Development of Hydrogen Deuterium Exchange Mass Spectrometry Methodologies for the Analysis of Bacterial Nano-compartments

Thomas Ole Tandrup Lambert

PhD in Chemistry

The University of Edinburgh

2021

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

Acknowledgements

I would like to acknowledge both my supervisors Dr David Clarke and Dr Logan Mackay for providing me with mentorship and support through the PhD. Through your supervision I have been able to evolve into the chemist I am today. From David I learnt to appreciate the necessity of impeccable scientific rigor and that Denmark is definitely still in the European Union. From Logan, I learnt that the application of technical expertise must be coupled with a deep understanding of the theory. I also learnt the limits of my technical expertise in regards to the consumption of hot sauce. I would also like to acknowledge the Clarke group, in particular Dr Jennifer Ross and Dr Kelly Gallagher for providing consistently high-quality samples for my analysis.

Furthermore, thank you to Sam Hughes, Lulu Tucker, Lavrentis Galanopoulous, Kiani Jeacock and Maria Jimenez. You all contributed greatly in generating vibrant, entertaining and challenging debates around our respective fields of research.

My family has been extremely supportive of me with each of them providing help in their own unique way. My mother and father provided the support, insight and lasagnes. My siblings, Sophia and Max, provided a healthy dose of perspective, reality checking and curries. I would also like to thank both Mike and Martin for the extra help they provided for my viva. The confidence you gave me was invaluable and I am forever grateful for it. To the miscellaneous friends, acquaintances and colleagues who have supported me. Your help has been essential for me to finish, and you are all fantastic people. In no particular order: Dom “The Wise One” Jamie “The Running Prole”, Sophie “Lady Planktonia”, Zena “Improvised Ice Skater”, Jay “It’ll be fine…”, Luke “Hides in plain sight”, Rebecca “The calm in the chaos”, Annie ”The chaos”, Hannah “Bikes, Cleats and Beats”, Yas “Arts and Lovecrafts”, Weronika “The best thing about Reading”, Lou “True star of London” and last but not least Kyle “The comedian’s dream!”.

Finally, I would like to thank the SIRCAMS facility managers Alan Taylor and Dr Faye Cruickshank for providing high quality technical support and maintenance of the MS instrumentation.

Abstract

Hydrogen deuterium exchange mass spectrometry (HDX-MS) is a versatile biophysical technique used in a variety of applications within structural proteomics. The key advantages of HDX-MS are that it can provide insight into conformational changes in protein structure as well as allowing investigation of the details of protein-protein interactions, protein-small molecule interactions and protein folding. For these reasons it is an often-used, complementary technique when paired with other biomolecular techniques such as X-ray crystallography, cryogenic electron microscopy (Cryo-EM) and native mass spectrometry.

Encapsulin nano compartments (Enc) or encapsulins are a family of bacterial proteins which form large icosahedral protein cages. These protein cages are of a large enough size (25-30 nm) to allow for the encapsulation of smaller cargo proteins. It has been found that compartmentalization plays an important physiological role within bacteria. One recently discovered ENC, encapsulated ferritin (EncFtn), is known to act as an iron storage system.

Although details of Enc-EncFtn structure are beginning to emerge, several questions remain regarding dynamics and interactions of this complex. This thesis focuses on the application of HDX-MS for the study of these important bacterial nano-compartments.

This body of work initially explores the assembly pathway of EncFtn derived from *Haliangium ochraceum* using a bottom-up HDX-MS workflow. It was previously shown that EncFtns characteristically exist as a decameric ‘pentamer-of-dimers’ arrangement. HDX was used to determine the details of the assembly pathway and the formation of the decameric assembly via recruitment of specific dimer subunits.

This study’s second aim was to analyse the structures of both a loaded encapsulin (Enc-Loaded) and empty encapsulin (Enc-Empty). Differential HDX-MS was used to reveal significant differences in deuteration exchange

between both the empty and loaded forms. The regions of higher deuterium uptake in Empty encapsulin were localized within the interior of the encapsulin and surrounded a hypothesised cargo binding sequence that been previously determined via Cryo-EM. This HDX analysis also highlighted the dynamic and flexible nature the putative entry pore within the ENC.

Finally, a novel automated HDX method was developed for top-down HDX analysis of proteins. This strategy allows for single residue resolution of a deuterium uptake event with the potential of proteoform isolation and selection. Fourier Transform Ion Cyclotron Resonance (FT-ICR MS) was coupled with an automated HDX platform and Electron Capture Dissociation (ECD). This workflow was applied to study EncFtns, and direct comparison with traditional bottom-up data demonstrated the feasibility and potential advantages of this new workflow.

Lay summary

Proteins are essential biomolecules and vital components in the biological make up of organisms. They are the active molecules in living systems, acting as catalysts, transporters, and motile elements in organs. Understanding the functions which proteins play within the body and the molecular details of how individual proteins perform these functions is crucial when trying to understand biological processes.

To this end, it is important to consider both the behaviour and structure of proteins. Mass spectrometry is one such tool which allows for the study of protein structure. This thesis focuses on studying protein structure through hydrogen deuterium exchange mass spectrometry (HDX-MS).

Hydrogen is the smallest element within the periodic table and consists of 1 proton and 1 electron. Weighing 1.008amu, Hydrogen is one of the four most common elements in living organisms with others being Carbon, Oxygen and Nitrogen. For this reason, proteins heavily use hydrogens within their structure which makes them ideal candidates to monitor when analysing proteins. Hydrogen can also incorporate neutrons to form multiple hydrogen isotopes with the most common being deuterium. Deuterium is formed through hydrogen recruiting 1 neutron which gives the new isotope an atomic weight of 2.014 amu.

Interestingly, when a protein is exposed to solution of deuterium select groups of hydrogen will readily exchange with the deuterium isotopes in solution. For each successfully exchanged hydrogen to deuterium there will be a mass increase of 1.006 amu. Differences in protein structure or the presence of an interaction site, will slow or increase the rate at which this exchange occurs. HDX-MS monitors this phenomenon by measuring and comparing the rate of deuteration across a protein’s structure.

This thesis has successfully utilized HDX-MS to identify crucial structural characteristics of the encapsulin and encapsulated ferritin protein families and produced novel HDX-MS methodologies.

List of Figures

List of Tables

Contents

Abbreviations

DDA Data dependant acquisition

DIA Data independent acquisition

CCS Collision cross section

CEM Chain ejection model

CID Collision-induced dissociation

CRM Charge residue model

Cryo-EM Cryogenic electron microscopy

ECD electron capture dissociation

EM Electron microscopy

Enc Encapsulin

EncFtn Encapsulated ferritin

ESI Electrospray ionisation

ETD Electron transfer dissociation

EXD electron transfer or capture dissociation

FID Free induction decay

FOC Ferroxidase centre

FT-ICR Fourier transform-ion cyclotron resonance

HDX Hydrogen deuterium exchange mass spectrometry

H. ochraceum Haliangium ochraceum

Hoch-EncFtn Haliangium ochraceum association encapsulated ferritin

ID Isotopically depleted

IEM Ion evaporation model

IM-MS Ion-mobility mass spectrometry

LC-MS Liquid chromatography mass spectrometry

LS Localization sequence

kDa Kilodalton

m/z Mass-to-charge ratio

MS Mass spectrometry

MW Molecular weight

Non-FOC: EncFtn dimer subcomplex which does not contain the FOC region

NMR Nuclear magnetic resonance

PDB Protein Data Bank

R. rubrum Rhodospirillum rubrum

RDU Relative deuterium uptake

RFU Relative fractional uptake

Rru-EncFtn Rhodospirillum rubrum association encapsulated ferritin

SDS-PAGE Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SEC Size exclusion chromatography

SIM Selected ion monitoring

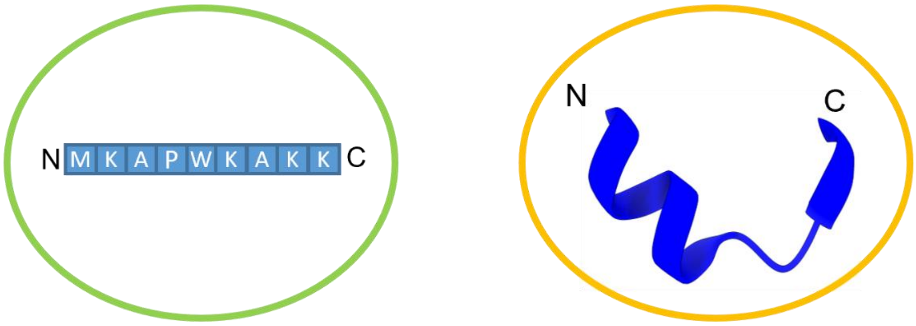
TOF Time-of-flight

WT Wild type

Chapter 1 Introduction

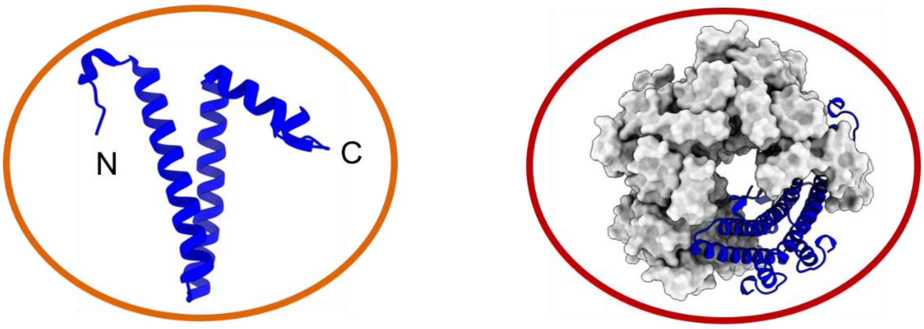
1.1 Protein structure

Proteins are essential biomolecular polymers which play important roles in most biological processes within the cellular environment. Proteins have a wide array of functions with roles in mediating signalling, storage of small molecules and acting as catalysts. The chemical versatility of proteins is a result of their complex hierarchical structure. Protein structure consists of four distinct levels, the primary, secondary, tertiary and quaternary. The primary structure relates to a protein's amino acid sequence, the secondary structure consists of the initial folds within the protein into alpha helices and beta sheets, the tertiary structure describes how these secondary structure elements are arranged and describes the overall conformation of a protein monomer. Finally, the quaternary structure describes how multiple protein polypeptide chains can associate to form higher order protein complexes (Figure1.1).



Primary structure

Secondary structure



Tertiary structure

Quaternary structure

Figure 1.1: Protein structure hierarchy

1.2 Traditional Methods for studying protein structure

There is an expansive list of available techniques able to study protein structure. Below is a selection of traditional methods which can complement mass spectrometry (MS) based techniques.

X-Ray crystallography has held a long pedigree as the standard for solving the 3D protein conformation at an atomic resolution[1]. Using X-rays, which have wavelengths comparable to the ~1Å range of chemical bonds, it is possible to determine the 3-dimensional molecular structure of a protein from its crystalline form.

Biomolecular X-Ray crystallography was first implemented 1958 by Kendrew et al, who solved the structures of whale hemoglobin and myoglobin[1]. This initially work laid the foundations for turning X-ray crystallography into the widely utilized technique it is today. To collect the proteins atomic coordinates, X-ray crystallography relies on the protein structure being static through crystallization into a regular repeating lattice. This subsequently makes discerning protein dynamics difficult. One solution would be to utilizes the molecular disorder within the crystal (B-factors) as an estimate of protein dynamics, but this is not an ideal solution[2]. Furthermore, crystallography is reliant on possessing a viable protein crystal which is not possible for every protein.

Cryogenic-Electron Microscopy (Cryo-EM) is a technique which focuses on studying biomolecular structures at near atomic resolution. The modern form of Cryo-EM was developed in 1982 by Dubochet *et al*[3]. Their study was focused on observing biomolecules at near native conditions through transmission electron microscopy. They achieved this by introducing liquid nitrogen temperatures into the workflow which would preserve the structure of a biomolecule of interest allowing it to be observed by the microscope[3]. This provided the initial benefit over X-ray crystallography as it was not necessary to have 3D crystal to allow for analysis. Likewise, it can be used to study large biomolecules and is capable of analysis into the mega-Dalton mass range[4]. There are drawbacks associated with Cryo-EM analysis of biomolecules. Very low signal to noise ratio is arguably the largest limitation, the technique is reliant on sufficient electron absorption to provide an acceptable image contrast[5]. However, the carbon, oxygen hydrogen and nitrogen, which are the main constituent atoms of biomolecules, demonstrate poor electron absorption. A low image contrast results in structural features becoming difficult to elucidate.

Finally, nuclear magnetic resonance (NMR) spectroscopy is a widely used technique for studying proteins. NMR spectroscopy is possible because of nuclei absorbing radio frequencies within a magnetic field. By monitoring an interaction’s NMR within electromagnetic field, it is possible determine both its nuclear identity and environment. NMR spectroscopy has a long history of studying Hydrogen Deuterium Exchange (HDX) profiles of biomolecules[6]. Unlike hydrogen, deuterium is not detectable by NMR. The result is for every incorporated deuterium there is a loss of a proton signal within the amide backbone of a polypeptide. This can be monitored, allowing for analysis of protein dynamics and structure. Although powerful, NMR is limited by analyte size with viable protein sizes being at the 40–50 kDa[7].

1.3 Biomolecular Mass Spectrometry

Over the 20th century mass spectrometry has evolved to become an essential addition to the large toolkit available for protein analysis. The advent of structural mass spectrometry was heralded by the discovery of soft ionisation techniques, generating a large assortment of structural MS strategies[8]. Figure 1.2 gives an overview of MS based strategies.

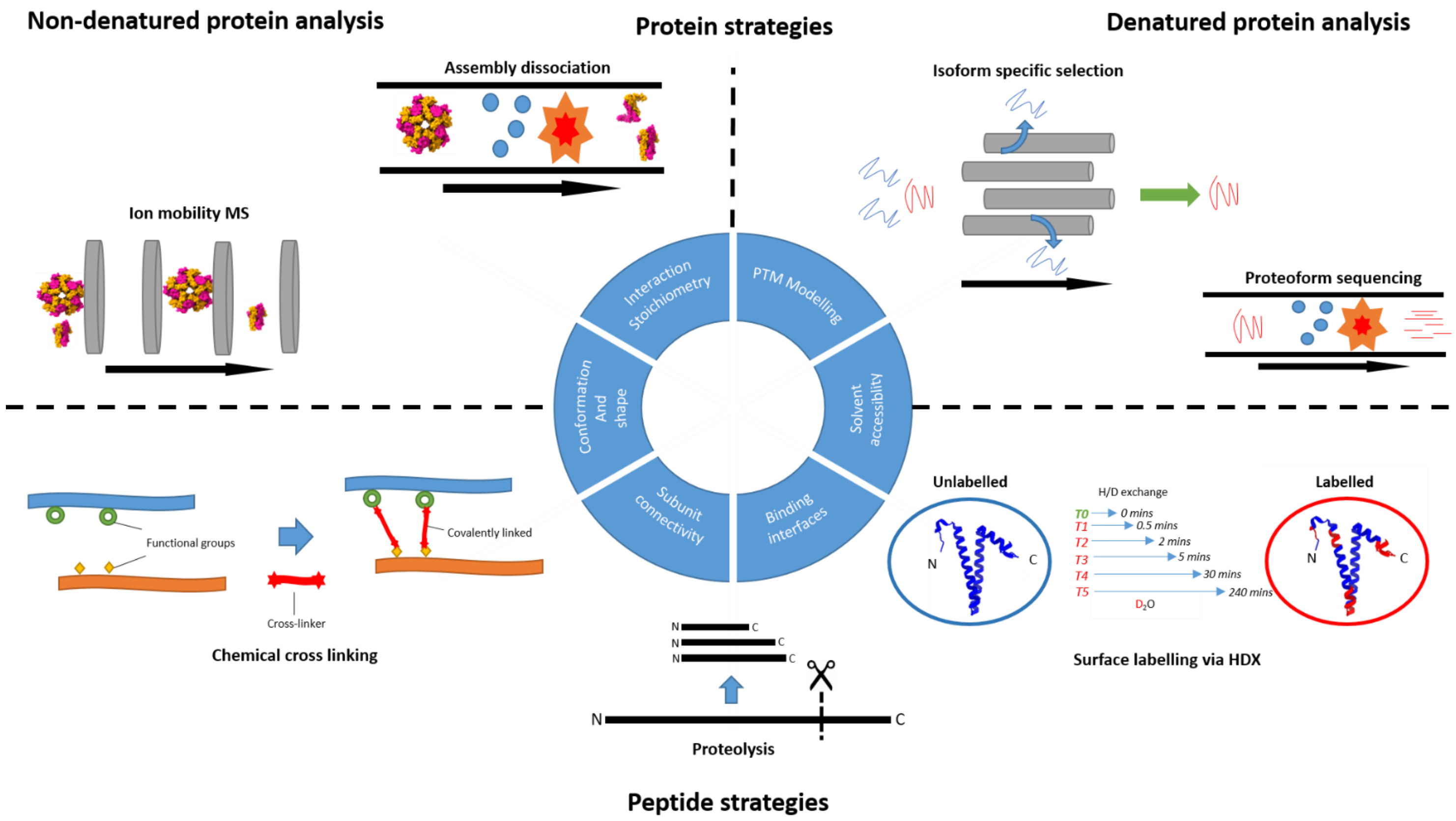


Figure 1.2: Overview of mass spectrometry strategies for structural and molecular biology. Each technique featured focuses on the respective nearest segments of the central circle. The protein centric strategies (top panel) have been sub divided into non-denatured and denatured protein analysis. The bottom panel focuses on the analysis of peptides

1.3.1 Native Mass Spectrometry

Native MS describes experimental parameters which allows for insights into protein behaviour under physiological conditions[9]. In Native MS large biomolecules in the liquid phase can be transferred into the gas phase whilst preserving the biomolecules native state. This process is reliant on using conditions that preserve the biomolecule’s structure (governed by non-covalent interactions) of the biomolecule during the ionisation process, with electrospray ionisation (ESI) being a major technique to facility this[9]. To perform intact native MS the sample must be transferred from physiological buffer into a volatile buffer to aid the electrospray[10]. This factor increases the complexity of the experimental design as the buffer conditions and susceptibility to spray is sample dependent, with the conditions needing to be optimized for each sample[11]. The ESI sprayer ionizes and expels analytes into the gas phase whilst retaining the protein or protein complex’s integrity (see section 1.4.2 for details). Once in the gas phase, information regarding the analyte’s stoichiometry, protein interactions or the topology of protein assemblies can be assessed. Advance native MS techniques such as Ion Mobility Mass Spectrometry (IM-MS) have extended the application of native MS as a technique and are discussed below.

1.3.2 Ion Mobility Mass Spectrometry

Ion mobility mass spectrometry (IM-MS) is a separation MS technique that allows for multidimensional characterization of biomolecules[12]. IM-MS rapidly separates ionized analytes in an inert gas whilst under the influence of an electric field. The analytes travel through the system and separate from one another based on their mobility[12]. In drift-time IM-MS, ions will enter a drift tube containing neutral buffer gas at a specific pressure. As the ions migrate through the drift tube, smaller ions will separate from the large ions due to the difference in collisional cross section (CCS). After separation, the ions will reach the MS detector which records both the ion's *m/z* and drift time through the ion mobility device. By careful calibration of the instrument the recorded drift time of an ion can be converted to its CCS; therefore IM-MS can be used to infer some structural information on native protein ions. Although versatile, the low resolution of IM-MS does limit its application as analyte peaks can overlap[13]. This can be mitigated through the implementation of an liquid chromatographic (LC) separation step prior to electrospray[13].

Furthermore, recent studies have focused on developing high resolution IM-MS approaches. An example of this can be seen in a 2019 study by Jakub Ujma *et al,* where they showed that higher resolution IM-MS was possible through the use of a cyclic ion mobility separator (cIM)[14]. They demonstrated that greater separation of ions would occur after increasing the number of cycles the ions would travel around the cIM separator[14]. With each cycle the drift time of ions would be extended allowing for a great degree of separation and higher resolution data. Although this does increase the resolution, there was a marked decrease in sensitivity further demonstrating the limitation of IM-MS approaches. Furthermore, cyclic IM-MS can result in in higher mobility ions overtaking low mobility ions in a phenomena known as roll over. This roll over effect can lead to an overlap in data and artefacts[15].

1.3.3 Covalent labelling and Crosslinking

Covalent labelling is a technique used to explore protein conformation. It relies on the introduction of a covalent labelling agent reacting with functional groups of residues within a protein’s sequence. Exposed residues will react rapidly, and protected residues will react slowly[16].

A commonly used covalent label are hydroxyl radicals ($^{\bullet}$OH). Hydroxyl labelling benefits from being able to interact with a large assortment of residues whilst remaining highly reactive[16]. Once introduced to a protein sample the $^{\bullet}$OH radicals will react (<1µs) and modify, via oxidation, exposed residues. These factors make it ideal for studying short lived structural folds within a protein’s structure[17]. Following labelling an enzymatic digest is performed just prior to the LC-MS, allowing for quantitative analyses of the oxidation patterns. This technique can use to study protein-protein interactions[18,19], protein-ligand interactions[19] and binding behaviours of membrane bound proteins. There are limitations to covalent labelling as they may alter the protein’s structure[20].

Covalent Crosslinking utilizes binding interactions (crosslinks) between two proteins. The interaction relies on both proteins having viable side chains at an appropriate distance from one another, interacting with a crosslinking agent[21]. There is a large library of crosslink agents available with the majority consisting of two alkyl chain reactive sites. Lysine and Cysteine are common targets for crosslinking forming homo-bifunctional linkers (Lys-Lys) or (hetero-bifunctional linkers) (Lys-Cys)[22]. Like labelling, once the crosslinks have been formed, the modified proteins will undergo a proteolytic digest just prior to entering the MS for analysis. Crosslinking can also be used to study protein-protein interactions[22], protein-ligand interactions[23] and binding behaviours of membrane bound proteins[24]. However, the analysis is reliant on separating modified peptides from the unmodified forms which is further complicated by the presence of multiple peptide forms being present within a spectrum.

1.3.4 Hydrogen Deuterium Exchange Mass spectrometry

Hydrogen/Deuterium Exchange-MS (HDX-MS) describe the equilibrium reaction in which a covalently bonded hydrogen is replaced by a deuterium (H→D) isotope or the inverse (D→H)[25]. This reaction can be used to study hydrogen containing complexes such proteins and peptides[25]. When a peptide which contains exchangeable hydrogens is exposed to deuterium the exchange will occur[25]. However, extensive research has been conducted exploring how altering conditions such as temperature and pH impact the dynamics of HDX[25–28] As the HDX reaction is in equilibrium, the concentration of the deuterium labelling buffer should be significantly higher than the exchangeable hydrogens on the peptide's amide backbone[27]. In principle, a HDX experiment consists of a timecourse allowing the rate of the H→D exchange to be measured as it moves to equilibrium.

A review of the legacy of HDX by *S.W.Englander* names *Kaj Linderstrøm-Lang* as the father of HDX[29]. In 1954 *Linderstrøm-Lang* first demonstrated the exchange of hydrogen protons in insulin with deuterium in aqueous solutions[30]. He would later go on to explore deuterium exchange of poly-DL-alanine[31] and myoglobin[32]. Years later it was found that the analytical capabilities of techniques such as NMR and MS when coupled to HDX would allow for protein hydrogen exchange studies[33]. Although NMR-HDX coupling was the first to allow protein HDX studies, HDX-MS has become the mainstream technique for protein hydrogen exchange studies and was first demonstrated by Anderegg *et al*[34]. A detailed discussion of HDX-MS can be found in section 1.5.

1.4 Mass Spectrometry Theory and Instrumentation

1.4.1 Liquid Chromatography

Chromatography is a widely used separation science that differentiates species within a complex mixture based on physical, chemical or biochemical properties. The technique relies on analytes interacting with two phases, a mobile and stationary phase respectively[35]. The mobile phase provides an environment that flows or “carries” the compound over the stationary phase. The species within the mixture will separate out based on their affinity to each of the phases; with species which have either low affinity to stationary phase or a high affinity to the mobile phase and will thus travel faster than those with the opposite affinities.

With the advent of ESI-MS and Liquid Chromatography-MS (LC-MS) intact proteins became viable for analysis[36]. Protein chromatography typically uses hydrophobicity as the separation criteria. Separation by hydrophobicity uses the differing hydrophobic nature of individual proteins based on their amino acid composition[37].

1.4.2 Electrospray Ionization

Prior to MS analysis, an ionization process is necessary, this poses two unique ionization challenges. For MS analysis of intact proteins their integrity has to be maintained, therefore degradation from the ionization process has to be minimal[38]. Coupled to this, there are significant challenges in evaporating liquid buffers. To overcome this soft ionization methods are used, with the most common being Electrospray ionization *(ESI)* [38].

The first documented uses of ESI was in 1968 by Dole[39], in his original experiment a sample is first dissolved in solution and is then passed through a metal capillary. A large potential (typically, 3-5 kV) is applied to the electrospray emitter, this large electric potential preferentially draws out ions with an opposite charge which are release from the capillary as a nebulized stream of droplets. ESI allows for both positive and negative ion selection. For example, in positive mode, the potential draws positively charged ions[39,8]. The ESI source typically is heated to encourage the evaporation of the droplets. As the droplets volume decreases, charge density and repulsive force between ions increases[8]. The repulsive force of the ions works against the surface tension as the droplet evaporates. Eventually the repulsive force exceeds the surface tension of the droplet and coulombic fissions occur, this is known as the Rayleigh limit[8]. This process repeats until, eventually, gas phase ions are produced.

Although Dole first described ESI it wasn't until 1984 that ESI was coupled to MS (ESI-MS) by Fenn *et al*[8,40]. Initially ESI-MS allowed for the coupling of liquid separation techniques such as LC with MS instrumentation. Fenn's research group would also go on to demonstrate the first successful ionization and MS spectra of proteins such as insulin (bovine), myoglobin and cytochrome C amongst many others[40]. This was made possible due to ESI being a "soft ionization" approach where little energy is imparted on the target analyte. The low energy minimizes fragmentation, preserving protein structure and integrity.

In modern ESI-MS, the sample droplets get uniformly charged from the metal capillary's voltage. Nebulizer gas flows axially around the outside of the capillary to assist with nebulisation[41].

As the analyte leaves the capillary it forms a Taylor cone from with a jet of charged particles forms. The jet then breaks up into a plume of charged droplets.

As the sample droplets are being sprayed, a heated source encourages evaporation and therefore columbic fissions events, which results in the formation of smaller droplets. This process repeats until the ions are in the gaseous phase. An example of this can be seen in Figure 1.3[42].

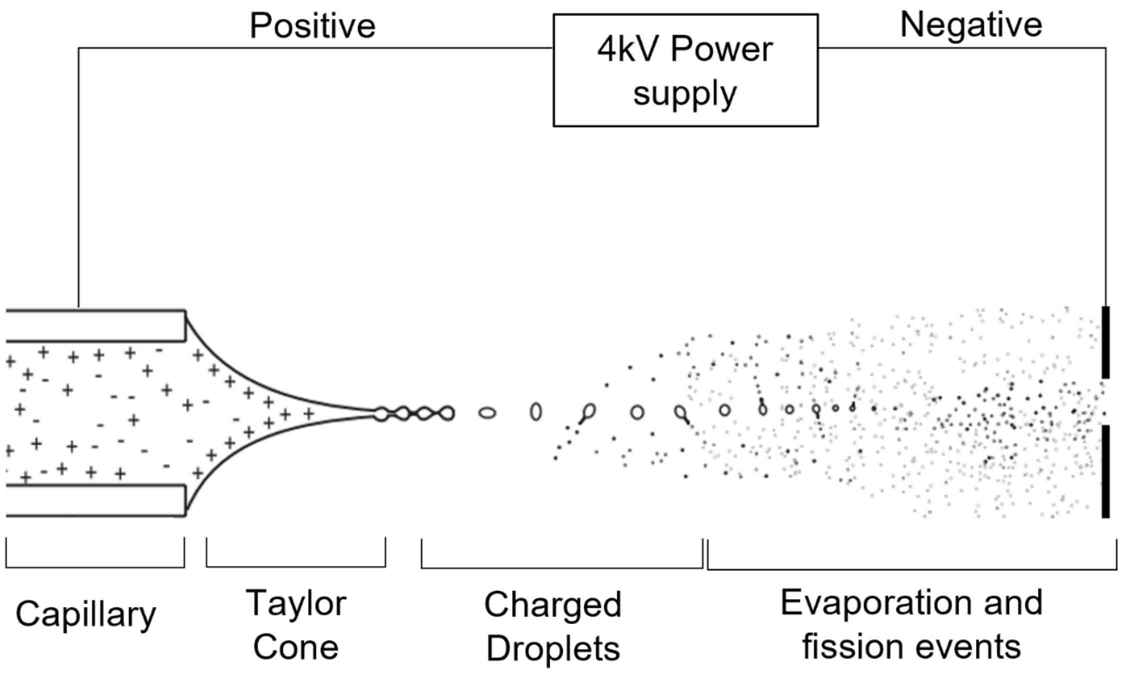


Figure 1.3: Overview of an electrospray capillary, detailing the ejection of charged droplets and the subsequent fission events. A potential difference is applied between the ESI capillary and the electrode distorting the meniscus into a Taylor cone resulting in ejection of charged droplets. Ion release is mediated via heat induced evaporation and Coulombic repulsion. Image modified from P. Kebarle and U. Verkerk[42]*.*

The development of ESI-MS heralded a large step forward in biological mass spectrometry. However, questions remain about how the final ionization step of the process works. Three competing models have risen to prominence and are currently central to this debate, the ion evaporation model (IEM)[43,44,] the charge residue model (CRM)[45] and more recently chain ejection model (CEM).

The IEM model proposes that the solvent droplet shrinks from evaporation to a size of <10 nm via Coulomb repulsion. Once at this size, direct ion emission will take over from the columbic repulsion as the means of ion production. Ion emission is the product of the surface electrical strength which will increase as the droplet shrinks and the ions repel. This electrical strength will reach a threshold at which ions will begin to escape from the droplet gaining their charge from the field strength on the surface (Figure 1.4).

The CRM model was initially proposed by Dole and posits that it's the solvent's evaporation to dryness which is solely responsible for ion production[39]. In this model the solvent droplets would continually undergo fission until reaching "dryness" releasing the ions retaining some of the droplet's charge (Figure 1.4).

Recent work has also provided a third model for the final ionization step CEM which relates to the ESI ionization of unfolded proteins. In CEM proteins are exposed to an acidic mobile phase which causes the protein to unfold and switch in its properties from a compact and hydrophilic structure to an accessible unfolded and hydrophobic structure[46]. The new hydrophobic structure will migrate to the droplets surface. When this occurs in a droplet close to its Rayleigh limit, it is thought that the one of the protein's termini will be first be expelled from the droplet followed sequentially by the remaining protein structure[47] (Figure 1.4).

Each model has publications providing evidence for their validity with no definitive study proving one model over the other[41]. The general consensus is that the large multi protonated molecules will be undergoing CRM[48], small singularly charge molecules will be produced via IEM[49] and unfolded proteins will be ionized through CEM[47,50,51]. Although there has been recent evidence of protein ionization through the IEM[52] further emphasising the ongoing discourse surrounding this subject.

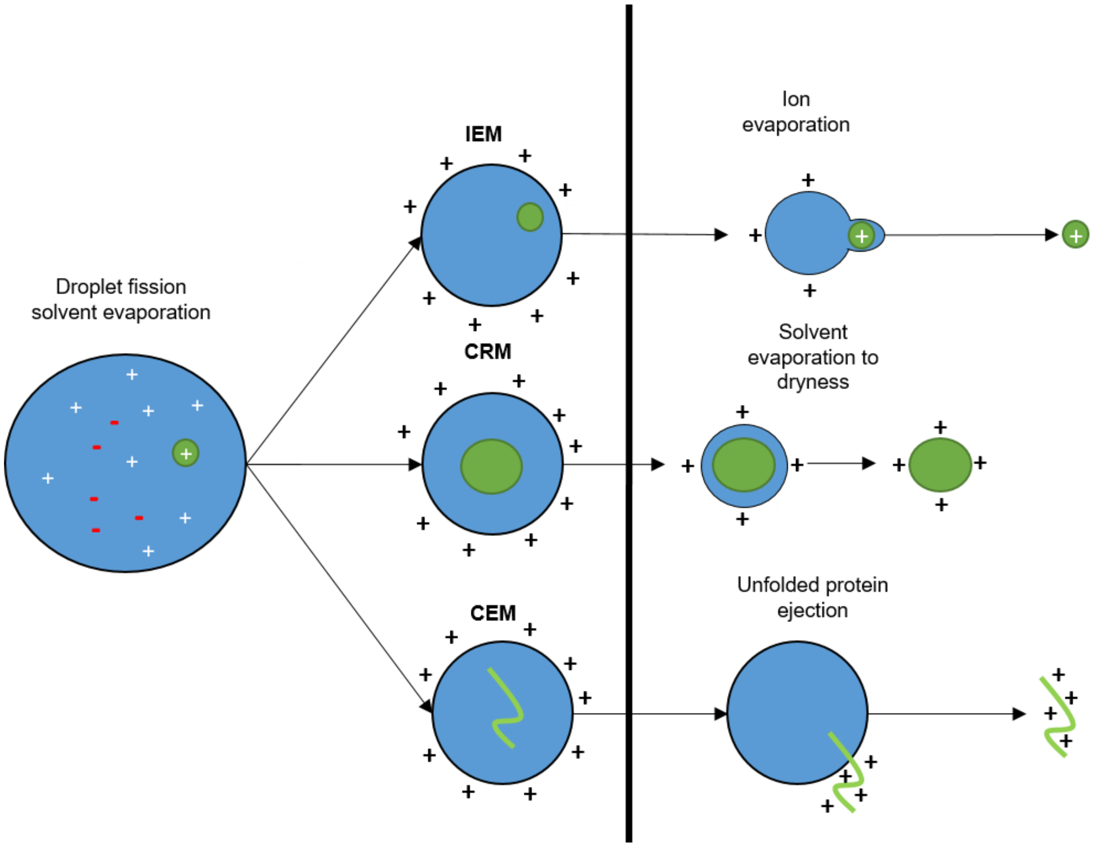


Figure 1.4: A comparison of protein ionisation by IEM, CRM and CEM. The black line represents where the two models begin to differ. The IEM model proposes that the solvent droplet shrinks from evaporation to a size of <10 nm via Coulomb repulsion. Once at this size direct ion emission will take over from the columbic repulsion as the means of ion production. The CRM model proposes that solvent evaporation to dryness is solely responsible for ion production

1.4.3 Mass analysers

1.4.3.1 Time Of Flight

A time of flight (TOF) MS determines an ion's *m/z* ratio by measuring the time it takes for an ion, with known kinetic energy (E_k), to travel through a field free region of a fixed length to reach a detector[53].

The ions are accelerated through an electric potential before they enter a field-free drift region. The velocity *(v)* of an ion is the result of the application of the acceleration voltage (*U)* where the electric potential energy of the ion E_{el} is converted into kinetic energy E_k (Equation 1.1).

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

$$E_{el} = qU$$

$$E_{el} = E_k$$

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

Equation 1.1:

The potential energy of a charged ion is related to the charge of the ion and to the strength of the electric field where *(*E_{el}*)* is electric potential energy, (*q*) is the charge of the ion, and *U* is the acceleration voltage. When the charged ion is accelerated its E_{el} is converted to kinetic energy *(*E_k*)*. Rearranging this equation will result in the velocity, or distance *(d)* over time *(t)*, of the ion being related to its *m/z* value. Thus, by accurately measuring the time of flight through the field free region, of known distance (*d*), the m/z ratio can be determined (Equation 1.2).

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

$$\frac{1}{2}m\ \frac{d^2}{t^2} = qU$$

$$t = \ d\sqrt{\frac{m}{2qU}}$$

$$\frac{m}{q} = 2U\frac{t^2}{d^2}$$

Equation 1.2:

In theory, accelerating each ion with the same potential will leave the *m/z* as the only discriminating factor between each ion. However, differences in the ion's spatial position in the acceleration region and the spread of kinetic energy imparted by the ionisation process will leave slight differences between ions. This results in a decrease in mass resolution as the arrival time of the ions reaching the detector will be different.

It is possible to overcome this problem by using a reflectron ion mirror. The mirror corrects for the kinetic energy distribution of the ions with a retarding electric field. More excited fast-moving ions will travel deeper into the mirror's field, taking a longer path to reach the detector[54]. The slower, less excited ions will take a quicker path which will compensate for the ions' slow speed allowing both sets of ions to meet reach the detector at the same time, thus increasing the mass resolution[54]*.* Furthermore, the reflection has the added benefit of increasing the drift time of ions allowing greater separation between analytes improving resolution. One key limitation to using reflectrons is their use on large ions such as proteins. The electric field generated by a mirror can interfere with their trajectory inhibiting detection.

A comparison between a reflectron and regular TOF can be seen in Figure 1.5. Detection is produced from an ion detector which amplifies the electrical current of ions as they strike it, which is then measured and converted into mass spectrum[54].

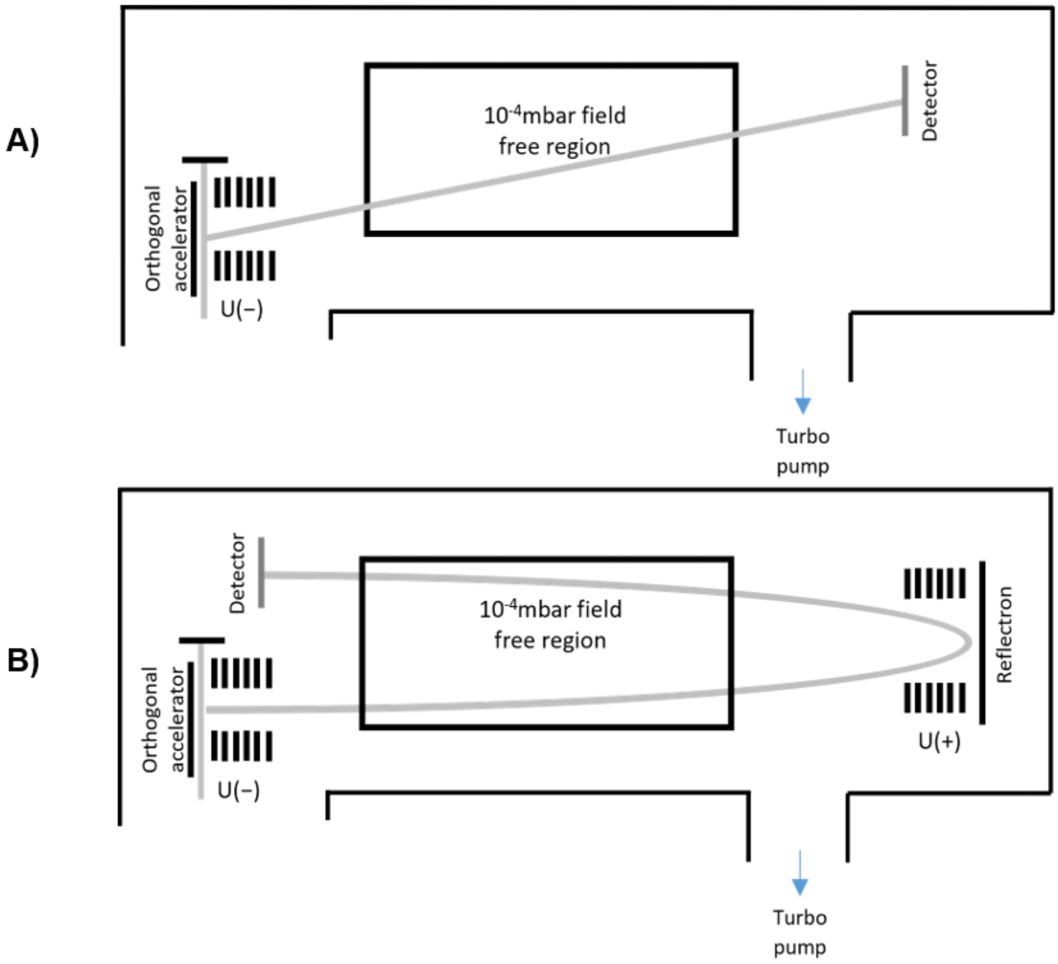


Figure 1.5: A time of flight (TOF) MS setup. *(A) In a standard TOF the ion path goes straight to the detector without a focusing reflectron mirror. A high voltage Vtof(−) (negative polarity for positive ions) is applied to the orthogonal accelerator accelerating the ions toward the detector*[54] *The accelerator is orthogonal to the inlet of ions, this provides the same starting positions for each ion*[54]*. (B) A time of flight (TOF) MS setup with a reflectron mirror. In a reflectron Tof the ion path goes focusing reflectron mirror and then travels to the detector. The accelerator is orthogonal to the inlet of ions, this provides the same starting positions for each ion*[54]

1.4.3.2 Fourier-transform ion cyclotron resonance

Fourier transform-ion cyclotron resonance mass spectrometry (FT-ICR) is a high-resolution mass spectrometry technique. Developed by Comisarow and Marshall in the 1970s they deduced that the *m/z* of an ion could be calculated by measuring it's cyclotron frequency (ω) generated when in the presence of a magnetic field[55]. Ions are contained within a voltage gated cell known as a Penning trap[56]. The Penning trap sits in a spatially uniform magnetic field *(B)*. Trapped ions are forced into a circular motion perpendicular to the magnetic field and the frequency of an ion's rotation is correlated to their respective *m/z* ratio[56]. The Penning trap uses trapping plate electrodes which generate an electric field, preventing the ions from escaping axially[56]. A further two pairs of opposing electrodes surround the cells and are the excitation and detections plates respectively (Figure 1.6)[56].

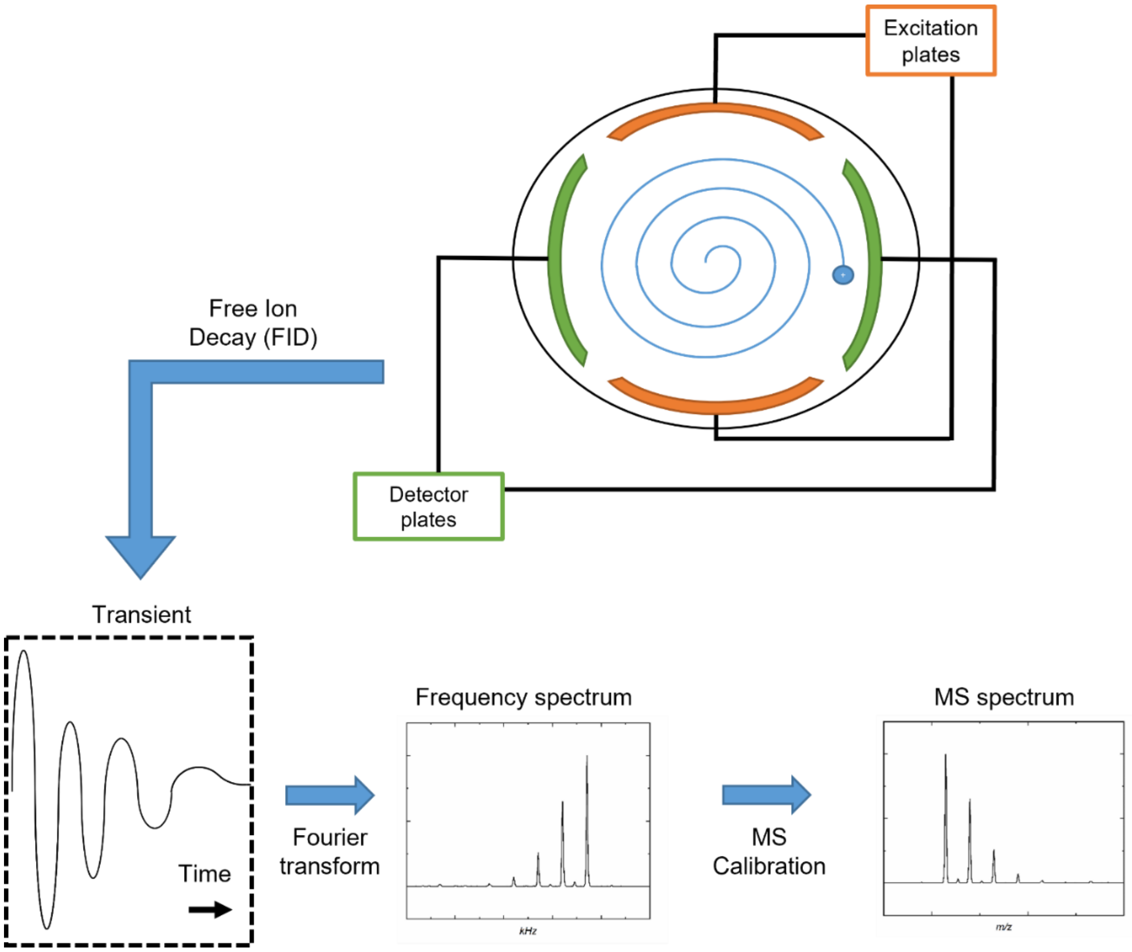


Figure 1.6: The cyclotron motion of ions is governed by the magnetic field. Ions are held axially within the cell by voltages applied to the trapping plates at either end of the ICR cell. An rf sweep is applied in the form of a chirp, this provides a range of RF values which correspond with a user specified range of cyclotron frequencies. This will excite ions into a larger orbit. As ions pass the detection plates an alternative current is generated and detected. The generated current is detected as a transient signal (the free induced decay) which in turn is converted into an MS spectra via Fourier transform

Ions moving in the presence of a spatially uniform magnetic field are subject to the Lorentz force. Mass, velocity, and charge are signified by *m, v* and *q* respectively (Equation 1.3)

$$F = m\frac{dv}{dt} = qvB$$

Equation 1.3:

The Lorentz force acts at an angle perpendicular to the plane determined by *v* and *B.* The motion of the trapped ions, in the absence of collisions, will be at a constant speed with the magnetic field bending the ion's path into a circular radius. This can be seen in Figure 1.7B.

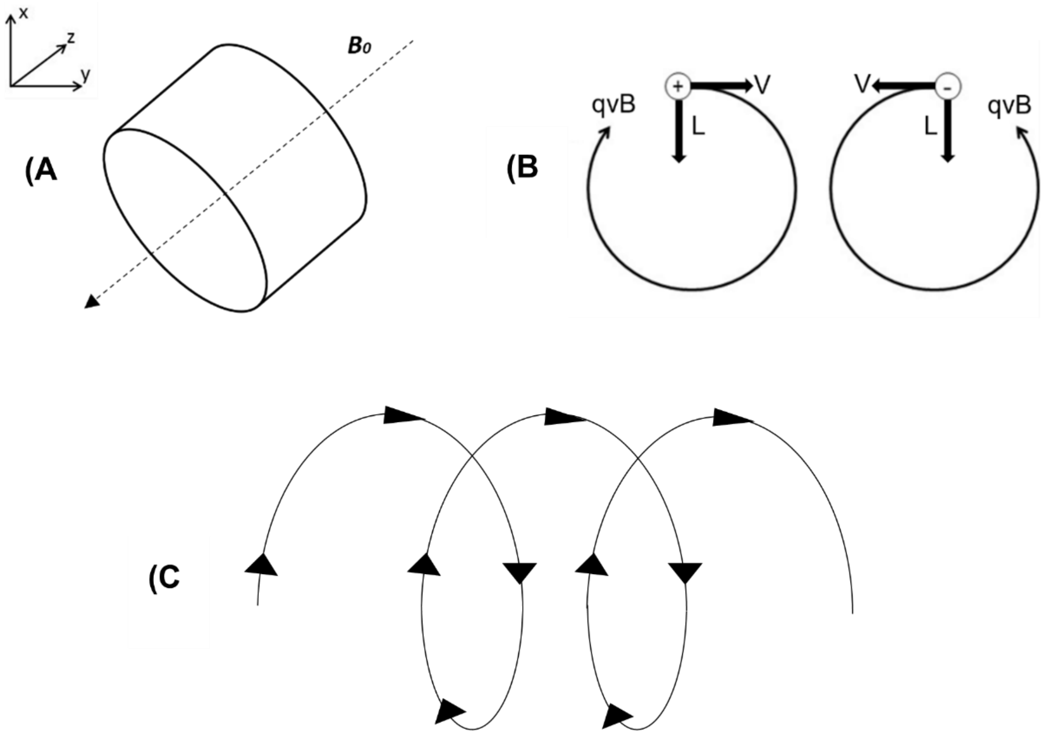


Figure 1.7: Mechanics behind Lorentz motion of ions within a magnetic field. (A) Cell orientation and direction of the magnetic field B_0 represents the ions entering the cell. (B) Showing the Lorentz motion two possible charge dependent orbits of ion clouds along an xy plane. (L) Represents the direction of the Lorentz force. (C) Circular motion around the xy plane with a defined radius

Considering the ion's velocity across the *xy* plane will be perpendicular to *B* (Equations 1.4)

$$v_{xy} = \sqrt{v_x^2 + v_y^2}$$

Equations 1.4:

And that the angular acceleration is equivalent by

$$\frac{dv}{dt} = \frac{v^2{}_{xy}}{r}$$

Equation 1.5:

A substitution can be made to Equation 1.3.

$$\frac{mv^2{}_{xy}}{r} = qv_{xy}B$$

Equation 1.6:

The angular velocity (ω) from Figure 1.7B can used to further simplify the Lorentz equation providing the angular cyclotron frequency (ωc) as shown in Equation 1.7.

$$\omega_c = \frac{v_{xy}}{r}$$

Equation 1.7:

This can be adapted into the ion cyclotron equation, (Equation 1.8).

$$m\omega_c = qB$$

Equation 1.8:

Which can be rearranged into Equation 1.9

$$\omega_c = \frac{qB}{m}$$

Equations 1.9:

Equation 1.9 reveals that the ions with the same *m/q* will share the same ICR frequency and when held in a magnetic field. Thus, an ion's ω_c frequency can be used to determine their respective *m/z.*

In order to detect this signal, an application of RF is first applied to the ions in the cell using the excite electrodes. This excites the ions orbit towards the edge of the cell, increasing their radius and bringing the ions in closer proximity to the detection plates. The orbiting ions will pass in proximity with detection plates inducing an alternating current as they travel between the detection plates generating an ICR signal or free induced decay (FID), often referred to as the transient. Fourier transform of the ICR signal produces a frequency spectrum which in turn can be converted into the mass spectrum using Equation 1.9.

The measurement of an ion's cyclotron frequency means FT-ICR MS can perform high resolution mass measurement of analytes. This is because a frequency can be measured more accurately than time. Furthermore, the frequency will provide many measurements over a single analysis.

1.4.3.3 Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) is a multi-step process comprising (i) *m/z* selection from a population of ions; (ii) fragmentation of the selected m/z ion; (iii) detection of the resulting fragment ions. This is most commonly achieved using an instrument containing two (or more) mass analysers. One example of an MS/MS setup would be a quadrupole time of flight mass spectrometer (q-TOF-MS). The 1st mass analyser is a quadrupole which is used to filter out ions based on their *m/z.* The selected ions (precursor ions) then undergo ion dissociation in a collision cell, resulting in product ions. The *m/z* values of these product ions are then measured in the second mass analyser. The product ions then detected in the TOF. A schematic overview of a tandem–qTOF instrument can be seen in Figure 1.8.

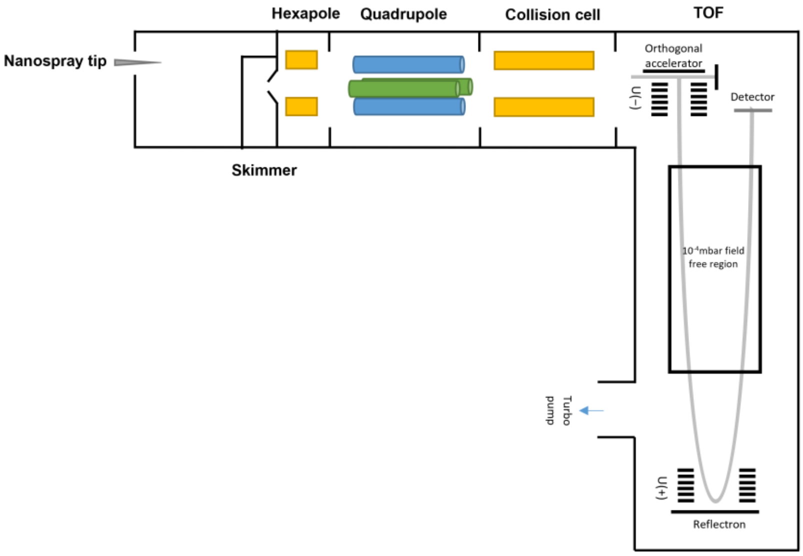


Figure 1.8: Overview of qTOF-MS. The quadrupole mass analyser filters ions before entering the TOF-MS for quantification and detection

Tandem MS/MS can also be incorporated into an FT-ICR MS workflow by exchanging the TOF-MS with the FTICR Cell (qFT-ICR MS) as shown in Figure 1.9.

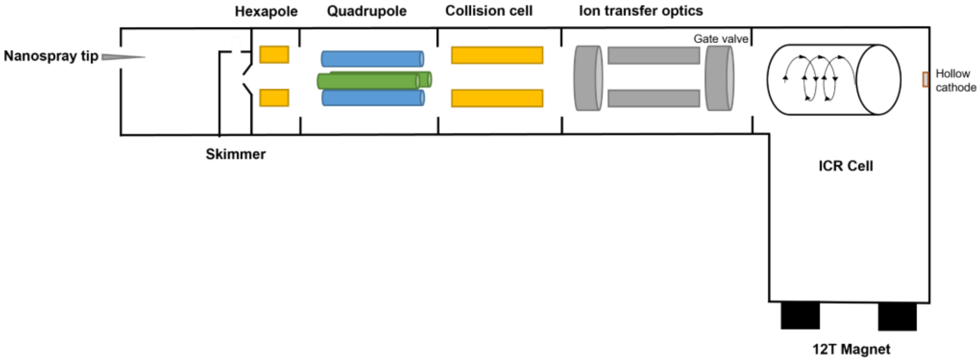


Figure 1.9: Overview of qFT-ICR-MS. The quadrupole mass analyser filters ions before entering the FTICR Cell for quantification and detection

1.4.4 Gas Phase fragmentation methods

In the bottom-up approach, a protein is cleaved into peptides via proteolytic digestion prior to LC-MS injection[57]. The peptides ion masses are measured and are then fragmented via MS/MS[57]

1.4.4.1 Protein Fragmentation Strategies

There are two primary protein fragmentation MS strategies available. Top-down proteomics which analyses fragments from intact proteins and bottom-up protein analysis, which targets peptide products of proteins after proteolytic digest (Figures 1.10 and 1.11)[57].

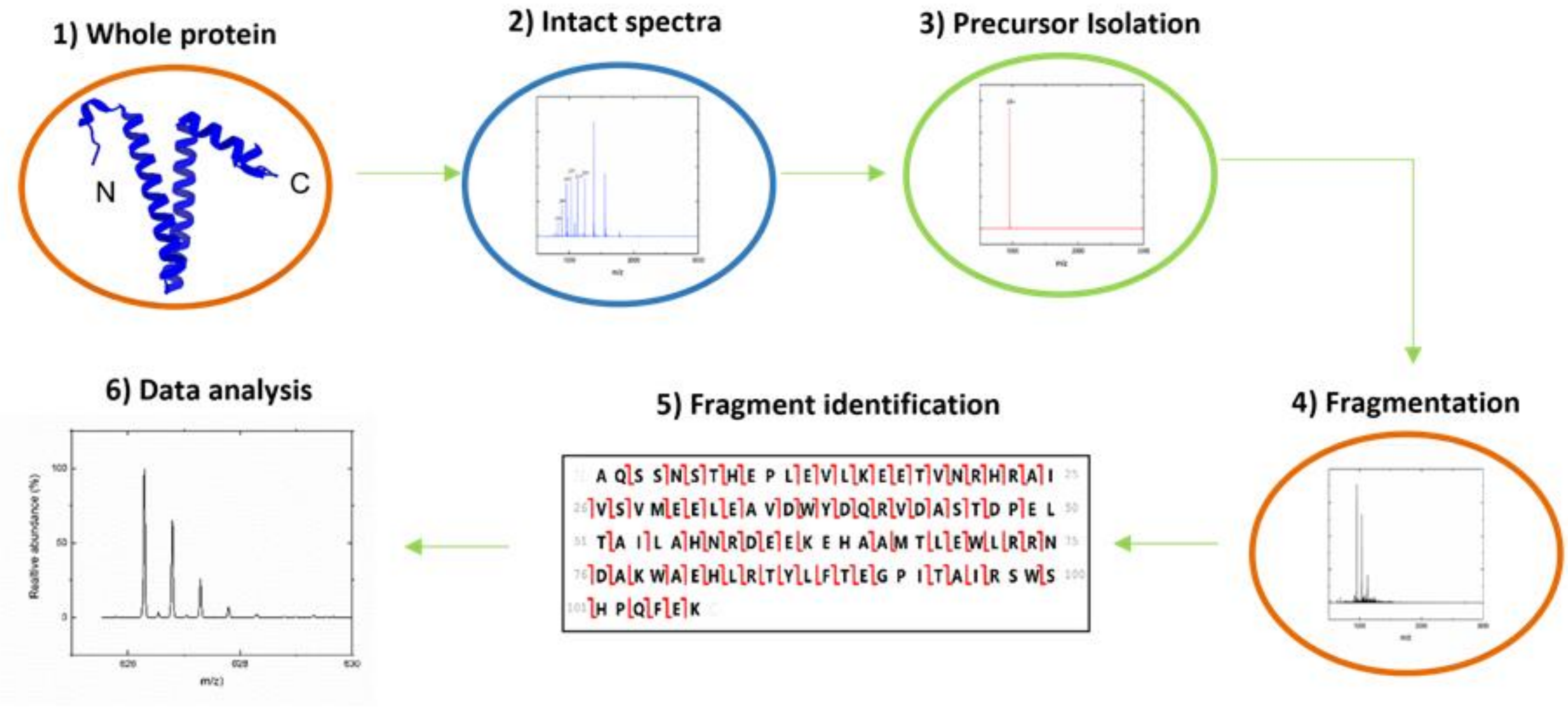


Figure 1.10: Top-down fragmentation of a protein, showing the isolation of a selected charge state, fragmentation, assignment and analysis

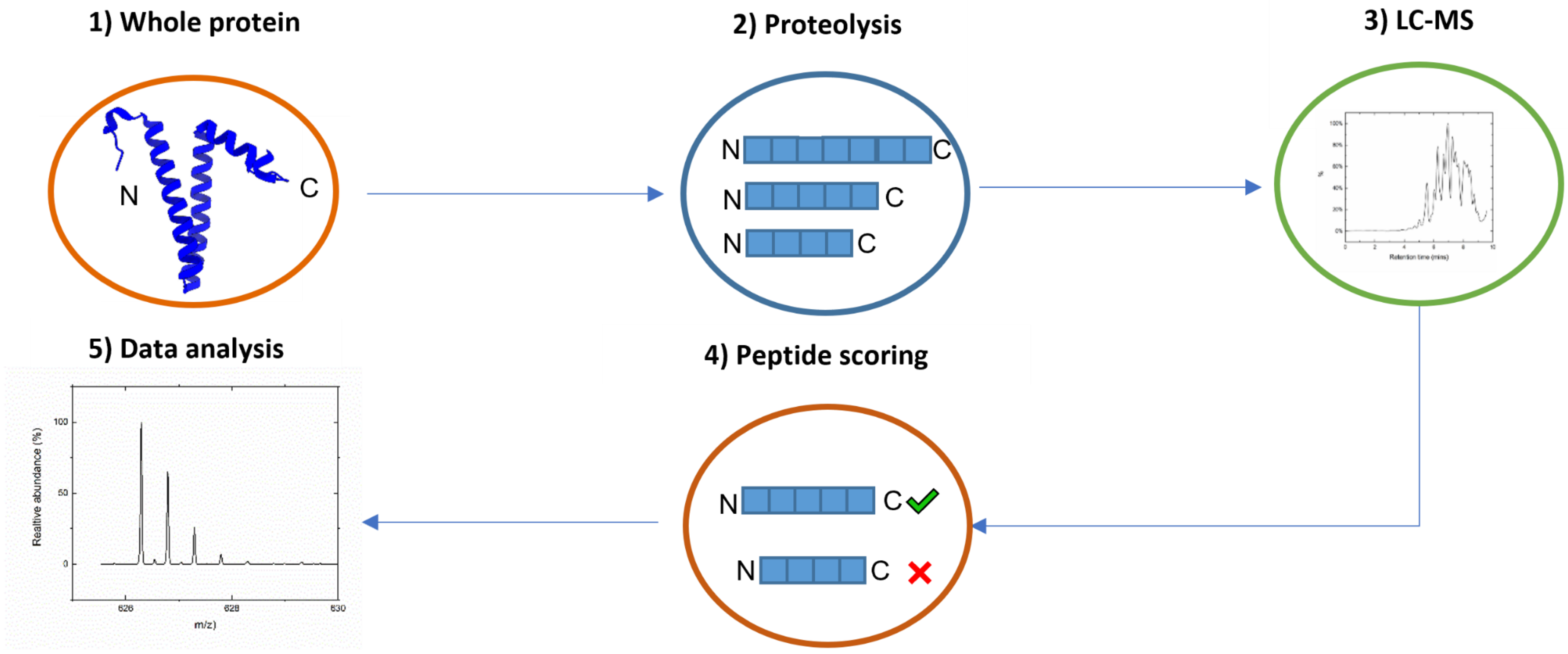


Figure 1.11: Bottom-up protein analysis, the initial proteolytic digest forming peptides. The peptides are separated by liquid chromatography. After separation, the peptides undergo MS/MS analysis to identify the mass spectra of each peptide. These spectra are used in a database search to identify the amino acid sequence of each peptide[57]

1.4.4.2 Collison-induced dissociation

Collison-induced dissociation (CID) describes the interaction of an ion colliding with a neutral heavy gas. The kinetic energy of the collision is converted into the internal energy of the ion resulting in bond vibration and subsequent bond cleavage[58]. When a peptide undergoes dissociation, the vibrational energy generated from the kinetics of the collision will break the weakest bonds first and, in the case of gas phase polypeptides, the amide bonds are the weakest bond resulting in characteristic *b* and *y* fragment ions. Because of the predicable nature of peptide fragmentation, protein and peptide sequencing via CID has become commonplace[59].

Dissociated fragment ions of polypeptides are categorized in relation to whether their charge is N or C terminal derived. If the charge is associated with the N terminal they are considered (a_n,b_n,c_n,d_n) or if C terminal (x_n, y_n,z_n,v_n,w_n). Figure 1.12 shows the most common fragment ions for dissociated peptides[58].

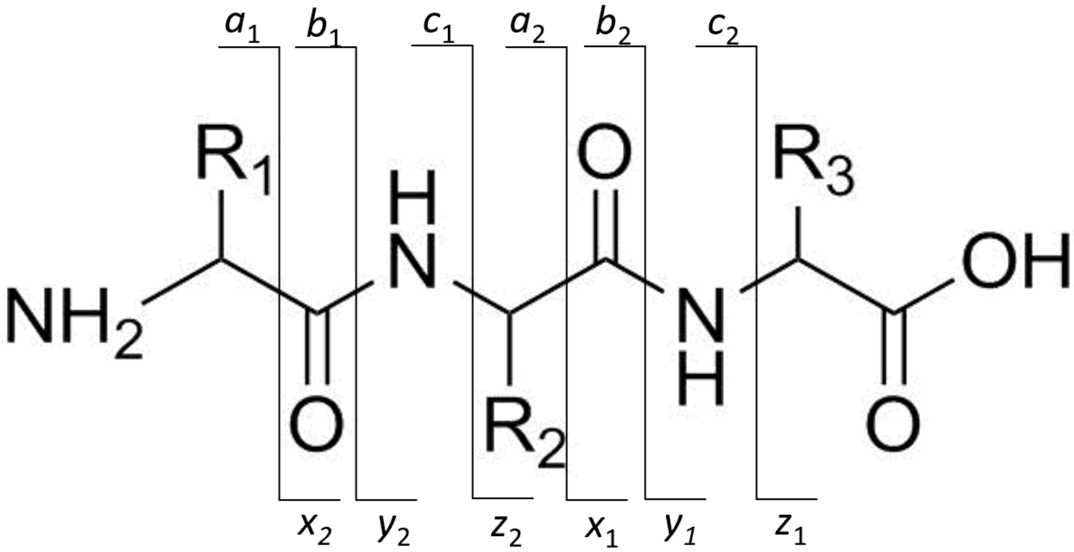


Figure 1.12: Common ions formed from the fragmentation of polypeptides

1.4.4.3 Electron capture Dissociation

Electron Capture Dissociation (ECD) is an ion activation technique which has become a commonly used fragmentation method when coupled with FT-ICR MS. First discovered in 1998, ECD allowed for the selective dissociation of large proteins, this conveyed a significant advantage over vibrationally energetic fragmentation techniques such as CID[60]. Typically, CID may result in not only fragmentation of the amides but may potentially result in internal fragments and loss or migration of any potential post translational modifications (PTMs)[61]. ECD approaches do not experience these problems as readily as it’s dissociation mechanism is considered “non-ergodic”; where the energy it exerts is not free to relax intramolecularly[61].

In the context of polypeptide dissociation, ECD fragmentation occurs by irradiating multiply protonated peptide ions with low-energy electrons (<2 eV), typically via heated filament electron gun[62]. ECD can easily compliment FT-ICR MS workflow as the ICR cell is able to introduce electrons to trapped ions. A limitation of ECD is the requirement that an isolated precursor ion must carry two or more charges to facilitate an electron capture event[62]. Fortunately, polypeptide ions generated through ESI will carry multiple charges.

The mechanism behind ECD fragmentation begins with the capture of an electron converting the charged polypeptide into an odd-electron ion $[M + nH]^{(n-1)+}$. Next the rearrangement of an $H^{\cdot}$ to a carbonyl group on the peptide backbone results in bond cleavage and *c* and *z* type fragment ions. This method of fragmentation allows for both a very high sequence coverage but also due its non-ergodic nature allows labile PTMs to be retained to a far greater degree[60]. An alternative method to ECD is electron transfer dissociation (ETD) which follows the principles as ECD; however, the electrons are generated by radical anions[63]. One key advantage ETD has over ECD is that can it be performed in an ion trap, thus negating the large instrumentation cost that comes with FT-ICR MS.

There is still some debate regarding how the process of electron capture works, with two mechanisms proposed[64]. The Cornell mechanism proposes that electron capture can occur at the protonated side chains (e.g.lysine or arginine) which will form a hypervalent radical[60]. An unoccupied orbital of NH^{3+} group captures the electron forming an electronically excited state. This results in an H transfer from the proton amide to a backbone amide carbonyl group behind the N-C bond, forming *c* and *z* ions. In contrast, the Utah-Washington proposes that it occurs directly into amide antibonding orbitals. This mechanism theorizes that the lowest unoccupied molecular orbital of the amide bond undergoes an electron capture event which precedes amide backbone cleavage. This would result in proton transfer from a protonated side chain occurring after fragmentation[65,66].

There have been various publications supporting both mechanisms[64]. For example, an observation of *c* and *z* fragmentation from metal-cationised peptides, which lack free mobile hydrogens, contradicts the Cornell method[67]. Conversely, the necessity for protons rather than metal ions as a source of positive charge does support the Cornell mechanism[68].

1.4.4.4 Data dependent acquisition

The conventional method for running LC-MS/MS is known as Data Dependent Acquisition (DDA)[69,70]. As previously stated in proteomics LC-MS, peptides will separate out over a chromatography gradient, with different species eluting out into the MS at different timepoints[69,70]. DDA will perform two scans of the data, MS1 and MS2 respectively. The MS1 precursor scan will screen eluting peptides and monitor peptide abundance to identify potential fragmentation targets. From the survey scan precursor ions are selected based upon the highest abundance[69,70]. After a precursor ion is selected it will undergo fragmentation forming product ions for further analysis (MS2). There are multiple fragmentation methods available[71] with the most common being CID[72]. A key advantage of DDA analysis is that the product ions produced can be directly associated with the selected ion from the isolation step[72]. However, this is extremely processing intensive on the MS system but also is limited by the precursor ion scan's time delay resulting in missing assignable fragments[73]. The MS and MS/MS data acquired is post processed and searched against a protein database to detect for peptide ion signatures which can be matched to the original polypeptide sequence[73].

1.4.4.5 Data Independent acquisition

Data Independent Acquisition (DIA) methods are another form of tandem mass spectrometry. There are multiple available DIA methods available such as Sequential Windowed Acquisition of All Theoretical Fragment-ion Spectra (SWATH) and MS^e [69].

MS^e removes the isolation step from the traditional DDA approach and instead dynamically alternates between high and low energy conditions within the collusion cell. This produces 'intact peptide mass spectra' (low energy) and 'fragmentation mass spectra' (high energy) from all the precursor ions within a single chromatographic timepoint[74].

SWATH relies on sequential acquisition of all theoretical fragment ions within given mass windows[75]. SWATH will divide mass ranges into separate mass windows from the intentional precursor mass. Whereas in DDA from the isolation of the most abundant precursor ion and subsequent fragmentation, SWATH will fragment all precursor ions within a selected mass window.

For a detailed overview of the DIA workflow used within this thesis' bottom-up experiments see section 2.2.3.

1.5 HDX-MS Theory and Application

HDX-MS is a technique that can be used to analyse protein dynamics and detect conformational changes within a protein's structure. It is typically used to study protein folding, protein-protein interactions and protein small molecule interactions[25]. HDX-MS relies on hydrogens found on the backbones of peptide residues readily exchanging with Deuterium, which results in an overall increase of mass within the peptide of 1.006 Da per deuterium exchange[76]. By observing the variable H→D exchange rates throughout a protein sequence it is possible to elucidate critical information regarding the protein's structure. Depending on the experiment interpretation of regions which exchange more readily can indicate regions possessing high conformational dynamics and/or solvent exposure[77]. Conversely, regions which exchange at a slow rate could indicate a potential interaction site within a protein-protein or protein-ligand binding experiment[77].

Although most HDX-MS studies focus the analysis on amide backbone hydrogens, there are several labile hydrogens found within protein sidechains - including acidic and basic hydrogens, alcohol and thiol hydrogens and amide sidechain hydrogens. However, due to the rapid forward and backward exchange profiles of side chain groups these provide limited information in a structural context[78]. Furthermore, aliphatic carbon bound hydrogens do not readily exchange within a suitable timeframe to be targets for standard HDX-MS studies[79–81]. An overview of the different classes of hydrogen can be seen in Figure 1.13.

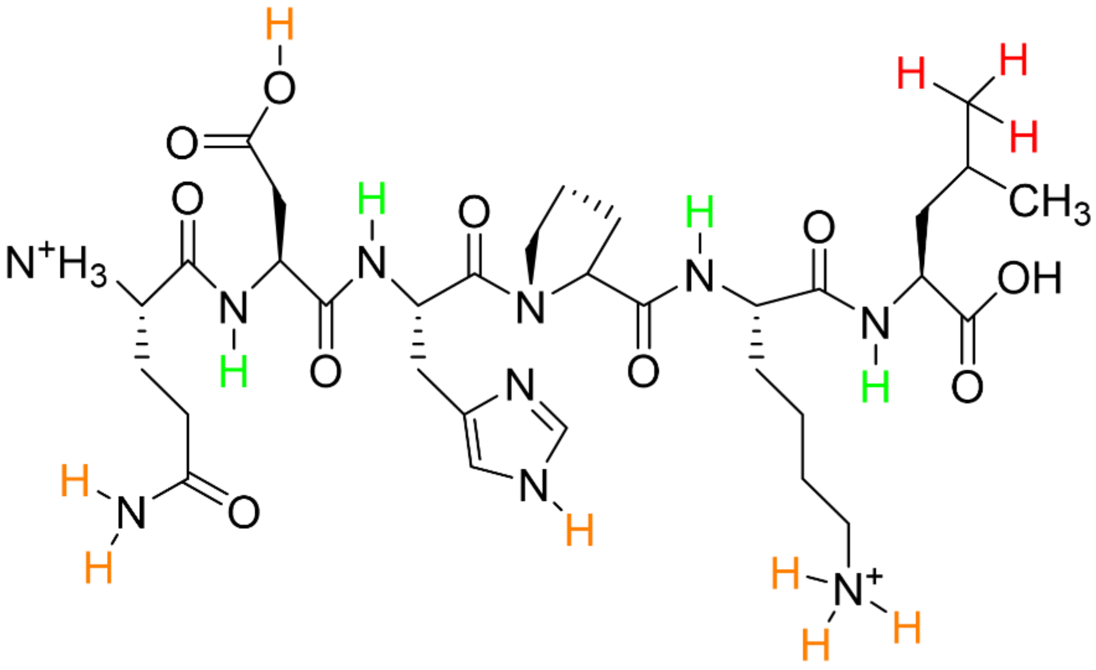


Figure 1.13: Gln-Asp-His-Pro-Lys-Leu peptide showing different hydrogen groups. The green backbone hydrogens are the traditional targets sites of HDX analysis. The red aliphatic carbon-bound–hydrogens do not readily exchange under HDX conditions. The brown hydrogens are considered to exchange too rapidly for analysis. The Gln associated free N-terminal amino group will continue to exchange even after quench as it is exposed to free hydrogens in the LC-MS solvents. The other brown hydrogens belong to amino acids with side chains that have pKa values; Asp pKa 3.5 ± 1, His pKa 6.6 ± 1. and Lys pKa 10.5 ± 1.1[81]. These values rest above the pH of the quenching solution (2.5) meaning they will continue to undergo exchange when exposed to acidic LC solvent[25]

1.5.1 Deuterium incorporation into a polypeptide

When simplified, a traditional bottom-up HDX-MS workflow consists of a protein sample being exposed to deuterium in favourable pH and temperature conditions[26]. Following labelling an acidic quenching solution is introduced to the sample to arrest the forward H→D exchange rate[26].

Following quenching samples undergo proteolytic digestion to produce peptides which are then separated and analysed via LC-MS Figure 1.14.

To determine the rate of deuterium uptake (DU) in individual peptides, the MS data from multiple timepoints need to be considered. The DU is calculated for each timepoint using the by observing the *m/z* shift in isotopic distribution increasing as more deuterium is incorporated into the peptide or fragment. To calculate the DU) of a peptide, first the average mass for an undeuterated peptide (T0) is calculated and then compared with the average mass of deuterated form of the same peptide at a given timepoint. The mass difference between the averages represents the peptide's DU at the given timepoint. The *m/z* shift can be observed in the stacked spectra in Figure 1.14.

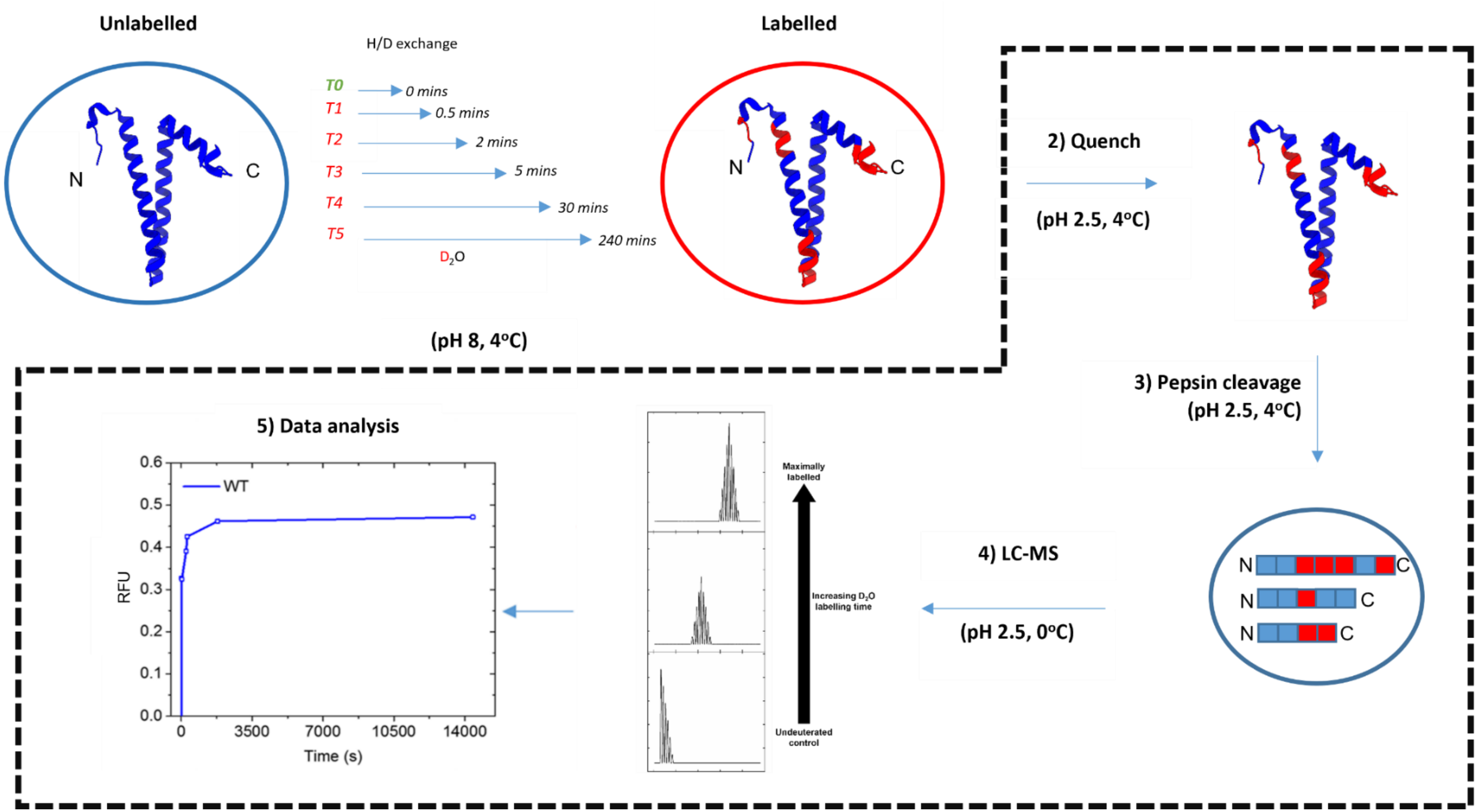


Figure 1.14: A summary of the bottom-up HDX-MS experimentation. The black dashed lines highlight the additional steps that are found in HDX-MS experimentation. Within the peptide structure the only observable sites of HDX are the hydroxyl or amine group exchange protons found within the amide backbone and the exchange rate is dependent on temperature and pH conditions making them highly quenchable[25]

The HDX experiment setup consist of a time-course analysis where the deuterium exposure time is altered between timepoints. This allows the determination of the rate of deuterium uptake. An HDX timecourse will start with a T0 or where the protein is placed into equilibration buffer (EB) which contains no deuterium. The T0 acts as a control and provides information on the protein with zero seconds of deuterium exposure. The T0 will be used to assess the level of deuterium uptake in subsequent timepoints. The length of exposure time can be altered although each timepoint, and will have a logical increase in exposure time. An example of HDX timecourse could consist of a T0, a T1 (10 second exposure to deuterium labelling buffer (LB), T2 2 minutes LB exposure, T3 three minutes of minutes LB exposure, a T4 15 minutes of LB exposure, a T5 with 2 hours of LB exposure and a Tmax with 24 hours of exposure time[82]. The Tmax is the “full exposure timepoint” where the length of exposure time is long enough that the protein rate of exchange will have reached equilibrium at the time of quenching. Once a protein is introduced to deuterium the amides within its structure will commence H→D isotopic exchange but at different rates. These rates are a consequence of variables governing an amide’s ability to exchange such as solvent accessibility and hydrogen bonding strength[82]. This is particularly useful when determining natively folded proteins which contain a significant degree of secondary structure, which is stabilised by hydrogen bonding. This closes off amides by lowering their accessibility to solvent, thus slowing their rate of exchange. It is through a combination of k_{op} (opening), k_{cl} (closing) and k_{ch} (chemical exchange rate constants) which determines the rate of exchange as shown in the Linderstrøm-Lang Equation 1.10[82].

$$N - H_{C\,losed} \quad \underset{k_{cl}}{\overset{k_{op}}{\rightleftarrows}} \quad N - H_{open} \quad \xrightarrow[D_2O]{k_{ch}} \quad exchanged$$

Equation 1.10:

1.5.2 EX1 and EX2 exchange regimes

The rate of exchange *(k_{ex})* can be derived from the Linderstrøm-Lang equation using Equation 1.11.

$$k_{ex} = \frac{k_{op}k_{ch}}{k_{cl} + k_{op} + k_{ch}}$$

Equation 1.11:

EX1 and EX2 describe two exchange profiles associated with HDX-MS spectral data. EX2 is the more common of the two profiles when observing native proteins at physiological pH[83].

Assuming the protein is in a folded native state, k_{cl} will be significantly larger than k_{op}.

Within an EX2 profile, opening and closing events are short lived (i.e k_{cl} is significantly larger than k_{ch}). Therefore, Equation 1.11 can be simplified to.

$$k_{ex} = \frac{k_{op}k_{ch}}{k_{cl}}$$

Equation 1.12:

This results in unimodal isotopic distribution that increases in *m/z* with increasing deuterium labelling times[83].

Conversely, in the EX1 regime the opening/closing events are considered to be long lived (i.e. k_{ch} is larger than k_{cl}). Therefore, the k_{ex} of EX1 will approximate k_{op}. This can produce a bimodal isotopic distribution which shifts to higher *m/z* over increasing deuterium labelling times this will continue until it forms a unimodal distribution once it is fully deuterated[83]. Examples of EX1 and EX2 exchange regimes can be seen in Figure 1.15.

Figure 1.15: EX1 and EX2 exchange regimes

1.5.3 Impact of pH and temperature on HDX

HDX exchange rates are governed by four key factors; pH, temperature, hydrogen bonding and solvent accessibility. Hydrogen bonding and solvent accessibility will be sample dependent whereas both temperature and pH can be accounted for in the experiment's design. Both of these factors can degrade the quality of exchange if not properly controlled[25]. Thus, regulating both temperature and pH are of paramount concern when devising a HDX experiment. Fortunately, the exchange of the hydrogen within the amide backbone can be either acid-catalysed or base-catalysed. This chemical exchange's rate constant (k_{ch}) can be seen in Equation 1.13. The equation consists of the intrinsic rate constants for the acid- ($k_{int,H}$), and base-catalysed reactions ($k_{int,OH}$).

$$k_{ch} = k_{int,H}[H^{+}] + k_{int,OH}[OH^{-}]$$

Equation 1.13:

There are limitations in HDX analysis, for example aliphatic and aromatic bound hydrogens will only exchange under catalysis due to their strong covalent characteristics. Certain sites will exchange too rapidly for HDX analysis; these include hydrogens from the amino and carboxy termini and side chain groups containing –SH, –OH –NH_2, –COOH, and –$CONH_2$[84]. An overview of different hydrogen exchange sites can be seen in Figure 1.13. The HDX viability of these hydrogen groups is determined by their respective pK_a values. In HDX experiments deuterium labelling is conducted in pH conditions between pH 6.5-8.5. These are ideal conditions for backbone amide exchange measurement as their hydrogens will exchange with deuterium at a measurable rate, and can be quenched to arrest the exchange. The slowing of the exchange through acid quench also minimizes the impact of D→H back-exchange after sample is exposed to LC solvents.

HDX rates are also temperature dependant. Increasing temperature will increase the exchange rates in both forward and back-exchange conditions.

This primarily a result of temperature impacting the water ionization constant, resulting in an increase in OH^- concentration. If unregulated, this will generate significant inconsistencies in exchange profiles thus inhibiting accurate analysis.

1.5.4 HDX back-exchange

Back-exchange (D→H) refers to the exchange of deuterons attached to the amide backbones with free protons in solution after quenching. HDX analysis relies on the amide backbones retaining the deuterons making uncontrolled back-exchange key source of data degradation. The loss of deuterium label inhibits detailed and comprehensive analysis of protein structure and interaction. back-exchange is often cited as the most significant challenge with solution-phase H/D exchange[85,86].

Optimization of HDX conditions with the purpose of minimization is a key area of research within the HDX community[85,86]. Many studies have helped to optimize conditions to allow for more reliable exchange in the pursuit of standardizing pH, temperature and buffer reagents[85,86].

Ideally HDX analysis would have 0% back-exchange, however it is not possible to stop the exchange reaction fully. This is due to the large number of free protons in the LC-MS reagents which will continue to exchange with the deuterated protein even under optimal conditions. For this reason, 20-30% system associated back-exchange values are considered suitable. This level of back-exchange is considered to not compromise the validity of the experiment[85].

Back-exchange is affected by the same conditions as normal exchange i.e. pH, temperature and time. Acid quenching the forward exchange (H→D) after an exposure timepoint can be used as a means of decreasing back-exchange[85,86]. By lowering the pH and decreasing the temperature, the rate of back-exchange is decreased sufficiently to allow for LC MS/MS analysis.

Equation 1.14 can be used to calculate the effect of back-exchange of a model peptide

$$back\ exchange = \left(1 - \frac{m_{100\%} - m_{0\%}}{N\ x\ D_{frac}}\right) \text{x } 100$$

Equation 1.14:

D is the deuterium uptake; $m_{0\%}$ and $m_{100\%}$ is the centroid mass for the undeuterated control and fully deuterated control respectively. N is the number of exchangeable amides in the peptide and D_{frac} represents the fraction of D/H present in the labelling buffer i.e. 0.95.

Through use of a maximally labelled control *(m_{100}),* it is possible to estimate the level of back-exchange for each analysed peptide equation 1.15[38,87,88].

$$D_{corr} = \frac{m_{expt} - m_{0\%}}{m_{100\%} - m_{0\%}}\ x\ N$$

Equation 1.15:

D_{corr} is the back-exchange corrected fractional value of a deuterium within a peptide, m_{expt} is the experimentally determined centroid mass of the peptide, m_0 is the centroid mass of the peptide in the undeuterated control, m_{100} is the centroid mass of the peptide when fully deuterated (corrected the percentage deuterium used in the experiment), and N is the number of exchangeable amide hydrogens in the peptide[87].

1.5.5 Top-down HDX-MS

Since its inception, HDX-MS has been developed into a highly versatile technique. The majority of published HDX-MS literature is reliant on the previously mentioned bottom-up HDX-MS workflow (Figure 1.14).

As time has progressed there have been many developments focused on improving the bottom-up workflow. However, there still exist several key drawbacks which are discussed below. This has meant that HDX-MS is still largely considered to be a complementary technique for other higher resolution approaches such as X-ray crystallography or NMR spectroscopy[88].

The primary drawback of bottom-up HDX-MS is its reliance on sufficient sequence coverage provided by efficient digestion. One solution to overcome this limitation would be to use top-down HDX-MS. This approach would use gas phase fragmentation, in place of proteolytic digestion, to provide greater sequence coverage at the amide level. This would not only help address the problems associated with proteolytic digestion but would also allow for higher amide level resolution. This in turn allows for the analysis of residue deuteration profiles.

One further benefit of top-down HDX-MS would be the ability to analyze two individual proteoforms within a single sample. An example of this would be the analysis of a solution containing a mixture of unmodified and PTM modified protein forms. Bottom-up HDX-MS would treat peptides from both forms as the same.

In contrast, a top-down HDX-MS approach would include the incorporation of an isolation step on specific protein ions. This will allow precursor ions of both the modified and unmodified proteins to be isolated prior to gas phase fragmentation. This would provide users greater certainty regarding the origin of the fragments being analyzed.

1.5.6 Fragmentation induced Hydrogen Scrambling

Top-Down HDX and gas phase fragmentation does hold a lot of promise. However, there is a significant amount of method optimization involved with Top-Down HDX. Vibrational fragmentation techniques such as CID can result in HDX-MS data degradation though hydrogen scrambling. Scrambling refers to fragmentation induced internal rearrangement of labile hydrogen atoms along the protein sequence[89]. This rearrangement of hydrogen atoms will also cause the deuterium labelled amino acid to shift thereby negating single amino acid resolution. Currently there is no consensus on how to overcome or minimize this, with each experiment needing suitable optimization of fragmentation conditions and isolation parameters[90]. To assess scrambling, work performed by the Rand group has resulted in a peptide probe for the occurrence of hydrogen (1H/2H) scrambling upon gas-phase fragmentation. This can be used to assess the degree of scrambling that is occurring under specific mass spectral conditions[91].

1.6 Encapsulin-Encapsulated Ferritin complexes

Due to iron’s role in several key biological functions, such as oxygen transportation and metabolism, its storage and regulation has been critical for most lifeforms. The most widely known biologically relevant form of iron is arguably heme-bound iron found within red blood cells. However, iron can also directly bind onto proteins at select amino acids such as histidine, cysteine and glutamate without the presence of a heme unit. Due to the heterogeneity of metabolic pathways, there are a vast array of protein families involved in iron transportation with one of the more prominent families being ferritins[92–94].

1.6.1 Ferritins and iron storage

Ferritins are a large family of intracellular proteins involved in iron oxidation and storage. Ferritins oxidise iron from Fe(II) to Fe(III) then store them within their spherical quaternary structure. This is an important biological function as free iron within the cellular space acts as a catalyst in the formation of free radicals a source oxidative damage impacting the health of the cell and organism[92–94].

First described in 1937 by the French scientist Laufberger and adapted by Granick, he observed that serum harvested from a horse’s spleen contained 23% by dry weight of iron and hypothesised the presence of proteins involved with iron storage within the serum. These iron storage proteins were named ferritins accordingly[95,96].

Over the subsequent years many advancements were made in characterizing ferritins in the pursuit of being able to accurately quantify ferritin from ferritin serum. This work culminated in the development of sensitive immunoassay techniques utilizing ferritin and anti-ferritin antibodies. This discovery paved the way for clinical techniques to assess the iron levels within patients to rapidly detect iron deficiencies[97].

Ferritins exist within nearly all eukaryotic and prokaryotic life and extensive research has been performed on animal, plant, and bacterial ferritins. Within eukaryotes there are two primary types of ferritins: cytosolic ferritins (CFs), which exist within the intracellular space, and mitochondrial ferritins (MFs); both perform iron storage and protect from oxidative damage amongst other regulatory roles. One example of this varied functionality can be seen in two separate downregulation studies. The first study demonstrated that MFs may confer neurological protection and demonstrated that the downregulation of MF within mice resulted in β-amyloid induced memory impairment and neuronal apoptosis[98]. Whereas a separate study has shown that down regulation of CF impacts the ability of macrophages to store and detoxify exogenously added iron resulting in oxidative stress and cell death[99]. Furthermore, Ferritins have been found to reversibly dissociate and reassemble at different pH values making alone an interesting potential nanocarrier platform[100]

These studies are a glimpse into the versatility in ferritin functions; and research into this important class of proteins is still highly active[92–94]. There are two primary types of ferritins within prokaryotes heme-containing bacterioferritins and the heme-free ferritins. Although distantly related both variants share similar structures, being hollow globular proteins with external diameters of approximately 120 Å and an internal iron storage cavity being 80 Å (Figure 1.16)[101].

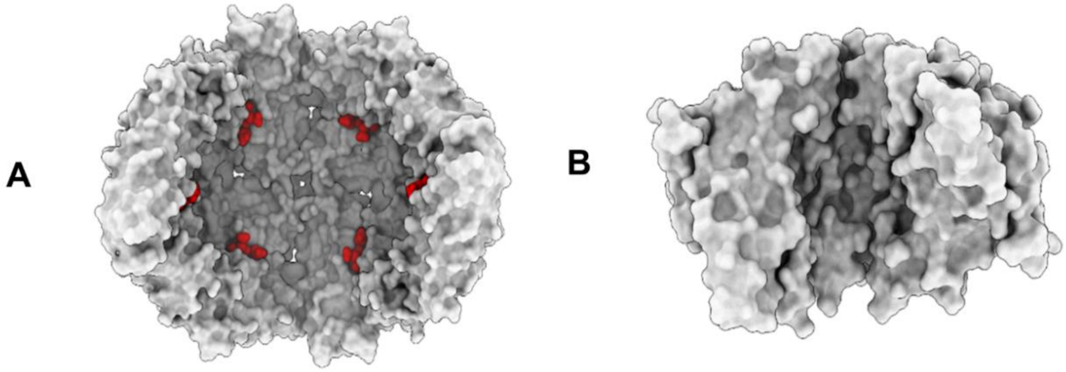


Figure 1.16: Bacterioferritin structure (A) Cross section of heme-containing Escherichia coli Bacterioferritin PDB ID: 2VXI heme units are highlighted in red. (B) Cross-section of a heme-free ferritin PDB ID: 52A5

There are many potential applications of ferritin structures and as such there is great interest in developing ferritin like structures. One hypothesised application would be to repurpose the ferritin’s hollow interior core into a novel drug transport system[102,103]. Material scientists have already begun to use the ferritin system as a template for synthetic molecular transport systems ranging from carrying chemotherapeutics[102,104] to therapeutic approaches for treating rheumatoid arthritis[102,105]. Outside of the ferritin’s interior, the exterior surface can be modified without disrupting the assembly of the ferritins structure. This work may form the basis of a targeted delivery system in which the exterior of the synthesised ferritin transport system may be modified to allow for a unique trigger mechanism in which the sequestered carrier molecule will only be released once the ferritin molecule is in the presence of an appropriate trigger. This would provide a precise method of delivery of chemotherapeutics to cancer cell targets. The primary benefit of this system would mean less damage would be caused to healthy non-cancerous cells resulting in less tissue damage and a low chance of organ failure. However, all of these potential avenues of research are reliant on a greater understanding of how EncFtns assemble[102].

1.6.2 Classical Ferritins

Classical ferritins are multimeric structures which consist of 24 polypeptide chains which configure into a polyhedron protein cage. The protein cage contains an interior pocket into which iron is sequestered[93]. For non-heme-free ferritins the coordination of iron is controlled via key iron-binding amino acids in the ferroxidase site interfaces (Figure 1.17)[106].

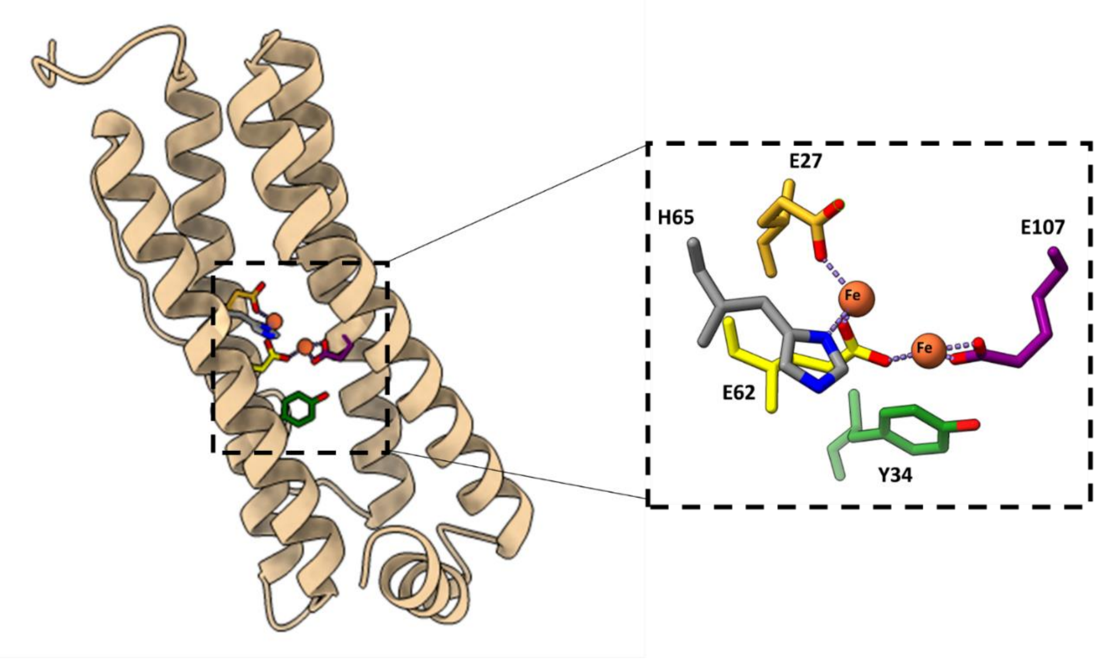


Figure 1.17: Monomer of Human Ferritin (Classical Ferritin) PDB ID:4YKD showing the coordination of iron (orange) within the feroxidase centre (FOC) by essential amino acids within a E27 (gold), E62 (yellow), H65 (grey) and E107 (purple). Y34 (green has been highlighted due as it is believed to assist iron oxidation)

It has been theorized that the structure of classical ferritins is reliant on the self-assembly of multiple protein subunits. A 2011 study demonstrated that ferritin self-assembly is pH dependent and that mutations of key residues would disrupt the assembly preventing formation of higher order structures[107].

On a cellular level ferritin play a significant role in preventing damage, ferritins are located both within the cell's nuclear and cytoplasmic environments. The responsibility for iron storage typically falls upon cytoplasmic ferritin which is composed of heteropolymers of two heavy (H) and light (L) chain subunits[108]. Once assembled, Ferritins can sequester 4500 iron atoms. The H subunit contain the ferroxidase activity which converts Fe2+ to Fe3+ for storage within the ferritin assembly. Conversely L-subunits facilitate the uptake of iron into the assembly and help stabilize its structure[108].

Ferritin-mediated iron oxidation occurs within a ferroxidase centre (FOC) however the mechanism is still poorly understood. One possible hypothesis proposes that the implementation of an oxidising agent, such as O_2, facilitates the oxidation. Other questions remain regarding whether the presence two Fe^{2+} ions is critical to the oxidation to occur. There are currently three proposed mechanisms for Fe^{2+} oxidation as shown in Equations 1.16A-C[109].

A) H_2O_2 detoxification model: $2\ Fe^{2+} + H_2O_2 + 2\ H_2O \rightarrow 2\ FeOOH_{core} + 4H^+$

B) Crystal growth model: $4\ Fe^{2+} + O_2 + 6\ H_2O \rightarrow 4\ FeOOH_{core} + 8\ H^+$

C) Protein Catalysis model: $2Fe^{2+} + O_2 + 4\ H_2O \rightarrow 2FeOOH_{core} + H_2O_2 + 4\ H_+$

Equation 1.16:

Although there is still much debate regarding whether one or a combination of these mechanisms results in oxidation, there is a consensus that the FOC region does facilitates Fe^{2+} oxidation.

1.6.3 Encapsulated ferritins

Encapsulated Ferritins (EncFtn) are a subfamily of classical ferritins that are sequestered into larger macrostructure Encapsulin capsids (Enc)[110]. Discovered in 2009, whilst performing sequence analysis of a ferritin-like protein found in *Nitrosomonas europaea* (NE0167) two distinct adjacent open reading frames (ORF) were found to be conserved. The first ORF shared sequence similarities to classical ferritin[110]. The second ORF shared an association with the large capsid protein HK97[111]. These findings suggested that there existed an association between the two gene products with the hypothesis being that the larger capsid like protein would encapsulate the smaller ferritin-like protein and thus the capsid like protein was named Enc[112,113].

Operating similarly in function to non-encapsulated bacterial ferritins, EncFtn catalytically oxidise Fe(II) to Fe(III) and the resulting Fe(III) ion is stored within a larger capsid-like Enc nanocompartment (Enc-EncFtn)[110]. EncFtns are directed into the larger Enc via a short LS motif at the C-terminus of EncFtn. Once the EncFtn has entered the interior of Enc the LS is thought to interact with the inner surface of Enc interior wall where the LS docks into a binding pocketing forming the Enc-EncFtn complex.[110,114].

Recent studies look at the sequence heterogeneity between different EncFtn proteins have found that share the similar LS motifs consisting of residues (LT/GI/VGLSK)[110,114]. This suggests the mechanism behind encapsulation is the same across the subfamily[110,115]. Along with sharing the similar localization motifs, previous work has found that EncFtns are known to assemble as a decameric ‘pentamer-of-dimers’ arrangement, and this is a common structural fold for the family (Figure 1.18)[110,116].

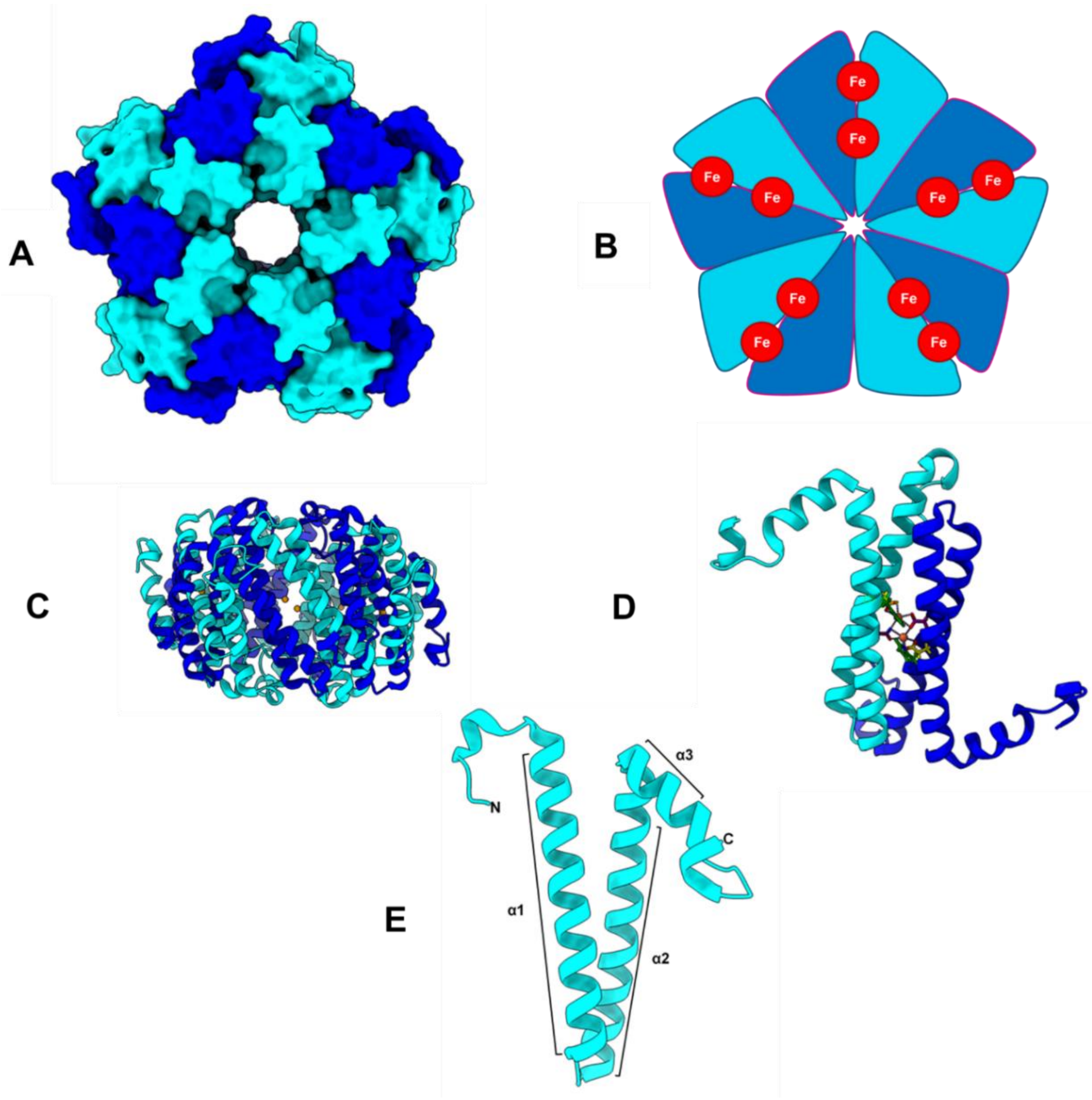


Figure 1.18: EncFtn decamer assembly (A) Graphical representation of the D3 symmetry decamer of an EncFtn derived from Rhodospirillum rubrum PDB ID: 5DA5. The ten alternating monomer subunits have been coloured blue and cyan respectively. (B) A schematic overview of an EncFtn showing the dimer subunits along with iron binding sites. (C) A cartoon representation of the decamer structure with iron molecules present (D) Dimer subunit showing the iron coordination via amino acids Glu32, Glu62 and His65. (E) A cartoon representation of the EncFtn monomer subunit showing its secondary structure

The decameric assembly of EncFtn is formed through the coordination of 10 monomers running antiparallel to one another. The secondary structure of an example EncFtn monomer can be seen in Figure 1.19B.

1.6.4 Ferritin Iron coordination and dimer recruitment

Both EncFtns and classical ferritin coordinate iron via the FOC motif. However, the conformation of FOC does differ between both. In classical ferritins, the FOC rests within the centre of its monomer. In contrast, EncFtn’s FOC is held in between a dimer subunit (Figure 1.19).

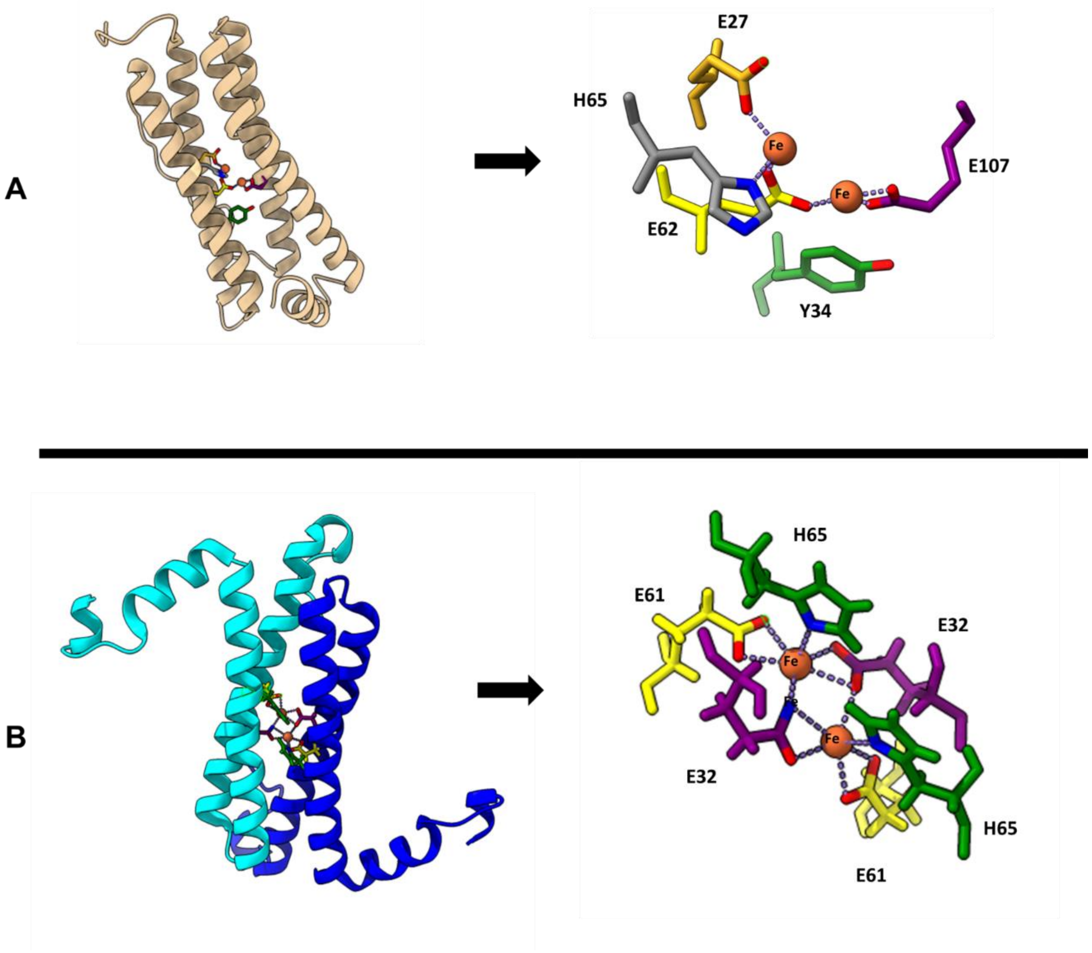


Figure 1.19: (A) Classical ferritin monomer subunit with the FOC held within its centre PDB ID:4YKD. (B) EncFtn dimer with the FOC held between two monomer subunits PDB ID: 5DA5

Considering EncFtn's FOC is reliant on the interactions between two monomer subunits there has been much debate regarding whether the FOC of EncFtn assists the recruitment of dimer subunits to form the decameric assembly. Recent work has explored the impact of altering essential amino acids involved with the FOC, finding that it dramatically decreased the presence of higher order structure above dimer suggesting that the FOC is playing a significant role[117]. For a more detailed analysis regarding dimer recruit see chapter 4.

1.6.5 Bacterial Nanocompartments

Ferritins are also utilized by prokaryotes, however, due to a lack of internal membrane-bound organelles within their cytoplasm, prokaryotes have evolved to use micro and nano-compartments to aid homeostatic and metabolic functions[118,119]. Structurally, both bacterial micro-compartments (BMCs) and bacterial nano-compartments (BNCs) are classes of large multimeric protein cages. Their primary function is to house partner enzymes and prevent the cytosol environment from being exposed to harmful by-product formed by the enzymatic reactions within their lumen[118,119]. Although sharing similar function, there are radical size differences between BMCs (1000Å) and BNCs (250Å) as shown in Figure 1.20. With advances in instrumentation, it has become possible to perform detailed structural analysis on BNCs, allowing for the study of both their assemblies and any associated interaction mechanisms. Recent studies also suggest that the Enc BNC also plays a role protecting the cytosol environment from dangerous reactions[118,119].

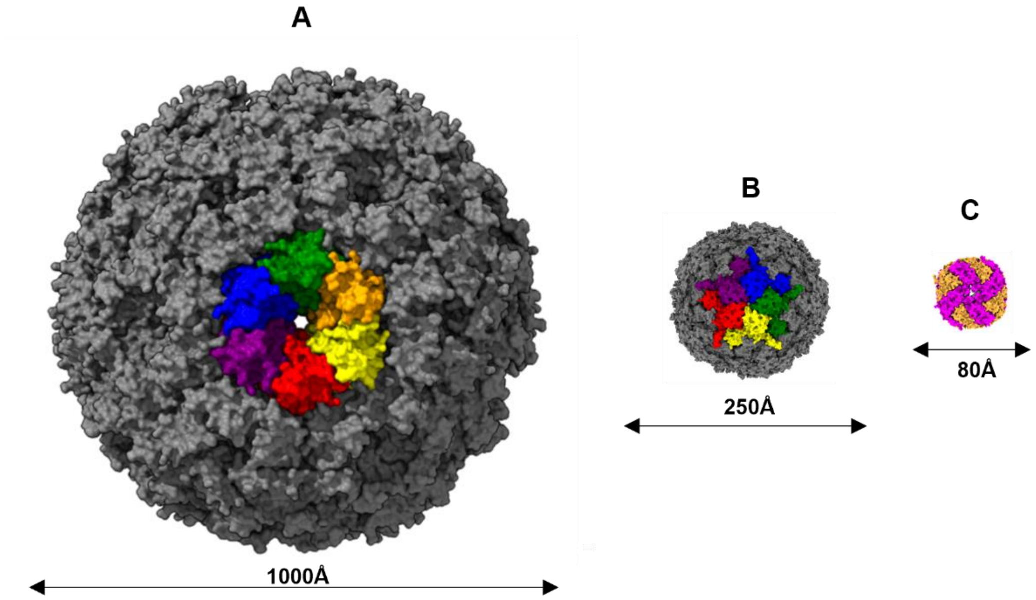


Figure 1.20: Size of BMCs, BNCs, and classical ferritin A) Carboxysome BMC with entry pore highlighted PDB ID 6OWG. (B) Homology model of H. ochraceum derived Enc with proposed 5-fold entry pore highlighted. (C) Full assembly of human ferritin PDB ID 1FHA

Encapsulins assemble from a single capsid monomer into one of three forms: 60 subunits (T=1 capsid symmetry), 180 subunits (T=3 capsid symmetry) or 240 subunits (T=4 symmetry)[120]. As each of these symmetries are generated through the same repeating subunits, the same icosahedral geometry is retained. One key characteristic that Icosahedron shapes have is the presence of 3 unique symmetries axes. These are the 3-, 2- and 5-fold symmetries respectively (Figure 1.21)[112]. As each of these symmetries represent the intersection of proteins chains, it makes them likely candidates to form poresites within Enc which sequester substrates into the interior lumen.

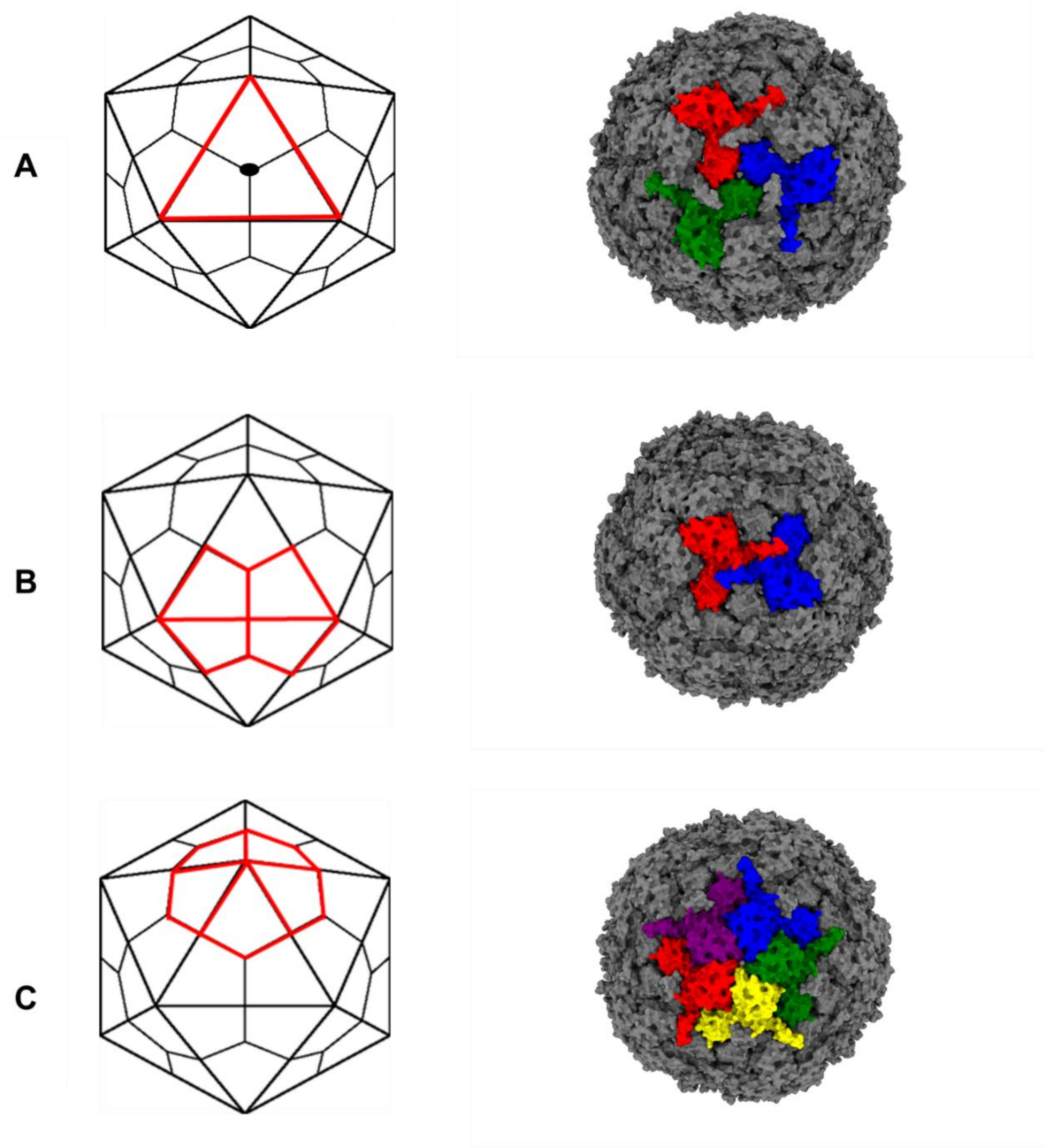


Figure 1.21: Icosahedral symmetries of the Enc BNC on the mapped onto the H. ochraceum Enc homology model. (A) Three-fold symmetry subunit. (B) Two-fold symmetry. (C) Five-fold symmetry

1.7 Structural relationship between Encapsulins and EncFtn Cargo protein

As previously stated, the structural interactions between cargo proteins such as EncFtns and the larger Enc noncompartments is an ongoing field of research[110,119,113]. However, great strides have been made in elucidating the Enc-EncFtn complex. A recent study investigating the EncFtn derived from *R. rubrum* (RruEnc) revealed the high resolution structure of the EncFtn protein and proposed a model for the interaction of the enzyme with its encapsulin protein partner (RruEnc)[110]. The study also found the RruEnc assembly consisted of 60-mer icosahedral shell with diameter of 24 nm[110] consistent with the 20-42nm size range of all encapsulins studied to date[112,121]. However, questions remained regarding the Enc-EncFtn assembly and how the EncFtn cargo protein orientates itself within the interior space. It is thought that this interplay between the Enc and EncFtn complex and the pH mediated dissociation and assembly of ferritin can be repurposed in to an efficient nanocarrier platform. Several uses have been proposed for similar Nanocarrier platforms – such development of novel transport platforms for targeted delivery of chemotherapeutics, utility as a biomarkers or as gene editing complexes (Figure1.22)[102,103,105,122]. To this end, uncovering how both the Enc-EncFtn complex interact, and their assembly is crucial to understanding how these systems can be repurposed for clinical applications[102,103,105,122].

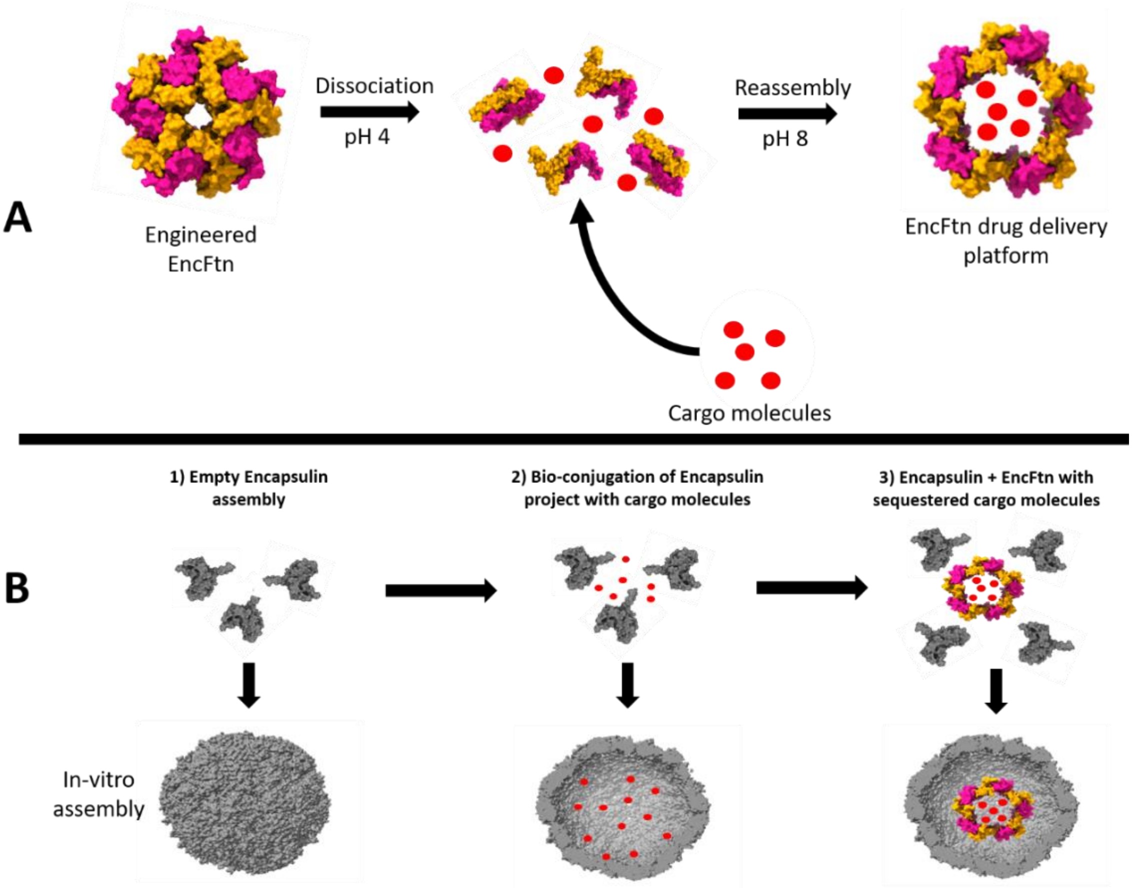


Figure 1.22: (A) Fabrication and assembly of an engineered EncFtn containing therapeutic cargo molecules through reversible dissociation and reassembly via pH regulation. (B) Proposed encapsulin based drug delivery platfroms, (B1) is an empty encapsulin assembly without any sequestered materials. (B2) is a bio-conjugated encapsulin where the therapeutic agents are retained on the Enc protein's structure. (B3) is the Enc/EncFtn delivery vehicle with the therapeutics agents held within the EncFtn which in turn is sequestered into the larger enapsulin

1.7.1 EncFtn cargo protein localization

Prior work by Altenburg *et al,* has explored peptide-shell interactions in Enc nanocompartments through the implementation of force-field modelling and particle fluorescence measurements. This was performed on two Enc complexes, *Thermotoga maritima* PDB ID:3DKT and *Myxococcus xanthus* PDB ID:4PT2[123]. It was found that coordination of C terminal targeting peptides are essential in the encapsulation process and highlighted key binding regions, along the interior Enc surface. These binding regions correspond to the LS on the EncFtn cargo protein anchoring it in place. These initial findings allude to a universal system for cargo protein sequestration across the Enc family[117]. Although promising, the findings lacked site specific data confirming the conformational interactions between cargo protein and the Enc interior. Furthermore, due to the reliance of fluorescence on labelled protein, it is possible that the label may inhibit the sequestering of cargo protein and alter conformational pathways within the interior space[124,125]. For these reasons, there are grounds to use alternative analytical techniques to further explore Enc-EncFtn protein interactions.

There has been some evidence that the coordination of the cargo enzymes into the interior of the Enc is mediated via a N-termini localization sequence (LS) which directs and binds the cargo enzyme to the interior walls of the Enc[126,127]. Furthermore, recent preliminary cryo-EM studies from our research have shown the tetrahedral arrangement of four EncFtn cargo protein within the interior space of an encapsulin (Figure 1.23)[128].

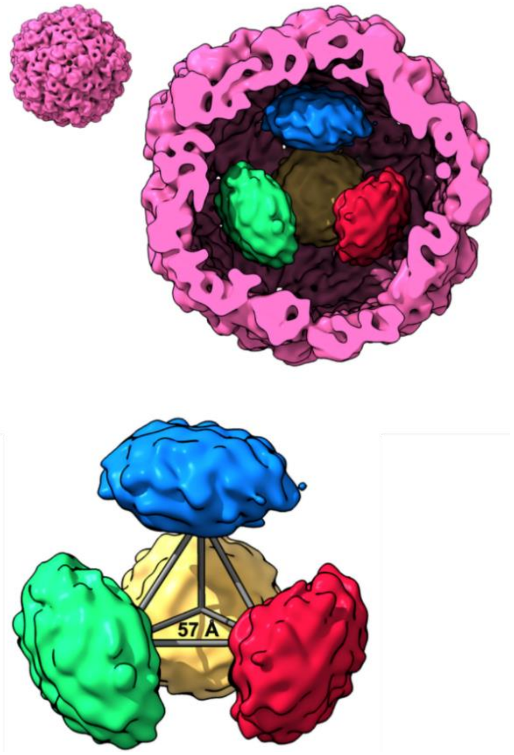


Figure 1.23: Cryo-EM imaging of the tetrahedral arrangement of EncFtn cargo proteins sequestered within the interior space of Enc, image adapted from Jenn Ross et al[128]

The Cryo-EM data suggests that the coordination of the EncFtn cargo protein is governed by both the steric effects of the Enc interior surface residues and the fixed number of EncFtn decamers sequestered within. This implies that the loading of the Enc nanocompartment is limited by the steric effect of both engaged and free Enc binding pockets and EncFtn's LS (Figure 1.24)[128].

One hypothesis posits that the cargo protein orientates itself in front of one of the potential poresites (5-Fold) of the Enc shell[110]. This hypothesis proposed a model where the Enc-EncFtn complex was reliant on the unique structural characteristic of the EncFtn cargo protein and the orientations of its FOC active site to align is to cargo binding sites on the interior of Enc surface.

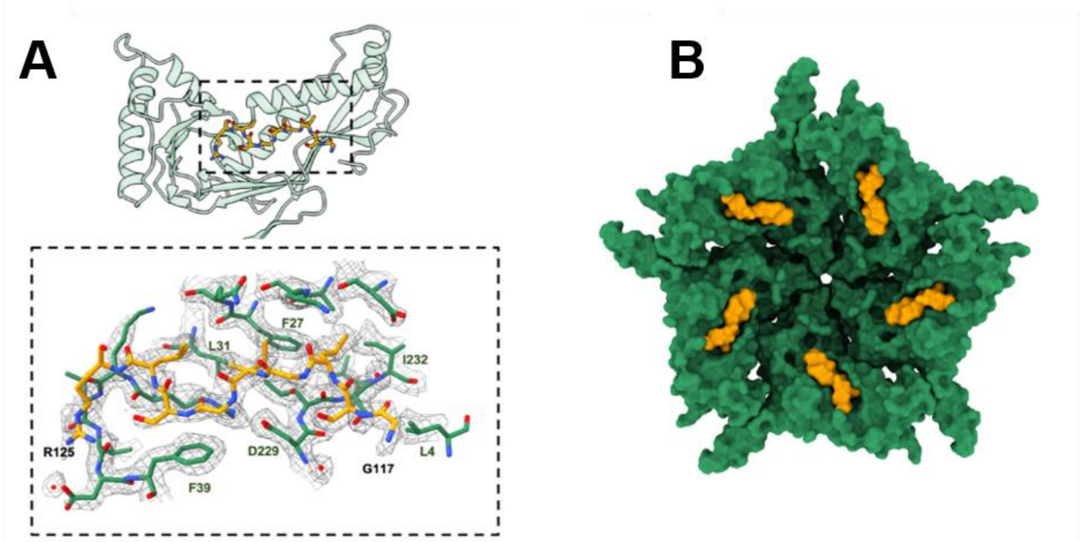


Figure 1.24: Enc binding pocket and EncFtn Localization Residues from Haliangium ochraceum. (A) Localization residues from EncFtn (Yellow sticks) binding to the interior wall of the Enc nanocompartment residues (Green stick). Lower panel is the Enc binding pocket for the LS. The Enc starting residue (G117) and the end residue (R125) are highlighted in black. (B) The closed conformation pentamer (green) with a LS (yellow) shown on each monomer. Image adapted from Jenn Ross et al[128]

1.8 Project aims

The aim of this investigation is to explore and expand upon prior studies whose focus has been on the structural interaction and assembly of Enc and encapsulated ferritins through the implementation of HDX-MS workflows. The HDX results provided crucial data confirming the EncFtn assembly pathways are reliant on retaining its functional FOC centre. Furthermore HDX-MS also confirmed of the dynamic behaviour of a key poresite located within the 5-fold poresite of Enc’s icosahedral structure. Finally, HDX analysis provided data on a potential binding interaction between the interior wall of the Enc macro structure and sequestered EncFtn cargo proteins.

The scope of the project was further expanded to develop novel Top-down HDX-MS methodologies. This study included development of an LC-Top-down

HDX workflow as a means of simplifying data complexity. The incorporation of the LC step removed the rapidly exchanging side chain hydrogens which are a source of increased data complexity.

The use of isotopic depletion (ID) to overcome deuteration induced signal suppression was also explored. ID reduces the isotopic ratio within both precursor and fragment ion's isotope distribution. This results in a significant boost in the monoisotopic peak and a narrowing of the distribution. ID had a further benefit of increasing the number of assignable fragments which were detectable after deuterium labelling. This represents a potentially leap in Top-down HDX methodology and could increase its viability as an alternative HDX-MS approach.

In simpler terms, the aims of this project were to use HDX-MS to explore the Enc-EncFtn complex, whilst also developing novel top-down HDX methodologies for this purpose.

Chapter 2 Materials and Methods

All chemicals were supplied by Sigma-Aldrich and Fischer Scientific. Solvents were all LC-MS grade and purchased from the previously stated companies. All Enc / EncFtn samples where provided by Dr Jennifer Ross detailed protein preparation and purification protocols can found in the provided references[110,117].

2.1 Sample preparation

In all HDX-MS experiment three buffer solutions were prepared. Equilibration buffer (4.7 mM K_2HPO_4, 0.3 mM KH_2PO_4 and 100 ml H_2O), adjusted to pH 8.0 with formic acid. Labelling buffer (4.7 mM K_2HPO_4, 0.3 mM KH_2PO_4 and 90.55 ml D_2O), adjusted to pH 8.0 with DCl. Quench buffer (50 mM K_2HPO_4, 50 mM KH_2PO_4 and 100 ml H_2O) adjusted to pH 2.3 with formic acid. For results chapters 3 and 5 the labelling and equilibration buffers final pH values were adjusted to pH 8.5.

32 µM protein solutions were prepared in equilibration buffer. Timecourse experiments consisted of 4-6 timepoints: T0 (0 minute; undeuterated control), T1 (0.1 minutes), T2 (0.5 minutes), T3 (2 minutes), T4 (5 minutes), T5 (30 minutes) and T6 (240 minutes) unless otherwise stated, with each timepoint being performed in triplicate[88]. Sample preparation consisted of 5 µl protein solution, 57 µl equilibrium buffer (T0) or labelling buffer (t1-6). Therefore, the final concentration of deuterium during the labelling step was 91.2%. Exchange was allowed to proceed at 4^{o}C. To arrest the exchange reaction, 50 µl of quench buffer was added to this initial solution just prior to sample injection into the LC-MS.

2.2 Bottom-up HDX-MS experimental conditions

LC-MS experiments were performed on a Synapt G2 MS system coupled to an ACQUITY UPLC M-Class UPLC with the HDX manager module (Waters Corporation, Manchester, UK). For improved reliability and precision, a custom-built CTC analytics automated platform was utilised in all sample preparation and injections prior to HDX-MS analysis.

2.2.2 Bottom-up LC-MS conditions

After injection, samples underwent proteolytic digestion on a 2.1 × 30 mm Waters Enzymate BEH pepsin column for 3 minutes at 200 µL/min. After digestion, the peptide digest was loaded on a 2.1 ×5.0 mm Acquity BEH C18 VanGuard 1.7µm C18 trapping column to pre-concentrate the sample for 3 minutes at 200 µL/min. Following trapping, the digests were separated through a 2.1 × 5.0 mm Acquity BEH 1.7 µm C18 analytical column prior to MS/MS analysis via the Water Synapt G2 MS system equipped MassLynx v4.1 software (Waters Corporation, Manchester, UK). The separation gradient was 5-95% acetonitrile with 0.1% formic acid over 10 minutes at 40 µl/min. Both the trapping and LC separation were performed at 1^{o}C to minimize back-exchange.

The Synapt G2 was calibrated with sodium formate using the Acquity's intellistart software. To apply mass accuracy correction, Leucine Enkephalin (LeuEnk 200 pg/µl, Sigma) was used as a lock mass flowing at 10 µl/min.

The Synapt G2 has configurable ionization modes, ESI source voltages, temperature control, gas flowrates and Acquisition methods (Table 2.1). The experimental MS settings were adapted from the HDX System Suitability Test by Waters Corporation[27].

Setting	**Configurations**
Ionization mode	ESI positive
ESI source voltages (V)	Capillary 2500 Sampling cone 30 Source offset 30
Temperature control (^oC)	Source 80 Desolvation 250
Gas flowrates	Cone gas (L/h) = 100 Desolvation Gas (L/h) = 80 Nebuliser (Bar) = 6.5
Acquisition method	MS^e continuum

Table 2.1: Synapt G2 configurations for the bottom-up HDX-MS workflow

The G2 flight tube was set to resolution mode and a lock mass of 200 pg/ul of Leucine Encephalin (556.2771 Da) was injected into the MS at 10 ul/min for system real-time recalibration by correction of *m/z* shifts[27]. The temperature of the source was set to 80^oC with the desolvation (nitrogen) gas heated to 250^oC. The Acquisition mode was set to MS^e to allow for post processing via the PLGS and DynamX software respectively. The acquisition time was set to 10 minutes with a mass range of 50 Da to 2000 Da with a scan time of 0.3 sec.

2.2.3 Bottom-up HDX-MS Data Analysis

Bottom-up post-processing and HDX quantification was performed using ProteinLynx Global Server 3.0.3 and Dynamx 3.0 software using processing parameters detailed in the waters HDX system suitability test document[27]. ProteinLynx global server (PLGS). PLGS reads DIA collected by Water unique MS^e workflow[74]. Once data acquisition was completed the raw data files were processed by ProteinLynx Global SERVER (PLGS) which scrutinizes the data by three different algorithms[74]. The first is the Apex3D intensity filter which removes any ion species below a configurable intensity threshold[75]. Next, PEP3D identifies ions which represent charge states and isotopes of a peptide and collates them into an exact mass retention time pair (EMRT)[75]. PEP3D then removes all EMRTs that have an intensity below a configurable intensity threshold. Ion accounting or identityE is the final algorithm and is used for protein identification. Ion accounting runs the Raw. MS data against the online protein database or against a custom FASTA file to elucidate which peptides and proteins are present in the data[74].

PLGS has configurable settings can identify charge states and isotopes of a peptide. The settings for all Bottom-up HDX-MS experiment within this project were adapted from the HDX System Suitability Test (Waters, Milford, MA)[27]. The PLGS system has two configurable parameters; processing and workflow respectively. The .raw files containing the spectral data of the T0 repeats are taken from data analysis and are upload to the PLGS software for processing (Tables 2.2 & 2.3).

Setting:	**User designated parameter**
Chromatographic peak width	Automatic
MS TOF resolution	Automatic
Lock Mass	556.2771 Da (*Leucine enkephalin*)
Lockmass window	0.25 Da
Low energy threshold	135.0 counts
Elevated energy threshold	30.0 counts
Intensity threshold	750 counts

Table 2.2: The list of processing parameters filled out to the specification for the analysis of myoglobin adapted from the on the HDX System Suitability Test by Waters Corporation. The Lock mass of for Leucine Encephalin was entered with lock mass window of 0.25 Da. The low and energy thresholds for the MS^e acquisition were entered at 135.0 counts and 30.0 counts respectively

Setting:	**User designated parameter**
Databank	Protein sample Pepsin sequence Protein sample Shuffle Pepsin shuffle
Peptide Tolerance	Automatic
Fragment Tolerance	Automatic
Min fragment ion Matches Per Peptide	2
Min fragment ion Matches Per Protein	7
Min Peptide Matches Per Protein	3
Maximum Hits to Return	20
Maximum Protein Digest	250000
Primary Digest Reagent	Non-specific
Secondary Digest Reagent	None
Missed Cleavages	1
Varible Modifer Reagents	Oxidation M
False Discovery Rate	100

Table 2.3: The list of workflow parameters filled out to the specification for the analysis of myoglobin adapted from the HDX System Suitability Test by Waters Corporation. Databank refers to the generated FASTA file against which the MS data is screened. Minimum fragment ion Matches per peptide was set to 2, Minimum fragment ion Matches per protein was set to 7 and Minimum peptide Matches per protein was set to 3. The primary digestion was set to non-specific as this projected used pepsin as a proteolytic enzyme

Deuterium incorporation was calculated by DynamX 3.0 (Waters, Milford, MA). Here the scored peptides from the PLGS analysis are compared to the deuterated forms of the same peptides found within the respective labelling timepoint (T1-6). DynamX identifies assignable peaks which are likely to be associated with a deuterated isotopic distribution of the undeuterated peptide provided by the PLGS scoring list. Processing parameters were taken from waters HDX System Suitability Test (Table 2.4)[27].

Settings:	**Parameter:**
Minimum intensity	3000
Minimum product ions per amino acid	0.3
Maximum MH^+ error	5 ppm
File threshold	3 (out of 4)

Table 2.4: DynamX Processing parameters

2.3 Top-Down HDX-MS experimental conditions

Top-down HDX-MS experiments were conducted on a SolariX 2XR 12 T FT-ICR mass spectrometer. Experimental parameters are shown in Tables 2.5-7.

Sample preparation from section 2.1 were integrated without alteration into the Top-Down HDX-MS workflow.

Top-Down LC-MS flowrates, temperatures and timing remained the same to allow for comparison to bottom-up analysis. The pepsin column was replaced with 70um silca tubing leading directly to the trap column. An Acquity BEH C4 VanGuard 1.7 µm C4 Trapping column was installed to pre-concentrate the sample for 3 minutes at 200 µL/min. Following trapping, protein samples through a 2.1 x 5.0 mm Acquity BEH 1.7 µm C4 analytical column prior to FT-ICR MS analysis.

Source Optics Settings	**Configurations**
Capillary exit (V)	210
Deflector plate (V)	160
Funnel 1 (V) Funnel 2 (V)	100 6
Skimmer1 (V)	20-60

Table 2.5: Source optic settings

Gas phase Fragmentation settings	**Configurations**
Fragmentation method	ECD
Bias (V)	1.2
Lens (V)	15
Pulse length (ms)	10

Table 2.6: Fragmentation settings

Acquisition settings	**Configurations**
Q 1 Isolation (*m/z*)	955.6
Mass selection window (*m/z*)	10
Acquired scans	2
Ion accumulation time (secs)	0.600
Transient data size	2 Mword

Table 2.7: Synapt Acquisition settings

2.3.1 Top-down HDX-MS Data Analysis

The top-down post-processing was performed using DataAnalysis software (Bruker Version 4.4) and ECD fragment assignment was performed by Autovectis Pro (version 8.8). Developed by the Kilgour group, Autovectis process the transicent signal from FTICR data to generate a FTICR spectrum[129–132]. Autovectis will then autocorrect for the “phase shift”, typically found within the FTICR spectrum, to produce an absorption mass spectrum.

From this corrected data Autovectis can identify and assign fragments forming comprehensive fragment databases[129].

Recent updates to the Autovectis software allow users to define the isotopic abundances of samples. This setting allows for the analysis of isotopically depleted protein and assignment of subsequent product ions after ECD induced fragmentation[130].

HDX quantification was performed manually with relevant equations detailed in section 6.3.4.

2.4 Calculating Deuterium incorporation of a polypeptide

To determine the rate of deuterium uptake in individual peptides, the MS data from all 6 timepoints were considered. To calculate the deuterium uptake (DU) of a peptide, first the average m/z for an undeuterated peptide (T0) is calculated (Equation 2.1). Next, the DU is calculated, for each timepoint using the equation below (Equation 2.2). With the DU of both forms of peptide at a given timepoint calculated, the relative fractional uptake (RFU) of deuterium uptake was calculated (Equation 2.3).

$$average\ m/z = \frac{\sum(\frac{m}{z})\ xA}{\sum A}$$

Equation 2.1: m represents the mass, z the charge and A is peak area.

When this equation was applied to the undertreated control (*T0*) and a deuterated timepoint the difference in average mass can be used to calculate the uptake of deuterium at that timepoint. Equation 2.2 demonstrates how to derive DU from average masses.

$$DU = \ (average\ m/z\ (tX) - \ average\ m/z\ (t0))\ \times\ z$$

Equation 2.2:

In Equation 2.2, tX represents the deuterated time point, t0 represent the undeuterated control. This can be further processed to assess the relative fractional uptake (RFU) which also considers the number of potential exchangeable residues within a select peptide (Equation 2.3).

$$relative\ fraction\ uptake\ (RFU) = \frac{DU}{number\ of\ exchangeable\ sites}$$

Equation 2.3:

Relative Fractional uptake (RFU) is the ratio of exchanged hydrogens to all exchangeable hydrogens in the given peptide. This represents the percentage exchange that has occurred against the theoretical maximum number of exchangeable sites within the peptide. RFU is calculated by dividing the deuterium level (in Da) by the total number of backbone amide hydrogens that could have become deuterated.

In differential HDX experiments comparing two protein states, it is possible to use the RFU difference (RFU^{diff}) value. This value consequently describes the susceptibility of a given protein state to HDX. The equation used to calculate the RFU^{diff} can be seen in Equation 2.4.

$$RFU^{diff} = RFU^{A} - RFU^{B}$$

Equation 2.4:

RFU^{diff} represents the difference in fractional uptake in deuterium. RFU^{A} represents the RFU of deuterium in protein state A and RFU^{B} RFU^{A} represents the RFU of deuterium in protein state B.

Chapter 3 Validation of the HDX-MS automated platform

3.1 Validation experiments

A custom-built automated platform was used in both the top-down and bottom-up HDX-MS. It was responsible for providing a high degree of accuracy and reproducibility during sample preparation and handling. The following are a series of experiments to validate the automated platform.

3.1.1 Manual HDX-MS analysis of aMb- and hMb-Myoglobin

Myoglobin is a monomeric protein consisting of 153 amino acids. It is an essential oxygen-binding protein found in muscle tissue and is responsible for oxygen storage in muscle cells. Structurally myoglobin consists of alpha helices surrounding a Fe^{2+} heme group which acts as the binding site for oxygen (Figure 3.1)[133].

The heme group within myoglobin is bound between two histidines (His) at positions 64 and 93 in the sequence[134]. This manual HDX study focused on comparing two forms of myoglobin, Holomyoglobin (hMb) which contains an heme group and Apomyoglobin (aMb) which lacks it[1]. The objective of the study was to identify the regions within myoglobin’s structure that are affected by the loss of the heme group.

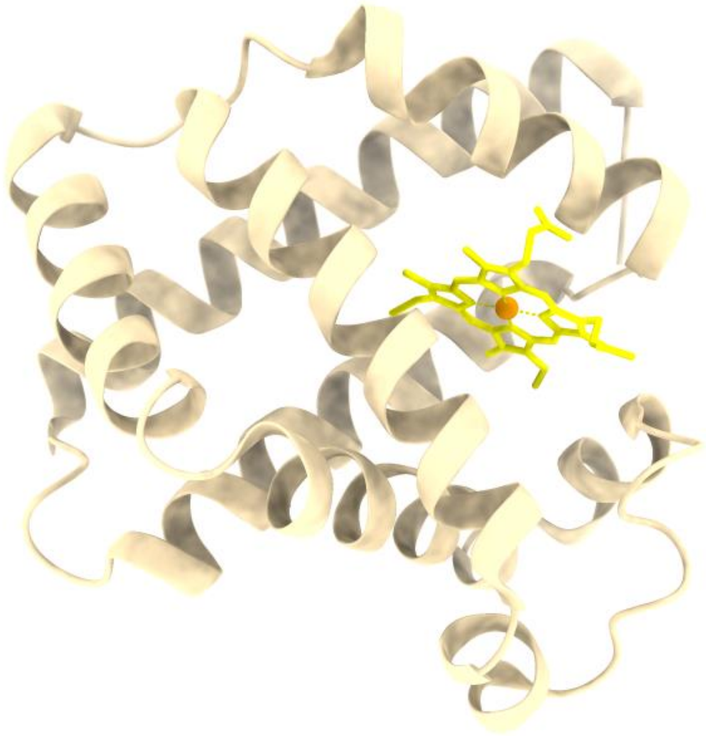


Figure 3.1: Myoglobin derived from equine heart containing the heme unit (yellow), generated using PyMol PDB ID: 1WLA

Two manual HDX timecourse experiments (0s, 30s, 2mins, 5mins, 30mins and 240mins) were performed to create an HDX profile of myoglobin with and without a heme group.

After on column pepsin digestion the myoglobin peptides were separated and analysed by LC-MS, the ion chromatograms obtained for the T0 are shown in Figure 3.2.

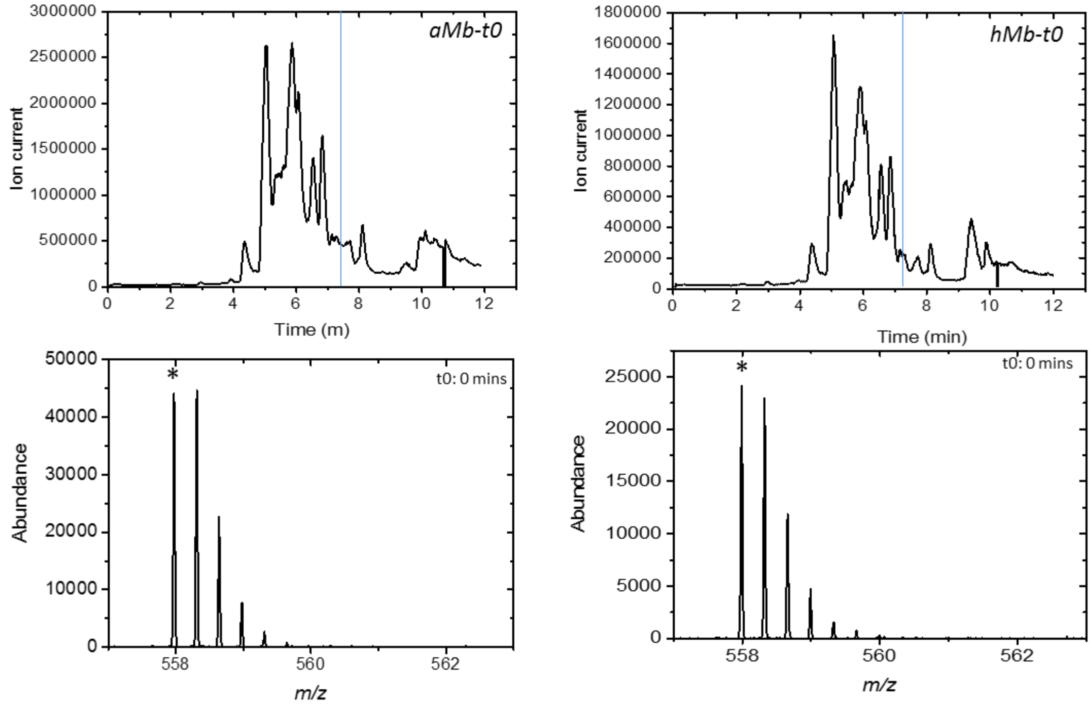


Figure 3.2: Total ion chromatograms for both hMb and aMb digests. The spectra of the [M+3H]$^{3+}$ ^{138}FRNDIAAKYKELG^{150} peptide is shown below their respective chromatograms. The Blue line signifies the 7.30-minute retention time of the ^{138}FRNDIAAKYKELG^{150} peptide

From the PLGS/DynamX data analysis of hMb and aMb, 21 peptides were identified in both timecourses. The sequences and their start and end points are shown in Table 3.1 & Figure 3.3.

Protein	Start	End	Sequence	MHP	RT	RT SD
Myoglobin	1	8	GLSDGEWQ	891.38	6.98	0.00
Myoglobin	12	19	NVWGKVEA	902.47	6.53	0.01
Myoglobin	12	20	NVWGKVEAD	1017.50	6.37	0.18
Myoglobin	12	29	NVWGKVEADIAGHGQEVL	1921.98	8.12	0.00
Myoglobin	14	20	WGKVEAD	804.39	6.41	0.00
Myoglobin	14	29	WGKVEADIAGHGQEVL	1708.87	6.85	0.04
Myoglobin	20	29	DIAGHGQEVL	1038.52	6.25	0.02
Myoglobin	21	29	IAGHGQEVL	923.49	5.72	0.01
Myoglobin	30	40	IRLFTGHPETL	1283.71	6.91	0.00
Myoglobin	33	40	FTGHPETL	901.44	5.85	0.01
Myoglobin	41	55	EKFDKFKHLKTEAEM	1880.96	5.52	0.01
Myoglobin	55	69	MKASEDLKKHGTVVL	1655.92	5.25	0.01
Myoglobin	56	69	KASEDLKKHGTVVL	1524.87	4.96	0.04
Myoglobin	60	69	DLKKHGTVVL	1109.67	5.38	0.12
Myoglobin	70	76	TALGGIL	644.40	7.98	0.00
Myoglobin	70	86	TALGGILKKKGHHEAEL	1802.03	5.35	0.01
Myoglobin	110	131	AIIHVLHSKHPGDFGADAQGAM	2272.13	6.02	0.01
Myoglobin	124	134	GADAQGAMTKA	1020.48	4.59	0.04
Myoglobin	138	150	FRNDIAAKYKELG	1524.82	5.97	0.02
Myoglobin	138	151	FRNDIAAKYKELGF	1671.89	7.27	0.02
Myoglobin	143	153	AAKYKELGFQG	1211.64	6.78	0.02

Table 3.1: The peptide sequence list observed in both the aMb and hMb technical repeats The MHP (Monoisotopic mass in Da) of the peptide and its chromatographic retention time. The 138FRNDIAAKYKELG150 peptide of interest has been highlighted in yellow

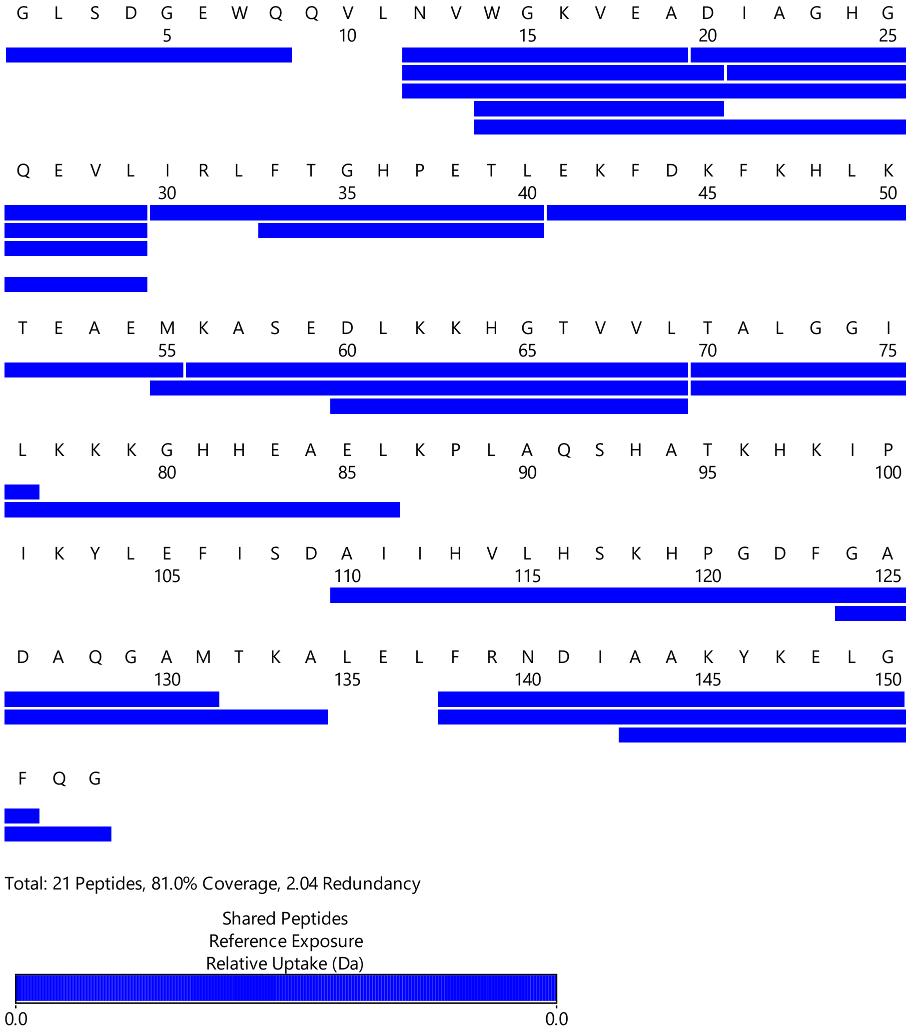


Figure 3.3: Coverage produced from both the hMb and aMb PLGS data. Each of the 21 peptides are represented as blue boxes. The 21 peptides provided an overall protein coverage of 81% with the unidentified areas being amino acids 9-11, 87,109 and 135-137 of the myoglobin's sequence. Importantly, regions covered by different peptides increases the resolution of deuteration maps in subsequent analysis

The peptide mapping revealed a sequence coverage of 81% (Figure 3.3). With peptide mapping completed, deuterium labelled sample data was loaded into DynmaX to assess each peptides deuterium uptake. The ^{138}FRNDIAAKYKELG^{150} has been used demonstrate shift in m/z as a result of deuteration. The peptide's spectral data across the timecourse can be seen in Figure 3.4.

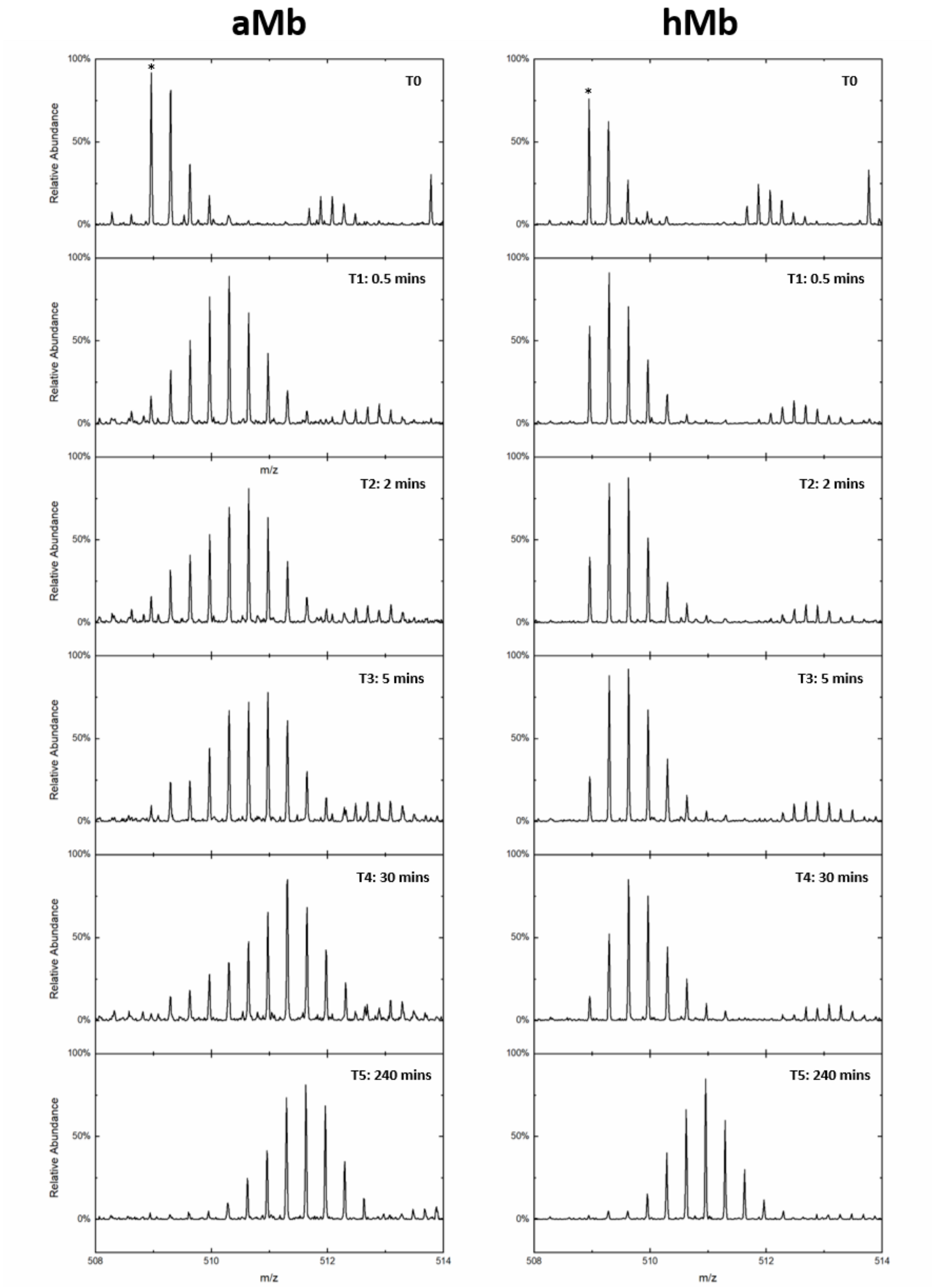


Figure 3.4: Stacked spectra of hMb peptide [M+3H]$^{3+}$ FRNDIAAKYKELG of aMb peptide [M+3H]$^{3+}$ 138FRNDIAAKYKELG150 across all timepoints. The star symbol in the t0 spectra represents the monoisotopic peak

Dynmax used the spectral to plot the relative deuterium uptake (RDU). The rate of exchange of both the hMb and aMb timecourse respectively are show in the deuterium uptake plot in Figure 3.5.

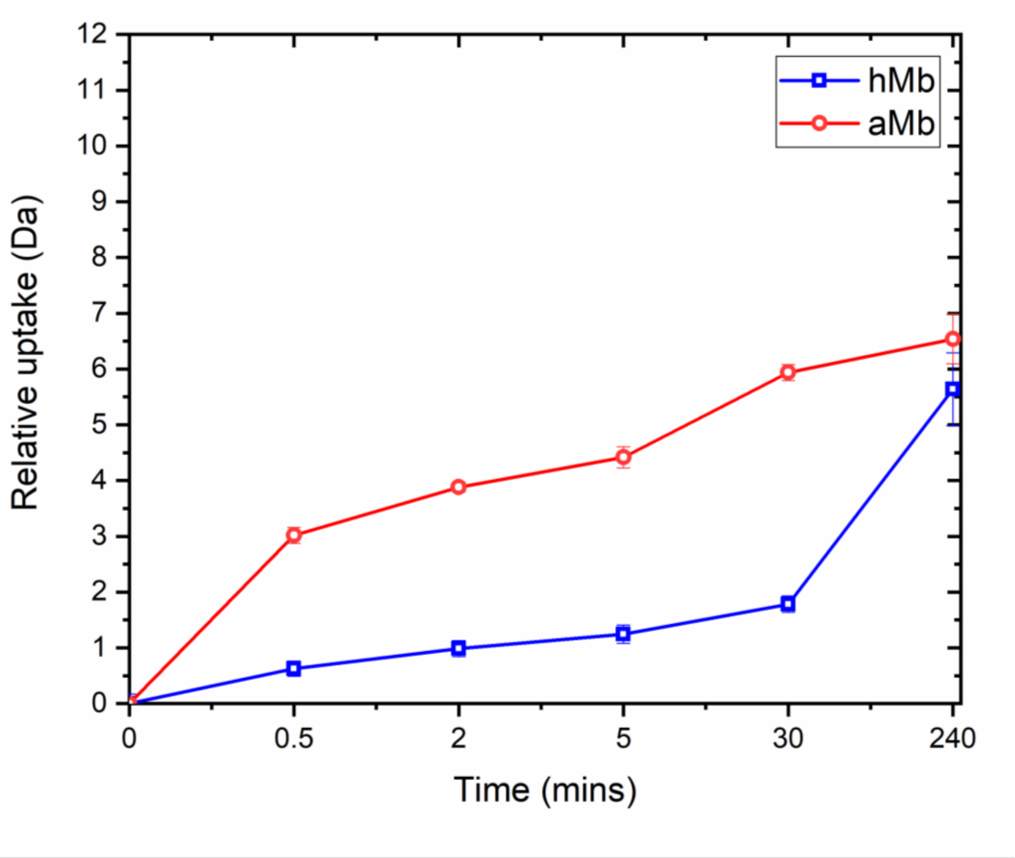


Figure 3.5: Relative uptake of deuterium of the [M+3H]$^{3+}$ ^{138}FRNDIAAKYKELG^{150} peptide calculated by DynamX.

The data from Figure 3.5 shows that the removal of the Heme group has resulted in a dramatic change in the rate of deuteration for the aMb form of the [M+3H]$^{3+}$ FRNDIAAKYKELG peptide. The aMb timecourse exchanges at a faster rate over the first few timepoints (T0-T3) than the hMb timecourse. By T4 (30 mins) the rate of deuterium uptake seems similar to the (T0-T3). However, by T5 (240 minutes) there is only marginal difference in RDU between protein states and the rate has dramatically dropped appearing to reach equilibrium.

To demonstrate how DynamX processed the data to generate the RDU plot for Figure 3.5 the aMb & hMb T2 timepoint data has been tabulated below (Tables 3.2 & 3.3*).*

hMb Sequences	**Start**	**End**	**Relative Uptake (Da)**	**Relative Uptake SD**	**Num Exchangers**	**Rel Fract Uptake**	**Rel Fract Uptake SD**
GLSDGEWQ	1.00	8.00	1.27	0.10	7.00	18.21	1.41
NVWGKVEA	12.00	19.00	0.85	0.05	7.00	12.12	0.71
NVWGKVEAD	12.00	20.00	0.83	0.67	8.00	10.36	8.36
NVWGKVEADIAGHGQEVL	12.00	29.00	3.12	0.14	17.00	18.35	0.83
WGKVEAD	14.00	20.00	0.78	0.07	6.00	13.07	1.11
WGKVEADIAGHGQEVL	14.00	29.00	2.26	0.34	15.00	15.07	2.26
DIAGHGQEVL	20.00	29.00	1.38	0.29	9.00	15.30	3.20
IAGHGQEVL	21.00	29.00	0.60	0.12	8.00	7.55	1.46
IRLFTGHPETL	30.00	40.00	0.17	0.21	9.00	1.93	2.36
FTGHPETL	33.00	40.00	0.66	0.09	6.00	11.01	1.45
EKFDKFKHLKTEAEM	41.00	55.00	2.33	0.06	14.00	16.66	0.40
MKASEDLKKHGTVVL	55.00	69.00	0.71	0.11	14.00	5.07	0.79
KASEDLKKHGTVVL	56.00	69.00	0.84	0.10	13.00	6.45	0.75
DLKKHGTVVL	60.00	69.00	0.44	0.10	9.00	4.84	1.15
TALGGIL	70.00	76.00	0.50	0.05	6.00	8.41	0.79
TALGGILKKKGHHEAEL	70.00	86.00	0.93	0.04	16.00	5.84	0.25
AIIHVLHSKHPGDFGADAQGAM	110.00	131.00	3.89	0.11	20.00	19.43	0.54
GADAQGAMTKA	124.00	134.00	1.83	0.08	10.00	18.34	0.80
FRNDIAAKYKELG	138.00	150.00	0.99	0.14	12.00	8.21	1.18
FRNDIAAKYKELGF	138.00	151.00	1.41	0.08	13.00	10.85	0.58
AAKYKELGFQG	143.00	153.00	1.72	0.13	10.00	17.15	1.26

Table 3.2: DynamX data for hMb at deuterium 2 minutes . The table includes the peptide names, the start and end location of the peptide within the protein sequence, the relative uptake which represents the level of deuteration of the peptide, the relative uptake, the number of exchangeable hydrogens within the peptide, the relative fraction uptake which the percentage of exchange. The M+3H]$^{3+}$ ^{138}*FRNDIAAKYKELG*^{150} *has been highlighted yellow. Standard deviations have been calculated from the triplicate technical repeats*

aMb Sequences	**Start**	**End**	**Relative Uptake (Da)**	**Relative Uptake SD**	**Num Exchangers**	**Rel Fract Uptake**	**Rel Fract Uptake SD**
GLSDGEWQ	1.00	8.00	1.52	0.04	7.00	21.71	0.52
NVWGKVEA	12.00	19.00	0.84	0.02	7.00	12.00	0.32
NVWGKVEAD	12.00	20.00	0.93	0.06	8.00	11.58	0.71
NVWGKVEADIAGHGQEVL	12.00	29.00	3.03	0.08	17.00	17.84	0.47
WGKVEAD	14.00	20.00	0.77	0.10	6.00	12.89	1.64
WGKVEADIAGHGQEVL	14.00	29.00	2.12	0.30	15.00	14.11	2.00
DIAGHGQEVL	20.00	29.00	1.26	0.04	9.00	14.02	0.48
IAGHGQEVL	21.00	29.00	0.56	0.02	8.00	7.02	0.25
IRLFTGHPETL	30.00	40.00	0.89	0.06	9.00	9.85	0.63
FTGHPETL	33.00	40.00	0.95	0.07	6.00	15.86	1.19
EKFDKFKHLKTEAEM	41.00	55.00	2.68	0.22	14.00	19.15	1.54
MKASEDLKKHGTVVL	55.00	69.00	1.22	0.05	14.00	8.68	0.37
KASEDLKKHGTVVL	56.00	69.00	1.07	0.04	13.00	8.21	0.27
DLKKHGTVVL	60.00	69.00	0.98	0.07	9.00	10.86	0.75
TALGGIL	70.00	76.00	0.79	0.02	6.00	13.23	0.34
TALGGILKKKGHHEAEL	70.00	86.00	2.17	0.06	16.00	13.57	0.38
AIIHVLHSKHPGDFGADAQGAM	110.00	131.00	4.02	0.12	20.00	20.08	0.58
GADAQGAMTKA	124.00	134.00	2.19	0.05	10.00	21.92	0.52
FRNDIAAKYKELG	138.00	150.00	3.88	0.08	12.00	32.34	0.71
FRNDIAAKYKELGF	138.00	151.00	4.79	0.06	13.00	36.85	0.49
AAKYKELGFQG	143.00	153.00	4.60	0.10	10.00	46.00	1.01

Table 3.3: DynamX data for aMb at deuterium 2 minutes. The table includes the peptide names, the start and end location of the peptide within the protein sequence, the relative uptake which represents the level of deuteration of the peptide, the relative uptake, the number of exchangeable hydrogens within the peptide, the relative fraction uptake which the percentage of exchange. The M+3H]$^{3+}$ ^{138}FRNDIAAKYKELG^{150} has been highlighted yellow. Standard deviations have been calculated from the triplicate technical repeats

From Tables 3.2 and 3.3 it is possible to see the variation in RDU of each of the 21 peptides at 2 minutes of deuteration for both hMb and aMb. The RDU of peptides from the aMb 2 min timepoint ranged between 0-4.78 Da whereas the hMb peptides range 0-3.11 Da. The RDU of each peptide is shown above the protein coverage maps of both the hMb and aMb timecourses (Figure 3.6).

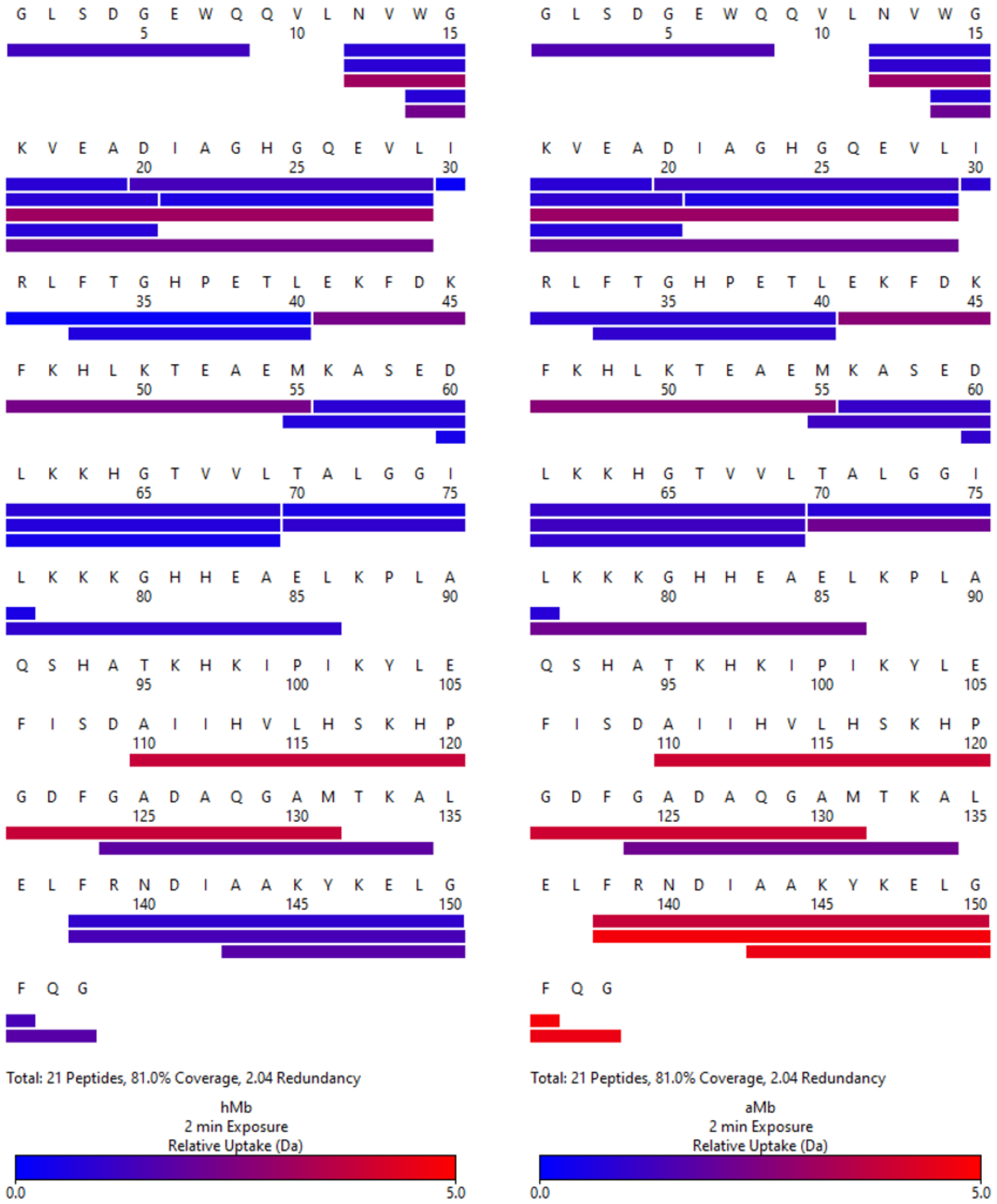


Figure 3.6: Myoglobin coverage maps hMb and aMb at 2 minutes of deuterium exposure. In the C-terminus of aMb timecourse there is an increase in RDU There is a similar level of exchange at regions near the heme binding His64 residue

From Figure 3.6, the removal of the heme group results in a substantial increase in RDU in the C-terminus of aMb timecourse, which is not present in the hMb. These coverage maps were then processed onto the 3D crystal structure of myoglobin PDB ID:1MBN (Figure 3.7).

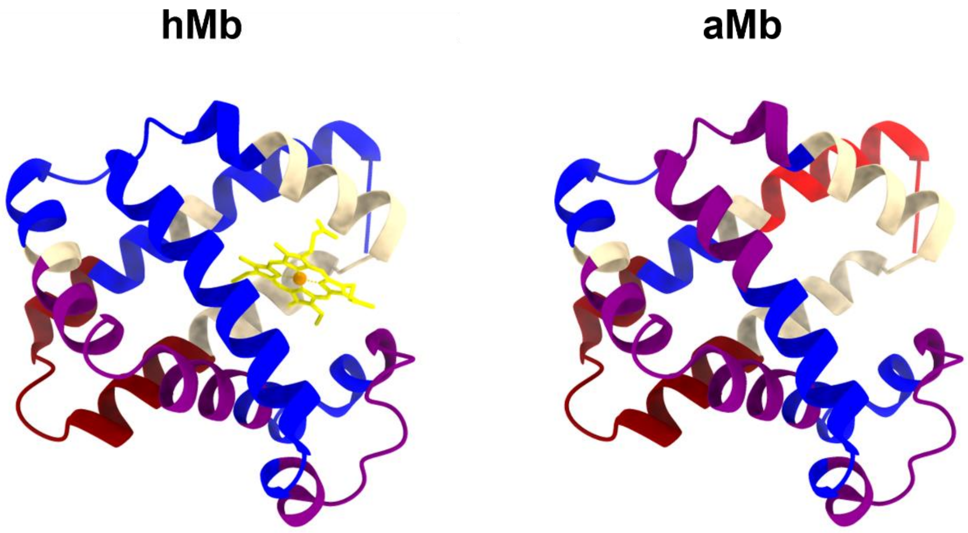


Figure 3.7: Visual representation of relative uptake in hMb and aMb at 2 minutes of deuterium exposure. Regions coloured red show highest exchange in the protein while purple and blue show intermediate and low exchange respectively. Areas coloured cream are unaccounted for in the experiment

Although promising it is important to state that this data represents the change in RDU for one timepoint. Furthermore, RDU values only represent levels of deuteration and do not provide a percentage of deuterated viable exchange sites within the peptides. To factor in a peptide’s viable exchange sites, the relative fractional uptake (RFU) can be calculated using Equation 2.3. The RFU values for both aMb and hMb are plotted on the butterfly plot in Figure 3.8.

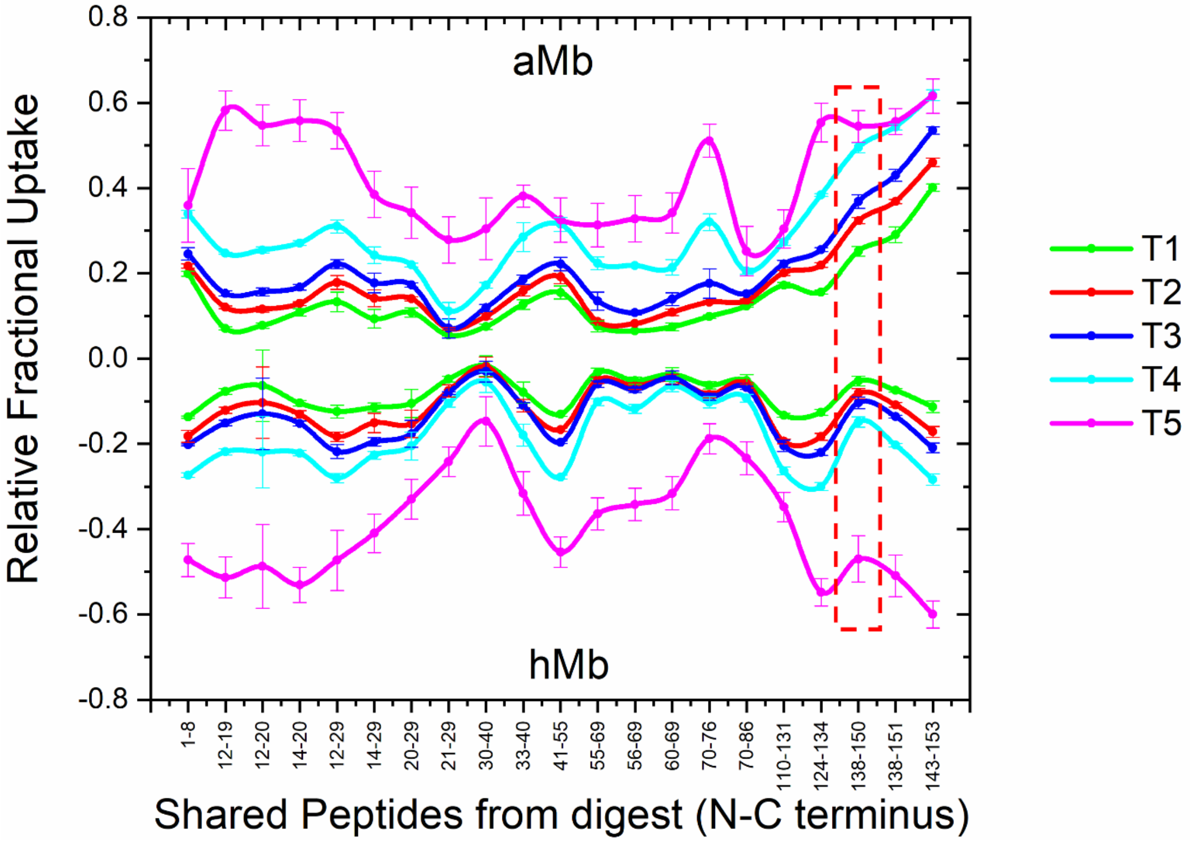
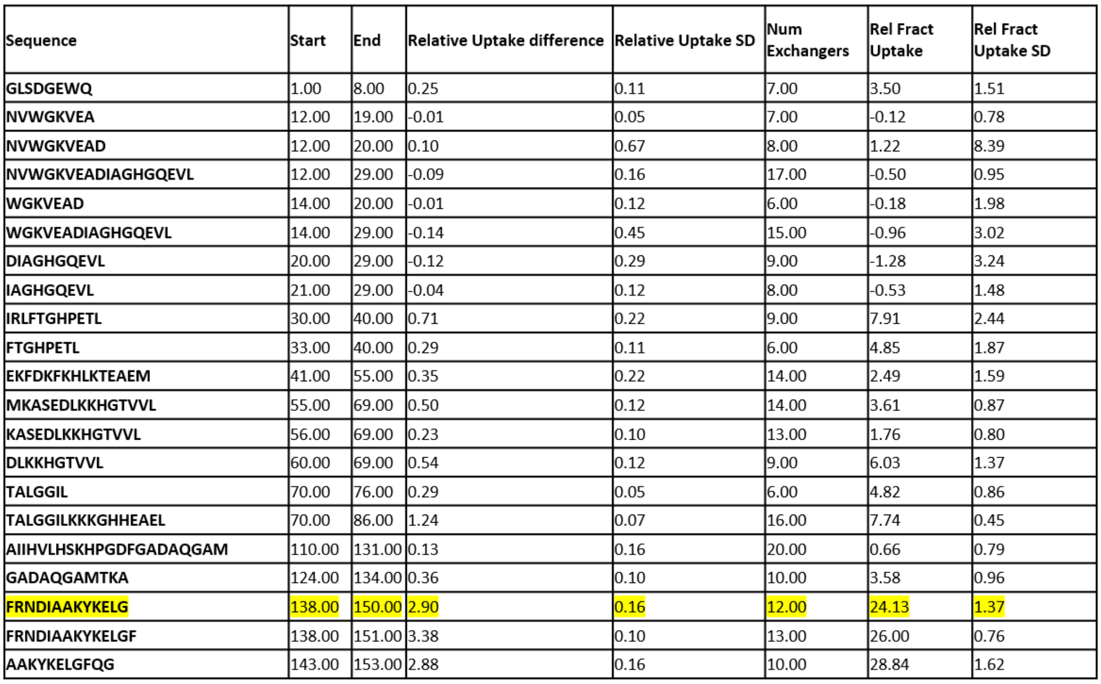


Figure 3.8: Butterfly plot produced from the combined hMb and aMb data showing relative fractional uptake against the myoglobin primary sequence. aMb HDX shown in the positive RFU region and hMb HDX in the negative RFU region. Each coloured line corresponds to one timepoint. Each vertical series of dots within the lines represent a peptide. The highlighted red box shows the ^{138}*FRNDIAAKYKELG*^{150} *peptide across all timepoints*

Figure 3.8 shows that there is substantial difference in level of deuteration around the C-terminus with the removal of the heme group. The RFU^{Diff} between the aMb and hMb has been calculated for each peptide at the T2 timepoint. The RFU^{Diff} values for T2 timepoints have been tabulated (Table 3.4), and the results have been placed atop the peptide coverage map (Figure 3.9).



Sequence	Start	End	Relative Uptake difference	Relative Uptake SD	Num Exchangers	Rel Fract Uptake	Rel Fract Uptake SD
GLSDGEWQ	1.00	8.00	0.25	0.11	7.00	3.50	1.51
NVWGKVEA	12.00	19.00	-0.01	0.05	7.00	-0.12	0.78
NVWGKVEAD	12.00	20.00	0.10	0.67	8.00	1.22	8.39
NVWGKVEADIAGHGQEVL	12.00	29.00	-0.09	0.16	17.00	-0.50	0.95
WGKVEAD	14.00	20.00	-0.01	0.12	6.00	-0.18	1.98
WGKVEADIAGHGQEVL	14.00	29.00	-0.14	0.45	15.00	-0.96	3.02
DIAGHGQEVL	20.00	29.00	-0.12	0.29	9.00	-1.28	3.24
IAGHGQEVL	21.00	29.00	-0.04	0.12	8.00	-0.53	1.48
IRLFTGHPETL	30.00	40.00	0.71	0.22	9.00	7.91	2.44
FTGHPETL	33.00	40.00	0.29	0.11	6.00	4.85	1.87
EKFDKFKHLKTEAEM	41.00	55.00	0.35	0.22	14.00	2.49	1.59
MKASEDLKKHGTVVL	55.00	69.00	0.50	0.12	14.00	3.61	0.87
KASEDLKKHGTVVL	56.00	69.00	0.23	0.10	13.00	1.76	0.80
DLKKHGTVVL	60.00	69.00	0.54	0.12	9.00	6.03	1.37
TALGGIL	70.00	76.00	0.29	0.05	6.00	4.82	0.86
TALGGILKKKGHHEAEL	70.00	86.00	1.24	0.07	16.00	7.74	0.45
AIIHVLHSKHPGDFGADAQGAM	110.00	131.00	0.13	0.16	20.00	0.66	0.79
GADAQGAMTKA	124.00	134.00	0.36	0.10	10.00	3.58	0.96
FRNDIAAKYKELG	138.00	150.00	2.90	0.16	12.00	24.13	1.37
FRNDIAAKYKELGF	138.00	151.00	3.38	0.10	13.00	26.00	0.76
AAKYKELGFQG	143.00	153.00	2.88	0.16	10.00	28.84	1.62

Table 3.4: DynamX data for relative fractional difference of deuteration at 2 minutes. The table includes the peptide names, the start and end location of the peptide within the protein sequence, the RFU^{Diff}, the number of exchangeable hydrogens within the peptide and the RFU Standard deviations have been calculated from the triplicate technical repeats

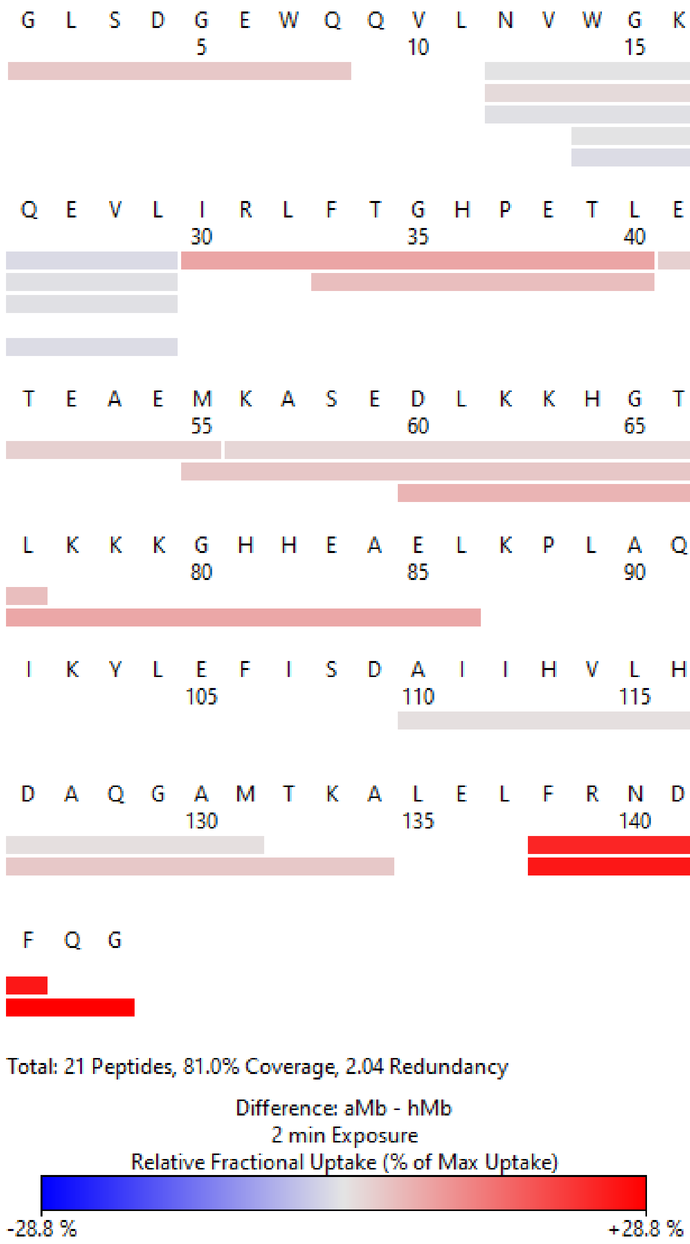


Figure 3.9: Myoglobin RFU^{diff} coverage map between aMb (red) and hMb (blue) at 2 minutes exposure. There is a modest increase in deuterium uptake favouring aMb across the central region 30-90. There is a large increase in deuteration in near the C-terminus peptides favouring hMb

It can be seen from Figure 3.9 that the exchange is much greater in aMb as the RFU is substantially higher at the C-terminus with aMb deuterating 34% at 30 minutes of deuteration. This suggests the increase in RFU is linked to the structural difference caused by the loss of the heme group. This result indicates that the absence of the heme group would allow for the protein structure to be less tightly bound and thus give rise to more HDX.

Myoglobin binds to the heme group at positions His64 and His93 in the sequence. Therefore, the free His64 and His93 in aMb may be responsible for the increase in deuterium uptake. One possibility is that myoglobin loses its tertiary structure as the pore containing the heme group collapses. This collapse may result in the C-terminal helix becoming more solvent accessible even though the C-Terminus is located 12 Angstroms away from the Heme group.

This theory is further supported by the low RFU values in the regions surrounding the His64 and His93 binding sites. If the pore collapses, there might not be a large increase in solvent accessibility in these regions. However, the pore collapse may result in the C-terminus become more solvent accessible as the tertiary structure changes conformation.

3.1.2 Validation of the HDX-MS automated platform

Automation is commonly used as a means to reduce back-exchange and increase experimental throughput[135]. A typical automated platform for HDX-MS consists of a cartesian robot arm with a syringe, a cooler and computer through which the arm can be programmed. Once programmed the arm can be used for sample preparation and liquid handling purposes.

This section details the assembly, programming, and initial validation experiments of the automated platform.

The automated platform consists of two CTC HTC PAL robotic arms working in conjunction with the waters HDX manager LC system and the Synapt G2 MS. Figure 3.10: details an overview of the processes that comprise the automated system.

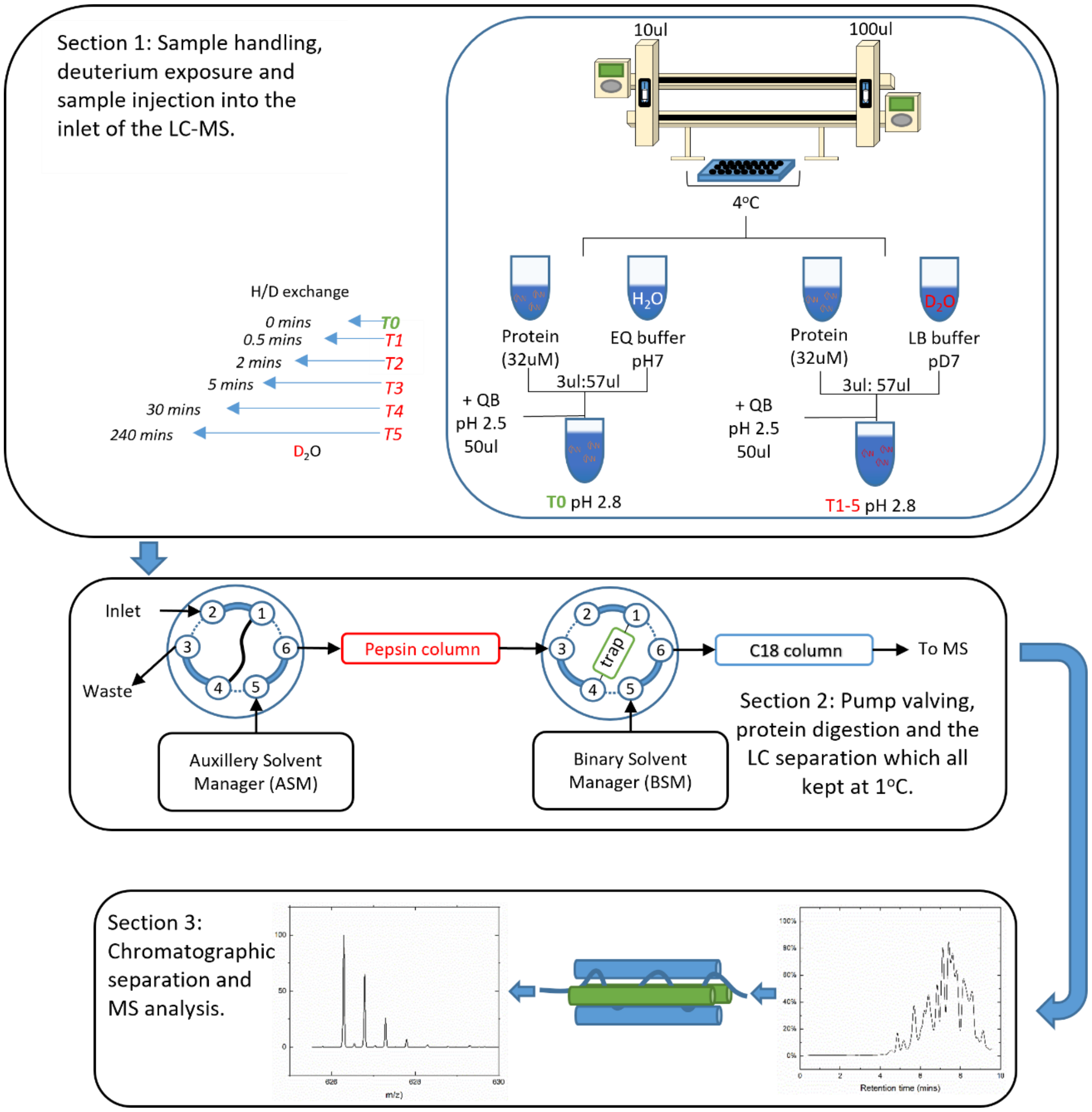
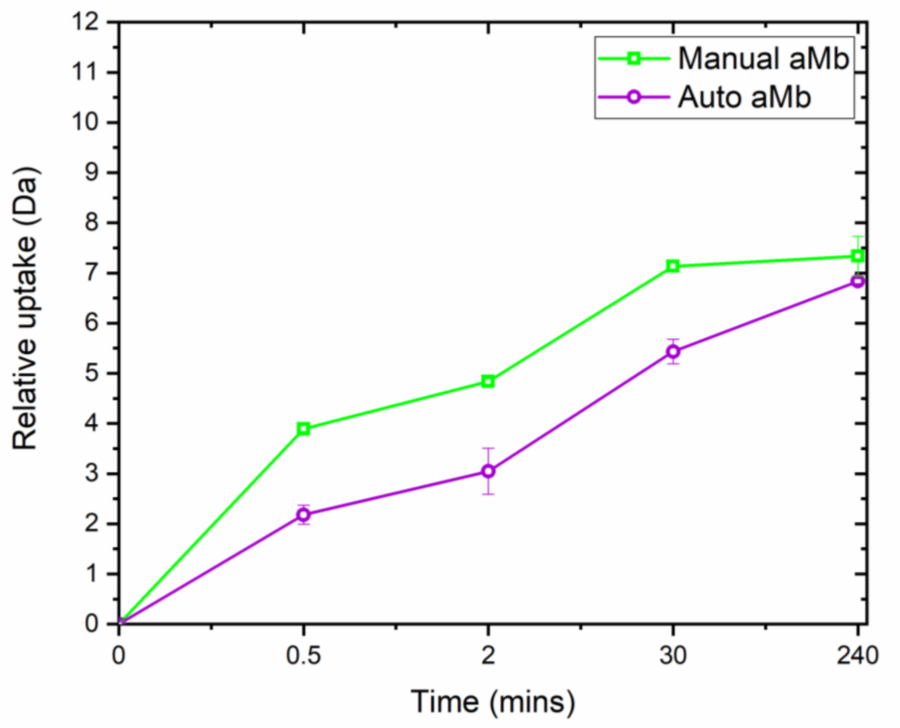
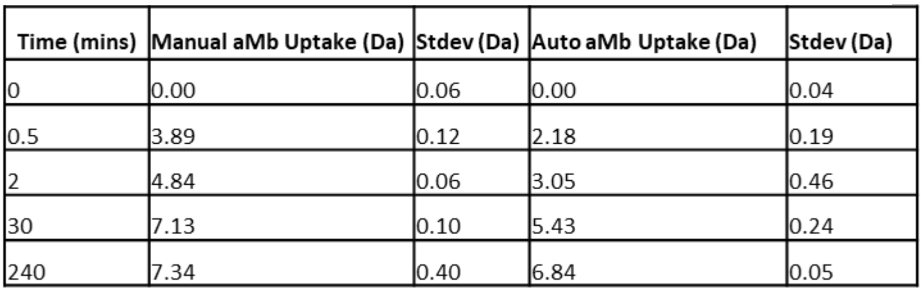


Figure 3.10: Technical overview of the automated bottom up HDX-MS workflow. Section 1: An overview of the generation of timepoints including liquid handling and sample preparation. The blue highlighted area represents the operational area for CTC HTS Robotic arm. Section 2: an overview of valving and pumps within the waters HDX manager. The filled lines on the valving represent the "Inject" valve position and the dash lines represent "Load" valve position. The flow rates through the ASM were 200ul/min and 40ul/min in the BSM Section 3: chromatographic separation and MS analysis.

3.1.3 Automated vs manual HDX of aMb

The automated hMb timecourse using the same experimental criteria as the manual experiment. Figure 3.11 shows the relative deuterium uptake plots of $[M+3H]^{3+}$ FRNDIAAKYKELG generated from both manual and automated experimentation.





Time (mins)	Manual aMb Uptake (Da)	Stdev (Da)	Auto aMb Uptake (Da)	Stdev (Da)
0	0.00	0.06	0.00	0.04
0.5	3.89	0.12	2.18	0.19
2	4.84	0.06	3.05	0.46
30	7.13	0.10	5.43	0.24
240	7.34	0.40	6.84	0.05

Figure 3.11: Relative uptake of deuterium of the manual and automated timecourse experiment of $[M+3H]^{3+}$ FRNDIAAKYKELG

Figure 3.11 shows that the automated platform offers a slower rate of exchange over individual time points. This is a result of the cold labelling step as described in Figure 3.10.

However, the maximal exchanged controls were exchanging at similar level (Manual = 7.34± 0.4 and Auto = 6.84 ± 0.05 at 240 mins labelling). The slow rate of exchange provides unique benefits namely the ability to observe the rapidly exchanging lower time points < 30 seconds without any significant increase in back-exchange. This trend can be seen across the protein analysis Figure 3.12. Finally there is a high degree of variation between timepoint replicates in the manual over the automated datasets. This further demonstrates the improved reliability of an automated HDX-MS setup.

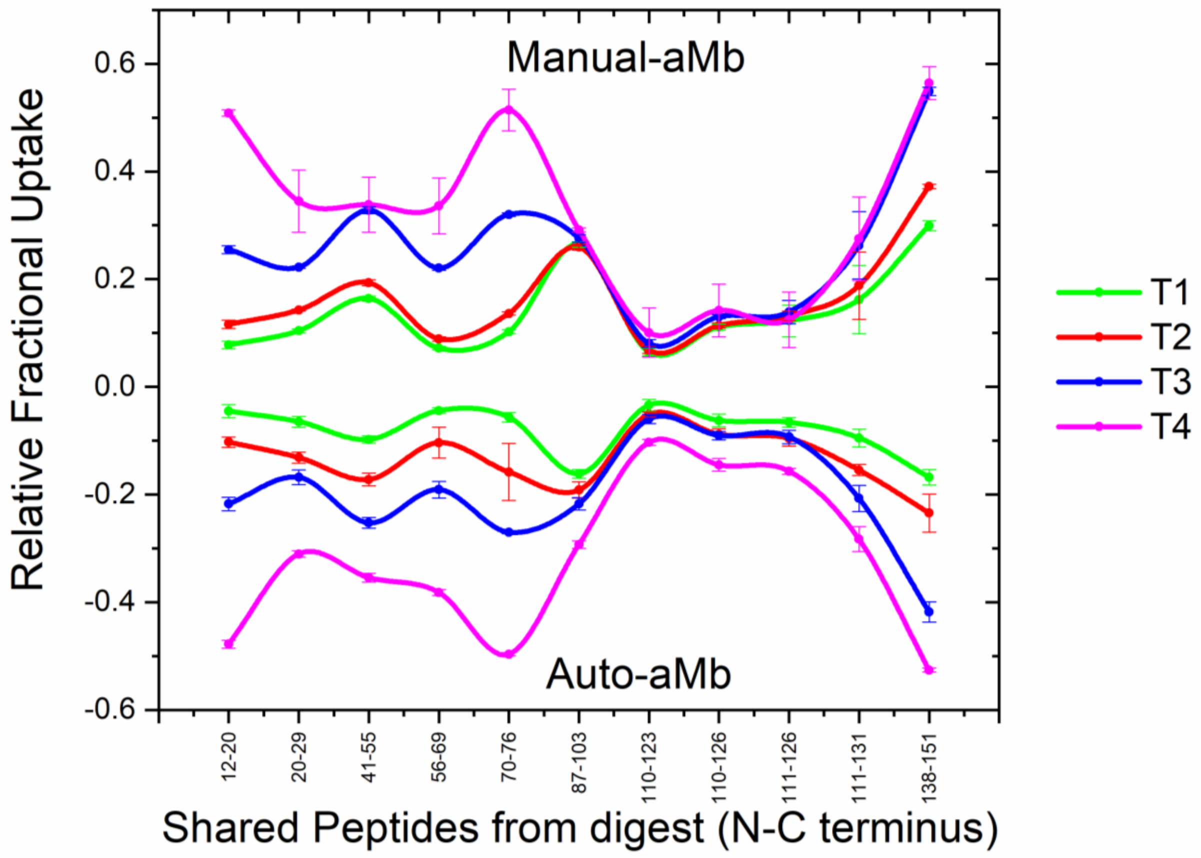


Figure 3.12: Butterfly plot produced from the manual and automated aMb timecourses showing RFU against the aMb primary sequence. Manual aMb

HDX shown in the positive RFU region and automated aMb HDX in the negative RFU region. Each coloured line corresponds to one timepoint

3.2 System associated back-exchange

One of the largest concerns of HDX-MS workflows is the degradation of data quality via D→H back-exchange. As previously discussed, back-exchange can be mitigated through a combination of regulating pH and temperature. However, due to the involvement of a chromatographic separation step it is inevitable that a deuterated sample will be in contact with solvents involved with the mobile phase during separation. To this end, it is important that any HDX-MS platform be validated to ensure that the % back-exchange inherent from the workflow falls within a suitable threshold. Recent publications have provided suitable recommendations on how to assess back-exchange with new system[88]. Typically, acceptable levels of back-exchange associated from a bottom-up HDX workflow rests between 25-40% although many have reported back-exchange levels below these parameters[88].

3.2.1 Alpha synuclein back-exchange assessment

There have been several recommendations regarding back-exchange standards. These range from peptide assessments with linearized peptide such as substance P to model proteins. Substance P can assess the level of exchange as an individual peptide. Alternatively, an inherently disorder protein which lacks higher order conformation can be used. One such example would be the α-Synuclein which is an ideal candidate for back-exchange assessment not only because it lacks high order conformation but it also contains significant regions which lack secondary structure[136]. Previously published reports have used also used α-Synuclein as a model protein to explore both protein aggregation and folding dynamics[137,138].

It is from these regions that system associated back-exchange is assessed by the maximally deuterated peptides from the Tmax Figure 3.13.

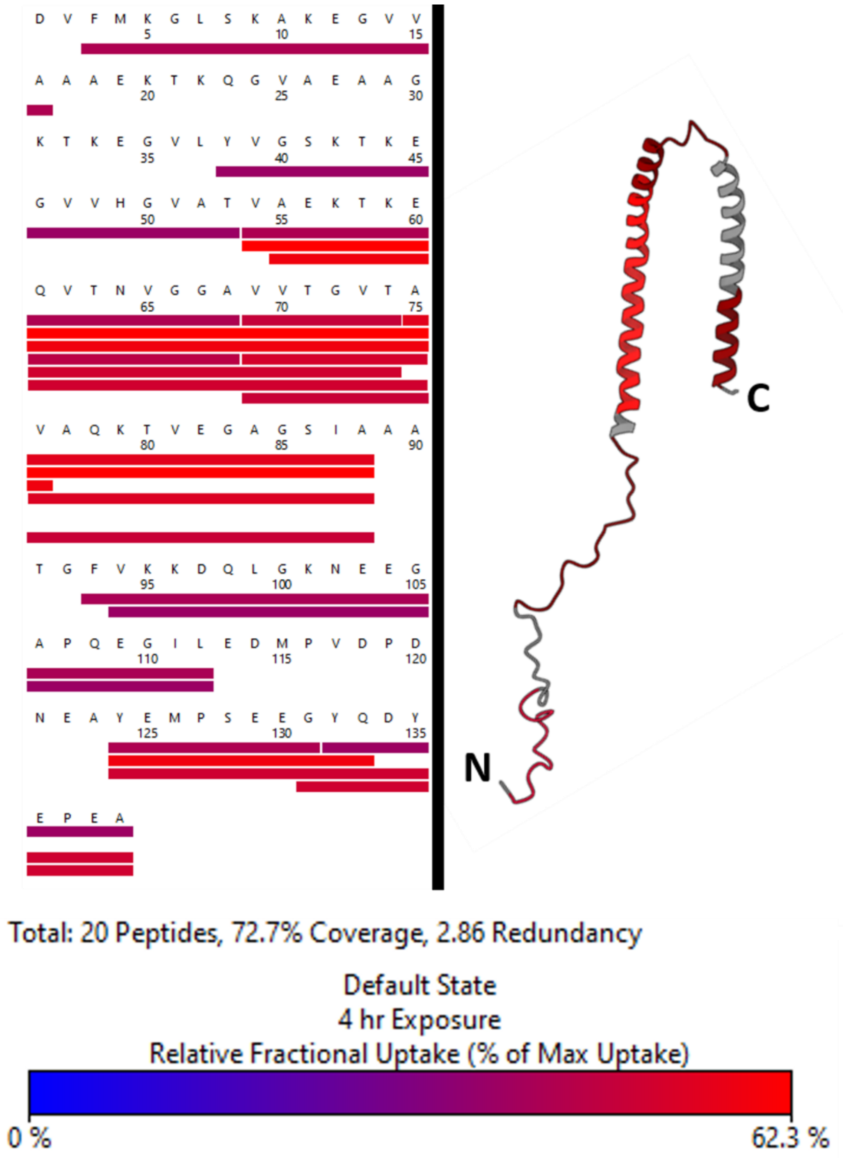


Figure 3.13: Relative fractional Coverage map of Alpha synuclein at 4 hours of deuterium labelling, the coverage map is also overlain the crystal structure of α-Synuclein. Grey regions represents no coverage PDB ID: 1XQ8

Utilizing the workflow previously detailed in Figure 3.10, the deuterium concentration for the labelling step was calculated to be 95% this back-exchange study demonstrated that peptides were deuterating between 37.92%-62.3% (Table 3.5).

Sequence	**% Uptake @ 4 hours labelling**	**% Back-exchange**
FMKGLSKAKEGVVA	42.09	52.91
YVGSKTKEGVVHGVAT	37.96	57.04
VAEKTKEQVTNVGGA	42.59	52.41
VAEKTKEQVTNVGGAVVTGVTAVAQKTVEGAGSIA	62.31	32.69
AEKTKEQVTNVGGAVVTGVTAV	58.63	36.37
QVTNVGGA	45.34	49.66
VVTGVT	48.09	46.91
VVTGVTA	51.80	43.20
VVTGVTAVAQKTVEGAGSIA	57.94	37.06
AVAQKTVEGAGSIA	55.31	39.69
VAQKTVEGAGSIA	53.67	41.33
FVKKDQLGKNEEGAPQEGIL	40.89	54.11
VKKDQLGKNEEGAPQEGIL	37.92	57.08
YEMPSEEG	42.51	52.49
YEMPSEEGYQDYEPEA	50.23	44.77
GYQDYEPEA	50.51	44.49
YQDYEPEA	38.27	56.73
	Avg % backexchange =	**47.00**

Table 3.5: % Uptake and % back-exchange values from the 17 Alpha synuclein peptides at 4 hours of deuteration. Back-exchange values were adjusted to the 95% deuterium present during the labelling step. Back-exchange values which fall below 50% have been highlighted blue. Back-exchange values which are above the 50% threshold have been highlighted orange

These findings fall within the recommendations for good HDX practice which demonstrate that fully labelled peptide should have %back-exchange lying between 30-40% back-exchange. From these findings several peptides demonstrate back-exchange values which fall between the 30-40% target. Likewise, the writers state that most peptides should have back-exchange values below 50%[88]. These findings show that the average back-exchange 47.00% with 10 of the 17-peptide having back exchange values below 50%.

Chapter 4 Unravelling the assembly pathway of EncFtn decamers using comparative HDX-MS

Mass spectrometry reveals the assembly pathway of encapsulated ferritins and highlights a dynamic ferroxidase interface

Data presented in this chapter was published in Chemical Communications[117]

EncFtns are a subfamily of classical ferritins that are sequestered into larger macrostructure Enc capsids[110]. Operating similarly in function to unencapsulated bacterial ferritins, EncFtn catalytically oxidise Fe(II) to Fe(III) and the resulting Fe(III) ion is stored within a larger encapsulating (Enc-EncFtn nanocompartment. As stated in section 1.7, EncFtns are directed into the larger Encs via a short C-terminal LS within EncFtn[128]. It is thought that this interplay between the EncFtn and the Enc can be repurposed into transport platforms such as drug delivery systems for chemotherapeutics, utility as a biomarkers or gene editing complexes[122]. To this end, uncovering how both the Enc-EncFtn complex interact, and how they assemble is crucial to understanding how these systems can be repurposed for clinical applications.

4.1 Introduction to EncFtn assembly pathways via Dimer recruitment

Previous work has found that EncFtns assemble in a decameric ‘pentamer-of-dimers’ arrangement. This has been found to be a common structural fold for the EncFtn family (Figure 4.1)[110,116]. The decameric “pentamer of dimers” has been identified from the solved crystal structures of *Rhodospirillum rubrum* (Rru-EncFtn) PDB ID: 5DA5, *Pyroccocus furiosus* (Pfc-EncFtn) PDB ID: 5N5E and *Haliangium ochraceum* (Hoch-EncFtn) PDB ID: 5N5F[110,116]. Due to the arrangement of the assembly, there are two distinct dimer interfaces present in the Hoch-EncFtn decamer - one containing the catalytic ferroxidase centre that binds to two iron ions (FOC dimer) coordinated by three amino acid (Glu30, Glu60 and His63) (Figure 4.1B). The second dimer (the non-FOC dimer) is held together by hydrophobic interfaces and hydrogen bonds (Figure 4.1C).

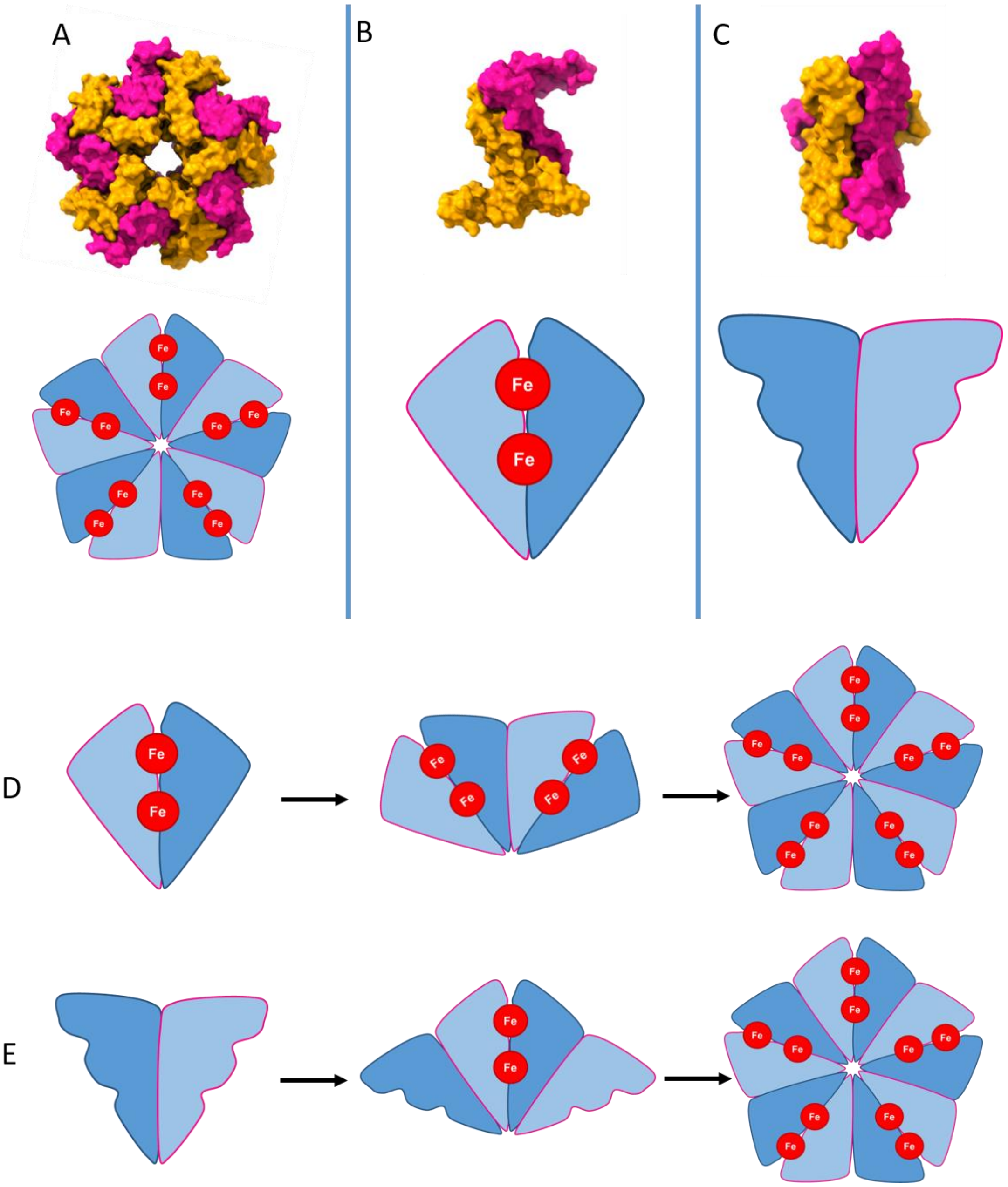


Figure 4.1: The EncFtn decamer structure and the two dimer subcomplexes found within Hoch-EncFtn assembly PDB ID: 5N5F. A schematic diagram of each structure is beneath each assembly. (A) Hoch-EncFtn decamer, (B) the FOC dimer and (C) Non-FOC dimer. The two possible assembly pathways of EncFtn via dimer association; the FOC assembly pathway (D) and the Non-FOC assembly pathway (E)

4.2 Native MS analysis of EncFtn Assembly

Recent native MS studies from our group found that wild type Hoch-EncFtn (EncFtn-WT) exists as a series of subcomplexes comprising dimer, tetramer and hexamer and decamer[117] (Figure 4.2B, top). This observation led to the hypothesis that the decamer structure is assembled from the association of dimer subunits.

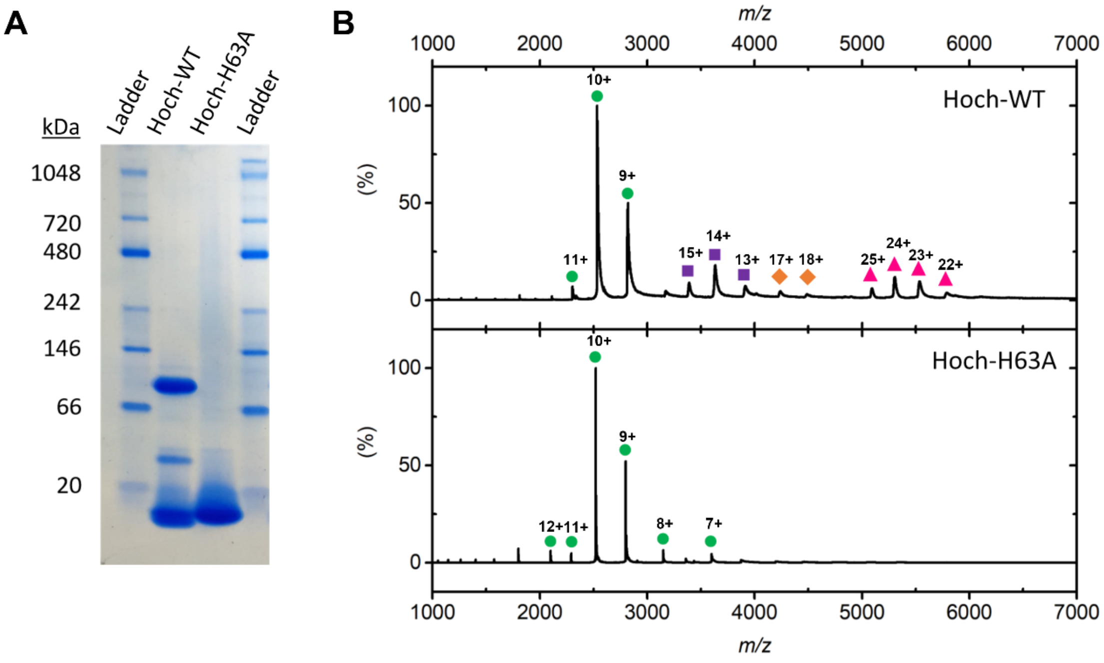


Figure 4.2: Native mass spectrum of EncFtn-WT demonstrating the various oligomerisation states present at pH 8.5 and 6.5 respectively. (A) Native PAGE of EncFtn-WT and EncFtn-H63A (B) Native nanoelectrospray ionization spectra of EncFtn-WT and EncFtn-H63A with gas phase oligomerization stressed with coloured shapes. Dimer as green circles; tetramer as purple squares; hexamer as orange diamonds and decamer as pink triangles. Data collected by Dr J. Ross

Thus, it has been proposed that the characteristic pentamer of dimers' structure is assembled by the addition of dimer subunits. If this is the case, then two possible assembly pathways can be proposed (Figure 4.1 D & E). The first is by sequential addition of FOC dimers (Figure 4.1D); the second by addition of non-FOC dimer units (Figure 4.1E).

In a related study by our group investigating the structure of a series of EncFtn single point variants, native MS analysis and gel electrophoresis suggested that an H63A EncFtn variant was unable to form higher order structure above dimers (Figure 4.2B). In addition, native ion mobility analysis revealed that the H63A dimer and the WT dimer subcomplex have very similar collision cross sections (20.02nm^2 and 20.15 nm^2 respectively). suggesting that they share the same topology (i.e. are the same dimer form)[117].

Because of the inability to form higher order structures, the H63A variant should possess a dimer interface that is significantly more solvent exposed than the WT protein. Therefore, it was proposed that differential HDX-MS analysis may be able to identify the dominant dimer form present and thus ascertain which assembly pathway is favoured in the formation of the EncFtn decamer structure.

4.3 Aim

The aim of this study was to use bottom-up differential HDX-MS to analyse Hoch EncFtn (EncFtn-WT) and the H63A variant (EncFtn-H63A). Analysis of the HDX data in the context the Solvent Accessible Surface Area (SASA) analysis of the FOC and Non-FOC dimer subcomplexes will be used to confirm which dimer interface are solvent exposed in solution and thus shed light on the assembly of EncFtn decameric structure.

4.4 Proteolytic digest and LC separation of EncFtn

EncFtn-WT and EncFtn-H63A were recombinantly produced in *E. coli* and purified by Dr. Jennifer Ross. After automated HD exchange and quenching protein samples were loaded onto a pepsin column for digestion and the resulting peptides were separated by reverse phase chromatography (see section 2.1 for details).

The digestion conditions used were successful for both WT and H63A EncFtn with most peptides eluting between RT 5.5 mins and RT 8.25 mins and no undigested protein being retained (Figure 4.3).

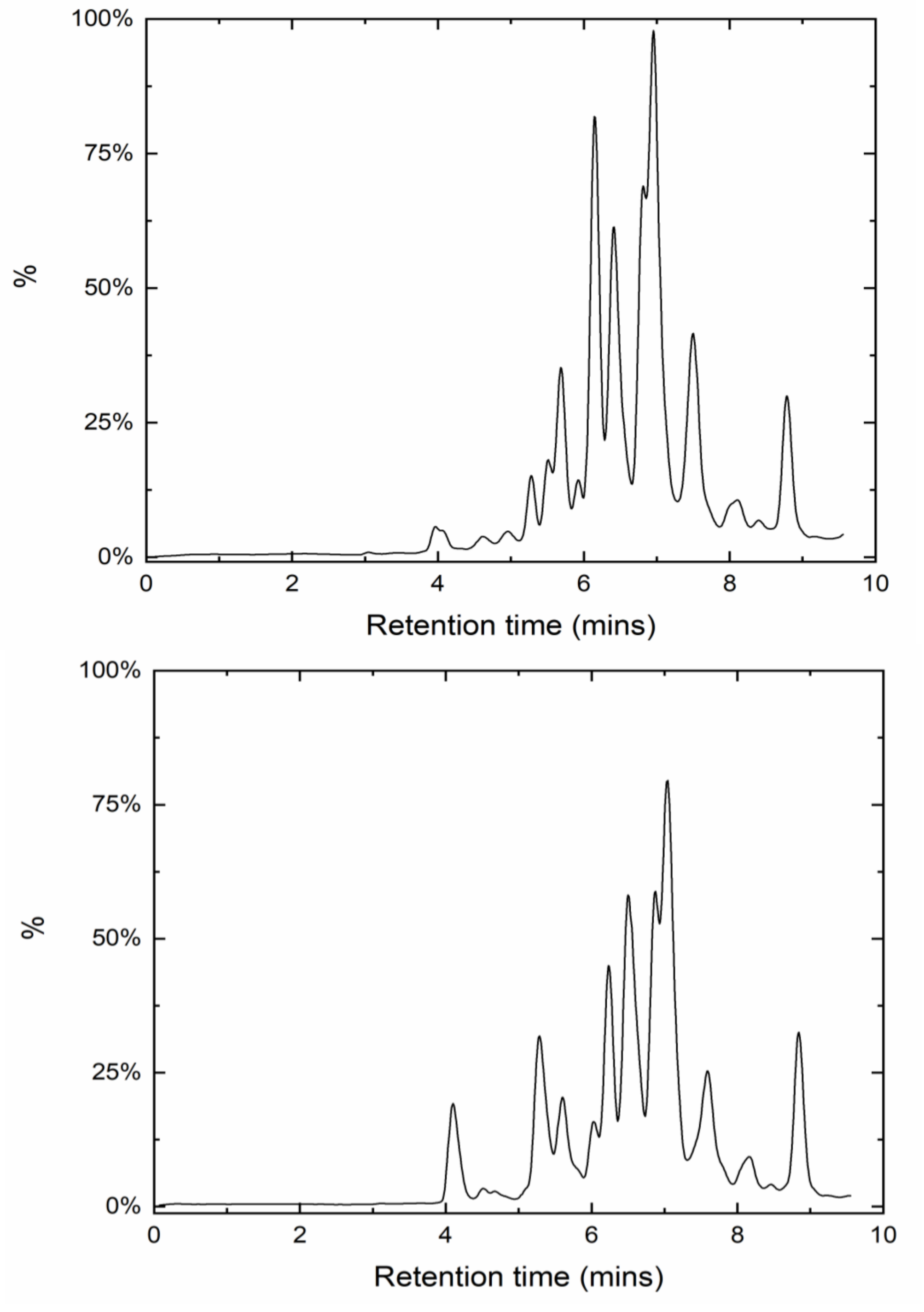


Figure 4.3: Total ion chromatograms for both H63A and WT samples. Separation was performed on a Waters Acquity C18 column using a 10 minute step gradient from 5%-95% water – acetonitrile

4.5 Differential HDX-MS analysis of WT and H63A EncFtn dimer

The collected HDX-MS raw data analysed using the ProteinLynx Global SERVER (PLGS; Waters) to assign peptides from the MS^E data. Total protein coverages for both the WT and H63A were >80%. This passed the threshold coverage to proceed to perform the deuteration assessment using DynamX HDX Data Analysis Software (Waters). The overall sequence coverage after digestion was found to be 94.3% from 25 peptides in the EncFtn-WT and 100% from 51 peptides in the EncFtn-H63A samples across all time-points. The coverage maps for each timecourse can be seen in Figure 4.4.

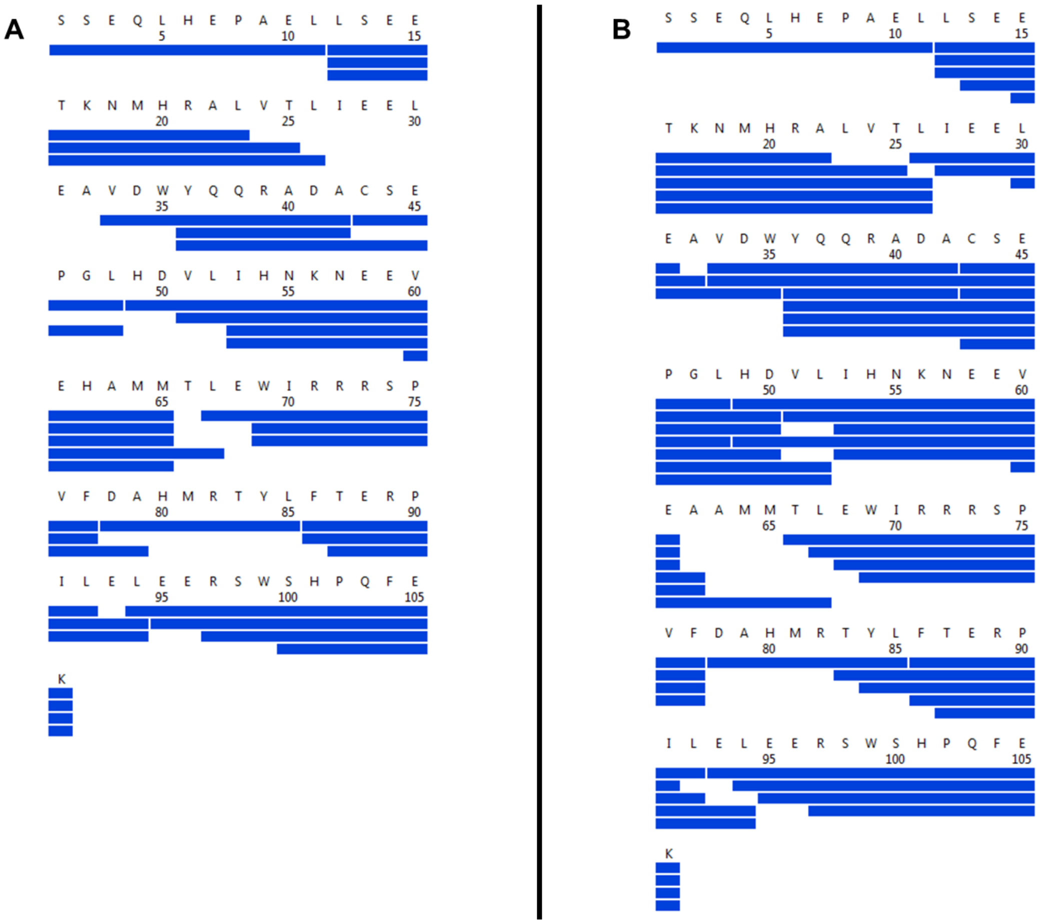


Figure 4.4: HDX Coverage maps of the peptides found both within filled and empty EncFtn samples. Each map has had the starting methionine removed

for the sake of data processing. (A) Coverage for the 94.3% from 25 peptides in the WT Peptides. (B) 51 peptides in the H63A samples. Analysis of peptides between timecourses resulted in 13 comparable peptides providing a total of 87.7% coverage with 1.51 redundancy

A comparison between the two timecourses revealed that both digests shared a total of 16 comparable peptides providing a total of 87.7% coverage with 1.51% redundancy. Due to the mutation, it was not possible to do a 1 to 1 comparison of a peptide containing the H63A residue. The uptake chart for each of the 16 shared peptides can be seen in Figure 4.5.

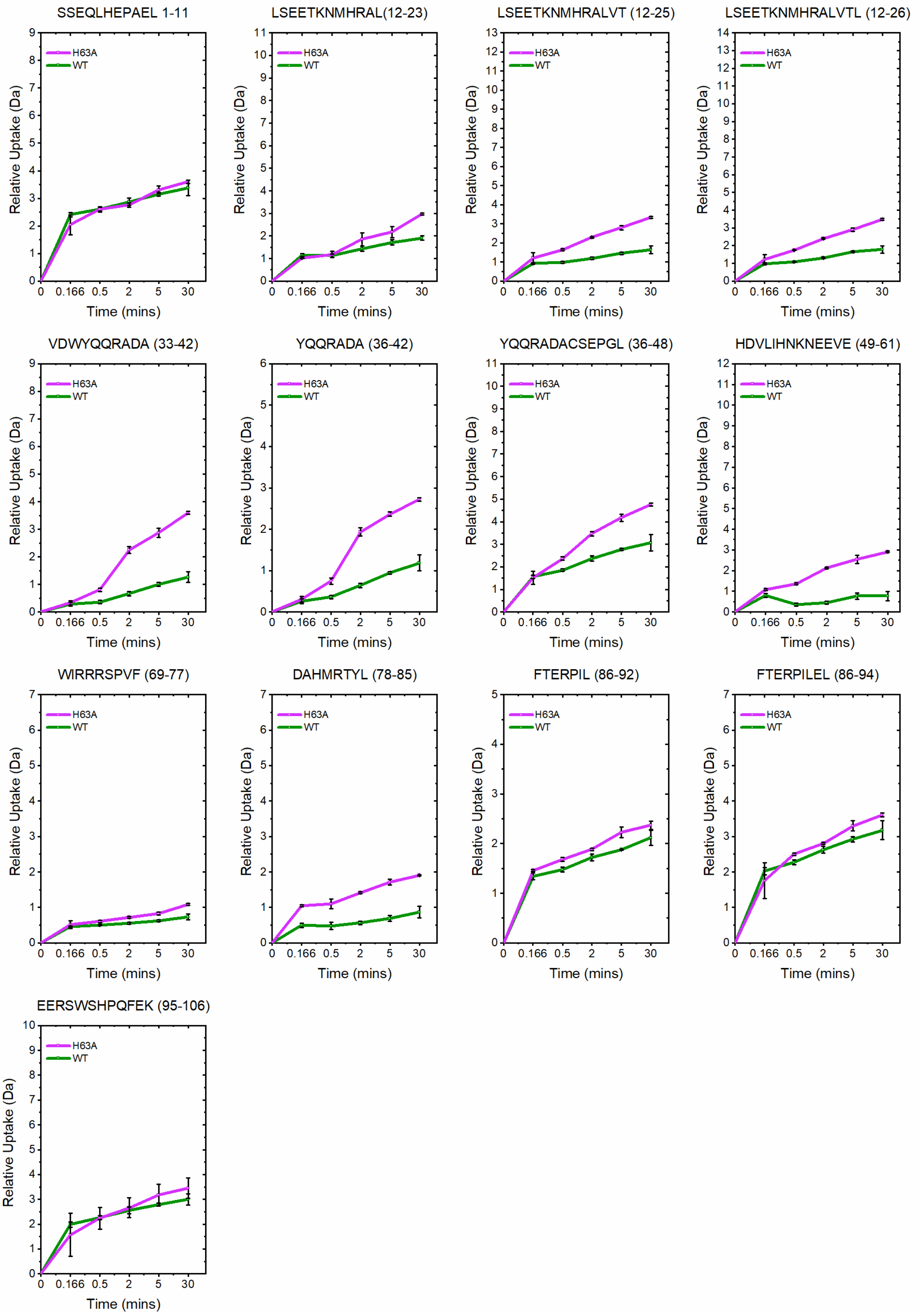


Figure 4.5: HDX Exchange kinetics of shared peptides between EncFtn-WT (Green) and EncFtn-H63A (Magenta). Black points represent an individual timepoint (T1: 10s, T2: 30s, T3: 2mins, T4: 5mins, T5: 30mins)

The DynamX results were used to calculate the deuterium uptake difference between the two the variants across all timepoints. This data was then processed into a butterfly plot (Figure 4.6).

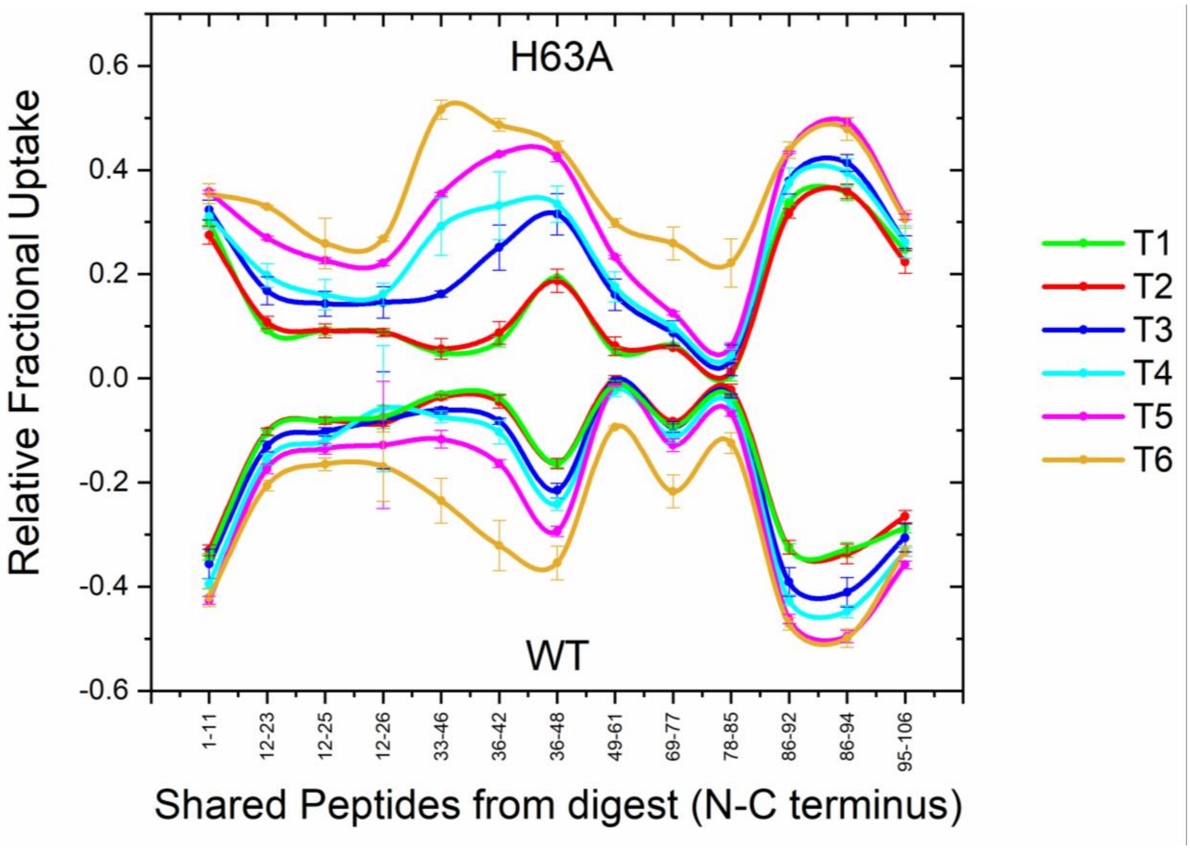


Figure 4.6: Butterfly plot of relative fractional uptake of all peptide across all timepoints 10s, 30s, 2 mins, 5mins and 30 mins respectively. An additional 240 minute timepoint was taken which is represented as T6. Comparability between 13 peptides common to EncFtn-H63A and EncFtn-WT (along the x-axis). The colour code for the timepoints T1 (10 seconds) T2 (30 seconds) T3 (2 minutes) T4 (5 minutes) T5 (30 minutes) and T6 (240 minutes). The Relative fractional uptake was calculated by dividing the deuterium level (in Da) by the total number of labile backbone amide hydrogens (the number of amino acids, minus prolines residues and minus 1 for the N-terminal amide)

A subsequent difference plot was produced to assess the level of difference in deuteration (Figure 4.7A).

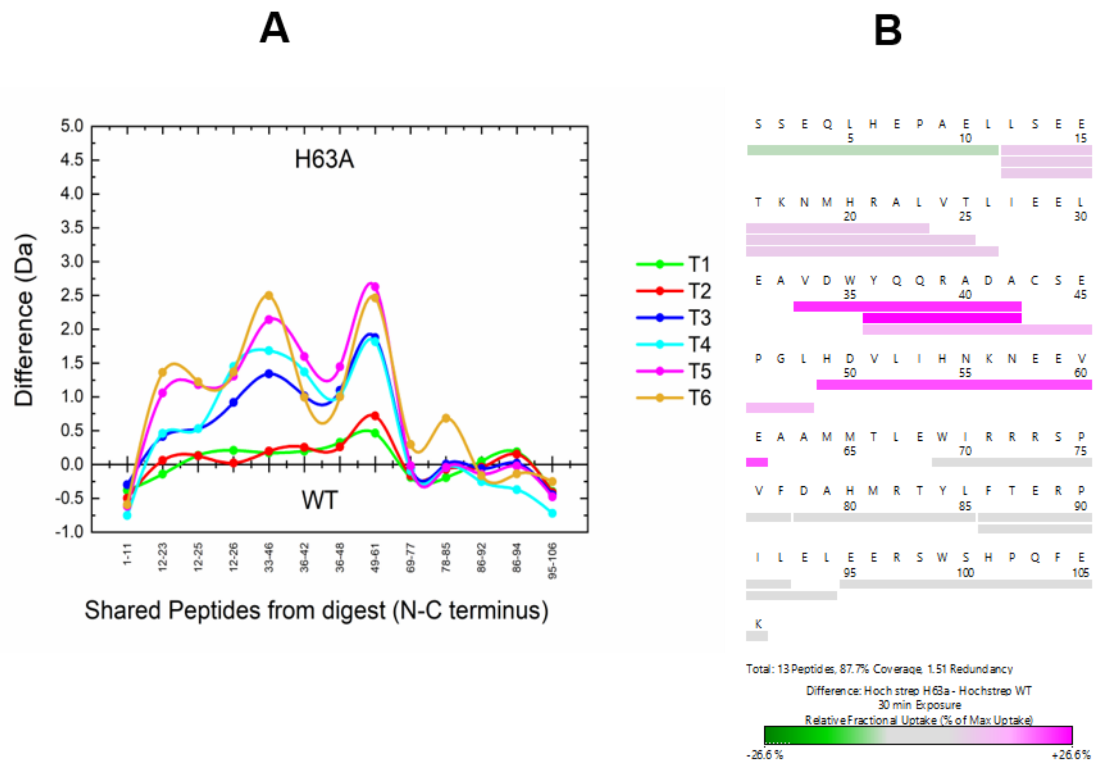


Figure 4.7: (A) Difference plot between EncFtn-WT vs EncFtn-H63A across all timepoints 10s, 30s, 2 mins, 5mins and 30 minutes respectively. An additional Tmax time point was taken at 4 hours which is represented as T6. (B) A protein coverage map illustrating the fractional difference in deuterium incorporation (EncFtn-H63A – EncFtn-WT) at 30 minutes deuteration

4.6 Structural analysis of the EncFtn differential HDX

The differential analysis revealed that there was a slight increase in deuteration in the H63A samples at the N-termini and centre region (SI Table 4.1 & 4.2). Significance testing using unpaired student t-test found most peptides were not significantly different at 30 minutes of labelling (SI Table 4.3).

However, there were three peptides, which demonstrated a significant difference in deuteration between H63A and WT. The peptide corresponding to residues ^{33}VDWYQQRADA^{42} had an RFU of 0.14 ± 0.021 in WT and 0.4 ±

0.006 in H63A. Similarly, the peptide ^{49}HDVLIHNKNEEVE^{61} displayed significantly higher exchange rate in the H63A variant - RFU values of 0.06 ± 0.007 in EncFtn-WT and 0.25 ± 0.014 in the EncFtn-H63A (Figure 4.7). For a complete overview of the HDX data, see (SI Figure 4.1-4) (SI Tables 4.1 and 3.2). A slight increase in deuteration was observed within the H63A variant of 0.15 RFU near the N-termini across the ^{12}LSEETKNMHRA^{223} peptide

The levels of deuteration were highly reproducible with three technical repeats for each timepoint and across three biological repeats for both WT and H63A samples. To assess the level of significant difference between WT and EncFtn-H63A samples an unpaired T-test was performed on each highlighted peptide after 30 minutes of labelling. The results of the T-test revealed that the differences between the ^{49}HDVLIHNKNEEVE^{61} and ^{33}VWQQRADA^{42} peptides were statistically significant, where across the rest of protein the exchange was found not to be statistically significant different. These T-tests were applied across all the biological and technical repeats (Table SI 4.3).

The differential HDX analysis clearly demonstrates that two specific regions of the H63A variant display increased deuterium uptake when compared to WT protein. To try and understand the structural reasons for this difference differential coverage was mapped onto the crystal structure of the EncFtn-WT monomer (Figure 4.8).

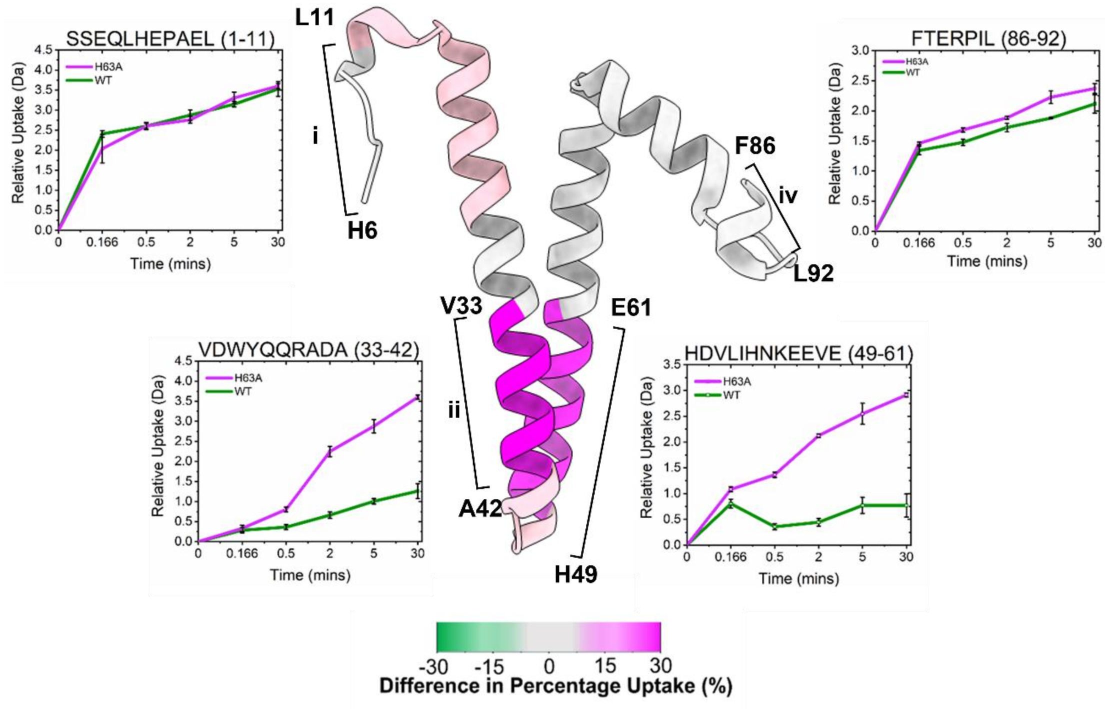


Figure 4.8: The difference in H/D exchange at 30 minutes of D_2O labelling mapped on to the monomer crystal structure of EncFtn-WT. Regions displaying increased exchange in H63A are coloured pink. HDX uptake plots are shown for the four peptides of interest (i–iv). PDB ID: 5N5F

The overlain data showed that there were similar levels of uptake within the N- and C-termini within both dimer forms after 30 minutes of deuterium labelling (Figure 4.8, i and iv). For examples, at the N-termini, SSEQLHEPAEL (1-11) demonstrated a RFU of 0.37 ± 0.031 within EncFtn-WT and 0.39 ± 0.007 in EncFtn-H63A (Figure 4.8 i). Likewise, at the C-termini peptide, FTERPIL (residues 86-92), there was a 0.42 ± 0.03 increase in WT and 0.47 ± 0.016 in H63A (Figure 4.8 iv). The mapped structure revealed that the regions of increased deuteration in the H63A variant covered two sections of the α-helices either side of the central hairpin suggesting these regions deprotected from exchange in the H63A variant.

4.6.1 Solvent accessible surface area analysis of FOC and Non-FOC interfaces

To obtain a more detailed analysis of the solvent accessibility of both the FOC and Non-FOC dimer subcomplexes, we analysed both interfaces using PDBePISA Solvent accessible surface area (SASA) tool[139,140].

The PISA software uses the PDB crystal structure to identify crystal interfaces and predict the formation of higher order structures. The software also analyses the assemblies present within the crystal structure determines the residues within each interface, and calculates interfacial areas. In doing so it can be used to calculate the surface exposure of all residues within the two distinct dimer subcomplexes. Sites which are highly exposed are considered solvent accessible whereas sites which are highly buried are considered solvent inaccessible. Thus, it is possible to calculate the solvent accessibility of the protein sequence for both the FOC and Non-FOC dimer forms present within the crystal structure of the EncFtn decamer (Figure 4.11 and SI Tables 4.).

PISA analysis of the FOC dimer subcomplex of EncFtn revealed that the N- (1-10) and C-termini (76-97) of this assembly show a high degree of solvent exposure, whereas the rest of the FOC dimer assembly (residues 16-75) shows high degrees of buried residues (Figure 4.9 and SI table 4.4).

In contrast, the inverse is observed for the Non-FOC dimer interface. In this assembly a high number of buried amino acids are situated in the N- and C-termini, and the central region (45-57) displays high solvent exposure (Figure 4.9 and SI Table 4.5).

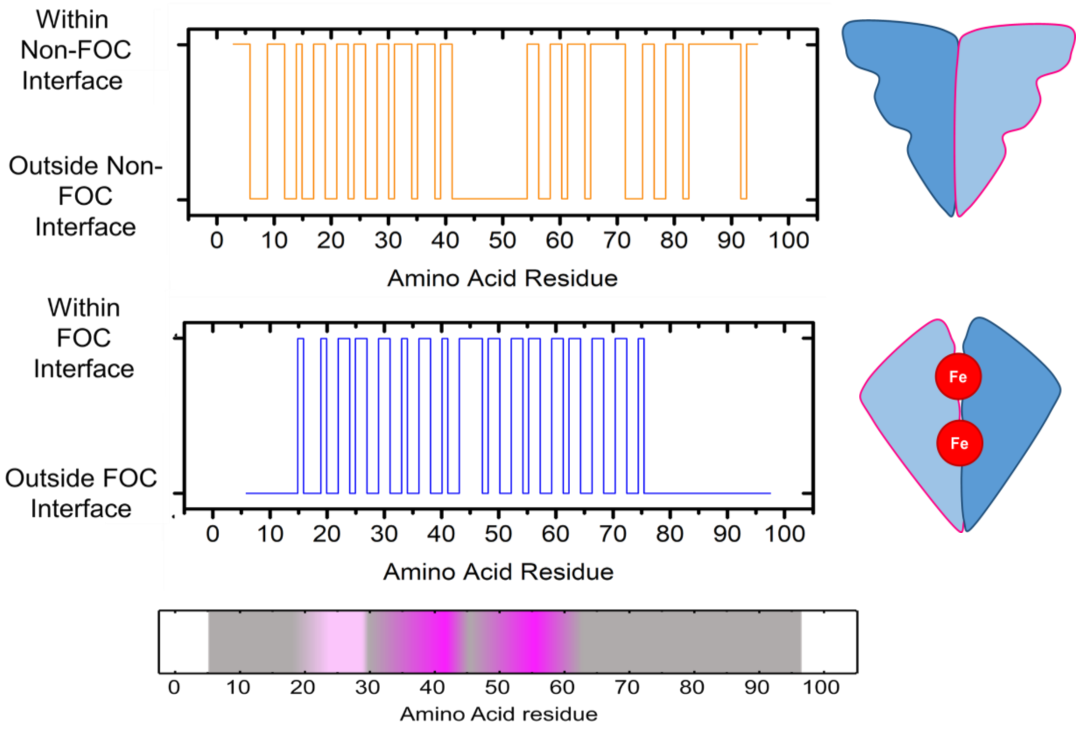


Figure 4.9: SASA analysis of for both the FOC and Non-FOC dimer forms. The difference in exchange at 30 minutes of D2O labelling has been placed between interface maps. Regions displaying increased exchange in H63A are coloured pink

With the SASA analysis collected, the next step was to map the data across the crystal structures of both dimer forms and the differential HDX-MS data to see if the regions of higher solvent exposure aligned with the region of higher deuteration in the H63A variant (Figure 4.10 & 4.11).

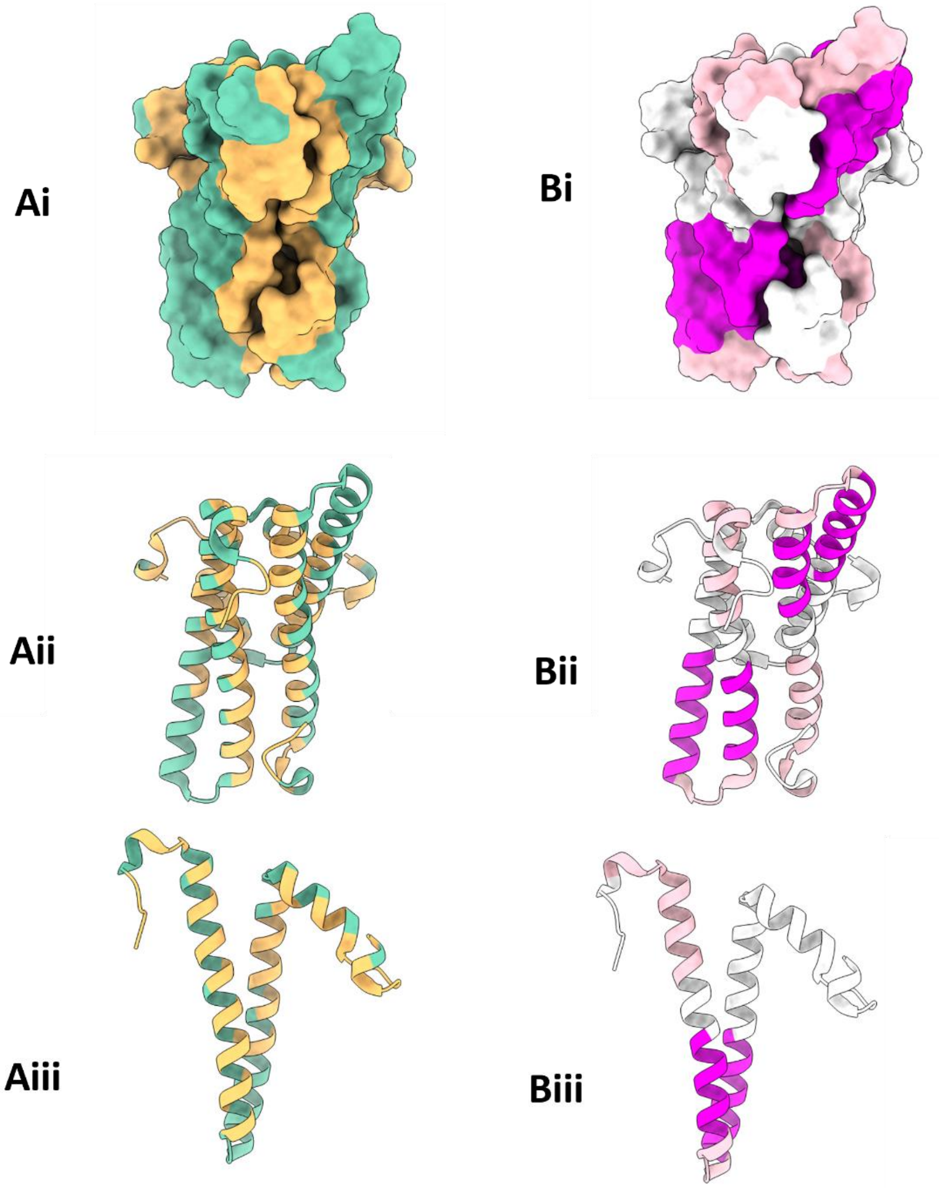


Figure 4.10: (Ai) SASA analysis of the Non-FOC dimer subcomplex overlain the crystal structure of the non-FOC dimer subunit, green represents solvent exposed residues, yellow represent buried residues, (Aii) a cartoon representation of the non-FOC dimer subcomplex (Aiii) the monomer subunit. Bi) differential HDX analysis with key regions of increased H63A deuteration highlighted in magenta, overlaid the non-FOC dimer subcomplex. (Bii) a cartoon representation of the non-FOC dimer subcomplex (Biii) the monomer subunit PDB ID: 5N5F

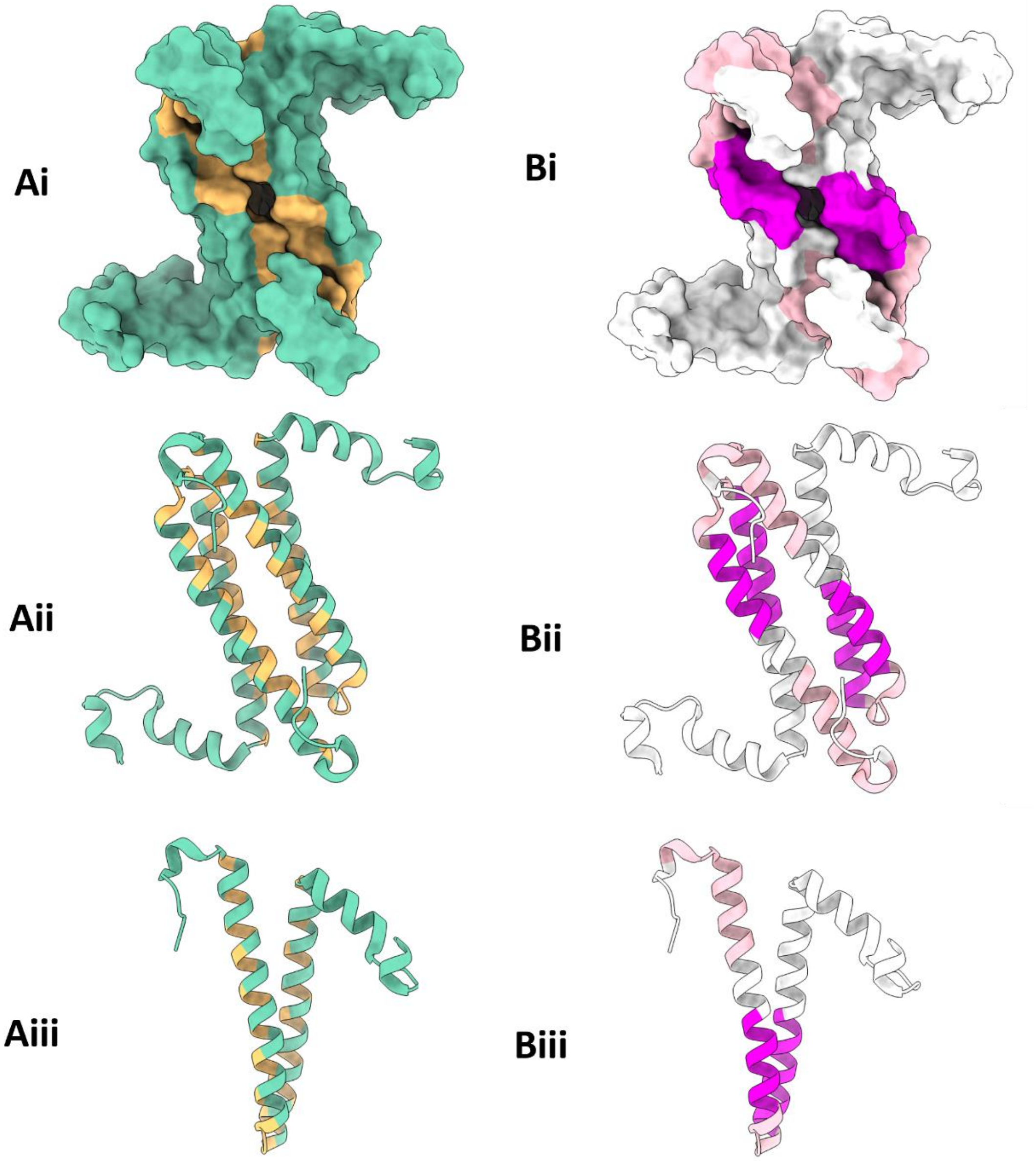


Figure 4.11: (Ai) FOC Dimer SASA analysis overlain the crystal structure of the FOC dimer subunit green represents solvent exposed residues, yellow represent buried residues, (Aii) a cartoon representation of the FOC dimer interface (Aiii) the monomer subunit. (Bi) differential HDX analysis with key regions of increased H63A deuteration highlighted in magenta, overlain the FOC dimer subunit (Bii) a cartoon representation of the FOC dimer subunit (Biii) the monomer subunit PDB ID: 5N5F

Next the crystal structure region of significantly different exchange ^{33}VDWYQQRADACSEPGLHDVLIHNKNEEVE^{61} was isolated and each dimer subcomplex form was compared to the differential HDX data (Figure 4.12). From this data an extended region of exposed residue ranging from A45-H57 in the Non-FOC subcomplex aligns with the region of increased deuteration in EncFtn-H63A. In contrast, these residue in the FOC dimer are mainly buried and dramatically less exposed.

Thus, it appears that exposed residues in the non-FOC dimer subcomplex are deprotected in the H63A variant. This observation strongly suggests that the Non-FOC dimer is the dimer form present within the EncFtn-H63A. This suggests the formation of the FOC between monomer subunits is followed by association of Non-FOC dimers which then form the EncFtn Decamer complex.

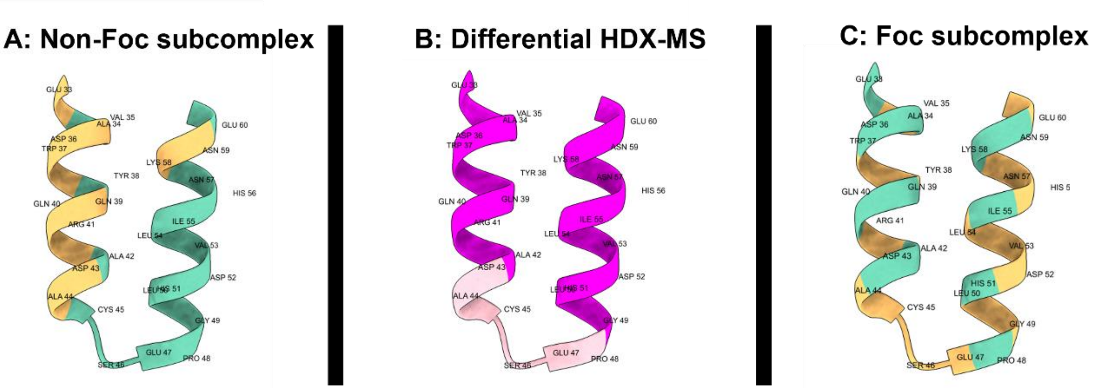


Figure 4.12: The isolated 33VDWYQQRADACSEPGLHDVLIHNKNEEVE60 aa region of Hoch-EncFtn. The first alpha helical region spans 33VDWYQQRADA42 the hair pin loop spans 43CSEPGL48 and the second alpha helical region spans 49HDVILIHNKNEEVE60. (A) SASA analysis of the Non-FOC interface overlay. (B) Differential HDX-MS data overlay (C) SASA analysis of the FOC interface overlay PDB ID: 5N5F. Numbering has been offset of two N-terminal amino acids

4.7 Conclusion

The differential HDX-MS analysis of WT and H63A EncFtn supports the hypothesis that the decameric assembly of EncFtn progresses via the addition of the Non-FOC dimers. These findings are supported by native ion mobility studies that revealed that the dimer form present in the native MS analysis of both WT and H63A EncFtn have CCS values similar to the calculated CCS of the non-FOC dimer subcomplex (Figure 4.13)[117].

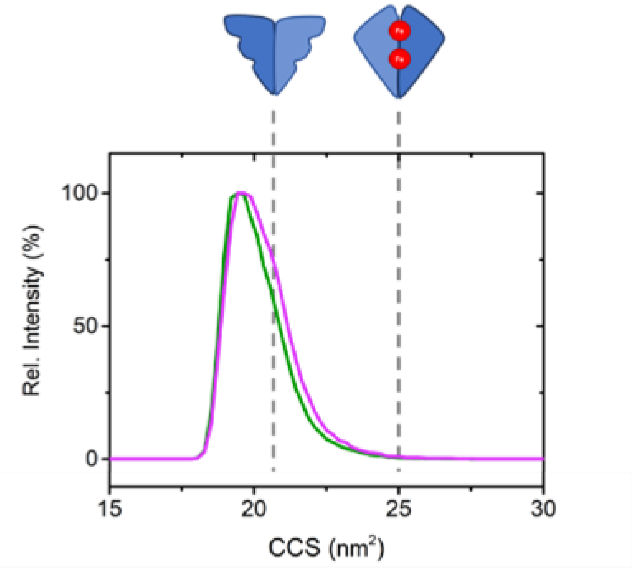


Figure 4.13: Ion mobility drift time profiles of EncFtn-WT (green) and EncFtn-H63A (purple). Non-FOC and FOC dimer are shown in dashed grey lines and were calculated using IMPACT software. Data collected by Dr J. Ross

It is possible that Non-FOC dimer association occurs through the coordination of iron by FOC ligands, strengthening the FOC interface.

These results highlight the dynamic nature of the FOC region within EncFtn proteins and suggests a dual role in both mediating the oligomerisation of the complex and catalytic activity of these enzymes. There are many classes of enzymes whose active sites rest between the subunits of a larger oligomerization state. Furthermore, many metalloproteins rely on metal mediated multimerization[141–143].However, this is the first instance of a ferritin whose assembly is mediated by iron-binding. Although the HDX-MS data does provide a great deal of insight regarding ferritin assembly it is still limited by the peptide level resolution provided by bottom HDX-MS analysis. If single amino acid level resolution were obtained it would be interesting to identify which amino acids are becoming exposed after the single point mutation. To achieve this, a top-down HDX workflow could be utilized[144–146].

Chapter 5 Structural HDX-MS study of an Encaspulin: EncFtn complex

Pore dynamics and asymmetric cargo loading in an encapsulin nanocompartment.

Data presented in this chapter is currently in preprint and approved for publication in Science Advances[128]

5.1 HDX Analysis on viral capsid and BNCs

Although BNCs are a relatively recent discovery, significant HDX-MS work has been conducted on viral capsids[147,148]. Whilst structural information is typically derived via crystallography and electron microscopy both techniques face challenges monitoring dynamic assemblies such as capsids. For this reason, HDX-MS implementation has been used to provide an effective solution. In the context of viral capsid studies HDX-MS has successfully identified critical regions of dynamic conformation change[147–150], measured local reorganization throughout a viral capsid's structure and can reveal conformational heterogeneity[88]. Geometrically, BNCs share an association with viral capsids with both having icosahedral shapes and sharing the same rotational axes. The primary difference between a BNC-like Enc and viral capsids is size. BNCs are significantly larger in diameter (21-42nm) than viral capsids (4-8nm)[126]. The increase in size presents unique challenges and require significant optimization of the labelling, digestion, and separation steps.

5.1.1 Symmetry axis associated poresites

The icosahedral geometry of T=1 encapsulins provides the Enc shell 3 unique rotational symmetry axes. It is thought that these axes may be acting as poresites to allow for the transport of substrates into the interior Enc lumen[151]. The rotational symmetries consist of 5-fold, 3-fold and 2-fold symmetries as previously shown in Figure 1.21.

Recent Cryo-EM work from our group has determined the structure of the EncFtn complex to ca. 3.5Å. This study also identified that the 5-fold site as a region that displays a significant degree of conformational flexibility[128] and the encapsulin can adopt both an open and closed conformation of the 5-fold pore (Figure 5.1). However, there was not sufficient evidence to accurately characterize this dynamic behaviour. Here we use HDX-MS analysis to support these Cryo-EM studies.

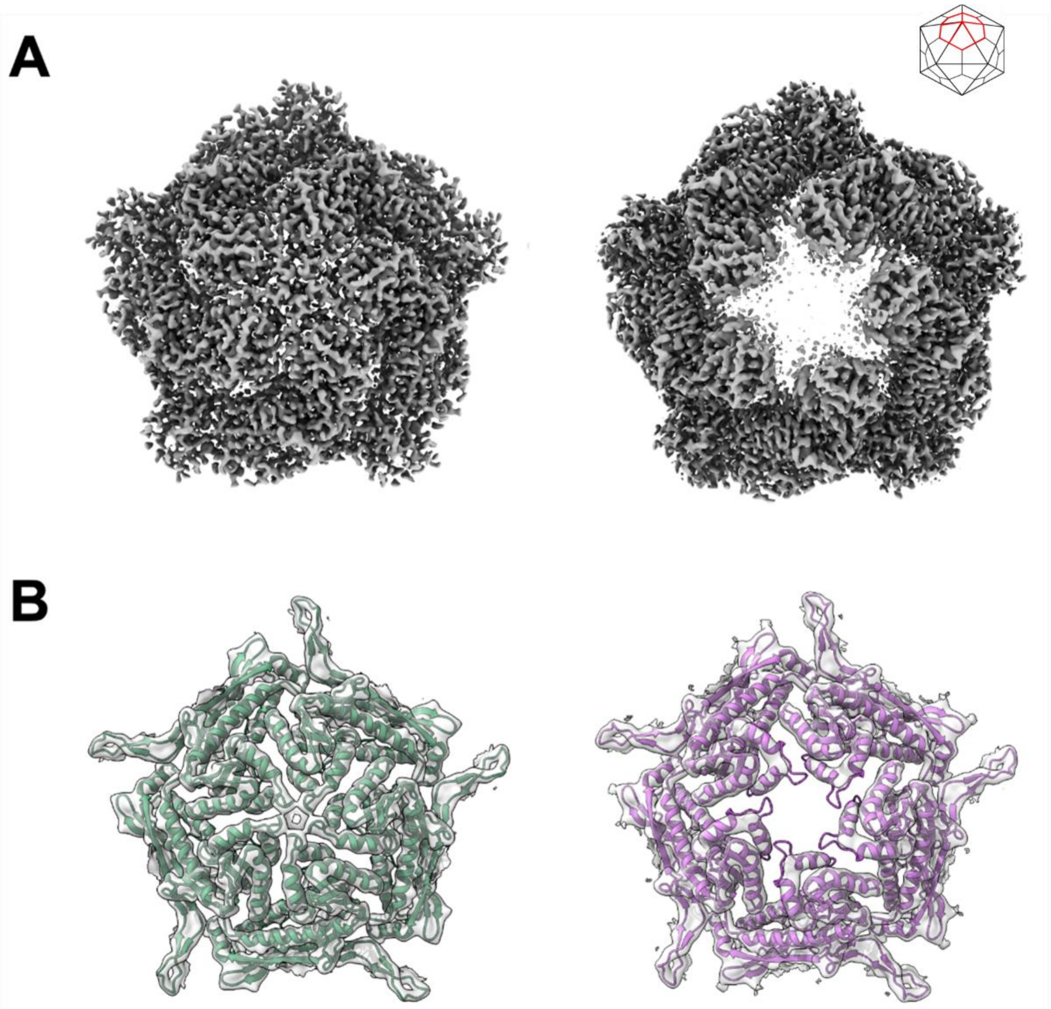


Figure 5.1: Electronic potential maps and models from masked 3D-classification and refinement centred on the five-fold symmetry axis of the symmetry expanded icosahedral reconstruction A: Cryo-EM maps of the ‘closed’ (left) and ‘open’ (right) pentamer conformations from symmetry expansion of the icosahedral five-fold axis. B: The Gaussian smoothed cryo-EM maps with docked models of the ‘closed’ (green, left) and ‘open’ (purple, right) conformations. Smoothed EM maps allow easy visualization of the docked secondary structure. Cryo EM images taken by Dr Jennifer Ross[128]

5.1.2 EncFtn cargo protein orientation

As mentioned in section 1.7.1, questions remained regarding the Enc-EncFtn assembly and how the EncFtn cargo protein orientates itself within the Enc interior space. The tetrahedral symmetry results in two distinct EncFtn environments within the encapsulin interior. The first appears to align at the symmetry axes of the 5-fold pore. The second environment rests between the 3- and 5-fold axes where the Cryo-Em data reveals the presences of two EncFtn decamers. The presence of two alternative EncFtn environments challenges previous hypothesis of the arrangement being localised the 5-fold poresite (Figure 5.2).

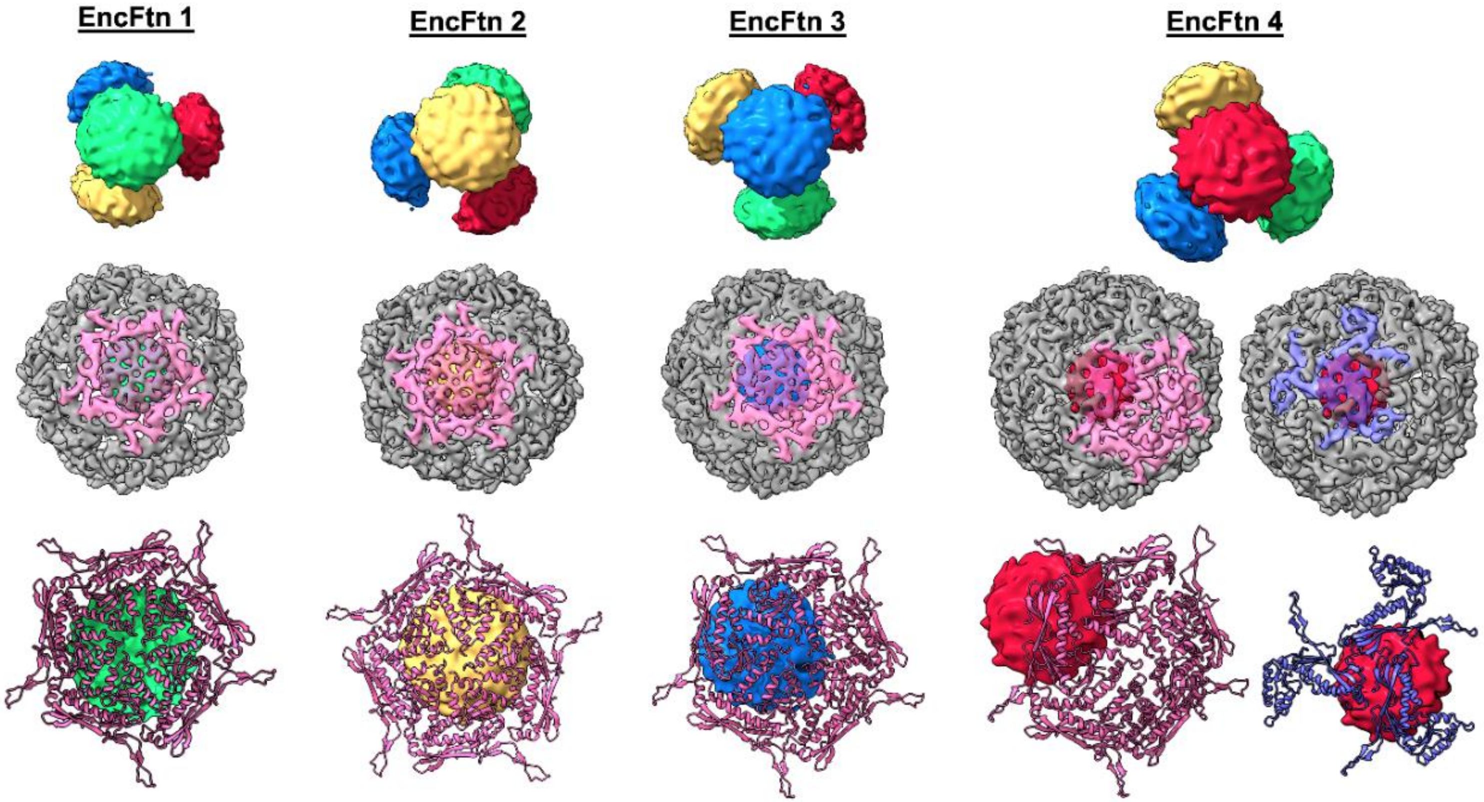


Figure 5.2: Visualization of the distinct EncFtn environments within the icosahedral encapsulin nanocompartment. The four orientations of the EncFtn tetrahedral arrangement. Top panels: the four EncFtn cargo proteins found within Enc

nanocompartment interior. Middle panels: an outside view of the Enc shell with a pink overlay highlighting the 5-fold poresite. The top panel EncFtns place in their corresponding orientations within the Enc interior. The 3-fold poresite has been highlighted with a blue overlay EncFtn 4 does not align with the 5-fold poresite and may interact with the 3-fold poresite. Bottom panel: shows the potential interactions of each Encftn cargo proteins and the highlighted poresites. *Models generated by Dr Jennifer Ross*[128]

Considering the critical role that Enc-EncFtn plays in iron transport, these initial findings raise questions regarding the oxidation and storage of iron within the Enc-EncFtn interior space. This is due to the placement of the EncFtn decamer being in non-equivalent positions with some of the decamers being in closer proximity to the pores which potentially limit the diffusion of substrate in the interior space. EncFtn decamers closer to the poresite will exist in a different chemical environment to those which are placed further away. The presence of engaged and unengaged LSs will also impact the diffusion of substrates by providing a “soft” steric barrier to the diffusion of substrates. Previous work has already identified that the ferroxidase activity of encapsulated EncFtn is greater than isolated EncFtn protein[110]. When combined, these findings demonstrated that the complex interactions that occur within the Enc-EncFtn complex enhance the iron oxidation activity of the EncFtn cargo protein. However, more work is required to further elucidate how the mechanics behind this enhanced activity.

5.2 Aim

This chapter details the implementation of bottom-up HDX-MS to explore the exterior and interior of encapsulin nanocompartment.HDX-MS was initially used to further investigate the dynamics of potential poresites previously identified through Cryo-EM analysis. Subsequent differential HDX-MS was performed on Loaded and Empty-encapsulinin order to identify and characterise the interaction between the EncFtn cargo protein's localization sequence and the interior surface of encapsulin nanocompartment.

5.3 Proteolytic digest and LC separation of EncFtn

Both 'Empty' Enc and the "Loaded" EncFtn-Enc complex were expressed and purified in *E. coli* and the Enc assembly was verified by size exclusion chromatography and electron microscopy analysis.

The encapsulin samples (32µM) were then subject to automated HDX-MS using the methodology described in section 2.1. For these experiments 7 timepoints were used. After automated HD exchange and quenching, protein samples were loaded onto a pepsin column for digestion. The resulting peptides were separated by reverse phase chromatography using a gradient of 10 minutes. The digestion conditions used were successful for both Loaded-Enc and Empty-Enc with most peptides eluting between RT 5.5 mins and RT 8.25 mins and no undigested protein being retained (Figure 5.3).

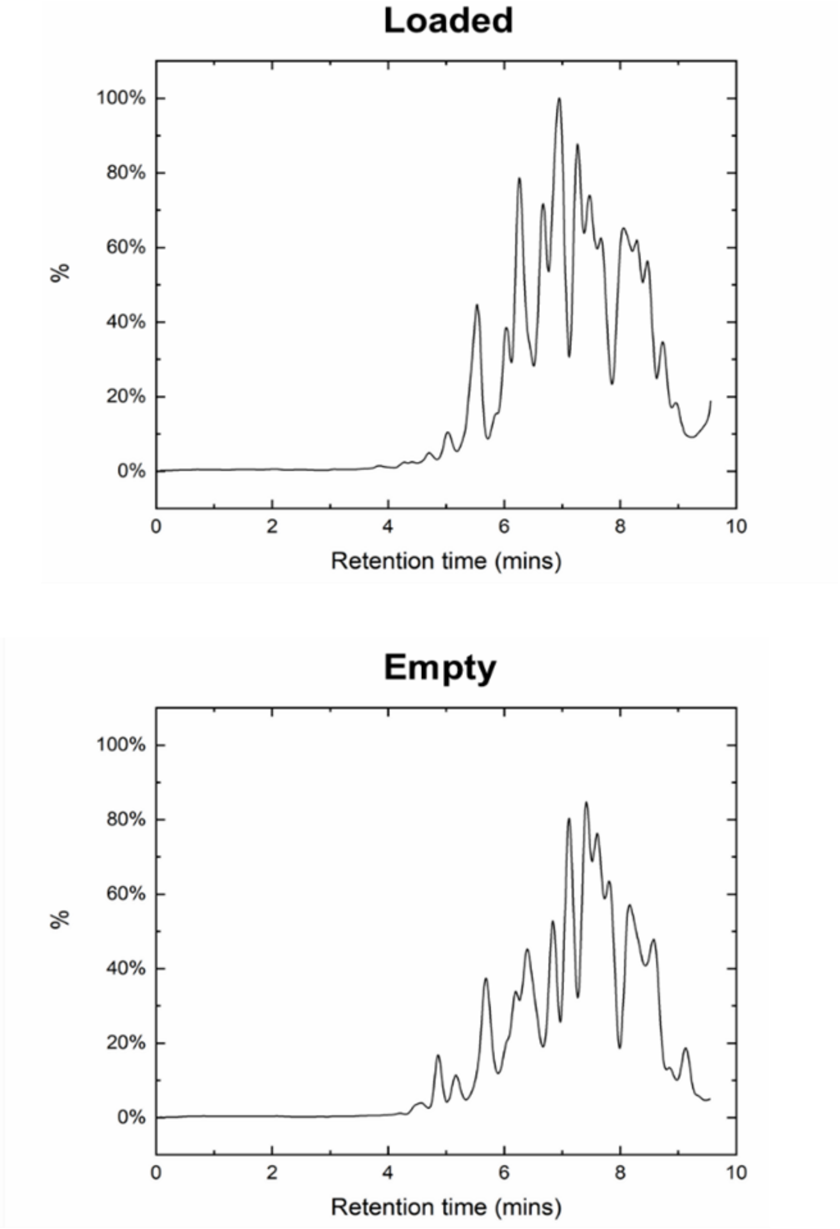


Figure 5.3: Total ion chromatograms for both Loaded and Empty samples. Separation was performed on a Waters Acquity C18 column using a 12 minute step gradient from 5%-95% water – acetonitrile

An undeuterated control underwent digestion to assess the digestion efficiency for both Empty and Loaded Enc (SI Figure 5.1A & 5.1B). These datasets were combined to generate a peptide map that consisted of peptides which were shared between both samples Figure 5.4.

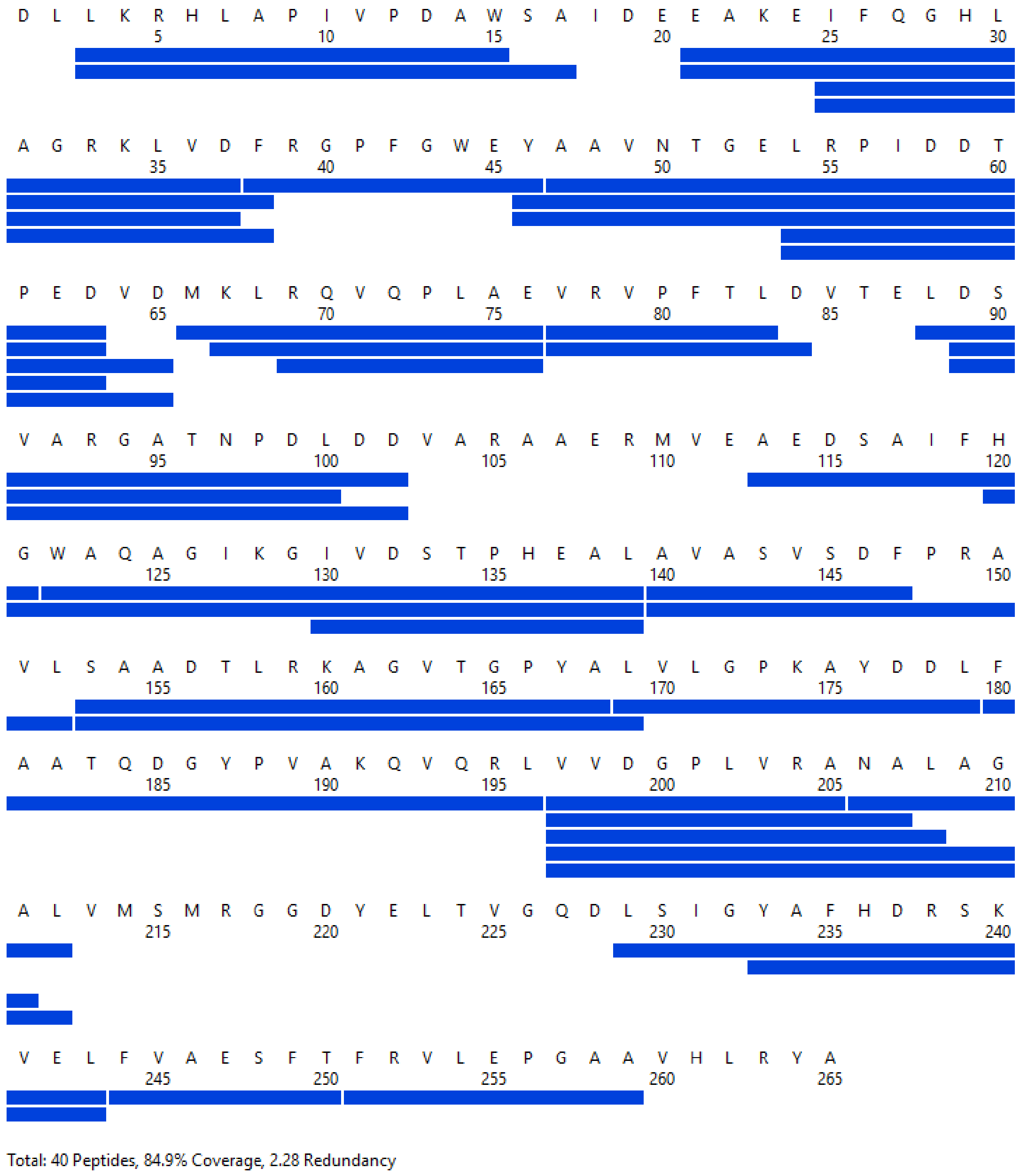


Figure 5.4: HDX Coverage map of the peptides found both within Loaded-Enc and Empty-Enc digests. Peptides observed by MS in all HDX timepoints are represented by blue bars. Pepsin digestion of both variants resulted in 40 shared peptides providing 84.9% protein sequence coverage with 2.28 redundancy

5.4 Poresite dynamics revealed by HDX-MS

Following analysis of the T0 unlabeled controlled, the deuterium uptake for the subsequent labelling timepoints *(T1: 10s, T2: 30s, T3: 2 minutes, T4: 5 minutes, T5 30 minutes and T6 240 minutes)* was generated for the bottom-up data across the full timecourse (see SI Table 5.1). The 240 minutes deuteration coverage map of Loaded-Enc can be seen in Figure 5.5.

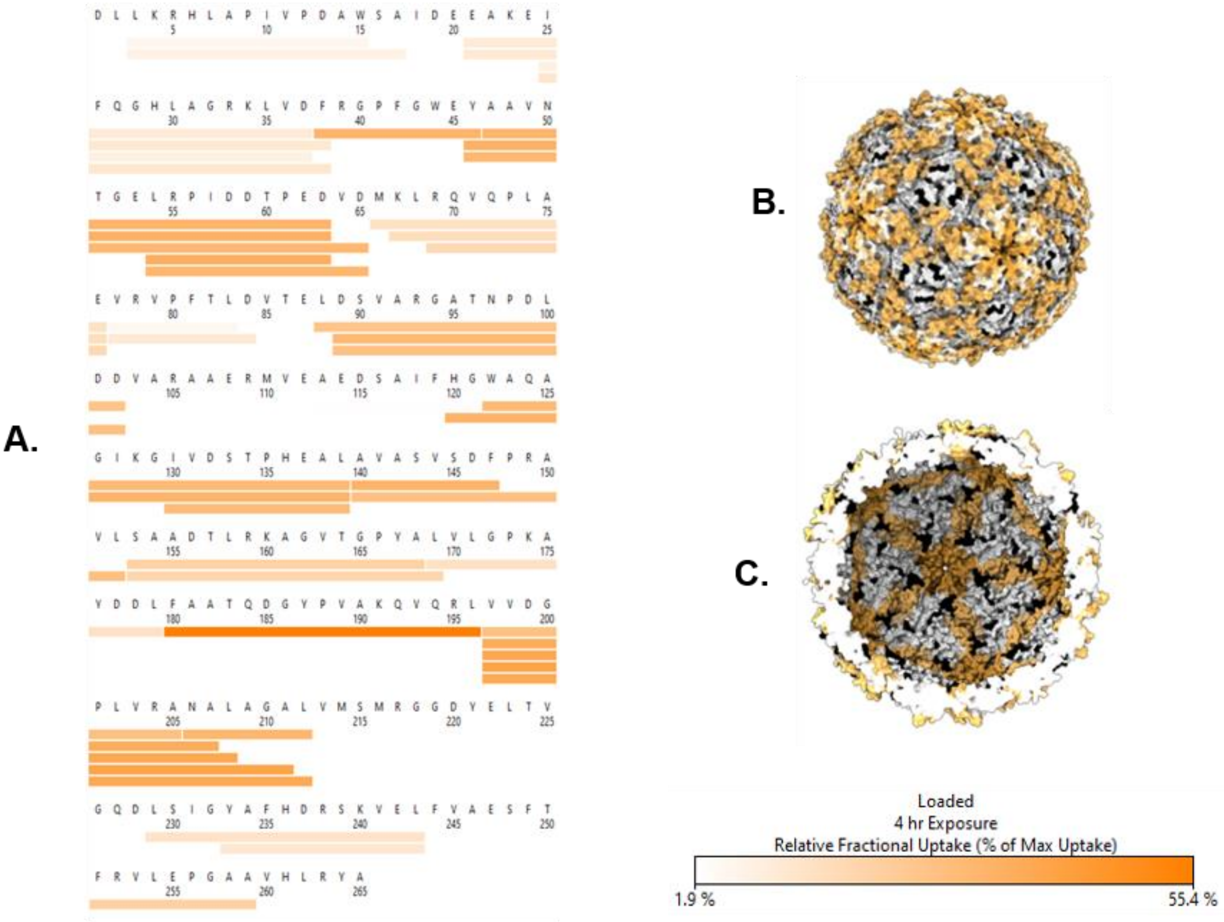


Figure 5.5: (A) HDX Coverage map showing the relativedeuterium uptake for Loaded-Enc after 240 minutes of deuterium labelling. There are 40 observed peptides common to both states Loaded and Empty-Enc, providing 84.9% protein sequence coverage with 2.35 redundancy. (B) The relative fractional uptake of each peptide displayed on the exterior of the 60mer Enc nanocompartment reconstruction. (C) The relative fractional uptake of each peptide displayed on the interior of the 60mer Enc nanocompartment

5.4.1 HDX-MS analysis of Poresite conformational flexibility

Both the interior and exterior deuteration profiles were examined for the 3 theoretical poresite locations within the Enc macro structure (Figure 5.6).

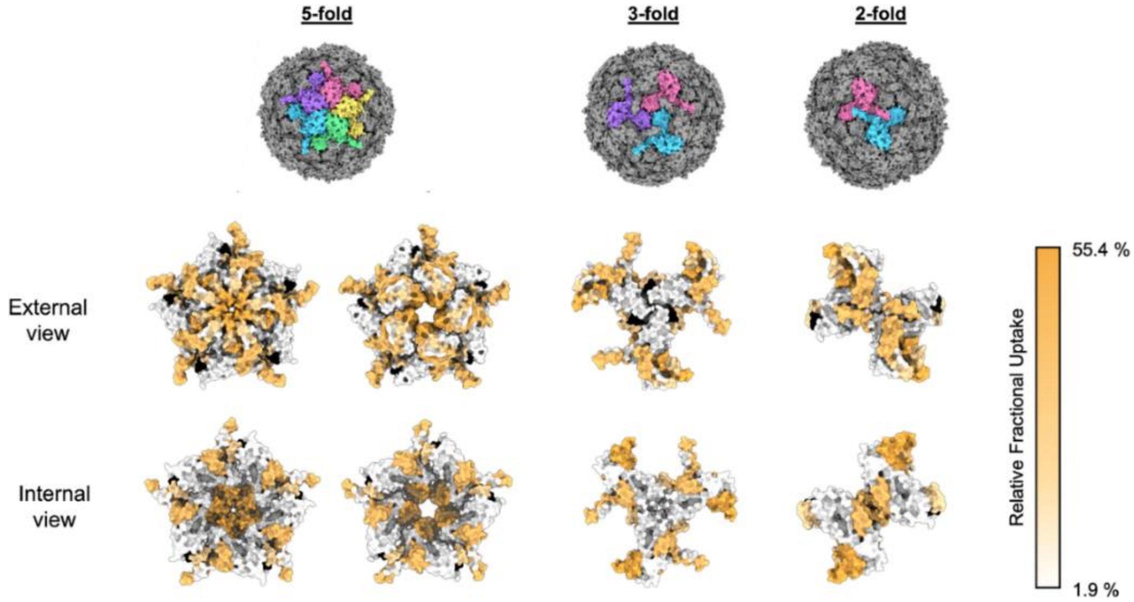


Figure 5.6: HDX-MS fractional uptake at 240 minutes of deuteration mapped to the symmetry axes of the icosahedral reconstruction of the H. ochraceum Enc complex. From top to bottom: The top row depicts molecular surface models of the Enc in three orientations to show the 5-fold, 3-fold and 2-fold symmetry axes. Monomers have been coloured individually (pink, yellow, green, blue and purple) to highlight the symmetry axes. The final two rows display the relative fractional uptake of each peptide from HDX experiments displayed on the 5-fold, 3-fold and 2-fold pores of the Enc nanocompartment. Peptide fractional uptake is shown on a white-to-orange colour scale with a colour key show on the right-hand side of the figure. Areas coloured black correspond to no peptide coverage

5.4.2 Five-fold Poresite analysis

The first proposed poresite, which rests at the centre of the 5-fold symmetry, demonstrated a rapid rate of deuteration both on the exterior and within the interior. The deuteration map of the 240 minutes timepoint has been overlaid on the crystal structure of this symmetry in Figure 5.7.

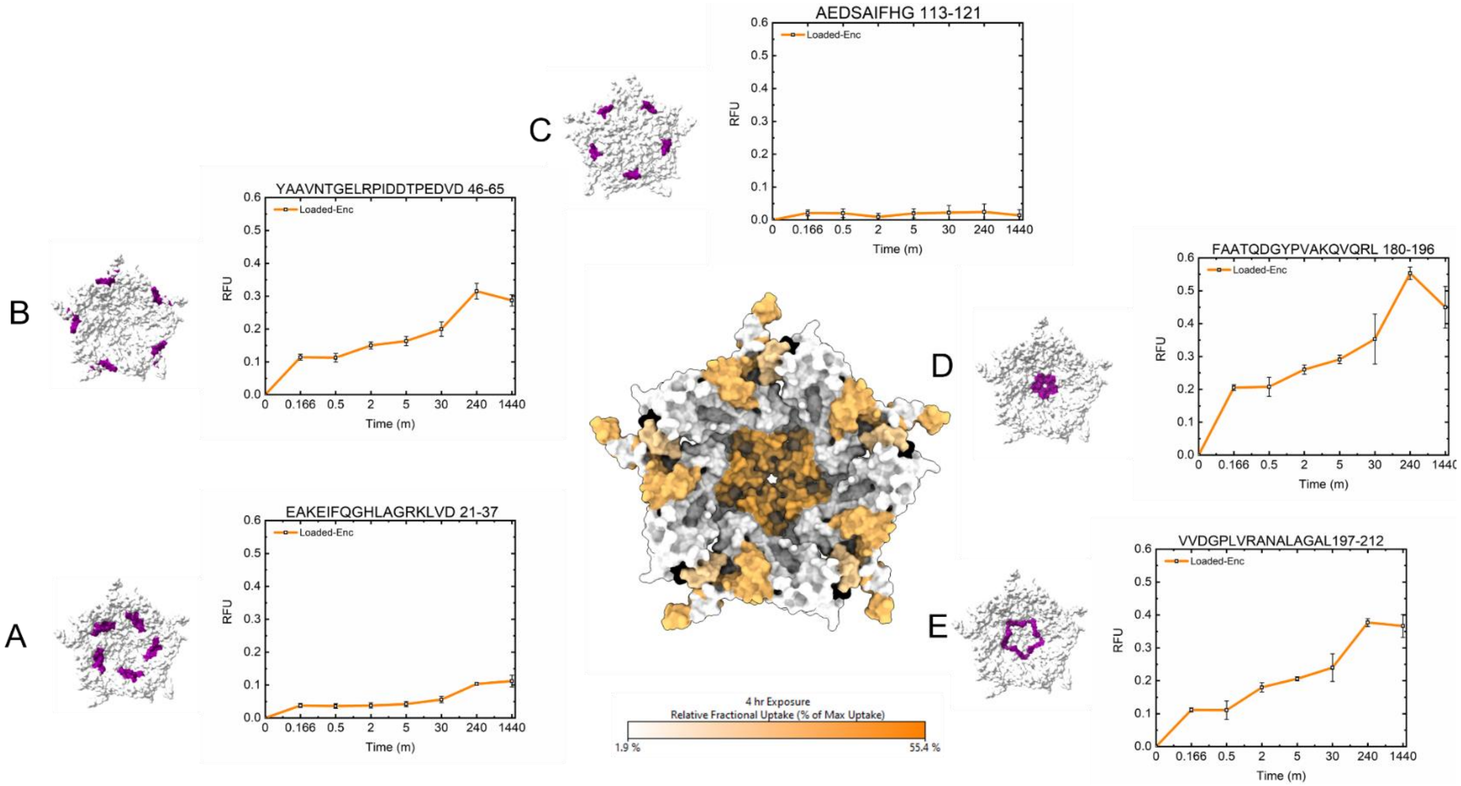


Figure 5.7: Interior view of HDX-MS of Loaded-Enc 5-fold poresite showing the amount of deuterium incorporation after 240 min exposure, coloured according to RFU. (A) residue 21-38, (B) residue 46-65, (C) residue 113-121, (D) residue 180-196, (E) interior residue 197-212

From Figure 5.7 (A) residues 21-38 whose proximity is close to the proposed LS for EncFtn 11.2 ± 0.2%, (B) residues 46-65 which overlays the potential poresite within 2-fold symmetry 31.7 ± 2.2%, (C) Location of residues 113-121 a highly buried exterior exposed peptide 0.2 ± 0%, (D) residues 180-196 the potential 5-fold symmetry poresite 55.4 ± 2.2%, (E) interior residues 197-212 near the potential 5-fold symmetry poresite 37.7 ± 1.2%

The deuteration map show regions of intense deuteration at the centre of the 5-fold pore with the rate decreasing from the centre outwards. Interestingly there are higher rates of deuteration within the interior than the exterior near the centre of the model. These data suggests that this region is highly exposed to labelling solution. This result coupled with Cryo EM data is indicative of a highly dynamic region and based on the diameter of the pore in the open conformation (24 Å), strongly suggests that this region acts as a poresite to allow substrates to enter and interact with EncFtn cargo protein.

Variable exchange rates were measured across the Enc structure with high levels of deuteration occurring around pentameric vertices (Figures 5.4 and 5.5). The highest deuteration spans across the region between amino acids ^{180}FAATQDGYPVAKQVQRL^{196} and lies within the centre of the 5-fold pore. This region consists of a large degree of alpha-alpha hairpin loops with extensive regions which lack any notable secondary structure ^{183}TQDGYPVA^{190}. The lack of secondary structure will result in a lower degree of solvent protection which potentially could result in the higher degree of deuteration across the region. However, looking at the surrounding structure both on the interior and exterior there is a high degree of deuteration at the centre of the poresite (Figure 5.7). This supports the hypothesis that the poresite is highly dynamic and is allowing for deuterium to enter Enc's interior space and is deuterating from the entry site outwards into the interior space.

5.4.3 Three-fold Poresite analysis

The 3-fold poresite demonstrated very low levels of deuteration both on the interior and exterior facing residues. This is indicative of a highly protected region which has a very low level of solvent accessibility and low structural dynamics. These findings suggest that the 3-fold symmetry does not have a highly dynamic nature is probably not acting as a poresite due to high degree protection from deuteration. From Figure 5.8 (A) residues 21-38 whose proximity is close to the proposed LS for EncFtn 11.2 ± 0.2%. (B) residues 46-65 which overlays the proposed poresite within potential 2-fold symmetry 31.7 ± 2.2% (see section 4.4.4), (C) Location of residues 113-121 a highly buried exterior exposed peptide 0.2 ± 0%, (D) residues 180-196 the potential 5-fold symmetry poresite 55.4 ± 2.2%, (E) interior residues 197-212 near the potential 5-fold symmetry poresite 37.7 ± 1.2%

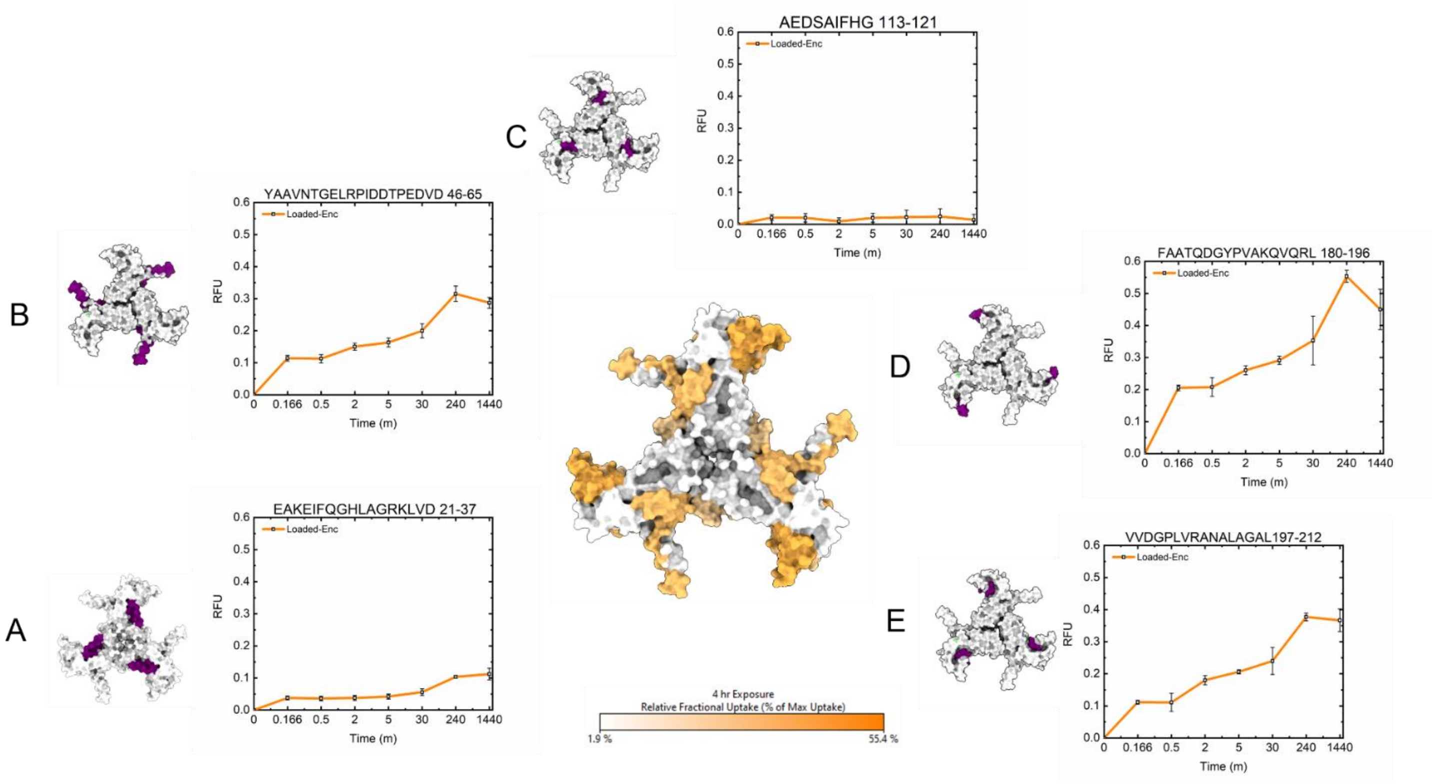


Figure 5.8: Interior view of HDX-MS of Loaded-Enc of 3-fold poresite highlighting the amount of deuterium incorporation after 240 min exposure, coloured according to uptake. (A) residue 21-38, (B) residue, (C) Location of residue, (D) residue 180-196, (E) interior residue 197-212

5.4.4 Two-fold Poresite analysis

The poresite located across the 2-fold symmetry demonstrated higher degrees of deuteration in the interior whilst the exterior demonstrated a lower level of deuteration in comparison to the 5 fold pore (Figure 5.9). Due to the proximity of peptides with high levels of deuteration to the centre of the 2-fold symmetry it is likely that there is a pore-like structure allowing for deuterium to enter and saturate the surrounding region. These results suggest highlight the possibility of a pore in the 2-fold symmetry site. However, the dynamics in this region are less than those displayed by the five-fold poresite.

From Figure 5.9 (A) residues 21-38 whose proximity is close to the proposed LS for EncFtn 11.2 ± 0.2%, (B) residues 38-46, 32.9 ± 0.9%, (C) residues 46-65 which rests at the centre of the potential 2-fold symmetry poresite 31.7 ± 2.2%, (D) residues 180-196, 55.4 ± 2.2%, (E) interior residues 197-212 near the potential 5-fold symmetry poresite 37.7 ± 1.2%.

The 2-fold pore does demonstrate a higher degree of deuteration however the region of higher exchange consists of ^{180}FAATQDGYPVAKQVQRL^{196} and which lays both on the Enc’s exterior and interior spaces. These are large peptides which also span the 5-fold pore as well but unlike the 5-fold pore there is lack of deuterating regions located within the 2-fold pore. For this reason, the deuteration profile cannot be adequately differentiated from the 5-fold poresite, thus its dynamic nature remains to be confirmed.

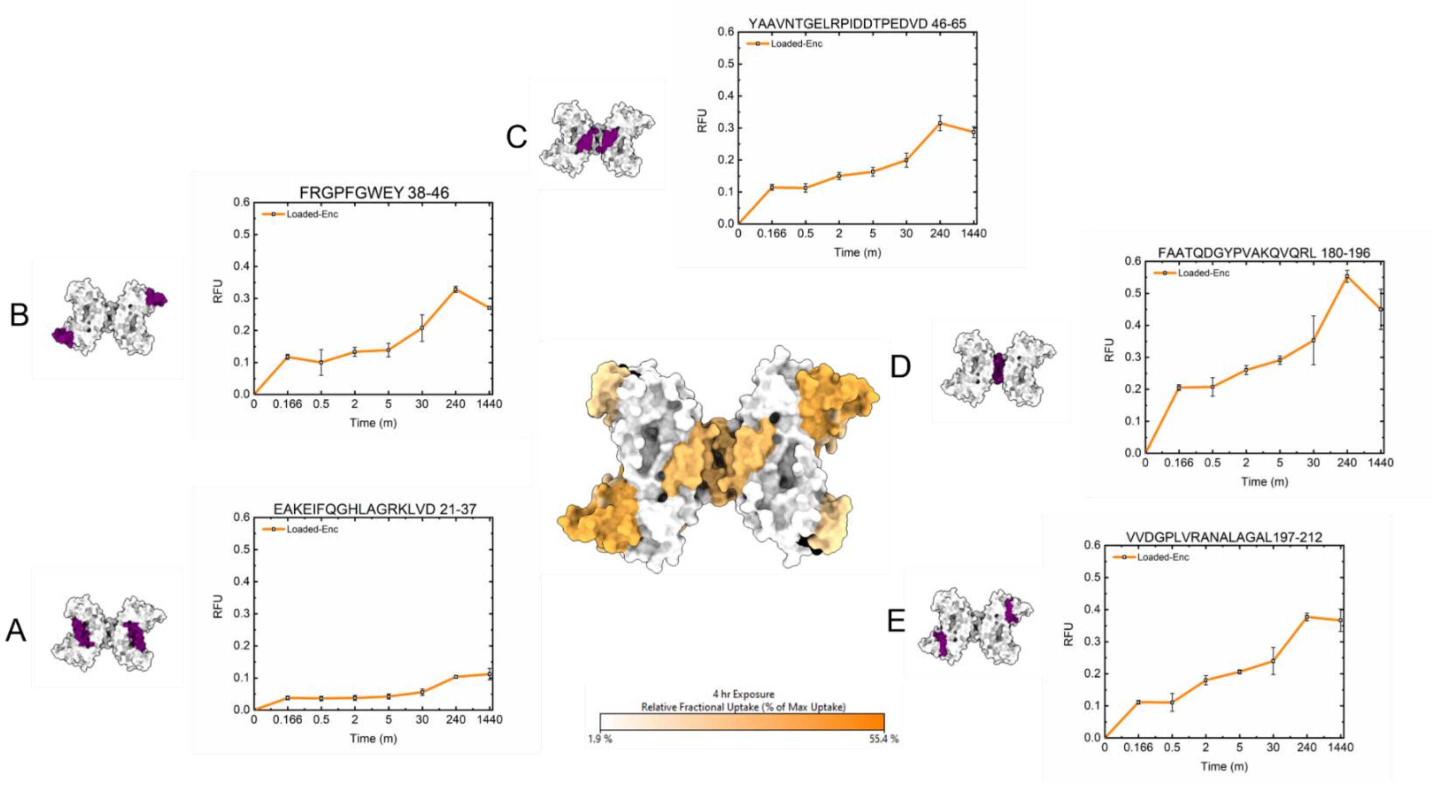


Figure 5.9: Interior view of HDX-MS of encapsulin 2-fold poresite highlighting the amount of deuterium incorporation after 240 min exposure, colored according to uptake. (A) Location of residue 21-38 (B) residue 38-46, (C) residue 46-65, (D) residue 180-196, (E) interior residue 197-212

5.5 Differential HDX-MS study of cargo site localization

The differential HDX-MS analysis of the Empty and Loaded Enc revealed similar exchange profiles between forms (Figure 5.10). It appears that the presence of cargo protein does not dramatically alter the overall dynamics and architecture of the Enc nanocompartment. However, although no significant difference in exchange is observed at earlier timepoints (T1-5), later time points 240mins revealed there is a modest difference in exchange. At 4 hours of labelling Loaded-Enc demonstrates a reduction in exchange in comparison to Empty-Enc. A full overview of the exchange across all timepoints can be seen in a butterfly plot in Figure 5.11.

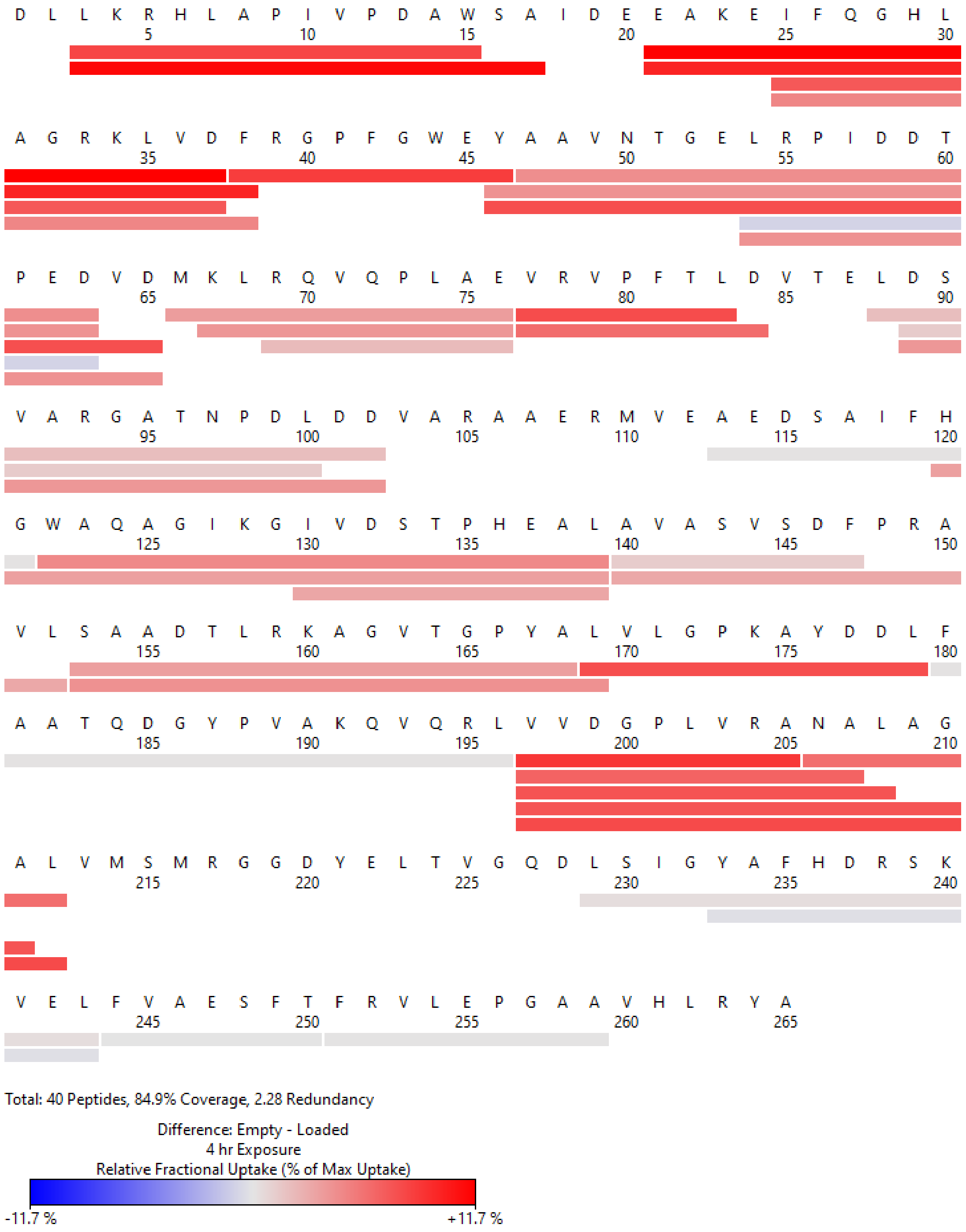


Figure 5.10: Protein coverage map illustrating the fractional difference in deuterium incorporation (Empty-Enc – Loaded-Enc) at 4 hours deuteration

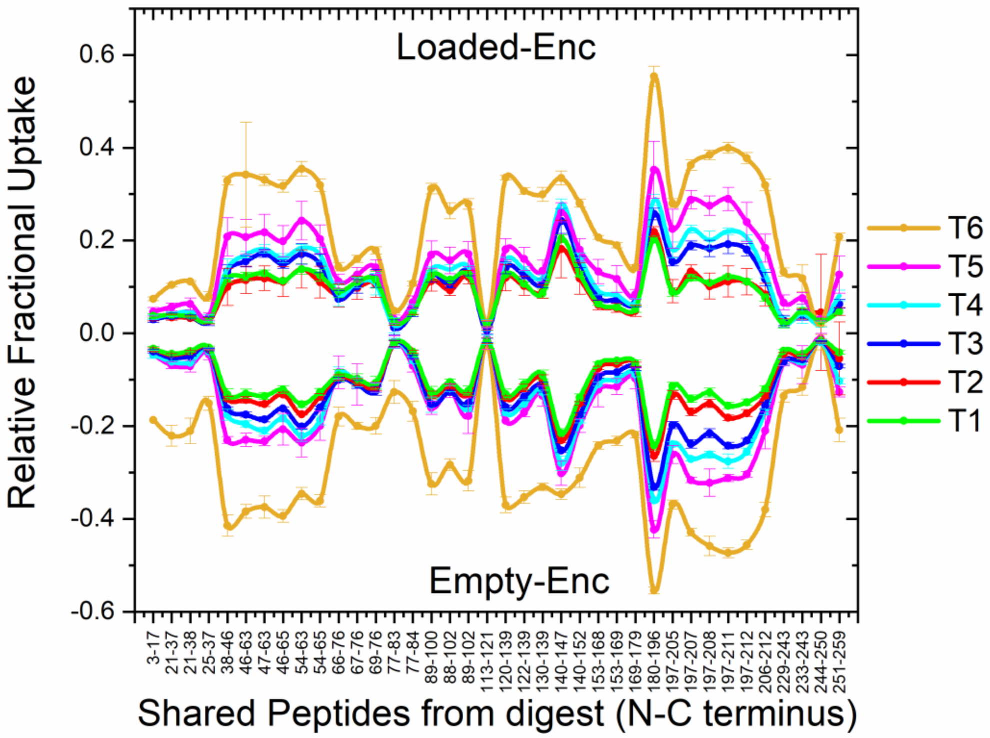


Figure 5.11: Loaded vs Empty-Enc relative fractional uptake butterfly plot of all peptides across all timepoints 10s, 30s, 2 mins, 5mins, 30mins and 240mins respectively. Comparability between 13 peptides common to Loaded-Enc and Empty-Enc (along the x-axis). The colour code for the timepoints T1 (10 seconds) T2 (30 seconds) T3 (2 minutes) T4 (5 minutes) T5 (30 minutes) and T6 (240 minutes). The Relative fractional uptake was calculated by dividing the deuterium level (in Da) by the total number of labile backbone amide hydrogens (the number of amino acids, minus proline residues and minus 1 for the N-terminal amide)

In particular, the regions spanning ^{3}LKRHLAPIVPDAWSA^{17} and ^{21}EAKEIFQGHLAGRKLVD^{37} demonstrated a 12% reduction in exchange within Loaded-Enc and was found to be significantly different to Empty after an unpaired student T-test (SI Tables 5.3-5).

When mapped on the Enc structure this region is located on the interior face of the Enc complex near the 5-fold pore and covers the proposed binding site of the EncFtn LS (Figures 5.10 & 5.11). This further supports the hypothesis that the EncFtn cargo protein may orientate itself to align with this potential poresite (see section 1.6.4). The deuteration profile has also been modelled onto the closed confirmation of the 5-fold poresite with detailed peptides of interest Figure 5.12.

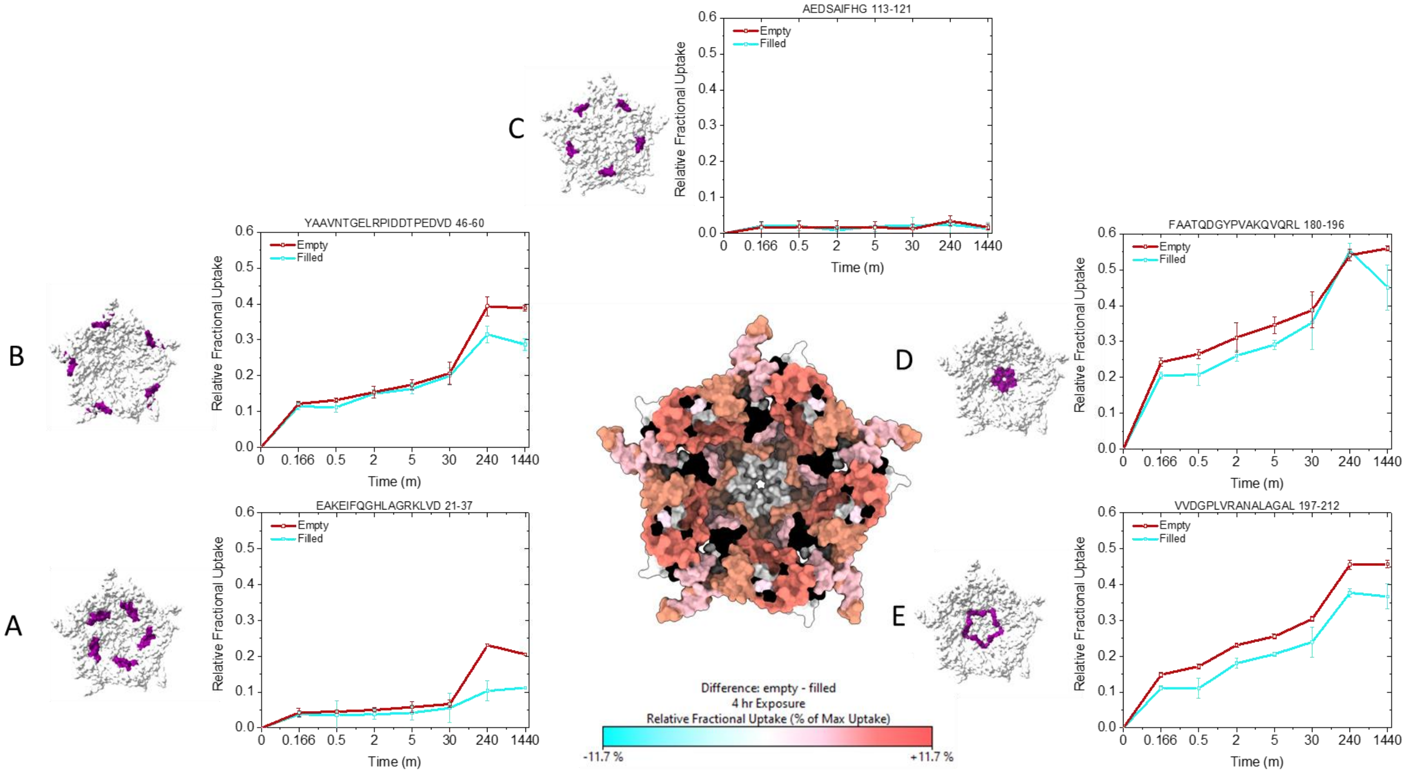


Figure 5.12: Differential HDX-MS/MS of Empty and Loaded Enc showing the difference in relative fractional uptake of deuterium after 240 min exposure, colored according to uptake difference. (A) Location of residue 21-37, (B) residue 46-65 (C) Location of residue 113-121, (D) residue 180-196, (E) interior residue 197-212

From Figure 5.12 (A) Location of residue 21-37 difference of +11.7 ± 1.5%, (B) residue 46-65 with a difference of +7.6 ± 1.8% (C) Location of residue 113-121with a difference of 0.1 ± 0%, (D) residue 180-196 with a difference of 0 ± 0.3%, (E) interior residue 197-212 with a difference of 7.9 ± 1.7%.

The differential HDX-MS analysis revealed that the exchange profiles of the Empty and Loaded Enc peptides did not appear to be significantly different from one another in the earlier timepoints (T1-5). However, at later timepoints (T6-7) a moderate amount of difference is present. The differential exchange profile of the Enc binding pocket is shown in Figure 5.13.

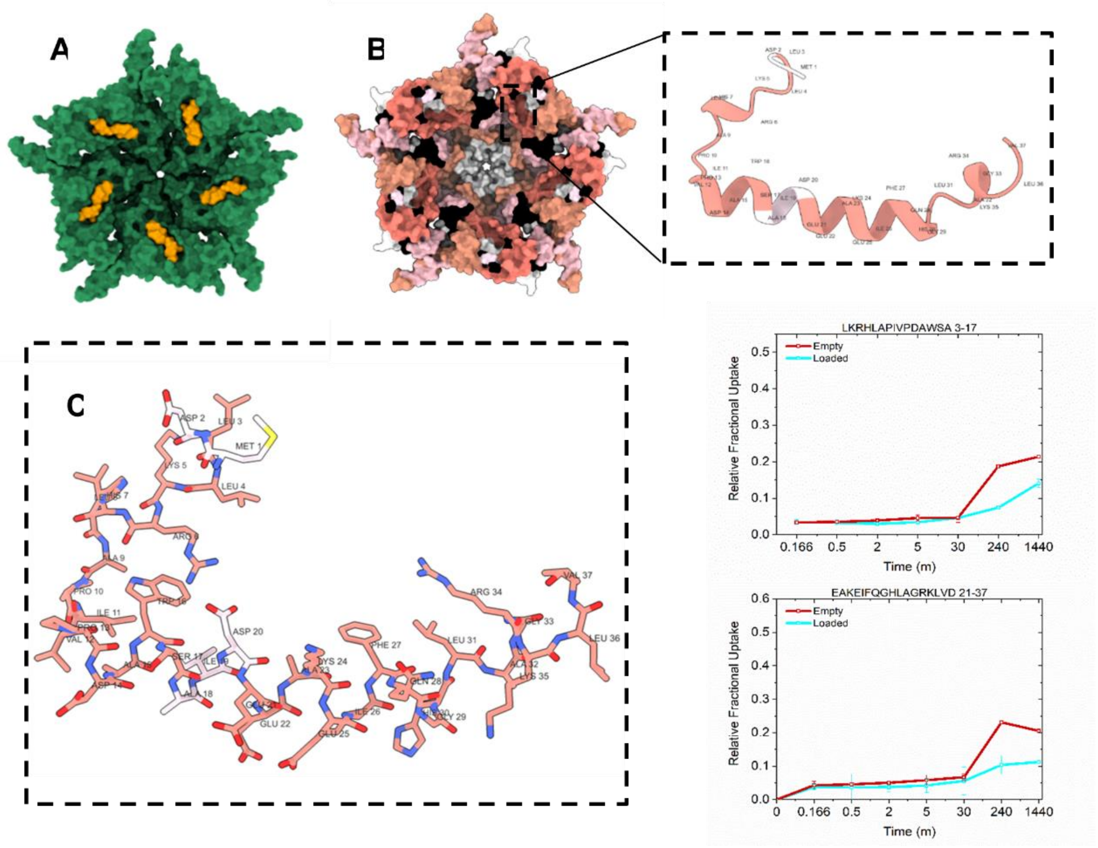
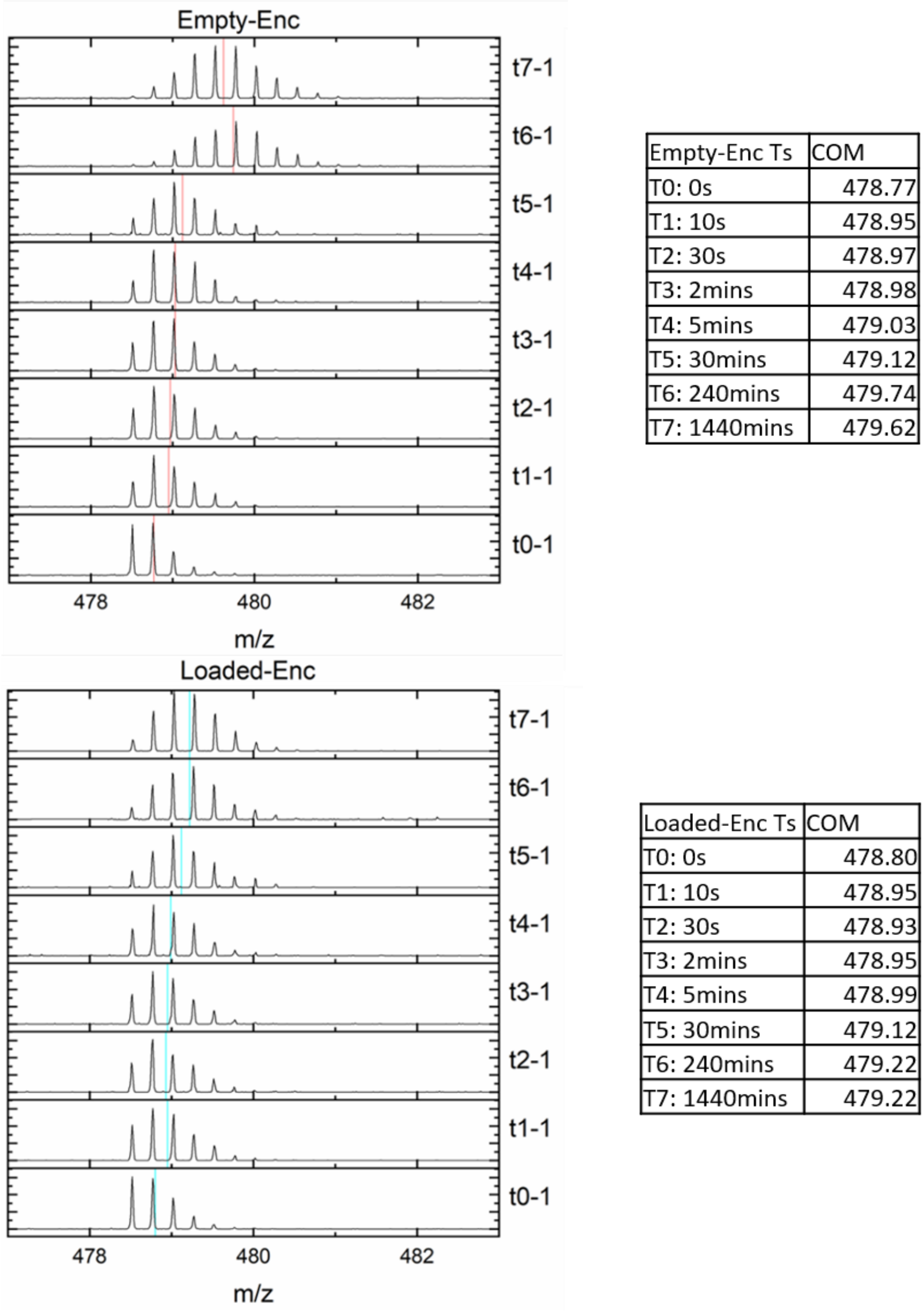


Figure 5.13 Empty vs loaded-Enc differential HDX analysis of the binding pocket (A) Closed conformation of the encapsulin 5-fold pore. (B) 240 minute deuteration profile of Loaded-Enc overlain the closed 5-fold pore with the cartoon representation of the 3-37 residues associated with the binding pocket extrapolated out. (C) Stick representation of the 3-37 residues of the localization binding pocket with the RFU uptake plots of the associated peptides

The significance for both ^{3}LKRHLAPIVPDAWSA^{17} and ^{21}EAKEIFQGHLAGRKLVD^{37} was calculated via an unpaired student T-test which confirmed this region is significantly different at the T6 (240mins) and T7 (1440mins) timepoints (SI Tables 5.4 & 5.5). This demonstrates the presence of EncFtn does impact the solvent accessibility of the Enc binding pocket at longer timepoints.

Looking at the coverage map overlaid the crystal structure in Figure 5.12 it becomes apparent that the exchange profiles are the same at the central vertices of the 5-fold pore for Empty and Loaded-Enc. These findings demonstrate the presence of cargo protein within the interior lumen does not affect the solvent accessibility of this region, thus the conformational flexibility of the region is not being compromised by the presence of cargo protein.

The prior Cryo-EM data of the Loaded-Enc revealed that the interior Lumen appears to a have a maximum loading capacity of 4 cargo proteins which carry a tetrameric arrangement. This loading capacity means that the interior binding pockets will have a population of engaged and unengaged sites. For this reason, it was hypothesised that there would be a difference in the exchange profiles between Empty and Loaded-Enc. It was predicted that the Empty-Enc would display EX2 profile with all of the binding pockets being unoccupied. Conversely, Loaded-Enc may display an EX1 profiles across regions which are involved with the binding pocket due the presence of both occupied and unoccupied binding pockets. However, analysis of the spectral data reveals that only EX2 exchange was visible in both Empty and Loaded-Enc which suggests that the exchange profile is uniform across the each of the binding pockets in both states Figure 5.14.



Empty-Enc Ts	COM
T0: 0s	478.77
T1: 10s	478.95
T2: 30s	478.97
T3: 2mins	478.98
T4: 5mins	479.03
T5: 30mins	479.12
T6: 240mins	479.74
T7: 1440mins	479.62

Loaded-Enc Ts	COM
T0: 0s	478.80
T1: 10s	478.95
T2: 30s	478.93
T3: 2mins	478.95
T4: 5mins	478.99
T5: 30mins	479.12
T6: 240mins	479.22
T7: 1440mins	479.22

Figure 5.14: Empty-Enc and Loaded-Enc timecourse spectra for the ^{21}EAKEIFQGHLAGRKLVD^{37} 2+ peptide. The coloured reference lines highlight the centroid m/z for each isotopic distribution and the values are show in the adjacent table for each Loaded and Empty data respectively

The difference in uptake of this region is still relatively low with only an 11% difference in Empty and Loaded-Enc occurring at the later timepoints. It is possible that the exchange profile may still be different, but it is not significantly different enough to distinguish the two isotopic distribution associated with an EX1 profile (see section 1.5.2).

5.6 Conclusion

Previous work has highlighted the 5-fold pore of Enc shell regions may have a dynamic region due to it loss of resolution during Cryo-EM imaging. To this end, HDX-MS analysis of Loaded-Enc was used to confirm the presence of this dynamic region across the 5-fold, 3-fold and 2-fold poresites.

In contrast, lower rates of HDX are observed in both the 2-fold and 3-fold pores suggesting a lack of an equivalent dynamic nature within their regions as with the 5-fold. The 3-fold pore demonstrated very low level of deuteration both across regions which span both the interior and exterior of the Enc shell. These findings are indicative of a protein with highly buried and protected residues. Furthermore, they revealed the dynamic nature of the 5-fold poresite and are in agreement with the prior Cryo-EM analysis, which had initially highlighted the presence of a potentially dynamic nature of the 5-fold poresite[128].

The differential HDX-MS analysis demonstrated that the presence of the EncFtn cargo protein does significantly alter the exchange profile of the interior Enc peptides with an association with the EncFtn LS. Whilst significant, the levels of difference in deuteration between the Empty and Loaded-Enc’s lumen are still considerably low revealing that the binding pocket is still well protected in both Empty and Loaded-Enc. The low exchange is likely due to the pocket retaining a large proportion of protected residues. Looking at the structure the residues associated with this region run in-between the interior and exterior of the Enc shell. The interwoven structure will convey a high degree of protection due to extensive Hydrogen network and low solvent accessibility.

Although the bottom-up HDX-MS data does provide a great deal of insight regarding structural aspects of the Enc and its cargo protein relationship, it is still limited by the lack of single amino acid resolution across the LS. Using a robust top-down HDX-MS workflow would help to further confirm the interaction mechanics within this region by revealing precisely which amino acids within the binding pocket are interacting with the EncFtn LS.

Chapter 6 Development of a top-down HDX-MS approach for the analysis of EncFtn

As previously demonstrated, MS analysis of amide hydrogen deuterium exchange is important for characterizing protein structure and dynamics. The traditional bottom-up HDX-MS workflow consists of deuterium isotope labelling of a protein followed by quenching under acidic conditions to arrest the D→H reaction. Next the protein undergoes proteolysis and then the subsequent peptides are separated by LC into the MS.

A benefit of the bottom-up approach is the lack of an upper mass limit of the protein so long as the resulting peptides can be efficiently analyzed[152]. However, the spatial resolution of the bottom-up approach is limited by the size of the peptides produced[88]. This limitation can be mitigated somewhat via the analysis using overlapping peptides through manipulation of digestion conditions[88]. Published work has demonstrated that using alternative proteases with varying cleavage preferences can result in many overlapping peptides, although this requires multiple time course repeats which can be a time consuming and impractical solution[153,154]. Furthermore, there has been evidence that the conversion of a backbone amide into the free H_2N-peptide terminus residue cannot retain its deuterium label along with the adjacent amide linkage. This results in the loss of data associated with two residues. Whether this should be accounted for in the calculation of deuterium incorporation is still an area of much debate[80]. One alternative approach to overcoming these limitations would be to utilise gas phase fragmentation in place of proteolysis, which can be found in a top-down MS approach (see section 1.5.5). This philosophy has already been adopted in “middle down” HDX-MS studies where bottom-up HDX incorporates ETD induced fragmentation of the peptide products after proteolysis. The added fragmentation step allows for increased spatial resolution[135,155]. However, in both to the bottom-up and middle-down approaches there is a lack of a gas phase isolation step of the intact mass precursor ions of the protein sample of interest.

6.1 Gas Phase Fragmentation: an alternative to Proteolysis

One well reported alternative to proteolysis is to utilise gas phase fragmentation techniques to generate fragments from the exchanged intact protein i.e., without proteolytic digest. An overview of both top-down and bottom-up workflows can be seen in Figure 6.1. Previously reports have demonstrated successful generation of deuterated fragmentation via CID and EXD techniques[90,156]. A detailed technical overview for each gas phase fragmentation techniques has been discussed in section 1.4.4. Early studies utilizing CID in place of the proteolytic digest step found that although partially successful in producing deuterated fragments data, CID often produces a high degree of H/D scrambling, resulting in deuteration profile being randomized across the fragment's residues[156]. The scrambling effect is caused by the vibration induced generation *of* the *b* and *y* fragments from the precursor ion. In contrast, both ECD and ETD are low scrambling alternative gas fragmentation techniques. These EXD approaches are particularly suited for HDX-MS workflows as there is significant evidence that they are nonergodic, therefore cleavage of the amide backbone occurs in the absence of excitation energy, limiting the extent of H→D scrambling[90,135,157]. Several reports have successfully demonstrated the implementation of both ETD and ECD top-down HDX workflows[90,158]. A further benefit provided by the gentler EXD fragmentation methods allows for the analysis of fragments retaining a PTM which are more often lost under CID fragmentation[159]. Both factors mean that HDX-LCMS-EXD are suitable for the analysis of intact protein without the necessity of a proteolysis step an overview can be seen in Figure 6.1.

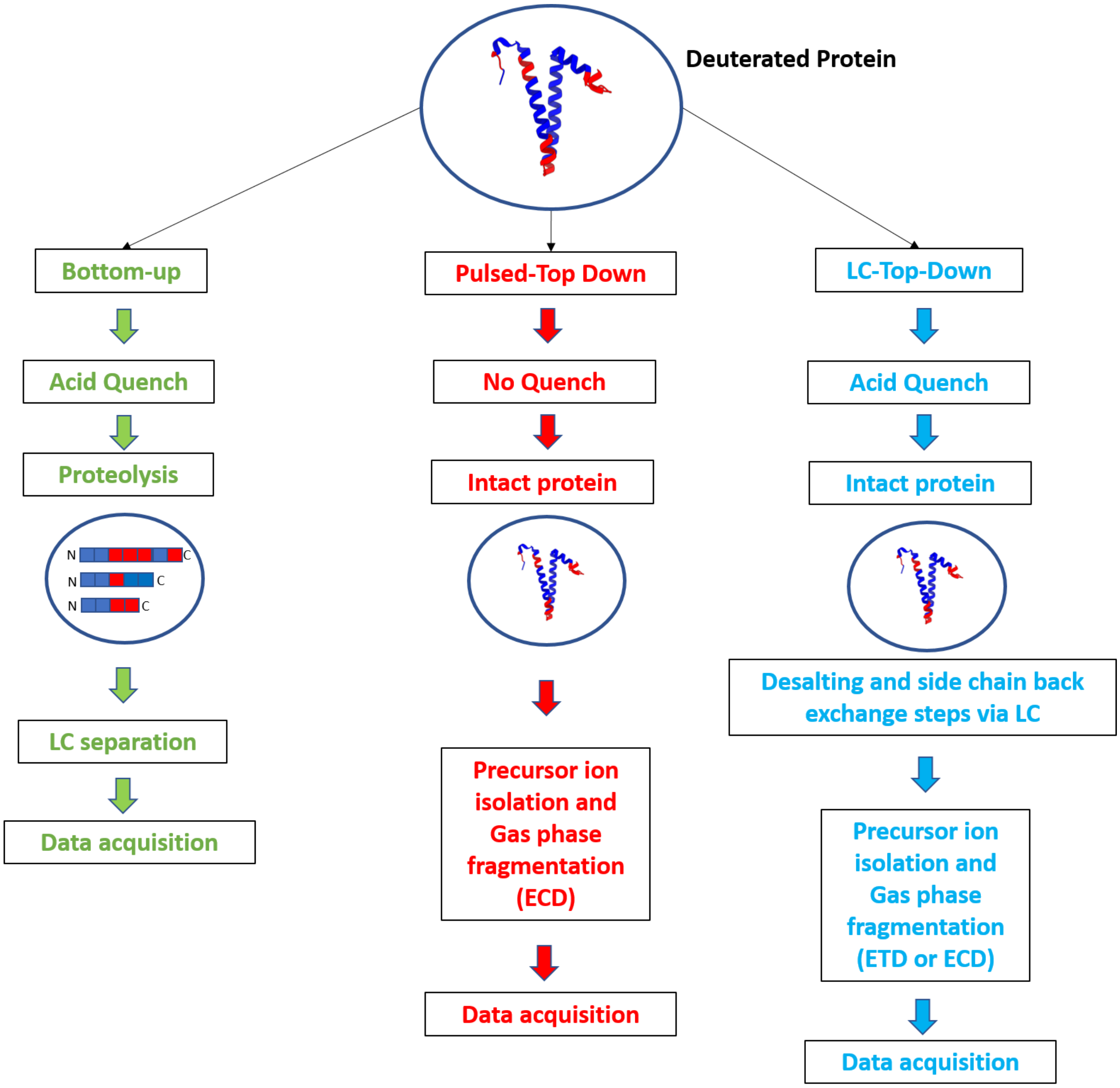


Figure 6.1: Overview of the bottom-up (green), Pulsed top-down HDX (red) and LC Top-down HDX-MS strategies

6.1.1 Isolation of proteoforms

One unique benefit of top down HDX-MS is its ability to isolate the intact protein precursor ion prior to gas phase fragmentation. This step allows for the selection and analysis of different proteoforms present in samples with protein heterogeneity[152,160,161].

Traditional bottom-up HDX-MS will assess the isotopic distributions of deuterated peptides. When confronted with samples which contain heterogeneous protein population, such as a WT form and single residue mutant. Bimodal EX1 distributions which would be indicative of the peptic region being associated with modified and non-modified protein states. However, the bottom-up approach will be unable to discern which distribution comes from which state. This coupled with the fact the EX1 distributions are often difficult to observe may further complicate protein characterization[76].

Top down HDX-MS can overcome this by initially assessing the *m/z* values of the intact protein precursor peaks prior to isolation (proteoform profiling). This allows the user to select the peak associated with the proteoform of interest. Furthermore, the isolation step can be utilised in the study of protein PTM interactions within a mixed population. Using the same principle, the precursor ion of a modified or unmodified protein can be isolated, fragmented and analysed individually. Any subsequent EX1 data found will then be associated with the conformational dynamics of the selected proteoform and not the presence of two competing proteoforms[160]. There are inherent limitations associated with top-down MS strategies. Top-down data is very complex with many fragment ions of overlapping isotope distributions. When exposed to deuterium these distributions will broaden further, resulting lowering the signal of individual isotopologues. Top-down HDX studies have been largely limited to small protein (<200 kDa)[89,158,162]. This in part due to the intractability of top-down MS analysis of proteins larger than ~500 residues[163]. Taken together these limitations present a significant challenge for instruments performance.

6.1.2 The pulsed HDX-MS approach

To date, recent studies utilizing an EXD top-down HDX-MS approaches have focused on rapid "pulsed" HDX instrumentation [137,164]. In these workflows labelling and quenching steps are performed at rapid speeds (0.1s) and in the absence of proteolysis or the LC separation step of the subsequent peptides. The lack of the LC step results in the both the amide backbone hydrogens and those associated with the sidechains remaining deuterated. These factors allow the "pulsed" top-down HDX workflow to characterises and observe the deuteration of side-chain hydrogens of individual residues. Furthermore, Pulsed HDX allows prolonged direct infusion of deuterated sample at a given labelling timepoint for extended data acquisition times. This allows for real time analysis of data and alteration of experimental conditions allowing for rapid optimization of the MS parameters.

Pulsed HDX was initially developed by the Englander group in 1988 to perform kinetic refolding experiments with Cytochrome c using HDX-NMR[29]. Their approach began with the protein sample being initially denatured followed by three rapid mixing events to explore its subsequent refolding dynamics. The first mixer introduced a 6-fold dilution with 0.1 M acetate at pH 6.2 in H_2O to provide stable refolding conditions. This was followed by a waiting step wherein the protein was allowed to time to undergo refolding[29]. The sample would then enter a second mixer which would introduce D_2O labelling solution and was left to deuterate for a user specified time. Finally, the third mixer would introduce quenching buffer to the sample to arrest the forward $H \rightarrow D$ exchange prior to immediate injection into the NMR instrument. Their findings were highly successful and the novel pulse HDX methodology offered unique benefits in terms of refolding kinetics, such as allowing for NH sites to be monitored directly at constant pH. The pulsed HDX approach was adapted for MS utility in 2005 by the Konermann group who implemented this approach to study ubiquitin unfolding. This initial setup lacked a quenching step, as arresting of the forward exchange was reliant on sample desolvation by ESI which ended the labelling step[165]. However, in 2012 Konermann further adapted this by adding a third mixer which would introduce quenching solution just prior to sample ionization. An overview of the 3 mixer pulsed HDX system can be seen in Figure: 6.2[166].

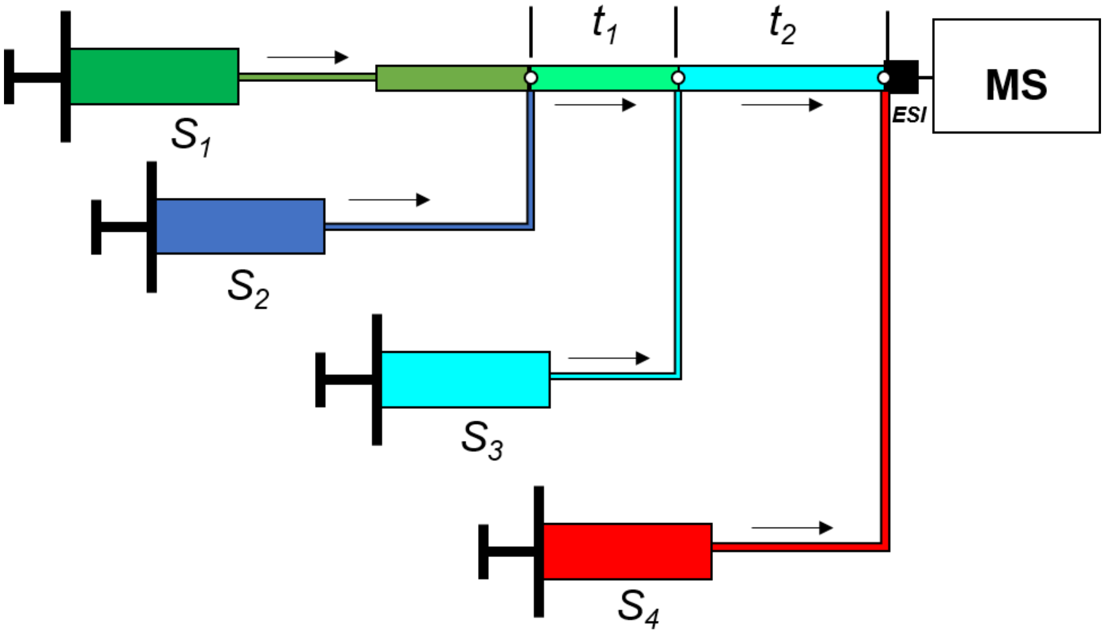


Figure 6.2: A schematic of the rapid mixing pulsed online HDX-MS setup. The black arrows indicate the direction of flow, and white spheres represent mixers. Each syringe is controlled by its respective syringe pump. S_1 contains the protein sample under denaturing conditions and is mixed with S_2 which contains refolding buffer. The sample then enters variable time span (t_1) to allow for refolding. S_3 containing D_2O is introduced via the second mixer beginning the variable deuteration time (t_2) just prior to entering the ionization source. S_4 contains a quenching solution to arrest the forward H→D exchange just prior to the sample entering the ESI[166]

Since Konermann's initial experiment Pulsed top-down HDX has been used in many studies ranging for exploration of folding dynamics[167], aggregation studies[137,168] and elucidating assembly pathways[169].

The pulsed HDX approach's lack of an LC step conveys an advantage in that it minimizes the impact of D→H back-exchange. This is because fragments are not exposed to the LC solvents containing exchangeable hydrogens. Furthermore, recent works has demonstrated that exposure to the hydrophobic C18 media can also accelerate the back-exchange step[170] and also allows for the prolonged data acquisition times previously mentioned.

Finally, as peptides are separated on an LC time scale, peptides which take longer to elute may have a slightly higher degree of back-exchange than smaller fast eluting peptides. However, although there is a significant benefit to removing the LC step there are several limitations associated with its absence. These limitations include the removal of a desalting step which can negatively impact sample ionization[171]. Furthermore, the deuteration of the side chains will need to be considered during fragment assignment. As the side chains are considered rapid exchangers, they will be deuterated proportional to the % D_2O used during the labelling step. This results in the addition of a further side chain data reduction step during data analysis prior to calculating the amide backbone exchange profiles. Within this study, a top-down HDX-MS workflow was designed to incorporate the LC step both to eliminate deuterated side chains but also to allow for a direct comparison with a traditional bottom-up analysis. An overview of the LC top-down HDX-MS workflow can be seen in Figure 6.3.

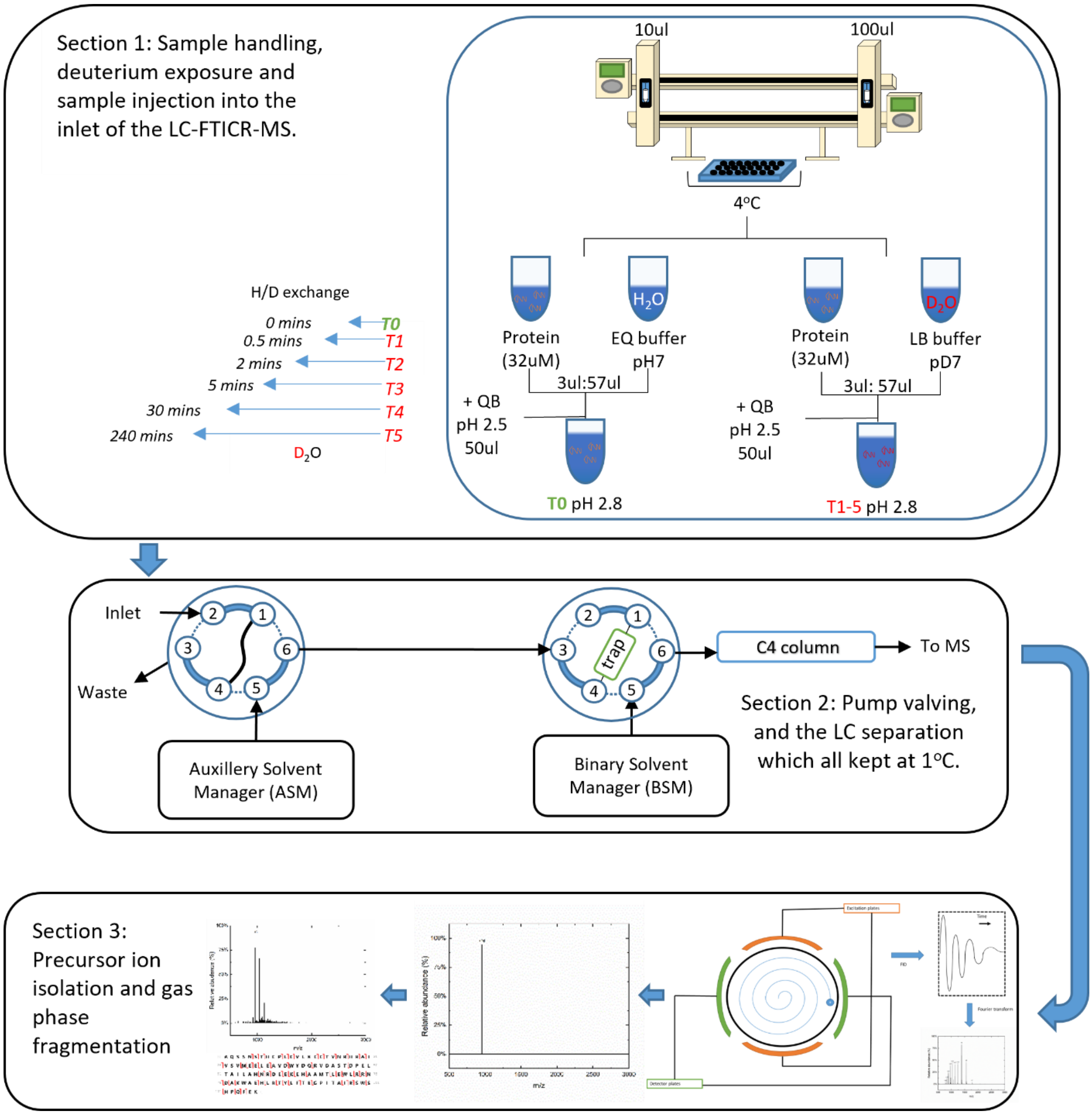


Figure 6.3: Technical overview of the automated (LC)Top-Down HDX-MS workflow. Section 1: An overview of the generation of timepoints including liquid handling and sample preparation. Section 2: an overview of valving and pumps within the waters HDX manager. The filled lines on the valving represent the "Inject" valve position and the dashed lines represent "Load" valve position Section 3: Precursor ion isolation and gas phase fragmentation by ECD and subsequent fragment assignment

Top-down HDX is reliant on the observation of both the undeuterated and deuterated isotopic distributions of a target fragment. To achieve sequential amino acid residue spatial resolution of deuteration event, ideally every residue should undergo fragmentation excluding N-terminal proline fragments. The large number of data points will in turn result in fragments saturating the spectral data. This problem is minimized in Bottom-up workflow as peptides will be separated by LC and analysed in relation their retention time. Unfortunately, as fragmentation of the intact protein occurs in the gas phase, LC separation of the fragments is not possible. The resulting fragment saturation will lead to significant peak overlap and there will be loss of fragment ion signal. As the fragment incorporates deuterium its isotopic distribution will broaden lowering the signal of the individual isotopologues with many being lost in the noise. This results in a significant loss in fragment coverage as the necessary S/N ratio must be collected for accurate peak assignment. To effectively combat this limitation a significant amount of time must be taken to optimize the ECD condition, where a balance between signal intensity, fragmentation coverage and scrambling must be found.

6.2 Aim

This chapter discussed the development of a novel strategy for Top-Down LC-HDX-MS to mitigate the challenge of isotopic heterogeneity in the characterization of an encapsulated ferritin (Rru-EncFtn). An isotopically depleted protein variant (ID-Rru-EncFtn) was implemented as a strategy to increase the accuracy of peak assignment.

By using isotopically depleted protein, this study demonstrates that peak broadening can be minimized resulting in more deuterated fragments being assignable over longer labelling timepoints. Furthermore, isotopically depleted protein has the added benefit of decreasing the prevalence of overlapping fragments further aiding accurate peak assignments. This represents a significant step forward in top-down FT-ICR MS HDX with data that is directly comparable with traditional Bottom-up data without the necessity of a proteolysis step and inclusion of intact protein precursor ion isolation utility.

6.3 Top-down HDX-MS of Rru-EncFtn using FT-ICR MS

Incorporation of FT-ICR MS and ECD into a HDX workflow is reliant on a significant degree of system optimization. This is primarily due to the delicate balance of three essential criteria. First, the HDX labelling and quenching conditions need to be optimized enough to provide good signal quality for the detection of fragments. Second, the fragmentation conditions need to be suitable enough to allow for enough *c* and *z* ions to provide maximal sequential protein coverage. Finally, the fragmentation and source settings cannot induce a significant amount of deuterium isotope scrambling decreasing overall data quality.

6.3.1 Improving top-down HDX-MS data quality using Isotopic depleted proteins

Both signal overlap of assigned fragments and loss of signal due to deuteration induced peak broadening significantly impact data quality. An alternative strategy would be to improve a fragment's signal by altering the protein's isotopic distribution prior to the MS workflow. This can be achieved by utilizing isotopic depleted forms of a target protein. First described in 1997 by Marshall *et al* they successfully observed a change in the isotopologue distribution of a 12 kDa protein after being grown in minimal medium in which the only source of carbon is 13C-depleted glucose (99.9%12C) and the only source of nitrogen is 15N-depleted ammonium sulfate (99.95%14N)[172].

Subsequent work has demonstrated the usefulness of isotopic depletion (ID) in the context of improve CID fragmentation, namely by increasing the number of assignable fragments through improved S/N ratios within the spectra[173]. More recently the advantages of using an isotope depletion strategy for greatly improved top-down fragmentation sequence coverage using ECD have been demonstrated by Gallagher et al[174].

Coupling ID to an HDX-MS workflow was first demonstrated in 2010 by Bou-Assaf *et al* [131]. This report demonstrated that ID reduces the mass spectral complexity by improving the S/N ratio of proteolysis products after deuteration. These initial findings form the basis of how ID could improve the signal quality of top-down HDX-MS associated fragmentation products, namely through ECD, in a similar fashion. One key aspect the 2010 study did not explore was the impact of many deuterated peptides co-eluting. This problem is overcome in a bottom-up workflow due to separation of peptides via LC after digestion. In a top-down experiment fragmentation occurs within the gas phase of MS, meaning the collected spectral data will be dense with fragments each containing varying rates of deuteration. This can lead to a high number of isotopic distributions overlapping. ID could be a potential solution to minimizing the impact of this and maybe necessary for the validity of future top-down HDX-MS studies.

6.3.2 Online isolation of Rru-EncFtn precursor ion

In the absence of an LC step, side chains will remain deuterated which can increase the spectral complexity. For this reason, an LC step was introduced into this study’s top-down workflow.

To allow the collected top-down HDX-MS data to be comparable with a bottom-up approach, the same LC conditions and settings were used along with the same solvents and reagents for sample preparation see sections 2.1 and 2.2.2. For the top-down workflow, a Waters Acuity C4 trap and analysis column was used to compensate for retention of protein. Data from the TIC trace was processed to assess data quality and to identify a suitable precursor ion to be isolated for subsequent fragmentation (Figure 6.4).

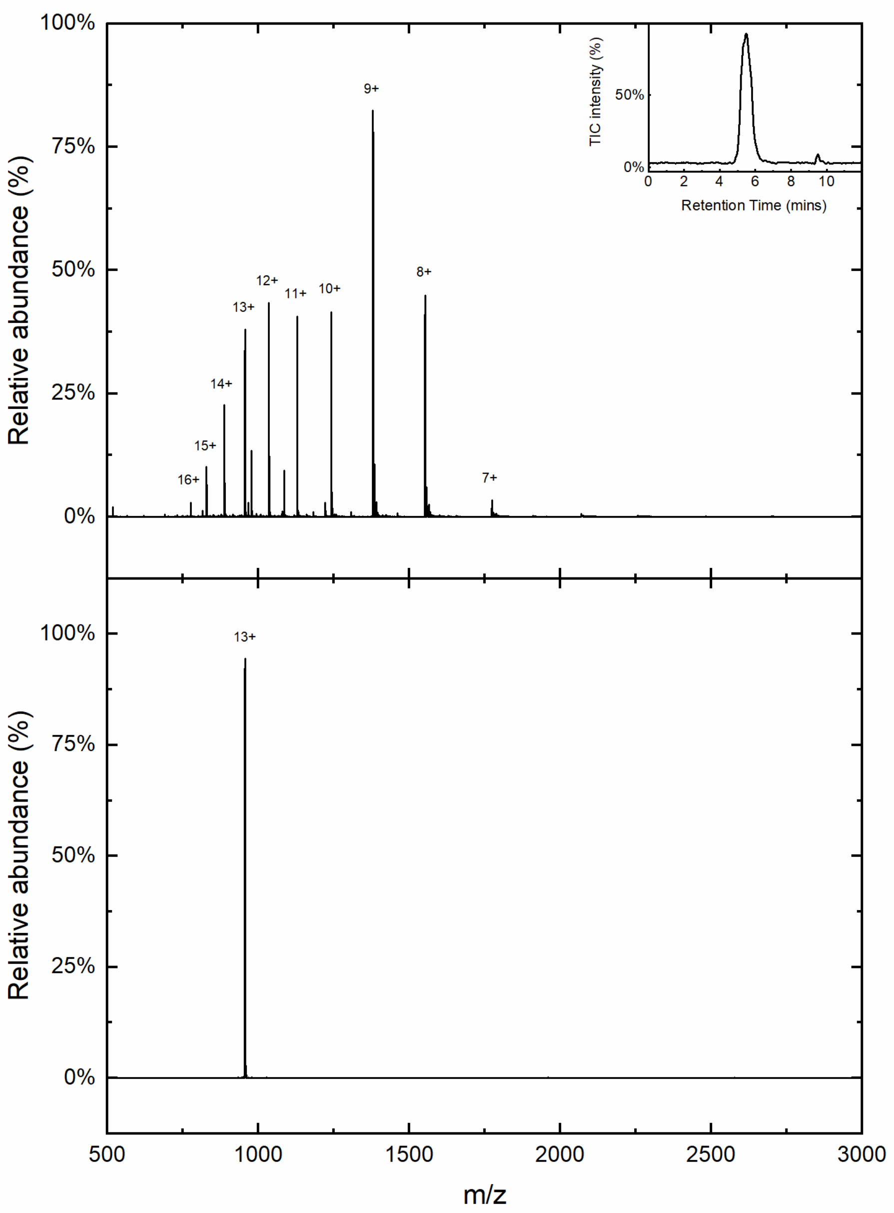


Figure 6.4: Native mass spectrum of the Rru-EncFtn, (top panel) without isolation (bottom panel) Isolated Rru-EncFtn 13+ ion isolation window set to 5 m/z. 2 M-words Ion accumulation time 0.600s. (corner panel) Total ion count chromatogram for intact Rru-EncFtn. Separation was performed on a Waters Acquity C4 column using a 12-minute step gradient from 5%-95% water – acetonitrile. The elution time of Rru-EncFtn was found to be at 5.2-6.16 minutes

Analysis of Rru-EncFtns was successful and a clear series of charge states were observed, as shown in Figure 6.4. From this charge state distribution, the 13+ charge state (955 *m/z*) was selected for quadrupole isolation and ECD fragmentation. This decision was taken to demonstrate that the top-down workflow was not reliant on using the highest available charge state as seen in the bottom panel of Figure 6.4.

6.3.3 ECD Optimization for Rru-EncFtn Fragmentation

Prior to performing HDX analysis it was necessary to optimize the ECD conditions using undeuterated controls. The ideal ECD parameters would need to provide enough assignable fragments to allow for sequential residue analysis throughout the protein’s sequence. One of the challenges facing this optimization is maintaining a balance between data quality, the number fragments and introducing new sources of error, e.g., in form of lack of fragment coverage or isotope scrambling. The primary parameters which can induce scrambling are electron pulse length, electron beam bias and source voltages i.e., skimmer 1. From these recommendations the ECD conditions were as follows: electron pulse length of 0.010s and an electron beam bias of 1.3. These parameters were selected as a compromise between the “harsh” and “soft” conditions previously reported (Figure 6.5).

Fragmentation of the precursor ion did produce a selection of fragments, although not enough to provide substantial coverage (Figure 6.6A). To increase the fragmentation efficiency, the source voltage (skimmer 1) was increased stepwise (Figures 6.5 & 6.6).

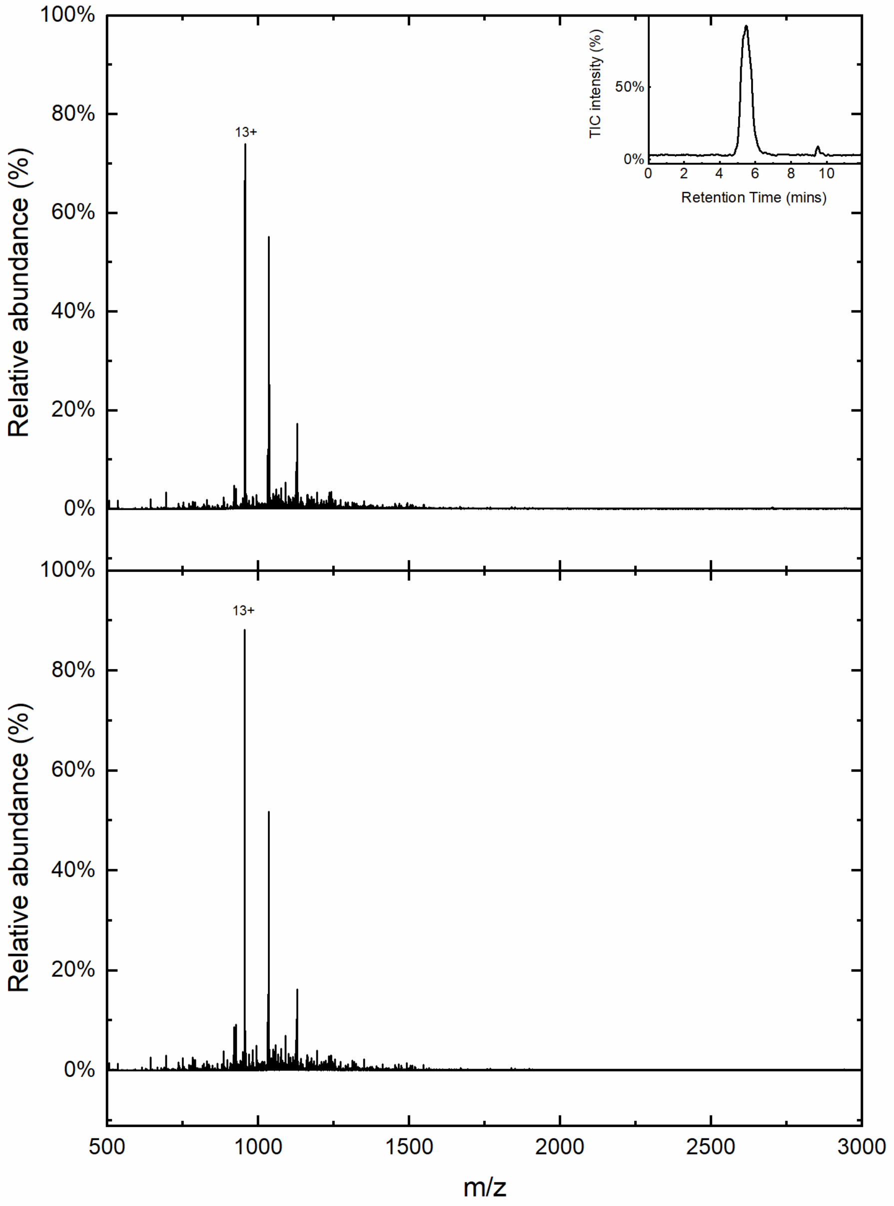
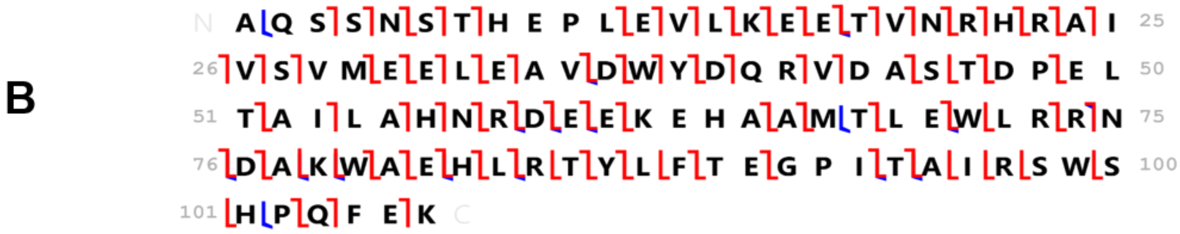


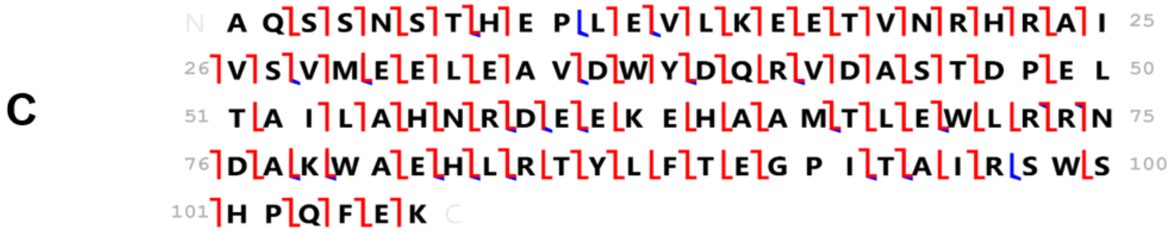
Figure 6.5: Isolated denatured Rru-EncFtn 13+ at 955 under ECD fragmentation conditions of both Rru-EncFtn (top panel) and ID-Rru-EncFtn (bottom panel)

Figure 6.5 shows the impact of ECD activation upon the 13+ precursor ion in both protein states. *c and z* fragment ions immediately appear as indicated by the new fragment peaks surrounding the initial precursor peak. These new fragment peaks then underwent assignment and identification (Figure 6.6).

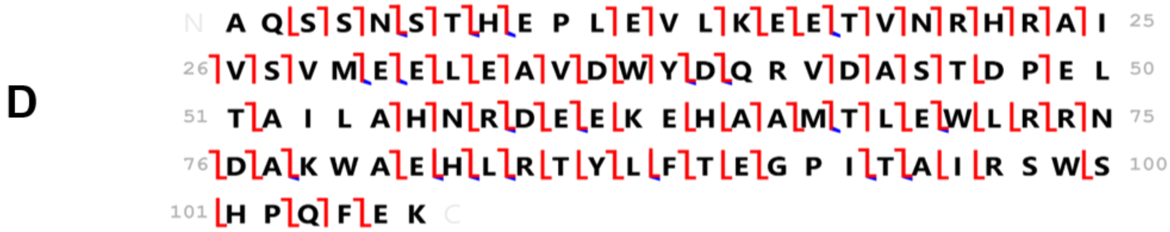
Skimmer 1 At 20V



Skimmer 1 At 25V



Skimmer 1 At 30V



Skimmer 1 At 40V

Figure 6.6: ECD Fragmentation maps of Rru-EncFtn produced at different skimmer voltages ranging from (A) 20V, (B) 25V, (C) 30V, (D) 40 volts. c and z ions are highlighted in red and b and y in blue

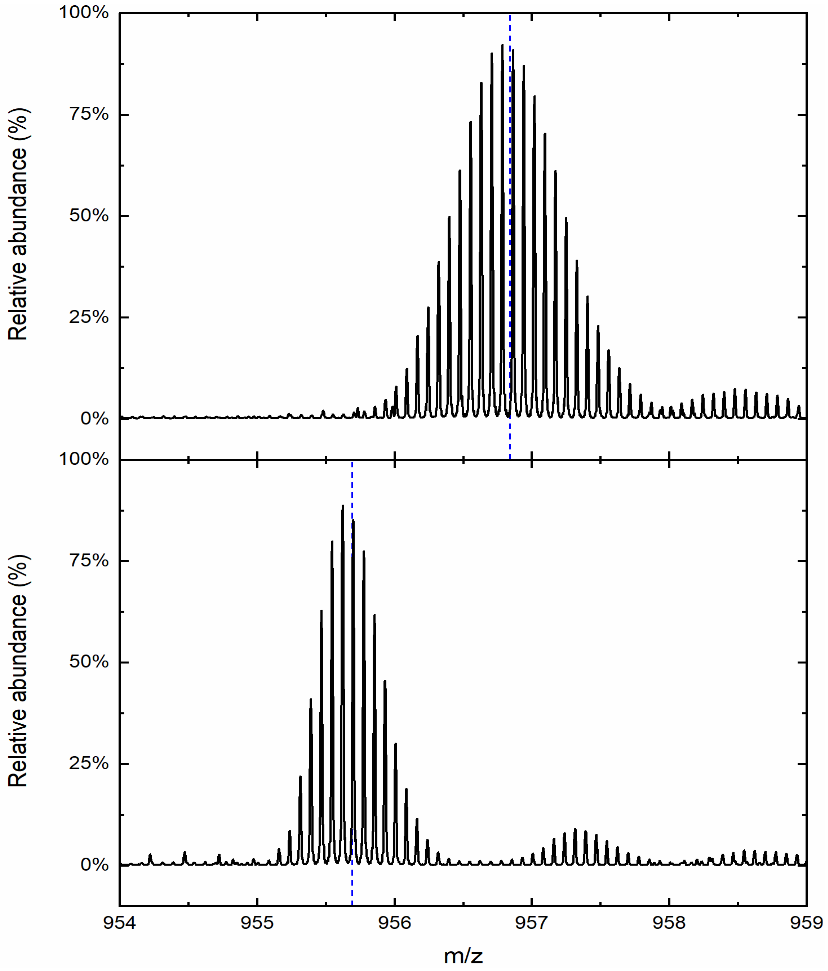
To maintain the balance between scrambling and assignable fragments the number of *b* and *y* ions present in the fragmentation map was used as a benchmark. The presence of *b* and *y* ions would be indicative a high degree of vibrational energy during fragmentation which can induced scambling. The number of each type of fragment for each condition can be seen in Table 6.1.

Condition	***c* ions**	***z* ions**	***b* ions**	***y* ions**
Skimmer 1 20V	17	7	1	0
Skimmer 1 25V	69	56	1	16
Skimmer 1 30V	64	48	0	19
Skimmer 1 40V	75	56	2	24

Table 6.1: Total number of each type of fragment present in the spectral data at each skimmer 1 voltage. In order minimize back-exchange the number of b and y fragments were taken into consideration due to their presence suggesting the conditions maybe too harsh and may induced isotope scrambling. It was decided that the Skimmer 1 voltage be set to 25V. This value was selected as it provided a significant number of fragments to allow for sequential fragment analysis, whilst having fewer b and y ions than the high voltages. The 20V setting was disregarded due to the low number of c and z ions.

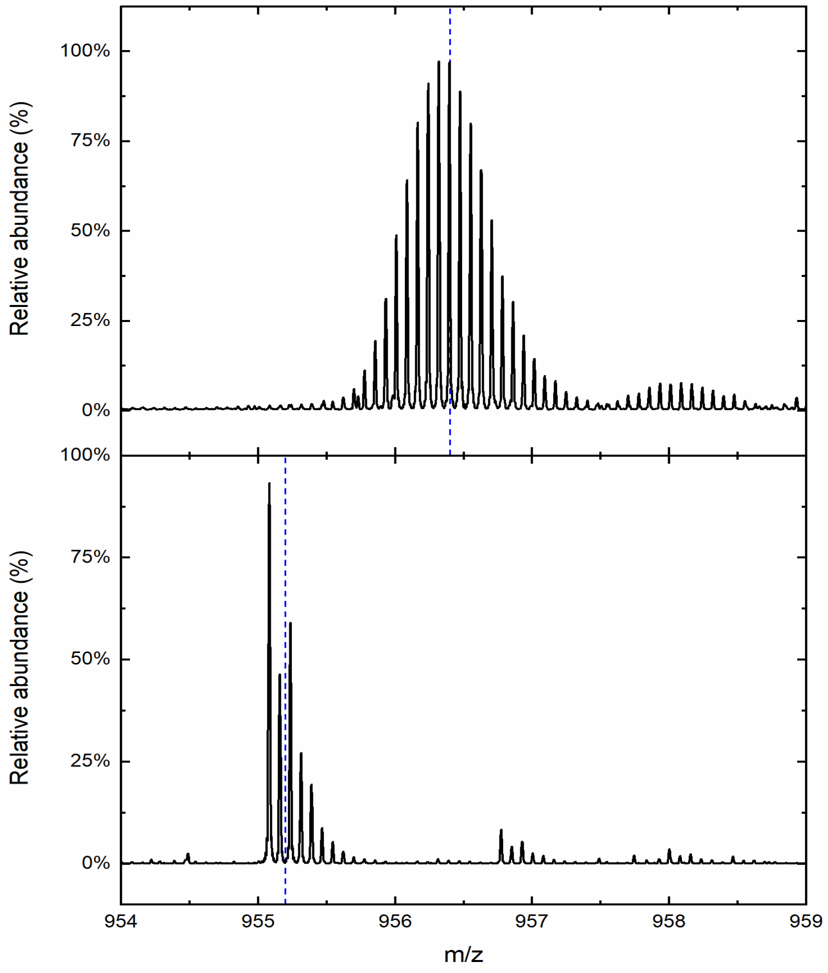
6.3.4 Deuteration and fragment assignment

With the ECD conditions and a suitable degree of fragments assigned, the next step was to make sure there were detectable fragments after deuteration. For the sake of the limiting the impact data processing would have on the workflow, a timecourse comprised 3 timepoints was employed. The timecourse consisted of an unlabelled control (T0) a short timepoint (30s) and a long timepoint (5 mins). The data collected from the deuterating timepoints would then be used to scrutinise the initial list of assignable fragments and eliminate fragments that cannot be accurately assigned. If fragments were no longer accurately detectable by 5 minutes, they were removed from the coverage list. Both isotopically natural and isotopically depleted (ID) forms of the EncFtn were used for this analysis – allowing us to assess the advantages of ID in top-down HDX experiments. Prior to fragment assignment the amount of total deuteration was calculated from the precursor peaks with the 5 minutes timepoint acting as the Tmax. The spectra both the Rru-EncFtn and ID-Rru-EncFtn and are shown in Figures 6.7 & 6.8 respectively.



Timecourse	Centroid *m/z*	Difference (*m/z*)	Mass difference (Da)	%Relative deuteration
5 mins labeling	957.02	1.33	17.29±2.22	17.12±2.19
T0	955.69	0.00		

Figure 6.7: The mass spectra of Rru-EncFtn 13+ charge state precursor ions at 5 minutes of labelling (top panel) and T0 (bottom panel) timepoints respectively. The blue coloured reference lines highlight the centroid m/z for each isotopic distribution and the values are show in the adjacent table for each T0 and 5 minutes of labelled timepoints respectively. The 5 minute labelling time point consists of 27 assignable isotopologues



Timecourse	Centroid *m/z*	Difference (*m/z*)	Mass difference (Da)	%Relative deuteration
5 mins labeling	956.40	1.19	15.50	15.35
T0	955.69	0.00		

Figure 6.8: The mass spectra of ID-Rru-EncFtn 13+ charge state precursor ion at 5 minutes of labelling (top panel) and T0 (bottom panel) timepoints respectively. The blue coloured reference lines highlight the centroid m/z for each isotopic distribution and the values are show in the adjacent table for each T0 and T3 (5 mins) timepoints respectively. The 5 minute labelling time point consists of 21 assignable isotopologues

Interestingly, a slight increase in deuteration is observed in the depleted form of the protein (1.77 % accounting for 1.78 exchange events). The 13+ charge state of the two forms of EncFtn were then subject to ECD after 30 sec and 5 minutes of HDX. ID-Rru-EncFtn. The hypothesis was that removing the number of isotopologues present in fragments will limit the impact of peak broadening attributed to the addition of deuterium during the labelling steps. Peak assignment of the isotopically depleted sample was performed by the AutoVectis peak assignment software Ver.8.0. developed by the Kilgour group. Autovectis allows for the assignment of regular and ID peaks found within raw FT-ICR MS data (Figure 6.9)[129,130]. For the overview of parameters used in this dataset see materials and methods section 2.3.

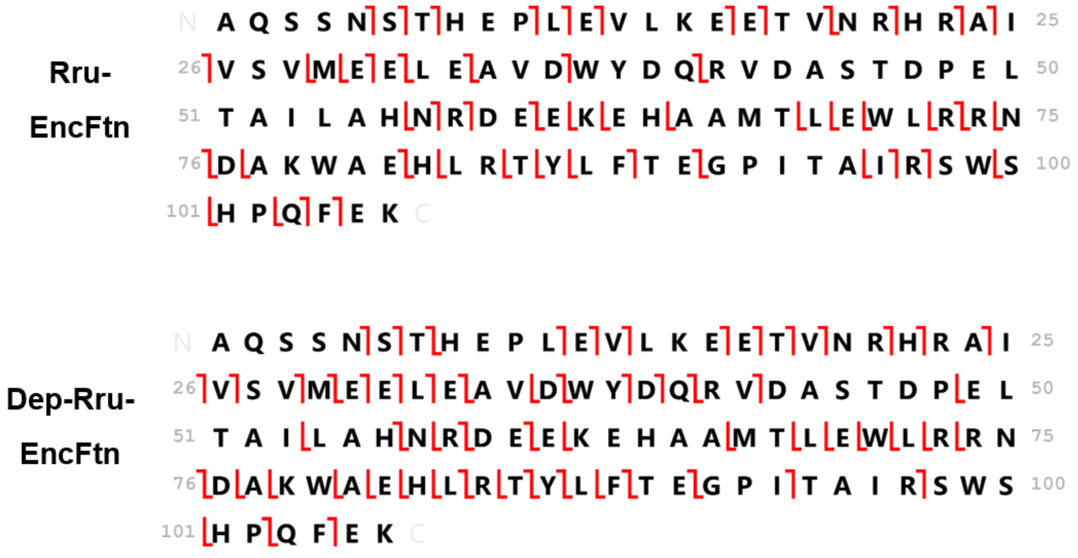


Figure 6.9: Online LC-MS/MS analysis of Depleted and Non-Depleted Rru-EncFtn fragments(Top panel) c and z Fragment ion assignment of Rru-EncFtn derived from the fragmentation of the 13+ charge state. (Bottom panel) c and z Fragment ion assignment of ID-Rru-EncFtn derived from the fragmentation of the 13+ charge state

When comparing the ECD coverage obtained from both EncFtn and ID-EncFtn, there is a marked increase in fragments assigned within the ID-Rru-EncFtn over Rru-EncFtn. There were 36 *c* and 40 *z* ions present in the ID-Rru-EncFtn samples. In contrast there were 31 *c* and 30 *z* ions present in the Rru-EncFtn. The increase of assignable fragments in the ID-Rru-EncFtn data is consistent with the findings reported by Gallagher et al[174]. Many of the ID-Rru-EncFtn fragments had a higher S/N ratio due to the narrowing of the isotopic distribution than their non-depleted counter parts, resulting in a 24.5% increase in assignable fragments in the depleted sample. Even considering the slight increase in deuteration in the Rru-EncFtn timecourse, this is a significant increase in assignable fragments within the ID protein state. There was a significant decrease in distribution overlap within the ID-Rru-EncFtn dataset (Figure 6.10). Taken, together the improved data quality and increase in assignable fragments makes a strong case for using isotopic depleted samples in future Top-down HDX-MS studies.

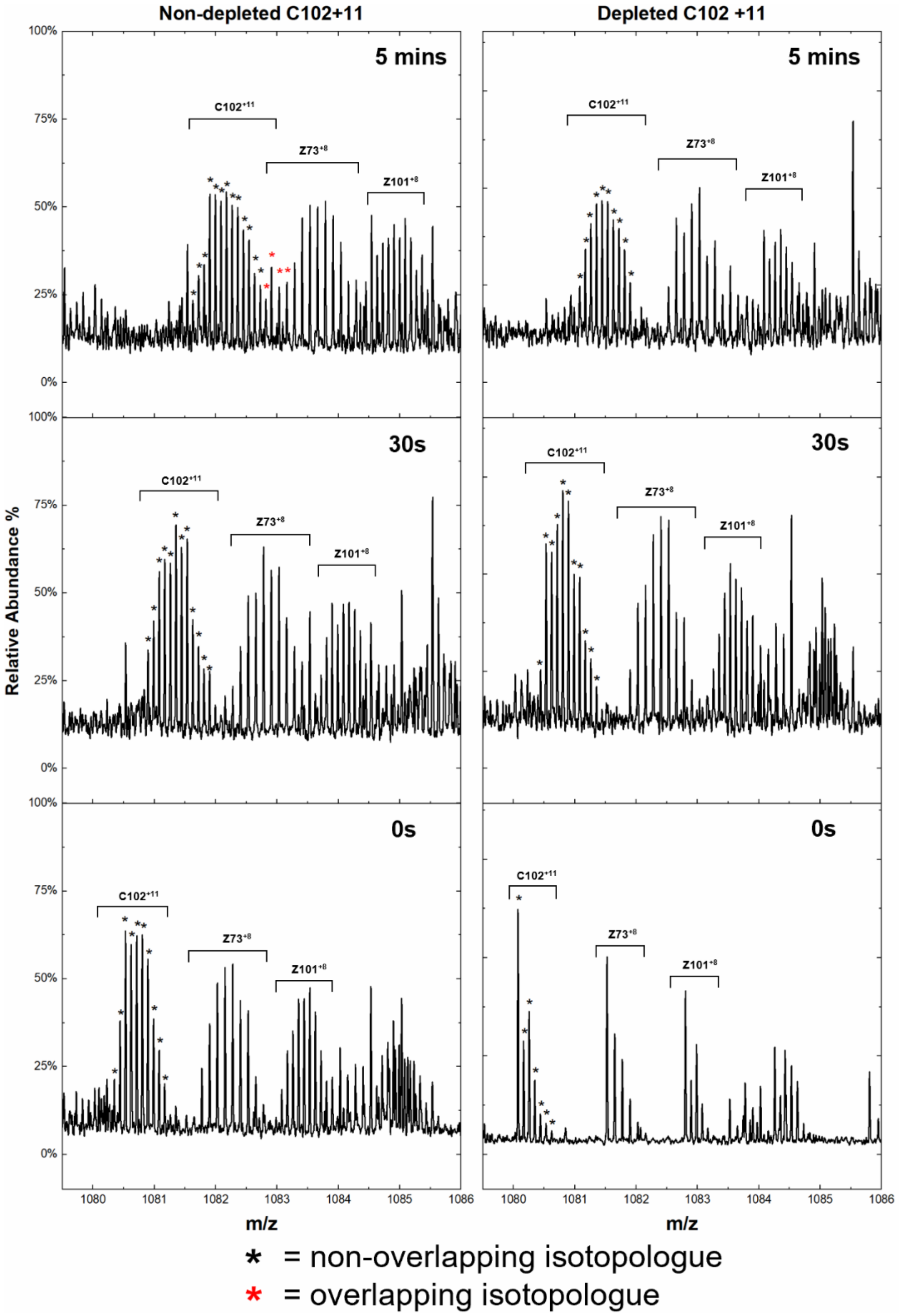


Figure 6.10: Non-depleted and ID-Rru-EncFtn spectra of the c102+11 fragment at 0, 30s and 5 minutes of labelling. Peaks used in the calculation of the centroid of the isotope distribution have been highlighted with and colour coded in relation to whether they are intersecting with an adjacent fragment distribution. Black for non-intersecting and red for intersecting

The % relative deuteration all assignable fragment across both the 30s- and 5-minutes timepoints of the ID-Rru-EncFtn can be seen in SI Tables 6.1 & 6.2.

Unfortunately, there is not an established DIA software capable of analysing top-down LC-HDX-MS data. However, it was possible to manually calculate the amide deuteration status (*D)* of individual residues (SI Tables 6.3-6.5). The *D* value was used to determine deuteration events at a single amide resolution, which will lie in the range between 0 (fully protected) and 1 (fully deuterated). *D* values were calculated in two discrete ways from both sequential fragments (Equation 6.1) and nonsequential (Equation 6.2). These equations were adapted from previously published equations used to calculate *D* values from pulsed top-down HDX-MS data[90].

$$D = N_{ex}\left(c_q\right) - N_{ex}(c_r)$$

Equation 6.1:

For sequential fragments individual amide residue *D* values were determined using the deuterium uptake (N_{ex}) of a target fragment which contained the addition of a residue $c(_q)$ subtracted by the deuterium uptake of the previous recorded fragment $c(_r)$.

For regions with no fragment coverage, it was necessary to factor in the residues with no fragment coverage which spanned across two fragments and use an averaged deuterium status (D_{avg}) to account for these regions (equation 6.2).

$$D_{avg} = \frac{N_{ex}\left(c_q\right) - N_{ex}(c_r)}{Deuterable\ sites\ \left(c_q\right) - Deuterable\ sites\ (c_r)}$$

Equation 6.2:

To further demonstrate the implementation of these equations, a worked through example can be seen below.

c fragments	D_2O exchange (Da)	% relative uptake	Deuteratable sites	D-value *c* ions
28.00	4.97	19.11	26.00	0.06
27.00	*Missing fragment*	*Missing fragment*	25.00	0.06
26.00	4.86	20.23	24.00	0.01
25.00	4.85	21.08	23.00	0.21
24.00	4.64	21.07	22.00	0.08

🟧 Average deuteration status across residues
🟩 sequential Amide deuteration status

Table 6.2: Example data to demonstrate both sequential D and D_{avg} equations

From Table 6.2, *c* fragment 26 can be used as an example of sequential equation 6.1 and is shown below (Equation 6.3).

$$4.86 - 4.85 = \ 0.01\ D$$

Equation 6.3:

The D_{avg} equation (Equation 6.2) can demonstrated through the calculation of the *D*-value of *c* fragment 27. This fragment was not assignable after 5 minutes of labelling. However, it can be accounted for in the analysis through averaging across the region to the next assignable fragment. In this example the next assignable fragment is *c* fragment 28 so the *D* value will be calculated using equation below (Equation 6.4)

$$\frac{(4.97 - 4.86)}{(26 - 24)} = \ 0.06\ D_{avg}$$

Equation 6.4:

The newly obtained D_{avg} value is then applied across the region of no fragment coverage.

The *D* values of the amides through the sequence of ID-Rru-EncFtn can be seen in Figure 6.11 and SI Tables 6.3 & 6.4.

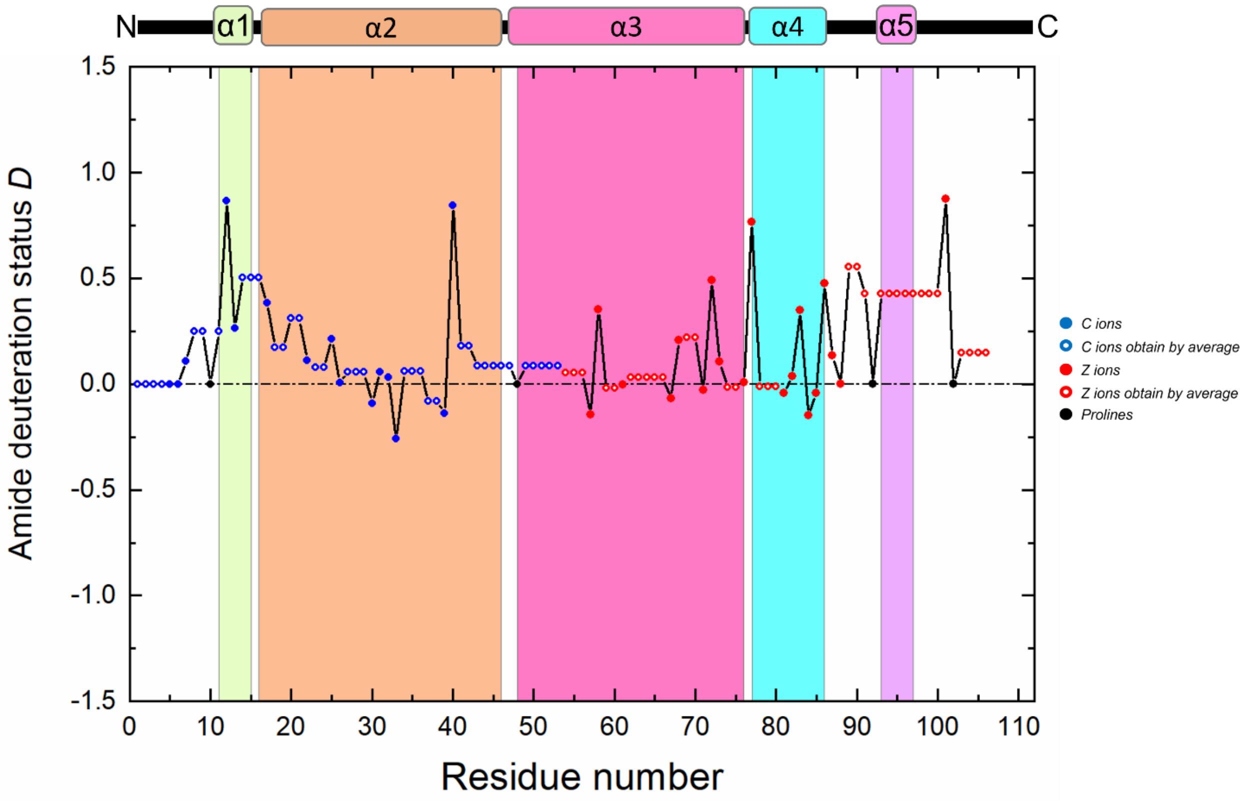


Figure 6.11: Deuteration status of backbone amide sites of ID-Rru-EncFtn c (Blue) and z (red) ions respectively, at 5 minutes of deuterium labelling. The secondary structure of Rru-EncFtn PDB ID: 5DA5 with its 5 helices (α1-green, α2-orange, α-3 Magenta, α-4 cyan and α-5 purple) are displayed along the top of the figure and align with the residue numbers on the x axis. A D value of 0 represents a complete protection and a value of 1 represents total deuteration so can be assumed to be unprotected

As previously stated, single amino acid spatial resolution can only be reliably obtained from sequential fragments. The data contained 40 sequential residues. These residues rested between aa 10-32 and 70-90 the remaining region’s residues could only be calculated through averaged *D* status and thus cannot observed at single amino acid spatial resolution.

There are a significant number of protected residues which span throughout the central sequence of the ID-Rru-EncFtn protein. However, key residues do demonstrate near 1 *D* values. These individual residues (E12, Q40*,* A77 and

H101) lie throughout the sequence both at the N and C terminals but also the central region.

Furthermore, the data quality provided by the depleted samples has allowed for the sequential analysis coverage regions which contained unprotected fragments.

Although these finding are promising it was necessary to see if the data was then compared to a bottom-up analysis of the same protein to see if the datasets aligned.

6.3.5 The impact of ECD induced deuterium isotope scrambling

One of the largest sources of error associated with Top-down HDX-MS is induced deuterium scrambling. As previously outlined in section 1.5.6 deuterium scrambling can severely impact the data quality a collected spectral data of a deuterated proteoform to the extent of inhibiting single amino acid spatial resolution being one of critical benefit of a top-down approach. One way to assess the impact of scrambling is to perform a fragmentation with vibrational excitation on a deuterated form of a target protein and then compare the levels of deuteration with the non-vibrational ECD fragmentation approach. In the context of this study, it was decided that the CID fragmentation of IID-Rru-EncFtn at 5 minutes of labelling would be performed and would provide comparative deuterium profile data with the ECD soft condition. Figure 6.12 shows the number of assignment *y* fragments after 5 minutes of labelling.

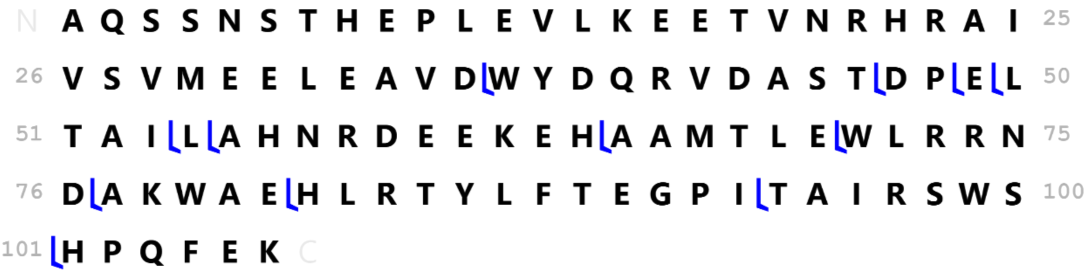


Figure 6.12: Assignable y fragments after 5 minutes of deuteration used in the scrambling assessment

The resulting scrambling assessment resulted in 13 assignment *y* ions after 5 minutes of labelling.

The hypothesis being that the harsh vibration energy introduced would provide completely scrambled *b* and *y* ions. These ions would then carry the same % deuteration as the precursor ion, namely 15%. An overview of the collected “harsh” CID conditions *y* ion product and “soft” ECD conditions *c* and *z* ions can be seen in Figure 6.13 and SI Table 6.5.

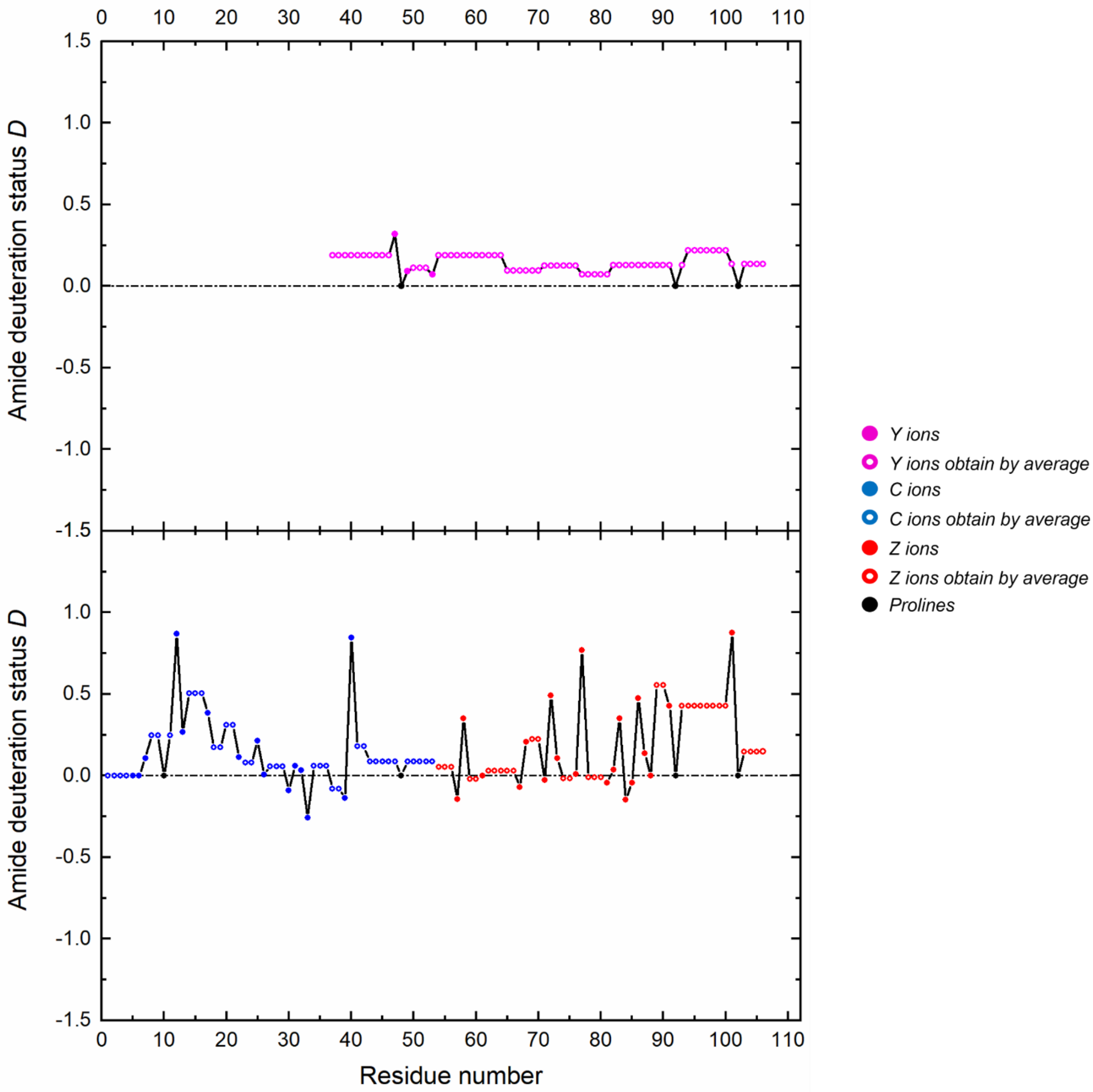


Figure 6.13: Relative % deuteration for different fragment ions of Rru-EncFtn under different conditions at 5 minutes of labelling y ions from the CID y ion products from Rru-EncFtn (top-panel) ECD ions from ID-Rru-EncFtn + Fe, c ions and z ions (bottom panel)

The amide deuteration analyses of the *y* ions show that there was a smaller range of *D* values when compared to the *c* and *z* ions produced from the ID-Rru-EncFtn dataset. Histograms showing frequency of *D* values in both the scrambled and ID-Rru-EncFtn dataset can be seen Figures 6.14 & 6.15 respectively.

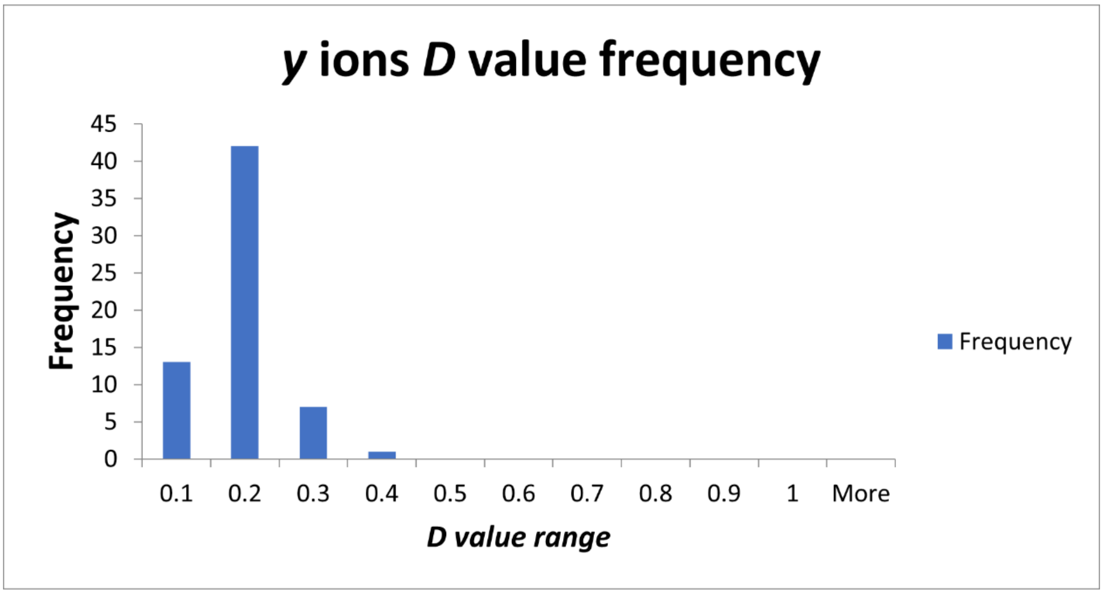


Figure 6.14: Histogram showing the frequency of y ions produced from the scrambling dataset.

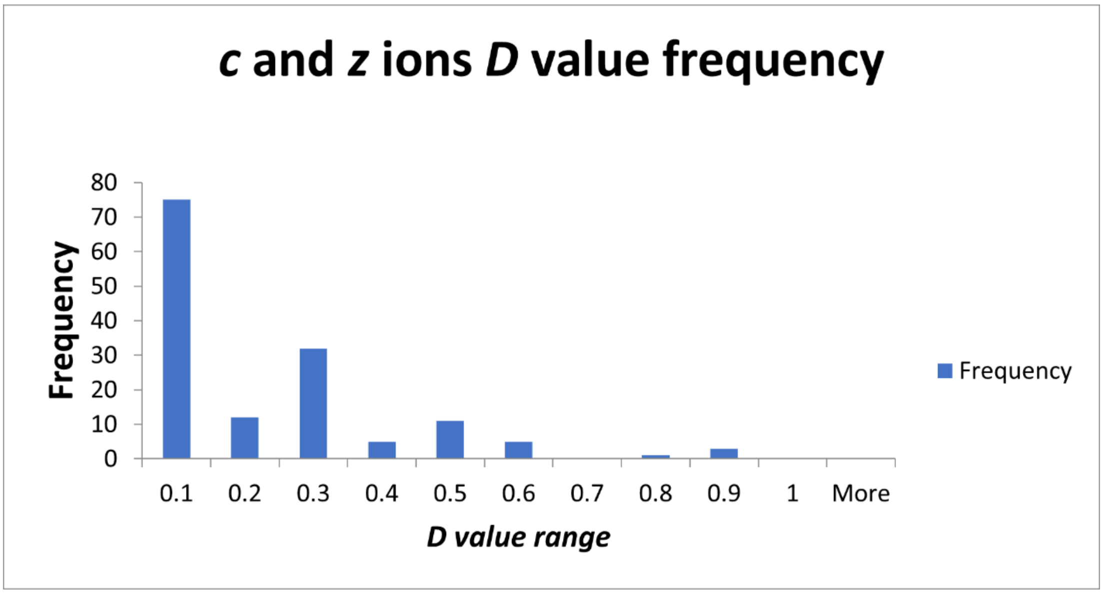


Figure 6.15: Histogram showing the frequency of c and z ions produced from the scrambling dataset.

Total frequency of the *D*-values from both data sets can be seen in Table 6.3

Top-down *c* and *z* ions	
D-value	*Frequency*
0.1	75
0.2	12
0.3	32
0.4	5
0.5	11
0.6	5
0.7	0
0.8	1
0.9	3
1	0
More	0

Scrambled *y* ions	
D-value	*Frequency*
0.1	13
0.2	42
0.3	7
0.4	1
0.5	0
0.6	0
0.7	0
0.8	0
0.9	0
1	0
More	0

Table 6.3: Total frequency for each D-value from both the ID-Rru-EncFtn c and z ion and scrambled y ions datasets

These findings were promising, the *y* ion's d-values were largely held around 0.2. Conversely, the *c* and *z* ions were spread across wide array of values (0.1-0.9). Decrease in protected residues (0.1) and lack of highly deuterating residues (>0.5) in the scrambled dataset is indicative of the uniformity in deuteration throughout its sequence coverage which is a clear marker of HD scrambling. Conversely the *c* and *z ions* do not demonstrate a clear degree of uniformity. This confirms that 100% scrambling is not occurring with the ECD conditions being used on the ID-Rru-EncFtn state.

Although the data can confirm the lack of 100% scrambling, partial scrambling may still be occurring which can bring into question the reliability of this study's findings. To this end it, would be worth performing a dedicated scrambling experiment to further stress test optimal ECD conditions. In 2014 the Rand group developed a synthesized scrambling probe[91,175]. This peptide-based solution is a unique tool to explore impact of gas phase fragmentation on deuterated samples and would be great tool to further stress test the impact of scrambling on this workflow[91,175].

6.4 Bottom-Up analysis of Rru-EncFtn

In order assess the efficacy and success of the top-down workflow a bottom HDX-MS workflow is necessary to provide comparative data. Furthermore, the bottom-up analysis will identify key regions of increased deuteration at the peptide spatial resolution. The exchange profiles of these regions should align with the deuteration profile produced from the top-down analysis, which is at a single amino acid spatial resolution.

After automated HD exchange and quenching Rru-EncFtn samples were loaded onto a pepsin column for digestion and the resulting peptides were separated by reverse phase chromatography using a 10-minute gradient, as detailed in section 2.2. The digestion conditions used proved successful for Rru-EncFtn with most peptides eluting between 4 mins and 7 mins and no undigested protein being retained (Figure 6.16).

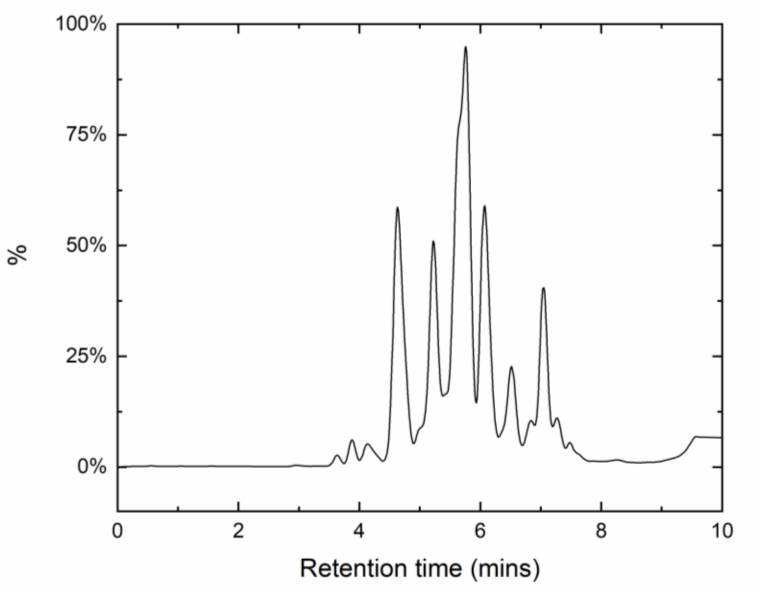


Figure 6.16: Total ion chromatograms for Rru-EncFtn. Separation was performed on a Waters Acquity C18 column using a 10-minute gradient from 5%-95% water – acetonitrile

The bottom-up HDX-MS raw data was analysed using the ProteinLynx Global SERVER (PLGS; Waters) to assign peptides from the MS^E data. Total protein coverage for both Rru-EncFtn were greater >80%. This passed the threshold of acceptable coverage allowing for deuteration assessment using DynamX HDX Data Analysis Software (Waters) to proceed. The overall sequence coverage after digestion from the T0 control was found to be 90.6% from 18 peptides (SI Figure 6.1).

Following analysis of the T0 unlabeled controlled, the deuterium uptake for the subsequent labelling timepoints *(T1: 10s, T2: 30s, T3: 2 minutes, T4: 5 minutes and T5 30 minutes)* was plotted for the bottom-up data. The 5 minutes deuteration coverage map of Rru-EncFtn can be seen in Figure 6.17.

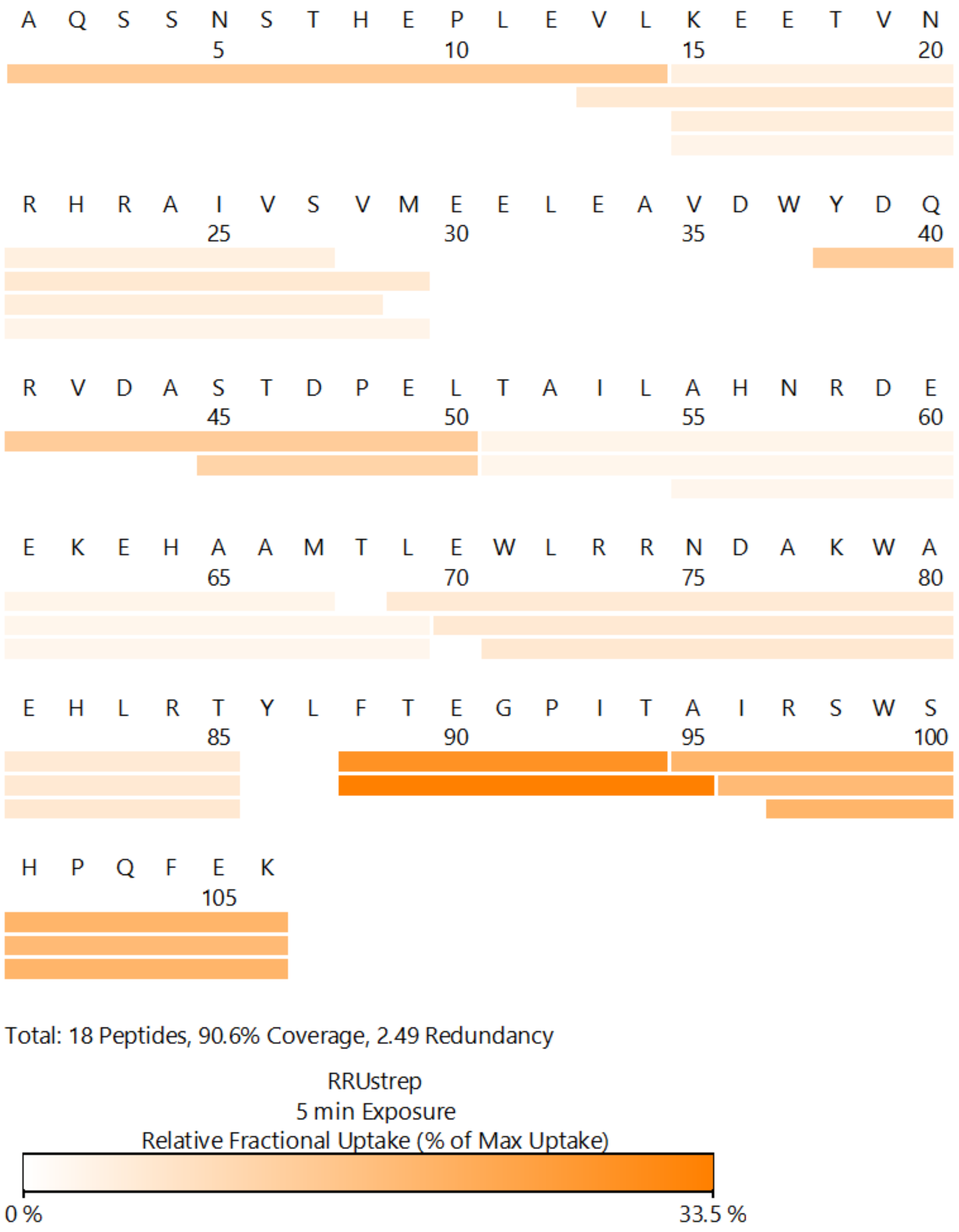


Figure 6.17: HDX Coverage map showing the relative fractional deuterium uptake for Rru-EncFtn after 5 minutes deuterium labelling. There are 18 observed peptides providing 90.6% protein sequence coverage with 2.49 redundancy

The deuteration map show regions of high deuteration rate at N-terminus (^{1}AQSSNSTHEPLEVL^{14} 14±1%), the region within the centre of the monomer (^{38}YDQRVDASTDPEL^{50} 13.3±1.5%) and the C-Terminus (^{88}FTEGPITA^{95} 33.5±1.4%). The bottom-up data at 5 minutes will act as a reference for the subsequent top-down data to assess which residues within the deuterating sites are deuterating and which are protected.

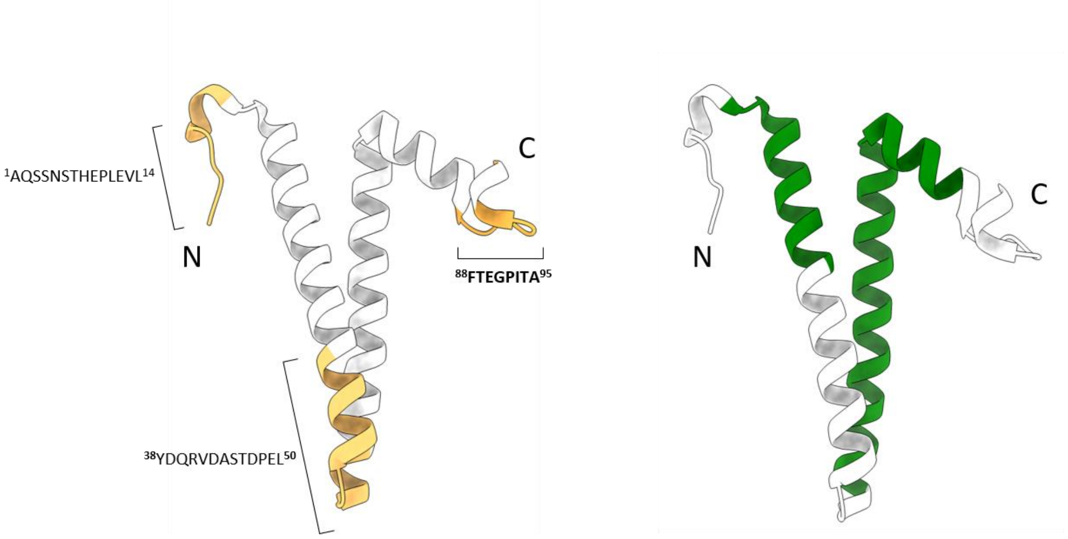


Figure 6.18 (A) HDX Coverage map showing the relative fractional deuterium uptake for Rru-EncFtn after 5 minutes of deuterium labelling overlain the crystal structure of the Rru-EncFtn monomer PDB ID 5L89 (B) Regions of proposed protection highlighted in green

The exchange profile of all peptides over all timepoints can be seen in a butterfly plot which details the uptake across the entire protein sequence in Figure 6.19 and the relative exchange profile of the deuterating peptides is shown in Figure 6.20.

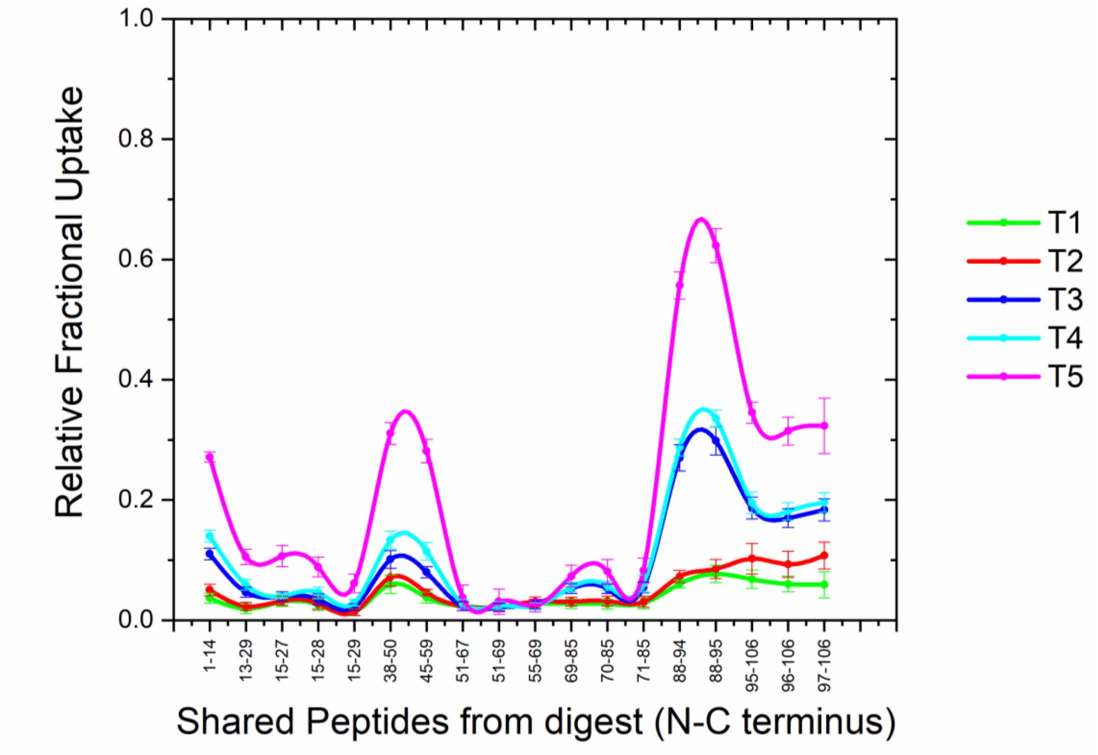


Figure 6.19: Butterfly plot of relative fractional uptake of all Rru-EncFtn *peptides* across *all timepoints 10s, 30s, 2 mins, 5mins and 30mins***.** *Comparability between 13 peptides common to Loaded-Hoch and Empty-Hoch (along the x-axis). The colour code for the timepoints T1 (10 seconds) T2 (30 seconds) T3 (2 minutes) T4 (5 minutes) and T5 (30 minutes)*

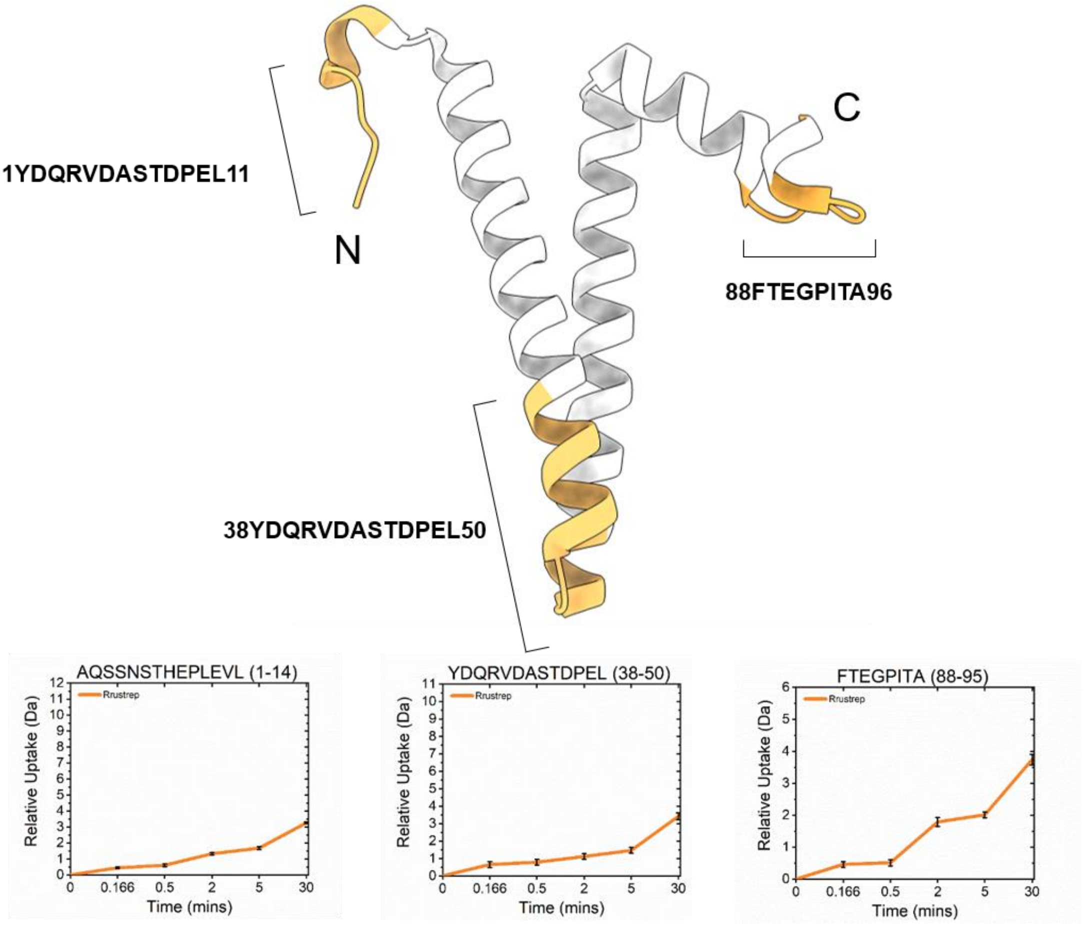


Figure 6.20: HDX-MS/MS of Rru-EncFtn. Timecourse relative deuterium uptake charts, indicated in yellow. (A) Location of residue 1-14 located at the N-terminus 11.2 ± 0.2%, (B) residue 38-50 Located at the center of the monomer 38YDQRVDASTDPEL95 13.3 ± 1.5% (C) Location of residue 88-95 located at the C-terminus

6.4.1 Comparison between Bottom-up and Top-down analysis of Rru-EncFtn

Bottom-up HDX analysis of Rru-EncFtn highlighted 3 key regions which display higher exchange rate (greater than 10% at 5 minutes of labelling).

Analysis of these regions in the top-down HDX data should allow us to identify which amino acids are undergoing a deuteration event within these select peptides. A comparison between the top-down HDX-MS data and the bottom-up analysis at 5 minutes of labelling can be seen in Figure 6.21.

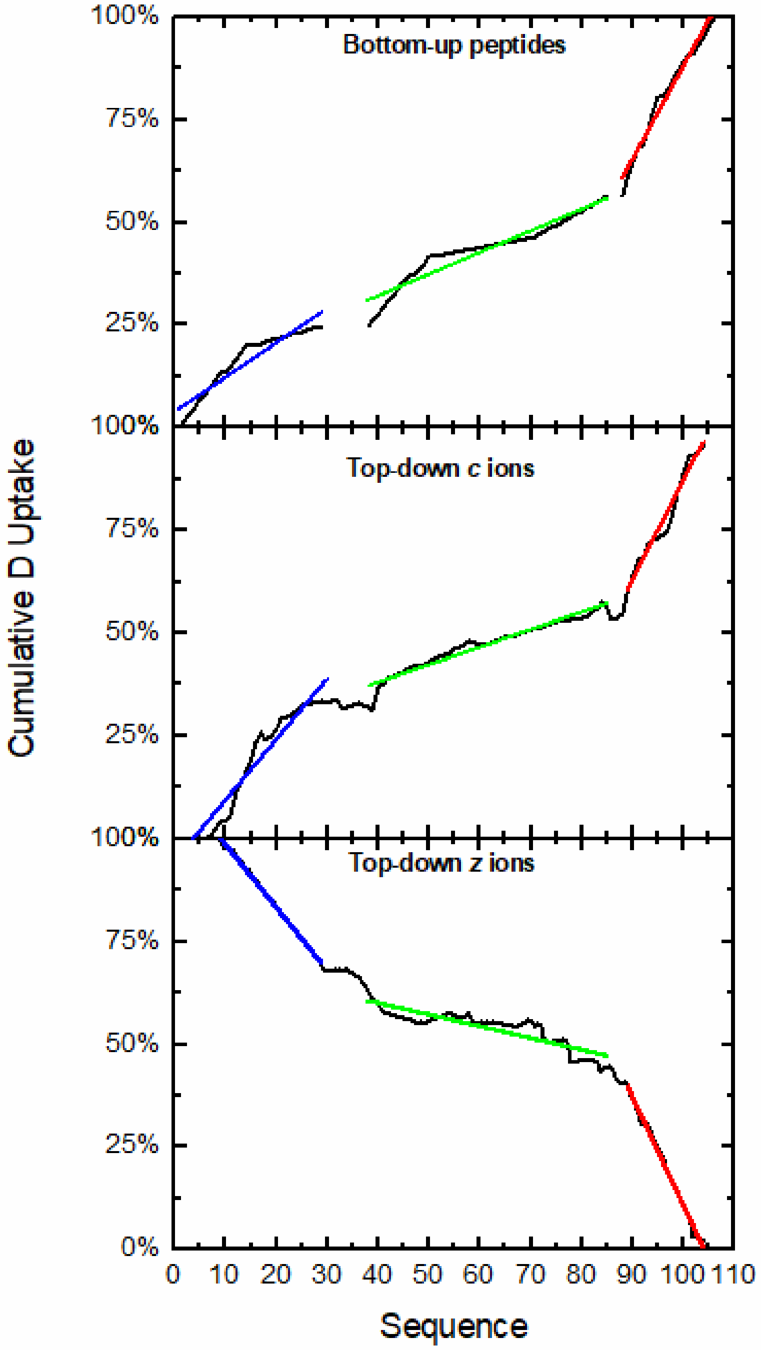


Figure 6.21: Cumulative uptake of deuterium of the bottom-up analysis RRUStrep, and c ion and z ion fragments produced from the top-down analysis. Trend lines have been places over regions of coverage which are shared in all three data sets. The blue trendline covers residues 0-29, the green trend line covers residues 38-85 and the red trendline covers residues 88-106

Figure 6.21 shows culmination of both the bottom up and top down HDX-MS approaches. The trendlines of regions shared between datasets revealed that locations of slow and fast exchanging residues appear identical in both workflows.

The N-terminal region (Blue trendline) contains the ^{1}AQSSNSTHEPLEVL^{14} peptide from the bottom-up data. This peptide experienced 1.7 ± 0.1 Da mass increase after 5 minutes of deuteration.

Direct top-down analysis of the c13 ion demonstrated a similar deuterium uptake of 1.98 Da. However, more detailed information can be extracted from the top-down HDX data – for example, there was little to no exchange in the first 6 amino acids. Both the c5 and c6 showed deuterium uptake of 0.05 and 0.08Da, respectively. In addition, this study successfully identified individual residues displaying high exchange within the same region. This is evidenced by the presence of a highly exchanging E12 residue - which displayed a D value of 0.86. To extract this level of spatial resolution requires a high degree of top-down fragmentation sequence coverage.

In the central region (Green trendline) all datasets produced trendlines which signified relatively slow rate of exchange. Looking at the region in detail, the bottom-up data contained the ^{38}YDQRVDASTDPEL^{50} peptide showing a total uptake of 1.5 ± 0.2 Da at 5 minutes of deuteration. This suggests this region has a high degree of protection from solvent exposure. The trend of slow exchange is mirrored in the top down dataset as well.

Interestingly, the top-down data revealed that Q40 possessed a D value of 0.84 through sequential calculation. This is clear evidence that Q40 is a highly deuterating residue within this region. Although, promising most residue D-values across this region were calculated via D_{avg}, inhibiting direct analysis of individual residues. This further highlights the limitations associated with top-down analysis with low sequence coverage.

The C-terminal region (Red trendline) displayed a high rate of deuteration in both bottom up and top-down datasets. In the bottom-up data, this region contained highest deuterating peptide ^{88}FTEGPITA^{95} (displaying a total uptake of 2 ± 0.1Da). In the top-down analysis, this region was comprised of D_{avg} residues with few sequential residues. However, much of the averaging occurred across fragment which only span one or two residues. This does provide some high-resolution information at a near amide level. For example, the residues T89 and E90 display a D_{avg} of 0.55. From this information it is likely that one of these two amides are responsible for a large proportion of the deuteration in this region. Other deuterating amide will likely rest across the D_{avg} of the 91-100 residues who share a D value of 0.42.

An amino acid of particular interest was A77 which exists as a single amino acid linking loop between two regions of alpha helical secondary structure. The peptide region spanning ^{69}LEWLRRNDAKWAEHLRT^{85} held a relative deuteration of 0.9 ± 0.1 da meaning it is likely that there is one exchange event occurring through this entire region. The subsequent top-down HDX MS analysis revealed that A77 was the only highly deuterating amino acid (D =0.77) within this region. This finding identifies that the linking loop between the alpha helices is solvent exposures and could possess a dynamic and flexible conformation. This finding can also be seen scene in Figure 6.11 where there is a highly exchanging residue which rests in-between α4 and α5 helices.

This study clearly demonstrates the advantages of being able to complement bottom-up HDX-MS data with the increased resolution provided by gas phase fragmentation. However, it also highlights the necessity of having a high sequence coverage allow for greater sequential fragment analysis.

6.5 Conclusion

Previous studies have demonstrated the viability of top-down HDX and the potential for both greater resolution and flexibility in the analysis of proteoforms. Furthermore, previous work has also demonstrated that both ECD and ETD are viable alternative to proteolysis without inducing H→D scrambling. This body of work builds upon this foundation by demonstrating the improved data quality using ID protein forms and introducing an automated quench step and LC step, allowing some degree of automation in the data collection. These findings have successfully demonstrated that it is possible to obtain good sequence coverage during online top-down analysis of deuterated proteins.

Manual analysis of the data demonstrated that the top-down workflow proceeded with minimal HD scrambling and was in complete agreement with data collected using more established bottom-up techniques. However, the sequential fragments obtained in the top-down analysis allowed deuteration rates of individual residues to be determined. This offers the possibility of spatial resolution not possible in bottom-up approaches.

This represents a promising foundation for the workflow to be further developed. The next stage will be to further address the limitations found within this preliminary study. One key limitation is top-down’s reliance on sufficient sequence coverage. Thus, further optimization of ECD fragmentation would be logical next step to take. It would also be interesting to utilize this workflow to explore larger proteins. This would further stress the benefit of using isotopic depletion as larger proteins provide a greater number of fragments ions, further increasing the mass spectral complexity.

Manual analysis of deuterated top-down data is also a limitation. This due to the necessity of manual assessing each fragment’s non-deuterated and deuterated isotopic distributions at each labelling timepoint. This places a significant time constraint on data analysis and inhibits top-down HDX from broader appeal. Further development of software which can provide fragment scoring and DU calculation, like those available for tradition bottom-up HDX-MS workflows, will help address this.

Considering Autovectis already comes with the ability to assign isotopically depleted fragments. Further collaboration with the Kilgour lab could provide a solution which allows for the assignment of deuterated fragments, using user assignable parameters. Example parameters would be: % of deuterium used in the labelling solution, a Tmax assessment of the precursor ion and suitable user control of signal to noise thresholds for peak picking.

Chapter 7 Conclusion

Over the turn of the century, developments in HDX-MS methodologies and technologies have turned it into a key structural technique. During this thesis HDX-MS has been pushed to the limits of what is currently thought possible within the scope of the technique.

To begin with a custom-built bottom-up HDX-MS platform was developed and successfully utilized to investigate the assembly pathway of EncFtn’s decameric assembly. The differential HDX-MS experiment compared the exchange profiles of H63A to Wild Type (WT) EncFtn. The results supported the hypothesis that the decameric assembly requires coordination of iron by FOC residues, including Histidine 63, to allow for the addition of the non-FOC dimers forming the higher order oligomeric states.

With the assembly pathway identified, a larger HDX-MS study was used to explore the larger Enc-EncFtn complex. The first objective was to gather critical structural information of the Enc shell to identify any potential poresites, through which, material could be sequestered into the interior space.

A HDX experiment was devised to identify potential poresites within both the 5-, 3-, and 2-fold symmetry of the icosahedral structure of the Enc marcostructure. The poresite analysis results show that there is an increase in deuteration around the 5-fold symmetry which supports the hypothesis that this is a functioning and dynamic poresite. There was also a significant uptake of deuterium within its 2-fold symmetry which was a previously unidentified poresite. However, the region of significant uptake is also shared with the 5-fold poresite, so it could not be conclusively identified as a distinct functioning poresite.

The project’s next focus was to examine the structures of both a loaded encapsulin (Enc-Loaded) and empty encapsulin (Enc-Empty). Differential HDX-MS revealed regions of significant difference between both the empty and loaded forms. It was found that at later timepoints there were regions with a higher degree of deuteration within the Empty-Enc than the Loaded-Enc. These regions of difference are localized within the interior of the Enc and surround a hypothesised cargo binding region, previously shown in Cryo-EM data.

The project's final focus was to develop a novel top-down HDX-MS workflow. This involved developing a system which coupled a custom-built automated HDX module with LC capability to an FT-ICR MS. This strategy allowed for single residue resolution of a deuterium uptake event with the added utility of proteoform isolation and selection. The setup utilized gas phase fragmentation (ECD) in lieu of proteolysis. This unique approach successfully resulted in the generation of sequential fragments from an EncFtn and achieved single residue resolution of amide deuteration. Furthermore, the use of LC within this workflow mitigated the impact of deuterated side chains on data analysis streamlining data analysis. Finally, the study investigated the use of isotopically depleted EncFtn protein as means of improving data quality. We found that using isotopic depletion resulted in a higher number of assignable fragments being retained within the time course. This minimized the effect of broadening isotopic disruption induced signal suppression.

7.1 Future Direction

HDX-MS is still evolving into a useful tool for structural biology. There several key areas the technique is expanding into, but incorporation of HDX-MS into top-down workflows is a critical one. To this end there are several hurdles that need to be overcome. Further studies should explore the use of isotopic depletion to combat signal suppression from deuteration induced peak broadening. It would be interesting to explore a range of different protein samples with alternative characteristics. Rru-EncFtn provided a good benchmark protein, as it contains a substantial amount of secondary structure. A dedicated experiment focused on a mixed population of post-translationally modified proteins would further demonstrate the benefits of accurate proteoform selection.

Bibliography

(1) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666. https://doi.org/10.1038/181662a0.

(2) Bryn Fenwick, R.; Van Den Bedem, H.; Fraser, J. S.; Wright, P. E. Integrated Description of Protein Dynamics from Room-Temperature X-Ray Crystallography and NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (4), E445. https://doi.org/10.1073/pnas.1323440111.

(3) Dubochet, J.; Lepault, J.; Freeman, R.; Berriman, J. A.; Homo, J. -C. Electron Microscopy of Frozen Water and Aqueous Solutions. *J. Microsc.* **1982**, *128* (3), 219–237. https://doi.org/10.1111/j.1365-2818.1982.tb04625.x.

(4) Murata, K.; Wolf, M. Cryo-Electron Microscopy for Structural Analysis of Dynamic Biological Macromolecules. *Biochimica et Biophysica Acta - General Subjects*. Elsevier B.V. February 1, 2018, pp 324–334. https://doi.org/10.1016/j.bbagen.2017.07.020.

(5) Glaeser, R. M.; Hall, R. J. Reaching the Information Limit in Cryo-EM of Biological Macromolecules: Experimental Aspects. *Biophysical Journal*. Biophysical Society May 18, 2011, pp 2331–2337. https://doi.org/10.1016/j.bpj.2011.04.018.

(6) Sattler, M.; Fesik, S. W. Use of Deuterium Labeling in NMR: Overcoming a Sizeable Problem. *Structure* **1996**, *4* (11), 1245–1249. https://doi.org/10.1016/S0969-2126(96)00133-5.

(7) Kainosho, M.; Miyanoiri, Y.; Terauchi, T.; Takeda, M. Perspective: Next Generation Isotope-Aided Methods for Protein NMR Spectroscopy. *J. Biomol. NMR* **2018**, *71* (3), 119–127. https://doi.org/10.1007/s10858-018-0198-x.

(8) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459. https://doi.org/10.1021/j150664a002.

(9) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J. Am. Soc. Mass Spectrom.* **2017**, *28* (1), 5. https://doi.org/10.1007/S13361-016-1545-3.

(10) Van Duijn, E. Current Limitations in Native Mass Spectrometry Based Structural Biology. *J. Am. Soc. Mass Spectrom. 2* **2011**, *21* (6), 971–978. https://doi.org/10.1016/j.jasms.2009.12.010.

(11) SP, C.; JS, P. Liberating Native Mass Spectrometry from Dependence on Volatile Salt Buffers by Use of Gábor Transform. *Chemphyschem* **2019**, *20* (4), 519–523. https://doi.org/10.1002/CPHC.201900022.

(12) Zhong, Y.; Hyung, S.-J.; Ruotolo, B. T. Ion Mobility–Mass Spectrometry for Structural Proteomics. *Expert Rev. Proteomics* **2012**, *9* (1), 47. https://doi.org/10.1586/EPR.11.75.

(13) Cumeras, R.; Figueras, E.; Davis, C. E.; Baumbach, J. I.; Gràcia, I. Review on Ion Mobility Spectrometry. Part 1: Current Instrumentation. *Analyst* **2015**, *140* (5), 1376. https://doi.org/10.1039/C4AN01100G.

(14) Ujma, J.; Ropartz, D.; Giles, K.; Richardson, K.; Langridge, D.; Wildgoose, J.; Green, M.; Pringle, S. Cyclic Ion Mobility Mass Spectrometry Distinguishes Anomers and Open-Ring Forms of Pentasaccharides. *J. Am. Soc. Mass Spectrom. 2019 306* **2019**, *30* (6), 1028–1037. https://doi.org/10.1007/S13361-019-02168-9.

(15) Eldrid, C.; Thalassinos, K. Developments in Tandem Ion Mobility Mass Spectrometry. *Biochem. Soc. Trans.* **2020**, *48* (6), 2457. https://doi.org/10.1042/BST20190788.

(16) Johnson, D. T.; Stefano, L. H. Di; Jones, L. M. Fast Photochemical Oxidation of Proteins (FPOP): A Powerful Mass Spectrometry–Based Structural Proteomics Tool. *J. Biol. Chem.* **2019**, *294* (32), 11969–

11979. https://doi.org/10.1074/JBC.REV119.006218.

(17) Gau, B. C.; Chen, J.; Gross, M. L. Fast Photochemical Oxidation of Proteins for Comparing Solvent-Accessibility Changes Accompanying Protein Folding: Data Processing and Application to Barstar. *Biochim. Biophys. Acta - Proteins Proteomics* **2013**, *1834* (6), 1230–1238. https://doi.org/10.1016/J.BBAPAP.2013.02.023.

(18) Jones, L. M.; Sperry, J. B.; Carroll, J. A.; Gross, M. L. Fast Photochemical Oxidation of Proteins for Epitope Mapping. *Anal. Chem.* **2011**, *83* (20), 7657–7661. https://doi.org/10.1021/AC2007366.

(19) Zhang, H.; Gau, B. C.; Jones, L. M.; Vidavsky, I.; Gross, M. L. Fast Photochemical Oxidation of Proteins for Comparing Structures of Protein−Ligand Complexes: The Calmodulin−Peptide Model System. *Anal. Chem.* **2010**, *83* (1), 311–318. https://doi.org/10.1021/AC102426D.

(20) Xu, Y.; Falk, I. N.; Hallen, M. A.; Fitzgerald, M. C. Mass Spectrometry- and Lysine Amidination-Based Protocol for Thermodynamic Analysis of Protein Folding and Ligand Binding Interactions. *Anal. Chem.* **2011**, *83* (9), 3555–3562. https://doi.org/10.1021/AC200211T.

(21) Klockenbusch, C.; O'Hara, J. E.; Kast, J. Advancing Formaldehyde Cross-Linking towards Quantitative Proteomic Applications. *Anal. Bioanal. Chem.* **2012**, *404* (4), 1057–1067. https://doi.org/10.1007/S00216-012-6065-9.

(22) Fischer, L.; Chen, Z. A.; Rappsilber, J. Quantitative Cross-Linking/Mass Spectrometry Using Isotope-Labelled Cross-Linkers. *J. Proteomics* **2013**, *88*, 120–128. https://doi.org/10.1016/J.JPROT.2013.03.005.

(23) Becker, J. M.; Naider, F. Cross-Linking Strategies to Study Peptide Ligand-Receptor Interactions. *Methods Enzymol.* **2015**, *556*, 527–547. https://doi.org/10.1016/BS.MIE.2014.12.001.

(24) Larance, M.; Kirkwood, K. J.; Tinti, M.; Murillo, A. B.; Ferguson, M. A. J.; Lamond, A. I. Global Membrane Protein Interactome Analysis Using In Vivo Crosslinking and Mass Spectrometry-Based Protein Correlation Profiling. *Mol. Cell. Proteomics* **2016**, *15* (7), 2476. https://doi.org/10.1074/MCP.O115.055467.

(25) Jensen, P. F.; Rand, K. D. Hydrogen Exchange: A Sensitive Analytical Window into Protein Conformation and Dynamics. *Hydrog. Exch. Mass Spectrom. Proteins* **2016**, *32* (22), 1–17. https://doi.org/10.1002/9781118703748.ch1.

(26) Hudgens, J. W.; Huang, R. Y.; Ambro, E. D. Method Validation and Standards in Hydrogen Exchange Mass Spectrometry. **2016**, 1–22.

(27) Laboratories, C. I.; Waters. HDX System Suitability Test. **2015**, No. December, 1–16.

(28) Zhang, T.; Scheffler, K.; Josephs, J. Optimization of a Fully Automated Hydrogen / Deuterium Exchange Mass Spectrometry Platform for Probe Protein Conformation / Conformation Dynamics. *Thermo Fish. Sci. ASMS 2015* **2015**.

(29) Englander, S. W.; Mayne, L.; Bai, Y.; Sosnick, T. R. Hydrogen Exchange: The Modern Legacy of Linderstrøm-Lang. *Protein Sci.* **1997**, *6* (5), 1101–1109. https://doi.org/10.1002/pro.5560060517.

(30) Hvidt, A.; Linderstrøm-Lang, K. Exchange of Hydrogen Atoms in Insulin with Deuterium Atoms in Aqueous Solutions. *Biochim. Biophys. Acta* **1954**, *14*, 574–575. https://doi.org/10.1016/0006-3002(54)90241-3.

(31) Berger, a; Linderstrom-Lang, K. Deuterium Exchange of Poly-DL-Alanine in Aqueous Solution. *Arch. Biochem. Biophys.* **1957**, *69*, 106–118. https://doi.org/10.1016/0003-9861(57)90478-2.

(32) Benson, E. E.; Linderstrom-Lang, K. Deuterium Exchange between Myoglobin and Water. *Biochim. Biophys. Acta* **1959**, *32*, 579–581. https://doi.org/10.1016/0006-3002(59)90649-3.

(33) Wei, H.; Mo, J.; Tao, L.; Russell, R. J.; Tymiak, A. A.; Chen, G.; Iacob, R. E.; John R. Engen. Hydrogen/Deuterium Exchange Mass Spectrometry for Probing Higher Order Structure of Protein Therapeutics: Methodology and Applications. *Drug Discov Today* **2014**, *19* (1), 95–102. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations.

(34) Anderegg, R. J.; Wagner, D. S. Mass Spectrometric Characterization of a Protein—Ligand Interaction. *J. Am. Chem. Soc.* **1995**, *117* (4), 1374–1377. https://doi.org/10.1021/ja00109a022.

(35) Karpievitch, Y. V.; Polpitiya, A. D.; Anderson, G. A.; Smith, R. D.; Dabney, A. R. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *Ann. Appl. Stat.* **2010**, *4* (4), 1797–1823. https://doi.org/10.1214/10-AOAS341.

(36) El-Shazly, M.; Tai, C. J.; Wu, T. Y.; Csupor, D.; Hohmann, J.; Chang, F. R.; Wu, Y. C. Use, History, and Liquid Chromatography/Mass Spectrometry Chemical Analysis of Aconitum. *J. Food Drug Anal.* **2016**, *24* (1), 29–45. https://doi.org/10.1016/j.jfda.2015.09.001.

(37) Mant, C. T.; Chen, Y.; Yan, Z.; Popa, T. V; Kovacs, J. M.; Mills, J. B.; Tripet, B. P.; Hodges, R. S. HPLC Analysis and Purification of Peptides. *Methods Mol. Biol.* **2007**, *386*, 3–55. https://doi.org/10.1007/978-1-59745-430-8_1.

(38) Nadler, W. M.; Waidelich, D.; Kerner, A.; Hanke, S.; Berg, R.; Trumpp, A.; Rösli, C. MALDI versus ESI: The Impact of the Ion Source on Peptide Identification. *J. Proteome Res.* **2017**, *16* (3), 1207–1215. https://doi.org/10.1021/acs.jproteome.6b00805.

(39) Dole, M.; Mack, L. L.; Hines, R. L.; Chemistry, D. O.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular Beams of Macroions. *J. Chem. Phys.* **1968**, *49* (5), 2240–2249. https://doi.org/10.1063/1.1670391.

(40) Fenn, J. B.; Mann, M.; Meng, C. K. A. I.; Wong, S. F.; Whitehouse, C.

M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science (80-.).* **1989**, *246* (6), 64–71.

(41) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9. https://doi.org/10.1021/ac302789c.

(42) Kebarle, P.; Verkcerk, U. H. Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know Now. *Mass Spectrom. Rev.* **2009**, *28* (6), 898–917. https://doi.org/10.1002/mas.20247.

(43) Iribarne, J. V. On the Evaporation of Small Ions from Charged Droplets. *J. Chem. Phys.* **1976**, *64* (6), 2287. https://doi.org/10.1063/1.432536.

(44) Thomson, B. A.; Iribarne, J. V. Field Induced Ion Evaporation from Liquid Surfaces at Atmospheric Pressure. *J. Chem. Phys.* **1979**, *71* (11), 4451–4463. https://doi.org/10.1063/1.438198.

(45) Konermann, L. A Simple Model for the Disintegration of Highly Charged Solvent Droplets during Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 496–506. https://doi.org/10.1016/j.jasms.2008.11.007.

(46) Jay, F. & P. Biochemistry : A Review with Questions and Explanations. *Biochemistry* **1987**, *50* (1), 50–1572.

(47) Donor, M. T.; Ewing, S. A.; Zenaidee, M. A.; Donald, W. A.; Prell, J. S. Extended Protein Ions Are Formed by the Chain Ejection Model in Chemical Supercharging Electrospray Ionization. *Anal. Chem.* **2017**, *89* (9), 5107–5114. https://doi.org/10.1021/acs.analchem.7b00673.

(48) Konermann, L.; Metwally, H.; Duez, Q.; Peters, I. Charging and Supercharging of Proteins for Mass Spectrometry: Recent Insights into the Mechanisms of Electrospray Ionization. *Analyst* **2019**, *144* (21), 6157–6171. https://doi.org/10.1039/C9AN01201J.

(49) Banerjee, S.; Mazumdar, S. Electrospray Ionization Mass

Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* **2012**, *2012*, 1–40. https://doi.org/10.1155/2012/282574.

(50) Metwally, H.; Duez, Q.; Konermann, L. Chain Ejection Model for Electrospray Ionization of Unfolded Proteins: Evidence from Atomistic Simulations and Ion Mobility Spectrometry. *Anal. Chem.* **2018**, *90* (16), 10069–10077. https://doi.org/10.1021/ACS.ANALCHEM.8B02926.

(51) Peters, I.; Metwally, H.; Konermann, L. Mechanism of Electrospray Supercharging for Unfolded Proteins: Solvent-Mediated Stabilization of Protonated Sites During Chain Ejection. *Anal. Chem.* **2019**. https://doi.org/10.1021/ACS.ANALCHEM.9B01470.

(52) Aliyari, E.; Konermann, L. Formation of Gaseous Proteins via the Ion Evaporation Model (IEM) in Electrospray Mass Spectrometry. *Anal. Chem.* **2020**, *92* (15), 10807–10814. https://doi.org/10.1021/ACS.ANALCHEM.0C02290.

(53) Savaryn, J. P.; Toby, T. K.; Kelleher, N. L. A Researcher's Guide to Mass Spectrometry-Based Proteomics. *Proteomics.* **2016**, *16* (18), 2435–2443. https://doi.org/10.1515/jci-2013-0007.Targeted.

(54) Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A. The Mass-Reflectron, a New Nonmagnetic Time-of-Flight Mass Spectrometer with High Resolution. *Sov Phys JETP* **1973**, *37* (1), 45–48.

(55) Comisarow, M. B.; Marshall, A. G. Theory of Fourier Transform Ion Cyclotron Resonance Mass Spectroscopy. I. Fundamental Equations and Low-pressure Line Shape. *J. Chem. Phys.* **1976**, *64* (1), 110–119. https://doi.org/10.1063/1.431959.

(56) Van Dyck, R. S.; Wineland, D. J.; Ekstrom, P. A.; Dehmelt, H. G. High Mass Resolution with a New Variable Anharmonicity Penning Trap. *Appl. Phys. Lett.* **1976**, *28* (8), 446–448.

https://doi.org/10.1063/1.88793.

(57) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; R., Y. I. J. Protein Analysis by Shotgun/Bottom-up Proteomics Yaoyang. *chem.rev* **2009**, *6* (4), 247–253. https://doi.org/10.1111/j.1743-6109.2008.01122.x.Endothelial.

(58) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol.* **2005**, *402* (1993), 148–185. https://doi.org/10.1016/S0076-6879(05)02005-7.

(59) Han, X.; Aslanian, A.; Yates, J. Mass Spectrometry for Proteomics. *Curr Opin Chem Biol.* **2008**, *12* (5), 483–490. https://doi.org/10.1016/j.cbpa.2008.07.024.Mass.

(60) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266. https://doi.org/10.1021/ja973478k.

(61) Zhang, H.; Ge, Y. Comprehensive Analysis of Protein Modifications by Top-down Mass Spectrometry. *Circulation. Cardiovascular genetics*. NIH Public Access 2011, p 711. https://doi.org/10.1161/CIRCGENETICS.110.957829.

(62) Cooper, H. J.; Håkansson, K.; Marshall, A. G. The Role of Electron Capture Dissociation in Biomolecular Analysis. *Mass Spectrometry Reviews*. March 2005, pp 201–222. https://doi.org/10.1002/mas.20014.

(63) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528. https://doi.org/10.1073/PNAS.0402700101.

(64) Jones, A. W.; Cooper, H. J. Probing the Mechanisms of Electron Capture Dissociation Mass Spectrometry with Nitrated Peptides. *Phys. Chem. Chem. Phys.* **2010**, *12* (41), 13394.

https://doi.org/10.1039/C0CP00623H.

(65) Syrstad, E. A.; Tureček, F. Toward a General Mechanism of Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (2), 208–224. https://doi.org/10.1016/j.jasms.2004.11.001.

(66) Sobczyk, M.; Anusiewicz, I.; Berdys-Kochanska, J.; Sawicka, A.; Skurski, P.; Simons, J. Coulomb-Assisted Dissociative Electron Attachment: Application to a Model Peptide. *J. Phys. Chem. A* **2005**, *109* (1), 250–258. https://doi.org/10.1021/jp0463114.

(67) Iavarone, A. T.; Paech, K.; Williams, E. R. Effects of Charge State and Cationizing Agent on the Electron Capture Dissociation of a Peptide. *Anal. Chem.* **2004**, *76* (8), 2231–2238. https://doi.org/10.1021/ac035431p.

(68) Schneeberger, E. M.; Breuker, K. Replacing H+ by Na+ or K+ in Phosphopeptide Anions and Cations Prevents Electron Capture Dissociation. *Chem. Sci.* **2018**, *9* (37), 7338–7353. https://doi.org/10.1039/c8sc02470g.

(69) Zhu, X.; Chen, Y.; Subramanian, R. Comparison of Information-Dependent Acquisition, SWATH, and MSAlltechniques in Metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2014**, *86* (2), 1202–1209. https://doi.org/10.1021/ac403385y.

(70) Li, G. Z.; Vissers, J. P. C.; Silva, J. C.; Golick, D.; Gorenstein, M. V.; Geromanos, S. J. Database Searching and Accounting of Multiplexed Precursor and Product Ion Spectra from the Data Independent Analysis of Simple and Complex Peptide Mixtures. *Proteomics* **2009**, *9* (6), 1696–1719. https://doi.org/10.1002/pmic.200800564.

(71) Volmer, Y.; A. Dietrich. Electron-Based Fragmentation Methods in Mass Spectrometry an Overview. *Mass Spectrom. Rev.* **2017**, *36*, 4–15. https://doi.org/10.1002/mas.

(72) Johnson, A. R.; Carlson, E. E. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. **2015**. https://doi.org/10.1021/acs.analchem.5b01543.

(73) Neuhauser, N.; Nagaraj, N.; McHardy, P.; Zanivan, S.; Scheltema, R.; Cox, J.; Mann, M. High Performance Computational Analysis of Large-Scale Proteome Data Sets to Assess Incremental Contribution to Coverage of the Human Genome. *J. Proteome Res.* **2013**, *12* (6), 2858–2868. https://doi.org/10.1021/pr400181q.

(74) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. UPLC/MSE; a New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation. *Rapid Commun. Mass Spectrom.* **2006**, *20* (13), 1989–1994. https://doi.org/10.1002/rcm.2550.

(75) Law, K.; Pin Lim, Y. Recent Advances in Mass Spectrometry: Data Independent Analysis and Hyper Reaction Monitoring. *Expert Rev. Proteomics* **2013**, *10*, 551–566.

(76) Wales, T. E.; Eggertson, M. J.; Engen, J. R. Considerations in the Analysis of Hydrogen Exchange Mass Spectrometry Data. *Methods Mol. Biol.* **2013**, *1007*, 263–288. https://doi.org/10.1007/978-1-62703-392-3_11.

(77) Tsirigotaki, A.; Elzen, R. Van; Veken, P. Van Der; Lambeir, A.-M.; Economou, A. Dynamics and Ligand-Induced Conformational Changes in Human Prolyl Oligopeptidase Analyzed by Hydrogen/Deuterium Exchange Mass Spectrometry. *Sci. Reports 2017 71* **2017**, *7* (1), 1–13. https://doi.org/10.1038/s41598-017-02550-1.

(78) Rand, K. D.; Pringle, S. D.; Morris, M.; Brown, J. M. Site-Specific Analysis of Gas-Phase Hydrogen/Deuterium Exchange of Peptides and Proteins by Electron Transfer Dissociation. *Anal. Chem.* **2012**, *84* (4), 1931–1940. https://doi.org/10.1021/AC202918J.

(79) Rudowska, M.; Wojewska, D.; Kluczyk, A.; Bąchor, R.; Stefanowicz, P.; Szewczuk, Z. The Hydrogen–Deuterium Exchange at α-Carbon Atom in N,N,N-Trialkylglycine Residue: ESI-MS Studies. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (6), 1024. https://doi.org/10.1007/S13361-012-0359-1.

(80) Atzrodt, J.; Derdau, V.; Fey, T.; Zimmermann, J. The Renaissance of H/D Exchange. *Angew. Chemie Int. Ed.* **2007**, *46* (41), 7744–7765. https://doi.org/10.1002/ANIE.200700039.

(81) Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. Protein Ionizable Groups: PK Values and Their Contribution to Protein Stability and Solubility. *J. Biol. Chem.* **2009**, *284* (20), 13285–13289. https://doi.org/10.1074/jbc.R800080200.

(82) Konermann, L.; Rodriguez, A. D.; Sowole, M. A. Type 1 and Type 2 Scenarios in Hydrogen Exchange Mass Spectrometry Studies on Protein-Ligand Complexes. **2014**. https://doi.org/10.1039/c4an01307g.

(83) Debra M. Ferraro; Noel D. Lazo, ‡ and; Robertson*, A. D. EX1 Hydrogen Exchange and Protein Folding†. *Biochemistry* **2003**, *43* (3), 587–594. https://doi.org/10.1021/BI035943Y.

(84) Balen, B.; Krsnik-rasol, M.; Zamfir, A. D.; Zadro, I.; Vakhrushev, S. Y. AB R F Assessment Of. **2007**, *18* (3), 162–172.

(85) Walters, B. T.; Ricciuti, A.; Mayne, L.; Englander, S. W. Minimizing Back Exchange in the Hydrogen Exchange Experiment. *J Am Soc Mass Spectrom* **2012**, *23* (12), 1–9. https://doi.org/10.1007/s13361-012-0476-x.MINIMIZING.

(86) Zhang, Z.; Zhang, A.; Xiao, G. Improved Protein Hydrogen/Deuterium Exchange Mass Spectrometry Platform with Fully Automated Data Processing. *Anal. Chem.* **2012**, *84* (11), 4942–4949. https://doi.org/10.1021/ac300535r.

(87) Hoofnagle, A. N.; Resing, K. A.; Ahn, N. G. Protein Analysis by

Hydrogen Exchange Mass Spectrometry. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32* (1), 1–25. https://doi.org/10.1146/annurev.biophys.32.110601.142417.

(88) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Brier, S.; Bou-assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M. Europe PMC Funders Group Recommendations for Performing , Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments. **2019**, *16* (7), 595–602. https://doi.org/10.1038/s41592-019-0459-y.Recommendations.

(89) Abzalimov, R. R.; Kaplan, D. A.; Easterling, M. L.; Kaltashov, I. A. Protein Conformations Can Be Probed in Top-Down HDX MS Experiments Utilizing Electron Transfer Dissociation of Protein Ions Without Hydrogen Scrambling. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1514–1517. https://doi.org/10.1016/J.JASMS.2009.04.006.

(90) Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. Hydrogen/Deuterium Exchange Mass Spectrometry with Top-down Electron Capture Dissociation for Characterizing Structural Transitions of a 17 KDa Protein. *J. Am. Chem. Soc.* **2009**, *131* (35), 12801–12808. https://doi.org/10.1021/ja904379w.

(91) Rand, K. D.; Zehl, M.; Jørgensen, T. J. D. Measuring the Hydrogen/Deuterium Exchange of Proteins at High Spatial Resolution by Mass Spectrometry: Overcoming Gas-Phase Hydrogen/Deuterium Scrambling. *Acc. Chem. Res.* **2014**, *47* (10), 3018–3027. https://doi.org/10.1021/ar500194w.

(92) Andrews, S. C.; Robinson, A. K.; Rodríguez-Quiñones, F. Bacterial Iron Homeostasis. *FEMS Microbiology Reviews*. 2003. https://doi.org/10.1016/S0168-6445(03)00055-X.

(93) Andrews, S. C. Iron Storage in Bacteria. *Adv. Microb. Physiol.* **1998**.

https://doi.org/10.1016/s0065-2911(08)60134-4.

(94) Saito, H. Metabolism of Iron Stores. *Nagoya Journal of Medical Science*. 2014. https://doi.org/10.18999/nagjms.76.3-4.235.

(95) MV, L. Sur La Cristallisation de La Ferritine. *Bull Soc Chim Biol* **1937**, *19:* (1575.).

(96) Granick, S. FERRITIN: I. PHYSICAL AND CHEMICAL PROPERTIES OF HORSE SPLEEN FERRITIN. *J. Biol. Chem.* **1942**, *146* (2), 451–461. https://doi.org/10.1016/S0021-9258(18)44965-4.

(97) Halterman, J. S.; Segel, G. B. Anemia, Iron Deficiency. In *Pediatric Clinical Advisor*; Elsevier Inc., 2007; pp 31–31. https://doi.org/10.1016/B978-032303506-4.10019-7.

(98) Wang, P.; Wu, Q.; Wu, W.; Li, H.; Guo, Y.; Yu, P.; Gao, G.; Shi, Z.; Zhao, B.; Chang, Y. Z. Mitochondrial Ferritin Deletion Exacerbates β - Amyloid-Induced Neurotoxicity in Mice. *Oxid. Med. Cell. Longev.* **2017**. https://doi.org/10.1155/2017/1020357.

(99) Mesquita, G.; Silva, T.; Gomes, A. C.; Oliveira, P. F.; Alves, M. G.; Fernandes, R.; Almeida, A. A.; Moreira, A. C.; Gomes, M. S. H-Ferritin Is Essential for Macrophages' Capacity to Store or Detoxify Exogenously Added Iron. *Sci. Rep.* **2020**. https://doi.org/10.1038/s41598-020-59898-0.

(100) Sahandi Zangabad, P.; Karimi, M.; Mehdizadeh, F.; Malekzad, H.; Ghasemi, A.; Bahrami, S.; Zare, H.; Moghoofei, M.; Hekmatmanesh, A.; Hamblin, M. R. Nanocaged Platforms: Modification, Drug Delivery and Nanotoxicity. Opening Synthetic Cages to Release the Tiger. *Nanoscale*. 2017. https://doi.org/10.1039/c6nr07315h.

(101) Carrondo, M. A. Ferritins, Iron Uptake and Storage from the Bacterioferritin Viewpoint. *EMBO Journal*. John Wiley & Sons, Ltd May 1, 2003, pp 1959–1968. https://doi.org/10.1093/emboj/cdg215.

(102) Chiou, B.; Connor, J. R. Emerging and Dynamic Biomedical Uses of Ferritin. *Pharmaceuticals*. 2018. https://doi.org/10.3390/ph11040124.

(103) Wang, Z.; Gao, H.; Zhang, Y.; Liu, G.; Niu, G.; Chen, X. Functional Ferritin Nanoparticles for Biomedical Applications. *Frontiers of Chemical Science and Engineering*. Higher Education Press December 1, 2017, pp 633–646. https://doi.org/10.1007/s11705-017-1620-8.

(104) Senapati, S.; Mahanta, A. K.; Kumar, S.; Maiti, P. Controlled Drug Delivery Vehicles for Cancer Treatment and Their Performance. *Signal Transduction and Targeted Therapy*. 2018. https://doi.org/10.1038/s41392-017-0004-3.

(105) Colombo, F.; Durigutto, P.; De Maso, L.; Biffi, S.; Belmonte, B.; Tripodo, C.; Oliva, R.; Bardini, P.; Marini, G. M.; Terreno, E.; Pozzato, G.; Rampazzo, E.; Bertrand, J.; Feuerstein, B.; Javurek, J.; Havrankova, J.; Pitzalis, C.; Nuñez, L.; Meroni, P.; Tedesco, F.; Sblattero, D.; Macor, P. Targeting CD34+ Cells of the Inflamed Synovial Endothelium by Guided Nanoparticles for the Treatment of Rheumatoid Arthritis. *J. Autoimmun.* **2019**. https://doi.org/10.1016/j.jaut.2019.05.016.

(106) Lawson, D. M.; Treffry, A.; Artymiuk, P. J.; Harrison, P. M.; Yewdall, S. J.; Luzzago, A.; Cesareni, G.; Levi, S.; Arosio, P. Identification of the Ferroxidase Centre in Ferritin. *FEBS Lett.* **1989**, *254* (1–2), 207–210. https://doi.org/10.1016/0014-5793(89)81040-3.

(107) Kim, M.; Rho, Y.; Jin, K. S.; Ahn, B.; Jung, S.; Kim, H.; Ree, M. PH-Dependent Structures of Ferritin and Apoferritin in Solution: Disassembly and Reassembly. *Biomacromolecules* **2011**, *12* (5), 1629–1640. https://doi.org/10.1021/bm200026v.

(108) Harrison, P. M.; Arosio, P. The Ferritins: Molecular Properties, Iron Storage Function and Cellular Regulation. *Biochimica et Biophysica Acta - Bioenergetics*. Elsevier B.V. July 31, 1996, pp 161–203.

https://doi.org/10.1016/0005-2728(96)00022-9.

(109) Bou-Abdallah, F. The Iron Redox and Hydrolysis Chemistry of the Ferritins. *Biochimica et Biophysica Acta - General Subjects*. Biochim Biophys Acta August 2010, pp 719–731. https://doi.org/10.1016/j.bbagen.2010.03.021.

(110) He, D.; Hughes, S.; Vanden-Hehir, S.; Georgiev, A.; Altenbach, K.; Tarrant, E.; Mackay, C. L.; Waldron, K. J.; Clarke, D. J.; Marles-Wright, J. Structural Characterization of Encapsulated Ferritin Provides Insight into Iron Storage in Bacterial Nanocompartments. *Elife* **2016**, *5* (AUGUST). https://doi.org/10.7554/eLife.18972.

(111) Helgstrand, C.; Wikoff, W. R.; Duda, R. L.; Hendrix, R. W.; Johnson, J. E.; Liljas, L. The Refined Structure of a Protein Catenane: The HK97 Bacteriophage Capsid at 3.44 Å Resolution. *J. Mol. Biol.* **2003**, *334* (5), 885–899. https://doi.org/10.1016/j.jmb.2003.09.035.

(112) Giessen, T. W.; Orlando, B. J.; Verdegaal, A. A.; Chambers, M. G.; Gardener, J.; Bell, D. C.; Birrane, G.; Liao, M.; Silver, P. A. Large Protein Organelles Form a New Iron Sequestration System with High Storage Capacity. *Elife* **2019**, *8*. https://doi.org/10.7554/eLife.46070.

(113) McHugh, C. A.; Fontana, J.; Nemecek, D.; Cheng, N.; Aksyuk, A. A.; Heymann, J. B.; Winkler, D. C.; Lam, A. S.; Wall, J. S.; Steven, A. C.; Hoiczyk, E. A Virus Capsid-like Nanocompartment That Stores Iron and Protects Bacteria from Oxidative Stress. *EMBO J.* **2014**. https://doi.org/10.15252/embj.201488566.

(114) Rurup, W. F.; Snijder, J.; Koay, M. S. T.; Heck, A. J. R.; Cornelissen, J. J. L. M. Self-Sorting of Foreign Proteins in a Bacterial Nanocompartment. *J. Am. Chem. Soc.* **2014**, *136* (10), 3828–3832. https://doi.org/10.1021/ja410891c.

(115) Piergentili, C.; Ross, J.; He, D.; Gallagher, K. J.; Stanley, W. A.; Adam, L.; Logan Mackay, C.; Baslé, A.; Waldron, K. J.; Clarke, D. J.; Marles-

Wright, J. Dissecting the Structural and Functional Roles of a Putative Metal Entry Site in Encapsulated Ferritins. *J. Biol. Chem.* **2020**, *295* (46), 15511–15526. https://doi.org/10.1074/jbc.RA120.014502.

(116) He, D.; Piergentili, C.; Ross, J.; Tarrant, E.; Tuck, L. R.; Logan Mackay, C.; McIver, Z.; Waldron, K. J.; Clarke, D. J.; Marles-Wright, J. Conservation of the Structural and Functional Architecture of Encapsulated Ferritins in Bacteria and Archaea. *Biochem. J.* **2019**, *476* (6), 975–989. https://doi.org/10.1042/BCJ20180922.

(117) Ross, J.; Lambert, T.; Piergentili, C.; He, D.; Waldron, K. J.; Mackay, C. L.; Marles-Wright, J.; Clarke, D. J. Mass Spectrometry Reveals the Assembly Pathway of Encapsulated Ferritins and Highlights a Dynamic Ferroxidase Interface. *Chem. Commun.* **2020**, *56* (23), 3417–3420. https://doi.org/10.1039/c9cc08130e.

(118) Kerfeld, C. A.; Aussignargues, C.; Zarzycki, J.; Cai, F.; Sutter, M. Bacterial Microcompartments. *Nat. Rev. Microbiol.* **2018**, *16* (5), 277–290. https://doi.org/10.1038/nrmicro.2018.10.

(119) Gabashvili, A. N.; Chmelyuk, N. S.; Efremova, M. V.; Malinovskaya, J. A.; Semkina, A. S.; Abakumov, M. A. Encapsulins—Bacterial Protein Nanocompartments: Structure, Properties, and Application. *Biomolecules* **2020**, *10* (6), 1–12. https://doi.org/10.3390/BIOM10060966.

(120) Giessen, T. W.; Silver, P. A. Widespread Distribution of Encapsulin Nanocompartments Reveals Functional Diversity. *Nat. Microbiol. 2017 26* **2017**, *2* (6), 1–11. https://doi.org/10.1038/NMICROBIOL.2017.29.

(121) Akita, F.; Chong, K. T.; Tanaka, H.; Yamashita, E.; Miyazaki, N.; Nakaishi, Y.; Suzuki, M.; Namba, K.; Ono, Y.; Tsukihara, T.; Nakagawa, A. The Crystal Structure of a Virus-like Particle from the Hyperthermophilic Archaeon Pyrococcus Furiosus Provides Insight into the Evolution of Viruses. *J. Mol. Biol.* **2007**, *368* (5), 1469–1483.

https://doi.org/10.1016/j.jmb.2007.02.075.

(122) He, D.; Marles-Wright, J. Ferritin Family Proteins and Their Use in Bionanotechnology. *N. Biotechnol.* **2015**, *32* (6), 651–657. https://doi.org/10.1016/j.nbt.2014.12.006.

(123) Altenburg, W. J.; Rollins, N.; Silver, P. A.; Giessen, T. W. Exploring Targeting Peptide-Shell Interactions in Encapsulin Nanocompartments. *Sci. Rep.* **2021**, *11* (1), 1–9. https://doi.org/10.1038/s41598-021-84329-z.

(124) Sánchez-Rico, C.; Voith von Voithenberg, L.; Warner, L.; Lamb, D. C.; Sattler, M. Effects of Fluorophore Attachment on Protein Conformation and Dynamics Studied by SpFRET and NMR Spectroscopy. *Chem. - A Eur. J.* **2017**, *23* (57), 14267–14277. https://doi.org/10.1002/chem.201702423.

(125) Luitz, M. P.; Barth, A.; Crevenna, A. H.; Bomblies, R.; Lamb, D. C.; Zacharias, M. Covalent Dye Attachment Influences the Dynamics and Conformational Properties of Flexible Peptides. *PLoS One* **2017**, *12* (5), e0177139. https://doi.org/10.1371/journal.pone.0177139.

(126) Sutter, M.; Boehringer, D.; Gutmann, S.; Günther, S.; Prangishvili, D.; Loessner, M. J.; Stetter, K. O.; Weber-Ban, E.; Ban, N. Structural Basis of Enzyme Encapsulation into a Bacterial Nanocompartment. *Nat. Struct. Mol. Biol.* **2008**, *15* (9), 939–947. https://doi.org/10.1038/nsmb.1473.

(127) Tamura, A.; Fukutani, Y.; Takami, T.; Fujii, M.; Nakaguchi, Y.; Murakami, Y.; Noguchi, K.; Yohda, M.; Odaka, M. Packaging Guest Proteins into the Encapsulin Nanocompartment from Rhodococcus Erythropolis N771. *Biotechnol. Bioeng.* **2015**, *112* (1), 13–20. https://doi.org/10.1002/BIT.25322.

(128) Ross, J.; McIver, Z.; Lambert, T.; Piergentili, C.; Gallagher, K. J.; Bird, J. E.; Cruickshank, F. L.; Zarazúa-Arvizu, E.; Horsfall, L. E.; Waldron,

K. J.; Wilson, M. D.; Mackay, C. L.; Baslé, A.; Clarke, D. J.; Marles-Wright, J. Pore Dynamics and Asymmetric Cargo Loading in an Encapsulin Nanocompartment. *bioRxiv* **2021**, 2021.04.15.439977. https://doi.org/10.1101/2021.04.15.439977.

(129) Kilgour, D. P. A.; Hughes, S.; Kilgour, S. L.; Mackay, C. L.; Palmblad, M.; Tran, B. Q.; Goo, Y. A.; Ernst, R. K.; Clarke, D. J.; Goodlett, D. R. Autopiquer - a Robust and Reliable Peak Detection Algorithm for Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (2), 253–262. https://doi.org/10.1007/s13361-016-1549-z.

(130) Kilgour, D. P. A.; Wills, R.; Qi, Y.; O'Connor, P. B. Autophaser: An Algorithm for Automated Generation of Absorption Mode Spectra for FT-ICR MS. *Anal. Chem.* **2013**, *85* (8), 3903–3911. https://doi.org/10.1021/ac303289c.

(131) Bou-Assaf, G. M.; Chamoun, J. E.; Emmett, M. R.; Fajer, P. G.; Marshall, A. G. Advantages of Isotopic Depletion of Proteins for Hydrogen/Deuterium Exchange Experiments Monitored by Mass Spectrometry. *Anal. Chem.* **2010**, *82* (8), 3293–3299. https://doi.org/10.1021/ac100079z.

(132) Hussong, R.; Tholey, A.; Hildebrandt, A. Efficient Analysis of Mass Spectrometry Data Using the Isotope Wavelet. In *AIP Conference Proceedings*; American Institute of PhysicsAIP, 2007; Vol. 940, pp 139–149. https://doi.org/10.1063/1.2793396.

(133) Vanek, T.; Kohli, A. *Biochemistry, Myoglobin*; StatPearls Publishing, 2019.

(134) Enyenihi, A. A.; Yang, H.; Jimmy Ytterberg, A.; Lyutvinskiy, Y.; Zubarev, R. A. Heme Binding in Gas-Phase Holo-Myoglobin Cations: Distal Becomes Proximal? *J. Am. Soc. Mass Spectrom.* **2011**, *22* (10), 1763–1770. https://doi.org/10.1007/s13361-011-0182-0.

(135) Landgraf, R. R.; Chalmers, M. J.; Griffin, P. R. Automated

Hydrogen/Deuterium Exchange Electron Transfer Dissociation High Resolution Mass Spectrometry Measured at Single-Amide Resolution. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (2), 301–309. https://doi.org/10.1007/s13361-011-0298-2.

(136) Del Mar, C.; Greenbaum, E. A.; Mayne, L.; Englander, S. W.; Woods, V. L. Structure and Properties of α-Synuclein and Other Amyloids Determined at the Amino Acid Level. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15477–15482. https://doi.org/10.1073/pnas.0507405102.

(137) Illes-Toth, E.; Rempel, D. L.; Gross, M. L. Pulsed Hydrogen-Deuterium Exchange Illuminates the Aggregation Kinetics of α-Synuclein, the Causative Agent for Parkinson’s Disease. *ACS Chem. Neurosci.* **2018**, *9* (6), 1469–1476. https://doi.org/10.1021/acschemneuro.8b00052.

(138) Suchanova, B.; Tuma, R. Folding and Assembly of Large Macromolecular Complexes Monitored by Hydrogen-Deuterium Exchange and Mass Spectrometry. *Microbial Cell Factories*. BioMed Central April 4, 2008, pp 1–20. https://doi.org/10.1186/1475-2859-7-12.

(139) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **2007**, *372* (3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.

(140) Institute., E. B. Protein Interfaces, Surfaces and Assemblies.

(141) He, M. M.; Clugston, S. L.; Honek, J. F.; Matthews, B. W. Determination of the Structure of Escherichia Coli Glyoxalase I Suggests a Structural Basis for Differential Metal Activation. *Biochemistry* **2000**, *39* (30), 8719–8727. https://doi.org/10.1021/bi000856g.

(142) Kihlken, M. A.; Leech, A. P.; Le Brun, N. E. Copper-Mediated Dimerization of CopZ, a Predicted Copper Chaperone from Bacillus Subtilis. *Biochem. J.* **2002**, *368* (3), 729–739.

https://doi.org/10.1042/BJ20021036.

(143) Partin, A. C.; Ngo, T. D.; Herrell, E.; Jeong, B. C.; Hon, G.; Nam, Y. Heme Enables Proper Positioning of Drosha and DGCR8 on Primary MicroRNAs. *Nat. Commun.* **2017**, *8* (1). https://doi.org/10.1038/s41467-017-01713-y.

(144) Sanguantrakun, N.; Chanthamontri, C.; Gross, M. L. Top-Down Analysis of In-Source HDX of Native Protein Ions. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (5), 1151–1154. https://doi.org/10.1021/jasms.9b00149.

(145) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down Proteomics: Facts and Perspectives. *Biochem. Biophys. Res. Commun.* **2015**, *445* (4), 683–693. https://doi.org/10.1016/j.bbrc.2014.02.041.

(146) Brodie, N. I.; Huguet, R.; Zhang, T.; Viner, R.; Zabrouskov, V.; Pan, J.; Petrotchenko, E. V.; Borchers, C. H. Top-Down Hydrogen-Deuterium Exchange Analysis of Protein Structures Using Ultraviolet Photodissociation. *Anal. Chem.* **2018**, *90* (5), 3079–3082. https://doi.org/10.1021/acs.analchem.7b03655.

(147) Tuma, R.; Coward, L. U.; Kirk, M. C.; Barnes, S.; Prevelige, P. E. Hydrogen-Deuterium Exchange as a Probe of Folding and Assembly in Viral Capsids. *J. Mol. Biol.* **2001**, *306* (3), 389–396. https://doi.org/10.1006/jmbi.2000.4383.

(148) Guttman, M.; Lee, K. K. Isotope Labeling of Biomolecules: Structural Analysis of Viruses by HDX-MS. In *Methods in Enzymology*; Academic Press Inc., 2016; Vol. 566, pp 405–426. https://doi.org/10.1016/bs.mie.2015.05.021.

(149) Gertsman, I.; Gan, L.; Guttman, M.; Lee, K.; Speir, J. A.; Duda, R. L.; Hendrix, R. W.; Komives, E. A.; Johnson, J. E. An Unexpected Twist in Viral Capsid Maturation. *Nature* **2009**, *458* (7238), 646–650.

https://doi.org/10.1038/nature07686.

(150) Gertsman, I.; Fu, C. Y.; Huang, R.; Komives, E. A.; Johnson, J. E. Critical Salt Bridges Guide Capsid Assembly, Stability, and Maturation Behavior in Bacteriophage HK97. *Mol. Cell. Proteomics* **2010**, *9* (8), 1752–1763. https://doi.org/10.1074/mcp.M000039-MCP201.

(151) Williams, E. M.; Jung, M.; Coffman, J. L.; Lutz, S. Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synth. Biol.* **2018**, *7* (11), 2514–2517. https://doi.org/10.1021/acssynbio.8b00295.

(152) Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Paša-Tolić, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N. Best Practices and Benchmarks for Intact Protein Analysis for Top-down Mass Spectrometry. *Nat. Methods* **2019**, *16* (7), 587–594. https://doi.org/10.1038/s41592-019-0457-0.

(153) Möller, I. R.; Slivacka, M.; Hausner, J.; Nielsen, A. K.; Pospíšilová, E.; Merkle, P. S.; Lišková, R.; Polák, M.; Loland, C. J.; Kádek, A.; Man, P.; Rand, K. D. Improving the Sequence Coverage of Integral Membrane Proteins during Hydrogen/Deuterium Exchange Mass Spectrometry Experiments. *Anal. Chem.* **2019**, *91* (17), 10970–10978. https://doi.org/10.1021/acs.analchem.9b00973.

(154) Nirudodhi, S. N.; Sperry, J. B.; Rouse, J. C.; Carroll, J. A. Application of Dual Protease Column for HDX-MS Analysis of Monoclonal Antibodies. *J. Pharm. Sci.* **2017**, *106* (2), 530–536. https://doi.org/10.1016/j.xphs.2016.10.023.

(155) Masson, G. R.; Maslen, S. L.; Williams, R. L. Analysis of Phosphoinositide 3-Kinase Inhibitors by Bottom-up Electron-Transfer

Dissociation Hydrogen/Deuterium Exchange Mass Spectrometry. *Biochem. J.* **2017**, *474* (11), 1867–1877. https://doi.org/10.1042/BCJ20170127.

(156) Jørgensen, T. J. D.; Gårdsvoll, H.; Ploug, M.; Roepstorff, P. Intramolecular Migration of Amide Hydrogens in Protonated Peptides upon Collisional Activation. **2005**. https://doi.org/10.1021/ja043789c.

(157) Abzalimov, R. R.; Kaltashov, I. A. Controlling Hydrogen Scrambling in Multiply Charged Protein Ions during Collisional Activation: Implications for Top-down Hydrogen/Deuterium Exchange Ms Utilizing Collisional Activation in the Gas Phase. *Anal. Chem.* **2010**, *82* (3), 942–950. https://doi.org/10.1021/ac9021874.

(158) Pan, J.; Zhang, S.; Borchers, C. H. Comparative Higher-Order Structure Analysis of Antibody Biosimilars Using Combined Bottom-up and Top-down Hydrogen-Deuterium Exchange Mass Spectrometry. *Biochim. Biophys. Acta - Proteins Proteomics* **2016**, *1864* (12), 1801–1808. https://doi.org/10.1016/j.bbapap.2016.08.013.

(159) Sobott, F.; Watt, S. J.; Smith, J.; Edelmann, M. J.; Kramer, H. B.; Kessler, B. M. Comparison of CID Versus ETD Based MS/MS Fragmentation for the Analysis of Protein Ubiquitination. *JAM* **2009**, *20*, 1652–1659. https://doi.org/10.1016/j.jasms.2009.04.023.

(160) Patrie, S. M. Top-down Mass Spectrometry: Proteomics to Proteoforms. In *Advances in Experimental Medicine and Biology*; Springer New York LLC, 2016; Vol. 919, pp 171–200. https://doi.org/10.1007/978-3-319-41448-5_8.

(161) Boone, C.; Adamec, J. Top-Down Proteomics. *Proteomic Profiling Anal. Chem. Crossroads Second Ed.* **2016**, 175–191. https://doi.org/10.1016/B978-0-444-63688-1.00010-0.

(162) Going, C. C.; Xia, Z.; Williams, E. R. Real-Time HD Exchange Kinetics of Proteins from Buffered Aqueous Solution with Electrothermal

Supercharging and Top-Down Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (6), 1019–1027. https://doi.org/10.1007/S13361-016-1350-Z.

(163) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons. *Science (80-.).* **2006**, *314* (5796), 109–112. https://doi.org/10.1126/SCIENCE.1128868.

(164) Illes-Toth, E.; Meisl, G.; Rempel, D. L.; Knowles, T. P. J.; Gross, M. L. Pulsed Hydrogen–Deuterium Exchange Reveals Altered Structures and Mechanisms in the Aggregation of Familial Alzheimer's Disease Mutants. *ACS Chem. Neurosci.* **2021**. https://doi.org/10.1021/ACSCHEMNEURO.1C00072.

(165) Pan, J.; Wilson, D. J.; Konermann, L. Pulsed Hydrogen Exchange and Electrospray Charge-State Distribution as Complementary Probes of Protein Structure in Kinetic Experiments: Implications for Ubiquitin Folding. *Biochemistry* **2005**, *44* (24), 8627–8633. https://doi.org/10.1021/bi050575e.

(166) Konermann, L.; Pan, J.; Liu, Y.-H. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40* (3), 1224–1234. https://doi.org/10.1039/C0CS00113A.

(167) Benhaim, M.; Lee, K. K.; Guttman, M. Tracking Higher Order Protein Structure by Hydrogen-Deuterium Exchange Mass Spectrometry. *Protein Pept. Lett.* **2019**, *26* (1), 16. https://doi.org/10.2174/0929866526666181212165037.

(168) Zhang, Y.; Rempel, D. L.; Zhang, J.; Sharma, A. K.; Mirica, L. M.; Gross, M. L. Pulsed Hydrogen–Deuterium Exchange Mass Spectrometry Probes Conformational Changes in Amyloid Beta (Aβ) Peptide Aggregation. *Proc. Natl. Acad. Sci.* **2013**, *110* (36), 14604–

14609. https://doi.org/10.1073/PNAS.1309175110.

(169) Du, Y.; Duc, N. M.; Rasmussen, S. G. F.; Hilger, D.; Kubiak, X.; Wang, L.; Bohon, J.; Kim, H. R.; Wegrecki, M.; Asuru, A.; Jeong, K. M.; Lee, J.; Chance, M. R.; Lodowski, D. T.; Kobilka, B. K.; Chung, K. Y. Assembly of a GPCR-G Protein Complex. *Cell* **2019**, *177* (5), 1232. https://doi.org/10.1016/J.CELL.2019.04.022.

(170) Sheff, J. G.; Rey, M.; Schriemer, D. C. Peptide-Column Interactions and Their Influence on Back Exchange Rates in Hydrogen/Deuterium Exchange-MS. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (7), 1006–1015. https://doi.org/10.1007/s13361-013-0639-4.

(171) Metwally, H.; McAllister, R. G.; Konermann, L. Exploring the Mechanism of Salt-Induced Signal Suppression in Protein Electrospray Mass Spectrometry Using Experiments and Molecular Dynamics Simulations. *Anal. Chem.* **2015**, *87* (4), 2434–2442. https://doi.org/10.1021/ac5044016.

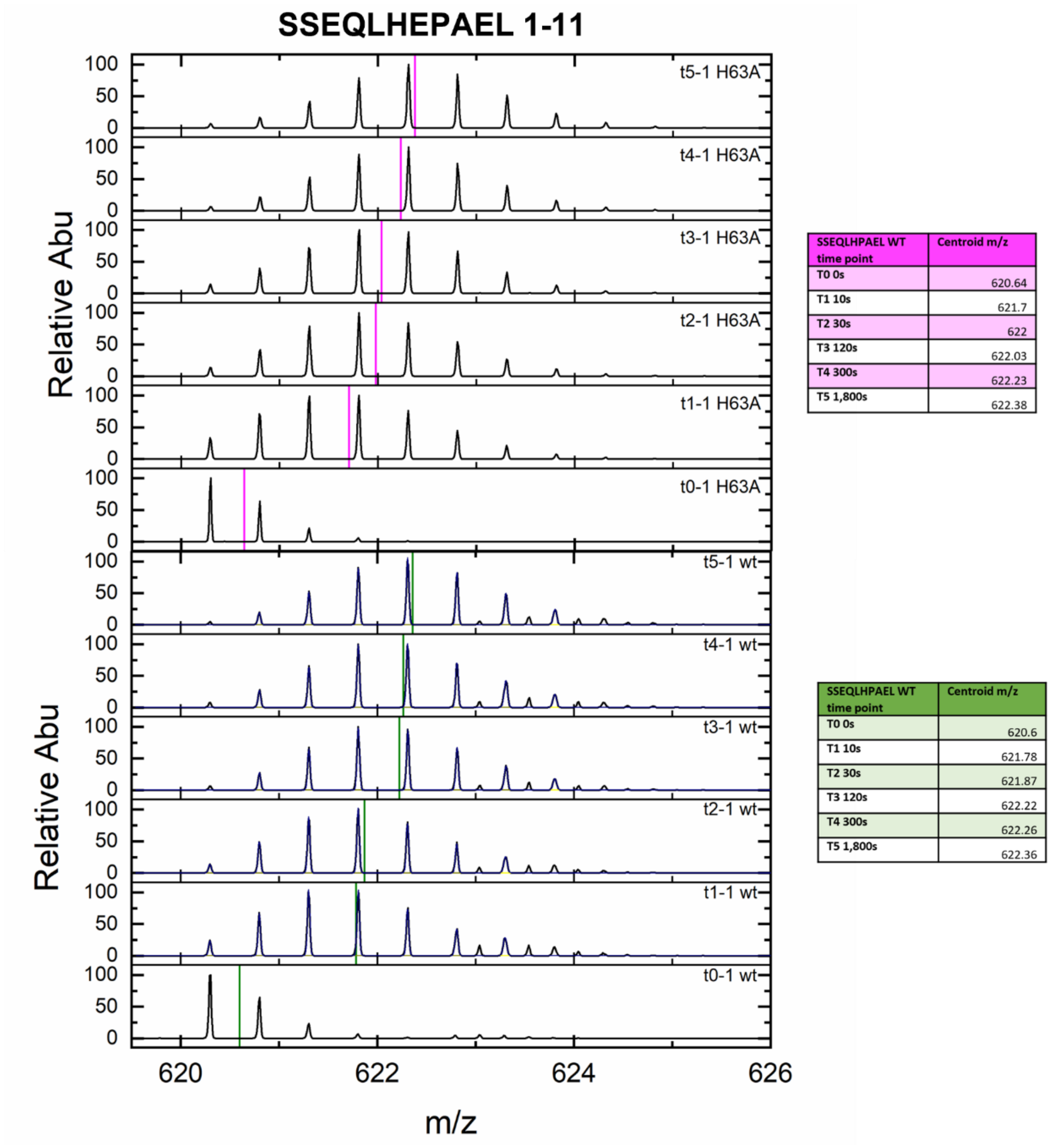
(172) Marshall, A. G.; Senko, M. W.; Li, W.; Li, M.; Dillon, S.; Guan, S. . L. T. M. Protein Molecular Mass to 1 Da By C-13, N-15 Double-Depletion and Ft-Icr Mass Spectrometry. *, T. M. J. Am. Chem. Soc., ,* **1997**, *119*, 433−434.

(173) Akashi, S.; Takio, K.; Matsui, H.; Tate, S.; Kainosho, M. Collision-Induced Dissociation Spectra Obtained by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Using a 13C,15N-Doubly Depleted Protein. *J. Am. Soc. Mass Spectrom* **1998**, *70* (15), 3333–3336.

(174) Gallagher, K. J.; Palasser, M.; Hughes, S.; Mackay, C. L.; Kilgour, D. P. A.; Clarke, D. J. Isotope Depletion Mass Spectrometry (ID-MS) for Accurate Mass Determination and Improved Top-Down Sequence Coverage of Intact Proteins. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (3), 700–710. https://doi.org/10.1021/JASMS.9B00119.

(175) Kasper D. Rand†, ‡ and; Thomas J. D. Jørgensen*, ‡. Development of a Peptide Probe for the Occurrence of Hydrogen (1H/2H) Scrambling upon Gas-Phase Fragmentation. *Anal. Chem.* **2007**, *79* (22), 8686–8693. https://doi.org/10.1021/AC0710782.

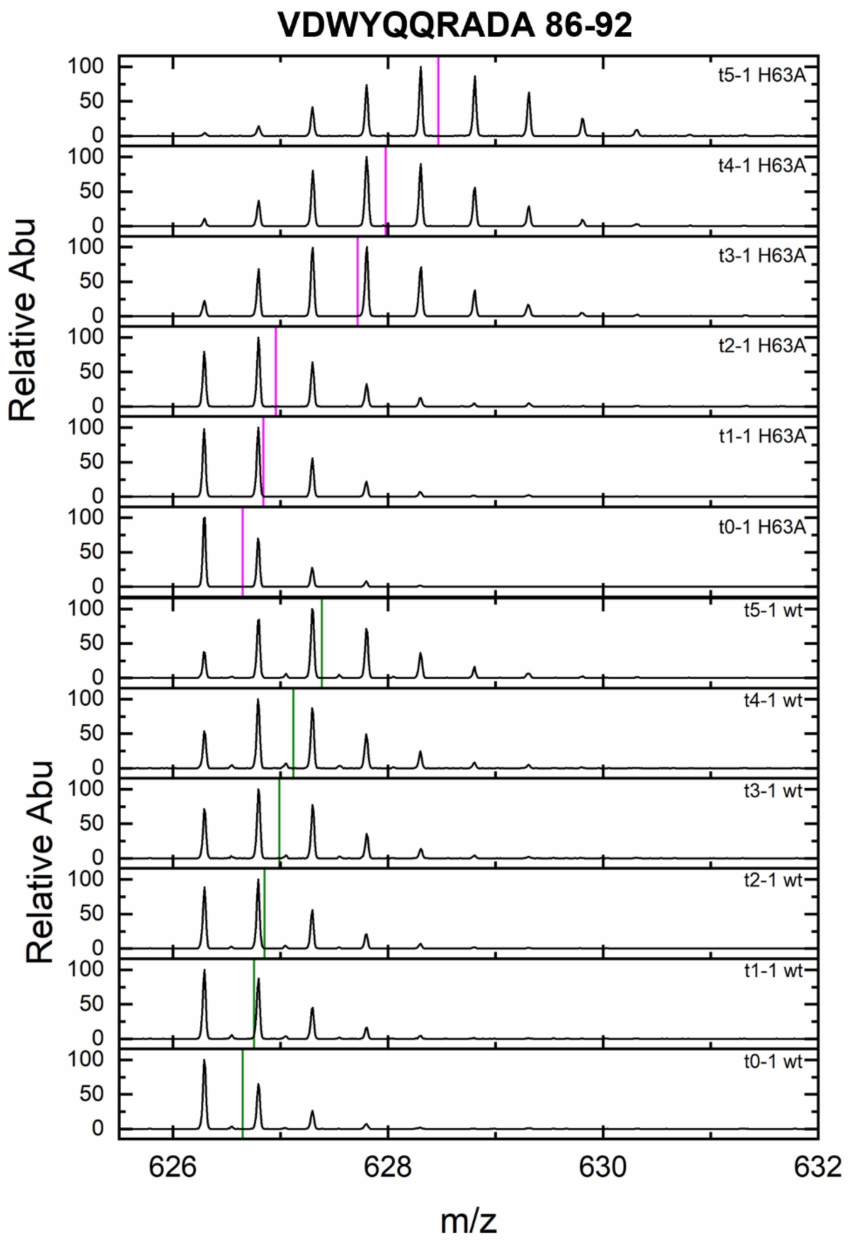
Appendix i: Supplementary Figures



SSEQLHPAEL WT time point	Centroid m/z
T0 0s	620.64
T1 10s	621.7
T2 30s	622
T3 120s	622.03
T4 300s	622.23
T5 1,800s	622.38

SSEQLHPAEL WT time point	Centroid m/z
T0 0s	620.6
T1 10s	621.78
T2 30s	621.87
T3 120s	622.22
T4 300s	622.26
T5 1,800s	622.36

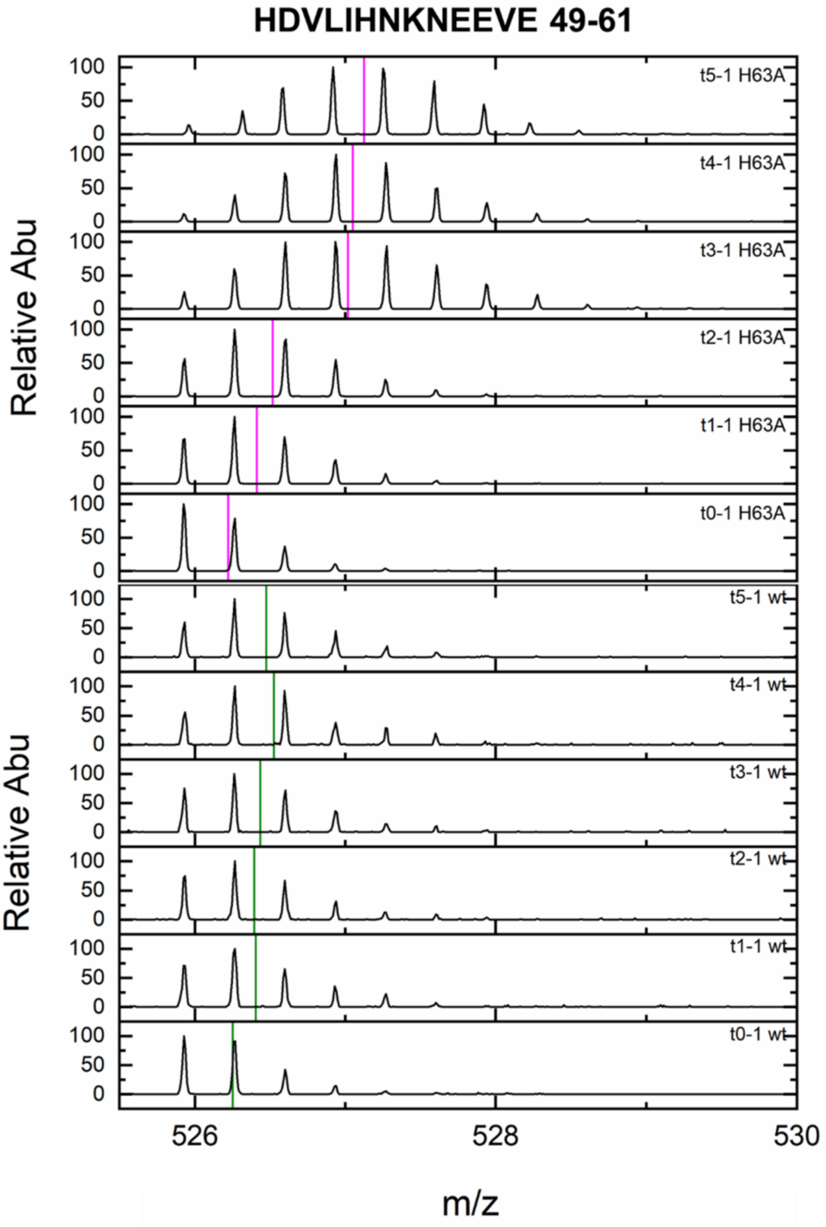
SI Figure 4.1: H63A and WT timecourse spectra for the SSEQLHEPAE L 2+ peptide. The coloured reference lines highlight the centroid m/z for each isotopic distribution and the values are show in the adjacent table for each H63A and WT timepoint respectively



VDWYQQRADA H63A time point	Centroid m/z
T0 0s	626.65
T1 10s	626.84
T2 30s	626.96
T3 120s	627.72
T4 300s	627.98
T5 1,800s	628.47

VDWYQQRADA WT time point	Centroid m/z
T0 0s	626.65
T1 10s	626.75
T2 30s	626.85
T3 120s	626.99
T4 300s	627.11
T5 1,800s	627.36

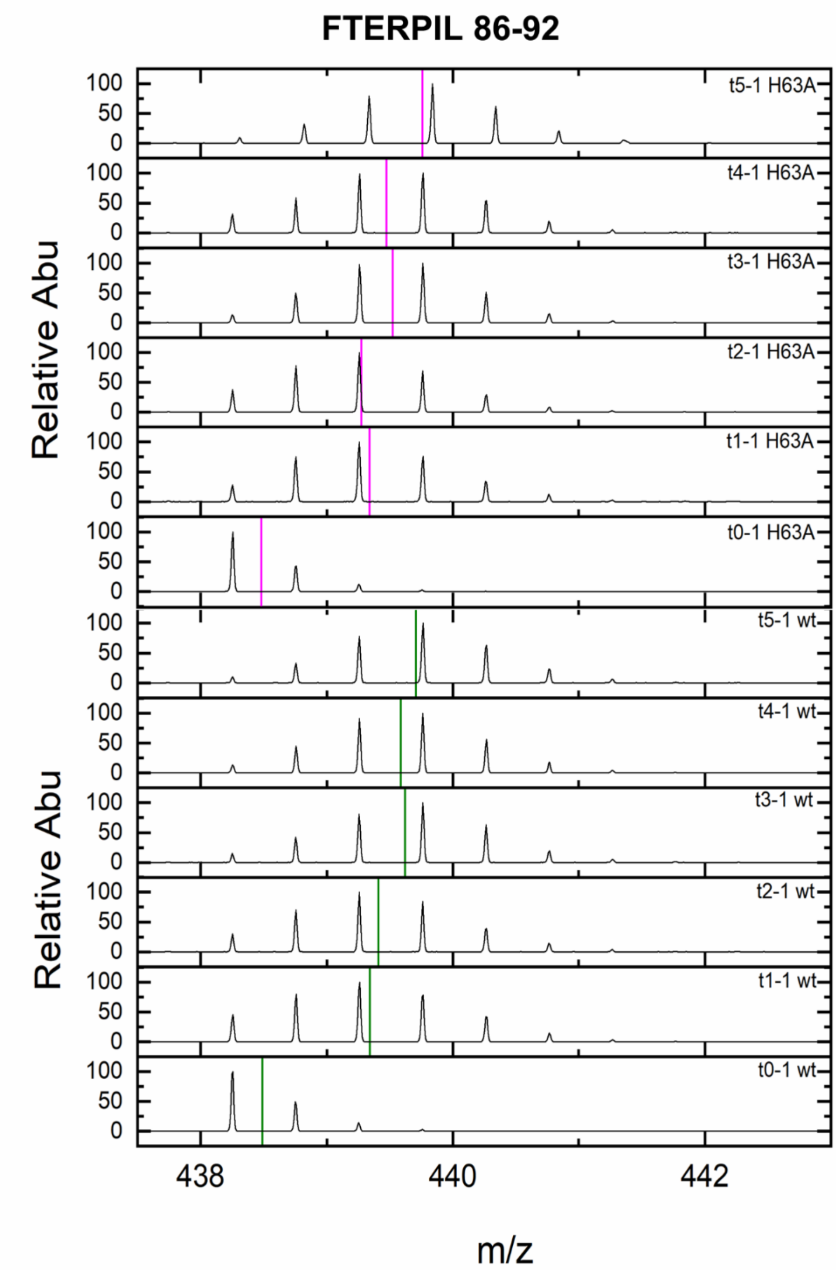
SI Figure 4.2: H63A and WT timecourse spectra for the VDWYQQRADA 2+ peptide. The coloured reference lines highlight the centroid m/z for each isotopic distribution and the values are show in the adjacent table for each H63A and WT timepoint respectively

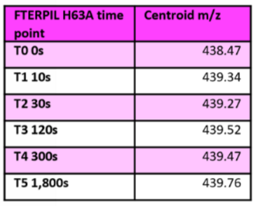


HDVLIHNKNEEEVE H63A time point	Centroid m/z
T0 0s	526.22
T1 10s	526.41
T2 30s	526.51
T3 120s	527.01
T4 300s	527.05
T5 1,800s	527.12

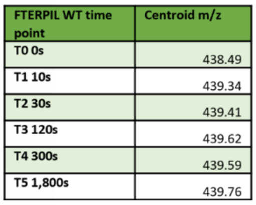
HDVLIHNKNEEEVE WT time point	Centroid m/z
T0 0s	526.25
T1 10s	526.4
T2 30s	526.39
T3 120s	526.44
T4 300s	526.52
T5 1,800s	526.47

SI Figure 4.3: H63A and WT timecourse spectra for the HDVLIHNKNEEVE 3+ peptide. The coloured reference lines highlight the centroid m/z for each isotopic distribution and the values are show in the adjacent table for each H63A and WT timepoint respectively



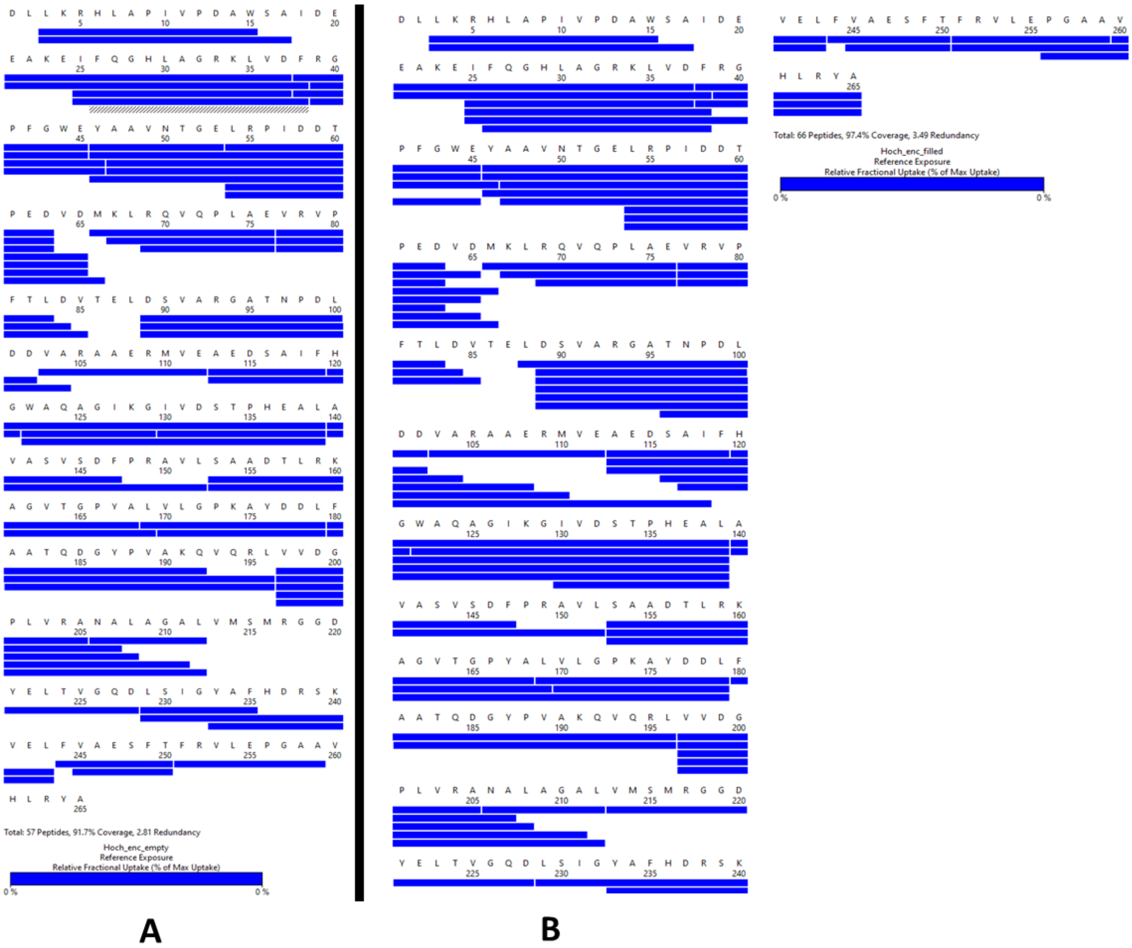


FTERPIL H63A time point	Centroid m/z
T0 0s	438.47
T1 10s	439.34
T2 30s	439.27
T3 120s	439.52
T4 300s	439.47
T5 1,800s	439.76

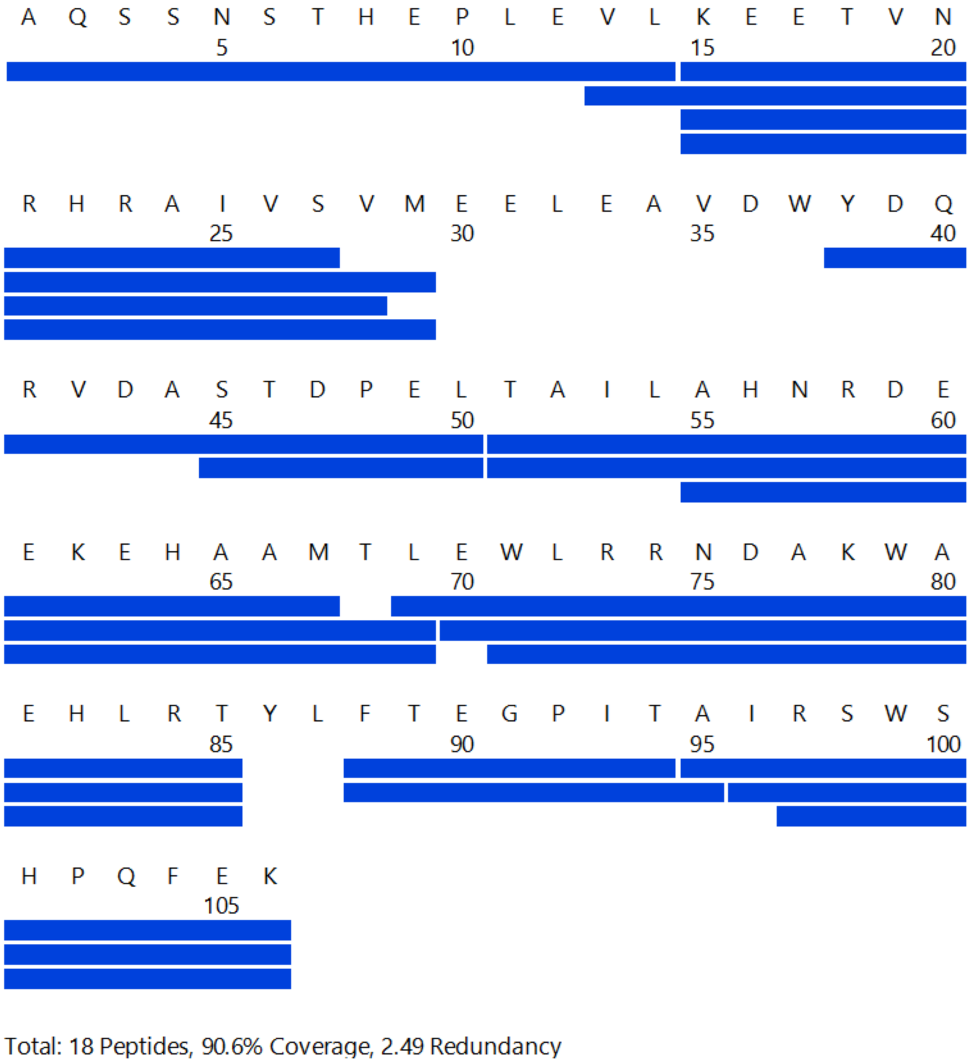


FTERPIL WT time point	Centroid m/z
T0 0s	438.49
T1 10s	439.34
T2 30s	439.41
T3 120s	439.62
T4 300s	439.59
T5 1,800s	439.76

SI Figure 4.4: H63A and WT timecourse spectra for the FTERPIL peptide. The coloured reference lines highlight the centroid m/z for each isotopic distribution and the values are show in the adjacent table for each H63A and WT timepoint respectively



SI Figure 5.1. HDX Coverage maps of the peptides found within Empty and Loaded encapsulin samples. (A) Coverage map of empty Enc peptides resulting in 91.7% protein coverage from 57 peptides. (B) Coverage map of loaded Enc peptides resulting in 97.4% protein coverage from 66 peptides



SI Figure 5.2: HDX Coverage map of the peptides found within the Rru-EncFtn digests. Peptides observed by MS in all HDX timepoints are represented by blue bars. Pepsin digestion resulted in 18 peptides providing 90.6% protein sequence coverage with 2.49% redundancy

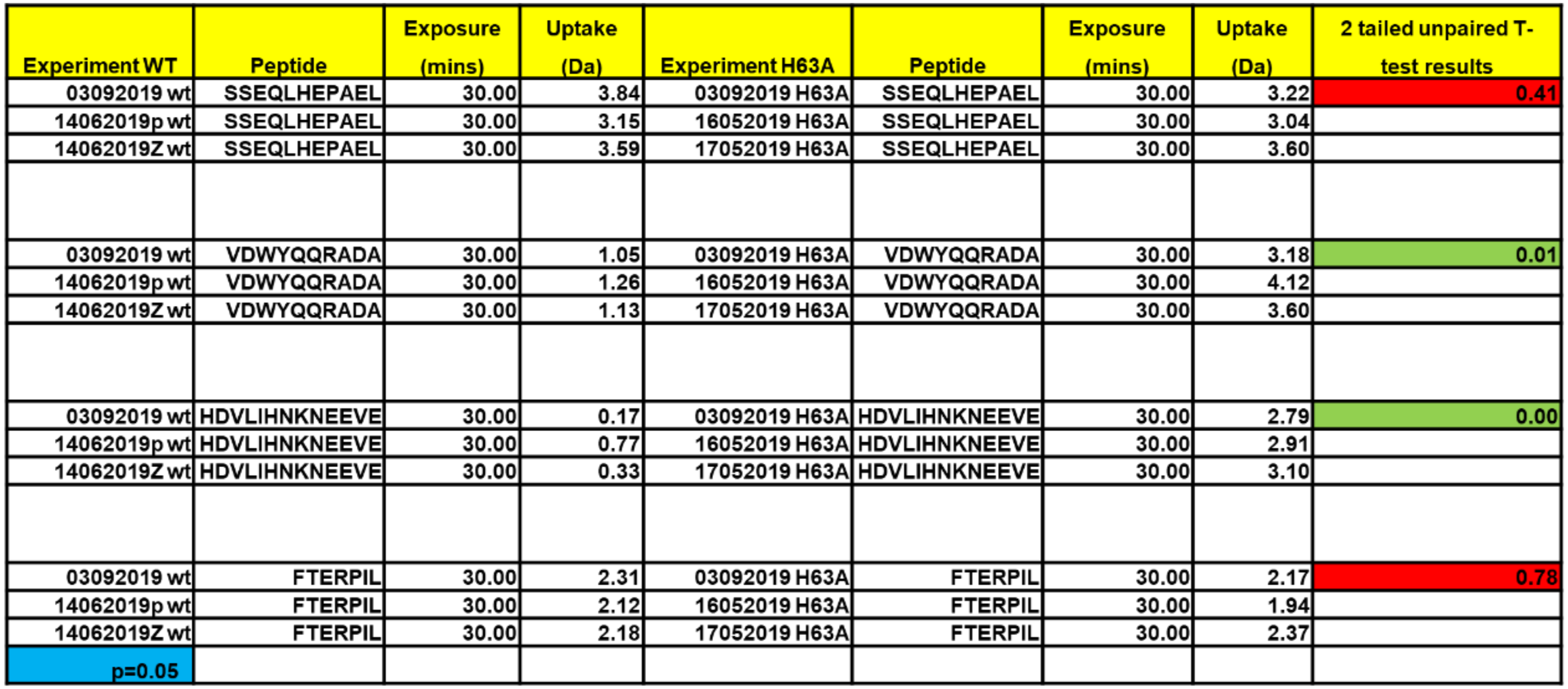
Appendix ii: Supplementary Tables

				EncFtn-WT Deuterium uptake (Da)									
Sequence	**Start (aa)**	**End (aa)**	**Num Exchangers**	**(10s)**	**(10s) SD**	**(30s)**	**(30s) SD**	**(120s)**	**(120s) SD**	**(500s)**	**(500s) SD**	**(1800s)**	**(1800s) SD**
SSEQLHEPAEL	1	11	9	2.41	0.08	2.60	0.09	2.88	0.13	3.15	0.07	3.38	0.29
LSEETKNMHRAL	12	23	11	0.88	0.09	0.97	0.08	1.23	0.09	1.43	0.07	1.67	0.31
LSEETKNMHRALVT	12	25	13	0.92	0.04	0.97	0.05	1.20	0.06	1.46	0.06	1.64	0.21
LSEETKNMHRALVTL	12	26	14	0.97	0.06	1.08	0.03	1.32	0.05	1.64	0.04	1.78	0.21
VDWYQQRADA	33	42	9	0.29	0.06	0.36	0.06	0.66	0.08	1.01	0.07	1.26	0.19
YQQRADA	36	42	6	0.26	0.06	0.36	0.04	0.64	0.05	0.95	0.03	1.19	0.19
YQQRADACSEPGL	36	48	11	1.56	0.07	1.86	0.05	2.37	0.11	2.77	0.04	3.07	0.36
CSEPGL	43	48	4	0.83	0.07	0.96	0.05	1.13	0.04	1.22	0.04	1.41	0.10
HDVLIHNKNEEVE	49	61	12	0.63	0.03	0.33	0.09	0.34	0.01	0.60	0.08	0.98	0.09
HDVLIHNKNEEVEHAMM	49	65	16	0.37	0.12	0.43	0.12	0.58	0.12	0.76	0.12	0.81	0.18
VLIHNKNEEVEHAMM	51	65	14	0.20	0.05	0.24	0.02	0.31	0.03	0.41	0.01	0.42	0.09
IHNKNEEVEHAMM	53	65	12	0.20	0.04	0.23	0.02	0.28	0.03	0.35	0.03	0.42	0.05
IHNKNEEVEHAMMTL	53	67	14	0.19	0.06	0.20	0.06	0.30	0.05	0.35	0.07	0.39	0.09
VEHAMM	60	65	5	0.00	0.03	0.02	0.03	0.03	0.03	0.02	0.03	0.02	0.04
LEWIRRRSPVF	67	77	9	0.52	0.06	0.58	0.06	0.63	0.05	0.67	0.06	0.75	0.09
WIRRRSPVF	69	77	7	0.46	0.02	0.50	0.02	0.56	0.03	0.62	0.02	0.73	0.08
WIRRRSPVFDA	69	79	9	0.41	0.06	0.50	0.06	0.62	0.16	0.57	0.04	0.62	0.08
DAHMRTYL	78	85	7	0.49	0.06	0.48	0.10	0.57	0.05	0.69	0.08	0.87	0.17
FTERPIL	86	92	5	1.34	0.07	1.48	0.05	1.72	0.07	1.88	0.01	2.12	0.15
FTERPILEL	86	94	7	2.02	0.10	2.27	0.07	2.62	0.10	2.92	0.07	3.18	0.27
TERPILEL	87	94	6	1.77	0.10	2.02	0.07	2.36	0.10	2.67	0.08	2.88	0.24
LEERSWSHPQFEK	94	106	11	2.27	0.10	2.55	0.08	2.90	0.14	3.16	0.03	3.43	0.28
EERSWSHPQFEK	95	106	10	1.98	0.11	2.26	0.08	2.56	0.14	2.79	0.05	3.00	0.23
RSWSHPQFEK	97	106	8	1.64	0.10	1.85	0.03	2.16	0.14	2.28	0.05	2.49	0.16
SHPQFEK	100	106	5	1.17	0.11	1.41	0.14	1.54	0.12	1.71	0.09	2.00	0.19

SI Table 4.1: The Deuterium uptake and standard deviation of the triplicate data for all labelling timepoints for EncFtn-WT. The peptide's sequence, start and end points are present in the table; and the four peptides referred to in the main text have been highlighted yellow

				EncFtn-H63A Deuterium uptake (Da)									
Sequence	Start (aa)	End (aa)	Num Exchangers	(10s)	(10s) SD	(30s)	(30s) SD	(120s)	(120s) SD	(500s)	(500s) SD	(1800s)	(1800s) SD
SSEQLHEPAEL	1	11	9	2.04	0.36	2.60	0.07	2.76	0.08	3.30	0.14	3.60	0.07
LSEETKNMHRA	12	22	10	1.10	0.26	1.47	0.04	2.03	0.04	2.55	0.07	3.14	0.03
LSEETKNMHRALVT	12	25	13	1.22	0.28	1.65	0.05	2.29	0.04	2.80	0.11	3.33	0.05
LSEETKNMHRALVTL	12	26	14	1.22	0.28	1.75	0.02	2.41	0.03	2.90	0.09	3.48	0.06
SEETKNMHRALVTL	13	26	13	1.12	0.25	1.54	0.08	2.24	0.03	2.69	0.17	3.14	0.20
ETKNMHRALVTL	15	26	11	0.45	0.10	0.57	0.04	0.89	0.06	1.27	0.07	1.55	0.04
LIEELE	26	31	5	0.05	0.07	0.19	0.09	0.22	0.04	0.28	0.04	0.29	0.02
IEELEA	27	32	5	0.12	0.01	0.12	0.01	0.13	0.01	0.13	0.01	0.13	0.01
LEAVDW	30	35	5	0.05	0.03	0.09	0.01	0.14	0.01	0.20	0.04	0.29	0.03
VDWYQQRADA	33	42	9	0.33	0.08	0.80	0.06	2.25	0.13	2.88	0.17	3.60	0.06
VDWYQQRADACSEPGLHD	33	50	16	2.68	0.34	3.33	0.16	3.53	0.16	4.07	0.22	4.33	0.17
YQQRADA	36	42	6	0.31	0.07	0.75	0.07	1.93	0.10	2.37	0.06	2.72	0.04
YQQRADACSEPGL	36	48	11	1.52	0.28	2.37	0.07	3.47	0.10	4.18	0.16	4.77	0.07
YQQRADACSEPGLHD	36	50	13	1.61	0.36	2.16	0.59	3.68	0.08	4.49	0.23	5.24	0.04
YQQRADACSEPGLHDVL	36	52	15	1.29	0.36	2.23	0.08	3.87	0.12	4.50	0.19	5.18	0.09
CSEPGL	43	48	4	0.73	0.16	1.05	0.04	1.14	0.02	1.36	0.08	1.47	0.02
CSEPGLHD	43	50	6	0.98	0.25	1.36	0.09	1.48	0.03	1.85	0.10	2.08	0.12
CSEPGLHDVL	43	52	8	0.78	0.19	1.20	0.08	1.81	0.04	2.34	0.09	2.79	0.10
HDVLIHNKNEEVE	49	61	12	0.46	0.04	0.89	0.08	2.27	0.10	2.52	0.16	3.10	0.17
HDVLIHNKNEEVEA	49	62	13	0.35	0.16	0.81	0.18	1.98	0.18	2.43	0.15	2.93	0.15
VLIHNKNEEVE	51	61	10	0.41	0.28	0.77	0.24	1.41	0.25	1.63	0.25	1.89	0.24
IHNKNEEVE	53	61	8	0.31	0.05	0.59	0.05	1.02	0.02	1.21	0.04	1.43	0.03
IHNKNEEVEA	53	62	9	0.24	0.06	0.55	0.09	0.91	0.02	1.08	0.08	1.28	0.11
VEAAMMTL	60	67	7	0.12	0.02	0.38	0.02	0.76	0.08	0.81	0.02	0.95	0.04
TLEWIRRRSPVF	66	77	10	0.35	0.09	0.54	0.09	0.58	0.06	0.64	0.07	0.97	0.14
LEWIRRRSPVF	67	77	9	0.50	0.12	0.66	0.02	0.72	0.03	0.81	0.02	1.05	0.03
EWIRRRSPVF	68	77	8	0.48	0.11	0.66	0.03	0.73	0.03	0.83	0.06	1.06	0.04
WIRRRSPVF	69	77	7	0.52	0.11	0.60	0.03	0.72	0.03	0.83	0.04	1.09	0.03
DAHMRTYL	78	85	7	1.04	0.03	1.10	0.14	1.42	0.03	1.71	0.08	1.91	0.02
TYLFTERPI	83	91	7	1.54	0.43	2.44	0.10	2.77	0.08	3.25	0.13	3.67	0.15
YLFTERPIL	84	92	7	1.22	0.27	1.80	0.14	2.19	0.02	2.39	0.05	3.06	0.07
FTERPIL	86	92	5	1.46	0.02	1.68	0.04	1.89	0.02	2.23	0.11	2.37	0.08
FTERPILEL	86	94	7	1.76	0.50	2.50	0.03	2.81	0.03	3.30	0.14	3.61	0.06
TERPILEL	87	94	6	1.72	0.01	2.06	0.02	2.25	0.07	2.94	0.12	3.19	0.09
ELEERSWSHPQFEK	93	106	12	2.35	0.31	2.97	0.03	3.24	0.06	3.78	0.15	4.03	0.06
LEERSWSHPQFEK	94	106	11	2.26	0.36	2.60	0.25	3.02	0.18	3.50	0.23	3.82	0.18
EERSWSHPQFEK	95	106	10	1.57	0.87	2.24	0.44	2.66	0.41	3.17	0.44	3.45	0.41
RSWSHPQFEK	97	106	8	1.52	0.20	1.93	0.04	2.10	0.03	2.51	0.07	3.00	0.14

SI Table 4.2: The Deuterium uptake and standard deviation of the triplicate data for all labelling timepoints for EncFtn-H63A. The peptide's sequence, start and end points are present in the table; and the four peptides referred to in the main text have been highlighted yellow



Experiment WT	Peptide	Exposure (mins)	Uptake (Da)	Experiment H63A	Peptide	Exposure (mins)	Uptake (Da)	2 tailed unpaired T-test results
03092019 wt	SSEQLHEPAEL	30.00	3.84	03092019 H63A	SSEQLHEPAEL	30.00	3.22	0.41
14062019p wt	SSEQLHEPAEL	30.00	3.15	16052019 H63A	SSEQLHEPAEL	30.00	3.04	
14062019Z wt	SSEQLHEPAEL	30.00	3.59	17052019 H63A	SSEQLHEPAEL	30.00	3.60	
03092019 wt	VDWYQQRADA	30.00	1.05	03092019 H63A	VDWYQQRADA	30.00	3.18	0.01
14062019p wt	VDWYQQRADA	30.00	1.26	16052019 H63A	VDWYQQRADA	30.00	4.12	
14062019Z wt	VDWYQQRADA	30.00	1.13	17052019 H63A	VDWYQQRADA	30.00	3.60	
03092019 wt	HDVLIHNKNEEVE	30.00	0.17	03092019 H63A	HDVLIHNKNEEVE	30.00	2.79	0.00
14062019p wt	HDVLIHNKNEEVE	30.00	0.77	16052019 H63A	HDVLIHNKNEEVE	30.00	2.91	
14062019Z wt	HDVLIHNKNEEVE	30.00	0.33	17052019 H63A	HDVLIHNKNEEVE	30.00	3.10	
03092019 wt	FTERPIL	30.00	2.31	03092019 H63A	FTERPIL	30.00	2.17	0.78
14062019p wt	FTERPIL	30.00	2.12	16052019 H63A	FTERPIL	30.00	1.94	
14062019Z wt	FTERPIL	30.00	2.18	17052019 H63A	FTERPIL	30.00	2.37	
p=0.05								

SI Table 4.3: Twin tailed unpaired student T-test results comparing the WT-EncFtn and H63A-EncFtn biological repeats of the highlighted peptides after 30 minutes of deuterium labelling. The null hypothesis was that the sample were not significantly different

<table>
<tr><th></th><th>Structure 1</th><th>HSDC</th><th>ASA</th><th>BSA</th><th>ΔiG</th></tr>
<tr><td>1</td><td>D:GLN 6</td><td></td><td>165.60</td><td>5.14 |</td><td>-0.07</td></tr>
<tr><td>2</td><td>D:LEU 7</td><td></td><td>73.55</td><td>7.60 ||</td><td>0.12</td></tr>
<tr><td>3</td><td>D:HIS 8</td><td>H</td><td>178.02</td><td>90.42 ||||||</td><td>0.09</td></tr>
<tr><td>4</td><td>D:GLU 9</td><td></td><td>94.68</td><td>50.77 ||||||</td><td>0.02</td></tr>
<tr><td>5</td><td>D:PRO 10</td><td></td><td>63.38</td><td>0.00</td><td>0.00</td></tr>
<tr><td>6</td><td>D:ALA 11</td><td></td><td>50.51</td><td>0.00</td><td>0.00</td></tr>
<tr><td>7</td><td>D:GLU 12</td><td></td><td>138.48</td><td>0.00</td><td>0.00</td></tr>
<tr><td>8</td><td>D:LEU 13</td><td></td><td>125.33</td><td>38.21 ||||</td><td>0.12</td></tr>
<tr><td>9</td><td>D:LEU 14</td><td></td><td>39.45</td><td>30.23 ||||||||</td><td>0.48</td></tr>
<tr><td>10</td><td>D:SER 15</td><td>H</td><td>68.76</td><td>8.54 ||</td><td>0.02</td></tr>
<tr><td>11</td><td>D:GLU 16</td><td></td><td>110.07</td><td>0.00</td><td>0.00</td></tr>
<tr><td>12</td><td>D:GLU 17</td><td></td><td>74.36</td><td>0.00</td><td>0.00</td></tr>
<tr><td>13</td><td>D:THR 18</td><td></td><td>64.45</td><td>58.65 ||||||||||</td><td>0.31</td></tr>
<tr><td>14</td><td>D:LYS 19</td><td></td><td>29.12</td><td>0.00</td><td>0.00</td></tr>
<tr><td>15</td><td>D:ASN 20</td><td></td><td>54.24</td><td>0.00</td><td>0.00</td></tr>
<tr><td>16</td><td>D:MET 21</td><td></td><td>31.25</td><td>7.72 |||</td><td>0.12</td></tr>
<tr><td>17</td><td>D:HIS 22</td><td>S</td><td>43.23</td><td>29.38 |||||||</td><td>-0.85</td></tr>
<tr><td>18</td><td>D:ARG 23</td><td></td><td>130.96</td><td>0.00</td><td>0.00</td></tr>
<tr><td>19</td><td>D:ALA 24</td><td></td><td>11.79</td><td>0.00</td><td>0.00</td></tr>
<tr><td>20</td><td>D:LEU 25</td><td></td><td>63.97</td><td>63.13 ||||||||||</td><td>1.00</td></tr>
<tr><td>21</td><td>D:VAL 26</td><td></td><td>37.05</td><td>4.66 ||</td><td>0.07</td></tr>
<tr><td>22</td><td>D:THR 27</td><td></td><td>51.84</td><td>0.00</td><td>0.00</td></tr>
<tr><td>23</td><td>D:LEU 28</td><td></td><td>20.15</td><td>15.96 ||||||||</td><td>0.26</td></tr>
<tr><td>24</td><td>D:ILE 29</td><td></td><td>93.56</td><td>80.51 |||||||||</td><td>1.28</td></tr>
<tr><td>25</td><td>D:GLU 30</td><td></td><td>102.76</td><td>0.00</td><td>0.00</td></tr>
<tr><td>26</td><td>D:GLU 31</td><td></td><td>23.44</td><td>0.00</td><td>0.00</td></tr>
<tr><td>27</td><td>D:LEU 32</td><td></td><td>85.48</td><td>85.31 ||||||||||</td><td>1.22</td></tr>
<tr><td>28</td><td>D:GLU 33</td><td></td><td>71.16</td><td>22.83 ||||</td><td>0.12</td></tr>
<tr><td>29</td><td>D:ALA 34</td><td></td><td>16.05</td><td>0.00</td><td>0.00</td></tr>
</table>

<table>
<tr><td>30</td><td>D:VAL 35</td><td></td><td>30.54</td><td>28.36 ||||||||||</td><td>0.45</td></tr>
<tr><td>31</td><td>D:ASP 36</td><td>S</td><td>71.79</td><td>66.77 ||||||||||</td><td>-0.20</td></tr>
<tr><td>32</td><td>D:TRP 37</td><td></td><td>125.34</td><td>45.40 ||||</td><td>0.32</td></tr>
<tr><td>33</td><td>D:TYR 38</td><td></td><td>48.85</td><td>0.00</td><td>0.00</td></tr>
<tr><td>34</td><td>D:GLN 39</td><td></td><td>57.67</td><td>26.67 |||||</td><td>0.34</td></tr>
<tr><td>35</td><td>D:GLN 40</td><td>H</td><td>104.80</td><td>104.80 ||||||||||</td><td>-0.22</td></tr>
<tr><td>36</td><td>D:ARG 41</td><td></td><td>128.42</td><td>38.13 |||</td><td>-0.33</td></tr>
<tr><td>37</td><td>D:ALA 42</td><td></td><td>1.72</td><td>0.00</td><td>0.00</td></tr>
<tr><td>38</td><td>D:ASP 43</td><td>H</td><td>107.85</td><td>60.71 ||||||</td><td>-0.32</td></tr>
<tr><td>39</td><td>D:ALA 44</td><td></td><td>87.42</td><td>56.50 |||||||</td><td>0.74</td></tr>
<tr><td>40</td><td>D:CYS 45</td><td></td><td>34.42</td><td>0.00</td><td>0.00</td></tr>
<tr><td>41</td><td>D:SER 46</td><td></td><td>118.25</td><td>0.00</td><td>0.00</td></tr>
<tr><td>42</td><td>D:GLU 47</td><td></td><td>86.39</td><td>0.00</td><td>0.00</td></tr>
<tr><td>43</td><td>D:PRO 48</td><td></td><td>97.40</td><td>0.00</td><td>0.00</td></tr>
<tr><td>44</td><td>D:GLY 49</td><td></td><td>38.51</td><td>0.00</td><td>0.00</td></tr>
<tr><td>45</td><td>D:LEU 50</td><td></td><td>89.30</td><td>0.00</td><td>0.00</td></tr>
<tr><td>46</td><td>D:HIS 51</td><td></td><td>68.83</td><td>0.00</td><td>0.00</td></tr>
<tr><td>47</td><td>D:ASP 52</td><td></td><td>86.70</td><td>0.00</td><td>0.00</td></tr>
<tr><td>48</td><td>D:VAL 53</td><td></td><td>78.50</td><td>0.00</td><td>0.00</td></tr>
<tr><td>49</td><td>D:LEU 54</td><td></td><td>20.08</td><td>0.00</td><td>0.00</td></tr>
<tr><td>50</td><td>D:ILE 55</td><td></td><td>46.19</td><td>0.00</td><td>0.00</td></tr>
<tr><td>51</td><td>D:HIS 56</td><td></td><td>104.18</td><td>0.00</td><td>0.00</td></tr>
<tr><td>52</td><td>D:ASN 57</td><td></td><td>54.16</td><td>0.00</td><td>0.00</td></tr>
<tr><td>53</td><td>D:LYS 58</td><td></td><td>41.44</td><td>4.26 ||</td><td>0.05</td></tr>
<tr><td>54</td><td>D:ASN 59</td><td></td><td>85.12</td><td>43.45 ||||||</td><td>-0.12</td></tr>
<tr><td>55</td><td>D:GLU 60</td><td></td><td>75.71</td><td>0.00</td><td>0.00</td></tr>
<tr><td>56</td><td>D:GLU 61</td><td></td><td>34.38</td><td>0.00</td><td>0.00</td></tr>
<tr><td>57</td><td>D:VAL 62</td><td></td><td>59.53</td><td>59.20 ||||||||||</td><td>0.95</td></tr>
<tr><td>58</td><td>D:GLU 63</td><td>HS</td><td>106.55</td><td>67.54 |||||||</td><td>-0.04</td></tr>
<tr><td>59</td><td>D:HIS 64</td><td></td><td>78.38</td><td>0.00</td><td>0.00</td></tr>
</table>

<table>
<tr><td>60</td><td>D:ALA 65</td><td></td><td>0.34</td><td>0.34 ||||||||||</td><td>0.01</td></tr>
<tr><td>61</td><td>D:MET 66</td><td></td><td>95.60</td><td>94.26 ||||||||||</td><td>2.17</td></tr>
<tr><td>62</td><td>D:MET 67</td><td></td><td>122.01</td><td>72.44 ||||||</td><td>1.09</td></tr>
<tr><td>63</td><td>D:THR 68</td><td></td><td>20.77</td><td>0.00</td><td>0.00</td></tr>
<tr><td>64</td><td>D:LEU 69</td><td></td><td>33.47</td><td>31.80 ||||||||||</td><td>0.51</td></tr>
<tr><td>65</td><td>D:GLU 70</td><td>H</td><td>55.33</td><td>55.33 |||||||||||</td><td>0.54</td></tr>
<tr><td>66</td><td>D:TRP 71</td><td></td><td>93.62</td><td>19.44 |||</td><td>0.29</td></tr>
<tr><td>67</td><td>D:ILE 72</td><td></td><td>0.00</td><td>0.00</td><td>0.00</td></tr>
<tr><td>68</td><td>D:ARG 73</td><td></td><td>67.08</td><td>42.76 |||||||</td><td>-0.18</td></tr>
<tr><td>69</td><td>D:ARG 74</td><td>H</td><td>179.33</td><td>93.74 ||||||</td><td>-1.68</td></tr>
<tr><td>70</td><td>D:ARG 75</td><td></td><td>119.63</td><td>0.00</td><td>0.00</td></tr>
<tr><td>71</td><td>D:SER 76</td><td></td><td>13.67</td><td>0.00</td><td>0.00</td></tr>
<tr><td>72</td><td>D:PRO 77</td><td></td><td>112.74</td><td>0.00</td><td>0.00</td></tr>
<tr><td>73</td><td>D:VAL 78</td><td></td><td>90.75</td><td>55.05 |||||||</td><td>0.88</td></tr>
<tr><td>74</td><td>D:PHE 79</td><td></td><td>69.27</td><td>65.83 ||||||||||</td><td>1.05</td></tr>
<tr><td>75</td><td>D:ASP 80</td><td></td><td>30.43</td><td>0.00</td><td>0.00</td></tr>
<tr><td>76</td><td>D:ALA 81</td><td></td><td>58.53</td><td>0.00</td><td>0.00</td></tr>
<tr><td>77</td><td>D:HIS 82</td><td></td><td>107.57</td><td>70.80 |||||||</td><td>0.29</td></tr>
<tr><td>78</td><td>D:MET 83</td><td></td><td>27.24</td><td>27.24 |||||||||||</td><td>0.59</td></tr>
<tr><td>79</td><td>D:ARG 84</td><td></td><td>172.37</td><td>26.12 ||</td><td>-0.32</td></tr>
<tr><td>80</td><td>D:THR 85</td><td></td><td>52.06</td><td>0.00</td><td>0.00</td></tr>
<tr><td>81</td><td>D:TYR 86</td><td>H</td><td>84.85</td><td>80.64 ||||||||||</td><td>0.38</td></tr>
<tr><td>82</td><td>D:LEU 87</td><td></td><td>84.67</td><td>77.68 ||||||||||</td><td>1.24</td></tr>
<tr><td>83</td><td>D:PHE 88</td><td>H</td><td>203.57</td><td>141.85 |||||||</td><td>2.02</td></tr>
<tr><td>84</td><td>D:THR 89</td><td>H</td><td>54.49</td><td>28.37 ||||||</td><td>-0.31</td></tr>
<tr><td>85</td><td>D:GLU 90</td><td></td><td>152.98</td><td>11.22 |</td><td>0.02</td></tr>
<tr><td>86</td><td>D:ARG 91</td><td>H</td><td>133.00</td><td>6.44 |</td><td>-0.06</td></tr>
<tr><td>87</td><td>D:PRO 92</td><td></td><td>84.16</td><td>4.19 |</td><td>0.07</td></tr>
<tr><td>88</td><td>D:ILE 93</td><td></td><td>128.42</td><td>100.08 ||||||||</td><td>1.60</td></tr>
<tr><td>89</td><td>D:LEU 94</td><td></td><td>163.05</td><td>23.58 ||</td><td>0.38</td></tr>
</table>

90	D:GLU 95	99.89	0.00	0.00
91	D:LEU 96	23.70	0.34 |	0.01
92	D:GLU 97	169.42	8.01 |	0.02

SI Table 4.4: The PISA analysis for the C-D non-Foc complex assembly. The N and C termini contain a large number off interfacing residues, whereas the centre of the assembly C45-N57 contain no interfacing residues

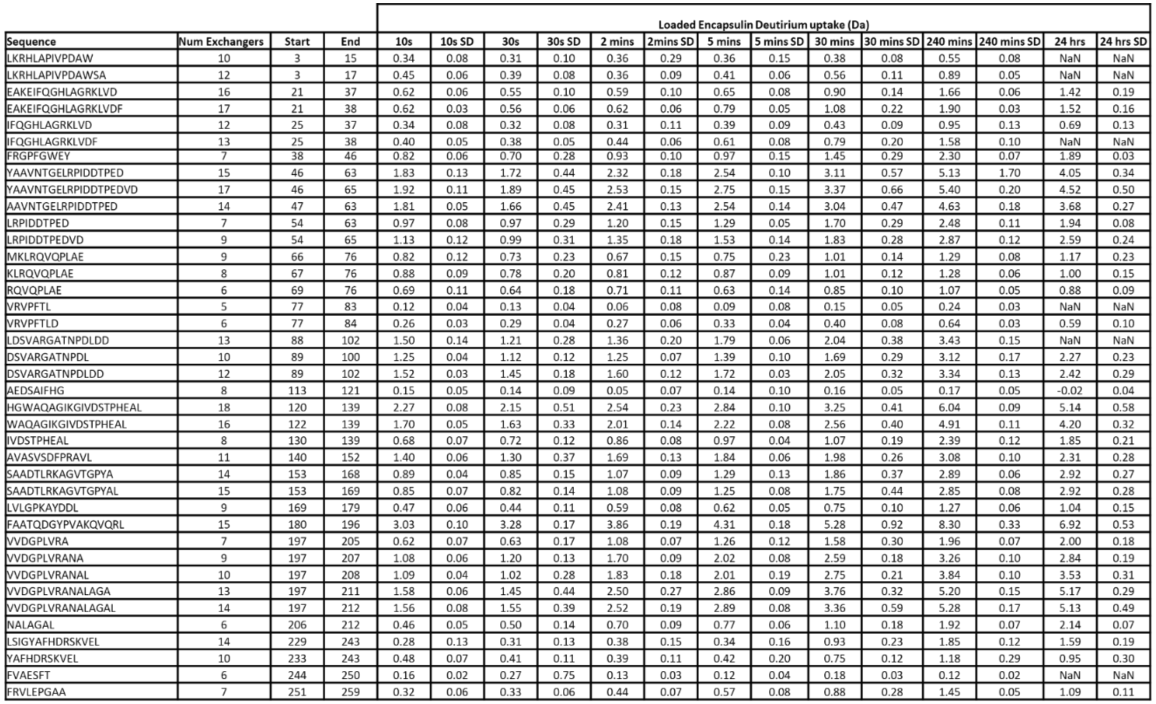
##	Structure 1	HSDC	ASA	BSA	ΔiG
1	H:GLN 6		189.84	0.00	0.00
2	H:LEU 7		71.41	0.00	0.00
3	H:HIS 8		176.33	0.00	0.00
4	H:GLU 9		96.74	0.00	0.00
5	H:PRO 10		62.38	0.00	0.00
6	H:ALA 11		47.16	0.00	0.00
7	H:GLU 12		140.59	0.00	0.00
8	H:LEU 13		128.33	0.00	0.00
9	H:LEU 14		43.83	0.00	0.00
10	H:SER 15		63.20	0.00	0.00
11	H:GLU 16		97.32	15.96 ||	-0.00
12	H:GLU 17		90.55	0.00	0.00
13	H:THR 18		65.03	0.00	0.00
14	H:LYS 19		40.59	0.00	0.00
15	H:ASN 20	H	61.29	45.78 ||||||||	-0.40
16	H:MET 21		28.59	0.00	0.00
17	H:HIS 22		51.17	0.00	0.00

18	H:ARG 23		133.01	81.62	|||||||	-0.02
19	H:ALA 24		10.48	10.48	|||||||||||	0.16
20	H:LEU 25		61.10	0.00		0.00
21	H:VAL 26		50.55	18.07	||||	0.29
22	H:THR 27		56.32	55.65	||||||||||	0.34
23	H:LEU 28		16.89	0.00		0.00
24	H:ILE 29		90.62	0.00		0.00
25	H:GLU 30	HS	106.73	70.93	|||||||	-0.44
26	H:GLU 31	H	22.69	21.53	||||||||||	0.17
27	H:LEU 32		87.63	0.00		0.00
28	H:GLU 33		69.19	0.00		0.00
29	H:ALA 34		16.10	0.67			0.01
30	H:VAL 35		32.33	0.00		0.00
31	H:ASP 36		73.69	0.00		0.00
32	H:TRP 37		120.59	22.64	||	0.36
33	H:TYR 38		55.84	52.40	||||||||||	-0.15
34	H:GLN 39		60.70	0.00		0.00
35	H:GLN 40		104.57	0.00		0.00
36	H:ARG 41	HS	125.08	90.22	||||||||	-0.78
37	H:ALA 42		1.97	0.00		0.00
38	H:ASP 43		104.50	0.00		0.00
39	H:ALA 44		91.39	31.59	||||	0.16
40	H:CYS 45		32.67	12.58	||||	0.30
41	H:SER 46		111.88	50.09	|||||	0.29
42	H:GLU 47	S	93.99	58.41	|||||||	0.38
43	H:PRO 48		97.71	0.00		0.00
44	H:GLY 49		37.42	4.67	||	0.07
45	H:LEU 50		87.76	87.59	||||||||||	1.39
46	H:HIS 51		66.45	0.00		0.00

<table>
<tr><td>47</td><td>H:ASP 52</td><td></td><td>90.20</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>48</td><td>H:VAL 53</td><td></td><td>82.07</td><td>41.19</td><td>||||||</td><td>0.66</td></tr>
<tr><td>49</td><td>H:LEU 54</td><td></td><td>23.82</td><td>23.49</td><td>||||||||||</td><td>0.38</td></tr>
<tr><td>50</td><td>H:ILE 55</td><td></td><td>46.34</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>51</td><td>H:HIS 56</td><td></td><td>110.62</td><td>15.23</td><td>||</td><td>0.10</td></tr>
<tr><td>52</td><td>H:ASN 57</td><td></td><td>58.81</td><td>57.85</td><td>||||||||||</td><td>-0.44</td></tr>
<tr><td>53</td><td>H:LYS 58</td><td></td><td>33.71</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>54</td><td>H:ASN 59</td><td></td><td>83.25</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>55</td><td>H:GLU 60</td><td>S</td><td>72.02</td><td>39.47</td><td>||||||</td><td>-0.29</td></tr>
<tr><td>56</td><td>H:GLU 61</td><td>S</td><td>35.09</td><td>32.79</td><td>||||||||||</td><td>-0.40</td></tr>
<tr><td>57</td><td>H:VAL 62</td><td></td><td>60.31</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>58</td><td>H:GLU 63</td><td></td><td>104.03</td><td>3.30</td><td>|</td><td>-0.04</td></tr>
<tr><td>59</td><td>H:HIS 64</td><td>S</td><td>71.81</td><td>70.27</td><td>||||||||||</td><td>-0.30</td></tr>
<tr><td>60</td><td>H:ALA 65</td><td></td><td>0.33</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>61</td><td>H:MET 66</td><td></td><td>93.52</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>62</td><td>H:MET 67</td><td></td><td>124.81</td><td>54.47</td><td>|||||</td><td>0.90</td></tr>
<tr><td>63</td><td>H:THR 68</td><td></td><td>18.97</td><td>17.81</td><td>||||||||||</td><td>0.10</td></tr>
<tr><td>64</td><td>H:LEU 69</td><td></td><td>31.65</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>65</td><td>H:GLU 70</td><td></td><td>55.14</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>66</td><td>H:TRP 71</td><td></td><td>86.54</td><td>69.20</td><td>||||||||</td><td>0.71</td></tr>
<tr><td>67</td><td>H:ILE 72</td><td></td><td>0.00</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>68</td><td>H:ARG 73</td><td></td><td>64.72</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>69</td><td>H:ARG 74</td><td></td><td>169.32</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>70</td><td>H:ARG 75</td><td>HS</td><td>117.45</td><td>29.66</td><td>|||</td><td>-0.88</td></tr>
<tr><td>71</td><td>H:SER 76</td><td></td><td>9.89</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>72</td><td>H:PRO 77</td><td></td><td>112.31</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>73</td><td>H:VAL 78</td><td></td><td>90.55</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>74</td><td>H:PHE 79</td><td></td><td>69.94</td><td>0.00</td><td></td><td>0.00</td></tr>
<tr><td>75</td><td>H:ASP 80</td><td></td><td>30.80</td><td>0.00</td><td></td><td>0.00</td></tr>
</table>

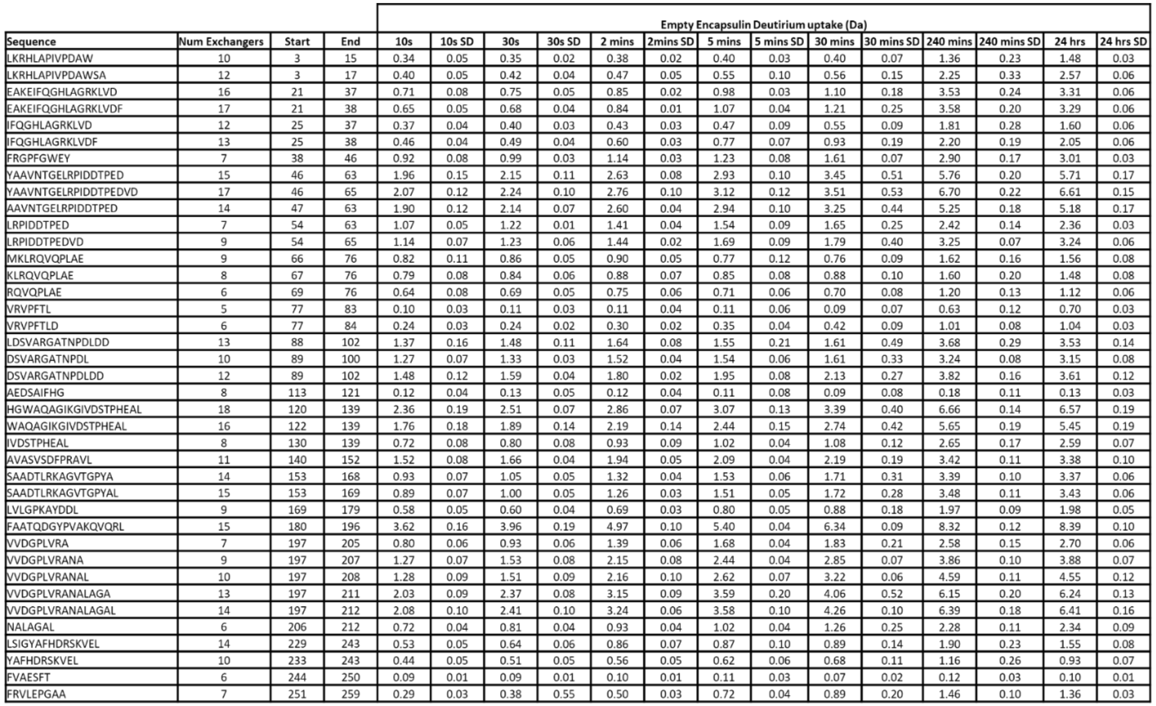
76	H:ALA 81	54.21	0.00	0.00
77	H:HIS 82	108.80	0.00	0.00
78	H:MET 83	34.40	0.00	0.00
79	H:ARG 84	179.84	0.00	0.00
80	H:THR 85	54.76	0.00	0.00
81	H:TYR 86	98.37	0.00	0.00
82	H:LEU 87	84.65	0.00	0.00
83	H:PHE 88	208.63	0.00	0.00
84	H:THR 89	66.38	0.00	0.00
85	H:GLU 90	168.37	0.00	0.00
86	H:ARG 91	134.14	0.00	0.00
87	H:PRO 92	83.48	0.00	0.00
88	H:ILE 93	125.30	0.00	0.00
89	H:LEU 94	174.36	0.00	0.00
90	H:GLU 95	98.32	0.00	0.00
91	H:LEU 96	58.98	0.00	0.00

SI Table 4.5: The PISA analysis for the C-D FOC complex assembly. The N and C termini contain a large number of solvent exposed residues, whereas the centre of the assembly C45-N57 a high number of buried residues



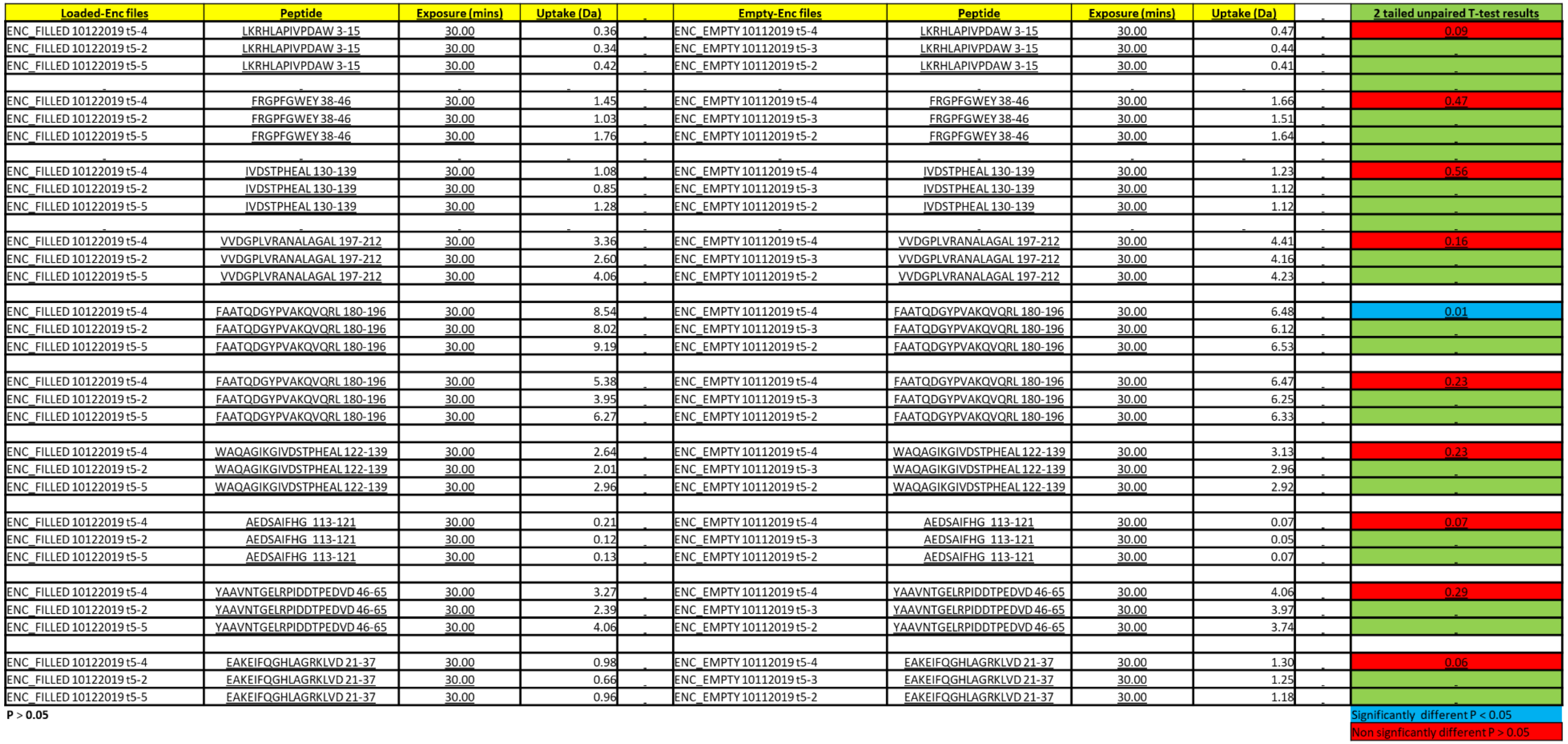
				Loaded Encapsulin Deutirium uptake (Da)													
Sequence	Num Exchangers	Start	End	10s	10s SD	30s	30s SD	2 mins	2mins SD	5 mins	5 mins SD	30 mins	30 mins SD	240 mins	240 mins SD	24 hrs	24 hrs SD
LKRHLAPIVPDAW	10	3	15	0.34	0.08	0.31	0.10	0.36	0.29	0.36	0.15	0.38	0.08	0.55	0.08	NaN	NaN
LKRHLAPIVPDAWSA	12	3	17	0.45	0.06	0.39	0.08	0.36	0.09	0.41	0.06	0.56	0.11	0.89	0.05	NaN	NaN
EAKEIFQGHLAGRKLVD	16	21	37	0.62	0.06	0.55	0.10	0.59	0.10	0.65	0.08	0.90	0.14	1.66	0.06	1.42	0.19
EAKEIFQGHLAGRKLVDF	17	21	38	0.62	0.03	0.56	0.06	0.62	0.06	0.79	0.05	1.08	0.22	1.90	0.03	1.52	0.16
IFQGHLAGRKLVD	12	25	37	0.34	0.08	0.32	0.08	0.31	0.11	0.39	0.09	0.43	0.09	0.95	0.13	0.69	0.13
IFQGHLAGRKLVDF	13	25	38	0.40	0.05	0.38	0.05	0.44	0.06	0.61	0.08	0.79	0.20	1.58	0.10	NaN	NaN
FRGPFGWEY	7	38	46	0.82	0.06	0.70	0.28	0.93	0.10	0.97	0.15	1.45	0.29	2.30	0.07	1.89	0.03
YAAVNTGELRPIDDTPED	15	46	63	1.83	0.13	1.72	0.44	2.32	0.18	2.54	0.10	3.11	0.57	5.13	1.70	4.05	0.34
YAAVNTGELRPIDDTPEDVD	17	46	65	1.92	0.11	1.89	0.45	2.53	0.15	2.75	0.15	3.37	0.66	5.40	0.20	4.52	0.50
AAVNTGELRPIDDTPED	14	47	63	1.81	0.05	1.66	0.45	2.41	0.13	2.54	0.14	3.04	0.47	4.63	0.18	3.68	0.27
LRPIDDTPED	7	54	63	0.97	0.08	0.97	0.29	1.20	0.15	1.29	0.05	1.70	0.29	2.48	0.11	1.94	0.08
LRPIDDTPEDVD	9	54	65	1.13	0.12	0.99	0.31	1.35	0.18	1.53	0.14	1.83	0.28	2.87	0.12	2.59	0.24
MKLRQVQPLAE	9	66	76	0.82	0.12	0.73	0.23	0.67	0.15	0.75	0.23	1.01	0.14	1.29	0.08	1.17	0.23
KLRQVQPLAE	8	67	76	0.88	0.09	0.78	0.20	0.81	0.12	0.87	0.09	1.01	0.12	1.28	0.06	1.00	0.15
RQVQPLAE	6	69	76	0.69	0.11	0.64	0.18	0.71	0.11	0.63	0.14	0.85	0.10	1.07	0.05	0.88	0.09
VRVPFTL	5	77	83	0.12	0.04	0.13	0.04	0.06	0.08	0.09	0.08	0.15	0.05	0.24	0.03	NaN	NaN
VRVPFTLD	6	77	84	0.26	0.03	0.29	0.04	0.27	0.06	0.33	0.04	0.40	0.08	0.64	0.03	0.59	0.10
LDSVARGATNPDLDD	13	88	102	1.50	0.14	1.21	0.28	1.36	0.20	1.79	0.06	2.04	0.38	3.43	0.15	NaN	NaN
DSVARGATNPDL	10	89	100	1.25	0.04	1.12	0.12	1.25	0.07	1.39	0.10	1.69	0.29	3.12	0.17	2.27	0.23
DSVARGATNPDLDD	12	89	102	1.52	0.03	1.45	0.18	1.60	0.12	1.72	0.03	2.05	0.32	3.34	0.13	2.42	0.29
AEDSAIFHG	8	113	121	0.15	0.05	0.14	0.09	0.05	0.07	0.14	0.10	0.16	0.05	0.17	0.05	-0.02	0.04
HGWAQAGIKGIVDSTPHEAL	18	120	139	2.27	0.08	2.15	0.51	2.54	0.23	2.84	0.10	3.25	0.41	6.04	0.09	5.14	0.58
WAQAGIKGIVDSTPHEAL	16	122	139	1.70	0.05	1.63	0.33	2.01	0.14	2.22	0.08	2.56	0.40	4.91	0.11	4.20	0.32
IVDSTPHEAL	8	130	139	0.68	0.07	0.72	0.12	0.86	0.08	0.97	0.04	1.07	0.19	2.39	0.12	1.85	0.21
AVASVSDFPRAVL	11	140	152	1.40	0.06	1.30	0.37	1.69	0.13	1.84	0.06	1.98	0.26	3.08	0.10	2.31	0.28
SAADTLRKAGVTGPYA	14	153	168	0.89	0.04	0.85	0.15	1.07	0.09	1.29	0.13	1.86	0.37	2.89	0.06	2.92	0.27
SAADTLRKAGVTGPYAL	15	153	169	0.85	0.07	0.82	0.14	1.08	0.09	1.25	0.08	1.75	0.44	2.85	0.08	2.92	0.28
LVLGPKAYDDL	9	169	179	0.47	0.06	0.44	0.11	0.59	0.08	0.62	0.05	0.75	0.10	1.27	0.06	1.04	0.15
FAATQDGYPVAKQVQRL	15	180	196	3.03	0.10	3.28	0.17	3.86	0.19	4.31	0.18	5.28	0.92	8.30	0.33	6.92	0.53
VVDGPLVRA	7	197	205	0.62	0.07	0.63	0.17	1.08	0.07	1.26	0.12	1.58	0.30	1.96	0.07	2.00	0.18
VVDGPLVRANA	9	197	207	1.08	0.06	1.20	0.13	1.70	0.09	2.02	0.08	2.59	0.18	3.26	0.10	2.84	0.19
VVDGPLVRANAL	10	197	208	1.09	0.04	1.02	0.28	1.83	0.18	2.01	0.19	2.75	0.21	3.84	0.10	3.53	0.31
VVDGPLVRANALAGA	13	197	211	1.58	0.06	1.45	0.44	2.50	0.27	2.86	0.09	3.76	0.32	5.20	0.15	5.17	0.29
VVDGPLVRANALAGAL	14	197	212	1.56	0.08	1.55	0.39	2.52	0.19	2.89	0.08	3.36	0.59	5.28	0.17	5.13	0.49
NALAGAL	6	206	212	0.46	0.05	0.50	0.14	0.70	0.09	0.77	0.06	1.10	0.18	1.92	0.07	2.14	0.07
LSIGYAFHDRSKVEL	14	229	243	0.28	0.13	0.31	0.13	0.38	0.15	0.34	0.16	0.93	0.23	1.85	0.12	1.59	0.19
YAFHDRSKVEL	10	233	243	0.48	0.07	0.41	0.11	0.39	0.11	0.42	0.20	0.75	0.12	1.18	0.29	0.95	0.30
FVAESFT	6	244	250	0.16	0.02	0.27	0.75	0.13	0.03	0.12	0.04	0.18	0.03	0.12	0.02	NaN	NaN
FRVLEPGAA	7	251	259	0.32	0.06	0.33	0.06	0.44	0.07	0.57	0.08	0.88	0.28	1.45	0.05	1.09	0.11

SI Table 5.1: Recorded uptake of deuterium for each peptide and timepoint observed in HDX-MS of Loaded-Enc for each peptide observed in the HDX MS analysis, the deuterium uptake (in Da) and standard deviation of triplicate data is shown for each timepoint (10s, 30s, 2mins, 5mins, 240mins and 24 hours) for each peptide. The number of exchangeable backbone amide hydrogens is also stated (Num Exchangers).



				Empty Encapsulin Deutirium uptake (Da)													
Sequence	**Num Exchangers**	**Start**	**End**	**10s**	**10s SD**	**30s**	**30s SD**	**2 mins**	**2mins SD**	**5 mins**	**5 mins SD**	**30 mins**	**30 mins SD**	**240 mins**	**240 mins SD**	**24 hrs**	**24 hrs SD**
LKRHLAPIVPDAW	10	3	15	0.34	0.05	0.35	0.02	0.38	0.02	0.40	0.03	0.40	0.07	1.36	0.23	1.48	0.03
LKRHLAPIVPDAWSA	12	3	17	0.40	0.05	0.42	0.04	0.47	0.05	0.55	0.10	0.56	0.15	2.25	0.33	2.57	0.06
EAKEIFQGHLAGRKLVD	16	21	37	0.71	0.08	0.75	0.05	0.85	0.02	0.98	0.03	1.10	0.18	3.53	0.24	3.31	0.06
EAKEIFQGHLAGRKLVDF	17	21	38	0.65	0.05	0.68	0.04	0.84	0.01	1.07	0.04	1.21	0.25	3.58	0.20	3.29	0.06
IFQGHLAGRKLVD	12	25	37	0.37	0.04	0.40	0.03	0.43	0.03	0.47	0.09	0.55	0.09	1.81	0.28	1.60	0.06
IFQGHLAGRKLVDF	13	25	38	0.46	0.04	0.49	0.04	0.60	0.03	0.77	0.07	0.93	0.19	2.20	0.19	2.05	0.06
FRGPFGWEY	7	38	46	0.92	0.08	0.99	0.03	1.14	0.03	1.23	0.08	1.61	0.07	2.90	0.17	3.01	0.03
YAAVNTGELRPIDDTPED	15	46	63	1.96	0.15	2.15	0.11	2.63	0.08	2.93	0.10	3.45	0.51	5.76	0.20	5.71	0.17
YAAVNTGELRPIDDTPEDVD	17	46	65	2.07	0.12	2.24	0.10	2.76	0.10	3.12	0.12	3.51	0.53	6.70	0.22	6.61	0.15
AAVNTGELRPIDDTPED	14	47	63	1.90	0.12	2.14	0.07	2.60	0.04	2.94	0.10	3.25	0.44	5.25	0.18	5.18	0.17
LRPIDDTPED	7	54	63	1.07	0.05	1.22	0.01	1.41	0.04	1.54	0.09	1.65	0.25	2.42	0.14	2.36	0.03
LRPIDDTPEDVD	9	54	65	1.14	0.07	1.23	0.06	1.44	0.02	1.69	0.09	1.79	0.40	3.25	0.07	3.24	0.06
MKLRQVQPLAE	9	66	76	0.82	0.11	0.86	0.05	0.90	0.05	0.77	0.12	0.76	0.09	1.62	0.16	1.56	0.08
KLRQVQPLAE	8	67	76	0.79	0.08	0.84	0.06	0.88	0.07	0.85	0.08	0.88	0.10	1.60	0.20	1.48	0.08
RQVQPLAE	6	69	76	0.64	0.08	0.69	0.05	0.75	0.06	0.71	0.06	0.70	0.08	1.20	0.13	1.12	0.06
VRVPFTL	5	77	83	0.10	0.03	0.11	0.03	0.11	0.04	0.11	0.06	0.09	0.07	0.63	0.12	0.70	0.03
VRVPFTLD	6	77	84	0.24	0.03	0.24	0.02	0.30	0.02	0.35	0.04	0.42	0.09	1.01	0.08	1.04	0.03
LDSVARGATNPDLDD	13	88	102	1.37	0.16	1.48	0.11	1.64	0.08	1.55	0.21	1.61	0.49	3.68	0.29	3.53	0.14
DSVARGATNPDL	10	89	100	1.27	0.07	1.33	0.03	1.52	0.04	1.54	0.06	1.61	0.33	3.24	0.08	3.15	0.08
DSVARGATNPDLDD	12	89	102	1.48	0.12	1.59	0.04	1.80	0.02	1.95	0.08	2.13	0.27	3.82	0.16	3.61	0.12
AEDSAIFHG	8	113	121	0.12	0.04	0.13	0.05	0.12	0.04	0.11	0.08	0.09	0.08	0.18	0.11	0.13	0.03
HGWAQAGIKGIVDSTPHEAL	18	120	139	2.36	0.19	2.51	0.07	2.86	0.07	3.07	0.13	3.39	0.40	6.66	0.14	6.57	0.19
WAQAGIKGIVDSTPHEAL	16	122	139	1.76	0.18	1.89	0.14	2.19	0.14	2.44	0.15	2.74	0.42	5.65	0.19	5.45	0.19
IVDSTPHEAL	8	130	139	0.72	0.08	0.80	0.08	0.93	0.09	1.02	0.04	1.08	0.12	2.65	0.17	2.59	0.07
AVASVSDFPRAVL	11	140	152	1.52	0.08	1.66	0.04	1.94	0.05	2.09	0.04	2.19	0.19	3.42	0.11	3.38	0.10
SAADTLRKAGVTGPYA	14	153	168	0.93	0.07	1.05	0.05	1.32	0.04	1.53	0.06	1.71	0.31	3.39	0.10	3.37	0.06
SAADTLRKAGVTGPYAL	15	153	169	0.89	0.07	1.00	0.05	1.26	0.03	1.51	0.05	1.72	0.28	3.48	0.11	3.43	0.06
LVLGPKAYDDL	9	169	179	0.58	0.05	0.60	0.04	0.69	0.03	0.80	0.05	0.88	0.18	1.97	0.09	1.98	0.05
FAATQDGYPVAKQVQRL	15	180	196	3.62	0.16	3.96	0.19	4.97	0.10	5.40	0.04	6.34	0.09	8.32	0.12	8.39	0.10
VVDGPLVRA	7	197	205	0.80	0.06	0.93	0.06	1.39	0.06	1.68	0.04	1.83	0.21	2.58	0.15	2.70	0.06
VVDGPLVRANA	9	197	207	1.27	0.07	1.53	0.08	2.15	0.08	2.44	0.04	2.85	0.07	3.86	0.10	3.88	0.07
VVDGPLVRANAL	10	197	208	1.28	0.09	1.51	0.09	2.16	0.10	2.62	0.07	3.22	0.06	4.59	0.11	4.55	0.12
VVDGPLVRANALAGA	13	197	211	2.03	0.09	2.37	0.08	3.15	0.09	3.59	0.20	4.06	0.52	6.15	0.20	6.24	0.13
VVDGPLVRANALAGAL	14	197	212	2.08	0.10	2.41	0.10	3.24	0.06	3.58	0.10	4.26	0.10	6.39	0.18	6.41	0.16
NALAGAL	6	206	212	0.72	0.04	0.81	0.04	0.93	0.04	1.02	0.04	1.26	0.25	2.28	0.11	2.34	0.09
LSIGYAFHDRSKVEL	14	229	243	0.53	0.05	0.64	0.06	0.86	0.07	0.87	0.10	0.89	0.14	1.90	0.23	1.55	0.08
YAFHDRSKVEL	10	233	243	0.44	0.05	0.51	0.05	0.56	0.05	0.62	0.06	0.68	0.11	1.16	0.26	0.93	0.07
FVAESFT	6	244	250	0.09	0.01	0.09	0.01	0.10	0.01	0.11	0.03	0.07	0.02	0.12	0.03	0.10	0.01
FRVLEPGAA	7	251	259	0.29	0.03	0.38	0.55	0.50	0.03	0.72	0.04	0.89	0.20	1.46	0.10	1.36	0.03

SI Table 5.2: Recorded uptake of deuterium for each peptide and timepoint observed in HDX-MS of Empty-Enc for each peptide observed in the HDX MS analysis, the deuterium uptake (in Da) and standard deviation of triplicate data is shown for each timepoint (10s, 30s, 2mins, 5mins, 240mins and 24 hours) for each peptide. The number of exchangeable backbone amide hydrogens is also stated (Exchangers)



Loaded-Enc files	Peptide	Exposure (mins)	Uptake (Da)		Empty-Enc files	Peptide	Exposure (mins)	Uptake (Da)		2 tailed unpaired T-test results
ENC_FILLED 10122019 t5-4	LKRHLAPIVPDAW 3-15	30.00	0.36		ENC_EMPTY 10112019 t5-4	LKRHLAPIVPDAW 3-15	30.00	0.47		🟥 0.09
ENC_FILLED 10122019 t5-2	LKRHLAPIVPDAW 3-15	30.00	0.34		ENC_EMPTY 10112019 t5-3	LKRHLAPIVPDAW 3-15	30.00	0.44		
ENC_FILLED 10122019 t5-5	LKRHLAPIVPDAW 3-15	30.00	0.42		ENC_EMPTY 10112019 t5-2	LKRHLAPIVPDAW 3-15	30.00	0.41		
ENC_FILLED 10122019 t5-4	FRGPFGWEY 38-46	30.00	1.45		ENC_EMPTY 10112019 t5-4	FRGPFGWEY 38-46	30.00	1.66		🟥 0.47
ENC_FILLED 10122019 t5-2	FRGPFGWEY 38-46	30.00	1.03		ENC_EMPTY 10112019 t5-3	FRGPFGWEY 38-46	30.00	1.51		
ENC_FILLED 10122019 t5-5	FRGPFGWEY 38-46	30.00	1.76		ENC_EMPTY 10112019 t5-2	FRGPFGWEY 38-46	30.00	1.64		
ENC_FILLED 10122019 t5-4	IVDSTPHEAL 130-139	30.00	1.08		ENC_EMPTY 10112019 t5-4	IVDSTPHEAL 130-139	30.00	1.23		🟥 0.56
ENC_FILLED 10122019 t5-2	IVDSTPHEAL 130-139	30.00	0.85		ENC_EMPTY 10112019 t5-3	IVDSTPHEAL 130-139	30.00	1.12		
ENC_FILLED 10122019 t5-5	IVDSTPHEAL 130-139	30.00	1.28		ENC_EMPTY 10112019 t5-2	IVDSTPHEAL 130-139	30.00	1.12		
ENC_FILLED 10122019 t5-4	VVDGPLVRANALAGAL 197-212	30.00	3.36		ENC_EMPTY 10112019 t5-4	VVDGPLVRANALAGAL 197-212	30.00	4.41		🟥 0.16
ENC_FILLED 10122019 t5-2	VVDGPLVRANALAGAL 197-212	30.00	2.60		ENC_EMPTY 10112019 t5-3	VVDGPLVRANALAGAL 197-212	30.00	4.16		
ENC_FILLED 10122019 t5-5	VVDGPLVRANALAGAL 197-212	30.00	4.06		ENC_EMPTY 10112019 t5-2	VVDGPLVRANALAGAL 197-212	30.00	4.23		
ENC_FILLED 10122019 t5-4	FAATQDGYPVAKQVQRL 180-196	30.00	8.54		ENC_EMPTY 10112019 t5-4	FAATQDGYPVAKQVQRL 180-196	30.00	6.48		🟦 0.01
ENC_FILLED 10122019 t5-2	FAATQDGYPVAKQVQRL 180-196	30.00	8.02		ENC_EMPTY 10112019 t5-3	FAATQDGYPVAKQVQRL 180-196	30.00	6.12		
ENC_FILLED 10122019 t5-5	FAATQDGYPVAKQVQRL 180-196	30.00	9.19		ENC_EMPTY 10112019 t5-2	FAATQDGYPVAKQVQRL 180-196	30.00	6.53		
ENC_FILLED 10122019 t5-4	FAATQDGYPVAKQVQRL 180-196	30.00	5.38		ENC_EMPTY 10112019 t5-4	FAATQDGYPVAKQVQRL 180-196	30.00	6.47		🟥 0.23
ENC_FILLED 10122019 t5-2	FAATQDGYPVAKQVQRL 180-196	30.00	3.95		ENC_EMPTY 10112019 t5-3	FAATQDGYPVAKQVQRL 180-196	30.00	6.25		
ENC_FILLED 10122019 t5-5	FAATQDGYPVAKQVQRL 180-196	30.00	6.27		ENC_EMPTY 10112019 t5-2	FAATQDGYPVAKQVQRL 180-196	30.00	6.33		
ENC_FILLED 10122019 t5-4	WAQAGIKGIVDSTPHEAL 122-139	30.00	2.64		ENC_EMPTY 10112019 t5-4	WAQAGIKGIVDSTPHEAL 122-139	30.00	3.13		🟥 0.23
ENC_FILLED 10122019 t5-2	WAQAGIKGIVDSTPHEAL 122-139	30.00	2.01		ENC_EMPTY 10112019 t5-3	WAQAGIKGIVDSTPHEAL 122-139	30.00	2.96		
ENC_FILLED 10122019 t5-5	WAQAGIKGIVDSTPHEAL 122-139	30.00	2.96		ENC_EMPTY 10112019 t5-2	WAQAGIKGIVDSTPHEAL 122-139	30.00	2.92		
ENC_FILLED 10122019 t5-4	AEDSAIFHG 113-121	30.00	0.21		ENC_EMPTY 10112019 t5-4	AEDSAIFHG 113-121	30.00	0.07		🟥 0.07
ENC_FILLED 10122019 t5-2	AEDSAIFHG 113-121	30.00	0.12		ENC_EMPTY 10112019 t5-3	AEDSAIFHG 113-121	30.00	0.05		
ENC_FILLED 10122019 t5-5	AEDSAIFHG 113-121	30.00	0.13		ENC_EMPTY 10112019 t5-2	AEDSAIFHG 113-121	30.00	0.07		
ENC_FILLED 10122019 t5-4	YAAVNTGELRPIDDTPEDVD 46-65	30.00	3.27		ENC_EMPTY 10112019 t5-4	YAAVNTGELRPIDDTPEDVD 46-65	30.00	4.06		🟥 0.29
ENC_FILLED 10122019 t5-2	YAAVNTGELRPIDDTPEDVD 46-65	30.00	2.39		ENC_EMPTY 10112019 t5-3	YAAVNTGELRPIDDTPEDVD 46-65	30.00	3.97		
ENC_FILLED 10122019 t5-5	YAAVNTGELRPIDDTPEDVD 46-65	30.00	4.06		ENC_EMPTY 10112019 t5-2	YAAVNTGELRPIDDTPEDVD 46-65	30.00	3.74		
ENC_FILLED 10122019 t5-4	EAKEIFQGHLAGRKLVD 21-37	30.00	0.98		ENC_EMPTY 10112019 t5-4	EAKEIFQGHLAGRKLVD 21-37	30.00	1.30		🟥 0.06
ENC_FILLED 10122019 t5-2	EAKEIFQGHLAGRKLVD 21-37	30.00	0.66		ENC_EMPTY 10112019 t5-3	EAKEIFQGHLAGRKLVD 21-37	30.00	1.25		
ENC_FILLED 10122019 t5-5	EAKEIFQGHLAGRKLVD 21-37	30.00	0.96		ENC_EMPTY 10112019 t5-2	EAKEIFQGHLAGRKLVD 21-37	30.00	1.18		

P > 0.05

🟦 Significantly different P < 0.05

🟥 Non signficantly different P > 0.05

SI Table 5.3: Twin tailed unpaired student T-test results comparing the Loaded-Enc and Empty-Enc biological repeats of the highlighted peptides after 30 minutes of deuterium labelling. The null hypothesis was that the samples were not significantly different

Loaded-ENC files	Peptide	Exposure (mins)	Uptake (Da)		Empty-Enc files	Peptide	Exposure (mins)	Uptake (Da)		2 tailed unpaired T-test results
hoch enc enc ftn 03032019 t6-1	LKRHLAPIVPDAW 3-15	240.00	0.54		ENC_EMPTY 10112019 t6-4	LKRHLAPIVPDAW 3-15	240.00	1.53		🟥 0.05
hoch enc enc ftn 03032019 t6-2	LKRHLAPIVPDAW 3-15	240.00	0.55		hoch enc 1mg 03012019 t6-4 rpt	LKRHLAPIVPDAW 3-15	240.00	1.28		
hoch enc enc ftn 03032019 t6-3	LKRHLAPIVPDAW 3-15	240.00	0.57		hoch enc 1mg 03012019 t6-5 rpt	LKRHLAPIVPDAW 3-15	240.00	0.96		
hoch enc enc ftn 03032019 t6-1	FRGPFGWEY 38-46	240.00	0.54		ENC_EMPTY 10112019 t6-4	FRGPFGWEY 38-46	240.00	1.53		🟥 0.05
hoch enc enc ftn 03032019 t6-2	FRGPFGWEY 38-46	240.00	0.55		hoch enc 1mg 03012019 t6-4 rpt	FRGPFGWEY 38-46	240.00	1.28		
hoch enc enc ftn 03032019 t6-3	FRGPFGWEY 38-46	240.00	0.57		hoch enc 1mg 03012019 t6-5 rpt	FRGPFGWEY 38-46	240.00	0.96		
hoch enc enc ftn 03032019 t6-1	IVDSTPHEAL 130-139	240.00	2.56		ENC_EMPTY 10112019 t6-4	IVDSTPHEAL 130-139	240.00	2.54		🟥 0.09
hoch enc enc ftn 03032019 t6-2	IVDSTPHEAL 130-139	240.00	2.29		hoch enc 1mg 03012019 t6-4 rpt	IVDSTPHEAL 130-139	240.00	2.69		
hoch enc enc ftn 03032019 t6-3	IVDSTPHEAL 130-139	240.00	2.34		hoch enc 1mg 03012019 t6-5 rpt	IVDSTPHEAL 130-139	240.00	2.91		
hoch enc enc ftn 03032019 t6-1	VVDGPLVRANALAGAL 197-212	240.00	5.45		ENC_EMPTY 10112019 t6-4	VVDGPLVRANALAGAL 197-212	240.00	6.48		🟦 0.00
hoch enc enc ftn 03032019 t6-2	VVDGPLVRANALAGAL 197-212	240.00	5.10		hoch enc 1mg 03012019 t6-4 rpt	VVDGPLVRANALAGAL 197-212	240.00	6.12		
hoch enc enc ftn 03032019 t6-3	VVDGPLVRANALAGAL 197-212	240.00	5.31		hoch enc 1mg 03012019 t6-5 rpt	VVDGPLVRANALAGAL 197-212	240.00	6.53		
hoch enc enc ftn 03032019 t6-1	FAATQDGYPVAKQVQRL 180-196	240.00	8.54		ENC_EMPTY 10112019 t6-4	FAATQDGYPVAKQVQRL 180-196	240.00	8.30		🟥 0.72
hoch enc enc ftn 03032019 t6-2	FAATQDGYPVAKQVQRL 180-196	240.00	8.02		hoch enc 1mg 03012019 t6-4 rpt	FAATQDGYPVAKQVQRL 180-196	240.00	8.13		
hoch enc enc ftn 03032019 t6-3	FAATQDGYPVAKQVQRL 180-196	240.00	9.19		hoch enc 1mg 03012019 t6-5 rpt	FAATQDGYPVAKQVQRL 180-196	240.00	8.85		
hoch enc enc ftn 03032019 t6-1	WAQAGIKGIVDSTPHEAL 122-139	240.00	5.06		ENC_EMPTY 10112019 t6-4	WAQAGIKGIVDSTPHEAL 122-139	240.00	5.40		🟦 0.01
hoch enc enc ftn 03032019 t6-2	WAQAGIKGIVDSTPHEAL 122-139	240.00	4.82		hoch enc 1mg 03012019 t6-4 rpt	WAQAGIKGIVDSTPHEAL 122-139	240.00	5.60		
hoch enc enc ftn 03032019 t6-3	WAQAGIKGIVDSTPHEAL 122-139	240.00	4.85		hoch enc 1mg 03012019 t6-5 rpt	WAQAGIKGIVDSTPHEAL 122-139	240.00	5.78		
hoch enc enc ftn 03032019 t6-1	AEDSAIFHG 113-121	240.00	0.15		ENC_EMPTY 10112019 t6-4	AEDSAIFHG 113-121	240.00	0.11		🟥 0.69
hoch enc enc ftn 03032019 t6-2	AEDSAIFHG 113-121	240.00	0.18		hoch enc 1mg 03012019 t6-4 rpt	AEDSAIFHG 113-121	240.00	0.35		
hoch enc enc ftn 03032019 t6-3	AEDSAIFHG 113-121	240.00	0.19		hoch enc 1mg 03012019 t6-5 rpt	AEDSAIFHG 113-121	240.00	0.17		
hoch enc enc ftn 03032019 t6-1	YAAVNTGELRPIDDTPEDVD 46-65	240.00	5.62		ENC_EMPTY 10112019 t6-4	YAAVNTGELRPIDDTPEDVD 46-65	240.00	6.58		🟦 0.00
hoch enc enc ftn 03032019 t6-2	YAAVNTGELRPIDDTPEDVD 46-65	240.00	5.27		hoch enc 1mg 03012019 t6-4 rpt	YAAVNTGELRPIDDTPEDVD 46-65	240.00	6.52		
hoch enc enc ftn 03032019 t6-3	YAAVNTGELRPIDDTPEDVD 46-65	240.00	5.28		hoch enc 1mg 03012019 t6-5 rpt	YAAVNTGELRPIDDTPEDVD 46-65	240.00	7.02		
hoch enc enc ftn 03032019 t6-1	EAKEIFQGHLAGRKLVD 21-37	240.00	1.67		ENC_EMPTY 10112019 t6-4	EAKEIFQGHLAGRKLVD 21-37	240.00	3.26		🟦 0.01
hoch enc enc ftn 03032019 t6-2	EAKEIFQGHLAGRKLVD 21-37	240.00	1.64		hoch enc 1mg 03012019 t6-4 rpt	EAKEIFQGHLAGRKLVD 21-37	240.00	3.78		
hoch enc enc ftn 03032019 t6-3	EAKEIFQGHLAGRKLVD 21-37	240.00	1.69		hoch enc 1mg 03012019 t6-5 rpt	EAKEIFQGHLAGRKLVD 21-37	240.00	3.70		

P > 0.05

🟦 Significantly different P < 0.05
🟥 Non signficantly different P > 0.05

SI Table 5.4: Twin tailed unpaired student T-test results comparing the Loaded-Enc and Empty-Enc biological repeats of the highlighted peptides after 240 minutes of deuterium labelling. The null hypothesis was that the samples were not significantly different

Loaded-ENC files	Peptide	Exposure (mins)	Uptake (Da)		Empty-Enc files	Peptide	Exposure (mins)	Uptake (Da)		2 tailed unpaired T-test results
hoch enc enc ftn 03032019 t6-1	LKRHLAPIVPDAW 3-15	240.00	0.54		ENC_EMPTY 10112019 t6-4	LKRHLAPIVPDAW 3-15	240.00	1.53		🟥 0.05
hoch enc enc ftn 03032019 t6-2	LKRHLAPIVPDAW 3-15	240.00	0.55		hoch enc 1mg 03012019 t6-4 rpt	LKRHLAPIVPDAW 3-15	240.00	1.28		
hoch enc enc ftn 03032019 t6-3	LKRHLAPIVPDAW 3-15	240.00	0.57		hoch enc 1mg 03012019 t6-5 rpt	LKRHLAPIVPDAW 3-15	240.00	0.96		
hoch enc enc ftn 03032019 t6-1	FRGPFGWEY 38-46	240.00	0.54		ENC_EMPTY 10112019 t6-4	FRGPFGWEY 38-46	240.00	1.53		🟥 0.05
hoch enc enc ftn 03032019 t6-2	FRGPFGWEY 38-46	240.00	0.55		hoch enc 1mg 03012019 t6-4 rpt	FRGPFGWEY 38-46	240.00	1.28		
hoch enc enc ftn 03032019 t6-3	FRGPFGWEY 38-46	240.00	0.57		hoch enc 1mg 03012019 t6-5 rpt	FRGPFGWEY 38-46	240.00	0.96		
hoch enc enc ftn 03032019 t6-1	IVDSTPHEAL 130-139	240.00	2.56		ENC_EMPTY 10112019 t6-4	IVDSTPHEAL 130-139	240.00	2.54		🟥 0.09
hoch enc enc ftn 03032019 t6-2	IVDSTPHEAL 130-139	240.00	2.29		hoch enc 1mg 03012019 t6-4 rpt	IVDSTPHEAL 130-139	240.00	2.69		
hoch enc enc ftn 03032019 t6-3	IVDSTPHEAL 130-139	240.00	2.34		hoch enc 1mg 03012019 t6-5 rpt	IVDSTPHEAL 130-139	240.00	2.91		
hoch enc enc ftn 03032019 t6-1	VVDGPLVRANALAGAL 197-212	240.00	5.45		ENC_EMPTY 10112019 t6-4	VVDGPLVRANALAGAL 197-212	240.00	6.48		🟦 0.00
hoch enc enc ftn 03032019 t6-2	VVDGPLVRANALAGAL 197-212	240.00	5.10		hoch enc 1mg 03012019 t6-4 rpt	VVDGPLVRANALAGAL 197-212	240.00	6.12		
hoch enc enc ftn 03032019 t6-3	VVDGPLVRANALAGAL 197-212	240.00	5.31		hoch enc 1mg 03012019 t6-5 rpt	VVDGPLVRANALAGAL 197-212	240.00	6.53		
hoch enc enc ftn 03032019 t6-1	FAATQDGYPVAKQVQRL 180-196	240.00	8.54		ENC_EMPTY 10112019 t6-4	FAATQDGYPVAKQVQRL 180-196	240.00	8.30		🟥 0.72
hoch enc enc ftn 03032019 t6-2	FAATQDGYPVAKQVQRL 180-196	240.00	8.02		hoch enc 1mg 03012019 t6-4 rpt	FAATQDGYPVAKQVQRL 180-196	240.00	8.13		
hoch enc enc ftn 03032019 t6-3	FAATQDGYPVAKQVQRL 180-196	240.00	9.19		hoch enc 1mg 03012019 t6-5 rpt	FAATQDGYPVAKQVQRL 180-196	240.00	8.85		
hoch enc enc ftn 03032019 t6-1	WAQAGIKGIVDSTPHEAL 122-139	240.00	5.06		ENC_EMPTY 10112019 t6-4	WAQAGIKGIVDSTPHEAL 122-139	240.00	5.40		🟦 0.01
hoch enc enc ftn 03032019 t6-2	WAQAGIKGIVDSTPHEAL 122-139	240.00	4.82		hoch enc 1mg 03012019 t6-4 rpt	WAQAGIKGIVDSTPHEAL 122-139	240.00	5.60		
hoch enc enc ftn 03032019 t6-3	WAQAGIKGIVDSTPHEAL 122-139	240.00	4.85		hoch enc 1mg 03012019 t6-5 rpt	WAQAGIKGIVDSTPHEAL 122-139	240.00	5.78		
hoch enc enc ftn 03032019 t6-1	AEDSAIFHG 113-121	240.00	0.15		ENC_EMPTY 10112019 t6-4	AEDSAIFHG 113-121	240.00	0.11		🟥 0.69
hoch enc enc ftn 03032019 t6-2	AEDSAIFHG 113-121	240.00	0.18		hoch enc 1mg 03012019 t6-4 rpt	AEDSAIFHG 113-121	240.00	0.35		
hoch enc enc ftn 03032019 t6-3	AEDSAIFHG 113-121	240.00	0.19		hoch enc 1mg 03012019 t6-5 rpt	AEDSAIFHG 113-121	240.00	0.17		
hoch enc enc ftn 03032019 t6-1	YAAVNTGELRPIDDTPEDVD 46-65	240.00	5.62		ENC_EMPTY 10112019 t6-4	YAAVNTGELRPIDDTPEDVD 46-65	240.00	6.58		🟦 0.00
hoch enc enc ftn 03032019 t6-2	YAAVNTGELRPIDDTPEDVD 46-65	240.00	5.27		hoch enc 1mg 03012019 t6-4 rpt	YAAVNTGELRPIDDTPEDVD 46-65	240.00	6.52		
hoch enc enc ftn 03032019 t6-3	YAAVNTGELRPIDDTPEDVD 46-65	240.00	5.28		hoch enc 1mg 03012019 t6-5 rpt	YAAVNTGELRPIDDTPEDVD 46-65	240.00	7.02		
hoch enc enc ftn 03032019 t6-1	EAKEIFQGHLAGRKLVD 21-37	240.00	1.67		ENC_EMPTY 10112019 t6-4	EAKEIFQGHLAGRKLVD 21-37	240.00	3.26		🟦 0.01
hoch enc enc ftn 03032019 t6-2	EAKEIFQGHLAGRKLVD 21-37	240.00	1.64		hoch enc 1mg 03012019 t6-4 rpt	EAKEIFQGHLAGRKLVD 21-37	240.00	3.78		
hoch enc enc ftn 03032019 t6-3	EAKEIFQGHLAGRKLVD 21-37	240.00	1.69		hoch enc 1mg 03012019 t6-5 rpt	EAKEIFQGHLAGRKLVD 21-37	240.00	3.70		

P > 0.05

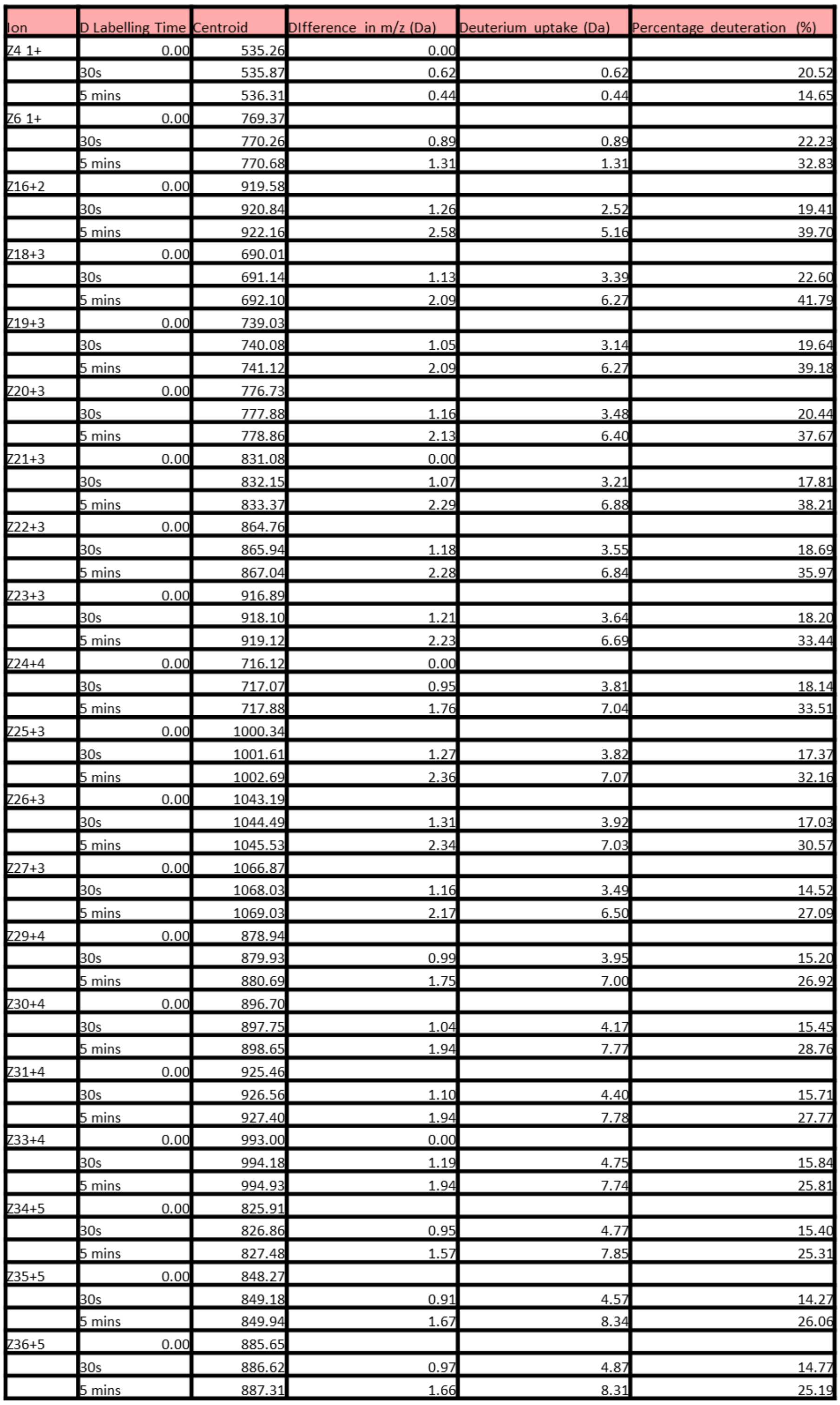
🟦 Significantly different P < 0.05
🟥 Non signficantly different P > 0.05

SI Table 5.5: Twin tailed unpaired student T-test results comparing the Loaded-Enc and Empty-Enc biological repeats of the highlighted peptides after 1440 minutes of deuterium labelling. The null hypothesis was that the samples were not significantly different

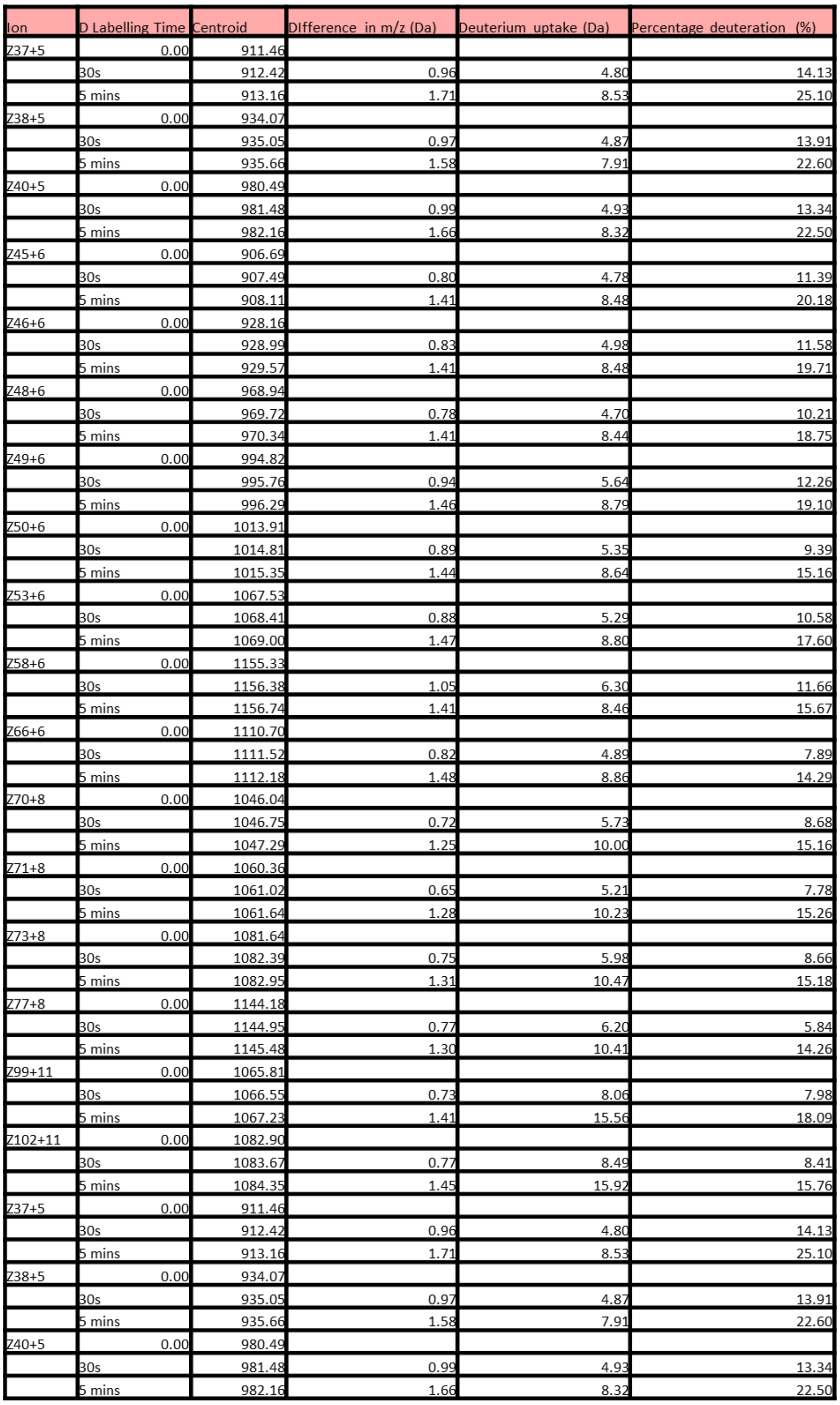
Ion	D Labelling Time	Centroid	DIfference in m/z (Da)	Deuterium uptake (Da)	Percentage deuteration (%)
C5 1+	0.00	505.23	0.05	0.00	
	30s	505.23		0.00	0.00
	5 mins	505.28	0.00	0.05	1.01
C6 1+	0.00	592.26	0.08		
	30s	592.26		0.00	0.00
	5 mins	592.34		0.08	1.65
C7 1+	0.00	693.31			
	30s	693.38	0.07	0.07	1.42
	5 mins	693.42	0.11	0.11	2.16
C11 2+	0.00	585.27	0.00		
	30s	585.40	0.13	0.25	2.79
	5 mins	585.82	0.43	0.85	9.45
C12 2+	0.00	649.79			
	30s	649.97	0.18	0.36	3.57
	5 mins	650.65	0.86	1.72	17.18
C13 1+	0.00	1397.65			
	30s	1398.45	0.81	0.81	7.32
	5 mins	1399.63	1.98	1.98	18.03
C16+2	0.00	884.44	0.00		
	30s	885.11	0.67	1.35	9.62
	5 mins	886.19	1.75	3.50	24.98
C17+2	0.00	948.96			
	30s	949.62	0.66	1.32	8.83
	5 mins	950.90	1.94	3.88	25.89
C18+2	0.00	999.48			
	30s	1000.19	0.71	1.42	8.85
	5 mins	1001.26	1.78	3.57	22.31
C19+2	0.00	1049.01	0.00		
	30s	1049.95	0.94	1.88	11.05
	5 mins	1050.89	1.87	3.74	22.02
C21+3	0.00	789.73			
	30s	790.29	0.57	1.70	8.97
	5 mins	791.18	1.45	4.36	22.97
C22+3	0.00	835.41			
	30s	835.98	0.57	1.72	8.58
	5 mins	836.90	1.49	4.48	22.38
C24+4	0.00	683.59			
	30s	684.03	0.44	1.75	7.96
	5 mins	684.75	1.16	4.64	21.07
C25+4	0.00	711.86			
	30s	712.38	0.51	2.05	8.90
	5 mins	713.08	1.21	4.85	21.08
C26+3	0.00	981.84			
	30s	982.49	0.65	1.96	8.16
	5 mins	983.46	1.62	4.86	20.23
C28+3	0.00	1043.87			
	30s	1044.58	0.70	2.11	8.13
	5 mins	1045.53	1.66	4.97	19.11
C29+3	0.00	1087.55			
	30s	1088.22	0.67	2.01	7.44
	5 mins	1089.23	1.67	5.02	18.61
C30+4	0.00	848.18			
	30s	848.66	0.48	1.93	6.64
	5 mins	849.41	1.23	4.93	17.02
C31+4	0.00	880.44			
	30s	880.79	0.36	1.43	4.92
	5 mins	881.68	1.25	4.99	17.22
C32+3	0.00	1211.27			
	30s	1211.94	0.67	2.01	6.70
	5 mins	1212.95	1.67	5.02	16.75

Ion	D Labelling Time	Centroid	Difference in m/z (Da)	Deuterium uptake (Da)	Percentage deuteration (%)
C36+4	0.00	1012.25			
	30s	1012.81	0.56	2.24	6.79
	5 mins	1013.49	1.24	4.95	14.99
C38+3	0.00	1465.97			
	30s	1466.66	0.69	2.07	5.92
	5 mins	1467.60	1.62	4.87	13.90
C39+3	0.00	1504.22			
	30s	1504.97	0.75	2.25	6.08
	5 mins	1505.89	1.67	5.00	13.52
C40+4	0.00	1160.39			
	30s	1161.19	0.80	3.19	8.40
	5 mins	1161.86	1.46	5.85	15.39
C42+5	0.00	979.48			
	30s	979.95	0.48	2.38	5.94
	5 mins	980.72	1.24	6.20	15.50
C56+6	0.00	1055.51			
	30s	1056.06	0.55	3.27	6.17
	5 mins	1056.55	1.03	6.21	11.71
C57+7	0.00	921.16			
	30s	921.68	0.52	3.65	6.75
	5 mins	922.21	1.05	7.34	13.59
C58+7	0.00	943.46			
	30s	943.95	0.49	3.44	6.26
	5 mins	944.54	1.08	7.59	13.79
C60+7	0.00	978.33			
	30s	978.86	0.53	3.73	6.54
	5 mins	979.39	1.06	7.41	12.99
C75+7	0.00	1244.95			
	30s	1245.50	0.56	3.90	5.35
	5 mins	1246.13	1.18	8.25	11.46
C81+8	0.00	1176.96			
	30s	1177.45	0.49	3.94	5.05
	5 mins	1178.02	1.06	8.44	10.82
C83+7	0.00	1208.24			
	30s	1208.75	0.51	4.60	5.75
	5 mins	1209.24	1.00	8.97	11.21
C85+9	0.00	1102.68			
	30s	1103.21	0.52	4.71	5.74
	5 mins	1103.66	0.98	8.81	10.74
C86+10	0.00	1008.81			
	30s	1009.22	0.41	4.06	4.90
	5 mins	1009.64	0.83	8.33	10.03
C88+9	0.00	1149.70			
	30s	1150.15	0.45	4.04	4.76
	5 mins	1150.65	0.95	8.55	10.06
C90+10	0.00	1057.81			
	30s	1058.33	0.52	5.17	5.94
	5 mins	1058.81	1.00	9.99	11.48
C93+9	0.00	1204.95			
	30s	1205.52	0.57	5.13	5.77
	5 mins	1206.19	1.23	11.10	12.47
C97+11	0.00	1026.16			
	30s	1026.68	0.52	5.19	5.58
	5 mins	1027.33	1.16	11.62	12.49
C102+11	0.00	1080.19			
	30s	1080.84	0.65	7.11	7.33
	5 mins	1081.49	1.30	14.31	14.75
C104+11	0.00	1105.21			
	30s	1105.89	0.69	7.57	7.64
	5 mins	1106.55	1.34	14.75	14.90

SI Table 6.1: Percentage deuteration of assignable c fragments from the isotopic depleted Rru-EncFtn at 30s and 5 minutes of deuterium labelling



Ion	D Labelling Time	Centroid	DIfference in m/z (Da)	Deuterium uptake (Da)	Percentage deuteration (%)
Z4 1+	0.00	535.26	0.00		
	30s	535.87	0.62	0.62	20.52
	5 mins	536.31	0.44	0.44	14.65
Z6 1+	0.00	769.37			
	30s	770.26	0.89	0.89	22.23
	5 mins	770.68	1.31	1.31	32.83
Z16+2	0.00	919.58			
	30s	920.84	1.26	2.52	19.41
	5 mins	922.16	2.58	5.16	39.70
Z18+3	0.00	690.01			
	30s	691.14	1.13	3.39	22.60
	5 mins	692.10	2.09	6.27	41.79
Z19+3	0.00	739.03			
	30s	740.08	1.05	3.14	19.64
	5 mins	741.12	2.09	6.27	39.18
Z20+3	0.00	776.73			
	30s	777.88	1.16	3.48	20.44
	5 mins	778.86	2.13	6.40	37.67
Z21+3	0.00	831.08	0.00		
	30s	832.15	1.07	3.21	17.81
	5 mins	833.37	2.29	6.88	38.21
Z22+3	0.00	864.76			
	30s	865.94	1.18	3.55	18.69
	5 mins	867.04	2.28	6.84	35.97
Z23+3	0.00	916.89			
	30s	918.10	1.21	3.64	18.20
	5 mins	919.12	2.23	6.69	33.44
Z24+4	0.00	716.12	0.00		
	30s	717.07	0.95	3.81	18.14
	5 mins	717.88	1.76	7.04	33.51
Z25+3	0.00	1000.34			
	30s	1001.61	1.27	3.82	17.37
	5 mins	1002.69	2.36	7.07	32.16
Z26+3	0.00	1043.19			
	30s	1044.49	1.31	3.92	17.03
	5 mins	1045.53	2.34	7.03	30.57
Z27+3	0.00	1066.87			
	30s	1068.03	1.16	3.49	14.52
	5 mins	1069.03	2.17	6.50	27.09
Z29+4	0.00	878.94			
	30s	879.93	0.99	3.95	15.20
	5 mins	880.69	1.75	7.00	26.92
Z30+4	0.00	896.70			
	30s	897.75	1.04	4.17	15.45
	5 mins	898.65	1.94	7.77	28.76
Z31+4	0.00	925.46			
	30s	926.56	1.10	4.40	15.71
	5 mins	927.40	1.94	7.78	27.77
Z33+4	0.00	993.00	0.00		
	30s	994.18	1.19	4.75	15.84
	5 mins	994.93	1.94	7.74	25.81
Z34+5	0.00	825.91			
	30s	826.86	0.95	4.77	15.40
	5 mins	827.48	1.57	7.85	25.31
Z35+5	0.00	848.27			
	30s	849.18	0.91	4.57	14.27
	5 mins	849.94	1.67	8.34	26.06
Z36+5	0.00	885.65			
	30s	886.62	0.97	4.87	14.77
	5 mins	887.31	1.66	8.31	25.19



Ion	D Labelling Time	Centroid	DIfference in m/z (Da)	Deuterium uptake (Da)	Percentage deuteration (%)
Z37+5	0.00	911.46			
	30s	912.42	0.96	4.80	14.13
	5 mins	913.16	1.71	8.53	25.10
Z38+5	0.00	934.07			
	30s	935.05	0.97	4.87	13.91
	5 mins	935.66	1.58	7.91	22.60
Z40+5	0.00	980.49			
	30s	981.48	0.99	4.93	13.34
	5 mins	982.16	1.66	8.32	22.50
Z45+6	0.00	906.69			
	30s	907.49	0.80	4.78	11.39
	5 mins	908.11	1.41	8.48	20.18
Z46+6	0.00	928.16			
	30s	928.99	0.83	4.98	11.58
	5 mins	929.57	1.41	8.48	19.71
Z48+6	0.00	968.94			
	30s	969.72	0.78	4.70	10.21
	5 mins	970.34	1.41	8.44	18.75
Z49+6	0.00	994.82			
	30s	995.76	0.94	5.64	12.26
	5 mins	996.29	1.46	8.79	19.10
Z50+6	0.00	1013.91			
	30s	1014.81	0.89	5.35	9.39
	5 mins	1015.35	1.44	8.64	15.16
Z53+6	0.00	1067.53			
	30s	1068.41	0.88	5.29	10.58
	5 mins	1069.00	1.47	8.80	17.60
Z58+6	0.00	1155.33			
	30s	1156.38	1.05	6.30	11.66
	5 mins	1156.74	1.41	8.46	15.67
Z66+6	0.00	1110.70			
	30s	1111.52	0.82	4.89	7.89
	5 mins	1112.18	1.48	8.86	14.29
Z70+8	0.00	1046.04			
	30s	1046.75	0.72	5.73	8.68
	5 mins	1047.29	1.25	10.00	15.16
Z71+8	0.00	1060.36			
	30s	1061.02	0.65	5.21	7.78
	5 mins	1061.64	1.28	10.23	15.26
Z73+8	0.00	1081.64			
	30s	1082.39	0.75	5.98	8.66
	5 mins	1082.95	1.31	10.47	15.18
Z77+8	0.00	1144.18			
	30s	1144.95	0.77	6.20	5.84
	5 mins	1145.48	1.30	10.41	14.26
Z99+11	0.00	1065.81			
	30s	1066.55	0.73	8.06	7.98
	5 mins	1067.23	1.41	15.56	18.09
Z102+11	0.00	1082.90			
	30s	1083.67	0.77	8.49	8.41
	5 mins	1084.35	1.45	15.92	15.76
Z37+5	0.00	911.46			
	30s	912.42	0.96	4.80	14.13
	5 mins	913.16	1.71	8.53	25.10
Z38+5	0.00	934.07			
	30s	935.05	0.97	4.87	13.91
	5 mins	935.66	1.58	7.91	22.60
Z40+5	0.00	980.49			
	30s	981.48	0.99	4.93	13.34
	5 mins	982.16	1.66	8.32	22.50

SI Table 6.2: Percentage deuteration of assignable z fragments from isotopically depleted Rru-EncFtn at 30s and 5 minutes of deuterium labelling

Residue no	*C* fragments	exchange events (Da)	% relative uptake	Deuteratable sites	D-value *c* ions
56	56.00	7.34	13.59	53.00	🟧 0.09
41-42	42.00	6.21	11.71	40.00	🟧 0.18
40	40.00	5.85	15.39	38.00	🟩 0.85
39	39.00	5.00	13.52	37.00	🟩 -0.14
37-38	38.00	4.87	13.90	36.00	🟧 -0.08
34-36	36.00	4.95	14.99	34.00	🟧 0.06
33	33.00	4.76	15.37	31.00	🟩 -0.26
32	32.00	5.02	16.75	30.00	🟩 0.03
31	31.00	4.99	17.22	29.00	🟩 0.06
30	30.00	4.93	17.02	28.00	🟩 -0.09
29	29.00	5.02	18.61	27.00	🟩 0.06
28	28.00	4.97	19.11	26.00	🟧 0.06
26	26.00	4.86	20.23	24.00	🟩 0.01
25	25.00	4.85	21.08	23.00	🟩 0.21
23-24	24.00	4.64	21.07	22.00	🟧 0.08
22	22.00	4.48	22.38	20.00	🟩 0.11
20-21	21.00	4.36	22.97	19.00	🟧 0.31
19	19.00	3.74	22.02	17.00	🟩 0.17
18	18.00	3.57	22.31	16.00	🟩 -0.31
17	17.00	3.88	25.89	15.00	🟩 0.39
14-16	16.00	3.50	24.98	14.00	🟧 0.50
13	13.00	1.98	18.03	11.00	🟩 0.27
12	12.00	1.72	17.18	10.00	🟩 0.87
8-11	11.00	0.85	9.45	9.00	🟧 0.25
7	7.00	0.11	2.16	6.00	🟩 0.11
6	6.00	0.00	0.00	5.00	🟩 0.00
0-5	5.00	0.00	0.00	4.00	🟧 0.00

🟦 Proline or N terminal non exchange sites
🟧 Average deuteration status across residues
🟩 sequential Amide deuteration status

Si Table 6.3: ID-Rru-EncFtn c ion D-value calculations

Residue no	*Z* fragments	exchange events (Da)	% relative uptake	Deuteratable sites	D-value *z* ions
106-103	4.00	0.44	14.65	3.00	🟧 0.15
102	5.00				🟦 0.00
101	6.00	1.31	32.83	4.00	🟩 0.87
100-91	16.00	5.16	39.70	13.00	🟧 0.43
90-89	18.00	6.27	41.79	15.00	🟧 0.55
88	19.00	6.27	39.18	16.00	🟩 0.00
87	20.00	6.40	37.67	17.00	🟩 0.13
86	21.00	6.88	38.21	18.00	🟩 0.47
85	22.00	6.84	35.97	19.00	🟩 -0.04
84	23.00	6.69	33.44	20.00	🟩 -0.15
83	24.00	7.04	33.51	21.00	🟩 0.35
82	25.00	7.07	32.16	22.00	🟩 0.04
81	26.00	7.03	30.57	23.00	🟩 -0.04
80-78	29.00	7.00	26.92	26.00	🟧 -0.01
77	30.00	7.77	28.76	27.00	🟩 0.77
76	31.00	7.78	27.77	28.00	🟩 0.01
75-74	33.00	7.74	25.81	30.00	🟧 -0.02
73	34.00	7.85	25.31	31.00	🟩 0.11
72	35.00	8.34	26.06	32.00	🟩 0.49
71	36.00	8.31	25.19	33.00	🟩 -0.03
70	37.00	8.53	25.10	34.00	🟩 0.22
69-67	40.00	8.32	22.50	37.00	🟧 -0.07
66-62	45.00	8.48	20.18	42.00	🟧 0.03
61	46.00	8.48	19.71	43.00	🟩 0.00
60-59	48.00	8.44	18.75	45.00	🟧 -0.02
58	49.00	8.79	19.10	46.00	🟩 0.35
57	50.00	8.64	15.16	47.00	🟩 -0.14
56-54	53.00	8.80	17.60	50.00	🟧 0.05
53-49	58.00	8.46	15.67	55.00	🟧 -0.07

🟦 Proline or N terminal non exchange sites
🟧 Average deuteration status across residues
🟩 sequential Amide deuteration status

SI Table 6.4: ID-Rru-EncFtn z ion D-value calculations

Residue no	***y* fragment**	**Exchange Events (Da)**	**% relative uptake of the c(*q*)**	**Deuteratable sites**	**D-value z ions**
106-94	13.00	2.07	15.61	11.00	🟧 0.22
93-82	25.00	3.49	15.87	22.00	🟧 0.13
81-77	30.00	3.84	14.24	27.00	🟧 0.07
76-71	36.00	4.59	13.91	33.00	🟧 0.12
70-65	42.00	5.16	13.23	39.00	🟧 0.09
64-54	53.00	7.22	14.45	50.00	🟧 0.19
53.00	54.00	7.29	14.59	51.00	🟧 0.07
52-50	57.00	7.63	14.13	54.00	🟧 0.11
49.00	58.00	7.72	14.57	55.00	🟩 0.09
48.00	59.00	7.72	14.57	56.00	🟦 0.00
47.00	60.00	8.04	14.35	56.00	🟩 0.32
46-37	70.00	9.90	14.78	66.00	🟧 0.19

Avg= 14.52

🟦 Proline or N terminal non exchange sites
🟧 Average deuteration status across residues
🟩 sequential Amide deuteration status

SI Table 6.5: Rru-EncFtn y ion D-value calculations

Appendix iii: Published materials

HDX-MS results detailed in chapter 3 were previously published in the Journal Chem Communications on the 13th February 2020. The manuscript is included on next page.

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Commun.*, 2020, **56**, 3417

Received 16th October 2019,
Accepted 13th February 2020

DOI: 10.1039/c9cc08130e

rsc.li/chemcomm

Mass spectrometry reveals the assembly pathway of encapsulated ferritins and highlights a dynamic ferroxidase interface†

Jennifer Ross,[a] Thomas Lambert,[a] Cecilia Piergentili,[b] Didi He,[c]
Kevin J. Waldron,[d] C. Logan Mackay,[a] Jon Marles-Wright*[b] and
David J. Clarke*[a]

Encapsulated ferritins (EncFtn) are a recently characterised member of the ferritin superfamily. EncFtn proteins are sequestered within encapsulin nanocompartments and form a unique biological iron storage system. Here, we use native mass spectrometry and hydrogen–deuterium exchange mass spectrometry to elucidate the metal-mediated assembly pathway of EncFtn.

Ferritins are ubiquitous enzymes used to oxidise and store iron in almost all life forms.[1] In classical ferritins, a 2-iron ferroxidase centre catalyses the oxidation of Fe(II) and the resulting Fe(III) is stored within their spherical quaternary nanocage structure.[2] A recently characterised sub-family of ferritins, encapsulated ferritins (EncFtn), are sequestered within 20–50 nm icosahedral encapsulin nanocompartments (Enc).[3–5] EncFtn proteins are directed into the interior of the encapsulin shell by a C-terminal localisation sequence, forming an encapsulin system.[3,6] These proteins work together to form a competent iron-store, with the EncFtn responsible for iron oxidation, while iron is mineralised and stored within the encapsulin nanocompartment.[7,8] Due to their greater size, Enc–EncFtn systems have a much greater iron storage potential than classical ferritins and DNA-binding Protein from Starved cells (DPS) nanocages.[4]

Using X-ray crystallography, we have recently shown that members of the EncFtn family adopt a conserved annular quaternary structure (Fig. 1A).[3,9] This is formed from a pentamer of dimers, the topology of which results in two distinct dimer interfaces. One dimer is structurally similar to the four helix-bundle of the classical ferritin monomer[9] and is formed by the two main helices of two EncFtn monomers. This dimer contains the di-iron ferroxidase centre (FOC) active site and is known as the FOC dimer (Fig. 1B). Two glutamates (Glu30 and Glu60) and one histidine (His63) per monomer coordinate the iron ions at the interface of the FOC dimer.[3] In the pentamer of dimers assembly, FOC dimers are arranged in an annular architecture with a second dimer interface, the non-FOC dimer interface (Fig. 1C) connecting five FOC dimers (Fig. 1C). The non-FOC dimer interface is more extensive and mediated *via* a mixture of hydrophobic interactions, hydrogen bonds, and salt bridges.[9]

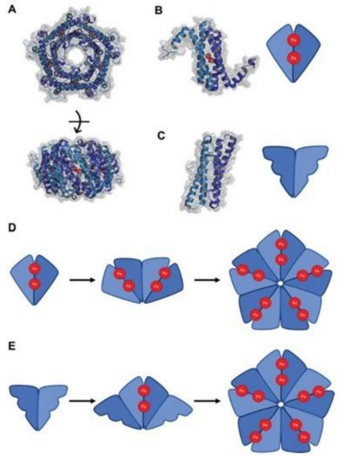


Fig. 1 (A) Annular pentamer of dimers structure of EncFtn (PDB ID: 5DA5) with transparent solvent accessible surface area (SASA) over cartoon secondary structure. (B) The FOC dimer containing two iron ions bound in the FOC shown with cartoon secondary structure and transparent SASA, and a simplified representation. (C) The non-FOC dimer in cartoon and simplified representation. The two possible assembly pathways of EncFtn *via* dimer association; the FOC assembly pathway (D) and the non-FOC assembly pathway (E).

---

[a] *EaStCHEM School of Chemistry, The University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ, Scotland, UK. E-mail: Dave.clarke@ed.ac.uk*
[b] *School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. E-mail: Jon.marles-wright1@ncl.ac.uk*
[c] *Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Roosvelt Drive, Oxford, OX3 7DQ, UK*
[d] *Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cc08130e

Although these crystallographic studies afford us insight into the atomic details of the EncFtn protein structure, these high resolution 'snapshots' offer limited insight into the dynamics and assembly pathways of protein complexes. In contrast, structural mass spectrometry methods such as native mass spectrometry (MS) and ion mobility (IM) can report on protein structural/conformational heterogeneity and structural dynamics in a relatively unbiased manner. Consequently, MS-based techniques have proven highly complementary to traditional 'high-resolution' structural biology methods.[10]

We have previously used native MS to demonstrate that removal of the conserved histidine FOC-ligand in the EncFtn from *Rhodospirillum rubrum* results in loss of the decametric quaternary structure; as a Histidine to Alanine variant of EncFtn exists as only monomer and dimer oligomerisation states.[3] In addition, prior native MS studies have also revealed that the EncFtn from the halophile *Haliangium Ochraceum* exists as a dynamic structure in multiple even-numbered oligomerisation states (dimer, tetramer, hexamer and decamer).[9] This observation implies that higher order oligomers are formed by the association of dimer units; presumably either *via* FOC dimer association, or non-FOC dimer association (Fig. 1D and E).

Herein, we build on these previous findings and employ several complimentary structural proteomic techniques to investigate the dimer association pathway leading to the annular EncFtn structure. We produced both wild type sequence and H63A variants of *H. Ochraceum* EncFtn (Hoch-WT and Hoch-H63A) as truncated variants (lacking the C-terminal encapsulin localization sequence), by heterologous expressed in *Escherichia coli*.[11] After purification, the masses of both Hoch-WT and Hoch-H63A were verified by LC-MS (Table S2, ESI†).

We first analysed the oligomerisation states of Hoch-WT and Hoch-H63A using native MS (Fig. 2). As previously reported, native MS analysis of Hoch-WT reveals a series of oligomerisation states.[9] The major species is dimer (+9 to +11) (Fig. 2A, green circles), with some tetramer (13+ to 15+), hexamer (17+ to 18+) and decamer (22+ to 25+) species also observed (Fig. 2A, purple squares, orange diamonds and pink triangles respectively). Native MS analysis of Hoch-H63A shows that it exists solely as a dimer (7+ to 12+) (Fig. 2A, green circles), demonstrating that the H63A mutation in the *H. Ochraceum* EncFtn disrupts the formation of higher order EncFtn structures in a similar manner to previously reported for *R. rubrum* EncFtn.[3] This was supported by using non-denaturing PAGE where Hoch-WT appeared in multiple oligomerisation states of varying size; whereas Hoch-H63A was only observed in one oligomerisation state (Fig. S2, ESI†).

Having established that dimer assemblies were present in both Hoch-WT and Hoch-H63A, native ion mobility MS was used to investigate their conformation and compare their observed collision cross section (CCS).[12] The IM-MS arrival time profiles of the dimer charge states of Hoch-WT and Hoch-H63A were highly similar – after CCS calculation both Hoch-WT and H63A dimers displayed CCS values of ~ 19.6 nm^2 and presented with similar widths in the CCS profiles (FWHM ~ 2.1 and 2.3 nm^2; Fig. 2C). The similarity of the mobility profiles of the Hoch-WT and Hoch-H63A dimers suggest that they adopt the same dimer form. Moreover, when theoretical CCS values of the two distinct dimer subcomplexes were calculated from the crystal structure of Hoch-WT,[13] it is clear that they display significant differences in theoretical CCS. The theoretical CCS of the non-FOC dimer is 21.4 nm^2 and for the FOC dimer is 25.3 nm^2. Comparison of the experimentally determined CCS values suggests that the dimer form present in both Hoch-WT and Hoch-H63A is the non-FOC dimer (Fig. 2C).

Molecular mass measurement by native MS can also inform on metal content and stoichiometry in metalloproteins. The 10+ charge state of the dimer of Hoch-WT displays several adducts of varying molecular mass; however, none of these adducts are consistent with the coordination of iron (Fe^{2+}:

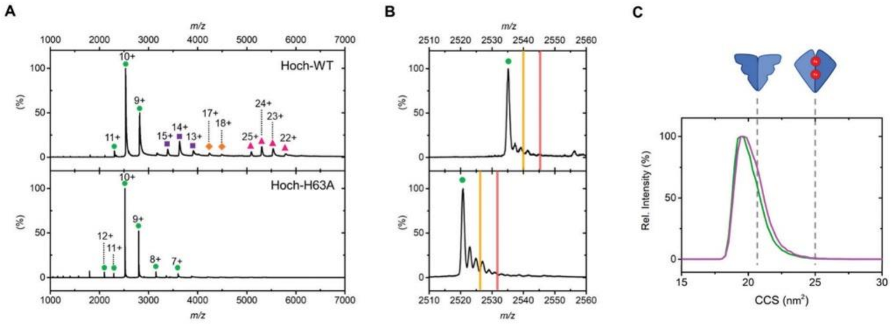


Fig. 2 (A) Native nanoelectrospray ionization mass spectra of Hoch-WT and Hoch-H63A with gas phase oligomerization states stressed with coloured shapes. Dimer as green circles; tetramer as purple squares; hexamer as orange diamonds and decamer as pink triangles. (B) 10+ dimer charge states of Hoch-WT (top panel) and Hoch-H63A (bottom panel). Theoretical *m/z* values of dimers with no metal bound; dimers with one iron bound and two irons bound are highlighted by green circles, yellow lines and red lines respectively. (C) Ion mobility drift time profiles of the 10+ charge states of Hoch-WT (green) and Hoch-H63A (purple). The theoretical collision cross sections of the non-FOC and FOC dimer are shown in dashed grey lines and were calculated using IMPACT software.[13]

theoretical monoisotopic Δmass +54 Da); and we attribute these to non-specific sodium and potassium adduction (commonly observed in native MS). Similarly, the 10+ Hoch-H63A dimer displays no iron coordination (Fig. 2B). The absence of Fe coordination supports the ion mobility data suggesting non-FOC dimer assembly in both variants (Fig. 1C). The charge states of higher order species of Hoch-WT present as wide peaks, due to overlapping signals of adducted species and the limited mass resolving power at molecular mass >40 kDa. Thus, metal coordination could not be determined for these charge states. Inductively coupled plasma (ICP) MS analysis of Hoch-H63A revealed no substantial iron content (5.6%, Table S4, ESI†). In contrast, significant iron content (albeit at sub-stoichiometric levels) was observed in Hoch-WT (29.1%, Table S4, ESI†), suggesting that iron may be associated with the higher order oligomers of Hoch-WT.

In order to investigate iron-loading in these higher molecular mass species and overcome the mass resolution limitation, Hoch-WT was analysed by native MS on a high-resolution FT-ICR MS platform. This instrument affords isotopic resolution, even at high mass, allowing discrimination of overlapping adduct species and more confident assignment of coordinated metals. This analysis confirmed no iron coordination in the dimer forms of Hoch-WT and Hoch-H63A (Fig. S3, ESI†). FT-ICR MS analysis of the 13+ tetrameric Hoch-WT allowed us to distinguish several adducted forms – which could be assigned as a prominent potassiated series, in addition an isotope distribution corresponding to the coordination of $2Fe^{2+}$ (Hoch-$WT_4Fe_2{}^{2+}$) was clearly present (Fig. 3). Interestingly, no signal was observed to suggest coordination of a single iron, suggesting the presence of a single di-nuclear FOC centre in the tetrameric species. Taken together these observations match the predicted stoichiometry of the non-FOC assembly pathway (Fig. 1E) *i.e.* that higher order EncFtn species are assembled from non-FOC dimer association, and that assembly of two non-FOC dimers result in the formation of one di-iron containing FOC.

This hypothesis is strengthened by titration of the Hoch-WT with Fe^{2+} prior to native MS analysis, which resulted in an increased intensity of higher oligomerisation states (tetramer and hexamer in particular) (Fig. S4, ESI†); suggesting that the association of dimers to form higher order subcomplexes is iron dependent.

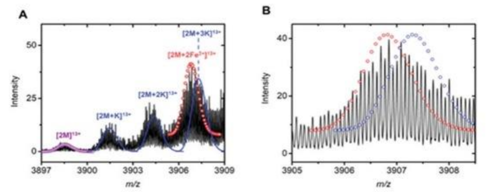


Fig. 3 (A) High resolution FT-ICR MS analysis of 13+ charge state of the Hoch-WT tetramer. The theoretical isotope distributions of Hoch-WT_4 and Hoch-$WT_4Fe_2{}^{2+}$ species are highlighted by purple and red circles respectively. Hoch-WT potassiated series shown by blue lines. (B) Region of the spectra which contains overlapping isotope distributions consistent with Hoch-$WT_4Fe_2{}^{2+}$ (shown in red) and Hoch-$WT_4K_3{}^{+}$ (shown in blue).

To complement our native MS experiments, we employed hydrogen deuterium exchange (HDX) MS (see ESI† for further details). By comparing the rate of backbone amide hydrogen exchange in Hoch-WT and Hoch-H63A, we could ascertain regions which differ in solvent exposure and/or dynamics. This analysis highlighted two regions of the Hoch-H63A variant which undergo exchange significantly faster than the wild-type protein – both are located in the central region of the primary sequence and span aa33–42 and aa46–61 (Fig. 4A, peptides ii and iii).

These regions of the primary sequence, protected from exchange in Hoch-WT, were mapped onto the higher order structure in order to determine their relationship to the two discrete dimer forms of the assembly. First, we delineated the regions of the Hoch-WT protein primary sequence which lie at the interfaces of the two dimers in the decamer assembly using PDBePISA[14] to analyse the Hoch-WT structure (PDB 5N5F, Fig. 4B). Although the helical nature of the protein results in partitioning of the side chains at either face of the helical bundles, it was possible to determine regions of the primary sequence associated with each interface. In general, amino acids which reside in the FOC dimer interface are located in

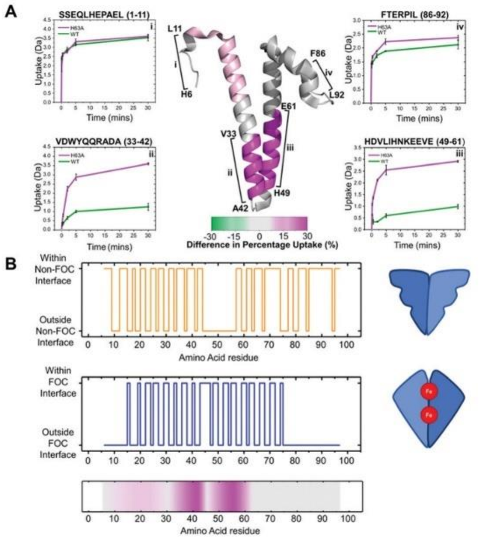


Fig. 4 (A) The difference in HD exchange at 30 minutes of D_2O labelling mapped on to the monomer crystal structure of Hoch-WT. Regions displaying increased exchange in H63A are coloured pink. HDX uptake plots are shown for four peptides (i–iv) highlighted in the main figure. (B) Structural analysis of Hoch-WT to determine the interface regions for both the FOC (blue) and non-FOC (orange) dimer subcomplexes. The differential HDX data is represented below mapped onto the primary sequence of Hoch-WT (colour scheme used as in (A)).

the central region of the primary sequence – between aa16–aa75; with extended regions at both the N- and C-termini that do not present in the interface (Fig. 4B, blue). In contrast, the non-FOC interface is more extensive and involves residues at both the N- and C-termini. However, we note a central extended region (aa45–aa57) which is not present at the non-FOC interface (Fig. 4B, orange). In light of this structural interface analysis, it is clear that the regions of the primary sequence protected from exchange in Hoch-WT span regions of the protein that correlate with the FOC–dimer interface (Fig. 4B); most evident from the region spanning aa46–61. This is consistent with the FOC interface being more accessible in Hoch H63A when compared to Hoch-WT.

These findings suggest that Hoch-H63A is present exclusively as the non-FOC dimer; as such, while the non-FOC dimer interface is shielded from exchange, the FOC dimer interface is exposed. In comparison, our native MS analysis reveals that the wild-type protein is present as dimers and higher order assemblies of dimers. Therefore, although it shares the same shielded non-FOC dimer interface as Hoch-H63A, the FOC interface, exposed in Hoch-H63A, will be somewhat shielded by forming these higher order species. Thus, HDX analysis further supports the hypothesis that the predominant dimeric form of Hoch-WT is the non-FOC dimer, and decamer formation is *via* assembly of these non-FOC dimer subunits.

In conclusion, the solution and gas phase studies of *H. Ochraceum* EncFtn presented here elucidate the assembly pathway of the characteristic pentamer of dimers arrangement found in EncFtns. The decameric assembly is associated by the addition of the non-FOC dimers *via* the coordination of iron by FOC residues, including Histidine 63, and the formation of the FOC interface. Our results highlight the dynamic nature of the FOC in the EncFtn proteins and its dual role in mediating the oligomerisation and enzymatic activity of these proteins.

Protein oligomerisation is energetically and evolutionarily favourable and many important classes of enzyme have active sites that are split between subunits in both homo- and hetero-oligomeric structures.[15,16] Metal-mediated multimerization is also widespread in metalloproteins.[17–20] In this work we have demonstrated the first instance of a ferritin whose assembly is mediated by iron-binding and the formation of the ferroxidase centre between EncFtn monomers. This assembly pathway of the EncFtn decamer, from non-FOC dimers, is distinct to other members of the ferritin family. While the FOC interface closely resembles the four-helix bundle FOC of the classical and bacterioferritins; the role of the FOC in stabilising the functional quaternary structure is unique to the EncFtn class of the ferritin-like proteins.

This work was financially supported by the BBSRC (Grant Number BB/N005570/1 and BB/R013993/1). J. R. is funded *via* the BBSRC EastBIO doctoral training programme (BB/M010996/1). T. L. is funded *via* an EPSRC PhD studentship and the School of Chemistry, UoE.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 F. Bou-Abdallah, *Biochim. Biophys. Acta, Gen. Subj.*, 2010, **1800**, 719–731.
2 X. Liu and E. C. Theil, *Acc. Chem. Res.*, 2005, **38**, 167–175.
3 D. He, S. Hughes, S. Vanden-Hehir, A. Georgiev, K. Altenbach, E. Tarrant, C. L. Mackay, K. J. Waldron, D. J. Clarke and J. Marles-Wright, *eLife*, 2016, **5**, e18972.
4 C. A. McHugh, J. Fontana, D. Nemecek, N. Cheng, A. A. Aksyuk, J. B. Heymann, D. C. Winkler, A. S. Lam, J. S. Wall, A. C. Steven and E. Hoiczyk, *EMBO J.*, 2014, **33**, 1896–1911.
5 F. Akita, K. T. Chong, H. Tanaka, E. Yamashita, N. Miyazaki, Y. Nakaishi, M. Suzuki, K. Namba, Y. Ono, T. Tsukihara and A. Nakagawa, *J. Mol. Biol.*, 2007, **368**, 1469–1483.
6 T. W. Giessen and P. A. Silver, *Nat. Microbiol.*, 2017, **2**, 17029.
7 M. Sutter, D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, N. Ban, A. Steven, E. Hoiczyk, N. Moriarty, R. Oeffner, R. Read, D. Richardson, J. Richardson, T. Terwilliger and P. Zwart, *Nat. Struct. Mol. Biol.*, 2008, **15**, 939–947.
8 W. F. Rurup, J. Snijder, M. S. T. Koay, A. J. R. Heck and J. J. L. M. Cornelissen, *J. Am. Chem. Soc.*, 2014, **136**, 3828–3832.
9 D. He, C. Piergentili, J. Ross, E. Tarrant, L. R. Tuck, C. Logan Mackay, Z. McIver, K. J. Waldron, D. J. Clarke and J. Marles-Wright, *Biochem. J.*, 2019, **476**, 975–989.
10 A. J. R. Heck, *Nat. Methods*, 2008, **5**, 927–933.
11 S. V. Iverson, T. L. Haddock, J. Beal and D. M. Densmore, *ACS Synth. Biol.*, 2016, **5**, 99–103.
12 E. Jurneczko and P. E. Barran, *Analyst*, 2011, **136**, 20–28.
13 E. G. Marklund, M. T. Degiacomi, C. V. Robinson, A. J. Baldwin and J. L. P. Benesch, *Structure*, 2015, **23**, 791–799.
14 E. Krissinel and K. Henrick, *J. Mol. Biol.*, 2007, **372**, 774–797.
15 H. J. Chen, T. P. Ko, C. Y. Lee, N. C. Wang and A. H. J. Wang, *Structure*, 2009, **17**, 517–529.
16 K. R. Malley, O. Koroleva, I. Miller, R. Sanishvili, C. M. Jenkins, R. W. Gross and S. Korolev, *Nat. Commun.*, 2018, **9**, 765.
17 A. C. Partin, T. D. Ngo, E. Herrell, B. C. Jeong, G. Hon and Y. Nam, *Nat. Commun.*, 2017, **8**, 1737.
18 M. M. He, S. L. Clugston, J. F. Honek and B. W. Matthews, *Biochemistry*, 2000, **39**, 8719–8727.
19 M. A. Kihlken, A. P. Leech and N. E. Le Brun, *Biochem. J.*, 2002, **368**, 729–739.
20 R. Dutoit, T. Van Gompel, N. Brandt, D. Van Elder, J. Van Dyck, F. Sobott and L. Droogmans, *J. Biol. Chem.*, 2019, **294**, 17777–17789.