A Study of Mass Spectrometry-based Footprinting Techniques for the Determination of Higher-order Structure in Biological Systems.

Jake Busuttil-Goodfellow

School of Molecular and Cellular Biology

Astbury Centre for Structural Molecular Biology

University of Leeds

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

March 2023

Declaration

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.

© 2023 The University of Leeds and Jake Busuttil-Goodfellow

Covid-19 impact statement

I have experienced a significant delay in my research due to:

1) Lack of access to the laboratory due to covid and significant lack of access to instruments following phased reopening of the laboratory did not allow me to undertake any experimental work for the entire covid closure period and very limited experimental work for several months following on from reopening due to limited laboratory time. Given my project is entirely laboratory based this allowed for no work to be performed.
2) A 9 month delay in the delivery of krypton fluoride gas for our FPOP laser due to covid. This allowed me to undertake no FPOP experiments, which is one of two experimental methods I solely used for my PhD. Given access to the other method (HDX) is significantly limited, this put a giant restriction on my ability to perform these experiments even after the covid reopening.

Acknowledgements

I would like to thank my supervisors, Frank and James, for their ongoing support throughout a very unpredictable few years. With the incidence of the COVID-19 pandemic, many plans and opportunities were changed or otherwise impacted, and their ongoing support to get through and help with finding alternative and meaningful work during the pandemic.

I would also like to thank Alison Ashcroft for initially bringing me to Leeds and her ongoing support throughout the years.

I'd like to thank my supervisors from Croda inc, especially Steve Rumbelow, for his advice and direction, and helping me with his wealth of knowledge on polysorbates.

I'd also like to thank Lars, Anna and Marcus for their help and advice with the MtrC work, and Owen Cornwell, members of the Sobott group and Calabrese group for helpful discussions and direction with the other chapters.

Table of Contents

Figures

Tables

Equations

Abbreviations

API - Active Pharmaceutical ingredients

CCS – collisional cross section

CID – collision induced dissociation

CMC – Critical Micelle Concentration

Da – Daltons

DC – direct current

DDA – data dependant acquisition

DIA - data independent acquisition

DTT – dithiothreitol

ECD – electron capture dissociation

ESI – electrospray ionisation

ETD – electron transfer dissociation

Fab – fragment antigen binding region

Fc – fragment crystallisable region

FPOP – fast photochemical ionisation of proteins

FTICR – Fourier-transform ion cyclotron resonance

FT-IR – fourier transform infrared spectroscopy

FWHM – full width half maximum

GC – gas chromatography

HILIC – Hydrophobic interaction liquid chromatography.

HDX – Hydrogen deuterium exchange

IMS – ion mobility spectrometry

IEX Ion-exchange chromatography

LC – liquid chromatography

m/z – mass to charge ratio

mAb – monoclonal antibody

MALDI – matrix-assisted laser absorption ionisation

MS – Mass spectrometry

MW – Molecular weight

NMR – nuclear magnetic resonance

NPF – normalised protection factor

OM – outer membrane

PEG – polyethylene glycol

POE – sorbitan polyoxyethylates

PS – Polysorbate

RF – radio frequency

SASA – solvent accessible surface area

SBSE-GC – stir-bar assisted sorptive extraction gas chromatography

TEAB – tetraethyl ammonium bromide

TIC- total ion count

TOF – time of flight

TWIMS – travelling wave IMS

XF – X-ray footprinting

XIC – extracted ion chromatogram

Abstract

Protein footprinting is a technique used to determine changes in higher order structure in proteins. Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS), Fast photochemical oxidation of proteins (FPOP) and synchrotron footprinting are common footprinting methods utilising different mechanisms to achieve elucidation of higher order structure. This thesis aims to probe the functionality of these methods, and to improve understanding of the mechanisms of each approach. HDX-MS is used to probe the action of polysorbates on antibodies, the effect of local residues is probed in FPOP, and the viability of synchrotron footprinting is probed for binding of TiO_2 to MtrC.

Polysorbates are commonly used excipients in pharmaceutical formulations, utilised to increase stability of active pharmaceutical ingredients (APIs). Despite this, the molecular mechanism of their stabilizing effect is not well described in literature. When used above their critical micelle concentration (CMC), they can protect against aggregation, in particular shielding hydrophobic patches of proteins. In protein formulations, polysorbates 20 and 80 are typically above the polysorbate CMC.

Here, using a common and abundant model protein, myoglobin, as well as a set of monoclonal antibodies with varying tendencies to aggregate, the principle of this protective effect in molecular detail is examined. Myoglobin was selected as a model protein as it is well characterised and not aggregation prone. Experiments are carried out with polysorbate excipient added both above and below the CMC, and long term (storage) stability and the extent of possible oxidative damage are also investigated.

Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) is a structural mass spectrometry technique that utilises the exchange of solvent accessible backbone hydrogens to determine

changes in protein structure, through differences in the uptake of deuterium based on solvent accessibility. These can be used to infer structural changes within the protein between states. Further, as access to solvent accessibility is affected by protein binding, HDX-MS can be utilised to map binding sites for both protein-protein interactions and binding with synthetic molecules.

By utilising HDX-MS, structural changes on proteins due to polysorbate exposure can be determined, and regions of protection and deprotection identified.

Initially, myoglobin samples with and without polysorbates were analysed without incubation by HDX to determine the veracity of the method for polysorbate analysis. This analysis showed no difference in protein structure with the addition of fresh polysorbate 20 and 80. This suggests polysorbate does not immediately perturb structure on addition to myoglobin both locally and globally, which is anticipated with a stable protein such as myoglobin.

Investigations of the antibodies WFL and STT, with WFL being more aggregation prone than STT, showed polysorbate 80 perturbs the structure of WFL in comparison to polysorbate 20 and the control. This is likely due to polysorbate 80 having high affinity to hydrophobic pockets on the fc domain of WFL.

FPOP is another structural mass spectrometry technique that allows for determination of conformational changes, similar to HDX. However, the effect of the local environment on labelling has not been rigorously scrutinised. Here, I look at six peptides, with varying distance between tryptophan and phenylalanine residues, and determined the total oxidation of each peptide.

Here, we see for the first time evidence to suggest fundamentally that as two highly oxidisable residues are spatially located closer, total oxidation decreases. It is hypothesised this is due to competition between two reactive residues for a limited number of hydroxyl radicals. As the distance increased between the residues, it is proposed the local competition between residues for

the limited number of hydroxyl radicals is decreased due to spatial distance, and thus more labelling can occur at each reactive residue.

Finally, MtrC binding with TiO_2 is interrogated through synchrotron footprinting. Here for the first time I determine that the location of binding of TiO_2 to MtrC is primarily due to electrostatic interactions, as opposed to chemical binding.

This provides a solution for rational design of light harvesting particles, and how to determine binding of these molecules for future molecular battery design.

Thus, each footprinting method is interrogated and understanding of each method has been advanced within this thesis.

Chapter 1:

Introduction

1. Introduction

An excipient, as defined in the European pharmacopeia, is a component that has no active effect pharmaceutically, but is present within the drug product (European Directorate for the Quality of Medicines 2002). These components act, alone or in conjunction with other components, as carriers of the Active pharmaceutical ingredient (API) functioning to increase stability, biopharmaceutical profile (the adsorption, distribution, metabolism and excretion (ADME)) , appearance and ease of API manufacture (European Directorate for the Quality of Medicines 2002).

Protein therapeutics are the fastest growing class of therapeutic drugs (Leurs et al., 2015), and encompass multiple classes from peptide drugs such as glucagon-like peptides, with molecular weights (MW) in the region of 3 kDa (de Graaf et al., 2016) to larger proteins such as asparaginase, MW 32 kDa (Lagassé et al., 2017), and monoclonal antibody therapies such as obiltoxaximab, 146 kDa (Lagassé et al., 2017). These therapies, whilst providing novel treatment opportunities, have a number of problems regarding their usage as standalone APIs. Proteins can unfold and aggregate during production, storage and transport (Wang, 1999; Brader et al., 2015). Lack of structural robustness, and thus unfolding and aggregation, can lead to a lack of efficacy as a treatment, due to drugs being unable to be absorbed and bind, and in some cases, induction of an immune response (Roberts, 2014). Lack of efficacy may be determined by both patient feedback and medical testing (Roberts, 2014). Aggregation of protein therapeutics has been linked with allergic responses, including anaphylaxis (Lundahl et al., 2021). Given the potential of therapeutic proteins to unfold and aggregate, excipients which can limit or eliminate these processes are of significant value.

1.1 Types of protein excipient

Protein excipients can be divided broadly into two discreet categories: those which increase the stability of the protein in its native state (Arakawa, T. and Timasheff, 1985; Santoro et al., 1992), and excipients which prevent aggregation without affecting protein stability (Arakawa, Tsutomu and Tsumoto, 2003; Mustafi et al., 2008). Common stabilising excipients include sugars, polyols, amino

acids and amines (Arakawa, T. and Timasheff, 1985; Mishra et al., 2005; Yancey, 2005; Ishrat et al., 2018), common non-stabilising excipients include arginine and proteins (Arakawa, Tsutomu et al., 2007; Pivovarova et al., 2007; Yoshizawa et al., 2017) such as heat shock protein 27 (HSP27). This protein yields its anti-aggregation effects by forming complexes with the target protein in solution (Pivovarova et al., 2007) .

Protein aggregation can occur both as a result of the protein unfolding (Vagenende et al., 2009), or through association between native proteins due to molecular crowding (Batra et al., 2009). Therefore, selection of the excipient depends on several factors including the protein, protein concentration, storage temperature, storage pH and desired viscosity.

1.1.1 Protein thermodynamics and kinetic stability

Thermodynamic stability of a protein is as an equilibrium between folded, partially folded and unfolded states, and this is governed by the change in Gibbs free energy required to transition between these states. (Colón et al., 2017). Kinetic stability is energy difference in the activation energy required for unfolding, shown in figure 1.1. When a protein has a high kinetic stability, the amount of free energy required to unfold the protein from its native conformation is higher compared to a lower kinetic stability. The required energy for overcoming either thermodynamic or kinetic stability can be catalysed to reduce this requirement, or increased through the use of additives. Such that, by increasing the activation energy ($\Delta G_U‡$), the protein is effectively forced into it's native state.

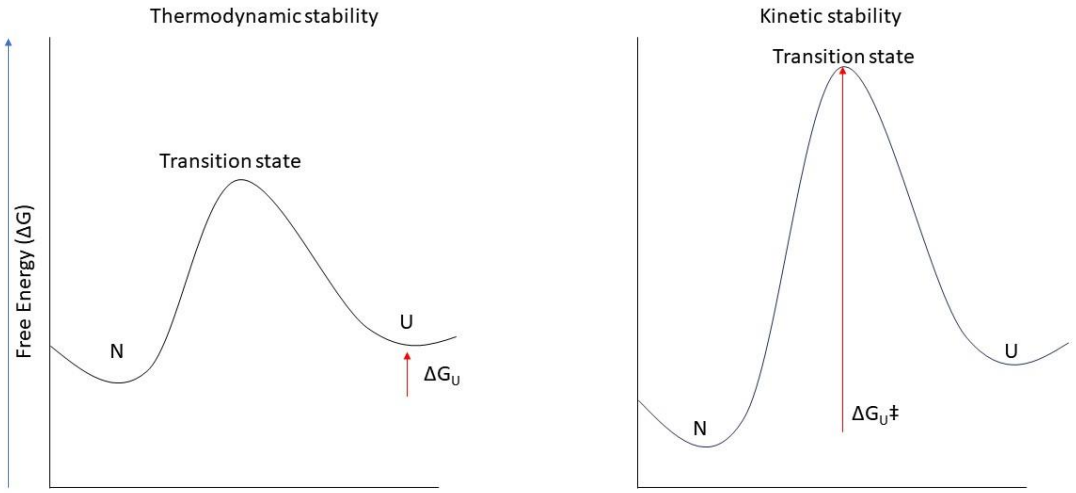


Figure 1.1: differences between thermodynamic stability and kinetic stability adapted from (Colón et al., 2017). N is the native protein, and U is the unfolded protein. ΔG_U is the difference in free energy, whilst $\Delta G_U‡$ is the difference in activation energy for the protein. A high thermodynamic stability in the native state requires more Gibbs free energy to shift the equilibrium from the native state to the unfolded state. A highly kinetic stable protein requires a higher activation energy to shift from the native state to the unfolded state.

1.1.2 Mechanisms of stability

Stabilising excipients have been reported to produce their stabilising effect through preferential exclusion (Schellman, J. A., 1997; Timasheff, Serge N., 1998; Timasheff, Serge N., 2002; Schellman, John A., 2003; Barnett et al., 2016). It is important to note that this implies these additives are not binding to the API, instead, they make activation energy harder to achieve to change conformation from the native state (as shown in figure 1.1) . The primary mechanism of stability from stabilising excipients is the excluded volume effect, however other mechanisms such as cohesive force with water and unfavourable interactions with peptide bonds have also been described (Ohtake et al., 2011).

Thermodynamically, protein folding can be displayed as a folding funnel (figure 1.2), where the folding of a protein into its native structure is explained in terms of entropy (ΔS) and free energy (ΔG). Conformational entropy is defined as the number of possible conformations of a molecule, whilst free energy is defined as the internal free energy of the molecule, encompassing bonds, and hydrophobic and solvation energies.

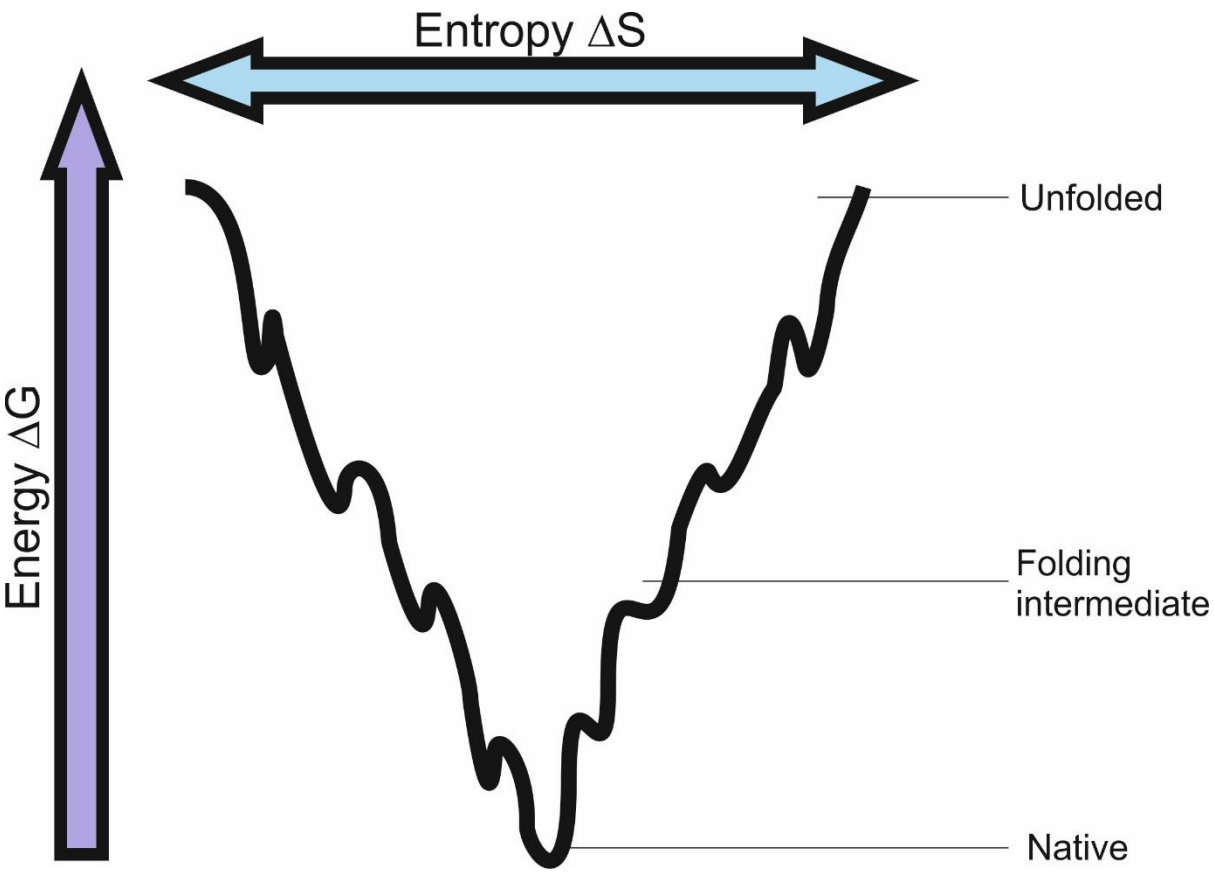


Figure 1.2: Folding funnel, adapted from (Wolynes et al., 1995). A folding funnel is a visualisation of protein folding energy landscape. As free energy in the system decreases, the protein is forced into a native state where free energy is lowest. Entropy is defined as the number of possible conformations of the molecule.

The excluded volume effect enhances protein stability by formation of a layer of water between the protein and the excipient. This discrete layer is formed due to exclusion of the excipient caused by its hydrodynamic radius. This has a significant effect by increasing the activation energy for unfolding

from the native state (Ohtake et al., 2011), meaning the protein is in effect forced to maintain its native structure (Ohtake et al., 2011).

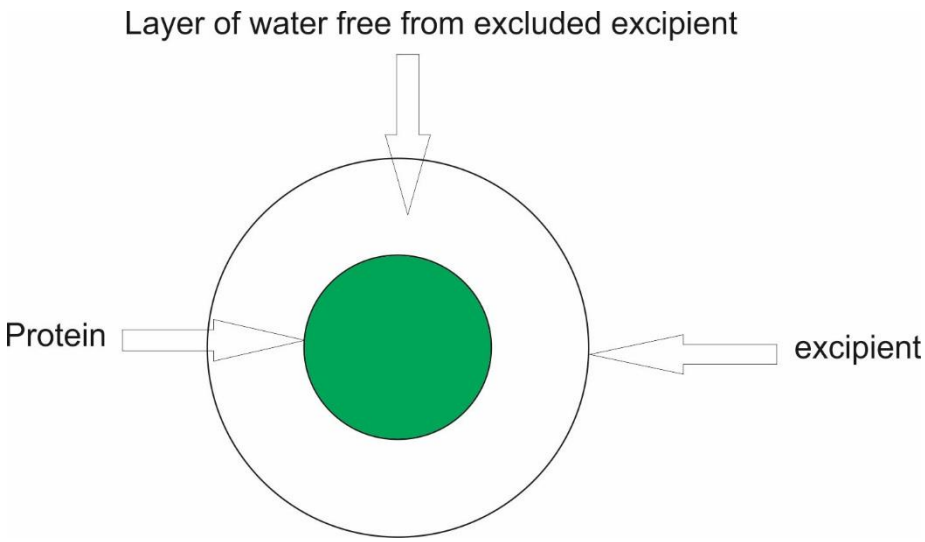


Figure 1.3: Diagram of the excluded volume effect, showing how the repulsive interactions between the excipient and protein cause a layer of water to be formed between both components. Diagram redrawn from (Ohtake et al., 2011). By forming a water layer, hydrophobic residues are folded inside the protein, adopting its native conformation.

Cohesive force was originally termed attraction pressure by Traube in 1909 (Traube, 1909). This concept causes an increase in the surface tension of water by the exertion of a force from the excipients on the water molecules. Whilst Traube did not specify the mechanism of cohesive force, he observed that salts in particular were excluded from the protein surface (Ohtake et al., 2011).

Bolen, along with Yi and Auton, (Liu, Y. and Bolen, 1995; Auton and Bolen, 2004; Auton and Bolen, 2007) identified that protein stabilising additives can yield their effects through unfavourable interactions between the excipient and peptide bonds, which link amino acids together in a protein.

Determination of these mechanisms could be determined through measuring excipient binding, conformational changes, and covalent and non-covalent changes to protein structure.

1.2 Polymers

Polymers are molecules which consist of a number of repeating subunits (Jenkins et al., 1996). A number of polymers have been shown to increase protein stability in solution (Sasahara et al., 2003; Wang et al., 2008; Rawat et al., 2010). Polymers are often used as excipients for other functions such as surfactants (Kishore et al., 2011a). Given the option to increase and decrease the number of repeating units to affect the physical properties of the polymer, polymers offer the option to be tailored for their specific function as an excipient.

Polymers produce their stabilising effects primarily through the molecular crowding effect (Laurent, 1963). This is essentially an extension of the excluded volume effect described above, whereby exclusion from the hydrodynamic radius and repulsion forces between the polymer and the protein force the protein into its native state. Given this polymer exclusion is increased for larger polymers, this provides a solution to the observation that larger polymers generally confer greater stabilising effects (Minton, 2005).

1.2.1 Polyethylene glycol (PEG) (Poly(oxyethylene))

PEGs are polymers with the structure H-[O-CH_2-CH_2]$_n$-OH (Bekale et al., 2015). This structure gives PEGs hydrophilic properties, proving useful for their biopharmaceutical applications. Specifically, this allows PEGs to be water soluble and interact with proteins effectively (Brady et al., 2017). It is important to note however, given the non-polar aliphatic chain component of PEGs, that they interact differently with proteins than entirely hydrophilic polymers (Timasheff, S N, 1993). PEG is utilised in protein formulations in a broad range of sizes, from PEG 200 to upwards of 20,000, with the number following the prefix denoting the average MW of the individual components (Lee and Lee, 1987; Wu et al., 2014). Given this breadth of size, PEGs have different properties reported in literature. Lee and Lee (Lee and Lee, 1987) showed PEG 200, 400, 1000 and 4000 actually decreased the thermal stability of β-lactoglobulin, chymotrypsinogen and lysozyme, especially at

concentrations higher than 20%. This is likely due to the non-polar moiety of PEG interacting preferentially with the unfolded conformation of these proteins (Lee and Lee, 1987).

Figure 1.4: Chemical structure of PEG, with the non polar moiety shown as the two $(CH_2)_2$ groups within the repeating $O(CH_2)_2$ unit

1.2.2 Polysorbates

Polysorbates (PS) are non-ionic surfactants utilised in drug formulations to reduce aggregation of protein drugs such as monoclonal antibodies (mAb) (Li, Y. et al., 2014). PS are comprised of a hydrophilic poly(oxyethylene) and sugar head, and a fatty acid hydrophobic tail. These chemical groups give PS their desirable properties of protein stability and aggregation prevention by competitive adsorption to interfaces and direct protein binding (Martos et al., 2017). PS 20 and 80 specifically are commonly used to stabilise protein therapeutics in solution. These two molecules differ in the *m* position (figure 3). PS 20 has a lauric acid group in this position whereas PS 80 has an oleic acid group (Kishore et al., 2011a). The oleic acid carbon chain is significantly longer than the lauric acid chain, making PS 80 more hydrophobic than PS 20. Further, from (figure 1.3), the sum of $w + x + y + z$ is equal to 20 (Kishore et al., 2011a) m denotes the particular tail present, defining the identity of the PS.

Commercial sources of PS are provided as mixtures containing a number of polymers including PEG, Sorbitan polyoxyethylates (POE), and sorbitan POE fatty acid esters (Ayorinde et al., 2000; Frison-Norrie and Sporns, 2001). This is due to PS synthesis originating from sorbitol and then sorbitan esters. Sorbitan esters can form alternative synthesis products with a similar structure to polysorbates that are present in the final product (Frison-Norrie and Sporns, 2001).

Polysorbates have been widely used as protein stabilising agents. This is attributed to their high hydrophile-lypophile balance numbers and contrastingly their low critical micelle concentration values (Kishore et al., 2011b). This gives PS effective surface activity at relatively low concentrations (Kishore et al., 2011b), which provides their protein stabilising effects.

Hydrophile-lipophile balance is a term used to describe how hydrophilic or lipophilic a molecule is, and is theoretically calculated by looking at how much of the molecule is hydrophilic (by molecular mass), in comparison to the molecular mass of the whole molecule (Griffin, 1954).

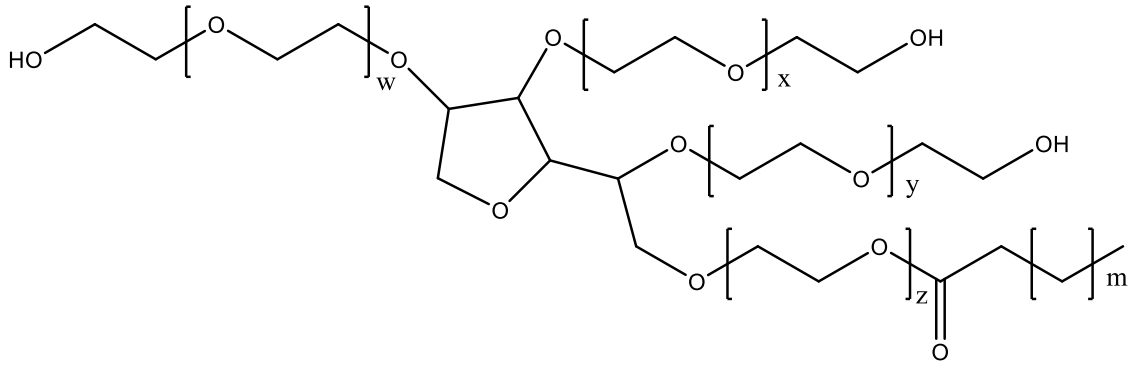


Figure 1.5: Structure of polysorbates, redrawn from (Kishore et al., 2011a). A polysorbate is defined from a fatty acid tail, described by *m* in the diagram, with a number of (poly)oxyethylene groups, *the sum of w + x + y + z equalling 20 for Polysorbate 20.*

Yi et al. (Li, Y. et al., 2014) utilised a 2D liquid chromatography – mass spectrometry (LC-MS) method to interrogate the stability of PS 20 in the presence of mAbs. PS can degrade in aqueous solution due to hydrolysis of the fatty acid ester (Martos et al., 2017). The PS used in the Yi et al contained a number of different PS esters in the PS preparation. This experiment showed that whilst all PS esters degraded, different PS esters degraded at different rates. This is important to note as degradation of the polymer may diminish its ability to stabilise the protein, although this has been disputed (Kishore et al., 2011a), as so long as the total concentration of PS remains above a stabilising threshold, degradation may not impact stability. Further, degradation can cause the solution to appear cloudy, which is undesirable for medical professionals as they cannot easily distinguish between discolouration due to aggregation and discolouration due to fatty acid ester hydrolysis**.**

Further, PS can also break down due to auto-oxidation, and this degradation is more prevalent than hydrolysis at common pharmaceutical storage conditions (Kishore et al., 2011a). Oxygen forms peroxide radicals, which leads to further radical formation and thus auto-oxidation (Martos et al., 2017).

Kishore et al. (Kishore et al., 2011a) utilised a number of methods including Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS), nuclear magnetic resonance (NMR), FT-IR and stir-bar-assisted sorptive extraction gas chromatography-mass spectrometry (SBSE-GC-MS). Their work investigated the link between degradation of both PS 20 and PS 80 and protein stability. Degradation products including fatty acids and polyoxyethylene esters were identified as a direct product of oxidation and hydrolysis of PS. 36 degradation products were identified from both PS 20 and 80, with a range of logP values between 0.75 and 4.25. LogP is a measure of solubility in two immiscible phases, usually a hydrophobic phase such as octanol and a hydrophilic phase such as water (figure 1.6).

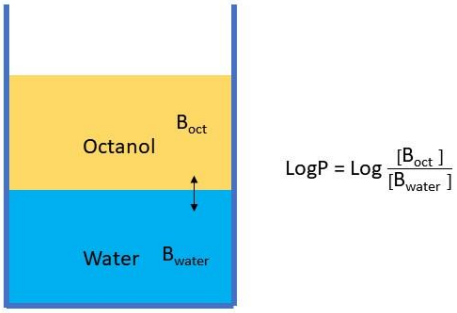


Figure 1.6: LogP calculation redrawn from (Bienta). More lipophilic molecules will have a higher affinity for octanol, and more hydrophilic molecules have an affinity for the water phase. The higher

the logP, the more lipophilic the molecule is, with a positive number being lipophilic and a negative number being hydrophilic

The logP values for PS 20 and 80 indicate a range of solubilities of the degradants, with some of them being poorly soluble in water. This may account for the cloudy appearance observed in some medicines stabilised with PS.

Crucially, Kishore determined that despite PS degradation occurring, the degradants have no effect on protein stability, so long as PS concentration does not drop below the threshold required to stabilise the protein.

Labrenz (Labrenz, 2014) suggested a lipid specific mechanism can hydrolyse oleate esters, however this effect is only observed within PS 80 not PS 20. Therefore Labrenz suggested an alteration to PS 80 molecules to include a nonhydrolysing bond between the head and the tail of the molecule, although this presents issues in manufacturing and regulatory approval.

Dixit and colleagues (Dixit et al., 2016) looked at the degradation of PS 20. Specifically, they looked at PS 20 loss from a sulfatase drug product formulation, and performed this analysis via LC-MS. Further, they studied the effect of common products produced from cell lines and their effect on PS 20 degradation. This included sulfatase, sulfatase inhibitors and phospholipase B-like 2. This study showed these impurities from the drug production process catalysed PS degradation, and thus may affect storage conditions and longevity for certain drugs. Phospholipase B-like 2 in particular appeared to have hydrolytic activity. They also deduced that despite the degradation of PS 20, given the concentration of intact PS 20 is above the threshold required for stability, the degradation did not impact the stabilising qualities of the polymer.

Ha et al. (Ha et al., 2002) analysed peroxide formation in PS 80 samples at differing storage conditions, and the effect on a model protein IL-2 mutein. Peroxide formation from PS 80 is important as it leads to oxidation of the protein which can affect its viability. He et al. analysed both

neat PS 80 samples and 20% PS 80 samples and identified that both produced peroxide radicals, which is attributed to autoxidation and degradation of the PS 80 molecule. However, peroxide formation in polysorbates can be reduced by the addition of an antioxidant (Chang and Bock, 1980). He et al. noted that peroxide formation from PS 80 occurred more quickly at higher temperatures and attributed this to more rapid radical formation.

To summarise, polysorbates can be degraded by two primary methods, hydrolysis and oxidation. This degradation occurs both in PS 20 and PS 80, which are the two commonly used polysorbates for protein stability. Whilst this degradation is well characterised, specifically by Kishore et al. (Kishore et al., 2011a), it does not appear to have a significant effect on either protein stability or protein function. However, some of these degradation products are poorly soluble, which can lead to particulate formation in drug solution vials. This presents a significant issue when attempting to diagnose the viability of the sample. Thus, should the mechanisms of how polysorbates specifically interact with protein molecules be characterised, structural changes to these molecules could be suggested to reduce degradation without affecting solubility and protein stabilising effects.

Thus, MS techniques, which will be described in the following chapters, specifically involving protein footprinting, should provide a format to study complex pharmaceutical mixtures and determine higher order structural changes.

1.3 Protein mass spectrometry

Mass Spectrometry (MS) is a technique which involves the separation of charged species by their mass to charge ratio (m/z). A mass spectrometer has three unique components to achieve this: the ion source, the mass analyser, and the detector (de Hoffmann and Stroobant, 2007). Given proteins are usually defined as peptides with at least 50 amino acids (IUPAC, 1997), their m/z values are higher than small molecules such as drugs, metabolites and lipids. Further, there is a requirement to ionise the molecules without fragmenting the protein. Thus, for protein analysis, soft-ionisation techniques are preferred (Mehmood et al., 2015).

1.3.1 Ionisation sources

Two ionisation methods are often utilised for protein analysis: electrospray ionisation (ESI), and matrix-assisted laser absorption ionisation (MALDI). Whilst both methods are soft ionisation techniques both have very different mechanisms of ionisation.

1.3.2 MALDI

Laser desorption ionisation was first described by Tanaka and colleagues (without a matrix) (Tanaka et al., 1988), with Karas and Hillenkamp specifically outlining MALDI, (Karas and Hillenkamp, 1988) which involves combining a matrix with the analyte of interest, before putting this matrix mixture on a target and irradiation via a pulsed laser (Duncan et al., 2016). This irradiation ablates the matrix and sample, which is acidic in nature to protonate the analyte, and subsequently leads to ionisation of the sample molecule, allowing these ions to be analysed by mass spectrometry (de Hoffmann and Stroobant, 2007; Duncan et al., 2016).

MALDI utilises a pulsed laser to eject a sample from a matrix. Initially desorption occurs releasing the matrix and analyte, before ionisation. Due to the acidic nature of the matrix, ions are formed in solution. This is an efficient method to ionise proteins into the gaseous phase, making this method amenable for MS-Imaging (Gessel et al., 2014; Winter et al., 2017). Further, MALDI produces predominantly singly charged ions from the matrix, making data analysis significantly less complex than spectra containing multiply charged species of the same analyte.

However, MALDI cannot be coupled directly to a separation technique like liquid chromatography. Thus, inherently there is an issue with the separation of different peptides before ionisation following digestion of a protein. This has been worked around by a number of approaches. Fernandez-Puente and colleagues utilised a nanoscale LC approach to fractionate different peptides before performing mass spectrometry (Fernandez-Puente et al., 2014). This method utilised a robot to deposit LC eluent combined with MALDI matrix onto a matrix target. This significantly reduced the

manual work required to remove the LC fractions and apply them to target plates manually. A similar approach was utilised by Wiangnon et al. (Wiangnon and Cramer, 2016), with the addition of a liquid matrices approach to perform peptide fingerprinting of BSA.

Further drawbacks exist for protein analysis, specifically native analysis, with MALDI. Given the propensity of MALDI to ionise analytes with a single charge, this method is not suitable for the analysis of native proteins, as the m/z value is too high to be detected. Further, the low pH conditions required for matrix formation are likely to unfold native proteins within the solution. Therefore, this review will focus on ESI based ionisation sources.

1.3.3 ESI

ESI achieves ionisation through an alternative mechanism to MALDI. The initial design of ESI was proposed in 1917 by Zeleny (Zeleny, 1917), with John Fenn combining ESI and Mass Spectrometry in 1984 (Yamashita and Fenn, 1984). Following this, Fenn published his first ESI of proteins in 1989 (Fenn et al., 1989). ESI involves passing a sample through a capillary and across a high electric potential to generate a spray. The capillary flow and voltage cause charged droplets to be formed, which decrease in size as the droplets evaporate. Thus, both electrostatic repulsion and Coulombic repulsion contribute to the formation of a charged gas phase ion (de Hoffmann and Stroobant, 2007; Mehmood et al., 2015).

ESI's specific mechanism of ionisation is debated, with ion evaporation, charged residue and chain ejection all proposed as theories of how ionization proceeds (Dole et al., 1968; Iribarne and Thomson, 1976; Konermann, Lars et al., 2012) (figure 1.7). The coupling of ESI to liquid chromatography is prevalent in literature, especially for the analysis of digested protein samples, as the LC allows for the separation of peptides before MS analysis.

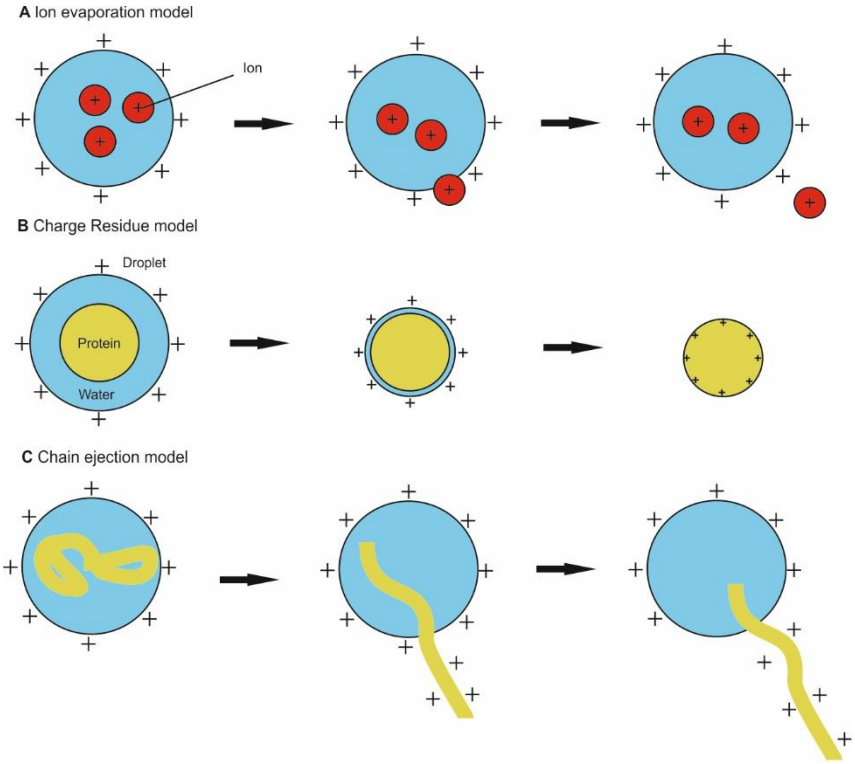


Figure 1.7: A diagram showing ESI ionisation. Diagram A shows the ion evaporation model, with B and C showing two other hypotheses for ion generation, the charge residue model and the chain ejection model. Image adapted from (Mehmood et al., 2015).

The ion evaporation model is based around the concept that the electric field generated from a Rayleigh-charged droplet is high enough to cause ejection of ions from the droplet surface (Iribarne and Thomson, 1976; Konermann, L. et al., 2013). Molecular dynamic simulations of this model have shown that a string of solvent molecules leave the droplet, which occurs when repulsion from the excess droplet charge overcomes attractive forces on the droplet from solvent polarisation (Ahadi and Konermann, 2011; Konermann, L. et al., 2013). This model has the most merit when describing the action of small molecules.

For folded "native" proteins, the charged residue model is generally accepted, and involves a single Rayleigh-charged nanodroplet evaporating until a single analyte is ionised with no solvent present. This ionisation comes from the passing of the charge from the solvent to the analyte (Iavarone and Williams, 2003).

For unfolded proteins and polymers, it is hypothesised that a chain ejection model is followed. This involves a long chain, when inside a Rayleigh-charged nanodroplet, migrating to the droplet surface and then having one end expelled from the surface through repulsive forces, similar to as described for the ion evaporation model. This process is repeated for the entire length of the protein, until the entire molecule is ejected from the droplet (Konermann, Lars et al., 2012; Konermann, L. et al., 2013)

ESI has the ability to maintain non-covalent interactions, which is not easily possible via MALDI. This is important for the analysis of proteins in their quaternary structure, which are comprised of multiple subunits non-covalently bound together. Light-Wahl et al., utilised ESI to look at tetramers of avidin, concanavlin and haemoglobin (Light-wahl, 1994). This work shows the capability of ESI to maintain native oligomer structure, something not reported with MALDI.

1.3.4 Mass analysers

The role of the mass analyser is to separate the ions based on their m/z ratio (Mehmood et al., 2015). This is usually achieved by use of electric fields, with several methods used to separate analytes, including quadrupole (Paul, 1953), Time of Flight (Cameron and Eggers, 1948; Wolff and Stephens, 1953) and Orbitrap (Makarov, 2000). The selection of the mass analyser is based on the specific requirements for the experiment, and include the m/z range required, the sensitivity required and the desired resolution. Scan speed, defined as the length of time it takes to acquire a spectrum, along with dynamic range and sensitivity of the instrument should also be taken into consideration.

Resolution can be defined as the ability to determine two different peaks of different m/z ratio.

This is determined by the equation

$$R = \frac{M}{\Delta M}$$

Equation 1.1: Equation for resolution in a mass spectrometer

Where R is the resolution, M is the mass of the second peak, and ΔM is the resolving power (Murray, 2022).

In practice, ΔM is calculated as the width of the peak at half of it's maximum height, known as full width half maximum (FWHM) (Murray, 2022).

Resolving power, is ability to distinguish between two different m/z ratios in a mass spectrum. This is measured by taking two points on the peak, commonly at 50% and 5% of the maximum peak height and using this to calculate the change in mass as a function ΔM (Murray, 2022).

Sensitivity is expressed a function of signal to noise, as a signal to noise ratio (S/N). Sensitivity can therefore be improved by either increasing the signal or reducing the noise. Sensitivity is split into limits of detection (LOD) and limits of quantification (LOQ), where LOD is defined as a signal to noise of ratio 3:1 and LOQ is defined as a signal to noise ratio of 10:1 (Li, C. et al., 2021).

LOD is the lowest concentration of an analyte that can be reliably detected (above a signal to noise ratio of 3:1), and the LOQ is the lowest concentration of an analyte that can be reliably quantified (above a signal to noise of 10:1). Therefore, the lower the LOD and LOQ of the instrument, the better the sensitivity.

Dynamic range is the m/z ratio range over which mass accuracy can be determined. The higher the dynamic range the wider the range of m/z ratios that can be determined by the mass analyser.

1.3.5 Quadrupole

A quadrupole mass analyser consists of four parallel rods, with each opposing rod pair having the same radio frequency (RF) and direct current (DC). When the RF and DC are applied on each rod, with non-opposing rods having differing values, only ions of a specific m/z are able to pass through.

Ions that have a m/z higher or lower than typically 1 m/z unit of the specified m/z do not have stable trajectories inside the quadrupole device and do not pass through the entire quadrupole (figure 1.8).

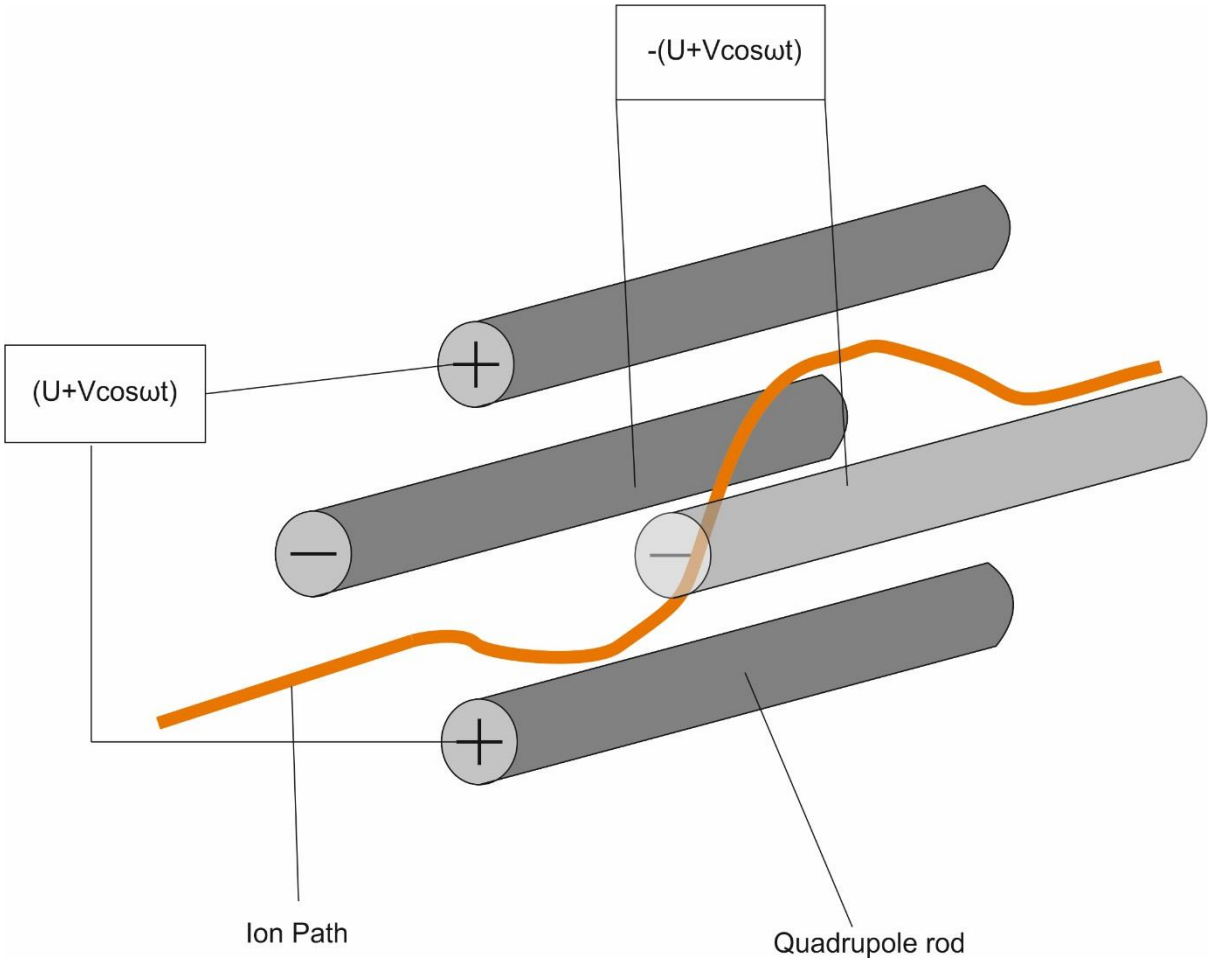


Figure 1.8: A diagram of a quadrupole mass analyser. Image adapted from (Savaryn et al., 2016). RF and DC current are applied to each rod, 4 rods in total. Opposing rods have +(U+Vcosωt) and –(U+Vcosωt). U is fixed potential and Vcosω is the RF of amplitude V.

Quadrouple mass analysers, whilst first developed as mass analysers in their own right, are now often placed in series in order to perform specific experiments. This is known as tandem quadrupole mass spectrometry, and involves placing quadrupole mass analysers either side of a central quadrupole, hexapole or octapole, which can be used for fragmentation experiments.

In a tandem quadrupole instrument, the first and final quadrupoles act as mass analysers, with the middle set of rods operating in Rf-mode only thus acting as a collision chamber. In this chamber, an inert gas is introduced in order to collide with and fragment the ions when their energy is increased, with the final quadrupole analysing the m/z values of the fragments produced (Yost and Enke, 1979).

The primary advantage of this arrangement is that specific fragmentation can be assessed, with the ability to select which ions enter the collision chamber, and which ions are scanned for following fragmentation.

1.3.6 Time of flight

A time of flight (TOF) mass spectrometer measures the m/z ratio of ions by acceleration from the pusher through a flight path of a set distance and measuring how long it takes for each ion to travel the flight path. To achieve this, ions are accelerated by a pulsed pusher, a device which uses an opposing voltage to repel the ions down the flight tube at a set frequency (Boesl, 2017). Often a reflectron (ion mirror) is incorporated, which increases the length of the flight path and corrects for kinetic energy differences for ions with the same m/z. This device reflects the ions back down the flight tube by utilising an opposing electric field. The higher the kinetic energy of the ion, the further it will penetrate into this electric field before being reflected. Given the ion is repelled by the same field it has penetrated, the kinetic energy is conserved upon leaving the reflectron, ensuring there is no effect on the m/z measurement at the detector (Boesl, 2017) (figure 1.9).

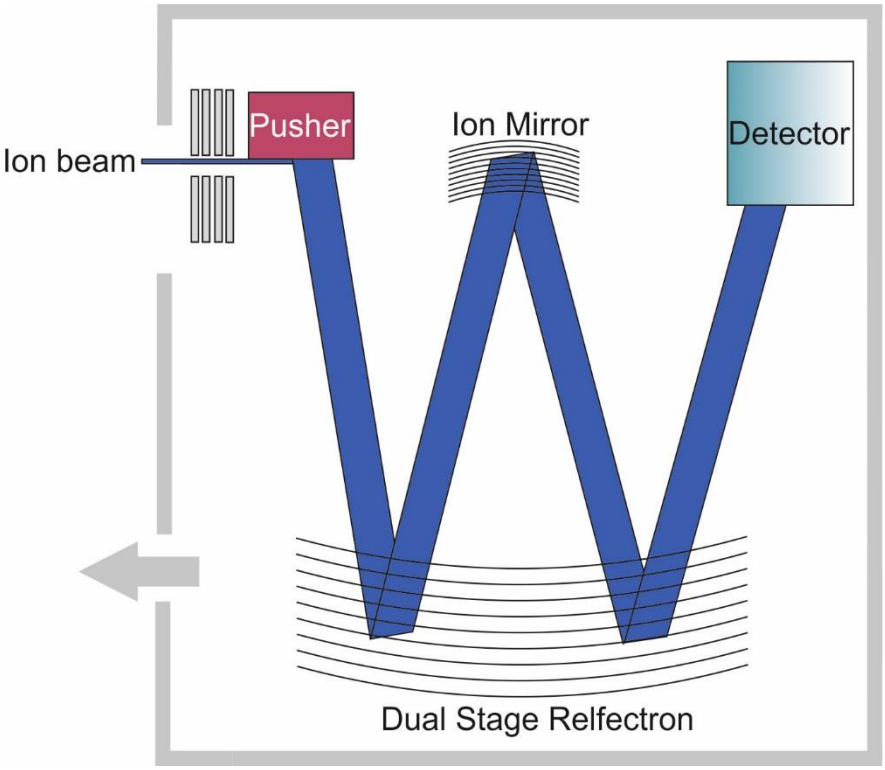


Figure 1.9: Schematic of a TOF flight tube showing the ion path and a dual stage reflectron. The ions enter as a beam, before being orthogonally accelerated down the flight tube by the pusher and separated by their m/z ratio as they travel through the reflectron before detection

Time of flight mass spectrometers have a number of advantages for protein analysis, the two primary ones being high mass resolution, with the incorporation of a reflectron, and a high m/z range, allowing for high mass accuracy peptide and intact protein analysis.

Indeed, TOF is governed through a conversion of kinetic energy to m/z. This is determined through the following equation, which shows the relationship between kinetic and potential energy for ions in the TOF.

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

Equation 1.2: relationship between potential energy and kinetic energy. zeV is potential energy, where z in the number of charges, e is the elementary charge and V is the pusher voltage. This is equal to kinetic energy, where m is ion mass, L is the length of the flight tube and t is flight time.

Where potential energy (zeV) (z being number of charges, e being elementary charge and V being pusher voltage) is equal to kinetic energy where m is ion mass, L is length of the flight tube and t is flight time.

This is generally rearranged such that

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

Equation 1.3: rearrangement of equation 1.2

Given the bracketed terms are constant in TOF:

$$t \propto \sqrt{\frac{m}{z}}$$

Equation 1.4: relationship between time and m/z in a TOF mass analyser

1.3.7 Orbitrap

An Orbitrap is an ion trapping instrument that consists of an inner spindle electrode and an outer barrel electrode. Ions oscillate around and along the spindle, with the barrel electrode split into left and right at either end of the spindle electrode. Frequency of longitudinal oscillation along the z axis is proportional to m/z, as shown in figure 1.10.

The Orbitrap instrument functions by initially trapping ions in a C-trap before sending the ion packet into the Orbitrap analyser. The analyser has strong radial and axial forces (Eliuk and Makarov, 2015) which cause the ions to rotate around the spindle and to oscillate along the z axis (figure 1.10). This oscillation frequency is inversely proportional to the m/z, and as such the m/z can be determined by

Fourier-transform methods. This gives Orbitrap analysers significantly higher resolution than either tandem quadrupoles or time of flight instruments (Eliuk and Makarov, 2015).

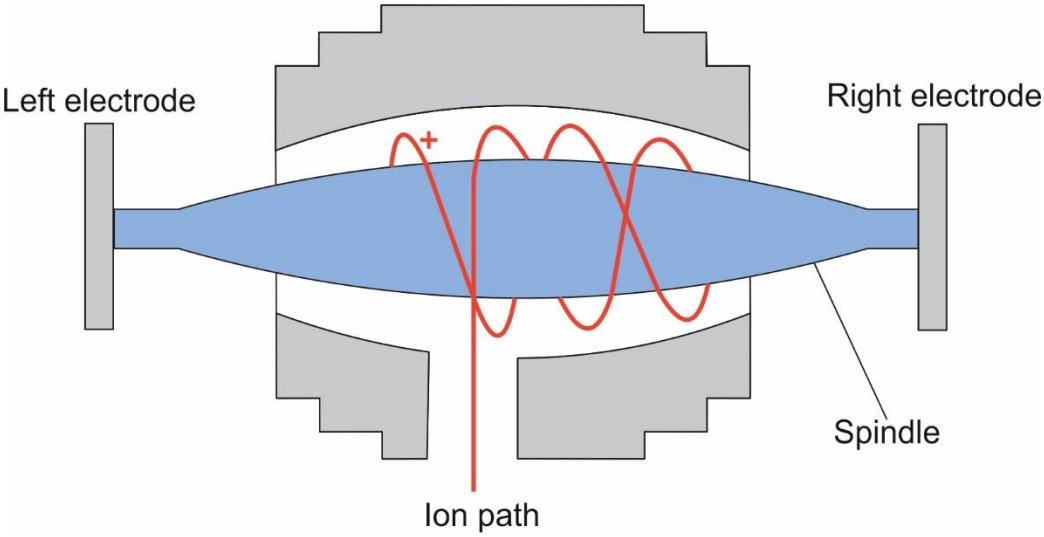


Figure 1.10: Diagram of an Orbitrap mass analyser. Ions enter along the ion path before oscillating around a central spindle. The frequency of oscillation is fourier-transformed to produce a mass spectrum. Diagram adapted from (Savaryn et al., 2016).

To summarise, whilst tandem quadrupoles offer significant capabilities with precursor and fragment ion scans, the higher mass range and resolution of Q-TOF and Orbitrap instruments make them more suitable for protein analysis, as they can be used for both intact and digested peptide samples.

Orbitraps, convert oscillation along the z axis to a frequency which can then be converted to a m/z ratio. This is determined by the following equation:

$$2\pi v = \sqrt{\frac{k}{m/z}}$$

Equation 1.5: relationship between oscillation and m/z in an orbitrap

Where v is the oscillation frequency, k is the field curvature and m/z is the mass to charge ratio.

1.3.8 Top-down vs bottom-up proteomics

Bottom-up proteomics is the traditional method of performing proteomics experiments (Catherman et al., 2014). A protein is digested either enzymatically or chemically, with either approach used to form peptides. Use of trypsin (Hildonen et al., 2014) and pepsin (López-Ferrer et al., 2011) is common. Typsin cleaves at the C-terminal peptide bond of arginine and lysine residues, with the exception of arginine-proline and arginine-lysine amino acid sequences, where cleavage is protected (Manea et al., 2007).

The trypsin digestion reaction is preceded with reduction of disulphide bridges, through reaction with dithiothreitol (DTT), before alkylating the cysteine residues with the addition of iodoacetamide (Walmsley et al., 2013). This step is required for bottom-up mass spectrometry to allow for the protein to unfold and full enzymatic digestion of the protein to occur. Following this step, the trypsin is added to digest the protein.

Following digestion of the protein into peptides of differing amino acid length, as determined by the distance between cleavable residues, the mixture is introduced to an LC-MS workflow.

LC for bottom-up proteomics is most commonly performed using reverse phase LC. In reverse phase LC, the polarity of the column is lower than the polarity of the mobile phase. The mobile phase is operated on a gradient, starting with a high polarity and reducing this polarity over time through the combination of a polar buffer such as water, and a non-polar buffer such as acetonitrile. Peptides have a higher affinity to the non-polar stationary phase than the initial polar mobile phase, and as such bind to the column. As the polarity of the mobile phase is reduced, peptides are sequentially eluted from the column, dependent on the specific affinity to the stationary phase of the individual peptide (Dupree et al., 2020).

MS of the peptides if performed following elution from the LC. The separation from the LC allows for sequential fragmentation within the MS on each peptide, allowing for MS-sequencing of each peptide.

Analysis of the sample is dependent on whether the protein amino acid sequence is known. If the sequence is known, sequenced peptides can be compared to the amino acid sequence to determine the coverage and validate the protein. For unknown proteins, given the cleavage parameters of the enzyme, and the sequencing of peptides through tandem MS, the protein can be sequenced and thus the amino acid sequence determined, termed a denovo sequence (Liu, X. et al., 2014). Tandem MS utilises two mass analysers in series with a collision cell between them. The first mass analyser (MS1) is used to select precursor ions of a specific m/z ratio, these ions are then fragmented in the collision cell, and the m/z ratio of the fragment ions measured in the second mass analyser (MS2). These fragments can be sequenced to determine the amino acid sequence of the precursor peptide in the case of bottom-up experiments.

Top-down proteomics involves introducing the protein to the mass spectrometer without any digestion or fragmentation, before using the mass spectrometer to determine the m/z of both the intact and fragments following collision induced dissociation (CID). CID is a fragmentation method, whereby precursor ions are accelerated into a cell containing a neutral gas, causing fragmentation of the precursor molecule through collisions with between ions and the gas.

Whilst these approaches are performed on individual proteins, it is common to look at the entire proteome using either top-down or bottom-up approaches. Top-down requires separation of a whole proteome before introduction into the MS due to the high complexity of the sample (Catherman et al., 2014). This separation can be performed through online or offline LC, however approaches such as Hydrophobic Interaction Liquid Chromatography (HILIC), Ion-exchange chromatography (IEX) and electrophoresis can also be implemented to facilitate this separation.

High resolution MS is required for top-down to determine specific protein features such as disulfide bridge formation, deamidation and trimethylation. This is due to the precursor and fragments being of a higher m/z ratio than in a bottom-up experiment. Further, sensitivity must be higher for top-down, as the charge envelope distribution is higher, meaning each individual m/z ratio peak is lower intensity (Catherman et al., 2014).

Top- down can maintain post-translational modifications and proteoform information. Further, lack of quantification reliability from analysing a subset of peptides can be circumvented by using top down and analysing the intact protein (Tholey and Becker, 2017). It is noted however, that data processing for top down is significantly more challenging, and bottom-up approaches using automated software currently routinely offer higher coverage than top-down approaches (Catherman et al., 2014).

Middle down proteomics is a further approach that encompasses aspects of both top-down and bottom-up. Middle down involves a protein digest, however this digest is limited, forming significantly longer peptides than in bottom-up (above 3kDa) (Cristobal et al., 2017). This increases the likelihood of detecting PTMs, as the peptides are longer in sequence, whilst still being able to take advantage of separation techniques. However, currently, middle-down is not commonly used for proteome wide studies (Cristobal et al., 2017).

1.3.9 Fragmentation methods

For both top-down and bottom-up proteomics, the protein and peptide respectively are fragmented to produce fragments that can be used to sequence the amino acid sequence of the analyte. The primary fragmentation methods used are collision induced dissociation (CID), electron capture dissociation (ECD) and electron transfer dissociation (ETD).

CID is the most common fragmentation method for proteomics. Ions are transmitted from the ion source through the electronic potential of the MS instrument into a collision cell containing an inert

gas, and are thus subjected to collisions with the inert gas molecules (Wells and McLuckey, 2005). Collisions increase the internal energy of the molecule, which is distributed uniformly across the molecule until the weakest bond is broken, initiating fragmentation. This is therefore described as an ergodic fragmentation method, as internal heating of the molecule is the mechanism of fragmentation. When applied to a peptide, this results in the cleavage of peptide bonds, and produces b and y fragment ions (figure 1.11) (Johnson and Carlson, 2015).

Alternatively, ECD and ETD can be used. These are both non-ergodic fragmentation methods, which means that fragmentation is not driven by the internal heating of the molecule, and the fragmentation occurs before this internal energy can be distributed over the molecule which would lead to fragmentation. ECD and ETD are facilitated by electronic excitation, and therefore are a faster method of fragmentation, and can be used to reduce scrambling in HDX (described in 1.4.2) (Brédy et al., 2022).

ECD involves colliding a charged precursor ion with a low energy electron beam. The capture of an electron causes backbone cleavage of, primarily, the N-C bond, producing c and z ions (Fort et al., 2018). ETD works in a similar fashion, however the electron transfer for fragmentation in this case is mediated by the introduction of anions, which allow for the backbone cleavage (Kim and Pandey, 2012). We again see c and z ions due to the cleavage of the N-C bond.

Further, a and x ions may be observed from the loss of CO_2 from b and y ions, as shown in figure 1.11 (Kim and Pandey, 2012).

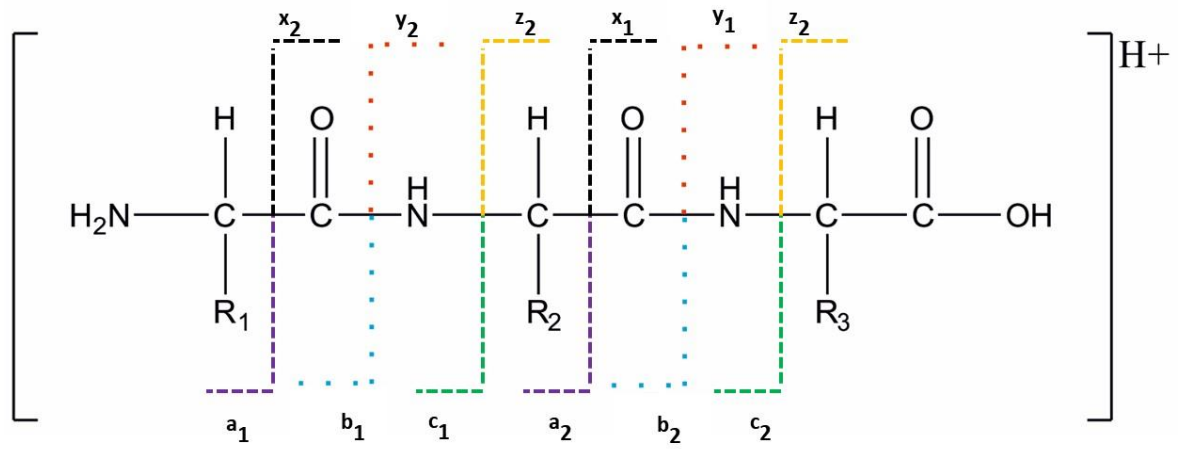


Figure 1.11: A schematic showing peptide ion nomenclature. b and y ions are commonly produced upon CID fragmentation, whilst c and z ions are commonly produced by non-ergodic fragmentation methods. a and x ions are commonly observed from the loss of CO_2 from b and y ions.

1.3.10 Ion mobility spectrometry

Ion mobility spectrometry (IMS) involves the separation of ions based on their mobility. The mobility of an ion is a measure of how quickly it moves through a buffer gas under the influence of a weak electric field. This allows for separation of ions based on their rotationally-averaged collision cross-section, mass and charge. When coupled with LC, separation by IMS gives additional resolution, allowing for separation by polarity (LC) and rotationally-averaged collisional cross section, mass and charge (IMS). This allows for the separation of isobaric and isomeric molecules prior to MS analysis (Lanucara et al., 2014).

A standard workflow is shown below, however it is noted that fragmentation can occur before ion mobility for mobility separation of fragment ions (Borotto and Graham, 2021).

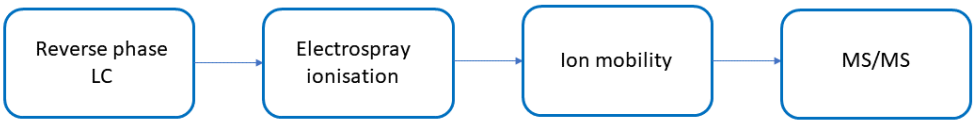


Figure 1.12: A standard LC-IMS-MS/MS workflow. Digested peptides are separated by their affinity to the stationary phase in LC. Reverse phase LC utilises a non-polar stationary phase and a polar mobile phase. Subsequently peptides are ionized through electrospray ionization, before separation according to rotationally-averaged collisional cross section, mass and charge in the ion mobility cell. Finally, peptides are fragmented and sequenced via MS/MS in the mass spectrometer.

Drift tube ion mobility involves introduction of the ions into a drift tube, with a uniform electric field and a buffer gas, commonly helium. The time taken for an ion to travel through the tube can be determined by its collisional-cross sectional area, mass and charge. Essentially, a more compact form of a molecule has less surface area, and thus experiences less collisions travelling through the cell than a less compact form. Collisional cross section (Ω) can be determined using the Mason-Schamp equation (Lanucara et al., 2014).

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

Equation 1.6: Calculation of collisional cross section. Collisional cross section is represented by Ω. K_0 is the measured mobility at a standard temperature and pressure. Z is the charge state of the ion, e is

the elementary charge, N is the density of the buffer gas, μ is the reduced mass of the ion-neutral drift gas pair, K_B is the Boltzmann constant and T is the gas temperature.

K_0 is the measured mobility at a standard temperature and pressure, z is the charge state, e is the charge, N is the density of the buffer gas, μ is the reduced mass of the ion-neutral drift gas pair, k_B is the Boltzmann constant and T is the gas temperature (Lanucara et al., 2014).

A schematic of a drift tube instrument can be found below.

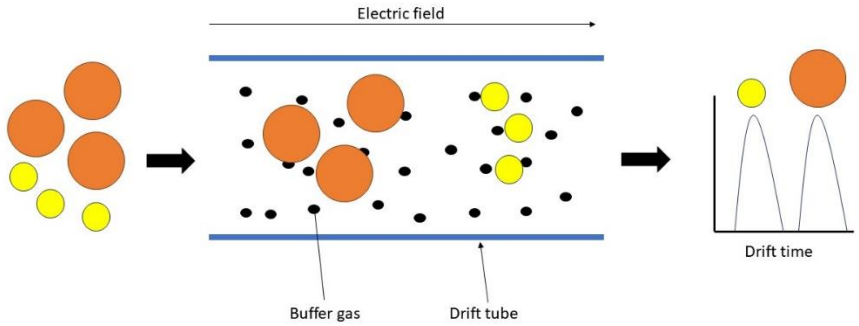


Figure 1.13 Diagram of drift cell mobility. Ionized molecules are guided into the drift tube and carried along with the electric field. The ions are separated via collisions with the buffer gas, with the larger ions by rotationally averaged collisional cross section having more collisions. This can be determined by the drift time at the end of the cell.

Travelling wave IMS (TWIMS) involves directing ions through a buffer gas filled RF-confining stacked ring ion guide (Richardson et al., 2021). The ions are separated using a series of DC waves. More compact ions will be carried by the wave, whereas more extended conformations will roll over the wave, causing separation through the longer period of time it takes for the larger ions to exit the IMS cell. By sending through a number of waves in quick succession, complex mixtures can be separated (Lanucara et al., 2014). Due to this, collisional cross section (CCS) cannot be directly determined,

however by using known calibrants for the cross section, CCS can be determined through a calibration curve.

However, when operated for separation of peptides, calibration is not required.

1.3.11 Acquisition modes

Determination of which ions are selected for fragmentation and detection within the mass spectrometer is essential for high sensitivity and accurate data. The two primary methods for data acquisition are data dependent acquisition (DDA) and data independent acquisition (DIA).

DDA is the most common method for data acquisition in proteomics experiments. Initially, peptides are eluted from the LC, ionized in the ion source and introduced to the MS. The ions introduced within a defined time period are scanned without fragmentation. Following this scan, the intensity of each ion is ranked by the instrument, and a user defined number of ions are selected for MS/MS analysis based on their intensity. This presents an inherent trade-off, as the more ions selected within the cycle, and the length of the cycle, reduces the amount of scans that can be completed over a given time frame.

DIA involves fragmentation of all of the peptide ions eluted from the LC by scanning through the entire mass range. This means all precursor ions in a selected window are fragmented at the same time. A rapid cycle of low energy fragmentation parameters for precursor ions and high energy for fragmentation of precursors allows for both precursor and fragment detection.

DIA offers significant advantages to DDA therefore, as all precursor ions are fragmented and analysed. Therefore, no ions are lost due to the precursor selection and fragmentation steps of DDA.

However, the drawbacks of DIA are the significant complexity of the data produced, as all of the precursors are selected and fragmented at the same time. Thus, at present, DDA is the popular option for proteomics experiments.

1.4 Structural Mass Spectrometry

Structural MS particularly pertains to the use of MS to determine the higher order structure of macromolecules, with a specific focus on proteins and protein complexes. A number of methods have been utilised for Structural MS including chemical cross-link, hydrogen-deuterium exchange (HDX) (Lento et al., 2017) and fast photochemical oxidation of proteins (FPOP) (Hambly and Gross, 2005). Within this, the labelling techniques of HDX and FPOP provide the opportunity to probe single residue structural information (Rand et al., 2009; Cornwell, O. et al., 2018a).

1.17.3 FPOP

FPOP utilises a 248 nm laser to form hydroxyl radicals through the hydrolysis of hydrogen peroxide. These radicals react with amino acid side chains to form covalent oxidative bonds with the side chain. This covalent addition of a hydroxyl radical is dependent on both the solvent accessibility and reactivity of the side chain. As described below, the sample is combined with hydrogen peroxide and a scavenger, commonly histidine or glutamine. This scavenger is utilised to react with free radical •OH lysed from hydrogen peroxide by the laser, to reduce the lifetime of the radicals. Following this, the sample is collected in an Eppendorf tube containing a quench solution of methionine and catalase, to decompose the hydrogen peroxide and remove remaining radicals.

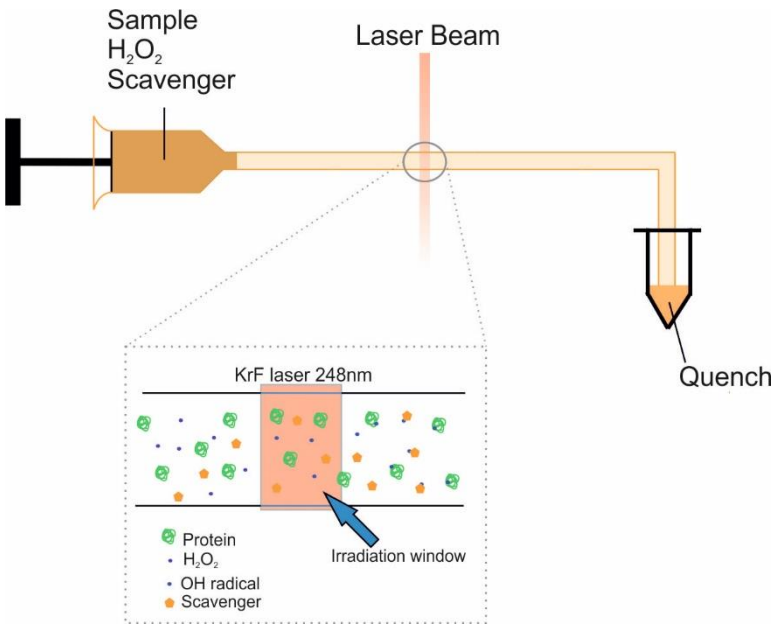


Figure 1.14: Diagram of the FPOP work flow redrawn from (Li, K.S. et al., 2018). The diluted sample is combined with hydrogen peroxide and a scavenger, usually histidine or glutamine. The mixture is flowed through a capillary, into the path of a 248nm laser beam, generated from an excimer laser. This lyses the hydrogen peroxide to OH radicals, with react with the sample. The scavenger also reacts with the lysed •OH, to reduce radical lifetime. The quench solution, containing methionine and catalase, mop up the remaining radicals and decompose the hydrogen peroxide.

FPOP has shown efficacy for protein labelling in a crowded environment, with in cell footprinting described by Jones and colleagues (Espino et al., 2015; Rinas et al., 2016). Cellular footprinting suffers with H_2O_2 decomposition through cellular catalase interactions, and this is combatted by ensuring the hydrogen peroxide is introduced to the sample as close to the irradiation window as possible.

FPOP was first proposed by Hambly and Gross in 2005 (Hambly and Gross, 2005), as a method to footprint solvent accessible side chains on a microsecond time scale. This microsecond timescale was also supported through statistical data (Gau et al., 2009) and subsequent observations via isotope dilution Gas Chromatography-Mass Spectrometry (GC/MS) (Niu, B. et al., 2015). This time scale has been disputed by Vahidi and Konermann (Vahidi and Konermann, 2016), who suggest that even though hydroxyl radicals have a lifetime on the microsecond time scale, the formation of secondary radicals with other constituents of the experiment such as the scavenger, may increase the lifetime of radicals to the millisecond timescale.

It is not clear from the work of Vahidi and Konermann, 2016, however, if secondary radical formation can cause an addition of +16, +32 or +48, which are the most common oxidation products seen in oxidative labelling techniques (Xu and Chance, 2005). Therefore, given the current evidence it is plausible to conclude that the majority of the labelling is undertaken on the microsecond timescale, through the reaction of hydrolysed hydroxyl radicals reacting with solvent exposed side chains.

Secondly, hydroxyl radicals react preferentially with different side chains. The reactivity of each sidechain is described by its rate constant with the hydroxyl radical. The different rate constants for each amino acid with hydroxyl radicals are shown below (table 1.1):

Table 1.1: Table showing the rate constants of amino acids with hydroxyl radicals, reproduced from (Xu and Chance, 2005).

Amino acid	**3 letter code**	**Rate constant ($m^{-1}s^{-1}$)**
Cysteine	Cys	3.5×10^{10}
Tryptophan	Trp	1.3×10^{10}
Tyrosine	Tyr	1.3×10^{10}
Methionine	Met	8.5×10^{9}
Phenylalanine	Phe	6.9×10^{9}
Histidine	His	4.8×10^{9}
Arginine	Arg	3.5×10^{9}
Cystine	n/a	2.1×10^{9}
Isoleucine	Ile	1.8×10^{9}
Leucine	Leu	1.7×10^{9}
Valine	Val	8.5×10^{8}
Proline	Pro	6.5×10^{8}
Glutamine	Gln	5.4×10^{8}
Threonine	Thr	5.1×10^{8}
Lysine	Lys	3.5×10^{8}
Serine	Ser	3.2×10^{8}
Glutamic acid	Glu	2.3×10^{8}
Alanine	Ala	7.7×10^{7}
Aspartic acid	Asp	7.5×10^{7}
Asparagine	Asn	4.9×10^{7}
Glycine	Gly	1.7×10^{7}

Further, the number of oxidations possible for each amino acid varies, depending on the chemical structure of each side chain. Cyclic side chains in particular can have multiple hydroxyl radicals on the same side chain. Whilst these modifications were identified through synchrotron hydroxyl footprinting, the chemistry is comparable to photochemically produced radical hydroxyl labelling. The possible primary oxidation products are shown below (table 1.2):

Table 1.2: Table showing side chain modifications for each amino acid residue, adapted from (Xu and Chance, 2005).

3 letter code	Side chain modification and mass changes
Cys	Sulfonic acid (+48), Sulfinic acid (+32), hydroxy (-16)
Trp	Hydroxy- (+16, +32, +48, +60, +72), pyrrol ring-open (+32)
Tyr	Hydroxy- (+16,+32)
Met	Sulfoxide (+16), sulfone (+32), aldehyde (-32)
Phe	Hydroxy- (+16, +32, +48)
His	Oxo- (+16), ring-open (-22, -10, +5)
Arg	Deguanidination (-43), hydroxy- (+16), carbonyl (+14)
Cystine	Sulfonic acid (+48), sulfinic acid (+32)
Ile	Hydroxy- (+16), carbonyl (+14)
Leu	Hydroxy- (+16), carbonyl (+14)
Val	Hydroxy- (+16), carbonyl (+14)
Pro	Hydroxy- (+16), carbonyl (+14)
Gln	Hydroxy- (+16), carbonyl (+14)
Thr	hydroxy- (+16), carbonyl (−2, or +16−H2O)
Lys	Hydroxy- (+16), carbonyl (+14)
Ser	Hydroxy- (+16), carbonyl (−2, or +16−H2O)
Glu	Deguanidination (-43), hydroxy- (+16), carbonyl (+14)
Ala	hydroxy- (+16)
Asp	Decarboxylation (−30), hydroxy- (+16)
Asn	Hydroxy- (+16)
Gly	n/a

Therefore, both reactivity and side chain composition play a vital part in both the intensity and the mass shift observed via FPOP, and thus residue level quantification needs to ensure that the correct mass shifts are being assigned to each residue and that the analysis is sensitive enough to detect small shifts in relative intensity.

It should be noted that FPOP data analysis is presented as fold change (Cornwell, O. et al., 2018a). Thus, even if a relatively unreactive residue has a small absolute change in oxidation, this fold change could indicate a large shift in solvent accessibility.

Further, the effect of neighbouring amino acids has not been comprehensively studied. Essentially, whilst the reactivity of single amino acids with hydroxyl radicals has been widely reported, as discussed previously, the effect of amino acid hydroxyl reactivity due to its environment has not been extensively examined. This effect has been used to explain differences in oxidation change and calculated solvent accessibility (Xie, B. et al., 2017), whereby both tertiary and quaternary structure

formation, along with highly reactive side chains being located close to each other sequentially, has been hypothesised to reduce reactivity of less reactive local side chains. This has further been proposed to explain a reduction in oxidation for a structurally more exposed phenylalanine, whereby the addition of a highly reactive tryptophan in the microenvironment gives a plausible solution to this observed drop in oxidation (Cornwell, O. et al., 2018a).

1.4.2 HDX

Hydrogen deuterium exchange (HDX) involves the exchange of solvent accessible labile hydrogens for deuterium. Contextually, N-H, S-H and O-H bonds, labile hydrogens found in solvent accessible regions of proteins, make HDX viable for the analysis of protein structures, and as such it has been used with both nuclear magnetic resonance (NMR) and MS for determination of protein conformation (Konermann, Lars et al., 2011).

In terms of protein MS, the most common method of performing this reaction involves a continuous labelling experiment, whereby D_2O is added to a sample for a set period of time before exchanged is ceased by addition of a low pH quench buffer and low temperature (Englander et al., 1996). Following this, proteolytic digestion occurs, before a short LC gradient and MS analysis.

Digestion times and LC gradients must be kept short due to the phenomena of back-exchange. As soon as the reaction is quenched, deuterium will exchange back to hydrogen. This occurs almost immediately for deuterium located on amino acid sidechains, and as such HDX in a MS setting probes amide hydrogens on the backbone only (Englander, 2006).

The +1 Da difference between hydrogen and deuterium, due to deuterium containing an extra neutron, allows for the difference in solvent accessible hydrogens to be probed easily with a change in m/z between two states. Thus, from the observed change in deuterium uptake we can infer change in protection due to either conformational changes or binding at the peptide level. However, when using collision induced dissociation (CID), deuterium can be scrambled to any position on the

peptide (Wollenberg et al., 2020). When trying to determine the location of which residue has exchanged, a non-ergodic fragmentation method must be used to reduce this scrambling effect. The majority of studies therefore are performed at the peptide level, due to further analysis and more complex fragmentation required for analysis.

The two primary factors dictating HDX labelling are solvent accessibility and hydrogen bonding. Solvent accessibility renders the biggest factor for labelling, as if the deuterium is unable to reach the residue in question, no exchange can take place. Further, protein folding is a primary driver of labelling, with folded protein amides having a 6-8 orders of magnitude less labelling than backbone amides in unfolded proteins (Wales and Engen, 2006; Konermann, Lars et al., 2011).

However, it is important to note, proteins are dynamic, and global and local transient conformational changes occur, as well as hydrogen bond separation and reformation. This leaves backbone amides alternating between states of being available for exchange and unexchangeable, which leads to EX1 and EX2 kinetics.

EX1 Kinetics occur when proteins open and fold (solvent exposed to non-solvent exposed) slow enough such that a single unfolding event is enough to completely exchange deuterium, whereas EX2 kinetics occur where multiple unfolding events are required for full deuterium uptake. EX1 kinetics are rarely observed at physiological pHs. Essentially, if a region of a protein is fully exposed to deuterated solvent, it can freely exchange with deuterium, undergoing EX1 kinetics. If a region is fully protected from interaction with the solvent no exchange can occur. However, if the protein is undergoing dynamic conformational change, partial uptake will occur in the region of interest and deuterium incorporation will increase over time, giving EX2 kinetics.

The difference in kinetics can be observed in the mass envelopes, and often a combination of EX1 and EX2 kinetics can often be observed in HDX experiments (Konermann, Lars et al., 2011).

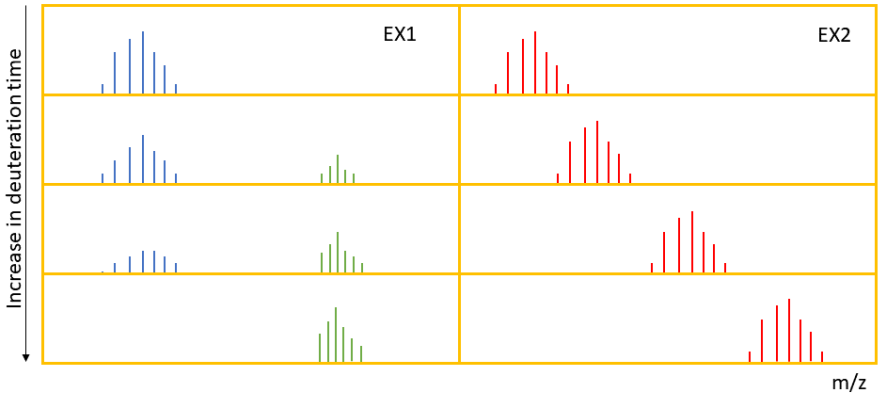
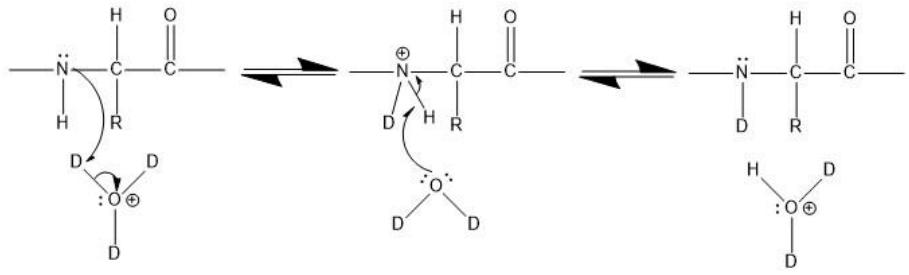


Figure 1.15: An example of EX1 and EX2 kinetics mass envelopes. With EX1 kinetics we see two mass envelopes as deuteration increases, this represents the undeuterated peptide and fully deuterated peptide existing in situ, where the peptide has full deuterium uptake or no uptake. The EX2 kinetics shows a slow shift towards full deuteration as partial deuterium uptake occurs. Redrawn from (Ozohanics and Ambrus, 2020).

Exchange rate of deuterium is also governed by pH, whereby both decreasing or increasing the pH from 2.5-3 will increase the exchange rate of deuterium. This is because exchange can be both acid catalysed and base catalysed depending on pH (Zhang and Smith, 1993).

Acid catalysed exchange



Base catalysed exchange

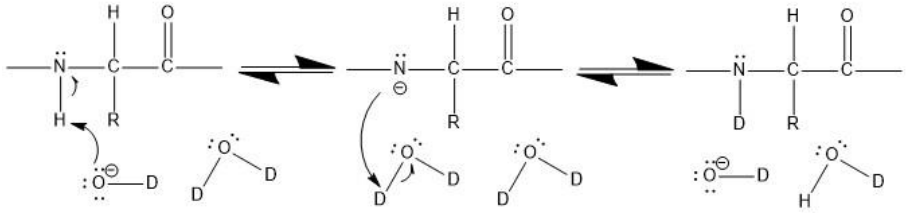


Figure 1.16: Mechanisms of base and acid catalysed exchange of deuterium with hydrogen during HDX. Below pH 2.5, acid catalysed exchange is used to exchange deuterium, whereas above ~pH 3, base catalysed exchange facilitates exchange.

1.4.3 Comparison of HDX and FPOP

Both HDX and FPOP are core footprinting technologies that can be used to elucidate higher order structure of proteins. However, as discussed in 1.4.1 and 1.4.2, the techniques achieve this through completely different mechanisms.

HDX labelling occurs through the exchange of hydrogen to deuterium, by immersing the protein in deuterated buffer, this mechanism is shown in figure 1.16. FPOP labelling is facilitated through generation of hydroxyl radicals and reaction with amino acid sidechains, an example of the mechanism is shown in figure 1.17.

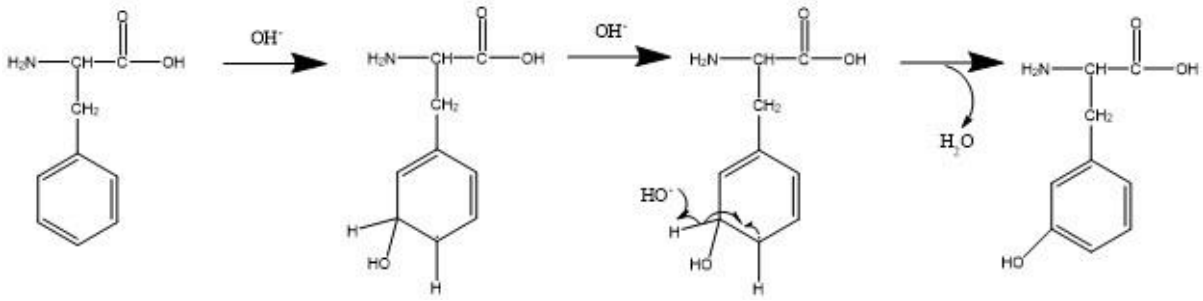


Figure 1.17 mechanism of labelling of phenylalanine for conversion to meta-tyrosine (Liu, X.R. et al., 2019). Hydroxyl radicals generated from lysis of H_2O_2 initiate the reaction, and attack the benzene ring of the phenylalanine residue. A second radical reacts with the radical on the ring forming two stable products

From these mechanisms, the first difference highlighted is HDX labelling is a reversable process, which is facilitated by an excess of deuterium in the buffer, whereas the OH radical labelling is covalent and thus irreversible. This distinction has implications for the downstream processing and analysis of proteins, and thus the applications for analysis.

As soon as the HDX reaction is quenched, back exchange from deuterium to hydrogen is initiated, with the amide backbone deuterium most protected from back exchange due to the location within the structure and hydrogen bonding (Scrosati et al., 2021). This requires rapid digestion and chromatography before introduction into the MS, to ensure back exchange is limited. Chromatographic columns are kept at 2°C and the pH at 2.5, to limit back exchange. Therefore, digestion occurs online, and the enzyme used must be functional at these conditions. Selection of the digestion enzyme is therefore limited, as is the methods to reduce disulphide bridges to allow for full digestion. Rapid online digestion with a fixed pepsin column is often utilised, and reduction of the disulphide bridges is commonly facilitated with the use of Tris (2-carboxyethyl) phosphine (TCEP) (Zhang, H.M. et al., 2010) (Cornwell, O. et al., 2018a).

Given the label is covalent for FPOP, the limitations caused by back-exchange are mitigated. Thus, reduction of disulphide bridges can be performed with dithiothreitol (DTT) and iodoacetamide over

a longer incubation time, increasing reduction efficiency. Secondly, selection of the digestion enzyme is not limited by pH, and thus a more application specific enzyme can be selected, and again incubated for a longer period. Finally, chromatography can be performed for a longer period, increasing chromatographic resolution.

The primary factor for labelling of a residue in FPOP is the rate of reactivity of the sidechain with OH radicals, with the least reactive residues having a reactivity rate of 3 orders of magnitude lower than the most reactive residues. Thus, at the peptide level, changes in labelling are dominated by reactive residues. Given the primary driver of labelling in HDX is solvent accessibility, there is no discrepancy between peptides due to reactivity of residues. Thus, at the peptide level, HDX can report on direct changes in labelling without bias from the individual reactivity of residues that compose the peptide.

At the residue level, when utilising CID fragmentation, the location of deuterium along the peptide backbone can be scrambled due to the internal energy produced by CID (Wollenberg et al., 2020). Thus, it is not possible to compare labelling at the residue level for HDX unless non-ergodic fragmentation methods are used. This issue is avoided with FPOP, as the covalent label cannot leave the residue that has been labelled. Thus, at the residue level, FPOP is viable with CID to locate differences in labelling at the residue level.

In HDX, the dosage of deuterium is not limited, thus, in the case of complex mixtures such as drug formulations, labelling is not affected by components within the formulation. However, for FPOP, as radical generation is limited, should other components in a mixture react with hydroxyl radicals, the labelling of the protein may be reduced, causing an erroneous reduction in labelling of the protein.

HDX is performed a time course, and deuteration of peptides plotted as a curve. This can be used to compare dynamic changes in structure over time. This is in contrast to FPOP which probes solvent accessibility for a single point. Therefore, HDX may be more suitable for monitoring changes in protein structure over a longer time course, and can be used to compare regions with faster and slower rates of exchange.

Another difference between HDX and FPOP is the rate of labelling. FPOP labels on the µs to ms timescale, which is significantly faster than labelling from commercial HDX systems, which label from 30s at a minimum (Vahidi and Konermann, 2016; Masson et al., 2019). Thus, FPOP provides a viable solution for probing highly dynamic proteins, such as intrinsically disordered proteins (IDPs).

Both footprinting techniques offer advantages and disadvantages, and the method selected should be tailored to the specific question required and the protein and solution being analysed.

1.5 Analysis of polysorbates and protein structure by mass spectrometry

Whilst polysorbate degradation has been measured via mass spectrometry (Kishore et al., 2011a), and thermal stability of proteins in solution with polysorbates (Agarkhed et al., 2013), little work has been done outlining the specific changes that polysorbates have on protein structure and how they stabilise proteins. Whilst Li and colleagues (Li, Y. et al., 2014) looked at polysorbate and protein samples using mass spectrometry, a 2D-LC separation system was used to interrogate polysorbate degradation and changes in degradation profiles based on their interaction with PS 20. Thus, although it has been shown polysorbates have a positive impact on the stability of several proteins (Ha et al., 2002; Kishore et al., 2011a), further work is required to investigate the specific mechanisms of PS interaction.

1.6. Objectives of the thesis

As presented in the introduction, a number of interesting areas have been described which warrant further investigation within the field of protein footprinting, and specifically in the advancement of these techniques for pharmaceutical applications.

Objective one is to improve the understanding of FPOP neighbourhood effects, by comparing peptides with reactive residues further apart and closer together in primary sequence. Initially, this experiment aims to prove that local environment is a factor in the labelling efficiency if individual

residues. Should this be the case, this should allow for a set of rules to be determined to determine the relationship between residue distance and reactivity, specifically for the FPOP labelling technique.

The second objective is to use HDX to determine how appropriate this labelling technique for determining changes in protein structure in complex environments, specifically for pharmaceutical applications. Comparing how polysorbates can perturb protein structure, and detecting these changes by HDX should allow for a thorough evaluation of this method, and if it is suitable for these applications. Further, this may have applications for selection of polysorbates in drug formulations and inclusion of a HDX assay during formulation design.

The final objective is to determine the binding sites of TiO_2 with MtrC. TiO_2 provides a specific opportunity as a candidate to facilitate electron transfer to the MtrCAB complex, and as such, determining how it binds may be important for the development of biological batteries and solar energy harvesting. However on a larger level, this aims to provide an evaluation of using synchrotron radiation for labelling of proteins in this context, and how it can be applied in the future.

The overarching theme, therefore, is to utilise different footprinting techniques, to compare and contrast their use for the determination of higher-order structure in biological systems.

Chapter 2:

Materials and methods

2 Materials and Methods

For each chapter, a brief methods section has been included. Comprehensive methods are listed in this chapter.

2.1 FPOP of peptides

2.1.1 Peptide acquisition

Six peptides were synthesised by Peptide Protein Research ltd (Bishops Waltham, UK). The structures are shown in Chapter 3. The sequences are listed below:

Table 2.1 Sequence of the peptides utilised in the labelling experiment. The first three peptides, W0F, W1F and W5F contain tryptophan and phenylalanine residues, separated by 0, 1 and 5 residues respectively, and capped with lysine residues. F0F, F1F and F5F substitute the tryptophan for a second phenylalanine residue

Peptide abbreviation	Amino acid sequence
W0F	KGWFGGGGGGK
W1F	KGWGFGGGGGK
W5F	KGWGGGGGFGK
F0F	KGFFGGGGGGK
F1F	KGFGGGGGFGK
F5F	KGFGGGGGFGK

2.1.2 FPOP conditions

Peptides were provided lyophilised and were dissolved in 50 mM potassium phosphate at pH 7. 100 µl sample preparations were made with 0.18 mg/mL peptide and 5 mM L-Histidine. A quench solution of 100 mM L-methionine and 1µM catalase was dissolved in 50 mM potassium phosphate at pH7. All these reagents were from Sigma Aldrich (Sigma Aldrich, UK).

2.1.3 FPOP protocol

100 µL of each sample were drawn up into a Hamilton 100 µL 710N Syringe **(**Hamilton, USA). This was injected into a 100 µm fused silica diameter capillary at 20 $\mu L min^{-1}$, and flowed in the path of a

110 mj 15hz laser beam, generated by a Compex 50 pro Kr-F 248 nm Excimer laser (Coherent Inc, Ely, UK). Following irradiation, the sample was directed into an Eppendorf with 20 µL quench solution (100 mM L-methionine, 1 µM catalase).

2.1.4 Mass Spectrometry

A vanquish neo liquid chromatography instrument was coupled to a Thermo Orbitrap Eclipse Tribrid EMR Mass Spectrometer (Thermo Fisher, Bremen, Germany). An EasySpray 50 cm x 75 µm, 2 µm particle size column was used for LC separation.

3 µL of 1 pmol concentration of FPOPed peptide was injected onto the column, dissolved in 0.1% Trifluoroacetic acid. The mobile phases used were water and acetonitrile, with 0.1% formic acid. The following LC separation was used:

Table 2.2: LC mobile phase gradient

Minutes	% acetonitrile
0.0	2
0.1	2
35.1	40
35.2	80
45.0	80

The flow rate used was 0.25 µL/min.

Mass Spectrometry Instrument settings

Table 2.3: instrument settings for exploris mass spectrometer

Method Duration	45 mins
Spray voltage	1500V
Default Charge State	2
Advanced peak determination	True
Cycle time	2.5 seconds
Desired minimum points across peak	9
MSn Level	1
Detector type	Orbitrap
Orbitrap Resolution	120K
Scam Range (m/z)	350-2000
Maximum Injection Time	50 ms
AGC target	1000000

Normalised AGC target	250%
Microscans	1
Maximum injection time type	Custom
RF Lens (%)	30
Use ETD internal calibration	False
Polarity	Positive
Source Fragmentation	False

Table 2.3: instrument settings

Filter MIPS

Table 2.4: MIPS settings

Relax Restrictions when too few precursors are found	True
MIPS mode	Peptide

Table 2.4: MIPS settings

Filter Intensity Threshold

Table 2.5: Intensity settings

Maximum intensity	1E+20
Minimum intensity	50000
Relative Intensity Threshold	20
Intensity filter type	Intensity Threshold

Filter Charge State

Table 2.6: Charge state filters

Include Charge State(s)	2-7
Include undermined charge states	False

Filter Dynamic Exclusion

Table 2.7: Dynamic exclusion settings

Exclude after n times	3
Exclusion duration	5s
Mass tolerance	Ppm
Mass tolerance low	10
Mass tolerance high	10
Use common settings	False
Exclude isotopes	true

Data dependent properties – Scan ddMsnScan

Table 2.8: Data dependent settings

Desired minimum points across the peak	9
MSn level	2
Collision Energy Type	Normalised
Isolation mode	Quadrupole
Enable intelligent product acquisition for MS2 isolation	False
Isolation Window	1.2
Isolation Offset	Off
Reported Mass	Original Mass
Multi-notch Isolation	False
Scan Range Mode	Define First Mass
First Mass	100
Scan Priority	1
Activation type	HCD
Collision Energy Mode	Fixed
Collision Energy (%)	30
Detector Type	Orbitrap
Orbitrap Resolution	30K
Maximum injection time	54ms
AGC target	50000
Inject ions for all available parallelizable time	False
Normalised AGC target	100%
Microscans	1
Maximum injection time type	Custom
Use ETD internal calibration	False

2.1.5 Data analysis

Extracted ion chromatograms (XICs) were produced in Freestyle version 1.4 (ThermoFisher, Bremen, Germany). These were produced for the unmodified peptide, +14, +16, +30, +32 and +48, for both the +1, +2 and +3 charge states. Each XIC was produced with a 0.25 atomic mass unit error. Each XIC was manually inspected for signal to noise.

Following this, the detect in active plot tool was used to determine peak areas. These peak areas were exported to excel and total oxidation was calculated using the following equation.

To calculate oxidation at the peptide level, the following equation was used:

$$\%\, modified = \frac{\sum all\ Modified\ peaks}{unmodified + \sum all\ Modified\ peaks}$$

Equation 2.1: calculation of modification of peptides

To calculate oxidation for individual residues the following equation was used:

$$\%\, modified = \frac{\sum Modified\ peaks}{unmodified + \sum all\ Modified\ peaks}$$

Equation 2.2: calculation of modification of individual residues

2.2 HDX Methods

2.2.1 WFL and STT antibodies

MEDI1912_WFL and MEDI1912_STT were provided by Medimmune on dry ice, thawed, aliquoted and snap frozen, in 125 mM arginine and 20 mM sodium succinate, at pH 6.0.

2.2.2 Other materials

Holo-myoglobin was purchased from Sigma Aldrich (Sigma Aldrich, UK). Polysorbate 20 and 80 were provided by Croda (Croda Inc, New Castle, DE).

2.2.3 Polysorbate conditions

For stability studies, 1 mg/ml of myoglobin, dissolved in HPLC grade water (fisher Scientific, UK), was combined with 0.5% w/v of either tween 20 or 80, and stored at 23®C for one month.

For WFL and STT stability studies, 1 mg/ml WFL and STT were combined with 0.5% w/v polysorbate 20 and 80, and dissolved in 125 mM arginine and 20 mM sodium succinate, at pH 6.0. These were stored at 23®C for one month.

For experiments with fresh polysorbate, the same process was followed, with incubation at room temperature for 30 minutes before analysis.

2.2.4 HDX conditions

HDX was performed at the mass spectrometry facility at the University of Leeds.

HDX-MS experiments were carried out using an automated HDX robot (LEAP Technologies, Fort Lauderdale, FL, USA) coupled to an M-Class Acquity LC and HDX manager (Waters Ltd., Wilmslow, Manchester, UK). Protein was diluted to 10 µM in equilibration buffer (50 mM potassium phosphate, pH 7. 5 µl sample was added to 95 µl deuterated buffer (50 mM potassium phosphate, pD 7) and incubated at 4 °C for 2, 5 or 10 min. Following the labelling reaction, samples were quenched by adding 75 µl of the labelled solution to 75 µl quench buffer (50 mM potassium phosphate, 0.05% DDM pH 2, 4 M guanidine 0.7 M TCEP for antibody samples) giving a final quench pH ~ 2.5. 50 µl of quenched sample were passed through a home-packed pepsin column using agarose immobilised pepsin (Thermo Fisher Scientific) at 40 µl min−1 (20 °C) and a VanGuard Pre-column Acquity UPLC BEH C18 (1.7 µm, 2.1 mm × 5 mm, Waters Ltd., Wilmslow, Manchester, UK) for 3 min in 0.3% formic acid in water. The resulting peptic peptides were transferred to a C18 column (75 µm × 150 mm, Waters Ltd., Wilmslow, Manchester, UK) and separated by gradient elution of 0–40% MeCN (0.1% v/v formic acid) in H_2O (0.3% v/v formic acid) over 7 min at 40 µl min^{-1}. Trapping and gradient elution of peptides was performed at 0 °C. The HDX system was interfaced to a Synapt G2Si mass spectrometer (Waters Ltd., Wilmslow, Manchester, UK). $HDMS_e$ and dynamic range extension modes (Data Independent Analysis (DIA) coupled with IMS separation) were used to separate peptides prior to CID fragmentation in the transfer cell. HDX data were analyzed using PLGS (v3.0.2) and DynamX (v3.0.0) software supplied with the mass spectrometer. Restrictions for identified peptides in DynamX were as follows: minimum intensity: 10000, minimum products per MS/MS spectrum: 3, minimum products per amino acid: 0.3, maximum sequence length: 18, maximum ppm error: 10, file threshold: 3/3.

Butterfly plots, plots comparing the deuterium uptake of each peptide, were produced in DynamX to show each time point for each experiment (2, 5 and 10m). These were manually inspected to determine congruence between all three time points.

Following manual curation of the data, summary plots were generated using Deuteros 2.0 (Lau et al., 2021).

In Deuteros, a P value of < 0.02 was used, with hybrid statistics. Additionally, an arbitrary cutoff for significance was used of 0.5 Da.

2.2.5 Data plotting

For myoglobin, significantly different changes were plotted onto the crystal structure from protein data bank file 1MBN.

Antibody structures were labelled on a generic antibody structure from (Spiteri et al., 2021).

2.3 MtrC Methods

2.3.1 MtrC production and synchrotron labelling

MtrC was produced by Dr Anna Stikane (Stikane, 2020), and all synchrotron experiments run by her at beamline 3.2.1 at the Advanced Light Source at the Lawrence Berkley National laboratory.

MtrC and MtrC with TiO_2 were labelled at 5 different exposure times, 0, 12.5, 25, 50 and 75 ms.

2.3.2 Mass spectrometry

The digestion of peptides was performed by Dr Rachel George at the University of Leeds, as described in Dr Anna Stikane's thesis (Stikane, 2020). This has been covered in section 5.2.

2.3.3 Data analysis

Peptides were de-novo sequenced using PEAKS X software (bioinformatics solutions Inc, waterloo, ON, Canada). Peptides with a +16 modification and +32 were identified through this software. For

peptides where modifications were identified in the software, XICs were produced using MassLynx 4.1, for each charge state and modification identified in PEAKS.

From Masslynx, peak areas for each of the XICs (unmodified, +16, +32) were identified and exported to Excel for each fraction.

At the peptide level, the total oxidation for each fraction was calculated using the below equation: To calculate oxidation at the peptide level, the following equation was used:

$$\%\, modified = \frac{\sum all\ Modified\ peaks}{unmodified + \sum all\ Modified\ peaks}$$

Equation 2.3 calculation of modified peptides for synchrotron oxidation

These fractions were converted to a ratio such that:

$$Fraction\ unmodified = 1 - fraction\ modified$$

Equation 2.4: Calculation of fractions modified in synchrotron oxidation

For each fraction, a dose response curve was produced (found in supplemental information), by using this ratio (a peptide rate constant) by a first order fit (y=exp(-A*x)) in originpro (OriginLab Corporation, Northampton, MA, USA).

R values were determined at the peptide level by calculating a ratio of MtrC to MtrC + TiO_2. For each peptide, if the ratio was above 2.5, it was determined significant. A value of 2 has been used previously in literature for significant difference in residue level calculations (Fukushima et al., 2017). However, here, the requirement for statistical significance was increased at the peptide level due to the higher variability in peptide measurements in comparison to residue level measurements.

This same approach was used at the residue level. However, for determination of the site at the peptide level, MS/MS was used to identify the location of the modified residue before undertaking the same analysis.

2.3.4 Electrostatic calculations

Electrostatic calculations were performed by collaborator Marcus J Edwards (University of Essex).

The c-type hemes of MtrC (minus iron atoms) were parameterised using PRODRG (Schüttelkopf and van Aalten, 2004). The charges and radii of atoms were calculated using PDB2PQR (Dolinsky et al., 2007) using the Amber force field at pH 8. The linearized Poisson–Boltzmann equation was solved with APBS (Jurrus et al., 2018) using cubic spline charge discretization, with dielectric constants of 2.00 for the solute and 78.54 for the solvent at 298.15 K and including ions corresponding to 5 mM NH_4Cl. Electrostatic potentials are displayed as surface potential maps (± 5 KbT/e) using Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.).

Chapter 3:

FPOP of peptides

3. FPOP of peptides

3.1 introduction

Due to amino acid residues having different reactivity towards OH radicals, the total amount of labelling is highly dependent on the specific residue. Data from synchrotron radiation oxidation experiments has shown a 3 order of magnitude difference between the most reactive residue (cysteine), and the least reactive residue (glycine)(Xu, Guozhong and Chance, Mark R., 2007). However, no similar work has been performed from oxidation via FPOP.

Observations in the literature suggest that proximity of reactive residues within a sequence can effect labelling of the residues. Xie et al(Xie, Boer et al., 2017) examined the correlation between normalised protection factor (NPF), and fractional solvent accessible surface area (SASA). Whilst highly and moderately reactive residues shared a high correlation between NPF and fractional SASA, low reactivity residues correlate poorly. This indicates that sequence composition is important for labelling, and the local neighbourhood can have an impact on the total oxidation of a single residue.

Sharp et al(Sharp and Tomer, 2006) analysed the oxidation of leucine and aspartic acid residues in small peptides separated by an increasing number of unreactive glycine residues. This work identified significant fold change differences for several peptides, dependent in part on residue distance.

Further, Cornwell et al(Cornwell, O. et al., 2018b) reported a threefold decrease in labelling for Valine 27 in the ΔN6 variant of ß$_2$microglobulin compared to the wild type despite the crystal

structure suggesting an approximately 22-fold increase in SASA. The author suggests this unexpected decrease is due to competition for radicals by the exposed phenylalanine within close proximity.

Highly reactive residues in close proximity has not been probed in the literature, however it is fundamental to determine if two highly reactive residues effect total oxidation percentage. This is particularly pivotal at the peptide level, where more reactive residues contribute more to the total global oxidation percentage of the peptide than less reactive residues. Thus, if highly reactive residues reduce total peptide oxidation, protein wide FPOP experiments at the peptide level could yield erroneous results. Tryptophan and phenylalanine are both highly reactive with OH• radicals. Furthermore, they both are aromatic in nature, with tryptophan containing a pyrrole and benzene ring, and phenylalanine containing a benzene ring. Both of these amino acid groups occupy a relatively large area in space in comparison to non-aromatic residues. Therefore, should there be interactions between these two groups due to steric effects, study of the stereochemistry, for example isomerism could allow for the determination of this effect, as a particular isomer may be favoured if there is interactions between the two.

The distance between residues is important. When residues are directly neighbouring in primary amino acid sequence, the r groups oppose each other in three-dimensional space. Amino acids placed two residues apart, i and i+2, face the same direction in three-dimensional space. Therefore, by probing the reactivity of neighbouring residues and those at positions i and i +2, it can be determined whether three-dimensional structure has any implication to labelling in comparison to primary amino acid sequence. This does not take into account secondary structure formation of the peptides, especially in longer chains, and thus the specific distance in space in solution.

Finally, should amino acid sequence be the primary driving factor, by having the residues more than one residue apart, for example five residues, it can be determined that when amino acids are further apart, the neighbourhood effect is reduced and labelling is increased across the peptide.

Herein, we attempt to probe the factors responsible for labelling. By changing the distance between reactive residues by zero, one and five residues, it is hoped the change in oxidation due to distance in amino acid sequence can be defined. Further, by interrogating aromatic residues such as tryptophan and phenylalanine, it is hoped interactions such as pi stacking can be probed. Tryptophan has a larger r group than phenylalanine, constituting a pyrrole and benzene ring compared to the benzene ring of phenylalanine. Thus, it is hoped to determine if the larger more reactive r group interacts with residues further in distance than the smaller less reactive phenylalanine. Further, the constitution of oxidation of the positional isomers can also be used to determine changes in reactivity due to location, as the isomers closer in space to other reactive residues may see a decrease in oxidation.

Utilising the unreactive residue glycine as a spacer between reactive residues should allow for the competition between these residues to be probed without radical dosage being decreased by reaction with glycine.

Variation of the reactive residue selected allows for findings to be confirmed. By using two of the same residue, if the effect is duplicated as observed with different residues, the rules surrounding neighbourhood effects can be extrapolated to other residues, and should allow for a prediction of rules for labelling effects.

Here we oxidise by FPOP 6 peptides, containing tryptophan (W) and phenylalanine (F) spaced zero, one and five residues apart with glycine (G) as the spacer. Lysine (K) residues at each end of the peptides increased solubility.

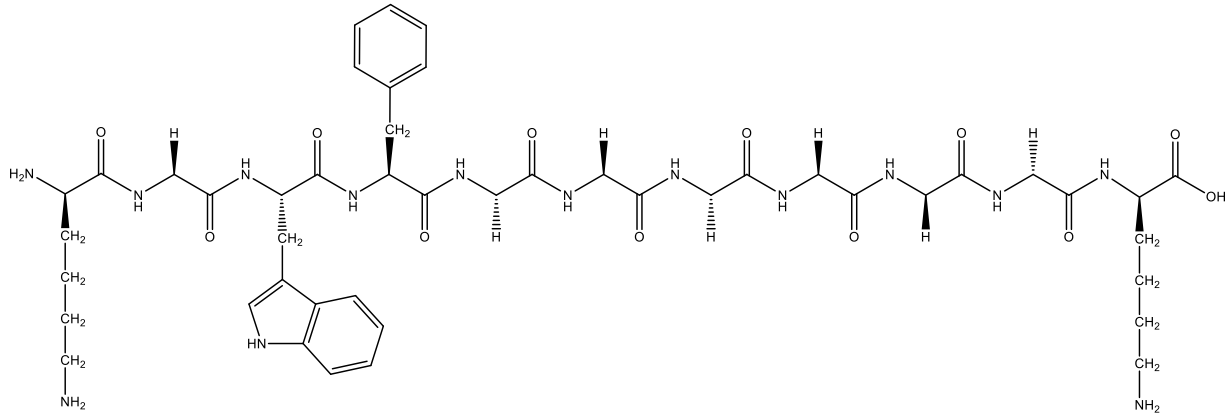


Figure 3.1: Peptide KGWFGGGGGGK abbreviated to W0F. This peptide has W and F residues sequentially in the peptide. The peptide is capped by lysine residues to improve solubility

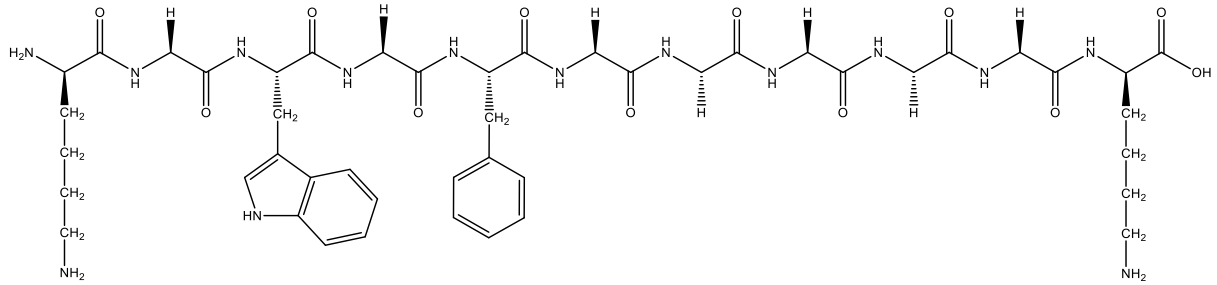


Figure 3.2: peptide KGWGFGGGGGK abbreviated to W1F This peptide has W and F residues separated by a single glycine residue in the peptide. The peptide is capped by lysine residues to improve solubility

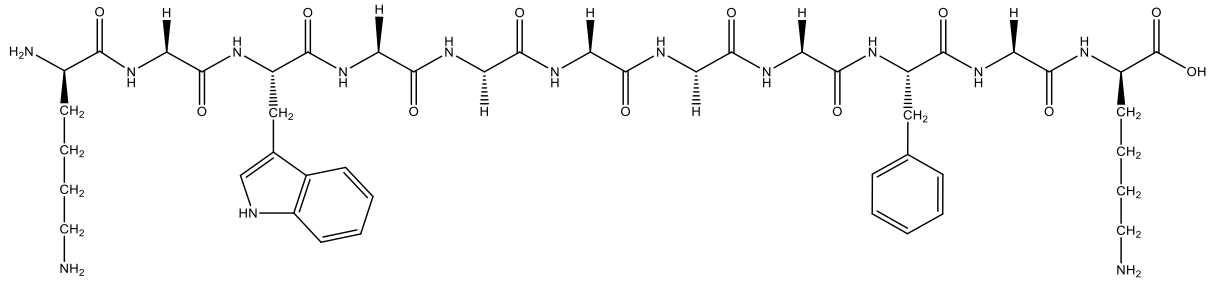


Figure 3.3: Peptide KGWGGGGGFGK abbreviated to W5F. This peptide has W and F residues separated by 5 glycine residue in the peptide. The peptide is capped by lysine residues to improve solubility

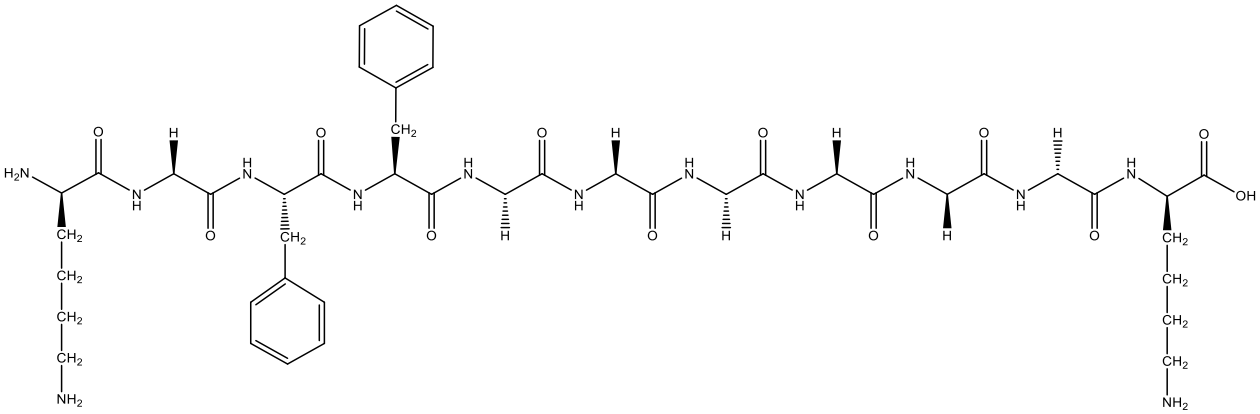


Figure 3.4: Peptide KGFFGGGGGGK abbreviated to F0F. This peptide has two F residues sequentially in the peptide. The peptide is capped by lysine residues to improve solubility

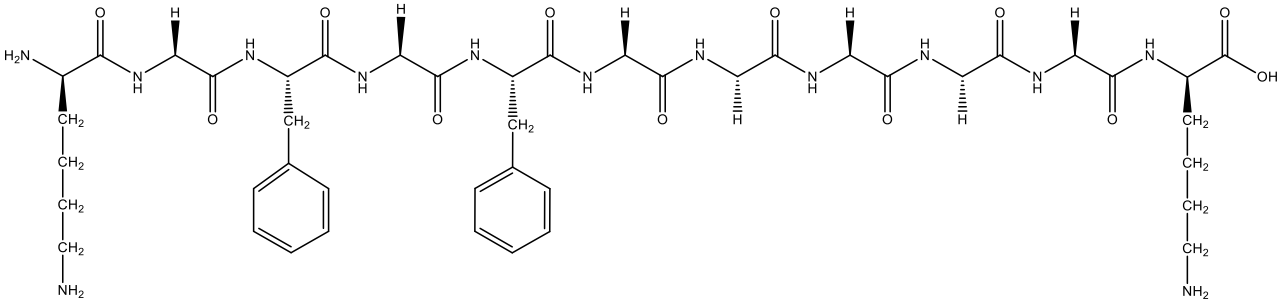


Figure 3.5: peptide KGFGFGGGGGK abbreviated to F1F. This peptide has two F residues separated by a single glycine residue in the peptide. The peptide is capped by lysine residues to improve solubility

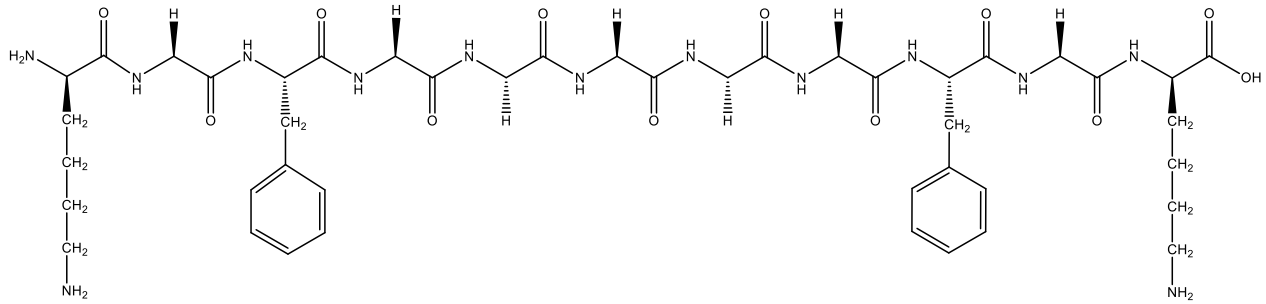


Figure 3.6: Peptide KGFGGGGGFGK abbreviated to F5F. This peptide has two F residues separated by five glycine residues in the peptide. The peptide is capped by lysine residues to improve solubility

Tryptophan forms five +16 side chain isomers following oxidation, one on the pyrrole ring and four on the benzene ring(Xu, Guozhong and Chance, Mark R., 2007) (figure 3.7). Further, phenylalanine can have three +16 modifications in the ortho, meta and para positions on the benzene ring(Xu, Guozhong and Chance, Mark R., 2007) (figure 3.7).

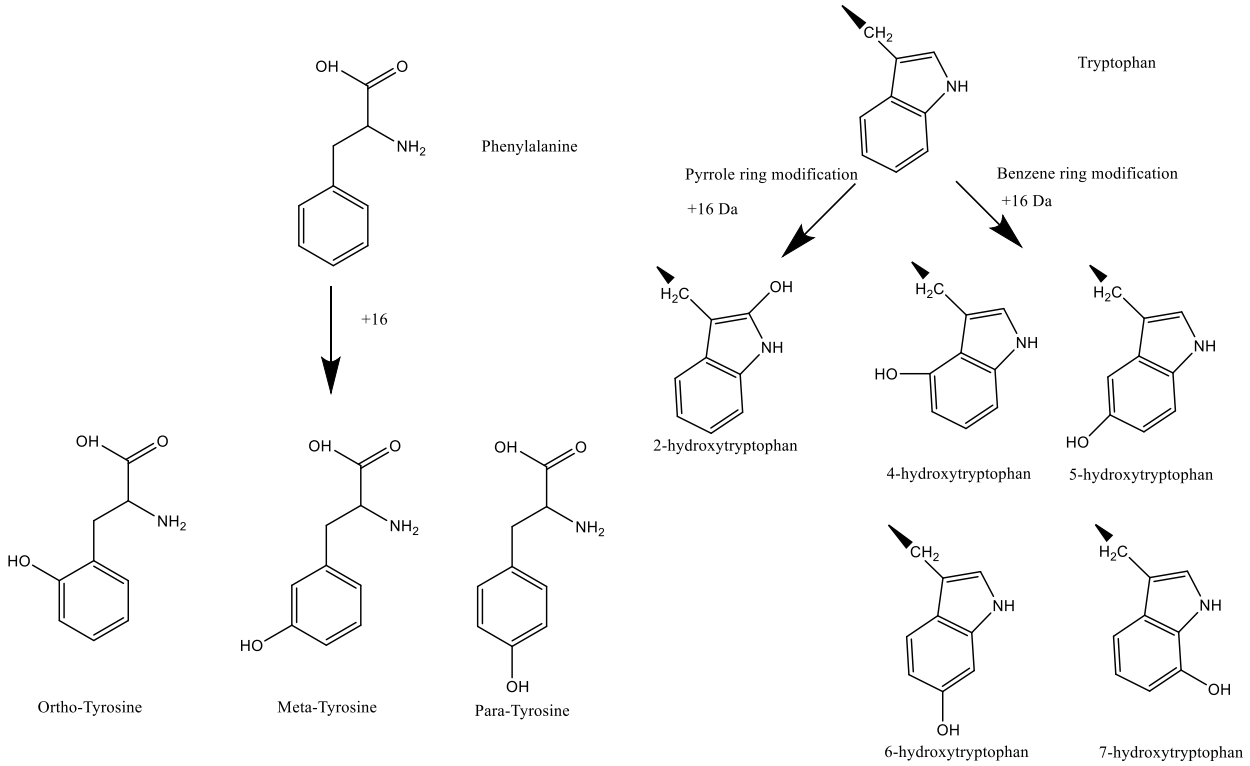


Figure 3.7: oxidation pathways of tryptophan and phenylalanine. Upon reaction with hydroxyl radicals, phenylalanine forms three distinct iosmers, ortho-tyrosine, meta-tyrosine and para-tyrosine. Tryptophan forms five isomers, one on the pyrole ring and four on the benzene ring.

Through Liquid Chromatography Mass Spectrometry (LC-MS), oxidised peptides can be separated, including positional isomers(Cornwell, O. et al., 2018b), and the location of the modification identified within the peptide chain via tandem MS (Cornwell, O. et al., 2018b; Cornwell, Owen et al., 2021). Therefore, by isolating each chromatogram peak, identifying modification location, and determining the area of each peak, the total oxidation of each amino acid in the chain can be determined.

3.2 Methods

3.2.1 Peptide acquisition

Peptides were synthesised by Peptide Protein Research ltd (Bishops Waltham, UK).

3.2.2 FPOP

The FPOP experiment was performed as described previously(Calabrese et al., 2015)**.** Briefly, 1 µL 5% v/v H_2O_2 was added to 100 µL of 0.18 mg/mL peptide, dissolved in potassium phosphate buffer pH 7.2 and 20mM L-histidine. This was injected into a 100 µm diameter capillary at 20 $\mu L min^{-1}$ and flowed in the path of a 110 mJ 15 Hz laser beam, generated by a Compex 50 pro Kr-F 248 nm Excimer laser (Coherent Inc, Ely, UK). Following irradiation, the sample was directed into an Eppendorf with 20 µL quench solution (100 mM L-methionine, 1 µM catalase). Three replicates and a control were labelled per peptide.

3.2.3 LC-MS

A Vanquish Neo liquid chromatography instrument was coupled to a Thermo Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher, Bremen, Germany). A EasySpray 50 cm x 75 µm, 2 µm particle size column was used for LC separation.

MS was performed in DDA mode, with the top 3 most intense ions selected for fragmentation. Ions were excluded for reselection for 3 s following selection.

3 µl of 1 pmol concentration was injected onto the column, dissolved in 0.1% Trifluoroacetic acid. The mobile phases used were water and acetonitrile, with 0.1% formic acid. Modification identification

Modifications were manually curated, and data manually quantified using Freestyle software version 1.4 (ThermoFisher, Bremen, Germany)**.** Data was quantified by extracting ion chromatograms (XICs) for all 2+ charge states for the unmodified peptide, +14, +16, +30, +32 and +48 modification states.

To calculate oxidation at the peptide level, the following equation was used:

$$\%\, modified = \frac{\sum all\ Modified\ peaks}{unmodified + \sum all\ Modified\ peaks}$$

Equation 3.1: calculation of modification of peptides

To calculate oxidation for individual residues the following equation was used:

$$\%\, modified = \frac{\sum Modified\ peaks}{unmodified + \sum all\ Modified\ peaks}$$

Equation 3.2: calculation of modification of individual residues

3.3 Results

3.3.1 Peptide level

Due to a low signal and signal to noise in the 1+ and 3+ charge states, only the 2+ charge state was used.

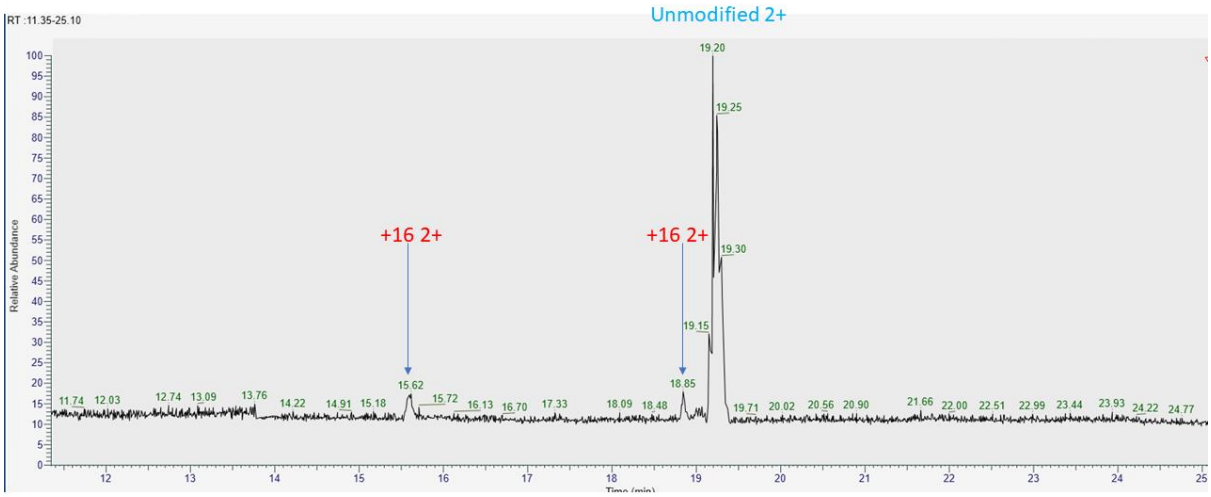


Figure 3.8. Base peak intensity chromatogram for W0F. Only the +2 charge states can be identified above the signal to noise. Two distinct +16 peaks are shown in the chromatogram, with the unmodified +2 charge state

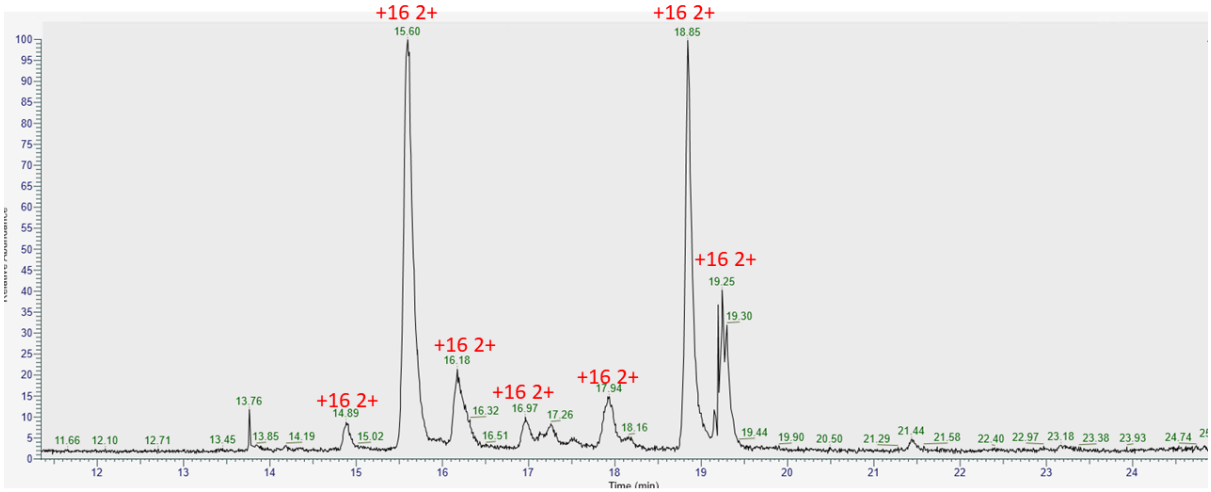


Figure 3.9: XIC of 512.25 for the W0F sample. The +2 charge state for the +16 modification from figure 3.8. Seven +16 modified peaks can be identified in the XIC, showing the isomers for both phenylalanine and tryptophan.

W0F showed a total oxidation percentage of 48.8% and a standard deviation of 3.2% (n=3). W1F showed a total oxidation of 54.8% and a standard deviation of 2.59% and W5F showed a total oxidation of 62.6% and a standard deviation of 3.7% (Figure 3.10).

We see an increase of oxidation at the peptide level as the residues get further apart in primary sequence, with W0F having 13.8% less total oxidation in comparison to W5F, exceeding standard deviation for both samples (Figure 3.10).

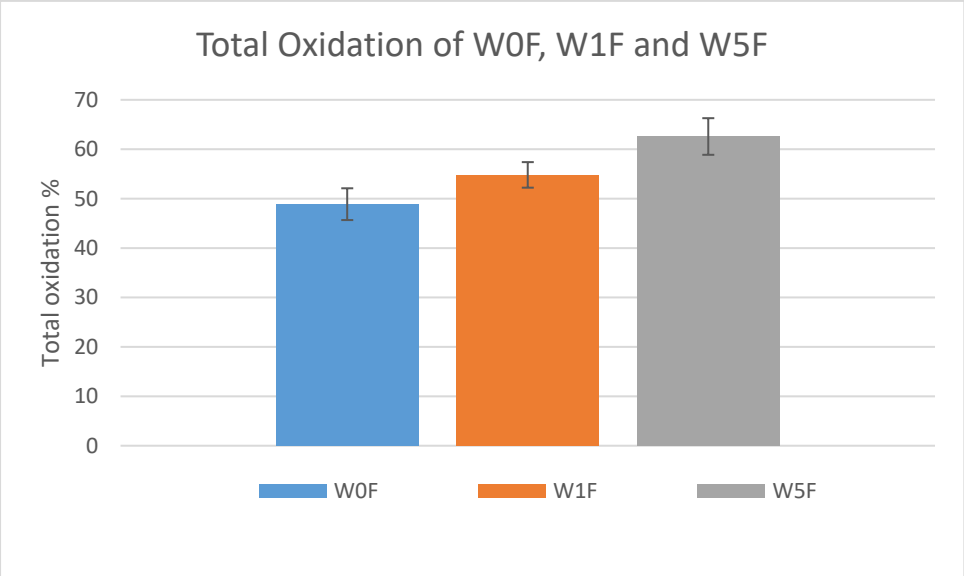


Figure 3.10: Increase of total oxidation for peptides with increasing distance between W and F. W0F had a total oxidation of 48.8%, W1F 54.8% and W5F a total oxidation of 62.6%. The difference in oxidation between W0F and W5F was above the standard deviation.

F0F had a 26.5% total oxidation with 6.0% standard deviation, F1F had 41.8% total oxidation with 3.3% standard deviation and F5F has 45.3% total oxidation with 3.9% standard deviation (Figure 3.11).

F5F has 18.8% total oxidation more than F0F, which is significantly different. F1F has 15.3% total oxidation more than F0F, which is significantly different.

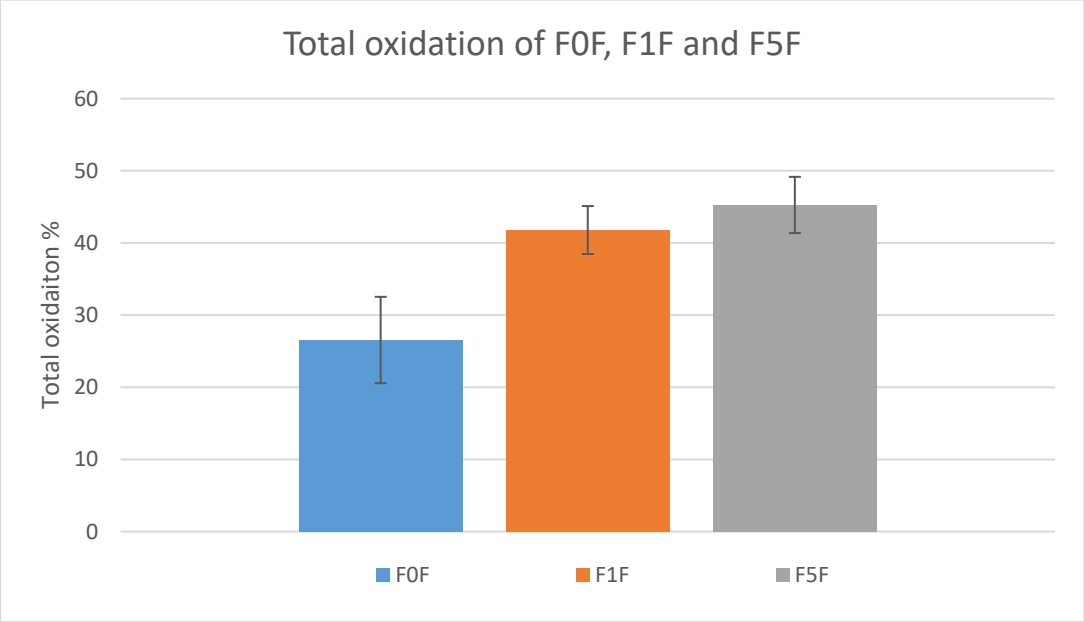


Figure 3.11: increase of oxidation for residues with increasing distance between F and F. F0F had a total oxidation of 26.5%, F1F had a total oxidation of 41.8% and F5F had a total oxidation of 45.3%. Both F1F and F5F are significantly different to F0F.

One way ANOVA was performed on the data in figure 3.10 and 3.11, (table 8.9 and 8.10 in the supplemental information). These showed significant differences between W0F and W5F, and between F0F and F1F and F5F.

For the oxidation distribution of W0F (Figure 3.12), +14 constituted 4.65% of the total oxidation, +16 constituted 39.96%, +30 constituted 34.91% +32 constituted 16.87% and +48 constituted 4.65%.

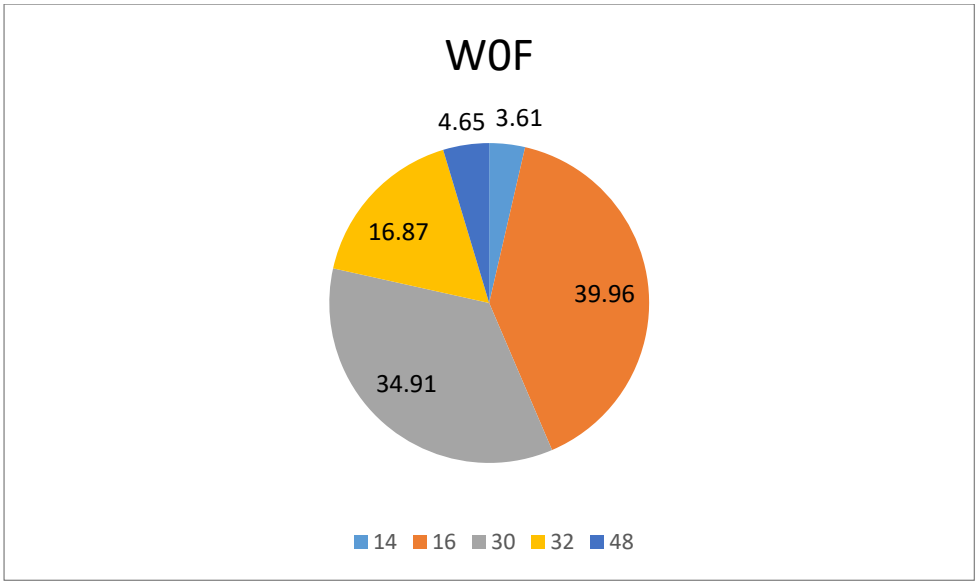


Figure 3.12: Oxidation products of the W0F peptide, with percentages for different types of observed mass increases (+14, +16, +30, +32 and +48 Da). The +16 modification contributes 39.96% to the total oxidation, with +30 modification contributing 34.91% and +32 contributing 16.87%.

For the oxidation distribution of W1F (figure 3.13), +14 constituted 3.55% of the total oxidation, +16 constituted 41.57%, +30 constituted 29.90% +32 constituted 21.29% and +48 constituted 3.70%.

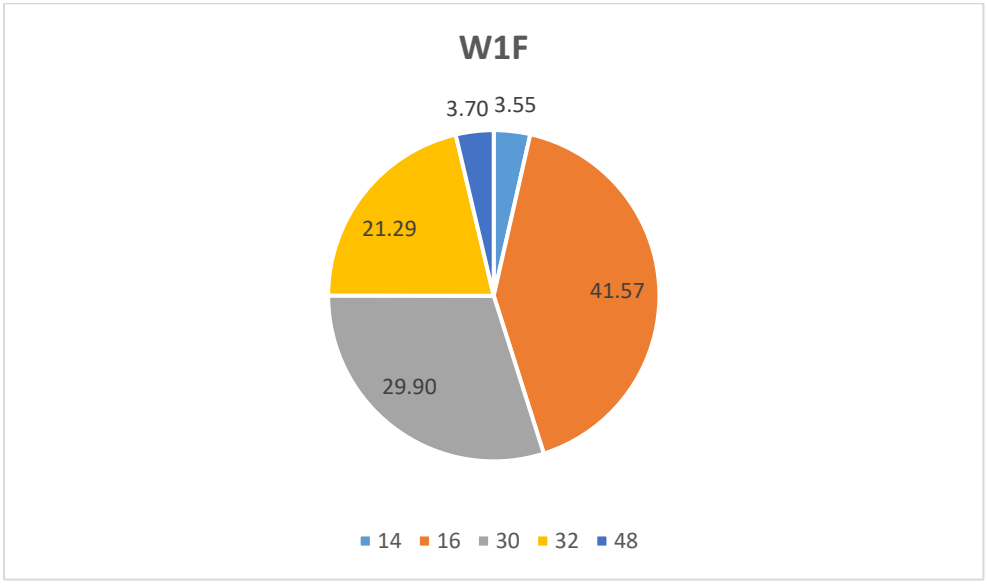


Figure 3.13: Constitution of total oxidation of W1F, with percentages for different types of observed mass increases (+14, +16, +30, +32 and +48 Da). The +16 modification contributes 41.57% to the total oxidation, with +30 contributing 29.90% and +32 contributing 21.29%.

For the oxidation distribution of W5F (figure 3.14), +14 constituted 3.34% of the total oxidation, +16 constituted 44.72%, +30 constituted 35.01% +32 constituted 12.93% and +48 constituted 4.01%.

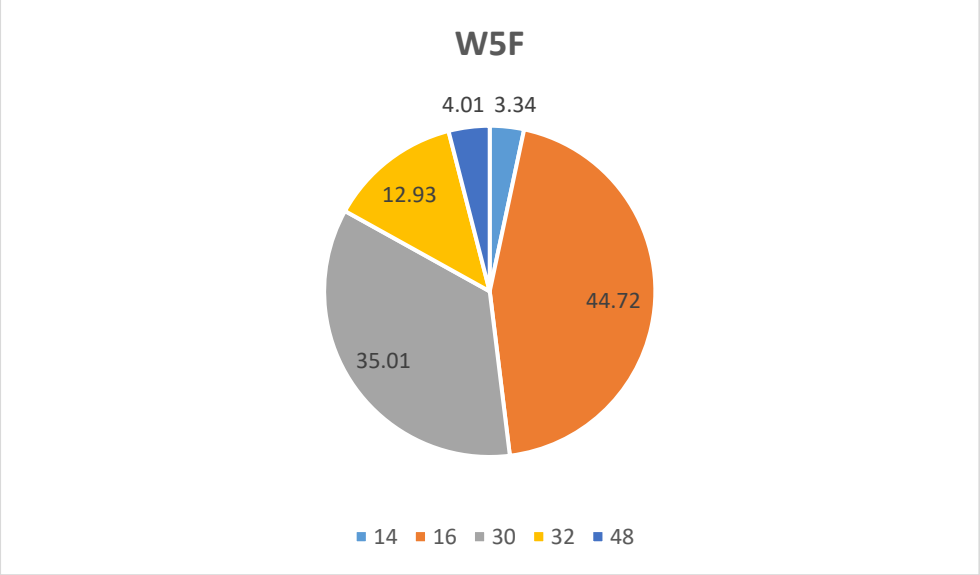


Figure 3.14: Constitution of total oxidation of W5F, with percentages for different types of observed mass increases (+14, +16, +30, +32 and +48 Da). The +16 modification contributes 44.72% to the total oxidation, with +30 contributing 35.01% and +32 contributing 12.93%.

For the oxidation distribution of F0F (figure 3.15), +14 constituted 9.29% of the total oxidation, +16 constituted 66.83%, +30 constituted 17.44% and +32 constituted 6.44%.

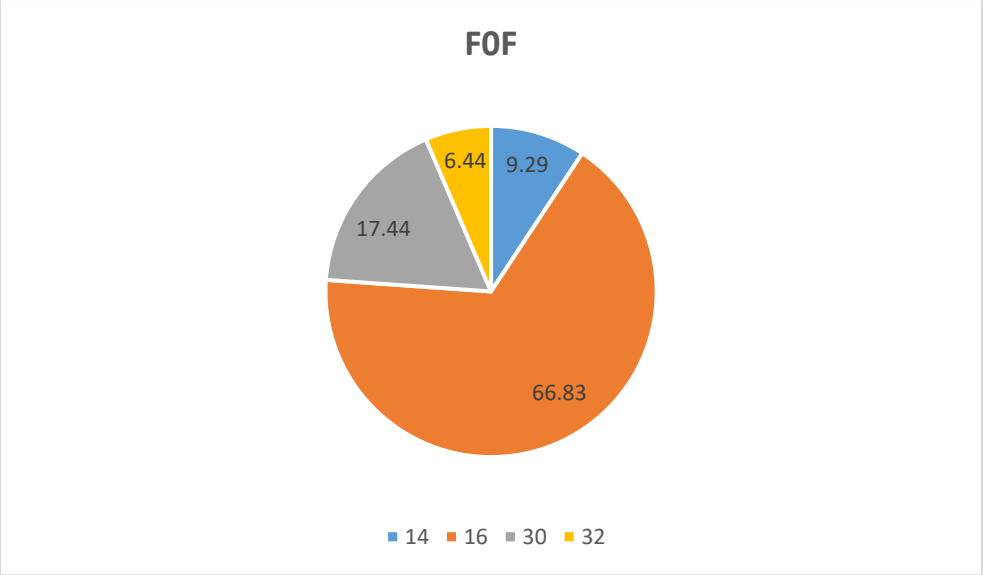


Figure 3.15: Constitution of total oxidation of F0F, with percentages for different types of observed mass increases (+14, +16, +30 and +32 Da). The +16 modification contributes 66.83% to the total oxidation, with +30 contributing 17.44% and +32 contributing 6.44%.

For the oxidation distribution of F1F (figure 3.16), +14 constituted 14.67% of the total oxidation, +16 constituted 56.45%, +30 constituted 18.04% and +32 constituted 10.84%.

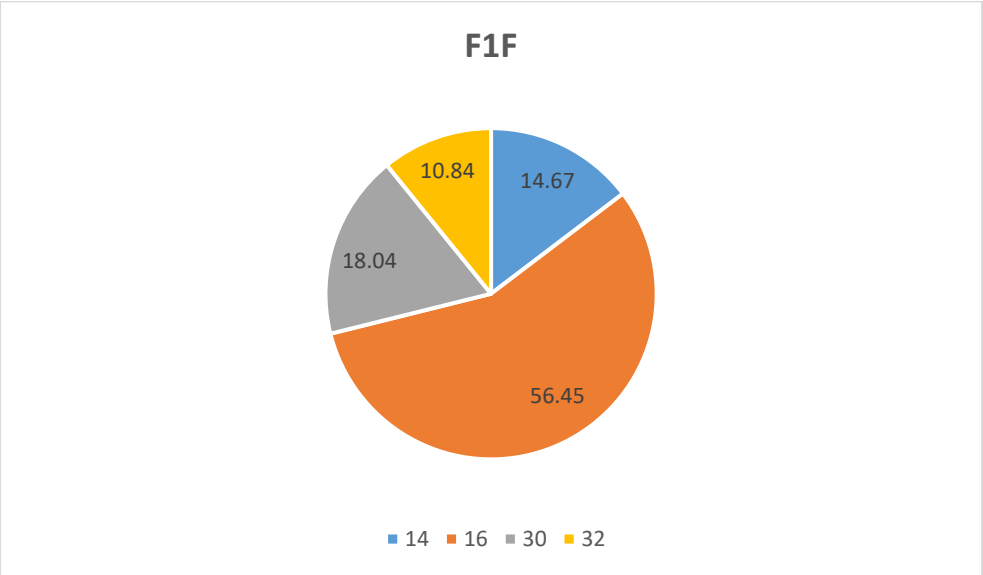


Figure 3.16: Constitution of total oxidation of F1F, with percentages for different types of observed mass increases (+14, +16, +30 and +32 Da). The +16 modification contributes 56.45% to the total oxidation, with +30 contributing 18.04% and +32 contributing 10.84%.

For the oxidation distribution of F5F (figure 3.17), +14 constituted 15.49% of the total oxidation, +16 constituted 57.83%, +30 constituted 14.73% and +32 constituted 11.95%.

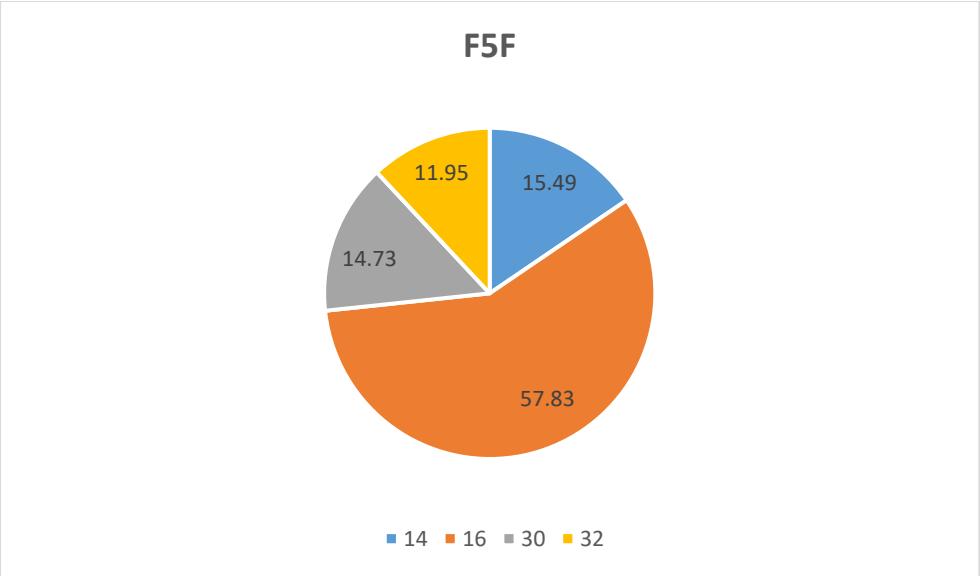


Figure 3.17: Constitution of total oxidation of F5F, with percentages for different types of observed mass increases (+14, +16, +30 and +32 Da). The +16 modification contributes 57.83% to the total oxidation, with +30 contributing 14.73% and +32 contributing 11.95%.

3.3.2 Individual residues

As discussed in the methods, determination of the location of each modification was determined through CID and location of the +16 Da increases. The analysis was performed using DDA, with the three most intense ions being selected. Following selection three times within a 30 s time frame, the m/z was excluded from selection for five s.

The fragment ions for the unmodifed are shown in the supplemental information. For peptide W0F, +16 isomers eluted as seven discrete peaks (Figure 3.18), indicating co elution of several isomers. The +16 XIC for the peaks at 15.5, 16.8, 17.4, 17.9 and 18.6 (figure 3.18) show modification of the tryptophan residue. The MS/MS fragmentation ions for unmodified peptide are shown in the supplemental information**.** Within the spectrum, m/z 636 (y8) identifies no modification to F, m/z 489 (y7) identifies no modification for K11 and m/z 186 (b2) shows no modification for K1. m/z 877 (b10+16) identifies this modification is located on W, as this mass shift is present after loss of K11.

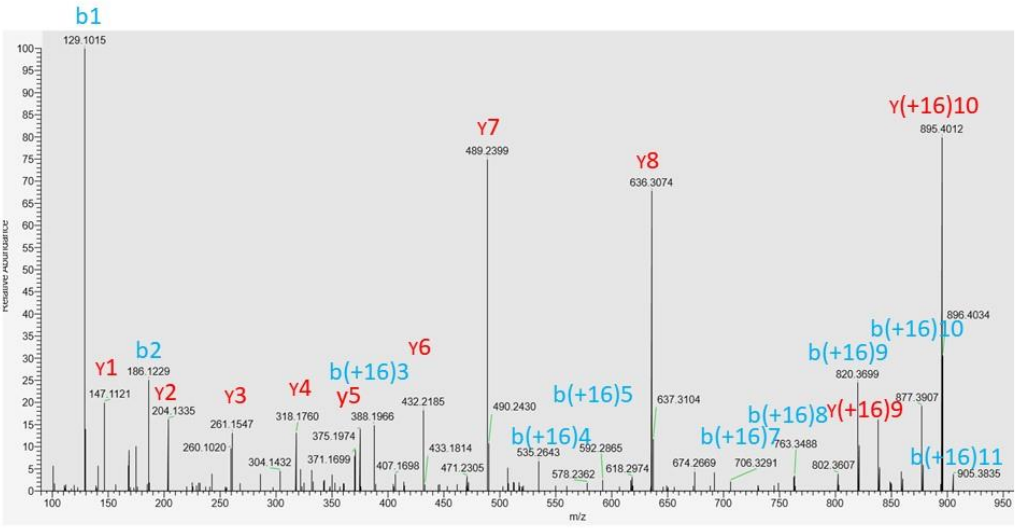


Figure 3.18: 512.25 m/z ms/ms mass spectrum for W modifcation identification. Within the spectrum, m/z 636 (y8) identifies no modification to F, m/z 489 (y7) identifies no modification for K11 and m/z 186 (b2) shows no modification for K1. m/z 877 (b10+16)

For the peak at 16.1 minutes (figure 3.19), the +16 mass spectrum shows elution of modified phenylalanine. Within the spectrum, m/z 489 (y7) identifies no modification for K11, and m/z 186 (b2) shows no modification to K1. m/z 652 (y8 + 16) with the presence of m/z 489 (y7) identifies a

modification to F. the lack of presence of m/z 636 (y8), identifies no modification to W. Therefore, it is concluded there is a modification to F.

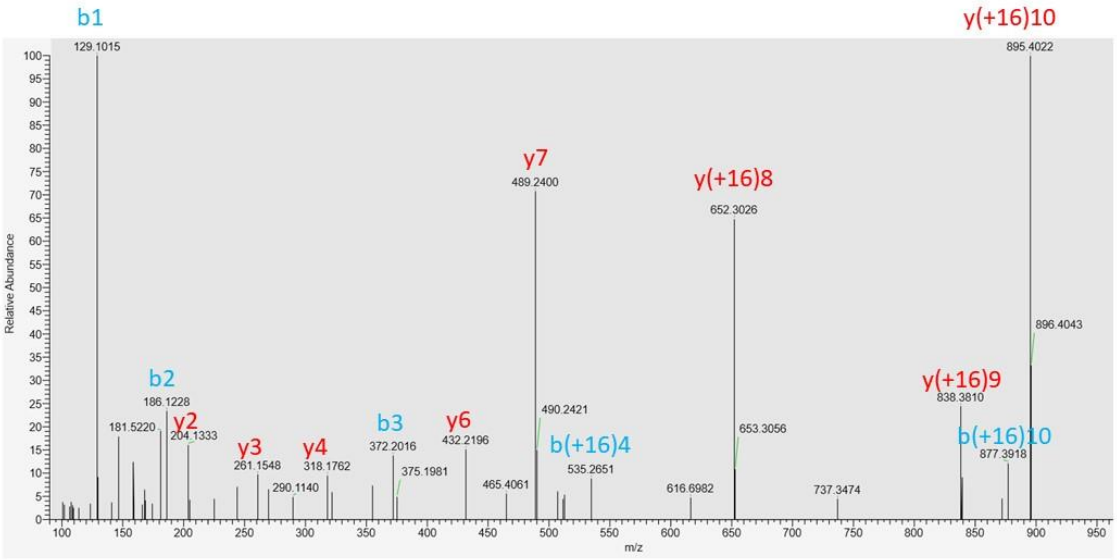


Figure 3.19: 512.25 MS/MS mass spectrum for F modification identification. Within the spectrum, m/z 489 (y7) identifies no modification for K11, and m/z 186 (b2) shows no modification to K1. m/z 652 (y8 + 16) with the presence of m/z 489 (y7) identifies a modification to F. the lack of presence of m/z 636 (y8), identifies no modification to W

For the peak at 15.1 (figure 3.20), the +16 XIC shows coelution of modified tryptophan and phenylalanine. Within the spectrum, M/z 489 (y7) identifies no modification for K2, and m/z 186 (b2) shows no modification to K1. m/z 652 (y8 + 16) with the presence of m/z 489 (y7) identifies a modification to F. However, the presence of m/z 636 (y8), coupled with m/z 895 (y10 +16) identifies a modification to W. Therefore, we conclude there is coelution of modifications to W and F.

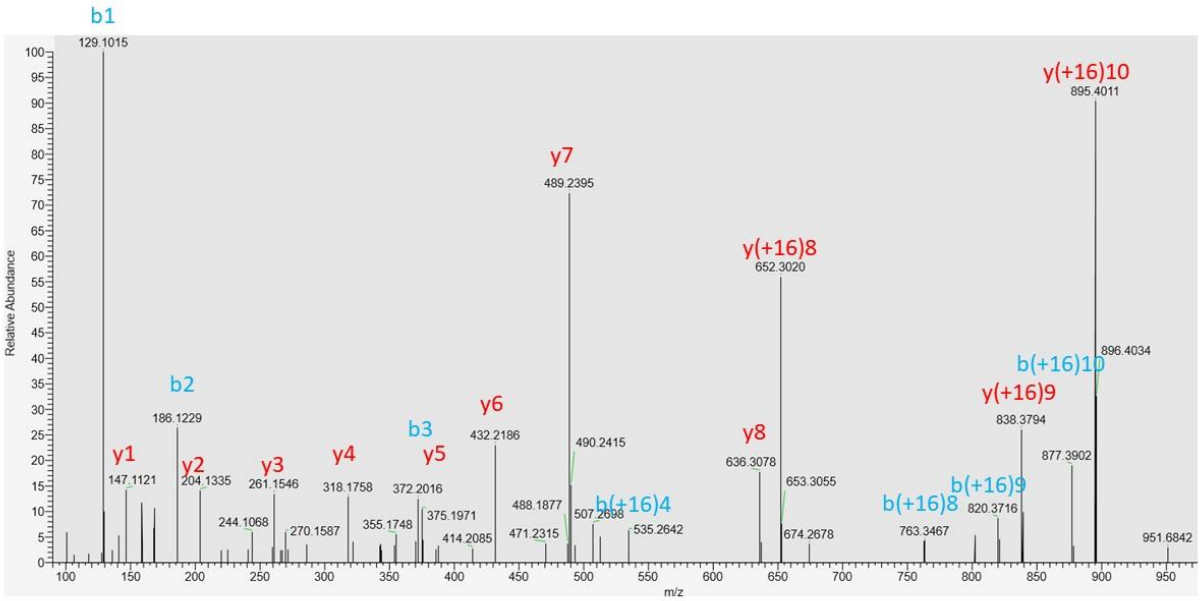


Figure 3.20: 512.25 ms/ms mass spectrum for W and F coelution. Within the spectrum, m*/z 489 (y7) identifies no modification for K2, and m/z 186 (b2) shows no modification to K1. m/z 652 (y8 + 16) with the presence of m/z 489 (y7) identifies a modification to F. However, the presence of m/z 636 (y8), coupled with m/z 895 (y10 +16) identifies a modification to W*

For the W0F +16 five individual tryptophan peaks can be observed, with one for phenylalanine and one peak coeluting both W and F (Figure 3.21). The peaks are not fully resolved, and a coleulting peak can observed with both W and F modified residues.

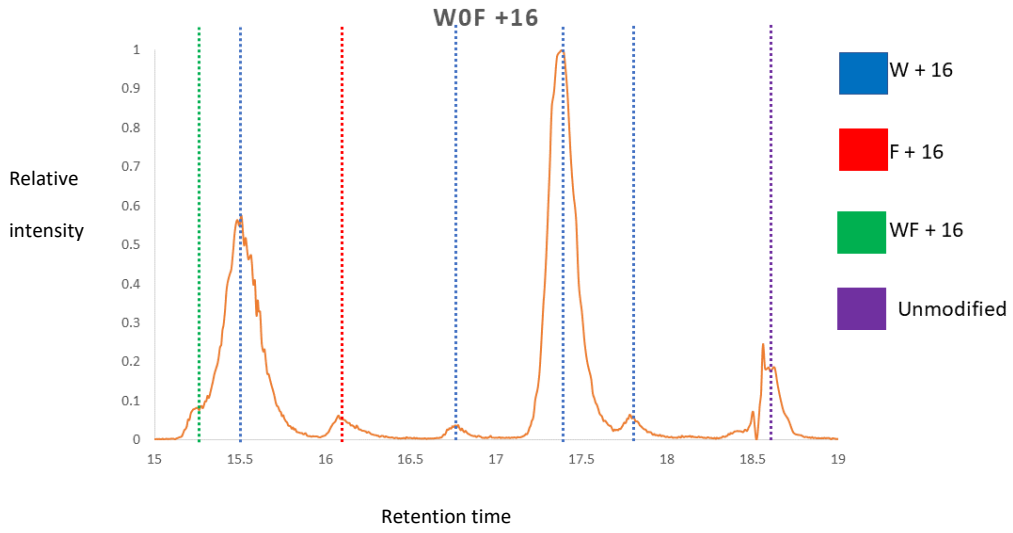


Figure 3.21: Chromatogram for W0F. Seven discrete peaks are shown, four for a modification on tryptophan, one for a modifcation on phenylalaine and a coeluting peak. The peaks at 15.2-15.7 are coeluting, identifying a lack of resolution with the chromatographic method.

For the W1F +16 four individual tryptophan peaks can be observed, with two for phenylalanine and one peak coeluting both W and F (Figure 3.22).

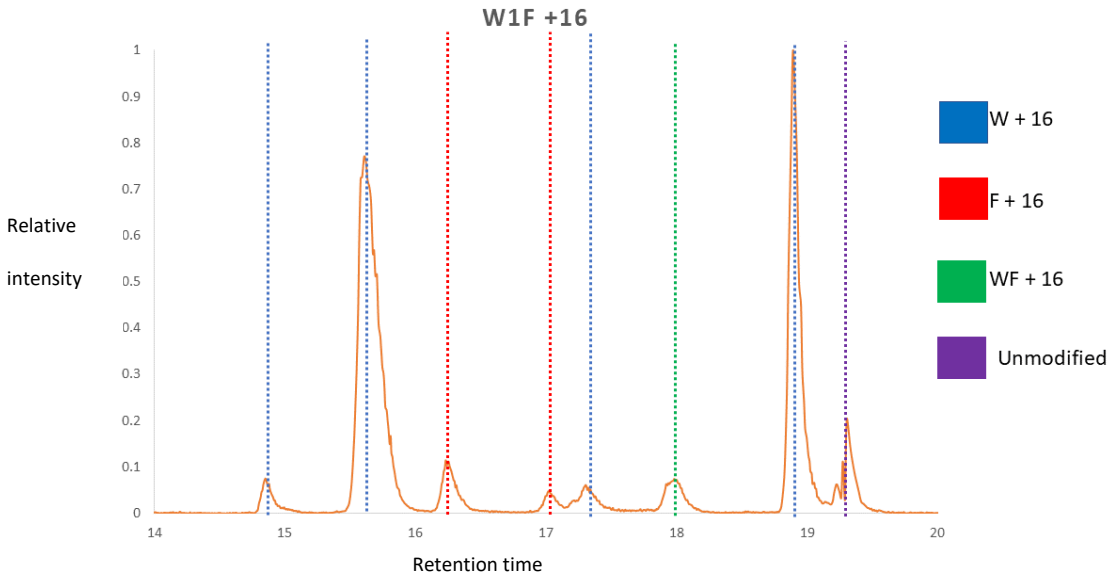


Figure 3.22: Chromatogram for W1F. Eight discrete peaks are shown, three for a modification on tryptophan, two for a modifcation on phenylalaine and a coeluting peak. All peaks are chromatographically resolved. However, the peak containing WF coelution shows a lack of resolution in the chromatographic method

For the W5F +16 three individual tryptophan peaks can be observed, with one for phenylalanine and one peak coeluting both W and F (Figure 3.23).

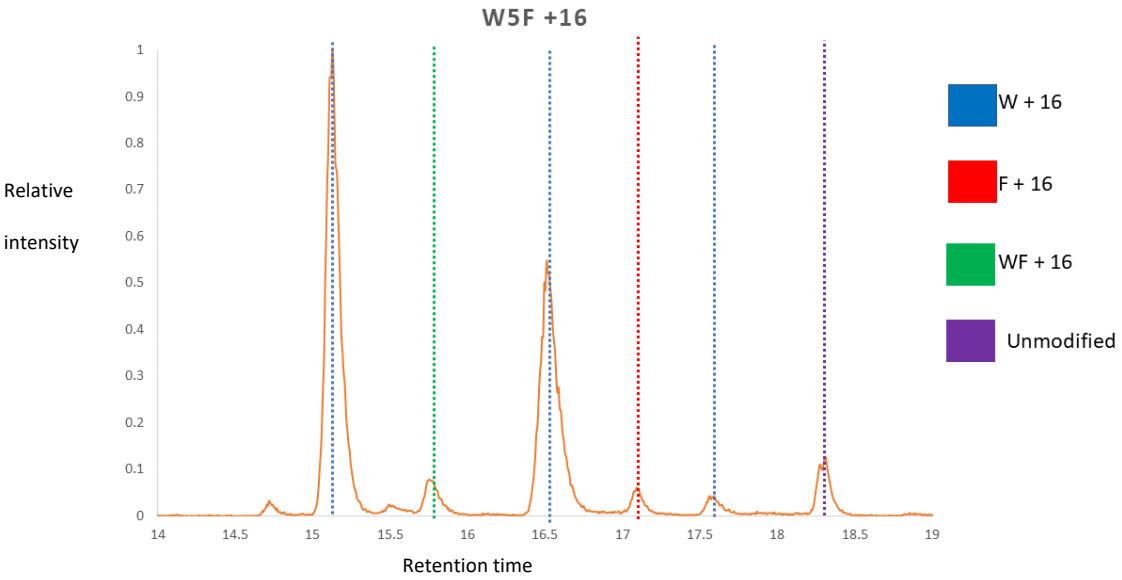


Figure 3.23: Chromatogram for W5F. Six discrete peaks are shown, three for a modification on tryptophan, one for a modifcation on phenylalaine and a coeluting peak. All peaks are chromatographically resolved. However the peak containing WF coelution shows a lack of resolution in the chromatographic method

For the F0F +16 two individual F1 peaks can be observed, with two for F2 and one peak coeluting both W and F (Figure 3.24).

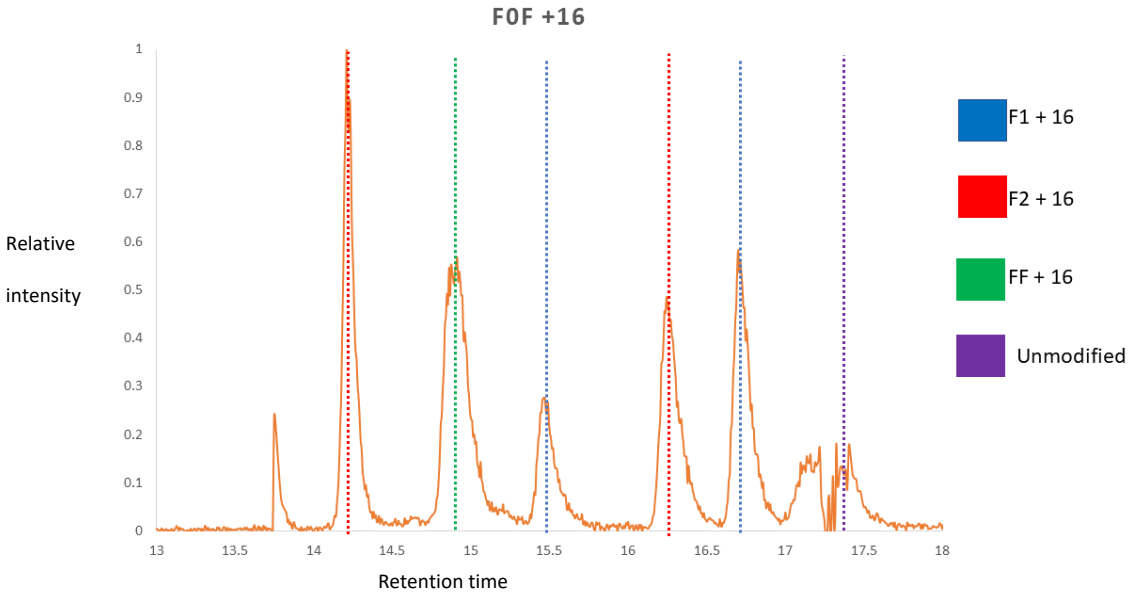


Figure 3.24: Chromatogram for F0F. Six discrete peaks are shown, two for a modification on the first phenylalanine, F1, two for a modifcation on the second phenylalaine, F2, and a coeluting peak. All peaks are chromatographically resolved. However the peak containing WF coelution shows a lack of resolution in the chromatographic method. There is significant peak tailing of the unmodifed peak at 17.5

For the F1F +16 three individual F1 peaks can be observe and three for F2 (Figure 3.25).

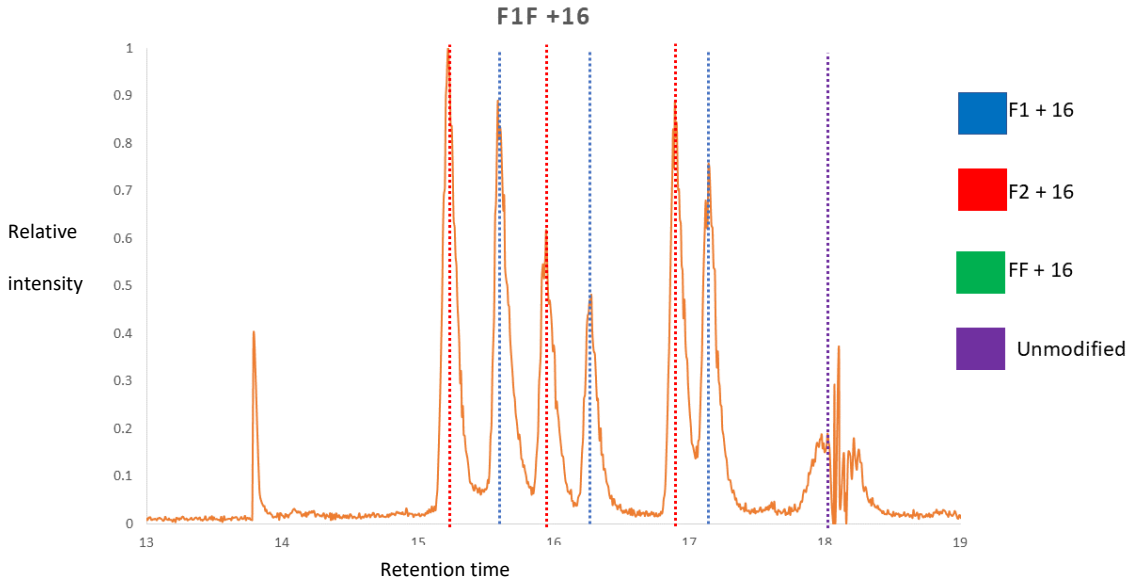


Figure 3.25: Chromatogram for F1F. Seven discrete peaks are shown, three for a modification on the first phenylalanine, F1, three for a modifcation on the second phenylalaine, F2, and a coeluting peak. The two peaks at RT 17.0 are not chromatographically resolved.

For the F1F +16 three individual F1 peaks can be observed and three for F2 (figure 3.26).

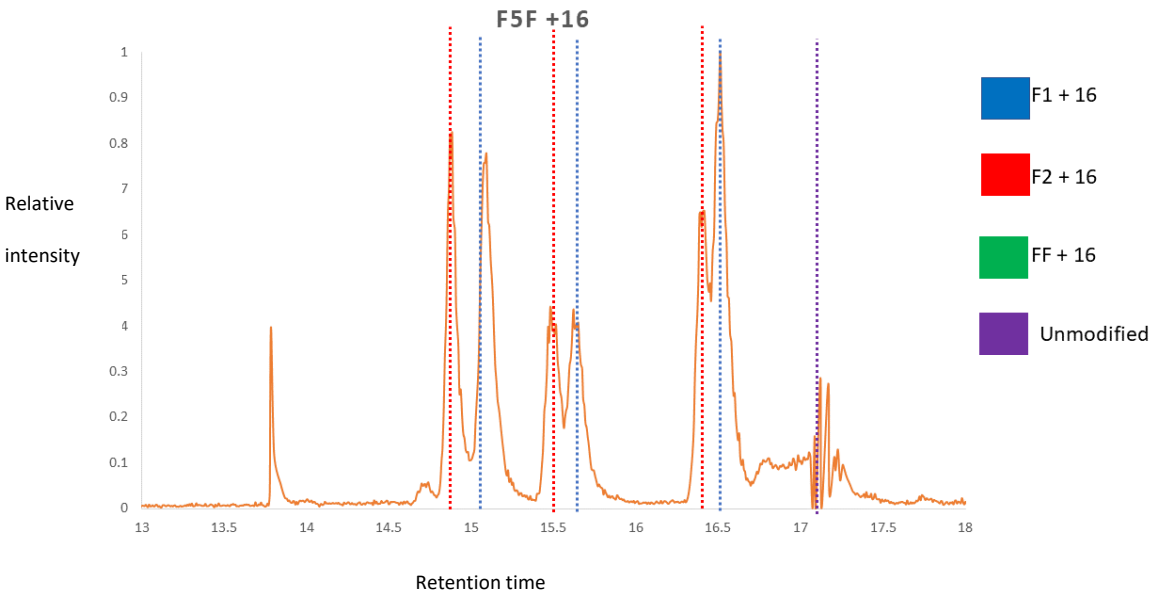


Figure 3.26: Chromatogram for F5F. Seven discrete peaks are shown, three for a modification on the first phenylalanine, F1, three for a modifcation on the second phenylalaine, F2. The peaks at RT 15.5 and 16.5 not chromatographically resolved.

3.4 Discussion

3.4.1 Peptide level calculations

WOF, W1F and W5F showed an increase of oxidation at the peptide level as the residues get further apart, with W0F having 13.8% less total oxidation in comparison to W5F, exceeding standard deviation for both samples. Whilst we do see an increase across all three samples, these are within the standard deviation from W0F to W1F and W1F to W5F.

In terms of distribution of oxidation sites, there is no significant difference between oxidation constitution for all three samples, aside from the +30 modification for W0F and W5F for which we see a significantly different change of 3.93%.

This general preservation of oxidation site distribution suggests the increase of oxidation across W0F to W5F is not due to an increase in a single oxidation state, and is instead due to a general increase in total oxidation across all oxidation states.

FxF similarly saw an increase in oxidation between F0F and F5F, with both F1F and F5F having a significantly higher oxidation than F0F.

For the +16 modifications, F1F and F5F are significantly differently constituted, with F1F having 10.38% and F5F 8.99% less constituent +16 modification as compared to F0F.

The +32 modification howed F0F having 5.5% less constituent oxidation than F5F, a significant difference.

Further, it is noted that the +48 modification is not included in the analysis for FxF residues due to the lack of conversion of phenylalanine to a +32 residue, meaning the abundance of the +48 peak is significantly less likely (Chea and Jones, 2018).

When a one-way ANOVA is performed on the data (Table 8.7 and 8.8 in the supplemental information), an increase in oxidation is significant from W0F to W5F, and from F0F to F1F and F5F. This provides statistical evidence that increasing the distance between reactive residues in primary sequence causes an increase in total peptide oxidation. Further, this presents a possibility that for peptides containing two phenyalanines, there is a shift in the oxidation states observed as F0F is statistically less oxidised than F1F, whereas this is not the case for the WxF samples. It is plausible this is due to Pi stacking of the benzine rings of the F0F phenylalanine residues, further decreasing distance between the side chains and thus increasing competition for radicals.

It is well reported in literature that radical dose is the primary driver of oxidative labelling within FPOP experiments (Niu, Ben et al., 2015), and further the lifetime of radicals has been reported to be within μs to ms timescale (Hambly and Gross, 2005; Vahidi and Konermann, 2016). This timescale is

contested as the scavenger used, in this case histidine, forms an intermediate radical before being quenched, such that:

$$\bullet OH + His \rightarrow H_2O + \bullet His$$

it has been proposed these metastable secondary radicals increase the oxidative labelling window by ~10,000x when compared to hydroxyl radicals alone (Vahidi and Konermann, 2016).

Previous work from synchotron radiation (Xu, G. and Chance, M. R., 2007) shows the rate constants with OH radicals for W and F as $1.3x10^{10}$ M^{-1} s^{-1} and $6.9x10^{9}$ M^{-1} s^{-1} respectively. Therefore, as observed with the difference in total global oxidation between WxF and FxF, the lower reactivity of phenylalanine in comparison to tryptophan accounts for the lower average oxidation as shown in figures 3.10 and 3.11.

The 2+ charge state was used in these calculations due to the significant lack of 1+ and 3+ ions (<5% of total ions.

As discussed previously, it has been described in a number of previous publications (Xie, B. et al., 2017; Cornwell, Owen et al., 2021) that residues in close proximity may decrease the oxidation of one or both of the residues. Further, sequence composition has a distinct effect on oxidation of residues (Sharp and Tomer, 2006).

Sharp and colleagues (Xie, B. et al., 2017) placed increasing numbers of glycine residues between leucine and aspartic acid residues, and observed a number of changes to each oxidation state as the distance between these residues increased.

However, there are a number of issues with this approach which could have been improved in this experiment. Firstly, by using direct infusion, it is not possible to separate isomers. Further, by not using an LC approach, it is not possible to summarise the total oxidation of each modification as a total. Thus, it is not possible to determine how much increasing the distance between residues affects the total oxidation of the peptide.

3.4.2 Individual residues

For each +16 modification, the most dominant modification for each peptide, the XICs were assigned based on the location of the modification. For each peak, where an MS/MS was taken, as described in results, a peak was assigned to either a tryptophan or phenylalanine residue for WF, or a phenylalanine residue for FF.

Due to the higher TICs for +16s for WF samples, we see smoother curves than for the FF samples.

Whilst every attempt has been made to separate every individual peak using this method, coelution of peaks is observed in the +16 charge state of every sample, however it is particularly pertinent in the WF samples. This is likely due to the very similar polarity of the isomers. It may have been possible to separate these out using higher resolution chromatographic techniques including different chromatographic columns or longer chromatographic runs, however, it was not possible to resolve peptides with different oxidation sites within our experiments.

The retention time of oxidised amino acid isomers has been identified in the literature, both for residues within peptides and for single residues (Hensley et al., 1999; Du et al., 2004). This identification can be achieved through using UV detection to determine which isomer is eluted within a given fraction. herein, whilst we can speculate which isomer eluted In which order, it is not possible to be certain due to both colelution of peptides in the chromatograms presented and influence of the other amino acid residues in retention time.

The retention time for free phenylalanine for reversed phase LC, is para < meta < ortho (Du et al., 2004). Therefore, it is hypothesized this is the expected order of detection observed in this experiment.

Thus, given the above data, characterisation of the representation of each residue at the +16 oxidation is not possible, and would be interesting to explore at a later stage.

3.5 Conclusion

This work, at the peptide level, provides significant evidence that the location of reactive residues in close proximity to other reactive residues reduces oxidation. Indeed, this appears to be the first definitive evidence that having two reactive residues in close proximity reduces total global oxidation. Total global oxidation is defined as the total oxidation of the molecule being interrogated, in this case the peptide.

Given the short lifetime of the radical, within the ms to μs timescale, it is proposed that where there is a competition of radicals below a dosage threshold, competition of more than one reactive residue is detrimental to the total labelling across the peptide before cessation of the labelling event. This is important as there may be an erroneously low result for the oxidation of a peptide in a comparison experiment should these residues be substituted. This work has not probed the effect of dosage on oxidation of reactive residues. If the radical dosage is high enough, it is plausible that there would be no impact on the labelling efficiency and thus mitigate the local effects.

It is therefore proposed, that for oxidation of the peptide studied, residue reactivity, residue distance, radical dose and radical lifetime all play significant parts within the equation for global oxidation.

As the ANOVA statistical tests show (in the supplemental Information), the increase in oxidation is statistically significant for W0F to W5F, and from F0F to F1F and F5F. Therefore, the evidence is strong that increasing reactive residue distance in primary amino acid sequence in these peptides increases oxidation at the peptide level.

This work does not determine the limit that increasing distance in amino acid sequence between reactive residues increases the labelling effect. It is only established that increasing the distance between WF or FF residues does increase total oxidation. Thus, should more experiments be performed, increasing the distance above five residues sequentially would allow for determination

of the increase in total oxidation at the peptide level until this increase is no longer observed at *n* number of residues. Should this work be performed, these initial data should form a foundation for determination of the reach of this reductive effect in amino acid sequence.

It is further noted, that the peptides interrogated are not defined in their secondary structure, and are thus free to form secondary structural elements. Thus, whilst reactive residues may be further away in primary amino acid sequence, the secondary structures formed may cause reactive residues to be closer in 3D space than would be anticipated by the amino acid sequence. Thus, by using peptides in a helix with a defined secondary structure, it would be possible to predict the distance between the reactive residues in solution.

However, given advances in prediction of secondary structure through in silico methods. Prediction of the secondary structure of the peptides used could be possible, and this would help determine the location of the reactive residues used in these experiments.

At the residue level, the resolution of the chromatography was not high enough to separate each individual isomer. As mentioned previously, the retention time of phenylalanine isomers is well documented in literature. However, given the complexity of the peptides, it is not apparent that this order would be maintained for the peptides interrogated. By increasing chromatographic resolution, the use of UV-VIS spectroscopy in-line, before introduction to the MS, would allow for determination of which isomer is present. Should this be the case, ratios of each isomer present could be determined.

Coupled with modelling data, the ratio of each isomer would allow us to further validate the theory that the local neighbourhood is important for the reduction in labelling, as if labelling is increased for isomers that are of increased distance spatially from competing residues, it validates the hypothesis that residues closer in space face more competition for radicals.

Finally, only one radical dosage was probed, it is therefore plausible that if radical dosage is high enough, neighbourhood effects would be limited, as all residues would receive ample dosage. However, dosage limitation is essential in FPOP experiments, as comparisons between solvent accessibility are dependent on the total oxidation.

Chapter 4:

HDX of polysorbate formulations

4. HDX of polysorbate formulations

4.1 introduction

Polysorbates are a group of nonionic surfactants incorporated into drug formulations, specifically protein drug formulations, to protect against surface induced aggregation and to stabilise proteins (Singh, Surinder M. et al., 2017; Schwartzberg and Navari, 2018). A number of mechanisms have been reported to yield these effects.

Firstly, polysorbates compete with proteins for access to surfaces by binding with hydrophobic pockets on the protein and therefore decreasing surface exposure and thus intermolecular interactions (Randolph and Jones, 2002; Chou et al., 2005; Singh, Surinder M. et al., 2017). Secondly, through an increase in the activation energy required to induce unfolding (Bam et al., 1996) and finally through binding with unfolded proteins to inhibit self-association (Singh, Surinder M. et al., 2017).

Polysorbates 20 and 80 are commonly used in antibody drug formulations, in part due to their low critical micelle concentrations (CMC), 0.006% w/v for polysorbate 20 and 0.001% w/v for polysorbate 80 respectively (Patist et al., 2000). Thus, in formulations polysorbates are used significantly above their CMCs (Kerwin, 2008).

When above the critical micelle concentration, surfactants self-assemble into spherical aggregates called micelles, with the hydrophobic tails assembling a core, and the hydrophilic head group forming the edge (Perinelli et al., 2020). It is therefore plausible that this formation of micelles is critical to the protection offered by polysorbates.

The knowledge of the effect of polysorbates on protein structure, specifically that of MAbs, is poorly understood. Singh and colleagues (Singh, S. M. et al., 2017) attempted to determine the effects of polysorbate on higher order structure using 2D-NMR with a mAb, Fab and Fc fragments and both

polysorbates 20 and 80. This work identified binding of the polysorbates, and significant changes in tertiary structure due to this.

HDX provides an option to study the conformation changes of antibodies in solution. Thus, it is possible to determine how the protein is affected by polysorbates in native conditions, and to monitor any changes due to polysorbate interaction.

HDX has been used previously to analyse polysorbates with antibodies, and to assess protein changes. Kerr and colleagues (Kerr et al., 2019) assessed 4 MAbs in their production formulation by use of HDX-MS. Of these, two had polysorbate 20 in their formulations, one polysorbate 80 and a control antibody with no polysorbate. This study highlights the ability to utilise a HDX-MS workflow for characterisation of Mab structure in the presence of polysorbate.

Given the use of polysorbates as excipients in antibody drugs, and the previous studies listed, highlighting how polysorbates interact with proteins, studying exactly how polysorbates interact with antibodies to elucidate their protective effects is of interest for both drug design and drug formulation questions. Understanding the fundamental structural changes to a protein upon addition of polysorbates can also allow for excipient selection dependent on the amino acid sequence and higher order structure of the protein of interest.

For this work, two antibodies, MEDI1912_STT, termed STT and MEDI1912_WFL, termed WFL were used (Willis et al., 2018). The names WFL and STT refer to residues 30,31 and 56 of the heavy chain of the antibody. With WFL having W30, F31 and L56, and STT having these residues substituted such that these are S30, T31 and T56 respectively. These changes are located in CDR1 and CDR2, in the V_H domain (Dobson, J. et al., 2017).

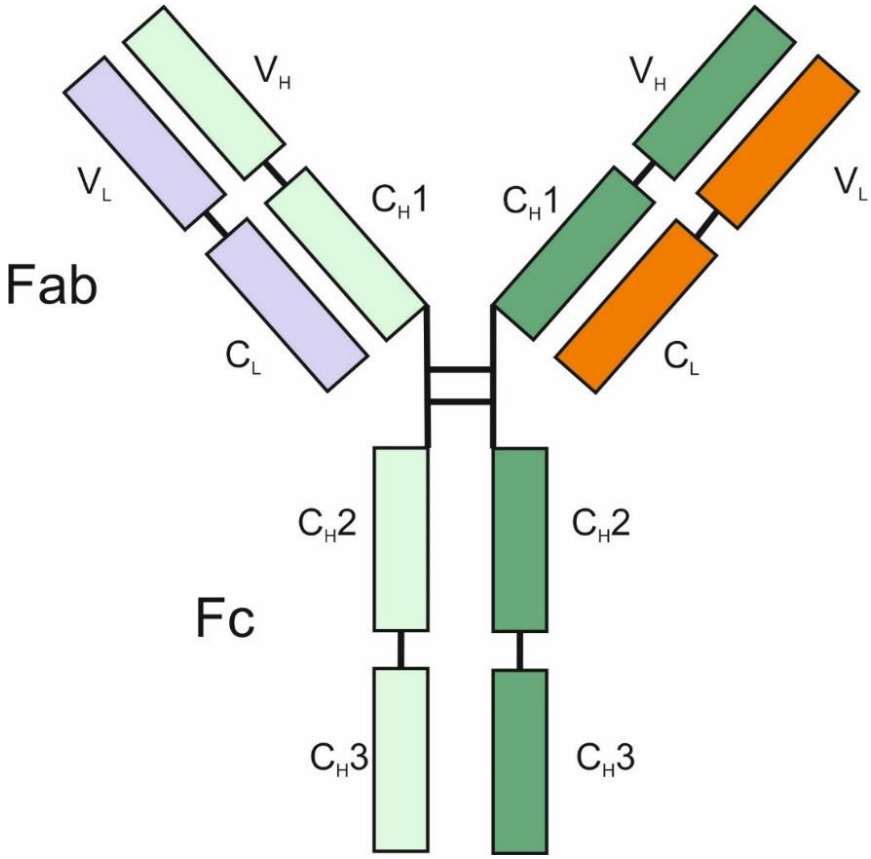


Figure 4.1: Cartoon of antibody, with each domain and heavy and light chains highlighted. Fab stands for fragment antigen-binding region, Fc stands for fragment crystallizable region. V indicates variable region and C constant region, with H and L referring to heavy or light chain.

Both proteins have a large hydrophobic region observed on the surface in the region where the WFL and STT substitutions are located, however, WFL has a larger and more hydrophobic patch on the surface (Dobson, J. et al., 2017).

Here, we utilise HDX to interrogate the effect of polysorbates on the structure of both WFL and STT, and interrogate whether changes in the protective effect of polysorbates can be observed, in the event this protective effect yields a change in structure.

4.2 Methods

The methods were described in 2.2

4.3 Results

Initially, myoglobin was mixed with polysorbate 20 and 80, 30 minutes before HDX analysis. No significant difference in protection or deprotection was observed above the significance cutoff of 0.5 Da, as shown in the 10 minute time point (Figure 4.2). Sequence coverage was 100% for both additions (Butterfly plots shown in the supplementary information figures 8.1 and 8.2). Additionally, PS20 was compared to PS80, and no significant difference was observed (figure 8.3 in the supplemental information.

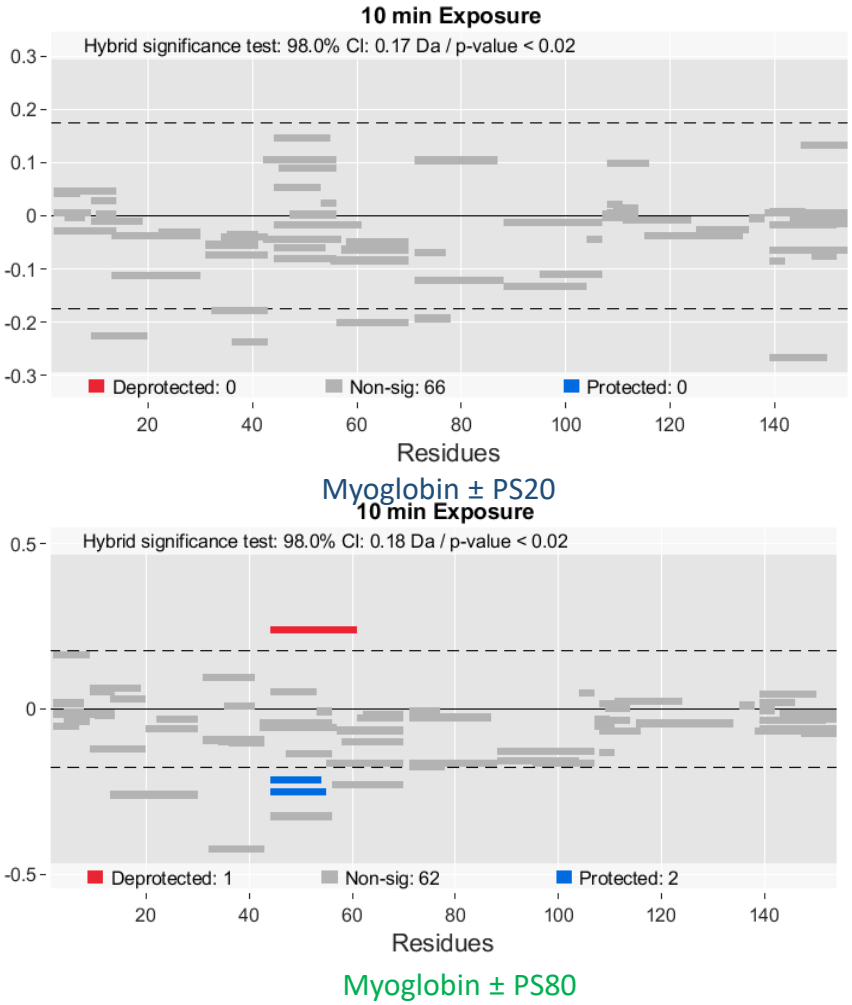


Figure 4.2: 10 min time point of myoglobin with and without polysorbate 20 and 80 added 30 minutes before analysis. The arbitrary cutoff for significance is 0.5 Da. Therefore there is no significant difference shown. (The butterfly plot for all time points is shown in supplementary information figure 8.1 and 8.2)

Following the test of addition of polysorbate 30 minutes before analysis, myoglobin was incubated with both polysorbate 20 and 80 for 1 month at 25 °C. These samples were then compared to the samples where fresh polysorbate and myoglobin were analysed as discussed in the paragraph above, to clarify, the data from the one month incubation were compared to the data from the fresh addition. Myoglobin with polysorbate 20 had 95% coverage (figure 4.3)**.** Myoglobin with polysorbate 80 saw 53% coverage, a decrease of 42%.

For myoglobin with polysorbate 20 (figure 4.3) (butterfly plots shown in the supplementary information (figure 8.4)), Two peptides were identified as significantly deprotected, and one peptide as protected. Residues 19-28 and 33-41 are deprotected, and residues 40-54 protected. These have been shown on a structure in figure 4.4, 1MBN from Protein Data Bank. These regions of protection and deprotection are located close to the N-terminus, in the first and second helices, with the protection concentrated in the second helix from the n-terminus.

There was no significant difference observed for the polysorbate 80 sample, although it is noted that as discussed previously the coverage was lower, at 53%.

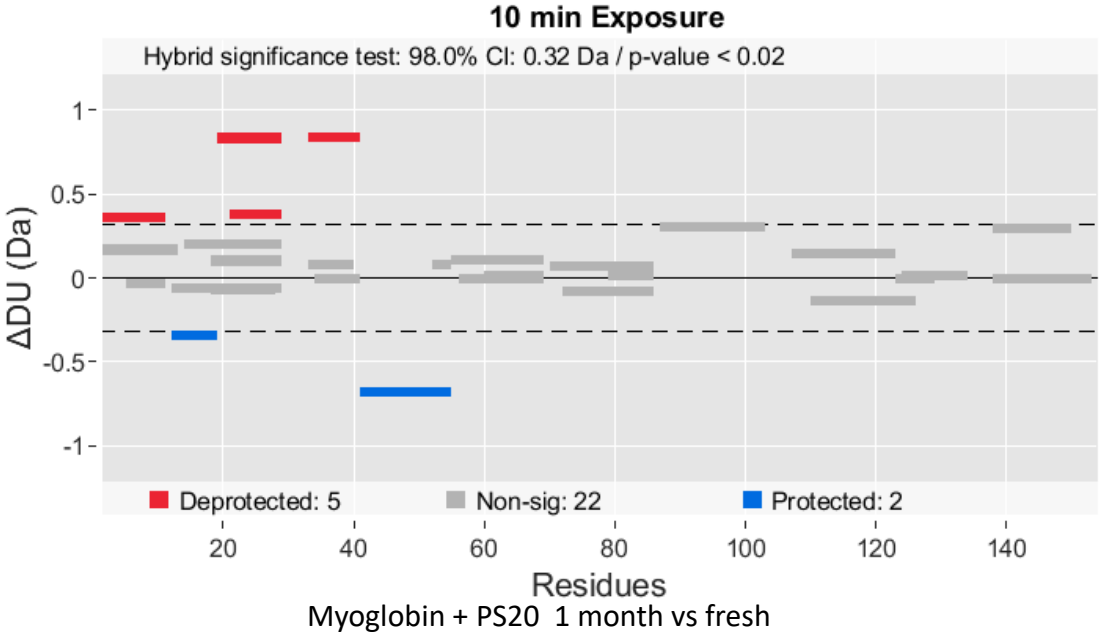


Figure 4.3: Myoglobin with polysorbate 20 1 month incubation. The arbitrary cutoff for significance is 0.5da. Three peptides fulfil this criteria, residues 19-28 and 33-41 are deprotected, and residues 40-

54 protected. The butterfly plots and peptide uptake graphs are shown in the supplementary information figures 8.4, 8.5 and 8.6.

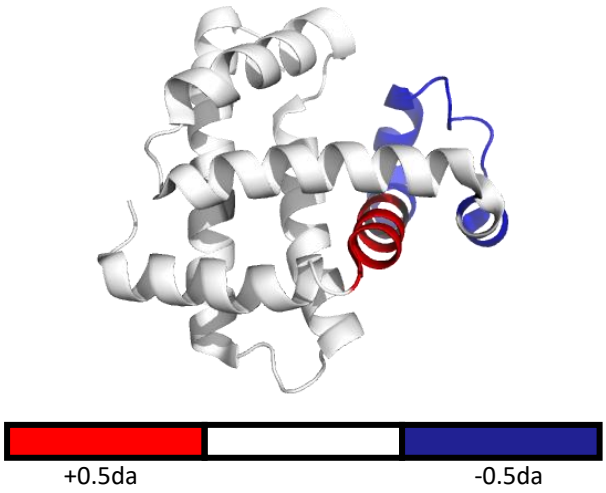


Figure 4.4: structure of myoglobin 1 month incubation vs fresh addition with polysorbate 20. Deprotection is shown in red, and protection in blue. Deprotection and protection is observed at the C-terminus of the molecule, occurring in the penultimate and final helices within the protein structure. The butterfly plots and uptake graphs are in the supplementary information figures 8.4. 8.5 and 8.6

To assess the protective effect of polysorbates on myoglobin, myoglobin was incubated with and without polsorbates at 25°C for 1 month. Each sample was compared, with no polysorbate, polysorbate 20 and polysorbate 80 (Figure 4.5). Sequence coverage for all three samples was 73%, however only 16 peptides were identified. No significant different was observed with or without polysorbate, and between polysorbate 20 and 80.

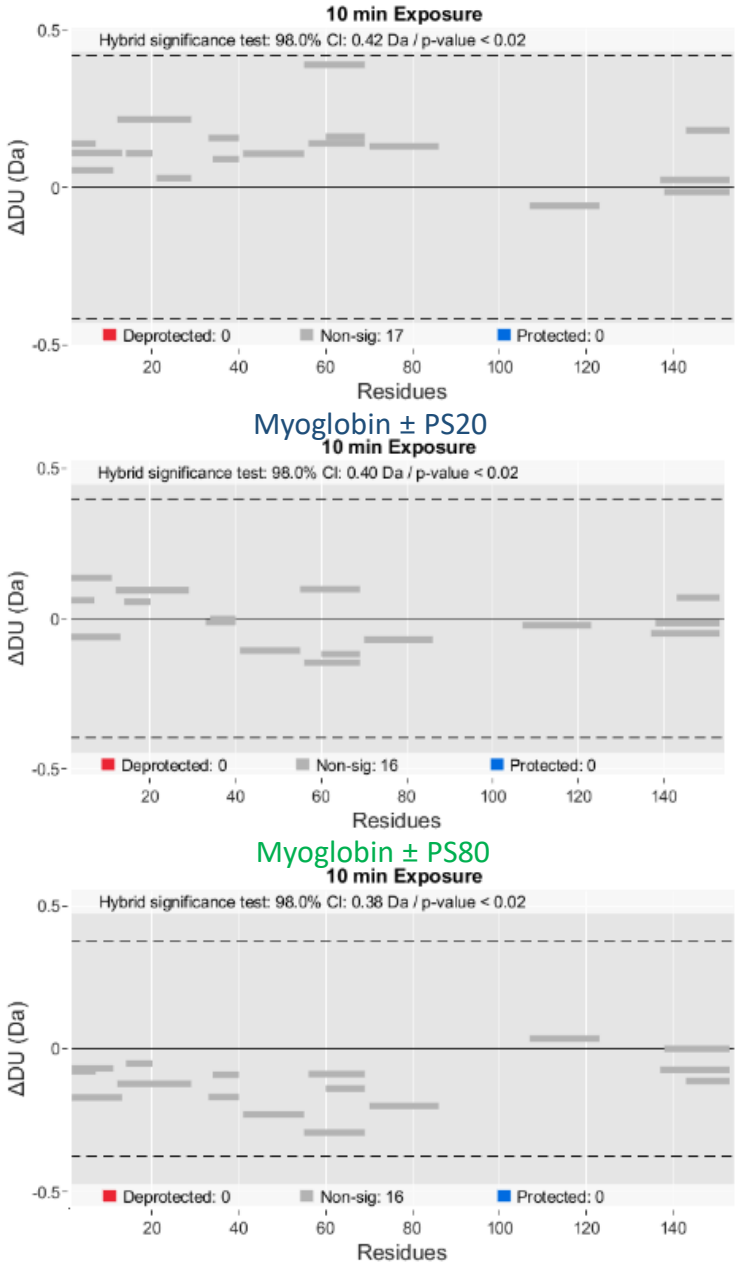
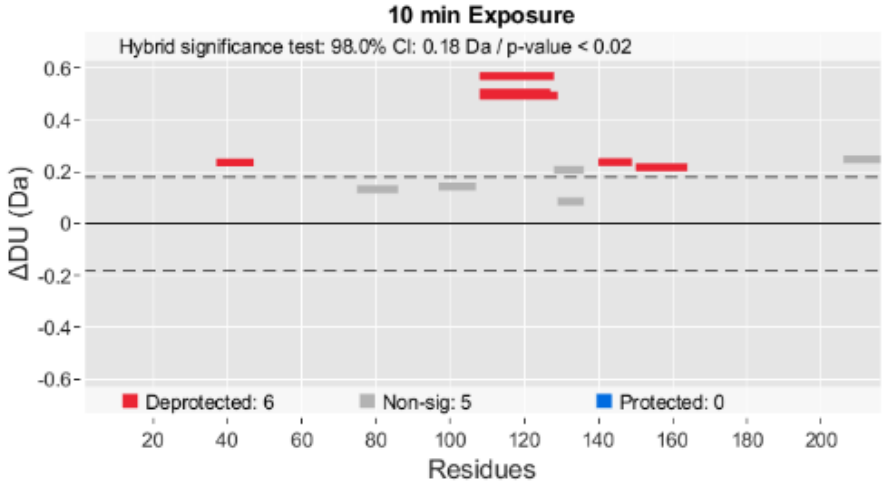


Figure 4.5: protective effect of polysorbate 20 and 80 vs no polysorbate over 1 month for myoglobin. No significant differences were observed between all three samples. 16 peptides were identified. The butterfly plots are shown in the supplemental information figures 8.7-8.9.

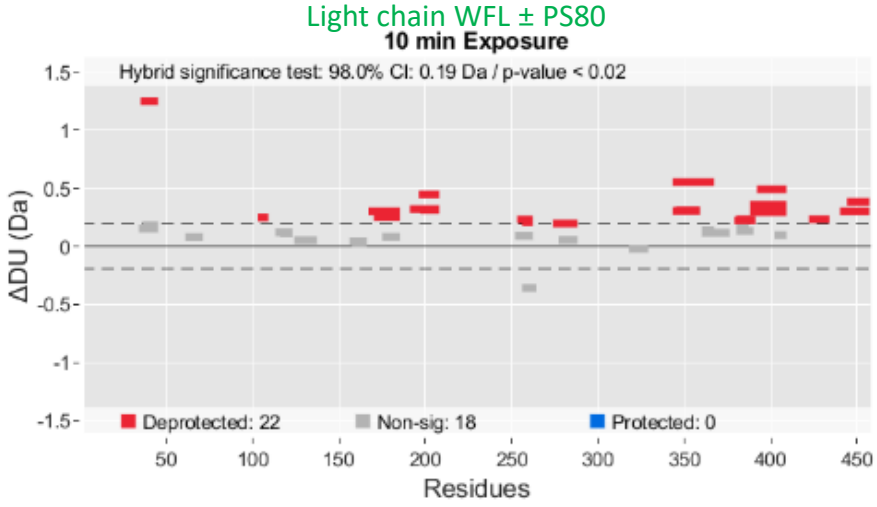
Following these experiments performed on myoglobin, WFL and STT were analysed, with polysorbate added 30 minutes before analysis.

For STT, sequence coverage was 65%. No significant difference was observed between the control and the addition of polysorbate 20 and 80.

For WFL, coverage was 50%. For the control and polysorbate 20, no significant difference was observed.. However, for the control and polysorbate 80 a number of significantly different regions were identified. Three significantly deprotected peptides were identified on the light chain between residues 108-129. On the heavy chain, two significantly deprotected peptides were observed between residues 348-356 and 396-405 (Figure 4.6). The butterfly plots are shown in the supplemental information (Figures 8.10-8.23), along with the uptake plots (Figures 8.22-8.24).



Light chain WFL ± PS80



Heavy chain of WFL ± PS80

Figure 4.6: WFL control vs polysorbate 80. The cutoff for significance is 0.5 Da. Deprotection was shown between residues 108-129 on the light chain and 348-356 and 396-405 on the heavy chain. The butterfly plots are in the supplemental information figures 8.18-19, and the uptake plots figures 8.22-24.

When comparing the additions of polysorbate 20 and polysorbate 80 (Figure 4.7), a significant difference in deprotection is observed in the light chain between peptides 108-125.

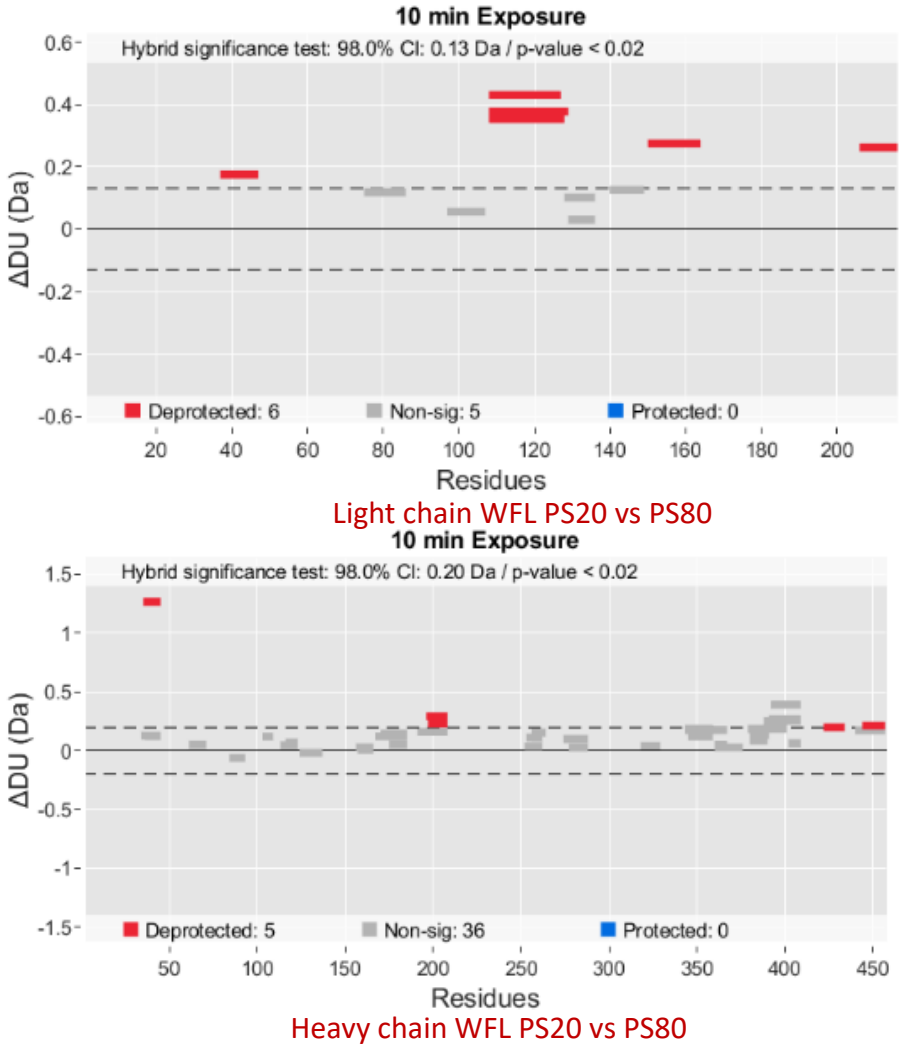


Figure 4.7: WFL ps20 vs ps80 with polysorbate added 30m before analysis. The arbitrary cutoff for significance was 0.5 Da. Deprotection was shown in the light chain between peptide 108-125. The butterply plots are shown in the supplemental information figures 8.20-21, and the uptake plots figures 8.24-26.

These changes were plotted on a generic antibody structure, to allow for interpretation (Figures 4.9 and 4.10).

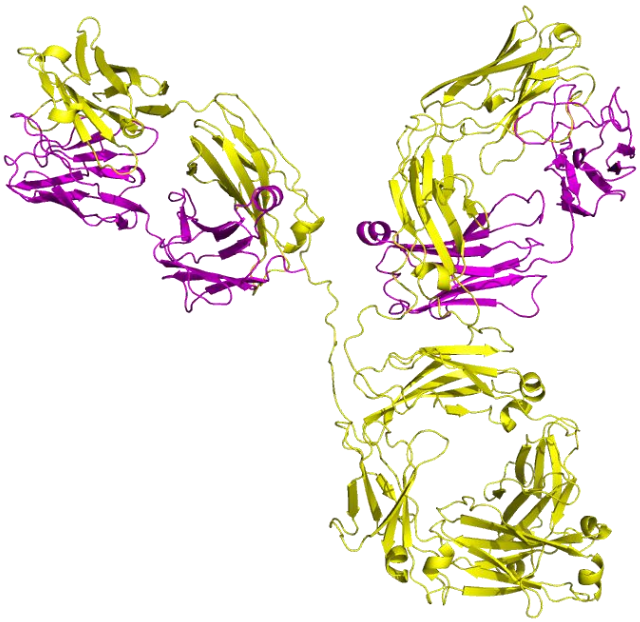


Figure 4.8: Identification of light chain and heavy chain of the antibody. The heavy chain is labelled in yellow, with the light chain labelled in magenta. The CDR region substitutions from WFL to STT, as shown in figure 4.9 and 4.10 are located on the heavy chain.

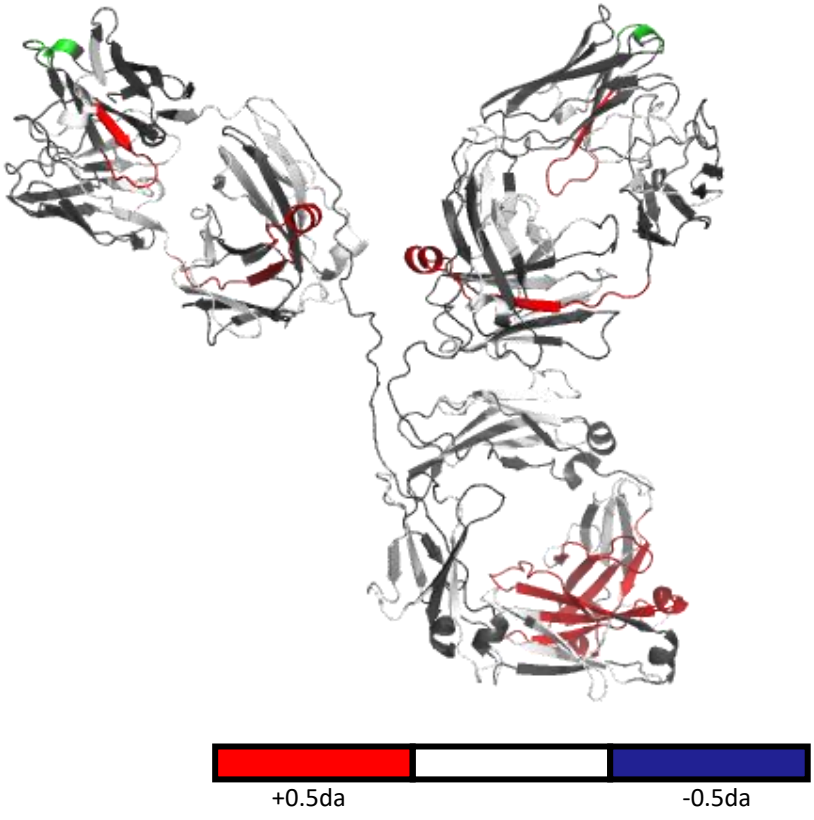


Figure 4.9: WFL control vs polysorbate 80 labelled structure showing deprotection in red, structure from (Spiteri et al., 2021). Three significantly deprotected peptides were identified on the light chain between residues 108-129. On the heavy chain, two significantly deprotected peptides were observed between residues 348-356 and 396-405. The substitutions from STT to WFL are shown in green, with the hydrophobic pockets forming in the CDR regions located around the substitution. Peptides with no coverage are shown in black, and and non-significantly different peptides with sequence coverage are shown in white.

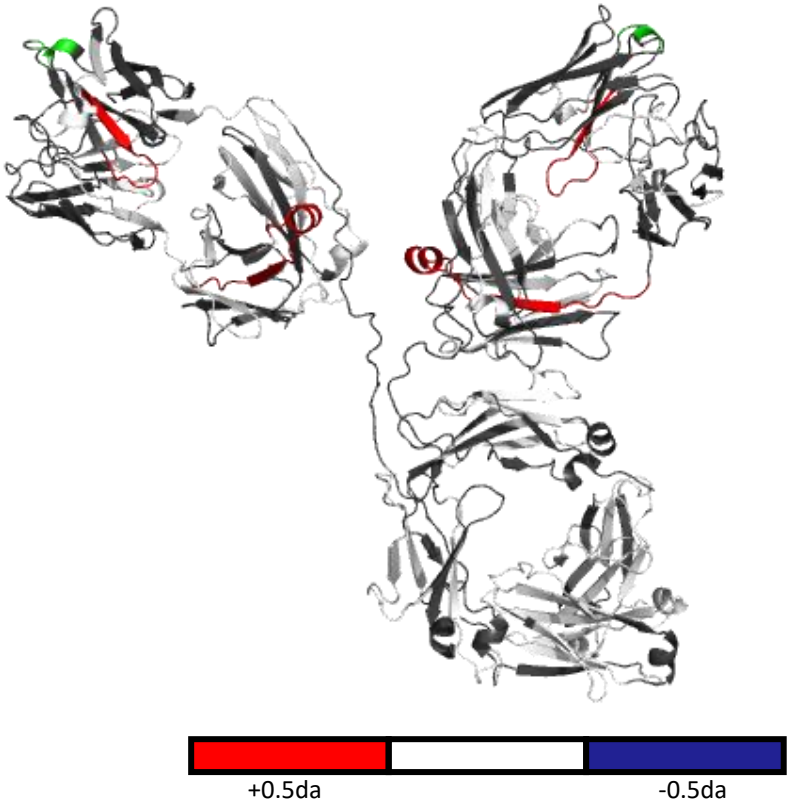


Figure 4.10: WFL polysorbate 20 vs polysorbate 80 structure showing deprotection in red, structure from (Spiteri et al., 2021). Significant deprotection observed between the control and polysorbate 20 in the light chain is observed between peptides 108-125. The substitutions from STT to WFL are shown in green, with the hydrophobic pockets forming in the CDR regions located around the substitution. Peptides with no coverage are shown in black, and non-significantly different peptides with sequence coverage are shown in white.

4.4 Discussion

Initially, the method was probed by interrogating myoglobin with polysorbates. This was done to ensure the viability of the method for the analysis of proteins with polysorbates. Polysorbate was added 30 minutes before the experiment to allow for equilibrium of the dispersion of the polysorbate around the protein before analysis was undertaken, to ensure that the system was comparable with the aged samples.

For each experiment, time points of 2 minutes, 5 minutes and 10 minutes were used. In the results, only the time points for 10 m are shown, however butterfly plots for each time point are shown in the supplemental data. Butterfly plots show a stacked representation of each time point, which display the changes in deuteration for each sample at each time point. The time point selected for analysis was 10 minutes, as this provided the longest time point for the deuterium to exchange with the hydrogen atoms. Thus, given that the incubation between the polysorbate took place over 1 month or 30 minutes respectively, the longer time point allows for full uptake of deuterium for partially open states, and gives a more representative view of the 30 minute and 1 month incubation. However, as stated, butterfly plots for each timepoint are shown in the supplemental data, and these have general conformity with the data presented here.

For myoglobin, coverage was 100% with the addition of both polysorbate 20 and 80. Pepsin has been readily used for online digestion of myoglobin previously (Zhang, H.-M. et al., 2010), and online digestion of polysorbate containing mixtures has been successful (Kerr et al., 2019). Given this, 100% coverage (figure 4.2) proves the viability of this method for a model protein.

Following this, as discussed, myoglobin was incubated for one month, and some degradation was observed in the first and second helices for polysorbate 20 (figure 4.4). Interestingly, coverage was significantly lower for the myoglobin sample with polysorbate 80 than polysorbate 20 (42% lower). It is plausible that the sample with polysorbate 80 has degraded significantly during incubation and thus the total protein remaining in the vial means lower abundance peptides from digestion are

under the analytical threshold when analysed. By using a pelleting assay or similar technique, it would have been possible to test this hypothesis.

Alternatively, as the polysorbate 80 degrades, the protein may become oxidised through the degraded species (Doyle Drbohlav et al., 2019; Larson et al., 2020). Thus, these peptides will not be found through the analytical workflow. (Wang et al., 2008). This is due to the generalised workflow not including modification detection. It would be beneficial in the future to include oxidised species, as this would provide a further measure of degradation. This would require a change to the general HDX workflow for this application, as oxidation PTMs are not currently included in this HDX analysis pipeline.

No control was ran with no polysorbate to compare to polysorbate 20 and 80 results for the one month incubation of myoglobin against the fresh incubation. This would have allowed for a baseline to be determined of how protective polysorbate is in comparison to no polysorbate. However, given coverage was significantly lower for the polysorbate 80 samples, it is plausible the control would begin aggregation, lowering the amount of peptides for comparison as shown in figure 8.5 where, in a separate experiment, all three samples were incubated for 1 month and only 16 peptides were available for comparison between all three states.

For polysorbate 20, the 1st helix from the N-terminus shows deprotection with the 2nd helix showing protection.

When all three aged myoglobin samples are compared (figure 4.5), no difference in protection is observed between the control, polysorbate 20 and polysorbate 80. This implies there is no protective effect from polysorbate for the degradation of myoglobin over one month. However, compared to other studies on polysorbate, the conditions for this incubation were less accelerated (Kerr et al., 2019)**,** and given the relative stability of myoglobin, a forced aggregation approach may have yielded more difference in protein protection.

Following the success of these experiments, this work followed on by looking at two different Mabs, STT and WFL.

For both the STT and WFL no sequence coverage for the CDR regions is observed, where the sequence change for WFL to STT takes place. As observed with myoglobin, the control with polysorbate 20 had higher coverage than with the polysorbate 80 (by 17%). It may be that there is not full removal from the CDR regions of the polysorbate, which impacts the digestion of the CDR regions.

For WFL, we see a significant difference between the control and polysorbate 20 both in the light chain and the heavy chain. And this heavy chain deprotection is conserved when polysorbate 20 and 80 are compared. (Singh, Surinder M. et al., 2017; Larson et al., 2020). Previous studies have identified that polysorbates exert their effects primarily on the fab domain and not the fc (Singh, Surinder M. et al., 2017), however, the conserved regions are in the fc domain. However, due to the ~50% coverage it is possible effects on the fab are not reported.

Whilst all polysorbates have affinity to hydrophobic residues, polysorbate 80 has a higher affinity than polysorbate 20 due to the longer fatty acid tail, which also causes the lower CMC of polysorbate 80 than 20 (Singh, Surinder M. et al., 2017).

The change of STT to WFL causes an enlarged hydrophobic patch to be present on the WFL MAb structure in the fab domain (Dobson, C.L. et al., 2016). This hydrophobic pocket is likely to attract higher binding from polysorbate 80 than 20, however binding from both is expected (Singh, Surinder M. et al., 2017). This more significant binding in WFL may account for the deprotection observed both vs the control and vs polysorbate 20. We do not see the conformational changes in the same regions as the hydrophobic regions, suggesting these deprotections are due to longer range conformation changes. Conformational changes in tertiary structure due to polysorbates have been reported (Singh, Surinder M. et al., 2017), and this is due to binding.

It is important to note that there has been numerous examples of polysorbate not only stabilising protein structure but also perturbing changes in protein conformation (Webb et al., 2002; Wang et al., 2008). Indeed, polysorbates have been shown to have both positive and negative effects on protein stability and aggregation based on both the protein and conditions (Wang et al., 2008).

The polysorbate concentrations used were within the range of current market monoclonal antibodies (Singh, Surinder M. et al., 2017).

4.5 Conclusions

The concept for polysorbate analysis using the method described was shown with myoglobin. This method was then shown to be applicable to complex mixtures containing excipients, such as those used in drug formulations.

The aim of this work was to determine the effect of polysorbates on two antibodies of varying aggregation propensity. Of particular interest was the difference in deprotection between polysorbate 20 and 80 for the WFL sample. This supports the idea that hydrophobic regions offer high affinity for polysorbate 80 in particular.

The perturbations in structure shown for WFL and STT when incubated with polysorbate 20, polysorbate 80 or in the absence of polysorbate shows that selection of excipient is important for maintaining higher-order structure. Should a wider experiment be conducted, determining the location of hydrophobic regions may predict whether structural perturbation is more likely from polysorbate 80. However a large scale screen of a library of antibodies with the addition of polysorbate 20 and 80 should allow for more interpretation of how both molecules effect antibodies with a broad spectrum of chemical environments on the surface.

Currently, LC-MS/MS analysis is routinely used in pharmaceutical development, to determine the presence of covalent modifications to antibody structure, such as oxidation and post-translational modifications. This application applies to forced degradation studies which are performed routinely

in pharmaceutical environments. This entails stressing pharmaceutical molecules through agitation, freeze thaw cycles and storage at different temperatures. The stressed molecules are then analysed by use of bottom-up MS workflows, to look for covalent changes in protein structure, such as increased oxidation or deamidation, in comparison to a control or previous timepoint. The limitation of such approaches is that changes in higher order structure that do not involve a covalent modification cannot be determined.

Thus, there may be structural changes, such as protein binding and conformational changes, to pharmaceutical proteins as they are stressed that are not determined by bottom-up analysis.

By use of HDX-MS/MS analysis we have a direct view of the higher-order structure of the protein in solution, and can see structural changes, such as protein-protein interactions, not identified by current commonly used techniques within the pharmaceutical industry.

Therefore, establishment of HDX-MS/MS as a standard workflow in the pharmaceutical industry would allow for determination of conformational changes within a complex mixture such as a drug formulation. Determination of protein-protein interactions, and increases of excipient binding over time could further be probed, allowing these changes to be tracked as the molecule ages. Given these interactions are non-covalent, they are not determined during the conventional workflows.

Finally, by determining which regions of proteins are dynamic, and monitoring changes in dynamics as molecules age, strategies can be formulated to engineer molecules which are more resistant to structural degradation of important regions of the molecule, such as binding sites.

Thus, a HDX-MS/MS workflow, in conjunction with the current techniques mentioned, allows for a holistic approach, which takes into account all structural and covalent changes of an antibody molecule in a complex drug formulation.

Chapter 5:

Determination of TiO_2 binding to MtrC by X-ray Protein Footprinting and Electrostatic Calculations

5 Determination of TiO_2 binding to MtrC by X-ray Protein Footprinting and electrostatic calculations

5.1 Introduction

Labelling was undertaken by Dr Anna Stikane, and Mass Spectrometry performed by Dr Rachel George. All data analysis, including initial peaks data, manual assignments, manual processing at both residue and peptide level was performed by Jake Busuttil-Goodfellow, aside from molecular dynamics simulations which were performed by Dr Marcus J Edwards.

Renewable electricity generation using solar energy as a source is a significant area of research for solving the current worldwide clean energy demand. Within this, using biological photosynthesising systems has been touted as a solution for energy capture and harvesting (Faunce et al., 2013). Photosynthesis in nature requires specific arrangements of light harvesting molecules to create conductance pathways which are able to harvest light for an electrical output (in this case, electrons). There is considerable interest in building such photosystems artificially from the bottom up and connecting them with electrodes and standard electrical components, in order to create technological solutions to meet ever growing energy demands (Mony et al., 2022).

The use of light harvesting inorganic materials has proven to be a possible method to transfer solar energy, by excitation of heme-containing proteins and electron transfer through nanoparticles (Simonneaux and Bondon, 2005). Coated titanium compounds have been rigorously studied as a conduit for this transfer, and this includes the use of MtrCAB which facilitates anaerobic respiration in *Shewanella* (White et al., 2013). The MtrCAB complex is an outer membrane spanning, 20-heme containing, protein complex consisting of three protein subunits, the periplasmic decaheme cytochrome MtrA, a transmembrane 26-strand β-barrel MtrB, and another decaheme cytochrome MtrC which binds non-covalently to MtrAB on the extracellular side of the cell membrane (Edwards, M. J. et al., 2020).

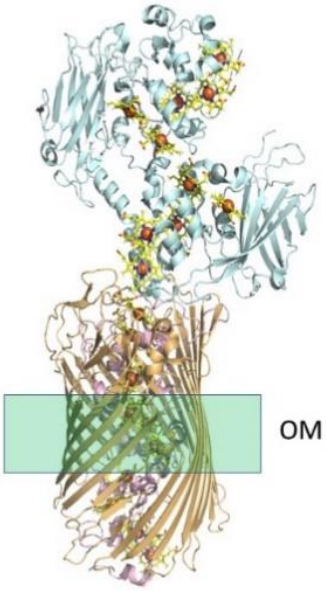


Figure 5.1 MtrCAB complex, adapted from (Jiang et al., 2020). MtrC is colored in blue, with MtrA pink contained in the barrel of MtrB, in gold. An 18-heme chain runs directly through the molecule and is essential for the electron transfer from the outer membrane (OM) subunit MtrC through the barrel to the cell. The outer bacterial membrane is shown in green. Hemes are shown in yellow with heme iron atoms shown as orange spheres.

Electron transfer into the cell via the MtrCAB complex occurs through a transmembrane 18-heme pathway from MtrC at the top of the barrel through to the base of MtrA. The heme groups in MtrA in particular are arranged in a shift-stacked and T shaped motif, allowing for van der Waals forces between hemes (Hwang et al., 2015). It is generally accepted that electrons are transported through contacts between adjacent heme groups ("electron hopping") and not through heme rearrangement ((Breuer et al., 2014; Blumberger, 2015; Jiang et al., 2019)), and therefore location of the hemes is critical to efficiently facilitate electron transfer. While Stikane and others have previously shown that components of the MtrCAB complex, when combined with a titanium based nanoparticle or carbon

dot, can facilitate electron transfer from these excited molecules (Hwang et al., 2015; Stikane et al., 2019), it remains unknown how exactly the protein interacts with the particle surface.

X-ray footprinting mass spectrometry (XF-MS) analysis is proposed here as a plausible technique to determine the protein surfaces of MtrC which bind to TiO_2 nanoparticles. In XF-MS, an x-ray beam lyses (splits) water in solution to produce hydroxyl radicals *in situ* which can covalently modify proteins or other biomolecules in solvent-accessible regions on a millisecond timescale (Gupta et al., 2016). These modifications and their sites are detected using standard proteolytic digestion and liquid chromatography mass spectrometry (LC-MS/MS) sequencing. XF-MS can in principle achieve single amino acid resolution and all 20 naturally occurring amino acids can form side chain oxidation products. The inherent reactivity towards ·OH radicals varies considerably over three orders of magnitude, with large hydrophobic residues (e.g. W, Y, F, H) or sulphur-containing side chains (C, M) showing the highest intrinsic rates of oxidation when exposed. α-carbons are rarely attacked resulting in fragmentation of the polypeptide backbone, as these positions are usually sterically shielded by the amino acid side chains (Xu, G. and Chance, M. R., 2007). By far the most common modifications are straightforward oxidations (+O, +15.99 Da), but carbonyl formation (+O-2H, +13.98 Da) in combination with decarboxylation (-CO-2H, -29.99 Da) and others are also found, depending on the amino acid, and multiple events also occur (Xu and Chance, 2005).

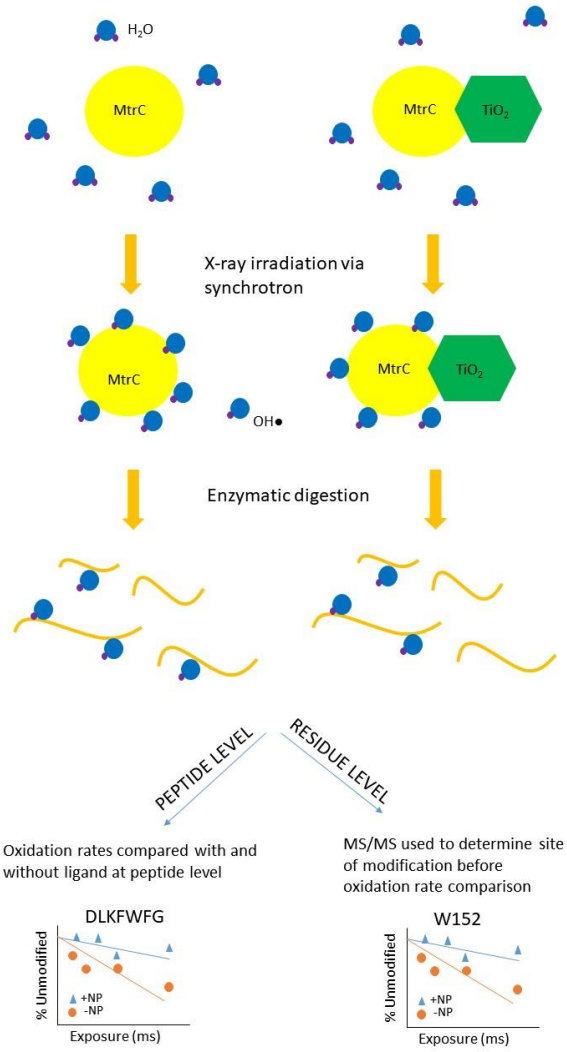


Figure 5.2: Graphical representation of the x-ray footprinting workflow. The protein of interest (MtrC) is exposed to x-ray synchrotron radiation, with and without TiO_2 nanoparticle binding, over increasing amounts of time. The differential modification pattern between the two states is analysed using proteolytic digestion followed by LC-MS/MS analysis of oxidation sites.

XF-MS and other covalent labelling approaches are often used differentially, in comparative studies between two scenarios such as with and without a ligand (Gupta et al., 2016). Protein labelling can be done *in situ*, for example in living cells (Adilakshmi et al., 2009), and also with and without binding to macroscopic objects. Indeed, this approach was used for identifying binding sites of a MtrC homolog, MtrF, with Fe_2O_3 nanoparticles (Fukushima et al., 2017). Due to the covalent nature of the label, downstream enzymatic digestion of the protein can be performed without the requirement for

rapid digestion and low temperatures, such as required for hydrogen-deuterium exchange (HDX)-MS. Further, different types of orthogonal digestion can be performed and pooled to achieve improved peptide composition and measurements.

Following digestion, the mixture of peptides is analysed in a standard reversed phase liquid chromatography and tandem MS (LC-MS/MS) proteomics workflow. The percentage of modification due to labelling can be probed at the level of the whole peptide or individual residue. In both instances, extracted ion chromatograms (XICs) are produced at the expected mass-to-charge (m/z) values of both the unmodified and modified peptides; with different types of modification including multiple oxidation events being considered. At the intact peptide level it is usually found that the retention time changes slightly upon oxidation, so that individual types and sites of oxidation are more or less well separated prior to MS analysis. The areas of these chromatographic peaks represent the relative amounts of individual and, taken together, total modification; whereas at the individual level each peak in the XIC is scrutinised through analysis of the MS/MS spectrum to determine the location of the modification on the peptide. During a comparative experiment, both the peptide level and residue level information provide valuable insight into areas of protection due to nanoparticle binding. Residue level information has the benefit of taking into account the rate constant for each individual residue, whereas at the peptide level, residues with a high modification rate will dominate overall changes in peptide modification state, and changes in less reactive residues will be overshadowed by minor changes in significantly more reactive residues. While all individual amino acid oxidations contribute to the overall peptide oxidation, it can happen that the individual oxidation sites cannot be determined due to poor MS/MS fragmentation.

Herein I use synchrotron x-ray pulses of different duration to generate dose response curves for the oxidative labelling of MtrC protein with and without TiO_2 nanoparticles. I developed a new approach by analysing differential protection patterns both at the intact peptide and individual residue level.

By comparing two different states of labelled protein, for example with and without ligand, solvent accessible surfaces can be probed to determine binding sites of non-covalent interactions.

5.2 Methods

5.2.1 Expression and purification of MtrC

Purified MtrC was provided by Professor Julea Butt (University of East Anglia). α-TiO_2 was purchased from Nanostructured & Amorphous Materials, Inc. (Katy, USA). Protein concentration was measured by UV-Visible spectroscopy prior to measurements. The protocol is as described in (Fukushima et al., 2017).

5.2.2 Synchrotron x-ray footprinting

Performed by Dr Anna Stikane, MtrC samples with and without TiO_2 nanoparticles were footprinted at the Advanced Light Source (ALS, Lawrence Berkeley National Laboratory) beamline 3.2.1 as described previously (Fukushima et al., 2017). Briefly, 2 µM MtrC with 2 nmol α-TiO_2 where required were mixed in 5 mM Na_4HCl buffer. Samples were flowed through a flowing set up as described previously. Samples were irradiated for 0, 12.5, 25, 50 and 75 ms before collection in a sample tube containing 220 mM methionine amide. Offline desalting was performed directly following labelling for samples containing α-TiO_2, using 7k MWCO spin columns (Zeba, Thermo Fisher). Samples were spin concentrated before being snap frozen and stored at -80°C.

5.2.3 Digestion and LC-MS/MS

The samples were proteolytically digested with Glu-C and Asp-N prior to analysis. For this, 50 µL of sample were diluted with 50 µL solubilising buffer, containing 5% sodium dodecyl sulfate and 50 mM tetraethyl ammonium bromide (TEAB). Following this, dithiothreitol (DTT) was added to a final concentration of 20 mM before heating to 90^{o}C for 10 minutes. Reductive alkylation of cysteines was carried out with iodoacetic acid added to 40 mM concentration, before incubation for 30 minutes at 20^{o}C.

Following this, 7.8 µL 12% phosphoric acid were added, along with 1200 µL binding buffer (90% methanol, 100 mM TEAB). 2 µg of Glu-C was added and the sample moved to an S-trap (Protifi, NY USA), and centrifuged at 4,000x g for 30 min. The S-trap was washed three times with 150 µL binding buffer before the addition of 0.5 µg Glu-C and incubation for 60 min 47^{o}C. Peptides were eluted and dried before being resuspended in 50 µL of 50 mM ammonium biocarbonate. 0.2 µg Asp-N (Promega, UK) was added before incubation for 18 hours at 37^{o}C before quenching with 5 µL 1% trifluoracetic acid.

LC separation of the resulting peptides was performed on a ACQUITY M-Class UPLC (Waters, Manchester UK) by gradient elution of 1-60% solvent B (0.1% formic acid in acetonitrile) in A (0.1% formic acid in water) over 30 min at a flow rate of 0.3 µL/min. The column eluent was directly infused into a quadrupole-orthogonal time of flight mass spectrometer (Xevo G2-XS Q-TOF, Waters UK, Manchester) via a Z-spray nanoflow electrospray source. The instrument was operated in positive ion mode using a capillary voltage of 3.0 kV, cone voltage of 40 V, source offset of 80 V, and backing pressure of 3.58 mbar. The source temperature was 80°C. Argon was used as the buffer gas at a pressure of 8.6.

5.2.4 MS analysis

Peptides were identified using PEAKS X (Bioinformatics Solutions Inc, Waterloo Canada). Single and double oxidations (+15.99 Da and +31.98 Da modifications) were considered for each peptide. Peptides where a modification was found were quantified for the different conditions and time points to determine total modification. XICs were produced for unmodified and modified peaks across all charge states. The peak area of each XIC was determined by integrating under the curve using the MassLynx software (Waters, Manchester UK). Each peak was investigated manually to verify correct m/z values and to remove low abundance peaks with signal-to-noise (S/N) ratio of less than 3. For each exposure, total modification for each peptide was calculated from the sum of all

modified peptide peaks (i.e. single and double oxidations and all detected charges states, z) using the below equation:

$$Modified = \frac{\sum Modified\ peptide\ peaks}{\sum unmodified\ +\ \sum modified\ peptide\ peaks}$$

Equation 5.1: calculation of modified peptides for synchrotron oxidation

This % modification was then converted to fraction unmodified by the equation:

$$1 - modified$$

Equation 5.2: calculation of fraction unmodified in synchrotron oxidation

These values (percentages) were used to calculate the degree of modification for each peptide and time point with and without TiO_2. This was performed by plotting the fraction unmodified vs each exposure using a first order exponential fit in Originpro 2019 (OriginLab Corporation, Northampton, MA, USA). Ratios of the rate constant with and without nanoparticle were then identified by:

$$\frac{Rate\ constant\ unbound\ MtrC}{Rate\ constant\ MtrC + Nanoparticle}$$

Equation 5.3: calculation of rate constants

Peptides with a ratio of 2.5 or higher were identified as significantly different labelling between unbound and bound MtrC. This value was chosen as a value of two has been used previously in literature at the residue level (Fukushima et al., 2017). And such this value has been increased to account for the higher variability in peptide level calculations.

5.2.5 Electrostatic calculations

The c-type hemes of MtrC (minus iron atoms) were parameterised using PRODRG (Schüttelkopf and van Aalten, 2004). The charges and radius atoms were calculated using PDB2PQR (Dolinsky et al., 2007) using the Amber force field at pH 8. The linearized Poisson–Boltzmann equation was solved with APBS (Jurrus et al., 2018) using cubic spline charge discretization, with dielectric constants of

2.00 for the solute and 78.54 for the solvent at 298.15 K and including ions corresponding to 5 mM NH4Cl. Electrostatic potentials are displayed as surface potential maps (± 5 KbT/e) using Pymol 2.0 (Schrodinger LLC).

5.3 Results and discussion

5.3.1 Peptide level analysis

Initially, synchrotron labelling of MtrC against MtrC + TiO_2 was compared at the peptide level. Total ion counts (TIC) for each peptide both with and without TiO_2 reduced as the exposure time increased. For MtrC only, TIC was reduced from 2.94e8 for 0 msec exposure to 6.64e7 for the 75 msec exposure. For MtrC + TiO_2, the TIC was reduced from 1.95e8 for the 0 msec exposure to 3.09e7 for the 75 msec exposure. This reduction of TIC is likely due to a breakdown of the protein due to overexposure of X-ray dose. However, the decrease may be explained due to increases in aggregation due to higher levels of oxidation.

For the comparison between states, 18 modified peptides were selected for analysis. The reasoning for selection has been shown in supplemental information**.** Regions of protection are shown in figure 5.3.

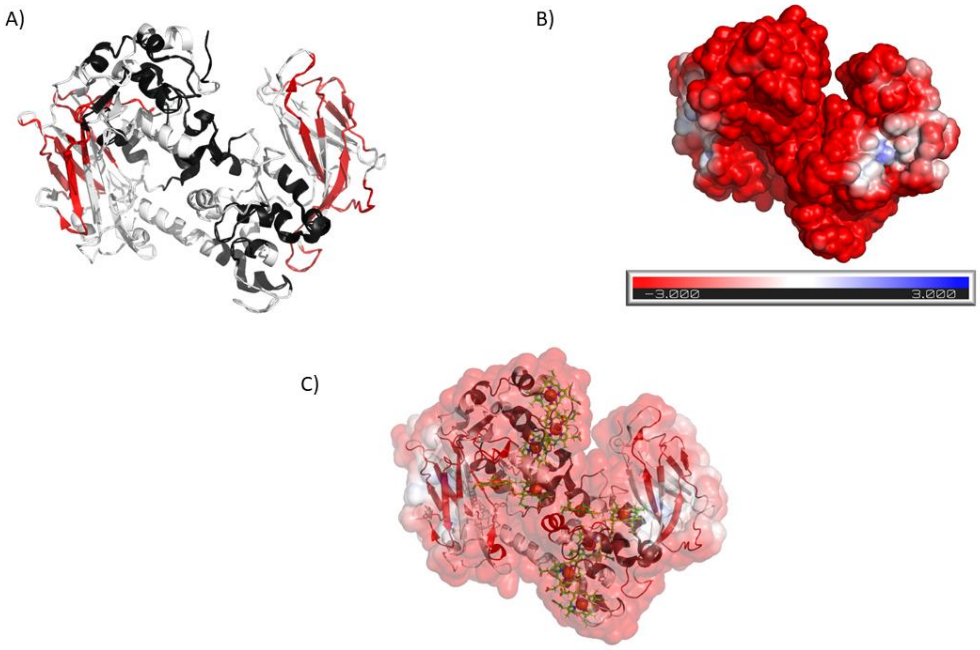


Figure 5.3: MtrC structure(Edwards, Marcus J. et al., 2015) showing significant differences in oxidation at the peptide level. A) Both exposed sides of the protein show a significant difference in labelling (above a 2.5x ratio). B) Electrostatic map showing negative regions in red and positive regions in blue. C) Ribbon diagram superimposed on the electrostatic diagram showing that differential labelling occurs mainly in the positively charged regions on opposite sides of the protein. The uptake curves for each peptide are shown in the supplemental information figures 8.27-39, and the information for each peptide table 8.9-10.

Significantly different regions of oxidation, i.e., protection upon nanoparticle binding, were found between residues 73-152 and 354-487. These sequences, shown in figure 5.4, correspond to two highly exposed regions on opposite sides of the MtrC protein, as it is orientated into the barrel as shown in figure 5.1. As MtrC is approximately 8 nm x 7 nm x 4 nm in size, with the TiO_2 nanoparticles being approximately 5 nm (Edwards, Marcus J. et al., 2015), it is apparent that one particle cannot bind both epitopes at the same time. I therefore assume that both positively charged patches on opposite sides of the protein are alternate sites of binding, although concurrent binding of two TiO_2

particles, one on either side, would also be possible. In either case, (partial) protection from oxidation will be observed in these regions, as is indeed observed in the peptide level data.

The ten heme groups within MtrC are covalently bound to CXXCH- residues, however no modifications were found within these regions. However, given that this procedure occurs post labelling, solvent accessibility to these peptides is likely blocked by the bound haem c groups.

Higher modification was found, as expected, in longer exposure samples.

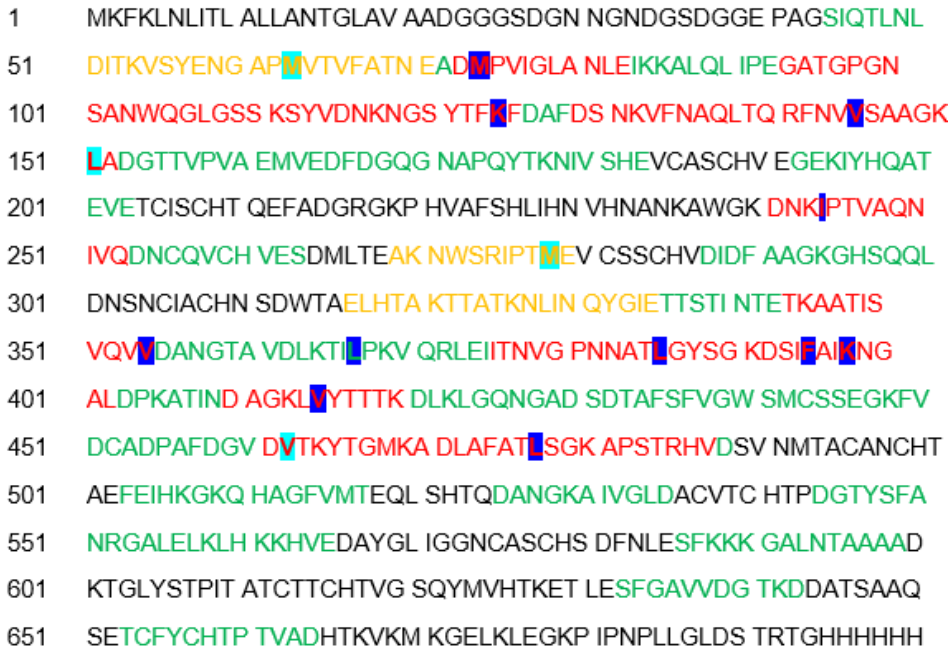


Figure 5.4 MS/MS analysis of MtrC oxidation products with and without TiO_2, at both the peptide and single amino acid level. Residues in black are not covered in the peptide digest LC-MS/MS spectra. Peptide sequences in green are covered but showed no modification. Residues in yellow were modified but not significantly, whereas residues in red were significantly modified (at least two-fold change in oxidation). Additionally, single residues in cyan were modified but not significantly, and residues in blue were significantly different in oxidation (i.e., protected upon nanoparticle binding).

5.3.2 Residue level analysis

Using a different strategy in addition to the peptide-level analysis described above and in the Methods section 2.3.3, an additional search was undertaken for individual oxidised amino acids (single and double oxidations, +15.99 Da and +31.98 Da, were considered), in order to pinpoint the sites of differential modification more accurately. 15 residues (shown boxed in cyan or blue in Figure 5.4) were found to differ in oxidation, these are shown in the supplemental. For 11 out of 13 peptides the difference was found to be significant, i.e. at least two-fold. This relatively low number of residues is due to the high stringency criteria used, as only those residues were considered where the exact site of oxidation was clearly identified in the MS/MS spectra. Overall there is good agreement with most prominently modified amino acids also picked via oxidation of their corresponding peptides, as expected, apart from L367 which is a significantly different residue located within an unmodified peptide. Of particular interest is the region located between residues 354-480, which shows seven significantly different residues. These are identified in Figure 5.5**.**

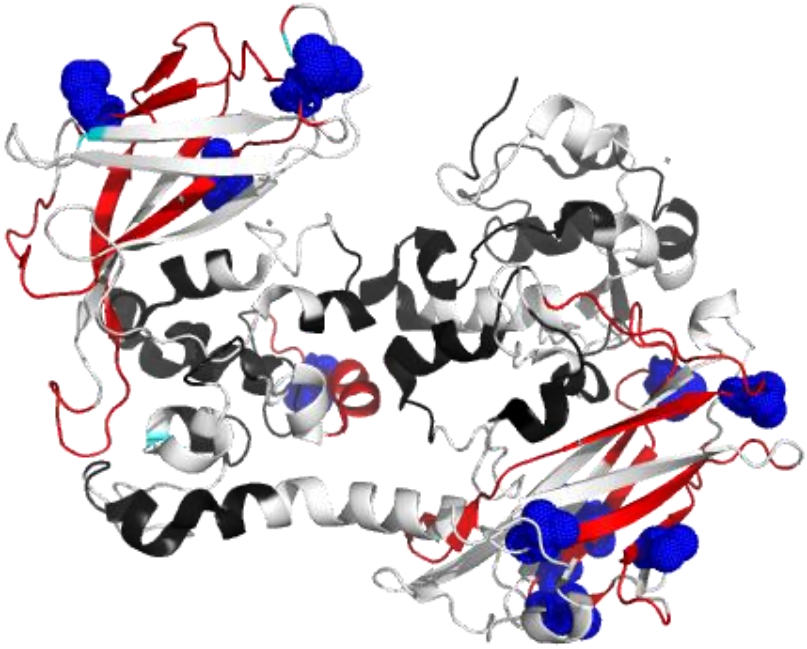


Figure 5.5: labelled structure of MtrC from Figure 5.3 with single residues superimposed. Significantly different residues are shown in deep blue with not significantly different residues in cyan. The significantly different residues have strong conformity with the significantly different residues identified in figure 5.3 (in red).

5.3.3 Electrostatic calculations

I see two specific positive regions at each end of the molecule, in the orientation in the complex, as shown in Figure 3b. TiO_2 has shown affinity to positively charged residues within peptides previously (Chen et al., 2008; Limo et al., 2018), specifically lysine and arginine. It is therefore plausible a large positively charged region would be likely to undertake binding through electrostatic interactions with electrostatically positive charged regions.

As shown in figure 5.3, there is a distinct correlation between deprotection at the peptide level and regions of electrostatic. The region between peptide 350-480 is superimposed in figure 5.3, and we

can see significantly different single residues, significantly different peptide residues and positive electrostatic regions within the same region. This strongly suggests the location of the TiO_2 binding occurs within this region, with similar results for the region between peptide 73-152.

As discussed previously, due to the possibility of multiple partial binding locations, the localised positive regions at each end of the protein provide a general large area where binding can be expected. However, utilising a technique such as scanning transmission electron microscopy (STEM) imaging or a fluorescence quenching assay, it would be possible to confirm that the TiO_2 particle is binding, and that conformational changes are due to binding and not due to allosteric changes from interaction with the TiO_2 molecule.

These findings show the location of binding at each end of the molecule, a significant distance away from the haem chain at the top of the molecule. Whilst this makes sense chemically, suggesting electrostatic interactions are the primary driver of the binding of TiO_2. This poses problems for the design of future engineered biohybrid systems, as the location of the nanoparticle is not optimal for electron transfer. Further, focus on the specific chemistry of binding may be of particular use for the design of biohybrids moving forward.

5.4 Conclusions

These experiments identify two possible binding sites for TiO_2 on the MtrC molecule, one on either side of the orientation shown in the MtrCAB complex. Both peptide level and residue level significantly different regions were observed, suggesting TiO_2 has two discrete binding sites on the molecule. This is further supported by electrostatic calculations, which show two discrete regions of less negativity, in the same regions identified by XRFP experiments.

To determine whether binding is site specific, and conformational changes are occurring due to binding, as mentioned above, fluorescence quenching or STEM imaging could be used. Additionally,

HDX could be used to monitor the dynamic regions of the protein over a longer time course, to determine to determine if the arrangements observed in this experiment are conserved over time.

Further work is required to determine whether the binding is specific to each region and is maintained at the complex and cellular level. This would require performing the technique either in cell, or in a larger complex. Further, recording electron transfer rates would be useful to see how good TiO_2 is a conduit for electron transfer.

6. Conclusions and future work

This thesis explores the use of three different footprinting techniques to study biomolecules; FPOP of peptides, HDX of antibodies and X-ray footprinting of MtrC. All three chapters have sought to improve the knowledge of each individual technique, and aim to improve the application of footprinting in each case. The FPOP experiments have advanced knowledge of localised neighbourhood effects, the HDX experiments an improvement in analysis of antibodies in complex formulations, and the X-ray footprinting experiments offer applications of this technique for determining binding of TiO_2 to proteins for advancement in biomolecular battery technology.

The primary objective with the FPOP experiment was to enable the determination of a set of rules for neighbourhood effects from the primary sequence of a peptide, should neighbourhood effects be detected. Whilst a rule hasn't been determined, this thesis has provided evidence to push towards a defining rule combining primary sequence and changes in oxidation.

To the authors' knowledge this is the first time a neighbourhood effect has been conclusively proven with regard to oxidative labelling. This is significant for the future interpretation of FPOP data, specifically for comparison experiments where changes in amino acid sequence occur, as labelling of the local environment of the peptide may be influenced by a substitution in amino acid sequence. However, further work, as discussed in chapter 2, should be to probe structures with fixed or rigid dimensions, such that the primary sequence rule can be confirmed to be applicable to rigid

structures with neighbourhood effects, and thus be extended to an interpretation of the tertiary structure and not just primary sequence as interrogated within these experiments.

It is noted that the LC-MS separation within these experiments was not precise enough to determine the individual composition of each peak, and thus only interpretation at the peptide level was possible. Further work to separate each individual isomer and determine the oxidation of each constituent would be advantageous.

Further, due to the lack of solubility of phenylalanine and tryptophan in a glycine peptide, lysine residues were required to dissolve the peptides to a suitable level to perform analysis. Further work involving the removal of these residues would simplify MS/MS sequencing.

Studying peptides with different amino acid residue compositions would determine the relationship between highly reactive residues and less reactive residues. In addition hydroxyl radical labelling reactivity for FPOP were determined through synchrotron experiments, and further investigation of the reaction rates of residues through the FPOP system would be useful for the FPOP community in general.

The aim for the HDX chapter was to determine the suitability of this method for complex samples, and this aim was broadly achieved in this thesis. Antibodies in complex mixtures were probed and changes identified through a HDX-MS assay. Further increases in coverage would have determined the suitability of this approach for use within analysis of pharmaceuticals that have been stressed over time. Specifically, an increase of the coverage for the CDR regions of the antibodies which were not covered by this method would be required for routine pharmaceutical assays.

However, this method did identify significant differences between WFL and STT in the presence of polysorbate 20 and 80. Thus the primary criteria for this set of experiments was met, however further improvement of this method would allow for this to be transitioned into an industry standard in the future.

The broad aim for the MtrC chapter was to determine the binding locations of TiO_2. This goal was met, as two binding locations were determined on the molecule. Further it was determined that this binding was primarily due to electrostatic conditions as opposed to chemical binding. This is the first reported incidence of how TiO_2 binds to MtrC, and has significant implications into biohybrid generation in the future. Further, this experiment shows the viability for oxidative footprinting in relation to determination of binding sites of biohybrids.

Given our investigations utilising all three footprinting techniques, a comparison between all three footprinting techniques can be drawn.

HDX-MS is a well characterised technique within the literature for interrogating dynamic regions and higher order structural changes within protein experiments. The exchange takes place along the backbone amides of the amino acid sequence, in comparison to both FPOP and X-Ray footprinting, which are labelled on the amino acid sidechain. This gives HDX a significant advantage in data analysis, as labelling is not dictated by the reactivity of the residue. Furthermore, software packages that currently exist for the analysis of HDX-MS data are significantly more advanced than for radical based labelling techniques.

Further, due to the entire deuteration of the molecule, the results from HDX are highly reproducible over multiple experiments, as the dosage is controlled. This is an advantage compared to FPOP, where the dosage can vary between experiments due to different buffers being used. However, advances have been made in this area, with the inclusion of a radical dosimeter, which utilises an inline UV detector to accurately determine the radical dose administered via FPOP. X-ray footprinting also has high reproducibility, as the radicals are generated from the water molecules within the sample vial, and not lysed from H_2O_2 as in an FPOP experiment, and the irradiation dosage is precisely controlled through the synchrotron.

HDX suffers a significant drawback in comparison to the covalent labelling methods with regard to digestion and chromatography before analysis by MS. As soon as the deuteration step is quenched,

back exchange of deuterium back to hydrogen is initiated. Thus, the digestion and chromatography must be performed rapidly to preserve the deuterated amide bonds. Breaking of disulphide bonds is of particular challenge, as TCEP must be used to rapidly break down these bonds prior to analysis. Further the enzyme used for digestion in HDX is commonly pepsin, which can digest at the low quench pH used for HDX. This enzyme is highly non-specific, and coupled with poor disulphide bond breaking can make analysis of complex samples challenging.

The label for both FPOP and X-ray footprinting is a non-reversable OH radical. This means downstream processing prior to analysis can be tailored to the molecule of interest. Selection of digestion enzyme and length of chromatography can be tailored to each experiment, to increase coverage and reduce data analysis complexity.

Finally, HDX labelling is not capable of recording events on the µs to ms timescale that radical labelling workflows can achieve. Thus, for analytes that have highly dynamic regions, HDX may not be able to capture information that is captured through FPOP and X-ray footprinting. By utilising both techniques, rapid conformational changes can be detected with FPOP, and HDX can probe changes over time, for less dynamic regions, building a complete assessment of both highly dynamic and less dynamic regions.

FPOP, in addition to the factors discussed, has other advantages, especially when compared to X-Ray footprinting. The equipment required to perform the experiment is more accessible, compared to the sparsity of beamlines that can facilitate X-ray footprinting. Further, FPOP is only performed at one time point. This could be considered both an advantage or disadvantage, whilst only one time point does not allow for a scale of time points in radical dosage, the data analysis required is reduced, as each individual time point for an X-ray footprinting experiment is the same amount of data analysis as an entire FPOP experiment. Furthermore, dosage can be tailored in an FPOP experiment by reducing the scavenger concentration or increasing hydrogen peroxide concentration.

This is less applicable to X-ray footprinting, whereby higher doses of X-rays may cause structural perturbations in the analyte.

The two drawbacks of FPOP are differences in reactivity of amino acid sidechains with OH radicals, and the inability to record dosage to the analyte of interest. When analysing complex mixtures, such as drug formulations with differing excipients, the excipients themselves can be labelled, reducing the radical dosage available for the protein. Thus, if one excipient labels at a higher propensity than in the comparative study, the labelling result for the protein may be erroneous.

Secondly, due to the difference in reactivity of the sidechains, a highly reactive sidechain can label order of magnitudes more than a less reactive sidechain. Thus, when analysing at the peptide level, changes in highly reactive residues such as tryptophan and phenylalanine dominate the total oxidation of the peptide, and as such relatively large changes in less reactive sidechains have little effect on the final data. This is a problem mirrored with X-ray footprinting, however is avoided in HDX workflows.

X-ray footprinting shares the negatives of FPOP with regard to sidechain reactivity. However, as no scavenger is used in the reaction, secondary radical formation from scavengers is mitigated in this case, which may decrease radical lifetime, however this requires further investigation. Further, as a number of time points are taken, the experiment can show dynamic changes over time, similar to an HDX experiment, and not a single timepoint as observed in FPOP.

This thesis has aimed to show the advantages and disadvantages of each method, and how each method may be applicable to pharmaceutical applications dependent on the formulation and analyte of interest.

However, further work is required to advance the ideas presented within this thesis. With regard to FPOP, by further understanding the rules behind labelling, it is hoped in the future this technique can be incorporated as a standardised method for higher-order structure analysis by mass

spectrometry. To do this, a number of steps would need to be implemented. Firstly, a more automated pipeline for analysis would be required to utilise FPOP within a pharmaceutical setting. Secondly, more understanding of the interactions between labelling and neighbourhood effects is required to determine whether measurements taken within the experiment are a true representation of higher-order structure. Finally, an accurate determination of the dosage of oxidation is required to enable comparison between multiple samples with different buffer productions.

With regard to the polysorbate HDX-MS experiments, it is clear that sequence coverage was an issue with this method. Thus, a standardised method for use in pharmaceutical workflows would require further improvement to ensure all quality control attributes are covered through higher sequence coverage. Further, more antibodies would need to be tested to ensure this method is suitable for other systems than the ones tested. Finally, a way to easily display the differences between different time points would be of interest.

With regard to the synchrotron experiments, further work is required to determine whether binding of the TiO_2 molecules allows for transport of electrons through the electron chain. This would allow us to determine whether this binding is viable for electron transport.

To conclude, these experiments have advanced knowledge of footprinting techniques for use in the pharmaceutical industry, and with the further experiments suggested, footprinting has the potential to be a cornerstone of analysis for pharmaceutical applications in the future.

References

Adilakshmi, T., Soper, S.F.C. and Woodson, S.A. 2009. Chapter 12 - Structural Analysis of RNA in Living Cells by In Vivo Synchrotron X-Ray Footprinting. *Methods in Enzymology.* Academic Press, pp.239-258.

Agarkhed, M., O'Dell, C., Hsieh, M.C., Zhang, J., Goldstein, J. and Srivastava, A. 2013. Effect of polysorbate 80 concentration on thermal and photostability of a monoclonal antibody. *AAPS PharmSciTech.* **14**(1), pp.1-9.

Ahadi, E. and Konermann, L. 2011. Ejection of Solvated Ions from Electrosprayed Methanol/Water Nanodroplets Studied by Molecular Dynamics Simulations. *Journal of the American Chemical Society.* **133**(24), pp.9354-9363.

Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y. and Timasheff, S.N. 2007. Suppression of protein interactions by arginine: A proposed mechanism of the arginine effects. *Biophysical Chemistry.* **127**(1), pp.1-8.

Arakawa, T. and Timasheff, S.N. 1985. The stabilization of proteins by osmolytes. *Biophysical Journal.* **47**(3), pp.411-414.

Arakawa, T. and Tsumoto, K. 2003. The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. *Biochemical and Biophysical Research Communications.* **304**(1), pp.148-152.

Auton, M. and Bolen, D.W. 2004. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry.* **43**(5), pp.1329-1342.

Auton, M. and Bolen, D.W. 2007. Application of the transfer model to understand how naturally occurring osmolytes affect protein stability. *Methods Enzymol.* **428**, pp.397-418.

Ayorinde, F.O., Gelain, S.V., Johnson, J.H., Jr. and Wan, L.W. 2000. Analysis of some commercial polysorbate formulations using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom.* **14**(22), pp.2116-2124.

Bam, N.B., Cleland, J.L. and Randolph, T.W. 1996. Molten globule intermediate of recombinant human growth hormone: stabilization with surfactants. *Biotechnol Prog.* **12**(6), pp.801-809.

Barnett, G.V., Razinkov, V.I., Kerwin, B.A., Blake, S., Qi, W., Curtis, R.A. and Roberts, C.J. 2016. Osmolyte Effects on Monoclonal Antibody Stability and Concentration-Dependent Protein Interactions with Water and Common Osmolytes. *The Journal of Physical Chemistry B.* **120**(13), pp.3318-3330.

Batra, J., Xu, K., Qin, S. and Zhou, H.-X. 2009. Effect of Macromolecular Crowding on Protein Binding Stability: Modest Stabilization and Significant Biological Consequences. *Biophysical Journal.* **97**(3), pp.906-911.

Bekale, L., Agudelo, D. and Tajmir-Riahi, H.A. 2015. The role of polymer size and hydrophobic end-group in PEG–protein interaction. *Colloids and Surfaces B: Biointerfaces.* **130**, pp.141-148.

Bienta. *LogD/LogP.* [Online]. [Accessed 01/09/2023]. Available from: https://bienta.net/logdlogp/

Blumberger, J. 2015. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chemical Reviews.* **115**(20), pp.11191-11238.

Boesl, U. 2017. Time-of-flight mass spectrometry: Introduction to the basics. *Mass Spectrom Rev.* **36**(1), pp.86-109.

Borotto, N.B. and Graham, K.A. 2021. Fragmentation and Mobility Separation of Peptide and Protein Ions in a Trapped-Ion Mobility Device. *Analytical Chemistry.* **93**(29), pp.9959-9964.

Brader, M.L., Estey, T., Bai, S., Alston, R.W., Lucas, K.K., Lantz, S., Landsman, P. and Maloney, K.M. 2015. Examination of Thermal Unfolding and Aggregation Profiles of a Series of Developable Therapeutic Monoclonal Antibodies. *Molecular Pharmaceutics.* **12**(4), pp.1005-1017.

Brady, J., Dürig, T., Lee, P.I. and Li, J.X. 2017. Chapter 7 - Polymer Properties and Characterization. In: Qiu, Y., et al. eds. *Developing Solid Oral Dosage Forms (Second Edition).* Boston: Academic Press, pp.181-223.

Brédy, R., Hervé, M., Boyer, A., Brown, J.M., Compagnon, I. and Lépine, F. 2022. Non-ergodic fragmentation of protonated reserpine using femtosecond laser activation. *International Journal of Mass Spectrometry.* **471**, p116729.

Breuer, M., Rosso, K.M. and Blumberger, J. 2014. Electron flow in multiheme bacterial cytochromes is a balancing act between heme electronic interaction and redox potentials. *Proceedings of the National Academy of Sciences.* **111**(2), pp.611-616.

Calabrese, A.N., Ault, J.R., Radford, S.E. and Ashcroft, A.E. 2015. Using hydroxyl radical footprinting to explore the free energy landscape of protein folding. *Methods.* **89**, pp.38-44.

Cameron, A.E. and Eggers, D.F. 1948. An Ion ``Velocitron''. *Review of Scientific Instruments.* **19**(9), pp.605-607.

Catherman, A.D., Skinner, O.S. and Kelleher, N.L. 2014. Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun.* **445**(4), pp.683-693.

Chang, H.W. and Bock, E. 1980. Pitfalls in the use of commercial nonionic detergents for the solubilization of integral membrane proteins: Sulfhydryl oxidizing contaminants and their elimination. *Analytical Biochemistry.* **104**(1), pp.112-117.

Chea, E.E. and Jones, L.M. 2018. Modifications generated by fast photochemical oxidation of proteins reflect the native conformations of proteins. *Protein Sci.* **27**(6), pp.1047-1056.

Chen, H., Su, X., Neoh, K.G. and Choe, W.S. 2008. Probing the interaction between peptides and metal oxides using point mutants of a TiO2-binding peptide. *Langmuir.* **24**(13), pp.6852-6857.

Chou, D.K., Krishnamurthy, R., Randolph, T.W., Carpenter, J.F. and Manning, M.C. 2005. Effects of Tween 20 and Tween 80 on the stability of Albutropin during agitation. *J Pharm Sci.* **94**(6), pp.1368-1381.

Colón, W., Church, J., Sen, J., Thibeault, J., Trasatti, H. and Xia, K. 2017. Biological Roles of Protein Kinetic Stability. *Biochemistry.* **56**(47), pp.6179-6186.

Cornwell, O., Ault, J.R., Bond, N.J., Radford, S.E. and Ashcroft, A.E. 2021. Investigation of D76N β2-Microglobulin Using Protein Footprinting and Structural Mass Spectrometry. *Journal of the American Society for Mass Spectrometry.*

Cornwell, O., Radford, S.E., Ashcroft, A.E. and Ault, J.R. 2018a. Comparing Hydrogen Deuterium Exchange and Fast Photochemical Oxidation of Proteins: a Structural Characterisation of Wild-Type and DeltaN6 beta2-Microglobulin. *J Am Soc Mass Spectrom.* **29**(12), pp.2413-2426.

Cornwell, O., Radford, S.E., Ashcroft, A.E. and Ault, J.R. 2018b. Comparing Hydrogen Deuterium Exchange and Fast Photochemical Oxidation of Proteins: a Structural Characterisation of Wild-Type and ΔN6 β(2)-Microglobulin. *J Am Soc Mass Spectrom.* **29**(12), pp.2413-2426.

Cristobal, A., Marino, F., Post, H., van den Toorn, H.W.P., Mohammed, S. and Heck, A.J.R. 2017. Toward an Optimized Workflow for Middle-Down Proteomics. *Analytical Chemistry.* **89**(6), pp.3318-3325.

de Graaf, C., Donnelly, D., Wootten, D., Lau, J., Sexton, P.M., Miller, L.J., Ahn, J.-M., Liao, J., Fletcher, M.M., Yang, D., Brown, A.J.H., Zhou, C., Deng, J. and Wang, M.-W. 2016. Glucagon-Like Peptide-1 and Its Class B G Protein–Coupled Receptors: A Long March to Therapeutic Successes. *Pharmacological Reviews.* **68**(4), pp.954-1013.

de Hoffmann, E. and Stroobant, V. 2007. *Mass Spectrometry: Principles and Applications, 3rd Edition.* Wiley-Interscience.

Dixit, N., Salamat-Miller, N., Salinas, P.A., Taylor, K.D. and Basu, S.K. 2016. Residual Host Cell Protein Promotes Polysorbate 20 Degradation in a Sulfatase Drug Product Leading to Free Fatty Acid Particles. *Journal of Pharmaceutical Sciences.* **105**(5), pp.1657-1666.

Dobson, C.L., Devine, P.W.A., Phillips, J.J., Higazi, D.R., Lloyd, C., Popovic, B., Arnold, J., Buchanan, A., Lewis, A., Goodman, J., van der Walle, C.F., Thornton, P., Vinall, L., Lowne, D., Aagaard, A., Olsson, L.-L., Ridderstad Wollberg, A., Welsh, F., Karamanos, T.K., Pashley, C.L., Iadanza, M.G., Ranson, N.A., Ashcroft, A.E., Kippen, A.D., Vaughan, T.J., Radford, S.E. and Lowe, D.C. 2016. Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. *Scientific Reports.* **6**(1), p38644.

Dobson, J., Kumar, A., Willis, L.F., Tuma, R., Higazi, D.R., Turner, R., Lowe, D.C., Ashcroft, A.E., Radford, S.E., Kapur, N. and Brockwell, D.J. 2017. Inducing protein aggregation by extensional flow. *Proceedings of the National Academy of Sciences.* **114**(18), pp.4673-4678.

Dole, M., Mack, L.L., Hines, R.L., Mobley, R.C., Ferguson, L.D. and Alice, M.B. 1968. Molecular Beams of Macroions. *The Journal of Chemical Physics.* **49**(5), pp.2240-2249.

Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G. and Baker, N.A. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research.* **35**(suppl_2), pp.W522-W525.

Doyle Drbohlav, L.M., Sharma, A.N., Gopalrathnam, G., Huang, L. and Bradley, S. 2019. A Mechanistic Understanding of Polysorbate 80 Oxidation in Histidine and Citrate Buffer Systems-Part 2. *PDA J Pharm Sci Technol.* **73**(4), pp.320-330.

Du, M., Wu, W., Ercal, N. and Ma, Y. 2004. Simultaneous determination of 3-nitro tyrosine, o-, m-, and p-tyrosine in urine samples by liquid chromatography–ultraviolet absorbance detection with pre-column cloud point extraction. *Journal of Chromatography B.* **803**(2), pp.321-329.

Duncan, M.W., Nedelkov, D., Walsh, R. and Hattan, S.J. 2016. Applications of MALDI Mass Spectrometry in Clinical Chemistry. *Clinical Chemistry.* **62**(1), pp.134-143.

Dupree, E.J., Jayathirtha, M., Yorkey, H., Mihasan, M., Petre, B.A. and Darie, C.C. 2020. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field. *Proteomes.* **8**(3).

Edwards, M.J., White, G.F., Butt, J.N., Richardson, D.J. and Clarke, T.A. 2020. The Crystal Structure of a Biological Insulated Transmembrane Molecular Wire. *Cell.* **181**(3), pp.665-673.e610.

Edwards, M.J., White, G.F., Norman, M., Tome-Fernandez, A., Ainsworth, E., Shi, L., Fredrickson, J.K., Zachara, J.M., Butt, J.N., Richardson, D.J. and Clarke, T.A. 2015. Redox Linked Flavin Sites in Extracellular Decaheme Proteins Involved in Microbe-Mineral Electron Transfer. *Scientific Reports.* **5**(1), p11677.

Eliuk, S. and Makarov, A. 2015. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu Rev Anal Chem (Palo Alto Calif).* **8**, pp.61-80.

Englander, S.W. 2006. Hydrogen exchange and mass spectrometry: A historical perspective. *J Am Soc Mass Spectrom.* **17**(11), pp.1481-1489.

Englander, S.W., Sosnick, T.R., Englander, J.J. and Mayne, L. 1996. Mechanisms and uses of hydrogen exchange. *Curr Opin Struct Biol.* **6**(1), pp.18-23.

Espino, J.A., Mali, V.S. and Jones, L.M. 2015. In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells. *Anal Chem.* **87**(15), pp.7971-7978.

European Directorate for the Quality of Medicines 2002. *European Pharmacopoeia 4th edition.* [Online]. Strasbourg: Council of Europe.

Faunce, T.A., Lubitz, W., Rutherford, A.W., MacFarlane, D., Moore, G.F., Yang, P., Nocera, D.G., Moore, T.A., Gregory, D.H., Fukuzumi, S., Yoon, K.B., Armstrong, F.A., Wasielewski, M.R. and Styring, S. 2013. Energy and environment policy case for a global project on artificial photosynthesis. *Energy & Environmental Science.* **6**(3), pp.695-698.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* **246**(4926), pp.64-71.

Fernandez-Puente, P., Mateos, J., Blanco, F.J. and Ruiz-Romero, C. 2014. LC-MALDI-TOF/TOF for shotgun proteomics. *Methods Mol Biol.* **1156**, pp.27-38.

Fort, K.L., Cramer, C.N., Voinov, V.G., Vasil'ev, Y.V., Lopez, N.I., Beckman, J.S. and Heck, A.J.R. 2018. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *Journal of Proteome Research.* **17**(2), pp.926-933.

Frison-Norrie, S. and Sporns, P. 2001. Investigating the Molecular Heterogeneity of Polysorbate Emulsifiers by MALDI-TOF MS. *Journal of Agricultural and Food Chemistry.* **49**(7), pp.3335-3340.

Fukushima, T., Gupta, S., Rad, B., Cornejo, J.A., Petzold, C.J., Chan, L.J.G., Mizrahi, R.A., Ralston, C.Y. and Ajo-Franklin, C.M. 2017. The Molecular Basis for Binding of an Electron Transfer Protein to a Metal Oxide Surface. *Journal of the American Chemical Society.* **139**(36), pp.12647-12654.

Gau, B.C., Sharp, J.S., Rempel, D.L. and Gross, M.L. 2009. Fast photochemical oxidation of protein footprints faster than protein unfolding. *Anal Chem.* **81**(16), pp.6563-6571.

Gessel, M.M., Norris, J.L. and Caprioli, R.M. 2014. MALDI imaging mass spectrometry: spatial molecular analysis to enable a new age of discovery. *J Proteomics.* **107**, pp.71-82.

Griffin, W.C. 1954. Calculation of HLB Values of Non-Ionic Surfactants. *Journal of the Society of Cosmetic Chemists.* **5**(4), pp.249-256.

Gupta, S., Feng, J., Chan, L.J., Petzold, C.J. and Ralston, C.Y. 2016. Synchrotron X-ray footprinting as a method to visualize water in proteins. *J Synchrotron Radiat.* **23**(Pt 5), pp.1056-1069.

Ha, E., Wang, W. and John Wang, Y. 2002. Peroxide formation in polysorbate 80 and protein stability. *Journal of Pharmaceutical Sciences.* **91**(10), pp.2252-2264.

Hambly, D.M. and Gross, M.L. 2005. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J Am Soc Mass Spectrom.* **16**(12), pp.2057-2063.

Hensley, K., Williamson, K.S., Maidt, M.L., Prasad Gabbita, S., Grammas, P. and Floyd, R.A. 1999. Determination of Biological Oxidative Stress Using High Performance Liquid Chromatography with Electrochemical Detection (HPLC-ECD). *Journal of High Resolution Chromatography.* **22**(8), pp.429-437.

Hildonen, S., Halvorsen, T.G. and Reubsaet, L. 2014. Why less is more when generating tryptic peptides in bottom-up proteomics. *Proteomics.* **14**(17-18), pp.2031-2041.

Hwang, E.T., Sheikh, K., Orchard, K.L., Hojo, D., Radu, V., Lee, C.-Y., Ainsworth, E., Lockwood, C., Gross, M.A., Adschiri, T., Reisner, E., Butt, J.N. and Jeuken, L.J.C. 2015. A Decaheme Cytochrome as a Molecular Electron Conduit in Dye-Sensitized Photoanodes. *Advanced Functional Materials.* **25**(15), pp.2308-2315.

Iavarone, A.T. and Williams, E.R. 2003. Mechanism of charging and supercharging molecules in electrospray ionization. *J Am Chem Soc.* **125**(8), pp.2319-2327.

Iribarne, J.V. and Thomson, B.A. 1976. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics.* **64**(6), pp.2287-2294.

Ishrat, M., Hassan, M.I., Ahmad, F. and Islam, A. 2018. Sugar osmolytes-induced stabilization of RNase A in macromolecular crowded cellular environment. *International Journal of Biological Macromolecules.* **115**, pp.349-357.

IUPAC. 1997. *Compendium of Chemical Terminology.* Oxford: Blackwell Scientific Publications.

Jenkins, A.D., Kratochvíl, P., Stepto, R.F.T. and Suter, U.W. 1996. *Glossary of basic terms in polymer science (IUPAC Recommendations 1996)*. Pure and Applied Chemistry. 68. p.2287. [Accessed 2018-11-06t11:17:36.793+01:00]. Available from: https://www.degruyter.com/view/j/pac.1996.68.issue-12/pac199668122287/pac199668122287.xml

Jiang, X., Burger, B., Gajdos, F., Bortolotti, C., Futera, Z., Breuer, M. and Blumberger, J. 2019. Kinetics of trifurcated electron flow in the decaheme bacterial proteins MtrC and MtrF. *Proceedings of the National Academy of Sciences.* **116**(9), pp.3425-3430.

Jiang, X., van Wonderen, J.H., Butt, J.N., Edwards, M.J., Clarke, T.A. and Blumberger, J. 2020. Which Multi-Heme Protein Complex Transfers Electrons More Efficiently? Comparing MtrCAB from

Shewanella with OmcS from Geobacter. *The Journal of Physical Chemistry Letters.* **11**(21), pp.9421-9425.

Johnson, A.R. and Carlson, E.E. 2015. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. *Analytical Chemistry.* **87**(21), pp.10668-10678.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D.W., Dolinsky, T., Konecny, R., Koes, D.R., Nielsen, J.E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.W., Holst, M.J., McCammon, J.A. and Baker, N.A. 2018. Improvements to the <scp>APBS</scp> biomolecular solvation software suite. *Protein Science.* **27**(1), pp.112-128.

Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry.* **60**(20), pp.2299-2301.

Kerr, R.A., Keire, D.A. and Ye, H. 2019. The impact of standard accelerated stability conditions on antibody higher order structure as assessed by mass spectrometry. *mAbs.* **11**(5), pp.930-941.

Kerwin, B.A. 2008. Polysorbates 20 and 80 used in the formulation of protein biotherapeutics: structure and degradation pathways. *J Pharm Sci.* **97**(8), pp.2924-2935.

Kim, M.S. and Pandey, A. 2012. Electron transfer dissociation mass spectrometry in proteomics. *Proteomics.* **12**(4-5), pp.530-542.

Kishore, R.S., Kiese, S., Fischer, S., Pappenberger, A., Grauschopf, U. and Mahler, H.C. 2011a. The degradation of polysorbates 20 and 80 and its potential impact on the stability of biotherapeutics. *Pharm Res.* **28**(5), pp.1194-1210.

Kishore, R.S., Pappenberger, A., Dauphin, I.B., Ross, A., Buergi, B., Staempfli, A. and Mahler, H.C. 2011b. Degradation of polysorbates 20 and 80: studies on thermal autoxidation and hydrolysis. *J Pharm Sci.* **100**(2), pp.721-731.

Konermann, L., Ahadi, E., Rodriguez, A.D. and Vahidi, S. 2013. Unraveling the mechanism of electrospray ionization. *Anal Chem.* **85**(1), pp.2-9.

Konermann, L., Pan, J. and Liu, Y.-H. 2011. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews.* **40**(3), pp.1224-1234.

Konermann, L., Rodriguez, A.D. and Liu, J. 2012. On the Formation of Highly Charged Gaseous Ions from Unfolded Proteins by Electrospray Ionization. *Analytical Chemistry.* **84**(15), pp.6798-6804.

Labrenz, S.R. 2014. Ester Hydrolysis of Polysorbate 80 in mAb Drug Product: Evidence in Support of the Hypothesized Risk After the Observation of Visible Particulate in mAb Formulations. *Journal of Pharmaceutical Sciences.* **103**(8), pp.2268-2277.

Lagassé, H.A.D., Alexaki, A., Simhadri, V.L., Katagiri, N.H., Jankowski, W., Sauna, Z.E. and Kimchi-Sarfaty, C. 2017. Recent advances in (therapeutic protein) drug development. *F1000Research.* **6**, p113.

Lanucara, F., Holman, S.W., Gray, C.J. and Eyers, C.E. 2014. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature Chemistry.* **6**(4), pp.281-294.

Larson, N.R., Wei, Y., Prajapati, I., Chakraborty, A., Peters, B., Kalonia, C., Hudak, S., Choudhary, S., Esfandiary, R., Dhar, P., Schöneich, C. and Middaugh, C.R. 2020. Comparison of Polysorbate 80 Hydrolysis and Oxidation on the Aggregation of a Monoclonal Antibody. *Journal of Pharmaceutical Sciences.* **109**(1), pp.633-639.

Lau, A.M., Claesen, J., Hansen, K. and Politis, A. 2021. Deuteros 2.0: peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry. *Bioinformatics.* **37**(2), pp.270-272.

Laurent, T.C. 1963. THE INTERACTION BETWEEN POLYSACCHARIDES AND OTHER MACROMOLECULES. 5. THE SOLUBILITY OF PROTEINS IN THE PRESENCE OF DEXTRAN. *Biochemical Journal.* **89**(2), p253.

Lee, L.L.Y. and Lee, J.C. 1987. Thermal stability of proteins in the presence of poly(ethylene glycols). *Biochemistry.* **26**(24), pp.7813-7819.

Lento, C., Zhu, S., Brown, K.A., Knox, R., Liuni, P. and Wilson, D.J. 2017. Time-resolved ElectroSpray Ionization Hydrogen-deuterium Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *J Vis Exp.* (122).

Leurs, U., Mistarz, U.H. and Rand, K.D. 2015. Getting to the core of protein pharmaceuticals – Comprehensive structure analysis by mass spectrometry. *European Journal of Pharmaceutics and Biopharmaceutics.* **93**, pp.95-109.

Li, C., Chu, S., Tan, S., Yin, X., Jiang, Y., Dai, X., Gong, X., Fang, X. and Tian, D. 2021. Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Frontiers in Chemistry.* **9**.

Li, K.S., Shi, L. and Gross, M.L. 2018. Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization. *Acc Chem Res.* **51**(3), pp.736-744.

Li, Y., Hewitt, D., Lentz, Y.K., Ji, J.A., Zhang, T.Y. and Zhang, K. 2014. Characterization and Stability Study of Polysorbate 20 in Therapeutic Monoclonal Antibody Formulation by Multidimensional Ultrahigh-Performance Liquid Chromatography–Charged Aerosol Detection–Mass Spectrometry. *Analytical Chemistry.* **86**(10), pp.5150-5157.

Light-wahl, K.J., Schwartz, B.L., Smith, R.D.,. 1994. Observation of the Noncovalent Quaternary Assiciations of Proteins by Elecrospray Ionization Mass Spectrometry. *Journal of the American Chemical Society.* **116**, pp.5271-5278.

Limo, M.J., Sola-Rabada, A., Boix, E., Thota, V., Westcott, Z.C., Puddu, V. and Perry, C.C. 2018. Interactions between Metal Oxides and Biomolecules: from Fundamental Understanding to Applications. *Chemical Reviews.* **118**(22), pp.11118-11193.

Liu, X., Dekker, L.J.M., Wu, S., Vanduijn, M.M., Luider, T.M., Tolić, N., Kou, Q., Dvorkin, M., Alexandrova, S., Vyatkina, K., Paša-Tolić, L. and Pevzner, P.A. 2014. De Novo Protein Sequencing by Combining Top-Down and Bottom-Up Tandem Mass Spectra. *Journal of Proteome Research.* **13**(7), pp.3241-3248.

Liu, X.R., Zhang, M.M., Zhang, B., Rempel, D.L. and Gross, M.L. 2019. Hydroxyl-Radical Reaction Pathways for the Fast Photochemical Oxidation of Proteins Platform As Revealed by (18)O Isotopic Labeling. *Anal Chem.* **91**(14), pp.9238-9245.

Liu, Y. and Bolen, D.W. 1995. The Peptide Backbone Plays a Dominant Role in Protein Stabilization by Naturally Occurring Osmolytes. *Biochemistry.* **34**(39), pp.12884-12891.

López-Ferrer, D., Petritis, K., Robinson, E.W., Hixson, K.K., Tian, Z., Lee, J.H., Lee, S.W., Tolić, N., Weitz, K.K., Belov, M.E., Smith, R.D. and Pasa-Tolić, L. 2011. Pressurized pepsin digestion in proteomics: an automatable alternative to trypsin for integrated top-down bottom-up proteomics. *Mol Cell Proteomics.* **10**(2), pM110.001479.

Lundahl, M.L.E., Fogli, S., Colavita, P.E. and Scanlan, E.M. 2021. Aggregation of protein therapeutics enhances their immunogenicity: causes and mitigation strategies. *RSC Chem Biol.* **2**(4), pp.1004-1020.

Makarov, A. 2000. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry.* **72**(6), pp.1156-1162.

Manea, M., Mezo, G., Hudecz, F. and Przybylski, M. 2007. Mass spectrometric identification of the trypsin cleavage pathway in lysyl-proline containing oligotuftsin peptides. *J Pept Sci.* **13**(4), pp.227-236.

Martos, A., Koch, W., Jiskoot, W., Wuchner, K., Winter, G., Friess, W. and Hawe, A. 2017. Trends on Analytical Characterization of Polysorbates and Their Degradation Products in Biopharmaceutical Formulations. *Journal of Pharmaceutical Sciences.* **106**(7), pp.1722-1735.

Masson, G.R., Burke, J.E., Ahn, N.G., Anand, G.S., Borchers, C., Brier, S., Bou-Assaf, G.M., Engen, J.R., Englander, S.W., Faber, J., Garlish, R., Griffin, P.R., Gross, M.L., Guttman, M., Hamuro, Y., Heck, A.J.R., Houde, D., Iacob, R.E., Jørgensen, T.J.D., Kaltashov, I.A., Klinman, J.P., Konermann, L., Man, P., Mayne, L., Pascal, B.D., Reichmann, D., Skehel, M., Snijder, J., Strutzenberg, T.S., Underbakke, E.S., Wagner, C., Wales, T.E., Walters, B.T., Weis, D.D., Wilson, D.J., Wintrode, P.L., Zhang, Z., Zheng, J., Schriemer, D.C. and Rand, K.D. 2019. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nature Methods.* **16**(7), pp.595-602.

Mehmood, S., Allison, T.M. and Robinson, C.V. 2015. Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annual Review of Physical Chemistry.* **66**(1), pp.453-474.

Minton, A.P. 2005. Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: macromolecular crowding and protein stability revisited. *Biophys J.* **88**(2), pp.971-985.

Mishra, R., Seckler, R. and Bhat, R. 2005. Efficient refolding of aggregation-prone citrate synthase by polyol osmolytes: how well are protein folding and stability aspects coupled? *J Biol Chem.* **280**(16), pp.15553-15560.

Mony, C., Kaur, P., Rookes, J.E., Callahan, D.L., Eswaran, S.V., Yang, W. and Manna, P.K. 2022. Nanomaterials for enhancing photosynthesis: interaction with plant photosystems and scope of nanobionics in agriculture. *Environmental Science: Nano.* **9**(10), pp.3659-3683.

Murray, K.K. 2022. Resolution and Resolving Power in Mass Spectrometry. *Journal of the American Society for Mass Spectrometry.* **33**(12), pp.2342-2347.

Mustafi, D., Smith, C.M., Makinen, M.W. and Lee, R.C. 2008. Multi-block poloxamer surfactants suppress aggregation of denatured proteins. *Biochimica et Biophysica Acta (BBA) - General Subjects.* **1780**(1), pp.7-15.

Niu, B., Zhang, H., Giblin, D., Rempel, D.L. and Gross, M.L. 2015. Dosimetry determines the initial OH radical concentration in fast photochemical oxidation of proteins (FPOP). *J Am Soc Mass Spectrom.* **26**(5), pp.843-846.

Niu, B., Zhang, H., Giblin, D., Rempel, D.L. and Gross, M.L. 2015. Dosimetry Determines the Initial OH Radical Concentration in Fast Photochemical Oxidation of Proteins (FPOP). *Journal of the American Society for Mass Spectrometry.* **26**(5), pp.843-846.

Ohtake, S., Kita, Y. and Arakawa, T. 2011. Interactions of formulation excipients with proteins in solution and in the dried state. *Advanced Drug Delivery Reviews.* **63**(13), pp.1053-1073.

Ozohanics, O. and Ambrus, A. 2020. Hydrogen-Deuterium Exchange Mass Spectrometry: A Novel Structural Biology Approach to Structure, Dynamics and Interactions of Proteins and Their Complexes. *Life.* [Online]. **10**(11).

Patist, A., Bhagwat, S.S., Penfield, K.W., Aikens, P. and Shah, D.O. 2000. On the measurement of critical micelle concentrations of pure and technical-grade nonionic surfactants. *Journal of Surfactants and Detergents.* **3**(1), pp.53-58.

Paul, W., Steinwedel, H.,. 1953. Quadrupole mass filter. *Zeitschrift für Naturforschung A.* **8**, p448.

Perinelli, D.R., Cespi, M., Lorusso, N., Palmieri, G.F., Bonacucina, G. and Blasi, P. 2020. Surfactant Self-Assembling and Critical Micelle Concentration: One Approach Fits All? *Langmuir.* **36**(21), pp.5745-5753.

Pivovarova, A.V., Chebotareva, N.A., Chernik, I.S., Gusev, N.B. and Levitsky, D.I. 2007. Small heat shock protein Hsp27 prevents heat-induced aggregation of F-actin by forming soluble complexes with denatured actin. *Febs j.* **274**(22), pp.5937-5948.

Rand, K.D., Zehl, M., Jensen, O.N. and Jorgensen, T.J. 2009. Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal Chem.* **81**(14), pp.5577-5584.

Randolph, T.W. and Jones, L.S. 2002. Surfactant-Protein Interactions. In: Carpenter, J.F. and Manning, M.C. eds. *Rational Design of Stable Protein Formulations: Theory and Practice.* Boston, MA: Springer US, pp.159-175.

Rawat, S., Raman Suri, C. and Sahoo, D.K. 2010. Molecular mechanism of polyethylene glycol mediated stabilization of protein. *Biochemical and Biophysical Research Communications.* **392**(4), pp.561-566.

Richardson, K., Langridge, D., Dixit, S.M. and Ruotolo, B.T. 2021. An Improved Calibration Approach for Traveling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. *Analytical Chemistry.* **93**(7), pp.3542-3550.

Rinas, A., Mali, V.S., Espino, J.A. and Jones, L.M. 2016. Development of a Microflow System for In-Cell Footprinting Coupled with Mass Spectrometry. *Anal Chem.* **88**(20), pp.10052-10058.

Roberts, C.J. 2014. Therapeutic protein aggregation: mechanisms, design, and control. *Trends in Biotechnology.* **32**(7), pp.372-380.

Santoro, M.M., Liu, Y., Khan, S.M.A., Hou, L.X. and Bolen, D.W. 1992. Increased thermal stability of proteins in the presence of naturally occurring osmolytes. *Biochemistry.* **31**(23), pp.5278-5283.

Sasahara, K., McPhie, P. and Minton, A.P. 2003. Effect of Dextran on Protein Stability and Conformation Attributed to Macromolecular Crowding. *Journal of Molecular Biology.* **326**(4), pp.1227-1237.

Savaryn, J.P., Toby, T.K. and Kelleher, N.L. 2016. A researcher's guide to mass spectrometry-based proteomics. *Proteomics.* **16**(18), pp.2435-2443.

Schellman, J.A. 1997. Temperature, stability, and the hydrophobic interaction. *Biophysical Journal.* **73**(6), pp.2960-2964.

Schellman, J.A. 2003. Protein Stability in Mixed Solvents: A Balance of Contact Interaction and Excluded Volume. *Biophysical Journal.* **85**(1), pp.108-125.

Schüttelkopf, A.W. and van Aalten, D.M. 2004. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr.* **60**(Pt 8), pp.1355-1363.

Schwartzberg, L.S. and Navari, R.M. 2018. Safety of Polysorbate 80 in the Oncology Setting. *Adv Ther.* **35**(6), pp.754-767.

Scrosati, P.M., Yin, V. and Konermann, L. 2021. Hydrogen/Deuterium Exchange Measurements May Provide an Incomplete View of Protein Dynamics: a Case Study on Cytochrome c. *Anal Chem.* **93**(42), pp.14121-14129.

Sharp, J.S. and Tomer, K.B. 2006. Effects of Anion Proximity in Peptide Primary Sequence on the Rate and Mechanism of Leucine Oxidation. *Analytical Chemistry.* **78**(14), pp.4885-4893.

Simonneaux, G. and Bondon, A. 2005. Mechanism of Electron Transfer in Heme Proteins and Models: The NMR Approach. *Chemical Reviews.* **105**(6), pp.2627-2646.

Singh, S.M., Bandi, S., Jones, D.N.M. and Mallela, K.M.G. 2017. Effect of Polysorbate 20 and Polysorbate 80 on the Higher-Order Structure of a Monoclonal Antibody and Its Fab and Fc Fragments Probed Using 2D Nuclear Magnetic Resonance Spectroscopy. *J Pharm Sci.* **106**(12), pp.3486-3498.

Singh, S.M., Bandi, S., Jones, D.N.M. and Mallela, K.M.G. 2017. Effect of Polysorbate 20 and Polysorbate 80 on the Higher-Order Structure of a Monoclonal Antibody and Its Fab and Fc Fragments Probed Using 2D Nuclear Magnetic Resonance Spectroscopy. *Journal of Pharmaceutical Sciences.* **106**(12), pp.3486-3498.

Spiteri, V.A., Goodall, M., Doutch, J., Rambo, R.P., Gor, J. and Perkins, S.J. 2021. Solution structures of human myeloma IgG3 antibody reveal extended Fab and Fc regions relative to the other IgG subclasses. *Journal of Biological Chemistry.* **297**(3).

Stikane, A. 2020. Development of bio-mimetic nano-compartments for solar energy capture.

Stikane, A., Hwang, E.T., Ainsworth, Emma V., Piper, S.E.H., Critchley, K., Butt, J.N., Reisner, E. and Jeuken, L.J.C. 2019. Towards compartmentalized photocatalysis: multihaem proteins as transmembrane molecular electron conduits. *Faraday Discussions.* **215**(0), pp.26-38.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. and 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**(8), pp.151-153.

Tholey, A. and Becker, A. 2017. Top-down proteomics for the analysis of proteolytic events - Methods, applications and perspectives. *Biochim Biophys Acta Mol Cell Res.* **1864**(11 Pt B), pp.2191-2199.

Timasheff, S.N. 1993. The Control of Protein Stability and Association by Weak Interactions with Water: How Do Solvents Affect These Processes? *Annual Review of Biophysics and Biomolecular Structure.* **22**(1), pp.67-97.

Timasheff, S.N. 1998. Control of Protein Stability and Reactions by Weakly Interacting Cosolvents: The Simplicity of the Complicated. In: Di Cera, E. ed. *Advances in Protein Chemistry.* Academic Press, pp.355-432.

Timasheff, S.N. 2002. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proceedings of the National Academy of Sciences.* **99**(15), pp.9721-9726.

Traube, J. 1909. The Theory of Attraction Pressure. *The Journal of Physical Chemistry.* **14**(5), pp.471-475.

Vagenende, V., Yap, M.G.S. and Trout, B.L. 2009. Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. *Biochemistry.* **48**(46), pp.11084-11096.

Vahidi, S. and Konermann, L. 2016. Probing the Time Scale of FPOP (Fast Photochemical Oxidation of Proteins): Radical Reactions Extend Over Tens of Milliseconds. *J Am Soc Mass Spectrom.* **27**(7), pp.1156-1164.

Wales, T.E. and Engen, J.R. 2006. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev.* **25**(1), pp.158-170.

Walmsley, S.J., Rudnick, P.A., Liang, Y., Dong, Q., Stein, S.E. and Nesvizhskii, A.I. 2013. Comprehensive Analysis of Protein Digestion Using Six Trypsins Reveals the Origin of Trypsin As a Significant Source of Variability in Proteomics. *Journal of Proteome Research.* **12**(12), pp.5666-5680.

Wang, W. 1999. Instability, stabilization, and formulation of liquid protein pharmaceuticals. *International Journal of Pharmaceutics.* **185**(2), pp.129-188.

Wang, W., Wang, Y.J. and Wang, D.Q. 2008. Dual effects of Tween 80 on protein stability. *International Journal of Pharmaceutics.* **347**(1), pp.31-38.

Webb, S.D., Cleland, J.L., Carpenter, J.F. and Randolph, T.W. 2002. A new mechanism for decreasing aggregation of recombinant human interferon‐γ by a surfactant: Slowed dissolution of lyophilized formulations in a solution containing 0.03% polysorbate 20. *Journal of Pharmaceutical Sciences.* **91**(2), pp.543-558.

Wells, J.M. and McLuckey, S.A. 2005. Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol.* **402**, pp.148-185.

White, G.F., Shi, Z., Shi, L., Wang, Z., Dohnalkova, A.C., Marshall, M.J., Fredrickson, J.K., Zachara, J.M., Butt, J.N., Richardson, D.J. and Clarke, T.A. 2013. Rapid electron exchange between surface-exposed bacterial cytochromes and Fe(III) minerals. *Proceedings of the National Academy of Sciences.* **110**(16), pp.6346-6351.

Wiangnon, K. and Cramer, R. 2016. Liquid MALDI MS Analysis of Complex Peptide and Proteome Samples. *J Proteome Res.* **15**(9), pp.2998-3008.

Willis, L.F., Kumar, A., Dobson, J., Bond, N.J., Lowe, D., Turner, R., Radford, S.E., Kapur, N. and Brockwell, D.J. 2018. Using extensional flow to reveal diverse aggregation landscapes for three IgG1 molecules. *Biotechnol Bioeng.* **115**(5), pp.1216-1225.

Winter, M., Tholey, A., Kristen, A. and Rocken, C. 2017. MALDI Mass Spectrometry Imaging: A Novel Tool for the Identification and Classification of Amyloidosis. *Proteomics.* **17**(22).

Wolff, M.M. and Stephens, W.E. 1953. A Pulsed Mass Spectrometer with Time Dispersion. *Review of Scientific Instruments.* **24**(8), pp.616-617.

Wollenberg, D.T.W., Pengelley, S., Mouritsen, J.C., Suckau, D., Jørgensen, C.I. and Jørgensen, T.J.D. 2020. Avoiding H/D Scrambling with Minimal Ion Transmission Loss for HDX-MS/MS-ETD Analysis on a High-Resolution Q-TOF Mass Spectrometer. *Analytical Chemistry.* **92**(11), pp.7453-7461.

Wolynes, P.G., Onuchic, J.N. and Thirumalai, D. 1995. Navigating the folding routes. *Science.* **267**(5204), pp.1619-1620.

Wu, J., Zhao, C., Lin, W., Hu, R., Wang, Q., Chen, H., Li, L., Chen, S. and Zheng, J. 2014. Binding characteristics between polyethylene glycol (PEG) and proteins in aqueous solution. *Journal of Materials Chemistry B.* **2**(20), pp.2983-2992.

Xie, B., Sood, A., Woods, R.J. and Sharp, J.S. 2017. Quantitative Protein Topography Measurements by High Resolution Hydroxyl Radical Protein Footprinting Enable Accurate Molecular Model Selection. *Sci Rep.* **7**(1), p4552.

Xie, B., Sood, A., Woods, R.J. and Sharp, J.S. 2017. Quantitative Protein Topography Measurements by High Resolution Hydroxyl Radical Protein Footprinting Enable Accurate Molecular Model Selection. *Scientific Reports.* **7**(1), p4552.

Xu, G. and Chance, M.R. 2005. Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Anal Chem.* **77**(14), pp.4549-4555.

Xu, G. and Chance, M.R. 2007. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem Rev.* **107**(8), pp.3514-3543.

Xu, G. and Chance, M.R. 2007. Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chemical Reviews.* **107**(8), pp.3514-3543.

Yamashita, M. and Fenn, J.B. 1984. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry.* **88**(20), pp.4451-4459.

Yancey, P.H. 2005. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *Journal of Experimental Biology.* **208**(15), pp.2819-2830.

Yoshizawa, S., Arakawa, T. and Shiraki, K. 2017. Thermal aggregation of human immunoglobulin G in arginine solutions: Contrasting effects of stabilizers and destabilizers. *International Journal of Biological Macromolecules.* **104**, pp.650-655.

Yost, R.A. and Enke, C.G. 1979. Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Analytical Chemistry.* **51**(12), pp.1251-1264.

Zeleny, J. 1917. Instability of Electrified Liquid Surfaces. *Physical Review.* **10**(1), pp.1-6.

Zhang, H.-M., McLoughlin, S.M., Frausto, S.D., Tang, H., Emmett, M.R. and Marshall, A.G. 2010. Simultaneous Reduction and Digestion of Proteins with Disulfide Bonds for Hydrogen/Deuterium Exchange Monitored by Mass Spectrometry. *Analytical Chemistry.* **82**(4), pp.1450-1454.

Zhang, H.M., McLoughlin, S.M., Frausto, S.D., Tang, H., Emmett, M.R. and Marshall, A.G. 2010. Simultaneous reduction and digestion of proteins with disulfide bonds for hydrogen/deuterium exchange monitored by mass spectrometry. *Anal Chem.* **82**(4), pp.1450-1454.

Zhang, Z. and Smith, D.L. 1993. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**(4), pp.522-531.

8. Supplemental information

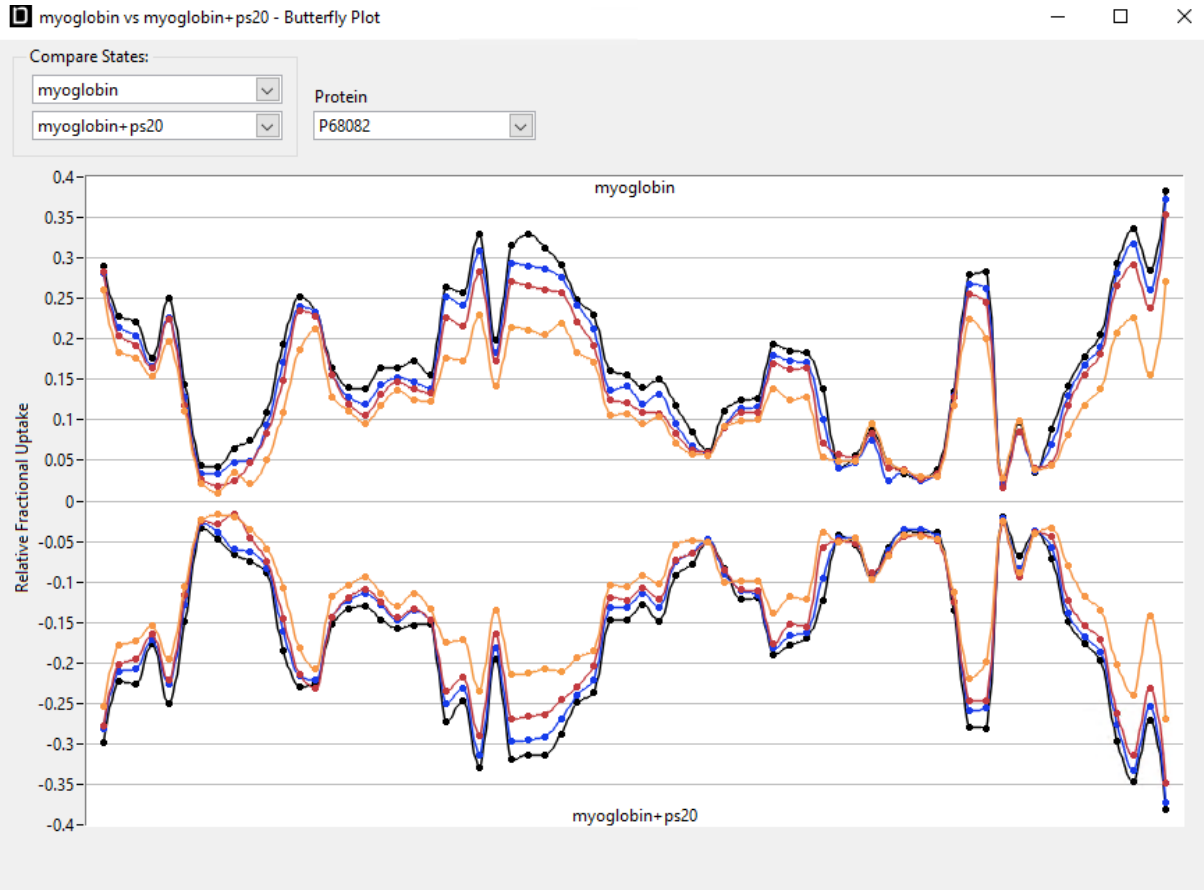


Figure 8.1 Butterfly plot for myoglobin no polysorbate vs PS20 fresh addition. Yellow lines are for the 0.5 minute time point, red for the 2 time point, blue for the 5 minute time point and black the 10 minute point. This data is discussed in figure 4.2.

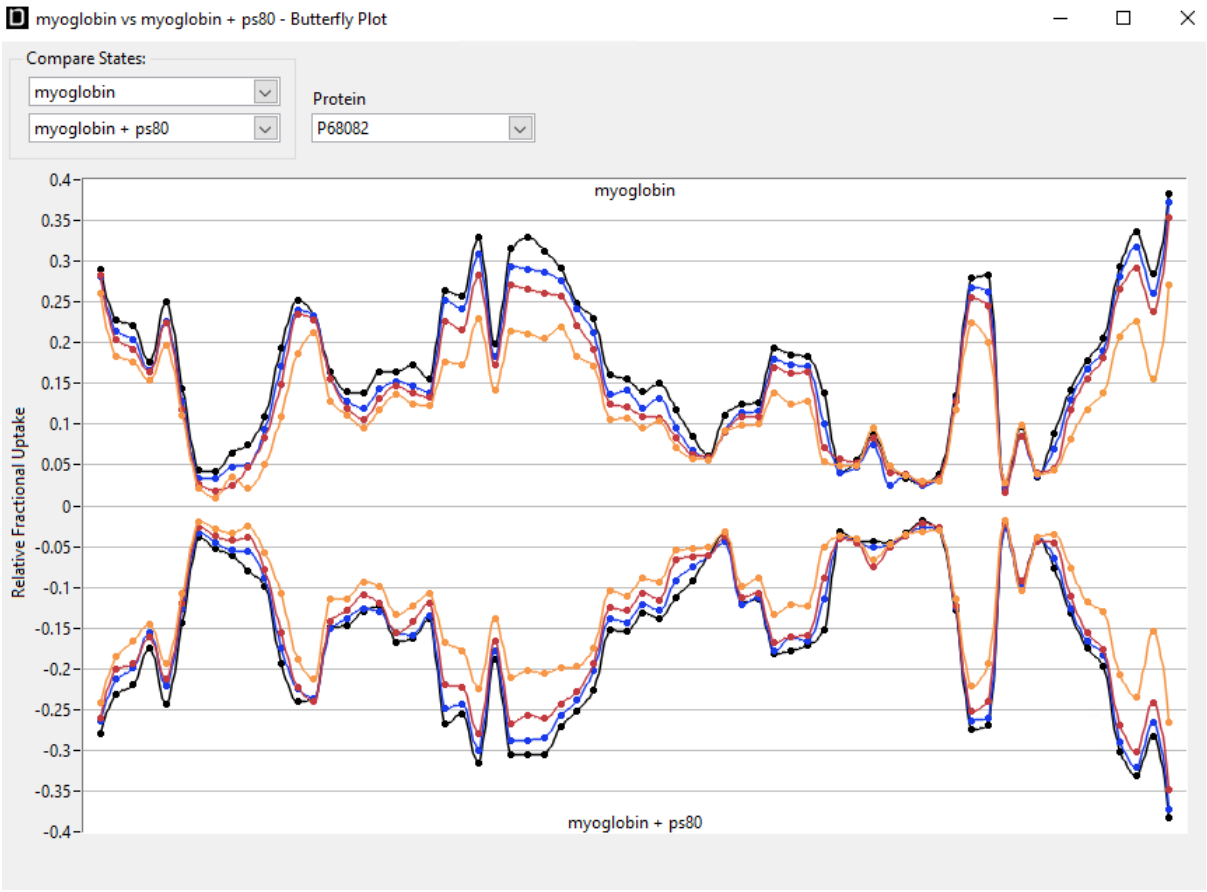


Figure 8.2 Butterfly plot for myoglobin no polysorbate vs PS80 fresh addition of polysorbate 30 minutes before MS. Yellow lines are for the 0.5 minute time point, red for the 2 time point, blue for the 5 minute time point and black the 10 minute point. This data is discussed in figure 4.2.

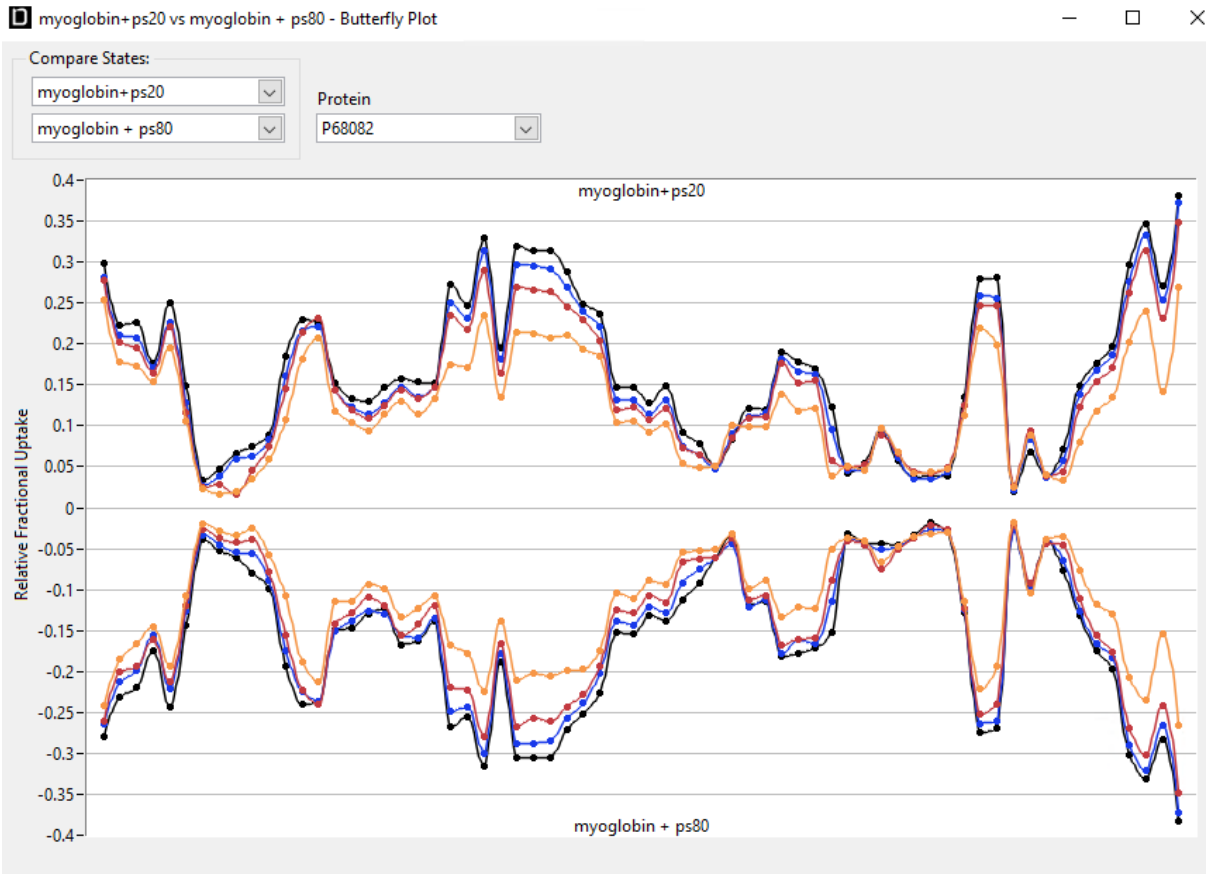


Figure 8.3 Butterfly plot for myoglobin no PS20 vs PS80 fresh addition of polysorbate 30 minutes before MS. Yellow lines are for the 0.5 minute time point, red for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point.

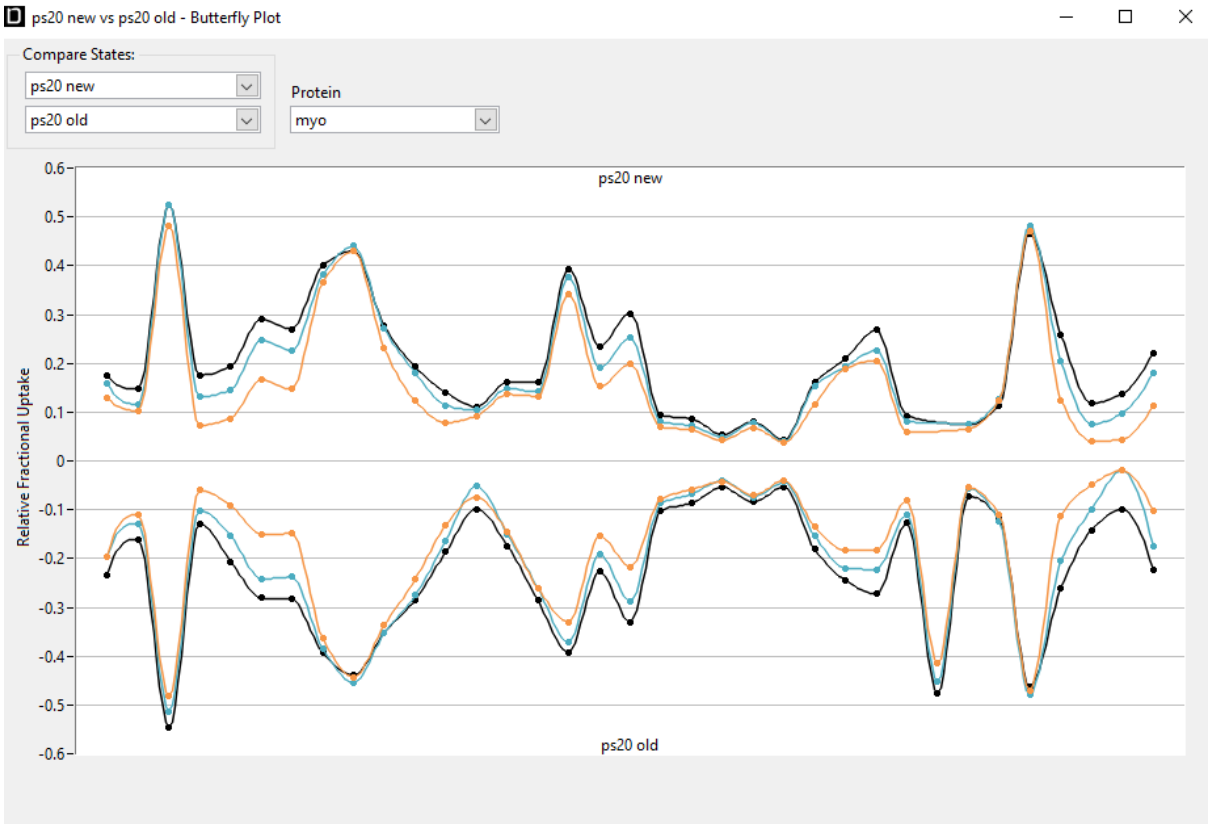


Figure 8.4 Butterfly plot for PS20 fresh vs PS20 one month aged sample. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. This data is discussed in figure 4.3 and 4.4

No butterfly plot is provided for the PS80 sample due to the low sequence coverage (53%).

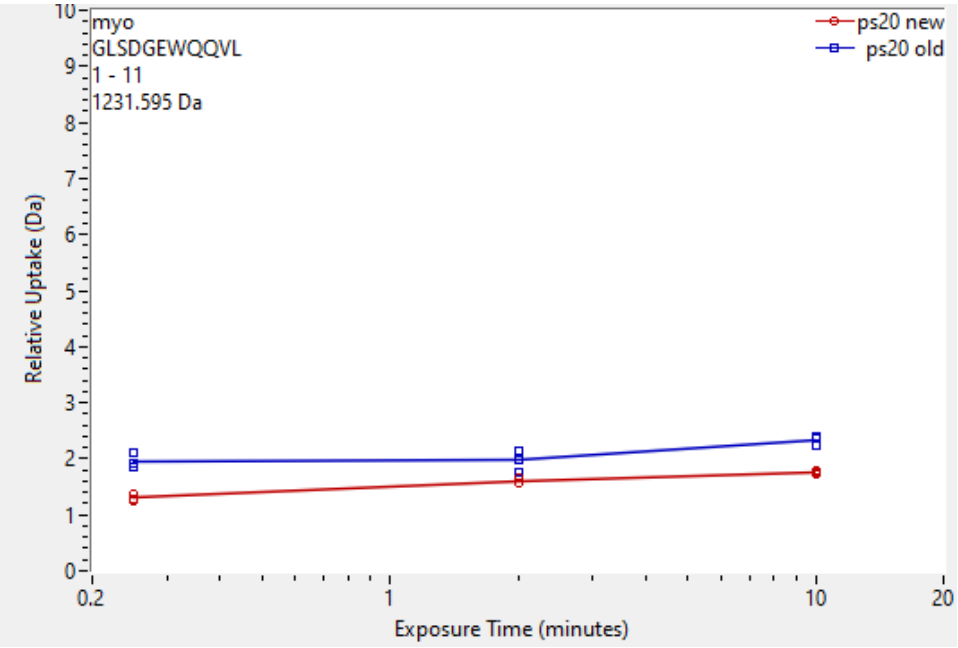


Figure 8.5 peptide GLSDGEWQQVL for PS20 fresh vs PS20 with 1 month incubation showing deprotection for the 1month incubation. This data is shown in figures 4.3 and 4.4

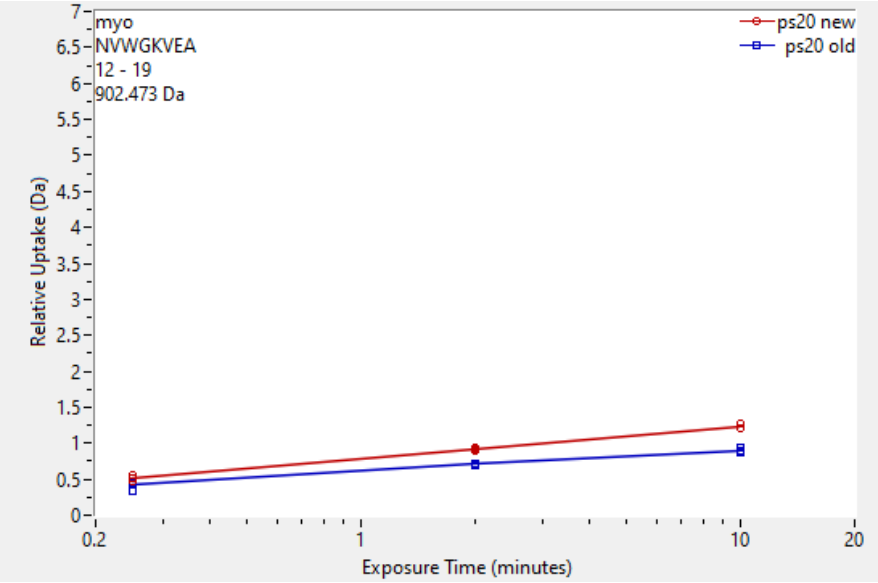


Figure 8.6 peptide NVWGKVEA for no PS20 Fresh vs PS20 1 month incubation showing protection for the 1 month incubation. This data is shown in figures 4.3 and 4.4

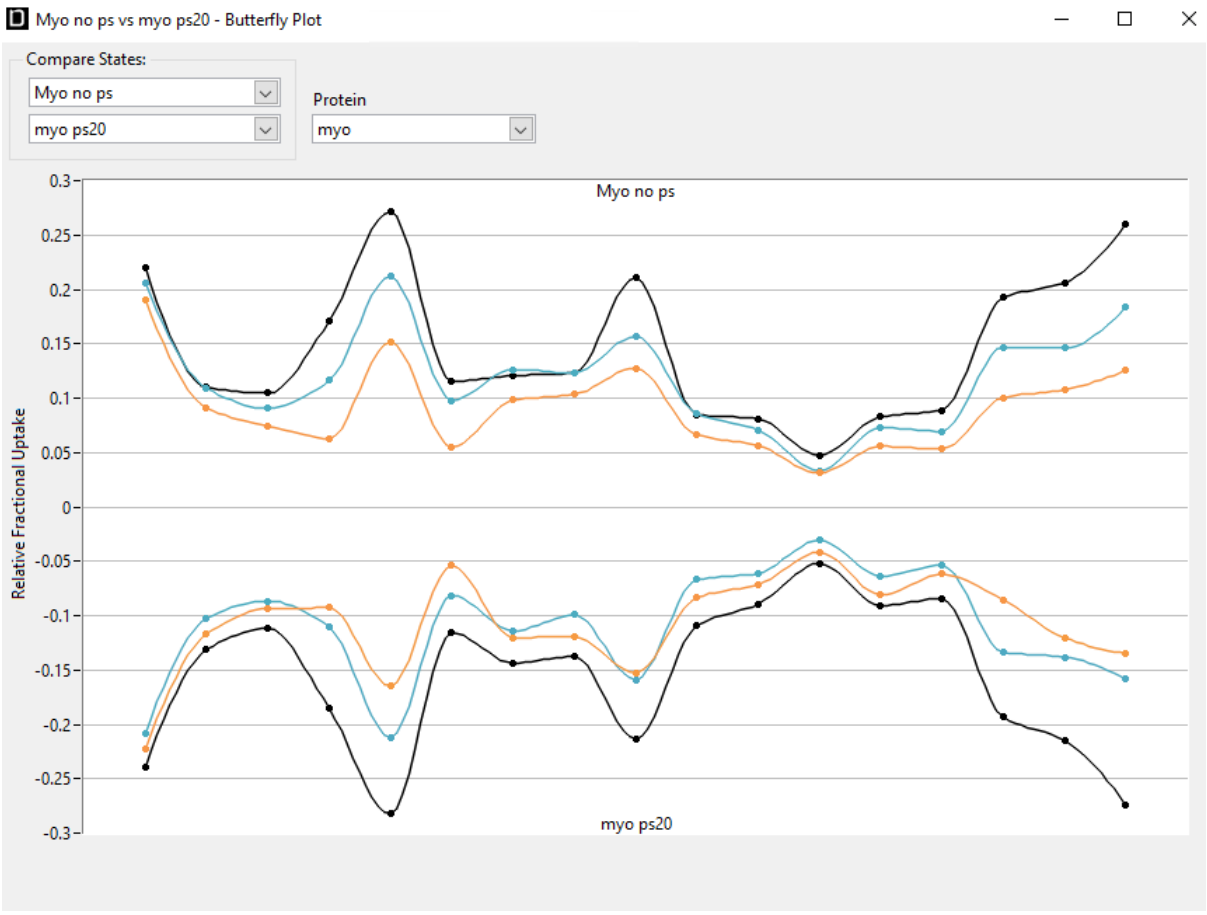


Figure 8.7 Butterfly plot of myoglobin incubated with no polysorbate vs myoglobin incubated with polysorbate 20 for 1 month at 25°C. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed. This data is presented in figure 4.5.

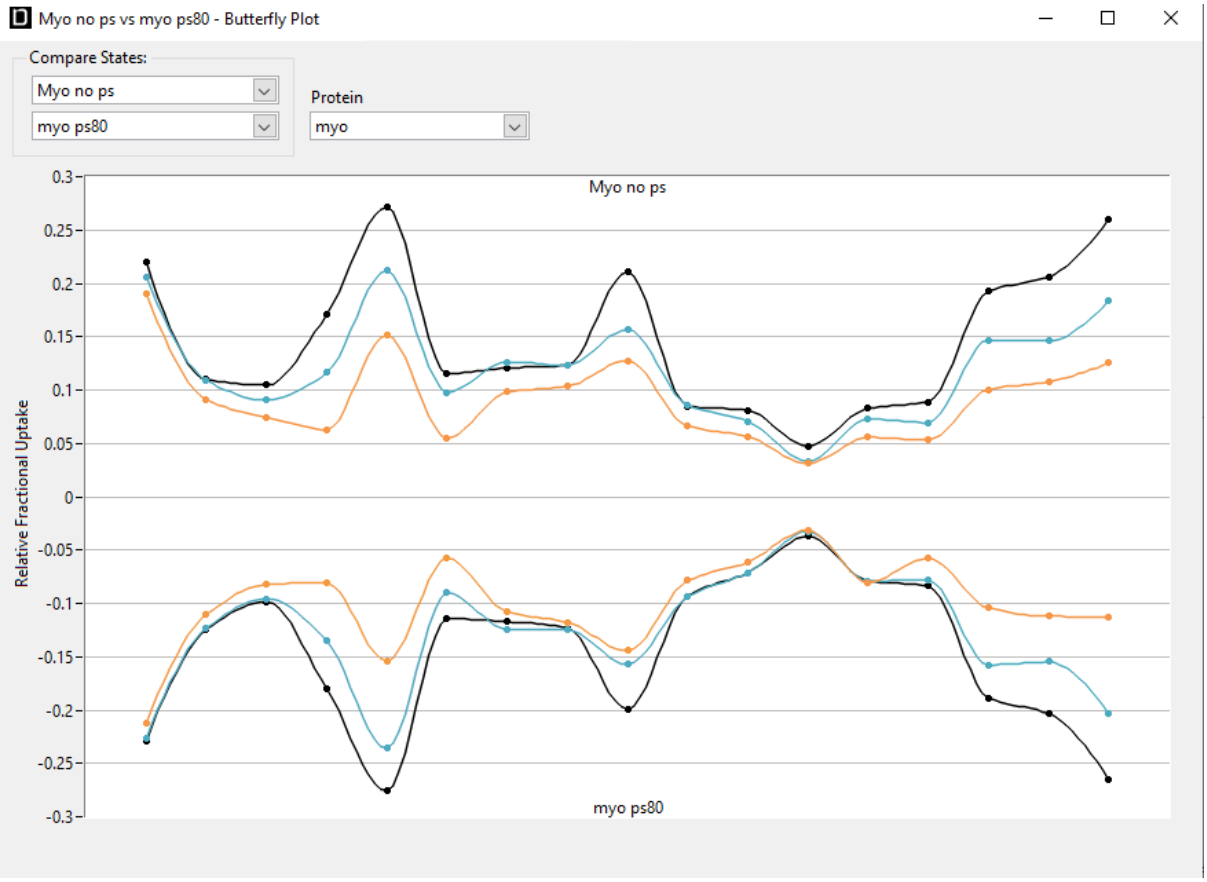


Figure 8.8: Butterfly of plot myoglobin incubated with no polysorbate vs myoglobin incubated with polysorbate 80 for 1 month at 25°C. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed. This data is presented in figure 4.5.

Figure 8.9: Butterfly plot of myoglobin incubated with polysorbate 20 vs myoglobin incubated with polysorbate 80 for 1 month at 25°C. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed. This data is presented in figure 4.5.

Figure 8.10: Butterfly plot for no PS vs PS20 fresh addition for STT light chain. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed at the 10 minute time point.

Figure 8.11: Butterfly plot for no PS vs PS20 fresh addition for STT heavy chain. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed.

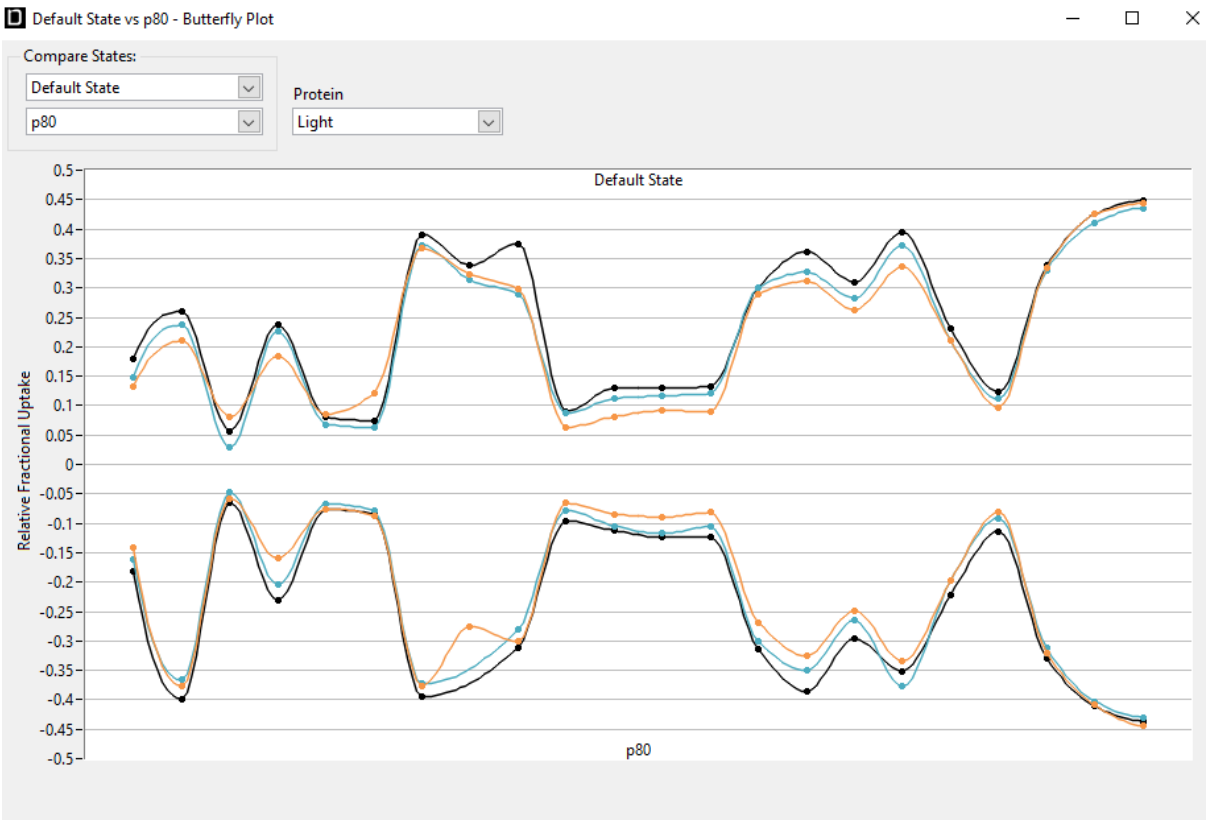


Figure 8.12**:** Butterfly plot for no PS vs PS80 fresh addition for STT light chain *Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed.*

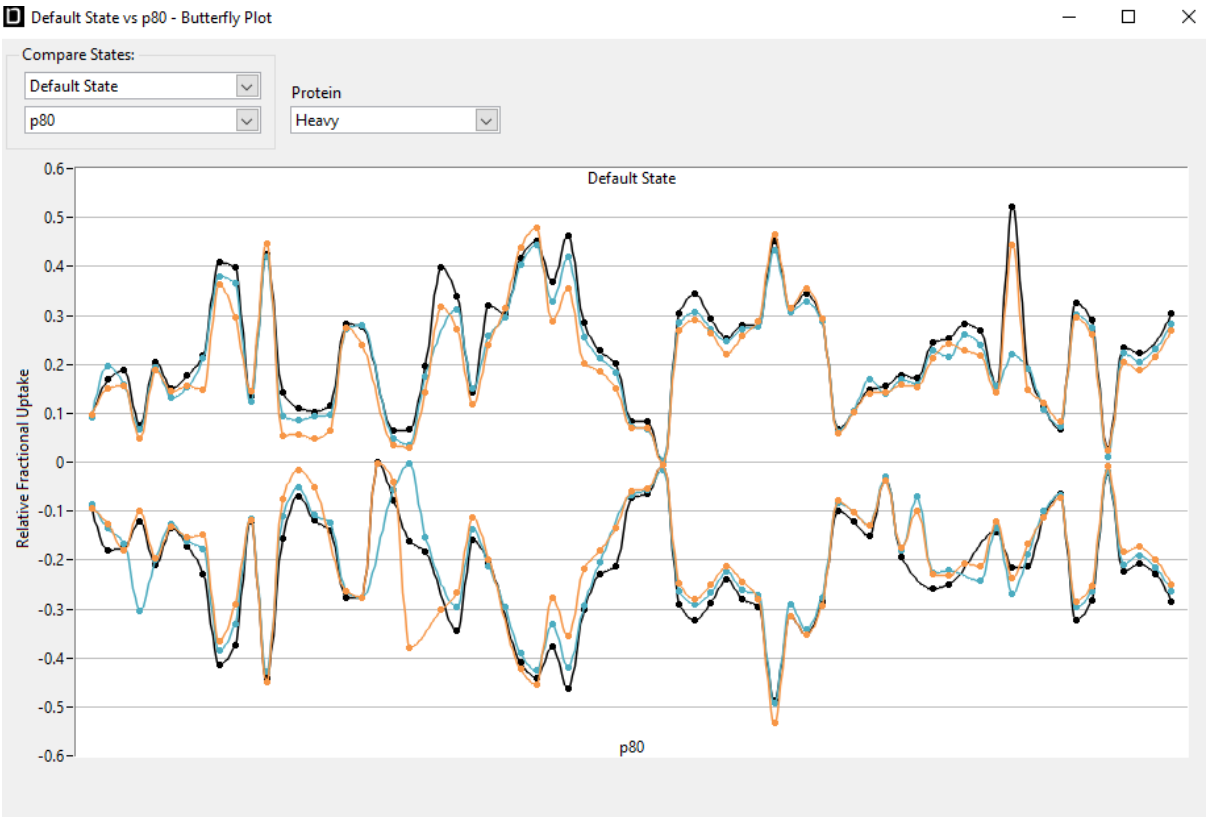


Figure 8.13: Butterfly plot for no PS vs PS80 fresh addition for STT heavy chain Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed.

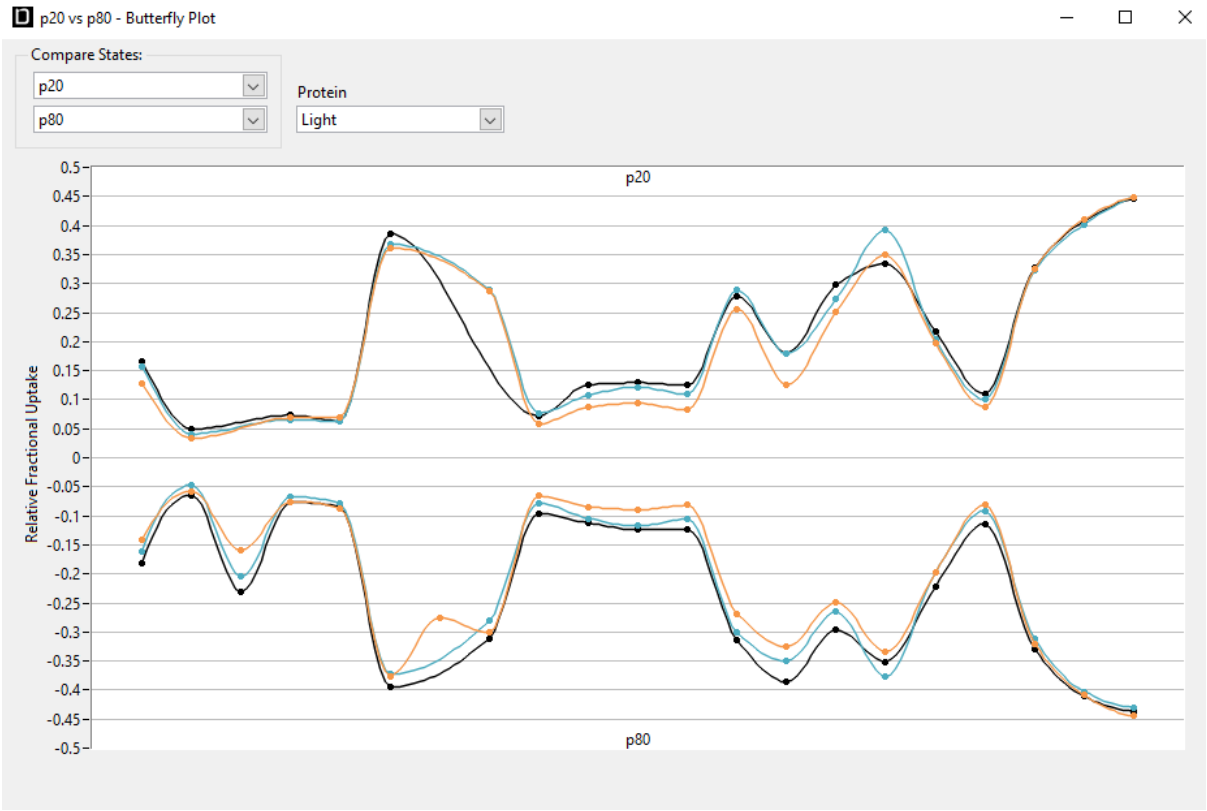


Figure 8.14: Butterfly plot for PS20 vs PS80 fresh addition STT light chain. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed.

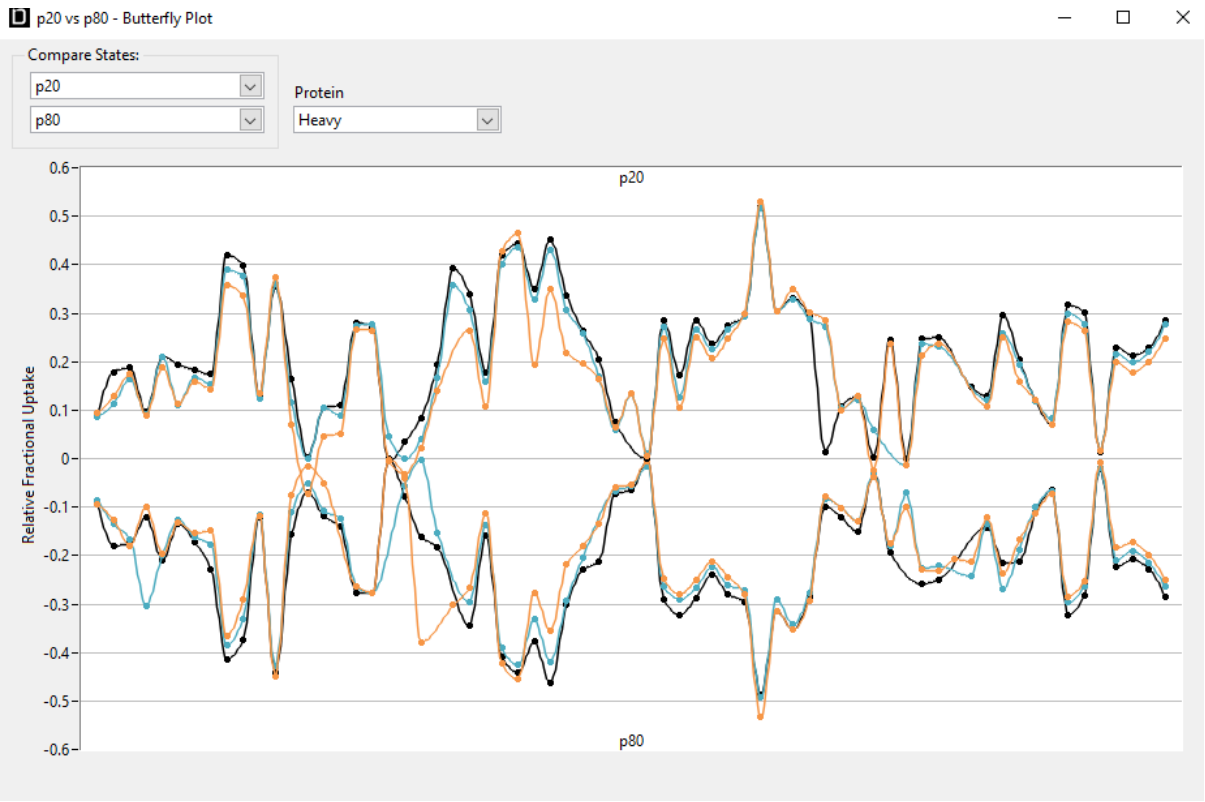


Figure 8.15: Butterfly plot for PS20 vs PS80 fresh addition STT heavy chain Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed.

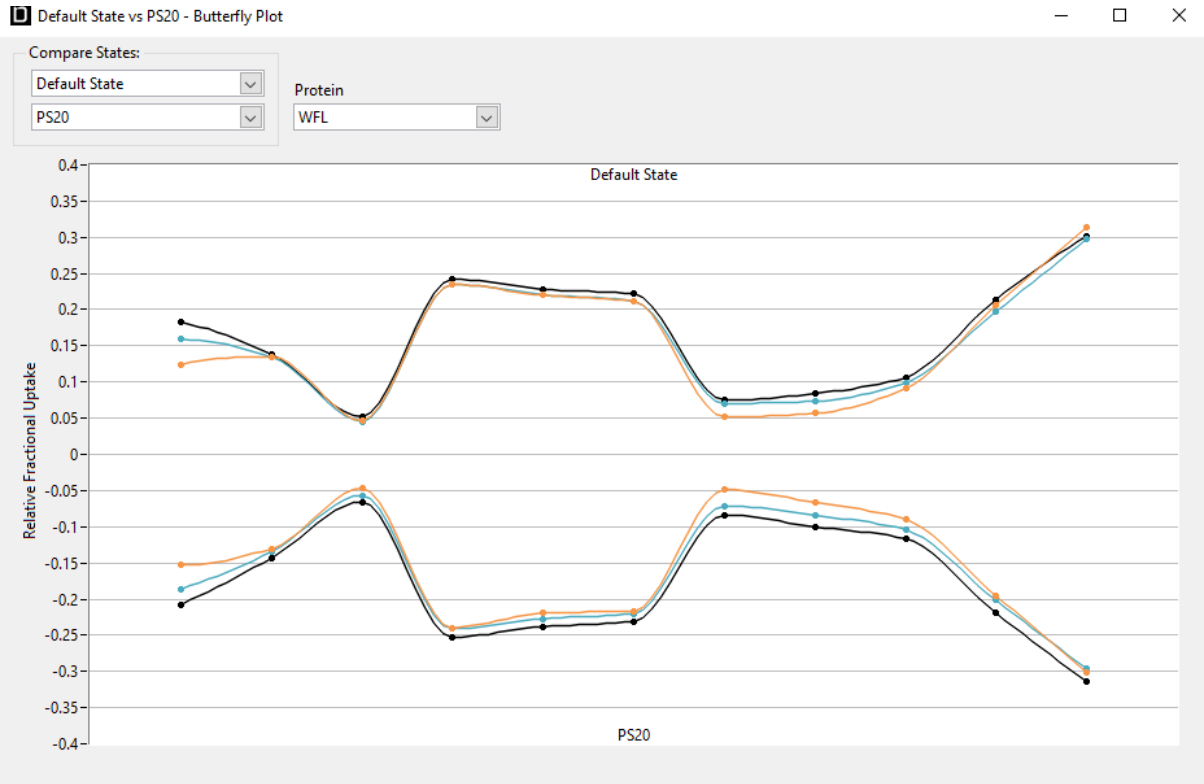


Figure 8.16: Butterfly plot for no PS vs PS20 fresh addition for WFL light chain. *Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. No significant difference was observed.*

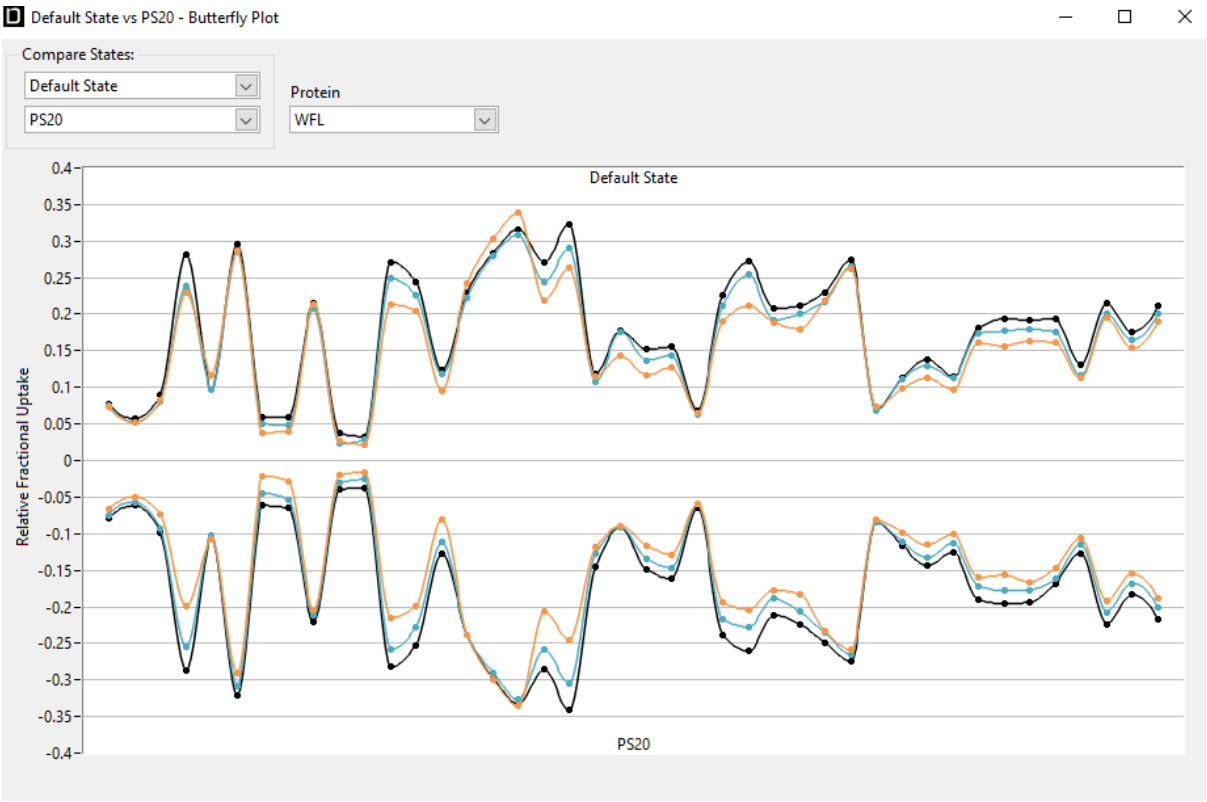


Figure 8.17: Butterfly plot for no PS vs PS20 fresh addition for WFL heavy chain. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. This data is discussed in figures 4.6 and 4.7.

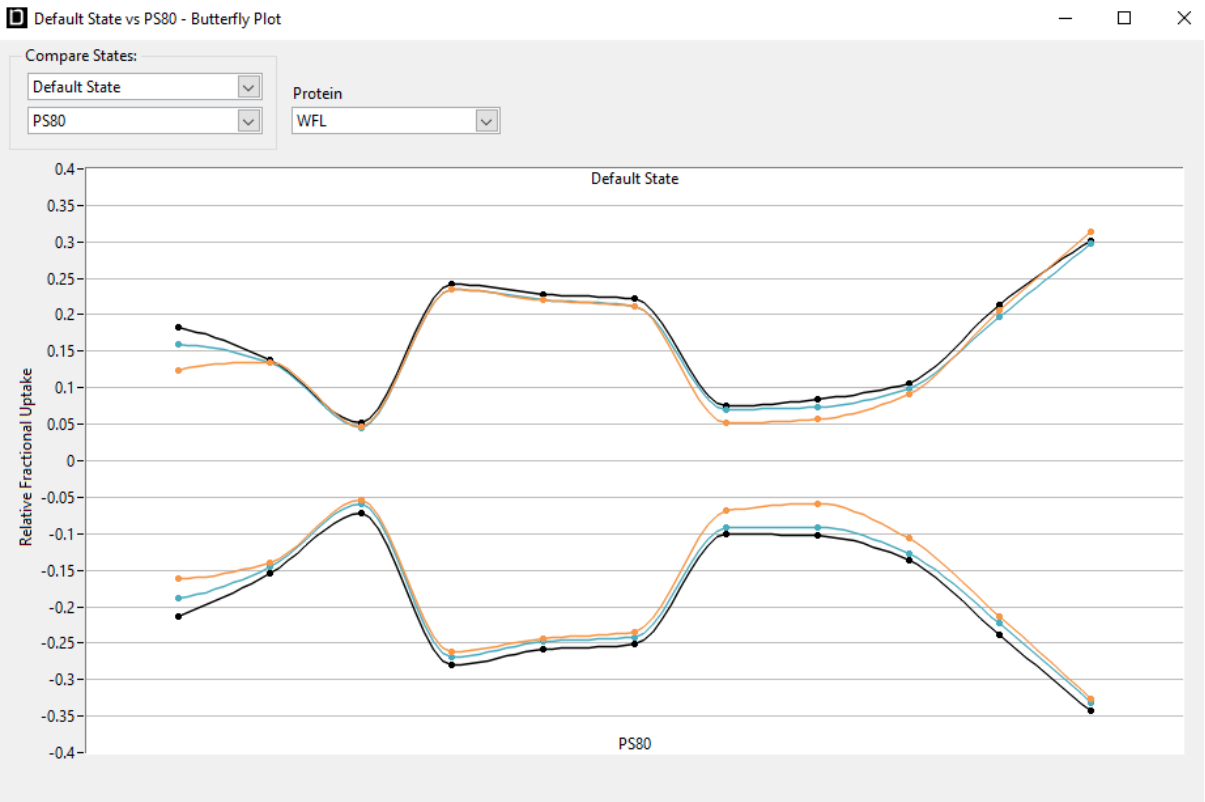


Figure 8.18: Butterfly plot for no PS vs PS80 fresh addition for WFL light *Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. This data is discussed in figures 4.6 and 4.7.*

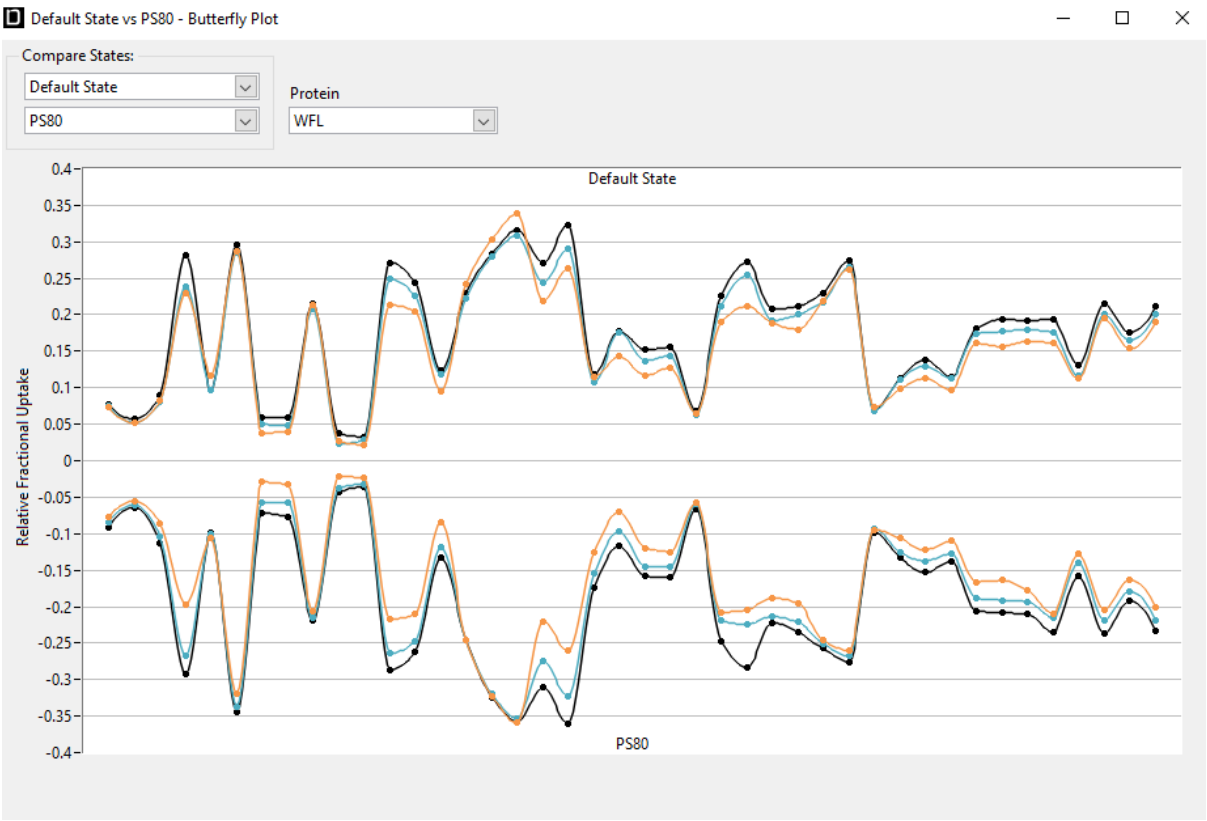


Figure 8.19 Butterfly plot for no PS vs PS80 fresh addition for WFL heavy chain. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. This data is discussed in figures 4.6 and 4.7.

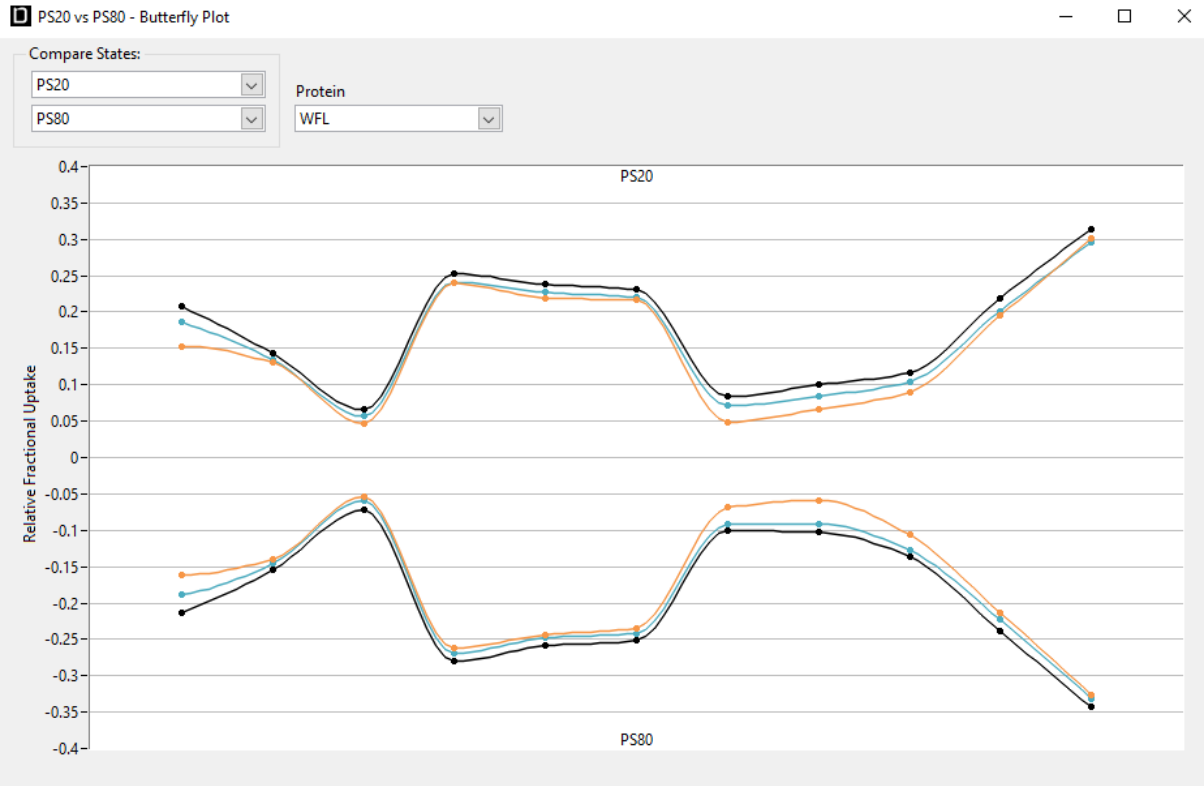


Figure 8.20: Butterfly plot for PS20 vs PS80 fresh addition for WFL light chain. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. This data is discussed in figures 4.6 and 4.7.

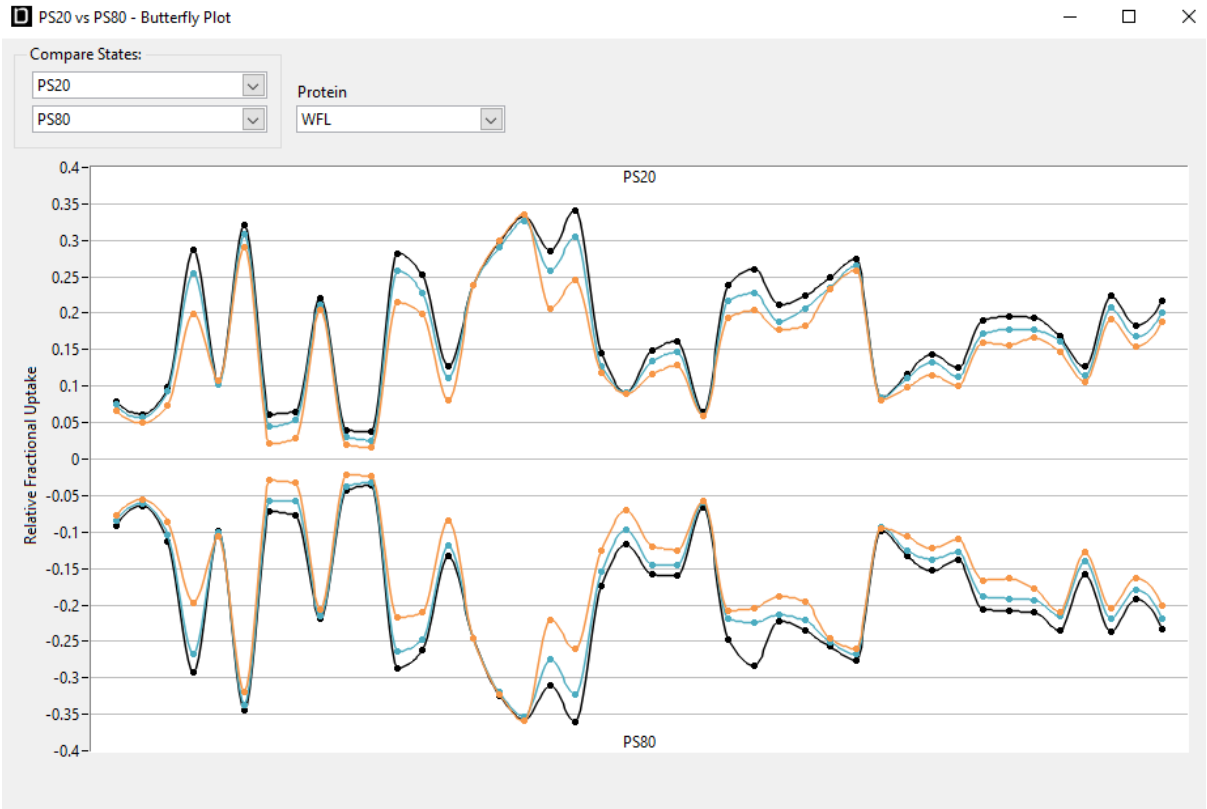


Figure 8.21: Butterfly plot for PS20 vs PS80 fresh addition for WFL heavy chain. Yellow lines are for the 2 minute time point, blue for the 5 minute time point and black the 10 minute point. This data is discussed in figures 4.6 and 4.7.

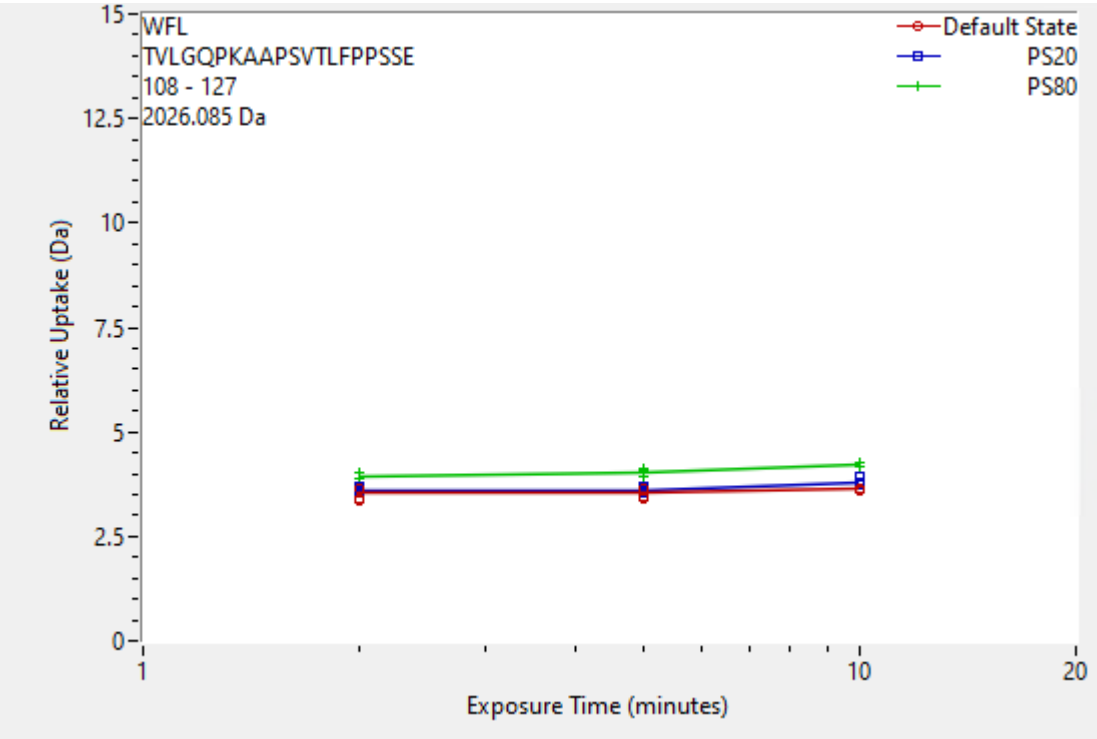


Figure 8.22: Peptide TVLGQPKAAPSVTLFPPSE showing deprotection of the PS80 peptide in comparison to the default state and PS20. This data is discussed in figures 4.6 and 4.7.

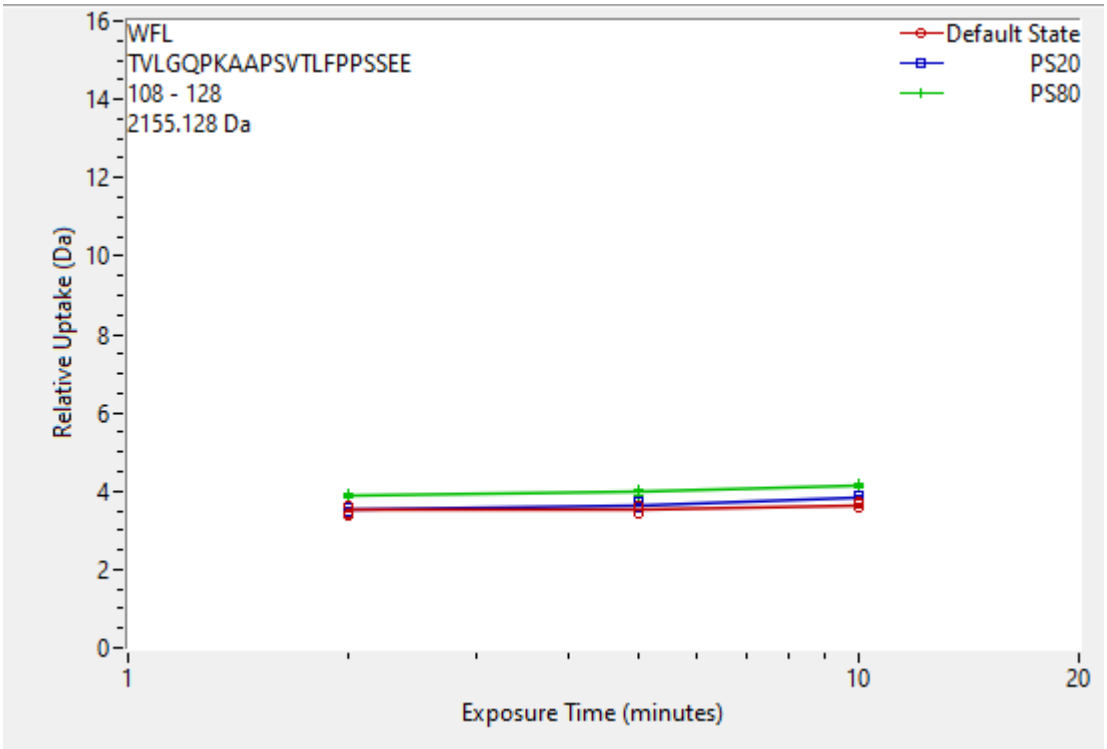


Figure 8.23: Peptide TVLGQPKAAPSVTLFPPSEE showing deprotection of the PS80 peptide in comparison to the default state and PS20 This data is discussed in figures 4.6 and 4.7.

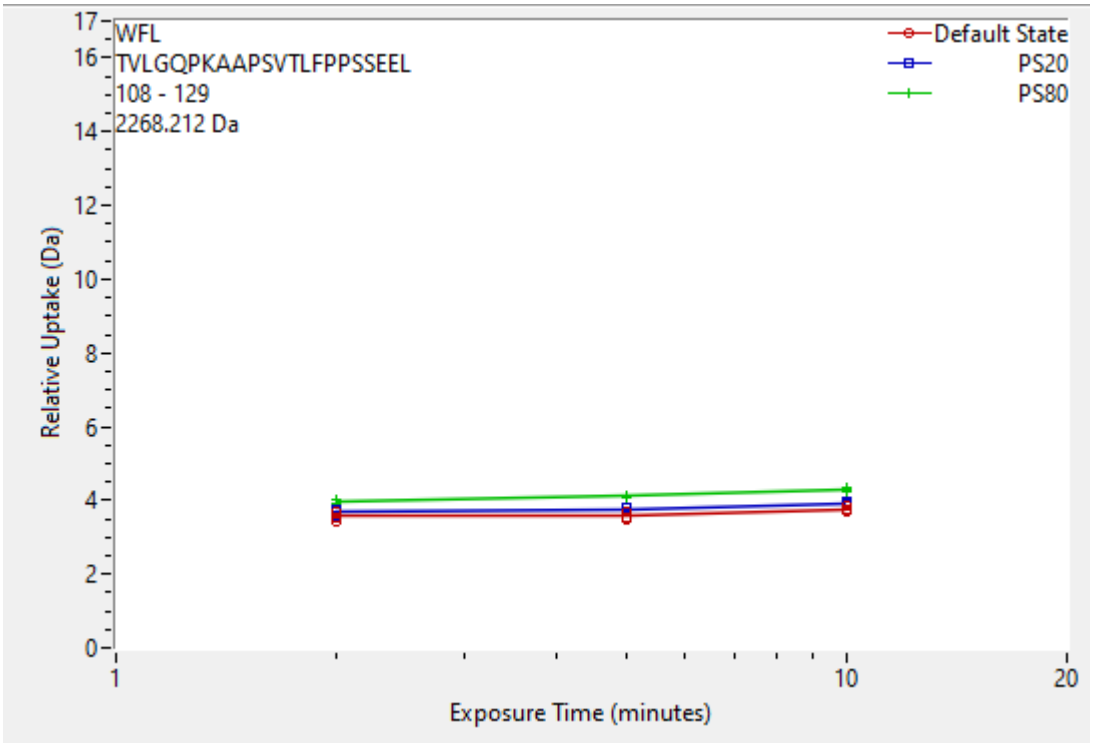


Figure 8.24: Peptide TVLGQPKAAPSVTLFPPSEL showing deprotection of the PS80 peptide in comparison to the default state and PS20. This data is discussed in figures 4.6 and 4.7.

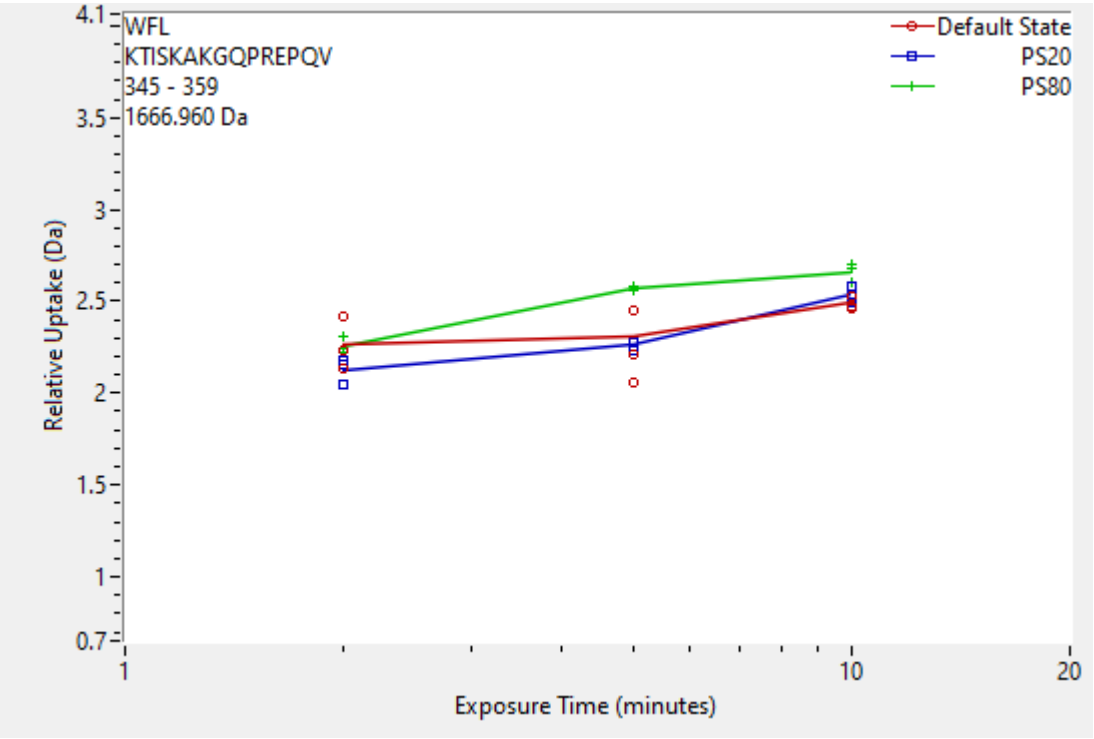


Figure 8.25: Peptide KTISKAKGQPREPQV showing deprotection of showing deprotection of the PS80 peptide in comparison to the default state and PS20. This data is discussed in figures 4.6 and 4.7.

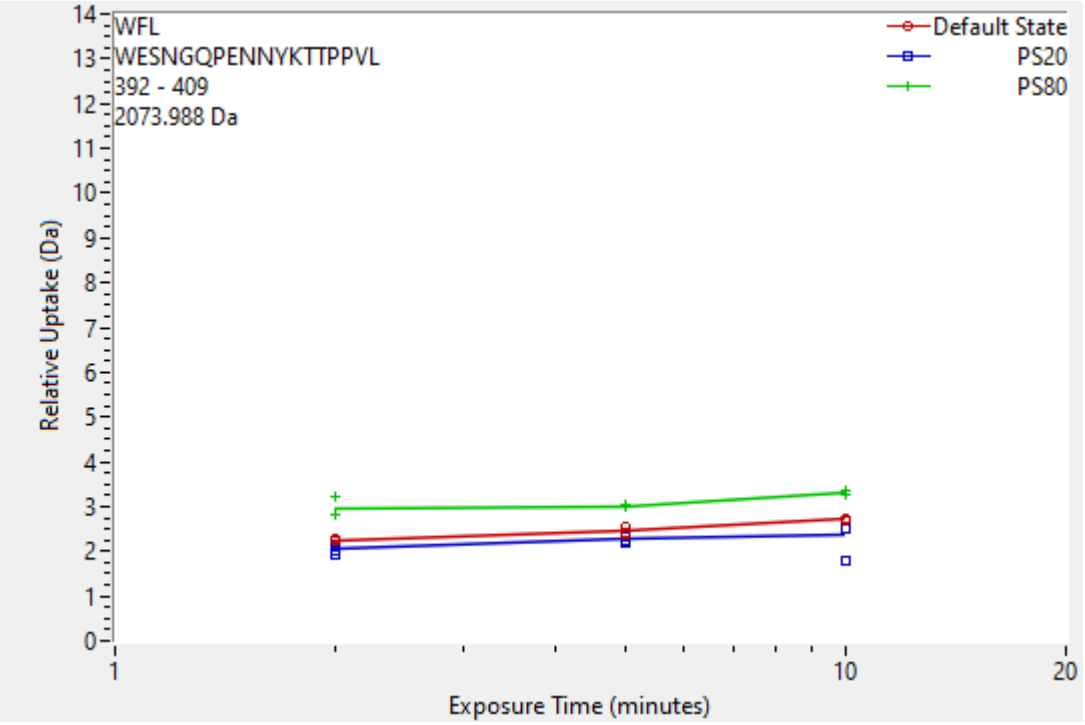


Figure 8.26: Peptide WESNGQPENNYKTTPPVL showing deprotection of showing deprotection of the PS80 peptide in comparison to the default state and PS20. This data is discussed in figures 4.6 and 4.7.

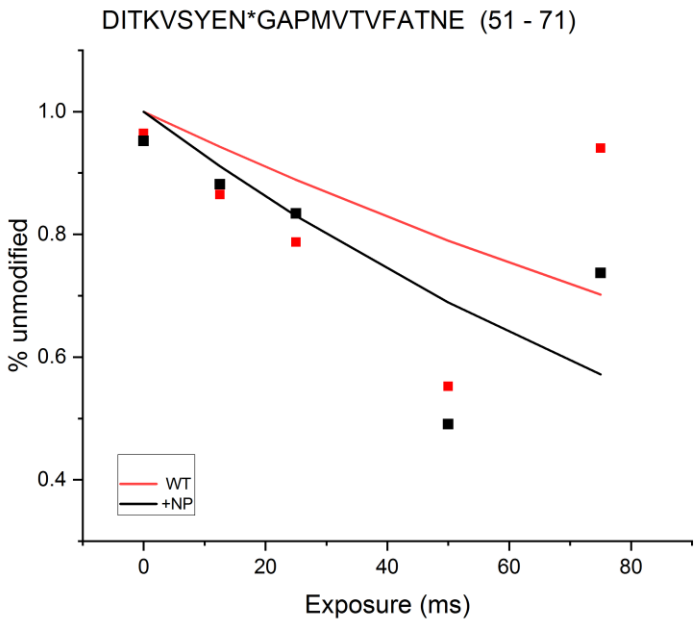


Figure 8.27: Rate constant curve for residues 51-71 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

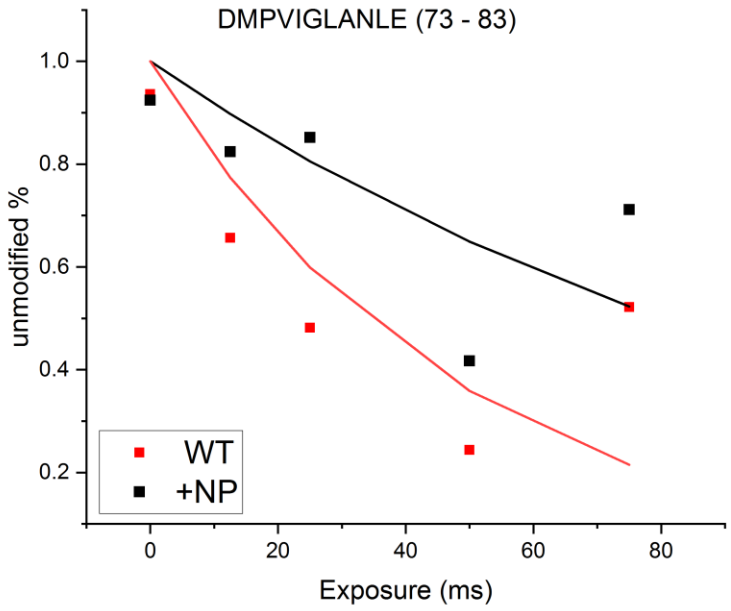


Figure 8.28: Rate constant curve for residues 73-83 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

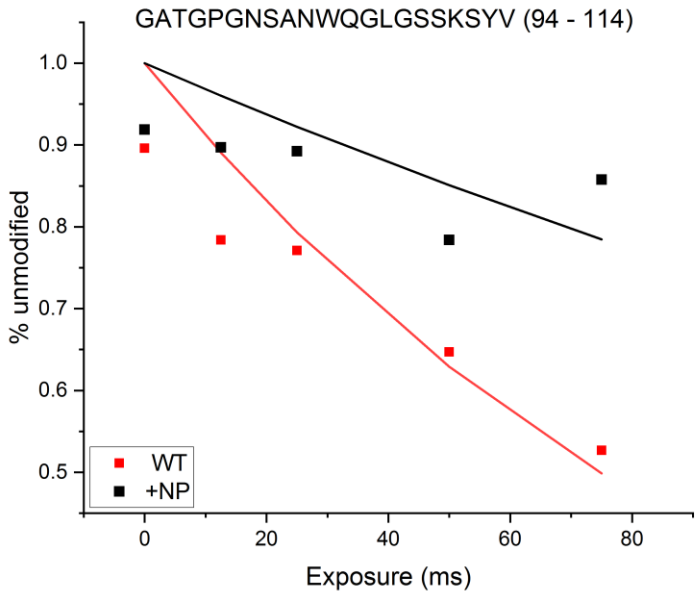


Figure 8.29: Rate constant curve for residues 94-114 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

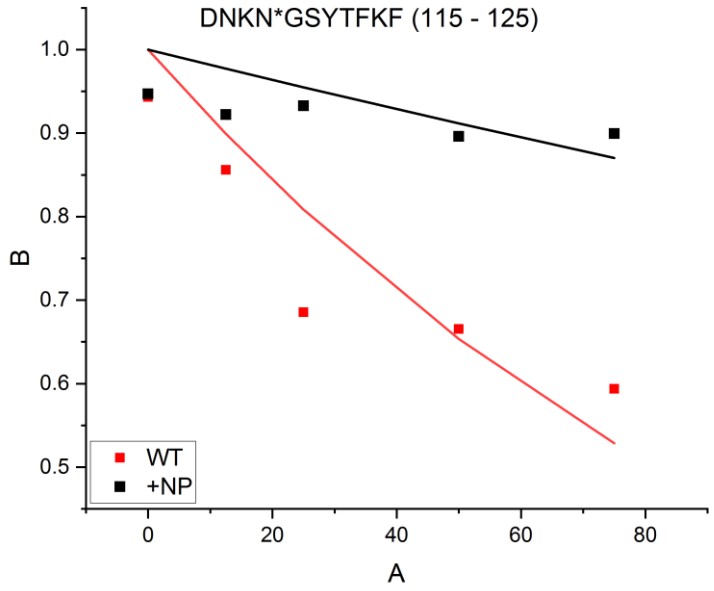


Figure 8.30: Rate constant curve for residues 115-125 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

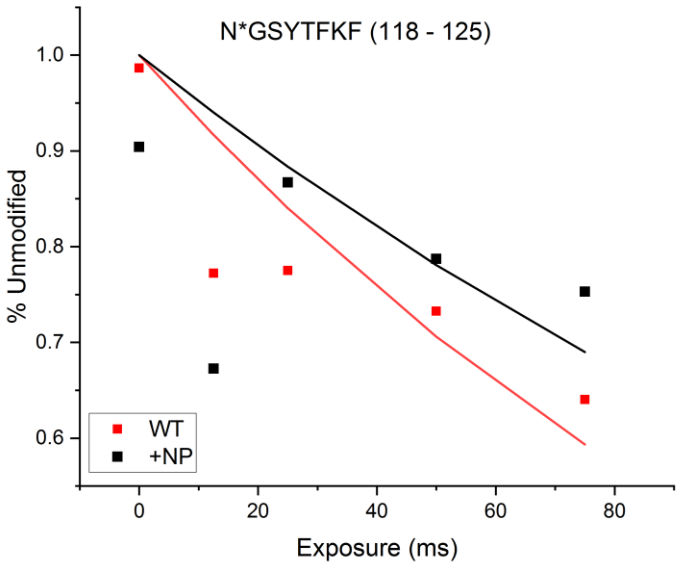


Figure 8.31: Rate constant curve for residues 118-125 *showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.*

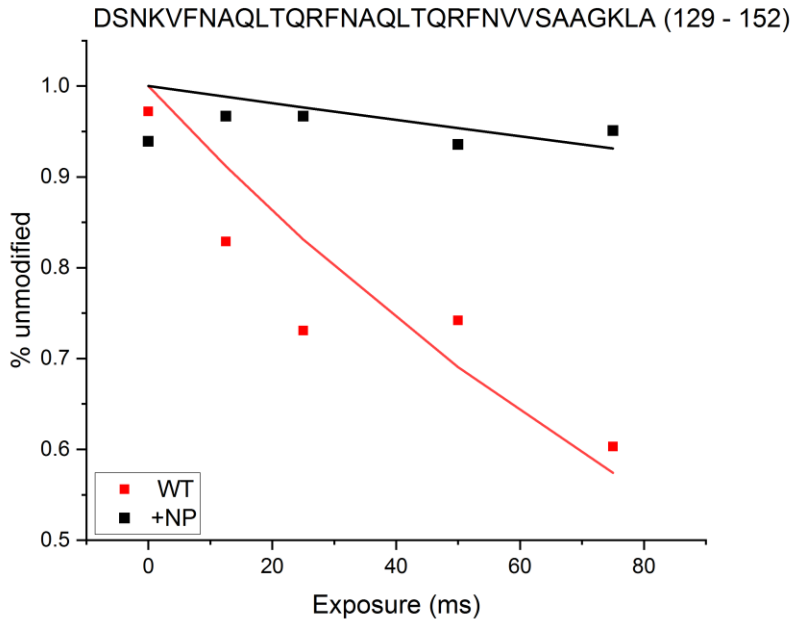


Figure 8.32: Rate constant curve for residues 129-152 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

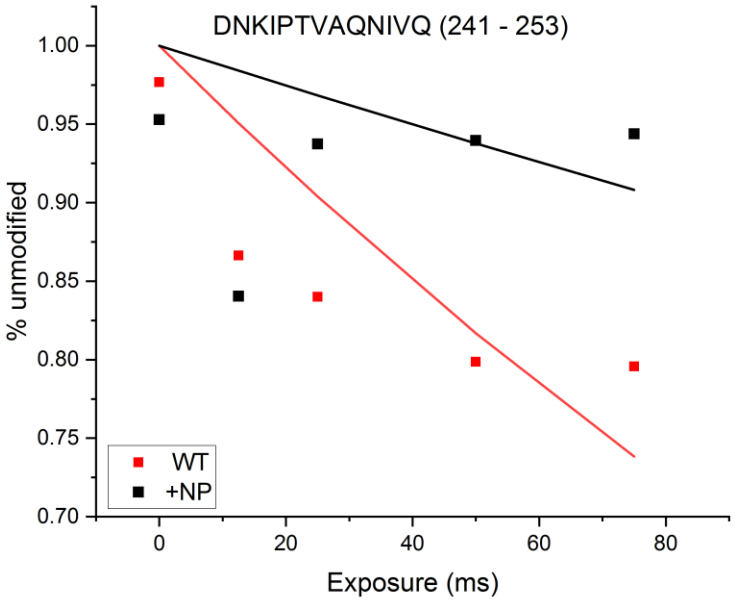


Figure 8.33: Rate constant curve for residues 241-253 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

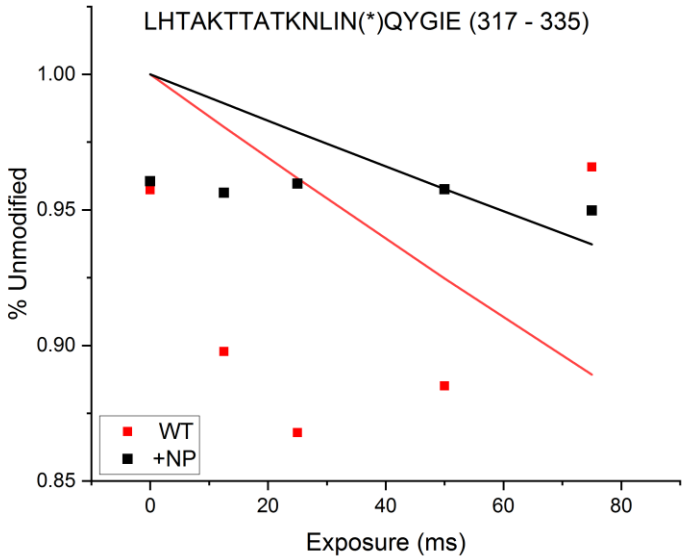


Figure 8.34: Rate constant curve for residues 317-335 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

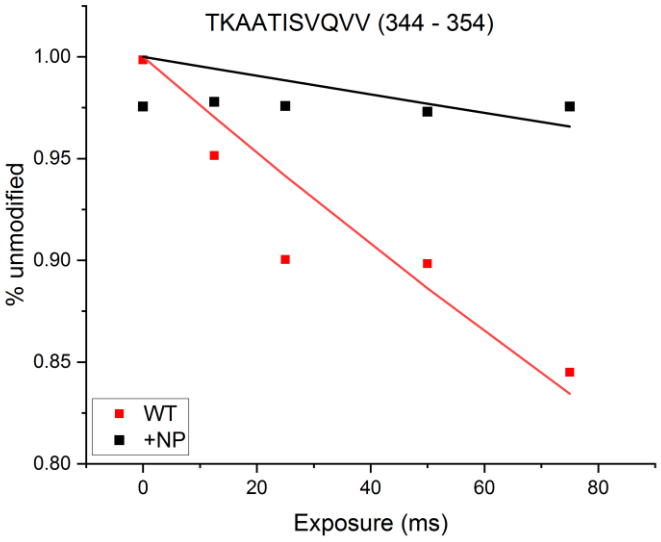


Figure 8.35: Rate constant curve for residues 344-354 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

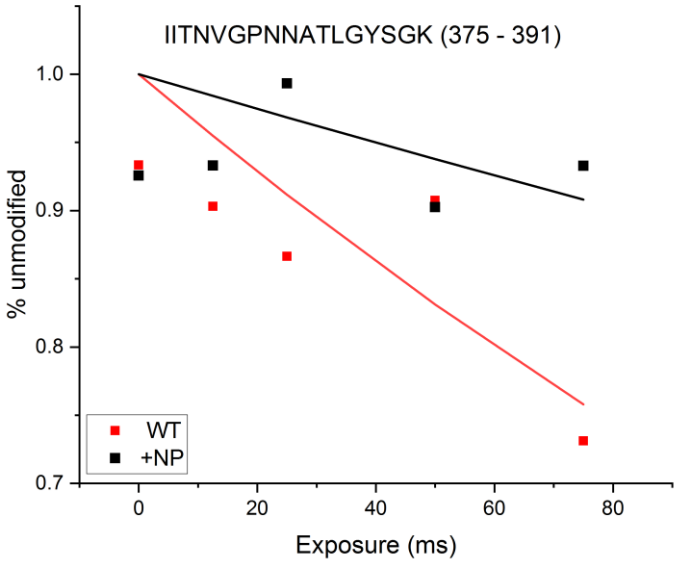


Figure 8.36: Rate constant curve for residues 375-391 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

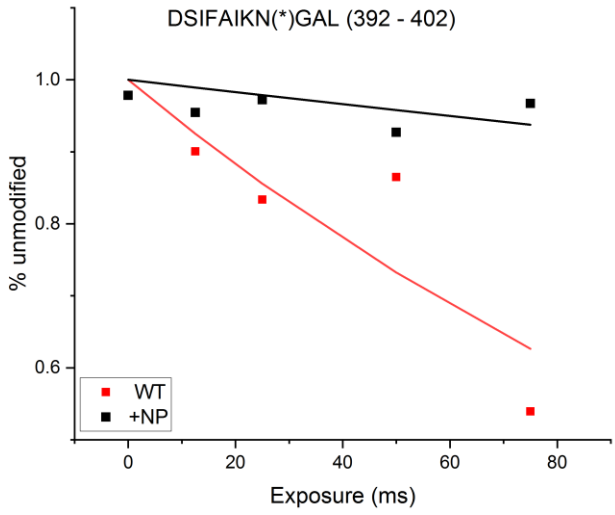


Figure 8.37: Rate constant curve for residues 392-402 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

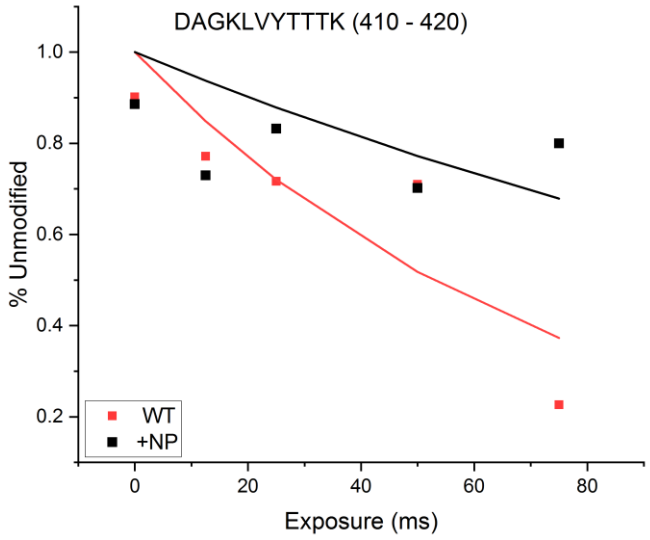


Figure 8.38: Rate constant curve for residues 410-420 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

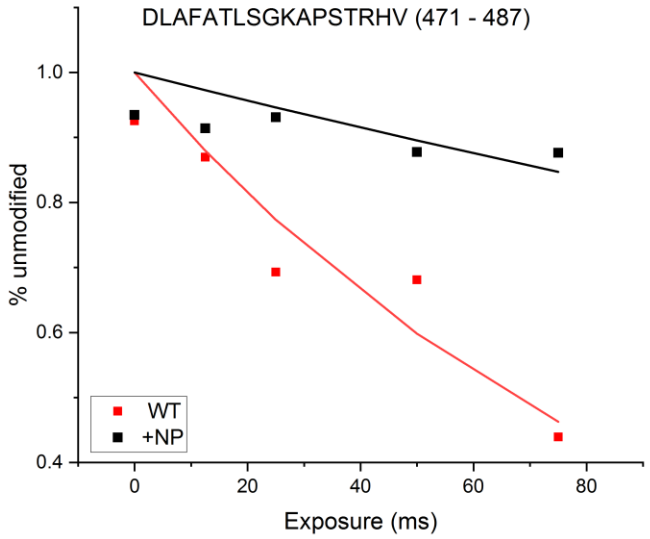


Figure 8.39: Rate constant curve for residues 471-487 showing higher modification for the unmodified in comparison to the nanoparticle. This is shown in figure 5.3.

The unmodified fragmentation patterns for the peptides in chapter 3 are shown below.

Table 8.1: MS/MS fragmentation ions for unmodified W0F

Residue	#	m/z of b ions	m/z of y ions
K	1	129.10	1007.51
G	2	186.12	879.41
W	3	372.20	822.39
F	4	519.27	636.31
G	5	576.29	489.24
G	6	633.31	432.22
G	7	690.34	375.20
G	8	747.36	318.18
G	9	804.38	261.16
G	10	861.40	204.13
K	11	989.50	147.11

Table 8.2: MS/MS fragmentation ions for unmodified W1F

Residue	#	m/z of b ions	m/z of y ions
K	1	129.10	1007.51
G	2	186.12	879.41
W	3	372.20	822.39
G	4	429.22	636.31
F	5	576.29	579.29
G	6	633.31	432.22
G	7	690.34	375.20
G	8	747.36	318.18
G	9	804.38	261.16
G	10	861.40	204.13
K	11	989.50	147.11

Table 8.3: MS/MS fragmentation ions for unmodified W5F

Residue	#	m/z of b ions	m/z of y ions
K	1	129.10	1007.51
G	2	186.12	879.41
W	3	372.20	822.39
G	4	429.22	636.31
G	5	486.25	579.29

G	6	543.27	522.27
G	7	600.289	465.25
G	8	657.31	408.22
F	9	804.38	351.20
G	10	861.40	204.13
K	11	989.50	147.11

Table 8.4: MS/MS fragmentation ions for unmodified F0F

Residue	#	m/z of b ions	m/z of y ions
K	1	129.10	968.49
G	2	186.12	840.40
F	3	333.19	783.38
F	4	480.26	636.31
G	5	537.28	489.24
G	6	594.30	432.22
G	7	651.325	375.20
G	8	708.35	318.18
G	9	765.37	261.16
G	10	822.39	204.13
K	11	950.48	147.11

Table 8.5: MS/MS fragmentation ions for unmodified F1F

Residue	#	m/z of b ions	m/z of y ions
K	1	129.10	968.49
G	2	186.12	840.40
F	3	333.19	783.38
G	4	390.21	636.31
F	5	537.28	579.29
G	6	594.30	432.22
G	7	651.325	375.20
G	8	708.35	318.18
G	9	765.37	261.16
G	10	822.39	204.13
K	11	950.48	147.11

Table 8.6: MS/MS fragmentation ions for unmodified F5F

Residue	#	m/z of b ions	m/z of y ions
K	1	129.10	968.49
G	2	186.12	840.40
F	3	333.19	783.38

G	4	390.21	636.31
G	5	447.24	579.29
G	6	504.26	522.27
G	7	561.28	465.25
G	8	618.30	408.22
F	9	765.37	351.20
G	10	822.39	204.13
K	11	950.48	147.11

Table 8.7 One-way ANOVA for WxF peptides. Significant differences were observed (P<0.05) between W0F and W5. These is explained to section 3.3.

	W0F	W1F	W5F
W0F		0.111900306	0.016848
W1F	0.111900306		0.07187
W5F	0.016848	0.07187	

Table 8.8 one-way ANOVA for FxF peptides. Significant differences were observed (P<0.05) between F0F and F1F, F0F and F5F. This is explained in section 3.3

	F0F	F1F	F5F
F0F		0.034557	0.020718
F1F	0.034557		0.391538
F5F	0.020718	0.391538	

Table 8.9: Selection of labelled peptides for peptide level analysis. Peptides in red have not been selected for analysis, peptides in green are under the threshold for significant different and peptides in blue are significantly different.

Peptide	Charge states used	Modifications used	Decision	Graph	Ratio NP v control
DGGEPAGSIQTLNL (37 - 50)	2	16,32	Not used, poor signal		
DITKVSY*E (51 - 58)	2	16,32	Not used, poor signal		
DITKVSYEN*GAPMVTVFATNE (51 - 71)	2,3	16,32	YES		0.63

DITKVSYEN*GAPMVTVFATNEA (51 - 72)	2	16	low intensity of +16		
N(*)GAPMVTVFATNE (59 - 71)	2	16	Ambiguous precursor		
N(*)GAPMVTVFATNEA (59 - 72)	2	16	Poor +16 signal		
ADMPVIGLANLE (72 - 83)	2	16	Ambiguous +16 signal		
DMPVIGLANLE (73 - 83)	2	16,32	YES		2.37
GATGPGNSANWQGLGSSKSYV (94 - 114)	2,3	16,32	YES		2.87
DNKN*GSYTFKF (115 - 125)	2,3	16	YES		4.59
N*GSYTFKF (118 - 125)	2	16	YES		
DSNKVFNAQLTQRFNAQLTQRFNVVSAAGKLA (129 - 152)	3,4	16	YES		7.81
DGTTVPVAEMVE (153 - 164)	2	16,32	NO, poor precursor signal		
DGTTVPVAE (153 - 161)	2	16,32	NO, poor +16 signal		
DGQGNAPQYTKNIVSHE (167 - 183)	2,3	16,32	NO, poor signal at higher exposures		
DNKIPTVAQNIVQ (241 - 253)	2,3	16,32	YES		3.14
AKNWSRIPTM (269 - 278)	2,3	16,32	NO, lack of clarity vs unmodified		
AKNWSRIPTME (269 - 279)	2,3	16,32	YES		1.17
DFAAGKGHSQQL (289 - 300)	2,3	16	NO, in source only		
LHTAKTTATKNLIN(*)QYGIE (317 - 335)	2,3,4	16	YES		1.8
TKAATISVQVV (344 - 354)	1,2	16,32	YES		51.8
DLKTILPKVQRLE (362 - 374)			NO		
IITNVGPNNATLGYSGK (375 - 391)	2,3	16,32	YES		2.89
DSIFAIKN(*)GAL (392 - 402)	2	16	YES		7.24
DAGKLVYTTTK (410 - 420)	2,3	16,32	YES		2.54
DLKLGQN*GA (421 - 429)	1	16	NO, signal intensity too low		
DLKLGQN*GADS (421 - 431)	1,2	16,32	NO, oxidation signal intensity too low		
DTAFSFVGWSM (432 - 442)	1,2,3	16,48	NO, poor precursor signal		
DTAFSFVGWSMC*SSE (432 - 446)	2	16,32	NO, large overlapping precursor		
DGVDVTKYTGMKA (458 - 470)	2,3	16	NO, poor ms/ms		
DVTKYTGMKA (461 - 470)	2,3	16,32	YES		2.91

DLAFATLSGKAP (471 - 482)	2,3,4	16,32	NO, poor precursor signal		
DLAFATLSGKAPSTRHV (471 - 487)	2,3,4	16,32	YES		4.6
DLAFATLSGKAPHVD (471 - 488)	n	16,32	Odd +32 signal		
DAN(*)GKAIVGL (525 - 534)	2	16,32	YES		1.6
DGTYSFANRGALE (544 - 556)	n	16	YES		2.43

Table 8.10: peptides that are significantly different at the peptide level. This is shown in figure 5.3

Peptide	Start	end
DMPVIGLANLE (73 - 83)	73	83
GATGPGNSANWQGLGSSKSYV (94 - 114)	94	114
DNKN*GSYTFKF (115 - 125)	115	125
N*GSYTFKF (118 - 125)	118	125
DSNKVFNAQLTQRFNAQLTQRFNVVSAAGKLA (129 - 152)	129	152
DNKIPTVAQNIVQ (241 - 253)	241	253
TKAATISVQVV (344 - 354)	344	354
IITNVGPNNATLGYSGK (375 - 391)	375	391
DSIFAIKN(*)GAL (392 - 402)	392	402
DAGKLVYTTTK (410 - 420)	410	420
DVTKYTGMKA (461 - 470)	461	470
DLAFATLSGKAPSTRHV (471 - 487)	471	487