Defining the mechanism behind the self-association of therapeutic monoclonal antibodies using mass spectrometric techniques

Paul William Andrew Devine

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

Astbury Centre for Structural Molecular Biology
University of Leeds

September 2016

The candidate confirms that the submitted work is his own and that appropriate credit has been given within the thesis where reference has been made 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.

© 2016 The University of Leeds and Paul William Andrew Devine

A note to the reader

With the aim of making this thesis easier to digest,
I have split the introduction into two parts:

i) An introduction to the theory of mass spectrometry, and
ii) An introduction into the field of structural biology.

Finally, the materials and methods section of this thesis is listed in
the order that they arise throughout this study.

Thank you

Acknowledgements

It is with great pleasure that I thank everyone who has helped throughout my PhD. With their support, I can sincerely say that my time in Leeds has been one of the most rewarding experiences.

Firstly, my deepest thanks go to my two supervisors **Prof. Sheena Radford** and **Prof. Alison Ashcroft** for their unbowed encouragement and constant guidance throughout my entire PhD. Their enthusiasm for science is nothing short of incredible and helped the completion of my project and this thesis, especially in the difficult and demanding times. I am incredibly grateful for the time and effort they have poured into this project, as well as helping me to continually develop as a scientist. I would also like to thank my industrial supervisors **Dr David Lowe** and **Dr Dan Higazi**. Their time and insight has helped overcome many a hurdle throughout this project.

My gratitude goes out to all members of the Radford and Ashcroft labs, both past and present, for the advice and support they have provided since day one; they have made the lab a great place to be. For the science related, I thank **Dr Claire Sarell** and **Dr Sophia Goodchild** for help with the ÅKTAmicro, **Dr Claire Windle** for help with PyMol at every turn and **Dr James Ault** for all things mass spectrometry related, regardless of how naive my questions were. For the not so science related, thanks to the Monday night gaming crew (and **Patrick** for hosting us!) for keeping up the much needed breaks from the science and thesis writing. I would also like to thank **Nasir Khan** who keeps the lab running smoothly all year round without fail, as well as providing an endless supply of biscuits and doughnuts.

I owe a tremendous amount of thanks to my collaborator **Dr Janet Saunders** who originally developed the *in vivo* assay used in this thesis. As well as providing the beginners guide to molecular biology, you helped me to obtain results in the final months that would have otherwise been impossible by myself. I hope every collaboration is as fun as this was!

Finally, my last and immeasurable thanks go to my parents **Moira** and **Stephen**, and my brother **Michael**. Their unwavering confidence and encouragement has shaped me into the

person I am today. I owe every accomplishment I have achieved to their love and support. I therefore dedicate this thesis to them. I hope that it makes you proud.

"If you're not going to play golf in the rain in Scotland,
you're not going to play golf at all"
-Stephen Devine

Table of contents

List of figures

List of tables

List of abbreviations

ADH = Alcohol dehydrogenase

ARM = Antibody-ribosome-mRNA

ATD = Arrival time distribution

AUC = Analytical ultracentrifugation

βla = β-lactamase

BPI = Base peak index

BS3 = Bis(sulfosuccinimidyl)suberate

CCA = Concanavalin A

CCS = Collision cross-sectional area

CDR = Complementarity-determining region

CEM = Chain ejection model

CID = Collision induced dissociation

CRM = Charged residue model

CsI = Caesium iodide

CTD = Charge transfer dissociation

dAbs = Single domain antibodies

DC = Direct current

DEPC = Diethylpyrocarbonate

DLS = Dynamic light scattering

ECD = Electron capture dissociation

EM = Electron microscopy

EMDB = Electron microscopy data bank

ESI = Electrospray ionisation

EtBr = Ethidium bromide

ETD = Electron transfer dissociation

Fab = Antigen binding fragment

FAB = Fast atom bombardment

Fc = Fragment crystallisable

FDA = Food and Drug Administration

fHbp = Factor H binding protein

FPOP = Fast photochemical oxidation of proteins

FT-ICR = Fourier transform ion cyclotron resonance

Fv = Variable fragment

FWHM = Full width at half maximum

HAT = Hypoxanthine-aminopterin-thymidine

HDX = Hydrogen-deuterium exchange

IEM = Ion evaporation model

ICR = Ion cyclotron resonance

Ig = Immunoglobulin

IHC = Immunohistochemistry

IL1β = Interleukin-1-beta

IL1 = Interleukin-1 receptor

IMS = Ion mobility spectrometry

IPTG = Isopropyl β-D-1-thiogalactopyranoside

LB = Luria-Bertani

LC = Liquid chromatography

mAb = Monoclonal antibody

MALDI = Matrix-assisted laser desorption/ionisation

MCD = Maximal cell dilution

MCP = Microchannel plate

MD = Molecular dynamics

MS = Mass spectrometry

MS/MS = Tandem mass spectrometry

PCR = Polymerase chain reaction

PDB = Protein data bank

PEG = Polyethylene glycol

PIC = Pre-initiation complex

PK = Pharmacokinetics

Ppm = Parts per million

PSA = Projection superposition approximation

PTM = Post translational modification

PyK = Pyruvate kinase

RF = Radio frequency

Rg = Radius of gyration

RMSD = Root mean square deviation

SAP = Spatial aggregation propensity

SASA = Solvent accessible surface area

SDS-PAGE = Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SEC = Size exclusion chromatography

TAE = Tris-acetate-ethylenediaminetetraacetic acid

TFA = Trifluoroacetic acid

TM = Trajectory method

ToF = Time of flight

TWIG = Travelling wave ion guide

TWIMS = Travelling wave ion mobility spectrometry

UV = Ultraviolet

VMD = Visual molecular dynamics

XL = Chemical cross-linking

List of amino acid abbreviations and their monoisotopic masses

A	Ala	alanine	71.037 Da
C	Cys	cysteine	103.009 Da
D	Asp	aspartate	115.026 Da
E	Glu	glutamate	129.042 Da
F	Phe	phenylalanine	147.068 Da
G	Gly	glycine	57.021 Da
H	His	histidine	137.058 Da
I	Ile	isoleucine	113.084 Da
K	Lys	lysine	128.094 Da
L	Leu	leucine	113.084 Da
M	Met	methionine	131.040 Da
N	Asn	asparagine	114.042 Da
P	Pro	proline	97.052 Da
Q	Gln	glutamine	128.058 Da
R	Arg	arginine	156.101 Da
S	Ser	serine	87.032 Da
T	Thr	threonine	101.047 Da
V	Val	valine	99.068 Da
W	Trp	tryptophan	186.079 Da
Y	Tyr	tyrosine	163.063 Da

Abstract

Protein aggregation is responsible for a vast array of life-threatening protein based diseases as well as being an economic hurdle in biopharmaceutical development and manufacturing. Monoclonal antibodies represent the fastest growing class of biotherapeutics, with 53 antibodies in late phase clinical trials as of late 2015. Antibodies serve as ideal therapeutics due to their exquisite specificity and favourable safety profile. However, further therapeutic antibody development is hamstringed by uncontrolled self-association and aggregation which can occur at all stages of biotherapeutic development. Therefore, there is an urgent need for methods to dissect the mechanisms that drive uncontrolled self-association and protein aggregation.

This thesis presents techniques which were applied to address the identification of aggregated material of a therapeutically relevant monoclonal antibody, and to characterise the mechanism responsible for driving oligomerisation. A combination of mass spectrometric techniques were employed to visualise the oligomeric species. Ion mobility spectrometry coupled to nanoelectrospray ionisation mass spectrometry was utilised to identify the oligomeric species formed under native conditions and to define the oligomers in terms of their mass and collision cross-sectional area.

To characterise the regions responsible for driving oligomer formation, chemical cross-linking was employed to capture the oligomeric species in solution which were then analysed using tandem mass spectrometry. The initial dimer interaction was modelled using distance restraints obtained from the chemical cross-linking results and a model proposed that explains the oligomerisation events, and how runaway polymerisation can occur at higher concentrations.

Finally, a powerful *in vivo* assay in the *E. coli* periplasm was developed to differentiate between aggregation and non-aggregation-prone sequences using single chain variable fragments (scFv) of the antibodies studied. The results presented demonstrate the applicability of the assay to molecules relevant to the biopharmaceutical sector; as an upstream platform for the identification of aggregation-prone sequences, prior to antibody production and development.

Overall, the work presented within this thesis describes techniques that can be successfully applied to define the mechanism that underpins the self-association of a therapeutically-relevant monoclonal antibody. Furthermore, the study presents a novel *in vivo* assay that can be used to identify aggregation-prone sequences, and to develop them further by mutagenesis, which could be useful in protein development in the biopharmaceutical sector.

Chapter 1:
Introduction I – The theory of mass spectrometry

1 Introduction I: The theory of mass spectrometry

1.1 A brief history of mass spectrometry

Mass spectrometry (MS) is an analytical technique that is used to study charged ions in the gas phase and measure them based upon their mass-to-charge ratio (*m/z*). It is a technique that is used primarily to analyse the mass of ions, but has been extended more recently to give in depth information on composition, as well as structure. For a sample to be analysed by MS it must first be introduced into the gas phase, which is achieved in the first stage of a mass spectrometer. The different stages of a mass spectrometer can be simplified and separated into ionisation, mass analysers and detection (**Figure 1.1**).

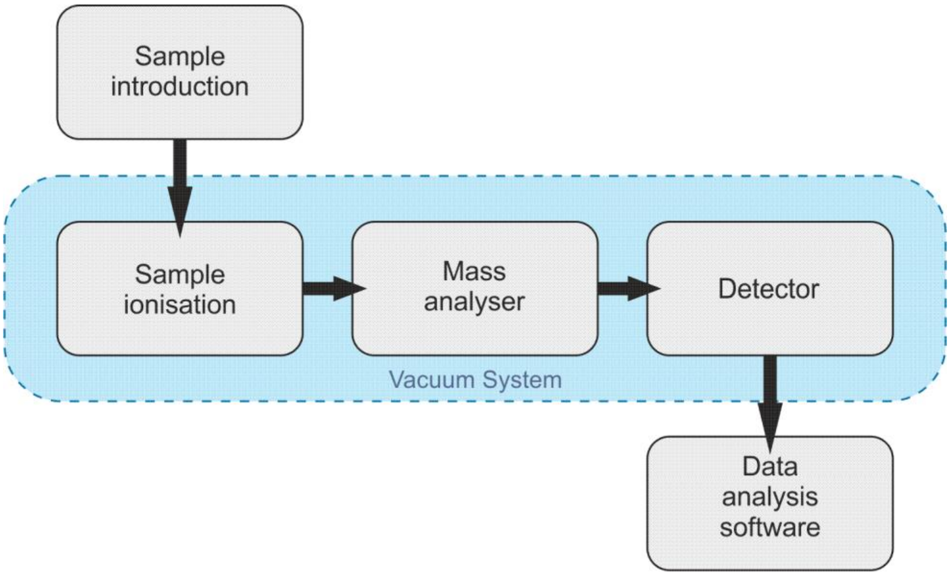


Figure 1.1: Schematic of sample analysis by use of mass spectrometry. A sample is introduced whereupon it is ionised and the ions separated in the mass analyser according to their *m/z* ratio. These ions are then measured by the detector where computer software can convert the data into a mass spectrum.

Modern mass spectrometers and their application have evolved a tremendous amount over the past century since their first use to measure the mass of an electron by J.J. Thomson in 1897 (Thomson, 1897). Thomson discovered that cathode rays, Lenard rays at the time, travelled at a speed much faster than one would expect for a particle of that size. He deduced that these negatively charged particles were 1,000 times lighter than the mass of a proton. Thomson denoted these particles as corpuscles, which were later renamed as electrons as G.S. Stoney had suggested before Thomson had ever made his discovery (Stoney, 1894). Thomson was awarded the Nobel Prize in physics for the "*great merits of his theoretical and experimental investigations on the conduction of electricity by gases"*. The first detection of elemental isotopes was also discovered by Thomson in 1913 (Thomson, 1913). Using magnetic and electric fields to manipulate the beam path of a stream of neon gas, he noticed the appearance of two distinct patches on the photographic plate used for detection. They concluded that there must be in fact two beams with different masses to give rise to the results observed, this was the first discovery of Ne-20 and Ne-22. This initial use of mass spectrometry was further developed by Aston who continued to measure isotopes of various different elements and proposed the "whole number system"; a system whereby the mass of all elements are whole number multiples of the mass of a proton, e.g. oxygen being 16 Da (Aston, 1935). In 1922, Aston was awarded the Nobel Prize in chemistry "f*or his discovery, by means of his mass spectrograph, of isotopes, in a large number of non-radioactive elements, and for his enunciation of the whole-number rule"*.

1.2 Ionisation

Ionisation can be achieved by a variety of methods whereupon the sample is analysed by merit of its *m/z* ratio. Ions of different *m/z* are separated from one another before the abundance of each ion is finally recorded by the detector. Modern day applications of mass spectrometry range from the analysis of small chemicals through to intact bio-molecules. The analysis of complex bio-molecules was limited at first, however, due to the methodologies used in sample ionisation. The first main breakthrough in this field was the development of fast atom bombardment (FAB) (Barber et al., 1981). FAB works by striking a sample containing a non-volatile surface matrix with high voltage atoms of an inert gas (typically argon or xenon). The sample in the matrix is then energised as the inert gas is neutralised when it collides with the surface and the protein ions are ejected in the form of $[M+H]^+$. FAB is a relatively "soft" ionisation technique in that it the energies impacted on the sample do not cause any fragmentation of the analyte. The technique is not ideal, however, since samples have to first be embedded in a matrix that is commonly made from glycerol or 3-nitrobenzyl alcohol.

1.2.1 Soft ionisation techniques

Native MS had its real breakthrough from the developments of the soft ionisation techniques matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) which result in no fragmentation of the analyte, and therefore is capable of retaining non-covalent interactions upon transition into the gas phase.

MALDI (Karas et al., 1985) first requires the analyte of interest to be mixed with a matrix, usually aromatic acids in nature (Fitzgerald et al., 1993), that is then spotted down onto a metal plate and left to crystallise together. Once co-crystallised, the samples are irradiated with a UV laser (337 nm) where ablation and desorption of the analyte from the matrix occurs (Knochenmuss, 2006) (**Figure 1.2**). The charged analytes, typically singly charged (in the form of $[M+H]^+$) are then introduced to the source of the mass spectrometer. Although suited for small peptides and proteins, MALDI is not necessarily the best choice for large biomolecules due to the energy that is imparted on the sample from the laser irradiation. Furthermore, the co-crystallisation of the sample with the acid based matrix is not ideal for retaining non-covalent interactions and so ESI tends to be preferred for the analysis of larger proteins because of its ability to maintain such interactions for analysis in the gas phase.

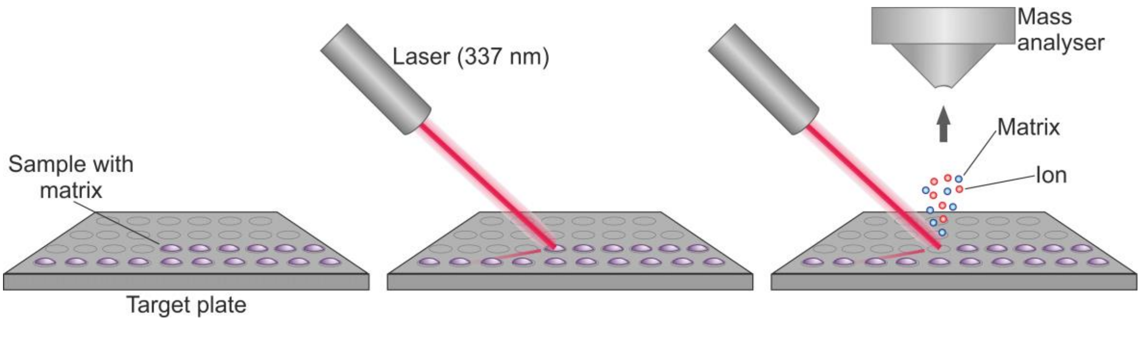


Figure 1.2: Schematic of the MALDI ionisation process. Samples are co-crystallised with the matrix onto a MALDI plate. A UV laser at 337 nm is used to irradiate the matrix where sample ablation and desorption from the matrix occurs. The gas phase matrix undergoes charge transfer to protonate the ions.

ESI is an ionisation technique that takes place at atmospheric pressure and is seen as the most gentle of ionisation methods since many more non-covalent interactions are retained. Furthermore, the technique has been shown to be applicable to a wide range of proteins from a few hundred Da to multiple MDa (Bothner and Siuzdak, 2004, Van Berkel et al., 2000, Kaddis et al., 2007, Snijder et al., 2013). The electrospray process occurs as a capillary, containing a sample of interest, has a high potential applied across it and an electrode placed at the entrance of the mass spectrometer. As the voltage is applied across the capillary, a Taylor cone is formed at the front of the charged sample-containing solvent (**Figure 1.3**).

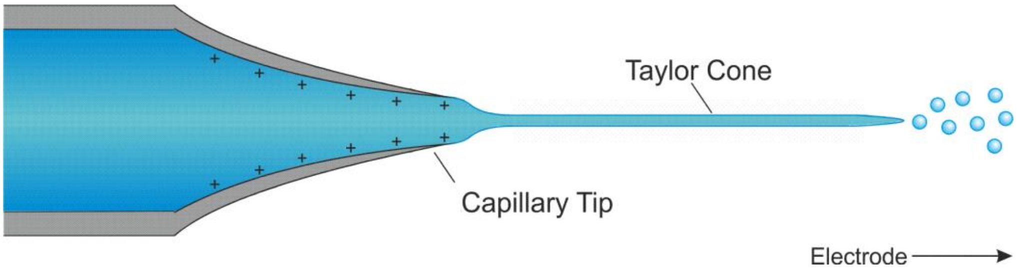


Figure 1.3: Mechanism of ionisation using electrospray ionisation. Upon applying a voltage between the capillary and the entrance of the instrument, a Taylor cone is formed from the build-up of charge at the end of the capillary where fine charged droplets are ejected.

The sample is then ejected in the form of an aerosol from the end of the filament formed by the Taylor cone (Taylor, 1964). The highly charged droplets in this aerosol contain the sample of interest. The droplets then decrease in size due to desolvation with the aid of a nebulising gas (typically nitrogen), at the front of the instrument. To further assist the desolvation process, a drying gas is sprayed orthogonally to the path of the droplets as they are attracted to the negatively charged source of the instrument (when working in positive ion mode). As the size of these droplets decrease, they experience an increase of columbic repulsion to a point where coulombic fission occurs due to the droplet reaching the Rayleigh limit (Taflin et al., 1989). From this stage, it is somewhat controversial as to how the proteins have the charge from the droplet deposited on their surface. There are two well accepted models for what happens to the droplets from this point onward, the charged residue model (CRM) (Dole et al., 1968) and the ion evaporation model (IEM) (Iribarne and Thomson, 1976) (**Figure 1.4**).

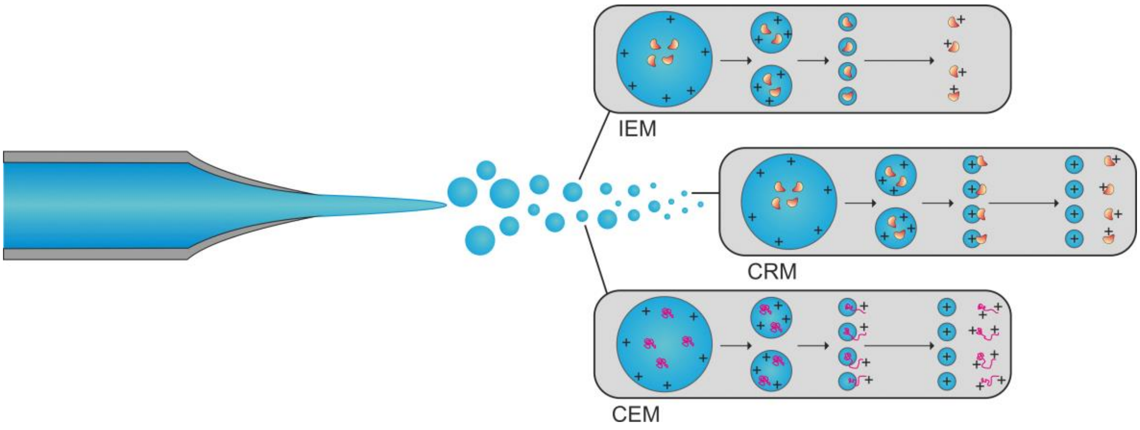


Figure 1.4: Schematic showing the different mechanisms of charge transfer onto the sample. The individual charged droplets decrease in size until the Raleigh limit is reached, and Coulombic fission of the droplets occurs. IEM = ion evaporation model, CRM = charged residue model and CEM = chain ejection model.

The CRM is a model that is expected to hold with large polar samples such as proteins and other bio-molecules. This model states that as the droplets continue to decrease in size, they reach a stage where protons are ejected via the ion evaporation model (IEM) to prevent further coulombic fission occurring. It is believed that the proteins remain near the centre of the droplets so that they can maximise their interactions at the water interface and the charges are transferred onto positively charged side-chain residues by charge carriers during the final moments of the process. The IEM proposed by Iribarne and Thomson proposes that as the droplets decrease in size, they come to a state where the charge across the surface of the droplet is enough to eject the analyte carrying protons. As the CRM is believed to hold true for proteins and larger molecules, there is literature that provides evidence that the IEM is the method of ionisation of small molecules (Kebarle and Verkerk, 2009, Znamenskiy et al., 2003). A third model has recently been proposed for the mechanism of ionisation for intrinsically disordered proteins, the chain ejection model (CEM) (Konermann et al., 2012). The CEM suggests that a region of the peptide chain is expelled from the droplet, since the disordered protein/peptide is unlikely remain at the centre of the desolvating droplet, as stated by the CRM. Therefore this model has some similarities with the IEM in that it states that a small region of the protein is expelled as a charged species from the droplet, sequentially followed by the rest of the protein until it is fully released into the gas phase.

The electrospray process was improved further in the field of native MS by the introduction of nanoelectrospray technology (nESI). With the use of volatile buffers (typically ammonium salts), this new technology allowed for slower flow rates (Wilm and Mann, 1996, Juraschek et al., 1999) meaning that a few microliters of sample could be sprayed from a nESI source for hours. Lower flow rates and lower voltage potentials applied (~1 kV) allowed for smaller droplets to form at the end of the capillary, made from borosilicate glass often coated with gold or palladium. These smaller droplets enhance the desolvation process and, due to lower energies, have greater capacity to retain the non-covalent interactions between bio-molecules. The work on ESI resulted in John Fenn (Whitehouse et al., 1989) being awarded the Nobel Prize in 2002 along with Koichi Tanaka (Tanaka et al., 1988), who developed laser desorption ionisation, for "*their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules*".

1.3 Mass analysers

Mass analysers function by separating ions created in the source of the instrument based on the *m/z* ratio. There are various different types of analysers: quadrupoles, time of flight (ToF), ion trapping and ion cyclotron resonance (ICR) analysers. Each of the different mass analysers outperform others in unique aspects and therefore the type of mass analyser chosen is dependent on the function required. Mass analysers are normally characterised by their resolving power, sensitivity, mass limitation and accuracy.

The resolving power of a mass analyser is an important characteristic and is typically measured in two ways. The mass resolution is defined by the capacity of the analyser to separate two peaks, close on the *m/z* scale. First, resolution can be defined as the distance a single peak covers on the *m/z* scale at half of its maximum height (Δm/z) otherwise known as its full width at half maximum (FWHM) (Hoffmann and Stroobant, 2001) (**Figure 1.5a**). The second definition is peaks are said to be resolved (for different types of analysers) when the valley between two adjacent peaks (Δ*m/z*) lies below a percentage threshold of the previous peak (**Figure 1.5b**). This threshold is distinct for different analysers. For a ToF analyser this is 50 % whereas an orbitrap (ion trap) analyser allows a percentile maximum of 10 %.

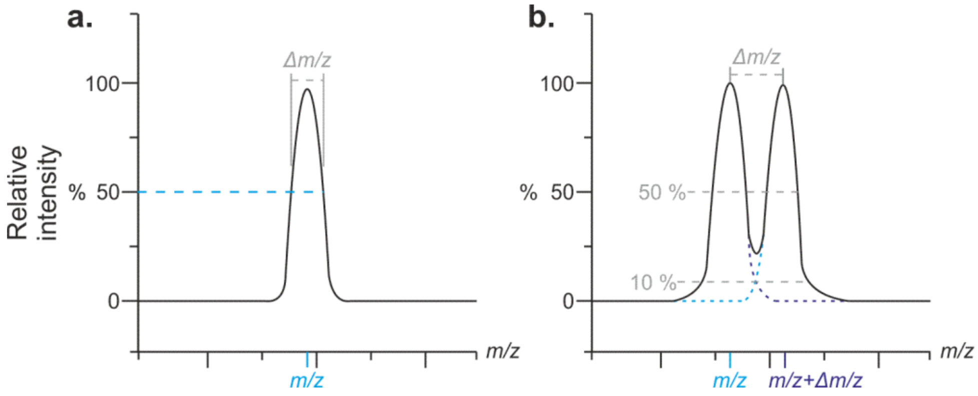


Figure 1.5: Schematic showing the different definitions of mass resolution. **(a)** Mass resolution is defined as the full width at half maximum of a given peak and **(b)** mass resolution of two peaks at a percentile, set by the analyser, of the peak maximum.

Sensitivity is a characteristic that is measured across an entire instrument with regards to ion loss at different stages throughout the instrument. However, it is measured as percentage of ions that reach the detector compared with the amount of ions created at the source. The mass limitation of an instrument, or mass range, is the full *m/z* scale at which an analyser can measure and separate ions. The mass limitation/range differs greatly for different analysers, a quadrupole analyser for example has a relatively low mass range whereas a ToF has a, theoretically, infinite mass limitation.

Lastly, the mass accuracy of a detector is measured in parts per million (ppm) and is defined as:

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

Equation 1.1

This measures how much the experimental mass of an ion diverges from its theoretical mass. An example of this is if an analyser has a mass accuracy of 10 ppm, we would expect the observed mass of a sample of predicted 1,000 Da to be within ±0.01 Da.

1.3.1 Quadrupole analysers

A quadrupole analyser (Miller and Denton, 1986, Paul, 1990, Dawson, 2013) exploits the resonating trajectory of ions at a particular *m/z* and uses this to separate them. The quadrupole consists of four metal (or metal coated, ceramic) rods, two pairs that sit perpendicular to each other (**Figure 1.6**).

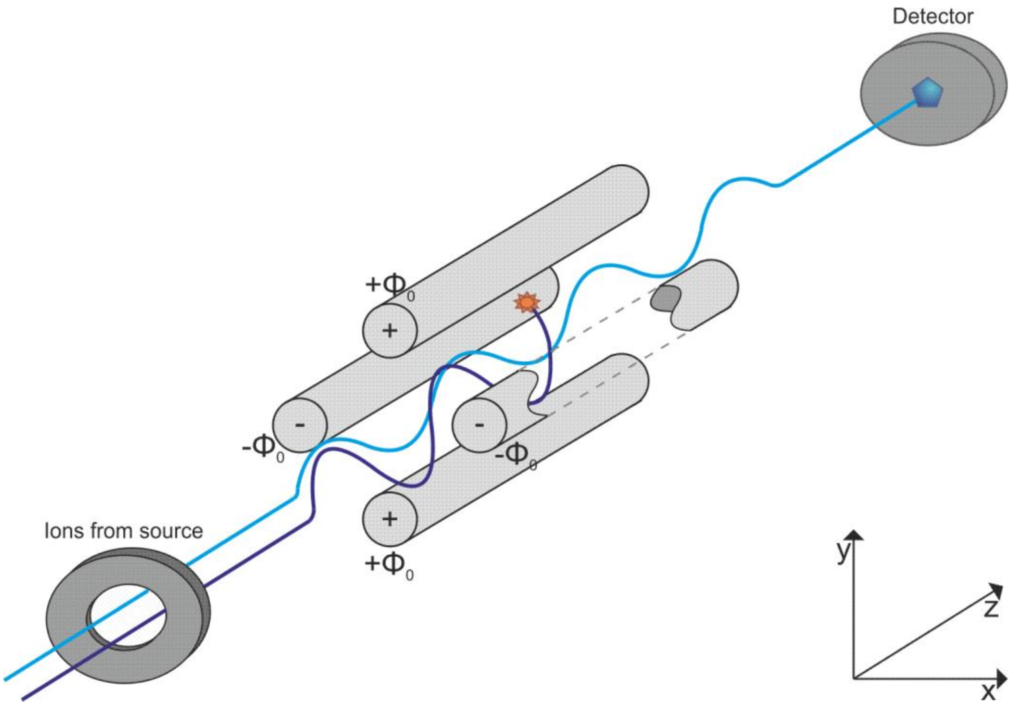


Figure 1.6: Schematic of a quadrupole analyser. The potentials of the rod pairs can be set to only allow transmission of ions of a given *m/z*. The voltages of the different rod pairs are altered to allow the transmission of ions of increasing *m/z*, which builds up the mass spectrum.

These pairs have electrical potentials applied to them, both a direct current (DC) potential and oscillating radio-frequency (RF) potential. The rod pairs are arranged so that each pair has an opposite polarity applied to it. Depending on the potentials applied to the poles at any given time, this dictates whether an ion of a given *m/z* will transmit through the analyser and reach the detector. The quadrupole is a scanning analyser and so allows ions of particular *m/z* through to the detector sequentially to build up a mass spectrum. Whether an ion with a certain *m/z* has a stable trajectory through the analyser is dependent on the potential applied to the poles at that given time which is defined by:

$$\varphi_T = DC + RFcosvt$$

Equation 1.2

And:

$$-\varphi_T = -(DC + RFcosvt)$$

Equation 1.3

where φ_T is the total potential, DC is the strength of the direct current, RF the oscillating frequency, v is the frequency and t is time.

As the ions pass through the analyser in the z plane, their motions in the x and y plane are influenced independently depending on the strength of the potentials impacted on the ions. If their trajectory in either of these planes is unstable, particularly due to the fields implemented by the oscillating potential, the ion will collide with the charged rod and be neutralised; therefore never reaching the detector. Simplistically, lighter ions experience a larger effect from the oscillating RF field and heavier ions are impacted largely from the DC voltage applied (March, 1997). The stability of an ion can be expressed in separate stability diagrams in terms of their x-z and y-z trajectories (**Figure 1.7a**). As mentioned, the voltages across the quadrupole are then increased linearly while maintaining the same ratio between the RF and DC potentials. This linear trend between the RF and DC values, can be plotted to illustrate how ions of different masses are detected over the range of potentials (**Figure 1.7b**).

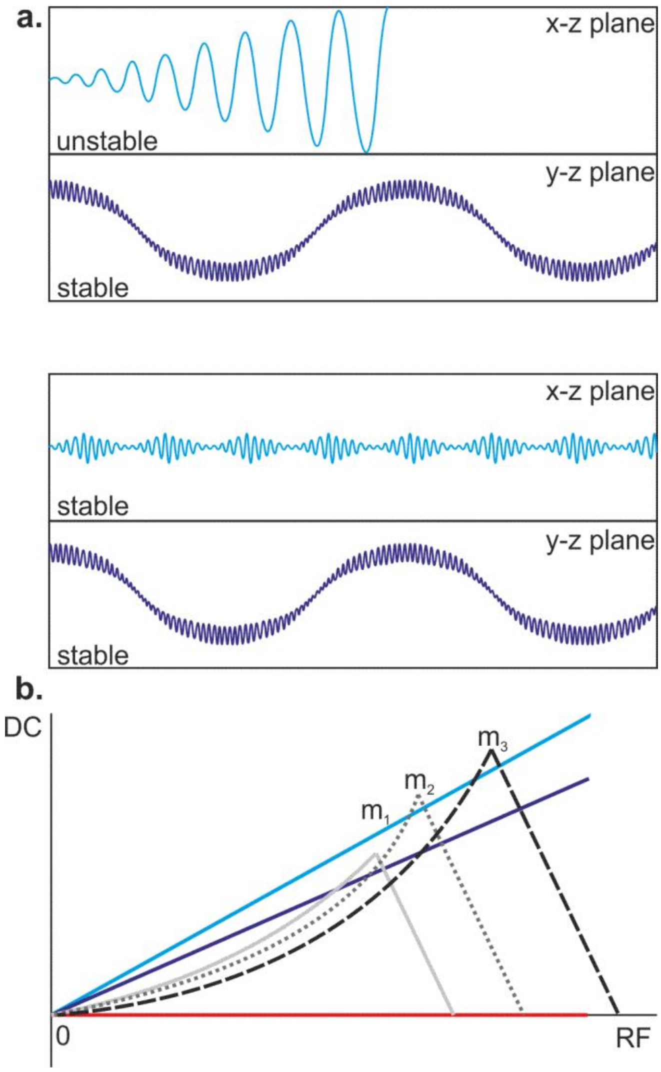


Figure 1.7: Stability diagrams for a quadrupole analyser. **(a)** Stability diagrams can be drawn to show the stability of an ion in the different planes of the rod pairs. The top image shows an unstable ion in the x-z plane while the bottom diagram shows an ion stable in both planes. **(b)** Stability areas are plotted against the RF and DC voltages, where the voltages are increased while maintaining a constant ratio between them. The blue line indicates a high resolution mode where ions close in mass can be detected, however m_1 is missed. The purple line shows a lower resolution mode where m_1 is resolved but now m_2 and m_3 are not, due to the overlapping profiles. The red line indicates where the quadrupole acts as a linear ion guide, when only and RF voltage is applied with no DC voltage. Figures redrawn from (Hoffmann and Stroobant, 2001).

Only when an ion has a stable trajectory, in both the x-z and y-z planes, and intersects the lines shown in **Figure 1.7b** is the ion able to pass through the analyser and reach the detector.

These lines plotted (**Figure 1.7b**) also highlight a quadrupole's resolution power, where the steeper the gradient of the line (while keeping the ratios constant) corresponds to greater resolution. This is shown for the blue and purple lines where the only the purple line allows the detection of all three ions (m_1, m_2 and m_3). However, the resolving power between m_2 and m_3 is reduced compared with the blue line. Finally, when only a RF potential is applied (DC = 0, red line) the resolving power of the analyser is lost and allows the transmission of all ions to the detector.

1.3.2 Time-of-flight (ToF) analysers

The ToF analyser is ideal due to its exquisite mass range. ToF analysers calculate an ion's *m/z* by measuring the time taken for an ion to transfer a field-free region after being accelerated by an applied field (Stephens, 1946). The time taken for an ion to reach the detector is based on two parameters, (1) the mass and (2) the charge of the ion. The mass of an ion is responsible for its final velocity since all ions experience the same amount of kinetic energy (E_K) per charge found on the ion since:

$$v^2 = \frac{2E_K}{m}$$

Equation 1.4

where v = velocity, E_k = kinetic energy and m = mass

Therefore, ions of larger mass will travel more slowly, ultimately taking a longer time to reach the detector. The *m/z* of an ion can be calculated from the time taken to traverse the analyser by starting with the potential energy (E_P) of an ion which is directly related to its charge (z) and the strength of the electric field (V):

$$E_P = zV$$

Equation 1.5

As the charged particle is accelerated into the flight tube, the potential energy is converted into kinetic energy and rearranging **Equation 1.4** gives:

$$E_K = \frac{1}{2}mv^2$$

Equation 1.6

as energy is conserved ($E_P = E_K$), we can combine equations **Equation 1.5** and **Equation 1.6**:

$$zV = \frac{1}{2}mv^2$$

Equation 1.7

since velocity is related to distance and time (v = d/t), we can substitute the value of velocity as:

$$zV = \frac{1}{2}m\left(\frac{L}{t}\right)^2$$

Equation 1.8

where L is the length of the flight tube (in m) and t is the flight time of the ion (s). This can now be re-arranged to give:

$$t^2 = \frac{m}{z}\left(\frac{L^2}{2V}\right)$$

Equation 1.9

Modern ToF analysers have been modified to accommodate some of the problems with resolution that linear ToFs had due to the natural differences of isobaric ions formed as a result of the electrospray process. The properties that result in these isobaric ions are the kinetic energy distribution, resulting in isobaric ions with a different initial energy from the source, and thus different velocity. The spatial distribution of isobaric ions results from ions being formed at even slightly different regions in the source, which will experience different voltage potentials for different lengths of time and therefore impacts their final velocities. Finally there is the temporal distribution which results in isobaric ions being formed at different times, therefore reaching the detector at different times as a consequence.

To accommodate for these natural and unavoidable phenomena which result in poor resolution, there are two methods that can be applied to restore high resolving power. By applying a delayed voltage pulse (delayed pulse extraction) immediately before the field-free region of the ToF this corrects the velocities and flight times and ultimately, the resolution. A second method for increasing ToF resolution is by using a reflectron ToF analyser.

For linear ToF analysers, the introduction of delayed pulse extraction (Wiley and McLaren, 1955, Brown and Lennon, 1995) increased the resolution since all ions experience a delayed accelerating voltage that allows for the slower isobaric ions to catch up with their faster counter parts (**Figure 1.8**).

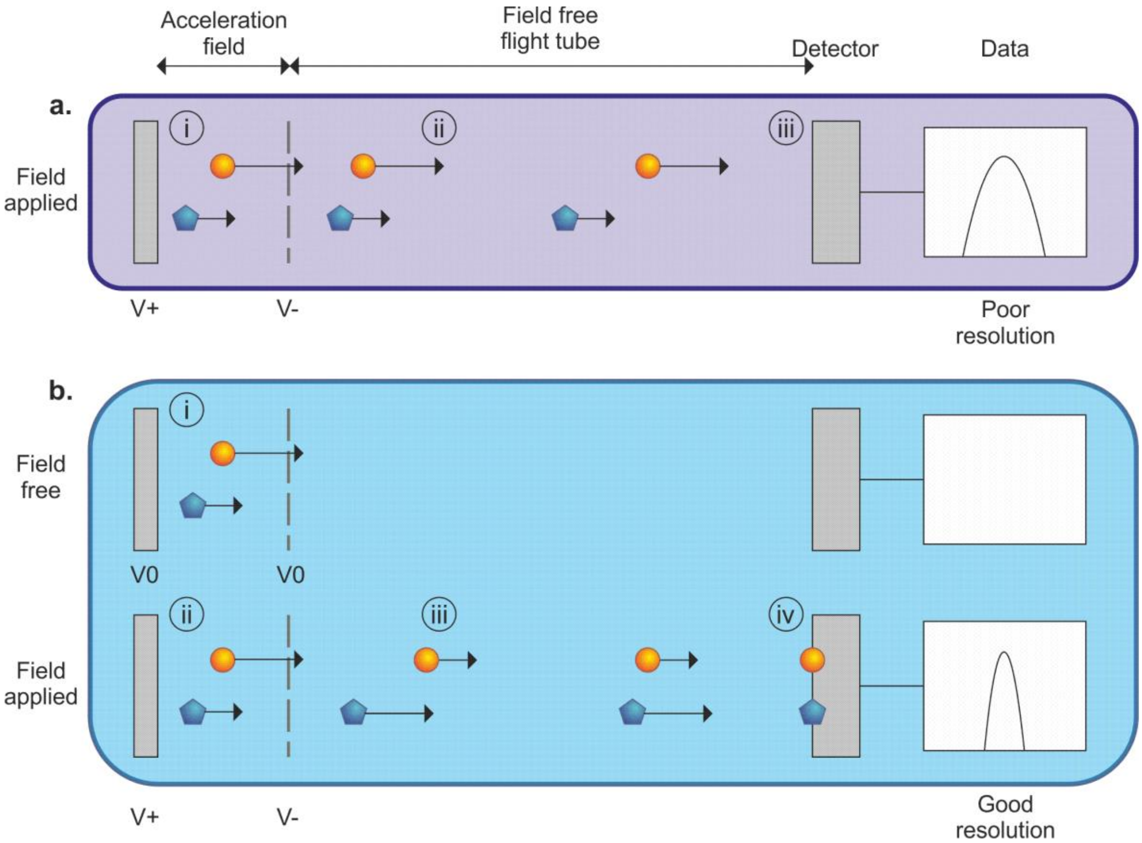


Figure 1.8: Schematic showing continuous pulse extraction against delayed pulse extraction. **(a)** Continuous pulse extraction of isobaric ions results the orange ion entering the field-free region with more energy (i) and therefore then travel the field-free region quicker (ii) and will reach the detector first (iii) resulting in poor resolution. **(b)** Delayed pulse extraction showing that two ions have (i) different energies, however upon applying a voltage, (ii) the blue ion experiences a greater pull and therefore is able to catch up in the field-free region (iii) and the two ions reach the detector at the same time (iv) resulting in greater resolution.

In a linear ToF, the isobaric ions that have higher energies have a greater velocity by the time that they reach the field-free region and so the ions reach the detectors at different times, resulting in the poor resolution. By having a delay before the pulse to accelerate the ions into the field-free region of the analyser, the slower isobaric ions experience a greater effect when the field is applied compared with the faster ions. This enhanced potential imparted on the slower ions increases their velocity in such a way that they catch-up with the isobaric ions and so reach the detector at the same time, increasing resolution.

The reflectron (Mamyrin et al., 1973, Mamyrin, 2001) is a number of ring electrodes stacked together so that they essentially function as an ion-mirror, repelling ions towards a detector perpendicularly placed from the ion beam (**Figure 1.9**). As the isobaric ions are accelerated towards the reflectron, the higher velocity ions will reach the reflectron first and penetrate the ion mirror further before being repelled back. As a result, the overall flight path for the higher energy isobaric ions is longer which gives time for the lower energy ions to catch-up and overtake them. Ions are then accelerated again as they leave the reflectron towards the detector. The ions reach their initial velocities again that allow for all ions to reach the detector at the same time.

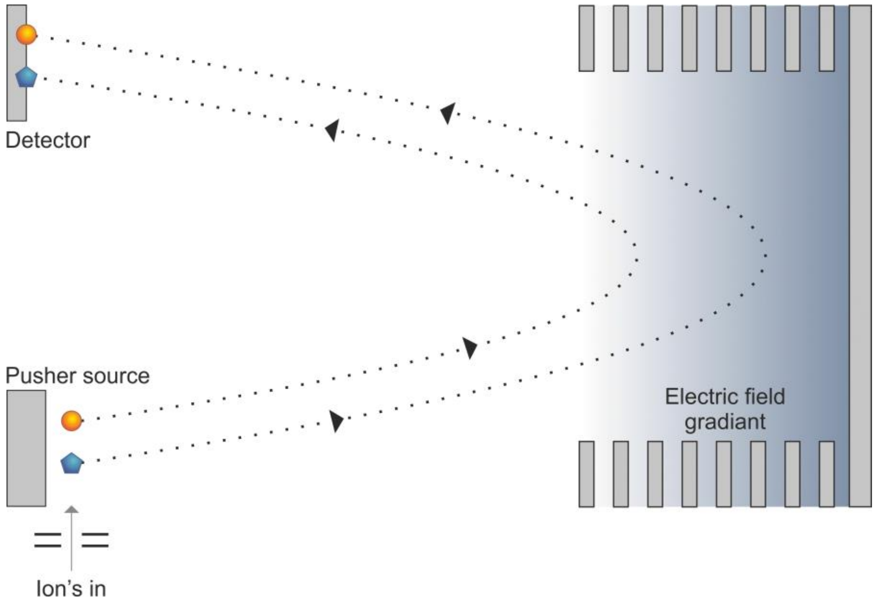


Figure 1.9: Schematic of the reflectron ToF analyser. Isobaric ions generated in the source with different energy potentials will travel faster and therefore penetrate further into the reflectron (blue) compared with the slower ions (orange). This results in the blue ion having a longer flight path, which due to its higher velocity, reaches the detector at the same time.

1.3.3 Orthogonal-acceleration ToF analysers

Coupling the ToF analyser with the pulsed ion extraction provided the perfect marriage between using ToF analysers and nESI methodology. Since ToF analysers need precise ionisation times this previously could not be achieved. The orthogonal acceleration ToF, developed in the late 1980's (Guilhaus et al., 2000), showed that it was possible to bring the two technologies together. The idea behind the orthogonal ToF was to take a continuous ion beam (as formed from nESI) and create pulsed packets of ions, with precisely defined times that could be introduced into a reflectron ToF (**Figure 1.10**).

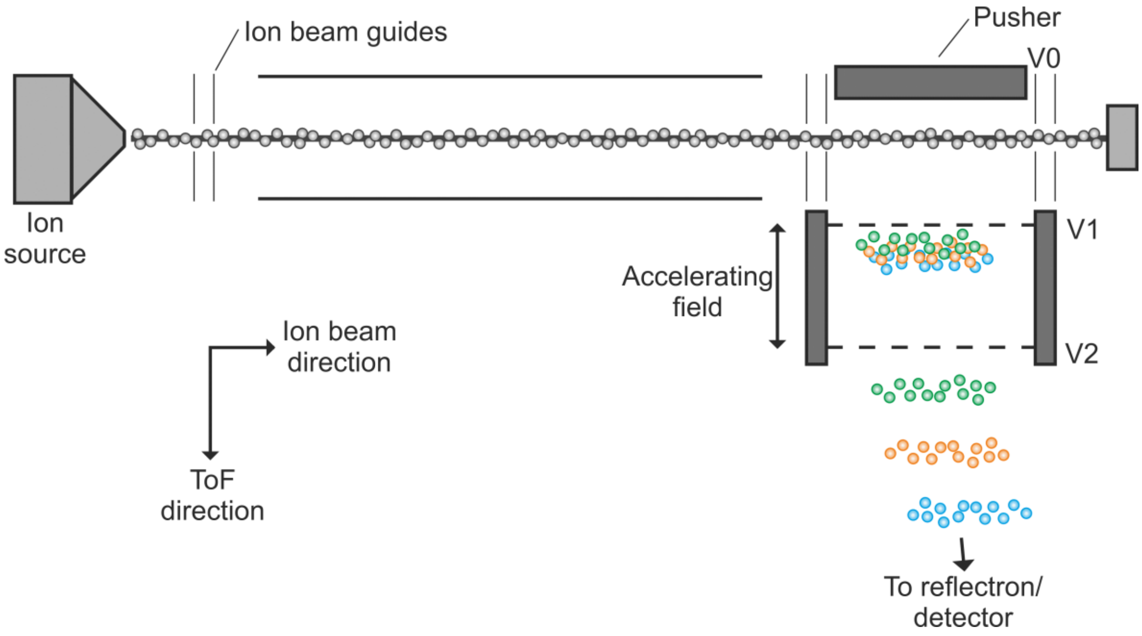


Figure 1.10: Schematic of an orthogonal ToF analyser. Resolution of different mass ions is increased by combining the delayed pulse extraction from the pusher with a reflectron.

By using a pusher source that sits perpendicular to the ion beam path and applying an electric potential at set time periods, packets of ions (in a field-free region) can be pushed towards the start of the reflectron. As the ions are pushed in the field-free region, they enter an accelerating field where they are propelled towards the reflectron and detector, retaining the high mass resolution previously mentioned with the reflectron ToF and pulsed extraction.

This orthogonal set up became paramount as it allowed for the introduction of another analyser before the ToF which became crucial for tandem mass spectrometry analysis. This kind of tandem geometry is now standard in many commercial instruments such as the Waters Q-ToF set up (**Figure 1.11**) that was used for the majority of the work presented in this thesis.

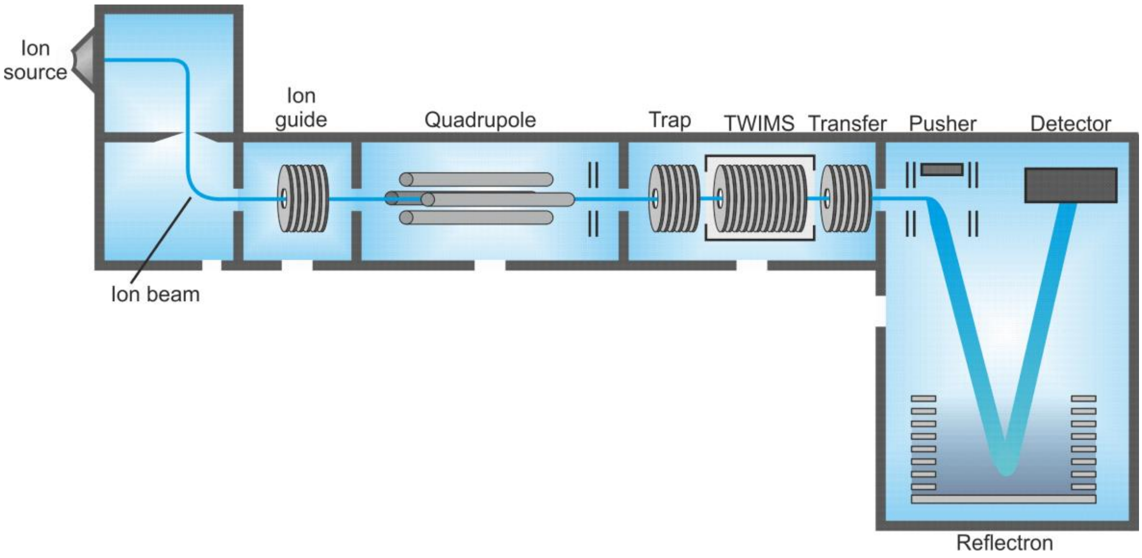


Figure 1.11: Schematic of the Synapt G1 instrument used in this study. This instrument consists of two mass analysers: a quadrupole and a delayed pulse reflectron ToF. The G1 instrument also allows for a travelling wave ion mobility spectrometry (TWIMS) with the inclusion of a travelling wave ion mobility cell for enhanced ion separation.

1.3.4 Ion trapping analysers

The most common form of ion trapping analysers is the orbitrap (Makarov, 2000, Hu et al., 2005). The orbitrap is a modification of the traditional ion trap analyser that was designed in the early 1920's (Kingdon, 1923). An orbitrap analyser works by trapping ions based on their attractions to an electrically charged spindle and their flight paths which is stabilised by the centrifugal forces they experience as they cycle around this spindle (**Figure 1.12**). FT-ICR analysers are also trapping analysers except that they also use a magnetic field to trap the ions in the analyser (Marshall and Hendrickson, 2002). The analysers are then detected by inducing a charge that is measured by a set of electrodes as the ions travel past them as a current. This signal then has to be Fourier transformed to produce a mass spectrum for the sample.

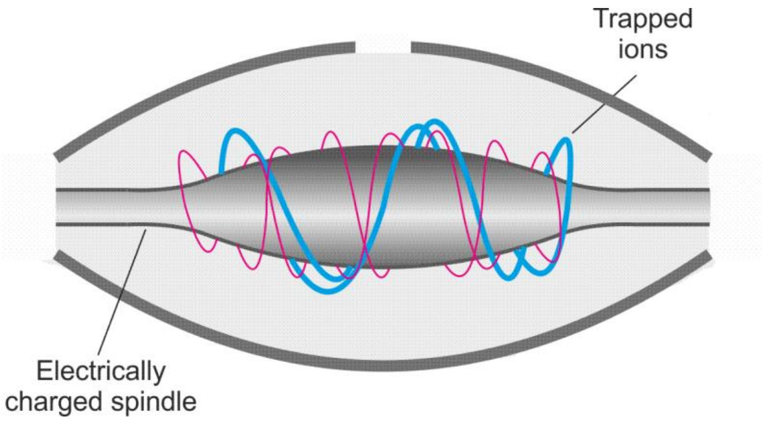


Figure 1.12: Schematic of an orbitrap ion trapping analyser. The trapping analyser functions by trapping ions by their attraction to the electrically charged spindle while separating them out based upon their different velocities.

1.4 Ion detection

The final stage of a mass spectrometer is detection of the ions after they have been separated by whatever analysers used in that instrument. There are a few types of ion detectors which are routinely used in most commercial instruments.

Ion detectors often take advantage of electron multipliers in order to amplify the signal received by the detector. An electron multiplier works by creating an electron cascade, through secondary emission, that is used to generate a mass spectrum after an ion has reached the detector (Hoffmann and Stroobant, 2001). There are various different types of electron multiplier detection systems; the most common of these is the microchannel plate (MCP) detector (Dubois et al., 1999) (**Figure 1.13**).

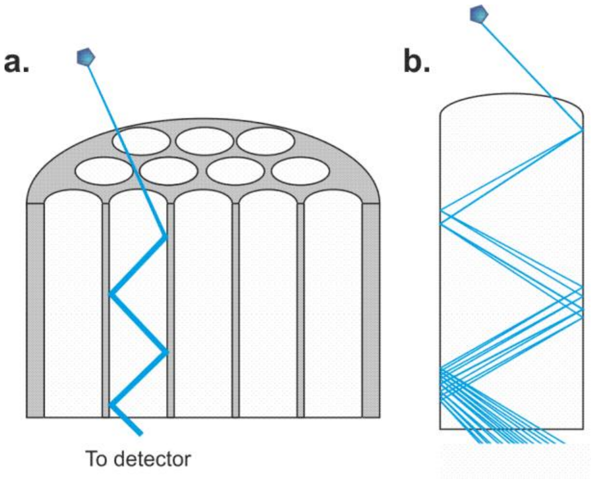


Figure 1.13: Schematic of the MCP electron multiplier detector. **(a)** The MCP Is made up from multiple continuous dynode detectors over a short distance to increase the precision of flight times. **(b)** Once an ion hits the dynode, this causes an electron cascade to occur increasing the signal reached at the end of the detector. Once a channel has been struck by an ion, it must recharge before it can be struck again.

The MCP is a continuous dynode electron multiplier in the form of a metal plate which consists of many small channels through the detector (Matsuura et al., 1985). As ions hit the wall of one of these channels, an electron is produced. This collision with the surface of the channel creates an electron cascade which increases the amount of electrons emitted with each impact. This increases the strength of the signal by several orders of magnitude before they reach the detector. The microchannel then needs time to recharge before it can detect another signal.

1.5 Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) utilises the power of two mass analysers within one instrument. By fitting a collision cell between two mass analysers, an ion can be selected (precursor ion) in the first analyser, fragmented in the collision cell and these fragmented products (product ions) are then analysed by the second mass analyser. This type of MS can be applied to small molecules (Kind and Fiehn, 2010) right up to large protein systems in order to study the individual sub-units within a larger complex (Sharon, 2010). A typical MS/MS experiment yields:

$$Precursor\ ion\ \rightarrow Product\ ions + Neutral\ losses$$

To fragment an ion, a variety of methods are now available including collision-induced dissociation (CID) (Jennings, 1968), electron-capture dissociation (ECD) (Zubarev et al., 1998), electron-transfer dissociation (ETD) (Syka et al., 2004) and surface-induced dissociation (SID) (Mabud et al., 1985). Each of these different methods of fragmentation have their own advantages, for example ETD does not fragment/cleave post-translational modifications (PTM) on the protein and therefore is an attractive technique for PTM mapping and profiling (Kim and Pandey, 2012). The most well studied and used method of fragmentation however is CID.

In the Waters Synapt G1 instrument (**Figure 1.11**), CID occurs as an ion is accelerated into a collision cell after mass selection in the quadrupole analyser. Both the trap and transfer cells before and after the IMS cell of the instrument can act as collision cells. These collision cells are filled with an inert gas that the selected ions collide with as they enter the cell. These collisions cause some of the kinetic energy of the ion to convert to vibrational energy which spreads throughout the ion. If the vibrational energy is large enough then it will cause a break in the peptide backbone of the ion and fragmentation occurs. The fragmentation that occurs has a predictable pattern based upon the type of mechanism used to form the product ions. CID fragmentation occurs through the breaking of a protein/peptide across the peptide bond. This cleavage yields two ions that are named the b- and y- ions, the N-terminal and C-terminal fragments respectively (**Figure 1.14**). These fragments then undergo further rounds of collisions and sequential cleavages across the amide bonds yielding signals in the acquired mass spectra that correspond to the peptide but differ by one amino acid either side. As the mass of each amino acid is known, the data can be analysed and used to build up the peptide sequence from the peaks that correspond to the mass of the next amino acid in the sequence.

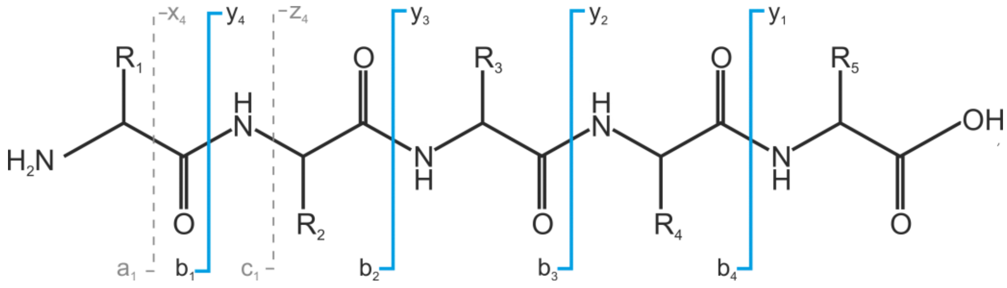


Figure 1.14: Schematic of a pentapeptide showing the different fragmentation patterns that can occur. Figure adapted from (Roepstorff and Fohlman, 1984).

As CID produces b- and y- ions, the other forms of fragmentation give rise to other product ion types. ETD and ECD for example yield c- and z- ions and the fragmentation event occurs on much faster timescales due to the mechanism of cleavage. The different types of fragmentation and their characteristics are shown in the table below (**Table 1.1**).

Table 1.1: Table of different fragmentation methodologies.

Fragmentation type	Timescale	Product ions produced
Collision-induced dissociation (CID)	Slow	b- and y- ions
Electron-transfer dissociation (ETD)	Fast	c- and z- ions
Electron-capture dissociation (ECD)	Fast	c- and z- ions
Surface-induced dissociation (SID)	Slow	b- and y- ions
Charge-transfer dissociation (CTD)	Fast	a- and x- ions

Depending on the data one wishes to acquire and what question they are wishing to address, there are a variety of different MS/MS experiments that can be performed depending on the scanning mode chosen.

1.5.1 MS/MS scanning modes

The different scanning modes used in MS/MS experiments allow for the acquisition of distinct data that can answer biological questions from a range of viewpoints. There are four main types of scan modes used which are summarised below (Louris et al., 1985) (**Figure 1.15**):

I. *Product ion scan*: A parent ion is selected using the first analyser in the mass spectrometer by virtue of its *m/z* ratio and then fragments in a collision cell. All product ions from this fragmentation are then analysed by the second mass analyser.

II. *Precursor ion scan*: Instead of the parent ion being selected, a specific product ion is selected in the second mass analyser. All precursor ions are allowed to traverse through the first analyser and undergo fragmentation in the collision cell. This technique is commonly used with multiple quadrupole instruments since the second analyser is also needed for selection.

III. *Neutral loss scan*: As with precursor ion scan, this technique is also not applicable to ToF analysers and so is common again with multiple quadrupole instruments. Both analysers are used to scan through a range of ions except the second analyser is looking for a specific loss in mass (from the neutral loss) during its scan. Only ions that correspond to the loss of mass in the second analyser reach the detector.

IV. *Selected reaction monitoring (SRM)*: Unlike the other reaction modes, SRM does not consist of a scan step and instead selects for a specific reaction. Ions are only detected by the second detector if the precursor that was selected gives the correct reaction being monitored for. One advantage to SRM is an increase in sensitivity due to fact it enables the focus on selected ions for longer timescales with both analysers, removing any need for scans as well as no interference from background ions.

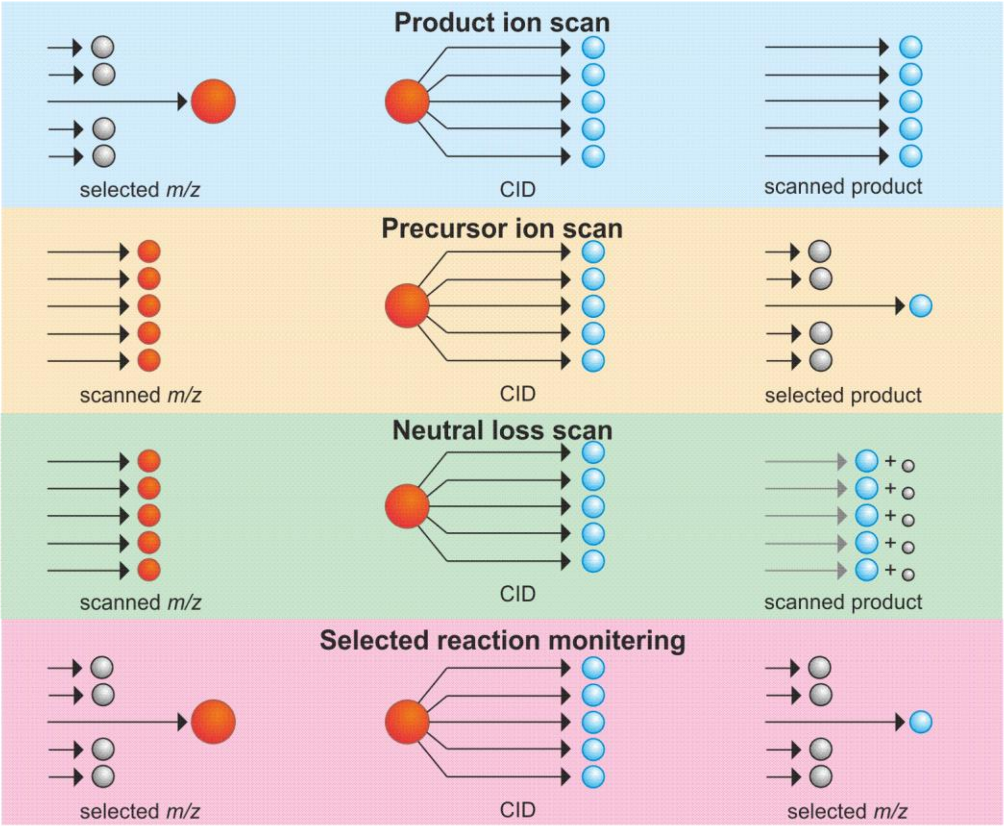


Figure 1.15: Illustration highlighting the different MS/MS scanning modes. Product ion scan involves isolating a precursor ion for fragmentation and scanning the products. Precursor ion scan allows fragmentation of all precursor ions and analysing for a selected product ion. The neutral loss scanning mode scans both the precursor and products while looking for a given mass loss. Selected reaction monitoring isolates a precursor for fragmentation and using the second analyser to isolate for a given product ion expected from the fragmentation. Figure redrawn and adapted form (Shepherd, 2012).

1.5.2 Collisional activation of protein complexes

As well as gaining sequence information, native-intact proteins and protein complexes can be activated in order to elucidate information about their higher order structure. Various studies have used collisional activation to understand protein complex assembly and structure (Benesch, 2009, Sharon, 2010, Zhou et al., 2008). Two main approaches have been used: CID, and more recently, SID. Both work on the same principal of using collisional energy to destabilise a protein complex until a subunit is ejected. The advantage of SID, when compared with CID, is the retention of a native-like conformation of the ejected monomer (Dongre et al., 1996) (**Figure 1.16**).

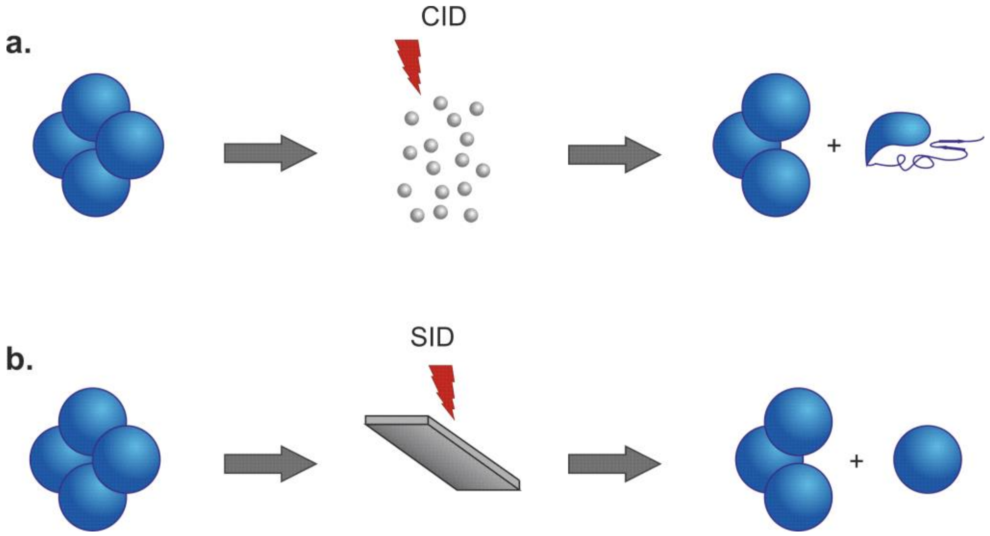


Figure 1.16: Schematic showing the mechanism of CID versus SID. **(a)** CID activation of a complex results in an unfolded monomer ejection. **(b)** SID activation results in ejection of a folded monomer, allowing determination of complex assembly. The surface used for SID is typically an alkanethiolate monolayer that has been self-assembled on a gold surface (Wysocki et al., 2008b).

The use of CID often results in a protein unfolding and gathering more charge as it is being ejected from the complex. By contrast, SID results in a more symmetrical charge state distribution after the collision and ion mobility has been used in conjunction with SID to show that the ejected protein remains in a native-like conformation. These techniques can be used to probe the quaternary structure of protein complexes and reveal structural information (Wysocki et al., 2008a, Wysocki et al., 2008b, Blackwell et al., 2011, Zhou et al., 2013).

1.6 Ion mobility spectrometry (IMS) mass spectrometry

To further aid in yielding higher resolution structural information about proteins and protein complexes, mass spectrometry can be combined with ion mobility spectrometry (IMS) as an extra dimension of separation. So far, measuring ions has only been discussed in terms of their *m/z* ratio. IMS-MS adds an extra scope of measurement by monitoring an ion's drift time (t_D) as well as its *m/z*. IMS works by separating ions based on their physical size and dimensions as well as the amount of charges the ion carries. The ions are pulled through a neutral buffer gas filled cell and will have different velocities, due to collisions with the gas molecules, which can be related to their shape and size via their rotationally averaged collision cross-sectional area (CCS). These experiments are carried out in a drift cell region of the instrument of which there are generally two types: a linear drift tube and a travelling wave (TW) device.

1.6.1 Linear drift tube IMS

A linear drift tube is the simplest form of IMS and information regarding a protein's size and shape can be directly related to its measured t_D. Ions are introduced into a drift tube under a constant electric field to which the ions are pulled through. As the ions traverse the cell they collide with the buffer gas which impedes their mobility (**Figure 1.17**). Different ions experience different levels of retardation based on their overall size and shape.

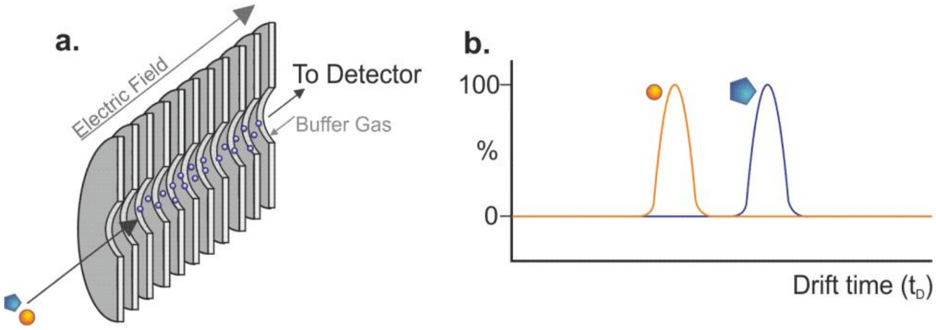


Figure 1.17: Schematic of a linear drift tube used for IMS separation. **(a)** Illustration of the drift cell showing the injection of buffer gas in the opposite direction of the electric field, resulting in collision with the ions. **(b)** Plot of the drift time against ion intensity showing the differences in drift time of the two ions due to the additional collisions with the buffer gas of the larger ion (blue).

As stated previously, the CCS of a protein or analyte can be directly measured from the time it takes to traverse the buffer gas filled cell. Where the mobility, K, of an ion is related to its velocity, v_i, and the electric field, E.

$$K = \frac{v_i}{E}$$

Equation 1.10

As the length, L, of the drift tube is known, and the drift time, t_D, can be measured, this gives:

$$K = \frac{L}{t_D E}$$

Equation 1.11

In order to standardise the equation, to allow for the comparison of results, the reduced mobility of an ion is reported. The reduced mobility of an ion (K_0) is normalized for standard pressure and temperature, P (760 torr) and T (273.2 K) respectively.

$$K_o = K\frac{273.2}{T}\frac{P}{760}$$

Therefore,

$$K_0 = \frac{L}{t_D E}\frac{273.2}{T}\frac{P}{760}$$

Equation 1.12

where L is the length of the drift tube and E is the electrical field. This is aided by the Mason-Schamp equation which relates the mobility of an ion to its CCS (Ω) (Mason and Schamp, 1958).

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

Equation 1.13

where z is the charge state of the analyte, e is elementary charge (1.6022x10^{-19} C), N is the density of the buffer gas used in the drift tube, μ is the reduced mass between the ion and the buffer gas and k_B is the Boltzmann constant (1.381x10^{-23} JK^{-1}). By substituting for reduced mobility (**Equation 1.12**) and re-arranging the equation for CCS (Ω) we can achieve.

$$\Omega = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{ze}{(k_B T)^{\frac{1}{2}}} \left[\frac{1}{m_i} + \frac{1}{m_n}\right]^{\frac{1}{2}} \frac{760}{P} \frac{T}{273.2} \frac{1}{N} \frac{t_D E}{L}$$

Equation 1.14

where m_i and m_n are the mass of the ion and the buffer gas used, respectively. As mentioned above, we can clearly see from **Equation 1.14** that a larger cross-section correlates with a longer drift time since the length (L) of the drift tube remains constant.

1.6.2 Travelling wave ion mobility spectrometry

The majority of modern mass spectrometers no longer use a linear drift tube, due to the ion losses suffered, and one other method of ion mobility instruments have adopted is the use of travelling wave ion mobility spectrometry (TWIMS) (Pringle et al., 2007, Giles et al., 2011). An example of a modern instrument that is equipped with a TWIMS device is the Waters Synapt mass spectrometer which was used for the work presented in this thesis (**Figure 1.18**).

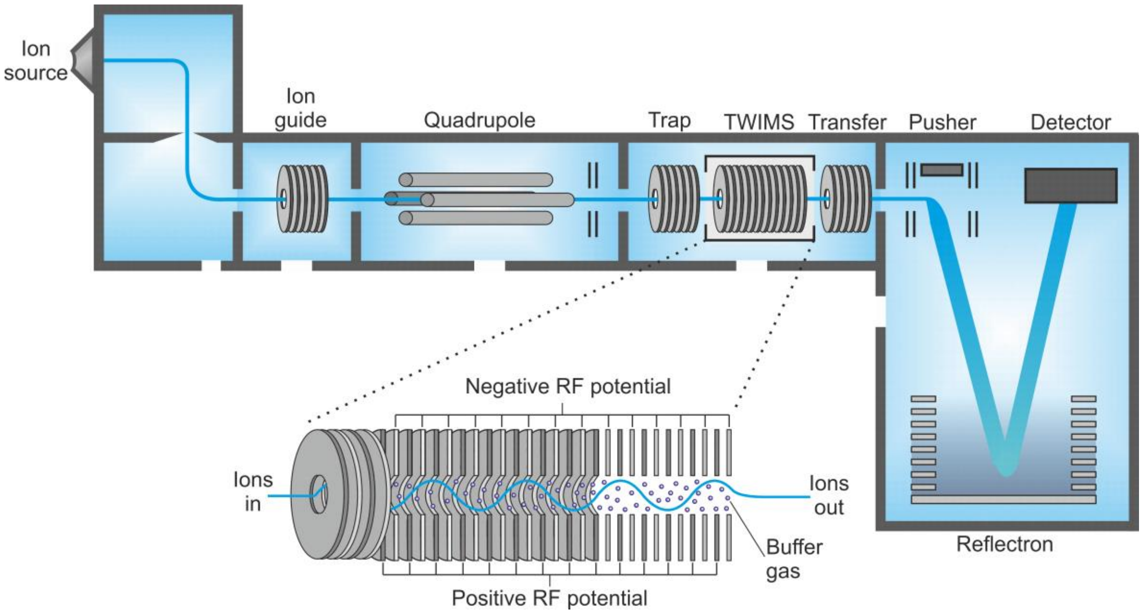


Figure 1.18: Schematic of the Synapt G1 instrument, highlighting the TWIMS device located between the two mass analysers.

In the Synapt instruments, the TWIMS device is situated between a quadrupole and ToF analyser which allows the selection of precursor ions which may or may not be collisionally activated in the trap region. The region made up of the trap, TWIMS and transfer of these instruments is known as the "TriWave" section; each of these cells are made up of a series of ring electrodes called travelling wave ion guides (TWIGS).

A series of ring electrodes are used to make up the TWIGS to which different electric potentials are applied in order to guide the ions through to the ToF analyser. A constant RF potential is applied to the electrodes which aids in confining the ions and minimises any ion loss. A DC voltage is applied to the first electrode in the series and moves from one ring electrode to the next sequentially which is where the "travelling wave" term is derived. Without this travelling wave potential, the ions would remain static in the TWIG due to the radially confining RF voltages applied (**Figure 1.19a**). For the TWIMS region, the TWIGS are surrounded in a cell which is filled with a neutral buffer gas, typically nitrogen or argon. Here, as the travelling wave pushes the ions along through the TWIG they collide with the neutral buffer gas and their mobility becomes retarded. An ion with a large enough cross-section becomes impeded to such an extent that it falls behind the travelling wave and therefore must wait for the subsequent wave to push it further along the cell (**Figure 1.19b**). Using the premise that the ions are separated on mobility through the cell by their CCS, this process can be tuned by applying different DC voltages superimposed on the TWIG which is also termed the wave height.

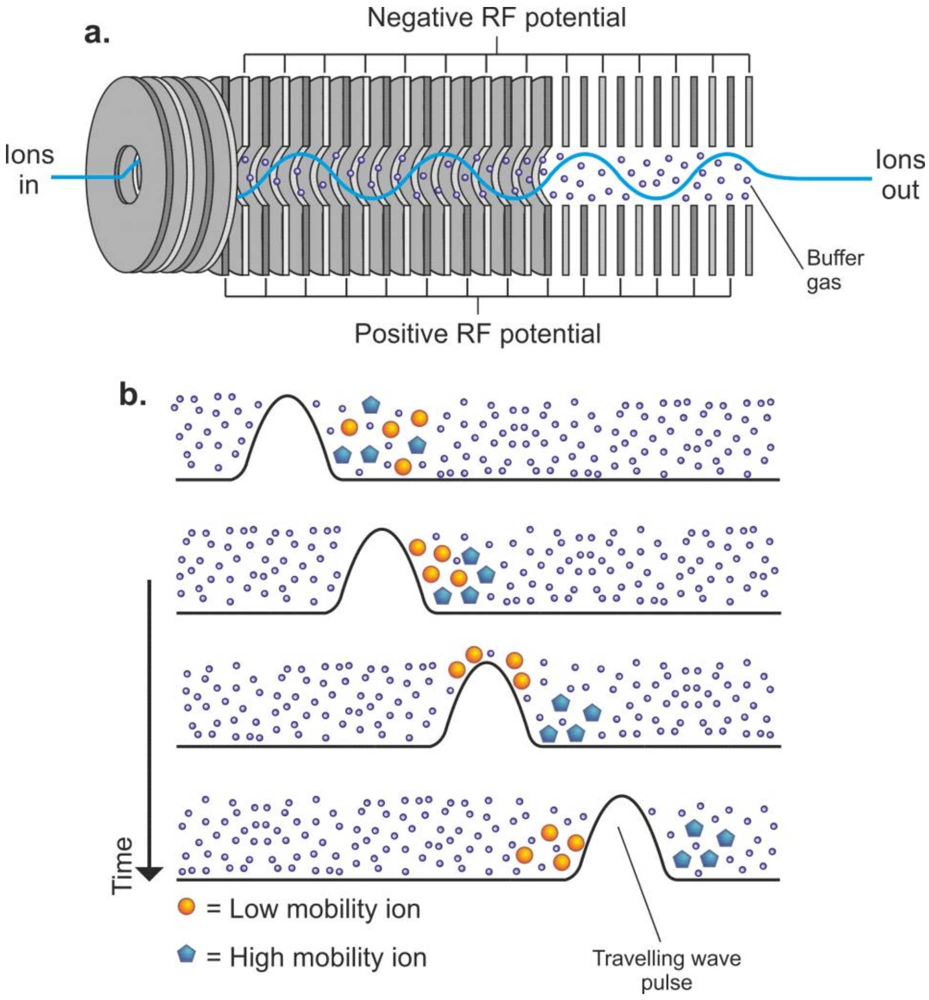


Figure 1.19: Illustration of the TWIMS device and its mechanism of separating ions based on their CCS. **(a)** Ring electrode with different electrical potentials prevent ions transmitting without the help of an additional travelling wave. **(b)** Low mobility ions (orange) collide with the buffer gas more frequently and so roll over the back of the wave and must await additional wave to reach the end of the device. While high mobility ions (blue) can ride the wave to the end of the device.

As the ions are now being propelled through the TWIMS cell, the previous method of calculating the CCS based upon an ions t_D no longer applies. In the absence of a uniform electric field, the difficulty in establishing the CCS of an analyte of interest increases. To address this, a calibration approach is adopted by taking proteins with known cross-sections using a linear drift tube from a database that has been provided by the Bush group (Bush et al., 2010).

The previous equation (**Equation 1.14**) used for calculating CCS values can be modified in order to account for the non-uniform electric field:

$$\Omega = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{ze}{(k_B T)^{\frac{1}{2}}} \left[\frac{1}{m_i} + \frac{1}{m_n}\right]^{\frac{1}{2}} \frac{760}{P} \frac{T}{273.2} \frac{1}{N} X t_D{}^{y}$$

Equation 1.15

where X and y are correction factors for the electric field applied in the TWIMS device and its non-uniform and non-linear existence, respectively. Furthermore, the reduced cross-section (Ω') value which is independent of mass and charge can be expressed by dividing the above equation by reduced mass $\left(\left[\frac{1}{m_i} + \frac{1}{m_n}\right]\right)$ and charge (ze):

$$\Omega' = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{1}{(k_B T)^{\frac{1}{2}}} \frac{760}{P} \frac{T}{273.2} \frac{1}{N} X t_D{}^{y}$$

Equation 1.16

After the correction for charge and reduced mass along with the correction factors, a series of constant factors remain and the equation can now be further simplified by uniting these constants as one new constant (X'):

$$\Omega' = X' t_D{}^{y}$$

Equation 1.17

This new simplified equation is now in the form of $y = ax^b$ and can be plotted as such. By plotting the reduced cross-section (Ω') against t_D a calibration curve can be created for proteins of known CCS values from linear drift tube measurements (Ruotolo et al., 2008). From this calibration, we can then estimate the reduced cross-section for unknown species and proteins of interest; for which CCS values can then be established by multiplying by charge and reduced mass as had previously been removed from the procedure:

$$\Omega = ze\left[\frac{1}{m_i} + \frac{1}{m_n}\right]^{\frac{1}{2}} \Omega'$$

Equation 1.18

With the plethora of structural information that is now available through the protein databank (PDB), it is now routine to compare the measured CCS against a theoretically derived CCS value – if a 3D structure of your protein of interest is available.

1.6.3 Theoretical CCS estimations

The information obtained from the cross-section of an analyte of interest is often insufficient alone to derive useful structural information. As mentioned, it is now commonplace to compare experimental results to theoretical CCS projections and to analyse the extent to which the results are in agreement. Furthermore, theoretical CCS estimations are extremely useful in model generation. The ability to calculate the theoretical CCS of a potential model and compare the results with an experimental result has not only aided in oligomer extension/formation (Smith et al., 2010), but has also supplied invaluable assistance in establishing the structure of protein complexes (Ruotolo et al., 2005). Alternatively, the measured CCS of a protein can be compared directly with its theoretically determined value from the PDB in order to establish whether the ion has retained a native-like structure, or whether it has adopted an alternative structure in the gas-phase.

The most common method for generating theoretical CCS values of a system is the use of the MOBCAL program, developed by Martin F. Jarrold (Mesleh et al., 1996, Shvartsburg, 1996). The main criticism of the MOBCAL system is how computational intensive the trajectory method (TM) (Mesleh et al., 1996) can be, which is widely known to be the most accurate prediction method. Advancements in the field, however, have led to the generation of other programs to calculate the CCS of proteins in a fraction of the time taken by MOBCAL. The recently developed IMPACT software (Marklund et al., 2015) has the ability to calculate the CCS of a submitted structure in a matter of seconds. As well as the IMPACT software, there has also been a recent modified projection approximation algorithm reported that has results comparable to the accuracy of the trajectory method, yet is much less computationally demanding (Benesch and Ruotolo, 2011). Briefly, these programs calculate the 2D projection of a submitted structure as a snapshot at one angle. The structure is then rotated around various different angles in order to build up a rotationally averaged collision cross-section, as it is accepted that the ions tumble randomly through the mass spectrometer. The different methods that contribute to accuracy and computing time all use this premise except that they differ on how they treat the gas molecules.

Chapter 2:

Introduction II – Structural biology and the role of mass spectrometry in understanding protein complex assembly

2 Introduction II: Structural biology and the role of mass spectrometry in understanding protein complex assembly

2.1 Protein folding, misfolding and aggregation

In order for a protein to function, a protein must fold correctly to adopt its correct three-dimensional structure. Since it was discovered by Anfinsen in the 1960's that ribonuclease A can refold after denaturation into a thermodynamically stable and functional state (Anfinsen et al., 1961), the question arose of how proteins are able to do this on a biologically relevant timescale? For a protein to randomly sample all possible conformations before adopting its native fold would take an immense amount of time, even for small proteins. Therefore, it was concluded that the information directing the folding of a protein must be contained within a protein's primary amino acid sequence (Levinthal, 1969). This paradox, of how proteins fold on a biologically relevant timescale, was first proposed by Levinthal, who suggested that specific and controlled mechanisms exist, in the forms of folding pathways, to reduce the time it takes for a protein to achieve its native fold. Through the rapid formation of local interactions, leading to a stable core or nucleus, the number of total conformations that need to be sampled decreases, aiding the correct minimum energy structure to be reached on an appropriate timescale.

This folding process is often interpreted as an energy landscape (Dill and Chan, 1997). The energy landscape presents a series of structural assemblies that may be sampled en route to a protein reaching its native fold, transitioning from a high-energy and unfolded state to its low-energy native conformation. These energy landscapes are plotted as the internal free energy of a protein conformation versus the potential conformational space, or entropy. In an idealised world, the amount of internal free-energy decreases linearly with the amount of conformational space available to the system until the native fold is reached, in this case, with rapid folding kinetics (**Figure 2.1**).

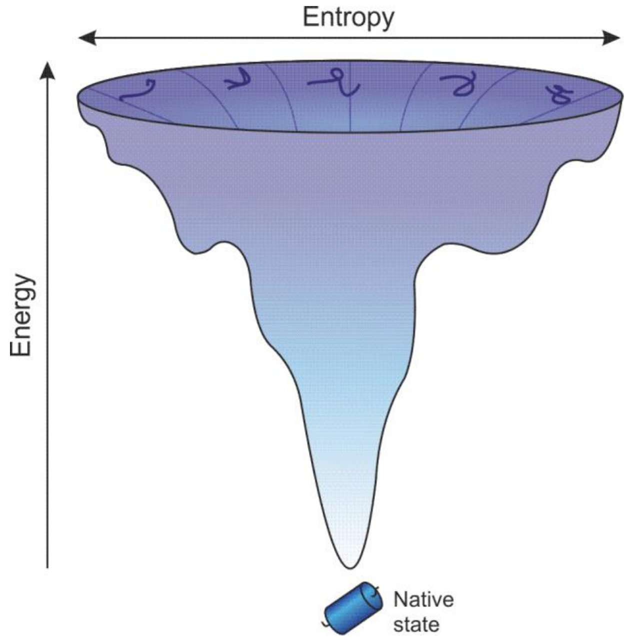


Figure 2.1: Idealised folding landscape. The internal free energy is represented on the vertical axis with the entropy represented by the width of the folding landscape. Figure redrawn and adapted from (Bartlett and Radford, 2009).

In reality however, the energy landscapes that proteins must traverse are very different (**Figure 2.2**) (Vendruscolo et al., 2003, Brockwell and Radford, 2007, Jahn and Radford, 2008). There are many intermediate states, en route to the native state, which can adopted where intra- and inter-molecular interactions can occur. These intermediate states give rise to a 'rough' energy landscape, where there are various high energy barriers with local low energy troughs a protein must navigate through to reach its conformation. Folding proteins are constantly in exchange between these intermediate states (Jahn and Radford, 2008), with small fluctuation's in a proteins conformation occurring in, and around, the minimal energy native fold, as well as intermediate folds. These intermediate states can be seen as 'on-pathway' conformations, forming local, intramolecular interactions that aid the folding into the native conformation. However as **Figure 2.2** shows, there are also low energy minima that can be reached in a misfolded conformation. The event in which a protein adopts non-native low-energy conformations forms the basis for protein misfolding; where significant reorganisation of the protein is required in order to reach the native state (Kim et al., 2013).

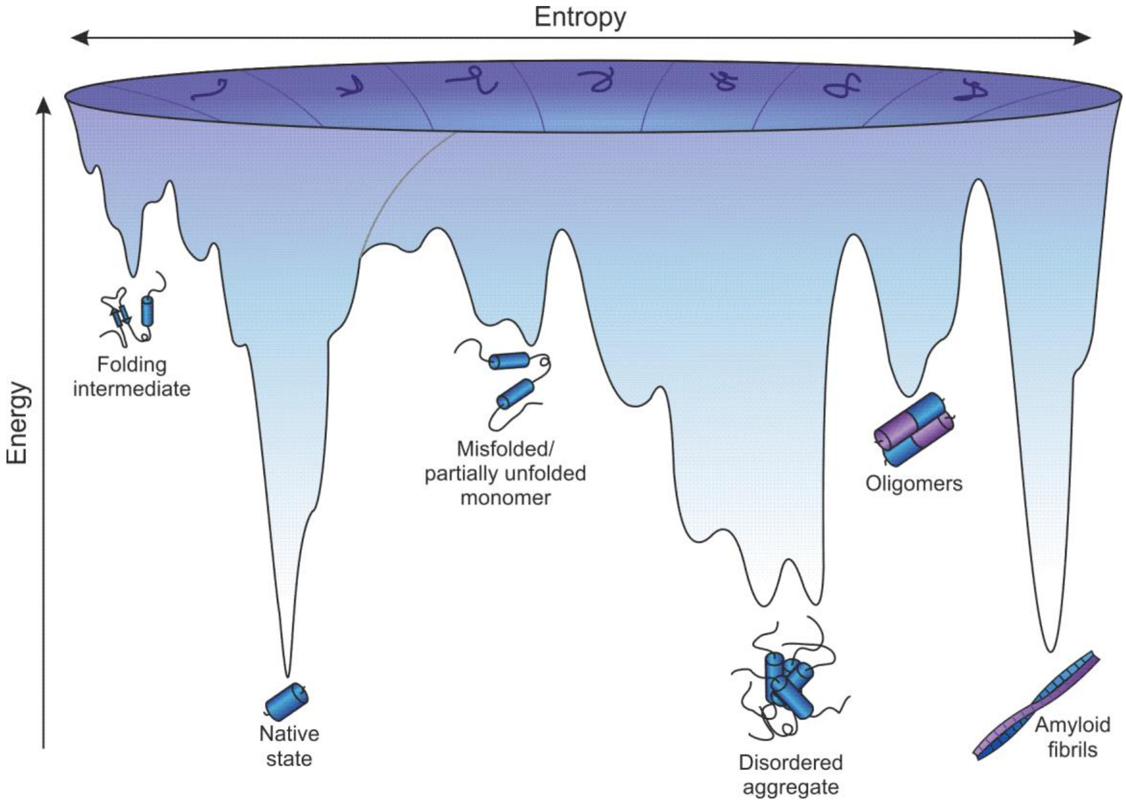


Figure 2.2: Schematic of protein folding energy landscape showing protein folding and aggregation. Unfolded polypeptide chains (blue lines) fold towards the low energy state by forming intra-molecular contacts. By contrast, inter-molecular contacts can lead to oligomers and aggregates being formed. Figure redrawn and adapted from (Jahn and Radford, 2008).

Various properties can increase the potential of a protein to become trapped in non-native or aggregation prone conformations. Destabilising factors in the proteins amino acid sequence, such as mutations of key residues, or even changes in the cellular environment *in vivo* can result in an increased propensity for proteins to adopt these intermediate species (Knowles et al., 2014). In these misfolded states, the hydrophobic core of the protein may be perturbed to an extent where hydrophobic residues are exposed and are able to form incorrect intra-molecular contacts, or in worse case scenarios, interact with other misfolded monomers and result in the formation of aggregates.

As well as misfolded and partially unfolded proteins leading to aggregation, small fluctuations in the native state can lead to regions of a protein becoming exposed that have the potential to drive aggregation (Dobson, 2003) (**Figure 2.3**). Therefore, tight regulation of the protein folding process is required. However, while there are chaperones present *in vivo* to assist the folding process and proteolytic systems in place to remove any misfolded species (Goldberg, 2003, Ellgaard and Helenius, 2003, Hingorani and Gierasch, 2014, Balchin et al., 2016, Gruebele et al., 2016), over-expression of misfolded proteins overload these systems, increasing the likelihood of protein misfolding and aggregation regardless.

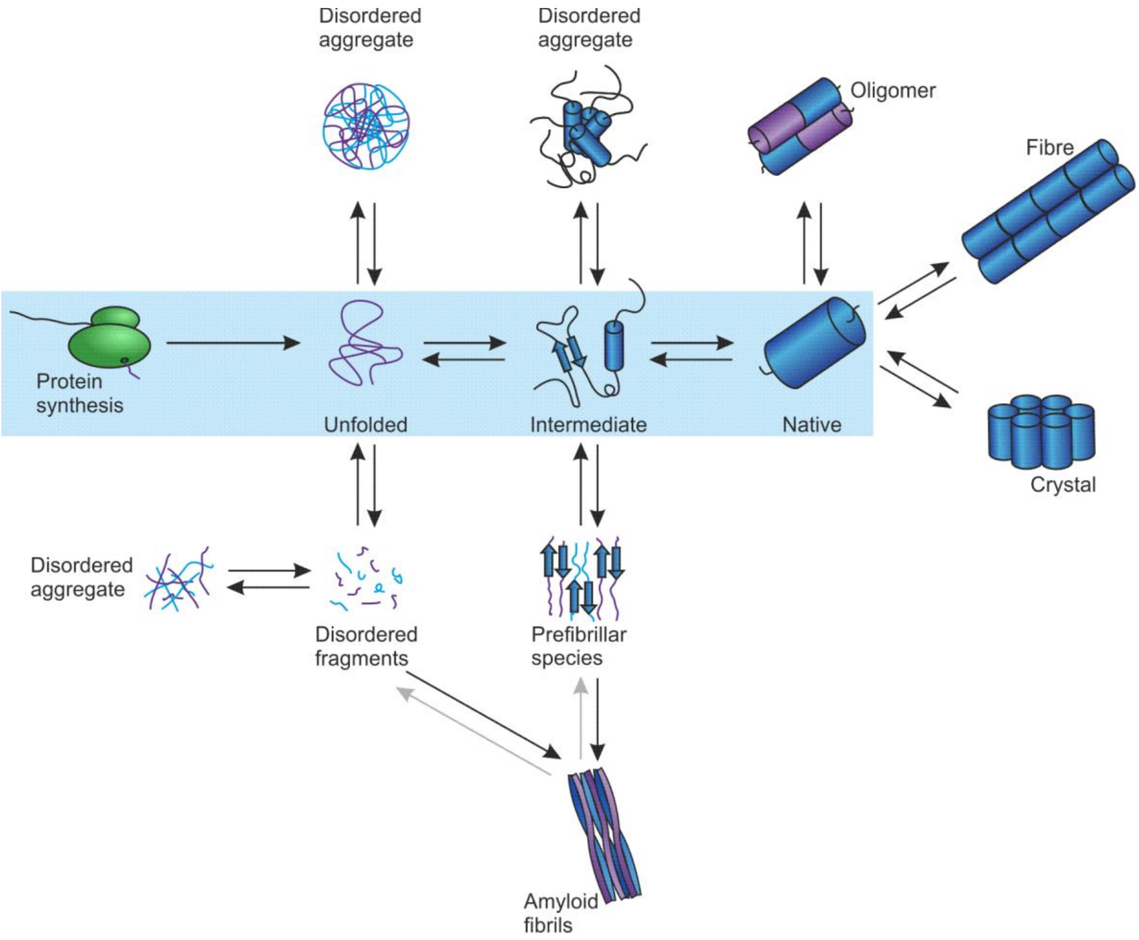


Figure 2.3: Schematic showing the protein folding pathway and the oligomers that can be formed en route to the native state. Figure redrawn and adapted from (Dobson, 2003).

Thus, it is imperative that techniques are available which are able to identify even small, transient fluctuations in a protein's conformation. Furthermore, it is important to be able to characterise these conformations once their presence is identified. Mass spectrometry is ideally suited to such analyses, able to determine protein conformations even in complex mixtures of species with a plethora of orthogonal techniques available.

2.2 Mass spectrometry in structural biology

The gold standard techniques for elucidating the structure of proteins and protein complexes are X-ray crystallography and NMR spectroscopy, yielding high resolution structures in atomic detail. Recently, advances in cryo-electron microscopy (EM) have resulted in cryo-EM yielding structures that are approaching atomic resolution (currently 1.8 Å (Merk et al., 2016)), close to that of crystallography, which has achieved resolutions of around 1 – 1.4 Å for relatively large proteins (~140 kDa) (Lubkowski et al., 2003) and even better (< 1 Å) for smaller proteins (10-15 kDa) (Antonyuk et al., 2011). However, the techniques above have limits to the systems they can analyse; crystallography can often encounter trouble when working with dynamic samples, while NMR has an upper limit on the size of proteins that can be studied (around 40-50 kDa for *de novo* structure determination). In addition, both techniques often require large amounts of protein and data analysis can be expensive in terms of time, although this latter point has been ameliorated to some extent by the development of software automation (Wishart, 2013). While not at the same resolving power to crystallography, NMR or EM, MS and its applications can reveal important information on the structure of a protein/protein complex (**Figure 2.4**). This has been expertly utilised in determining the mode of assembly in systems such as virus capsids (Uetrecht et al., 2011a, Shepherd et al., 2013) and amyloid oligomer formation (Smith et al., 2010, Young et al., 2014a, Woods et al., 2013), identifying novel drug targets (Hofstadler and Sannes-Lowery, 2006, Deng and Sanyal, 2006, Luchini et al., 2014), epitope mapping on the surface of antibodies (Yan et al., 2014a, Zhang et al., 2014) as well as identifying sites of aggregation (Ashcroft, 2010, Wang et al., 2012, Bronsoms and Trejo, 2015).

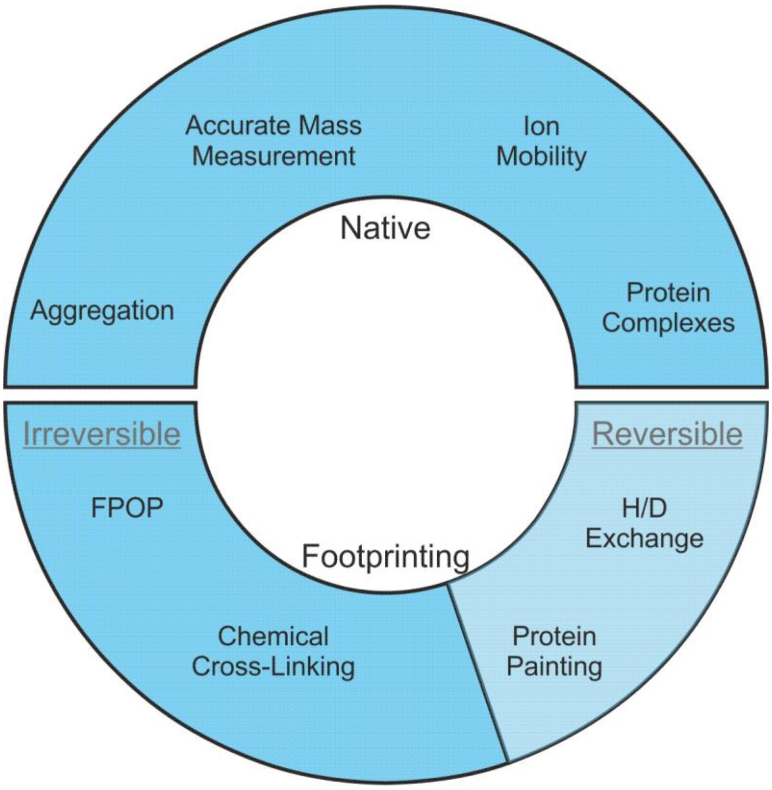


Figure 2.4: Schematic showing the different applications of mass spectrometry for analysis of protein conformation. One half of the schematic shows the application of native MS while the other shows the application footprinting methodologies.

2.3 Characterising protein structure and conformation using mass spectrometry

2.3.1 Charge state distribution is indicative of conformation

nESI-MS analyses have shown that the initial structure or conformation of a protein, folded or unfolded, can be related to the amount of charges that it acquires during the ionisation process. It has been well established that there is a clear link between the amount of charges that a protein acquired during the electrospray process, versus the solvent exposed surface area (SASA) (Chowdhury et al., 1990). More sites for protonation become available as the solvent accessible surface area of a protein increases (**Figure 2.5**) (Chowdhury et al., 1990, Kaltashov and Mohimen, 2005, Heck, 2008).

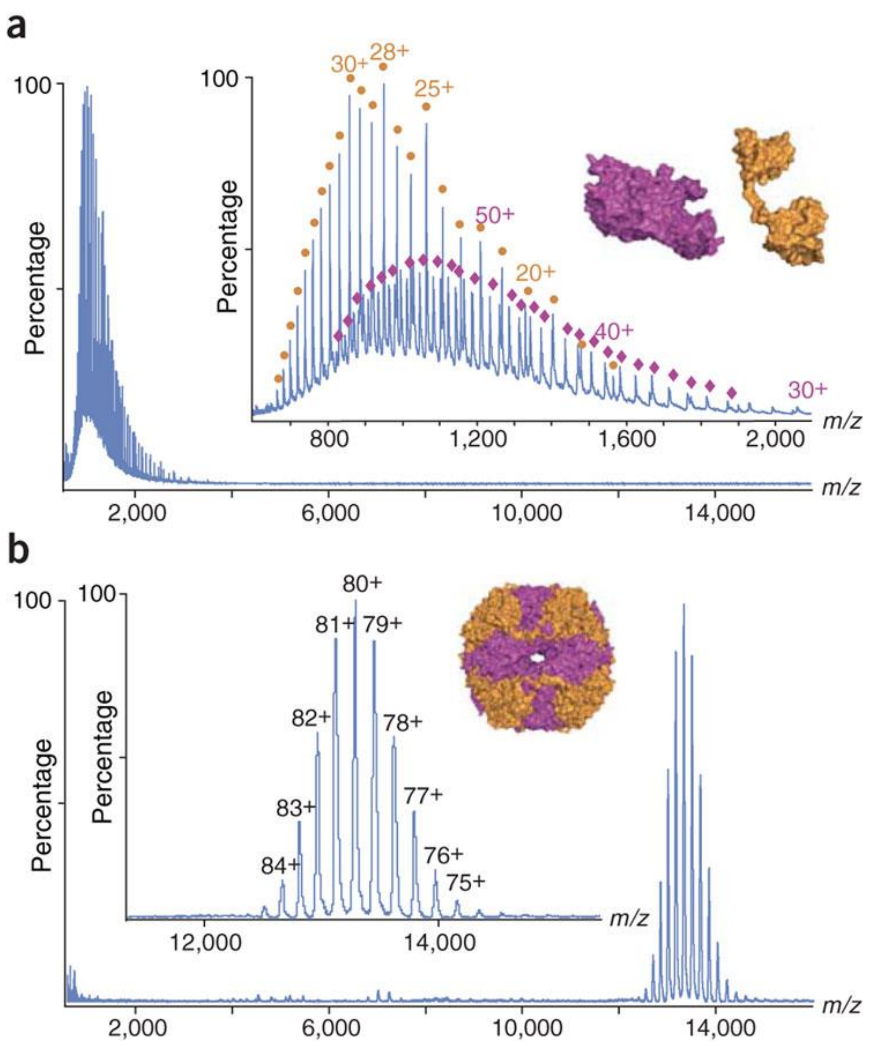


Figure 2.5: Figure highlighting the charge state distribution of a protein in a denatured and native state. **(a)** Urease was acid denatured in 50 % (v/v) aqueous acetonitrile containing 0.1 % (v/v) formic acid revealing a highly charged distribution of the different subunits of the protein. Orange = α subunit and magenta = β subunit. **(b)** Spectrum showing the same protein under native conditions showing a narrow charge state distribution. Figure taken from (Heck, 2008).

In a stable and folded state, the extent of protonation on a protein's surface is limited due to the burial of amino acid side chains. This results in a protein having a narrow charge state distribution at low charge. Conversely the same protein, once denatured, acquires a greater amount of charges during ionisation and populates a larger, highly charged distribution found at relatively low *m/z*. This transpires since as the SASA of a protein increases, during unfolding for example, the amount of sites for protonation also increase and higher charge states become dominant. A paradigm example of defining a protein's conformation by the observed charge state distribution was the analysis of β2-microglobulin, where three conformers are observed across different pH; and interestingly, it was known that the partially folded state is amyloidogenic (Borysik et al., 2004).

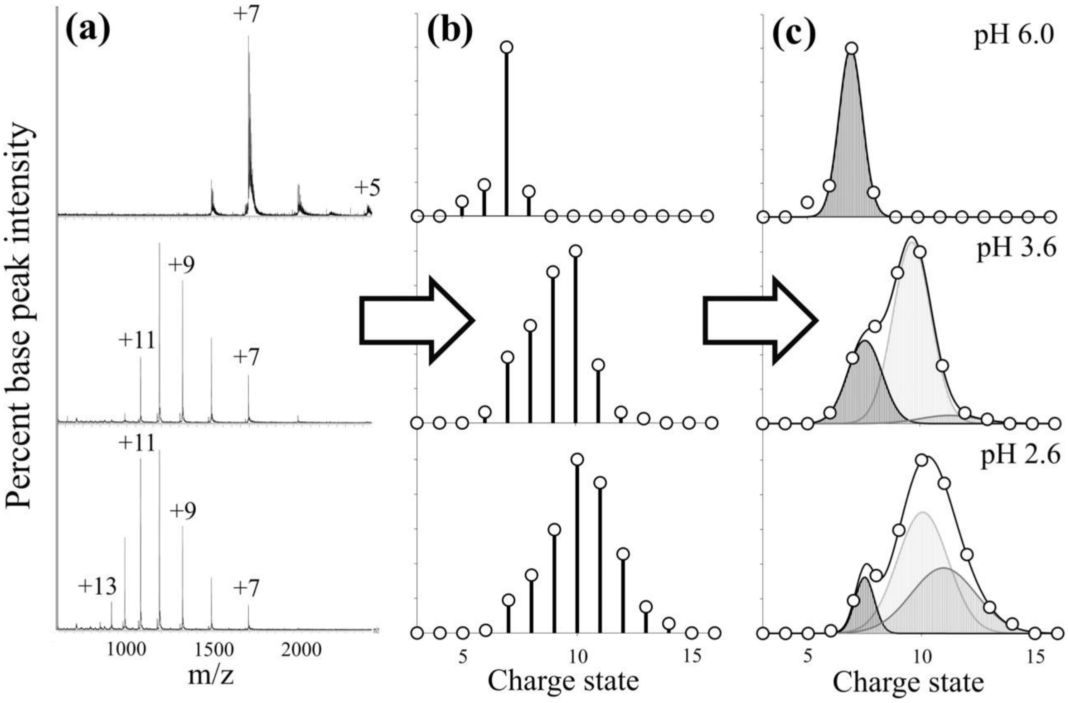


Figure 2.6: Figure showing the identification of different conformers by charge state distribution. **(a)** Spectra of β2m acquired at different pH values. pH 6 (top) where amyloid propensity is low, pH 3.6 (middle) where moderate amyloid propensity is observed and pH 2.6 (bottom) where the amyloid propensity is high. **(b)** Signal intensity is plotted against charge state. **(c)** Fitting of Gaussian distributions to the observed data showing that as pH decreases, a greater proportion of unfolded states are observed. Taken from (Borysik et al., 2004).

This observation of multiple conformations for a simple polypeptide chain has also been observed without the requirement of changing pH. Studies have shown that intrinsically disordered proteins are able to adopt multiple conformations under native-like conditions with multiple charge state distributions being observed upon ionisation (Frimpong et al., 2010). Furthermore along with IMS-MS, a framework has been established using the charge state distribution and CCSs populated for a given sample in order to elucidate structural information for an unknown protein, which again can separate intrinsically disordered and ordered proteins (Beveridge et al., 2014).

2.3.2 Protein and protein complex conformations defined using IMS-MS

The combination of MS with IMS has aided the understanding of protein conformation in the gas-phase. Since the development of ion mobility devices in commercially available instruments, the amount of publications using IMS-MS has increased dramatically (Lanucara et al., 2014). Initial studies with IMS-MS showed that a protein's native conformation was believed to be retained upon transition into the gas-phase by measuring CCS values that were in agreement with structure predictions, as outlined in **Section 1.6.3**. In general, the CCS of native proteins increases with charge, which has been attributed to Coulombic repulsion forces (Wu et al., 1998, Tolmachev et al., 2009). These initial studies propelled further studies using IMS-MS to understand protein conformation and complex assembly.

Using B2m as an example, IMS-MS has been used to show that the species involved in the oligomeric states are elongated in nature, rather than globular, defined by their CCS values (Smith et al., 2010). This kind of information is vital in understanding how amyloidogenic proteins, such as β2m, aggregate which will be pivotal in designing therapeutics which can target the oligomeric species responsible for driving aggregate formation. Finally as mentioned above, IMS-MS has been crucial in determining protein complex formation. An excellent example of this was determining virus capsid assembly using IMS-MS (**Figure 2.7**) (Uetrecht et al., 2011a, Shepherd et al., 2013). IMS-MS and CCS determination was used to show that the hexamer was an on-pathway intermediate to whole capsid assembly, formed by a nucleus exhibiting a five-fold symmetry. As mentioned with β2m, achieving a greater understanding of how the proteins involved in virus capsids assemble will aid the generation of future therapeutics that can target these assemblies.

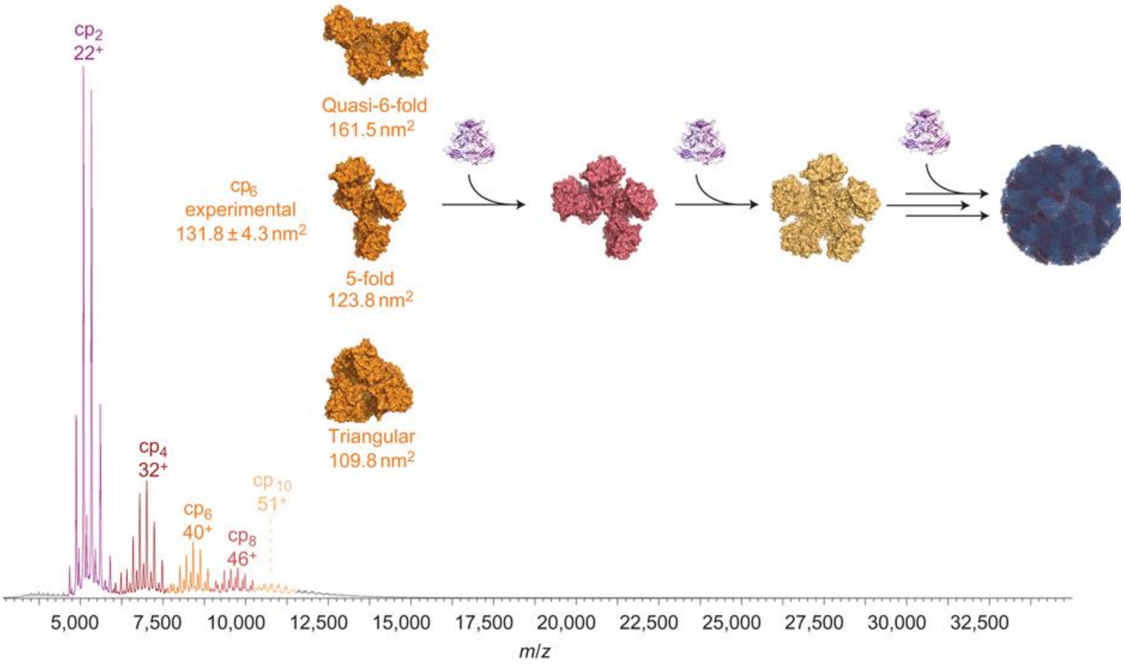


Figure 2.7: Figure showing the spectrum of a norovirus capsid under conditions which do not permit assembly. The inset shows the proposed model of the assembly pathway, indicating that the experimentally observed hexamer is an on-pathway intermediate. Figure adapted from (Uetrecht et al., 2011a).

2.3.3 Mass spectrometry of membrane proteins and their complexes

While membrane proteins are responsible for a wide range of biological process, as well as well as making up 60 % of drug targets (Lappano and Maggiolini, 2011, Dorsam and Gutkind, 2007, Arinaminpathy et al., 2009), there is significantly less structural information available for this class of proteins, only 2.9 % of all protein structures in the PDB. One of the causes for this is due to the traditional structural techniques such as X-ray crystallography and NMR requiring large amounts of soluble protein, which is often not possible with membrane proteins. However, as native MS is a rapid and sensitive technique it can provide critical details on membrane protein complexes, as well as documenting and bound lipids and molecules. As the analysis of membrane proteins by MS had previously been frustrated by the proteins being kept in detergent-containing buffers. Advances in purifications such as detergent micelles (Barrera et al., 2009, Barrera and Robinson, 2011, Konijnenberg et al., 2015), nanodiscs (Marty et al., 2012, Hopper et al., 2013) and amphipols (Leney et al., 2012, Calabrese et al., 2015b, Watkinson et al., 2015) have led to exciting discoveries of membrane protein structures and conformations by MS and IMS-MS.

2.3.4 Probing solution conformation with footprinting technologies

While native MS and IMS-MS yield information that relates to the conformation of proteins in the gas-phase, other techniques can be used in conjunction with MS that reveal information on protein conformation and dynamics in solution. Protein footprinting technologies are capable of examining conformational change and ligand binding in proteins by monitoring their response to chemical or enzymatic modifications. Protein footprinting as a complimentary technique to MS has evolved rapidly in the last decade (Kaltashov et al., 2013). Footprinting technologies include amide hydrogen-deuterium exchange (HDX) (Konermann et al., 2011, Pacholarz et al., 2012), hydroxyl-radical based labelling (otherwise known as fast photochemical oxidation of proteins (FPOP)) (Sclavi et al., 1998, Calabrese et al., 2015a), protein painting (Luchini et al., 2014), chemical labelling (Mendoza and Vachet, 2009, Zhou and Vachet, 2012) and chemical cross-linking (XL) (Sinz, 2006, Herzog et al., 2012).

2.3.4.1 Hydrogen-deuterium exchange (HDX)

HDX involves the exchange of hydrogen with deuterium (or vice versa) by a protein upon exposure to a deuterated buffer or solution. Due to the 1 Da mass difference of deuterium compared with hydrogen, HDX can be monitored in the resulting mass spectra, as a mass shift is observed as more exchange occurs. There are three types of exchangeable hydrogens on a protein: the hydrogens on side chain carbons, hydrogens bonded with heteroatoms and hydrogens found on backbone amides (Cao et al., 2013). The latter exchange is a reversible process and occurs on time-scales measurable using real time experiments using ESI-MS (Englander et al., 1996). Combined with liquid chromatography (LC) and MS/MS, region-specific information can be obtained based on the deuterium uptake of peptides. This can be used for the characterisation of protein-protein and protein-ligand interactions (Pirrone et al., 2015), epitope mapping (Chen, 2013) or even identifying regions of aggregation (Zhang et al., 2013) by searching for regions of protection from exchange upon oligomerisation. The example below shows the use of HDX exchange to define the epitope of factor H binding protein (fHbp) (**Figure 2.8**) which is a key virulence factor and vaccine antigen in *Neisseria meningitidis*. These experiments detected regions involved in binding that were missed using other epitope mapping techniques (Malito et al., 2013).

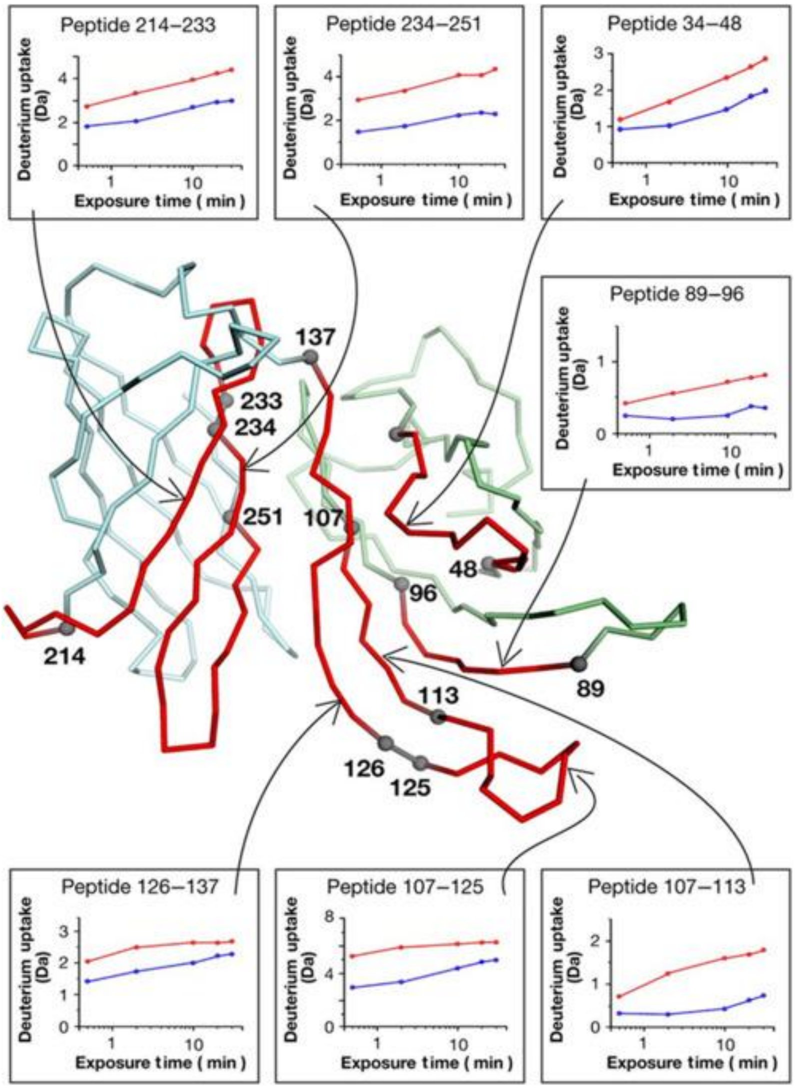


Figure 2.8: The results from epitope mapping of factor H binding protein (fHbp) by HDX-MS. The boxes show the deuterium uptake over 30 min in the absence (red) or presence (blue) of the binding mAb. The regions of protection are highlighted red on the structure. Figure taken from (Malito et al., 2013).

One of the major drawbacks to HDX, however, is the problem with back exchange. After deuterium exchange has occurred, the protein has to be held at low temperatures and pH to minimise any back exchange with hydrogen. While an extremely informative technique, this prohibits any further downstream treatment or modification of the protein as proteolysis requires that the sample is immediately taken for LC and MS/MS analysis. However, HDX is capable of monitoring protein dynamics and identifying rare conformations that may be missed by other techniques. Additionally, as mentioned in **Section 2.1**, while these rare conformations may be in a state of flux with native of partially native folds, it can be these rare states that are responsible for driving unwanted protein aggregation.

2.3.4.2 Fast photochemical oxidation of proteins (FPOP)

In this footprinting technology, highly reactive hydroxyl radicals (·OH) are generated from hydrogen peroxide, via laser photolysis, which react with amino acid side chains irreversibly (Xu and Chance, 2007). This approach generates ·OH in a few nanoseconds, with a laser pulse (248 nm) to a flowing solution which contains the protein sample of interest with a small amount of hydrogen peroxide present in the sample (typically 10-20 mM) (Zhang et al., 2011). The flowing solution ensures that the proteins are only exposed to the laser pulse once in the experiment. To control the labelling and limit the lifetime of the ·OH, radical scavengers are added into the solution in the form of free amino acids (typically glutamine or histidine). A schematic of the experimental set up can be seen below (**Figure 2.9**).

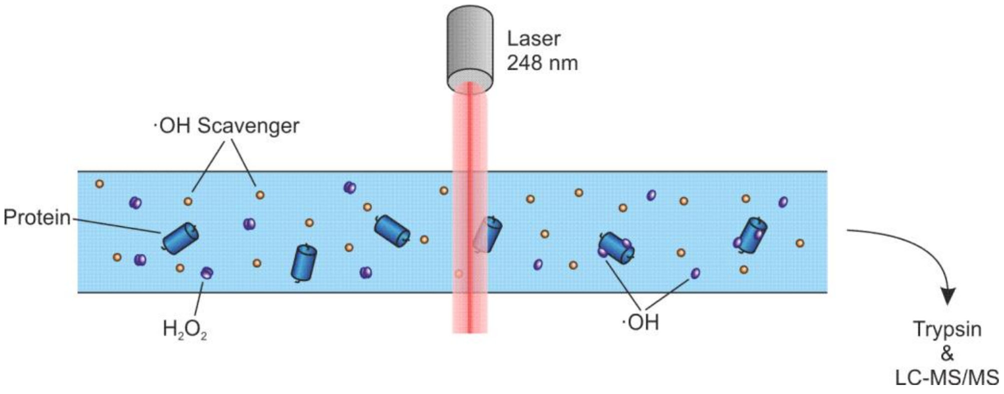


Figure 2.9: Schematic of the FPOP workflow. A solution containing the protein of interest, hydrogen peroxide (purple ovals) and a hydroxyl radical scavenger (orange spheres) are

flowed through a capillary that is pulse by a laser at 248 nm. This creates hydroxyl radicals (·OH) which are able to modify the side chains of solvent exposed amino acids.

As well as probing the surface of proteins (Zhang et al., 2010) and epitope mapping (Yan et al., 2014b), FPOP has been used to follow protein folding in real time (Chen et al., 2012). By initiating protein folding (of a cold denatured protein) with a temperature jump induced by a laser pulse, and then labelling at different time points after initiation, the formation of the hydrophobic core was monitored before the rest of the protein found its native conformation. However, a pitfall of the technique is the amount of accessible sites for modification. The reactivity of side chains is not only dictated by their solvent exposure, but also their chemical nature. The most reactive residues are the sulfur containing amino acids cysteine and methionine, followed by the aromatic side chains: tryptophan, tyrosine and phenylalanine. Other potential sides of modification include histidine, leucine, isoleucine, arginine, lysine, valine, proline, glutamine and glutamic acid (Takamoto and Chance, 2006, Xu and Chance, 2007, Konermann et al., 2008). Due to the lack of 100 % amino acid coverage, there is a chance interfaces may be missed.

2.3.4.3 Protein painting

Protein painting is one of the newer footprinting methodologies, being first published in 2014 (Luchini et al., 2014). This method uses small, organic dyes or 'paints' which bind to the surface of proteins, presumably by binding to lysine and arginine residues. The power of this technique is that the small molecule paints remain bound to the protein's surface in denaturing conditions. Thus, by taking a pair of interacting proteins, such as interleukin 1-beta (IL1β) and its receptor 1 (IL1), and allowing them to form the complex before adding the small molecule paints; the regions of the proteins responsible for the interaction are 'paint' free. Upon dissociation and denaturation, residues in the interaction interface can be cleaved with trypsin while the rest of the protein is protected from the proteolytic enzyme by the dyes (**Figure 2.10**). As only the regions involved in binding are digested by trypsin, upon LC and MS/MS analysis of the digested peptides, the regions protected from paint binding can be identified. While a promising technique further studies will be needed to determine its generic applicability to identifying and characterising protein-protein interactions and the conformations of protein complexes.

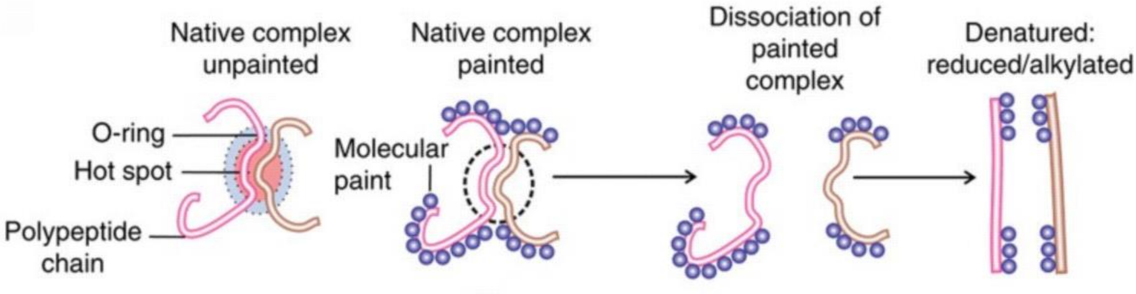


Figure 2.10: Schematic showing the protein painting workflow. Protein complexes are coated with small molecule 'paints'. The region of interaction is not accessible to the paints and therefore is protected from coating. Upon dissociation and denaturation of the complex only the previously protected sites are exposed for digestion by trypsin and will subsequently be analysed by MS. Figure adapted from (Luchini et al., 2014).

2.3.4.4 Chemical labelling

A plethora of chemical reagents are available that can be used to covalently modify the surface of a protein at specific amino acid side chains (Mendoza and Vachet, 2009). Examples include glycyl ethyl ester which reacts with carboxyl groups on a proteins surface (Liu et al., 2014), the use of iodoacetamide or N-ethylmalemides to modify the thiol group of cysteine residues (Su et al., 2011) and diethylpyrocarbonate (DEPC) which can be used to modify histidine, lysine, tyrosine, serine, threonine and cysteine residues, as well as the N-terminus (Borotto et al., 2015).

This method of labelling proteins is similar to that of FPOP, where solvent exposed amino acid side chains are selectively modified depending on the chemical probe being used. The advantage of this approach compared with FPOP is that a combination of probes can be used to modify a protein to increase coverage and it does not require a laser or expensive experimental set up. However, the disadvantage is that the reactions occur at much slower timescales compared with FPOP and therefore any transiently exposed sites may be missed.

2.3.4.5 Chemical cross-linking

Chemical cross-linking can be used to covalently link two interacting partners; with peptides from the interacting partners being sequenced by LC-MS/MS downstream (Sinz, 2006). Similar to the chemical labelling approach, the chemical cross-linkers are used to target specific surface-exposed residues on the proteins surface, most commonly lysine residues by the use of N-hydroxysuccinimide ester chemistry. However, there are cross-linkers which are also capable of reacting with carboxyl groups and non-specific cross-linkers which use diazirines (free radicals) as the reactive group (Dubinsky et al., 2012) (**Figure 2.11**)

As each of the chemical cross-linkers have set distances between their reactive groups, these distances can be used as restraints to drive modelling of the interacting proteins. This can be beneficial when little is known about a protein complex and its assembly. A paradigm example of this was the structure determination of the 1.5 MDa complex, RNA polymerase II transcription pre-initiation complex (PIC) by XL-MS and cryo-EM (Murakami et al., 2013).

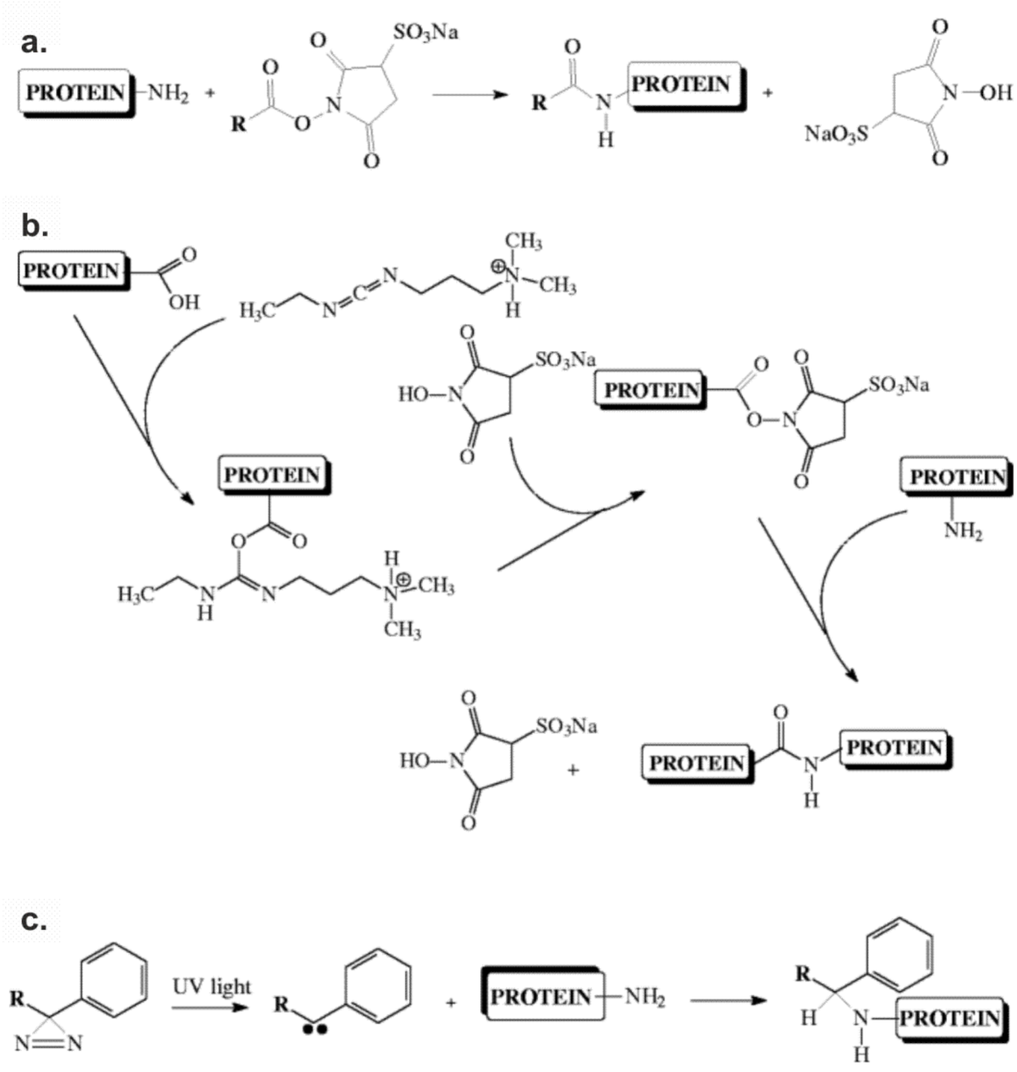


Figure 2.11: Reaction mechanisms for different types of cross-linkers. The reaction mechanisms for **(a)** NHS ester (amine reactive), **(b)** zero-length carbodiimide (carboxylic acid reactive) in conjunction with sulfo-NHS (amine reactive) and **(c)** diazerine (photoreactive) cross-linkers are shown. Figure adapted from (Sinz, 2010).

An advantage of chemical cross-linking is that the restraints achieved from cross-linking experiments allow for the validation of generated models. Furthermore, as with some of the other techniques, since cross-linking involves a covalent modification, a sample does not need to be handled as carefully post cross-linking compared to HDX. A caveat of chemical cross-linking is that the cross-linkers target specific residues. There is the potential for false-negatives if there are no cross-linkable residues near the binding site or sites of interaction.

The structural MS toolbox has evolved rapidly in the past two decades with the development of these various labelling techniques. However, choosing which methodology to use depends on the system being studied as each techniques has its own advantages and disadvantages (**Table 2.1**)

Table 2.1: Summary of the different footprinting techniques. Table adapted from (Dailing et al., 2015).

	Footprinting technique			
	Chemical Cross-linking	Hydrogen-deuterium exchange	Fast photochemical oxidation of proteins	Protein painting
Experimental configuration	No extra set-up required	Optimised for deuterium retention	Optimised for UV pulse < 1 µs	No extra set-up required
pH	Neutral-basic	Acidic (pH = 2)	Neutral- slightly basic	Neutral
Temperature	4 °C to room temperature	Room temperature, 4 °C	Room temperature	Room temperature
Timescale of experiment	0.5 – 2 h	Short (min)	< 1 µs	Short (min)
Software required	Specialised software	Specialised software	Standard LC-MS/MS software	Standard MS software
Format of results	Typically trypsin digested fragments containing peptides from each protein in interaction	Pepsin digested fragments, ~10 amino acids	Typically trypsin digested fragments	Typically trypsin digested fragments

Protein conformation	Pre-formed complexes, covalently cross-linked	Pre formed complexes, deuterated	Pre formed complexes, oxidised residues	Pre formed complexes, non-covalent coating with dyes
Identification	Binding partners are identified with low specificity for interface solvent excluded binding regions	Interaction regions are identified by a small 1.0073 Da shift in peptic fragment peptide mass	Interaction regions are identified by absence of oxidization	Interaction regions are identified by presence of tryptic peptides exclusive to the site of interaction
Negative outputs	Intra-molecular cross-links as well as cyclic and dead end cross-links	Insensitive to internal protein interactions	Treatment of the protein with a laser, may induce protein unfolding	Insensitive to internal protein interactions

As structural MS has evolved to be a prevalent player in the field of structural biology, the technique can be applied to targeting issues of complex formation and aggregation in one sector of biology that structural information is paramount: biopharmaceuticals.

2.4 Biopharmaceuticals: monoclonal antibodies

Monoclonal antibody (mAb) development was introduced in 1975 by Köhler and Milstein, in which immortal, mAb secreting, mouse cells are generated through the fusion with myeloma cells that secrete a single type of antibody (Köhler and Milstein, 1975); they were later presented with the Nobel Prize in medicine and physiology in 1984 for "*theories concerning the specificity in development and control of the immune system and the discovery of the principle for production of monoclonal antibodies*". Their findings swiftly led to the generation of the first therapeutic antibody that was approved for human use (Orthoclone OKT3, muromonab) by the US Food and Drug Administration (FDA) in 1986, used for the treatment of acute rejection in organ transplant patients (Smith, 1996). Now, 30 years after the first approval of Orthoclone OKT3, there has been an explosion in the market for mAbs and their fragments (Maggon, 2007, Liu, 2014), with mAbs now ranked as the highest grossing class of biopharmaceuticals. MAbs now contribute to 50 % of the top 10 selling drugs, as of 2014 (in the US), and more than 50 % of the generated revenue. This is a striking difference when compared with the top 10 selling drugs from 2000-2011, where only one mAb was present in the top 10 selling drugs (**Figure 2.12**).

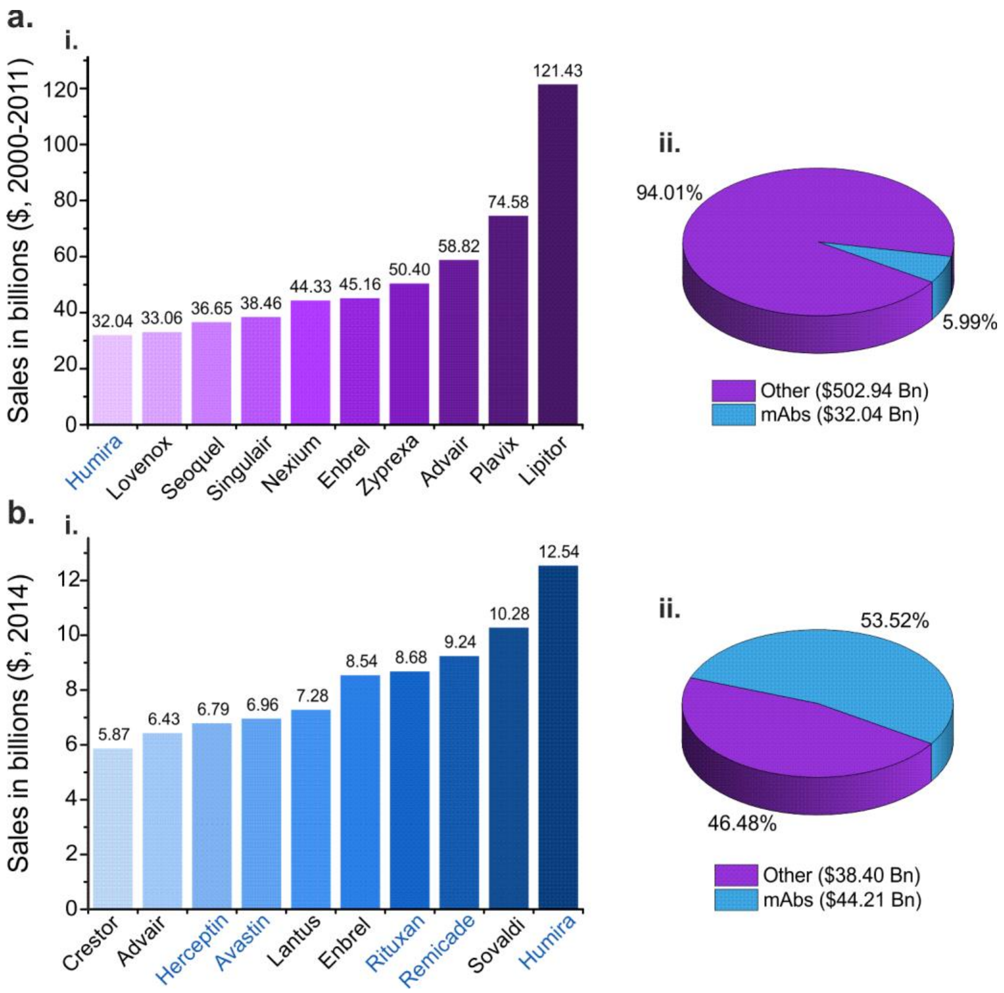


Figure 2.12: The top drug sales of the 21st century. **(a)** (i) Graph showing the top 10 selling drugs of from the year 2000-2011 with (ii) the percentage of monoclonal antibody drugs shown. **(b)** (i) Graph showing the top 10 selling drugs of 2014 with (ii) the percentage of monoclonal antibody drugs shown. Monoclonal antibody drugs = blue. Figure generated with data collated from (Philippidis, 2012, Philippidis, 2015) (www.genengnews.com) .

The success of these molecules, aside from the exquisite specificity, can be partly due to the advancements in the methodologies used for their generation and production: phage and ribosome display (Chames et al., 2009).

2.4.1 Monoclonal antibodies and their generation

Intact mAbs, or immunoglobulins (Ig), are 150 kDa plasma glycoproteins that are employed by the immune system to detect and neutralise foreign bodies (Van Dijk and Van De Winkel, 2001). Mab monomers have a characteristic 'Y' shape and consist of two pairs of polypeptide chains, connected through inter-molecular disulfide bonds: two identical heavy chains (~50 kDa) and two identical light chains (~25 kDa). There are five different classes of antibodies that exist in higher vertebrates: IgM, IgA, IgD, IgG and IgE. The different classes of antibodies have unique monomeric structures dictated by their heavy chain (Janeway et al., 1997). Out of these classes, the IgG is the most common isotype which has four different subclasses: IgG1, IgG2, IgG3 and IgG4 (Janeway et al., 1997); the most common of which is the IgG1 isoform (Van Dijk and Van De Winkel, 2001). Each chain consists of sequential immunoglobulin domains, structural subunits comprising of two β-sheets arranged in an immunoglobulin fold. The higher order structure of the mAb can be broken down into the top half, constituting of two antigen binding fragments (Fabs) 'arms' (**Figure 2.13a,b**), and the bottom half of the molecule, consisting of the crystallisable fragment (Fc) (**Figure 2.13a**). The variable regions or fragments (Fv) (**Figure 2.13a**), are the regions of the molecule responsible for recognising their cognate antigens through the complementarity-determining regions (CDR) (**Figure 2.13a, inset**). Each variable domain consists of three CDRs that make contacts with the specific antigen; the hypervariablity of these CDR loops gives rise to the tremendous diversity in antigen-recognition.

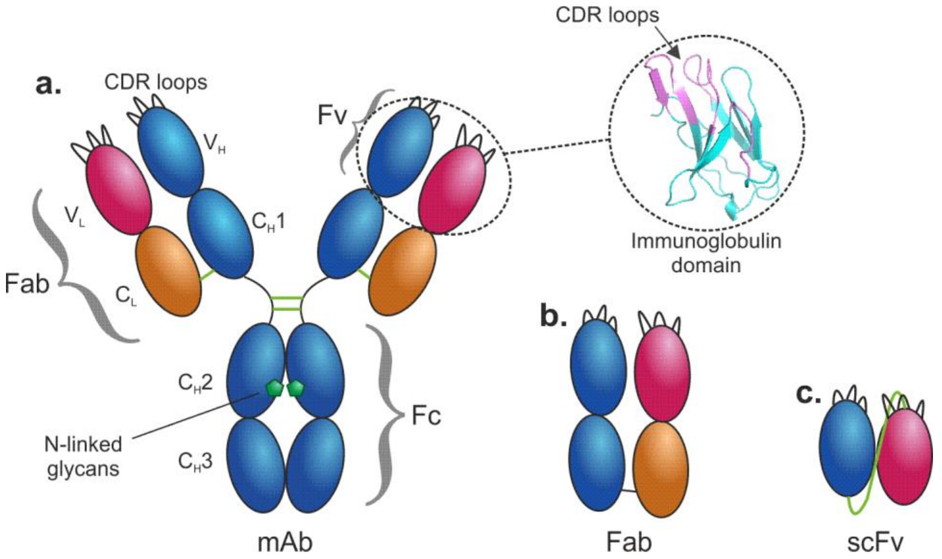


Figure 2.13: Structure of an IgG1 antibody. **(a)** Structure of an IgG1 monoclonal antibody with the different regions highlighted. Green lines = disulphide bonds. The inset shows an individual Ig domain with the CDR loops highlighted (pink). **(b)** Structure of an isolated antigen binding fragment (Fab). **(c)** Structure of a single chain variable region (scFv) where the green line = glycine/serine linker. Figure redrawn from (Saunders, 2014).

As described by Köhler and Milstein in 1975, mAbs were first generated using hybridoma technology (Köhler and Milstein, 1975). Briefly, an immortal hybridoma cell line is generated by fusing isolated spleen B-cells from immunised mice with an immortal myeloma cell line. The myeloma cells lack the HGPRT gene (hypoxanthine-guanine phosphoribosyltransferase), which makes them sensitive to aminopterin found in HAT (hypoxanthine-aminopterin-thymidine) medium. Therefore, only fused cells are able to survive as the myeloma cells cannot survive in HAT medium, while the B-cells have a naturally short life time (Nelson et al., 2000). However, as mentioned previously, this method of generating mAbs has generally been replaced by either phage or ribosome display technologies.

2.4.1.1 Phage display for the generation of mAbs

The generation of mAbs using phage (Smith, 1985) and ribosome display (Hanes and Pluckthun, 1997) methodologies is an attractive approach as it is an *in vitro* process that generates diverse libraries, 10^9-10^{11} and 10^{12}-10^{14} in size, respectively (Lu et al., Groves et al., 2014). The phage display technology has been adapted from the work of Georg P. Smith in 1985, showing that phages can display a peptide of interest, through the fusion of protein III of the M13 bacteriophage and the peptide of interest, creating a peptide::pIII protein fusion (Smith, 1985).

This principle was applied to display small mAb fragments, such as scFvs and Fabs, on the surface of phage (Carmen and Jermutus, 2002). The scFv or Fab sequence of interest is cloned into the phage plasmid between the promoter and phage coat protein to create a 'phagemid'. Phagemid libraries are used to transfect *E. coli* cultures to express the fusion proteins (Qi et al., 2012). Selected hits from the phagemid libraries can then be re-grafted back onto IgG scaffolds to generate the fully functioning mAb. A diverse library of phagemids can be used to express scFv protein fusions which can be displayed on the surface of phage (**Figure 2.14a**), allowing for 'panning' of the fusion proteins to identify sequences that bind the target antigen. Any scFvs which do not bind can be removed through wash steps (**Figure 2.14b**) before eluting the scFvs of interest (**Figure 2.14c**). To increase the affinity of the binding, the binding scFvs can be amplified in *E. coli*, through the introduction of a helper phage to enable correct phage assembly (Bazan et al., 2012) (**Figure 2.14d**). This process is then repeated another 2-3 times while altering conditions in order to increase the affinity of the scFv for its epitope, whereupon the sequence is analysed (**Figure 2.14e**).

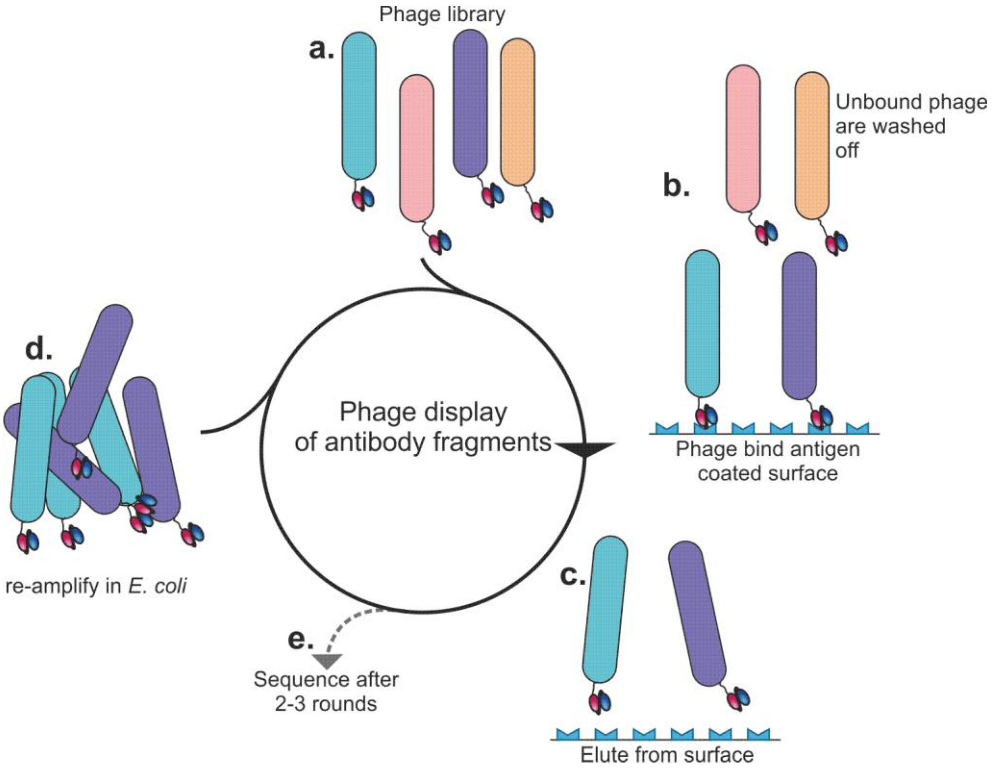


Figure 2.14: The process of scFv selection using phage display. **(a)** Phage containing phagemid library are selected and **(b)** added to the antigen being selected against, which is immobilised onto a surface. **(c)** Binding partners are eluted and **(d)** re-amplified in *E. coli* for sequential rounds of selection to pan for further antigen affinity. **(e)** Selected candidates are sequenced after 2-3 rounds. Redrawn from (Schirrmann et al., 2011).

Due to the stability of the filamentous phage used in this method, typically f1, fd or M13 bacteriophages (Bazan et al., 2012), a range of conditions including extreme pH, high and low temperatures and a range of ionic strengths can be tested to find the tightest binders (Brigati and Petrenko, 2005). The tightest binding sequences can then be transfected into *E. coli* cells to test for binding specificity through ELISA assays (Lee et al., 2007). A paradigm example of phage display for the generation of mAbs is the generation of humira (Kempeni, 1999). Humira was generated to bind to tumor necrosis factor-alpha (TNFα), inhibiting the inflammatory response and aiding the treatment of rheumatoid arthritis, psoriatic arthritis, Chron's disease, and psoriasis.

2.4.1.2 Ribosome display for the generation of mAbs

Ribosome display is another form of *in vitro* display technology for the generation of mAbs, with the advantage of larger library sizes. This methodology uses a similar 'panning' procedure, cycling through multiple rounds to increase antigen specificity to the mAb fragment being displayed. However, here the mAb fragment is displayed from the exit tunnel of a free ribosome while still attached to its cognate mRNA sequence (Hanes and Pluckthun, 1997). Starting with a pool of DNA sequences, each sequence is transcribed and translated *in vitro* to display the different polypeptide chains, typically scFvs. However, a spacer sequence is used, which lacks a stop codon, to halt the translation of the sequence while remaining attached to the peptidyl tRNA, prohibiting release factors from binding resulting in the protein of interest being displayed out of the ribosome exit tunnel, able to bind its target antigen. This presentation of the antibody fragment from the ribosome while still attached to its corresponding mRNA is called an antibody-ribosome-mRNA (ARM) complex (He and Taussig, 1997) (**Figure 2.15**).

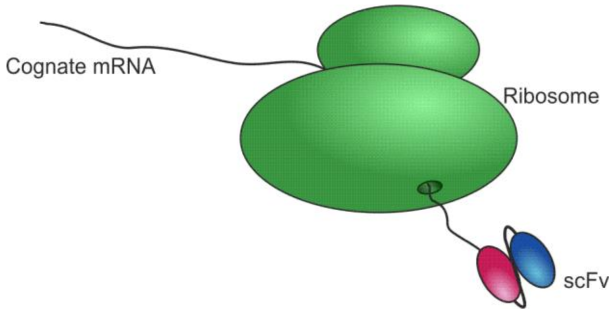


Figure 2.15: Schematic of an antibody-ribosome-mRNA complex.

After non-binding scFv sequences have been removed through a wash step, initial hits can be eluted from their antigen and the ARM complex dissociated and the mRNA-scFv complex cleaved to release the mRNA. Reverse transcription PCR is then used to generate a pool of DNA molecules which can be further mutated and sequential rounds of 'panning' carried out.

The evolution of these *in vitro* methods of selecting antibody fragments with exceptional binding affinity has led to the generation of mAbs unobtainable from conventional immunisation methods (Bradbury et al., 2011). However, one issue that is commonly encountered, which these display methods do not address, is the expression levels of selected sequences.

2.4.2 Antibody fragments and mimetics

It was shown some time ago that the full 150 kDa IgG is not required for antigen binding, and that binding can be retained upon production of the V_H and V_L domains connected through a peptide linker (Huston et al., 1988). These antibody fragments have since been denoted as single chain variable fragments (scFv) (**Figure 2.13**). The idea of using scFvs as therapeutics was an attractive approach as they are smaller in size (~25 kDa) than IgGs and can be expressed and purified in *E. coli* (Guglielmi and Martineau, 2009). This led to the interrogation of various antibody fragments, such as Fabs, scFvs and single domain antibodies (dAbs) (Ward et al., 1989) (**Figure 2.16b**), as potential therapeutic candidates. These fragments of antibodies can be modified to retain the parent Ig characteristics such as affinity, immunogenicity and also effector function (Nelson and Reichert, 2009). Furthermore, advancements within the last decade have led to the generation of small polypeptide antibody mimetics as therapeutic agents (**Figure 2.16a**) (Qiu et al., 2007, Tiede et al., 2014). These mimetics adopt a similar binding mechanism to that of mAbs, using randomised amino acid loop regions to identify the target antigen while being held in place by a scaffold protein.

However, the use of antibody fragments as therapeutic candidates have their own pros and cons. One of the major advantages to antibody fragments and mimetics are their size. Due to their relatively low molecular weight, they are able to infiltrate further into tissues and tumours compared with their mAb counterparts (Jain, 1990, Yokota et al., 1992). Furthermore, it has been hypothesised that due to their smaller size, a larger repertoire of epitopes are presented to antibody fragments with the possibility of binding to ‘cryptic’ epitopes readily available to fragments that would be un-achievable for full length mAbs. Indeed, single domain antibodies were shown to be able to act as potent enzyme inhibitors.(Lauwereys et al., 1998). Finally, antibody fragments are ideal as most can be produced in microbial systems as these fragments lack the glycosylation sites of full mAbs.

This therefore decreases the cost and timescale of their production compared with conventional mammalian expression systems.

The disadvantages of using small antibody fragments arise in their short circulating half-life *in vivo*, due to the lack of an Fc region and therefore antibody fragments are not shielded for filtering by the Fc mediated recycling system (Mitragotri et al., 2014). Although it is worth mentioning that various Fc fusion proteins (**Figure 2.16b**) have been developed that bridge the gap of utilising small fragment therapeutics, while retaining *in vivo* Fc derived recycling, such as Enbrel, which also treats autoimmune diseases by interfering with TNF and was the fifth top selling drug in 2014 (**Figure 2.12**) (Peppel et al., 1991). It has also been shown that the aggregation propensity of mAbs and their different fragments may vary widely (Bird et al., 1988) and antibody fragments may be less stable than full length mAbs; this therefore requires rigorous analysis of the fragments to ensure their stability as monomers. Finally, as full mAbs contain the Fc region they elicit the necessary downstream immune response and effector functions, such as complement-dependent cytotoxicity or antibody-dependant cell-mediated cytotoxicity (Hansel et al., 2010). This is, however, also elicited with the use of Fc fusion proteins. Thus, antibody fragments can function by either competing for the binding site of a receptor or ligand, simply blocking the action of the molecules, inhibiting a downstream response.

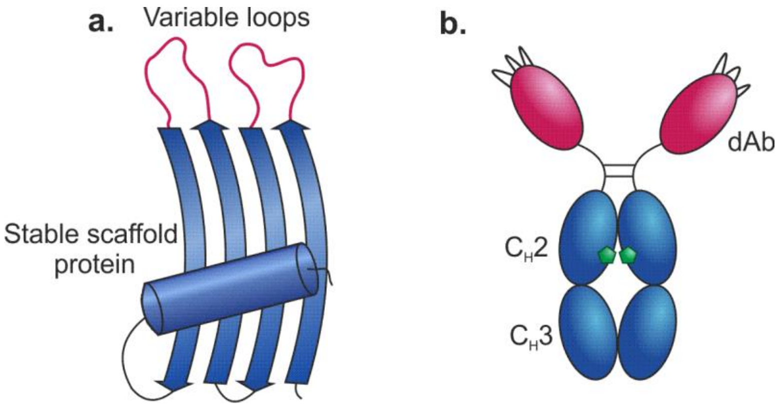


Figure 2.16: Examples of antibody mimetics and antibody fragments. **(a)** Schematic of an antibody mimetic with the two variable loops highlighted (pink). **(b)** Schematic of an antibody fragment, dAb-Fc-fusion protein.

In summary, both mAbs, antibody fragments and antibody mimetics have advantages and disadvantages depending on the antigen being targeted. However mAbs generally have a greater stability compared with their fragment counterparts and as mAbs are entirely native to the host, they elicit the appropriate downstream responses while having the required *in vivo* half-life, without the need for any further genetic manipulation.

2.4.3 Antibody and biopharmaceutical aggregation

As mentioned previously, protein aggregation is a problem that affects all fields of protein science; from being the nightmare of PhD students, through to contributing to life-threatening diseases. In biopharmaceutical development, protein aggregation can have multiple effects, from the stability of a candidature therapeutic through to jeopardising patient safety.

As outlined in **Section 2.1**, protein aggregation is a complex problem that can be difficult to interrogate due to the mechanisms that drive oligomerisation are often poorly understood. Furthermore, added difficulty in characterising protein aggregates can arise due to the different classes of protein aggregates which can form: soluble/insoluble, covalent/non-covalent, reversible/non-reversible and native/denatured (Cromwell et al., 2006, Vasquez-Rey and Lang, 2011). Due to the stresses impacted on biotherapeutics, aggregate formation can occur at any stage of protein development and manufacture (Mahler et al., 2009); from pH stress, temperature jumps, changes in ionic strength, storage at high concentration and shear/extensional flow forces (Bekard et al., 2011). Finally, while it is ideal that stable candidates are chosen during the initial selections, any oligomeric material present in the product must be removed at the earliest stage possible. If unnoticed and present in the final product, aggregates may invoke an immune response to the active therapeutic agent (Rosenberg, 2006). In worst case scenarios, neutralising antibodies can impact the efficacy of the therapeutic (Chirmule et al., 2012) or cross-reactivity generates neutralising antibodies to endogenous proteins, which can lead to auto-immunity (De Groot and Scott, 2007). Therefore two things are clear, that (1) candidate protein therapeutic agents need to be exceptionally stable molecules, and (2) methods need to be in place to rigorously characterise candidate sequences to identify the possibility of aggregated material.

Currently, the main approaches to characterising aggregates in a candidate solution are chromatographic methods as well as light scattering, amongst others (Den Engelsman et al., 2011). However, aside from protein aggregation prediction algorithms, such as Aggrescan (Conchillo-Solé et al., 2007) and the spatial aggregation propensity algorithm (Chennamsetty et al., 2009b), these techniques are set up to identify protein aggregates once they have formed, and then to process their removal. While necessary, these techniques do not yield any information into the underlying mechanisms that cause aggregates to form. With a greater understanding into the mechanisms that drive these aggregation events, methods could be established to minimise or ameliorate these effects earlier in the developmental pipeline, increasing the amount of potential candidates entering trials and ultimately, the clinic.

2.5 WFL and STT – A model pair of monoclonal antibodies

The study presented in this thesis focuses on two model antibodies that target nerve growth factor (NGF): a highly aggregation-prone mAb, WFL, and a non-aggregation-prone, but closely related variant STT that were generated by MedImmune plc. These two samples share a sequence identity of 99.6 %, with mutations found in the variable region of the heavy chain of the antibodies. STT was engineered from WFL with mutations W30S, F31T and L57T (**Figure 2.17**) which ameliorate the self-association properties of WFL.

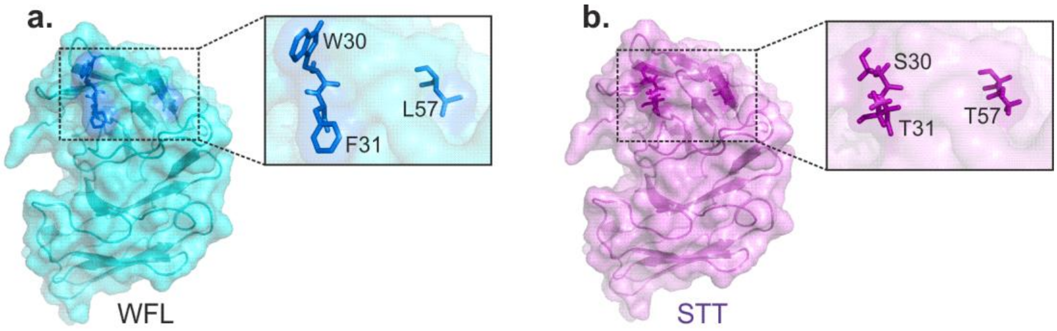


Figure 2.17: scFv structures of **(a)** WFL and **(b)** STT. Residues 30, 31 and 57 are represented as sticks in the insets. Homology models were generated from the parent scFv. Structures were kindly supplied by MedImmune plc.

WFL is the affinity matured version of its parent mAb that was derived from the parent sequence using phage display and has low picomolar (K_d = 1.6-9.8 pM) affinity for its target antigen. However, after the affinity maturation process, the poor biophysical characteristics that lead to the self-association and aggregation of WFL were realised. Along with poor yields in purification (< 30 %), absorption to filter membranes and colloidal instability, WFL eluted with a long retention time and asymmetric peak shape upon analysis by high performance size exclusion chromatography (HP-SEC), indicative of a high aggregation propensity. By contrast, while retaining its affinity for NGF (K_d = 1.8-8.3 pM), STT did not exhibit these poor biophysical characteristics (**Figure 2.18**). The reengineering WFL to STT at these positions (30, 31 and 57) were chosen as the parent sequence contained S30, T31 and T57 and did not display the poor biophysical characteristics, including aggregation. Thus, it was hypothesized that the mutations to hydrophobic residues during the affinity maturation were adding to the aggregation propensity of WFL.

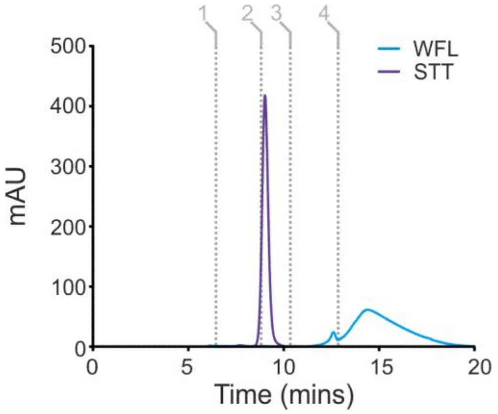


Figure 2.18: High performance size exclusion chromatography elution profiles of WFL (blue) and STT (purple). Grey lines = calibrant proteins: 1, Thyroglobulin (670 kDa); 2, IgG (158 kDa); 3, Ovalbumin (44 kDa) and 4, Vitamin B12 (1.35 kDa). Data acquired by MedImmune plc.

As well as demonstrating difficulties in purification, *in vivo* data showed that WFL has a significantly shorter half-life compared with its parent upon intravenous administration into healthy rats and cynomolgus monkeys. By contrast, upon carrying out the same pharmacokinetic (PK) studies using WFL and STT, STT exhibited a 2-fold improvement in half-life. (**Figure 2.19a**). Finally, to examine whether the enhanced clearance of WFL was driven by non-specific binding to tissues *in vivo*, binding of both WFL and STT were tested against a range of human tissue samples using immunohistochemistry (IHC) (**Figure 2.19b**). The results demonstrated that WFL was indeed binding non-specifically to the range of tissues that were tested against.

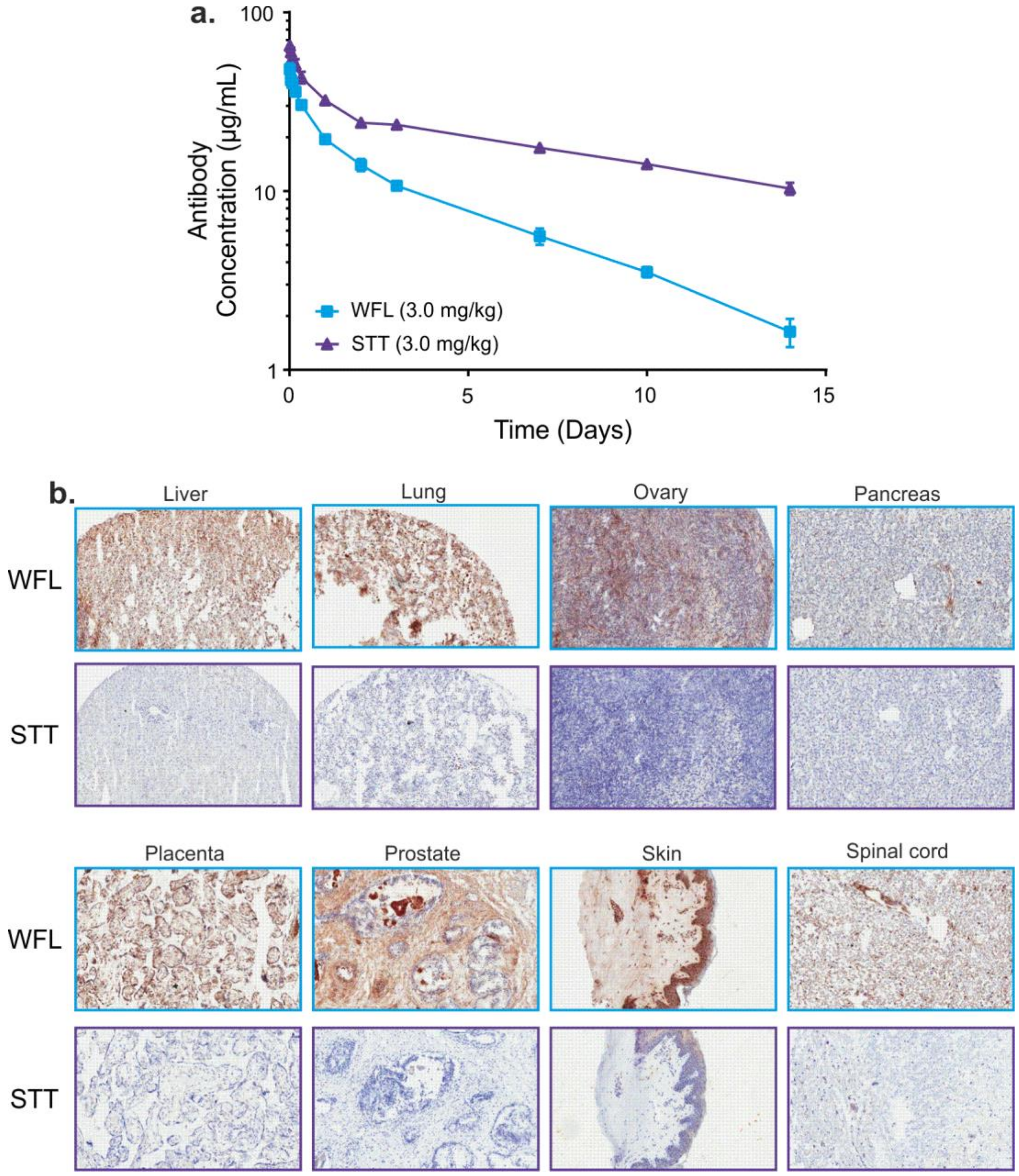


Figure 2.19: Pharmacokinetics and tissue studies of WFL and STT. **(a)** WFL and STT pharmacokinetic profiles showing faster clearance of WFL upon intravenous administration (of a 3 mg/kg dose) into healthy rats. **(b)** Non-specific binding of WFL to various tissue types while STT presented no evidence of staining any of the tissue types tested. Data acquired by MedImmune plc.

While the generation of STT ameliorates the self-association and aggregation of WFL, the underlying mechanism driving the self-association remained elusive. As WFL is one of the most extreme cases of antibody self-association experienced by MedImmune plc whilst STT differs in only three amino acids (per heavy chain) yet has substantially improved biophysical properties, it was concluded that this pair of mAbs would make an ideal pair to determine how subtle changes in sequence can result in catastrophic differences in their aggregation properties.

2.6 The aim of this study

Protein aggregation has sweeping negative effects throughout academia, biopharmaceutical production and all of protein science. However, the methods utilised within the biopharmaceutical industry are set up to identify and analyse protein aggregates once targets have been selected and antibodies generated, which is relatively late in the production pipeline. Furthermore, current established methods are able to detect and separate aggregated material from a product, but often do not yield site-specific information or insights into the mechanism underpinning the oligomerisation events. Therefore, there is a critical requirement for the development of techniques that can identify aggregation-prone sequences or candidates at the earliest stage possible. Such techniques have to be sensitive to small amounts of aggregated material and simple to equip and employ since predictions of aggregation-prone regions are only possible once a better understanding of the mechanisms are achieved. Thus, the aims of this study are two-fold: (i) to establish techniques that can identify the mechanisms responsible for driving the aggregation of a therapeutically relevant mAb and (ii) to characterise the regions responsible. Finally, this study aims to establish an *in vivo* method for the rapid detection of aggregation-prone sequences and candidates at early stages in candidate selection and biopharmaceutical product development.

The techniques described should:

- Be sensitive enough to identify small percentages of oligomeric/aggregated material
- Require small volumes and concentrations of samples
- Be amenable to a diverse range of proteins: different size, structures and aggregation mechanisms
- Yield high resolution information into the specific regions responsible for driving the oligomerisation/aggregation

And secondly, be able to:

- Accurately measure the extent of aggregation prone-sequences *in vivo*
- Have low false discovery rates
- Be rapid and cost effective in identifying aggregation prone sequences
- Be applicable to predicting the regions responsible for driving the aggregation events

All of these requirements are met throughout the studies described in this thesis focusing on the two closely related mAbs, WFL and STT, provided by MedImmune plc. Chapter 4 employs IMS-MS to identify and characterise small amounts of oligomeric material in terms of their mass and shape, via their collision cross-sectional area, to which models were built to evaluate potential mechanisms of aggregation. Chapter 5 utilises chemical cross-linking combined with mass spectrometry to capture the oligomeric species present in solution and identify inter-molecular cross-links, which were then used as restraints to derive high resolution models detailing the mechanism of self-association. Finally, Chapter 6 describes the development of an *in vivo* assay to identify aggregation-prone sequences in biotherapeutic candidates, which can be applied to the early stages of the biopharmaceutical developmental pipeline. Finally, Chapter 7 summarises the new insights gained from the work presented and highlights the future developments and impacts of the results presented in this thesis.

Chapter 3:
Materials and methods

3 Materials and Methods

3.1 Materials

3.1.1 Equipment

Mass Spectrometers

LCT Premier time-of-flight mass spectrometer (Waters Corp., Manchester, UK).

Synapt High Definition Mass Spectrometer (HDMS) quadrupole ion-mobility time-of-flight mass spectrometer (Waters Corp., Manchester, UK).

Synapt G2-Si HDMS quadrupole ion-mobility time-of-flight mass spectrometer (Waters Corp., Manchester, UK).

Xevo G2-XS quadrupole time-of-flight mass spectrometer (Waters Corp., Manchester, UK).

Centrifuges

Avanti J-26 XP Centrifuge	Backman Coulter, Brea, CA, USA
GenFuge 24D Centrifuge	Progen Scientific, London, UK
Eppendorf 5810R Centrifuge	Fisher Scientific, Loughborough, UK
Beckman Coulter XL-A Ultracentrifuge	Beckman Coulter Inc., CA, USA

Incubators, shakers and mixers

Gallenkamp Economy Incubator Size 1	Sanyo, Watford, UK
Innova 43 Shaker Incubator	New Brunswick Scientific, USA
Innova 44 Shaker Incubator	New Brunswick Scientific, USA
Stuart Oribtal Incubator S1600	Bibby Scientific, Stone, UK

Protein purification

ÄKTAprime Plus	GE Healthcare, Little Chalfont, UK
ÄKTAmicro	GE Healthcare, Little Chalfont, UK
HisTrap FF 5 mL Ni Sepharose	GE Healthcare, Little Chalfont, UK
Superdex Peptide 3.2/30 Column	GE Healthcare, Little Chalfont, UK

Gel electrophoresis equipment

Slab Gel Electrophoresis Chamber AE-6200	ATTO, Tokyo, Japan
Powerpac 3000	Bio-Rad Lab., Hercules, CA, USA
Powerpac Basic	Bio-Rad Lab., Hercules, CA, USA

Spectrophotometers

UltroSpex 2100 Pro UV/Visible Spectrophotometer	GE Healthcare, Little Chalfont, UK
NanoDrop 2000 Spectrophotometer	Thermo Scientific, Surrey, UK

Electron Microscope

FEI T12 Transmission Electron Microscope	FEI Inc., Hillsboro, OR, USA

Other Equipment

Bio-Rad T1000 Thermocycler	Bio-Rad Lab., Hercules, CA, USA
Grant JB1 Unstirred Waterbath	Grant Instruments, Shepreth, UK
Jenway 3020 Bench pH Meter	Bibby Scientific, Stone, UK
SnakeSkin Pleated Dialysis Tubing – 3,500 MWCO	Thermo Scientific, Surrey UK
Techne Dri-block Heater	Bibby Scientific, Stone, UK
C18 SepPak Cartridges	Waters corp., Wilmslow, UK

3.1.2 Chemicals

A

Acetic acid, glacial	Fischer Scientific, Loughborough, UK
Acetonitrile	Fischer Scientific, Loughborough, UK
Agar	Melford Laboratories, Suffolk, UK
Agarose	Melford Laboratories, Suffolk, UK
Ammonium Acetate, glacial	Fischer Scientific, Loughborough, UK
Ammonium bicarbonate	Fischer Scientific, Loughborough, UK
Ampicillin	Formedium, Norfolk, UK
Arabinose	Sigma Life Sciences, St. Louis, USA
L-Arginine	Sigma Life Sciences, St. Louis, USA

B

Bromphenol Blue	Sigma Life Sciences, St. Louis, USA
BS3 (bis(sulfosuccinimidyl)suberate)	Thermo Scientific, UK

C

Chloramphenicol	Sigma Life Sciences, St. Louis, USA
Caesium iodide	Sigma Life Sciences, St. Louis, USA

D

DTT (1,2-dithioltheritol)	Formedium, Norfolk, UK

E

Ethanol	Fischer Scientific, Loughborough, UK
Ethidium Bromide, EtBr	Sigma Life Sciences, St. Louis, USA
Ethylenediamine tetra acetic acid, EDTA	Sigma Life Sciences, St. Louis, USA

F

Formic acid	Fischer Scientific, Loughborough, UK

G

Glycerol	Fischer Scientific, Loughborough, UK
Glutaraldehyde	Sigma Life Sciences, St. Louis, USA

H

Hydrochloric acid	Fischer Scientific, Loughborough, UK

I

Instant blue stain	Expedeon, Harston, UK
Iodoacetamide	Sigma Life Sciences, St. Louis, USA

L

LB broth, granulated	Melford Laboratories, Suffolk, UK

R

L-Rhamnose	New England Biolabs, Hitchin, UK

S

Sodium chloride, NaCl	Fischer Scientific, Loughborough, UK
Sodium dodecyl sulphate, SDS	Sigma Life Sciences, St. Louis, USA
Sodium hydroxide, NaOH	Fischer Scientific, Loughborough, UK
Sodium phosphate monobasic, NaH_2PO_4	Sigma Life Sciences, St. Louis, USA
Sodium phosphate dibasic, Na_2HPO_4	Sigma Life Sciences, St. Louis, USA
Sodium succinate	Sigma Life Sciences, St. Louis, USA

T

Tetracycline	Formedium, Norfolk, UK
Tetramethylethylenediamine (TEMED)	Sigma Life Sciences, St. Louis, USA
Trifluoroacetic acid	Fischer Scientific, Loughborough, UK
Tris-(hydroxymethyl)-aminomethane, Tris	Melford Laboratories, Suffolk, UK
Tryptone	Melford Laboratories, Suffolk, UK

Y

Yeast extract	Melford Laboratories, Suffolk, UK

3.1.3 Enzymes used for molecular biology

Table 3.1: Enzymes used to carry out the molecular biology in this study. The buffers and additives of the corresponding enzymes supplied by the manufacturer were used.

Enzyme	Manufacturer
XhoI restriction endonuclease (20 U / µL)	New England Biolabs, Hitchin, UK
BamHI restriction endonuclease (20 U / µL)	New England Biolabs, Hitchin, UK
NdeI restriction endonuclease (20 U / µL)	New England Biolabs, Hitchin, UK
T4 DNA ligase (Quick Ligation™ Kit)	New England Biolabs, Hitchin, UK
Taq DNA polymerase (5 U / µL)	New England Biolabs, Hitchin, UK
Vent DNA polymerase (2 U / µL)	New England Biolabs, Hitchin, UK
Antartic phosphatase (5 U / µL)	New England Biolabs, Hitchin, UK
Lysozyme	Sigma Aldrich, St. Louis, USA
Deoxyribonuclease I (DNase I)	Sigma Aldrich, St. Louis, USA

3.1.4 Antibiotics

Table 3.2: Antibiotics used in this study.

Antibiotic	Solvent dissolved in	Stock concentration (mg/mL)	Working concentration (µg/mL)
Ampicillin	H_20	100	100
Chloramphenicol	Ethanol	30	25
Tetracycline	Ethanol	5	12.5

3.1.5 Media

Table 3.3: Media used in this study.

Media	Reagents	Weight / Volume
Luria-Bertani (LB) Media	Bacto-tryptone	10 g
	Yeast extract	5 g
	NaCl	10 g
	H_2O (18 MΩ)	Make up to 1 L
	Autoclave	20 min. at 121^{0}C, 15 psi
2xTY Media	Bacto-tryptone	16 g
	Yeast extract	10 g
	NaCl	5 g
	H_2O (18 MΩ)	Make up to 1 L
	Autoclave	20 min. at 121^{0}C, 15 psi

3.1.6 Markers and dyes

Table 3.4: Protein and DNA makers used in this study.

Marker	Manufacturer
Precision Plus Protein™ Dual Xtra Standard	Bio-Rad, Hemel Hempstead, UK
100 bp DNA ladder	Promega, Southampton, UK
1 kb DNA ladder	Promega, Southampton, UK

Table 3.5: Protein and DNA dyes used in this study

Dye	Manufacturer
Instant Blue Stain	Expedeon Protein Solutions, UK
Orange G Loading Dye (10x) for DNA	Sigma Aldrich, St. Louis, USA

3.1.7 Kits

Table 3.6: Commercially purchase kits used in this study.

Kit	Manufacturer
QIAquick Gel Extraction	QIAGEN, Crawley, UK
QIAquick PCR Purification	QIAGEN, Crawley, UK
QIAprep spin Miniprep	QIAGEN, Crawley, UK
Q5 Site-Directed Mutagenesis	New England Biolabs, Hitchin, UK

3.1.8 LB-agar plates

Table 3.7: LB-agar plates used in this study.

Antibiotic resistance required	Component
Tetracycline	2.5 g LB broth
	1.5 g Agar
	10 μg/mL tetracycline
	Make up to 100 mL in H_2O
Ampicillin-chloramphenicol	2.5 g LB broth
	1.5 g Agar
	100 μg/mL ampicllin
	30 μg/mL chloramphenicol
	Make up to 100 mL in H_2O

3.1.9 Buffers

Table 3.8: Buffers used in this study.

Buffer	Reagent
SDS-PAGE 2x loading buffer	50 mM Tris.HCl, pH 6.8
	100 mM DTT
	2 % (w/v) SDS
	0.1 % (w/v) bromophenol blue
	10 % (v/v) glycerol
Gel electrophoresis cathode buffer	200 mM Tris.HCl, pH 8.25
	200 mM Tricine
	0.2 % (w/v) SDS
Gel electrophoresis anode buffer	400 mM Tris.HCl, pH 8.8
Tris-acetate-EDTA (TAE) Buffer	40 mM Tris.HCl, pH 8
	20 mM acetic acid (glacial)
	1 mM EDTA, pH 7.5
	25 mM Tris.Hcl, pH 7.4

Cell lysis buffer	50 mg lysozyme
	10 mg DNase I
	300 µM PMSF
	1 mM EDTA
	Make up to 500 mL with H_2O
HisTrap™ binding buffer	50 mM Tris.HCl, pH 7.4
	300 mM NaCl
	40 mM Imidazole
HisTrap™ elution buffer	50 mM Tris.HCl, pH 7.4
	300 mM NaCl
	400 mM imidazole
Storage buffer	20 mM sodium succinate
	125 mM arginine, pH 6

3.2 Methods

3.2.1 Mass spectrometry

All IMS-MS measurements were made using a Synapt G1 high definition mass spectrometer (Waters Corp., Wilmslow, UK). Samples were introduced to the mass spectrometer using borosilicate capillaries pulled in-house (Sutter Instrument Company, Novato, CA, USA) and coated with palladium using a sputter coater (Polaron SC7620, Quorum Technologies Ltd, Kent, UK). All samples were analysed in the positive ion mode using nanoelectrospray ionisation. The *m/z* scale was calibrated using 10 mg/mL aqueous caesium iodide (CsI) clusters across the acquisition range (typically 500-15,000 *m/z*). Data were processed and analysed with the MassLynx v4.1 and Driftscope software, supplied with the mass spectrometer.

3.2.2 Sample preparation

Samples were dialysed against 150 mM ammonium acetate, pH 6 using 20 μL dialysis buttons (Hampton Research Crop., Aliso Viejo, CA, USA). Calibrant proteins β-lactoglobulin, avidin, concanavalin A, alcohol dehydrogenase, pyruvate kinase, glutamate dehydrogenase and GroEL were buffer exchanged into 150 mM ammonium acetate, pH 6 using Zeba spin desalting columns (3.5 kDa MWCO) (Thermo Scientific, UK) and used for IMS-MS calibration.

3.2.3 nESI-MS and IMS-MS of monoclonal antibodies

MAb samples were dialysed at 1 mg/mL before being infused into the Synapt G1 instrument. nESI-MS experiments were conducted under the following settings: capillary voltage, 1.5 kV; sample cone, 30 V; extraction cone, 4 V; source temperature, 60-80 ^{0}C; backing gas flow, 20 mL/min, IMS wave height (ramped), 5-30 V and travelling wave speed, 300 ms. For increased mass accuracy of mAb samples a trap voltage of 50-80 V was used.

3.2.4 TWIMS device calibration

The calibration of the TWIMS device was carried out using calibrant proteins from the Clemmer/Bush (Bush et al., 2010) database appropriate to the mass of the protein of interest. Typically used calibrant proteins were: beta-lactoglobulin, B-Lac; concanavalin A, CCA; alcohol dehydrogenase, ADH; pyruvate kinase, PyK and glutamate dehydrogenase, GluD. The calibrant proteins were acquired using the same conditions to that of the sample in question and a calibration curve was constructed using an established method, briefly outlined in **Section 1.6.2** (Ruotolo et al., 2008). The EDC correction coefficient used in the calibration curve construction was 1.57.

3.2.5 Theoretical CCS calculation

The MOBCAL software was used to calculate the theoretical cross-section for the samples studies in this thesis. The MOBCAL projection approximation value (Mesleh et al., 1996) was used to generate the projection superposition approximation (PSA) outlined by (Bleiholder et al., 2011b). The equation used for this was:

$$PSA = (PA - 81)\ x\ 1.299$$

Equation 3.1

The MOBCAL software was implemented using a Linux operating system.

3.2.6 LC-MS/MS

Peptide samples were separated and analysed on a nanoAcquity LC system (Waters Ltd., Wilmslow, UK) connected on-line to a Synapt G2-Si mass spectrometer (Waters Corp., Wilmslow, UK). 1 μL of sample was injected onto a Acquity M-Class column (C18, 75 μm x 150 mm) (Waters Ltd., Wilmslow, Manchester, UK) and subsequently separated by a 1-50 % gradient elution of solvent B (0.1 % (v/v) formic acid: acetonitrile) in solvent A (0.1 % (v/v) formic acid in water) over a 60 minute time period at a 0.3 μL/min flow rate. The instrument was operated in positive ion mode using collision-induced dissociation (CID) for fragmentation of selected ions. Data dependant MS/MS experiments were conducted in the trap region of the instrument using a 1 second scan with the five most intense ions being selected for fragmentation over a 350 – 2000 *m/z* window. Data were processed and analysed with the MassLynx (V4.1) software provided with the instrument, and PEAKS software (Zhang et al., 2012). Manually identified and low abundant cross-linked peptides were added to an inclusion list to be selected for sequencing.

3.2.7 Cross-linking analysis

LC-MS/MS spectra were exported from the PEAKS software in .MGF format to be analysed with the StavroX software (Götze et al., 2012). The search parameters used were as follows: maximum missed cleavages at C-terminal of K and R, 1; static modifications, C to B (alkylated cysteine) and variable modifications, M to m (oxidised methionine, max = 2); precursor precision, 3.0 ppm; fragment ion precision, 0.8 Da; lower mass limit, 200 Da; upper mass limit, 8000 Da; false discovery rate cut off, < 5 % with a score cut off of 10 and slow, precise scoring applied.

3.2.8 Manual data analysis for cross-linked species

To identify all candidates as potential inter-molecular cross-links the LC-MS peptide spectra were compared. Each spectrum from the dimer sample of 1912 were interrogated to search for potential cross-linked candidates. Once a candidate had been identified, the spectrum from the cross-linked monomer sample was analysed at the same retention time (± 1 minute) to confirm unique inter-molecular cross-links.

3.3 Modelling

3.3.1 MOBCAL modelling of mAbs

Approximate models for the mAb monomers and their oligomers were modelled using the MOBCAL software (Mesleh et al., 1996, Shvartsburg and Jarrold, 1996) by editing the MOBCAL code to define the Fab and Fc regions of a mAb as single atoms. Lines 581 and 583 of the MOBCAL code were changed to the mass of the Fab region of the molecule (47 kDa) and line 586 edited for the hydrodynamic radius (30 Å). Lines 591 and 593 were then given the mass of the Fc region (53 kDa) and line 596 edited for the Fc hydrodynamic radius (30 Å). These values were also implemented on lines beginning 2599 and 2609 of the code. The input .mfj file was then modified to manipulate the vector coordinates of the individual regions to re-build a monomeric mAb which predicted CCS value was in agreement with experimental data. Oligomeric species were then built to the models specified in Chapter 4 (see **Section 4.3.3**) and their predicted CCSs calculated.

3.3.2 Generation of spherical models

To generate spherical models, the Ruotolo method (Ruotolo et al., 2008) and Benesch methods (Benesch et al., 2007) were used. An average protein density of 0.44 Da Å^{-3} was used.

<u>Ruotolo method:</u>

$$CCS = M_i^{\frac{2}{3}} \times 2.435$$

Equation 3.2

Where M_i is the mass of the ion

<u>Benesch method:</u>

$$CCS = \pi((\frac{3M_i}{4\pi\rho})^{\frac{1}{3}} + R_g)^2$$

Equation 3.3

where M_i is the mass of the ion, ρ is the density of the protein (0.44 Da Å^{-3}) and R_g is the radius of the buffer gas (nitrogen = 1.55 Å)

3.3.3 Generation of the (127)$_5$ polyprotein structure

The I27 monomer subunit structure was taken from the PDB (PDB= 1TIT). The four linker domains connecting the subunits were then attached to the C-terminus of the previous sub-unit based upon the linkers used while purifying the proteins; the linker regions added were:

Table 3.9: Table of amino acid linkers added to the C-terminal of the I27 sub-units 1-4.

Linker	Amino acid composition
Linker 1	VEAR
Linker 2	LIEAR
Linker 3	LSSAR
Linker 4	LIEARA

The five PDB structures were then aligned manually with one another before the C-terminal amino acid from the linker region was connected to the N-terminal leucine of the sequential I27 sun-unit. The subunits were connected by Dr Claire Windle (University of Leeds) using the Coot software, operated under a Linux operating system.

3.3.4 In vacuo molecular dynamics (MD) simulations

MD simulations were run using the NAMD software (NAMD 2.9) (Phillips et al., 2005), using the CHARMM force field (Brooks et al., 1983). Structures were simulated in a solvent free system. For the simulation, a constant temperature of 300 K with Langevin thermostat was used and a time-step of 2.0 fs with a radial cut-off distance of 12 Å used throughout. Energy minimization *in vacuo* was implemented for a total of 0.5 ns before an equilibration of 10 ns, the cut-off distance, force field and time step remained as described above throughout the simulation. Visual Molecular Dynamics (VMD) (Humphrey et al., 1996) was then used to visualise the simulation and individual frames were saved as PDB coordinates in order to compute the CCS using MOBCAL. The VMD software was also used to calculate the root mean square deviation (RMSD) and radius of gyration (Rg). The CCS were calculated every 0.1 ns of the simulation to 1 ns and then every 1 ns thereafter. Analysis of the root-mean-square deviation (RMSD) revealed whether a protein had equilibrated by the end of the 10 ns simulation, any sample that had not finished equilibration was re-submitted for a further 10 ns until equilibration was reached. The NAMD and VMD software was operated under a Linux operating system. n = 1 for all MD minimisation and equilibration experiments.

3.3.5 Model generation for 1912 scFv dimer

The 1912 scFv homology model (from parent mAb, kindly provided by MedImmune plc.) was submitted to the HADDOCK webserver (van Zundert et al., 2016) with residues W30, F31 and L57 designated as the active residues to drive the docking procedure. The HADDOCK model was then refined in XPLOR-NIH (Schwieters et al., 2003) using a distance restraint set by the bis(sulfosuccinimidyl)suberate (BS3) cross-linker (11.4 Å) between residues Q1 of the heavy chain and K54 of the light chain with a square well energy potential and residues W30, F31 and L57 as sparse, highly ambiguous distance restraints (Clore and Schwieters, 2003). The model refinement in XPLOR-NIH was carried out with the help of Dr Theodoros Karamanos (University of Leeds).

3.4 Biochemical methods

3.4.1 Analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation (SV-AUC) of antibody samples was conducted on a Beckman Coulter XL-A instrument (Beckman Coulter Inc., CA, USA) at 20 °C. MEDI1912 and MEDI1912_STT were dialysed into 150 mM ammonium acetate, pH 6 or analysed in 'succinate-Arg buffer' (20 mM sodium succinate, 125 mM arginine, pH 6.0) at the same protein concentration (1 mg/mL, determined by A_{280}). A wavelength of 298 nm and rotor speed of 28,000 rpm were used. Data were analysed using SEDFIT (Schuck, 2000). AUC data were analysed with the help of Dr Clare Pashley (University of Leeds).

3.4.2 Negative stain transmission electron microscopy

mAb samples at 1 mg/mL, as measured using A_{280}, in storage buffer (**Table 3.8**). Samples were diluted and immediately spotted onto carbon coated electron microscopy grids (400 mesh, glow discharged 25 mA for 40 seconds) (Fisher 462 Scientific, Leicestershire, UK) and allowed to absorb for 30 seconds. The grids were then washed 463 with milliQ-H_2O with blotting on Whatman paper between wash steps, Stained with 1 % (w/v) uranyl 464 acetate, air dried under a lamp and imaged on a FEI T12 TEM (FEI Inc., Hillsboro, OR, USA) operating 465 at 120 keV.

3.4.3 Lys-C enzymatic digestion of mAb samples

Endoproeinase Lys-C from *Lysobater enzymogenes* (Sigma Aldrich, St. Louis, USA) was used to generate Fab and Fc regions from mAb samples. mAb samples were used at 1 mg/mL in 100 mM sodium phosphate buffer, pH 7.4 and Lys-C added at a 1:50 (protein:enzyme) ratio. The protein-enzyme reaction was incubated at 37 ^{0}C for 18 hours before the resulting digests being taken for dialysis and analysis via MS and IMS-MS.

3.4.4 Glutaraldehyde cross-linking

Protein samples were diluted to 1 mg/mL and buffer exchanged into 100 mM sodium phosphate buffer, pH 7.4, using zeba spin columns (Thermo Scientific, UK), for cross-linking with glutaraldehyde (Sigma Aldrich, St. Louis, USA). A stock solution of 1 % (w/v) glutaraldehyde was used to probe for optimal concentrations across the range of: 0.1 %, 0.05 %, 0.01 % and 0.001 % (w/v).

3.4.5 BS3 cross-linking

Protein samples were diluted to 1 mg/mL and buffer exchanged into 100 mM sodium phosphate buffer, pH 7.4, using zeba spin columns (Thermo Scientific, UK), for cross-linking with a mixture of deuterated (d4) and non-deuterated (d0) BS3 (Thermo Scientific, UK). 1 mg of both d0 and d4 BS3 was dissolved in 138.5 µL 20 mM sodium phosphate buffer, pH 7.4. Equimolar concentrations of the d0 and d4 cross-linker were added together to yield a 50:50 mix of d0:d4 as a stock solution. BS3 cross-linker was initially added in increasing ratios for optimisation: 1:10, 1:20, 1:50, 1:100, 1:200 and 1:500 (mol/mol) (protein: cross-linker). The cross-linking reaction was incubated on ice for 30 minutes before quenching with 50 mM Tris.HCl, pH 8 (final concentration). The resulting cross-linked species were then analysed using either digestion and peptide size-exclusion chromatography (SEC) or via SDS-PAGE analysis and in-gel digestion methods, see **Section 2.4.6** and **Section 2.4.7/Section 2.4.8,** respectively.

3.4.6 Enzymatic digestion, peptide purification and concentration

Prior to digestion, first the samples were denatured with an equal volume of 6 M Gdn.HCl. 7.5 mM DTT was added to the samples and incubated at 37 °C for 30 min. Samples were then cooled back down to room temperature and 15 mM of iodoacetamide added and the reaction incubated in the dark at room temperature for 30 min. The samples were then diluted to reduce the Gdn.HCl concentration below 1.5 M. The resulting BS3 cross-linked mixture was then digested with a 1:100 (w/w) ratio (enzyme:protein) with gold standard trypsin (Promega, UK) for 18 hours at 37 ^{0}C. 50 µL Formic acid was then added to quench the reaction, the reaction mixture was then made up to 1 mL with ddH$_2$0. The digested mixture was purified using C18 SepPak cartridges (Waters corp., Wilmslow, UK). The cartridges were equilibrated with 1 mL 100 % acetonitrile and washed with 2×1 mL wash buffer (95:5:0.1 (v/v/v), ddH$_2$0: acetonitrile: formic acid) before the digested mixture was passed through the cartridge. The bound mixture was washed with a further 2×1 mL of wash buffer before being eluted with 1 mL elution buffer (50:50:0.1, ddH$_2$0: acetonitrile: formic acid). Eluted peptides were then evaporated to completion using a Savant SpeedVac system (Thermo Scientific, UK). The peptide mixture was then reconstituted in 40 µL ddH$_2$0 and taken for peptide SEC, see **Section 2.4.6**.

3.4.7 Peptide size exclusion chromatography (SEC)

Reconstituted cross-linked peptide mixtures were pipetted into a 96-well plat for injection into an ÄKTAmicro (GE Healthcare, Little Chalfont, UK) using an online A-905 auto-injection system (GE Healthcare, Little Chalfont, UK) with a Superdex Peptide 3.2/30 column (GE Healthcare, Little Chalfont, UK) being used for the gel filtration. The column was washed and equilibrated with 30 % (v/v) acetonitrile, 0.1 % (v/v) trifluoroacetic acid (Buffer A). 40 µL of sample was injected using the pick-up injection method and the peptides eluted using a constant flow rate of 50 µL/min. 100 µL fractions were collected in a 96-well plate format with the elution of peptides being monitored by absorbance at 218, 254 and 280 nm. The collected fractions were evaporated to completion using a Savant SpeedVac system (Thermo Scientific, UK) before being taken for analysis by LC-MS/MS (see **Section 2.2.6**).

3.4.8 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Tris-tricine buffered sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate species based upon their molecular weight to analyse the formation of cross-linked species. Two glass plates were assembled following the manufacturer's directions with a 1.5 mm spacer between the plates to ensure a tight seal between the plates. The resolving and stacking gels were made up of the components shown in **Table 3.10**. Ammonium persulphate was added to the gel mixtures last before the resolving gel was immediately poured into the gel moulds. The stacking gel was poured on top of the resolving gel and a 12 well comb was added to create the wells needed once the gels had set. The gel cast was left to set for a minimum of 1 hour.

Table 3.10: Components used to create Tris-tricine SDS-PAGE gels. The volumes stated here allow for the casting of two gels (80 mm × 100 mm) using a 1.5 mm spacer.

Solution component	Resolving gel (mL)	Stacking gel (mL)
30 % (w/v) acrylamide: 0.8 % (w/v) bis-acrylamide	7.5	0.83
3 M Tris.HCl, 0.3 % (w/v) SDS pH 8.45	5.0	1.55
ddH_20	0.44	3.72
Glycerol	2.0	--
10 % (w/v) ammonium persulphate	0.05	0.10
Tetramethylethylenediamine (TEMED)	0.005	0.005

Protein samples for SDS-PAGE were diluted 2-fold with 2x loading buffer (see **Table 2.7**) and boiled for 3 minutes prior to loading onto gels. For identification of molecular weight species, the first lane was loaded with 7 µL of Precision Plus Protein™ Dual Xtra Prestained protein standard (Bio-Rad,UK). The inner and outer reservoirs were filled with cathode buffer (see **Table 2.7**) and anode buffer (see **Table 2.7**), respectively, prior to sample loading. 15 µL of samples were loaded into each well before begin electrophoresed. Gels were run with an initial current of 30 mA until all samples has entered into the resolving gel, where the current was increased to 75 mA until the dye front had reached the bottom of the resolving gel. For cross-linked mAb samples, the gels were run until the 25 kDa ladder had reached the bottom of the resolving gel in order to separate the oligomeric species. The gels were removed from there casts and stained using Instant Blue (Coomassie) stain (Expedeon, UK) and visualised using a Syngene InGenius gel documentation system (Syngene, UK) post gel de-staining in H_2O.

3.4.9 In-gel enzymatic digestion

A scalpel was used to remove the gel bands of interest and placed into 1.5 mL Eppendorfs. The gel pieces were subjected to three repeat rounds of hydration and dehydration with 200 µL 25 mM ammonium bicarbonate, pH 7.8 and 200 µL 50 % (v/v) acetonitrile/ 25 mM ammonium bicarbonate, respectively, to completely de-stain the excised gel pieces. Each solution was removed from the Eppendorf tube with a gel loading tip before the subsequent solution added Gel pieces were fully submerged in 100 % acetonitrile to shrink the gel piece down to completion. The acetonitrile was removed by incubating the open Eppendorfs in a laminar flow hood for a minimum of 45 minutes. Gel pieces were re-hydrated with 200 µL 10 mM DTT for 45 minutes at 37 ^{0}C and shaking at 300 RPM in a Thriller shaking incubator (WVR Peqlab, Germany) before free cysteines were subsequently alkylated with 200 µL 55 mM iodoacetamide in the dark at room temperature. A final dehydration step was then carried out using 200 µL 50 % (v/v) acetonitrile / 25 mM ammonium bicarbonate. The dehydration solution was again removed under a laminar flow hood for a minimum of 45 minutes. Dry gel pieces were re-hydrated with 0.1 µg/µL gold standard trypsin solution in the supplied re-suspension buffer (Promega, UK), ensuring that the trypsin solution was added to completely submerge the gel pieces. Gel pieces were incubated for 30 minutes at room temperature to allow the trypsin solution to be absorbed before adding 50 µL 25 mM ammonium bicarbonate and the reaction incubated at 37 ^{0}C for 18 hours, shaking at 300 RPM.

20 μL of formic acid was added to quench the reactions and the digested cross-linked peptides were purified as described above (see **Section 2.4.5**).

3.4.10 Polyethylene glycol (PEG) precipitation

A 40 % (w/v) stock solution of PEG (~8000 average molecular weight) in 20 mM sodium succinate, 125 mM arginine, pH 6 buffer was prepared as the titrant. ScFv stock solutions were prepared at 150 μM in the same buffer. Samples were prepared in triplicate and loaded into a 96-well, clear, flat-bottomed plate (Greiner Bio-One, Austria) using a multichannel pipette as follows: a) succ-Arg buffer was dispensed into each well with a volume of 100-180 μL. b) 0-180 μL of the 40 % (w/v) PEG stock was titrated into each well to give a series of PEG concentrations from 0-36 % (w/v). c) 20 μL of scFv stocks were dispensed into each well to give a final concentration of 15 μM. d) Samples were then mixed by slowly pipetting up and down the wells. The plates were examined for any air bubbles and nephelometry measurements were taken immediately using a NEPHELOstar Galaxy (BMG Labtech, Germany).

3.4.11 *In vivo* western blot analysis

A single colony from fresh *E. coli* BL21 (DE3) cells (transformed with the appropriate plasmid) was used to inoculate 100 mL sterile LB containing 10 μg/mL tetracycline. Cultures were incubated overnight at 37 °C with shaking (200 rpm). 1 mL of overnight culture was used to inoculate 100 mL sterile LB containing 10 μg/mL tetracycline and grown at 37 °C (shaking at 200 rpm) until an OD_{600} of 0.6 was reached. 10 mL of culture was removed for the uninduced sample and centrifuged at 4,000 g for 10 min (4 °C). Expression of the β-lactamase fusion construct was induced by the addition of filter-sterilized arabinose to a final concentration of 0.02 % (*w/v*). Cultures were incubated for 1 h (37 °C, 200 rpm) and 10 mL was removed from each (induced sample). The 10 mL cultures were harvested by centrifugation at 4,000 *g* for 10 min (4 °C). The cell pellets (uninduced and induced with arabinose) were resuspended in phosphate buffered saline (PBS, Dulbecco's PBS, Sigma) to obtain an OD_{600} of 5. For whole cell samples, 200 μL of the OD_{600} = 5 sample was combined with 200 μL PBS. 100 μL of 6× loading dye (150 mM Tris-HCl, pH 6.8, 300 mM DTT, 6 % (*v/v*) SDS, 0.3 % (*w/v*) bromophenol blue) was then added. For soluble samples, 200 μL of the OD_{600} = 5 sample was combined with 200 μL bacterial protein extraction reagent (ThermoFisher Scientific) and incubated with agitation for 10 min. The sample was then centrifuged at 16,000 *g* for 30 min (4 °C) and 100 μL of 6× loading dye was added to the supernatant. The mixtures were then incubated at 90 °C for 10 min.

Protein samples were separated on a BIORAD Mini-PROTEAN TGX precast protein and were transferred to a BIORAD 0.2 µm polyvinylidene fluoride membrane using a Trans-Blot Turbo Semi-Dry (Bio-Rad Ltd). Blocking was performed overnight at 4 °C using 5 % (*w/v*) milk powder in TBST (tris-buffered saline Tween; 20 mM Tris-HCl, 150 mM NaCl, 0.2 % (*v/v*) Tween-20). Membranes were incubated overnight with the anti-β-lactamase antibody (CSB-PA352353YA0IENL, Cusabio) diluted 1:10,000 in 5 % (*w/v*) milk powder in TBST. The membranes were washed for 3× 10 min in TBST. Membranes were then incubated with goat anti-rabbit IgG horseradish peroxidase conjugate (7074, New England Biolabs) diluted 1:10000 in TBST. Membranes were then washed 3× 10 min in TBST before incubation with SuperSignal™ western pico chemiluminescent substrate (Thermo Fisher Scientific). The emitted signal was detected with Amersham hyperfilm (GE Healthcare).

3.5 Molecular Biology

3.5.1 E. coli (BL21 (DE3)) transformation

50 µL of commercially competent BL21 (DE3) cells were transformed with 2 µL of plasmid DNA (100 ng/µL). Cells were incubated on ice for 10 minutes to allow the cells to thaw. 2 µL of DNA was added to the cells and the mixture incubated on ice for a further 2 minutes. The cells were then heat shocked at 42 ^{0}C for 10 seconds before being incubated back on ice for 2 minutes. After the 2 minute incubation on ice, the transformed cells were plated directly onto LB-agar plates containing 10 µg/mL tetracylcine. The cells were left to dry in a sterile environment for 20 minutes before being incubated overnight at 37 ^{0}C.

3.5.2 *In vivo* Assay

BL21 (DE3) cells (New England Biolabs, Hitchin, UK) were freshly transformed with the scFv-β-lactamase construct DNA and plated out onto 10 μg/mL tetracycline plates and incubated at 37 ^{0}C overnight. 100 mL overnight cultures were prepared in a 250 mL flask with 2.5 % (w/v) LB media continuing 10 μg/mL tetracycline which were inoculated with a single colony and grown for 18 hours at 37 ^{0}C, 200 RPM. Two 48 well agar plates containing 10 μg/mL tetracycline and increasing ampicillin antibiotic are made with the following protocol: 100 mL 1.5 % (w/v) agar is autoclaved and allowed to cool to < 50 ^{0}C before being poured into a sterile beaker. 333 μL tetracycline stock and 200 μL arabinose stock were added to give final concentrations of 10 μg/mL tetracycline and 0.02 % (w/v) arabinose. A multichannel pipette was used to draw up 900 μL and 300 μL agar was pipetted into the first column of each of the two plates; the last 300 μL was dispensed back into the beaker. The required volume of ampicillin was then added into the agar (see **Table 2.10**) and mixed thoroughly. 300 μL of agar was pipetted into the second column of each plate and the process is repeated until the plates were full. The plates were left to set in a sterile environment and then stored upside down with their lids on.

Table 3.11: Components for to creation of 48-well LB-agar plates used for the MIC assay. The volumes stated here allow for the creation of two 48-well LB-agar plates.

Final (Amp) concentration (µg/mL)	Agar vol.	10 mg/mL Amp. stock required (µL)
0	100	0
20	96.4	193
40	92.8	186
60	89.2	178
80	85.6	171
100	82	164
120	78.4	157
140	74.8	150

1 mL of overnight culture was used to inoculate pre-warmed 100 mL sterile LB media containing 10 µg/mL tetracycline. The cultures were grown at 37 ^{0}C, 200 RPM until an OD_{600} = 0.6 is reached. 200 µL arabinose stock (0.02 % (w/v) final concentration) as added to the 100 mL cultures to induce scFv-β-lactamase protein expression. 400 µL aliquots were moved into labelled 1.5 mL Eppendorf tubes and incubated for 1 hour at 37 ^{0}C, 200 RPM to allow the constructs to be expressed before plating out. Log_{10} dilutions were carried out by taking 20 µL of cell culture and pipetting into subsequent 180 µL 170 mM NaCl, mixing thoroughly and

repeating. 3 µL of each cell dilution was then plated onto each column of the 48-well ampicillin-containing agar plates. The plates were allowed to dry in a sterile environment before being incubated upside down at 37 ^{0}C for 18 hours.

3.5.2.1 Polymerase chain reaction

The polymerase chain reaction (PCR) was carried out to amplify selected DNA sequences *in vitro*. The primer sequences and the purpose of those sequences designed to amplify specific genes from the selected plasmids are listed in (**Table 2.11**).

Table 3.12: Oligonucleotide primer used in this study. NdeI and XhoI restrictions sites are shown in gold and red respectively. The hexa-histag sequence is shown in orange with the additional amino acids between the scFv and hexa-histag shown in blue. Start and stop codons are underlined.

Primer	Sequence	Use
WFL Forward	GGTGGTGACATATGCAGGTTCAGCTGGTTCAGAGTG	To clone the scFv gene out of the β-lactamase construct
WFL Reverse	GCATACTCGAGTTAGTGGTGATGGTGATGGTGCGCCGCCGCCCCCAGCACGGTGAGTTTGGTG	To clone the scFv gene out of the β-lactamase construct. Addition of a hexahistag at the C-terminus of the protein
STT Forward	GGTGGTGACATATGCAGGTTCAGCTGGTTCAG	To clone the scFv gene out of the β-lactamase construct
STT Reverse	GCATACTCGAGTTAGTGGTGATGGTGATGGTGCGCTGCCGCCAGAACGGTCAGTTTAGTGC	To clone the scFv gene out of the β-lactamase construct. Addition of a hexahistag at the C-terminus of the protein

The typical PCR reaction contained the following components:

Table 3.13: Table of components used for PCR reactions.

Component	Concentration / volume
Primers	2 µL (100 ng/µL stock)
dNTPs	0.25 mM
DMSO	1 µL
$MgSO_4$	2, 4 or 6 mM
Vent DNA polymerase (2000 U/mL)	1 U (0.5 µL)
Vent DNA polymerase buffer	1×
Nuclease free deionised ddH_2O	To make reactions up to 100 µL
DsDNA template	100ng

A reaction lacking the dsDNA template was used as a negative control to determine if any non-specific interactions between primers occurred.

The theoretical melting temperature (T_m) of the primers was calculate using the following equation (**Equation 2.2**):

$$T_m = (n_{AT} \times 2) + (n_{GC} \times 4)$$

Equation 2.2

Where n_{AT} is the number of AT nucleotide base pairs in the primer sequence and n_{GC} is the total number of GC nucleotide base pairs in the primer sequence. The temperature cycle for the PCR cycle is shown in (**Table 2.12**).

Table 3.14: Temperature cycle for the PCR reactions used in this study.

Step	Temperature (^{0}C)	Time (s)
Initial denaturation	95	300
Denaturation	95	30
Annealing	55	30
Elongation	72	45
Repeat denaturation, annealing and elongation x29		
Final elongation	72	300

The PCR products were then visualised by agarose gel electrophoresis (see **Section 2.5.1.2**) and excised from the gel using a sterilised scalpel. DNA extraction from the gel was performed using the QIAquick Gel Extraction Kit (QIAGEN, UK) as described by the instructions supplied by the manufacturer.

3.5.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out in Tris-acetate-EDTA (TAE) buffer (see Table 3.8). Gels were made by dissolving 1.5 % (w/v) agarose (Melford Laboratories Ltd, Ipswich, UK) in 1x TAE buffer and heating the solution in a microwave until the agarose had dissolved. Once cooled to < 50 ^{0}C, 0.5 µg/mL ethidium bromide (EtBr) was added. The gels were then poured into a 12 x 15 cm cast with a comb in place and allowed to set for a minimum of 20 minutes at 25 ^{0}C. DNA samples were diluted in 6x orange gel loading buffer (see **Table 3.5**). 30 µL of sample were loaded into the gel lanes, 5 µL of 1 kb and 100 bp DNA ladders (Promega, UK) were added to the first two lanes to allow for accurate sample size determination. Electrophoresis was carried out in 1x TAE buffer at 100 V until the DNA samples were suitably resolved. The results were visualised using ultra violet (UV) transillumination and photographed using a Syngene InGenius gel documentation system (Syngene, UK).

3.5.2.3 Restriction digestion of plasmid DNA and plasmid dephosphorylation

Site-specific restriction digestion of plasmid or PCR product DNA was carried out using enzymes and buffers acquired from New England Biolabs, USA (NEB). Restriction digestions contained the following components:

Table 3.15: Table of components used for restriction digests.

Component	Volume
Plasmid DNA or purified PCR product DNA	30 µL
10× CutSmart Buffer	4 µL
NdeI restriction enzyme (20 U/µL)	1 µL
XhoI restriction enzyme (20 U/µL)	1 µL
10× NEB bovine serum albumin	4 µL

A control reaction containing no restriction enzymes were carried out in tandem. Reactions were incubated for 1 hour at 37 ^{0}C. The restriction enzymes were then inactivated by incubation at 65 ^{0}C for 20 min.

Restriction enzymes, buffer components and unwanted by-products from the digestion protocol were removed through separating the DNA fragments by agarose gel electrophoresis (see **Section 2.5.2.2**). The required DNA fragments were then excised using a sterilised scalpel and extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen, UK) as described by the instruction supplied by the manufacturer.

In order to prevent unwanted pre-ligation of the plasmid DNA, the 5'-ends were dephosphorylated with Antarctic Phosphatase (NEB, USA). 1:10 (v/v) of 10x Antarctic Phosphatase Reaction Buffer (NEB, USA) and 0.01 U of enzyme per pmol of DNA ends was added to the completed digestion reaction. The reaction mixture was incubated for 15 min at 37 ^{0}C. The enzyme was then inactivated through heating at 65 ^{0}C for 5 min.

3.5.2.4 Ligation of DNA

DNA ligation was carried out using the NEB Quick Ligation Kit. A typical ligation reaction was set up with the following components:

Table 3.16: Table of components used for the ligation of DNA inserts into plasmid vector.

Component	Volume / concentration
Digested and dephosporylated DNA vector	100 ng
2x NEB Quick Ligation Buffer	10 µL
NEB T4 DNA Ligase (2000 U/µL)	2000 U (1 µL)
Digested insert DNA	up to 20 µL

Control reactions were carried out simultaneously using nuclease-free deionised water in place of insert DNA. Reactions were incubated for 15 minutes at 25 ^{0}C and then incubated on ice prior to transformation into pLemo cells (see **Section 2.5.2**)

3.5.3 Q5 mutagenesis of WFL and STT

To obtain all possible mutants, the WFL and STT sequences were mutated using a Q5 mutagenesis kit, following the manufacturer's protocol.

Table 3.17: Table of mutants generated with the template DNA and primer sequences listed. The mutated residue is highlighted in red. The annealing temperatures are listed.

Required mutant	Template	Primer F	Primer R	Annealing temp (°C)
W30S	β-lactamase-WFL	CGGTACGTTTTCCTTTGGGGCCTTTAC	CCACTCGCTTTGCAGCTA	62
F31T	β-lactamase-WFL	TACGTTTTGGACCGGGGCCTTTACTTGGG	CCGCCACTCGCTTTGCAG	67
L57T	β-lactamase-WFL	TATTTTCGGCACGACAAACCTGGC	GGGATAATGCCACCCATC	59
W30S, F31T	β-lactamase-STT	GATTTTTGGGCTGACGAACCTGGCC	GGGATGATACCGCCCATC	63
W30S, L57T	β-lactamase-STT	CACCTTCTCCTTTGGAGCGTTTACGTGGGTG	CCGCCGCTCGCTTTGCAA	68
F31T, L57T	β-lactamase-STT	GGCACCTTCTGGACCGGAGCGTTTA	GCCGCTCGCTTTGCAAGA	69

The reactions for the Q5 mutagenesis were as follows:

Table 3.18: Table of components and volumes used for the Q5 mutagenesis.

Component	Volume required (μL)	Final concentration
Q5 Hot start high-fidelity 2× master mix	12.5 μL	1x
10 μM forward primer	1.25 μL	0.5 μM
10 μM reverse primer	1.25 μL	0.5 μM
Template DNA (1-25 ng/μL)	1 μL	1-25 ng
Nuclease-free water	9 μL	-----

The PCR cycling conditions were:

Table 3.19: Temperature cycle for the PCR reactions for the Q5 mutagenesis.

Step	Temperature	time
Initial denaturation	98 °C	30 s
25 cycles	98 °C	10 s
	T_A of primer	10-30 s
	72 °C	20-30 s/kb

Final extension	72 °C	2 min
hold	4-10 ‘C	-----

3.5.4 E. coli (pLemo cells) transformation

50 µL of commercially competent pLemo cells (New England Biolabs, Hitchin, UK) were transformed with 2 µL of plasmid DNA (100 ng/µL). Cells were incubated on ice for 10 minutes to allow the cells to thaw. 2 µL of DNA was added to the cells and the mixture incubated on ice for a further 2 minutes. The cells were then heat shocked at 42 ^{0}C for 10 seconds before being incubated back on ice for 2 minutes. After the 2 minute incubation on ice, the transformed cells were plated directly onto LB-agar plates containing 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. The cells were left to dry in a sterile environment for 20 minutes before being incubated overnight at 37 ^{0}C.

3.5.5 Small scale expression trial

A small scale expression trial was carried out to identify the optimal amount of L-rhamnose concentration required to optimise expression of scFvs in the soluble fraction. Lemo21 (DE3) cells were transformed with either the 1912_scFv or STT_scFv plasmid (see **Appendix 8.1**) as described above (see **Section 2.5.2**). A single colony was taken from a plate of freshly transformed bacteria and used to inoculate a 2x1 mL LB overnight cultures containing 100 µg/mL ampicillin and 25 µg/mL chloramphenicol. 200 µL of overnight culture was used to inoculate a 10 mL sterile LB media, in falcon tubes, containing the same antibiotics and varying concentrations of L-rhamnose: 0, 100, 250, 500, 750, 1000 and 2000 µM. The cultures were grown at 37 ^{0}C with shaking at 200 RPM to an OD_{600} of 0.6. A 2x1 mL sample was taken before induction and stored at -20 ^{0}C. Protein expression was induced with 40 µL of 100 mM filter-sterilised Isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown for a further 4 hours (37 ^{0}C, 200 RPM) before further 2x1 mL samples were taken of the induced samples. Whole cell expressions were measured by harvesting the 1 mL samples by centrifugation (10 minutes, 13,000 RPM) and re-suspending in 2x SDS-PAGE loading buffer, before boiling and loaded for SDS-PAGE analysis (see **Section 2.4.7**).

3.5.6 Separation of the soluble and insoluble fractions

To gauge the expression levels of the scFvs in the soluble and insoluble fractions of *E. coli*, the cells were lysed and the two fractions separated. 1 mL cultures were sedimented by centrifugation (10 minutes, 13,000RPM) to form a pellet. The cell pellet was re-suspended in cell lysis buffer (see **Table 3.8**) and left shaking for 30 minutes at room temperature. The reaction was then centrifuged at 13,000 RPM on a benchtop centrifuge to separate the soluble and insoluble fractions. The soluble fraction was pelleted using acetone precipitation; ice cold acetone was added to the soluble fraction whereupon the mixture was vortexed and left on ice for 15 minutes. Following incubation on ice, the mixture was centrifuged at 13,000 RPM in a benchtop centrifuge to pellet the soluble protein. The acetone was carefully removed by pipette and the remaining pellet was dried in an incubator at 37 ^{0}C. Equal amounts of 2x SDS-PAGE loading buffer was added to both the insoluble and soluble pellets before the samples being analysed by SDS-PAGE (see **Section 2.4.7**).

3.5.7 Large-scale expression of scFv proteins

The expression of scFvs was scaled up in for *in vitro* characterisation of the different constructs. Lemo21 (DE3) cells were transformed with the relevant plasmid (see **Appenix 8.1**) as described above (see **Section 2.5.2**). A single colony was picked and used to inoculate at 100 mL overnight starter culture in a 250 mL flask containing 100 µg/mL ampicillin and 25 µg/mL chloramphenicol, incubated at 37 ^{0}C and shaking at 200 RPM. 10 mL of starter culture was used to inoculate each 1 L of 10× 1 L sterile LB media containing 100 µg/mL ampicillin and 25 μg/mL chloramphenicol and either 500 or 1000 μM L-rhamnose (depending on construct) in 2 L conical flasks. 10× 1 L cultures were incubated at 37 ^{0}C, shaking at 200 RPM to and OD_{600} of 0.6. ScFv protein expression was induced by the addition of 400 μM IPTG. The cultures were grown for another 4 hours before being harvested by centrifugation at 4,000 RPM (JLA 8.1 rotor) and the soluble and insoluble fractions separated as described by centrifugation at 15,000 RPM (JLA 16.25 rotor).

3.5.8 Purification of scFvs from the soluble fraction

The soluble fractions were filtered through a sterile 0.45 μM filter (Millipore Ltd, Watford, UK) prior to purification. The soluble fractions were then loaded onto a 5 mL HisTrap™ HP column (GE Healthcare, Little Chalfont, UK) using a Pump P-1 peristaltic pump (GE Healthcare, Little Chalfont, UK). The column was prepared by passing 5 column volumes of filter sterilised water followed by 5 column volumes of filtered binding buffer (see **Table 3.8**). A final 5 column volumes of binding buffer was passed through the column to remove any unbound material prior to elution. The column was then attached to an ÄKTAprime plus (GE Healthcare, Little Chalfont, UK) with buffer/lines A and B loaded with binding buffer and elution buffer, respectively. 2.5 mL fractions were collected at a flow rate of 2 ml/min and a maximum pressure of 0.15 MPa. A step-wise elution gradient was used with a 10 % increase of buffer B every 20 mL to a final concentration of 100 % B. Protein-containing fractions, measured by A_{280} were pooled together and concentrated to 10 mL, using 20 mL vivaspin columns (Sartorius UK Ltd, Epsom, UK), before the final protein concentration was measured by A_{280} and the protein dialysed into storage buffer (see **Table 3.8**). Using SnakeSkin pleated dialysis tubing (Thermo Scientific, UK; 3,500 MWCO), the protein samples were dialysed against 5 L of storage buffer (20 mM sodium succinate, 125 mM arginine, pH6). The buffer was exchanged three times with a minimum of 3 hours between changes. The protein was then divided into 1 mL aliquots in 1.5 mL Eppendorf tubes and snap frozen in liquid nitrogen before being stored at -80 ^{0}C.

Chapter 4:
Characterising monoclonal antibody behaviour in the gas-phase

4 Characterising monoclonal antibody behaviour in the gas-phase

4.1 Objectives

The initial ambition of this thesis was to characterise two closely related mAbs, one that is highly aggregation prone and its non-aggregation prone counterpart. Using nESI-IMS-MS, the aim was to monitor any oligomers formed and their assembly pathway, while characterising these species based upon their mass and collision cross-sectional area (CCS) values. Once established, modelling methods would be adopted to build models that could describe the oligomers that are visualised in the gas-phase. From these models, it was hoped to propose a potential mode of oligomer assembly.

4.2 Characterising mAb oligomerisation

As stated in **Section 2.4.3**, the aggregation of monoclonal antibodies and biopharmaceuticals is a major problem to the industry as a whole (Philo and Arakawa, 2009, Saluja and Kalonia, 2008, Wang, 2005). Characterising self-association and aggregation is a significant challenge since the majority of techniques that can yield high resolution information on binding interfaces and sites of aggregation are costly in time; additionally, many of these techniques do not lend themselves to heterogeneous samples and therefore struggle to yield atomistic detail with some systems, depending on their mode of oligomerisation. Other, more common methods of characterising mAb aggregation, such as dynamic light scattering (DLS), size-exclusion chromatography (SEC) and sedimentation velocity analytical ultra-centrifugation (AUC) (Amin et al., 2014, Arthur et al., 2009, Gabrielson et al., 2009) are all excellent methods of determining what species are present in the sample, however, yield no information about the mode of oligomerisation. It was proposed that that IMS-MS could be utilised to bridge the gap between generating higher resolution information while still analysing samples in a fast time scale. IMS-MS is also beneficial as a tool for characterising aggregation as most industries are equipped with mass spectrometers.

4.2.1 Characterising the aggregation propensity of WFL and STT in solution

Before characterising both mAbs in terms of their CCSs in the gas-phase, it was imperative to understand the species which were present in solution at the working concentration. AUC was carried out on both samples at a concentration of 1 mg/mL (6.74 µM) to establish the oligomeric species present in solution (**Figure 4.1**). The AUC experiment was carried out using storage buffer (20 mM sodium succinate, 125 mM arginine, pH 6).

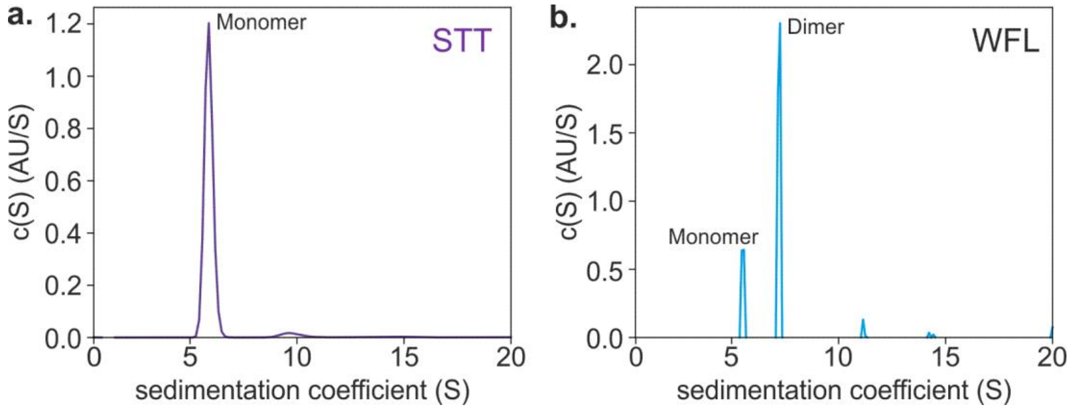


Figure 4.1: Analytical ultracentrifugation of **(a)** STT and **(b)** WFL at 1 mg/mL in storage buffer. n = 1.

The results of the AUC experiments showed that WFL (blue) adopts different conformations from STT (purple) at the working concentration of 1 mg/mL; while predominately dimer, a range of oligomers (up to and including a small amount of tetramer) is observed. By contrast, only a small proportion of STT forms higher order material. As mentioned before, although AUC can indicate which species are present in the sample, no information about the nature of the oligomers or how they assemble can be determined and thus an orthogonal technique is required.

Before analysing the two mAbs using nESI-IMS-MS, the AUC experiment was repeated in MS compatible buffer (150 mM ammonium acetate, pH 6) (**Figure 4.2**) that was chosen to match the same ionic strength as the storage buffer previously used, while remaining at the same pH to directly compare the results.

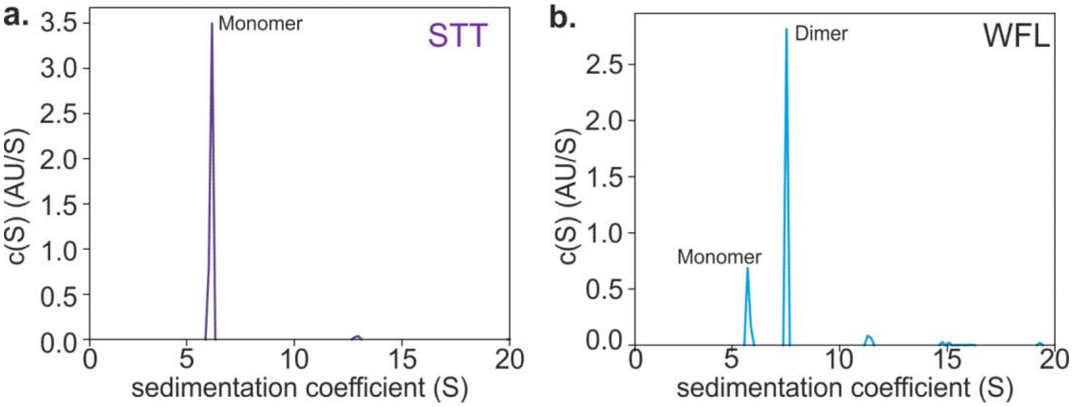


Figure 4.2: Analytical ultracentrifugation of **(a)** STT and **(b)** WFL at 1 mg/mL in ammonium acetate, pH 6 buffer. n = 1.

The AUC data indicates that the distribution of species present in a 1 mg/mL solution is the same in both of the two buffers tested. Since both mAb samples behave in a similar fashion in the two respective buffers, as judged by AUC, the samples were taken for further characterisation using mass spectrometry. Finally, to obtain insight into the possible conformations that WFL and STT might populate, electron microscopy (EM) was employed to visualise the oligomeric species present (**Figure 4.3**).

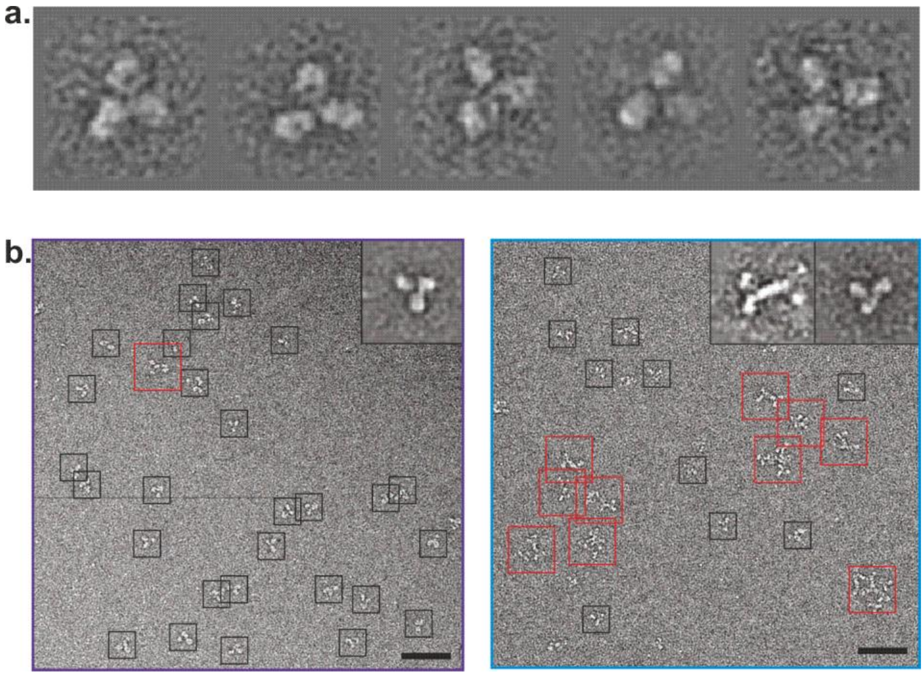


Figure 4.3: Negative stain transmission electron micrographs of samples of WFL and STT at 1 mg/mL in storage buffer. **(a)** Class averages obtained of monomeric mAb structures from STT. **(b)** Electron micrographs highlighting the difference in oligomeric species present in WFL (blue, right) compared to STT (purple, left). Insets show the class averages obtained for the monomeric species in both samples, as well as a dimer species. Red boxes = oligomeric species. Black scale bar = 50 nm. n = 2. The data were acquired by Dr Matt Iadanza (University of Leeds).

From the EM images, mAb images were obtained with sufficient resolution to observe the individual regions (Fab ad Fc) of the antibody (although not which was which) (**Figure 4.3a**). Multiple oligomeric species were observed for the WFL sample from dimeric species up to and including oligomeric tangles, with multiple monomers involved. Interestingly, sufficient dimeric species were present in the sample to generate a class average, shown in the inset of **Figure 4.3b** (right hand side). This class average suggests that the two WFL monomers may be interacting through the binding of two distinct nodes; while it was hypothesized that this may be a Fab-Fab interaction due to the location of the mutations from WFL > STT lie, the resolution is not adequate enough to conclude this confidently and other methodologies were required.

4.3 Monoclonal antibodies and their oligomers have a smaller CCS value than predicted

4.3.1 Native MS of mAbs WFL and STT

Once the buffer conditions had been optimised to reflect the aggregation of WFL and STT observed previously, the mAb samples were analysed by native MS. Independent samples of WFL and STT were diluted to 1 mg/mL and dialysed against 150 mM ammonium acetate over an 18 hours period prior to data acquisition. Spectra were recorded under native-like conditions (see **Section 3.2.3**) in attempt to maintain the oligomeric species present in the sample (**Figure 4.4**).

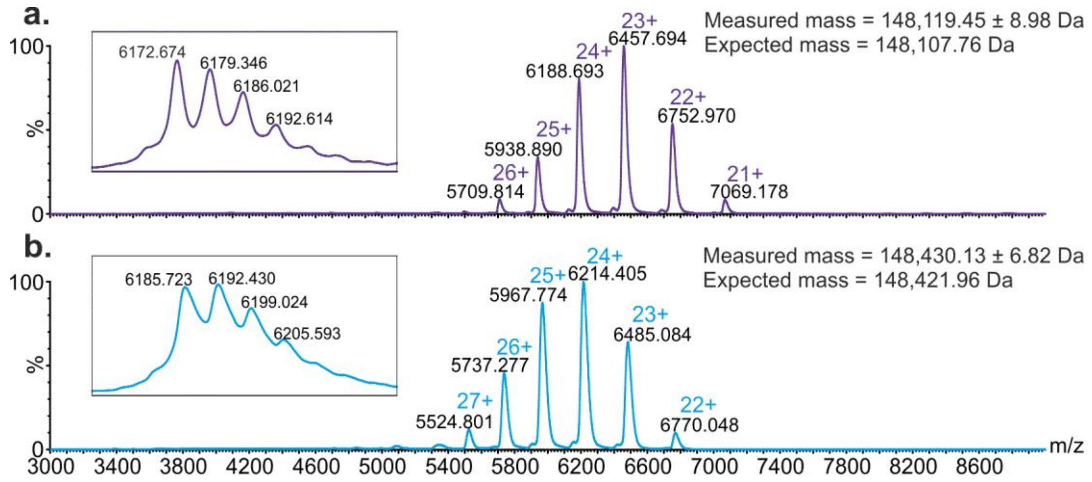


Figure 4.4: Native mass spectra showing the native-like conformation of **(a)** STT (purple, top) and **(b)** WFL (blue, bottom). Samples were analysed at a concentration of 1 mg/mL in 150 mM ammonium acetate, pH 6. Inset shows increased resolution spectrum, focusing on the 24+ charge state ions. Representative spectra of n = 3.

The initial conditions tested show that both WFL and STT exist in a native-like conformation with a narrow charge state distribution focused on the base peak of 23+/24+. However, analysis of both samples indicates that WFL and STT appear monomeric upon MS acquisition with a monomer mass of 148,430.13 ± 6.82 Da (expected = 148,421.96 Da) for WFL and 148,119.45 ±8.98 Da (expected mass = 148,107.76 Da). As AUC and EM experiments observed oligomeric species of WFL, the majority of which is dimer, exist under these buffer conditions in solution (**Figure 4.2**), it was hypothesized that the process of introducing samples into the gas-phase led to an abundance of monomeric species being observed. Alternatively, this may have also been predicted since the three mutations between the two proteins is the substitution of three hydrophobic residues (W30, F31 and L57T) and it has been reported that hydrophobic interactions become weaker in the gas-phase due to the removal of solvent (Loo, 1997). Since the AUC shows such strikingly different results in solution, we can presume that these three residues play a key role in driving the self-association of WFL into dimer and oligomeric species. Furthermore, as analysis by MS requires the removal of water from the ions, the driving force behind any hydrophobic interactions are removed during the transition into the gas-phase. Unless there are other interactions maintaining the interface between the oligomeric species, the visualisation of WFL oligomers would be lost without careful optimisation of the MS instrument to preserve these interactions.

To increase the transmission of oligomeric species into the mass spectrometer, the MS conditions were optimised accordingly. The backing pressure of the instrument was increased (from 3 to 5.5 Mbar) to aid in the transmission of high molecular weight ions while the trap voltages and DC bias were optimised in order to maintain the protein-protein interactions of WFL. Spectra were acquired for WFL and STT under the new conditions in order to monitor the amount of dimer and oligomeric species present in the samples (**Figure 4.5**).

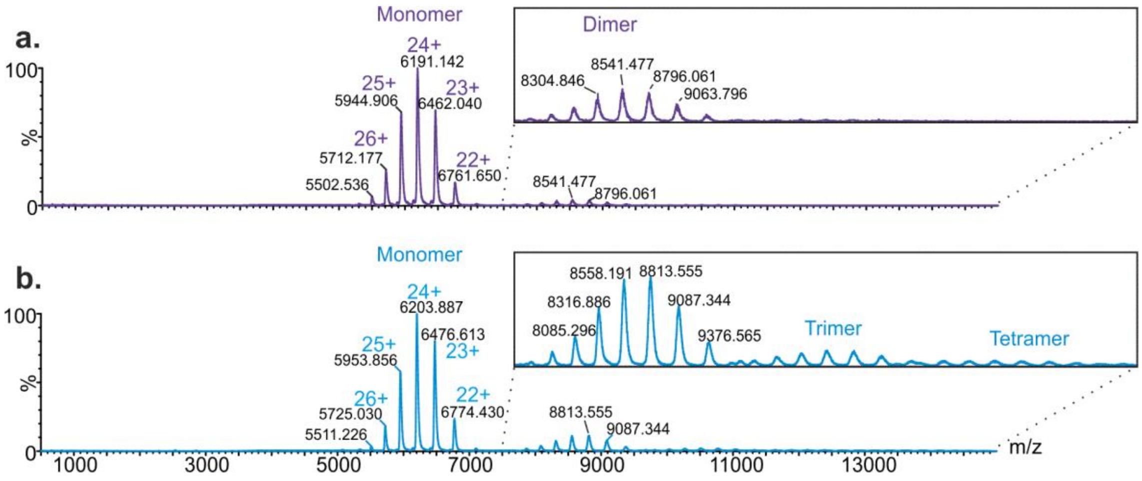


Figure 4.5: Native mass spectra of **(a)** STT (purple, top) and **(b)** WFL (blue, bottom). Oligomeric species of WFL, up to and including tetramer, can be observed for WFL (shown in inset). Representative spectra of n = 4.

Under these new MS conditions a greater proportion of oligomeric species are present in the spectra of WFL (**Figure 4.5**, bottom). While previously there was only a small amount of dimer present in the spectra, the dimeric species present in the spectra (ca. *m/z* 9000) is now > 10 % (of base peak index (BPI)), while higher order species up to and including tetramer are also readily observed. By contrast, only a small percentage of dimer was present in the STT sample, which is consistent with the solution-based AUC results obtained (**Figure 4.1+Figure 4.2**).

As conditions were optimised to visualise the oligomeric populations of WFL in the gas-phase, the mass spectrometer was set to ion mobility mode in order to identify whether there are single or multiple species of oligomers present, and to calculate the CCS of any species identified.

4.3.2 IMS-MS of mAbs WFL and STT

To fully characterise to oligomeric populations identified in the mass spectra, the CCSs were determined to enable the development of models to interrogate how WFL self-associates. Using the previous MS parameters established, IMS-MS measurements of the two samples were acquired along with calibrant proteins in order to estimate CCS values for the two proteins. As oligomeric species up to and including tetramer were identified (~600 kDa), calibrant proteins were chosen that encompass the appropriate MW distribution (150 – 600 kDa); the calibrant proteins used for this study were concanavalin A (CCA), 103 kDa; alcohol dehydrogenase (ADH), 148 kDa; pyruvate kinase (PyK), 237 kDa; glutamate dehydrogenase (GluD), 336 kDa and GroEL, 801 kDa.

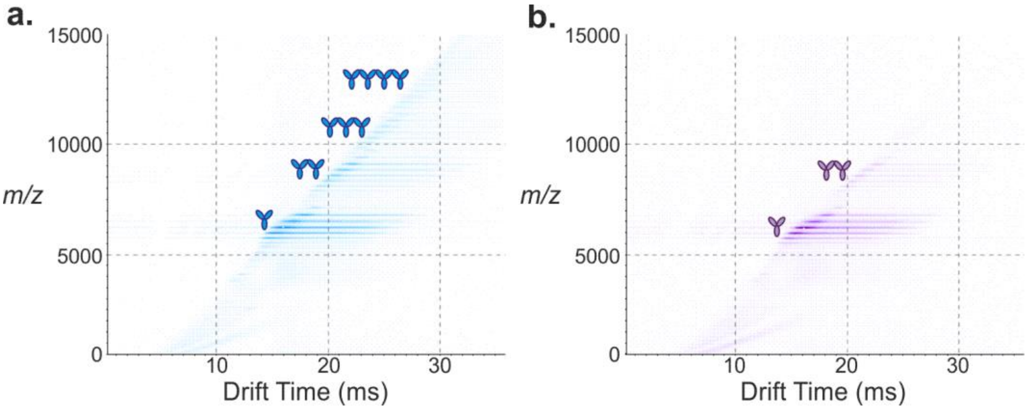


Figure 4.6: Ion mobility driftscope plots indicating the presence of oligomeric species in samples under native conditions. The driftscope plot of **(a)** WFL highlighting the presence of up to and including tetrameric species while only up to dimeric species can be observed for **(b)** STT. (Y) = IgG monomer. Spectra were acquired of WFL and STT in 150 mM ammonium acetate, pH6 with a trap voltage of 40 V, IMS gas flow of 1 mL/min and a ramped wave height from 5-30 V. Representative date of n = 3.

It is evident from the IMS-MS driftscope data (**Figure 4.6**) that there is only one conformation detected for each oligomeric species observed. Under the conditions used to optimise visualisation of the oligomeric species of the mAb samples, tailing is observed behind the main peaks in the driftscope plots, indicative of some unfolding of the proteins. To investigate this, a trap ramp experiment was used to attempt to fully unfold the protein and monitor any unfolding events and at what voltages these occur (**Figure 4.7**). By extracting all of the arrival time distributions (ATDs), and plotting these sequentially versus collision energy, a significant shift in the ATD indicative of an unfolding event is observed (**Figure 4.7**) (Niu and Ruotolo, 2015, Tian et al., 2015).

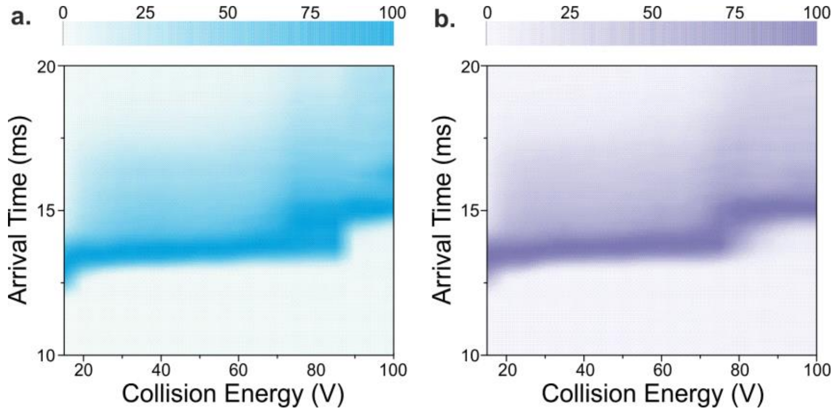


Figure 4.7: Contour plots displaying the unfolding profiles of **(a)** WFL and **(b)** STT in the gas-phase over a trap voltage range of 15-100 V. Unfolding events can be identified around 70-80 V. Representative data of n = 3.

Although tailing begins between 15-20 V, an unfolding event is not observed until around 70-80 V which is double the voltage employed for the IMS-MS experiments. This gave confidence that although there might be some unfolding occurring under the conditions used, the arrival time used for the CCS estimation is, presumably, reflective of the native species. The CCS values of WFL and STT were calculated for the 21+ charge state and found to be being 6820 Å^2 and 6843 Å^2, respectively. The CCS values of the oligomeric species were also calculated and plotted alongside the calibrant proteins used for the study (**Figure 4.8**).

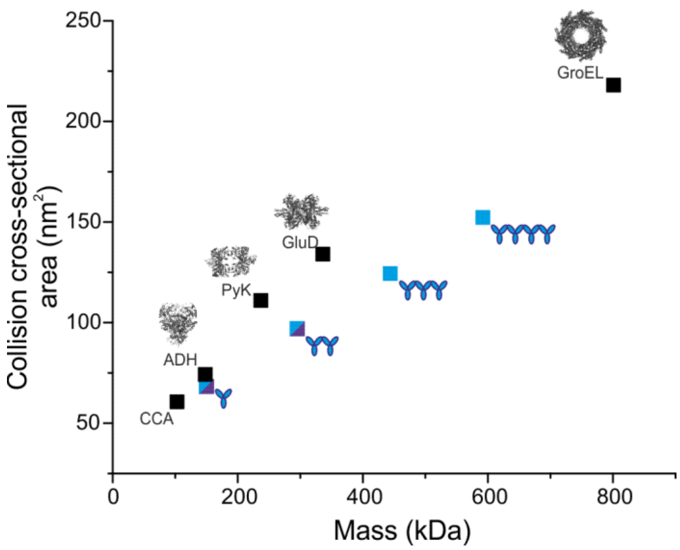
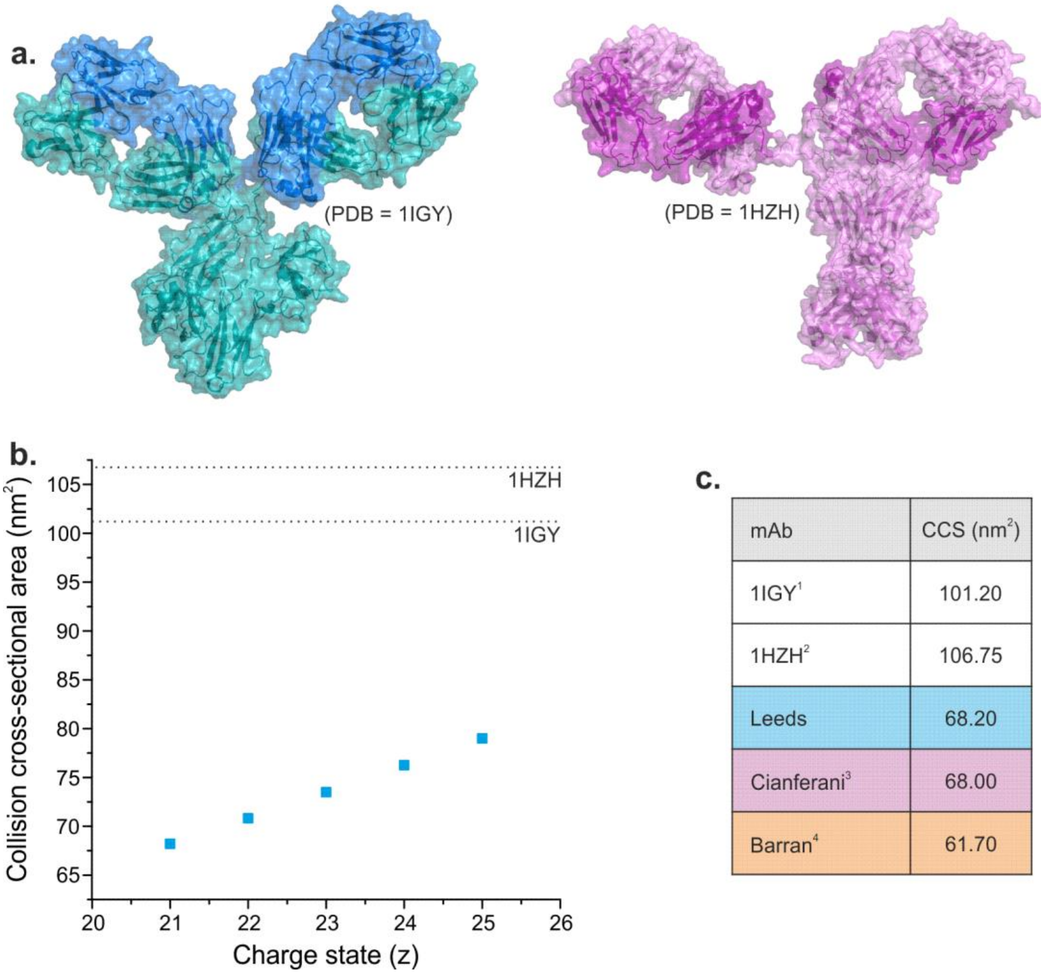


Figure 4.8: Plot of CCS against molecular mass of mAb samples tested and calibrant proteins used for CCS measurements. mAb CCS values up to and including tetramer for WFL are plotted in either blue (WFL) or purple (STT). Calibrant proteins are plotted in black with their respective structures shown. CCA = concanavalin A, ADH = alcohol dehydrogenase (PDB – 4W6Z (Raj et al., 2014)), PyK = pyruvate kinase (PDB – 2E28 (Suzuki et al., 2008)), GluD = glutamate dehydrogenase (PDB – 1HWZ (Smith et al., 2001)) and GroEl (PDB – 5DA8).

Interestingly, the CCSs measured for monomeric WFL and STT fall below that of ADH which has similar mass (148 kDa). At the beginning of this study, there were no published CCS data for IgGs; however, there has since been work from the Cianferani (Debaene et al., 2013, Debaene et al., 2014) and Barran (Pacholarz et al., 2014) groups reporting their CCS data for IgG proteins. Both report similar values (**Figure 4.9c**) which gives confidence in the values calculated in this study. However, due to the structure of an IgG1, one would expect the cross-section measured to be significantly larger. Furthermore, it was hypothesized that the CCS would have been larger than that of ADH since it has a similar mass (148 kDa), yet a globular structure.

The theoretically predicted CCS values are often reported alongside experimental results as an indicative measure of native structure in the gas-phase, with a higher degree of similarity between values suggesting a more native-like structure. Due to their multidomain structure, IgGs have been shown to populate an ensemble of structures (Clark et al., 2013) due to the inherent flexible hinge region of the molecule which is paramount for their function. Therefore, theoretical values were generated for a selection of IgG1 molecules from the PDB in order to establish a range of theoretical values for conformers that an IgG might populate. Using the MOBCAL software (Shvartsburg and Jarrold, 1996, Mesleh et al., 1996), the CCSs of two known IgG1 structures (PDB = 1IGY (Harris et al., 1998) and 1HZH (Saphire et al., 2001), **Figure 4.9a**) were calculated to compare with experimental values (**Figure 4.9b**). The CCSs of the two structures were calculated using the PSA scaling factor from the PA output of MOBCAL (see **Section 3.2.5**).



mAb	CCS (nm^2)
1IGY[1]	101.20
1HZH[2]	106.75
Leeds	68.20
Cianferani[3]	68.00
Barran[4]	61.70

Figure 4.9: Collisional cross-section measurements of mAb samples. **(a)** PDB structures of two IgG molecules used to calculate theoretical CCS values using the MOBCAL software. **(b)** Plot of CCS of monomeric WFL as a function of increasing charge state. The predicted values of the two PDB structures are represented by dotted lines. **(c)** CCS values from the theoretical predictions from references 1 = (Harris et al., 1998) and 2 = (Saphire et al., 2001), as well as experimental data from this study and references 3 = (Debaene et al., 2014) and 4 = (Pacholarz et al., 2014).

It is interesting to note that there is a significant difference in the predicted CCS of the two PDB structures which is again a testament to the inherent flexible nature of the molecules and their collapse in the gas-phase. The idea was to use multiple PDB structures in order to obtain a range of predicted CCS values so that could be compared with the experimental results. However, it is clear that the experimental results are far from agreement with the range between these two structures. **Figure 4.9b+c** indicate that there is a 30 % discrepancy in the measured cross-sections (Leeds and Cianferani data) compared with the predicted CCS of the 1IGY structure. Although there is also a small discrepancy between the experimental data (**Figure 4.9c**), it is important to note that both the Leed's and Cianferani group's data were measured using Waters instruments (Syanpt G1 and Synapt G2, respectively) with a TWIMS device. The Barran group have an in-house modified Water's instrument that has a traditional linear drift cell where the CCS of an ion can calculated directly from its drift time. However, the experimental values across the three groups are in agreement with each other overall, suggesting that the discrepancy in predicted and observed CCS does not depend on the antibody sequence or the mass spectrometer used.

4.3.3 Modelling the mAb oligomeric species

In an attempt to understand the data, simple models were built using the MOBCAL software to interrogate the experimental data (**Figure 4.10**). Each region of the IgG molecule (2xFab and Fc) was represented as a single spherical "atom" with the mass of the individual regions (**Figure 4.10a**) (see **Section 3.3.1**). These "atoms" were then arranged in 3D space, with the same centre on the z-plane and the x- and y-coordinates altered until a theoretical CCS value was obtained that was in closer agreement with measured experimental data. Once obtained, four different models were constructed that could potentially explain WFL's mode of self-assembly into larger oligomers (**Figure 4.10**). The four different models generated were two stacking models (one with the sequential antibodies in opposite orientations), an open "arms" model and a closed model (**Figure 4.10**). The rationale behind these models was that the all mutations between WFL and STT are all located in the VH domain and it was hypothesized that self-association was likely to be localised to this region of the mAb.

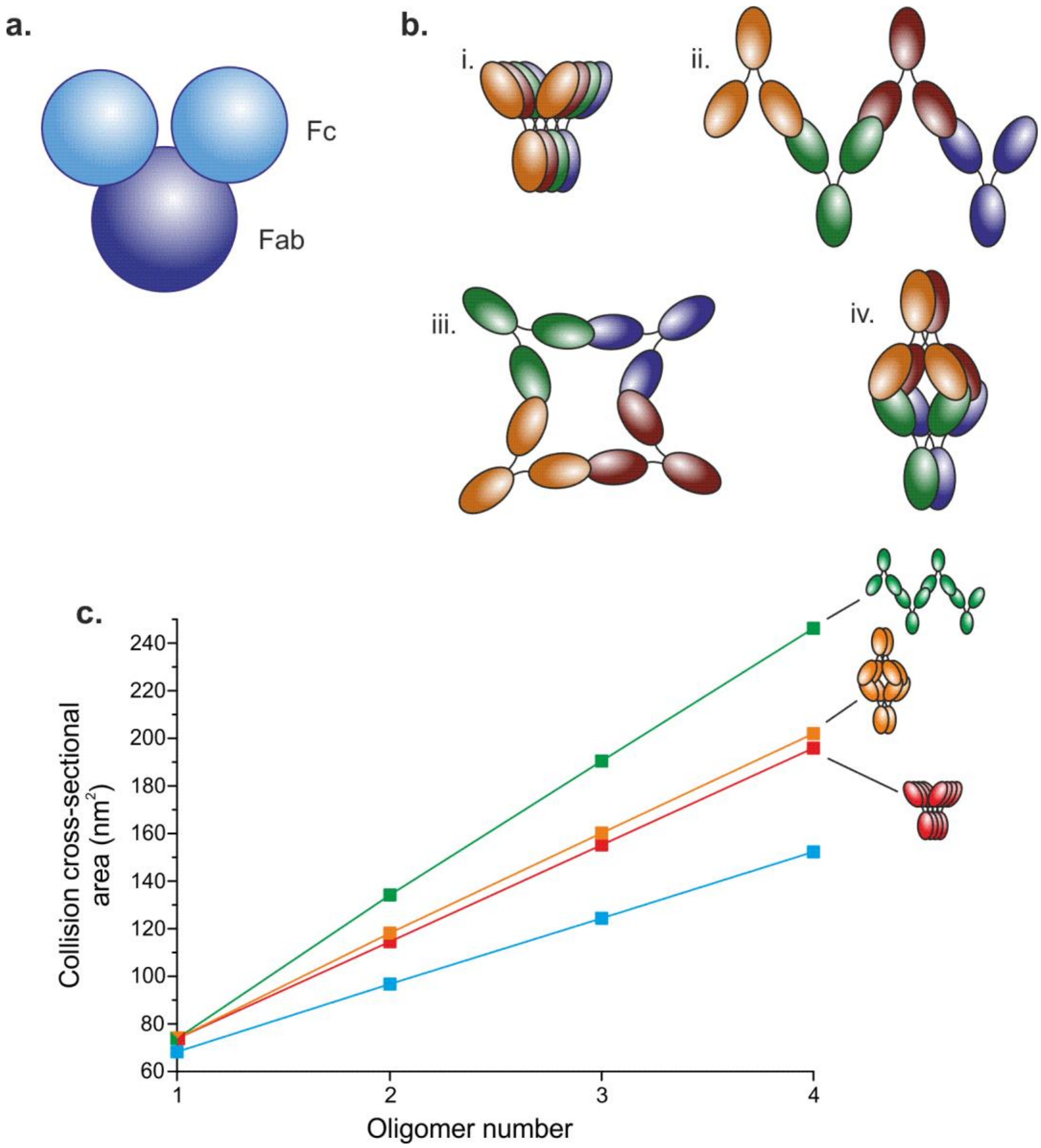


Figure 4.10: Modelling of the mAb oligomeric CCS data. **(a)** Model of mAb monomer showing the Fab and Fc regions of the mAb being represented as single "atoms". **(b)** Schematic representation of four models hypothesized for mAb aggregation: (i) stacked model, (ii) overlapping arms model, (iii) closed model and a (iv) opposite orientation stacked model. **(c)** Plot of CCS vs oligomer number for the different models generated. Stacked model = red, opposite orientation stacked model = orange and overlapping arms model = green. Measured CCS data are in blue.

Out of the four models generated, the closed model was ruled out since it was presumed that higher order oligomerisation would be difficult to achieve if the mechanism proposed was correct; a closed state would have to open for any further self-association, which seems unlikely. Upon comparing the data to the generated models, it is clear that none of these models explain the data observed adequately, with the stacking models being closest to the experimental data.

Interestingly, it has been reported in the literature that the measured CCS of a mAb monomer was in excellent agreement with a spherical model (Debaene et al., 2014), based upon a sphere with the given mass of the protein in question. To understand whether a spherical model can explain the experimental data generated, spherical models were built based upon the mass of WFL and the oligomeric species observed (**Figure 4.11**). As well using the method used in the published report (denoted here as the Ruotolo method (Ruotolo et al., 2008)), a second method of generating spherical models was also used (denoted as the Benesch model (Benesch et al., 2007)). The Rutolo method generates a spherical model by manipulating the molecular mass of a protein of interest using empirically derived factors (see **Section 3.3.2**, **Equation 3.2**). By contrast, the Benesch model utilises the mathematical basis of calculating the volume of a sphere based upon the molecular mass of a protein of interest (see **Section 3.3.2**, **Equation 3.3**).

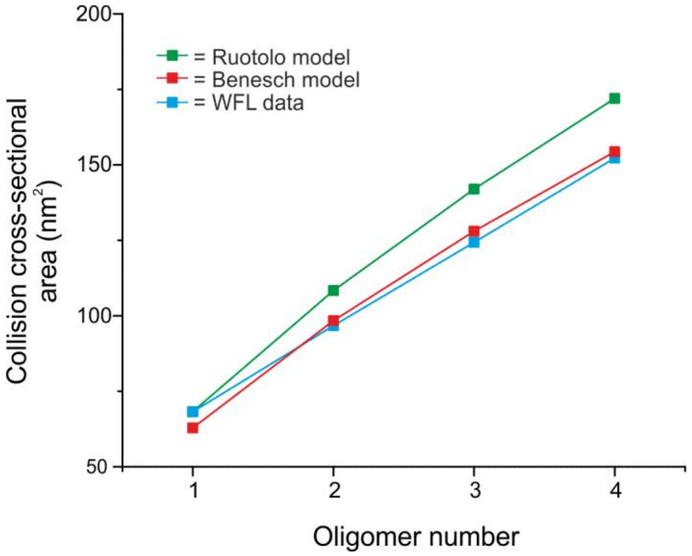


Figure 4.11: Plot of CCS against oligomer number for generated spherical models. Ruotolo method (Ruotolo et al., 2008) = green, Benesch model (Benesch et al., 2007) = red and measured CCS data =blue.

As indicated in the literature (Debaene et al., 2014), the spherical model of the monomer using the Ruotolo method is in excellent agreement with the measured data. However, it is clear that this does not hold when extrapolated to the oligomeric species, as it deviates further from the data with extending oligomer number. Intriguingly, although the Benesch model shows less agreement for the monomeric CCS, it agrees well with all of the oligomeric species measured. The agreement between the experimental data and the spherical model again suggests that there is a significant collapse of the IgGs into more compact structures in the gas-phase, resulting in CCS values that fall dramatically below the values expected, based on their 3D structure.

4.3.4 IgG collapse in the gas-phase

To understand the compaction observed in the gas-phase through the experimental CCS data, molecular dynamics (MD) was employed in order to probe the behaviour of the molecules in the absence of water. By carrying out MD simulation experiments in the absence of buffer or water, *in vacuo*, the conditions that the ions experience in the mass spectrometer can be mimicked, with the aim of achieving a glimpse of how the molecules might behave in the gas-phase.

As this methodology was being explored, work was published from the Barran group in Manchester using the same approach and showing that mAb structures do appear to collapse in the gas-phase (**Figure 4.12**), consistent with the hypothesis from the experimental data presented here.

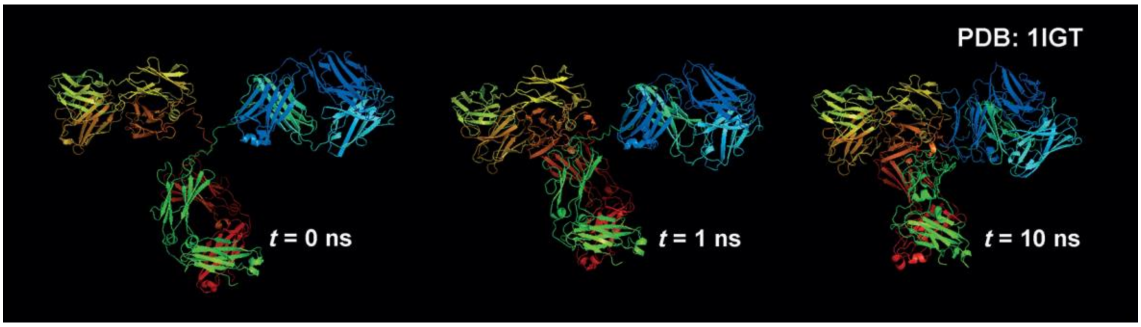


Figure 4.12: Visualisation of IgG collapse *in vacuo* over a 10 ns simulation. Figure was taken from (Pacholarz et al., 2014). Green and red = heavy chains, blue and yellow = light chains.

The data from the published work reveal that the extent of the collapse can be attributed to the flexible hinge region of the mAb structures tested (PDBs = 1IGY, IgG1; IHZH, IgG and 1IGT, IgG2 (Harris et al., 1997, Harris et al., 1998, Saphire et al., 2001)). It is perhaps unsurprising that the collapse observed is focused on the hinge region of the molecules, since they are known to be highly flexible regions of the proteins (Sandin et al., 2004, Janeway et al., 1997).

To scrutinize the effect of the hinge region on the observed collapse of the molecules in the gas-phase MD simulations, it was hypothesized that if the mAbs were separated into their different possible regions (Fab, Fc and F(ab')2), then the regions still containing the hinge would have a preponderant effect on compaction. Before examining the individual regions of the structures, it was important to assess whether the same degree of collapse was observed in simulations. An IgG1 crystal structure (1IGY) (Harris et al., 1998) was used to carry out a 10 ns MD equilibration experiment after minimization (**Figure 4.13**) (see **Section 3.3.4**).

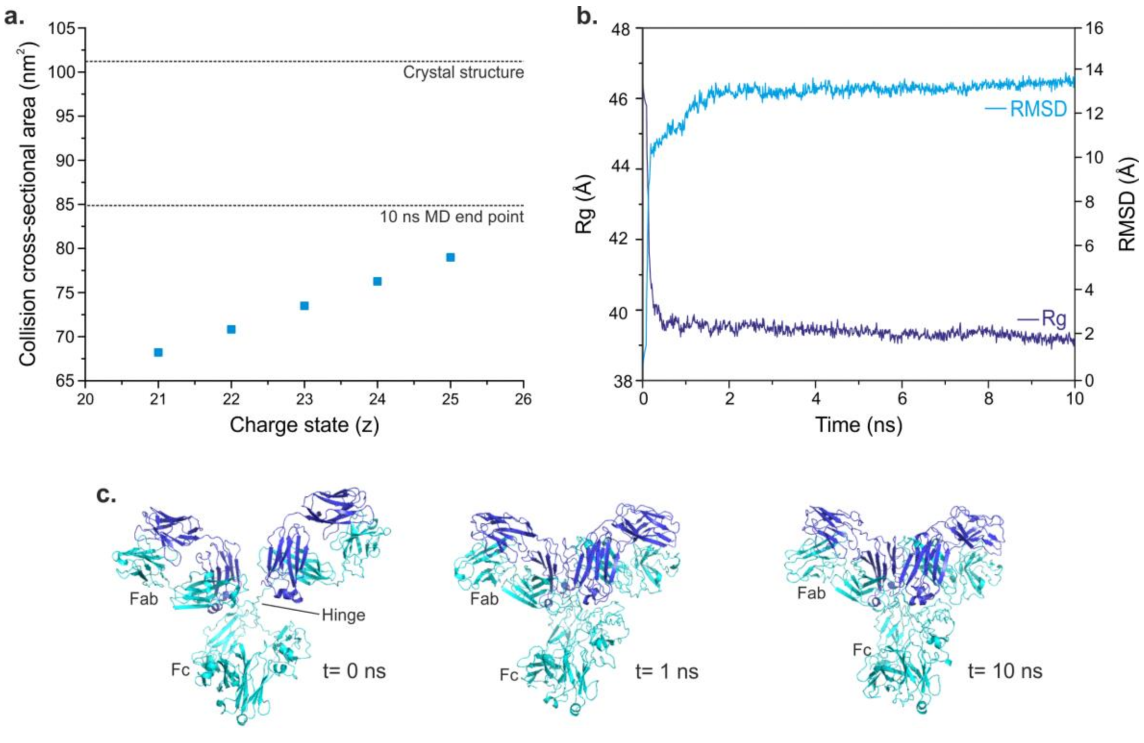


Figure 4.13: *In vacuo* collapse of mAbs in the gas phase. **(a)** Plot of CCS against the measured charge states of WFL. The predicted CCSs of the IgG crystal structure (1IGY) and the equilibrated structure, after 10 ns, are represented as dotted lines. Measured data are plotted as blue squares. **(b)** Plots of the RMSD and Rg of the 1IGY structure as a function of simulation time (in ns). **(c)** Visual representation of the structure at different time points of the simulation showing the collapse around the hinge region.

In the gas-phase MD simulations performed (see **Section 3.3.4**), the IgG structure collapses to a significantly smaller CCS in comparison to the value that is predicted from crystal structure (**Figure 4.13a**). Notably, the CCS at the end of the collapse (84.86 nm^2) is in agreement with the CCS established from the MD experiments carried out from the Barran group (84.13 nm^2) (Pacholarz et al., 2014). Furthermore, the RMSD plot (**Figure 4.13b**) indicates that the structure/system has reached equilibrium in the MD simulations (Daggett and Levitt, 1993, Walton and Vanvliet, 2006). Finally, the greatest extent of collapse observed from the *in vacuo* simulations involved the hinge region of the molecule, between the CH1 and CH2 domains (**Figure 4.13c**), in agreement with the published work.

4.4 In vacuo MD simulations of IgG fragments

To determine the exact role of the hinge region in the collapse of the molecule, the mAb was digested into its constituent regions, and the CCSs measured both experimentally and computationally.

Using various proteases, mAbs can be partially digested to yield the individual regions that make up the molecule. For example IdeS, which cuts specifically below the disulphide bridges in the hinge region, can be used to generate and purify F(ab')2 regions of antibodies (Chevreux et al., 2011). By contrast, Lys-C which cuts at the C-terminal of lysine residues above the hinge region, can be utilised to yield two Fab fragments, as well as generate an intact Fc region (Gadgil et al., 2006) (**Figure 4.14**).

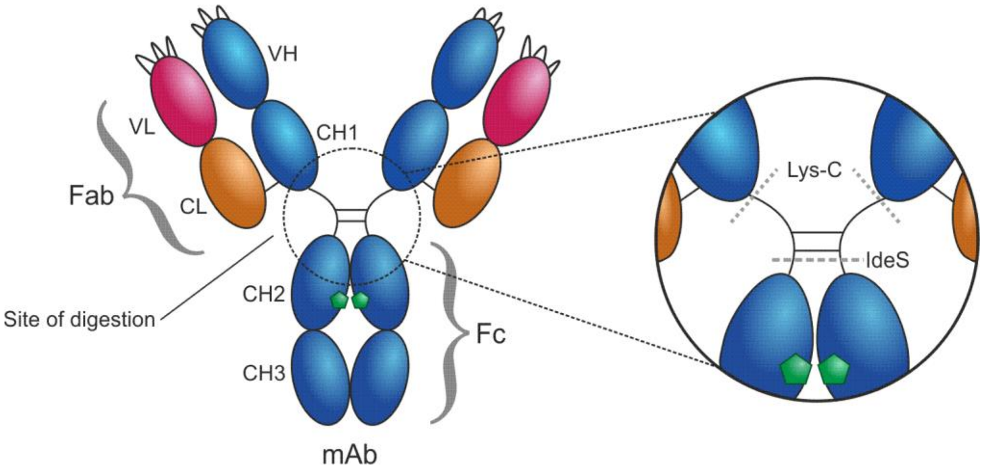


Figure 4.14: Sites of enzymatic digestion by IdeS and lys-C. Lys-c (grey dotted line) digests above the hinge region while IdeS (grey dashed line) cuts below the hinge region.

By utilising these enzymes, IgG samples can be digested and identical IMS-MS measurements carried out in order to gauge the effect the hinge region of the molecule has on the overall collapse of the IgG. Lys-C was used to digest WFL into its Fab and Fc region before dialysing into 150 mM ammonium acetate, pH 6 and IMS-MS measurements made (see **Section 3.4.3**) (**Figure 4.15**).

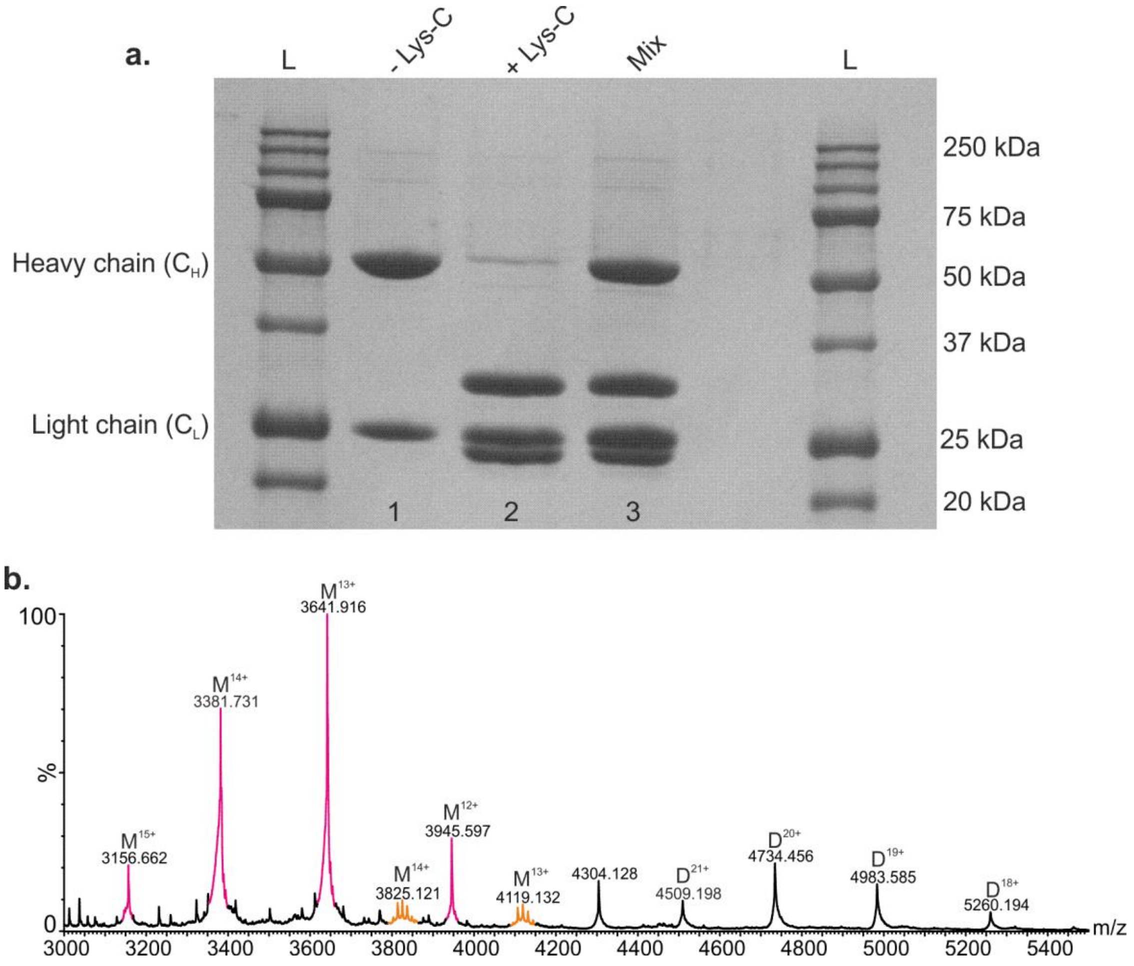


Figure 4.15: Enzymatic digestion of mAb WFL using endoproteinase Lys-C. **(a)** SDS-PAGE showing the specific cleavage of the heavy chain. L= MW ladder, lane 1 = WFL alone, lane 2 = WFL incubated with Lys-C, lane 3 = mix of digested and undigested mAb and lane 4 = buffer blank. **(b)** Mass spectrum showing the products generated from the lys-C digest. Pink = Fab and orange = Fc. Representative spectrum of n = 2.

There is a clear distinction of the two regions of the mAb, indicated in the denaturing SDS-PAGE (**Figure 4.15**) and also in the resulting mass spectrum (Fab mass = 47,338.26 ± 3.89 Da, Fc mass = 53,382.51 ± 7.57 Da). Interestingly, a Fab dimer is observed in the spectrum (**Figure 4.15b**, 4,400-5,300 *m/z*) indicating that it is the Fab region, where the mutations between WFL and STT lie, of the molecule that is responsible in driving the self-association. However, some Fab dimer was also observed in a digested sample of STT. As mentioned above, Lys-C digests above the hinge region of the molecule and thus the Fc region of the molecule still has regions of the flexible linker left intact (~20 residues). Therefore, it was hypothesized that the

Fc region of the molecule would experience a greater degree of collapse in the gas-phase in comparison to the Fab regions, as the latter regions lack a hinge region (see **Figure 4.14**). After confirming the digestion via MS and SDS-PAGE, the digested sample was analysed via IMS-MS and the cross-sections for the two regions estimated (**Figure 4.16**).

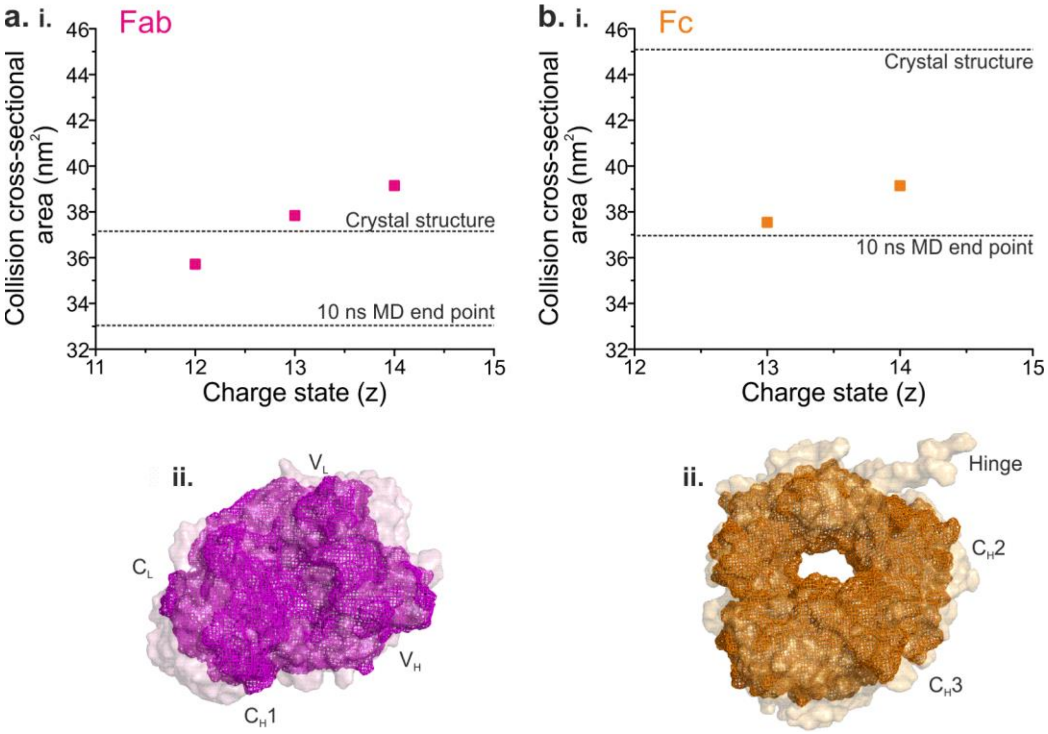


Figure 4.16: *In vacuo* collapse of the Fab and Fc regions of an IgG. **(a)** (i) Plot of CCS against charge state for the Fab region of an IgG. The predicted CCSs of the Fab structure (1IGY) and the equilibrated structure, after 10 ns, are represented as dotted lines. Measured data are plotted as pink squares. (ii) Visualisation of the collapsed structure is shown (solid mesh) against the structure before the equilibration (transparent surface). **(b)** (i) Plot of CCS against charge state for the Fc region of an IgG. The predicted CCSs of the Fab structure (1IGY) and the equilibrated structure, after 10 ns, are represented as dotted lines. Measured data are plotted as orange squares. (ii) Visualisation of the collapsed structure is shown (solid mesh) against the structure before the equilibration (transparent surface).

Similar to the full IgGs, we observe a collapse of both the Fab and Fc regions of the mAbs (**Figure 4.16**). The extent of collapse between the two molecules differs significantly however; a larger collapse is observed for the Fc region (crystal structure predicted CCS = 45.08 nm^2, MD end point predicted CCS = 36.96 nm^2) when compared to the collapse of the Fab region (crystal structure predicted CCS = 37.15 nm^2, MD end point predicted CCS = 33.03 nm^2). Upon comparing the experimental data with the predicted values, from both the crystal structure and MD end points, it is observed that the Fab CCS data is in closer agreement with the crystal structure prediction, while the Fc values are in closer agreement with the MD. This is consistent with the previous hypothesis that a greater extent of collapse would occur when the hinge remains partially present with the digested region of the molecule.

To explore this hypothesis further, the F(ab')2 region, as previously mentioned. By digestion below the hinge of the IgG, the two Fab arms remain connected by the inter-heavy chain disulphide bridges (see **Figure 2.13**). This leaves the majority of the hinge region intact between the two heavy chains, and so it is likely that a significant collapse would be observed. While the experimental CCS for this region of WFL was not calculated, the *in vacuo* MD simulations were carried out on the 1IGY structure to visualise the theoretical extent of collapse (**Figure 4.17**).

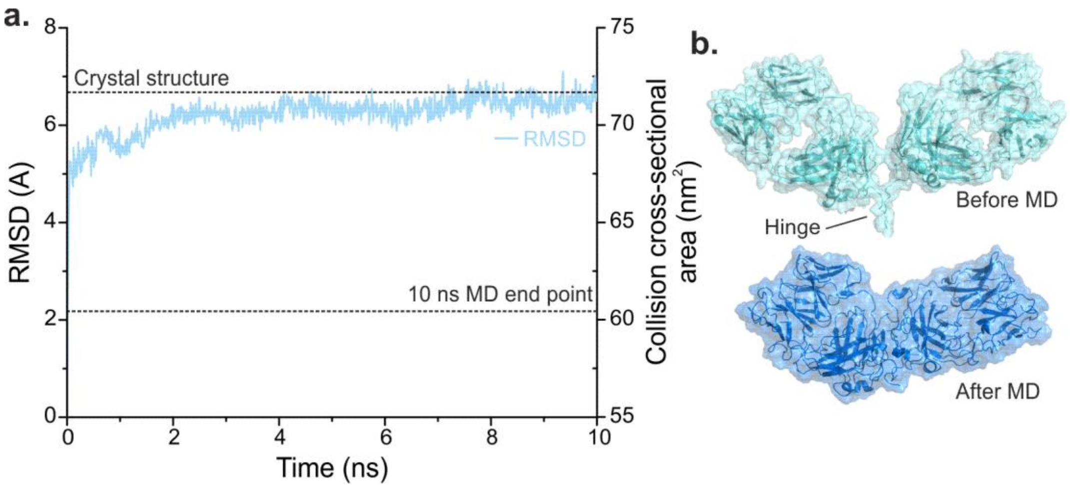


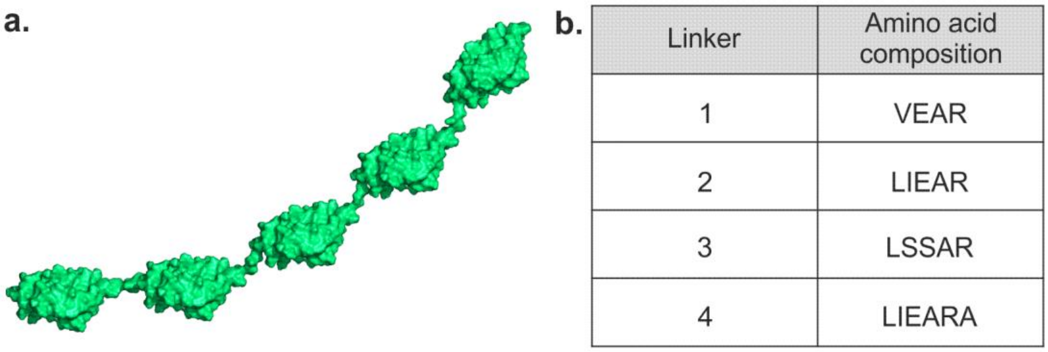
Figure 4.17: *In vacuo* collapse of the F(ab')2 structure. **(a)** Plot of RMSD against simulation time. The predicted CCSs of the F(ab')2 structure (1IGY) and the equilibrated structure, after 10 ns, are represented as dotted lines. **(b)** Visualisation of the collapsed structure is shown (solid mesh) with the structure before the equilibration shown below (transparent surface).

As predicted, **Figure 4.17** shows a wide discrepancy between the crystal structure and MD end point predicted CCS values with a difference of 11.25 nm^2 (**Figure 4.17a**, dashed lines). Upon visualising the MD results, it is transparent that although the Fab regions themselves collapse (see **Figure 4.16**), this large collapse can be mainly attributed to the “loss” of the hinge region in the molecule by the two Fabs closing together towards the centre of the molecule (**Figure 4.17**).

The results here, along with the full IgG data, indicate that both the Fab and Fc regions collapse to some extent in the gas-phase, but the Fc region collapses to a larger extent. To understand whether this was unique to mAbs, other constructs (while maintaining the Ig subunit) were examined both experimentally and through *in vacuo* simulations.

4.5 Characterising gas-phase collapse as a function of linker length and composition

In an attempt to understand the collapse of mAbs in the gas-phase, it was decided to characterise other molecules with flexible linkers using the same experimental procedure. For these experiments, an I27 concatamer consisting of five I27 subunits connected via small (4-6 residue) amino acid linker regions was used (denoted hereon as $(I27)_5$). The I27 protein is found in titin (Improta et al., 1996), an abundant protein found in striated muscle which aids the function of muscle contraction in sarcomeres (Herzog, 2014). $(I27)_5$ is a mechanically robust protein and has therefore been used for AFM and mechanical stability studies (Oberhauser et al., 1999, Brockwell et al., 2002). Here we use the concatamer to understand gas phase collapse as the subunits are connected through small flexible linkers (4-6 amino acids) (**Figure 4.18**).



Linker	Amino acid composition
1	VEAR
2	LIEAR
3	LSSAR
4	LIEARA

Figure 4.18: Generation of the $(I27)_5$ structure. **(a)** The $(I27)_5$ structure was generated by adding the required amino acid linkers onto the C-terminus of the monomeric subunit (PDB = 1TIT) before stitching the molecules together. Stitching together of the subunits was performed in COOT by Dr Claire Windle (University of Leeds). **(b)** Table of linker composition added to the C-terminus of the monomer subunits.

As the individual subunits of $(I27)_5$ are separated by amino acid linkers, it was hypothesized that this polymer of Ig domains may behave similarly to the mAb and its fragments; where the CCS value of the protein was expected to be lower than that predicted from the modelled structure. A sample of $(I27)_5$ (kindly provided by Dr David Brockwell, University of Leeds) was dialysed into 150 mM ammonium acetate and its properties in the gas-phase measured through nESI-MS and IMS-MS (**Figure 4.19**).

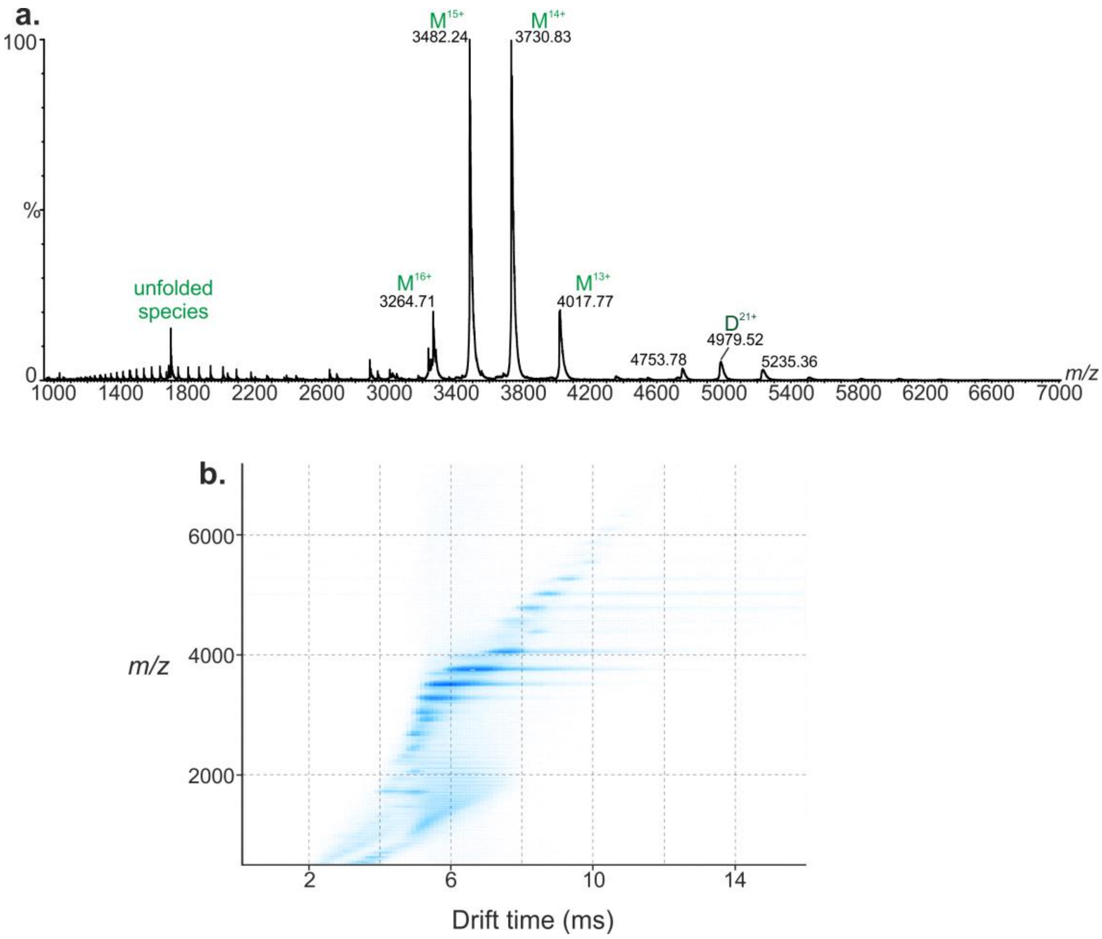


Figure 4.19: Native nESI-MS and IMS-MS spectra of (I27)$_5$ **(a)** nESI-MS spectrum of (I27)$_5$ showing monomeric (green) and dimeric (dark green). **(b)** Driftscope plot of the (I27)$_5$ protein indicating the drift time of the different species. Representative spectra of n = 3.

Figure 4.19 shows that a significant amount of (I27)$_5$ forms dimeric species which are also observed in the IMS-MS measurements. The measured mass of (I27)$_5$ was 52,076.74 ± 0.41 Da, in agreement with previously measured data (52,082 Da). Using the lowest molecular weight species present, the CCS was calculated for the (I27)$_5$ monomeric species (**Figure 4.20**). These data were used to test the hypothesis of gas-phase collapse being attributed to the subunits of proteins being joined through flexible linkers. Since no crystal or NMR structure exists of the (I27)$_5$ concatamer for the CCS prediction, a structure was generated through building the linker regions (see **Figure 4.20b**) and stitching the subunits together to create the full (I27)$_5$ (see **Section 3.3.3**).

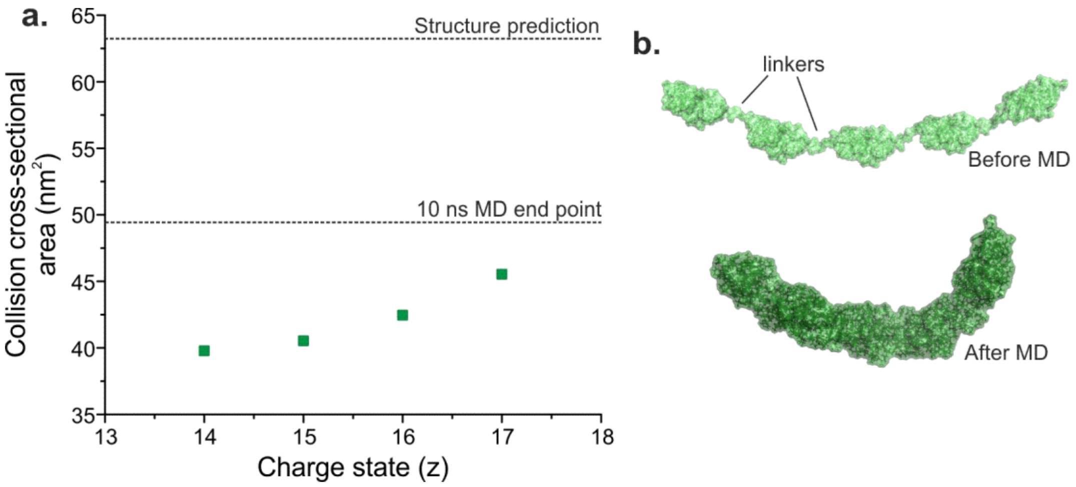


Figure 4.20: *In vacuo* collapse of the (I27)$_5$ structure. **(a)** Plot of CCS against charge state for the (I27)$_5$ concatamer. The predicted CCSs of the structure and the equilibrated structure, after 10 ns, are represented as dotted lines. Measured data are plotted as green squares. (ii) Visualisation of the collapsed structure is shown (solid mesh) against the structure before the equilibration (transparent surface).

As observed for the mAb samples and its constituent parts, the CCS of (I27)$_5$ is significantly below the value expected based upon the generated structure (**Figure 4.20b**); suggesting that (I27)$_5$ also collapses in the gas-phase. The MD simulations indicate that although the structure coils around itself, aiding in the collapse, the individual subunits compact towards the centre of the molecule through collapse around the linker regions. This agrees with the hypothesis that the collapse observed for biomolecules in the gas-phase is mainly attributed due to the flexible hinge regions present in the molecules. (I27)$_5$ has been also been reported as a flexible molecule, shown to adopt various conformations via electron microscopy (Li et al., 2001).

To understand the role that these flexible linker regions play in the observed gas-phase collapse, dumbbell-like models were created of two I27 monomers (PDB = 1TIT (Improta et al., 1996)) connected by linkers of varying length and composition; The proteins were connected through either glycine-serine (GS) or poly-proline (PP) linkers of varying length. The aim was to characterise the effect of linker composition, first through computation, and if a significant difference in CCS was observed to express and purify the proteins and measure their CCSs experimentally for comparison. Since poly-proline motifs are known to exhibit more rigid qualities (Kay et al., 2000, Moradi et al., 2009), it was predicted that the structures linked together through a poly-proline linker would experience less of a collapse. The structures were initially linked together with a two amino acid linker: a glycine-serine or proline-proline linker to test this hypothesis before moving onto longer linker lengths (**Figure 4.21**).

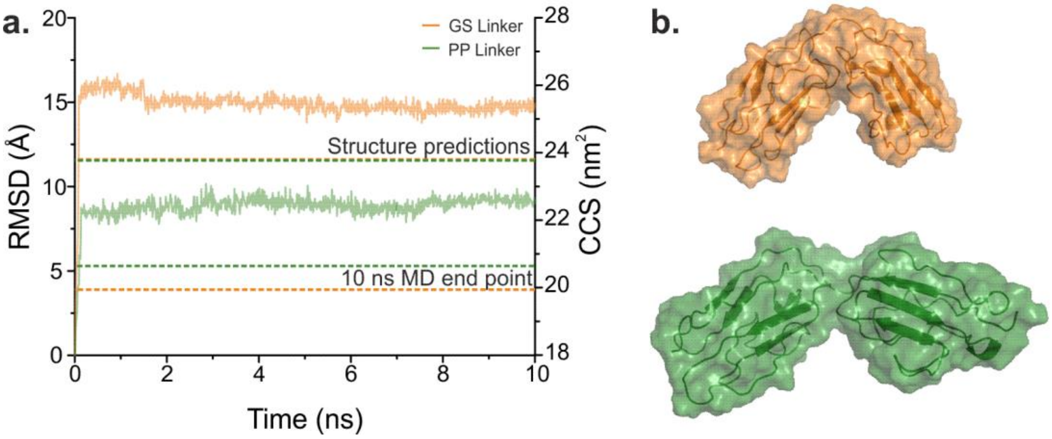


Figure 4.21: *In vacuo* collapse of two I27 domains connected through two amino acid linkers. **(a)** Plot of RMSD and CCS of the two structures as a function of simulation time. The RMSD values are plotted as lines whereas the CCS values are plotted as dotted lines. **(b)** Visualisation of the two structures after the 10 ns simulation. GS Linker = orange and PP linker = green.

The predicted CCS of the starting structures were 23.80 and 23.77 nm^2 for the GS and PP linker, respectively. As both predicted CCSs were of similar values, it extrapolates that any differences observed in CCS after the MD simulation are due to the linker composition. **Figure 4.21a** shows that a small difference in CCS is measured between the two constructs, 19.93 and 20.64 nm^2, suggesting a greater collapse may be possible a GS linker present, compared with a PP linker. The extra flexibility of the GS linker allows the two sub-units to form an arc-like structure, whereas the PP linker restricts the structure to a more linear orientation (**Figure 4.21b**).

As some difference could be observed when using a two amino acid linker between the I27 domains, further structures were generated consisting of either a ten or fourteen amino acid linker (**Figure 4.22**) to understand if further differences could be observed.

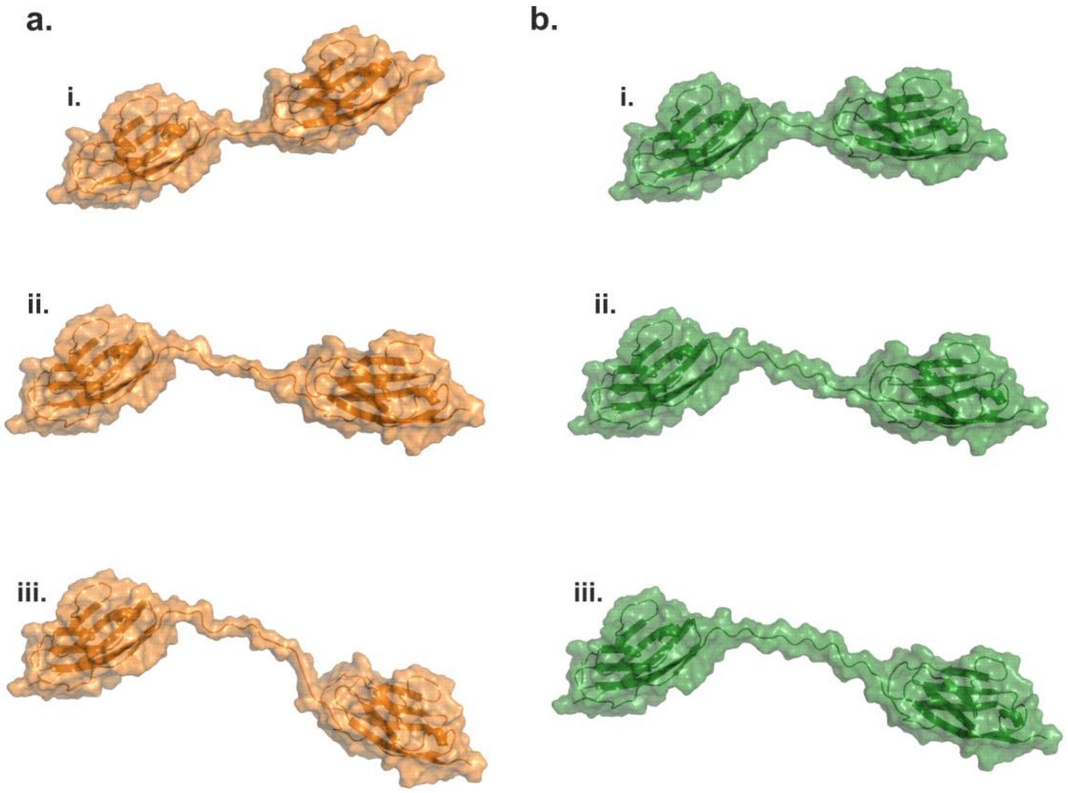


Figure 4.22: Structures of the I27 dimer that were generated to test the extent of gas-phase collapse as a function of linker length and composition. **(a)** Generated structures of the I27 dimer with GS linkers of different lengths. (i) = 2 amino acid, (ii) = 10 amino acid and (iii) = 14 amino acid linker. **(b)** Generated structures of the I27 dimer with PP linkers of different lengths. (i) = 2 amino acid, (ii) = 10 amino acid and (iii) = 14 amino acid linker.

In vacuo MD simulations were carried out on the ten and fourteen length structures and the CCS differences interrogated to examine if gas-phase collapse is enhanced in constructs containing longer linker lengths (**Figure 4.23**).

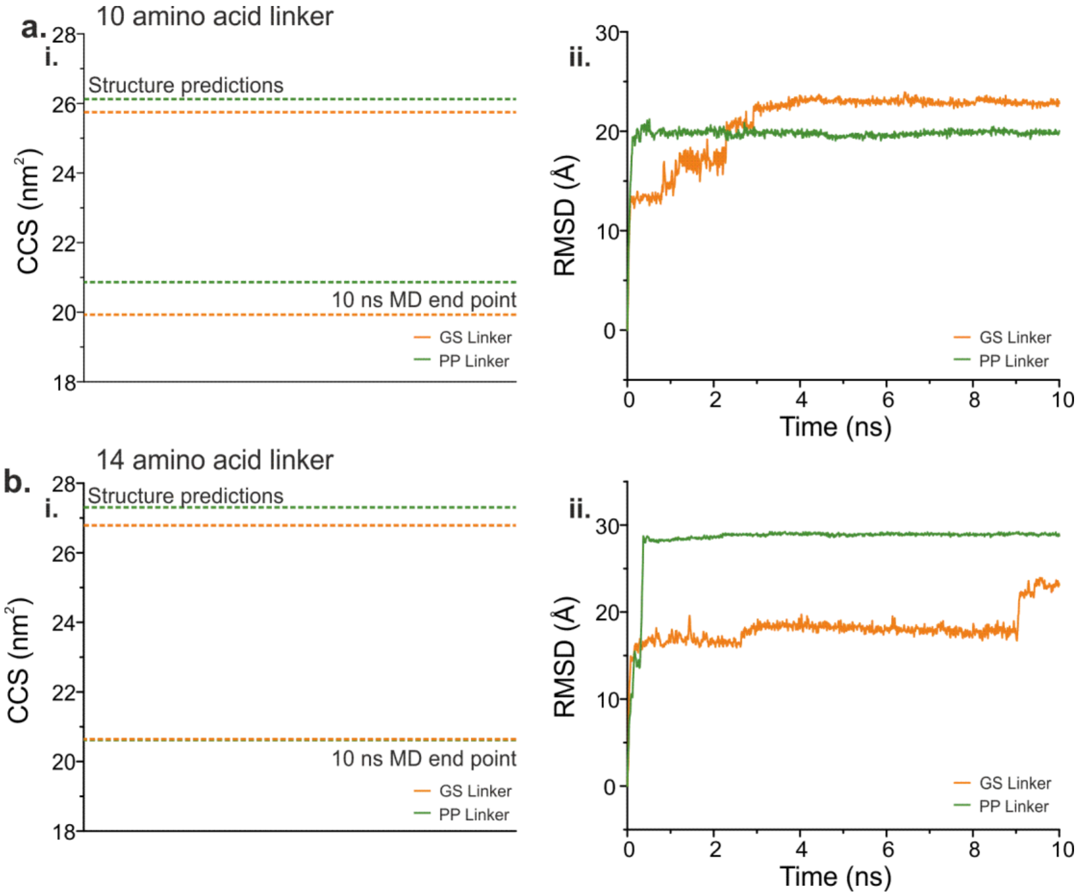


Figure 4.23: Plots showing the results from the *in vacuo* simulations of I27 dimer structures with 10 and 14 amino acid linkers. **(a)** (i) CCS plot highlighting the collapse of the two molecules with different linker composition. (ii) Plot showing the change in RMSD over the 10 ns simulation time. GS linker = orange and PP linker = green. **(b)** (i) CCS plot highlighting the collapse of the two molecules with different linker composition. (ii) Plot showing the change in RMSD over the 10 ns simulation time. GS linker = orange and PP linker = green.

As found with the two amino acid linker constructs, the initial predicted CCS values of the two structures are similar with differences of 0.38 nm^2 for the 10 amino acid linker (**Figure 4.23a**) and 0.52 nm^2 for the 14 amino acid linker (**Figure 4.23b**). Furthermore, for the structures with the 10 amino acid linkers there is an observed difference between the two end points, as observed for the initial experiments (**Figure 4.22**). However, when the linker length is increased to fourteen amino acids, similar to the length of the hinge region in IgG1s, both structures collapse to almost the same value by the end of the 10 ns simulation (20.64 nm^2 for GS and 20.62 nm^2 for PP). While initial results indicate that different composition of linkers may be able to be distinguished using IMS-MS, this needs further work and constructs made for experimental values to be acquired for direct comparison with the predicted results.

Although there are differences in the estimated CCS values for the constructs with the shorter linker lengths, whether these differences would be large enough to detect via direct IMS-MS measurements was uncertain. Therefore, instead of cloning and purifying these constructs it was decided to look for other protein structures that have flexible linker regions (such as mAbs and $(I27)_5$) to characterise, experimentally, how they behave in the gas-phase of a mass spectrometer.

4.6 Exploring the conformations of elongated structures in the gas-phase

As mentioned previously, native IMS-MS experiments have been well-used to characterise a plethora of globular proteins and protein complexes (Scarff et al., 2008, Leary et al., 2009, van Duijn et al., 2009, Hilton et al., 2010, Zhou et al., 2014). However, the technique has not been as implemented as often to characterise longer, elongated proteins. Two other proteins, with elongated structures, were chosen to study via IMS-MS: the surface protein G (SasG) and the POTRA domains of Bam A (**Figure 4.24**).

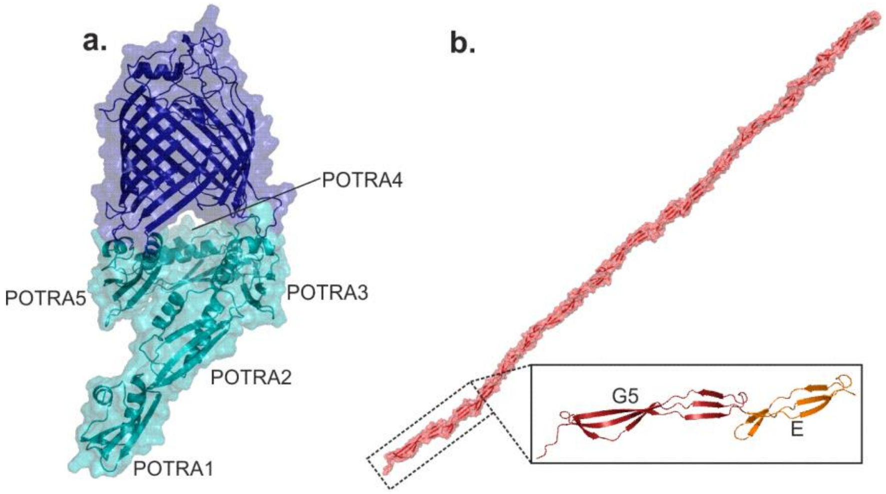


Figure 4.24: Structures of the BamA barrel with POTRA domains and SasG. **(a)** BamA from *E. coli* (PDB = 5D0O (Gu et al., 2016)) showing the five POTRA domains (cyan) from BamA (blue) and **(b)** SasG used to investigate the gas-phase collapse properties of elongated molecules. Inset shows the two repeat domains of SasG.

The POTRA domains were chosen for analysis since, similar to the $(I27)_5$, the protein consists of five POTRA subunits (POTRAS 1-5) that are connected through small linker regions (4-5 residues) between the domains (Knowles et al., 2008, Gatzeva-Topalova et al., 2010). While their function is not fully known, it is believed that the POTRA domains serve as loading machinery of client proteins into the BamA barrel and Bam complex for folding into the outer membrane (Fleming et al., 2016). SasG is responsible for the adherence and the biofilm formation of its host *Staphylococcus aureus (S. aureus)* (Corrigan et al., 2007). This protein was of particular interest as it consists of repeats of two domains (G5 and E), in which the C-terminus of the previous subunit is directly connected to the N-terminus of the subsequent G5/E subunit. Furthermore, SasG is responsible for biofilm formation the protein has been shown to form long, elongated fibrillar structures that maintain a highly extended conformation in solution (Gruszka et al., 2015). How the molecule behaves in the gas-phase was of interest since it elongated, yet does not have a ‘beads on a string’ architecture. The POTRA domains are constituents of the Bam complex that lie N-terminal to the BamA barrel (Gu et al., 2016).

The POTRA domains and SasG (kindly supplied by Robert Schiffrin, University of Leeds and Prof. Jennifer Potts, University of York, respectively) were dialysed against 150 mM ammonium acetate and characterised using nESI-MS (**Figure 4.25**).

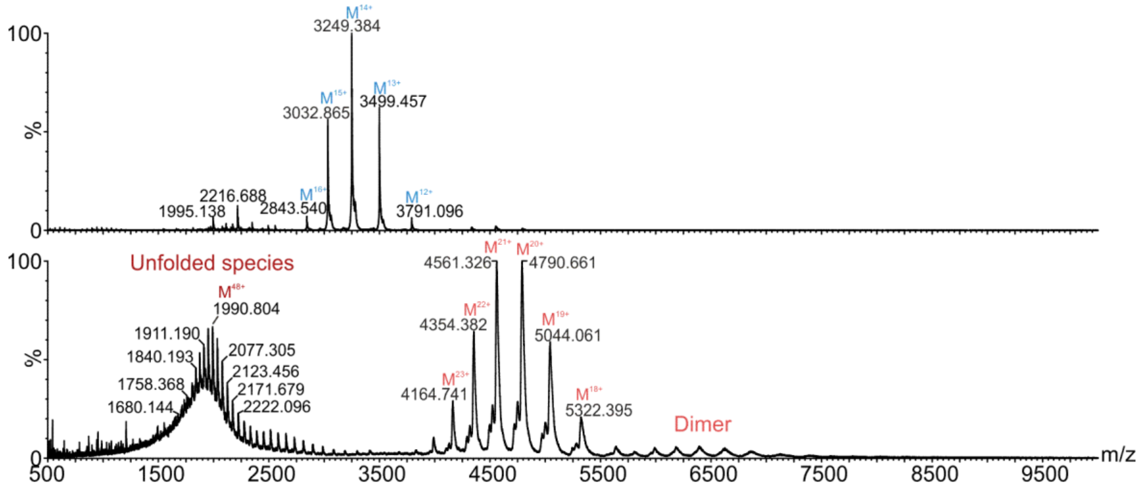


Figure 4.25: Native mass spectra of the POTRA domains (top) and SasG (bottom). A small proportion of dimeric species can be observed for SasG as well as a highly charged species at low *m/z*. Representative spectra of n =2.

The results showed that the POTRA domains and SasG predominantly exist as monomers, with measured masses of 45,480.61 ± 3.49 Da and 95,513.49 ± 8.08 Da, respectively. Interestingly, SasG exists in two monomeric conformations: a native-like charge state conformation (centred on 20+/21+) and a highly charged conformation (centred on 48+). As mentioned in the introduction to this study, this highly charged conformation is usually indicative of an unfolded state (see **Section 2.3.1**).

IMS-MS measurements were then taken for the two samples and their CCSs determined. As carried out before, these CCS values were compared directly to the predicted values from both the initial structures and the gas-phase equilibrated structures (**Figure 4.26**). It is important to note that the structure used for the SasG predictions and simulations was generated from SAXS data (Gruszka et al., 2015), which was kindly supplied by Prof. Jennifer Potts group (University of York, UK).

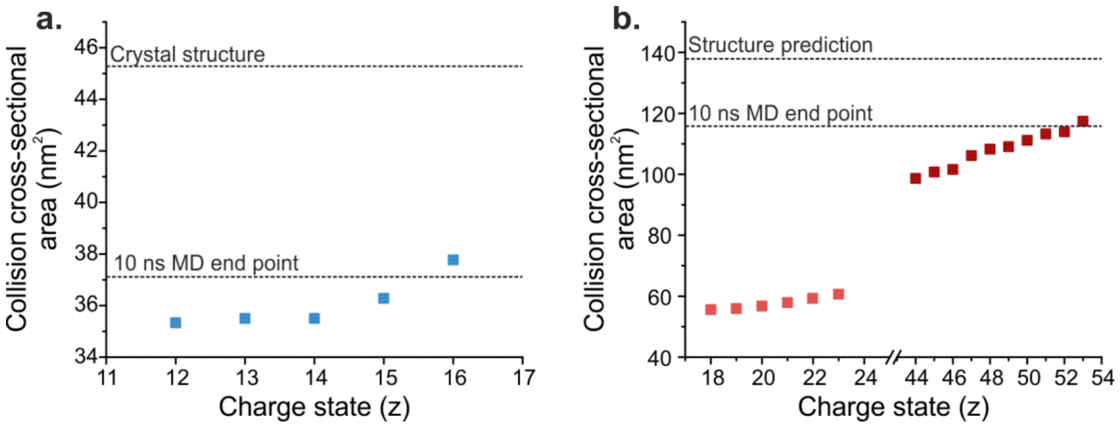


Figure 4.26: *In vacuo* collapse of the POTRA domains and SasG. **(a)** Plot of CCS against charge state for the POTRA domains. The predicted CCSs of the POTRA domains (5D0O) and the equilibrated structure, after 10 ns, are represented as dotted lines. Measured data are plotted as blue squares. **(b)** Plot of CCS against charge state for SasG. The predicted CCSs of the SasG structure and the equilibrated structure, after 10 ns, are represented as dotted lines. Measured data are plotted as red (native species) and dark red (highly charged species) squares.

Consistent with the previous structures examined, both the POTRA domains and SasG collapse in the gas-phase, with the experimental data in closer agreement with the *in vacuo* equilibrated structures. Interestingly, the native-like conformation measured by IMS-MS of SasG is significantly smaller than the predicted CCS of both the initial structure and the simulated structure (**Figure 4.26b**). Upon analysis of the SasG simulation, it appeared that the structure still had further potential for collapse since the RMSD of the simulation was not equilibrated after 10 ns (**Figure 4.27**).

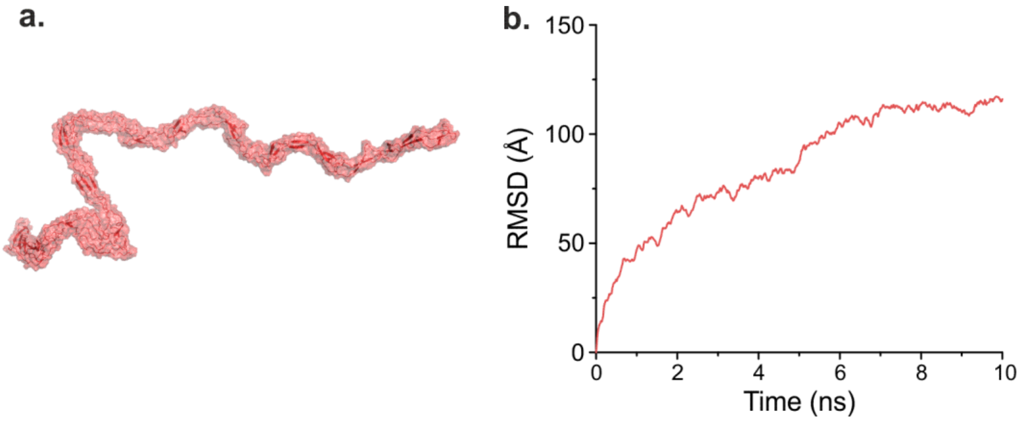


Figure 4.27: *In vacuo* collapse of SasG after 10 ns.**(a)** Visual representation of the SasG structure after a 10 ns equilibration experiment. **(b)** Plot of RMSD against simulation time indicating that the simulation has not reached equilibrium.

Unlike all of the other systems tested, the RMSD of SasG (**Figure 4.27**) indicates that the simulation had not yet reached equilibration after 10 ns as the values had not plateaued. Visualisation of the structure also made it clear to see that there was further potential space for the structure to collapse as only the N-terminus had altered significantly. Since the protein is made up of repeating subunits, there was no clear reason that only one region of the molecule should undergo collapse. To obtain further information of how the molecule might adapt in the gas-phase, the simulation of SasG was extended until the structure had reached equilibration (around 35 ns) so that the experimental results can be compared to the final equilibrated conformation (**Figure 4.28**).

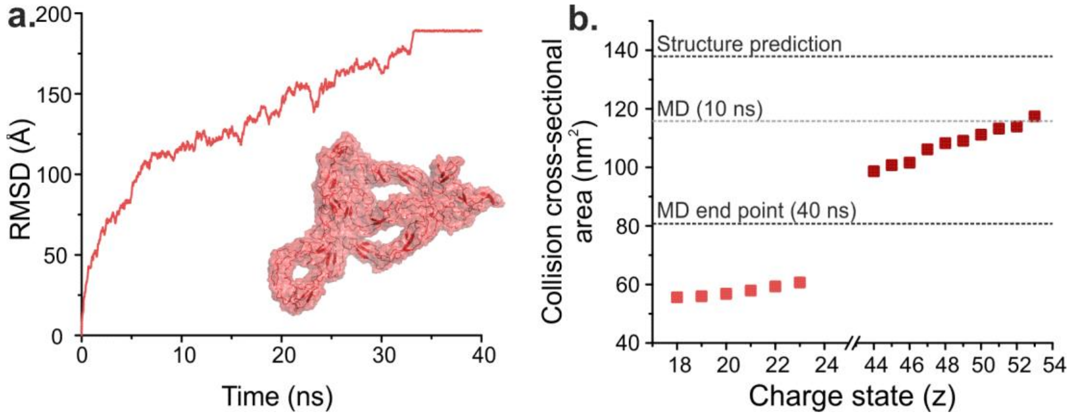


Figure 4.28: *In vacuo* collapse of SasG after 40 ns. (a) Plot of RMSD against simulation time showing that the simulation reaches equilibrium around 35 ns. A visual representation of the structure, at 40 ns, is also presented. (b) Plot of CCS against charge state for SasG. The predicted CCSs of the SasG and the equilibrated structure, after 10 and 40 ns, are represented as dotted lines. Measured data are plotted as red (native species) and dark red (highly charged species) squares.

After a 40 ns simulation, the molecule has appeared to reach equilibration and a larger compaction can be observed across the molecule (**Figure 4.28**). Unlike **Figure 4.27**, the C-terminal region of the molecule has also altered significantly whereas after 10 ns only changes significant changes around the N-terminus were observed. Intriguingly, comparison of the CCS from the 40 ns simulation with the experimental values (**Figure 4.28**) revealed that the experimental CCS is still lower than that predicted from the simulations. This was observed across multiple of the systems tested, most noticeably the IgG and SasG data, and it is still currently unknown what factors might attribute to this final collapse observed in the IMS-MS measurements.

4.6.1 The potential of gas-phase collapse across other biomolecules

Gas-phase collapse has been observed across a range of proteins in this thesis. However, gas-phase collapse is not limited to proteins. Data acquired by a previous student in the Ashcroft lab (Dr Henry Fisher), using model RNA molecules as the test system (Fisher, 2014), presented similar data.

Two candidate RNA structures were selected in this previous work, 2PCV (Jin et al., 2007) and 2DRB (Tomita et al., 2006). While both RNA molecules contain 35 nucleotides in length and have similar mass, 11, 217 Da for 2PCV and 11, 219 Da for 2DRB, the two molecules had distinct MOBCAL predicted CCS from their known crystal, using the PA method, of 1445 Å and 1146 Å f structures, respectively. These RNAs have also been analysed (by Dr. Henry Fisher) using IMS-MS in order to measure their CCSs experimentally (**Figure 4.29**).

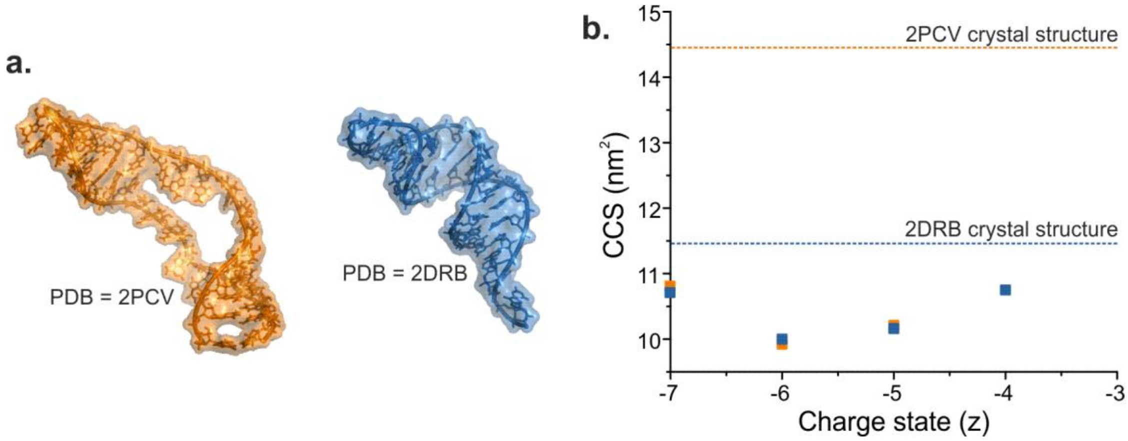


Figure 4.29: Observed collapse for the two RNA PDB structures. **(a)** Visual representation of the two RNA molecules showing that RNA 2PCV (orange) has a more extended conformation compared to RNA 2DRB (blue). **(b)** Plot of CCS against charge state showing that all measured CCSs fall below either of the expected CCS MOBCAL predictions. Measured CCS data were acquired by Dr Henry Fisher.

In contrast with the differences in their predicted CCS values, the measured CCSs of the two RNA molecules are indistinguishable from one another. Of particular interest, both RNA molecules collapse in the gas-phase IMS-MS experiments to values smaller than those predicted by their crystal structures.

This suggests that the phenomenon of gas-phase collapse is not unique to proteins alone and that it is plausible for other biomolecules with an expected elongated structure to be susceptible to this compaction within a mass spectrometer.

4.7 Discussion

Native IMS-MS as a technique has been widely used in the area of structural biology. IMS-MS has been used to characterise intermediates in virus particle assembly (Shepherd et al., 2013, Knapman et al., 2010, Uetrecht et al., 2011a), amyloid oligomeric intermediates (Bleiholder et al., 2011a, Bleiholder et al., 2013, Woods et al., 2013, Young et al., 2014b) and has been pivotal in the structural analysis of membrane proteins, which can be difficult to analyse using other structural techniques (Wang et al., 2010, Lanucara et al., 2014, Calabrese et al., 2015b, Watkinson et al., 2015). Interestingly, it is worth noting that native characterisation of WFL showed a decrease in the amount of dimer present in comparison to the results obtained from the AUC experiments (**Figure 4.2+4.5**); whilst a small amount of dimer was generated for STT under these conditions (**Figure 4.5a**). For STT, this suggests that there is a degree of non-specific interactions being formed, which may be attributed to concentration of the analyte or an artefact of the ionisation process (Robinson et al., 1996, Veenstra, 1999, Davidson et al., 2016). As mentioned previously, the observed decrease in the amount of dimer of WFL, in the native mass spectra, is consistent with the weakening of hydrophobic interactions upon transition into the gas phase (Loo, 1997, Breuker and McLafferty, 2008), which may explain the discrepancies observed between the AUC and native MS data. Therefore, the AUC data is a more likely representation of the degree of self-association in solution.

While IMS-MS can be useful to compare experimental results to theoretical values, the majority of work in the literature has focused on characterising globular proteins; whereas the structural study of elongated and linear proteins through IMS-MS is less well developed. In this Chapter, the results presented demonstrate that native IMS-MS measurements may be less applicable to such elongated structures. By firstly characterising mAb samples and moving on to other elongated systems, there appears to be a general trend where the molecules undergo collapse/compaction upon the transition to the gas-phase.

The notion of proteins undergoing collapse upon transition to the gas-phase is not a new concept. While not many, there have been reports of the gas-phase collapse of various protein structures including membrane proteins (Mehmood et al., 2014), molecular chaperones (Hogan et al., 2011), RNA binding proteins (Ruotolo et al., 2005), and of course, antibodies

(Pacholarz et al., 2014). The factors that underpin the causation of gas-phase compaction and collapse, however, are still unknown. Work in this Chapter attempts to address these factors through the characterisation of different proteins with elongated structures.

Pacholarz et al. showed that full mAbs undergo significant compaction in the gas-phase and suggested that the flexible nature of the hinge region may be responsible. By characterising the digested regions of an IgG (Fab, Fc and F(ab')2), significant collapse was observed predominantly in regions that contain or partially retain the hinge region. While it is clear that regions that have this flexible hinge region undergo the largest extent of collapse, it is unclear what drives this linker to compact. While this result occurs in the gas-phase, it is important to note that mAbs have been shown to adopt an ensemble of structures, both extended and collapsed, in solution. SAXS data have highlighted the flexibility in the hinge region of an IgG2 molecule, presenting an ensemble of structures that populate an Rg as small as 15.94 Å (Clark et al., 2013). While the data presented on IgGs in this Chapter indicate a compaction around the hinge region with the Fab arms and the Fc still existing on the same plane, SAXS data suggest that it is possible for an IgG2 to fold over its hinge region. This could explain the discrepancy that is observed in the IgG collapse data, both presented here (see **Figure 4.13**) and in the literature; although both agree that a collapsed is observed as evidenced by the *in vacuo* simulations, the experimental CCS values still fall below the predicted values. If the molecules were to fold around their hinge region, as well as compacting, one would expect a significantly smaller CCS that may be in closer agreement with the experimental values, which could also be the case for the F(ab')2 modelling (see **Figure 4.17**). Therefore, although it is still unknown what causes the significant collapse of mAbs in the gas phase, it can be concluded that the flexible hinge region is playing a critical role.

As the hinge region proves critical, it was decided that further clarity surrounding this collapse could be achieved by investigating molecules with similar flexible linker regions between domains. $(I27)_5$ was chosen as the molecule retains the Ig fold while orientating the subunits in a different global architecture. As observed with the IgG data, $(I27)_5$ undergoes compaction in the gas-phase which also involves the hinge regions of the protein (see **Figure 4.20**). While the study into linker length and composition remained inconclusive, the focus of altering linker composition for either increased distance or flexibility/rigidity is an approach that is of interest (Klein et al., 2014). Therefore, a focus on linker length and composition in the gas-phase is an area that needs to be explored further.

The correlation between proteins containing flexible linker regions and gas-phase collapse was striking and thus was further probed. IMS-MS has been used, in the vast majority of studies, to characterise globular proteins; while elongated and extended proteins been left on the side-lines. To understand how elongated structures behave in the gas phase, two other proteins were adopted for study using IMS-MS: SasG and the POTRA domains of BamA. Alike the other systems analysed in this study, both SasG and the POTRA domains collapse to significantly compact structures in the gas-phase (**Figure 4.26-4.29**). It was hypothesized that the POTRA domains would undergo gas-phase collapse due to the small linkers connecting the individual domains. However, SasG was an interesting discovery as the protein lacks any obvious linker regions; with the C-terminus of one subunit immediately connected to the N-terminus of the sequential subunit (**Figure 4.24**). Although being stable and extended in solution (Gruszka, 2012, Gruszka et al., 2015), the IMS-MS data show that the protein adopts a surprisingly compact structure in the gas-phase (~60 % smaller CCS than predicted). While an extended conformation is observed in solution, this result suggests that elongated structures struggle to retain their solution conformations in the transition to the gas-phase.

Finally, the study was extended to the analysis of other biomolecules. The latter results in this Chapter highlight that RNA undergoes similar gas-phase compaction (see **Figure 4.29**), as observed for various proteins. This was presented through the IMS-MS analysis of the two RNA molecules with PDB structures, 2PCV and 2DRB. The result showing that these two molecules both collapse to a state with a smaller CCS than predicted reveals that it is not only proteins that undergo collapse upon transition to the gas-phase; that this phenomenon is observed across various biomolecules.

In summary, the data presented here in Chapter 4 demonstrate the potential for biomolecules to collapse upon the transition from solution to the gas-phase. While the correlation between expected CCS of a protein and the measured value are in excellent agreement for many globular proteins, as evidenced throughout the literature (Clemmer et al., 1995, Jarrold, 1999, Benesch and Ruotolo, 2011, Illes-toth et al., 2015, Loo et al., 2005, Scarff et al., 2008, Wyttenbach and Bowers, 2011) as well as small molecules (Campuzano et al., 2011),this does not necessarily hold true for elongated structures. Chapter 4 demonstrates a potential and critical caveat of IMS-MS, which highlights for caution to be taken when measuring the CCS values for proteins or biomolecules with no known 3D structure.

Chapter 5:
Defining the site of self-association using chemical cross-linking

5 Defining the site of self-association using chemical cross-linking

5.1 Objectives

Chapter 4 provided evidence of the oligomeric species of WFL present in solution. Chapter 5 now sets out to characterise these oligomers and identify the region responsible for the self-association. Chemical cross-linking was employed to capture the oligomeric species and their conformations which were then interrogated using liquid chromatography coupled with tandem mass spectrometry.

5.2 Introduction

5.2.1 Chemical cross-linking

As described in **Section 2.2**, the structural characterisation of proteins and protein complexes has been dominated by techniques such as X-ray crystallography and NMR with cryo-EM now being raised to the same pedestal due to recent technical advancements (Bai et al., 2015, Nogales, 2016). However there are many systems which are not amenable to these techniques due to either protein size or a high degree of flexibility/dynamics in a system (Leitner, 2016). Mass spectrometry has been instrumental in bridging this structural gap (Rappsilber, 2011). There has been a surge of publications in recent years utilising mass spectrometry, coupled with structural proteomic techniques such as hydrogen-deuterium exchange (HDX), photo-oxidative labelling of proteins (FPOP) (Baud et al., 2016, Heinkel and Gsponer, 2016) and chemical cross-linking (XL) to achieve high resolution structural information of protein complexes (Navare et al., 2015, Walzthoeni et al., 2015, Liu et al., 2015, Nguyen-Huynh et al., 2015, Belsom et al., 2016, Politis et al., 2015, Packiam et al., 2015, Chen et al., 2016c, Chen et al., 2016b). One excellent example of this was the solution of a 3 MDa complex, the RNA polymerase II transcription pre-initiation complex (Murakami et al., 2013), using XL-MS in combination with cryo-EM.

XL-MS is an ideal technique for characterising aggregation-prone systems as it is able to "capture" any oligomeric species present in solution, insensitive of conformation, which can be analysed downstream. Furthermore, due to the covalent nature of the cross-links, unlike techniques such as HDX which involve careful sample handling after labelling, cross-linked samples can be taken for further reactions before analysis. With sophisticated chemistry, it is possible to target any residue on a protein's surface via photo activatable cross-linkers, such as diazirines, or target specific residues using N-hydroxysuccinimide (NHS) ester chemistry (amine group on lysine residues) or conjugation with carbodiimides (carboxyl groups) (Sinz et al., 2015, Liu and Heck, 2015, Chen et al., 2016a, Arlt et al., 2016, Tran et al., 2016, Faini et al., 2016).

Here we utilise a routinely used amine-specific cross-linker, BS3 (*bis*(sulfosuccinimidyl)suberate) (**Figure 5.1**) (d'Souza et al., 1988, Andrews et al., 1989, Palecanda et al., 1992, Nürnberger et al., 1995, Huang et al., 2004, Fischer et al., 2013), in order to cross-link solvent exposed lysine residues presented on the surface of WFL oligomers.

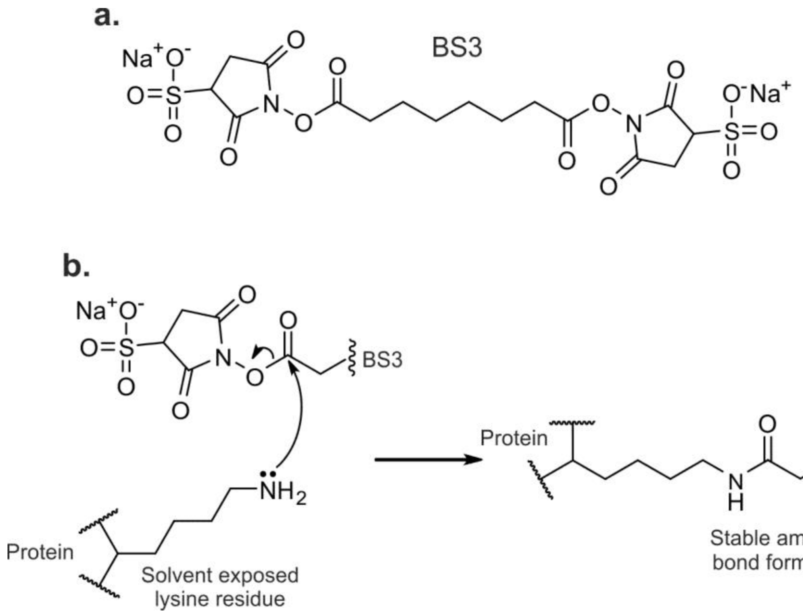


Figure 5.1: Structure of the chemical cross-linker used in this study. **(a)** Structure of BS3 and **(b)** its mechanism of action. The amine group of lysine residues attacks the ester group of BS3 and an amide bond is formed between the lysine residue and the cross-linker.

Unlike HDX experiments, which typically use pepsin to digest labelled proteins due to the low temperature and low pH requirements of the technique, any protease can be used to generate peptide fragments from a cross-linked sample. However, trypsin is most commonly used due to its valuable sequence coverage, specificity and its compatibility with mass spectrometry analyses (Olsen et al., 2004).

5.3 Analysing chemical cross-linking data

To analyse the results of the chemical cross-linking using MS, after digestion, it is imperative to purify the cross-linked peptides in order minimise the background noise arising from non-cross-linked peptides, the latter of which are in the majority. There are currently two methods that are used to purify cross-linked material: peptide size exclusion chromatography (SEC) (Leitner et al., 2014) and in-gel digestion (Shevchenko et al., 1996). The peptide SEC method was established by Prof. Reudi Aebersold's group at ETH Zurich (Switzerland). When carrying out SEC on a protein digest, a range of overlapping peaks are observed in the elution profile (**Figure 5.2a**). However, due to cross-linked peptides being heavier in mass on average, compared with their non-cross-linked counterparts, they elute earlier in the elution from a peptide SEC column (**Figure 5.2b**); thereby enriching the amount of cross-linked material in the earlier eluting fractions (Leitner et al., 2014) .

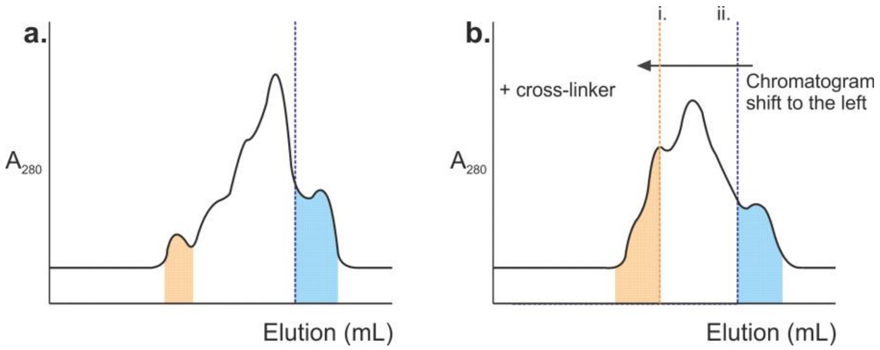


Figure 5.2: Example elution profiles for non-cross-linked and cross-linked samples. Example elution profile for a **(a)** non-cross-linked peptic digest and **(b)** cross-linked peptic digest. A shift in elution can be observed (blue (ii) to orange (i)) for cross-linked peptides.

Chemical cross-linking followed by in-gel digestion is routinely used to identify cross-linked fragments in the field (Morgner et al., 2015). This method has the advantage of isolating individual species from an SDS-PAGE gel where the user knows that inter-molecular cross-links are present in the sample due to the corresponding molecular weight on the gel (**Figure 5.3**).

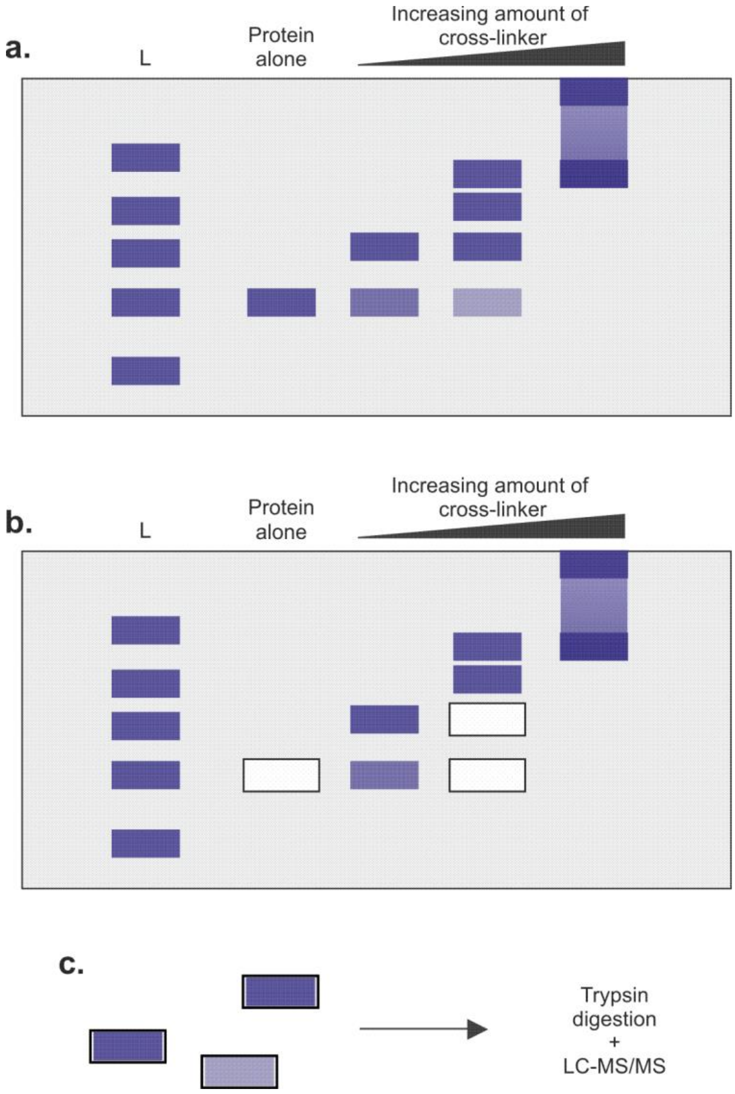


Figure 5.3: Schematic of the in-gel digestion workflow. **(a)** Protein samples with and without varying amounts of cross-links are analysed via SDS-PAGE. **(b)** Bands of interest are carefully excised from the gel and **(c)** taken forward for in-gel trypsin digestion, followed by LC-MS/MS. L = protein marker.

5.4 WFL oligomers can be captured using chemical cross-linking

5.4.1 Chemical cross-linking with glutaraldehyde

To probe the accessibility of WFL and STT to chemical cross-linking, solvent exposed lysines were first interrogated by reacting the protein with glutaraldehyde, targeting amine groups (Migneault et al., 2004). While glutaraldehyde polymerises in solution to form various linker lengths (Aso and Aito, 1962a, Aso and Aito, 1962b), this experiment established whether it was possible to capture the oligomeric species of WFL while specifically targeting lysine residues.

Briefly, WFL was cross-linked at 1, 2 and 4 mg/mL in 100 mM sodium phosphate, pH 7.4 with varying amounts of glutaraldehyde (0.1 – 0.005 % (*w/v*), 25 % (*w/v*) stock solution) (see **Section 3.4.4**). Cross-linking reactions were incubated on ice for 20 minutes before quenching with Tris. HCl, pH 8, at a final concentration of 50 mM. The resulting cross-linked species were then analysed via SDS-PAGE (see **Section 3.4.8**) (**Figure 5.4**).

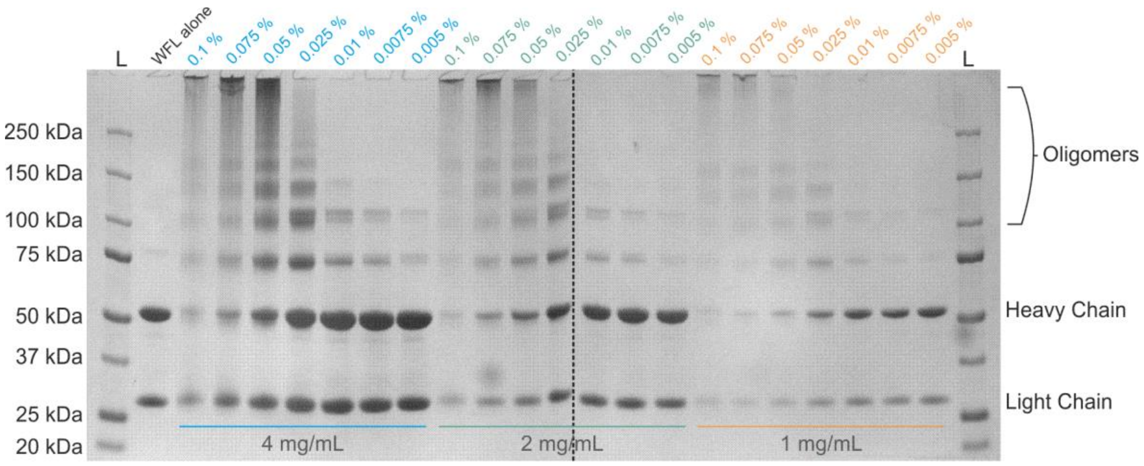


Figure 5.4: SDS-PAGE gel of WFL samples cross-linked with glutaraldehyde. WFL was cross-linked at protein concentrations of 4 (blue), 2 (green) and 1 (orange) mg/mL with decreasing amounts of glutaraldehyde, from 0.1 – 0.005 (*w/v*) %. L = protein marker. Stained with Instant Blue Coomassie stain.

Cross-linking with glutaraldehyde resulted in a significant proportion of WFL precipitating out of solution, from 0.1 – 0.05 % (*w/v*) glutaraldehyde. However, the laddering effect observed reveals that it was possible to cross-link WFL oligomers by exclusively targeting solvent exposed lysine residues.

While effectively cross-linking the IgG sample, glutaraldehyde is not an ideal chemical cross-linker for the follow-on sequencing in the mass spectrometer due to the heterogeneous lengths that glutaraldehyde forms. Therefore, as it had been concluded that targeting the lysine residue was a viable method, a cross-linker with equivalent specificity but homogeneous linker length was chosen for further cross-linking studies.

5.4.2 Chemical cross-linking with BS3

To ensure specificity in the cross-linking experiments and to allow for downstream mass spectrometry analysis, the amine specific cross-linker BS3 was chosen as it had an appropriate cross-linker length of 11.4 Å and has been historically and widely, but not exclusively, used in the field (Back et al., 2003). There was a concern that any longer length cross-linkers would increase the probability of forming uninformative intra-molecular cross-links due to the increased length. Furthermore, BS3 is commercially available in deuterated (d4) and non-deuterated (d0) forms, making downstream analysis easier for the user (Seebacher et al., 2006, Fischer et al., 2013). By using a 50:50 mix of deuterated and non-deuterated cross-linker, any site that is modified by BS3 has a 50 % probabilistic chance to be modified by either the deuterated or non-deuterated form (Schmidt and Robinson, 2014). This is reflected in the resulting mass spectra by the appearance of a fingerprint doublet peak that differs 4 Da in mass (**Figure 5.5**).

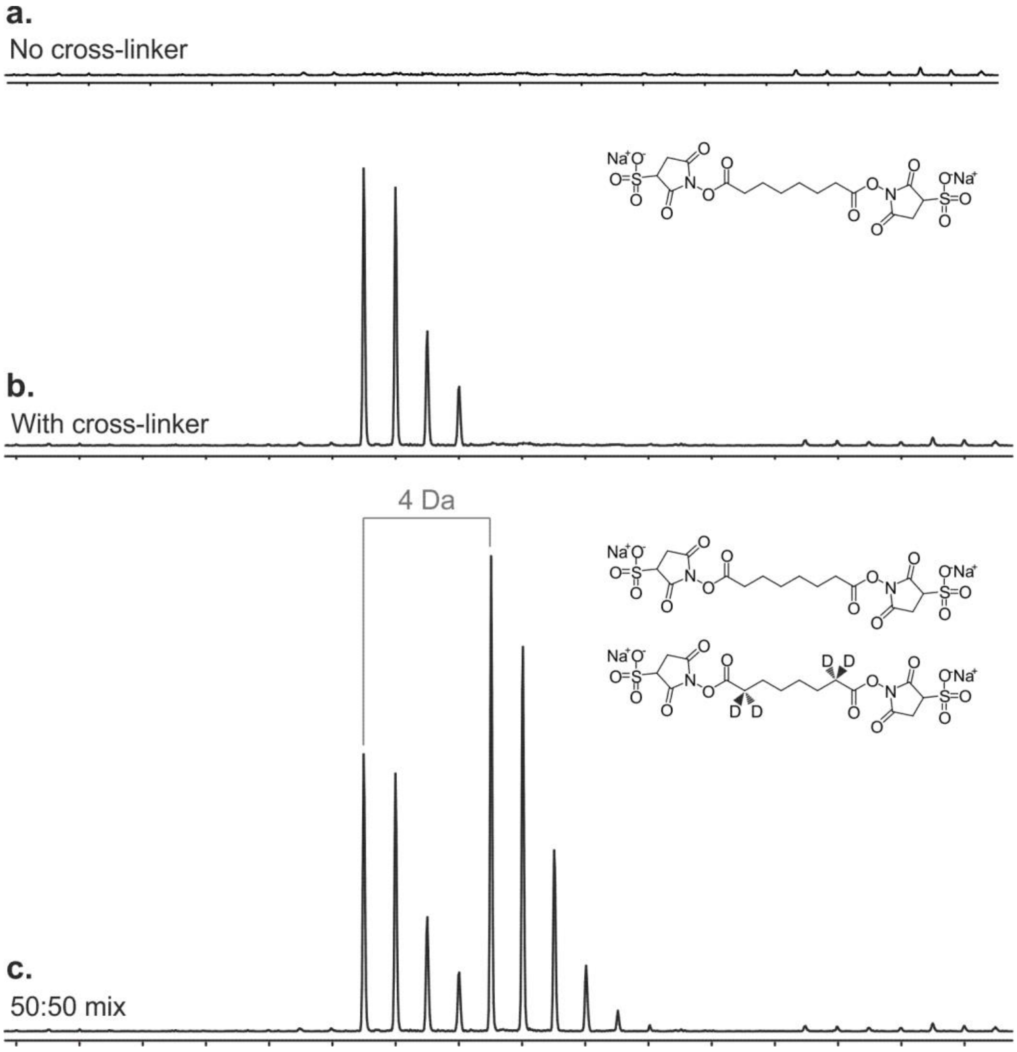


Figure 5.5: Schematic showing the MS-identification of cross-linked peptides. **(a)** No peaks are observed in this region of the mass spectrum, in the absence of cross-linker while **(b)** new peaks appear showing typical isotope distribution for a tryptic peptide upon addition of BS3. **(c)** A signature doublet set of peak is formed when cross-linking with a 50:50 mix of deuterated and non-deuterated BS3 cross-linker. The 4 Da difference indicating the d0/d4 mass difference in BS3 for the tryptic peptide is shown.

The mixture of deuterated and non-deuterated cross-links allows the unequivocal detection of cross-linked peptides when analysed using MS. Therefore, a comparative approach can be taken by searching for cross-links that are unique to the oligomeric species present in the sample.

Samples of WFL and STT were cross-linked with increasing (molar excess) amounts of deuterated and non-deuterated BS3 (d0/d4) (see **Section 3.4.5**). The IgGs were cross-linked at a constant antibody concentration of 1 mg/mL since WFL has been shown to be predominantly dimeric at this concentration, whilst STT is predominantly monomeric (see **Figure 4.1**). The reactions were performed on ice for 30 min before quenching with Tris.HCl, pH 8, at a final concentration of 50 mM. The resulting cross-linked mixtures were then analysed via SDS-PAGE (see **Section 3.4.8**) (**Figure 5.6**).

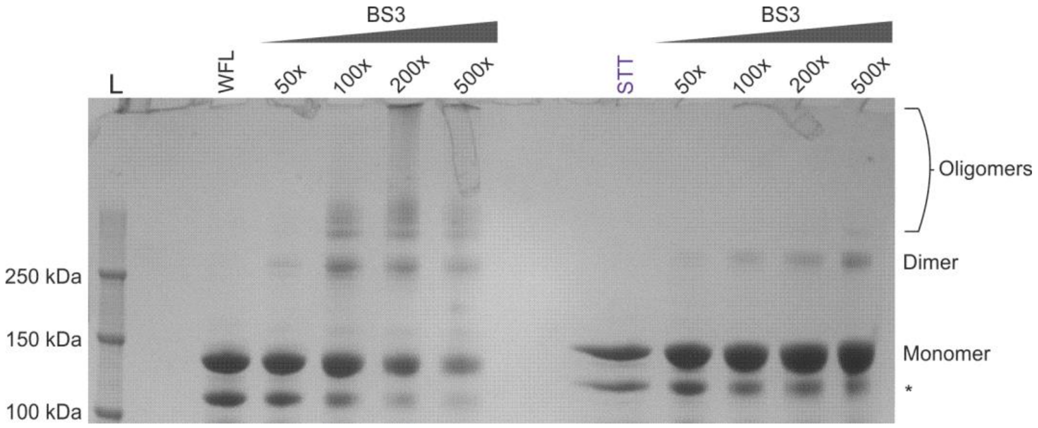


Figure 5.6: Non-reducing SDS-PAGE gel of WFL and STT samples cross-linked with BS3. Protein samples were cross-linked at 1 mg/mL with increasing amounts of BS3, from 50× to 500× molar excess. L = protein marker. * = corresponds to monomer with the loss of a single light chain. Representative gel of n = 5. Stained with Instant Blue Coomassie stain.

Similar to the cross-linking experiments with glutaraldehyde (**Figure 5.4**), WFL forms a range of oligomeric species in the presence of cross-linker while STT only forms a small amount of dimeric species. At cross-linker concentrations higher than 100× molar excess, WFL begins to precipitate out of solution and such high cross-linker concentrations were subsequently avoided. Importantly, the difference in cross-linking efficiencies observed reflect the behaviour of WFL and STT in solution (see **Figure 4.1**), with the majority of STT being monomeric.

As stated above, the cross-linked peptides required purification and enhancement to minimise the background noise arising from non-cross-linked peptides before sequencing in the mass spectrometer. Both the peptide SEC method and in-gel digestion method were applied to analyse which approach yields a larger number of cross-linked peptides, at greater purity.

5.4.2.1 Purifying cross-linked peptides via peptide SEC

For the peptide SEC, samples of WFL and STT were cross-linked with 100× molar excess of d0/d4 BS3 as described above and in **Section 3.4.5**. After cross-linking and quenching, mAb samples were denatured in 6 M guanidine hydrochloride (Gdn.HCl) for 30 min. Cross-linked samples were then treated with 10 mM DTT for 30 min at 37 °C and subsequently treated with 25 mM iodoacetamide for 20 min at 25 °C in the dark. 100 mM sodium phosphate, pH 7.4 was added to the reduced and alkylated samples to dilute the GdnHCl below 1.5 M. Trypsin was then added to the samples at a ratio of 1:50 (w/w) (enzyme:protein) and incubated at 37 °C for 18 h. The resulting cross-linked and non-cross-linked peptide mixtures were then purified using C18 SepPak cartridges (see **Section 3.4.6**) before being evaporated to completion in a SpeedVac system. Peptides were then re-suspended in 40 µL H_2O before being loaded onto a Superdex 3.2/30 peptide column (see **Section 3.4.7**) equilibrated in 30 % (v/v) acetonitrile and 0.1 % (v/v) trifluoroacetic acid (TFA), as described in (Leitner et al., 2014) and **Section 3.4.7**. Cross-linked and non-cross-linked samples were then analysed via peptide SEC for comparison (**Figure 5.7**).

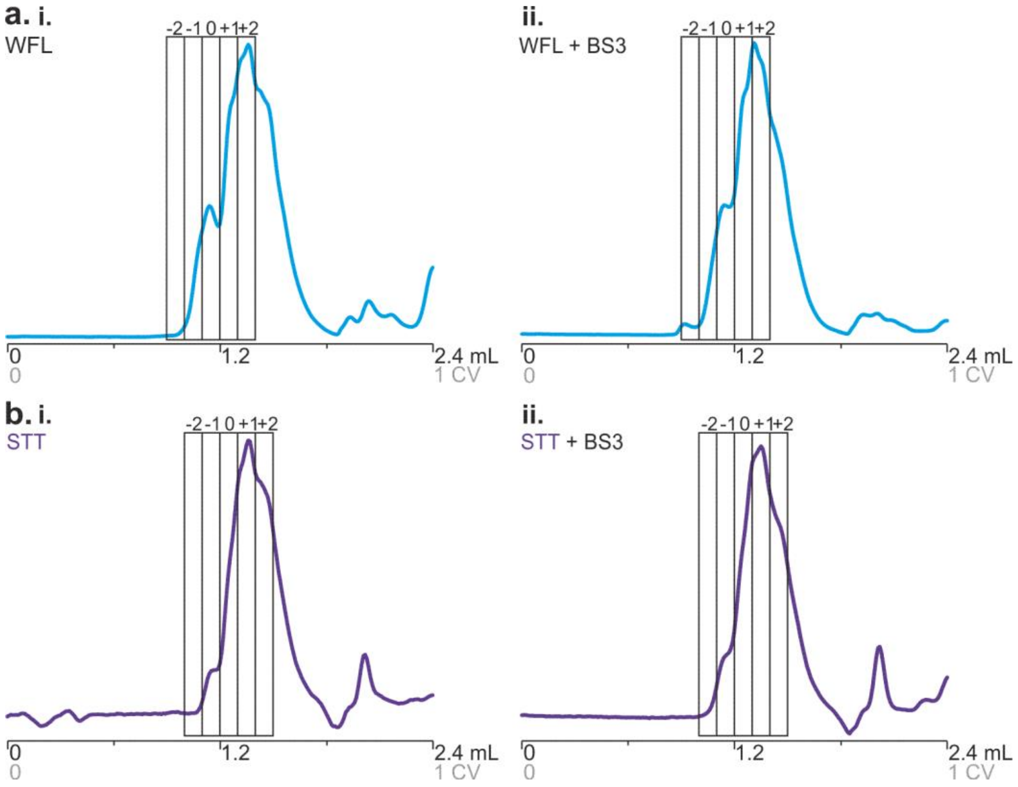


Figure 5.7: Peptide SEC elution profiles for non-cross-linked and cross-linked samples of WFL and STT. **(a)** Peptide SEC elution chromatograms of (i) non-cross-linked and (ii) cross-linked WFL. **(b)** Peptide SEC elution chromatograms of (i) non-cross-linked and (ii) cross-linked STT. Fractions from 1 mL (-2) to 1.5 mL (+2) are highlighted. Representative data of n = 3.

Peptide SEC reveals a shift in the chromatograms of the cross-linked samples (**Figure 5.7a.ii. and b.ii.**), indicative of cross-linked peptides being present. The fractions were collected and concentrated by evaporation in a SpeedVac system before being taken forward for MS/MS sequencing (see **Section 3.4.7**). The fractions were reconstituted in 0.1 % (v/v) formic acid then separated and infused into the G2Si mass spectrometer using an ACQUITY LC system (see **Section 3.2.6**).

Unfortunately, no intermolecular cross-linked peptides were identified in the samples sequenced when analysed with the StavroX software (Götze et al., 2012, Götze et al., 2015). It was a concern that while cross-linked peptides elute in earlier fractions, two small peptides connected through a cross-link may elute in later fractions; amongst the non-cross-linked material. Furthermore, as the sample contained monomer to tetrameric species the concentration of each cross-linked species would potentially be low in this complex mixture, which will include both intra- and inter-molecular cross-links. Finally, due to the amount of remaining monomeric species present in the cross-linked samples (see **Figure 5.6**) it was hypothesised that the inter-molecular cross-linked species would be low in abundance and ultimately, difficult to visualise. Therefore, it was decided to use the in-gel digestion method as this may minimise the background noise by only analysing peptides forming from the oligomeric species of interest.

5.4.2.2 Purifying cross-linked peptides via SDS-PAGE and in-gel digestion

The advantage of the in-gel digestion method was that as bands were excised from a non-reducing SDS-PAGE gel, the band corresponding to the dimer of WFL can be preferentially selected for digestion and subsequent sequencing. By only excising the dimer band this aids in two ways: by (1) minimising the background noise from intra-molecular cross-links from all other species that were in the sample and (2) as this species presents itself as a dimer, it was confidently concluded that inter-molecular cross-links had to have been formed.

Samples of WFL and STT were cross-linked with a 100× molar excess of d0/d4 BS3 as described in **Section 3.4.5** and separated on a non-reducing SDS-PAGE gel, with a non-cross-linked sample in tandem. Bands corresponding to the non-cross-linked monomer, cross-linked monomer and cross-linked dimer of both WFL and STT were excised from the SDS-PAGE gel, cut into small cubes (around 1 mm^3) and subjected to in-gel digestion with trypsin (see **Section 3.4.9**). Briefly, the gel pieces were dehydrated with 50 % (v/v) acetonitrile in 25 mM ammonium bicarbonate, pH 7.8 to remove the gel stain. The IgGs were then reduced by incubating the gel pieces with 10 mM DTT for 45 min at 37 °C. The DTT solution was removed and the gel pieces were subsequently incubated with 50 mM iodoacetamide to alkylate the free cysteine residues. A final dehydration step was carried out with 100 % acetonitrile before the solution removed and the gel pieces dried to completion in a SpeedVac system. The gel pieces were then re-solvated in a 0.1 μg/mL trypsin solution and incubated at 25 °C for 15 min to allow the trypsin solution to absorb into the gel pieces. Excess trypsin was removed from the samples and replaced with 50 µL (enough to cover the gel pieces) 25 mM ammonium bicarbonate, pH 7.8. The trypsin digest reactions were then incubated at 37 °C for 18 h. The tryptic peptides were recovered by 4 repeat rounds of de-hydrating the gel-pieces with 50 µL of 60 % (v/v) acetonitrile, 5 % (v/v) formic acid. The recovered peptides were concentrated by evaporating the solutions to completion, before re-suspending the peptides in 0.1 % (v/v) formic acid and loading them onto the ACQUITY LC system for peptide separation and infusion into the G2Si mass spectrometer.

Upon data analysis with the both the StavroX and MassMatrix (Xu and Freitas, 2007, Xu et al., 2010) software, while many intra-molecular cross-links were identified in both the cross-linked monomer samples of WFL and STT and the cross-linked dimer sample of WFL, only one, tenuous, inter-molecular cross-link was identified automatically (see **Appendix 8.37-8.39**). The only cross-link that was identified, unique to the dimeric species, was a cross-link at the bottom of the mAb in the CH3 domains (LTVDKSR-LTVDKSR). However, this was only identified in one repeat and not others, potentially suggesting that this may be a non-specific interaction. This raised a concern about using the automated data analysis software as there had to be inter-molecular cross-links present in the samples as evidenced by SDS-PAGE. It was subsequently hypothesised that due to the manner of how cross-linking programs function, by looking for corresponding peptides at the intact peptide level first, was where problems were being encountered. If the cross-linked species were low in abundance, which was highly likely, then the peptides may have not been being selected for sequencing and therefore the programs unable to identify them.

Therefore, to identify any inter-molecular cross-linked peptides the mass spectra were analysed manually.

5.5 Sequencing of WFL cross-links indicates Fab-Fab interactions

The mass spectra from the LC runs were analysed manually and in a comparative manner; looking for cross-linked peptides (doublet peaks in the spectra) that were unique to the WFL dimer species and only present in that sample (**Figure 5.8**).

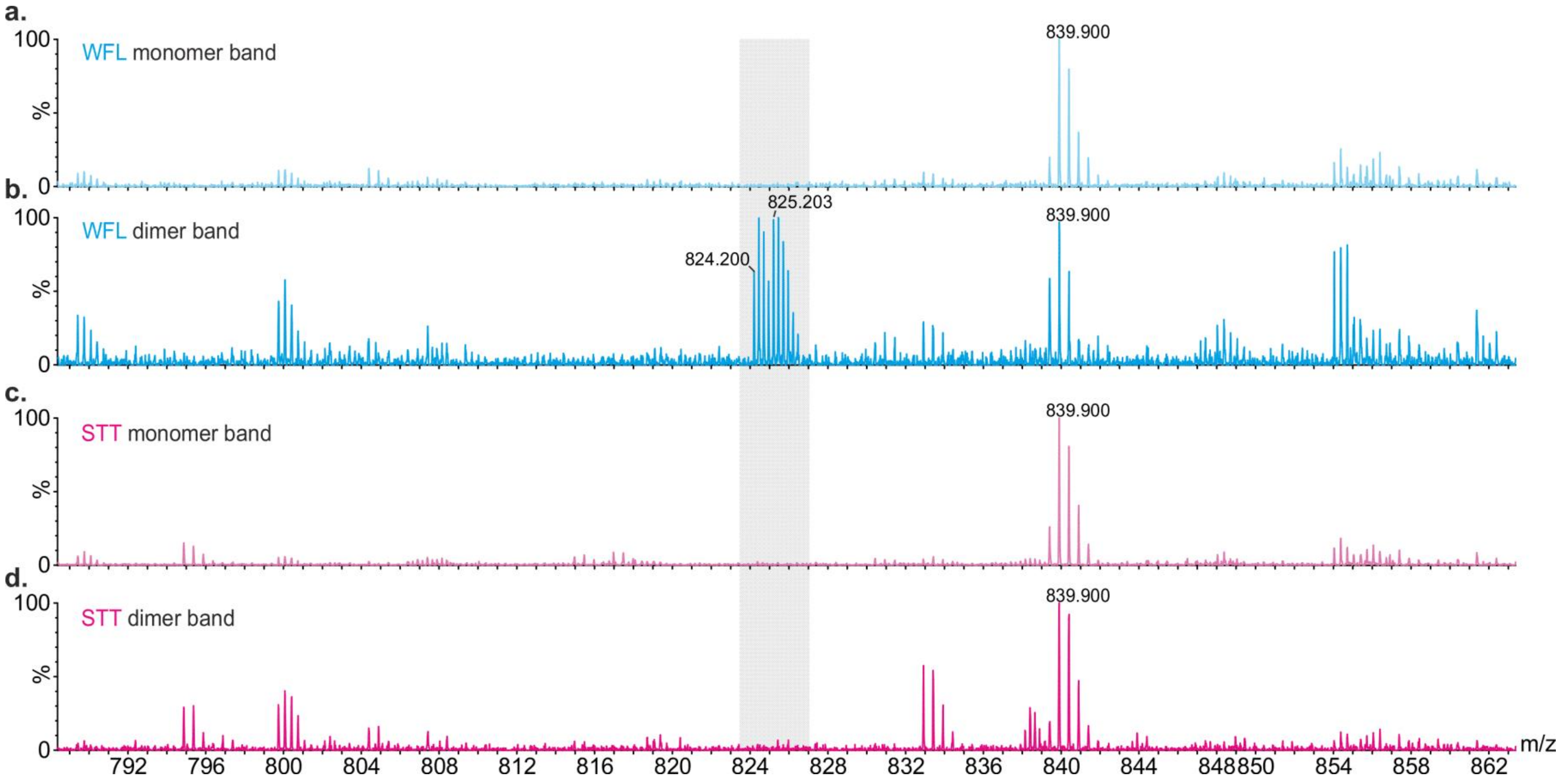


Figure 5.8: Mass spectra showing the identification of cross-link peptides unique to WFL. Mass spectra of the WFL **(a)** monomer band and **(b)** dimer band at the same retention time showing a unique cross-linked peptide. Mass spectra of the STT **(c)** monomer band and **(d)** dimer band at the same retention time showing no cross-linked peptide present in either spectrum. Shaded area = cross-linked peptide. Representative spectra of n = 3.

Manual analysis of the different spectra revealed a cross-link unique to the WFL dimer sample (**Figure 5.8b, *m/z 824-830***). As previously mentioned, this peptide was, presumably, not identified by the cross-linking programs as this peptide was never selected for sequencing due to its low ion intensity, potentially arising from low abundance in solution (see **Appendix 8.40**).

The mass spectrometer was tuned manually so that the quadrupole would only allow transmission of ions of 824.2 *m/z*. The cross-linked dimer sample of WFL was once again infused into the mass spectrometer in order for the identified cross-linked peptide to be sequenced by tandem mass spectrometry (**Figure 5.9**).

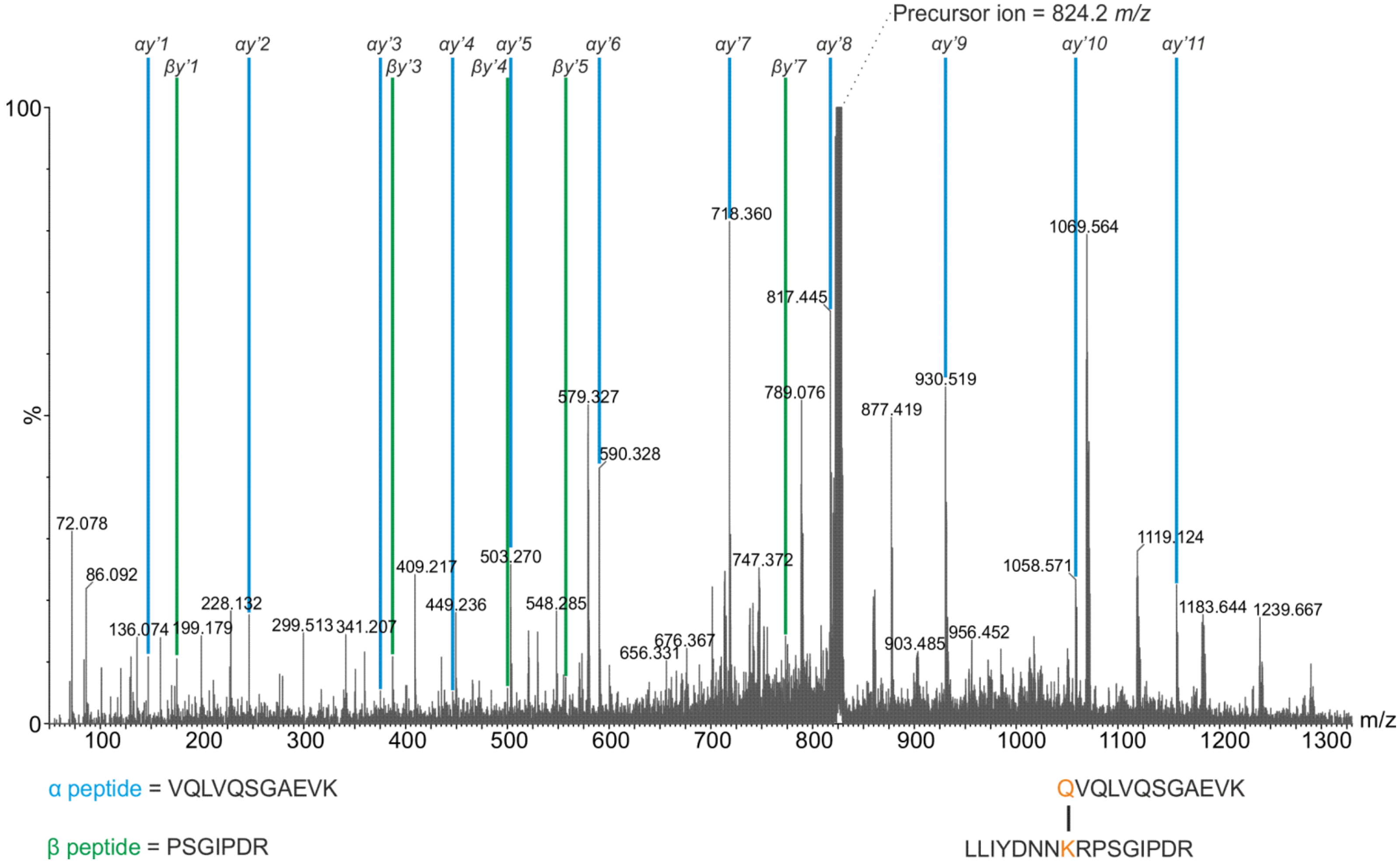


Figure 5.9: CID induced MS/MS spectrum of the unique cross-linked peptide identified in WFL. The y-ion series for both peptides (α, blue and β, green) are shown with the corresponding peptide sequence listed below. The two full length cross-linked peptides are highlighted (bottom right) with the cross-linked residues in orange. Precursor ion = 824.20 *m/z*.

Sequencing of the cross-linked peptide at 824.2 *m/z* revealed a cross-link between the N-terminus of the heavy chain to a lysine residue (K54) in the variable region of the light chain. While the collision voltages could be optimised further as the base peak in the spectrum remains as the precursor ion (824.2 *m/z*), the cross-linked peptide was fragmented sufficiently so that two partial peptides could be identified from their y-ion series. The two identified sequences were searched for in the full IgG sequence and were found to be unique peptide sequences, where the first was the N-terminal peptide from the heavy chain (QVQLVQSGAEVK) and the second was the peptide from the variable region of the light chain (LLIYDNNKRPSGIPDR).

Furthermore, due to the collision energies used for the sequencing, a series of doubly charged peptides could be identified at higher *m/z* in the spectrum (~1070 *m/z*). It was clear from the mass spectrum that these doubly charged species still correspond to two cross-linked peptides, as evidenced by the doublet peak from using the d0/d4 BS3 cross-linker (**Figure 5.9**). It was hypothesised that the doubly charged species would contain fragments of the two peptides that were being cross-linked, and would serve as further evidence for the cross-link existing across the two peptides shown above. As it appears sequencing of cross-linked peptides preferentially fragments one peptide before the other, with preferential fragmentation of specific sequences known (Qin and Chait, 1995), it was assumed that the first doubly charged peak at 1069 *m/z* would correspond to one intact peptide cross-linked to a fragment of the other. Therefore, the observed masses of the doubly charged species were compared with the expected masses of different cross-linked combinations (**Figure 5.10**).

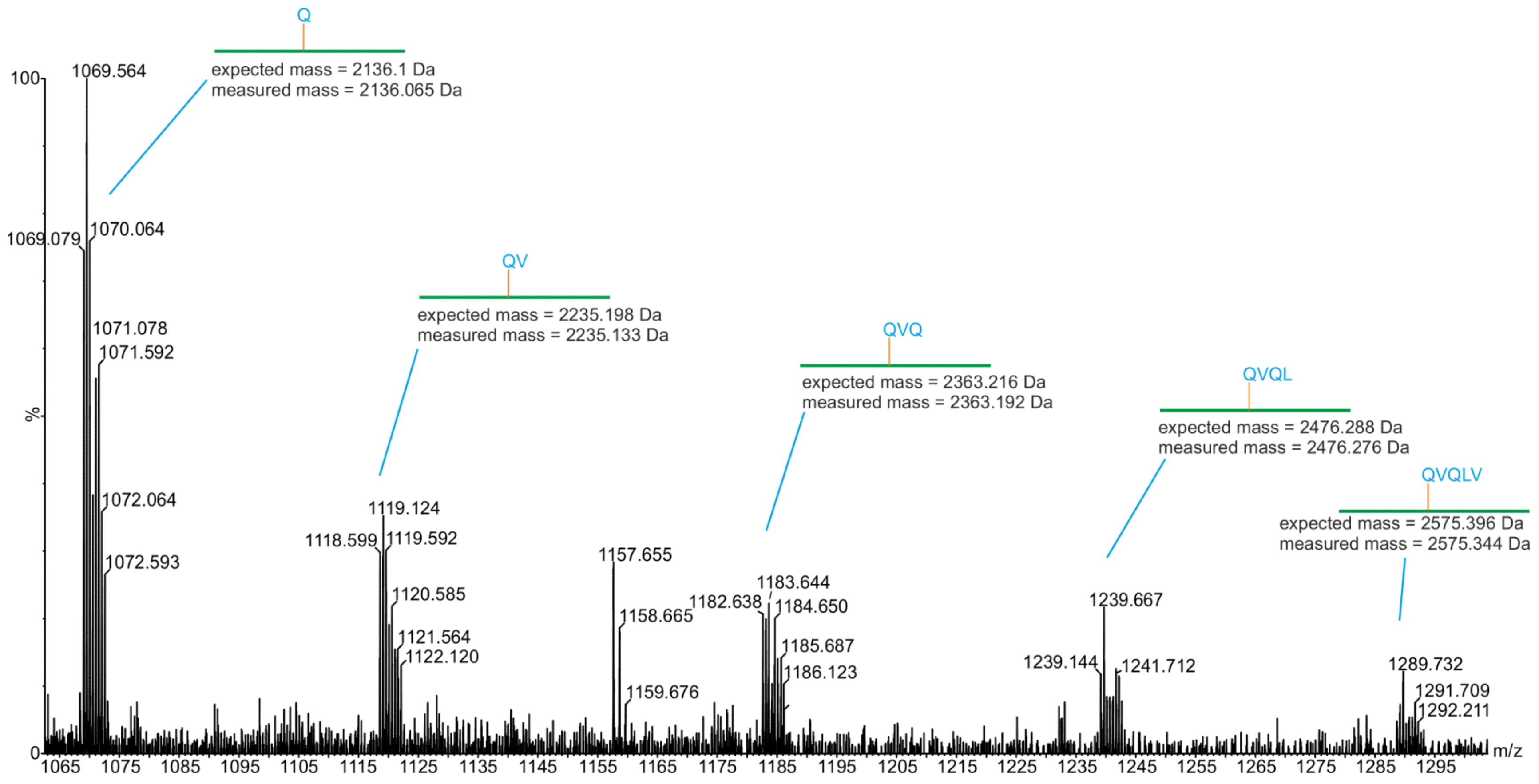


Figure 5.10: Analysis of the doubly charged species formed from the MS/MS spectrum of the cross-linked peptide at 824.2 *m/z*. The expected and measured masses are presented for each of the peaks showing the addition of sequential amino acids from the previously identified peptides. Green line = LLIYDNNKRPSGIPDR peptide.

As predicted, **Figure 5.10** reveals that the doubly charged species observed in the MS/MS spectrum correspond to peptides from the variable region of the light chain, cross-linked to fragments of the N-terminal heavy chain. This confirms that the cross-link is occurring at the N-terminus of the heavy chain to the lysine residue (K54) on the other peptide.

Strikingly, only one cross-link was identified from the 300 kDa complex of the WFL dimer. Although unexpected, this result can be explained as one of the cross-linked sites is at the N-terminus of the heavy chain, at a glutamine residue. An N-terminal glutamine or glutamic acid residue has been shown to be able to cyclise spontaneously to a pyroglutamate (Liu et al., 2011) (**Figure 5.11**).

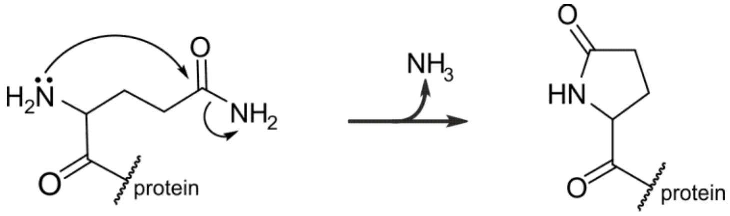


Figure 5.11: Cyclisation of glutamine to pyroglutamate.

Figure 5.11 shows that the amine group at the N-terminus is lost upon cyclisation to a pyroglutamate and therefore so is the cross-linkable site. It was suggested by MedImmune that around 95 % of the molecule is believed to form the pyroglutamate and would explain the difficulty in identifying the cross-linked peptides.

However, the identification of this cross-linked site suggests that the self-association of the WFL dimer is driven by a Fab-Fab interaction.

5.6 Visualising the cross-link in a structural model

In order to investigate the site responsible for the self-association of WFL and visualise the identified cross-link, two simplistic models were proposed (**Figure 5.12**). These models are similar to those explored in Chapter 4 (see **Section 4.3.3**).

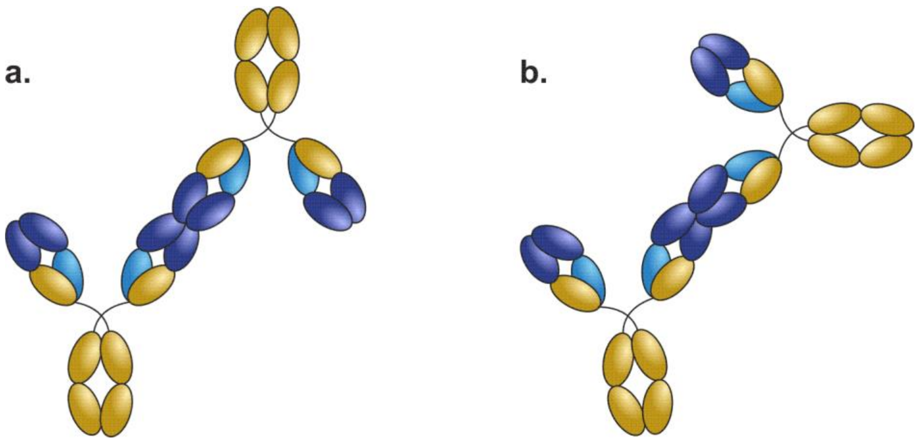


Figure 5.12: Hypothesised models of WFL self-association based on the chemical cross-linking data.

Both models are consistent with the cross-linking data and furthermore, would place residues W30, F31 and L57 at the heart of the binding interface, which was predicted, as mutating these residues ameliorates the self-association. Finally, these models are attractive as both leave a Fab arm free on each mAb either side of the binding interface, which would allow for further oligomer extension. However, to achieve higher resolution models a docking approach was employed to utilise the restraints that had been established from the chemical cross-linking data.

As the mutations from WFL to STT lie in the variable regions of the IgG, the scFv homology models (supplied by MedImmune plc.) of each mAb were used for docking and to generate high resolution models. The HADDOCK software was used to generate the docked structures (van Zundert et al., 2016). Firstly, models were generated by driving the docking procedure using residues W30, F31 and L57 of each molecule as the active residues (**see Section 3.3.5)** (**Figure 5.13**). Additionally, while these residues were specified as the active residues since they are mutated in STT, hydrogen-deuterium exchange (HDX) mass spectrometry data, obtained by Dr J.J. Phillips at MedImmune plc, shows protection at these residues, amongst others, upon dimer formation (see **Figure 5.17**). It was of interest to observe if the models generated were in agreement with the cross-linking data without further refinement.

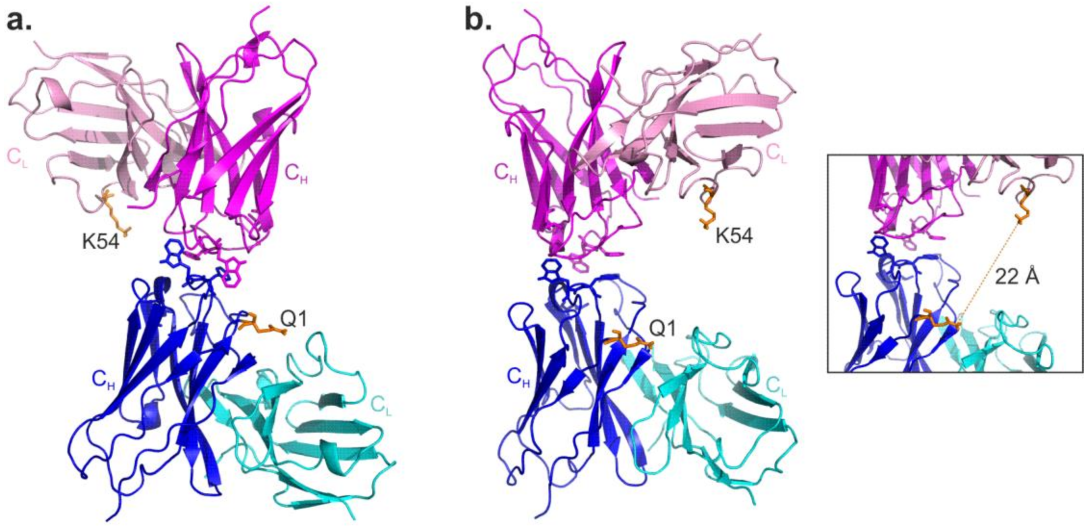


Figure 5.13: HADDOCK docked dimer structures of WFL from the top two clusters. **(a)** WFL dimer from the first cluster showing the two monomeric subunits coming together with a ~180 ° rotation from each other. **(b)** WFL dimer from the second cluster showing the two monomeric sub units coming together with the heavy and light chains parallel. Inset shows the distance between the identified cross-linked residues. Identified cross-linked residues are represented as orange sticks. As the cross-linked residues are 22 Å apart, this model needs refined.

As only three solvent exposed residues were set as the active residues to drive the docking procedure, models with different orientations were generated, centred on residues W30, F31 and L57 (**Figure 5.13**). The two top clusters show models with one scFv (pink) rotated around the other. The cross-linked residues, Q1 and K54, are indicated as orange sticks and the first structure (**Figure 5.13a**) shows that these two residues are at opposite sides of the dimer interface and therefore a cross-link would not be possible. Interestingly however, the second model (**Figure 5.13b**) has the two structures orientated in such a manner that the cross-linked residues are on the same side of the interface. However, the inset shows that the distance between these two residues is 22 Å. The distance restraint set by BS3 is 11.4 Å and therefore the model does not satisfy the cross-linking data.

In order to incorporate the cross-linking data a second docking procedure was carried out, this time adding a distance restraint of 11.4 Å between the identified cross-linked residues (see **Section 3.3.5**), Q1 (Cδ) and K54 (Nζ); while maintaining the dimer interface set by residues W30, F31 and L57 (**Figure 5.14**).

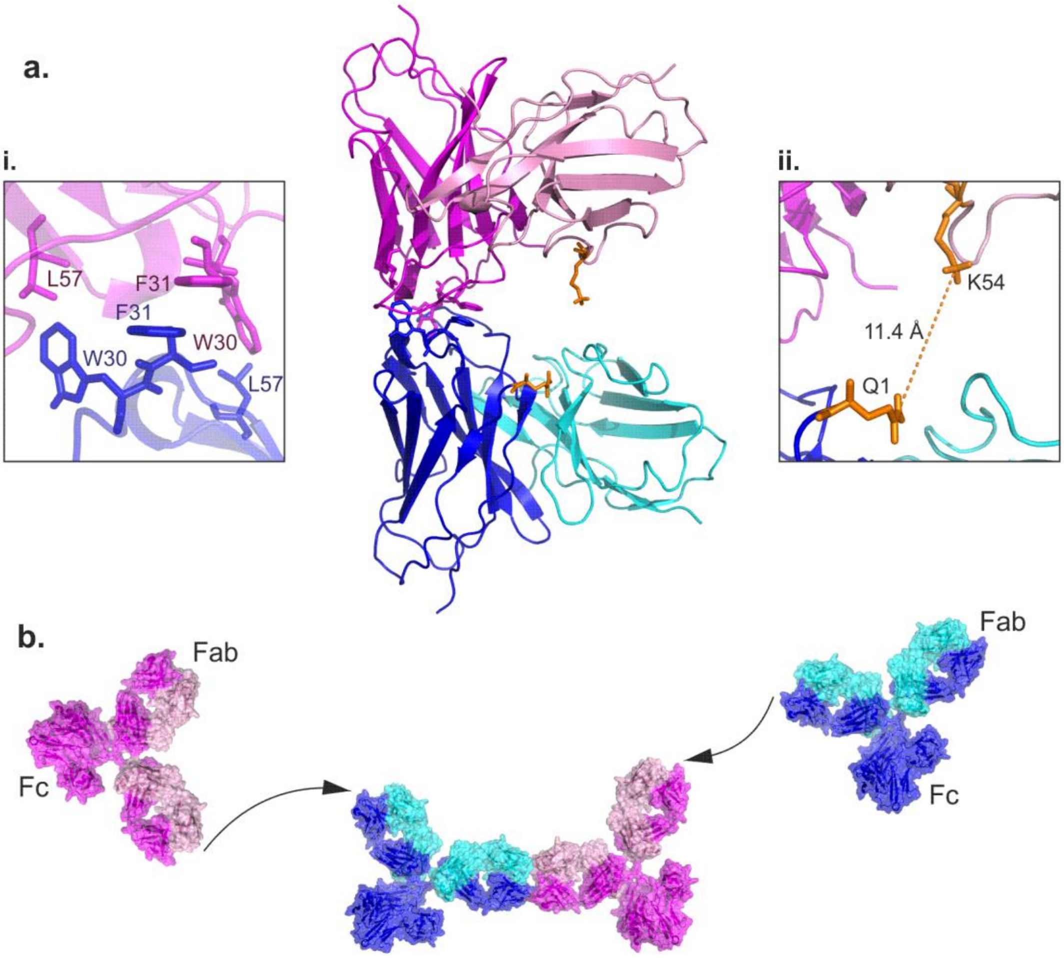


Figure 5.14: Proposed model for the self-association of WFL. **(a)** Atomistic model showing the refined docked dimer of WFL with (i) no disruption to the dimer interface upon refinement and (ii) the cross-linked residues within 11.4 Å of one another. **(b)** Proposed model for the oligomerisation of WFL, driven through Fab-Fab interactions with the potential binding sites for further oligomer extension.

The refined model reveals that it is possible to have the cross-linked residues within 11.4 Å (**Figure 5.14a.ii.**) while not perturbing the hydrophobic dimer interface formed by residues W30, F31 and L57 (**Figure 5.14a.i.**). With these data, a model was proposed which involves the monomeric IgGs of WFL self-associating via a hydrophobically driven, Fab-Fab interaction. This model also explains how larger oligomers can be formed at higher concentrations by the addition of further monomers onto the free Fab arm of each WFL mAb involved, leading to runaway polymerisation (**Figure 5.14b**). Furthermore, this model also facilitates the other cross-link that was identified by the StavroX software that was unique to the dimer (see **Section 5.4.2.2**). Due to the flexibility found within the hinge region, it would be possible for the two Fc domains to come into close contact and for the cross-link between the two CH3 regions to exist. This may be further suggested as the cross-link was not identified in all repeats, as stated before. Interestingly, this model reflects the structures that were earlier observed in the electron micrograph images (**Figure 4.3**).

5.7 Discussion

Mass spectrometry has seen a surge in the field of structural biology with the advances in labelling techniques, especially with systems that are too large or dynamic to be studied through conventional structural techniques (Konermann et al., 2014, Vandermarliere et al., 2015). While a structural proteomics approach will never achieve structures with resolution procurable to techniques such as X-ray crystallography, NMR or cryo-EM, it is the combination of MS with these techniques that proves invaluable. In Chapter 5 it was demonstrated that chemical cross-linking, coupled with MS, led to the determination and characterisation of the site responsible for the self-association and oligomerisation of a therapeutically relevant mAb.

An interesting question that arose from this study was why such few inter-molecular cross-link were identified? Cross-linking experiments typically yield multiple inter-molecular cross links that can be used as restraints to build models (Kosinski et al., 2015, Hofmann et al., 2015). Programs have been designed to aid in the visualisation of multiple cross-links across multi-sub-unit complexes (Combe et al., 2015), so why were only two cross-links identified here? Two possible explanations for this are that (1) there were no other cross-linkable sites due to the mechanism of self-association, and (2) that due to the complexity of working with a homologous system, inter-molecular cross-links were being masked by co-existing intra-molecular cross-links. The docked model was interrogated for any other cross-linkable sites, that were within the 11.4 Å (Nζ to Nζ) distance restraint set by BS3, that should have been identified by sequencing (**Figure 5.15**).

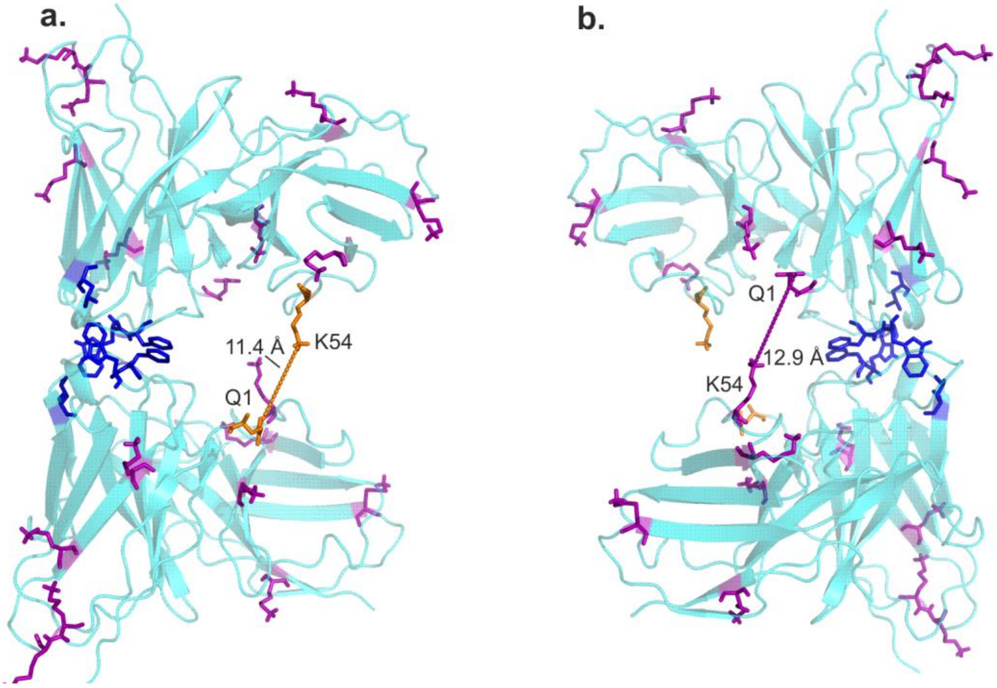


Figure 5.15: Structure showing all possible cross-linkable residues on the proteins surface. **(a)** Front and **(b)** back view of the dimer structure with all cross-linkable residues represented as purple sticks. Identified cross-linked residues are represented as orange sticks with the residues 30, 31 and 57 involved in the dimer interface are shown as blue sticks.

Analysis of other possible cross-links revealed only one other site around the dimer interface that lies slightly out of the distance restraint set (12.9 Å). However, the two residues that would be involved in this cross-link are the same as the identified cross-link, Q1 and K54, but from the other sub-units. Therefore, while it is possible this cross-link may have existed in solution, its impact would have been to add to the abundance of the cross-link previously identified. However, the conversion of glutamine to pyroglutamate lowers the cross-linking efficiency at this site and therefore some cross-links may be being missed. Subsequently as mentioned above, the other possible explanation would have been the masking of inter-molecular cross-links by the same intra-molecular cross-links (**Figure 5.16**).

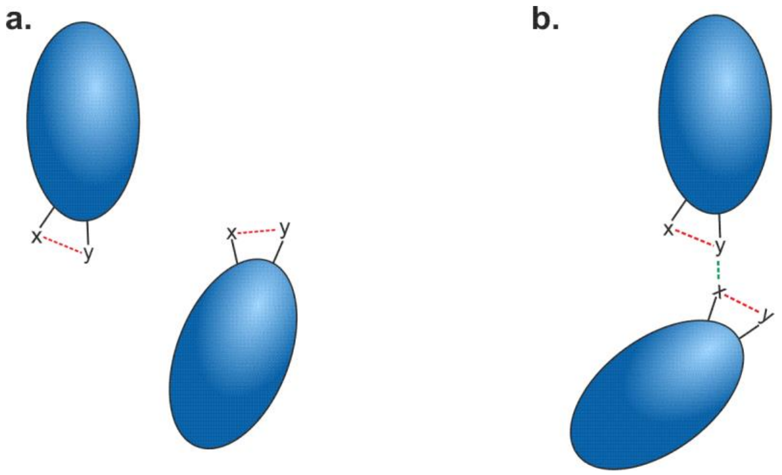


Figure 5.16: Schematic showing the potential for false negative cross-links. **(a)** Schematic showing that an intra-molecular cross-link may exist between residues x and y that also **(b)** exists inter-molecularly upon oligomerisation

If an inter-molecular cross-link is formed upon oligomer formation, but the same cross-link can be made intra-molecularly, there would be no way to differentiate these in the resulting mass spectra. To navigate this possibility, cross-linking oligomers formed by mixing labelled protein (^{15}N) with non-labelled protein, inter-molecular cross-links could be differentiated from their intra-molecular counter parts by their mass.

Although working with a complex homologous system, the approach described here was able to successfully identify an inter-molecular cross-link from which a model could be generated. This model (**Figure 5.14**) was also in agreement with HDX-MS and aggregation prediction data that were acquired by MedImmune (**Figure 5.17**).

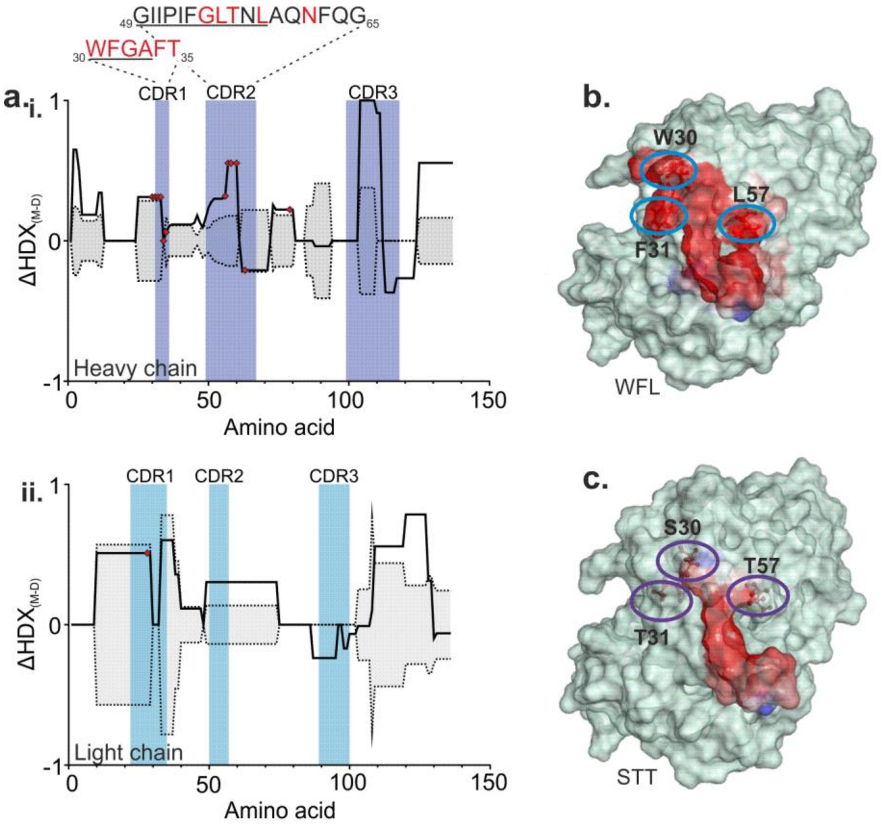


Figure 5.17: Sites of protection observed upon dimer formation and predicted areas for aggregation. **(a)** Localised sites of protection upon dimer formation for the (i) heavy chain and (ii) light chain variable regions of WFL. 99 % confidence limits for significant protection upon dimer formation are shown (grey shaded area, dashed lines). Residues of significant protection are underlined with residues mutated from parent sequence are highlighted in red. **(b)** Spatial aggregation propensity algorithm prediction of WFL for aggregation prone sites (exposed hydrophobic regions) are highlighted in red. **(c)** Spatial aggregation propensity algorithm prediction of STT for aggregation prone sites (exposed hydrophobic regions) are highlighted in red. Residues 30, 31 and 57 are highlighted in both structures. Data was acquired and analysed by Dr J.J. Phillps (MedImmune plc.).

Furthermore, the model presented in this thesis proposes that the WFL monomers are interacting through Fab-Fab interactions; it is hypothesised that these are the interactions that were being observed in class averages of the dimer in the EM micrographs shown in Chapter 4 (**Figure 4.3**). While the individual regions of the structures could not be classified due to the resolution of the images, the data here suggest that the micrographs were presenting the Fab-Fab interactions characterised in this Chapter.

Chemical cross-linking has aided in the characterisation of many protein complexes (Dyche Mullins and Pollard, 1999, Schraven et al., 1990, Brunner, 1993, Singh et al., 2010, Hall et al., 2016, Greenberg et al., 2016, Zheng et al., 2016, Yu et al., 2015). However, one area where cross-linking is less frequently applied is in biopharmaceutical characterisation, particularly in aggregation-prone candidates. This may be because such an approach often required computational workflows to aid data analysis and other structural data in order to yield reliable models. However, with the plethora of structures now available in the protein data bank (PDB), and the more recent EM data bank (EMDB), it is predicted that cross-linking combined with mass spectrometry will continue to be a useful technique in the study of the structure and topology of large dynamic protein systems (Leitner et al., 2016, Sinz et al., 2015). Therefore, it is proposed that the method utilised in this study, in combination with other structural MS techniques such as HDX-MS or hydroxyl radical foot printing (FPOP), could be used as a generic approach to characterise such systems. The promising advantages of utilising structural MS methodologies are their ability to characterise large, dynamic systems that otherwise cannot be studied through conventional approaches and finally, removing the requirement for large volumes of protein at high concentrations to undertake such studies. The combination of structural MS techniques could therefore aid in the re-engineering of aggregation-prone biologics, and assist in the improvement of their biophysical properties.

In summary, the site responsible for the self-association of a therapeutically relevant IgG was characterised here using XL-MS. However, one interesting question that arose from this study was whether all mutations of WFL to STT are needed to remove the propensity of the IgG to self-associate? Positions 30, 31 and 57 were selected for mutation due to these sites being altered from the parent sequence during the affinity maturation process, which unfortunately resulted in the poor biophysical characteristics and self-association; while not perturbing antigen binding. Furthermore, these residues were also highlighted upon analysing the sequence with an aggregation propensity algorithm (**Figure 5.17**) (spatial aggregation propensity, SAP (Voynov et al., 2009). However, it was unclear whether mutations at all three sites were needed, or if one residue was key into disrupting the propensity to aggregate. This was next interrogated using scFvs of different sequences and is described in Chapter 6.

Chapter 6:

Developing an *in vivo* screen to identify aggregation-prone sequences in biopharmaceuticals

6 Developing an *in vivo* screen to identify aggregation-prone sequences in biopharmaceuticals

6.1 Objectives

The work presented in Chapter 5 characterised the region responsible for the self-association of the IgG WFL, with the three positions (W30, F31 and L57) mutated to STT playing prominent roles in driving the self-association. However, it was unclear whether all three residue substitutions are needed to mitigate the self-association propensity of WFL. The final objective of this study was to investigate the residues involved individually, and in combination, to understand if all three mutations are critical to ameliorate the self-association. An *in vivo* system was employed to carry out the investigation which involved moving from full IgGs to scFv format.

6.2 Introduction

6.2.1 In vivo system for monitoring aggregation

Chapter 6 utilises a published *in vivo* method to monitor the aggregation of client proteins (Saunders et al., 2016). The assay exploits the natural antibiotic defence system of *E. coli* and uses β-lactamase as a reporter. In a split enzyme assay, the protein of interest is cloned into a glycine-serine (G/S) linker separating the two domains of β-lactamase (**Figure 6.1a+b**). This system was first utilised to characterise protein stability *in vivo* (Foit et al., 2009, Hailu et al., 2013). As the assay is rapid to perform and the analysis simple, it was postulated that this would be a convenient technique to use to examine the effects of multiple combinations of mutations on the aggregation propensity of scFv molecules.

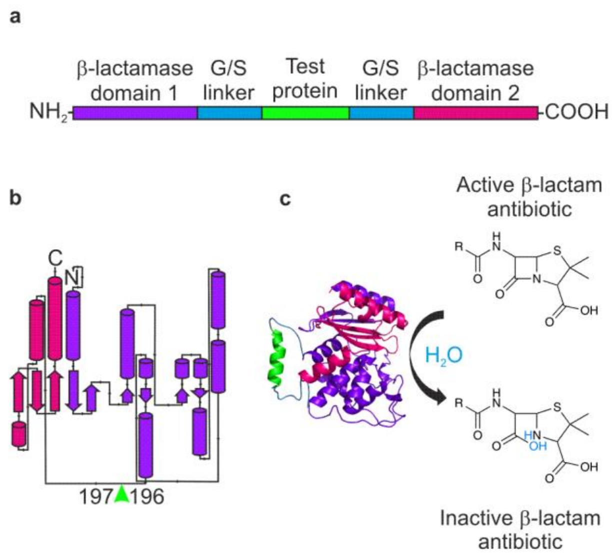


Figure 6.1: Schematic of the tri-partite β-lactamase construct. **(a)** The protein of interest (green) is inserted in between the two domains of β-lactamase (purple and pink) connected via glycine-serine (G/S) linkers (blue). **(b)** Structural schematic showing where the protein of interest lies in respect to the two β-lactamase domains. **(c)** Fully folded tri-partite construct can fold to form the functional enzyme and cleave β-lactam antibiotics. Figure taken and adapted from (Saunders et al., 2016).

Upon expression of the construct, the two halves of the β-lactamase can fold to form the fully functioning enzyme, which cleaves β-lactam antibiotics in the surrounding periplasm (**Figure 6.1**). The insertion of an aggregation prone protein into the construct causes the functioning enzyme to be pulled into the aggregate or degraded, resulting in a loss of β-lactam antibiotic resistance. The assay is performed in a 48-well agar plate, whereby the ability of the bacteria to grow at increasing concentrations of antibiotic is recorded. The assay is quantified by measuring the maximal dilution at which the cells grow across increasing ampicillin concentrations (**Figure 6.2**).

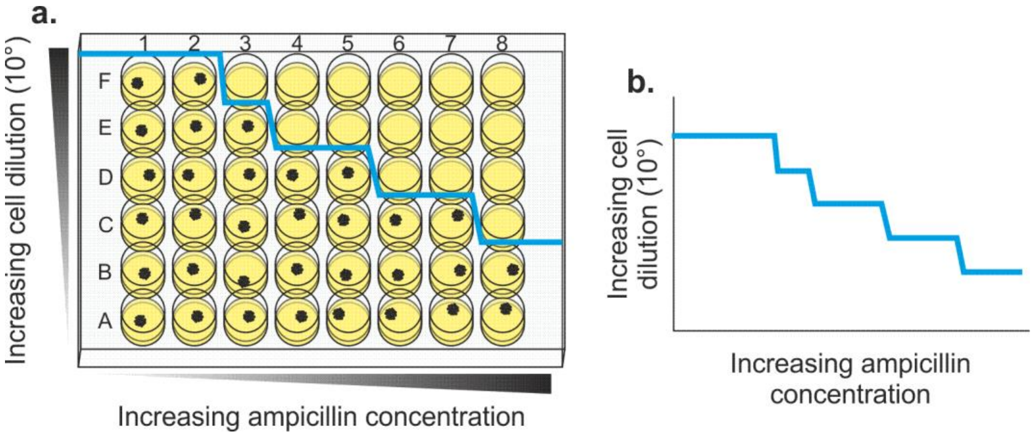


Figure 6.2: Schematic showing the *in vivo* assay work flow. **(a)** Cell dilutions are plated across a 48-well agar plate containing increasing ampicillin concentrations. Maximal cell growth at a given dilution is scored at each antibiotic concentration. **(b)** Data plotted in graph format for comparison between constructs.

The data, plotted as a graph of the maximal cell dilution the cells grow at against the increasing ampicillin concentration (**Figure 6.2a**), enabled comparison of the aggregation propensity of different test proteins. Furthermore, the results of multiple *in vivo* assays can be compared to one another by calculating the area under the plotted curves (**Figure 6.2b**) and represented as bar charts; making the comparison of multiple constructs easier for the analyst.

While the *in vivo* assay has been successfully applied to the characterisation of amyloidogenic proteins and identifying inhibitors for these systems (Saunders et al., 2016), this study aimed to extend the application of the *in vivo* assay to characterising proteins relevant to the biopharmaceutical sector.

6.2.2 Characterising biopharmaceuticals using the in vivo assay

The β-lactamase tripartite system has been shown to maintain function after incorporation of proteins up to 40 kDa in size (Foit et al., 2009). As for the characterisation of biopharmaceuticals, while this system is unable to characterise full IgGs, it is ideal for the analysis of the scFvs. This is desirable since it has been found that the aggregation of mAbs can often be due to the sequences found in the variable regions of the Fab arms (Wang et al., 2009, Wu et al., 2014). Furthermore, many industries have standard scaffolds onto which candidate IgG fragments are grafted for the formation of the functional mAb molecule (Renaut et al., 2012, Stewart et al., 2012). Therefore, it was hypothesised that β-lactam antibiotic resistance would be lost upon the insertion of an aggregation prone scFv into the β-lactamase tripartite system (**Figure 6.3a**), similar to the results published for the amyloidogenic proteins (Saunders et al., 2016).

As with the amyloid systems studied previously, it was hypothesised that upon expression in the periplasm β-lactamase function would persist when a stable scFv was inserted into the construct (**Figure 6.3b**). However, the insertion of an aggregation prone scFv, such as WFL, would cause self-association and pull the β-lactamase into the aggregate which would be reported through the loss of antibiotic resistance and, ultimately, cell death (**Figure 6.3c**).

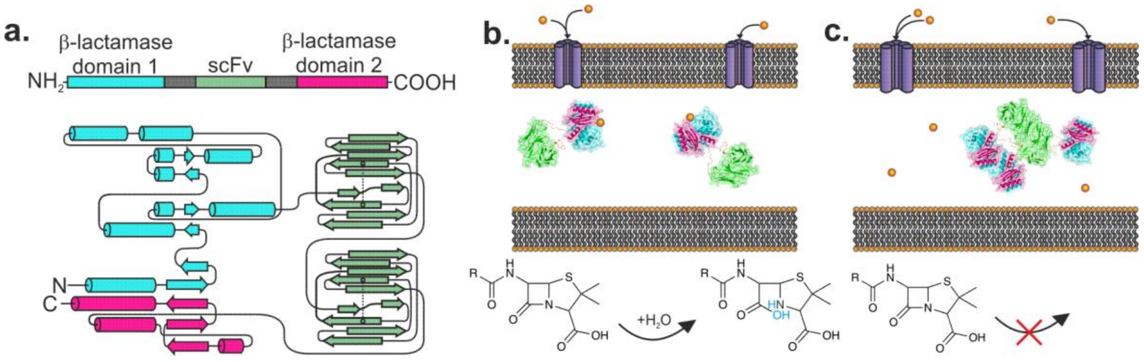


Figure 6.3: scFv-β-lactamase tri-partite system. **(a)** ScFv of interest is cloned into the linker region of β-lactamase. **(b)** β-lactamase remains monomeric and functional with a stable scFv insert and can cleave β-lactam antibiotics. **(c)** Aggregation prone sequences pull β-lactamase into the aggregate and host β-lactam antibiotic resistance is lost. Orange spheres = antibiotic.

To test the potential of the *in vivo* assay as a platform for identifying aggregation prone scFvs, the model pair at the focus of this study, WFL and STT, were used. The question of whether all three mutations were needed to ameliorate the self-association of WFL was of particular interest, in addition to the possibility of being able to identify differences between the different point mutations.

6.3 In vivo assay identifies differences of aggregation prone scFvs

6.3.1 In vivo analysis of candidate scFvs

For proof of concept it was important to analyse the effects that WFL and STT, in single-chain format, have on the tripartite β-lactamase system. Fortunately, the WFL_scFv and STT_scFv sequences had previously been cloned into the β-lactamase construct (creating βla-WFL and βla-STT) and screened in the *in vivo* assay (**Figure 6.4**).

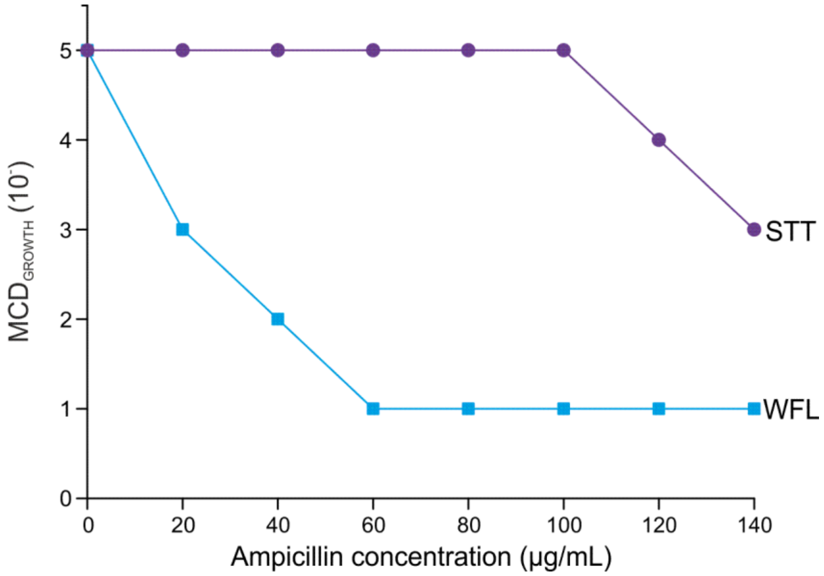


Figure 6.4: Comparison of *in vivo* assay data for WFL and STT sequences. The maximal dilution (MCD) at which cells still grow is monitored. Data obtained by Dr Janet Saunders (University of Leeds).

A striking difference was observed between the two constructs, with bacteria expressing the βla-STT outperforming βla-WFL significantly. Since a significant difference could be observed between the two constructs, it was decided to screen the individual point mutations between WFL and STT to understand the effect of these mutations on *in vivo* aggregation. Using either the WFL or STT plasmid DNA, Q5 mutagenesis was carried out to mutate the residues of interest (**Table 6.1**) (see **Section 3.5.3**).

Table 6.1: Table of constucts and mutated residues, from WFL, made to compare the aggregation of the scFvs in the MIC assay.

Construct	Mutations from WFL
WFL	-----
SFL	W30S
WTL	F31T
WFT	L57T
STL	W30S, F31T
SFT	W30S, L57T
WTT	F31T, L57T
STT	W30S, F31T, L57T

After confirmation of the correct sequences, the effects of the individual mutations were screened in the *in vivo* assay to analyse their effect on aggregation of the scFv-β-lactamase constructs (**Figure 6.5**, plates for WFL, SFL and STT can be seen in **Appendix 8.41**).

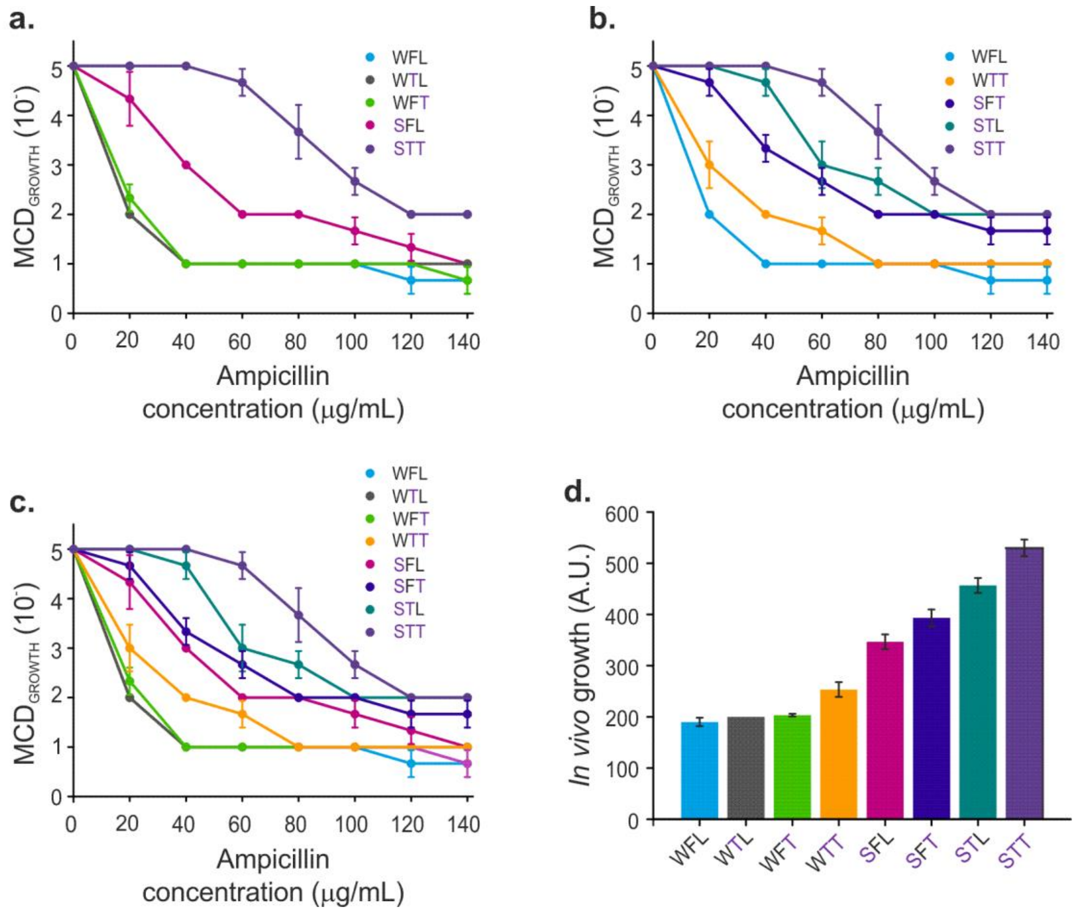


Figure 6.5: Comparison of *in vivo* assay results across all generated constructs. **(a)** Assay results for all single point mutations from WFL to STT. The parent WFL and STT results are plotted for comparison. **(b)** Assay results for single point mutations from STT to WFL. WFL and STT results are plotted for comparison. **(c)** Assay results for all constructs tested. **(d)** Bar chart showing the area under the curves (A.U.) for all constructs. Error bars show the standard error of the mean, n = 4 across all experiments, data collected with Dr Janet Saunders.

Strikingly, the *in vivo* assay was able to distinguish the aggregation propensity of the different constructs. Furthermore, this demonstrates the use of the *in vivo* assay for the characterisation of biopharmaceutical candidates which share an incredible sequence homology, yet have significantly different aggregation profiles.

The results of the assay show that tryptophan 30 has the largest effect on *in vivo* growth. **Figure 6.5a** highlights this with a significant increase in cell growth being observed only upon mutating W30 to S30. This result is reflected in **Figure 6.5b** where a significant decrease in cell growth is observed in mutants only containing tryptophan at position 30 (orange and blue). This result was confirmed upon analysis of the full data, where mutations of WFL at positions 31 (F31T, WTL – purple) and 57 (L57T, WFT – green) have no significant effect on cell growth (**Figure 6.5d**). While position 30 plays the most significant role in self-association, the double mutations indicate that F31 also plays a role in the self-association to a larger degree than L57 (**Figure 6.5b+d**). However, it is clear that all three mutations are needed to fully ameliorate the self-association of WFL.

In order to confirm that the results obtained in the MIC assay were due to protein self-association and aggregation, western blots were carried out on all of the constructs to ensure that the expression levels of the proteins were the same (**Figure 6.6**).

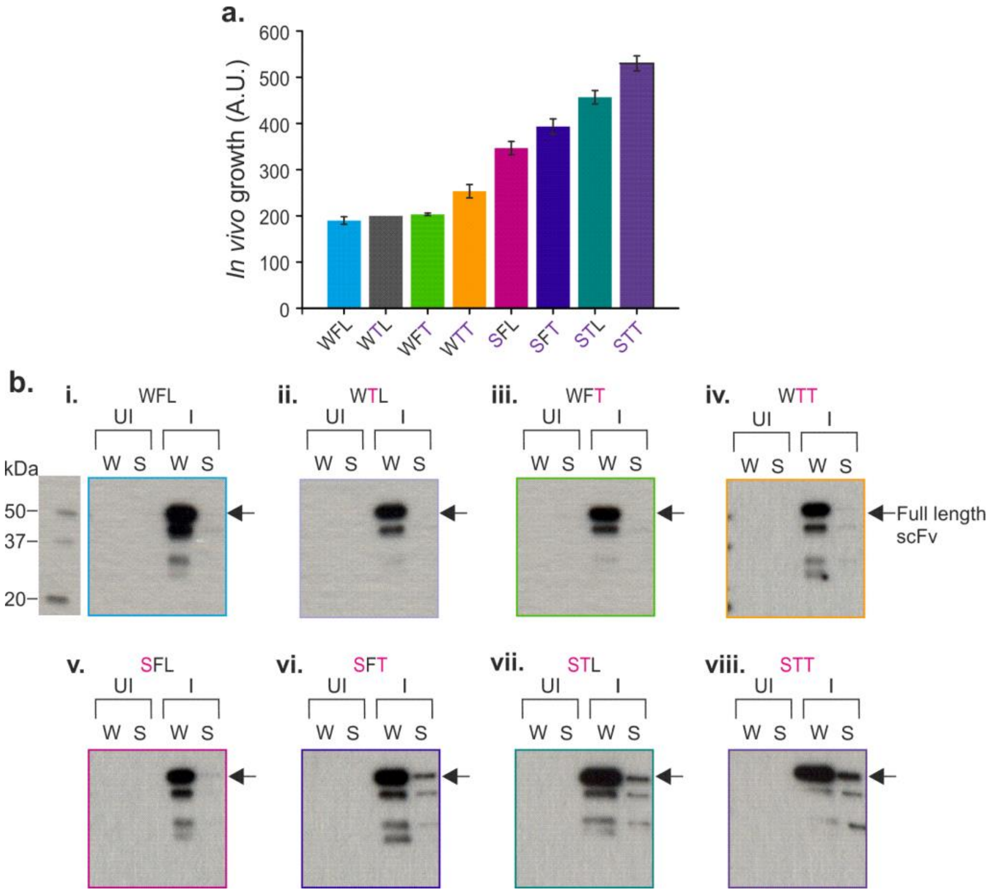


Figure 6.6: *In vivo* results are a result of protein aggregation and not a difference in expression levels. **(a)** Bar chart showing the area under the curves for all constructs. n = 4 **(b)** Western blots for the un-induced (UI) and induced (I) constructs. The whole (W) and soluble (S) fractions are shown for each construct. The arrows indicate the full βla-scFv constructs. An anti-β-lactamase antibody (CSB-PA352353YA0IENL) was used for blotting. Western blots were carried out by Dr Janet Saunders and Dr Matthew Jackson (University of Leeds).

Western blot analysis revealed that the total expression levels of the proteins was similar and therefore the results observed are due to the self-association and aggregation of the different constructs; rather than differences in expression levels. Furthermore, it can be observed that the amount of protein found in the soluble fraction of the cells correlates well with the *in vivo* growth observed, hence the less aggregation prone the sequence, the more protein is found in the soluble fraction of the cells.

6.3.2 Analysing the structural effects of the different point mutations

Homology models were built from the WFL_scFv structure (supplied by MedImmune plc) to analyse the effects of the different mutations on the surface properties of the scFv. The hydrophobic surfaces of the different mutants were analysed to interrogate how these patches may aid in the self-association (**Figure 6.7**).

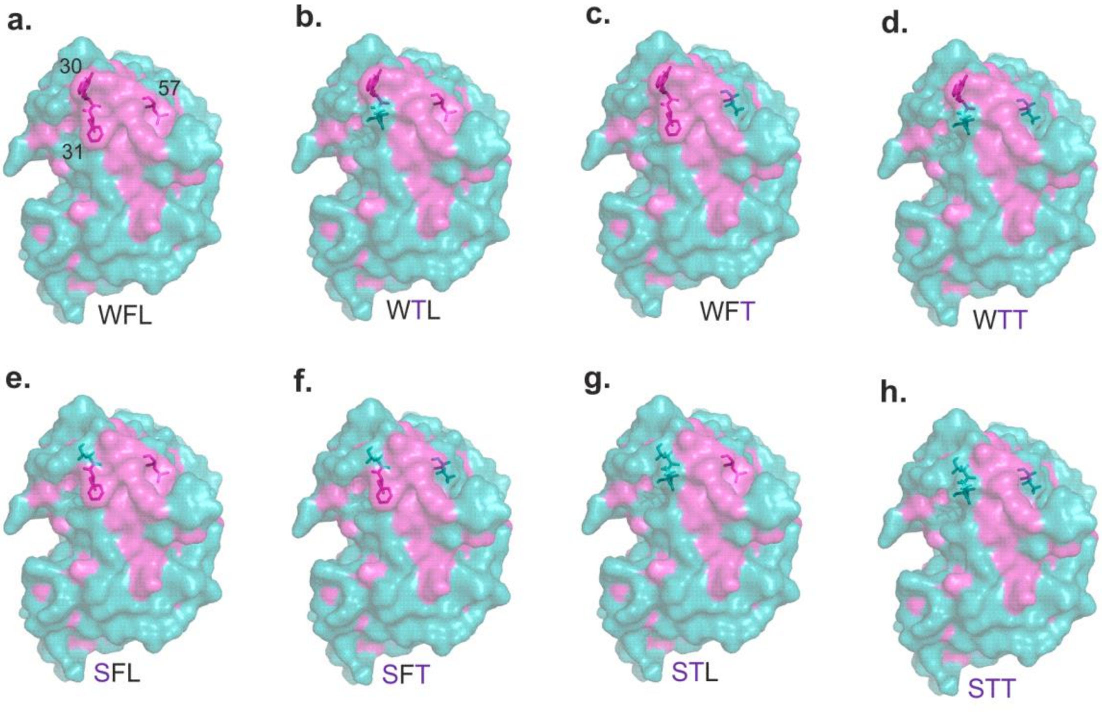


Figure 6.7: ScFv Homology models of the **(a)** WFL, **(b)** WTL, **(c)** WFT, **(d)** WTT, **(e)** SFL, **(f)** SFT, **(g)** STL and **(h)** STT. Residues 30, 31 and 57 are shown as sticks. Pink surface = hydrophobic areas, blue surface = rest of protein.

The mutations from WFL > STT reduce the area of the exposed hydrophobic patch and once again reveal how W30, F31 and L57 are aiding the self-association that was mapped in Chapter 5. Interestingly, the results from the MIC assay show that there is a substantial effect on cell growth upon mutation of W30S. However, a large proportion of the hydrophobic patch remains present in this construct. The structure of the SFL (W30S) mutant was investigated in further detail to observe the effects this mutation had on the proteins surface (**Figure 6.8**).

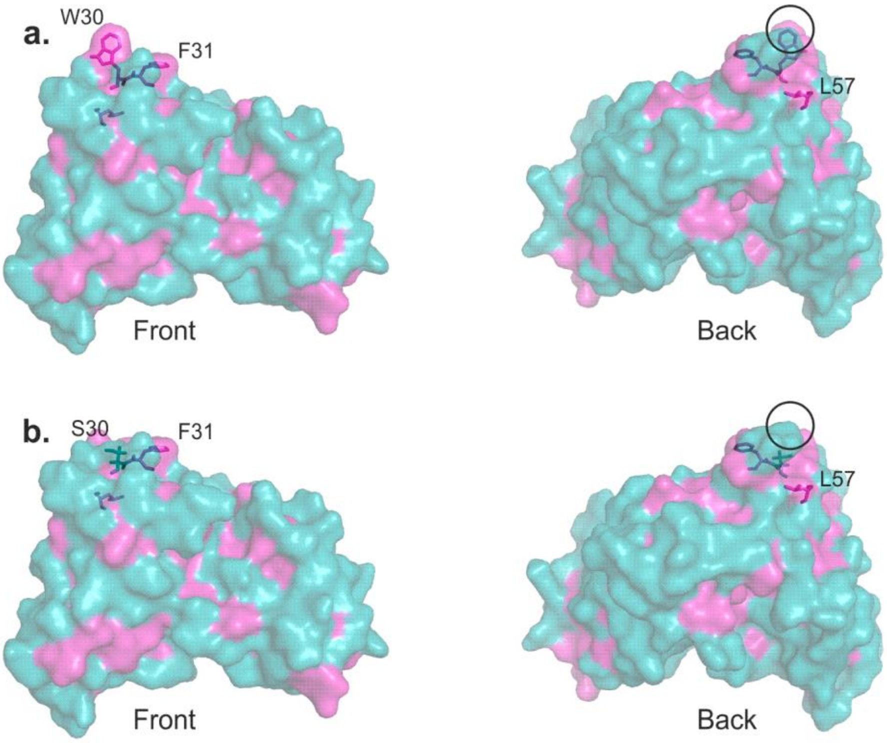


Figure 6.8: Structural differences between **(a)** WFL and **(b)** SFL as a result of mutation at position 30 (W30S). Circle = mutated tryptophan at position 30 between the two constructs.

Closer analysis of the SFL mutant revealed the removal of an important surface-exposed tryptophan residue that plays a strong role in the self-association of WFL, as also shown in Chapter 5. The results from the *in vivo* assay highlight how important this residue is in mediating the interaction between WFL monomers, with around 50 % of cell growth being recovered by this mutation alone (when compared with STT (**Figure 6.5**)).

While it is clear that individual mutations have a significant effect on restoring cell growth *in vivo*, it was important to test whether aggregation in the tripartite β-lactamase system results from the same mechanism of self-association as observed for the full IgGs *in vitro*. Therefore it was decided to sub-clone the scFvs out of the β-lactamase construct, so that the individual scFvs could be purified and their aggregation propensities characterises *in vitro*. As WFL and STT had been fully characterised as mAbs, these two sequences along with SFL (W30S) were selected for characterisation, since SFL has significant effects on cell growth *in vivo* this construct was of particular interest to analyse whether it aggregates in a similar manner to that of WFL.

6.4 In vitro characterisation reveals WFL_scFv self-associate through the same mechanism as IgG

6.4.1 Purification of WFL_scFv and STT_scFv

As WFL performed the poorest in the *in vivo* assay, protein preparation and purification procedures were established and optimised on this construct along with STT_scFv. It was anticipated that all other constructs could be expressed and purified using the same procedure that was optimised on WFL_scFv.

6.4.1.1 Cloning and protein expression trials

The results presented thus far indicate WFL aggregates both *in vitro* as an IgG and *in vivo* as a scFv. One key question regarding the observed aggregation was whether the mechanism of self-association and aggregation the same for the IgG and the equivalent scFv. To explore this, both WFL_scFv and STT_scFv were expressed and purified to enable characterisation of their self-association *in vitro*. Restriction digests of WFL_scFv and STT_scFv were used to sub-clone the scFvs out of the β-lactamase construct and a hexa-histag was added to the C-terminus of the scFvs to aid in purification (**Figure 6.9**) (see **Section 3.5.2.1**).

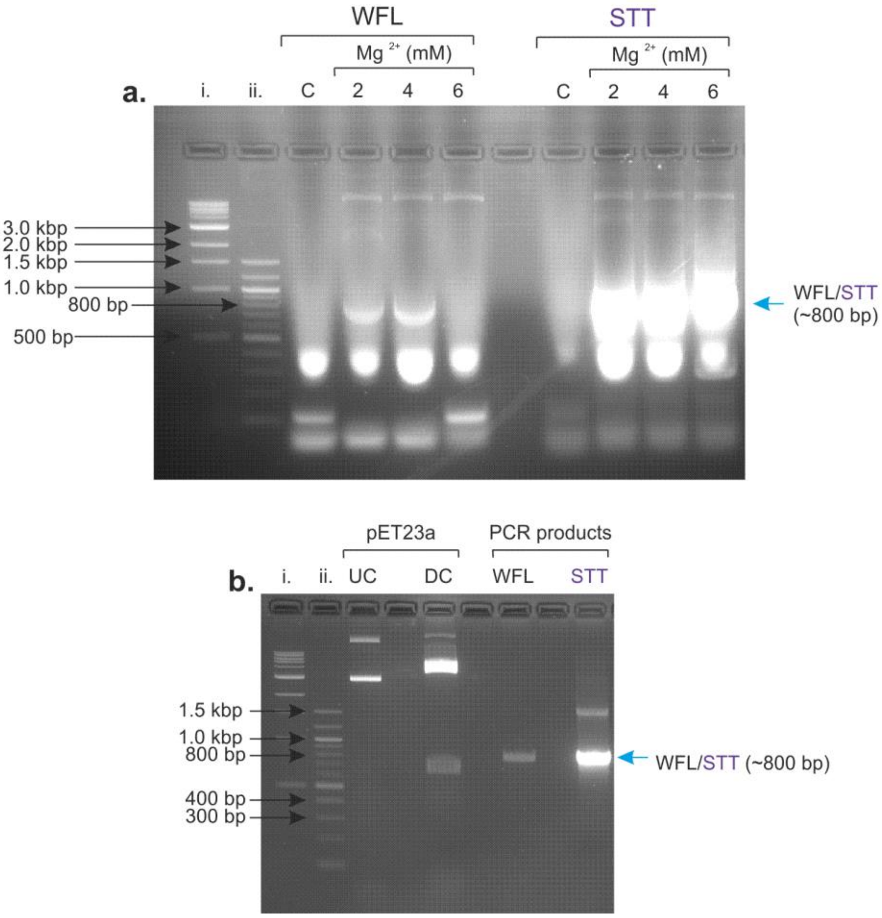


Figure 6.9: Results from PCR amplification. **(a)** Agarose gel showing the amplification of WFL and STT scFv sequences from the scFv-β-lactamase containing plasmid. C = control with no Mg^{2+} **(b)** Agarose gel showing the results of the double cut vector and the purified DNA sequences from PCR amplification. UC = uncut plasmid, DC = double cut. i = 1 kbp DNA ladder and ii = 100 bp DNA ladder.

WFL_scFv and STT_scFv DNA was amplified and ligated into a pET23a vector to be transformed into *E. coli* cells for expression and purification. As WFL has a strong propensity to aggregate, the pLemo cell line was chosen to express the two single-chain proteins. As less expression can often yield higher protein yields of difficult to express proteins (Rosano and Ceccarelli, 2014), the expression of the scFv containing plasmid can be controlled through the expression of lysozyme; under a separate and tuneable promoter, using L-rhamnose. This also aids the amount of protein found in the soluble fraction of the cells and is therefore attractive for purification.

The pET23a plasmids encoding WFL_scFv and STT_scFv (pET23a-WFL and pET23a-STT, see **Appendix 8.1+8.22**) were transformed into pLemo *E. coli* cells and plated out onto agar plates containing 100 μg/mL ampicillin and 30 μg/mL chloramphenicol (see **Section 3.5.4**). Single colonies from the plates were used to inoculate separate 10 mL of LB media containing 100 μg/mL ampicillin and 30 μg/mL chloramphenicol and grown for 18 h overnight, after which 200 μL of each were used to inoculate 7×10 mL LB media containing 0-2000 μM L-rhamnose containing 100 μg/mL ampicillin and 30 μg/mL chloramphenicol. Cultures were grown at 37 °C, 200 rpm until an OD_{600} of 0.6 was reached, where 2×1 mL aliquots were taken for the analysis of expression levels in the un-induced samples. Protein expression was induced with a final concentration of 400 μM IPTG and the bacteria grown for a further 4 h where a further 2×1 mL aliquots were taken for analysis of expression levels in the induced samples. All samples were corrected to the OD_{600} of the uninduced samples prior to the samples being pelleted by centrifugation. One set of the aliquots were then re-solubilised in 2× SDS-loading buffer (see **Section 3.5.5**) and then analysed via SDS-PAGE (see **Section 3.4.8**) to monitor whole cell protein expression (**Figure 6.10**).

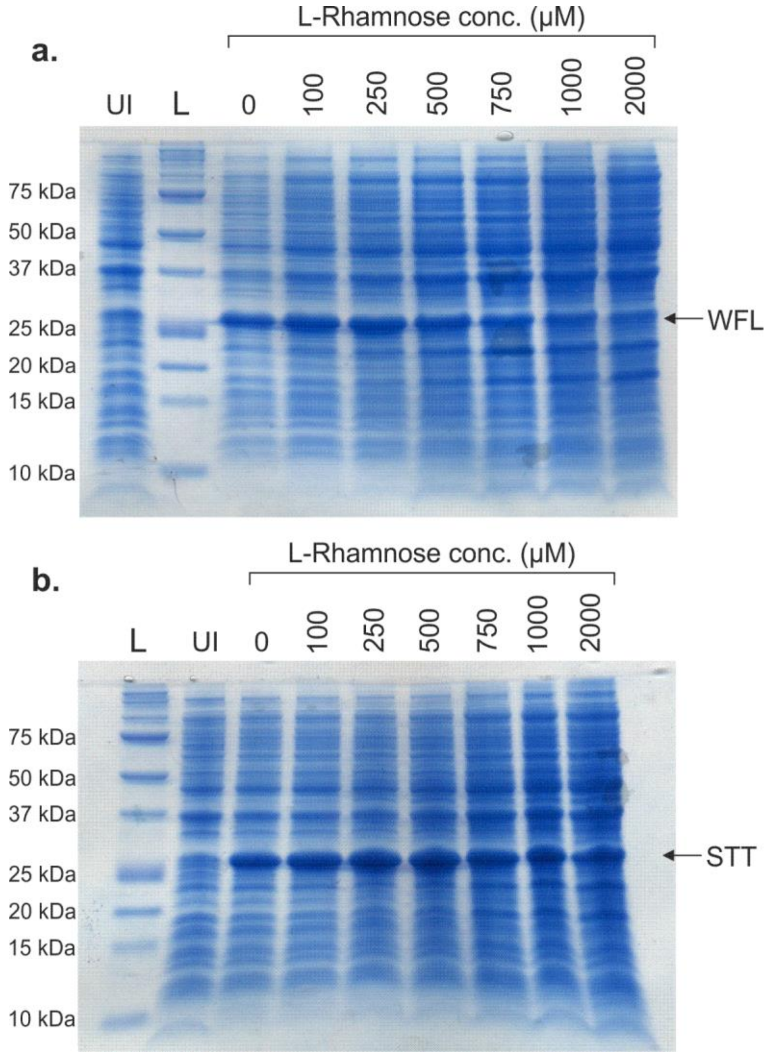


Figure 6.10: SDS-PAGE gel showing the tuneable expression of scFv constructs. **(a)** SDS-PAGE results of WFL_scFv whole cell expression across an increasing L-rhamnose concentration. **(b)** SDS-PAGE results of STT_scFv whole cell expression across an increasing L-rhamnose concentration. UI = uninduced samples, L = protein marker. Coomassie staining.

As the scFvs were to be purified from the soluble fractions, it was important to identify the L-rhamnose concentration at which the largest proportion of the scFvs remained soluble. The pelleted aliquots were used to analyse the amount of protein in the soluble fractions. The pellets were lysed to separate the soluble and insoluble fractions (see **Section 3.5.6**) and separated by centrifugation. The soluble fraction was collected and the protein concentrated by acetone precipitation (see **Section 3.5.6**) and centrifugation. The insoluble and soluble fractions were then re-solubilised in equivalent volumes of 2× SDS-loading buffer and analysed via SDS-PAGE (**Figure 6.11**).

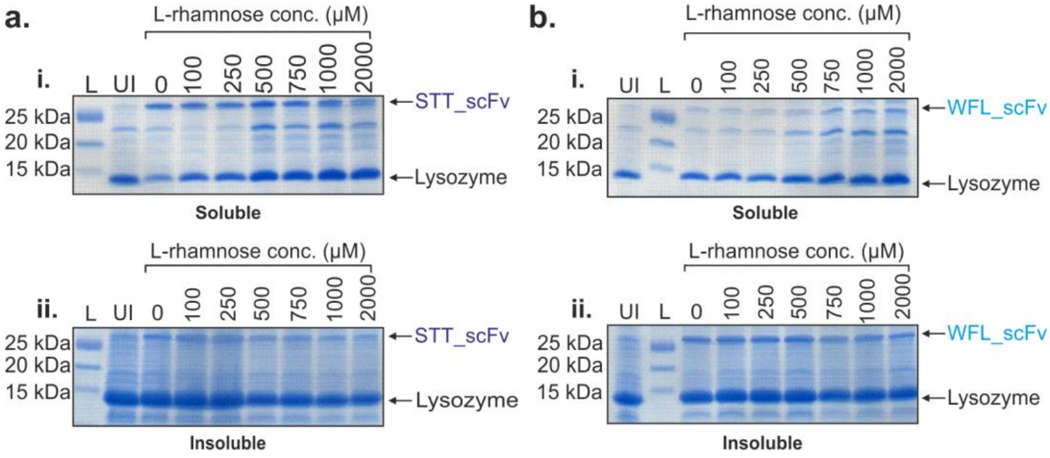


Figure 6.11: SDS-PAGE results showing the amount of protein found in the soluble and insoluble fractions. **(a)** SDS-PAGE showing the amount of STT_scFv found in the soluble (top) and insoluble (bottom) fractions of lysed cells across increasing L-rhamnose concentrations. **(b)** SDS-PAGE showing the amount of WFL_scFv found in the soluble (top) and insoluble (bottom) fractions of lysed cells across increasing L-rhamnose concentrations. UI= uninduced sample, L = protein marker. Coomassie staining.

As suggested from the western blot analysis of the constructs expression (**Figure 6.6**), more of the STT_scFv protein is found in the soluble fraction compared with WFL_scFv. Analysis of the soluble and insoluble fractions revealed that the greatest amount of soluble WFL_scFv was found at an L-rhamnose concentration of ≥ 750 μM, while the 100-500 μM L-rhamnose concentration yielded the largest amount of soluble STT_scFv. L-rhamnose concentrations of 1000 μM and 500 μM for WFL_scFv and STT_scFv were therefore used for subsequent protein preparations.

Once growth conditions had been optimised for maximal protein yields in the soluble fraction, the preparations were scaled up to produce enough scFv proteins for *in vitro* characterisation.

6.4.1.2 scFv purification

The WFL, STT and SFL scFvs were expressed as described in **Section 6.4.1.1** and purified by nickel affinity chromatography. Briefly, single colonies of pLemo competent cells transformed with either pET23a-WFL, pET23a-STT or pET23a-SFL plasmids were used to inoculate 200 mL of LB containing 100 µg/mL ampicillin and 30 µg/mL chloramphenicol in 500 mL flasks. These cultures were incubated for 18 h at 37 °C before 10 mL was used to inoculate each of ten 1 L of 2×TY media (see **Table 3.3**, **Section 3.5.7**) in 2 L flasks, containing 100 µg/mL ampicillin, 30 µg/mL chloramphenicol and either 1000 µM, 500 µM or 750 µM L-rhamnose for the respective constructs (WFL, STT and SFL). The cultures were incubated at 37 °C until the cells reached an OD_{600} of 0.6 (typically 4 h). ScFv protein expression was then induced by the addition of 400 µM IPTG. The cultures were grown for a further 4 h before harvesting. The cells were lysed, as described in **Section 3.5.8**, and the soluble and insoluble fractions separated. The soluble fractions from each of the cultures was taken forward for purification.

To purify the scFv proteins from the other solubly expressed host proteins, the whole fraction was loaded onto a HisTrap column for purification via nickel affinity chromatography. The column was washed with binding buffer (see **Table 3.8**, **Section 3.5.8**) and the soluble fraction was loaded onto the column before elution with a stepwise gradient of elution buffer (see **Table 3.8**, **Section 3.5.8**) (**Figure 6.12-6.13**). 2.5 mL fractions were collected throughout the elution and those containing the scFv proteins were identified via SDS-PAGE. The fractions corresponding to the scFv protein of interest were pooled together before being concentrated and subsequently dialysed against storage buffer (see **Table 3.8**, **Section 3.5.8**). The dialysed protein aliquoted into 1 mL samples, snap-frozen in liquid nitrogen and stored at -80 °C.

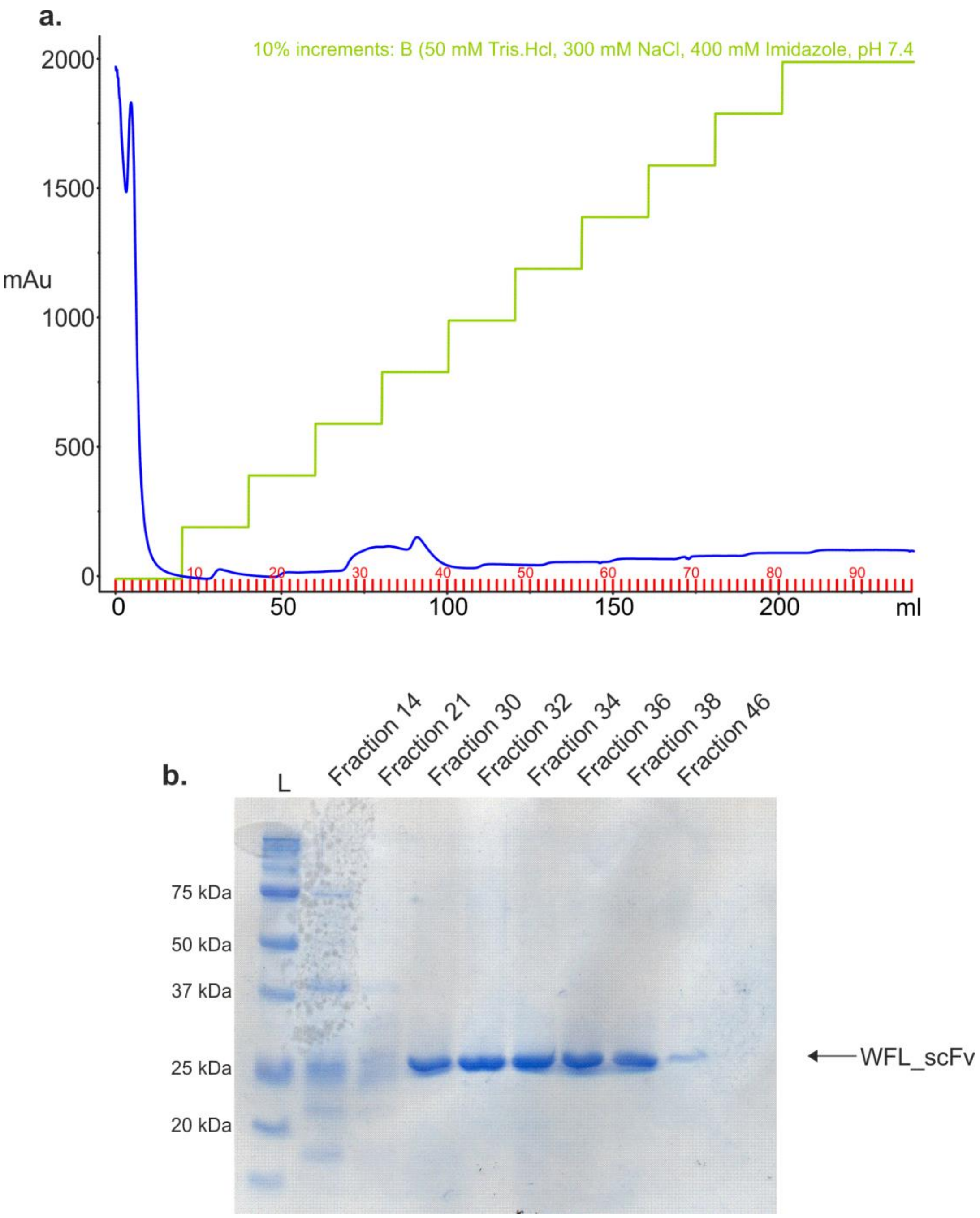


Figure 6.12: Purification of WFL_scFv. **(a)** Elution profile of WFL_scFv from the nickel affinity column. Blue = elution trace recorded at 280 nm, green = trace showing the increasing increments of buffer B. **(b)** SDS-PAGE gel of selected fractions from WFL_scFv elution. L = protein marker. Coomassie staining.

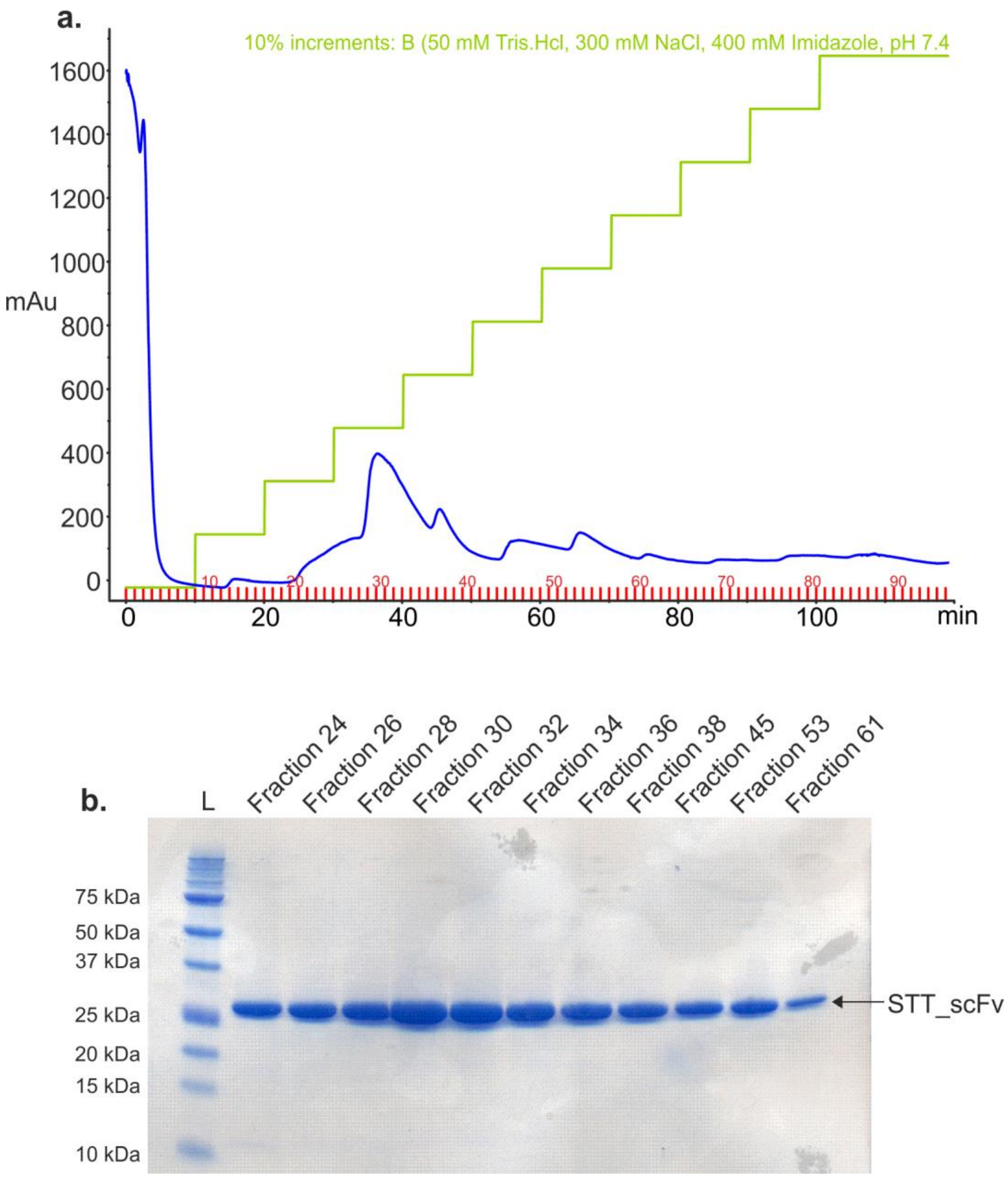


Figure 6.13: Purification of STT_scFv. **(a)** Elution profile of STT_scFv from the nickel affinity column. Blue = elution trace recorded at 280 nm, green = trace showing the increasing increments of buffer B. **(b)** SDS-PAGE gel of selected fractions from STT_scFv elution. L = protein marker. Coomassie staining.

The purified proteins were examined by mass spectrometry to confirm the purity of the samples. Samples of WFL_scFv and STT_scFv were dialysed against 150 mM ammonium acetate, pH 6 before being infused directly into the mass spectrometer for ESI-MS analysis

(**Figure 6.14**).

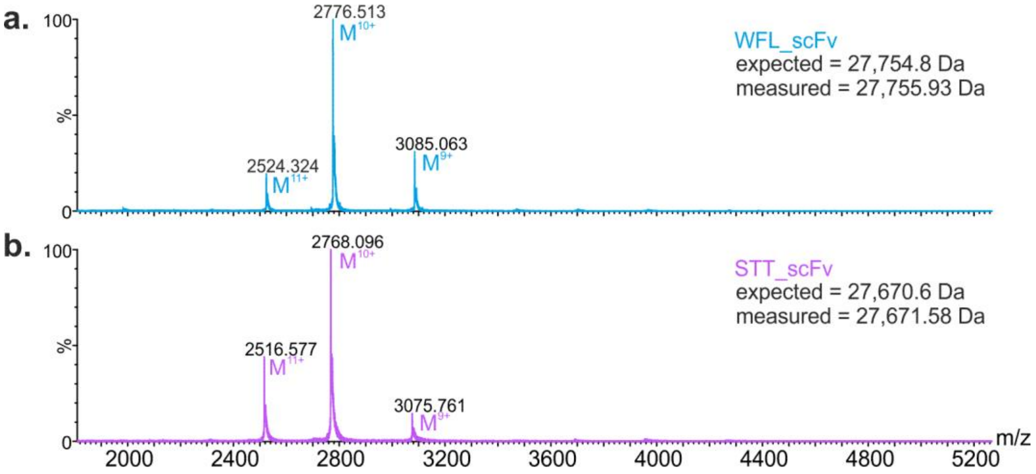


Figure 6.14: Mass spectra of the purified scFvs **(a)** WFL_scFv and **(b)** STT_scFv. Representative spectra of n = 2.

Mass spectrometry analysis confirmed the protein identity with measured mass of the WFL_scFv (27,755.93 Da) and STT_scFv (27671.58 Da) proteins agreeing well with the expected masses (27,754.8 Da and 27,670.6 Da, respectively).

As methods had been established for the large scale expression and purification of scFv constructs, the SFL_scFv (W30S) protein was expressed and purified so that a comparative analysis could be carried out across the scFv construct (**Appendix 8.4**). As expected from the *in vivo* results, STT_scFv yielded the most protein while WFL_scFv yielded the least (**Table 6.2**).

Table 6.2: Table of scFv constructs expressed and purified

Construct	Yield (mg/L)
WFL_scFv	~ 0.3
SFL_scFv	~ 0.5

STT_scFv	~ 2

6.4.2 In vitro analysis of scFv constructs

6.4.2.1 Polyethylene glycol precipitation

To assess the aggregation potential of the different scFvs *in vitro*, a polyethylene glycol (PEG) precipitation experiment was set up. PEG has been shown to cause proteins to precipitate from solution by acting as a water scavenger and increasing the apparent concentration of protein in the sample (Atha and Ingham, 1981, Vasquez-Rey and Lang, 2011, Hammerschmidt et al., 2015, Gagnon, 2008). The experiment was also chosen as MedImmune had acquired PEG precipitation data on the full length IgGs of WFL and STT (**Figure 6.15a**). Samples of WFL_scFv, SFL_scFv and STT_scFv were added to buffer with increasing amounts of PEG-8000 (average molecular weight = 8000) and the amount of protein aggregation was determined using nephelometry (see **Section 3.4.10**) (**Figure 6.15**).

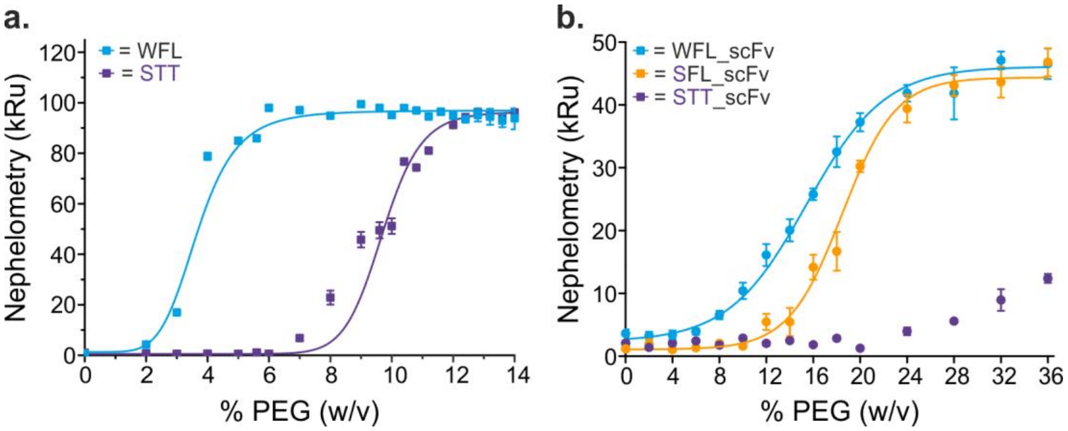


Figure 6.15: PEG precipitation assay showing WFL precipitates out of solution earlier than STT as both IgG and scFvs. **(a)** PEG precipitation results of WFL (blue) and STT (Purple) as full IgGs. Data collected by MedImmune. **(b)** PEG precipitation assay results of WFL (blue), SFL (orange) and STT (purple) as scFvs. IgG data = 6.74 μM, scFv data = 15 μM. n = 3.

The trend between the two precipitation experiments correlates well. It is not unexpected that the IgG samples require a lower percentage of PEG to precipitate since it has been demonstrated that larger proteins experience a greater effect from the PEG in this assay

(Ingham, 1977, Atha and Ingham, 1981). However, in all cases, proteins with the WFL sequence precipitate out of solution at significantly lower PEG concentrations than STT. Importantly, SFL_scFv requires a concentration of PEG-8000 intermediate to that required for WFL and STT (~13 % (*w/v*)) to precipitate out of solution; consistent with the *in vivo* assay results.

6.4.2.2 Chemical cross-linking of scFvs

As chemical cross-linking of the full length IgGs of WFL and STT resulted in identification of the site responsible for self-association, the technique was once again employed to characterise the self-association of the scFv proteins.

The three scFvs (WFL, SFL and STT) were dialysed into 100 mM sodium phosphate buffer, pH 7.4. As cross-linking of the IgGs was carried out at 6.74 μM and there are two scFvs per full IgG, the scFv samples were diluted to a concentration of 15 μM for the chemical cross-linking experiments (see **Section 3.4.5**). WFL_scFv, SFL_scFv and STT_scFv were cross-linked with increasing amounts of the 50:50 mixture of the non-deuterated (d0) and deuterated (d4) BS3 cross-linker and the reaction performed on ice for 30 min. The reactions were quenched with Tris.HCl, pH 8 at a final concentration of 50 mM and the results analysed via SDS-PAGE (see **Section 3.4.8**) (**Figure 6.16**)

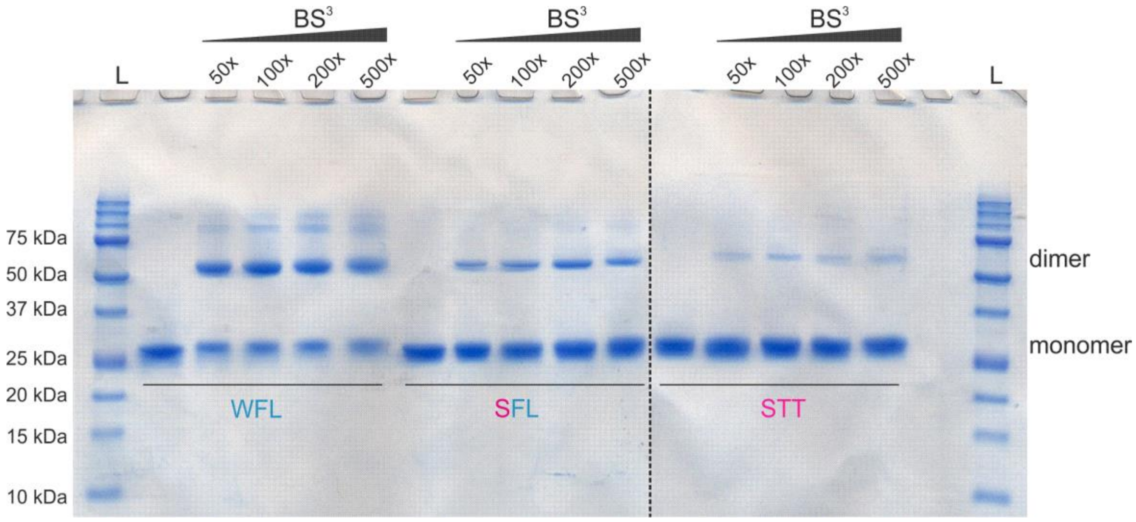


Figure 6.16: SDS-PAGE gel showing scFvs cross-link in a similar manner as full length IgGs across increasing cross-linker concentrations. L = Protein marker and dotted line = merging point of two separate gels. Coomassie staining. Representative gel of n = 3.

As observed for the full IgGs (See **Figure 5.6**), WFL_scFv forms oligomeric species, when cross-linked with BS3, while STT_scFv only forms a small amount of dimer across all cross-linker concentrations tested. Consistent with the PEG precipitation results, the data here are once again in accord with the *in vivo* assay results, where the extent of SFL_scFv self-association is intermediate between that of WFL and STT in that a significantly larger amount of dimer is present, compared with STT_scFv. The monomer and dimer bands, at 50× and 200× cross-linker, were excised from the gel and an in-gel digestion of the samples carried out (see **Section 3.4.9**).

The digested cross-linked material was purified and injected into the Waters G2Si mass spectrometer for LC-MS/MS analysis. As low abundance peptides were once again being searched for, the threshold for peptide to be selected for sequencing was lowered and the number of peptides sequenced in any one scan was increased to six. Comparative analysis of the resulting mass spectra revealed that the same cross-link which was found in the IgG data (**Figure 6.17-6.18**) was identified in WFL_scFv. Furthermore, a new cross-link between the N-termini of the two heavy chains was identified in the WFL and SFL samples (**Figure 6.19-6.20**).

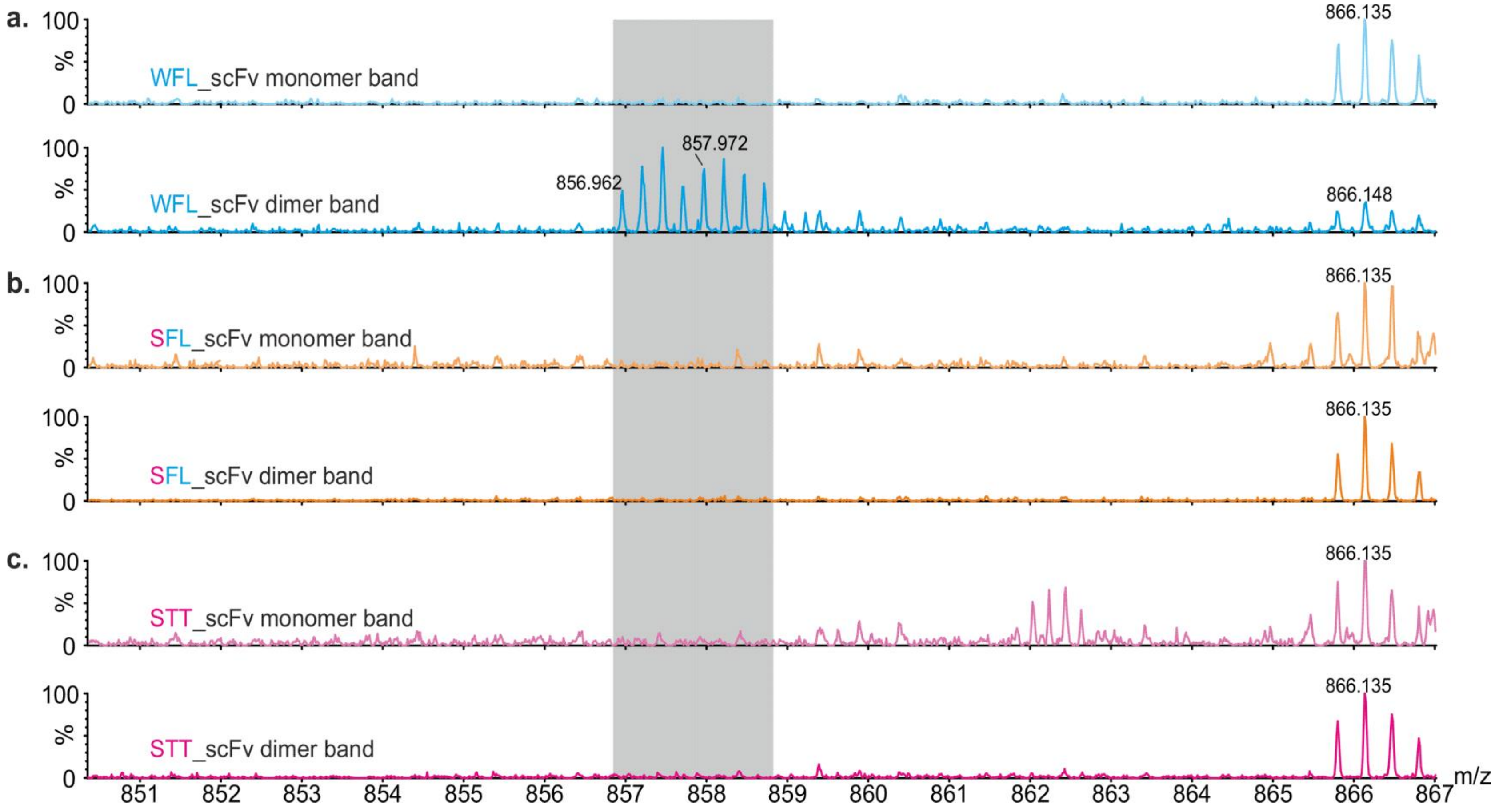


Figure 6.17: Mass spectra showing the identification of cross-link peptides unique to the WFL_scFv dimer band. **(a)** Mass spectra of the WFL_scFv monomer band (top) and dimer band (bottom) at the same retention time showing a unique cross-linked peptide. **(b)** Mass spectra of the SFL_scFv monomer band (top) and dimer band (bottom) at the same retention time showing no cross-linked peptide present. **(c)** Mass spectra of the STT_scFv monomer band (top) and dimer band (bottom) at the same retention time showing no cross-linked peptide present. Shaded area = cross-linked peptide. Cross-linked peptide = 3423.85 Da, 4+ charge state.

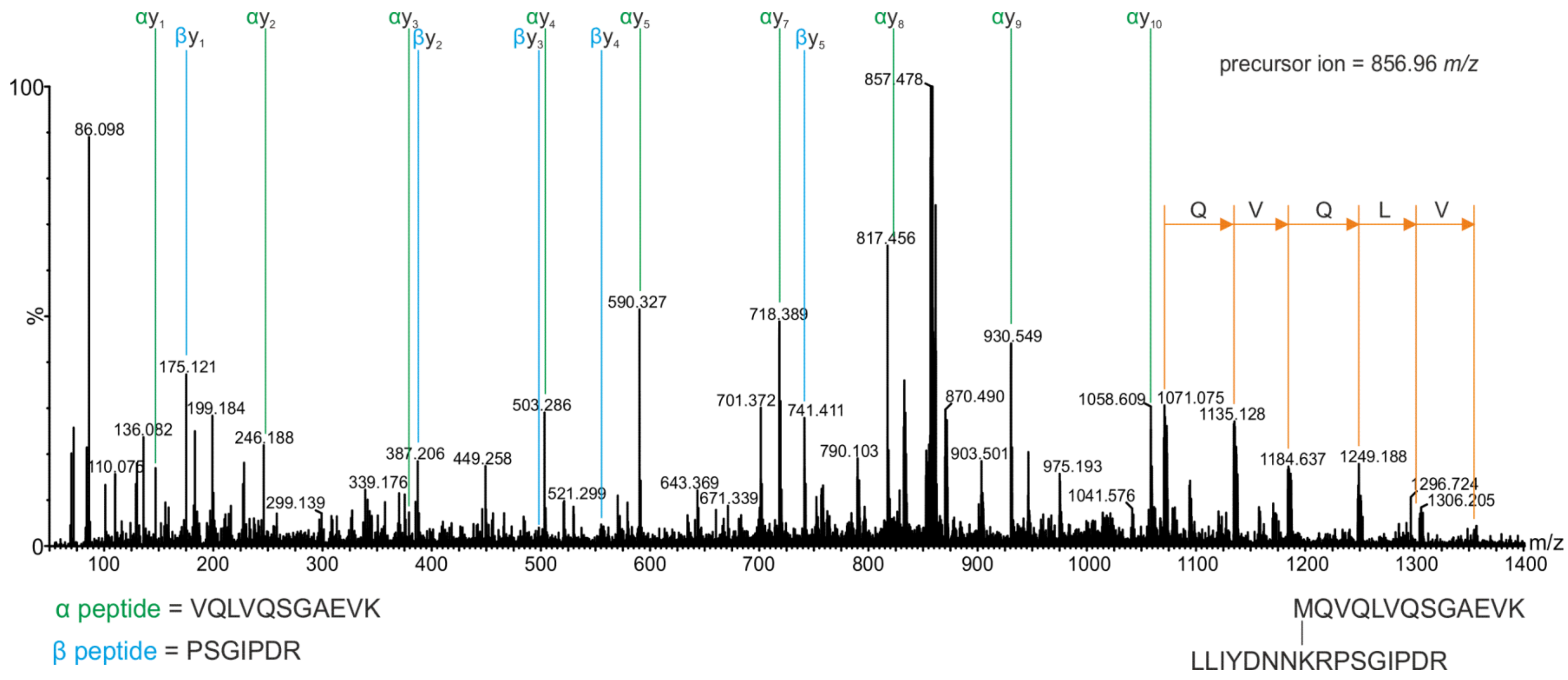


Figure 6.18: CID induced MS/MS spectrum of the unique cross-linked peptide identified in WFL_scFv. The y-ion series for both peptides (α, green and β, blue) are shown with the corresponding peptide sequence listed below. Doubly charged species with cross-linker still present are also highlighted (orange) showing the addition of sequential amino acids. Precursor ion = 856.96 *m/z*.

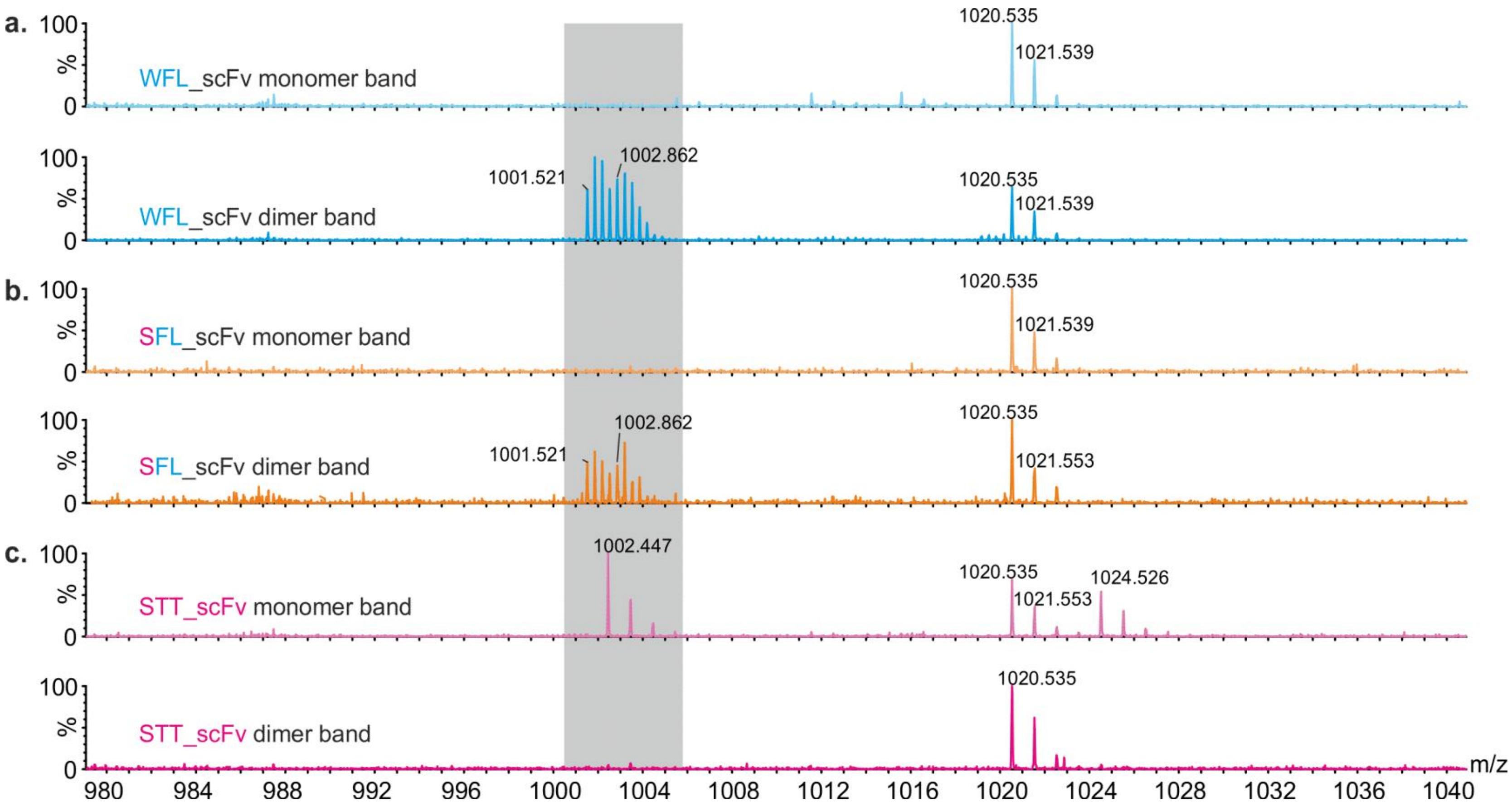


Figure 6.19: Mass spectra showing the identification of cross-link peptides in the WFL and SFL dimer bands. **(a)** Mass spectra of the WFL_scFv monomer band (top) and dimer band (bottom) at the same retention time showing a unique cross-linked peptide. **(b)** Mass spectra of the SFL_scFv monomer band (top) and dimer band (bottom) at the same retention time showing the same cross-linked peptide, as found in WFL_scFv. **(c)** Mass spectra of the STT_scFv monomer band (top) and dimer band (bottom) at the same retention time showing no cross-linked peptide present. Shaded area = cross-linked peptide. Cross-linked peptide = 3001.56 Da, 3+ charge state.

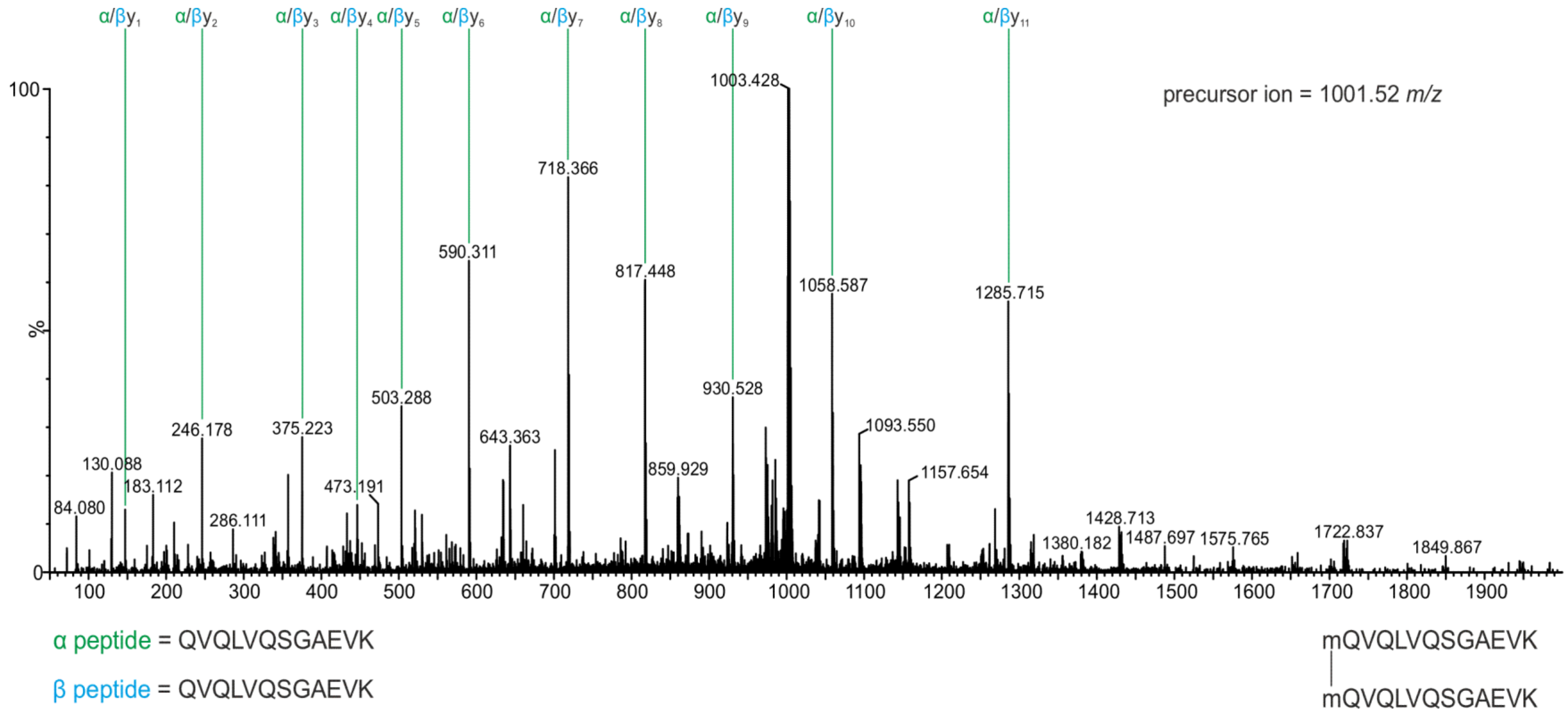


Figure 6.20: CID induced MS/MS spectrum of the unique cross-linked peptide identified in WFL_scFv and SFL_scFv (spectrum from WFL_scFv sample). The y-ion series for both peptides (α, green and β, blue) are shown with the corresponding peptide sequence listed below. Precursor ion = 1001.52 *m/z*. m = oxidised methionine.

The results from the chemical cross-linking experiment show that the same cross-linked peptide that was identified in the full IgGs was also identified in the scFv construct of WFL (**Figure 6.17-6.18**). This peptide was unique to WFL_scFv and indicates that the mechanism of self-association is the same irrespective of the format: either IgG or scFv. However, it is clear from the cross-linking gel (**Figure 6.16**) that SFL (W30S) forms a dimeric species that should be observed in the cross-linking analysis. A separate set of cross-linked peptides was identified for both WFL and SFL (W30S) that was only present in the spectra corresponding to the digested dimer bands and was also absent in either the STT monomer or dimer spectra (**Figure 6.19-6.20**).

An interesting question that this raised was why was this cross-link not identified in the full IgG data? While it could be possible that the mechanism of self-association had changed, which would explain why it was not observed, it was hypothesized that it was more likely due to the differences in the amino acid sequences of the two constructs. As these proteins had been expressed in *E. coli*, there was an additional methionine residue on the N-terminal of the scFv and therefore this cross-link may have been identified as the two N-termini were brought into closer proximity. However, as mentioned previously, the N-terminal glutamine residue in the full IgG undergoes conversion to a pyroglutamate residue (see **Section 5.5**); which is not possible in the scFvs as the N-terminal glutamine has been replaced with methionine. Therefore, another possible explanation as to why this cross-link was not identified in the IgGs is because it requires two Fab regions to dimerise that both have the non-pyroglutamate at position one, which has a low probabilistic chance.

In summary, while the unique cross-link identified for WFL_scFv was in agreement with the full length IgG data and suggests that the mechanism of aggregation of the IgG and the scFv is the same, it was essential to understand whether the second cross-link, between the two N-termini, agrees with the previous model generated in Chapter 5 or whether a new interface was being identified.

6.4.3 Modelling the self-association of scFvs

To understand the mechanism of self-association between the scFvs, the same docking procedure (see **Section 3.3.5** and **Section 5.6**) was utilised.

Residues W30, F31 and L57 were used to drive the docking procedure while setting 11.4 Å distance restraints between the same cross-linked atoms as previously established from the IgG data i.e. the N-terminal glutamine of the heavy chain to lysine 54 (K54) of the light chain, as well as the newly identified cross-link; N-terminal methionine to N-terminal methionine. The structure with the lowest energy across the entire molecule is shown in **Figure 6.21.**

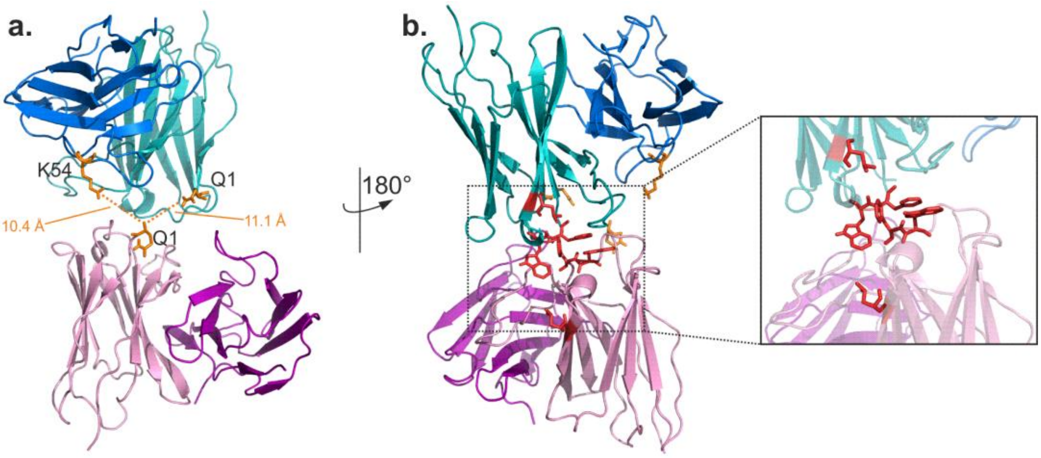


Figure 6.21: Structure of the WFL_scFv docked dimer. **(a)** Front view showing that all identified cross-linkable sites are within 11.4 Å and **(b)** back view showing the dimer interface. Inset = residues 30, 31 and 57 used to drive the docking procedure. Residues W30, F31 and L57 are shown as red sticks.

The modelled structure showed that it is possible for either cross-link to exist while maintaining a hydrophobic dimer interface. This structure has altered from the original model with the two leucine residues involved in the dimer interface (**Figure 6.21b**, red sticks) being located further away from the interaction site. However, this agrees with the *in vivo* data where the mutation of this leucine to threonine (L57 > T57) does not have any effect on *E. coli* cell growth (**Figure 6.5**).

One question that needed to be addressed was why were higher order oligomers (up to and including tetramer) observed for the scFv, when the previous model based on full IgGs would suggest only dimers would be possible? While there was a second available binding site on the IgG's Fab arm for oligomer extension, this model suggests that it would be difficult for higher order oligomers to form from scFvs. To investigate this, the hydrophobic surface of the dimer model was analysed to interrogate if there were any hydrophobic patches where other scFvs may potentially bind (**Figure 6.22**).

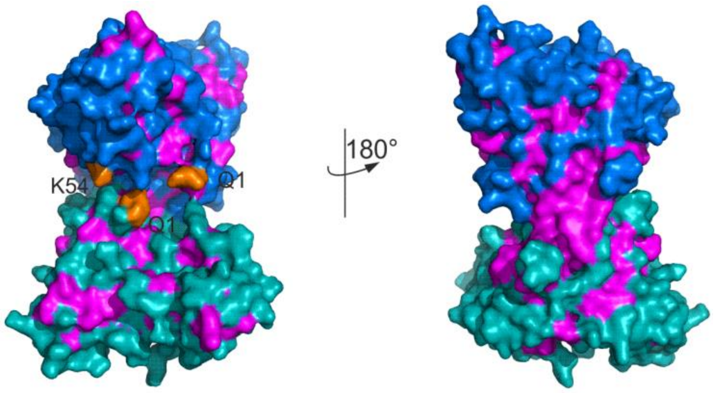


Figure 6.22: Surface representation of the docked dimer structure. The dimer interactions exposes a large hydrophobic region (pink) on the side of the docked structure. Labelled residues (orange) represent identified cross-linked residues. The two WFL monomers are represented as the teal and blue surfaces.

The interaction that leads to dimer formation results in a large hydrophobic patch being exposed near the dimer interface. This yields insight into how larger oligomeric species may arise and result in the loss of functional β-lactamase in the *in vivo* assay.

6.5 Comparing the aggregation of scFvs and IgGs

Finally, while the mechanism of aggregation between scFv and IgG has been shown to be the same across the two formats, it was of interest to understand how the *in vivo* data compared with aggregation of the mutated constructs as IgGs *in vitro*. For this, the results obtained from the β-lactamase *in vivo* assay were compared with retention time data from size-exclusion chromatography (SEC), acquired by MedImmune plc, for the eight mutants as IgGs (**Figure 6.23**).

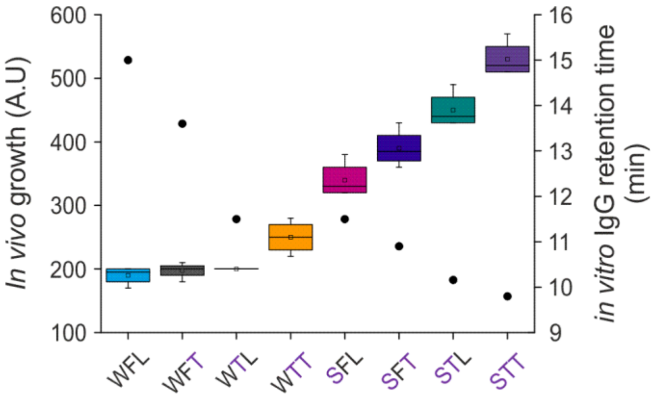


Figure 6.23: Comparison of *in vivo* scFv data and *in vitro* IgG data. The *in vivo* growth data = box plots, IgG retention times = circles.

A remarkable correlation can immediately be seen between the *in vivo* and *in vitro* data. Therefore, as well as the mechanism of self-association being conserved across scFv and IgG constructs analysed here, the data highlights that the β-lactamase *in vivo* assay can identify exactly the same aggregation prone sequences identified for the full IgGs *in vitro*.

6.6 Discussion

Early identification of aggregation-prone hotspots on the surface of proteins is required either in order to engineer these out in the context of potential therapeutics (Ratanji et al., 2014, Agrawal et al., 2011, Kayser et al., 2012), or to develop inhibitors of aggregation in the case of disease-relevant proteins (Weiss et al., 2008, Dickey et al., 2005, Ross and Poirier, 2004, David and Tayebi, 2014). This identification can often be difficult due to aggregation-prone sequences being buried or partially buried in a proteins structure and hence local or global unfolding may be required for aggregation to ensue. Furthermore, in terms of biotherapeutic development, significant economic cost may be spent during the development of a project before the aggregation propensity of a system is recognised, and any aggregates that do form may evoke an immunogenic response (Zurdo et al., 2011, Zurdo, 2013, Obrezanova et al., 2015). Therefore, there is a requirement for assays and screens that are able to identify aggregation prone sequences/regions as early as possible in the production pipeline.

Rational design has been favoured historically to remove unwanted self-association (Courtois et al., 2016, Camilloni et al., 2016, Geoghegan et al., 2016). However, introducing a stretch of mutations, or even a single point mutation, can have significant detrimental effects on a protein's stability, as well as having potential adverse effects on its ability to fold. The β-lactamase assay described in this Chapter was focused on addressing this question at the scFv level. As this is an *in vivo*, cell survivability assay, accurate protein folding as well as the ability to remain soluble is required for full β-lactamase function. Therefore, any adverse effects brought on by mutation would potentially be identifiable.

As the β-lactamase assay had been successfully applied previously to both amyloidogenic proteins as well as single-domain antibodies (dAbs) (Saunders et al., 2016), it was hypothesized that the assay would also readily identify aggregation-prone sequences in scFvs and hence could be used as a screen for proteins with robust bioprocess capability. However, concerns were raised about relating the aggregation of scFvs to the aggregation of full length IgGs, as they lack the stability of IgGs and hydrophobic hydrophobic surfaces between the two variable regions can be exposed, often leading to engineering being required to form stable scFv fragments (Glockshuber et al., 1990, Worn and Pluckthun, 2001, Miller et al., 2010). However, the results presented in this Chapter from the *in vivo* β-lactamase system beautifully follow the aggregation properties observed from the PEG precipitation and cross-linking results *in vitro*; as well correlating incredibly well with the full IgG data obtained by MedImmune plc. Finally, investigation of the self-association via chemical cross-linking analysis confirms that the mechanism of aggregation is conserved across the two formats, IgG and scFv, with only small changes being made to the original model upon the identification of an additional cross-link site.

The data presented here open up the exciting avenue for the β-lactamase assay to be used to pan for aggregation prone sequences early during biotherapeutic development. As candidate sequences are often screened against the target antigen using antibody fragments, such as scFvs or Fabs, the aggregation propensity of lead sequences could be evaluated before converting to full IgG format (**Figure 6.24**).

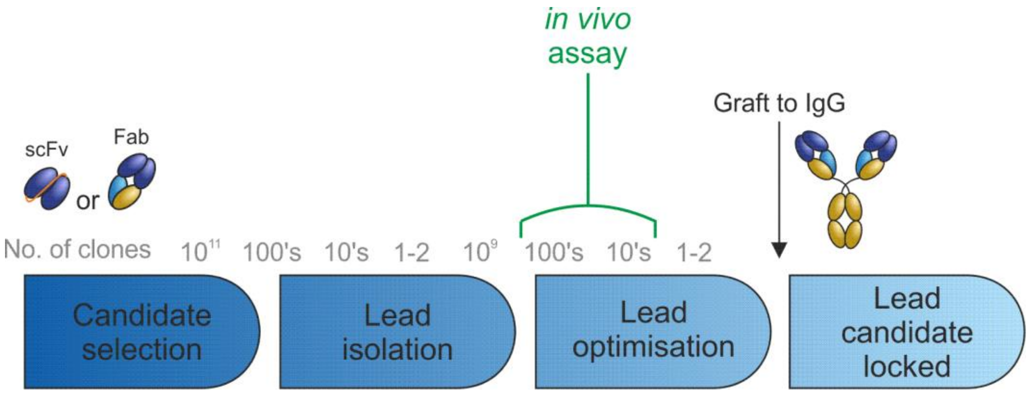


Figure 6.24: Schematic of therapeutic candidate selection, isolation and optimisation. The point in the developmental process at which the β-lactamase *in vivo* assay could be incorporated is shown.

Whereas screening IgGs for self-association requires *in vitro* analysis, typically SEC post expression and purification (Hong et al., 2012, Paul and Schwab, 2014, Steve and Chakrabarti, 2013), the β-lactamase *in vivo* screen allows for the same characterisation in a two-day assay that can be performed prior to IgG expression and purification. In summary, whilst reducing the need to screen for self-association and aggregation at the IgG level, the β-lactamase assay developed here allows for rapid identification of aggregation-prone sequences that can be removed from the industrial pipeline before creation of IgGs for further development.

Successful integration and accurate characterisation of scFvs by the β-lactamase assay could be used in other ways to facilitate biopharmaceutical development. Candidate sequences could be evolved and improved through random mutagenesis and screened at increasing ampicillin concentrations in order to select the least aggregation prone sequences. The model system presented in this work, WFL and STT, would be ideal for such a study. By performing random mutagenesis on the WFL sequence and carrying out an *in vivo* assay at antibiotic concentrations where the STT-β-lactamase construct permits cell growth but WFL does not, aggregation hotspots could be evolved out of the sequence. Ultimately, this would save time by removing the requirement for structural characterisation and analysis in order to determine the residues responsible for self-association and their removal through rational design. However, while selecting for monomer stability via random mutagenesis, there is potential for affinity to be lost by mutations affecting the binding site for the antigen. Therefore, careful sequencing analysis would be required in order to retain affinity while driving the propensity of a sequence to self-associate to a minimum.

To conclude, the *in vivo* β-lactamase system applied here was able to successfully disentangle aggregation prone mutants in therapeutically relevant scFvs sequences. This opens an exciting new avenue for the *in vivo* assay to be used in assessing the aggregation propensity of candidate sequences upstream in the developmental pipeline; thus eliminating the requirement for large scale IgG expression and purification of multiple variants.

Chapter 7:

Concluding remarks and future directions

7 Concluding remarks and future directions

7.1 Successes and limitations

Uncontrolled protein self-association and aggregation has implications in more than 50 known human diseases (Sipe et al., 2012), including Alzheimer's, Parkinson's and type II diabetes; with an ageing population in today's world, this has enormous social and economic burdens (Boyle et al., 2010). Furthermore, the underlying potential for any protein to self-associate is a major issue that negatively impacts the biopharmaceutical sector and its ability to develop new products to target disease (Frokjaer and Otzen, 2005, Gómez de la Cuesta et al., 2014, Zurdo et al., 2011). The capacity for oligomeric species to form various structures and morphologies, which can be large or small and soluble in nature (Fink, 1998), make their identification and furthermore, their characterisation, challenging. Hence, to alleviate even a fraction of these burdens, there is a critical need for the development of techniques that can not only characterise oligomeric species upon their formation, but to identify these species as early as possible in the design and development of proteins as biologics.

In the search for techniques that can identify and characterise oligomeric forms of a candidate protein, the demand is for a technique that has high sensitivity, requires low sample volumes and concentrations and finally, can shed light on the structure or morphology of any oligomeric species present in a complex mixture of species. Native mass spectrometry is equipped to answer all of these issues. With incredible sensitivity, mass spectrometry can detect small amounts of oligomeric material in heterogeneous mixtures, while requiring only microlitre volumes at micromolar concentrations. Additionally, with the coupling of IMS and MS, insights into the shape and conformation of each individual species can be established (Pinkse et al., 2003, Uetrecht et al., 2011b, Illes-toth et al., 2015), revealing insights into the mode of self-association. These approaches been applied successfully in various studies monitoring the assembly of protein complexes (Uetrecht et al., 2010, Zhong et al., 2012, Lanucara et al., 2014). While a logical approach to prevent aggregation is to target the protein at the monomer level, by either stabilising the monomeric form or eliminating aggregation-prone sequences, this is often not possible for multiple reasons. The first of these is the difficulty in identifying aggregation-prone sequences in a candidate protein. Prediction algorithms have been developed with some success (Wang et al., 2009, Chennamsetty et al., 2009a, Chennamsetty et al., 2009b). However, these algorithms are often not robust enough to rule out aggregation due to the complex nature of aberrant protein self-assembly. Orthogonal techniques are thus required to validate the predictions. Secondly, in order to predict the regions responsible for aggregation, information about the protein's structure in the aggregation-prone state is required. This is not always possible, especially when working with intrinsically disordered proteins or where the aggregation-prone species is rare or in dynamic equilibrium with multiple species

In the attempt to aid the biopharmaceutical industry and product development, the work in this thesis set out to identify and characterise a therapeutic mAb (WFL), with a strong tendency to self-associate, which was raised against nerve growth factor. A combination of mass spectrometric techniques were used to identify of the oligomeric species formed, and to determine the underlying mechanism self-association of WFL. Finally, an *in vivo* screen was developed and used to identify aggregation-prone sequences. In Chapter 4, native MS and IMS-MS were employed to visualise the oligomeric species of WFL, using their estimated CCSs to postulate models of its potential mode of self-association. Experimentally derived CCS values are commonly used to validate theoretical models, by comparison of measured and predicted CCS values and their agreement with one another. Strikingly, upon comparing the monomeric CCS of WFL to a predicted value from the crystal structure (1IGY), a 32 % decrease from the predicted value was observed. Taking the accepted error (3-5 %) of the IMS-MS measurements (Ruotolo et al., 2008, Smith et al., 2009), the measured value was clearly inconsistent with the crystal structure. However, this value was in agreement with previously published data of the CCSs of other IgGs (Debaene et al., 2014, Pacholarz et al., 2014), suggesting that the structure compacts, or collapses, upon transition to the gas phase. To explore this observation, *in vacuo* molecular dynamics simulations were carried out to visualise this compaction. At a similar time, the Barran group published their findings on other mAbs also showing this gas-phase collapse of IgGs (Pacholarz et al., 2014), and concluded that the hinge region appears to be the major player in this collapse. As there are relatively few studies reporting the gas-phase collapse of proteins, the gas-phase behaviour of three other proteins, two which share flexible hinge-like regions, were examined.

So how are proteins behaving in the gas-phase? The results presented in this thesis reveal that a number of proteins do collapse in the gas-phase, whether there are flexible linker regions between the domains or not, as exemplified in the case of SasG. *Yet what drives this collapse?* The structural re-arrangement of proteins in the gas-phase has been discussed, with a mention of side chain collapse within a picosecond timescale upon transition into the gas-phase (Counterman and Clemmer, 1999, Breuker and McLafferty, 2008). So perhaps the extent of collapse observed for IgGs and SasG is an extrapolation of these same effects. While it is unclear which forces that are causing this collapse in the gas-phase, the absence of water leads to the loss of hydrophobic interactions and an increase in electrostatic interactions. Therefore, it is possible that there are weak underlying interactions that become dominant in the gas-phase, leading to the observed collapse. Furthermore, the proteins selected for study in Chapter 4 were chosen based on their unusual structures: the majority of IMS-MS and CCS work in the literature has been carried out globular proteins, for which the majority of these globular proteins yield CCS values in agreement with the predicted values. *So why is collapse observed for only a small set of proteins*? Perhaps all proteins collapse in the gas-phase to some extent, as shown for the side chain collapse of the globular protein cytochrome c (Steinberg et al., 2008). However, with globular proteins already being compact structures, there is little room for further collapse of the tertiary and quaternary structure. By contrast, elongated structures with flexible hinge regions have the additional space for collapse to occur, as highlighted with the work this thesis. It is clear that the structure of proteins and their behaviour in the gas-phase has become an area of interest within the field in recent years (Ruotolo et al., 2005, Hogan et al., 2011, Mehmood et al., 2014, Pacholarz et al., 2014). The discovery that many proteins may collapse to some extent, whether that be a local or global collapse, leads to the particular interest in the pursuit of measuring the CCSs of various other elongated structures, in an attempt to obtain a greater understanding of the conformations that proteins can adopt upon the transition from solution to the gas-phase.

The discovery that mAbs collapse in the gas-phase ruled out the use of IMS-MS to identify the mode of self-association of WFL. Therefore, chemical cross-linking was employed in Chapter 5 as an alternative technique to identify the mechanism of WFL's aggregation. By cross-linking the WFL oligomers and identification of the interacting regions using LC-MS/MS, a model explaining the mechanism of self-association was generated. As mentioned in Chapter 1, the use of structural technique prior to MS, such as FPOP (Yan et al., 2014b), HDX (both solution and gas-phase (Pirrone et al., 2015, Beeston et al., 2015)), protein painting (Luchini et al., 2014), chemical labelling (Borotto et al., 2015) and chemical cross-linking (Chen et al., 2010) can be used in the search for high resolution structural data, especially when techniques such as NMR and X-ray crystallography are not possible. For general protein analysis as well as protein therapeutics, it is essential to have techniques at the researcher's disposal to enable identification of self-association interfaces as well as the mechanisms that drive aggregation. With a greater understanding of the mechanisms and forces that cause uncontrolled self-association, therapeutic candidates will be able to be developed with enhanced protection against aggregation, ultimately aiding their potential to enter clinical trials and hopefully the therapeutic market, where these candidate drugs are urgently needed.

The ability to characterise protein aggregates in high resolution is of primary importance once they are detected. One of the most logical approaches would be to establish a technique which can identify aggregation-prone sequences before they are selected for protein expression and production. Chapter 6 set out to address this by the development of an *in vivo* β-lactamase tripartite system that has previously been used to identify inhibitors of the aggregation of amyloid systems (Saunders et al., 2016). The ability of the assay to identify aggregation prone sequences in scFvs was examined, with the *in vivo* results obtained correlating remarkably with previous *in vitro* data obtained on the full length IgG variants. Prior to expression and purification of the intact, evolved IgG, the aggregation propensity of the WFL sequence previously remained hidden. Furthermore, there is still unexplored potential for the *in vivo* β-lactamase system and its use within the pharmaceutical sector. The assay could be used for the directed evolution of candidate sequences and the identification or removal of aggregation-prone hot-spots. Thus, the results presented in Chapter 6 have enormous implications in industry for screening for soluble biopharmaceutical sequences. This would impact both the time and cost that are currently invested into minimising aggregation-prone sequences for the development of promising biologics.

7.2 The protein aggregation predicament

As mentioned previously, protein aggregation is responsible for a plethora of life-threatening diseases, many of which are closely correlated with old age. With our life expectancy increasing year by year (Reynaud, 2010), protein aggregation is set to become an increasing problem. This therefore begs the question:

How can unwanted and uncontrolled protein self-association and aggregation be circumvented?

This intellectual question plagues all protein scientists. From a PhD project level, the aim will be to understand the mechanisms of aggregation, all the way through to protein therapeutic development and manufacturing, targeting the diseases that these aggregation events cause. The great challenge in controlling aggregation lies ultimately in obtaining a better understanding of the mechanisms that underpin aberrant protein self-assembly. With such knowledge, therapies targeting aggregation based protein diseases could be developed, as well as the design of aggregation resistant biotherapeutics, aiding the path to market.

The work detailed within this thesis presents the mechanism in which a therapeutic mAb, targeting NGF, self-associates. *However, is this enough to understand protein aggregation as a whole?* No, probably not. Nevertheless, to aid in the identification of aggregation-prone sequences and the development of future algorithms which can accurately identify such sequences, information on self-associating and aggregation-prone systems, like WFL, is required. Without it, the tremendous leaps forward that current aggregation prediction software have made in the past two decades would not have been possible (Buck et al., 2012). Furthermore, what can be achieved in the struggle against protein aggregation is the development of techniques for the rapid identification of aggregation-prone sequences, as well as techniques which can characterise them. The *in vivo* β-lactamase assay described herein may aid in this quest, whether that is through identifying aggregation-prone sequences that can be studied to understand the protein aggregation problem, or through its ability to aid in the development of next generation biopharmaceuticals. Overall therefore, MS, and MS based methods, are sure to see an increased use in the development of new understandings of protein aggregation and how to avoid aggregation in the development of new proteins as biologics.

8 Appendices

8.1 DNA protein sequences of β-lactamase (βla) fusion proteins used in the in vivo assay and their plasmid maps

8.1.1 βla-WFL

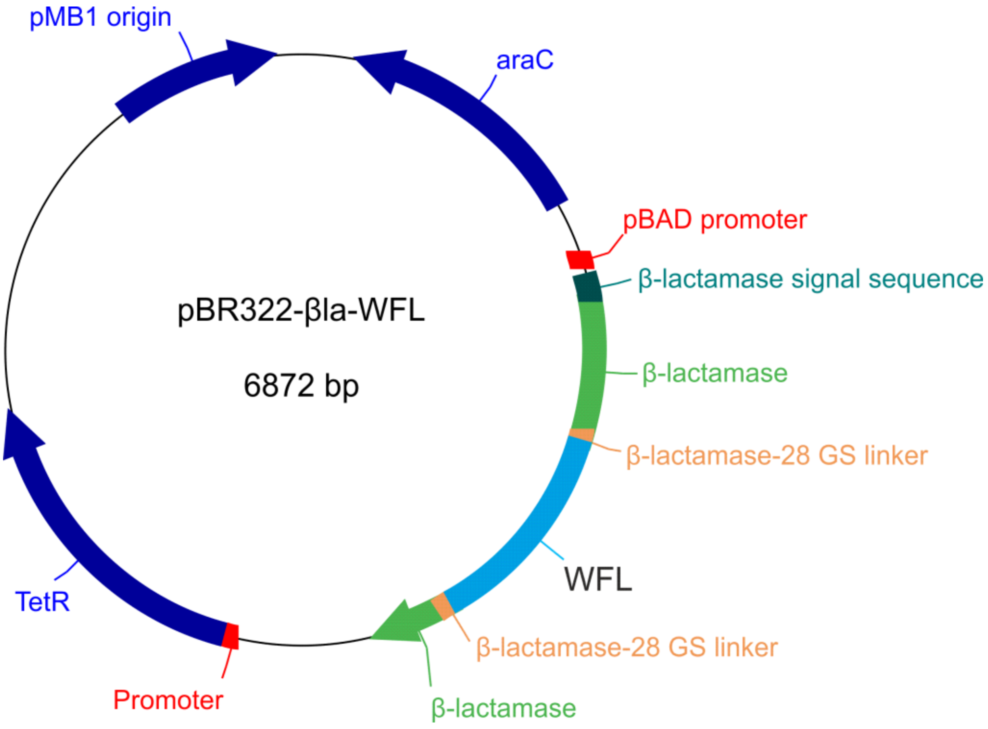
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT
GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC
CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG
CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTA**GG TGGTGGTGGT**
TCTGGTGGTG GTGGCTCGAG CCAGGTTCAG CTTGTGCAGA GCGGTGCGGA GGTCAAAAAA
CCCGGCAGCT CTGTAAAAGT TAGCTGCAAA GCGAGTGGCG GTACGTTTTG GTTTGGGGCC
TTTACTTGGG TTCGTCAAGC GCCGGGCCAG GGCTTGGAAT GGATGGGTGG CATTATCCCT
ATTTTCGGCC TCACAAACCT GGCGCAAAAC TTTCAAGGTC GCGTTACCAT TACGGCGGAC
GAAAGCACCA GTACCGTCTA TATGGAGCTG TCAAGCCTGC GCTCAGAAGA CACCGCAGTT
TACTACTGTG CGCGTAGCAG CCGCATTTAC GACTTGAATC CTAGCCTCAC AGCGTACTAC
GACATGGATG TGTGGGGGCA GGGCACCATG GTTACGGTGT CGAGTGGTGG TGGGAGCAGT
GGTGGAGGTG GGTCCGGGGG CGGCGGCGGC GCGCAAAGCG TATTAACTCA GCCGCCGAGC
GTGAGCGCAG CCCCTGGGCA GAAAGTCACC ATTTCATGCA GCGGCTCCTC CAGCGATATC
GGCAACAATT ACGTGTCCTG GTATCAGCAG CTGCCTGGCA CTGCGCCGAA GCTGTTGATT
TATGACAACA ATAAGCGTCC CTCGGGTATT CCAGATCGTT TTTCTGGCTC TAAAAGCGGG
ACATCAGCGA CACTGGGCAT CACCGGGCTG CAGACGGGGG ATGAAGCCGA TTATTACTGC
GGGACCTGGG ATAGTTCCCT GAGCGCGTGG GTGTTTGGCG GGGGCACCAA ACTCACCGTG
CTG**GGATCCG GGAGCGGTTC CGGAAGCGGA GGAGGTGGTT CAGGCGGAGG TGGAAGC**TTG
ACTCTAGCTA GCCGGCAGCA GCTCATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA
CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG
ATAGGTGCCT CACTGATTAA GCATTGGTAA

Appendix 8.1: DNA sequence of the β-lactamase WFL fusion construct. The periplasmic signal sequence is in purple. The WFL sequence is shown in blue with the XhoI and BamHI restriction sites shown in orange and red, respectively. The glycine-serine linker is highlighted in bold. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

MSIQHFRVALIPFFAAFCLPVFA – signal sequence

HPETLVKVKD AEDQLGARVG YIELDLNSGK ILESFRPEER FPMMSTFKVL
LCGAVLSRVD AGQEQLGRRI HYSQNDLVEY SPVTEKHLTD GMTVRELCSA
AITMSDNTAA NLLLTTIGGP KELTAFLHNM GDHVTRLDRW EPELNEAIPN
DERDTTMPAA MATTLRKLLT GEL**GGGGSGG** **GGSS**QVQLVQ SGAEVKKPGS
SVKVSCKASG GTFWFGAFTW VRQAPGQGLE WMGGIIPIFG LTNLAQNFQG
RVTITADEST STVYMELSSL RSEDTAVYYC ARSSRIYDLN PSLTAYYDMD
VWGQGTMVTV SSGGGSSGGG GSGGGGGAQS VLTQPPSVSA APGQKVTISC
SGSSSDIGNN YVSWYQQLPG TAPKLLIYDN NKRPSGIPDR FSGSKSGTSA
TLGITGLQTG DEADYYCGTW DSSLSAWVFG GGTKLTVL**GS** **GSGSGSGGGG**
SGGGGSLTLA SRQQLIDWME ADKVAGPLLR SALPAGWFIA DKSGAGERGS
RGIIAALGPD GKPSRIVVIY TTGSQATMDE RNRQIAEIGA SLIKHW

Appendix 8.2: Protein sequence of the β-lactamase WFL fusion construct. The glycine-serine linker is highlighted in bold. Residues 30, 31 and 57 are highlighted in yellow. The heavy and light chains are shown in blue and purple, respectively.



Appendix 8.3: Plasmid map of pBR322-βla-WFL. Generated and kindly provided by Dr Janet Saunders (University of Leeds)

8.1.2 βla-SFL

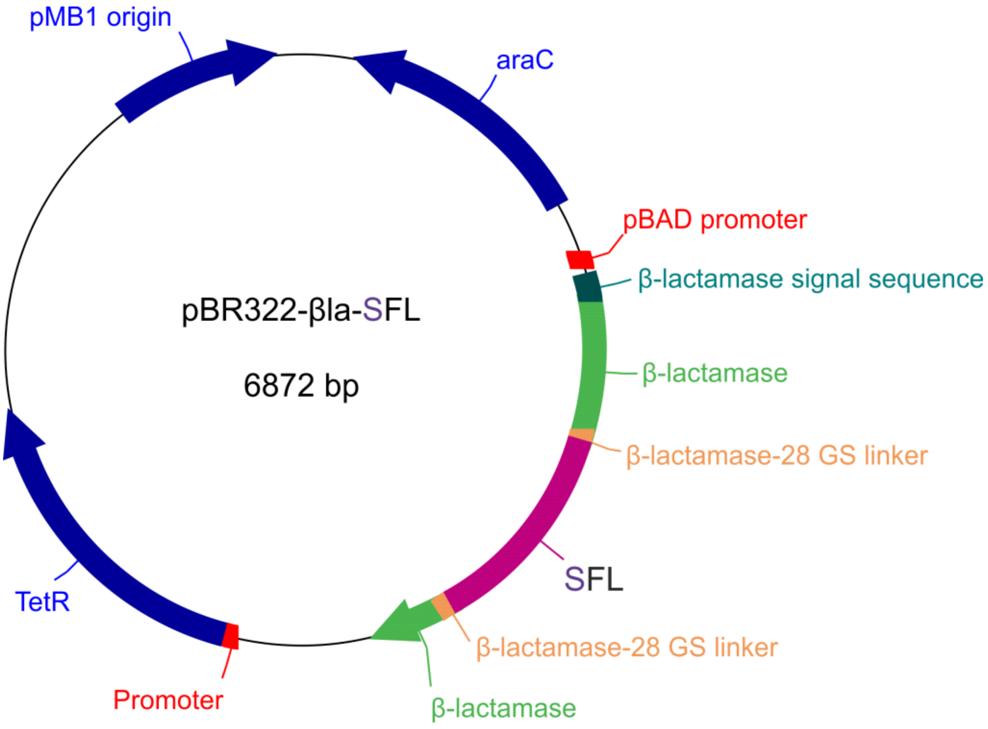
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT
GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC
CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG
CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTA**GG TGGTGGTGGT**
TCTGGTGGTG GTGGCTCGAG CCAGGTTCAG CTTGTGCAGA GCGGTGCGGA GGTCAAAAAA
CCCGGCAGCT CTGTAAAAGT TAGCTGCAAA GCGAGTGGCG GTACGTTTTC CTTTGGGGCC
TTTACTTGGG TTCGTCAAGC GCCGGGCCAG GGCTTGGAAT GGATGGGTGG CATTATCCCT
ATTTTCGGCC TCACAAACCT GGCGCAAAAC TTTCAAGGTC GCGTTACCAT TACGGCGGAC
GAAAGCACCA GTACCGTCTA TATGGAGCTG TCAAGCCTGC GCTCAGAAGA CACCGCAGTT
TACTACTGTG CGCGTAGCAG CCGCATTTAC GACTTGAATC CTAGCCTCAC AGCGTACTAC
GACATGGATG TGTGGGGGCA GGGCACCATG GTTACGGTGT CGAGTGGTGG TGGGAGCAGT
GGTGGAGGTG GGTCCGGGGG CGGCGGCGGC GCGCAAAGCG TATTAACTCA GCCGCCGAGC
GTGAGCGCAG CCCCTGGGCA GAAAGTCACC ATTTCATGCA GCGGCTCCTC CAGCGATATC
GGCAACAATT ACGTGTCCTG GTATCAGCAG CTGCCTGGCA CTGCGCCGAA GCTGTTGATT
TATGACAACA ATAAGCGTCC CTCGGGTATT CCAGATCGTT TTTCTGGCTC TAAAAGCGGG
ACATCAGCGA CACTGGGCAT CACCGGGCTG CAGACGGGGG ATGAAGCCGA TTATTACTGC
GGGACCTGGG ATAGTTCCCT GAGCGCGTGG GTGTTTGGCG GGGGCACCAA ACTCACCGTG
CTG**GGATCCG GGAGCGGTTC CGGAAGCGGA GGAGGTGGTT CAGGCGGAGG TGGAAGC**TTG
ACTCTAGCTA GCCGGCAGCA GCTCATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA
CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGCCAGATGG GTAAGCCCTC CCGTATCGTA
GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG
ATAGGTGCCT CACTGATTAA GCATTGGTAA

Appendix 8.4: DNA sequence of the β-lactamase SFL fusion construct. The periplasmic signal sequence is in purple. The WFL sequence is shown in blue with the XhoI and BamHI restriction sites shown in orange and red, respectively. The glycine-serine linker is highlighted in bold. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

MSIQHFRVALIPFFAAFCLPVFA – signal sequence

HPETLVKVKD AEDQLGARVG YIELDLNSGK ILESFRPEER FPMMSTFKVL
LCGAVLSRVD AGQEQLGRRI HYSQNDLVEY SPVTEKHLTD GMTVRELCSA
AITMSDNTAA NLLLTTIGGP KELTAFLHNM GDHVTRLDRW EPELNEAIPN
DERDTTMPAA MATTLRKLLT GEL**GGGGSGG** **GGSS**QVQLVQ SGAEVKKPGS
SVKVSCKASG GTFSFGAFTW VRQAPGQGLE WMGGIIPIFG LTNLAQNFQG
RVTITADEST STVYMELSSL RSEDTAVYYC ARSSRIYDLN PSLTAYYDMD
VWGQGTMVTV SSGGGSSGGG GSGGGGGAQS VLTQPPSVSA APGQKVTISC
SGSSSDIGNN YVSWYQQLPG TAPKLLIYDN NKRPSGIPDR FSGSKSGTSA
TLGITGLQTG DEADYYCGTW DSSLSAWVFG GGTKLTVL**GS** **GSGSGSGGGG**
SGGGGSLTLA SRQQLIDWME ADKVAGPLLR SALPAGWFIA DKSGAGERGS
RGIIAALGPD GKPSRIVVIY TTGSQATMDE RNRQIAEIGA SLIKHW

Appendix 8.5: Protein sequence of the β-lactamase SFL fusion construct. The glycine-serine linker is highlighted in bold. Residues 30, 31 and 57 are highlighted in yellow. The heavy and light chains are shown in blue and purple, respectively.



Appendix 8.6: Plasmid map of pBR322-βla-SFL.

8.1.3 βla-WTL

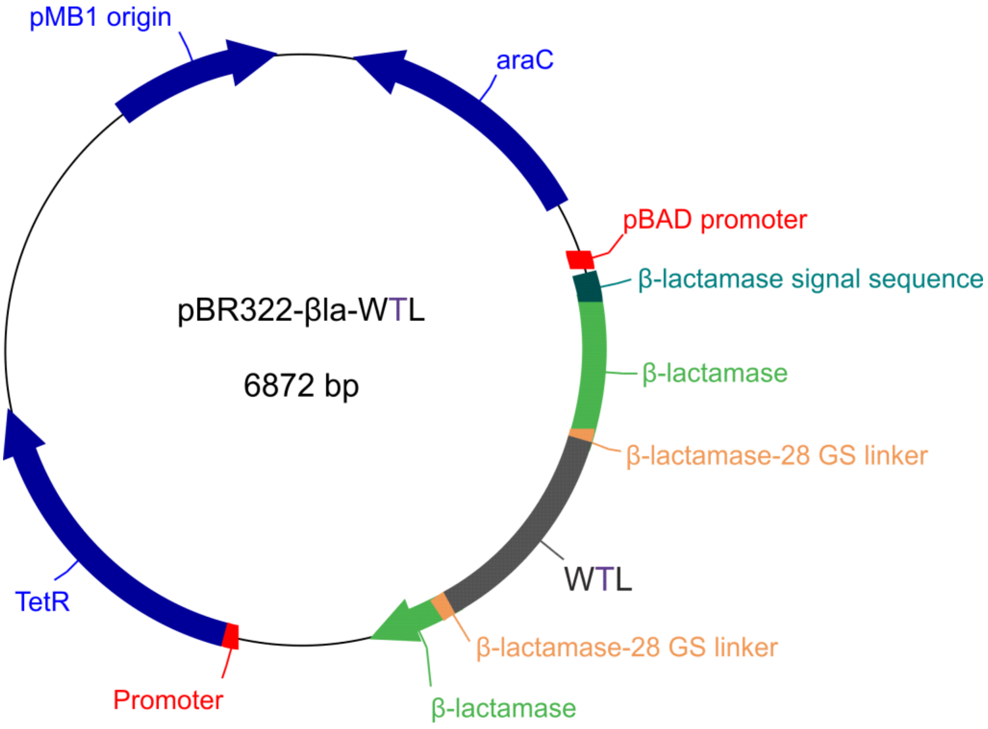
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT
GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC
CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG
CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTA**GG TGGTGGTGGT**
TCTGGTGGTG GTGGCTCGAG CCAGGTTCAG CTTGTGCAGA GCGGTGCGGA GGTCAAAAAA
CCCGGCAGCT CTGTAAAAGT TAGCTGCAAA GCGAGTGGCG GTACGTTTTG GACCGGGGCC
TTTACTTGGG TTCGTCAAGC GCCGGGCCAG GGCTTGGAAT GGATGGGTGG CATTATCCCT
ATTTTCGGCC TCACAAACCT GGCGCAAAAC TTTCAAGGTC GCGTTACCAT TACGGCGGAC
GAAAGCACCA GTACCGTCTA TATGGAGCTG TCAAGCCTGC GCTCAGAAGA CACCGCAGTT
TACTACTGTG CGCGTAGCAG CCGCATTTAC GACTTGAATC CTAGCCTCAC AGCGTACTAC
GACATGGATG TGTGGGGGCA GGGCACCATG GTTACGGTGT CGAGTGGTGG TGGGAGCAGT
GGTGGAGGTG GGTCCGGGGG CGGCGGCGGC GCGCAAAGCG TATTAACTCA GCCGCCGAGC
GTGAGCGCAG CCCCTGGGCA GAAAGTCACC ATTTCATGCA GCGGCTCCTC CAGCGATATC
GGCAACAATT ACGTGTCCTG GTATCAGCAG CTGCCTGGCA CTGCGCCGAA GCTGTTGATT
TATGACAACA ATAAGCGTCC CTCGGGTATT CCAGATCGTT TTTCTGGCTC TAAAAGCGGG
ACATCAGCGA CACTGGGCAT CACCGGGCTG CAGACGGGGG ATGAAGCCGA TTATTACTGC
GGGACCTGGG ATAGTTCCCT GAGCGCGTGG GTGTTTGGCG GGGGCACCAA ACTCACCGTG
CTG**GGATCCG GGAGCGGTTC CGGAAGCGGA GGAGGTGGTT CAGGCGGAGG TGGAAGC**TTG
ACTCTAGCTA GCCGGCAGCA GCTCATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA
CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG
ATAGGTGCCT CACTGATTAA GCATTGGTAA

Appendix 8.7: DNA sequence of the β-lactamase WTL fusion construct. The periplasmic signal sequence is in purple. The WFL sequence is shown in blue with the XhoI and BamHI restriction sites shown in orange and red, respectively. The glycine-serine linker is highlighted in bold. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

MSIQHFRVALIPFFAAFCLPVFA – signal sequence

HPETLVKVKD AEDQLGARVG YIELDLNSGK ILESFRPEER FPMMSTFKVL
LCGAVLSRVD AGQEQLGRRI HYSQNDLVEY SPVTEKHLTD GMTVRELCSA
AITMSDNTAA NLLLTTIGGP KELTAFLHNM GDHVTRLDRW EPELNEAIPN
DERDTTMPAA MATTLRKLLT GEL**GGGGSGG GGSS**QVQLVQ SGAEVKKPGS
SVKVSCKASG GTFWTGAFTW VRQAPGQGLE WMGGIIPIFG LTNLAQNFQG
RVTITADEST STVYMELSSL RSEDTAVYYC ARSSRIYDLN PSLTAYYDMD
VWGQGTMVTV SSGGGSSGGG GSGGGGGAQS VLTQPPSVSA APGQKVTISC
SGSSSDIGNN YVSWYQQLPG TAPKLLIYDN NKRPSGIPDR FSGSKSGTSA
TLGITGLQTG DEADYYCGTW DSSLSAWVFG GGTKLTVL**GS GSGSGSGGGG**
SGGGGSLTLA SRQQLIDWME ADKVAGPLLR SALPAGWFIA DKSGAGERGS
RGIIAALGPD GKPSRIVVIY TTGSQATMDE RNRQIAEIGA SLIKHW

Appendix 8.8: Protein sequence of the β-lactamase WTL fusion construct. The glycine-serine linker is highlighted in bold. Residues 30, 31 and 57 are highlighted in yellow. The heavy and light chains are shown in blue and purple, respectively.



Appendix 8.9: Plasmid map of pBR322-βla-WTL.

8.1.4 βla-WFT

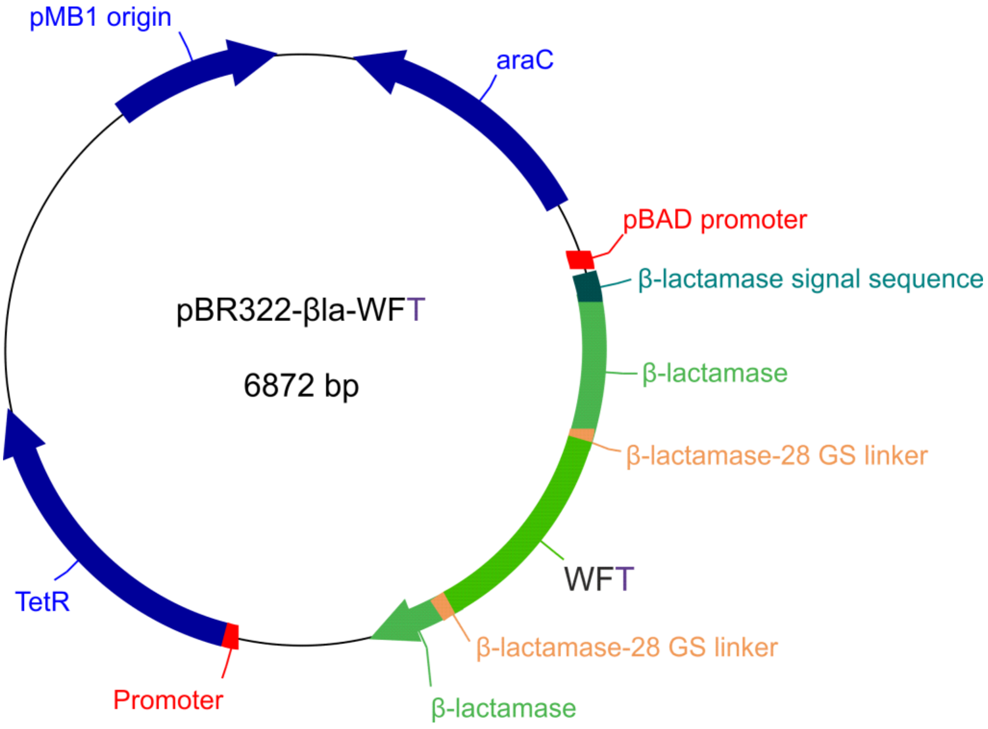
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT
GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC
CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG
CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTA**GG TGGTGGTGGT**
TCTGGTGGTG GTGGCTCGAG CCAGGTTCAG CTTGTGCAGA GCGGTGCGGA GGTCAAAAAA
CCCGGCAGCT CTGTAAAAGT TAGCTGCAAA GCGAGTGGCG GTACGTTTTG GTTTGGGGCC
TTTACTTGGG TTCGTCAAGC GCCGGGCCAG GGCTTGGAAT GGATGGGTGG CATTATCCCT
ATTTTCGGCA CGACAAACCT GGCGCAAAAC TTTCAAGGTC GCGTTACCAT TACGGCGGAC
GAAAGCACCA GTACCGTCTA TATGGAGCTG TCAAGCCTGC GCTCAGAAGA CACCGCAGTT
TACTACTGTG CGCGTAGCAG CCGCATTTAC GACTTGAATC CTAGCCTCAC AGCGTACTAC
GACATGGATG TGTGGGGGCA GGGCACCATG GTTACGGTGT CGAGTGGTGG TGGGAGCAGT
GGTGGAGGTG GGTCCGGGGG CGGCGGCGGC GCGCAAAGCG TATTAACTCA GCCGCCGAGC
GTGAGCGCAG CCCCTGGGCA GAAAGTCACC ATTTCATGCA GCGGCTCCTC CAGCGATATC
GGCAACAATT ACGTGTCCTG GTATCAGCAG CTGCCTGGCA CTGCGCCGAA GCTGTTGATT
TATGACAACA ATAAGCGTCC CTCGGGTATT CCAGATCGTT TTTCTGGCTC TAAAAGCGGG
ACATCAGCGA CACTGGGCAT CACCGGGCTG CAGACGGGGG ATGAAGCCGA TTATTACTGC
GGGACCTGGG ATAGTTCCCT GAGCGCGTGG GTGTTTGGCG GGGGCACCAA ACTCACCGTG
CTG**GGATCCG GGAGCGGTTC CGGAAGCGGA GGAGGTGGTT CAGGCGGAGG TGGAAGC**TTG
ACTCTAGCTA GCCGGCAGCA GCTCATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA
CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG
ATAGGTGCCT CACTGATTAA GCATTGGTAA

Appendix 8.10: DNA sequence of the β-lactamase WFT fusion construct. The periplasmic signal sequence is in purple. The WFL sequence is shown in blue with the XhoI and BamHI restriction sites shown in orange and red, respectively. The glycine-serine linker is highlighted in bold. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

MSIQHFRVALIPFFAAFCLPVFA – signal sequence

HPETLVKVKD AEDQLGARVG YIELDLNSGK ILESFRPEER FPMMSTFKVL
LCGAVLSRVD AGQEQLGRRI HYSQNDLVEY SPVTEKHLTD GMTVRELCSA
AITMSDNTAA NLLLTTIGGP KELTAFLHNM GDHVTRLDRW EPELNEAIPN
DERDTTMPAA MATTLRKLLT GEL**GGGGSGG** **GGSS**QVQLVQ SGAEVKKPGS
SVKVSCKASG GTFWFGAFTW VRQAPGQGLE WMGGIIPIFG TTNLAQNFQG
RVTITADEST STVYMELSSL RSEDTAVYYC ARSSRIYDLN PSLTAYYDMD
VWGQGTMVTV SSGGGSSGGG GSGGGGGAQS VLTQPPSVSA APGQKVTISC
SGSSSDIGNN YVSWYQQLPG TAPKLLIYDN NKRPSGIPDR FSGSKSGTSA
TLGITGLQTG DEADYYCGTW DSSLSAWVFG GGTKLTVL**GS** **GSGSGSGGGG**
SGGGGSLTLA SRQQLIDWME ADKVAGPLLR SALPAGWFIA DKSGAGERGS
RGIIAALGPD GKPSRIVVIY TTGSQATMDE RNRQIAEIGA SLIKHW

Appendix 8.11: Protein sequence of the β-lactamase WFT fusion construct. The glycine-serine linker is highlighted in bold. Residues 30, 31 and 57 are highlighted in yellow. The heavy and light chains are shown in blue and purple, respectively.



Appendix 8.12: Plasmid map of pBR322-βla-WTL.

8.1.5 βla-STL

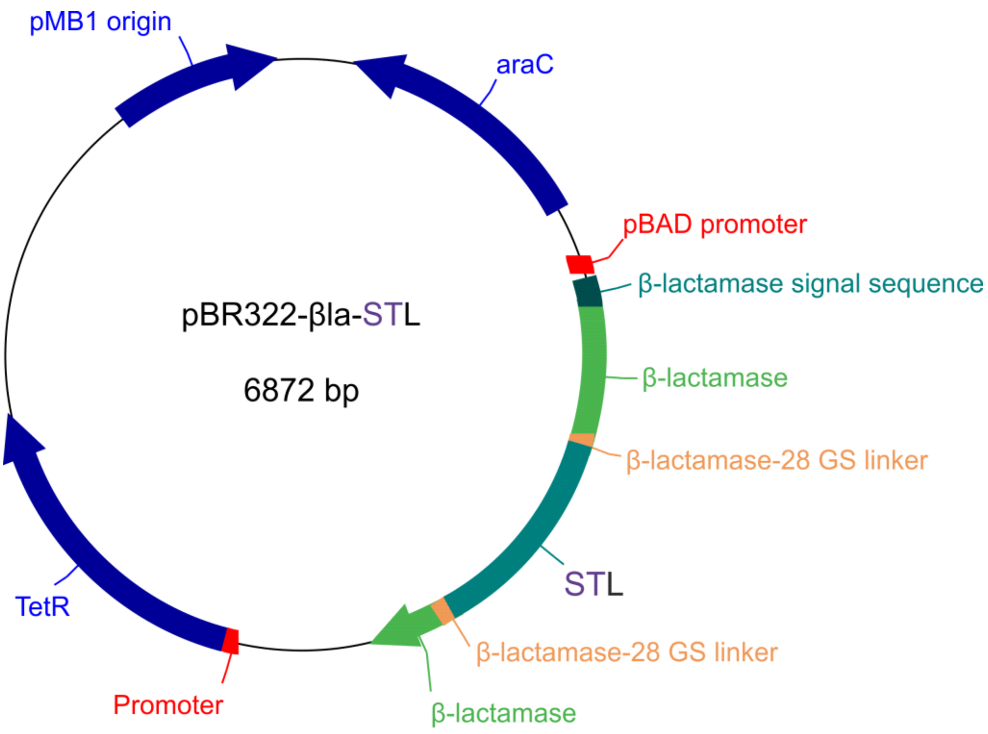
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT
GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC
CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG
CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTA**GG** **TGGTGGTGGT**
TCTGGTGGTG **GTGGCTCGAG** **C**CAGGTTCAG CTGGTTCAGA GTGGCGCGGA GGTTAAGAAA
CCGGGGTCAA GTGTTAAGGT GTCTTGCAAA GCGAGCGGCG GCACCTTCTC CACCGGAGCG
TTTACGTGGG TGCGTCAGGC GCCGGGCCAG GGTTTGGAAT GGATGGGCGG TATCATCCCG
ATTTTTGGGC TCACGAACCT GGCCCAAAAT TTCCAGGGGC GGGTGACGAT CACGGCAGAT
GAGAGTACAT CGACTGTGTA TATGGAACTG TCCAGCTTGC GCAGTGAAGA CACGGCGGTG
TATTACTGCG CCCGCTCCAG CCGGATCTAC GACTTGAACC CATCCCTGAC CGCCTATTAT
GATATGGATG TATGGGGCCA AGGCACGATG GTGACGGTGA GCTCTGGCGG CGGGTCAAGT
GGGGGAGGGG GTTCGGGGGG TGGTGGAGGC GCACAGTCTG TTTTAACCCA GCCCCCGAGT
GTGAGCGCCG CACCCGGGCA GAAAGTGACA ATCTCGTGTT CGGGTAGCTC TAGCGACATC
GGCAATAATT ATGTCAGCTG GTACCAACAG CTGCCGGGAA CTGCGCCTAA ATTGCTGATC
TACGATAATA ACAAACGTCC GAGTGGAATC CCTGACCGCT TCTCCGGGTC AAAAAGCGGC
ACCAGCGCCA CGTTAGGGAT TACTGGCCTG CAAACGGGGG ATGAAGCAGA CTATTATTGT
GGCACGTGGG ATAGCAGTTT AAGTGCCTGG GTTTTTGGGG GCGGCACTAA ACTGACCGTT
CTG**GGATCCG** **GGAGCGGTTC** **CGGAAGCGGA** **GGAGGTGGTT** **CAGGCGGAGG** **TGGAAGC**TTG
ACTCTAGCTA GCCGGCAGCA GCTCATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA
CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG
ATAGGTGCCT CACTGATTAA GCATTGGTAA

Appendix 8.13: DNA sequence of the β-lactamase STL fusion construct. The periplasmic signal sequence is in purple. The WFL sequence is shown in blue with the XhoI and BamHI restriction sites shown in orange and red, respectively. The glycine-serine linker is highlighted in bold. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

MSIQHFRVALIPFFAAFCLPVFA – signal sequence

HPETLVKVKD AEDQLGARVG YIELDLNSGK ILESFRPEER FPMMSTFKVL
LCGAVLSRVD AGQEQLGRRI HYSQNDLVEY SPVTEKHLTD GMTVRELCSA
AITMSDNTAA NLLLTTIGGP KELTAFLHNM GDHVTRLDRW EPELNEAIPN
DERDTTMPAA MATTLRKLLT GEL**GGGGSGG GGSS**QVQLVQ SGAEVKKPGS
SVKVSCKASG GTFSTGAFTW VRQAPGQGLE WMGGIIPIFG LTNLAQNFQG
RVTITADEST STVYMELSSL RSEDTAVYYC ARSSRIYDLN PSLTAYYDMD
VWGQGTMVTV SSGGGSSGGG GSGGGGGAQS VLTQPPSVSA APGQKVTISC
SGSSSDIGNN YVSWYQQLPG TAPKLLIYDN NKRPSGIPDR FSGSKSGTSA
TLGITGLQTG DEADYYCGTW DSSLSAWVFG GGTKLTVL**GS GSGSGSGGGG**
SGGGGSLTLA SRQQLIDWME ADKVAGPLLR SALPAGWFIA DKSGAGERGS
RGIIAALGPD GKPSRIVVIY TTGSQATMDE RNRQIAEIGA SLIKHW

Appendix 8.14: Protein sequence of the β-lactamase STL fusion construct. The glycine-serine linker is highlighted in bold. Residues 30, 31 and 57 are highlighted in yellow. The heavy and light chains are shown in blue and purple, respectively.



Appendix 8.15: Plasmid map of pBR322-βla-STL.

8.1.6 βla-SFT

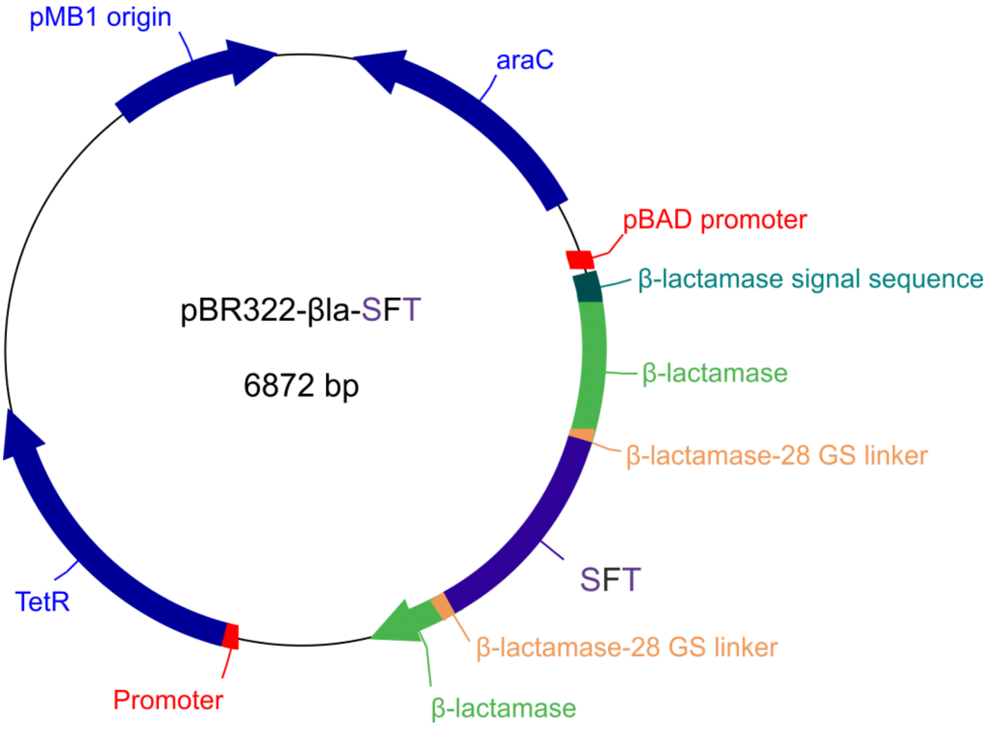
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT
GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC
CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG
CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTA**GG TGGTGGTGGT**
TCTGGTGGTG GTGGCTCGAG CCAGGTTCAG CTGGTTCAGA GTGGCGCGGA GGTTAAGAAA
CCGGGGTCAA GTGTTAAGGT GTCTTGCAAA GCGAGCGGCG GCACCTTCTC CTTTGGAGCG
TTTACGTGGG TGCGTCAGGC GCCGGGCCAG GGTTTGGAAT GGATGGGCGG TATCATCCCG
ATTTTTGGGA CGACGAACCT GGCCCAAAAT TTCCAGGGGC GGGTGACGAT CACGGCAGAT
GAGAGTACAT CGACTGTGTA TATGGAACTG TCCAGCTTGC GCAGTGAAGA CACGGCGGTG
TATTACTGCG CCCGCTCCAG CCGGATCTAC GACTTGAACC CATCCCTGAC CGCCTATTAT
GATATGGATG TATGGGGCCA AGGCACGATG GTGACGGTGA GCTCTGGCGG CGGGTCAAGT
GGGGGAGGGG GTTCGGGGGG TGGTGGAGGC GCACAGTCTG TTTTAACCCA GCCCCCGAGT
GTGAGCGCCG CACCCGGGCA GAAAGTGACA ATCTCGTGTT CGGGTAGCTC TAGCGACATC
GGCAATAATT ATGTCAGCTG GTACCAACAG CTGCCGGGAA CTGCGCCTAA ATTGCTGATC
TACGATAATA ACAAACGTCC GAGTGGAATC CCTGACCGCT TCTCCGGGTC AAAAAGCGGC
ACCAGCGCCA CGTTAGGGAT TACTGGCCTG CAAACGGGGG ATGAAGCAGA CTATTATTGT
GGCACGTGGG ATAGCAGTTT AAGTGCCTGG GTTTTTGGGG GCGGCACTAA ACTGACCGTT
CTG**GGATCCG GGAGCGGTTC CGGAAGCGGA GGAGGTGGTT CAGGCGGAGG TGGAAGC**TTG
ACTCTAGCTA GCCGGCAGCA GCTCATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA
CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG
ATAGGTGCCT CACTGATTAA GCATTGGTAA

Appendix 8.16: DNA sequence of the β-lactamase SFT fusion construct. The periplasmic signal sequence is in purple. The WFL sequence is shown in blue with the XhoI and BamHI restriction sites shown in orange and red, respectively. The glycine-serine linker is highlighted in bold. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

MSIQHFRVALIPFFAAFCLPVFA – signal sequence

HPETLVKVKD AEDQLGARVG YIELDLNSGK ILESFRPEER FPMMSTFKVL
LCGAVLSRVD AGQEQLGRRI HYSQNDLVEY SPVTEKHLTD GMTVRELCSA
AITMSDNTAA NLLLTTIGGP KELTAFLHNM GDHVTRLDRW EPELNEAIPN
DERDTTMPAA MATTLRKLLT GEL**GGGGSGG GGSS**QVQLVQ SGAEVKKPGS
SVKVSCKASG GTFSFGAFTW VRQAPGQGLE WMGGIIPIFG TTNLAQNFQG
RVTITADEST STVYMELSSL RSEDTAVYYC ARSSRIYDLN PSLTAYYDMD
VWGQGTMVTV SSGGGSSGGG GSGGGGGAQS VLTQPPSVSA APGQKVTISC
SGSSSDIGNN YVSWYQQLPG TAPKLLIYDN NKRPSGIPDR FSGSKSGTSA
TLGITGLQTG DEADYYCGTW DSSLSAWVFG GGTKLTVL**GS GSGSGSGGGG**
SGGGGSLTLA SRQQLIDWME ADKVAGPLLR SALPAGWFIA DKSGAGERGS
RGIIAALGPD GKPSRIVVIY TTGSQATMDE RNRQIAEIGA SLIKHW

Appendix 8.17: Protein sequence of the β-lactamase SFT fusion construct. The glycine-serine linker is highlighted in bold. Residues 30, 31 and 57 are highlighted in yellow. The heavy and light chains are shown in blue and purple, respectively.



Appendix 8.18: Plasmid map of pBR322-βla-SFT.

8.1.7 βla-WTT

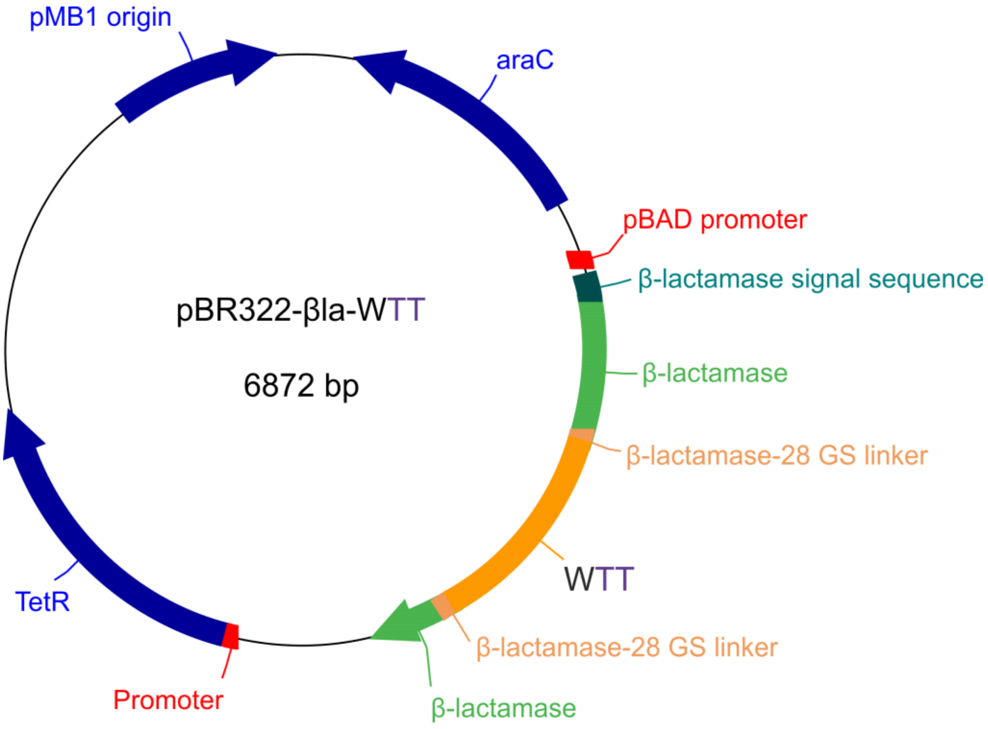
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT
GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC
CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG
CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTA**GG TGGTGGTGGT**
TCTGGTGGTG GTGGCTCGAG CCAGGTTCAG CTGGTTCAGA GTGGCGCGGA GGTTAAGAAA
CCGGGGTCAA GTGTTAAGGT GTCTTGCAAA GCGAGCGGCG GCACCTTCTG GACCGGAGCG
TTTACGTGGG TGCGTCAGGC GCCGGGCCAG GGTTTGGAAT GGATGGGCGG TATCATCCCG
ATTTTTGGGA CGACGAACCT GGCCCAAAAT TTCCAGGGGC GGGTGACGAT CACGGCAGAT
GAGAGTACAT CGACTGTGTA TATGGAACTG TCCAGCTTGC GCAGTGAAGA CACGGCGGTG
TATTACTGCG CCCGCTCCAG CCGGATCTAC GACTTGAACC CATCCCTGAC CGCCTATTAT
GATATGGATG TATGGGGCCA AGGCACGATG GTGACGGTGA GCTCTGGCGG CGGGTCAAGT
GGGGGAGGGG GTTCGGGGGG TGGTGGAGGC GCACAGTCTG TTTTAACCCA GCCCCCGAGT
GTGAGCGCCG CACCCGGGCA GAAAGTGACA ATCTCGTGTT CGGGTAGCTC TAGCGACATC
GGCAATAATT ATGTCAGCTG GTACCAACAG CTGCCGGGAA CTGCGCCTAA ATTGCTGATC
TACGATAATA ACAAACGTCC GAGTGGAATC CCTGACCGCT TCTCCGGGTC AAAAAGCGGC
ACCAGCGCCA CGTTAGGGAT TACTGGCCTG CAAACGGGGG ATGAAGCAGA CTATTATTGT
GGCACGTGGG ATAGCAGTTT AAGTGCCTGG GTTTTTGGGG GCGGCACTAA ACTGACCGTT
CTG**GGATCCG GGAGCGGTTC CGGAAGCGGA GGAGGTGGTT CAGGCGGAGG TGGAAGC**TTG
ACTCTAGCTA GCCGGCAGCA GCTCATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA
CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG
ATAGGTGCCT CACTGATTAA GCATTGGTAA

Appendix 8.19: DNA sequence of the β-lactamase WTT fusion construct. The periplasmic signal sequence is in purple. The WFL sequence is shown in blue with the XhoI and BamHI restriction sites shown in orange and red, respectively. The glycine-serine linker is highlighted in bold. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

MSIQHFRVALIPFFAAFCLPVFA – signal sequence

HPETLVKVKD AEDQLGARVG YIELDLNSGK ILESFRPEER FPMMSTFKVL
LCGAVLSRVD AGQEQLGRRI HYSQNDLVEY SPVTEKHLTD GMTVRELCSA
AITMSDNTAA NLLLTTIGGP KELTAFLHNM GDHVTRLDRW EPELNEAIPN
DERDTTMPAA MATTLRKLLT GEL**GGGGSGG GGSS**QVQLVQ SGAEVKKPGS
SVKVSCKASG GTFWTGAFTW VRQAPGQGLE WMGGIIPIFG TTNLAQNFQG
RVTITADEST STVYMELSSL RSEDTAVYYC ARSSRIYDLN PSLTAYYDMD
VWGQGTMVTV SSGGGSSGGG GSGGGGGAQS VLTQPPSVSA APGQKVTISC
SGSSSDIGNN YVSWYQQLPG TAPKLLIYDN NKRPSGIPDR FSGSKSGTSA
TLGITGLQTG DEADYYCGTW DSSLSAWVFG GGTKLTVL**GS GSGSGSGGGG**
SGGGGSLTLA SRQQLIDWME ADKVAGPLLR SALPAGWFIA DKSGAGERGS
RGIIAALGPD GKPSRIVVIY TTGSQATMDE RNRQIAEIGA SLIKHW

Appendix 8.20: Protein sequence of the β-lactamase WTT fusion construct. The glycine-serine linker is highlighted in bold. Residues 30, 31 and 57 are highlighted in yellow. The heavy and light chains are shown in blue and purple, respectively.



Appendix 8.21: Plasmid map of pBR322-βla-WTT.

8.1.8 βla-STT

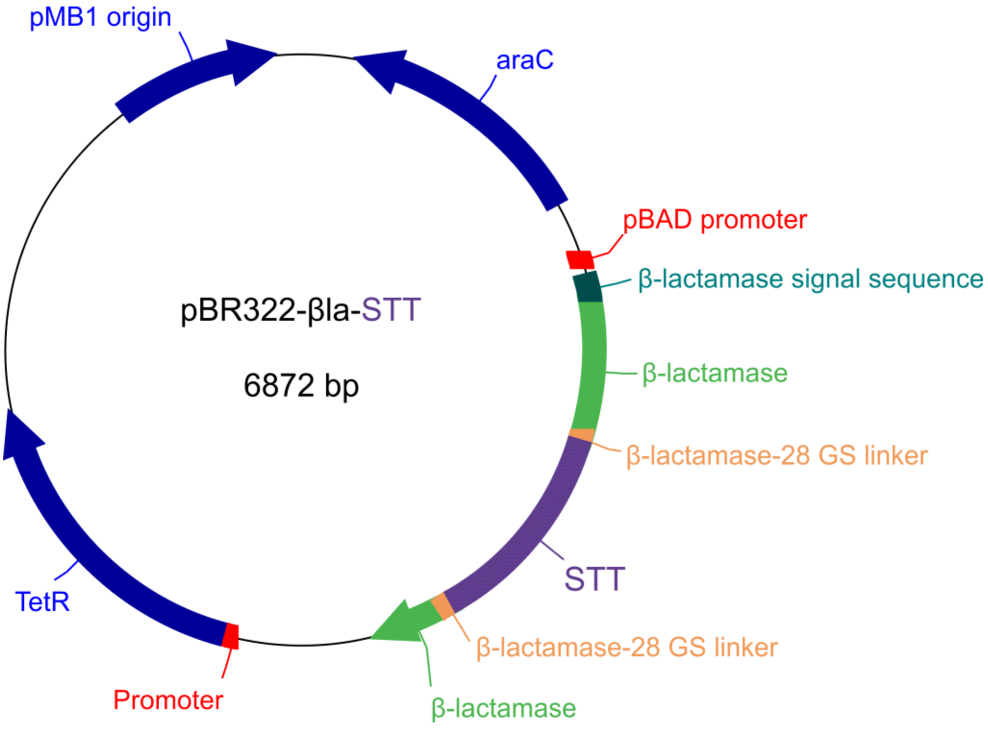
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT
GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC
CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG
CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTA**GG** **TGGTGGTGGT**
TCTGGTGGTG **GTGGCTCGAG** **C**CAGGTTCAG CTGGTTCAGA GTGGCGCGGA GGTTAAGAAA
CCGGGGTCAA GTGTTAAGGT GTCTTGCAAA GCGAGCGGCG GCACCTTCTC CACCGGAGCG
TTTACGTGGG TGCGTCAGGC GCCGGGCCAG GGTTTGGAAT GGATGGGCGG TATCATCCCG
ATTTTTGGGA CGACGAACCT GGCCCAAAAT TTCCAGGGGC GGGTGACGAT CACGGCAGAT
GAGAGTACAT CGACTGTGTA TATGGAACTG TCCAGCTTGC GCAGTGAAGA CACGGCGGTG
TATTACTGCG CCCGCTCCAG CCGGATCTAC GACTTGAACC CATCCCTGAC CGCCTATTAT
GATATGGATG TATGGGGCCA AGGCACGATG GTGACGGTGA GCTCTGGCGG CGGGTCAAGT
GGGGGAGGGG GTTCGGGGGG TGGTGGAGGC GCACAGTCTG TTTTAACCCA GCCCCCGAGT
GTGAGCGCCG CACCCGGGCA GAAAGTGACA ATCTCGTGTT CGGGTAGCTC TAGCGACATC
GGCAATAATT ATGTCAGCTG GTACCAACAG CTGCCGGGAA CTGCGCCTAA ATTGCTGATC
TACGATAATA ACAAACGTCC GAGTGGAATC CCTGACCGCT TCTCCGGGTC AAAAAGCGGC
ACCAGCGCCA CGTTAGGGAT TACTGGCCTG CAAACGGGGG ATGAAGCAGA CTATTATTGT
GGCACGTGGG ATAGCAGTTT AAGTGCCTGG GTTTTTGGGG GCGGCACTAA ACTGACCGTT
CTG**GGATCCG** **GGAGCGGTTC** **CGGAAGCGGA** **GGAGGTGGTT** **CAGGCGGAGG** **TGGAAGC**TTG
ACTCTAGCTA GCCGGCAGCA GCTCATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA
CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG
ATAGGTGCCT CACTGATTAA GCATTGGTAA

Appendix 8.22: DNA sequence of the β-lactamase STT fusion construct. The periplasmic signal sequence is in purple. The WFL sequence is shown in blue with the XhoI and BamHI restriction sites shown in orange and red, respectively. The glycine-serine linker is highlighted in bold. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

MSIQHFRVALIPFFAAFCLPVFA – signal sequence

HPETLVKVKD AEDQLGARVG YIELDLNSGK ILESFRPEER FPMMSTFKVL
LCGAVLSRVD AGQEQLGRRI HYSQNDLVEY SPVTEKHLTD GMTVRELCSA
AITMSDNTAA NLLLTTIGGP KELTAFLHNM GDHVTRLDRW EPELNEAIPN
DERDTTMPAA MATTLRKLLT GEL**GGGGSGG** **GGSS**QVQLVQ SGAEVKKPGS
SVKVSCKASG GTFSTGAFTW VRQAPGQGLE WMGGIIPIFG TTNLAQNFQG
RVTITADEST STVYMELSSL RSEDTAVYYC ARSSRIYDLN PSLTAYYDMD
VWGQGTMVTV SSGGGSSGGG GSGGGGGAQS VLTQPPSVSA APGQKVTISC
SGSSSDIGNN YVSWYQQLPG TAPKLLIYDN NKRPSGIPDR FSGSKSGTSA
TLGITGLQTG DEADYYCGTW DSSLSAWVFG GGTKLTVL**GS** **GSGSGSGGGG**
SGGGGSLTLA SRQQLIDWME ADKVAGPLLR SALPAGWFIA DKSGAGERGS
RGIIAALGPD GKPSRIVVIY TTGSQATMDE RNRQIAEIGA SLIKHW

Appendix 8.23: Protein sequence of the β-lactamase STT fusion construct. The glycine-serine linker is highlighted in bold. Residues 30, 31 and 57 are highlighted in yellow. The heavy and light chains are shown in blue and purple, respectively.



Appendix 8.24: Plasmid map of pBR322-βla-STT. Generated and kindly provided by Dr Janet Saunders (University of Leeds).

8.2 DNA and protein sequences of scFv constructs used for in vitro characterisation and their plasmid maps

8.2.1 WFL_scFv

```
CATATGCAGG TTCAGCTGGT TCAGAGTGGT GCGGAGGTCA AAAAACCCGG
CAGCTCTGTA AAAGTTAGCT GCAAAGCGAG TGGCGGTACG TTTTGGTTTG
GGGCCTTTAC TTGGGTTCGT CAAGCGCCGG GCCAGGGCTT GGAATGGATG
GGTGGCATTA TCCCTATTTT CGGCCTCACA AACCTGGCGC AAAACTTTCA
AGGTCGCGTT ACCATTACGG CGGACGAAAG CACCAGTACC GTCTATATGG
AGCTGTCAAG CCTGCGCTCA GAAGACACCG CAGTTTACTA CTGTGCGCGT
AGCAGCCGCA TTTACGACTT GAATCCTAGC CTCACAGCGT ACTACGACAT
GGATGTGTGG GGGCAGGGCA CCATGGTTAC GGTGTCGAGT GGTGGTGGGA
GCAGTGGTGG AGGTGGGTCC GGGGGCGGCG GCGGCGCGCA AAGCGTATTA
ACTCAGCCGC CGAGCGTGAG CGCAGCCCCT GGGCAGAAAG TCACCATTTC
ATGCAGCGGC TCCTCCAGCG ATATCGGCAA CAATTACGTG TCCTGGTATC
AGCAGCTGCC TGGCACTGCG CCGAAGCTGT TGATTTATGA CAACAATAAG
CGTCCCTCGG GTATTCCAGA TCGTTTTTCT GGCTCTAAAA GCGGGACATC
AGCGACACTG GGCATCACCG GGCTGCAGAC GGGGGATGAA GCCGATTATT
ACTGCGGGAC CTGGGATAGT TCCCTGAGCG CGTGGGTGTT TGGCGGGGGC
ACCAAACTCA CCGTGCTGGG GGCGGCGGCG CACCATCACC ATCACCACTA
ACTCGAG
```

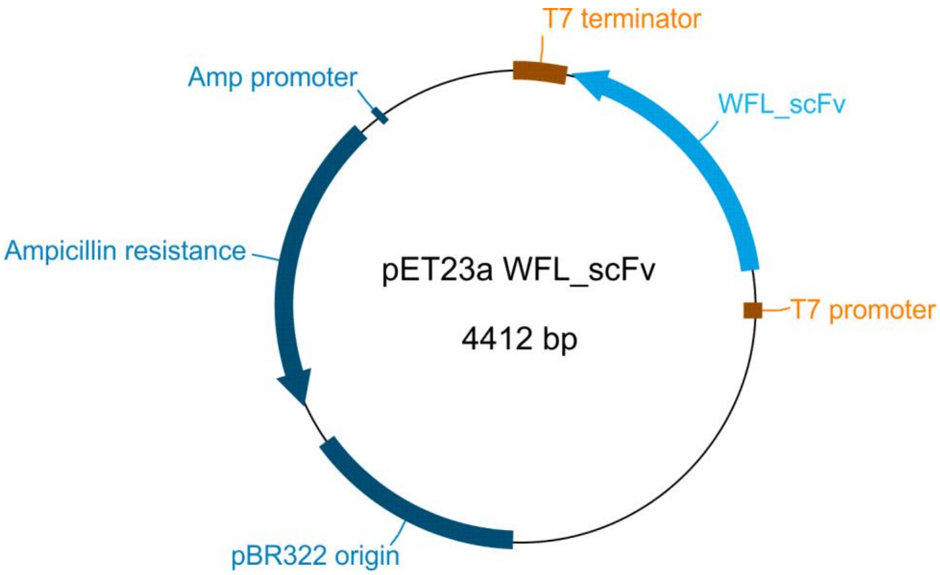
Appendix 8.25: DNA sequence of WFL_scFv. NdeI and XhoI restriction sites are shown in gold and orange, respectively. Additional GAAA amino acid codons are shown in pink with the hexa-histag shown in green. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

```
MQVQLVQSGA EVKKPGSSVK VSCKASGGTF WFGAFTWVRQ APGQGLEWMG
GIIPIFGLTN LAQNFQGRVT ITADESTSTV YMELSSLRSE DTAVYYCARS
SRIYDLNPSL TAYYDMDVWG QGTMVTVSSG GGSSGGGGSG GGGGAQSVLT
QPPSVSAAPG QKVTISCSGS SSDIGNNYVS WYQQLPGTAP KLLIYDNNKR
PSGIPDRFSG SKSGTSATLG ITGLQTGDEA DYYCGTWDSS LSAWVFGGGT
KLTVLGAAAH HHHHH
```

Appendix 8.26: Protein sequence of WFL_scFv. Additional GAAA amino acids are shown in pink with the hexa-histag shown in green. Residues 30, 31 and 57 are highlighted in yellow.

Appendix 8.27: Table of WFL_scFv properties.

Amino acids	Base pairs	MW	pI	ε
265	795	27,629.6 Da	6.57	55,140



Appendix 8.28: Plasmid map of pET23a WFL-scFv.

8.2.2 SFL_scFv

```
CATATGCAGG TTCAGCTGGT TCAGAGTGGT GCGGAGGTCA AAAAACCCGG
CAGCTCTGTA AAAGTTAGCT GCAAAGCGAG TGGCGGTACG TTTTCCTTTG
GGGCCTTTAC TTGGGTTCGT CAAGCGCCGG GCCAGGGCTT GGAATGGATG
GGTGGCATTA TCCCTATTTT CGGCCTCACA AACCTGGCGC AAAACTTTCA
AGGTCGCGTT ACCATTACGG CGGACGAAAG CACCAGTACC GTCTATATGG
AGCTGTCAAG CCTGCGCTCA GAAGACACCG CAGTTTACTA CTGTGCGCGT
AGCAGCCGCA TTTACGACTT GAATCCTAGC CTCACAGCGT ACTACGACAT
GGATGTGTGG GGGCAGGGCA CCATGGTTAC GGTGTCGAGT GGTGGTGGGA
GCAGTGGTGG AGGTGGGTCC GGGGGCGGCG GCGGCGCGCA AAGCGTATTA
ACTCAGCCGC CGAGCGTGAG CGCAGCCCCT GGGCAGAAAG TCACCATTTC
ATGCAGCGGC TCCTCCAGCG ATATCGGCAA CAATTACGTG TCCTGGTATC
AGCAGCTGCC TGGCACTGCG CCGAAGCTGT TGATTTATGA CAACAATAAG
CGTCCCTCGG GTATTCCAGA TCGTTTTTCT GGCTCTAAAA GCGGGACATC
AGCGACACTG GGCATCACCG GGCTGCAGAC GGGGGATGAA GCCGATTATT
ACTGCGGGAC CTGGGATAGT TCCCTGAGCG CGTGGGTGTT TGGCGGGGGC
ACCAAACTCA CCGTGCTGGG GGCGGCGGCG CACCATCACC ATCACCACTA
ACTCGAG
```

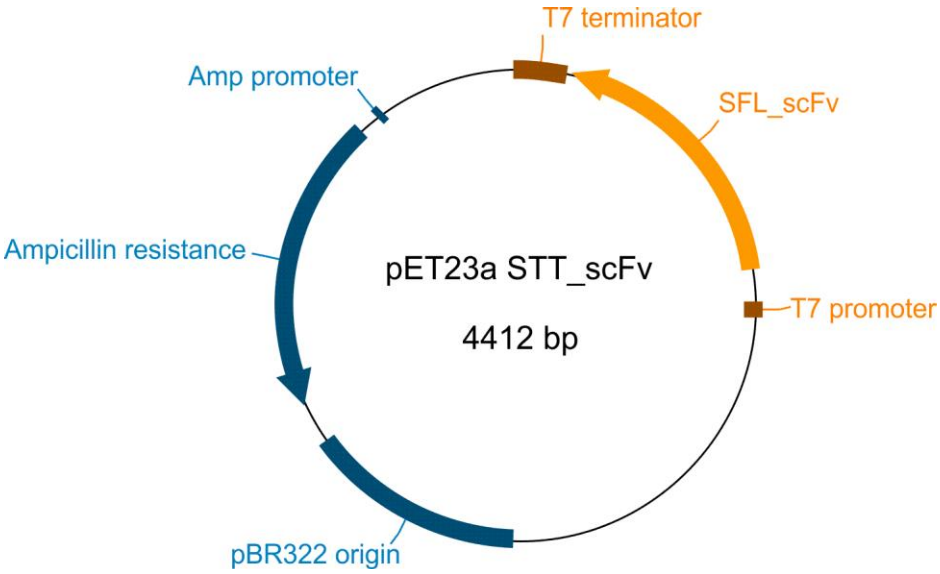
Appendix 8.29: DNA sequence of SFL_scFv. NdeI and XhoI restriction sites are shown in gold and orange, respectively. Additional GAAA amino acid codons are shown in pink with the hexa-histag shown in green. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

```
MQVQLVQSGA EVKKPGSSVK VSCKASGGTF SFGAFTWVRQ APGQGLEWMG
GIIPIFGLTN LAQNFQGRVT ITADESTSTV YMELSSLRSE DTAVYYCARS
SRIYDLNPSL TAYYDMDVWG QGTMVTVSSG GGSSGGGGSG GGGGAQSVLT
QPPSVSAAPG QKVTISCSGS SSDIGNNYVS WYQQLPGTAP KLLIYDNNKR
PSGIPDRFSG SKSGTSATLG ITGLQTGDEA DYYCGTWDSS LSAWVFGGGT
KLTVLGAAAH HHHHH
```

Appendix 8.30: Protein sequence of SFL_scFv. Additional GAAA amino acids are shown in pink with the hexa-histag shown in green. Residues 30, 31 and 57 are highlighted in yellow.

Appendix 8.31: Table of SFL_scFv properties.

Amino acids	Base pairs	MW	pI	ε
265	795	27,530.5 Da	6.57	49,640



Appendix 8.32: Plasmid map of pET23a SFL-scFv.

8.2.3 STT_scFv

```
CATATGCAGG TTCAGCTGGT TCAGAGTGGC GCGGAGGTTA AGAAACCGGG
GTCAAGTGTT AAGGTGTCTT GCAAAGCGAG CGGCGGCACC TTCTCCACCG
GAGCGTTTAC GTGGGTGCGT CAGGCGCCGG GCCAGGGTTT GGAATGGATG
GGCGGTATCA TCCCGATTTT TGGGACGACG AACCTGGCCC AAAATTTCCA
GGGGCGGGTG ACGATCACGG CAGATGAGAG TACATCGACT GTGTATATGG
AACTGTCCAG CTTGCGCAGT GAAGACACGG CGGTGTATTA CTGCGCCCGC
TCCAGCCGGA TCTACGACTT GAACCCATCC CTGACCGCCT ATTATGATAT
GGATGTATGG GGCCAAGGCA CGATGGTGAC GGTGAGCTCT GGCGGCGGGT
CAAGTGGGGG AGGGGGTTCG GGGGGTGGTG GAGGCGCACA GTCTGTTTTA
ACCCAGCCCC CGAGTGTGAG CGCCGCACCC GGGCAGAAAG TGACAATCTC
GTGTTCGGGT AGCTCTAGCG ACATCGGCAA TAATTATGTC AGCTGGTACC
AACAGCTGCC GGGAACTGCG CCTAAATTGC TGATCTACGA TAATAACAAA
CGTCCGAGTG GAATCCCTGA CCGCTTCTCC GGGTCAAAAA GCGGCACCAG
CGCCACGTTA GGGATTACTG GCCTGCAAAC GGGGGATGAA GCAGACTATT
ATTGTGGCAC GTGGGATAGC AGTTTAAGTG CCTGGGTTTT TGGGGGCGGC
ACTAAACTGA CCGTTCTGGC GGCAGCGCAC CATCACCATC ACCACTAACT
CGAG
```

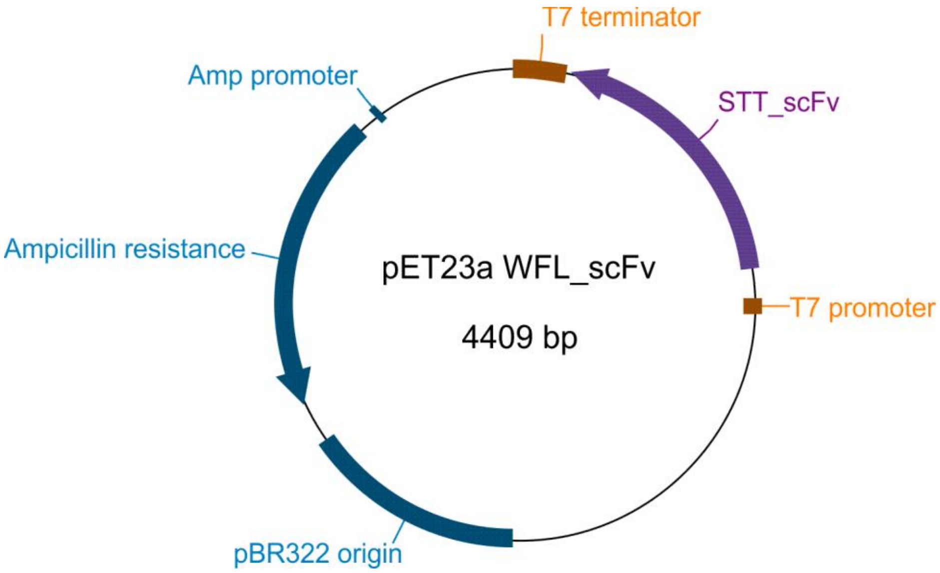
Appendix 8.33: DNA sequence of STT_scFv. NdeI and XhoI restriction sites are shown in gold and orange, respectively. Additional AAA amino acid codons are shown in pink with the hexa-histag shown in green. Codons for positions 30, 31 and 57 are highlighted in yellow. Start and stop codons are underlined.

```
MQVQLVQSGA EVKKPGSSVK VSCKASGGTF STGAFTWVRQ APGQGLEWMG
GIIPIFGTTN LAQNFQGRVT ITADESTSTV YMELSSLRSE DTAVYYCARS
SRIYDLNPSL TAYYDMDVWG QGTMVTVSSG GGSSGGGGSG GGGGAQSVLT
QPPSVSAAPG QKVTISCSGS SSDIGNNYVS WYQQLPGTAP KLLIYDNNKR
PSGIPDRFSG SKSGTSATLG ITGLQTGDEA DYYCGTWDSS LSAWVFGGGT
KLTVLAAAHH HHHH
```

Appendix 8.34: Protein sequence of SFL_scFv. Additional AAA amino acids are shown in pink with the hexa-histag shown in green. Residues 30, 31 and 57 are highlighted in yellow.

Appendix 8.35: Table of STT_scFv properties.

Amino acids	Base pairs	MW	pI	ε
264	792	27,415.3 Da	6.57	49,640



Appendix 8.36: Plasmid map of pET23a STT_scFv.

8.3 Cross-links identified using StavroX

Appendix 8.37: Cross-linked peptides identified using StavroX for the cross-linked monomer sample of WFL with a score above 100.

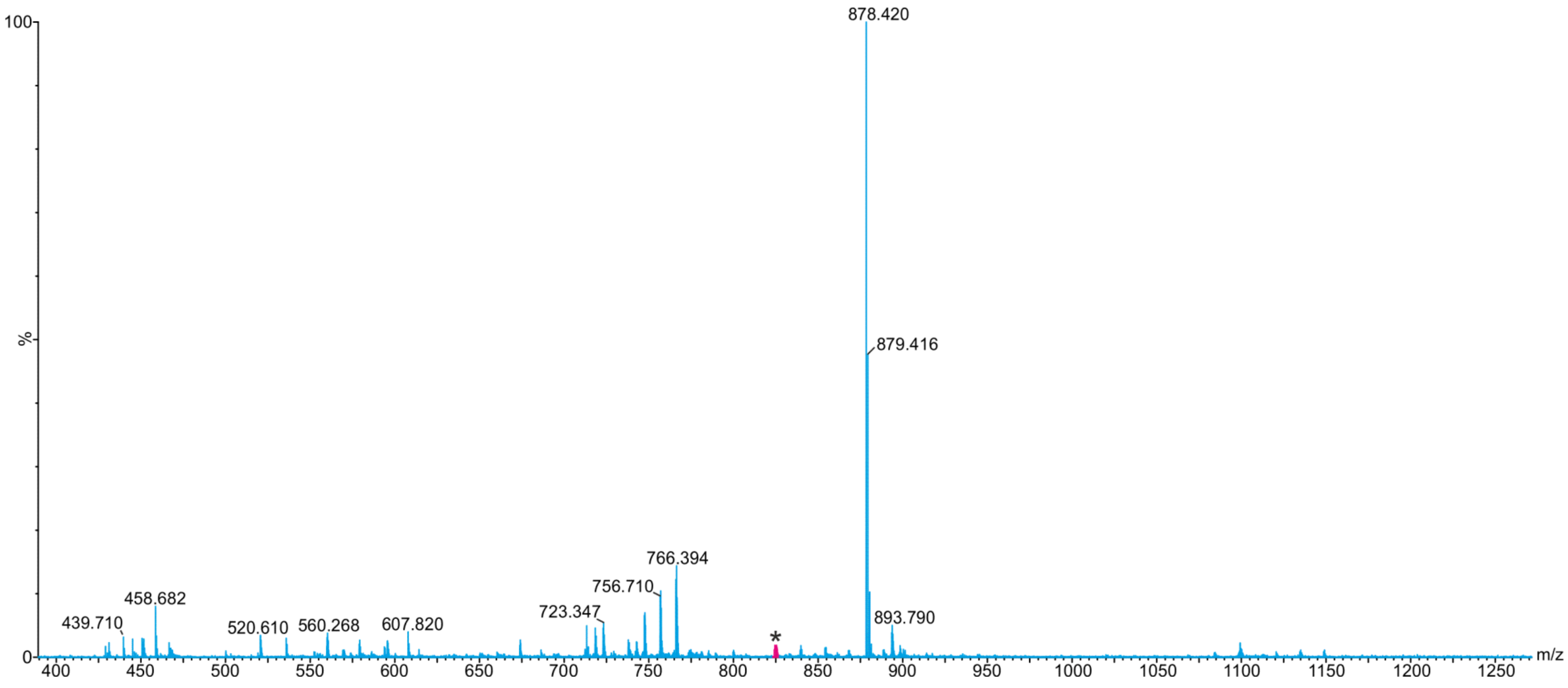
WFL cross-linked monomer	
Peptide 1	Peptide 2
ADSSPVK	AGVETTTPSK
ADSSPVKAGVETTPSK	YAASSYLSLTPEQWKSHR
ADSSPVKAGVETTPSK	SYSBQVTHEGSTVEK
ADSSPVKAGVETTPSK	GPSVFPLAPSSKSTSGGTAALGBLVK
GPSVFPLAPSSK	STSGGTAALGBLVK
LTVDKSR	ADSSPVKAGVETTPSK
LTVDKSR	EEMTKNQVSLTBLVK
QSNNK	YAASSYLSLTPEQWK
RPSGIPDR	LLIYDNNKR
SBDK	THTBPPBPAPEFEGGPSVFLFPPKPK
TISKAK	ADSSPVKAGVETTPSK
TISKAK	LTVDKSR
VDKR	ADSSPVKAGVETTTPSK
VEPKSBDK	ADSSPVKAGVETTPSK
VSNKALPASIEK	ADSSPVKAGVETTPSK

Appendix 8.38: Cross-linked peptides identified using StavroX for the cross-linked dimer sample of WFL with a score about 100. Blue = unique cross-link to the dimer sample (this was not identified in repeat samples).

WFL cross-linked dimer	
Peptide 1	Peptide 2
ADSSPVK	AGVETTTPSK
ADSSPVKAGVETTPSK	YAASSYLSLTPEQWKSHR
ADSSPVKAGVETTPSK	GPSVFPLAPSSKSTSGGTAALGBLVK
GPSVFPLAPSSK	STSGGTAALGBLVK
LTVDKSR	ADSSPVKAGVETTPSK
LTVDKSR	EEMTKNQVSLTBLVK
LTVDKSR	LTVDKSR
QSNNK	YAASSYLSLTPEQWK
RPSGIPDR	LLIYDNNKR
SBDK	THTBPPBPAPEFEGGPSVFLFPPKPK
TISKAK	ADSSPVKAGVETTPSK
VDKR	ADSSPVKAGVETTTPSK
VEPKSBDK	ADSSPVKAGVETTPSK
VSNKALPASIEK	ADSSPVKAGVETTPSK

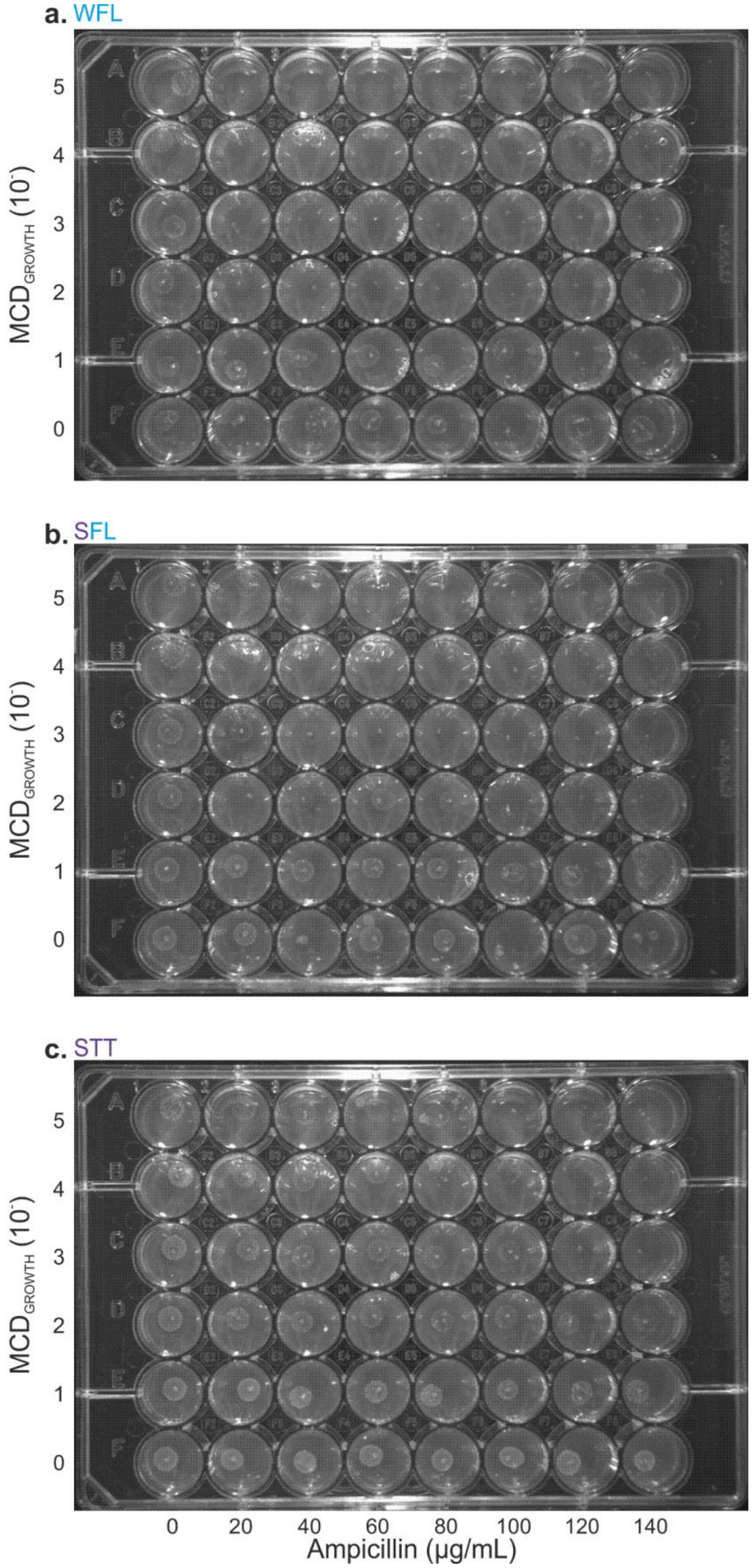
Appendix 8.39: Cross-linked peptides identified using StavroX for the cross-linked monomer sample of STT with a score above 100.

STT cross-linked monomer	
Peptide 1	Peptide 2
ADSSPVK	AGVETTTPSK
ADSSPVKAGVETTPSK	YAASSYLSLTPEQWKSHR
ADSSPVKAGVETTPSK	SYSBQVTHEGSTVEK
ADSSPVKAGVETTPSK	GPSVFPLAPSSKSTSGGTAALGBLVK
AKGQPR	ADSSPVKAGVETTPSK
GPSVFPLAPSSK	STSGGTAALGBLVK
KPGSSVKVSBK	ATLVBLISDFYPGAVTVAWKADSSPVK
LTVDKSR	EEMTKNQVSLTBLVK
QSNNK	YAASSYLSLTPEQWK
SBDK	THTBPPBPAPEFEGGPSVFLFPPKPK
TISK	ALPASIEK
TISKAK	ADSSPVKAGVETTPSK
TISKAK	LTVDKSR
VDKR	ADSSPVKAGVETTTPSK
VSBK	KPGSSVK
VSNKALPASIEK	ADSSPVKAGVETTPSK



Appendix 8.40: Mass spectrum from the LC-MS run of WFL in-gel digested peptides that elute at 21.37 min. * = inter-peptide cross-link at 824.2 *m/z* (magenta).

8.4 Results from the *in vivo* assay



Appendix 8.41: Plates showing the *in vivo* results of **(a)** WFL, **(b)** SFL and **(c)** STT.

References

AGRAWAL, N. J., KUMAR, S., WANG, X., HELK, B., SINGH, S. K. & TROUT, B. L. 2011. Aggregation in protein-based biotherapeutics: computational studies and tools to identify aggregation-prone regions. *Journal of Pharmaceutical Sciences,* 100**,** 5081-95.

AMIN, A., BARNETT, G. V., PATHAK, J. A., ROBERTS, C. J. & SARANGAPANI, P. S. 2014. Protein aggregation, particle formation, characterization & rheology. *Current Opinion in Colloid & Interface Science,* 19**,** 438-449.

ANDREWS, R. K., GORMAN, J. J., BOOTH, W. J., CORINO, G. L., CASTALDI, P. A. & BERNDT, M. C. 1989. Cross-linking of a monomeric 39/34-kDa dispase fragment of von Willebrand factor (Leu-480/Val-481-Gly-718) to the N-terminal region of the. alpha.-chain of membrane glycoprotein Ib on intact platelets with bis (sulfosuccinimidyl) suberate. *Biochemistry,* 28**,** 8326-8336.

ANFINSEN, C. B., HABER, E., SELA, M. & WHITE, F. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences,* 47**,** 1309-1314.

ANTONYUK, S. V., RUSTAGE, N., PETERSEN, C. A., ARNST, J. L., HEYES, D. J., SHARMA, R., BERRY, N. G., SCRUTTON, N. S., EADY, R. R. & ANDREW, C. R. 2011. Carbon monoxide poisoning is prevented by the energy costs of conformational changes in gas-binding haemproteins. *Proceedings of the National Academy of Sciences,* 108**,** 15780-15785.

ARINAMINPATHY, Y., KHURANA, E., ENGELMAN, D. M. & GERSTEIN, M. B. 2009. Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discovery Today,* 14**,** 1130-1135.

ARLT, C., GÖTZE, M., IHLING, C. H., HAGE, C., SCHAEFER, M. & SINZ, A. 2016. An Integrated Workflow for Structural Proteomics Studies based on Cross-linking/Mass Spectrometry with an MS/MS Cleavable Cross-linker. *Analytical Chemistry***,** 7930-7937.

ARTHUR, K. K., GABRIELSON, J. P., KENDRICK, B. S. & STONER, M. R. 2009. Detection of protein aggregates by sedimentation velocity analytical ultracentrifugation (SV-AUC): sources of variability and their relative importance. *Journal of Pharmaceutical Sciences,* 98**,** 3522-39.

ASHCROFT, A. E. 2010. Mass Spectrometry and the Amyloid Problem—How Far Can We Go in the Gas Phase? *Journal of the American Society for Mass Spectrometry,* 21**,** 1087-1096.

ASO, C. & AITO, Y. 1962a. Intramolecular-Intermolecular Polymerization of Glutaraldehyde. *Bulletin of the Chemical Society of Japan,* 35**,** 1426-1426.

ASO, C. & AITO, Y. 1962b. Studies on the polymerization of bifunctional monomers. II. Polymerization of glutaraldehyde. *Die Makromolekulare Chemie,* 58**,** 195-203.

ASTON, F. W. 1935. The story of isotopes. *Science,* 82**,** 235-240.

ATHA, D. H. & INGHAM, K. C. 1981. Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *The Journal of Biological Chemistry,* 256**,** 12108-17.

BACK, J. W., DE JONG, L., MUIJSERS, A. O. & DE KOSTER, C. G. 2003. Chemical cross-linking and mass spectrometry for protein structural modeling. *Journal of Molecular Biology,* 331**,** 303-313.

BAI, X.-C., MCMULLAN, G. & SCHERES, S. H. W. 2015. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences,* 40**,** 49-57.

BALCHIN, D., HAYER-HARTL, M. & HARTL, F. U. 2016. In vivo aspects of protein folding and quality control. *Science,* 353**,** aac4354.

BARBER, M., BORDOLI, R. S., SEDGWICK, R. D. & TYLER, A. N. 1981. Fast atom bombardment of solids (FAB): A new ion source for mass spectrometry. *Journal of the Chemical Society, Chemical Communications***,** 325-327.

BARRERA, N. P., ISAACSON, S. C., ZHOU, M., BAVRO, V. N., WELCH, A., SCHAEDLER, T. A., SEEGER, M. A., MIGUEL, R. N., KORKHOV, V. M. & VAN VEEN, H. W. 2009. Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nature Methods,* 6**,** 585.

BARRERA, N. P. & ROBINSON, C. V. 2011. Advances in the mass spectrometry of membrane proteins: from individual proteins to intact complexes. *Annual Review of Biochemistry,* 80**,** 247-271.

BARTLETT, A. I. & RADFORD, S. E. 2009. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nature Structural & Molecular Biology,* 16**,** 582-588.

BAUD, A., GONNET, F., SALARD, I., LE MIGNON, M., GIULIANI, A., MERCÈRE, P., SCLAVI, B. & DANIEL, R. 2016. Probing the solution structure of Factor H using hydroxyl radical protein footprinting and cross-linking. *Biochemical Journal,* 473**,** 1805-1819.

BAZAN, J., CAŁKOSIŃSKI, I. & GAMIAN, A. 2012. Phage display—A powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. *Human Vaccines & Immunotherapeutics,* 8**,** 1817-1828.

BEESTON, H. S., AULT, J. R., PRINGLE, S. D., BROWN, J. M. & ASHCROFT, A. E. 2015. Changes in protein structure monitored by use of gas-phase hydrogen/deuterium exchange. *Proteomics,* 15**,** 2842-50.

BEKARD, I. B., ASIMAKIS, P., BERTOLINI, J. & DUNSTAN, D. E. 2011. The effects of shear flow on protein structure and function. *Biopolymers,* 95**,** 733-745.

BELSOM, A., SCHNEIDER, M., FISCHER, L., BROCK, O. & RAPPSILBER, J. 2016. Serum albumin domain structures in human blood serum by mass spectrometry and computational biology. *Molecular & Cellular Proteomics,* 15**,** 1105-1116.

BENESCH, J. L. 2009. Collisional activation of protein complexes: picking up the pieces. *Journal of the American Society for Mass Spectrometry,* 20**,** 341-348.

BENESCH, J. L., RUOTOLO, B. T., SIMMONS, D. A. & ROBINSON, C. V. 2007. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chemical Reviews,* 107**,** 3544-67.

BENESCH, J. L. P. & RUOTOLO, B. T. 2011. Mass spectrometry: come of age for structural and dynamical biology. *Current Opinion in Structural Biology,* 21**,** 641-649.

BEVERIDGE, R., COVILL, S., PACHOLARZ, K. J., KALAPOTHAKIS, J. M., MACPHEE, C. E. & BARRAN, P. E. 2014. A mass-spectrometry-based framework to define the extent of disorder in proteins. *Analytical Chemistry,* 86**,** 10979-10991.

BIRD, R., HARDMAN, K., JACOBSON, J., JOHNSON, S., KAUFMAN, B., LEE, S., LEE, T., POPE, S., RIORDAN, G. & WHITLOW, M. 1988. Single-chain antigen-binding proteins. *Science,* 242**,** 423-426.

BLACKWELL, A. E., DODDS, E. D., BANDARIAN, V. & WYSOCKI, V. H. 2011. Revealing the quaternary structure of a heterogeneous noncovalent protein complex through surface-induced dissociation. *Analytical Chemistry,* 83**,** 2862-2865.

BLEIHOLDER, C., DO, T. D., WU, C., ECONOMOU, N. J., BERNSTEIN, S. S., BURATTO, S. K., SHEA, J. E. & BOWERS, M. T. 2013. Ion mobility spectrometry reveals the mechanism of amyloid formation of Abeta(25-35) and its modulation by inhibitors at the molecular level: epigallocatechin gallate and scyllo-inositol. *Journal of the American Chemical Society,* 135**,** 16926-37.

BLEIHOLDER, C., DUPUIS, N. F., WYTTENBACH, T. & BOWERS, M. T. 2011a. Ion mobility-mass spectrometry reveals a conformational conversion from random assembly to beta-sheet in amyloid fibril formation. *Nature Chemistry,* 3**,** 172-7.

BLEIHOLDER, C., WYTTENBACH, T. & BOWERS, M. T. 2011b. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method. *International Journal of Mass Spectrometry,* 308**,** 1-10.

BOROTTO, N. B., ZHOU, Y., HOLLINGSWORTH, S. R., HALE, J. E., GRABAN, E. M., VAUGHAN, R. C. & VACHET, R. W. 2015. Investigating Therapeutic Protein Structure with Diethylpyrocarbonate Labeling and Mass Spectrometry. *Analytical Chemistry,* 87**,** 10627-34.

BORYSIK, A. J. H., RADFORD, S. E. & ASHCROFT, A. E. 2004. Co-populated Conformational Ensembles of β2-Microglobulin Uncovered Quantitatively by Electrospray Ionization Mass Spectrometry. *Journal of Biological Chemistry,* 279**,** 27069-27077.

BOTHNER, B. & SIUZDAK, G. 2004. Electrospray ionization of a whole virus: analyzing mass, structure, and viability. *ChemBioChem,* 5**,** 258-260.

BOYLE, J. P., THOMPSON, T. J., GREGG, E. W., BARKER, L. E. & WILLIAMSON, D. F. 2010. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Population Health Metrics,* 8**,** 29.

BRADBURY, A. R. M., SIDHU, S., DUBEL, S. & MCCAFFERTY, J. 2011. Beyond natural antibodies: the power of in vitro display technologies. *Nature Biotechnology,* 29**,** 245-254.

BREUKER, K. & MCLAFFERTY, F. W. 2008. Stepwise evolution of protein native structure with electrospray into the gas phase, 10(−12) to 10(2) s. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 18145-18152.

BRIGATI, J. R. & PETRENKO, V. A. 2005. Thermostability of landscape phage probes. *Analytical and Bioanalytical Chemistry,* 382**,** 1346-1350.

BROCKWELL, D. J., BEDDARD, G. S., CLARKSON, J., ZINOBER, R. C., BLAKE, A. W., TRINICK, J., OLMSTED, P. D., SMITH, D. A. & RADFORD, S. E. 2002. The Effect of Core Destabilization on the Mechanical Resistance of I27. *Biophysical Journal,* 83**,** 458-472.

BROCKWELL, D. J. & RADFORD, S. E. 2007. Intermediates: ubiquitous species on folding energy landscapes? *Current Opinion in Structural Biology,* 17**,** 30-37.

BRONSOMS, S. & TREJO, S. A. 2015. Applications of mass spectrometry to the study of protein aggregation. *Methods in Molecular Biology,* 1258**,** 331-45.

BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S. & KARPLUS, M. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry,* 4**,** 187-217.

BROWN, R. S. & LENNON, J. J. 1995. Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Analytical Chemistry,* 67**,** 1998-2003.

BRUNNER, J. 1993. New photolabeling and crosslinking methods. *Annual Review of Biochemistry,* 62**,** 483-514.

BUCK, P. M., KUMAR, S., WANG, X., AGRAWAL, N. J., TROUT, B. L. & SINGH, S. K. 2012. Computational methods to predict therapeutic protein aggregation. *Methods in Molecular Biology,* 899**,** 425-51.

BUSH, M. F., HALL, Z., GILES, K., HOYES, J., ROBINSON, C. V. & RUOTOLO, B. T. 2010. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Analytical Chemistry,* 82**,** 9557-9565.

CALABRESE, A. N., AULT, J. R., RADFORD, S. E. & ASHCROFT, A. E. 2015a. Using hydroxyl radical footprinting to explore the free energy landscape of protein folding. *Methods,* 89**,** 38-44.

CALABRESE, A. N., WATKINSON, T. G., HENDERSON, P. J., RADFORD, S. E. & ASHCROFT, A. E. 2015b. Amphipols outperform dodecylmaltoside micelles in stabilizing membrane protein structure in the gas phase. *Analytical Chemistry,* 87**,** 1118-26.

CAMILLONI, C., SALA, B. M., SORMANNI, P., PORCARI, R., CORAZZA, A., DE ROSA, M., ZANINI, S., BARBIROLI, A., ESPOSITO, G., BOLOGNESI, M., BELLOTTI, V., VENDRUSCOLO, M. & RICAGNO, S. 2016. Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability. *Scientific Reports,* 6**,** 25559.

CAMPUZANO, I., BUSH, M. F., ROBINSON, C. V., BEAUMONT, C., RICHARDSON, K., KIM, H. & KIM, H. I. 2011. Structural characterization of drug-like compounds by ion mobility mass spectrometry: comparison of theoretical and experimentally derived nitrogen collision cross sections. *Analytical Chemistry,* 84**,** 1026-1033.

CAO, J., BURKE, J. E. & DENNIS, E. A. 2013. Using hydrogen/deuterium exchange mass spectrometry to define the specific interactions of the phospholipase A2 superfamily with lipid substrates, inhibitors, and membranes. *Journal of Biological Chemistry,* 288**,** 1806-1813.

CARMEN, S. & JERMUTUS, L. 2002. Concepts in antibody phage display. *Briefings in Functional Genomics & Proteomics,* 1**,** 189-203.

CHAMES, P., VAN REGENMORTEL, M., WEISS, E. & BATY, D. 2009. Therapeutic antibodies: successes, limitations and hopes for the future. *British Journal of Pharmacology,* 157**,** 220-233.

CHEN, F., GÜLBAKAN, B., WEIDMANN, S., FAGERER, S. R., IBÁÑEZ, A. J. & ZENOBI, R. 2016a. Applying mass spectrometry to study non-covalent biomolecule complexes. *Mass Spectrometry Reviews,* 35**,** 48-70.

CHEN, G. 2013. *Characterization of protein therapeutics using mass spectrometry*, Springer.

CHEN, J., REMPEL, D. L., GAU, B. C. & GROSS, M. L. 2012. Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level. *Journal of the American Chemical Society,* 134**,** 18724-18731.

CHEN, Z. A., FISCHER, L., TAHIR, S., BUKOWSKI-WILLS, J.-C., BARLOW, P. N. & RAPPSILBER, J. 2016b. Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes. *bioRxiv***,** 1-42.

CHEN, Z. A., JAWHARI, A., FISCHER, L., BUCHEN, C., TAHIR, S., KAMENSKI, T., RASMUSSEN, M., LARIVIERE, L., BUKOWSKI-WILLS, J. C., NILGES, M., CRAMER, P. & RAPPSILBER, J. 2010. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO Journal,* 29**,** 717-26.

CHEN, Z. A., PELLARIN, R., FISCHER, L., SALI, A., NILGES, M., BARLOW, P. N. & RAPPSILBER, J. 2016c. Structure of complement C3 (H2O) revealed by quantitative cross-linking/mass spectrometry and modelling. *Molecular & Cellular Proteomics***,** 1-48.

CHENNAMSETTY, N., HELK, B., VOYNOV, V., KAYSER, V. & TROUT, B. L. 2009a. Aggregation-prone motifs in human immunoglobulin G. *Journal of Molecular Biology,* 391**,** 404-13.

CHENNAMSETTY, N., VOYNOV, V., KAYSER, V., HELK, B. & TROUT, B. L. 2009b. Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences,* 106**,** 11937-11942.

CHEVREUX, G., TILLY, N. & BIHOREAU, N. 2011. Fast analysis of recombinant monoclonal antibodies using IdeS proteolytic digestion and electrospray mass spectrometry. *Analytical Biochemistry,* 415**,** 212-214.

CHIRMULE, N., JAWA, V. & MEIBOHM, B. 2012. Immunogenicity to therapeutic proteins: impact on PK/PD and efficacy. *The AAPS Journal,* 14**,** 296-302.

CHOWDHURY, S., KATTA, V. & CHAIT, B. 1990. Electrospray ionization mass spectrometric peptide mapping: A rapid, sensitive technique for protein structure analysis. *Biochemical and biophysical research communications,* 167**,** 686-692.

CLARK, N. J., ZHANG, H., KRUEGER, S., LEE, H. J., KETCHEM, R. R., KERWIN, B., KANAPURAM, S. R., TREUHEIT, M. J., MCAULEY, A. & CURTIS, J. E. 2013. Small-angle neutron scattering study of a monoclonal antibody using free-energy constraints. *The Journal of Physical Chemistry B,* 117**,** 14029-38.

CLEMMER, D. E., HUDGINS, R. R. & JARROLD, M. F. 1995. Naked protein conformations: cytochrome c in the gas phase. *Journal of the American Chemical Society,* 117**,** 10141-10142.

CLORE, G. M. & SCHWIETERS, C. D. 2003. Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from 1HN/15N chemical shift mapping and backbone 15N-1H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. *Journal of the American Chemical Society,* 125**,** 2902-2912.

COMBE, C. W., FISCHER, L. & RAPPSILBER, J. 2015. xiNET: Cross-link Network Maps With Residue Resolution. *Molecular & Cellular Proteomics,* 14**,** 1137-1147.

CONCHILLO-SOLÉ, O., DE GROOT, N. S., 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**,** 1.

CORRIGAN, R. M., RIGBY, D., HANDLEY, P. & FOSTER, T. J. 2007. The role of Staphylococcus aureus surface protein SasG in adherence and biofilm formation. *Microbiology,* 153**,** 2435-46.

COUNTERMAN, A. E. & CLEMMER, D. E. 1999. Volumes of individual amino acid residues in gas-phase peptide ions. *Journal of the American Chemical Society,* 121**,** 4031-4039.

COURTOIS, F., AGRAWAL, N. J., LAUER, T. M. & TROUT, B. L. 2016. Rational design of therapeutic mAbs against aggregation through protein engineering and incorporation of glycosylation motifs applied to bevacizumab. *MAbs,* 8**,** 99-112.

CROMWELL, M. E. M., HILARIO, E. & JACOBSON, F. 2006. Protein aggregation and bioprocessing. *The AAPS Journal,* 8**,** E572-E579.

D'SOUZA, S., GINSBERG, M. H., LAM, S. & PLOW, E. F. 1988. Chemical cross-linking of arginyl-glycyl-aspartic acid peptides to an adhesion receptor on platelets. *Journal of Biological Chemistry,* 263**,** 3943-3951.

DAGGETT, V. & LEVITT, M. 1993. Realistic simulations of native-protein dynamics in solution and beyond. *Annual Review of Biophysics and Biomolecular Structure,* 22**,** 353-80.

DAILING, A., LUCHINI, A. & LIOTTA, L. 2015. Unlocking the secrets to protein–protein interface drug targets using structural mass spectrometry techniques. *Expert Review of Proteomics,* 12**,** 457-467.

DAVID, M. A. & TAYEBI, M. 2014. Detection of Protein Aggregates in Brain and Cerebrospinal Fluid Derived from Multiple Sclerosis Patients. *Frontiers in Neurology,* 5**,** 251.

DAVIDSON, K. L., OBERREIT, D. R., HOGAN, C. J. & BUSH, M. F. 2016. Nonspecific aggregation in native electrokinetic nanoelectrospray ionization. *International Journal of Mass Spectrometry*.

DAWSON, P. H. 2013. *Quadrupole mass spectrometry and its applications*, Elsevier.

DE GROOT, A. S. & SCOTT, D. W. 2007. Immunogenicity of protein therapeutics. *Trends in Immunology,* 28**,** 482-490.

DEBAENE, F., BOEUF, A., WAGNER-ROUSSET, E., COLAS, O., AYOUB, D., CORVAIA, N., VAN DORSSELAER, A., BECK, A. & CIANFERANI, S. 2014. Innovative native MS methodologies for antibody drug conjugate characterization: High resolution native MS and IM-MS for average DAR and DAR distribution assessment. *Analytical Chemistry,* 86**,** 10674-83.

DEBAENE, F., WAGNER-ROUSSET, E., COLAS, O., AYOUB, D., CORVAIA, N., VAN DORSSELAER, A., BECK, A. & CIANFERANI, S. 2013. Time resolved native ion-mobility mass spectrometry to monitor dynamics of IgG4 Fab arm exchange and "bispecific" monoclonal antibody formation. *Analytical Chemistry,* 85**,** 9785-92.

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**,** 920-933.

DENG, G. & SANYAL, G. 2006. Applications of mass spectrometry in early stages of target based drug discovery. *Journal of Pharmaceutical and Biomedical Analysis,* 40**,** 528-538.

DICKEY, C. A., ERIKSEN, J., KAMAL, A., BURROWS, F., KASIBHATLA, S., ECKMAN, C. B., HUTTON, M. & PETRUCELLI, L. 2005. Development of a high throughput drug screening assay for the detection of changes in tau levels-proof of concept with Hsp90 inhibitors. *Current Alzheimer Research,* 2**,** 231-238.

DILL, K. A. & CHAN, H. S. 1997. From Levinthal to pathways to funnels. *Nature Structural Biology,* 4**,** 10-19.

DOBSON, C. M. 2003. Protein folding and misfolding. *Nature,* 426**,** 884-890.

DOLE, M., MACK, L., HINES, R., MOBLEY, R., FERGUSON, L. & ALICE, M. D. 1968. Molecular beams of macroions. *The Journal of Chemical Physics,* 49**,** 2240-2249.

DONGRE, A. R., SOMOGYI, A. & WYSOCKI, V. H. 1996. Surface-induced dissociation: an effective tool to probe structure, energetics and fragmentation mechanisms of protonated peptides. *Journal of Mass Spectrometry,* 31**,** 339-350.

DORSAM, R. T. & GUTKIND, J. S. 2007. G-protein-coupled receptors and cancer. *Nature Reviews Cancer,* 7**,** 79-94.

DUBINSKY, L., KROM, B. P. & MEIJLER, M. M. 2012. Diazirine based photoaffinity labeling. *Bioorganic & Medicinal Chemistry,* 20**,** 554-570.

DUBOIS, F., KNOCHENMUSS, R., ZENOBI, R., BRUNELLE, A., DEPRUN, C. & LE BEYEC, Y. 1999. A comparison between ion-to-photon and microchannel plate detectors. *Rapid Communications in Mass Spectrometry,* 13**,** 786-791.

DYCHE MULLINS, R. & POLLARD, T. D. 1999. Structure and function of the Arp2/3 complex. *Current Opinion in Structural Biology,* 9**,** 244-249.

ELLGAARD, L. & HELENIUS, A. 2003. Quality control in the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology,* 4**,** 181-191.

ENGLANDER, S. W., SOSNICK, T. R., ENGLANDER, J. J. & MAYNE, L. 1996. Mechanisms and uses of hydrogen exchange. *Current Opinion in Structural Biology,* 6**,** 18-23.

FAINI, M., STENGEL, F. & AEBERSOLD, R. 2016. The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *Journal of the American Society for Mass Spectrometry,* 27**,** 966-974.

FINK, A. L. 1998. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design,* 3**,** R9-R23.

FISCHER, L., CHEN, Z. A. & RAPPSILBER, J. 2013. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *Journal of Proteomics,* 88**,** 120-128.

FISHER, H. 2014. *Development of Techniques and Strategies for the Characterisation and Structure Elucidation of Oligonucleotides by Mass Spectrometry.* University of Leeds.

FITZGERALD, M. C., PARR, G. R. & SMITH, L. M. 1993. Basic matrices for the matrix-assisted laser desorption/ionization mass spectrometry of proteins and oligonucleotides. *Analytical Chemistry,* 65**,** 3204-11.

FLEMING, P. J., PATEL, D. S., WU, E. L., QI, Y., YEOM, M. S., SOUSA, M. C., FLEMING, K. G. & IM, W. 2016. BamA POTRA Domain Interacts with a Native Lipid Membrane Surface. *Biophysical Journal,* 110**,** 2698-709.

FOIT, L., MORGAN, G. J., KERN, M. J., STEIMER, L. R., VON HACHT, A. A., TITCHMARSH, J., WARRINER, S. L., RADFORD, S. E. & BARDWELL, J. C. 2009. Optimizing protein stability in vivo. *Molecular Cell,* 36**,** 861-71.

FRIMPONG, A. K., ABZALIMOV, R. R., UVERSKY, V. N. & KALTASHOV, I. A. 2010. Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: Conformational heterogeneity of α-synuclein. *Proteins: Structure, Function, and Bioinformatics,* 78**,** 714-722.

FROKJAER, S. & OTZEN, D. E. 2005. Protein drug stability: a formulation challenge. *Nature Reviews Drug Discovery,* 4**,** 298-306.

GABRIELSON, J. P., ARTHUR, K. K., KENDRICK, B. S., RANDOLPH, T. W. & STONER, M. R. 2009. Common excipients impair detection of protein aggregates during sedimentation velocity analytical ultracentrifugation. *Journal of Pharmaceutical Sciences,* 98**,** 50-62.

GADGIL, H. S., BONDARENKO, P. V., PIPES, G. D., DILLON, T. M., BANKS, D., ABEL, J., KLEEMANN, G. R. & TREUHEIT, M. J. 2006. Identification of cysteinylation of a free cysteine in the Fab region of a recombinant monoclonal IgG1 antibody using Lys-C limited proteolysis coupled with LC/MS analysis. *Analytical Biochemistry,* 355**,** 165-74.

GAGNON, P. 2008. Improved antibody aggregate removal by hydroxyapatite chromatography in the presence of polyethylene glycol. *Journal of Immunological Methods,* 336**,** 222-228.

GATZEVA-TOPALOVA, P. Z., WARNER, L. R., PARDI, A. & SOUSA, M. C. 2010. Structure and flexibility of the complete periplasmic domain of BamA: the protein insertion machine of the outer membrane. *Structure,* 18**,** 1492-501.

GEOGHEGAN, J. C., FLEMING, R., DAMSCHRODER, M., BISHOP, S. M., SATHISH, H. A. & ESFANDIARY, R. 2016. Mitigation of reversible self-association and viscosity in a human IgG1 monoclonal antibody by rational, structure-guided Fv engineering. *MAbs,* 8**,** 941-50.

GILES, K., WILLIAMS, J. P. & CAMPUZANO, I. 2011. Enhancements in travelling wave ion mobility resolution. *Rapid Communications in Mass Spectrometry,* 25**,** 1559-1566.

GLOCKSHUBER, R., MALIA, M., PFITZINGER, I. & PLUCKTHUN, A. 1990. A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry,* 29**,** 1362-7.

GOLDBERG, A. L. 2003. Protein degradation and protection against misfolded or damaged proteins. *Nature,* 426**,** 895-899.

GÓMEZ DE LA CUESTA, R., GOODACRE, R. & ASHTON, L. 2014. Monitoring Antibody Aggregation in Early Drug Development Using Raman Spectroscopy and Perturbation-Correlation Moving Windows. *Analytical Chemistry,* 86**,** 11133-11140.

GÖTZE, M., PETTELKAU, J., FRITZSCHE, R., IHLING, C. H., SCHÄFER, M. & SINZ, A. 2015. Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis. *Journal of the American Society for Mass Spectrometry,* 26**,** 83-97.

GÖTZE, M., PETTELKAU, J., SCHAKS, S., BOSSE, K., IHLING, C. H., KRAUTH, F., FRITZSCHE, R., KÜHN, U. & SINZ, A. 2012. StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *Journal of the American Society for Mass Spectrometry,* 23**,** 76-87.

GREENBERG, C. H., KOLLMAN, J., ZELTER, A., JOHNSON, R., MACCOSS, M. J., DAVIS, T. N., AGARD, D. A. & SALI, A. 2016. Structure of γ-tubulin small complex based on a cryo-EM map, chemical cross-links, and a remotely related structure. *Journal of Structural Biology,* 194**,** 303-310.

GROVES, M. A. T., AMANUEL, L., CAMPBELL, J. I., REES, D. G., SRIDHARAN, S., FINCH, D. K., LOWE, D. C. & VAUGHAN, T. J. 2014. Antibody VH and VL recombination using phage and ribosome display technologies reveals distinct structural routes to affinity improvements with VH-VL interface residues providing important structural diversity. *MAbs,* 6**,** 236-245.

GRUEBELE, M., DAVE, K. & SUKENIK, S. 2016. Globular Protein Folding In Vitro and In Vivo. *Annual Review of Biophysics,* 45.

GRUSZKA, D. T., WHELAN, F., FARRANCE, O. E., FUNG, H. K., PACI, E., JEFFRIES, C. M., SVERGUN, D. I., BALDOCK, C., BAUMANN, C. G., BROCKWELL, D. J., POTTS, J. R. & CLARKE, J. 2015. Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein. *Nature Communications,* 6**,** 7271.

GRUSZKA, D. T. E. A. 2012. Staphylococcal biofilm-forming protein has a contiguous rod-like structure. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** E1011-E1018.

GU, Y., LI, H., DONG, H., ZENG, Y., ZHANG, Z., PATERSON, N. G., STANSFELD, P. J., WANG, Z., ZHANG, Y., WANG, W. & DONG, C. 2016. Structural basis of outer membrane protein insertion by the BAM complex. *Nature,* 531**,** 64-9.

GUGLIELMI, L. & MARTINEAU, P. 2009. Expression of single-chain Fv fragments in E. coli cytoplasm. *Antibody Phage Display: Methods and Protocols***,** 215-224.

GUILHAUS, M., SELBY, D. & MLYNSKI, V. 2000. Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrometry Reviews,* 19**,** 65-107.

HAILU, T. T., FOIT, L. & BARDWELL, J. C. 2013. In vivo detection and quantification of chemicals that enhance protein stability. *Analytical Biochemistry,* 434**,** 181-6.

HALL, Z., SCHMIDT, C. & POLITIS, A. 2016. Uncovering the early assembly mechanism for amyloidogenic β2-microglobulin using cross-linking and native mass spectrometry. *Journal of Biological Chemistry,* 291**,** 4626-4637.

HAMMERSCHMIDT ET AL. 2015. Continuous polyethylene glycol precipitation of recombinantantibodies: Sequential precipitation and resolubilization. *Process Biochemistry,* 51**,** 325-332.

HANES, J. & PLUCKTHUN, A. 1997. In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 4937-42.

HANSEL, T. T., KROPSHOFER, H., SINGER, T., MITCHELL, J. A. & GEORGE, A. J. 2010. The safety and side effects of monoclonal antibodies. *Nature Reviews Drug Discovery,* 9**,** 325-338.

HARRIS, L. J., LARSON, S. B., HASEL, K. W. & MCPHERSON, A. 1997. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry,* 36**,** 1581-97.

HARRIS, L. J., SKALETSKY, E. & MCPHERSON, A. 1998. Crystallographic structure of an intact IgG1 monoclonal antibody. *Journal of Molecular Biology,* 275**,** 861-72.

HE, M. & TAUSSIG, M. J. 1997. Antibody-Ribosome-mRNA (ARM) Complexes as Efficient Selection Particles for in vitro Display and Evolution of Antibody Combining Sites. *Nucleic Acids Research,* 25**,** 5132-5134.

HECK, A. J. R. 2008. Native mass spectrometry: a bridge between interactomics and structural biology. *Nature Methods,* 5**,** 927-933.

HEINKEL, F. & GSPONER, J. 2016. Determination of protein folding intermediate structures consistent with data from oxidative footprinting mass spectrometry. *Journal of Molecular Biology,* 428**,** 365-371.

HERZOG, F., KAHRAMAN, A., BOEHRINGER, D., MAK, R., BRACHER, A., WALZTHOENI, T., LEITNER, A., BECK, M., HARTL, F.-U. & BAN, N. 2012. Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science,* 337**,** 1348-1352.

HERZOG, W. 2014. The role of titin in eccentric muscle contraction. *Journal of Experimental Biology,* 217**,** 2825-33.

HILTON, G. R., THALASSINOS, K., GRABENAUER, M., SANGHERA, N., SLADE, S. E., WYTTENBACH, T., ROBINSON, P. J., PINHEIRO, T. J., BOWERS, M. T. & SCRIVENS, J. H. 2010. Structural analysis of prion proteins by means of drift cell and traveling wave ion mobility mass spectrometry. *Journal of the American Society for Mass Spectrometry,* 21**,** 845-854.

HINGORANI, K. S. & GIERASCH, L. M. 2014. Comparing Protein Folding In vitro and In vivo: Foldability Meets the Fitness Challenge. *Current Opinion in Structural Biology,* 0**,** 81-90.

HOFFMANN, E. D. & STROOBANT, V. 2001. Mass spectrometry: principles and applications. England: Wiley.

HOFMANN, T., FISCHER, A. W., MEILER, J. & KALKHOF, S. 2015. Protein structure prediction guided by crosslinking restraints – A systematic evaluation of the impact of the crosslinking spacer length. *Methods,* 89**,** 79-90.

HOFSTADLER, S. A. & SANNES-LOWERY, K. A. 2006. Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. *Nature Reviews Drug Discovery,* 5**,** 585-595.

HOGAN, C. J., JR., RUOTOLO, B. T., ROBINSON, C. V. & FERNANDEZ DE LA MORA, J. 2011. Tandem differential mobility analysis-mass spectrometry reveals partial gas-phase collapse of the GroEL complex. *The Journal of Physical Chemistry B,* 115**,** 3614-21.

HONG, P., KOZA, S. & BOUVIER, W. S. P. 2012. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and Their Aggregates. *Journal of Liquid Chromatography & Related Technologies,* 35**,** 2923-2950.

HOPPER, J. T., YU, Y. T.-C., LI, D., RAYMOND, A., BOSTOCK, M., LIKO, I., MIKHAILOV, V., LAGANOWSKY, A., BENESCH, J. L. & CAFFREY, M. 2013. Detergent-free mass spectrometry of membrane protein complexes. *Nature Methods,* 10**,** 1206-1208.

HU, Q., NOLL, R. J., LI, H., MAKAROV, A., HARDMAN, M. & GRAHAM COOKS, R. 2005. The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry,* 40**,** 430-443.

HUANG, B. X., KIM, H.-Y. & DASS, C. 2004. Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *Journal of the American Society for Mass Spectrometry,* 15**,** 1237-1247.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics,* 14**,** 33-38.

HUSTON, J. S., LEVINSON, D., MUDGETT-HUNTER, M., TAI, M. S., NOVOTNÝ, J., MARGOLIES, M. N., RIDGE, R. J., BRUCCOLERI, R. E., HABER, E. & CREA, R. 1988. Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. *Proceedings of the National Academy of Sciences,* 85**,** 5879-5883.

ILLES-TOTH, E., RAMOS, M. R., CAPPAI, R., DALTON, C. & SMITH, D. P. 2015. Structural characterisation of high order alpha-synuclein oligomers capable of inducing intracellular aggregation. *Biochemical Journal*.

IMPROTA, S., POLITOU, A. S. & PASTORE, A. 1996. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure,* 4**,** 323-37.

INGHAM, K. C. 1977. Precipitation of Proteins with Polyethylene Glycol: Characterization of Albumin. *Archives of Biochemistry and Biophysics,* 186**,** 106-113.

IRIBARNE, J. & THOMSON, B. 1976. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics,* 64**,** 2287-2294.

JAHN, T. R. & RADFORD, S. E. 2008. Folding versus aggregation: polypeptide conformations on competing pathways. *Archives of Biochemistry and Biophysics,* 469**,** 100-117.

JAIN, R. K. 1990. Physiological Barriers to Delivery of Monoclonal Antibodies and Other Macromolecules in Tumors. *Cancer Research,* 50**,** 814-819.

JANEWAY, C. A., TRAVERS, P., WALPORT, M. & SHLOMCHIK, M. J. 1997. *Immunobiology: the immune system in health and disease*, Current Biology.

JARROLD, M. F. 1999. Unfolding, refolding, and hydration of proteins in the gas phase. *Accounts of Chemical Research,* 32**,** 360-367.

JENNINGS, K. R. 1968. Collision-induced decompositions of aromatic molecular ions. *International Journal of Mass Spectrometry and Ion Physics,* 1**,** 227-235.

JIN, H., LORIA, J. P. & MOORE, P. B. 2007. Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Molecular Cell,* 26**,** 205-15.

JURASCHEK, R., DÜLCKS, T. & KARAS, M. 1999. Nanoelectrospray—more than just a minimized-flow electrospray ionization source. *Journal of the American Society for Mass Spectrometry,* 10**,** 300-308.

KADDIS, C. S., LOMELI, S. H., YIN, S., BERHANE, B., APOSTOL, M. I., KICKHOEFER, V. A., ROME, L. H. & LOO, J. A. 2007. Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility. *Journal of the American Society for Mass Spectrometry,* 18**,** 1206-1216.

KALTASHOV, I. A., BOBST, C. E. & ABZALIMOV, R. R. 2013. Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Science,* 22**,** 530-544.

KALTASHOV, I. A. & MOHIMEN, A. 2005. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Analytical Chemistry,* 77**,** 5370-5379.

KARAS, M., BACHMANN, D. & HILLENKAMP, F. 1985. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry,* 57**,** 2935-2939.

KAY, B. K., WILLIAMSON, M. P. & SUDOL, M. 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB Journal,* 14**,** 231-241.

KAYSER, V., CHENNAMSETTY, N., VOYNOV, V., HELK, B., FORRER, K. & TROUT, B. L. 2012. A screening tool for therapeutic monoclonal antibodies: Identifying the most stable protein and its best formulation based on thioflavin T binding. *Biotechnology Journal,* 7**,** 127-32.

KEBARLE, P. & VERKERK, U. H. 2009. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews,* 28**,** 898-917.

KEMPENI, J. 1999. Preliminary results of early clinical trials with the fully human anti-TNFα monoclonal antibody D2E7. *Annals of the Rheumatic Diseases,* 58**,** I70-I72.

KIM, M. S. & PANDEY, A. 2012. Electron transfer dissociation mass spectrometry in proteomics. *Proteomics,* 12**,** 530-542.

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.

KIND, T. & FIEHN, O. 2010. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalytical Reviews,* 2**,** 23-60.

KINGDON, K. 1923. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Physical Review,* 21**,** 408.

KLEIN, J. S., JIANG, S., GALIMIDI, R. P., KEEFFE, J. R. & BJORKMAN, P. J. 2014. Design and characterization of structured protein linkers with differing flexibilities. *Protein Engineering, Design & Selection,* 27**,** 325-30.

KNAPMAN, T. W., MORTON, V. L., STONEHOUSE, N. J., STOCKLEY, P. G. & ASHCROFT, A. E. 2010. Determining the topology of virus assembly intermediates using ion mobility spectrometry-mass spectrometry. *Rapid Communications in Mass Spectrometry,* 24**,** 3033-42.

KNOCHENMUSS, R. 2006. Ion formation mechanisms in UV-MALDI. *Analyst,* 131**,** 966-986.

KNOWLES, T. J., JEEVES, M., BOBAT, S., DANCEA, F., MCCLELLAND, D., PALMER, T., OVERDUIN, M. & HENDERSON, I. R. 2008. Fold and function of polypeptide transport-associated domains responsible for delivering unfolded proteins to membranes. *Molecular Microbiology,* 68**,** 1216-27.

KNOWLES, T. P., VENDRUSCOLO, M. & DOBSON, C. M. 2014. The amyloid state and its association with protein misfolding diseases. *Nature Reviews Molecular Cell Biology,* 15**,** 384-396.

KÖHLER, G. & MILSTEIN, C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature,* 256**,** 495-497.

KONERMANN, L., AHADI, E., RODRIGUEZ, A. D. & VAHIDI, S. 2012. Unraveling the mechanism of electrospray ionization. *Analytical Chemistry,* 85**,** 2-9.

KONERMANN, L., PAN, J. & LIU, Y.-H. 2011. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews,* 40**,** 1224-1234.

KONERMANN, L., TONG, X. & PAN, Y. 2008. Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *Journal of Mass Spectrometry,* 43**,** 1021-1036.

KONERMANN, L., VAHIDI, S. & SOWOLE, M. A. 2014. Mass Spectrometry Methods for Studying Structure and Dynamics of Biological Macromolecules. *Analytical Chemistry,* 86**,** 213-232.

KONIJNENBERG, A., VAN DYCK, J. F., KAILING, L. L. & SOBOTT, F. 2015. Extending native mass spectrometry approaches to integral membrane proteins. *Biological Chemistry,* 396**,** 991-1002.

KOSINSKI, J., VON APPEN, A., ORI, A., KARIUS, K., MÜLLER, C. W. & BECK, M. 2015. Xlink Analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *Journal of Structural Biology,* 189**,** 177-183.

LANUCARA, F., HOLMAN, S. W., GRAY, C. J. & EYERS, C. E. 2014. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature Chemistry,* 6**,** 281-294.

LAPPANO, R. & MAGGIOLINI, M. 2011. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature Reviews Drug Discovery,* 10**,** 47-60.

LAUWEREYS, M., ARBABI GHAHROUDI, M., DESMYTER, A., KINNE, J., HÖLZER, W., DE GENST, E., WYNS, L. & MUYLDERMANS, S. 1998. Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *The EMBO Journal,* 17**,** 3512-3520.

LEARY, J. A., SCHENAUER, M. R., STEFANESCU, R., ANDAYA, A., RUOTOLO, B. T., ROBINSON, C. V., THALASSINOS, K., SCRIVENS, J. H., SOKABE, M. & HERSHEY, J. W. 2009. Methodology for measuring conformation of solvent-disrupted protein subunits using T-WAVE ion mobility MS: an investigation into eukaryotic initiation factors. *Journal of the American Society for Mass Spectrometry,* 20**,** 1699-1706.

LEE, C. M., IORNO, N., SIERRO, F. & CHRIST, D. 2007. Selection of human antibody fragments by phage display. *Nature Protocols,* 2**,** 3001-3008.

LEITNER, A. 2016. Cross-linking and other structural proteomics techniques: how chemistry is enabling mass spectrometry applications in structural biology. *Chemical Science*.

LEITNER, A., FAINI, M., STENGEL, F. & AEBERSOLD, R. 2016. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends in Biochemical Sciences,* 41**,** 20-32.

LEITNER, A., WALZTHOENI, T. & AEBERSOLD, R. 2014. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nature Protocols,* 9**,** 120-137.

LENEY, A. C., MCMORRAN, L. M., RADFORD, S. E. & ASHCROFT, A. E. 2012. Amphipathic polymers enable the study of functional membrane proteins in the gas phase. *Analytical Chemistry,* 84**,** 9841-7.

LEVINTHAL, C. 1969. How to fold graciously. *Mossbauer spectroscopy in biological systems,* 67**,** 22-24.

LI, H., OBERHAUSER, A. F., REDICK, S. D., CARRION-VAZQUEZ, M., ERICKSON, H. P. & FERNANDEZ, J. M. 2001. Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 10682-6.

LIU, F. & HECK, A. J. 2015. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Current Opinion in Structural Biology,* 35**,** 100-108.

LIU, F., RIJKERS, D. T., POST, H. & HECK, A. J. 2015. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nature Methods,* 12**,** 1179-1184.

LIU, H., ZHANG, H., KING, J. D., WOLF, N. R., PRADO, M., GROSS, M. L. & BLANKENSHIP, R. E. 2014. Mass spectrometry footprinting reveals the structural rearrangements of cyanobacterial orange carotenoid protein upon light activation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1837**,** 1955-1963.

LIU, J. K. H. 2014. The history of monoclonal antibody development – Progress, remaining challenges and future innovations. *Annals of Medicine and Surgery,* 3**,** 113-116.

LIU, Y. D., GOETZE, A. M., BASS, R. B. & FLYNN, G. C. 2011. N-terminal glutamate to pyroglutamate conversion in vivo for human IgG2 antibodies. *The Journal of Biological Chemistry,* 286**,** 11211-7.

LOO, J. A. 1997. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews,* 16**,** 1-23.

LOO, J. A., BERHANE, B., KADDIS, C. S., WOODING, K. M., XIE, Y., KAUFMAN, S. L. & CHERNUSHEVICH, I. V. 2005. Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *Journal of the American Society for Mass Spectrometry,* 16**,** 998-1008.

LOURIS, J. N., WRIGHT, L. G., COOKS, R. G. & SCHOEN, A. E. 1985. New scan modes accessed with a hybrid mass spectrometer. *Analytical Chemistry,* 57**,** 2918-2924.

LU, Z., DENG, S., HUANG, D., HE, Y., LEI, M., ZHOU, L. & JIN, P. 2012. Frontier of therapeutic antibody discovery: The challenges and how to face them. . *World Journal of Biological Chemistry,* 3**,** 187-196.

LUBKOWSKI, J., DAUTER, M., AGHAIYPOUR, K., WLODAWER, A. & DAUTER, Z. 2003. Atomic resolution structure of Erwinia chrysanthemi L-asparaginase. *Acta Crystallographica Section D: Biological Crystallography,* 59**,** 84-92.

LUCHINI, A., ESPINA, V. & LIOTTA, L. A. 2014. Protein painting reveals solvent-excluded drug targets hidden within native protein-protein interfaces. *Nature Communications,* 5**,** 4413.

MABUD, M. A., DEKREY, M. J. & GRAHAM COOKS, R. 1985. Surface-induced dissociation of molecular ions. *International Journal of Mass Spectrometry and Ion Processes,* 67**,** 285-294.

MAGGON, K. 2007. Monoclonal antibody "gold rush". *Current medicinal chemistry,* 14**,** 1978-1987.

MAHLER, H. C., FRIESS, W., GRAUSCHOPF, U. & KIESE, S. 2009. Protein aggregation: pathways, induction factors and analysis. *Journal of Pharmaceutical Sciences,* 98**,** 2909-2934.

MAKAROV, A. 2000. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry,* 72**,** 1156-1162.

MALITO, E., FALERI, A., LO SURDO, P., VEGGI, D., MARUGGI, G., GRASSI, E., CARTOCCI, E., BERTOLDI, I., GENOVESE, A., SANTINI, L., ROMAGNOLI, G., BORGOGNI, E., BRIER, S., LO PASSO, C., DOMINA, M., CASTELLINO, F., FELICI, F., VAN DER VEEN, S., JOHNSON, S., LEA, S. M., TANG, C. M., PIZZA, M., SAVINO, S., NORAIS, N., RAPPUOLI, R., BOTTOMLEY, M. J. & MASIGNANI, V. 2013. Defining a protective epitope on factor H binding protein, a key meningococcal virulence factor and vaccine antigen. *Proceedings of the National Academy of Sciences,* 110**,** 3304-3309.

MAMYRIN, B. 2001. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry,* 206**,** 251-266.

MAMYRIN, B., KARATAEV, V., SHMIKK, D. & ZAGULIN, V. 1973. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Soviet Journal of Experimental and Theoretical Physics,* 37**,** 45.

MARCH, R. E. 1997. An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry,* 32**,** 351-369.

MARKLUND, E. G., DEGIACOMI, M. T., ROBINSON, C. V., BALDWIN, A. J. & BENESCH, J. L. 2015. Collision cross sections for structural proteomics. *Structure,* 23**,** 791-799.

MARSHALL, A. G. & HENDRICKSON, C. L. 2002. Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *International Journal of Mass Spectrometry,* 215**,** 59-75.

MARTY, M. T., ZHANG, H., CUI, W., BLANKENSHIP, R. E., GROSS, M. L. & SLIGAR, S. G. 2012. Native mass spectrometry characterization of intact nanodisc lipoprotein complexes. *Analytical Chemistry,* 84**,** 8957-8960.

MASON, E. A. & SCHAMP, H. W. 1958. Mobility of gaseous lons in weak electric fields. *Annals of Physics,* 4**,** 233-270.

MATSUURA, S., UMEBAYASHI, S., OKUYAMA, C. & OBA, K. 1985. Characteristics of the newly developed MCP and its assembly. *IEEE Transactions on Nuclear Science,* 32**,** 350-354.

MEHMOOD, S., MARCOUX, J., HOPPER, J. T., ALLISON, T. M., LIKO, I., BORYSIK, A. J. & ROBINSON, C. V. 2014. Charge reduction stabilizes intact membrane protein complexes for mass spectrometry. *Journal of the American Chemical Society,* 136**,** 17010-2.

MENDOZA, V. L. & VACHET, R. W. 2009. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrometry Reviews,* 28**,** 785-815.

MERK, A., BARTESAGHI, A., BANERJEE, S., FALCONIERI, V., RAO, P., DAVIS, M. I., PRAGANI, R., BOXER, M. B., EARL, L. A. & MILNE, J. L. 2016. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell,* 165**,** 1698-1707.

MESLEH, M. F., HUNTER, J. M., SHVARTSBURG, A. A., SCHATZ, G. C. & JARROLD, M. F. 1996. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *Journal of Physical Chemistry,* 100**,** 16082-16086.

MIGNEAULT, I., DARTIGUENAVE, C., BERTRAND, M. J. & WALDRON, K. C. 2004. Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *BioTechniques,* 37**,** 12.

MILLER, B. R., DEMAREST, S. J., LUGOVSKOY, A., HUANG, F., WU, X., SNYDER, W. B., CRONER, L. J., WANG, N., AMATUCCI, A., MICHAELSON, J. S. & GLASER, S. M. 2010. Stability engineering of scFvs for the development of bispecific and multivalent antibodies. *Protein Engineering, Design & Selection,* 23**,** 549-57.

MILLER, P. E. & DENTON, M. B. 1986. The quadrupole mass filter: basic operating concepts. *Journal of Chemical Education,* 63**,** 617.

MITRAGOTRI, S., BURKE, P. A. & LANGER, R. 2014. Overcoming the challenges in administering biopharmaceuticals: formulation and delivery strategies. *Nature Reviews Drug Discovery,* 13**,** 655-672.

MORADI, M., BABIN, V., ROLAND, C., DARDEN, T. A. & SAGUI, C. 2009. Conformations and free energy landscapes of polyproline peptides. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 20746-51.

MORGNER, N., SCHMIDT, C., BEILSTEN-EDMANDS, V., EBONG, I.-O., PATEL, NISHA A., CLERICO, EUGENIA M., KIRSCHKE, E., DATURPALLI, S., JACKSON, SOPHIE E., AGARD, D. & ROBINSON, CAROL V. 2015. Hsp70 Forms Antiparallel Dimers Stabilized by Post-translational Modifications to Position Clients for Transfer to Hsp90. *Cell Reports,* 11**,** 759-769.

MURAKAMI, K., ELMLUND, H., KALISMAN, N., BUSHNELL, D. A., ADAMS, C. M., AZUBEL, M., ELMLUND, D., LEVI-KALISMAN, Y., LIU, X., GIBBONS, B. J., LEVITT, M. & KORNBERG, R. D. 2013. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science,* 342**,** 1238724.

NAVARE, A. T., CHAVEZ, J. D., ZHENG, C., WEISBROD, C. R., ENG, J. K., SIEHNEL, R., SINGH, P. K., MANOIL, C. & BRUCE, J. E. 2015. Probing the protein interaction network of Pseudomonas aeruginosa cells by chemical cross-linking mass spectrometry. *Structure,* 23**,** 762-773.

NELSON, A. L. & REICHERT, J. M. 2009. Development trends for therapeutic antibody fragments. *Nature Biotechnology,* 27**,** 331-337.

NELSON, P. N., REYNOLDS, G. M., WALDRON, E. E., WARD, E., GIANNOPOULOS, K. & MURRAY, P. G. 2000. Demystified ...: Monoclonal antibodies. *Molecular Pathology,* 53**,** 111-117.

NGUYEN-HUYNH, N. T., SHAROV, G., POTEL, C., FICHTER, P., TROWITZSCH, S., BERGER, I., LAMOUR, V., SCHULTZ, P., POTIER, N. & LEIZE-WAGNER, E. 2015. Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex. *Protein Science,* 24**,** 1232-1246.

NIU, S. & RUOTOLO, B. T. 2015. Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding. *Protein Science,* 24**,** 1272-81.

NOGALES, E. 2016. The development of cryo-EM into a mainstream structural biology technique. *Nature Methods,* 13**,** 24-27.

NÜRNBERGER, T., NENNSTIEL, D., HAHLBROCK, K. & SCHEEL, D. 1995. Covalent cross-linking of the Phytophthora megasperma oligopeptide elicitor to its receptor in parsley membranes. *Proceedings of the National Academy of Sciences,* 92**,** 2338-2342.

OBERHAUSER, A. F., MARSZALEK, P. E., CARRION-VAZQUEZ, M. & FERNANDEZ, J. M. 1999. Single protein misfolding events captured by atomic force microscopy. *Nature Structural Biology,* 6**,** 1025-8.

OBREZANOVA, O., ARNELL, A., DE LA CUESTA, R. G., BERTHELOT, M. E., GALLAGHER, T. R., ZURDO, J. & STALLWOOD, Y. 2015. Aggregation risk prediction for antibodies and its application to biotherapeutic development. *MAbs,* 7**,** 352-63.

OLSEN, J. V., ONG, S.-E. & MANN, M. 2004. Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Molecular & Cellular Proteomics,* 3**,** 608-614.

PACHOLARZ, K. J., GARLISH, R. A., TAYLOR, R. J. & BARRAN, P. E. 2012. Mass spectrometry based tools to investigate protein–ligand interactions for drug discovery. *Chemical Society Reviews,* 41**,** 4335-4355.

PACHOLARZ, K. J., PORRINI, M., GARLISH, R. A., BURNLEY, R. J., TAYLOR, R. J., HENRY, A. J. & BARRAN, P. E. 2014. Dynamics of intact immunoglobulin G explored by drift-tube ion-mobility mass spectrometry and molecular modeling. *Angewandte Chemie International Edition,* 53**,** 7765-9.

PACKIAM, M., WEINRICK, B., JACOBS, W. R. & MAURELLI, A. T. 2015. Structural characterization of muropeptides from Chlamydia trachomatis peptidoglycan by mass spectrometry resolves "chlamydial anomaly". *Proceedings of the National Academy of Sciences,* 112**,** 11660-11665.

PALECANDA, A., WALCHECK, B., BISHOP, D. K. & JUTILA, M. A. 1992. Rapid activation-independent shedding of leukocyte L-selectin induced by cross-linking of the surface antigen. *European Journal of Immunology,* 22**,** 1279-1286.

PAUL, A. J. & SCHWAB, K. H., F. 2014. Direct analysis of mAb aggregates in mammalian cell culture supernatant. *BMC Biotechnology,* 14**,** 1-11.

PAUL, W. 1990. Electromagnetic traps for charged and neutral particles. *Reviews of Modern Physics,* 62**,** 531.

PEPPEL, K., CRAWFORD, D. & BEUTLER, B. 1991. A tumor necrosis factor (TNF) receptor-IgG heavy chain chimeric protein as a bivalent antagonist of TNF activity. *The Journal of experimental medicine,* 174**,** 1483-1489.

PHILIPPIDIS, A. 2012. *Top 10 Best-Selling Drugs of the 21st Century* [Online]. Genetic Engineering & Biotechnology News. Available: http://www.genengnews.com/insight-and-intelligence/top-10-best-selling-drugs-of-the-21st-century/77899716/.

PHILIPPIDIS, A. 2015. *The Top 25 Best-Selling Drugs of 2014* [Online]. Genetic Engeinnering & Biotechnology News. Available: http://www.genengnews.com/keywordsandtools/print/3/37387/.

PHILLIPS, J. C., BRAUN, R., WANG, W., GUMBART, J., TAJKHORSHID, E., VILLA, E., CHIPOT, C., SKEEL, R. D., KALE, L. & SCHULTEN, K. 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry,* 26**,** 1781-802.

PHILO, J. S. & ARAKAWA, T. 2009. Mechanisms of Protein Aggregation. *Current Pharmaceutical Biotechnology,* 10**,** 348-351.

PINKSE, M. W. H., MAIER, C. S., KIM, J.-I., OH, B.-H. & HECK, A. J. R. 2003. Macromolecular assembly of Helicobacter pylori urease investigated by mass spectrometry. *Journal of Mass Spectrometry,* 38**,** 315-320.

PIRRONE, G. F., IACOB, R. E. & ENGEN, J. R. 2015. Applications of Hydrogen/Deuterium Exchange MS from 2012 to 2014. *Analytical Chemistry,* 87**,** 99-118.

POLITIS, A., SCHMIDT, C., TJIOE, E., SANDERCOCK, A. M., LASKER, K., GORDIYENKO, Y., RUSSEL, D., SALI, A. & ROBINSON, C. V. 2015. Topological models of heteromeric protein assemblies from mass spectrometry: application to the yeast eIF3: eIF5 complex. *Chemistry & Biology,* 22**,** 117-128.

PRINGLE, S. D., GILES, K., WILDGOOSE, J. L., WILLIAMS, J. P., SLADE, S. E., THALASSINOS, K., BATEMAN, R. H., BOWERS, M. T. & SCRIVENS, J. H. 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry,* 261**,** 1-12.

QI, H., LU, H., QIU, H.-J., PETRENKO, V. & LIU, A. 2012. Phagemid Vectors for Phage Display: Properties, Characteristics and Construction. *Journal of Molecular Biology,* 417**,** 129-143.

QIN, J. & CHAIT, B. T. 1995. Preferential Fragmentation of Protonated Gas-Phase Peptide Ions Adjacent to Acidic Amino Acid Residues. *Journal of the American Chemical Society,* 117**,** 5411-5412.

QIU, X.-Q., WANG, H., CAI, B., WANG, L.-L. & YUE, S.-T. 2007. Small antibody mimetics comprising two complementarity-determining regions and a framework region for tumor targeting. *Nature Biotechnology,* 25**,** 921-929.

RAJ, S. B., RAMASWAMY, S. & PLAPP, B. V. 2014. Yeast alcohol dehydrogenase structure and catalysis. *Biochemistry,* 53**,** 5791-5803.

RAPPSILBER, J. 2011. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of Structural Biology,* 173**,** 530-540.

RATANJI, K. D., DERRICK, J. P., DEARMAN, R. J. & KIMBER, I. 2014. Immunogenicity of therapeutic proteins: influence of aggregation. *Journal of Immunotoxicology,* 11**,** 99-109.

RENAUT, L., MONNET, C., DUBREUIL, O., ZAKI, O., CROZET, F., BOUAYADI, K., KHARRAT, H. & MONDON, P. 2012. Affinity maturation of antibodies: optimized methods to generate high-quality ScFv libraries and isolate IgG candidates by high-throughput screening. *Methods in Molecular Biology,* 907**,** 451-61.

REYNAUD, E. 2010. Protein misfolding and degenerative diseases. *Nature Education,* 3**,** 28.

ROBINSON, C. V., CHUNG, E. W., KRAGELUND, B. B., KNUDSEN, J., APLIN, R. T., POULSEN, F. M. & DOBSON, C. M. 1996. Probing the nature of noncovalent interactions by mass spectrometry. A study of protein-CoA ligand binding and assembly. *Journal of the American Chemical Society,* 118**,** 8646-8653.

ROEPSTORFF, P. & FOHLMAN, J. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry,* 11**,** 601.

ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology,* 5**,** 172.

ROSENBERG, A. S. 2006. Effects of protein aggregates: An immunologic perspective. *The AAPS Journal,* 8**,** E501-E507.

ROSS, C. A. & POIRIER, M. A. 2004. Protein aggregation and neurodegenerative disease. *Nature Medicine,* 10**,** S10-S17.

RUOTOLO, B. T., BENESCH, J. L., SANDERCOCK, A. M., HYUNG, S. J. & ROBINSON, C. V. 2008. Ion mobility-mass spectrometry analysis of large protein complexes. *Nature Protocols,* 3**,** 1139-52.

RUOTOLO, B. T., GILES, K., CAMPUZANO, I., SANDERCOCK, A. M., BATEMAN, R. H. & ROBINSON, C. V. 2005. Evidence for macromolecular protein rings in the absence of bulk water. *Science,* 310**,** 1658-61.

SALUJA, A. & KALONIA, D. S. 2008. Nature and consequences of protein-protein interactions in high protein concentration solutions. *International Journal of Pharmaceutics,* 358**,** 1-15.

SANDIN, S., OFVERSTEDT, L. G., WIKSTROM, A. C., WRANGE, O. & SKOGLUND, U. 2004. Structure and flexibility of individual immunoglobulin G molecules in solution. *Structure,* 12**,** 409-15.

SAPHIRE, E. O., PARREN, P. W., PANTOPHLET, R., ZWICK, M. B., MORRIS, G. M., RUDD, P. M., DWEK, R. A., STANFIELD, R. L., BURTON, D. R. & WILSON, I. A. 2001. Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science,* 293**,** 1155-9.

SAUNDERS, J. C. 2014. *An In Vivo Platform for Identifying Protein Aggregation Inhibitors.* University of Leeds.

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**,** 94-101.

SCARFF, C. A., THALASSINOS, K., HILTON, G. R. & SCRIVENS, J. H. 2008. Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Communications in Mass Spectrometry,* 22**,** 3297-3304.

SCHIRRMANN, T., MEYER, T., SCHÜTTE, M., FRENZEL, A. & HUST, M. 2011. Phage display for the generation of antibodies for proteome research, diagnostics and therapy. *Molecules,* 16**,** 412-426.

SCHMIDT, C. & ROBINSON, C. V. 2014. A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nature Protocols,* 9**,** 2224-36.

SCHRAVEN, B., SAMSTAG, Y., ALTEVOGT, P. & MEUER, S. C. 1990. Association of CD2 and CD45 on human T lymphocytes. *Nature,* 345**,** 71-74.

SCHUCK, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical Journal,* 78**,** 1606-1619.

SCHWIETERS, C. D., KUSZEWSKI, J. J., TJANDRA, N. & CLORE, G. M. 2003. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance,* 160**,** 65-73.

SCLAVI, B., SULLIVAN, M., CHANCE, M. R., BRENOWITZ, M. & WOODSON, S. A. 1998. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science,* 279**,** 1940-1943.

SEEBACHER, J., MALLICK, P., ZHANG, N., EDDES, J. S., AEBERSOLD, R. & GELB, M. H. 2006. Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. *Journal of Proteome Research,* 5**,** 2270-2282.

SHARON, M. 2010. How Far Can We Go with Structural Mass Spectrometry of Protein Complexes? *Journal of the American Society for Mass Spectrometry,* 21**,** 487-500.

SHEPHERD, D. A. 2012. *Characterisation of Virus Structural Proteins by Noncovalent Electrospray Mass Spectrometry.* University of Leeds.

SHEPHERD, D. A., HOLMES, K., ROWLANDS, D. J., STONEHOUSE, N. J. & ASHCROFT, A. E. 2013. Using Ion Mobility Spectrometry–Mass Spectrometry to Decipher the Conformational and Assembly Characteristics of the Hepatitis B Capsid Protein. *Biophysical Journal,* 105**,** 1258-1267.

SHEVCHENKO, A., WILM, M., VORM, O. & MANN, M. 1996. Mass Spectrometric Sequencing of Proteins Silver-Stained Polyacrylamide Gels. *Analytical Chemistry,* 68**,** 850-858.

SHVARTSBURG 1996. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters,* 261**,** 6.

SHVARTSBURG, A. A. & JARROLD, M. F. 1996. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters,* 261**,** 86-91.

SINGH, P., PANCHAUD, A. & GOODLETT, D. R. 2010. Chemical Cross-Linking and Mass Spectrometry As a Low-Resolution Protein Structure Determination Technique. *Analytical Chemistry,* 82**,** 2636-2642.

SINZ, A. 2006. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrometry Reviews,* 25**,** 663-682.

SINZ, A. 2010. Investigation of protein–protein interactions in living cells by chemical crosslinking and mass spectrometry. *Analytical and Bioanalytical Chemistry,* 397**,** 3433-3440.

SINZ, A., ARLT, C., CHOREV, D. & SHARON, M. 2015. Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. *Protein Science,* 24**,** 1193-1209.

SIPE, J. D., BENSON, M. D., BUXBAUM, J. N., IKEDA, S., MERLINI, G., SARAIVA, M. J. & WESTERMARK, P. 2012. Amyloid fibril protein nomenclature: 2012 recommendations from the Nomenclature Committee of the International Society of Amyloidosis. *Amyloid,* 19**,** 167-70.

SMITH, D. P., KNAPMAN, T. W., CAMPUZANO, I., MALHAM, R. W., BERRYMAN, J. T., RADFORD, S. E. & ASHCROFT, A. E. 2009. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *European Journal of Mass Spectrometry,* 15**,** 113-30.

SMITH, D. P., RADFORD, S. E. & ASHCROFT, A. E. 2010. Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 6794-8.

SMITH, G. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science,* 228**,** 1315-1317.

SMITH, S. L. 1996. Ten years of Orthoclone OKT3 (muromonab-CD3): a review. *Progress in Transplantation,* 6**,** 109-121.

SMITH, T. J., PETERSON, P. E., SCHMIDT, T., FANG, J. & STANLEY, C. A. 2001. Structures of bovine glutamate dehydrogenase complexes elucidate the mechanism of purine regulation. *Journal of Molecular Biology,* 307**,** 707-720.

SNIJDER, J., ROSE, R. J., VEESLER, D., JOHNSON, J. E. & HECK, A. J. R. 2013. Studying 18 Mega Dalton Virus Assemblies with Native Mass Spectrometry(). *Angewandte Chemie (International ed. in English),* 52**,** 4020-4023.

STEINBERG, M. Z., ELBER, R., MCLAFFERTY, F. W., GERBER, R. B. & BREUKER, K. 2008. Early Structural Evolution of Native Cytochrome c after Solvent Removal. *Chembiochem : a European journal of chemical biology,* 9**,** 2417-2423.

STEPHENS, W. A pulsed mass spectrometer with time dispersion. Physical Review, 1946. American Physical Society, One Physics Ellipse, College PK, MD 20740-3844 USA, 691-691.

STEVE, J. & CHAKRABARTI, A. 2013. Analysis of Monoclonal Antibody and Protein Aggregates Induced by Denaturation Using a Novel Size Exclusion Chromatography Column. *Chromatography Online: Special Issues,* 31.

STEWART, A., LIU, Y. & LAI, J. R. 2012. A strategy for phage display selection of functional domain-exchanged immunoglobulin scaffolds with high affinity for glycan targets. *Journal of Immunological Methods,* 376**,** 150-5.

STONEY, G. J. 1894. XLIX. Of the “electron,” or atom of electricity: To the editors of the Philosophical Magazine. *Philosophical Magazine,* 38**,** 1.

SU, D., DELAPLANE, S., LUO, M., REMPEL, D. L., VU, B., KELLEY, M. R., GROSS, M. L. & GEORGIADIS, M. M. 2011. Interactions of APE1 with a redox inhibitor: Evidence for an alternate conformation of the enzyme. *Biochemistry,* 50**,** 82-92.

SUZUKI, K., ITO, S., SHIMIZU-IBUKA, A. & SAKAI, H. 2008. Crystal structure of pyruvate kinase from Geobacillus stearothermophilus. *Journal of biochemistry,* 144**,** 305-312.

SYKA, J. E., COON, J. J., SCHROEDER, M. J., SHABANOWITZ, J. & HUNT, D. F. 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 9528-9533.

TAFLIN, D. C., WARD, T. L. & DAVIS, E. J. 1989. Electrified droplet fission and the Rayleigh limit. *Langmuir,* 5**,** 376-384.

TAKAMOTO, K. & CHANCE, M. R. 2006. Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annual Review of Biophysics and Biomolecular Structure,* 35**,** 251-276.

TANAKA, K., WAKI, H., IDO, Y., AKITA, S., YOSHIDA, Y., YOSHIDA, T. & MATSUO, T. 1988. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry,* 2**,** 151-153.

TAYLOR, G. Disintegration of water drops in an electric field. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1964. The Royal Society, 383-397.

THOMSON, J. J. 1897. XL. Cathode rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* 44**,** 293-316.

THOMSON, J. J. 1913. Rays of positive electricity. *Proceedings of the Royal Society of London.,* 89**,** 1-20.

TIAN, Y., HAN, L., BUCKNER, A. C. & RUOTOLO, B. T. 2015. Collision Induced Unfolding of Intact Antibodies: Rapid Characterization of Disulfide Bonding Patterns, Glycosylation, and Structures. *Analytical Chemistry,* 87**,** 11509-15.

TIEDE, C., TANG, A. A. S., DEACON, S. E., MANDAL, U., NETTLESHIP, J. E., OWEN, R. L., GEORGE, S. E., HARRISON, D. J., OWENS, R. J., TOMLINSON, D. C. & MCPHERSON, M. J. 2014. Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. *Protein Engineering Design and Selection,* 27**,** 145-155.

TOLMACHEV, A. V., CLOWERS, B. H., BELOV, M. E. & SMITH, R. D. 2009. Coulombic effects in ion mobility spectrometry. *Analytical Chemistry,* 81**,** 4778-4787.

TOMITA, K., ISHITANI, R., FUKAI, S. & NUREKI, O. 2006. Complete crystallographic analysis of the dynamics of CCA sequence addition. *Nature,* 443**,** 956-60.

TRAN, B. Q., GOODLETT, D. R. & GOO, Y. A. 2016. Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1864**,** 123-129.

UETRECHT, C., BARBU, I. M., SHOEMAKER, G. K., VAN DUIJN, E. & HECK, A. J. 2011a. Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nature Chemistry,* 3**,** 126-32.

UETRECHT, C., BARBU, I. M., SHOEMAKER, G. K., VAN DUIJN, E. & HECKALBERT, J. R. 2011b. Interrogating viral capsid assembly with ion mobility–mass spectrometry. *Nature Chemistry,* 3**,** 126-132.

UETRECHT, C., ROSE, R. J., VAN DUIJN, E., LORENZEN, K. & HECK, A. J. R. 2010. Ion mobility mass spectrometry of proteins and protein assemblies. *Chemical Society Reviews,* 39**,** 1633-1655.

VAN BERKEL, W. J., VAN DEN HEUVEL, R. H., VERSLUIS, C. & HECK, A. J. 2000. Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein Science,* 9**,** 435-439.

VAN DIJK, M. A. & VAN DE WINKEL, J. G. 2001. Human antibodies as next generation therapeutics. *Current Opinion in Chemical Biology,* 5**,** 368-374.

VAN DUIJN, E., BARENDREGT, A., SYNOWSKY, S., VERSLUIS, C. & HECK, A. J. 2009. Chaperonin complexes monitored by ion mobility mass spectrometry. *Journal of the American Chemical Society,* 131**,** 1452-9.

VAN ZUNDERT, G. C. P., RODRIGUES, J. P. G. L. M., TRELLET, M., SCHMITZ, C., KASTRITIS, P. L., KARACA, E., MELQUIOND, A. S. J., VAN DIJK, M., DE VRIES, S. J. & BONVIN, A. M. J. J. 2016. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology,* 428**,** 720-725.

VANDERMARLIERE, E., STES, E., GEVAERT, K. & MARTENS, L. 2015. Resolution of protein structure by mass spectrometry. *Mass Spectrometry Reviews***,** 1-13.

VASQUEZ-REY, M. & LANG, D. A. 2011. Aggregates in Monoclonal Antibody Manufacturing Processes. *Biotechnology and Bioengineering,* 108**,** 1494-1508.

VEENSTRA, T. D. 1999. Electrospray ionization mass spectrometry in the study of biomolecular non-covalent interactions. *Biophysical Chemistry,* 79**,** 63-79.

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,* 100**,** 14817-14821.

VOYNOV, V., CHENNAMSETTY, N., KAYSER, V., HELK, B. & TROUT, B. L. 2009. Predictive tools for stabilization of therapeutic proteins. *MAbs,* 1**,** 580-582.

WALTON, E. B. & VANVLIET, K. J. 2006. Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Physical Review E,* 74**,** 061901.

WALZTHOENI, T., JOACHIMIAK, L. A., ROSENBERGER, G., RÖST, H. L., MALMSTRÖM, L., LEITNER, A., FRYDMAN, J. & AEBERSOLD, R. 2015. xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nature Methods,* 12**,** 1185-1190.

WANG, G., JOHNSON, A. J. & KALTASHOV, I. A. 2012. Evaluation of Electrospray Ionization Mass Spectrometry as a Tool for Characterization of Small Soluble Protein Aggregates. *Analytical Chemistry,* 84**,** 1718-1724.

WANG, S. C., POLITIS, A., DI BARTOLO, N., BAVRO, V. N., TUCKER, S. J., BOOTH, P. J., BARRERA, N. P. & ROBINSON, C. V. 2010. Ion Mobility Mass Spectrometry of Two Tetrameric Membrane Protein Complexes Reveals Compact Structures and Differences in Stability and Packing. *Journal of the American Chemical Society,* 132**,** 15468-15470.

WANG, W. 2005. Protein aggregation and its inhibition in biopharmaceutics. *International Journal of Pharmaceutics,* 289**,** 1-30.

WANG, X., DAS, T. K., SINGH, S. K. & KUMAR, S. 2009. Potential aggregation prone regions in biotherapeutics: A survey of commercial monoclonal antibodies. *MAbs,* 1**,** 254-67.

WARD, E. S., GUSSOW, D., GRIFFITHS, A. D., JONES, P. T. & WINTER, G. 1989. Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. *Nature,* 341**,** 544-546.

WATKINSON, T. G., CALABRESE, A. N., GIUSTI, F., ZOONENS, M., RADFORD, S. E. & ASHCROFT, A. E. 2015. Systematic analysis of the use of amphipathic polymers for studies of outer membrane proteins using mass spectrometry. *International Journal of Mass Spectrometry,* 391**,** 54-61.

WEISS, A., KLEIN, C., WOODMAN, B., SATHASIVAM, K., BIBEL, M., REGULIER, E., BATES, G. P. & PAGANETTI, P. 2008. Sensitive biochemical aggregate detection reveals aggregation onset before symptom development in cellular and murine models of Huntington's disease. *Journal of Neurochemistry,* 104**,** 846-58.

WHITEHOUSE, C. M., DREYER, R., YAMASHITA, M. & FENN, J. 1989. Electrospray ionization for mass-spectrometry of large biomolecules. *Science,* 246**,** 64-71.

WILEY, W. & MCLAREN, I. H. 1955. Time-of-flight mass spectrometer with improved resolution. *Review of Scientific Instruments,* 26**,** 1150-1157.

WILM, M. & MANN, M. 1996. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry,* 68**,** 1-8.

WISHART, D. S. 2013. Characterization of biopharmaceuticals by NMR spectroscopy. *TrAC Trends in Analytical Chemistry,* 48**,** 96-111.

WOODS, L. A., RADFORD, S. E. & ASHCROFT, A. E. 2013. Advances in ion mobility spectrometry-mass spectrometry reveal key insights into amyloid assembly. *Biochimica et Biophysica Acta,* 1834**,** 1257-68.

WORN, A. & PLUCKTHUN, A. 2001. Stability engineering of antibody single-chain Fv fragments. *Journal of Molecular Biology,* 305**,** 989-1010.

WU, C., SIEMS, W. F., ASBURY, G. R. & HILL, H. H. 1998. Electrospray ionization high-resolution ion mobility spectrometry-mass spectrometry. *Analytical Chemistry,* 70**,** 4929-4938.

WU, H., KROE-BARRETT, R., SINGH, S., ROBINSON, A. S. & ROBERTS, C. J. 2014. Competing aggregation pathways for monoclonal antibodies. *FEBS letters,* 588**,** 936-41.

WYSOCKI, V. H., JONES, C. M., GALHENA, A. S. & BLACKWELL, A. E. 2008a. Surface-induced dissociation shows potential to be more informative than collision-induced dissociation for structural studies of large systems. *Journal of the American Society for Mass Spectrometry,* 19**,** 903-913.

WYSOCKI, V. H., JOYCE, K. E., JONES, C. M. & BEARDSLEY, R. L. 2008b. Surface-induced dissociation of small molecules, peptides, and non-covalent protein complexes. *Journal of the American Society for Mass Spectrometry,* 19**,** 190-208.

WYTTENBACH, T. & BOWERS, M. T. 2011. Structural stability from solution to the gas phase: native solution structure of ubiquitin survives analysis in a solvent-free ion mobility–mass spectrometry environment. *The Journal of Physical Chemistry B,* 115**,** 12266-12275.

XU, G. & CHANCE, M. R. 2007. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical Reviews,* 107**,** 3514-3543.

XU, H. & FREITAS, M. A. 2007. A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. *BMC Bioinformatics,* 8**,** 1-10.

XU, H., HSU, P.-H., ZHANG, L., TSAI, M.-D. & FREITAS, M. A. 2010. Database Search Algorithm for Identification of Intact Cross-Links in Proteins and Peptides Using Tandem Mass Spectrometry. *Journal of Proteome Research,* 9**,** 3384-3393.

YAN, Y., CHEN, G., WEI, H., HUANG, R. Y.-C., MO, J., REMPEL, D. L., TYMIAK, A. A. & GROSS, M. L. 2014a. Fast photochemical oxidation of proteins (FPOP) maps the epitope of EGFR binding to adnectin. *Journal of the American Society for Mass Spectrometry,* 25**,** 2084-2092.

YAN, Y., CHEN, G., WEI, H., HUANG, R. Y., MO, J., REMPEL, D. L., TYMIAK, A. A. & GROSS, M. L. 2014b. Fast photochemical oxidation of proteins (FPOP) maps the epitope of EGFR binding to adnectin. *Journal of the American Society for Mass Spectrometry,* 25**,** 2084-92.

YOKOTA, T., MILENIC, D. E., WHITLOW, M. & SCHLOM, J. 1992. Rapid Tumor Penetration of a Single-Chain Fv and Comparison with Other Immunoglobulin Forms. *Cancer Research,* 52**,** 3402-3408.

YOUNG, L. M., CAO, P., RALEIGH, D. P., ASHCROFT, A. E. & RADFORD, S. E. 2014a. Ion mobility spectrometry-mass spectrometry defines the oligomeric intermediates in amylin amyloid formation and the mode of action of inhibitors. *Journal of the American Chemical Society,* 136**,** 660-70.

YOUNG, L. M., CAO, P., RALEIGH, D. P., ASHCROFT, A. E. & RADFORD, S. E. 2014b. Ion Mobility Spectrometry–Mass Spectrometry Defines the Oligomeric Intermediates in Amylin Amyloid Formation and the Mode of Action of Inhibitors. *Journal of the American Chemical Society,* 136**,** 660-670.

YU, C., MAO, H., NOVITSKY, E. J., TANG, X., RYCHNOVSKY, S. D., ZHENG, N. & HUANG, L. 2015. Gln40 deamidation blocks structural reconfiguration and activation of SCF ubiquitin ligase complex by Nedd8. *Nature Communications,* 6.

ZHANG, H., CUI, W. & GROSS, M. L. 2014. Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS letters,* 588**,** 308-317.

ZHANG, H., GAU, B. C., JONES, L. M., VIDAVSKY, I. & GROSS, M. L. 2010. Fast photochemical oxidation of proteins for comparing structures of protein− ligand complexes: the calmodulin− peptide model system. *Analytical Chemistry,* 83**,** 311-318.

ZHANG, H., GAU, B. C., JONES, L. M., VIDAVSKY, I. & GROSS, M. L. 2011. Fast Photochemical Oxidation of Proteins (FPOP) for Comparing Structures of Protein/Ligand Complexes: The Calmodulin-peptide Model System. *Analytical Chemistry,* 83**,** 311-318.

ZHANG, J., XIN, L., SHAN, B., CHEN, W., XIE, M., YUEN, D., ZHANG, W., ZHANG, Z., LAJOIE, G. A. & MA, B. 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics,* 11**,** 1-8.

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 (Aβ) peptide aggregation. *Proceedings of the National Academy of Sciences,* 110**,** 14604-14609.

ZHENG, Q., ZHANG, H., WU, S. & CHEN, H. 2016. Probing Protein 3D Structures and Conformational Changes Using Electrochemistry-Assisted Isotope Labeling Cross-Linking Mass Spectrometry. *Journal of the American Society for Mass Spectrometry,* 27**,** 864-875.

ZHONG, Y., HYUNG, S.-J. & RUOTOLO, B. T. 2012. Ion mobility–mass spectrometry for structural proteomics. *Expert Review of Proteomics,* 9**,** 47-58.

ZHOU, M., JONES, C. M. & WYSOCKI, V. H. 2013. Dissecting the large noncovalent protein complex GroEL with surface-induced dissociation and ion mobility–mass spectrometry. *Analytical Chemistry,* 85**,** 8262-8267.

ZHOU, M., POLITIS, A., DAVIES, R. B., LIKO, I., WU, K.-J., STEWART, A. G., STOCK, D. & ROBINSON, C. V. 2014. Ion mobility–mass spectrometry of a rotary ATPase reveals ATP-induced reduction in conformational flexibility. *Nature Chemistry,* 6**,** 208-215.

ZHOU, M., SANDERCOCK, A. M., FRASER, C. S., RIDLOVA, G., STEPHENS, E., SCHENAUER, M. R., YOKOI-FONG, T., BARSKY, D., LEARY, J. A. & HERSHEY, J. W. 2008. Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proceedings of the National Academy of Sciences,* 105**,** 18139-18144.

ZHOU, Y. & VACHET, R. W. 2012. Diethylpyrocarbonate Labeling for the Structural Analysis of Proteins: Label Scrambling in Solution and How to Avoid It. *Journal of the American Society for Mass Spectrometry,* 23**,** 899-907.

ZNAMENSKIY, V., MARGINEAN, I. & VERTES, A. 2003. Solvated ion evaporation from charged water nanodroplets. *The Journal of Physical Chemistry A,* 107**,** 7406-7412.

ZUBAREV, R. A., KELLEHER, N. L. & MCLAFFERTY, F. W. 1998. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society,* 120**,** 3265-3266.

ZURDO, J. 2013. Surviving the Valley of Death. *European Biopharmaceutical Review*.

ZURDO, J., MICHAEL, R., STALLWOOD, Y., HENDMAN, K. & AASTRUP, T. 2011. Improving the Developability of Biopharmaceuticals. *Innopharma Pharmaceutical Technology***,** 34-40.