a peptide (rather than only the residues that can be resolved). It can be supplemented using the residue-level analysis seen in the above section. The power of this technique is apparent and there is strong evidence to support its continued use and further development to capture protein structural changes in response to flow forces.

5.5 Discussion

The peptide-level analysis performed in this chapter has been successful in displaying differences in oxidation between highly homologous mAbs, and also between the mAbs when stressed with hydrodynamic flow for a defined set of conditions. This peptide-level of analysis saves time for the user when manually extracting XIC values from the raw data, but sacrifices spatial resolution of the assignment of the modification. On the other hand, this can be more inclusive, as it captures the complete picture of the peptide's environment, rather than just those residues which can be resolved when analysing at amino-acid-level resolution.

The potential of hydrodynamic flow experiments coupled to FPOP-LC-MS/MS as a technique to investigate perturbations of the monomeric native state, and any higher order states, is demonstrated through the results in this chapter. It is assumed that the flow conditions used here generate mainly activated monomers, and that there are minimal higher order species present which could complicate the interpretation of the output. It was important that the FPOP labelling was performed within the shortest possible time between stressing and labelling, to avoid any refolding and to capture transiently populated states triggered by the hydrodynamic stress. This meant the solution could not be clarified before FPOP labelling (for example by pelleting out insoluble aggregates such as in Figure 3.16. From the hydrodynamic force data in Figure 3.16, when WFL (at 1 mg mL^{-1}) was stressed for 200 passes at 16 mm s^{-1}, ~24% monomer was lost to insoluble pelletable aggregate (whereas STT and 114 only lose ~6%). This higher percentage of presumably higher order pelletable aggregate could potentially be masking the oxidation information of monomer unfolding, as the results here are generated under the

assumption that the data reports on the average of all forms (soluble and insoluble) present at the time of labelling, 5 minutes after the final pass in the flow device. This assumes that the digestion protocol laid out in Chapter 4 is able to efficiently reduce, alkylate and digest the insoluble aggregates in the solution so that they contribute to the FPOP pattern seen.

The time delay between the flow event and labelling was unavoidable in the current experimental set up due to the physical location of the flow device and the FPOP laser on different floors of the building at the University of Leeds. Improvements to the set up could be explored, such as: creating a flow device to sit directly in the laser path; creating a microfluidic device which replicates the flow device on a small scale; using the microfluidic device to capture various time points along the flow pathway in one 'pass' through the extensional flow region. These ideas have been subjects of discussion and research throughout the course of this thesis, and great strides have been made in the development of this idea, which are further explored in the Future direction section below.

It is important to recognise that the flow experiments were performed with mAb samples at 1 mg mL^{-1}, and then the sample was diluted to 0.1 mg mL^{-1} for labelling. Dilution may cause dissociation of any transiently-formed oligomers, however this has not been tested. The experiment was initially designed this way as it has been hypothesised that the dilution aids with dissociating transient higher order species that could cloud the information from those that are just partially unfolded. The separation of transient higher order species would avoid the quenching effects suspected to happen due to the close proximity of (reactive) residues.

Additionally, It is known that WFL exhibits concentration-dependent self-association (C. L. Dobson *et al.*, 2016) and at 1 mg mL^{-1}, WFL is predominantly dimeric (>75% dimer) by AUC. At 0.1 mg mL^{-1}, WFL is monomeric (C. L. Dobson *et al.*, 2016). In this thesis, ultra-low-volume DLS was used to demonstrate that, in the FPOP buffer at 1 mg mL^{-1}, the average particle size of STT and 114 are 8.3 nm (expected for monomeric mAbs) whereas WFL had an average particle size of 13.1 nm, further suggesting a heterogeneous mixture at the beginning of the flow experiment. Therefore it was assumed that diluting from 1 mg mL^{-1} to 0.1 mg mL^{-1} between the flow and the FPOP procedures, any transiently-associated species of WFL would dissociate into monomer – how far this is true for activated monomer or partially unfolded species remains to be fully investigated. The influence of any irreversible aggregate on the output of the FPOP experiment has not been fully evaluated, and the discussion of the results in this thesis have been sensitive to this. Future work could include collecting the pelleted species after flow, resuspending in FPOP buffer, and testing its FPOP signature; this may help to determine whether it is composed

of species which are digestible in the procedure and if it makes a contribution to the signal from solution-phase monomer. Alternatively, the dilution step may be avoided in future experiments by further optimising the FPOP procedure to capture all the structures present in solution at 1 mg mL^{-1}, and following this with a clarification step to fractionate out monomer or higher-order species. This would require further tuning of scavenger and radical concentrations. The process could also be tuned to perform the flow experiment at lower concentrations (such as 0.2 mg mL^{-1}), which would require less of a dilution (with scavenger and H_2O_2 components essential for the FPOP section of the procedure) between the flow experiment and the labelling, to perhaps capture a more representative fingerprint of the species after flow.

Generally, the range of the oxidation percentage measurements for a peptide between duplicate datasets for the stressed data is higher than the control data (average range for the stressed and control data is ± 2.55% and ± 1.69%, respectively) (Figure 5.11 and Figure 5.12). This could be highlighting that there is variability in the structures populated in the unfolding pathways triggered by the flow experiment. As the flow experiment is an ensemble technique and there is no clarification between any format the mAbs may be in during or after the flow event, the inherent flexibility of the structures could have been captured by the large differences in % oxidised. This could be further understood through comparing more experimental repeats.

The global oxidation for 114 appears to be slightly higher for 9 out of 22 peptides in the control dataset; the mean percentage modification values for HC peptides 68-VTIT-87, 145-STSG-158, 286-FNWY-299, 313-VVSV-328, 356-EPQV-371, 382-GFYP-403, and LC peptides 1-Q(S/G)VL-17, 18-VTIS-46, 176-YAAS-190 are on average 6.7% higher than the average mean values for WFL and STT for the same peptide. It is unlikely that the differences in global oxidation of 114 being higher than WFL and STT have resulted from inaccuracies in the method (for example, a higher amount of peroxide added) because of my meticulousness. Additionally the n=2 replicates were as close to biological replicates as possible (same protein preparations, but separate aliquots taken before experimentation) and the variation in most samples are fairly narrow. However, as human error cannot be ruled out, it would be important to gather more replicates of these experiments again where statistical significance (from paired T-tests) may then be carried out. Additional control could be employed in the LC process: this work was done with C18-packed columns made in-house, where this could be further standardised by using a commercially-bought column specifically used for these FPOP experiments.

In an ideal FPOP experiment, each monomeric protein and the immediate solution solvating it, would be irradiated no more than once. It is important that sample does not get exposed too

many times as the introduction of non-native oxidations holds the potential to shift the protein structure to a non-native conformation (Konermann *et al.*, 2010). The ideal oxidation signal we aim for is around one third to one half of the signal being oxidised. This can be tuned using scavengers (such as amino acids or peptides) in the buffer, or adjusting the set-up, where we can assume typically that the radical lifetime is on average in a µsec timescale. As previously investigated (Cornwell, 2019), 20 µL min^{-1} flow rate was used in the set-up in The University of Leeds, and this was proven to be ideal for the majority of sample to be under single exposure conditions; 66% of the sample was under single exposure, 25% experiences no irradiation at all, leaving 9% to experience multiple exposures. All experiments used a laser firing frequency of 15 Hz. As such, I replicated these conditions to make my data as comparable to previously published results as possible. Similarly, previous results were taken in the formulation buffer Arginine Succinate. Arginine in itself is a free amino acid so acts as a scavenger (along with the L-histidine used in all experiments). Therefore, the amount of peroxide added was chosen based on the previous studies which specified how much was needed to generate radicals to get sufficient oxidised protein signal. These conditions were all taken into account when generating the results in this chapter.

Some of the predictions of amino acid distances in Cornwell's work were derived from homology model structures of a generic Fab domain and a generic Fc domain docked in the SWISS-MODEL server. My work is mapped onto an energy-minimised structure generated in Schrödinger (Figure 5.2), created from the exact Fv sequences of each mAb, along with the scaffold sequence used in the creation of the mAbs at AstraZeneca, in an environment tuned to be representative of pH 6.0 (the pH of the formulation/FPOP buffer). These newly created structures are more likely to be more representative of the true folds of the mAbs in solution at pH 6.0. Although the predictions were not made in combination with the buffer composition (Arginine Succinate), where it is commonly thought that Arginine has a blocking effect on hydrophobic interactions between mAbs (N. A. Kim *et al.*, 2016), the generated structure should be more representative of the accurate distances between amino acids. The energy-minimised structure does converge on a structure with a collapse of the Fab arm into the C_H2 region on one side of the mAb. This could tentatively highlight an area in the native mAb which is shielded from FPOP labelling, however these positions did not match up to the amino acids identified in Cornwell *et al.* (Cornwell *et al.*, 2019). More work could be done to validate any structural collapse in the native state here.

It is worth noting that the serum half-life and binding specificity *in vivo* has not been quantified for 114, unlike WFL and STT. The three substitutions, STT, improved the solution properties of

the mAb as well as the biophysical properties (C. L. Dobson *et al.*, 2016). 114's additional mutations F29S and I57N in the CDRs hold the potential to change the binding affinity and specificity of the mAb compared to WFL and STT. Both of these substitutions reduce the apparent hydrophobicity of the positions; it is unknown if this may act to change the functionality of binding, as CDRs are frequently enriched in hydrophobic residues, but their substitution may reduce non-specific interactions (Ausserwöger *et al.*, 2022). Additionally, Ebo *et al.* subjected the variable regions of WFL (as scFv format, scFv-WFL) to directed evolution and identified twelve 'hotspot' residues frequently mutated; these positions 29 and 57 were amongst those mutated. It would be prudent to test 114 for its binding properties *in vivo*, as its potential to reduce global hydrophobicity but retain W and F may improve mAb binding.

Overall, the results in this chapter have successfully highlighted the usefulness of radical oxidation for mapping structural changes at a peptide-resolution level. It has been clearly shown that the analysis can pick up differences between native protein and those which have been stressed by hydrodynamic force. This resolution could be improved using a combination of digestion enzymes, such as chymotrypsin and trypsin, to fragment into shorter peptides. This 'middle ground' analysis of shorter fragment peptide-level analysis over residue-level would allow faster manual analysis times coupled with more resolution to particular areas of sequence.

Chapter 6

Concluding remarks

6 Concluding remarks

The work in my thesis has outlined novel findings relating to the study of aggregation-prone proteins, how flow induces structural perturbations of mAbs, and the use of FPOP to capture this. In this last section, the overall findings of this thesis are discussed in a broader scientific context with regards to the initial aims set out in Section 1.4: The aims of the thesis.

Chapter 3 added a third variant to the WFL/STT story previously published (Cornwell *et al.*, 2019; C. L. Dobson *et al.*, 2016; Ebo, Saunders, *et al.*, 2020). The additional variant, 114, contains W30 and F31, which were previously identified as 'hotspot' positions (Ebo, Saunders, *et al.*, 2020). Similarly, of the four additional mutations in 114 (F29S (V_H), I52N (V_H), S2G (V_L), T77A (V_L)), both of the heavy chain mutations cluster near the hydrophobic patch and were also identified as mutational hotspot positions in WFL_scFv (Ebo, Saunders, *et al.*, 2020). The mutations F29S and I52N both change the properties of the residue's side chains from hydrophobic to polar uncharged, reducing hydrophobicity on the surface. In the TPBLA studies, 114_scFv performed better than STT (mean score of 1180 A.U. compared to 975 A.U.), demonstrating that the drop in hydrophobicity of the mutated residues decreases the scFv's overall aggregation potential. How the mutations in 114 translated to the characteristics of the full mAb format (compared to the scFv) was previously unknown. Therefore, this thesis was critical in increasing our understanding about the applicability of the TPBLA for predicting any mAb aggregation properties from the scFv scaffold.

Chapter 3 confirmed that the biophysical properties of 114 were largely unaffected by the presence of the WFL residues, and that the four additional mutations effectively rescue the biophysical behaviour to be similar to STT. 114 showed similar behaviour as STT through comparable elution times in SEC, HIC and SMAC. Similarly, 114 has a low measure of self-association propensity using AC-SINS, like STT. The work in Chapter 3 also investigated flow-induced insoluble aggregate formation using an extensional flow device (detailed in Section 1.1.4.4: Extensional Flow Device (EFD) to mimic aggregation under flow (J. Dobson *et al.*, 2017)), and demonstrated that the half time of aggregate formation for STT and 114 were similar, whereas WFL had a much quicker half time. Through this technique, 114 was shown to have a more similar loss of monomer from solution under flow stress to STT than WFL. Finally, HDX-MS showed that the variants are globally similar in uptake throughout the time course (30 minutes) but exhibited local dynamic differences in their V_H surface area, with the V_HCDR3 being the most highly dynamic in 114, over STT and WFL. This breadth of biophysical characterisation has

provided a great insight into how few mutations in the variable regions can influence mAb characteristics.

In Chapter 4, a systematic approach to generate a peptide preparation and separation technique suitable for analysing FPOP-modified peptides by LC-MS/MS was employed. STT was used to iteratively improve the process, where several methodologies with differing preparation, digestion, and peptide clean-up steps were explored. The sequence coverage was reliably improved from an un-optimised method which yielded a coverage of HC = 39%, LC = 62%, to HC = 81%, LC = 99%. The differing procedures explored here can provide valuable guidance for others setting up a proteolytic digest-LC-MS/MS method, where the use of a guard column may not be available in the laboratory.

Although the current LC-MS/MS workflow is largely effective, the enormous amount of data generated prevents FPOP-LC-MS/MS from being more widely used for biopharmaceutical structural analysis, and therefore it could be useful to explore alternative experimental strategies (such as a more 'native' approach) to increase the attractiveness of FPOP for industrial application. For example, 'top-down' sequencing of mAbs has progressed massively in recent years, where orbitrap instruments have been capable of using multiple ion activation processes to sequence whole mAbs (Fornelli *et al.*, 2018). This would prevent the need for optimising digestion protocols as the sample is introduced intact and then fragmented in the MS, although it would perhaps sacrifice instrument time due to the need to fragment, and therefore the MS/MS data obtained may not be sufficient for amino-acid resolution analysis. This highlights the applicability of using peptide-level FPOP assignment (mainly used in Chapter 5) as a quicker analysis method coupled to top-down mAb analysis.

Native-MS was used in Chapter 4 to screen for FPOP modification, where the overall shift in mass in the charge state distribution provides a crude but quick confirmation of oxidation compared to a no-laser control. The idea of using the average mass shift to check for successful labelling could be utilised further in the future, for screening dosage in different buffer conditions for different proteins intact, before the investment of time and materials for the full proteolytic-digestion-LC-MS/MS procedure.

In the search for techniques that can identify and characterise unfolding, and indeed flow-induced unfolding, the demand is for a technique that has high sensitivity, requires low sample volumes and concentrations, and can capture any solution-phase changes in structure. FPOP-LC-MS/MS holds this potential. From the results in Chapter 5 it is clear to see that, for the first

time, flow-induced structural changes to mAbs have been captured, and localised to peptide-resolution.

Generally, the highest amount of oxidation was found on the peptides spanning the CDRs, which are known to be highly conformationally flexible loops important for functional recognition, especially the long V_HCDR3 (D'Angelo *et al.*, 2018; North *et al.*, 2011). The biggest differences between the variants, and indeed in response to flow, are in the heavy chain CDRs 1 and 2, which surround the mutations of interest in this thesis.

Additionally, the advancements in FPOP modification detection to amino-acid residue level resolution of mAbs in recent years (Cornwell *et al.*, 2019) could provide an even deeper understanding of flow-induced amino acid side chain rearrangements with further repeats of the data. However, the residue-level FPOP determination process is time-consuming, and requires thorough MS/MS data to localise the modification to an amino acid, and so care would need to be taken to replicate the flow conditions. The flow device was designed using computational fluid dynamic modelling (J. Dobson *et al.*, 2017) to produce defined laminar flow fields, however it is not possible to fully control the exposure of all molecules to the same environment for the duration of the flow experiment. Even so, as this is an ensemble technique, the aggregation studies published previously have demonstrated the reproducibility of the monomer loss exhibited by the molecules that have been exposed to stress (J. Dobson *et al.*, 2017; Willis *et al.*, 2018, 2020). Work is ongoing for modelling the kinetics of the aggregation mechanism triggered by the extensional flow device (Willis et. al., manuscript in preparation). Overall, as it is yet unknown which regions of any of the mAbs studied may be firstly responsible for forming aggregation-prone species, or about the variability of the species formed by flow between different mAb variants, this thesis provides powerful first insights into conformational changes which may be common or unique to mAb variants.

Following on from this concept, an important consideration to be made of the current flow-FPOP-LC-MS/MS process is that FPOP is probing all states present in the heterogeneous mixture of flow-stressed sample. It is known that WFL reversibly self-associates (C. L. Dobson *et al.*, 2016; Willis *et al.*, 2018) and under flow stress in the timescale of the experiment it forms insoluble aggregates (3.4.2: Assessing mechanical stability under hydrodynamic force) (J. Dobson *et al.*, 2017; Willis, 2018; Willis *et al.*, 2018). Indeed, results in Chapter 3 also demonstrate that STT and 114 form insoluble aggregates but to a smaller extent. It is possible that any unfolding states captured in this current FPOP workflow in Chapter 5 could be convoluted with the signal from aggregates, and aggregates that have a distinct FPOP pattern to labelled partial-unfolded

monomer. This means that, because of the mutations between variants, that it is possible the inherent differences in their aggregation potential do not allow for a fair comparison. To account for the presence of aggregate, another stage to the workflow could be implemented, where insoluble aggregates are rapidly clarified from solution after flow and before FPOP (for example through centrifugation or in-line SEC). However, any delay between the two stages (flow and labelling) would allow further time for the progression of the unfolding and aggregation pathway, and therefore the FPOP could be capturing a different footprinting pattern at the later time point. Therefore, there is a key future direction for this thesis, where an in-line flow device or microfluidic device would allow the FPOP to capture the solution after one position of strain exposure (see Section 6.1: Future direction). This would require validation to ensure enough sample was being oxidised, so that the signal was identifiable by LC-MS/MS. Furthermore, time-resolved FPOP at different positions in the flow path may be useful for understanding the evolution of structural changes in the seconds after the flow event, to build up a flow-induced aggregation time line (this idea is explored in Section 6.1: Future direction).

Clearly, the flow-FPOP-LC-MS/MS workflow has advantages beyond the application to industrially relevant IgGs. Indeed, this protocol could be applied for studying any system which might be perturbed by flow, and this method has the potential to enlighten researchers further on the beginnings of unfolding and aggregation triggered by flow force. Overall, this technique could prove highly useful for generating a greater understanding of the mechanisms involved, and ultimately could have great implications for industry as a tool to screen mAbs for their aggregation potential.

6.1 Future direction

6.1.1 Time-resolved FPOP – expanding the capabilities of flow coupled to FPOP-LC-MS/MS

The research in this thesis demonstrates a strong potential to use the technique of FPOP to capture several conformational states along the unfolding and aggregation pathway of mAbs (and other protein models). The sub-millisecond reaction time scales can be utilised to create ‘snapshots’ or fingerprints of protein structure at the point of irradiation. Coupled to flow-induced unfolding, subtle changes which could be captured in a time-resolved manner may hold the potential to unpick structural changes which lead to aggregation, triggered by extensional flow.

The next stages of this project should be to vary the time the FPOP was performed after the flow event. For example, the 'dead time' of the experiments in this chapter (between the final pass of 200 and the beginning of the passage of sample through the FPOP syringe drive and capillary) was 5 minutes. The persistence of any conformational states could be explored through leaving sample for a variety of time points (for example, 20 minutes, 4 hours, 24 hours etc.). This could provide information on any long-lived structures or aggregated sample.

For capturing the very beginning stages of unfolding, the idea of creating a device which facilitated this was extensively discussed and explored during my time at The University of Leeds. The FPOP of samples exposed to flow force would be achievable by creating a smaller version of the flow device, to be situated in line with the laser system, reminiscent of the contraction experienced as a sample is moved from a syringe barrel to the capillary. Therefore, using computational fluid dynamic restraints, a design for this microfluidic device was created by colleagues Dr Ioanna Panagi and Professor Nikil Kapur in the School of Mechanical Engineering. The summary of this is seen in Figure 6.1, where the design features are detailed in Dr Panagi's PhD thesis (currently embargoed until 2024). This work provides a potential avenue to continue along the research into flow-induced protein unfolding investigated using FPOP-LC-MS/MS.

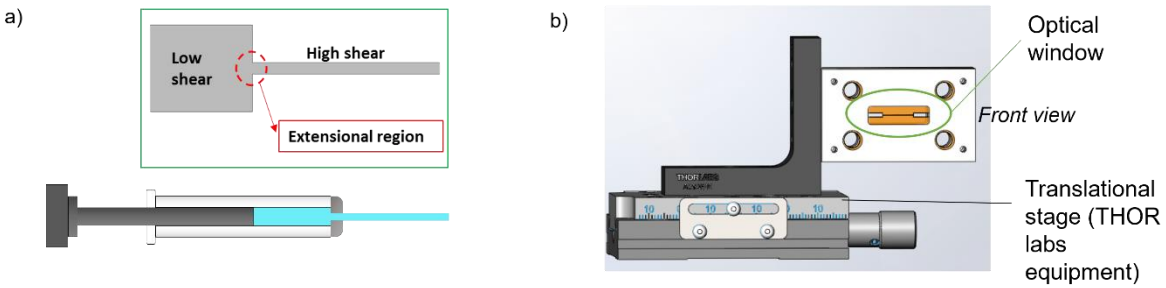


Figure 6.1 Summary of the microfluidic device.

a) The regions of low shear, high shear and the contraction ratio in the syringe was identified for the design of the device. b) An example of how the optical window of the device may be mounted in the path of the FPOP laser. Adapted from Dr Panagi.

Chapter 7

Appendices

7 Appendices

7.1 Related information for Introduction

HEAVY CHAIN

WFL	1	QVQLVQSGAEVKKPGSSVKVSCKASGGTFWFGAFTWVRQAPGQGLEWMGGIIPIFGLTNLAQNFQGRVTITADESTSTVYMELSSLRSEDTAVYYCARSSRIYDLNPSLTAYYDMDVWGQ	120
STT	1	----------ST--T----------	120
114	1	---S----N----------	120

WFL	121	GTMVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPC	240
STT	121	----------	240
114	121	----------	240

WFL	241	PAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVY	360
STT	241	----------	360
114	241	----------	360

WFL	361	TLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK	458
STT	361	----------	458
114	361	----------	458

LIGHT CHAIN

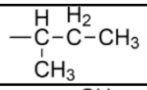
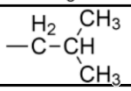
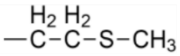
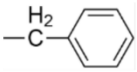
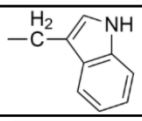
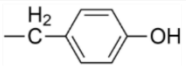
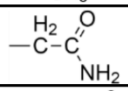
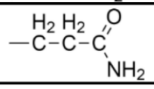
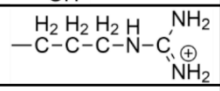
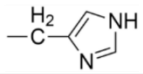
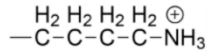
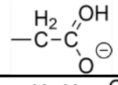
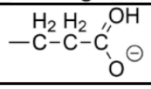
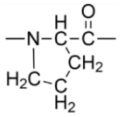
WFL	1	QSVLTQPPSVSAAPGQKVTISCSGSSSDIGNNYVSWYQQLPGTAPKLLIYDNNKRPSGIPDRFSGSKSGTSATLGITGLQTGDEADYYCGTWDSSLSAWVFGGGTKLTVLGQPKAAPSVT	120
STT	1	-----------------------------------	120
114	1	-G----------A----------	120

WFL	121	LFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKSHRSYSCQVTHEGSTVEKTVAPTECS	216
STT	121	----------	216
114	121	----------	216

Figure 7.1 Sequence alignments of the IgG sequences of WFL compared to STT and 114.

Residues that differ from WFL are noted in the alignment. Dash '-' represents conserved residues. CDRs are highlighted in yellow. Blue text = variable regions; black text = C_H1 domain; orange text (heavy chain only) = C_H2 domain; green text (heavy chain only) = C_H3 domain.

Table 7.1 Amino acids summary, including monoisotopic and average mass values for the condensed residue.

Name	Abbreviation	Side chain (R group)	Classification	Monoisotopic mass (Da)	Average mass (Da)
Alanine	Ala, A		Hydrophobic, small	71.04	71.0779
Isoleucine	Ile, I		Hydrophobic	113.08	113.1576
Leucine	Leu, L		Hydrophobic	113.08	113.1576
Methionine	Met, M		Hydrophobic, sulfur-containing	131.04	131.1961
Phenylalanine	Phe, F		Hydrophobic	147.07	147.1739
Tryptophan	Trp, W		Hydrophobic	186.08	186.2099
Tyrosine	Tyr, Y		Hydrophobic	163.06	163.1733
Valine	Val, V		Hydrophobic	99.07	99.1311
Asparagine	Asn, N		Polar uncharged	114.04	114.1026
Glutamine	Gln, Q		Polar uncharged	128.06	128.1292
Serine	Ser, S		Polar uncharged	87.03	87.0773
Threonine	Thr, T		Polar uncharged	101.05	101.1039
Arginine	Arg, R		Positive charge	156.10	156.1857
Histidine	His, H		Positive charge	137.06	137.1393
Lysine	Lys, K		Positive charge	128.09	128.1723
Aspartic acid	Asp, D		Negative charge	115.03	115.0874
Glutamic acid	Glu, E		Negative charge	129.04	129.114
Cysteine	Cys, C		Sulfur-containing	103.01	103.1429
Glycine	Gly, G		Small	57.02	57.0513
Proline	Pro, P		Uncharged	97.05	97.1152

7.2 Related information for Materials and Methods

Table 7.2 Example Tune File for Native MS (STT_control sample in Figure 4.10 a).

====== Ion Source: ======:	
Spray Voltage (V)	1516.3
Spray Current (µA)	0.26
Spray Current std. dev. (µA)	0.04
Capillary Temperature (°C)	249.91
Sheath gas flow rate	0
Aux gas flow rate	0.05
Sweep gas flow rate	0
Aux. Temperature (°C)	0.85
====== Ion Optics: ======:	
Capillary Voltage (V)	-0.3
Bent Flatapole DC (V)	2.2
Inj Flatapole A DC (V)	-150.5
Inj Flatapole B DC (V)	-149.6
Trans Multipole DC (V)	0.1
HCD Multipole DC (V)	-0.9
Inj. Flat. RF Amp (V)	699.2
Inj. Flat. RF Freq (kHz)	540.2
Bent Flat. RF Amp (V)	937.4
Bent Flat. RF Freq (kHz)	1098.87
RF2 and RF3 Amp (V)	898.8
RF2 and RF3 Freq (kHz)	751.68
Inter Flatapole DC (V)	5.71
Quad Exit DC (V)	-19.85
C-Trap Entrance Lens DC (V)	1.04
C-Trap RF Amp (V)	2950.1
C-Trap RF Freq (kHz)	2.163
C-Trap RF Curr (A)	0.53
C-Trap Exit Lens DC (V)	15.07
HCD Exit Lens DC (V)	34.91

====== Vacuum: ======:	
Fore Vacuum Sensor (mbar)	1.6
High Vacuum Sensor (mbar)	3.44E-09
UHV Sensor (mbar)	4.70E-10
Source TMP Speed	1000
UHV TMP Speed	1000
===== Temperatures: =====:	
Analyzer Temperature (°C)	23.59
Ambient Temperature (°C)	24.2
Ambient Humidity (%)	20.5
Source TMP Motor Temperature (°C)	44
Source TMP Bottom Temperature (°C)	36
UHV TMP Motor Temperature (°C)	42
IOS Heatsink Temp. (°C)	24.2
HVPS Peltier Temp. (°C)	39.56
Quad. Det. Temp. (°C)	28.87
==== Diagnostic Data: ====:	
Performance ld	0.24
Performance me	3055.875
Performance cy:	0
CTCD mV	14

7.3 Related information for Chapter 5

Table 7.3 Full details of the peptides searched for FPOP-LC-MS/MS experiments in this thesis.

Peptide	Peptide code (start-first four amino acids-end)	Variant	Position in sequence	*m/z*	Modifications identified
ASGGTF**WF**GAFTWVR	24-ASGG-38	WFL	(24-38) (HC, VH)	845.4148 (2+)	+16, +32
ASGGTF**ST**GAFTWVR		STT	(24-38) (HC, VH)	772.8787 (2+)	+16, +32
ASGGT**SWF**GAFTWVR		114	(24-38) (HC, VH)	815.3951 (2+)	+16, +32
QAPGQGLEWMGGIIPIFG**LT**NLAQNFQGR	39-QAPG-67	WFL	(39-67) (HC, VH)	1038.8732 (2+)	+16, +32
QAPGQGLEWMGGIIPIFG**TT**NLAQNFQGR		STT	(39-67) (HC, VH)	1034.8606 (2+)	+16, +32
QAPGQGLEWMGGI**N**PIFGLTNLAQNFQGR		114	(39-67) (HC, VH)	1039.1932 (2+)	+16, +32
VTITADESTSTVYMELSSLR	68-VTIT-87	ALL	(68-87) (HC, VH)	1102.0419 (2+) 740.3647 (3+)	+16, +32 +16
SEDTAVYYCAR	88-SEDT-98	ALL	(88-98) (HC, VH)	667.7879 (2+)	+16, +32
IYDLNPSLTAYYDMDVWGQGTMVTVSSASTK	102-IYDL-132	ALL	(102-132) (HC, VH)	1139.2048 (2+)	+16, +32
GPSVFPLAPSSK	133-GPSV-144	ALL	(133-144) (HC, CH1)	593.8281 (2+)	+16
STSGGTAALGCLVK	145-STSG-158	ALL	(145-158) (HC, CH1)	661.3435 (2+)	+16, +32
FNWYVDGVEVHNAK	286-FNWY-299	ALL	(286-299) (HC, CH2)	839.4055 (2+) 559.9392 (3+)	+16, +32 +16, +32
VVSVLTVLHQDWLNGK	313-VVSV-328	ALL	(313-328) (HC, CH2)	905.0077 (2+) 603.6737 (3+)	+16, +32 +16, +32
EPQVYTLPPSREEMTK	356-EPQV-371	ALL	(356-371) (HC, CH3)	952.9752 (2+) 635.6528 (3+)	+16 +16
NQVSLTCLVK	372-NQVS-381	ALL	(372-381 (HC, CH3)	581.3185 (2+)	+16, +32
GFYPSDIAVEWESNGQPENNYK	382-GFYP-403	ALL	(382-403) (HC, CH3)	849.0503 (3+)	+16, +32
TTPPVLDSDGSFFLYSK	404-TTPP-420	ALL	(404-420) (HC, CH3)	937.9673 (2+) 625.3145 (3+)	+16, +32 +16
Q**S**VLTQPPSVSAAPGQK	1-QGVL-17	WFL/STT	(1-17) (LC, VL)	847.9578 (2+)	+16, +32
Q**G**VLTQPPSVSAAPGQK		114	(1-17) (LC, VL)	832.9517 (2+)	+16
VTISCSGSSSDIGNNYVSWYQQLPGTAPK	18-VTIS-46	ALL	(18-46) (LC, VL)	1039.8274 (3+)	+16, +32
LLIYDNNKRPSGIPDR	47-LLIY-62	ALL	(47-62) (LC, VL)	624.6772 (3+) 468.5095 (4+)	+16 +16
LTVLGQPK	107-LTVL-114	ALL	(107-114) (LC, VL/CL)	428.2681 (2+)	+16
AAPSVTLFPPSSEELQANK	115-AAPS-133	ALL	(115- 133) (LC, CL)	994.0137 (2+)	+16
ATLVCLISDFYPGAVTVAWK	134-ATVL-153	ALL	(134-153) (LC, CL)	1106.5825 (2+)	+16, +32

				738.0572 (3+)	+16, +32
ADSSPVKAGVETTTPSK	154-ADSS-170	ALL	(154-170) (LC, CL)	837.9313 (2+) 558.9563 (3+)	+16 +16
YAASSYLSLTPEQWK	176-YAAS-190	ALL	(176-190) (LC, CL)	872.4342 (2+) 581.9583 (3+)	+16, +32 +16
SYSCQVTHEGSTVEK	194-SYSC-208	ALL	(194-208) (LC, CL)	856.3837 (2+) 571.2580 (3+)	+16, +32 +16, +32

Table 7.4 Full FPOP percentage modification values for control (native) and stressed FPOP-LC-MS/MS XIC measurements in this thesis.

Peptide code	Variant	Control or stressed	Total +16	Total +32	Overall	Mean overall % oxidation	Range (n=2)
ASGG	STT	C	3.21	4.03	7.23		
ASGG	STT	C	3.08	4.32	7.40	7.32	0.17
ASGG	STT	S	2.97	3.19	6.16		
ASGG	STT	S	2.86	4.10	6.96	6.56	0.80
ASGG	WFL	C	29.37	13.13	42.49		
ASGG	WFL	C	22.94	14.69	37.63	40.06	4.86
ASGG	WFL	S	7.76	7.62	15.38		
ASGG	WFL	S	16.32	9.90	26.23	20.80	10.85
ASGG	114	C	9.58	4.39	13.97		
ASGG	114	C	12.04	6.56	18.59	16.28	4.62
ASGG	114	S	10.02	5.40	15.41		
ASGG	114	S	8.85	5.37	14.22	14.81	1.20
QAPG	STT	C	13.22	4.44	17.67		
QAPG	STT	C	18.42	4.98	23.40	20.53	5.73
QAPG	STT	S	16.26	4.63	20.89		
QAPG	STT	S	16.32	4.56	20.88	20.89	0.01
QAPG	WFL	C	32.57	8.26	40.83		
QAPG	WFL	C	32.06	8.92	40.98	40.91	0.15
QAPG	WFL	S	38.88	11.32	50.20		
QAPG	WFL	S	15.59	4.38	19.97	35.09	30.23
QAPG	114	C	17.70	4.71	22.42		
QAPG	114	C	26.00	6.10	32.10	27.26	9.68
QAPG	114	S	35.02	7.99	43.01		
QAPG	114	S	23.49	7.12	30.62	36.81	12.40
VTIT	STT	C	15.10	0.28	15.38		
VTIT	STT	C	17.42	0.36	17.78	16.58	2.40
VTIT	STT	S	13.85	0.28	14.13		
VTIT	STT	S	18.72	0.35	19.07	16.60	4.94
VTIT	WFL	C	17.73	0.51	18.24		
VTIT	WFL	C	17.54	0.54	18.08	18.16	0.16
VTIT	WFL	S	17.67	0.19	17.85		
VTIT	WFL	S	20.55	0.36	20.91	19.38	3.06
VTIT	114	C	26.91	0.59	27.51		
VTIT	114	C	25.27	0.47	25.74	26.62	1.77
VTIT	114	S	27.60	0.52	28.12		
VTIT	114	S	28.23	0.63	28.86	28.49	0.74
SEDT	STT	C	0.37	0.39	0.76		
SEDT	STT	C	0.32	0.21	0.53	0.65	0.24
SEDT	STT	S	0.26	0.20	0.46		
SEDT	STT	S	0.35	0.36	0.71	0.58	0.25
SEDT	WFL	C	0.31	0.25	0.57		

SEDT	WFL	C	0.34	0.65	0.99	0.78	0.43
SEDT	WFL	S	0.21	0.55	0.76		
SEDT	WFL	S	0.45	0.33	0.78	0.77	0.01
SEDT	114	C	0.39	0.15	0.54		
SEDT	114	C	0.29	0.76	1.05	0.79	0.51
SEDT	114	S	0.34	0.53	0.88		
SEDT	114	S	0.41	0.12	0.53	0.70	0.35
IYDL	STT	C	32.33	13.44	45.77		
IYDL	STT	C	34.43	15.46	49.89	47.83	4.12
IYDL	STT	S	32.61	12.21	44.82		
IYDL	STT	S	30.69	15.83	46.52	45.67	1.70
IYDL	WFL	C	33.76	16.44	50.20		
IYDL	WFL	C	33.01	15.18	48.19	49.19	2.00
IYDL	WFL	S	30.98	28.50	59.47		
IYDL	WFL	S	33.80	18.08	51.88	55.68	7.60
IYDL	114	C	33.05	14.14	47.19		
IYDL	114	C	37.35	23.15	60.50	53.84	13.31
IYDL	114	S	35.78	18.70	54.48		
IYDL	114	S	32.42	17.49	49.91	52.20	4.57
GPSV	STT	C	1.37	0.00	1.37		
GPSV	STT	C	1.39	0.00	1.39	1.38	0.01
GPSV	STT	S	1.33	0.00	1.33		
GPSV	STT	S	1.32	0.00	1.32	1.32	0.01
GPSV	WFL	C	1.47	0.00	1.47		
GPSV	WFL	C	1.30	0.00	1.30	1.38	0.17
GPSV	WFL	S	1.11	0.00	1.11		
GPSV	WFL	S	1.26	0.00	1.26	1.19	0.15
GPSV	114	C	1.37	0.00	1.37		
GPSV	114	C	1.38	0.00	1.38	1.38	0.01
GPSV	114	S	1.45	0.00	1.45		
GPSV	114	S	1.80	0.00	1.80	1.63	0.35
STSG	STT	C	0.48	0.22	0.71		
STSG	STT	C	0.40	0.11	0.52	0.61	0.19
STSG	STT	S	0.40	0.08	0.48		
STSG	STT	S	0.50	0.23	0.72	0.60	0.25
STSG	WFL	C	0.47	0.15	0.63		
STSG	WFL	C	0.50	0.47	0.97	0.80	0.34
STSG	WFL	S	0.62	0.26	0.88		
STSG	WFL	S	0.64	0.15	0.78	0.83	0.10
STSG	114	C	1.63	0.05	1.68		
STSG	114	C	1.15	0.32	1.47	1.57	0.21
STSG	114	S	1.60	0.24	1.84		
STSG	114	S	1.26	0.07	1.34	1.59	0.50
FNWY	STT	C	4.45	2.73	7.18		
FNWY	STT	C	4.23	1.91	6.14	6.66	1.04
FNWY	STT	S	4.52	2.20	6.72		

FNWY	STT	S	5.20	2.26	7.46	7.09	0.74
FNWY	WFL	C	3.79	3.75	7.54		
FNWY	WFL	C	4.43	4.13	8.56	8.05	1.02
FNWY	WFL	S	7.51	7.83	15.33		
FNWY	WFL	S	6.24	6.98	13.21	14.27	2.12
FNWY	114	C	8.29	10.38	18.67		
FNWY	114	C	8.96	7.89	16.85	17.76	1.82
FNWY	114	S	8.91	10.54	19.45		
FNWY	114	S	11.06	12.12	23.18	21.32	3.73
VVSV	STT	C	4.03	10.99	15.03		
VVSV	STT	C	5.00	10.78	15.78	15.40	0.75
VVSV	STT	S	3.62	9.09	12.71		
VVSV	STT	S	3.74	11.86	15.60	14.16	2.89
VVSV	WFL	C	4.23	13.43	17.66		
VVSV	WFL	C	4.23	12.86	17.09	17.38	0.57
VVSV	WFL	S	5.56	16.41	21.97		
VVSV	WFL	S	5.47	12.79	18.26	20.12	3.71
VVSV	114	C	7.08	14.61	21.69		
VVSV	114	C	8.87	13.82	22.69	22.19	1.01
VVSV	114	S	7.49	14.89	22.37		
VVSV	114	S	8.88	15.59	24.47	23.42	2.10
EPQV	STT	C	19.94	0.00	19.94		
EPQV	STT	C	19.23	0.00	19.23	19.58	0.71
EPQV	STT	S	18.84	0.00	18.84		
EPQV	STT	S	20.84	0.00	20.84	19.84	2.00
EPQV	WFL	C	19.93	0.00	19.93		
EPQV	WFL	C	21.43	0.00	21.43	20.68	1.50
EPQV	WFL	S	24.37	0.00	24.37		
EPQV	WFL	S	26.04	0.00	26.04	25.20	1.68
EPQV	114	C	31.27	0.00	31.27		
EPQV	114	C	28.79	0.00	28.79	30.03	2.47
EPQV	114	S	29.45	0.00	29.45		
EPQV	114	S	32.65	0.00	32.65	31.05	3.20
NQVS	STT	C	0.72	0.48	1.21		
NQVS	STT	C	0.64	0.24	0.89	1.05	0.32
NQVS	STT	S	0.63	0.24	0.86		
NQVS	STT	S	0.67	0.34	1.01	0.94	0.14
NQVS	WFL	C	0.73	0.37	1.10		
NQVS	WFL	C	0.71	0.65	1.36	1.23	0.26
NQVS	WFL	S	0.87	0.58	1.44		
NQVS	WFL	S	1.04	0.18	1.21	1.33	0.23
NQVS	114	C	1.83	0.14	1.96		
NQVS	114	C	1.20	0.53	1.73	1.84	0.24
NQVS	114	S	1.53	0.51	2.04		
NQVS	114	S	1.37	0.21	1.57	1.80	0.47
GFYP	STT	C	2.31	2.51	4.81		

GFYP	STT	C	2.68	3.21	5.90	5.35	1.08
GFYP	STT	S	2.46	1.75	4.21		
GFYP	STT	S	2.44	2.05	4.49	4.35	0.28
GFYP	WFL	C	2.76	2.23	4.98		
GFYP	WFL	C	3.22	2.50	5.72	5.35	0.74
GFYP	WFL	S	4.02	3.36	7.39		
GFYP	WFL	S	3.98	3.71	7.68	7.53	0.30
GFYP	114	C	6.84	6.15	13.00		
GFYP	114	C	7.56	4.38	11.94	12.47	1.06
GFYP	114	S	5.74	6.25	11.99		
GFYP	114	S	7.41	7.84	15.26	13.62	3.27
TTPP	STT	C	2.71	0.05	2.76		
TTPP	STT	C	2.84	0.03	2.86	2.81	0.10
TTPP	STT	S	2.54	0.04	2.58		
TTPP	STT	S	2.42	0.02	2.44	2.51	0.14
TTPP	WFL	C	2.78	0.04	2.82		
TTPP	WFL	C	2.83	0.06	2.89	2.86	0.07
TTPP	WFL	S	2.92	0.06	2.99		
TTPP	WFL	S	3.10	0.06	3.16	3.07	0.17
TTPP	114	C	3.80	0.07	3.87		
TTPP	114	C	3.14	0.05	3.19	3.53	0.68
TTPP	114	S	3.74	0.07	3.80		
TTPP	114	S	3.89	0.08	3.97	3.89	0.17
Q(S/G)VL	STT	C	0.55	0.27	0.82		
Q(S/G)VL	STT	C	0.66	0.32	0.98	0.90	0.16
Q(S/G)VL	STT	S	0.62	0.24	0.87		
Q(S/G)VL	STT	S	0.67	0.40	1.08	0.97	0.21
Q(S/G)VL	WFL	C	0.73	0.31	1.04		
Q(S/G)VL	WFL	C	0.53	0.64	1.17	1.11	0.13
Q(S/G)VL	WFL	S	0.47	0.47	0.94		
Q(S/G)VL	WFL	S	0.36	0.16	0.53	0.74	0.41
Q(S/G)VL	114	C	2.04	0.00	2.04		
Q(S/G)VL	114	C	1.90	0.00	1.90	1.97	0.15
Q(S/G)VL	114	S	2.16	0.00	2.16		
Q(S/G)VL	114	S	1.08	0.00	1.08	1.62	1.08
VTIS	STT	C	3.86	2.10	5.96		
VTIS	STT	C	4.15	2.22	6.37	6.16	0.41
VTIS	STT	S	3.28	1.26	4.54		
VTIS	STT	S	3.92	1.91	5.83	5.18	1.30
VTIS	WFL	C	3.71	1.44	5.15		
VTIS	WFL	C	3.87	1.56	5.43	5.29	0.27
VTIS	WFL	S	5.42	2.61	8.03		
VTIS	WFL	S	5.14	2.58	7.73	7.88	0.30
VTIS	114	C	12.21	5.18	17.39		
VTIS	114	C	12.36	3.38	15.74	16.56	1.66
VTIS	114	S	12.83	5.10	17.93		

VTIS	114	S	16.71	6.28	22.99	20.46	5.06
LLIY	STT	C	2.28	0.00	2.28		
LLIY	STT	C	1.79	0.00	1.79	2.04	0.50
LLIY	STT	S	1.98	0.00	1.98		
LLIY	STT	S	1.91	0.00	1.91	1.94	0.07
LLIY	WFL	C	2.04	0.00	2.04		
LLIY	WFL	C	2.12	0.00	2.12	2.08	0.09
LLIY	WFL	S	2.53	0.00	2.53		
LLIY	WFL	S	3.18	0.00	3.18	2.86	0.65
LLIY	114	C	3.51	0.00	3.51		
LLIY	114	C	2.76	0.00	2.76	3.13	0.75
LLIY	114	S	2.87	0.00	2.87		
LLIY	114	S	2.92	0.00	2.92	2.90	0.05
LTVL	STT	C	0.48	0.00	0.48		
LTVL	STT	C	0.39	0.00	0.39	0.44	0.09
LTVL	STT	S	0.42	0.00	0.42		
LTVL	STT	S	0.40	0.00	0.40	0.41	0.02
LTVL	WFL	C	0.42	0.00	0.42		
LTVL	WFL	C	0.40	0.00	0.40	0.41	0.02
LTVL	WFL	S	0.36	0.00	0.36		
LTVL	WFL	S	0.40	0.00	0.40	0.38	0.04
LTVL	114	C	0.37	0.00	0.37		
LTVL	114	C	0.38	0.00	0.38	0.38	0.01
LTVL	114	S	0.39	0.00	0.39		
LTVL	114	S	0.34	0.00	0.34	0.37	0.05
AAPS	STT	C	1.08	0.00	1.08		
AAPS	STT	C	1.08	0.00	1.08	1.08	0.00
AAPS	STT	S	1.08	0.00	1.08		
AAPS	STT	S	1.16	0.00	1.16	1.12	0.08
AAPS	WFL	C	1.27	0.00	1.27		
AAPS	WFL	C	1.18	0.00	1.18	1.22	0.08
AAPS	WFL	S	1.06	0.00	1.06		
AAPS	WFL	S	1.16	0.00	1.16	1.11	0.09
AAPS	114	C	1.23	0.00	1.23		
AAPS	114	C	1.29	0.00	1.29	1.26	0.06
AAPS	114	S	1.26	0.00	1.26		
AAPS	114	S	1.74	0.00	1.74	1.50	0.48
ATLV	STT	C	9.22	4.53	13.75		
ATLV	STT	C	19.00	12.50	31.50	22.62	17.75
ATLV	STT	S	17.23	9.09	26.32		
ATLV	STT	S	11.86	7.74	19.60	22.96	6.72
ATLV	WFL	C	10.06	6.27	16.34		
ATLV	WFL	C	11.08	7.44	18.52	17.43	2.18
ATLV	WFL	S	12.30	7.48	19.78		
ATLV	WFL	S	11.15	6.62	17.77	18.77	2.01
ATLV	114	C	7.69	4.15	11.84		

ATLV	114	C	10.08	4.30	14.39	13.11	2.55
ATLV	114	S	13.39	6.95	20.34		
ATLV	114	S	10.11	5.07	15.18	17.76	5.16
ADSS	STT	C	2.01	0.00	2.01		
ADSS	STT	C	4.34	0.00	4.34	3.18	2.34
ADSS	STT	S	5.61	0.00	5.61		
ADSS	STT	S	3.79	0.00	3.79	4.70	1.82
ADSS	WFL	C	4.49	0.00	4.49		
ADSS	WFL	C	1.48	0.00	1.48	2.99	3.02
ADSS	WFL	S	1.49	0.00	1.49		
ADSS	WFL	S	3.21	0.00	3.21	2.35	1.72
ADSS	114	C	12.89	0.00	12.89		
ADSS	114	C	4.79	0.00	4.79	8.84	8.10
ADSS	114	S	2.16	0.00	2.16		
ADSS	114	S	23.61	0.00	23.61	12.89	21.44
YAAS	STT	C	7.21	0.35	7.56		
YAAS	STT	C	7.11	0.20	7.32	7.44	0.24
YAAS	STT	S	7.23	0.35	7.59		
YAAS	STT	S	6.48	1.95	8.43	8.01	0.84
YAAS	WFL	C	7.15	0.72	7.88		
YAAS	WFL	C	7.71	0.96	8.67	8.27	0.79
YAAS	WFL	S	9.49	1.31	10.80		
YAAS	WFL	S	8.84	0.22	9.06	9.93	1.74
YAAS	114	C	12.72	0.66	13.38		
YAAS	114	C	12.00	0.15	12.15	12.76	1.22
YAAS	114	S	12.59	0.50	13.09		
YAAS	114	S	18.98	0.42	19.41	16.25	6.32
SYSC	STT	C	3.55	1.80	5.34		
SYSC	STT	C	3.27	0.64	3.90	4.62	1.44
SYSC	STT	S	3.55	0.83	4.37		
SYSC	STT	S	3.55	1.18	4.72	4.55	0.35
SYSC	WFL	C	3.08	1.31	4.40		
SYSC	WFL	C	3.65	2.21	5.86	5.13	1.47
SYSC	WFL	S	1.10	1.60	2.70		
SYSC	WFL	S	4.39	0.98	5.36	4.03	2.66
SYSC	114	C	1.36	1.95	3.31		
SYSC	114	C	1.60	1.76	3.36	3.34	0.06
SYSC	114	S	1.19	2.19	3.38		
SYSC	114	S	0.45	2.03	2.48	2.93	0.89

Chapter 8

References

8 References

Abolhasani Khaje, N., Mobley, C. K., Misra, S. K., Miller, L., Li, Z., Nudler, E., & Sharp, J. S. (2018). Variation in FPOP Measurements Is Primarily Caused by Poor Peptide Signal Intensity. *Journal of the American Society for Mass Spectrometry*, *29*(9), 1901–1907. https://doi.org/10.1007/s13361-018-1994-y

Alberts, B., Johnson, A., & Lewis J, *et al.* (2002). *Molecular Biology of the Cell. 4th edition*.

Allison, T. M., Barran, P., Cianférani, S., Degiacomi, M. T., Gabelica, V., Grandori, R., Marklund, E. G., Menneteau, T., Migas, L. G., Politis, A., Sharon, M., Sobott, F., Thalassinos, K., & Benesch, J. L. P. (2020). Computational Strategies and Challenges for Using Native Ion Mobility Mass Spectrometry in Biophysics and Structural Biology. *Analytical Chemistry*, *92*(16), 10872–10880. https://doi.org/10.1021/acs.analchem.9b05791

Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science*, *181*(4096), 223.

Anfinsen, C. B., Haber, E., Sela, M., & White, F. H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, *47*(9), 1309–1314. https://doi.org/10.1073/pnas.47.9.1309

Aponte-Santamaría, C., Huck, V., Posch, S., Bronowska, A. K., Grässle, S., Brehm, M. A., Obser, T., Schneppenheim, R., Hinterdorfer, P., Schneider, S. W., Baldauf, C., & Gräter, F. (2015). Force-sensitive autoinhibition of the von willebrand factor is mediated by interdomain interactions. *Biophysical Journal*, *108*(9), 2312–2321. https://doi.org/10.1016/j.bpj.2015.03.041

Arakawa, T., & Tsumoto, K. (2003). The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. *Biochemical and Biophysical Research Communications*, *304*(1), 148–152. https://doi.org/10.1016/S0006-291X(03)00578-3

Ausserwöger, H., Schneider, M. M., Herling, T. W., Arosio, P., Invernizzi, G., Knowles, T. P. J., & Lorenzen, N. (2022). Non-specificity as the sticky problem in therapeutic antibody development. *Nature Reviews Chemistry*, *6*. https://doi.org/10.1038/s41570-022-00438-x

Baek, Y., & Zydney, A. L. (2018). Intermolecular interactions in highly concentrated formulations of recombinant therapeutic proteins. *Current Opinion in Biotechnology*, *53*, 59–64. https://doi.org/10.1016/j.copbio.2017.12.016

Bailly, M., Mieczkowski, C., Juan, V., Metwally, E., Tomazela, D., & Baker, J. (2020). Predicting Antibody Developability Profiles Through Early Stage Discovery Screening. *MAbs*, *12*(1), 1–28. https://doi.org/10.1080/19420862.2020.1743053

Baynes, B. M., Wang, D. I. C., & Trout, B. L. (2005). Role of arginine in the stabilization of proteins against aggregation. *Biochemistry*, *44*(12), 4919–4925. https://doi.org/10.1021/bi047528r

Beck, A., & Liu, H. (2019). Macro-and micro-heterogeneity of natural and recombinant IgG antibodies. *Antibodies*, *8*(1), 1–22. https://doi.org/10.3390/antib8010018

Bee, J. S., Davis, M., Freund, E., Carpenter, J. F., & Randolph, T. W. (2010). Aggregation of a monoclonal antibody induced by adsorption to stainless steel. *Biotechnology and Bioengineering*, *105*(1), 121–129. https://doi.org/10.1002/bit.22525

Bee, J. S., Randolph, T. W., Carpenter, J. F., Bishop, S. M., & Dimitrova, M. N. (2011). Effects of surfaces and leachables on the stability of biopharmaceuticals. *Journal of Pharmaceutical Sciences*, *100*(10), 4158–4170. https://doi.org/10.1002/jps.22597

Bee, J. S., Stevenson, J. L., Mehta, B., Svitel, J., Pollastrini, J., Platz, R., Freund, E., Carpenter, J. F., & Randolph, T. W. (2009). Response of a concentrated monoclonal antibody formulation to high shear. *Biotechnology and Bioengineering*, *103*(5), 936–943. https://doi.org/10.1002/bit.22336

Benhaim, M., Lee, K. K., & Guttman, M. (2020). Tracking Higher Order Protein Structure by Hydrogen-Deuterium Exchange Mass Spectrometry. *Protein and Peptide Letters*, *26*(1), 16–26. https://doi.org/10.2174/0929866526666181212165037.Tracking

Bergal, H. T., Jiang, Y., Yang, D., Springer, T. A., & Wong, W. P. (2022). Conformation of von Willebrand factor in shear flow revealed with stroboscopic single-molecule imaging. *Blood*, *140*(23), 2490–2499. https://doi.org/10.1182/blood.2022016969

Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G., Teplow, D. B., Shea, J., Brandon, T., Robinson, C. V, & Bowers, M. T. (2010). Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature Chemistry*, *1*(4), 326–331. https://doi.org/10.1038/nchem.247.Amyloid-

Bhambhani, A., Kissmann, J. M., Joshi, S. B., Volkin, D. B., Kashi, R. S., & Middaugh, C. R. (2012). Formulation design and high-throughput excipient selection based on structural integrity and conformational stability of dilute and highly concentrated IgG1 monoclonal antibody solutions. *Journal of Pharmaceutical Sciences*, *101*(3), 1120–1135. https://doi.org/10.1002/jps.23008

Bleakney, W. (1929). A New Method of Positive Ray Analysis and Its Application to the Measurement of Ionization Potentials in Mercury Vapor. *Physical Review*, *34*(1), 157–160.

Bou-Assaf, G. M., Budyak, I. L., Brenowitz, M., Day, E. S., Hayes, D., Hill, J., Majumdar, R., Ringhieri, P., Schuck, P., & Lin, J. C. (2022). Best Practices for Aggregate Quantitation of Antibody Therapeutics by Sedimentation Velocity Analytical Ultracentrifugation. *Journal of Pharmaceutical Sciences*, *111*(7), 2121–2133. https://doi.org/10.1016/j.xphs.2021.12.023

Bracewell, R. N. (1986). *The Fourier transform and its applications*.

Bradbury, A. R. M., Sidhu, S., Dübel, S., & McCafferty, J. (2011). *Beyond natural antibodies: the power of in vitro display technologies*. *29*(3), 245–254. https://doi.org/10.1038/nbt.1791.Beyond

Brader, M. L., Estey, T., Bai, S., Alston, R. W., Lucas, K. K., Lantz, S., Landsman, P., & Maloney, K. M. (2015). Examination of thermal unfolding and aggregation profiles of a series of developable therapeutic monoclonal antibodies. *Molecular Pharmaceutics*, *12*(4), 1005–1017. https://doi.org/10.1021/mp400666b

Brockwell, D. J., & Radford, S. E. (2007). Intermediates: ubiquitous species on folding energy landscapes? *Current Opinion in Structural Biology*, *17*(1), 30–37. https://doi.org/10.1016/j.sbi.2007.01.003

Brückl, L., Schröder, T., Scheler, S., Hahn, R., & Sonderegger, C. (2016). The Effect of Shear on the Structural Conformation of rhGH and IgG1 in Free Solution. *Journal of Pharmaceutical Sciences*, *105*(6), 1810–1818. https://doi.org/10.1016/j.xphs.2016.03.020

Brummitt, R. K., Nesta, D. P., Chang, L., Kroetsch, A. M., & Roberts, C. J. (2011). Nonnative aggregation of an IgG1 antibody in acidic conditions, part 2: Nucleation and growth kinetics with competing growth mechanisms. *Journal of Pharmaceutical Sciences*, *100*(6), 2104–2119. https://doi.org/10.1002/jps.22447

Bunc, M., Hadži, S., Graf, C., Bončina, M., & Lah, J. (2022). Aggregation Time Machine: A Platform for the Prediction and Optimization of Long-Term Antibody Stability Using Short-Term Kinetic Analysis. *Journal of Medicinal Chemistry*, *65*(3), 2623–2632. https://doi.org/10.1021/acs.jmedchem.1c02010

Buss, N. A. P. S., Henderson, S. J., Mcfarlane, M., Shenton, J. M., & Haan, L. De. (2012). Monoclonal antibody therapeutics : history and future. *Current Opinion in Pharmacology*, *12*(5), 615–622. https://doi.org/10.1016/j.coph.2012.08.001

Calabrese, A. N., Ault, J. R., Radford, S. E., & Ashcroft, A. E. (2015). Using hydroxyl radical footprinting to explore the free energy landscape of protein folding. *Methods*, *89*, 38–44. https://doi.org/10.1016/j.ymeth.2015.02.018

Callaway, E. (2020). Revolutionary cryo-EM is taking over structural biology. *Nature*, *578*, 201. https://doi.org/10.1038/d41586-020-00341-9

Chakroun, N., Hilton, D., Ahmad, S. S., Platt, G. W., & Dalby, P. A. (2016). Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment. *Molecular Pharmaceutics*, *13*(2), 307–319. https://doi.org/10.1021/acs.molpharmaceut.5b00387

Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M., & Patrick, R. (2011). Differential hydrogen/deuterium exchange mass spectrometry analysis of protein–ligand interactions. *Expert Reviews of Proteomics*, *8*(1), 43–59. https://doi.org/10.1586/epr.10.109.Differential

Chen, Y., & Barkley, M. D. (1998). Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, *37*(28), 9976–9982. https://doi.org/10.1021/bi980274n

Chennamsetty, N., Voynov, V., Kayser, V., Helk, B., & Trout, B. L. (2009). Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(29), 11937–11942. https://doi.org/10.1073/PNAS.0904191106

Chernushevich, I. V., & Thomson, B. A. (2004). Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Analytical Chemistry*, *76*(6), 1754–1760. https://doi.org/10.1021/ac035406j

Chi, E. Y., Krishnan, S., Randolph, T. W., & Carpenter, J. F. (2003). Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharmaceutical Research*, *20*(9), 1325–1336. https://doi.org/10.1023/A:1025771421906

Chisholm, C. F., Nguyen, B. H., Soucie, K. R., Torres, R. M., Carpenter, J. F., & Randolph, T. W. (2015). In Vivo Analysis of the Potency of Silicone Oil Microdroplets as Immunological Adjuvants in Protein Formulations. *Journal of Pharmaceutical Sciences*, *104*(11), 3681–3690. https://doi.org/10.1002/jps.24573

Chiti, F., & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, *75*, 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901

Chiti, F., & Dobson, C. M. (2017). Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annual Review of Biochemistry*, *86*, 27–68. https://doi.org/10.1146/annurev-biochem-061516-045115

Collings, B. A., & Douglas, D. J. (1997). An extended mass range quadrupole for electrospray mass spectrometry. *International Journal of Mass Spectrometry*, *162*, 121–127.

Conchillo-solé, O., Groot, N. S. De, Avilés, F. X., Vendrell, J., Daura, X., & Ventura, S. (2007). AGGRESCAN : a server for the prediction and evaluation of " hot spots " of aggregation in polypeptides. *BMC Bioinformatics*, *8*(65), 1–17. https://doi.org/10.1186/1471-2105-8-65

Connolly, B. D., Petry, C., Yadav, S., Demeule, B., Ciaccio, N., Moore, J. M. R., Shire, S. J., & Gokarn, Y. R. (2012). Weak interactions govern the viscosity of concentrated antibody solutions: High-throughput analysis using the diffusion interaction parameter. *Biophysical Journal*, *103*(1), 69–78. https://doi.org/10.1016/j.bpj.2012.04.047

Cornwell, O. (2019). *The development of structural mass spectrometry based techniques for the study of aggregation-prone proteins. December*.

Cornwell, O., & Ault, J. R. (2022). Fast photochemical oxidation of proteins coupled with mass spectrometry. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1870*(9), 140829. https://doi.org/10.1016/j.bbapap.2022.140829

Cornwell, O., Ault, J. R., Bond, N. J., Radford, S. E., & Ashcroft, A. E. (2021). Investigation of D76N β2-Microglobulin Using Protein Footprinting and Structural Mass Spectrometry. *Journal of the American Society for Mass Spectrometry*, *32*(7), 1583–1592. https://doi.org/10.1021/jasms.0c00438

Cornwell, O., Bond, N. J., Radford, S. E., & Ashcroft, A. E. (2019). Long-Range Conformational Changes in Monoclonal Antibodies Revealed Using FPOP-LC-MS/MS. *Analytical Chemistry*, *91*(23), 15163–15170. https://doi.org/10.1021/acs.analchem.9b03958

D'Angelo, S., Ferrara, F., Naranjo, L., Erasmus, M. F., Hraber, P., & Bradbury, A. R. M. (2018). Many routes to an antibody heavy-chain CDR3: Necessary, yet insufficient, for specific binding. *Frontiers in Immunology*, *9*, 1–13. https://doi.org/10.3389/fimmu.2018.00395

Demeule, B., Lawrence, M. J., Drake, A. F., Gurny, R., & Arvinte, T. (2007). Characterization of protein aggregation: The case of a therapeutic immunoglobulin. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1774*(1), 146–153. https://doi.org/10.1016/j.bbapap.2006.10.010

Den Engelsman, J., Garidel, P., Smulders, R., Koll, H., Smith, B., Bassarab, S., Seidl, A., Hainzl, O., & Jiskoot, W. (2011). Strategies for the assessment of protein aggregates in

pharmaceutical biotech product development. *Pharmaceutical Research*, *28*(4), 920–933. https://doi.org/10.1007/s11095-010-0297-1

Deslignière, E., Diemer, H., Erb, S., Coliat, P., Pivot, X., Detappe, A., Hernandez-Alba, O., & Cianférani, S. (2022). A Combination of Native LC-MS Approaches for the Comprehensive Characterization of the Antibody-Drug Conjugate Trastuzumab Deruxtecan. *Frontiers in Bioscience - Landmark*, *27*(10), 9–12. https://doi.org/10.31083/j.fbl2710290

Deslignière, E., Ley, M., Bourguet, M., Ehkirch, A., Botzanowski, T., Erb, S., Hernandez-Alba, O., & Cianférani, S. (2021). Pushing the limits of native MS: Online SEC-native MS for structural biology applications. *International Journal of Mass Spectrometry*, *461*, 116502. https://doi.org/10.1016/j.ijms.2020.116502

Devine, P. W. A., Fisher, H. C., Calabrese, A. N., Whelan, F., Higazi, D. R., Potts, J. R., Lowe, D. C., Radford, S. E., & Ashcroft, A. E. (2017). Investigating the Structural Compaction of Biomolecules Upon Transition to the Gas-Phase Using ESI-TWIMS-MS. *Journal of the American Society for Mass Spectrometry*, *28*(9), 1855–1862. https://doi.org/10.1007/s13361-017-1689-9

Dill, K. A., & Chan, H. S. (1997). From Levinthal to Pathways to Funnels: The “New View” of Protein Folding Kinetics. *Nat. Struct. Biol.*, *4*(1), 10–19.

Dill, K. A., Maccallum, J. L., & Folding, P. (2012). The Protein-Folding Problem , 50 Years On. *Science*, *338*, 1042–1047.

Dill, K. A., Ozkan, S., Shell, M., & Weikl, T. (2008). The Protein Folding Problem. *Annual Reviews in Biophysics*, *37*, 289–316. https://doi.org/10.1007/3-540-31618-3_6

Dimasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry : New estimates of R & D costs ☆. *Journal of Health Economics*, *47*, 20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

Dobson, C. L., Devine, P. W. A., Phillips, J. J., Higazi, D. R., Lloyd, C., Popovic, B., Arnold, J., Buchanan, A., Lewis, A., Goodman, J., Van Der Walle, C. F., Thornton, P., Vinall, L., Lowne, D., Aagaard, A., Olsson, L. L., Wollberg, A. R., Welsh, F., Karamanos, T. K., … Lowe, D. C. (2016). Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. *Scientific Reports*, *6*(1), 1–14. https://doi.org/10.1038/srep38644

Dobson, J., Kumar, A., Willis, L. F., Tuma, R., Higazi, D. R., Turner, R., Lowe, D. C., Ashcroft, A. E., Radford, S. E., Kapur, N., & Brockwell, D. J. (2017). Inducing protein aggregation by extensional flow. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(18), 4673–4678. https://doi.org/10.1073/pnas.1702724114

Dole, M., Mack, L. L., & Hines, R. L. (1968). Molecular Beams of Macroions. *Journal of Chemical Physics*, *49*, 2240.

Douglas, D. J., Frank, A. J., & Mao, D. (2005). Linear ion traps in Mass Spectrometry. *Mass Spectrometry Reviews*, *24*, 1–29. https://doi.org/10.1002/mas.20004

Duerkop, M., Berger, E., Dürauer, A., & Jungbauer, A. (2018). Impact of Cavitation, High Shear Stress and Air/Liquid Interfaces on Protein Aggregation. *Biotechnology Journal*, *13*(7). https://doi.org/10.1002/biot.201800062

Ebo, J. S., Guthertz, N., Radford, S. E., & Brockwell, D. J. (2020). Using protein engineering to understand and modulate aggregation. *Current Opinion in Structural Biology*, *60*, 157–166. https://doi.org/10.1016/j.sbi.2020.01.005

Ebo, J. S., Saunders, J. C., Devine, P. W. A., Gordon, A. M., Warwick, A. S., Schiffrin, B., Chin, S. E., England, E., Button, J. D., Lloyd, C., Bond, N. J., Ashcroft, A. E., Radford, S. E., Lowe, D. C., & Brockwell, D. J. (2020). An in vivo platform to select and evolve aggregation-resistant proteins. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-15667-1

Elgundi, Z., Reslan, M., Cruz, E., Sifniotis, V., & Kayser, V. (2017). The state-of-play and future of antibody therapeutics. *Advanced Drug Delivery Reviews*, *122*, 2–19. https://doi.org/10.1016/j.addr.2016.11.004

Englander, W. (2006). Hydrogen Exchange and Mass Spectrometry: A Historical Perspective. *Journal of The American Society for Mass Spectrometry*, *17*(11), 1481–1489. https://doi.org/10.1016/j.jasms.2006.06.006.Hydrogen

Esfandiary, R., Parupudi, A., Casas-Finet, J., Gadre, D., & Sathish, H. (2015). Mechanism of Reversible Self-Association of a Monoclonal Antibody: Role of Electrostatic and Hydrophobic Interactions. *Journal of Pharmaceutical Sciences*, *104*(2), 577–586. https://doi.org/10.1002/jps.24237

Espino, J. A., Mali, V. S., & Jones, L. M. (2015). In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells. *Analytical Chemistry*, *87*(15), 7971–7978. https://doi.org/10.1021/acs.analchem.5b01888

Espino, J. A., Zhang, Z., & Jones, L. M. (2020). Chemical Penetration Enhancers Increase Hydrogen Peroxide Uptake in C. elegans for in Vivo Fast Photochemical Oxidation of Proteins. *Journal of Proteome Research*, *19*(9), 3708–3715. https://doi.org/10.1021/acs.jproteome.0c00245

Faustino Jozala, A., Costa, D., Lacalendola, L., Araújo, V. De, Alexandre, C., Leite, S., Gava, P., Oliveira-nascimento, L. De, Rangel-yagui, C. D. O., Oliveira, P. De, Antonio, M., Oliveira, D., & Pessoa, A. (2016). Biopharmaceuticals from microorganisms : from production to purification. *Brazilian Journal of Microbiology*, *47*, 51–63. https://doi.org/10.1016/j.bjm.2016.10.007

Feige, M. J., Hendershot, L. M., & Buchner, J. (2010). How antibodies fold. *Trends in Biochemical Sciences*, *35*(4), 189–198. https://doi.org/10.1016/j.tibs.2009.11.005.How

Fekete, S., Guillarme, D., Sandra, P., & Sandra, K. (2016). Chromatographic, Electrophoretic, and Mass Spectrometric Methods for the Analytical Characterization of Protein Biopharmaceuticals. *Analytical Chemistry*, *88*(1), 480–507. https://doi.org/10.1021/acs.analchem.5b04561

Fekete, S., Veuthey, J. L., Beck, A., & Guillarme, D. (2016). Hydrophobic interaction chromatography for the characterization of monoclonal antibodies and related products. *Journal of Pharmaceutical and Biomedical Analysis*, *130*, 3–18. https://doi.org/10.1016/j.jpba.2016.04.004

Fenn, J. B., Mann, M., Meng, C. K. A. I., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, *246*(6), 64–71.

Filipe, V., Hawe, A., & Jiskoot, W. (2010). Critical evaluation of nanoparticle tracking analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. *Pharmaceutical Research*, *27*(5), 796–810. https://doi.org/10.1007/s11095-010-0073-2

Fornelli, L., Srzentić, K., Huguet, R., Mullen, C., Sharma, S., Zabrouskov, V., Fellers, R. T., Durbin, K. R., Compton, P. D., & Kelleher, N. L. (2018). Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Analytical Chemistry*, *90*(14), 8421–8429. https://doi.org/10.1021/acs.analchem.8b00984

Gabrielson, J., Arthur, K., Stoner, M., Winn, B., Kendrick, B., Razinkov, V., Svitel, J., Jiang, Y., Voelker, P., & Fernandes, CA Ridgeway, R. (2010). Precision of protein aggregation measurements by sedimentation velocity analytical ultracentrifugation in biopharmaceutical applications. *Analytical Biochemstry*, *396*(2), 231–241.

Garber, E., & Demarest, S. J. (2007). A broad range of Fab stabilities within a host of therapeutic IgGs. *Biochemical and Biophysical Research Communications*, *355*(3), 751–757. https://doi.org/10.1016/j.bbrc.2007.02.042

Geng, S. B., Cheung, J. K., Narasimhan, C., Shameem, M., & Tessier, P. M. (2014). Improving monoclonal antibody selection and engineering using measurements of colloidal protein interactions. *Journal of Pharmaceutical Sciences*, *103*(11), 3356–3363. https://doi.org/10.1002/jps.24130

Ghazvini, S., Kalonia, C., Volkin, D. B., & Dhar, P. (2016). Evaluating the Role of the Air-Solution Interface on the Mechanism of Subvisible Particle Formation Caused by Mechanical Agitation for an IgG1 mAb. *Journal of Pharmaceutical Sciences*, *105*(5), 1643–1656. https://doi.org/10.1016/j.xphs.2016.02.027

Gordon, K. B., Rosenbaum, J. T., Arikan, D., Lau, W. L., Li, P., & Faccin, F. (2019). Long-Term Safety of Adalimumab in 29,967 Adult Patients From Global Clinical Trials Across Multiple Indications: An Updated Analysis. *Advances in Therapy*, *37*, 364–380. https://doi.org/10.1007/s12325-019-01145-8

Graf, T., Heinrich, K., Grunert, I., Wegele, H., Haberger, M., Bulau, P., & Leiss, M. (2020). Recent advances in LC–MS based characterization of protein-based bio-therapeutics – mastering analytical challenges posed by the increasing format complexity. *Journal of Pharmaceutical and Biomedical Analysis*, *186*, 113251. https://doi.org/10.1016/j.jpba.2020.113251

Grassi, L., & Cabrele, C. (2019). Susceptibility of protein therapeutics to spontaneous chemical modifications by oxidation, cyclization, and elimination reactions. *Amino Acids*, *51*(10–12), 1409–1431. https://doi.org/10.1007/s00726-019-02787-2

Grigolato, F., & Arosio, P. (2020). Synergistic effects of flow and interfaces on antibody aggregation. *Biotechnology and Bioengineering*, *117*(2), 417–428. https://doi.org/10.1002/bit.27212

Guilhaus, M., Selby, D., & Mlynski, V. (2000). Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrometry Reviews*, *19*, 65–107.

Haberger, M., Leiss, M., Heidenreich, A. K., Pester, O., Hafenmair, G., Hook, M., Bonnington, L., Wegele, H., Haindl, M., Reusch, D., & Bulau, P. (2016). Rapid characterization of

biotherapeutic proteins by size-exclusion chromatography coupled to native mass spectrometry. *MAbs*, *8*(2), 331–339. https://doi.org/10.1080/19420862.2015.1122150

Hambly, D. M., & Gross, M. L. (2005). Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *Journal of the American Society for Mass Spectrometry*, *16*(12), 2057–2063. https://doi.org/10.1016/j.jasms.2005.09.008

Hoffmann, E., & Stroobant, V. (2007). *Mass Spectrometry: Principles and Applications, 3rd Edition*.

Hong, P., Koza, S., & Bouvier, E. S. P. (2012). Size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *Journal of Liquid Chromatography and Related Technologies*, *35*(20), 2923–2950. https://doi.org/10.1080/10826076.2012.743724

Houde, D., Berkowitz, S. A., & Engen, J. R. (2012). The Utility of Hydrogen/Deuterium Exchange Mass Spectrometry in Biopharmaceutical Comparability Studies. *Journal of Pharmaceutical Sciences*, *100*(6), 2071–2086. https://doi.org/10.1002/jps.22432.The

Inoue, N., Takai, E., Arakawa, T., & Shiraki, K. (2014). Arginine and lysine reduce the high viscosity of serum albumin solutions for pharmaceutical injection. *Journal of Bioscience and Bioengineering*, *117*(5), 539–543. https://doi.org/10.1016/j.jbiosc.2013.10.016

Ionescu, R. M., Vlasak, J., Price, C., & Kirchmeier, M. (2008). Contribution of variable domains to the stability of humanized IgG1 monoclonal antibodies. *Journal of Pharmaceutical Sciences*, *97*(4), 1414–1426. https://doi.org/10.1002/jps.21104

Iribarne, J. V., & Thomson, B. A. (1976). On the evaporation of small ions from charged droplets. *Journal of Chemical Physics*, *64*, 2287.

Jahn, T. R., & Radford, S. E. (2005). The Yin and Yang of protein folding. *FEBS Journal*, *272*(23), 5962–5970. https://doi.org/10.1111/j.1742-4658.2005.05021.x

Jahn, T. R., & Radford, S. E. (2008). Folding versus aggregation: Polypeptide conformations on competing pathways. *Archives of Biochemistry and Biophysics*, *469*(1), 100–117. https://doi.org/10.1016/j.abb.2007.05.015

Jain, T., Sun, T., Durand, S., Hall, A., Houston, N. R., Nett, J. H., Sharkey, B., Bobrowicz, B., Caffry, I., Yu, Y., Cao, Y., Lynaugh, H., Brown, M., Baruah, H., Gray, L. T., Krauland, E. M., Xu, Y., Vásquez, M., & Wittrup, K. D. (2017). Biophysical properties of the clinical-stage antibody landscape. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(5), 944–949. https://doi.org/10.1073/PNAS.1616408114/SUPPL_FILE/PNAS.1616408114.SD03.XLSX

Jefferis, R. (2016). Posttranslational modifications and the immunogenicity of biotherapeutics. *Journal of Immunology Research*, *2016*. https://doi.org/10.1155/2016/5358272

Jennings, K. R. (1968). Collision-induced decomposition of aromatic molecular ions. *International Journal of Mass Spectrometry and Ion Physics*, *1*, 227–235.

Jiskoot, W., Randolph, T. W., Volkin, D. B., Middaugh, C. R., Schöneich, C., Winter, G., Friess, W., Crommelin, D. J. A., & Carpenter, J. F. (2012). Protein instability and immunogenicity: Roadblocks to clinical application of injectable protein delivery systems for sustained

release. *Journal of Pharmaceutical Sciences*, *101*(3), 946–954. https://doi.org/10.1002/jps.23018

Johnson, D. T., Di Stefano, L. H., & Jones, L. M. (2019). Fast photochemical oxidation of proteins (FPOP): A powerful mass spectrometry–based structural proteomics tool. *Journal of Biological Chemistry*, *294*(32), 11969–11979. https://doi.org/10.1074/jbc.REV119.006218

Joubert, M. K., Luo, Q., Nashed-Samuel, Y., Wypych, J., & Narhi, L. O. (2011). Classification and characterization of therapeutic antibody aggregates. *Journal of Biological Chemistry*, *286*(28), 25118–25133. https://doi.org/10.1074/jbc.M110.160457

Kalonia, C. K., Heinrich, F., Curtis, J. E., Raman, S., Miller, M. A., & Hudson, S. D. (2018). Protein Adsorption and Layer Formation at the Stainless Steel-Solution Interface Mediates Shear-Induced Particle Formation for an IgG1 Monoclonal Antibody. *Molecular Pharmaceutics*, *15*(3), 1319–1331. https://doi.org/10.1021/acs.molpharmaceut.7b01127

Kan, Z., Walters, B. T., Mayne, L., & Englander, S. W. (2013). *Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis*. *110*(41), 16438–16443. https://doi.org/10.1073/pnas.1315532110

Kaplon, H., Chenoweth, A., Crescioli, S., & Reichert, J. M. (2022). Antibodies to watch in 2022. *MAbs*, *14*(1), 1–45. https://doi.org/10.1080/19420862.2021.2014296

Khristenko, N., Rosu, F., Largy, E., Haustant, J., Mesmin, C., & Gabelica, V. (2023). Native Electrospray Ionization of Multi-Domain Proteins via a Bead Ejection Mechanism. *Journal of the American Chemical Society*, *145*(1), 498–506. https://doi.org/10.1021/jacs.2c10762

Kim, J. Y., Kim, Y., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins : current state and further potential. *Applied Microbiology and Biotechnology*, *93*, 917–930. https://doi.org/10.1007/s00253-011-3758-5

Kim, N. A., Hada, S., Thapa, R., & Jeong, S. H. (2016). Arginine as a protein stabilizer and destabilizer in liquid formulations. *International Journal of Pharmaceutics*, *513*(1–2), 26–37. https://doi.org/10.1016/j.ijpharm.2016.09.003

Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., & Ulrich Hartl, F. (2013). Molecular chaperone functions in protein folding and proteostasis. *Annual Review of Biochemistry*, *82*, 323–355. https://doi.org/10.1146/annurev-biochem-060208-092442

Knight, M. J., Floret, L., Patel, N., Hara, J. O., Knight, M. J., Floret, L., Patel, N., Hara, J. O., & Rodriguez, E. (2022). The impact of forced degradation conditions on mAb dimer formation and subsequent influence on aggregation propensity. *MAbs*, *14*(1), e2127172 (16 pages). https://doi.org/10.1080/19420862.2022.2127172

Knowles, T. P. J., Vendruscolo, M., & Dobson, C. M. (2014). The amyloid state and its association with protein misfolding diseases. *Nature Reviews Molecular Cell Biology*, *15*(6), 384–396. https://doi.org/10.1038/nrm3810

Köhler, G., & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, *256*, 495–497.

Kohli, N., & Geddie, M. L. (2017). Novel HPLC-based screening method to assess developability of antibody-like molecules. *Methods in Molecular Biology*, *1575*(4), 189–196. https://doi.org/10.1007/978-1-4939-6857-2_11

Konermann, L., Ahadi, E., Rodriguez, A. D., & Vahidi, S. (2013). Unraveling the Mechanism of Electrospray Ionization. *Analytical Chemistry*, *85*(1), 2–9.

Konermann, L., Pan, J., & Liu, Y. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews*, *40*(3), 1224–1234. https://doi.org/10.1039/c0cs00113a

Konermann, L., Stocks, B. B., & Czarny, T. (2010). Laminar flow effects during laser-induced oxidative labeling for protein structural studies by mass spectrometry. *Analytical Chemistry*, *82*(15), 6667–6674. https://doi.org/10.1021/ac101326f

Kuriata, A., Iglesias, V., Pujols, J., Kurcinski, M., Kmiecik, S., & Ventura, S. (2019). Aggrescan3D (A3D) 2.0: Prediction and engineering of protein solubility. *Nucleic Acids Research*, *47*(W1), W300–W307. https://doi.org/10.1093/nar/gkz321

Leavy, O. (2010). Therapeutic antibodies: past, present and future. *Nature Publishing Group*, *10*, 2763. https://doi.org/10.1038/nri2763

Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique*, *65*, 44–45. https://doi.org/10.1051/jcp/1968650044

Li, J., Wei, H., Krystek, S. R., Bond, D., Brender, T. M., Cohen, D., Feiner, J., Hamacher, N., Harshman, J., Huang, R. Y., Julien, S. H., Lin, Z., Moore, K., Mueller, L., Noriega, C., Sejwal, P., Sheppard, P., Stevens, B., Chen, G., … Schneeweis, L. A. (2017). Mapping the Energetic Epitope of an Antibody/Interleukin-23 Interaction with Hydrogen/Deuterium Exchange, Fast Photochemical Oxidation of Proteins Mass Spectrometry, and Alanine Shave Mutagenesis. *Analytical Chemistry*, *89*, 2250–2258. https://doi.org/10.1021/acs.analchem.6b03058

Li, K. S., Chen, G., Mo, J., Huang, R. Y. C., Deyanova, E. G., Beno, B. R., O'Neil, S. R., Tymiak, A. A., & Gross, M. L. (2017). Orthogonal Mass Spectrometry-Based Footprinting for Epitope Mapping and Structural Characterization: The IL-6 Receptor upon Binding of Protein Therapeutics. *Analytical Chemistry*, *89*(14), 7742–7749. https://doi.org/10.1021/acs.analchem.7b01748

Li, K. S., Rempel, D. L., & Gross, M. L. (2016). Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1-42 Aggregation. *Journal of the American Chemical Society*, *138*(37), 12090–12098. https://doi.org/10.1021/jacs.6b07543

Li, Y., & Roberts, C. J. (2009). Lumry-eyring nucleated-polymerization model of protein aggregation kinetics. 2. Competing growth via condensation and chain polymerization. *Journal of Physical Chemistry B*, *113*(19), 7020–7032. https://doi.org/10.1021/jp8083088

Limpikirati, P., Pan, X., & Vachet, R. W. (2020). Covalent Labeling with Diethylpyrocarbonate is Sensitive to Residue Microenvironment, Providing Improved Analysis of Protein Higher Order Structure by Mass Spectrometry. *Analytical Chemistry*, *91*(13), 8516–8523. https://doi.org/10.1021/acs.analchem.9b01732.Covalent

Liu, H. F., Ma, J., Winter, C., & Bayer, R. (2010). Recovery and purification process development for monoclonal antibody production. *MAbs*, *2*(5), 480–499. https://doi.org/10.4161/mabs.2.5.12645

Liu, H., & May, K. (2012). Disulfide bond structures of IgG molecules: Structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs*, *4*(1), 17–23.

Liu, X. R., Rempel, D. L., & Gross, M. L. (2020). Protein higher-order-structure determination by fast photochemical oxidation of proteins and mass spectrometry analysis. *Nature Protocols*, *15*(12), 3942–3970. https://doi.org/10.1038/s41596-020-0396-3

Loo, J. A. (2000). Electrospray ionization mass spectrometry : a technology for studying noncovalent macromolecular complexes. *International Journal of Mass Spectrometry*, *200*, 175–186.

Lu, R., Hwang, Y., Liu, I., Lee, C., Tsai, H., Li, H., & Wu, H. (2020). Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science*, 1–30.

Lumry, R., & Eyeing, H. (1958). Conformation changes of proteins. *American Chemical Society*, *58*, 110–120.

Mahler, H. C., Friess, W., Grauschopf, U., & Kiese, S. (2009). Protein aggregation: Pathways, induction factors and analysis. *Journal of Pharmaceutical Sciences*, *98*(9), 2909–2934. https://doi.org/10.1002/jps.21566

Majumdar, R., Esfandiary, R., Bishop, S. M., Samra, H. S., Middaugh, C. R., Volkin, D. B., & Weis, D. D. (2015). Correlations between changes in conformational dynamics and physical stability in a mutant IgG1 mAb engineered for extended serum half-life. *MAbs*, *7*(1), 84–95. https://doi.org/10.4161/19420862.2014.985494

Makarov, A. (2000). Electrostatic Axially Harmonic Orbital Trapping : A High-Performance Technique of Mass Analysis. *Analytical Chemistry*, *72*(6), 1156–1162.

Mamyrin, B. A., Karataev, V. I., Shmikk, D. V, & Zagulin, V. A. (1973). The mass-reflectron A new nonmagnetic time-of-flight high resolution mass-spectrometer. *Zhurnal Ehksperimental'noj i Teoreticheskoj Fiziki*, *64*(1), 82–89.

Manzi, L., Barrow, A. S., Scott, D., Layfield, R., Wright, T. G., Moses, J. E., & Oldham, N. J. (2016). Carbene footprinting accurately maps binding sites in protein-ligand and protein-protein interactions. *Nature Communications*, *7*, 1–9. https://doi.org/10.1038/ncomms13288

Marichal-Gallardo, P. A., & Álvarez, M. M. (2012). State-of-the-art in downstream processing of monoclonal antibodies: Process trends in design and validation. *Biotechnology Progress*, *28*(4), 899–916. https://doi.org/10.1002/btpr.1567

Marklund, E. G., Degiacomi, M. T., Robinson, C. V., Baldwin, A. J., & Benesch, J. L. P. (2015). Collision cross sections for structural proteomics. *Structure*, *23*(4), 791–799. https://doi.org/10.1016/j.str.2015.02.010

Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K. A., Benesch, J. L. P., & Robinson, C. V. (2015). Bayesian deconvolution of mass and ion mobility spectra: From binary

interactions to polydisperse ensembles. *Analytical Chemistry*, *87*(8), 4370–4376. https://doi.org/10.1021/acs.analchem.5b00140

Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, *62*(6), 670–684. https://doi.org/10.1007/s00018-004-4464-6

Mazzer, A. R., Perraud, X., Halley, J., O'Hara, J., & Bracewell, D. G. (2015). Protein A chromatography increases monoclonal antibody aggregation rate during subsequent low pH virus inactivation hold. *Journal of Chromatography A*, *1415*, 83–90. https://doi.org/10.1016/j.chroma.2015.08.068

Menzen, T., & Friess, W. (2013). High-throughput melting-temperature analysis of a monoclonal antibody by differential scanning fluorimetry in the presence of surfactants. *Journal of Pharmaceutical Sciences*, *102*(2), 415–428. https://doi.org/10.1002/jps.23405

Meric, G., Robinson, A. S., & Roberts, C. J. (2017). Driving forces for nonnative protein aggregation and approaches to predict aggregation-prone regions. *Annual Review of Chemical and Biomolecular Engineering*, *8*, 139–159. https://doi.org/10.1146/annurev-chembioeng-060816-101404

Minton, A. P. (2016). Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. In *Analytical Biochemistry* (Vol. 501, pp. 4–22). Academic Press Inc. https://doi.org/10.1016/j.ab.2016.02.007

Mitragotri, S., Burke, P. A., & Langer, R. (2014). Overcoming the challenges in administering biopharmaceuticals: Formulation and delivery strategies. *Nature Reviews Drug Discovery*, *13*(9), 655–672. https://doi.org/10.1038/nrd4363

Molnár, I., & Horváth, C. (1976). Reverse-phase chromatography of polar biological substances: separation of catechol compounds by high-performance liquid chromatography. *Clinical Chemistry*, *22*(9), 1497-502.

Mouchahoir, T., & Schiel, J. E. (2018). Development of an LC-MS/MS peptide mapping protocol for the NISTmAb. *Analytical and Bioanalytical Chemistry*, *410*(8), 2111–2126. https://doi.org/10.1007/s00216-018-0848-6

Mullard, A. (2022). 2021 FDA approvals. *Nature Reviews. Drug Discovery*, *21*(2), 83–88. https://doi.org/10.1038/d41573-022-00001-9

Mullard, A. (2023). 2022 FDA approvals. *Nature Reviews Drug Discovery News*.

Murphy, C., & Kennedy, R. O. (2016). Technology advancements in antibody purification. *Antibody Technology Journal*, *2016*(6), 17–32.

North, B., Lehmann, A., & Dunbrack, R. L. (2011). A new clustering of antibody CDR loop conformations. *Journal of Molecular Biology*, *406*(2), 228–256. https://doi.org/10.1016/j.jmb.2010.10.030

Oliva, A., Llabrés, M., & Fariña, J. B. (2015). Fitting bevacizumab aggregation kinetic data with the Finke-Watzky two-step model: Effect of thermal and mechanical stress. *European Journal of Pharmaceutical Sciences*, *77*, 170–179. https://doi.org/10.1016/j.ejps.2015.06.011

Olsen, J. V, Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., & Horning, S. (2009). A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular and Cellular Proteomics*, *8*, 2759–2769. https://doi.org/10.1074/mcp.M900375-MCP200

Pacholarz, K. J., Peters, S. J., Garlish, R. A., Henry, A. J., Taylor, R. J., Humphreys, D. P., & Barran, P. E. (2016). Molecular Insights into the Thermal Stability of mAbs with Variable-Temperature Ion-Mobility Mass Spectrometry. *ChemBioChem*, *17*(1), 46–51. https://doi.org/10.1002/cbic.201500574

Pan, J., Zhang, S., Chou, A., Hardie, D. B., & Borchers, C. H. (2015). Fast Comparative Structural Characterization of Intact Therapeutic Antibodies Using Hydrogen − Deuterium Exchange and Electron Transfer Dissociation. *Analytical Chemistry*, *87*, 5884−5890. https://doi.org/10.1021/ac504809r

Pukala, T. L. (2023). Mass spectrometric insights into protein aggregation. *Essays in Biochemistry*, *EBC2022010*, 1–11.

Ratanji, K. D., Derrick, J. P., Dearman, R. J., & Kimber, I. (2014). Immunogenicity of therapeutic proteins: Influence of aggregation. *Journal of Immunotoxicology*, *11*(2), 99–109. https://doi.org/10.3109/1547691X.2013.821564

Rathore, N., & Rajan, R. S. (2008). Current perspectives on stability of protein drug products during formulation, fill and finish operations. *Biotechnology Progress*, *24*(3), 504–514. https://doi.org/10.1021/bp070462h

Raybould, M. I. J., Marks, C., Krawczyk, K., Taddese, B., Nowak, J., Lewis, A. P., Bujotzek, A., Shi, J., & Deane, C. M. (2019). Five computational developability guidelines for therapeutic antibody profiling. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 4025–4030. https://doi.org/10.1073/pnas.1810576116

Rayleigh, F. R. S. (1882). On the equilibrium of liquid conducting masses charged with electricity. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *14*(87), 184–186.

Razinkov, V. I., Treuheit, M. J., & Becker, G. W. (2013). Methods of high throughput biophysical characterization in biopharmaceutical development. *Current Drug Discovery Technologies*, *10*(1), 59–70.

Roepstorff, P., & Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry*, *11*(11). https://doi.org/doi: 10.1002/bms.1200111109

Roesch, A., Zölls, S., Stadler, D., Helbig, C., Wuchner, K., Kersten, G., Hawe, A., Jiskoot, W., & Menzen, T. (2022). Particles in Biopharmaceutical Formulations, Part 2: An Update on Analytical Techniques and Applications for Therapeutic Proteins, Viruses, Vaccines and Cells. *Journal of Pharmaceutical Sciences*, *111*(4), 933–950. https://doi.org/10.1016/j.xphs.2021.12.011

Saunders, J. C., Young, L. M., Mahood, R. A., Jackson, M. P., Revill, C. H., Foster, R. J., Smith, D. A., Ashcroft, A. E., Brockwell, D. J., & Radford, S. E. (2016). An in vivo platform for

identifying inhibitors of protein aggregation. *Nature Chemical Biology*, *12*(2), 94. https://doi.org/10.1038/NCHEMBIO.1988

Schmidt, E. M., Pudenzi, M. A., Santos, J. M., Angolini, C. F. F., Pereira, R. C. L., Rocha, Y. S., Denisov, E., Damoc, E., Makarov, A., & Eberlin, M. N. (2018). Petroleomics via Orbitrap mass spectrometry with resolving power above 1 000 000 at m/z 200. *RSC Advances*, *8*, 6183–6191. https://doi.org/10.1039/c7ra12509g

Schroeder, H. W., & Cavacini, L. (2010). Structure and Function of Immunoglobulins. *Journal of Allergy and Clinical Immunology*, *125*(202), S41–S52. https://doi.org/10.1016/j.jaci.2009.09.046

Sela-Culang, I., Alon, S., & Ofran, Y. (2012). A Systematic Comparison of Free and Bound Antibodies Reveals Binding-Related Conformational Changes. *The Journal of Immunology*, *189*(10), 4890–4899. https://doi.org/10.4049/jimmunol.1201493

Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas', A. F., & Gilmanshin, R. I. (1991). Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*, *31*(1), 119–128. https://doi.org/10.1002/bip.360310111

Shi, L., Liu, T., Gross, M. L., & Huang, Y. (2019). Recognition of Human IgG1 by Fcγ Receptors: Structural Insights from Hydrogen-Deuterium Exchange and Fast Photochemical Oxidation of Proteins Coupled with Mass Spectrometry. *Biochemistry*, *58*(8), 1074–1080. https://doi.org/10.1021/acs.biochem.8b01048

Shi, S., Semple, A., Cheung, J., & Shameem, M. (2013). DSF method optimization and its application in predicting protein thermal aggregation kinetics. *Journal of Pharmaceutical Sciences*, *102*(8), 2471–2483. https://doi.org/10.1002/jps.23633

Shukla, A. A., & Thömmes, J. (2010). Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends in Biotechnology*, *28*, 253–261. https://doi.org/10.1016/j.tibtech.2010.02.001

Simon, S., Krause, H. J., Weber, C., & Peukert, W. (2011). Physical degradation of proteins in well-defined fluid flows studied within a four-roll apparatus. *Biotechnology and Bioengineering*, *108*(12), 2914–2922. https://doi.org/10.1002/bit.23257

Snijder, J. (2013). Studying 18 Mega Dalton Virus Assemblies with Native Mass Spectrometry. *Angewandte Chemie (International Edition in English)*, *52*(14), 4020–4023. https://doi.org/10.1002/anie.201210197.Studying

Sormanni, P., Amery, L., Ekizoglou, S., Vendruscolo, M., & Popovic, B. (2017). Rapid and accurate in silico solubility screening of a monoclonal antibody library. *Scientific Reports*, *7*(8200), 1–9. https://doi.org/10.1038/s41598-017-07800-w

Sule, S. V., Dickinson, C. D., Lu, J., Chow, C. K., & Tessier, P. M. (2013). Rapid analysis of antibody self-association in complex mixtures using immunogold conjugates. *Molecular Pharmaceutics*, *10*(4), 1322–1331. https://doi.org/10.1021/mp300524x

Svilenov, H., & Winter, G. (2019). The ReFOLD assay for protein formulation studies and prediction of protein aggregation during long-term storage. *European Journal of*

Pharmaceutics and Biopharmaceutics, *137*, 131–139. https://doi.org/10.1016/j.ejpb.2019.02.018

Taylor, G. (1964). Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London Series A*, 383–397.

Telikepalli, S. N., Kumru, O. S., Kalonia, C., Esfandiary, R., Joshi, S. B., Middaugh, C. R., & Volkin, D. B. (2014). Structural characterization of IgG1 mAb aggregates and particles generated under various stress conditions. *Journal of Pharmaceutical Sciences*, *103*(3), 796–809. https://doi.org/10.1002/jps.23839

Terral, G., Beck, A., & Cianférani, S. (2016). Insights from native mass spectrometry and ion mobility-mass spectrometry for antibody and antibody-based product characterization. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *1032*, 79–90. https://doi.org/10.1016/j.jchromb.2016.03.044

Thomas, C. R., & Geer, D. (2011). Effects of shear on proteins in solution. *Biotechnology Letters*, *33*(3), 443–456. https://doi.org/10.1007/s10529-010-0469-4

Thomas, P., & Smart, T. G. (2005). HEK293 cell line : A vehicle for the expression of recombinant proteins. *Journal of Pharmacological and Toxicological Methods*, *51*, 187–200. https://doi.org/10.1016/j.vascn.2004.08.014

Thomson, J. J. (1913). Bakerian Lecture:—Rays of positive electricity. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, *89*, 1–20.

Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology*, *7*(420). https://doi.org/10.3389/fbioe.2019.00420

Ulrich Hartl, F., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, *475*(7356), 324–332. https://doi.org/10.1038/nature10317

Vallejo, D. D., Jeon, C. K., Parson, K. F., Herderschee, H. R., Eschweiler, J. D., Filoti, D. I., & Ruotolo, B. T. (2022). Ion Mobility-Mass Spectrometry Reveals the Structures and Stabilities of Biotherapeutic Antibody Aggregates. *Analytical Chemistry*, *94*, 6745–6753. https://doi.org/10.1021/acs.analchem.2c00160

Vallejo, D. D., Rojas, C., Parson, K. F., Han, Y., Gadkari, V. V, & Ruotolo, B. T. (2022). Mass Spectrometry Methods for Measuring Protein Stability. *Chemical Reviews*, *122*, 7690–7719. https://doi.org/10.1021/acs.chemrev.1c00857

Vendruscolo, M., Paci, E., Karplus, M., & Dobson, C. M. (2003). Structures and relative free energies of partially folded states of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 14817–14821. https://doi.org/10.1073/pnas.2036516100

Wagner, W. J., & Gross, M. L. (2022). Using mass spectrometry - based methods to understand amyloid formation and inhibition of alpha - synuclein and amyloid beta. *Mass Spectrometry Reviews*, 1–44. https://doi.org/10.1002/mas.21814

Walsh, G. (2003). *Biopharmaceuticals: Biochemstry and Biotechnology*. https://doi.org/10.1590/s1516-93322005000200017

Walsh, G., & Walsh, E. (2022). Biopharmaceutical benchmarks 2022. *Nature Biotechnology*, *40*(12), 1722–1760. https://doi.org/10.1038/s41587-022-01582-x

Wang, L., & Chance, M. R. (2017). Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Molecular & Cellular Proteomics*, *16*(5), 706–716. https://doi.org/10.1074/mcp.O116.064386

Wang, W., & Roberts, C. J. (2018). Protein aggregation – Mechanisms, detection, and control. In *International Journal of Pharmaceutics* (Vol. 550, Issues 1–2, pp. 251–268). Elsevier B.V. https://doi.org/10.1016/j.ijpharm.2018.08.043

Wang, W., Singh, S. K., Li, N., Toler, M. R., King, K. R., & Nema, S. (2012). Immunogenicity of protein aggregates - Concerns and realities. *International Journal of Pharmaceutics*, *431*(1–2), 1–11. https://doi.org/10.1016/j.ijpharm.2012.04.040

Wang, X., Singh, S. K., & Kumar, S. (2010). Potential aggregation-prone regions in complementarity-determining regions of antibodies and their contribution towards antigen recognition: A computational analysis. *Pharmaceutical Research*, *27*(8), 1512–1529. https://doi.org/10.1007/s11095-010-0143-5

Willis, L. F. (2018). *The effects of flow on therapeutic protein aggregation*.

Willis, L. F., Kumar, A., Dobson, J., Bond, N. J., Lowe, D., Turner, R., Radford, S. E., Kapur, N., & Brockwell, D. J. (2018). Using extensional flow to reveal diverse aggregation landscapes for three IgG1 molecules. *Biotechnology and Bioengineering*, *115*(5), 1216–1225. https://doi.org/10.1002/bit.26543

Willis, L. F., Kumar, A., Jain, T., Caffry, I., Xu, Y., Radford, S. E., Kapur, N., Vásquez, M., & Brockwell, D. J. (2020). The uniqueness of flow in probing the aggregation behavior of clinically relevant antibodies. *Engineering Reports*, *2*(5), 1–13. https://doi.org/10.1002/eng2.12147

Wilm, M., & Mann, M. (1996). Analytical Properties of the Nanoelectrospray Ion Source. *Analytical Chemistry*, *68*(1), 64–71.

Wilm, M. S., & Mann, M. (1994). Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes*, *136*, 167–180.

Winter, G., & Harris, W. J. (1993). Humanized antibodies. *Trends in Pharmacological Sciences*, *141*, 139–143.

Xie, B., Sood, A., Woods, R. J., & Sharp, J. S. (2017). Quantitative protein topography measurements by high resolution hydroxyl radical protein footprinting enable accurate molecular model selection. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-04689-3

Xu, G., & Chance, M. R. (2007). Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical Reviews*, *107*(8), 3514–3543. https://doi.org/10.1021/cr0682047

Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S., & Ventura, S. (2015). AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein

structures. *Nucleic Acids Research*, *43*(W1), W306–W313. https://doi.org/10.1093/NAR/GKV359

Zhang, B., Cheng, M., Rempel, D., & Gross, M. L. (2018). Implementing fast photochemical oxidation of proteins (FPOP) as a footprinting approach to solve diverse problems in structural biology. *Methods*, *144*, 94–103. https://doi.org/10.1016/j.ymeth.2018.05.016

Zhang, H., Cui, W., & Gross, M. L. (2014). Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Letters*, *588*(2), 308–317. https://doi.org/10.1016/j.febslet.2013.11.027

Zhang, M. M., Rempel, D. L., & Gross, M. L. (2019). A Fast Photochemical Oxidation of Proteins (FPOP) platform for free-radical reactions: the carbonate radical anion with peptides and proteins. *Free Radical Biology and Medicine*, *131*, 126–132. https://doi.org/10.1016/j.freeradbiomed.2018.11.031

Zhang, Y., Rempel, D. L., Zhang, J., Sharma, A. K., Mirica, L. M., & Gross, M. L. (2013). Pulsed hydrogen – deuterium exchange mass spectrometry probes conformational changes in amyloid beta (A β) peptide aggregation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14604–14609. https://doi.org/10.1073/pnas.1309175110

Zhang, Y., Wecksler, A. T., Molina, P., Deperalta, G., & Gross, M. L. (2017). Mapping the Binding Interface of VEGF and a Monoclonal Antibody Fab-1 Fragment with Fast Photochemical Oxidation of Proteins (FPOP) and Mass Spectrometry. *Journal of the American Society for Mass Spectrometry*, *28*(5), 850–858. https://doi.org/10.1007/s13361-017-1601-7

Zhu, S., Wilson, D. J., Liuni, P., Chen, T., Houy, C., & James, D. A. (2021). Epitope screening using Hydrogen / Deuterium Exchange Mass Spectrometry (HDX-MS): An accelerated workflow for evaluation of lead monoclonal antibodies. *Biotechnology*, *17*(2), e2100358. https://doi.org/10.1002/biot.202100358

Zwanzig, R., Szabo, A., & Bagchi, B. (1992). Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(1), 20–22. https://doi.org/10.1073/pnas.89.1.20