The Implementation and Impact of Isotope Depletion for Improved Protein Mass Spectrometry.

Kelly Jane Gallagher

PhD in Chemistry

The University of Edinburgh

2020

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.

Kelly Gallagher

Abstract

One of the fundamental challenges of high-resolution mass spectrometry (MS) analysis of protein occurs as the molecular weight increases. This is due to the increased chemical complexity of the protein analyte, as larger molecular weight proteins display a wider isotopologue distribution caused by increased incorporation of heavier isotopes. This phenomenon increases the inherent complexity of the protein mass spectrum. In top-down fragmentation workflows, this is observed as densely packed spectral regions of fragment ions with heavily overlapping isotopologue distributions. It also associated with a reduction in fragment ion signal, all of which contributes to hinder ion assignment. One potential method to improve protein analysis via mass spectrometry is to apply an isotope depletion strategy. This involves the recombinant expression of protein within a defined growth media, containing carbon and nitrogen sources depleted in heavy isotopes (notably ^{13}C and ^{15}N). Thus, the resulting recombinant proteins are expressed with limited heavier isotope incorporation.

An efficient and cost-effective method for isotope depletion was developed in *E. coli* and applied to several proteins of increasing molecular weight. After optimisation, the recombinant protein was expressed typically containing isotope abundances of ^{12}C 99.95% and ^{14}N 99.99%. The benefit of applying isotope depletion was assessed using top-down fragmentation, mainly electron-based, techniques on both 12T FT-ICR and orbitrap Eclipse instruments.

The change to the protein mass spectrum, when using the isotopic depletion strategy, is observed as a simplified isotopologue distribution; an increased net contribution of the monoisotopic peak intensity and an increase of the fragment ion signal. In top-down fragmentation spectra, this results in reduced overlap in fragment ion distribution and greater mass accuracy due to the presence of the abundant monoisotopic peak. As a result of these accumulative benefits, top-down fragmentation of isotopically depleted protein allows assignment of 2-3 times greater number of fragment ions. This-in-turn allows attainment of higher sequence coverage, particularly when the isotope depletion strategy is used in conjunction with other processes designed to increase top-down sequence coverage, like proton transfer charge reduction.

The increased number of assigned ions in isotopically depleted protein fragmentation spectra also permits a dramatic reduction in the number of averaged transients,

without a corresponding reduction in the sequence coverage. Therefore the isotope depletion strategy facilitates the acquisition of higher top-down fragmentation sequence coverage in LC-MS/MS workflows.

As the isotope depletion strategy causes consistent simplification to the isotopologue distribution, it has the potential to improve a range of protein MS analysis. This was further demonstrated using isotopically depleted protein with native protein analysis, which maintains the higher-order structure of proteins within the gas phase. With the isotopically depleted protein, it was possible to extend the feasible working mass limit of native analysis of 12T FT-ICR.

Lay Summary

Elemental isotopes are variants of a particular element that contain different numbers of neutrons within the atomic nucleus. This means different isotopes of the same element, like carbon, have small differences in molecular mass. Protein are composed of amino acid building blocks which contain multiples of carbon, nitrogen, oxygen, hydrogen and sulfur. Each of these elements can contribute isotopes to the protein molecule. This can be observed on a mass spectrometer, as a grouping of peaks for the same protein with very small differences in mass. As the molecular mass of our protein of interest get larger, the number of elements from which it is composed increases. This results in an increase in the number of different protein isotope peaks observed. Making analysis of larger protein by mass spectrometry more difficult.

This thesis demonstrates a new method for producing the protein sample for the improvement of mass spectrometry, even as the mass of the studied protein of interest becomes larger and analysis gets more complicated. Called the Isotope Depletion strategy. This involves removing unwanted ‘heavier’ isotopes of carbon and nitrogen from the protein during production, known as isotopically depleted protein. Reducing the total number of protein isotope peaks which are present in the mass spectrum. To test the usefulness of the Isotope Depletion strategy, the isotopically depleted protein was directly compared to the same protein of interest, but with natural isotopes. So, both isotopically depleted and natural protein undergo identical mass spectrometry analysis. The comparison was carried out with well-known protein at different mass and used common techniques for the analysis of protein; such as protein fragmentation, or native analysis which looks at the protein in smaller pieces or as a structural whole. The reduced number of isotope peaks in the isotopically depleted protein consistently simplifies and therefore improves protein mass spectrometry analysis.

Abbreviations

Alpha-N-gluconolylation	**α-NG**
Bovine carbonic anhydrase	**BCA**
Collision induced dissociation	**CID**
Electron capture dissociation	**ECD**
Electron dissociation techniques	**ExD**
Electrospray ionisation	**ESI**
Electron transfer dissociation	**ETD**
Encapsulated ferritin	**EncFtn**
Encapsulin	**Enc**
Fourier transform ion cyclotron resonance	**FT-ICR**
Free induction decay	**FID**
Full width half maximum	**FWHM**
Infrared multiphoton dissociation	**IRMPD**
kilo-dalton	**kDa**
Liquid chromatography	**LC**
Mass spectrometry	**MS**
Mass-to-charge ratio	***m/z***
Molecular weight	**MW**
Nano electrospray ionisation	**nESI**
Proton transfer charge reaction	**PTCR**
Post translational modification	**PTM**
Serine palmitoyltransferase	**SPT**
Signal-to-noise ratio	**S/N**
Sophisticated numerical annotation procedure	**SNAP**

Thorough high-resolution analysis of spectra by Horn	**THRASH**
Time of flight	**TOF**
Ultraviolet photodissociation	**UVPD**

Acknowledgments

Firstly, I would thank you who have taken the time to read this thesis. Whether through commitment or curiosity, this encompasses four years of hard work and that anyone takes the time to read it is incredible. I hope it doesn't disappoint.

The work carried out for this thesis were not conducted in a vacuum, so I must thank the various members of the Clarke group. Past members Sam and Lulu, for setting the example and being so supportive in helping me celebrate the successes and learn from the failures. A special thank you must be made to Jenn, who has been with me throughout the entire PhD process. Without whom I doubt I would have survived to reach this point. John White is also deserving of a special thank you for his advice, friendship and constant support throughout my PhD.

To the other members of the Clarke group; Lavrentis, Tom, Maria and Kiani this experience wouldn't have been the same without you all! Similarly, without Faye and Logan I imagine the mass spectrometry suite would have fallen to ruin a long time ago. It would also be truthful to admit that without Dave this project never would have happened. But I try not to hold it against him. None of the successes would have been possible either, so he definitely deserves many thanks for all his help and endless optimism.

Thanks to Janice, Heather and Martin, none of which directly impacted my project but have all been directly involved my training and ensured my demonstrating and teaching experiences were fun and interesting.

Similar thanks must be conveyed to my partner Duncan, my parents, my grandparents and siblings. None of them understand what I do, but they have always tried to be supportive and remind me that there is more to life than just work. Particularly during this long lockdown period, I would have been emotionally broken without them.

Much like any acknowledgements, I am sure there several more people who should be named here but are not. Fortunately for me it is very unlikely that any of them will ever read this. I hope I conveyed my gratitude for their help sufficiently at the time.

-Kelly Gallagher

Table of Contents

List of figures

List of Tables

1. Introduction

1.1 Mass Spectrometry

1.1.1 Biomolecular Mass Spectrometry

Mass spectrometry (MS) is a highly accurate analytical technique which is utilized worldwide in a wide range of applications and research. In its most simple sense, mass spectrometry is composed of an ionisation source which transfers the sample into the gaseous phase, separation of different ions as a result of their varying *m/z* and ultimately their detection. Biological samples can pose a range of challenges for analysis in the gaseous phase. The application of which only became feasible with the development of ionisation techniques, such as, electrospray ionisation (ESI) [1,2] and matrix-associated laser desorption ionisation (MALDI). [3,4] Facilitating the rapid improvement in biological MS analysis, which subsequently developed a fundamental role for structural and high throughput proteomic investigations.

1.1.2 Electrospray Ionisation

ESI is a conversion process of a liquid into a charged aerosol and is considered a 'soft' ionisation technique which can protonate non-volatile biomolecular samples, such as protein, without inducing fragmentation. The process, shown in Figure 1.1, produces ions in the form of $[M+nH]^{n+}$ where n can be a range of charges deposited on the protonated polypeptide.

A typical experimental set-up for ESI involves a sample inlay capillary and a counter electrode spaced around 2-3 cm away. A voltage, typically around 4 kV, is applied that creates a surplus of positive ions towards the solution meniscus. This distorts the solution into a Taylor cone [5] which will emit a spray of charged droplets from its' apex. The produced droplets decrease in size as they travel towards the negative electrode due to a combination of solvent evaporation and coulombic repulsion [6] where the coulombic force trying to keep the positive charges of the droplet spatially separate reaches a charge density sufficient to overcome the droplet surface tension, known as the Rayleigh limit [7] and the charged droplets undergoes fission events producing

daughter droplets of reduced diameter. Successive fission events will ultimately release ions into the gaseous state to enter the mass analyser.

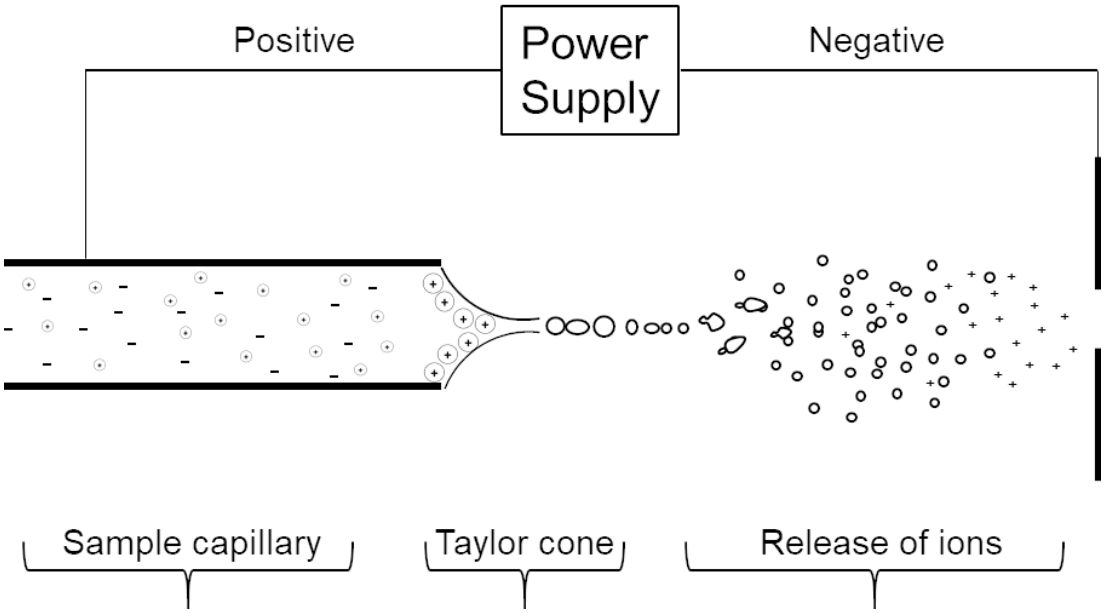


Figure 1.1: Schematic diagram of the electrospray ionisation process.

Application of a potential difference between the sample capillary and the electrode for ion entry distorts meniscus of the analyte solution into a Taylor cone which ejects charged droplets which reduce in size until the ions are released into the analyser.

Recreated from Kebarle and Verkerk. [7]

The release of analyte ions into the gaseous phase is described by three different release models. Smaller molecules are believed to migrate to the edge of the charged droplet, and the electric field promotes ejection of the analyte with charge once the droplet is sufficiently small (<10 nm), [8] this is known as the ion evaporation model (IEM) which is demonstrated in Figure 1.2a. In the IEM a single droplet will contain multiple small molecules which are emitted from the droplet. For intact globular proteins two competing mechanisms have been proposed. The first, the charged residue model (CRM) suggests that a single (often folded) protein is left inside the charged droplet which evaporates to dryness, leaving the charged folded protein in

the gaseous state (Figure 1.2b). In contrast, unfolded protein has been proposed to be ejected via the chain ejection model (CEM).[8] In this process, as higher order structure has been disrupted due to solvent conditions, residues which are normally buried within the folded structure become solvent accessible in the unfolded structure. As the internal residues of the folded polypeptide tend to be largely hydrophobic, these result in the polypeptide chain being forced to the to the edge of the droplet. Finally, much like the IEM, the Rayleigh limit causes ejection of the polypeptide chain with a wide range of charge states. [9]

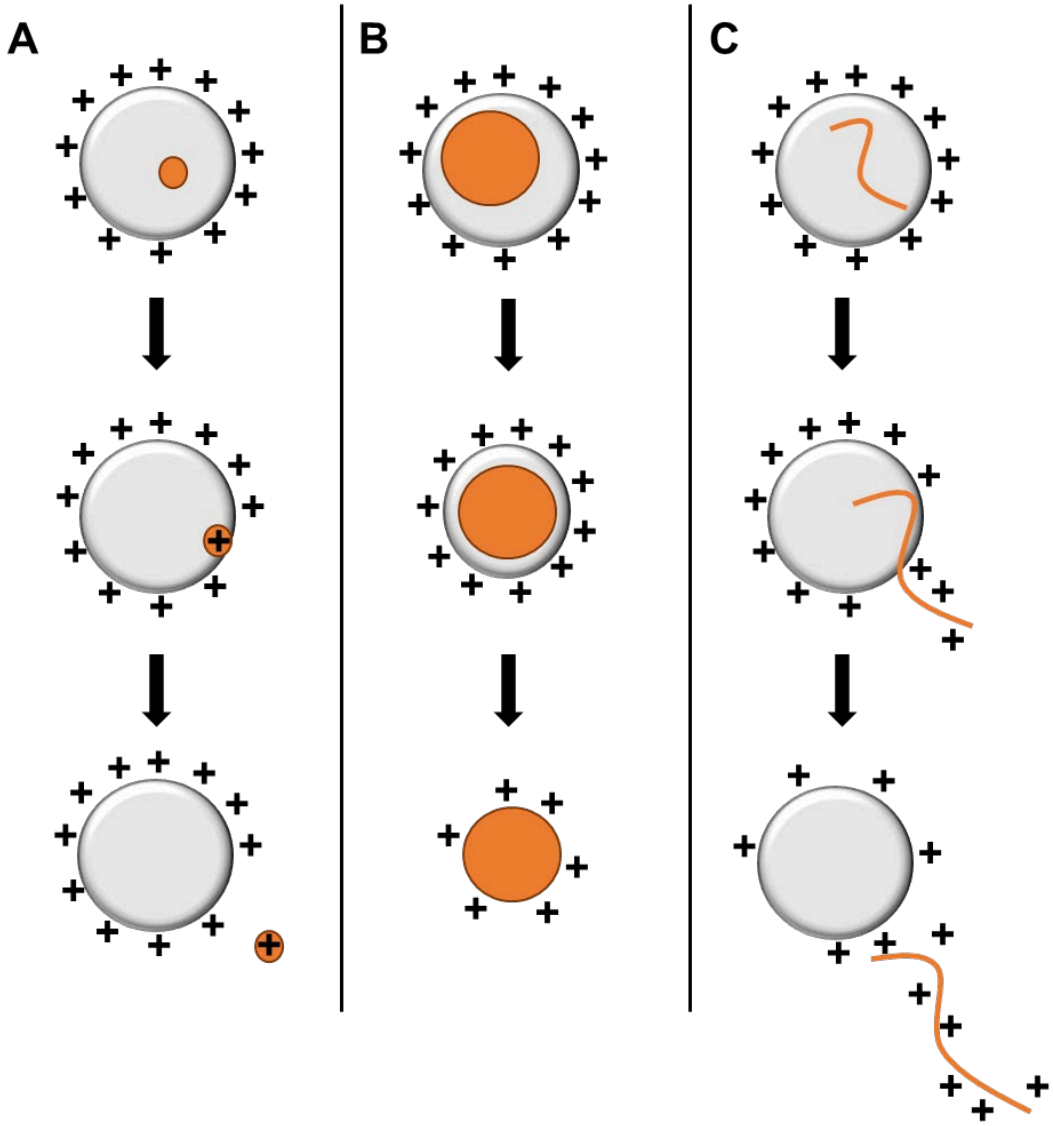


Figure 1.2: Mechanisms of ion transfer into the gaseous phase with electrospray ionisation.

Outline of the different electrospray mechanisms. Ion evaporation model (IEM) for small molecules (A), charged residue model (CRM) for folded protein (B) and the chain ejection model (CEM) for unfolded protein (C).

Recreated from Konnermann et al. [9]

Nanoelectrospray (nESI) [10,11] is a modified variation of ESI and follows the same ion ejection mechanisms. However, in nESI the ion is emitted from a capillary with a fine tip (~1 µm) producing nanoscale droplets which go through multiple fission events as

described above. The nESI is conducted typically lower at a lower voltage (~1 kV) resulting in a vastly lower flow rates and sample demand. [12] Similarly, the reduced initial droplet size is less likely to contain high salt concentrations, increasing sensitivity. This combined with lower temperatures, the absence of a nebuliser gas or sheath flow contribute to the maintenance of non-covalent interactions and higher order structure, if desired. Therefore, nESI is the preferred delivery methods for native MS structural MS of proteins and protein complexes.

1.1.3 Liquid chromatography

Liquid chromatography (LC) is a separation technique which separate analytes based on their affinity of a stationary solid phase and a liquid mobile phase. LC coupled to ESI is a rapid delivery method into a mass spectrometer, and LC-MS analysis is a common analytical technique. Reverse phase chromatography is most often used, with the exact hydrophobicity of the stationary phase tailored for the specific analyte. For example, the less hydrophobic C4 stationary phase will have weaker interactions with the analyte, so is better suited for the analysis of intact proteins. Conversely longer chained (more hydrophobic) C18 is better suited for small molecule, and protein digest or smaller peptide analysis.

In the context of biomolecular analysis, LC-MS has the added benefit of exchanging the analyte from a non-MS compatible buffer into the solvents required for efficient ESI. [2] During the electrospray process, the presence of physiological buffers, which contain millimolar concentration of non-volatile inorganic salts, leads to extensive adducted ion species (e.g. Na^+ and K^+ ion adducts) on the protein analyte. Consequently, in the resulting mass spectrum the intensity of the non-adducted ion is vastly reduced and an extended series of adduced species dominate the protein mass spectrum often having overlapping isotopologue distributions which complicate

analysis. [13,14] These issue are efficiently overcome during LC-MS analysis, as inorganic salts travel through the reverse phase LC column have little interaction with the stationary phase, and so are eluted first. Leaving the protein, now unfolded, to be eluted in the LC solvent mixture (acetonitrile: water: formic acid).

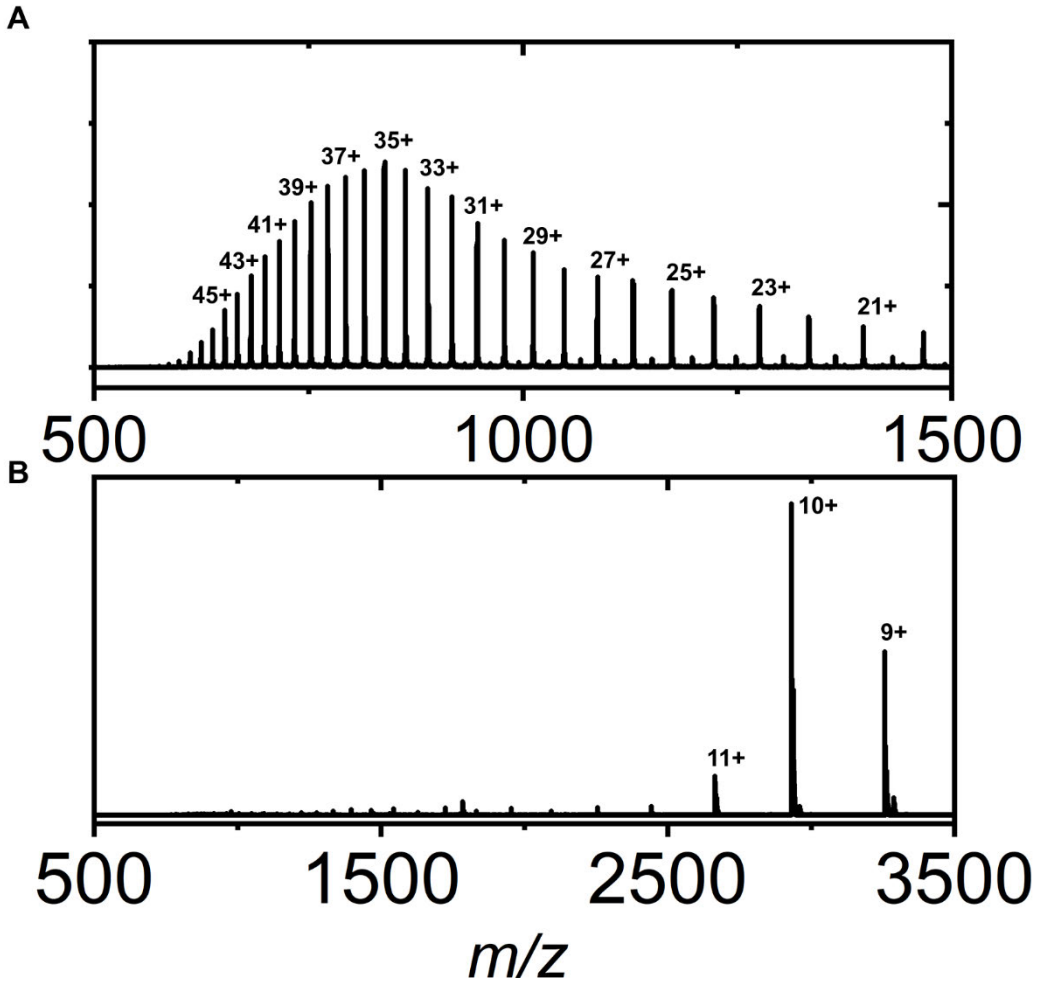


Figure 1.3: Example of protein charge state distribution

ESI of an unfolded (30 kDa) protein (A) has the potential to acquire a wide range of charges as it enters the gas state. Producing a charge state distribution that occupies a wide range of the m/z spectrum. Under native conditions, the same protein (B) displays a reduction in the total number and variation of the acquired charges.

ESI of a single intact unfolded protein, after LC chromatography (and other sample preparation) produce recognisable charge-state distributions. Displaying the same

protein across multiple m/z values with a range of acquired charges, as shown in Figure 1.3a. Native ESI of the same protein, acquires fewer numbers of charges and displays a reduced variation in charge (Figure 1.3b).

1.1.4 The isotopologue distribution

The isotopologue distribution is a natural observed phenomenon composing the individual charge states (Figure 1.3), as a result of the natural isotopic heterogeneity of the elements. Isotopes have the same number of protons and electrons, but a different number of neutrons so therefore contribute a different total mass to the polypeptide. Meaning that the isotopologue distribution shows the range of incorporation of heavier isotopes within a polypeptide chain, observed as the regularly spaced isotopologue peaks. For carbon, the most abundant isotope (^{12}C) exists with 98.93% abundance, and the heavier stable isotope (^{13}C) is naturally present at 1.07% abundance (Table 1.1). [15] In a peptide mass spectrum the monoisotopic peak (M) displays the lowest mass and is composed entirely of the lowest mass isotope (^{12}C). The next peak in the isotopologue distribution has the addition of approximately 1 Da due to an addition of a single ^{13}C isotope (M+1). Similarly, the addition of two ^{13}C isotopes results in a mass increase of approximately 2 Da and is observed as the third isotopologue peak in the distribution (M+2). In addition to carbon, all the organic elements present in proteins naturally exist with different isotope abundance (Table 1.1), all of which can contribute to the isotopic distribution of the peptide in a statistically predictable manner. [16,17]

Table 1.1 Natural isotope abundance of common protein elements

Element	Isotope	% Abundance	Mass	Increase in isotope mass (Da)
Hydrogen	^{1}H	99.99	1.007825	
(Deuterium)	^{2}H	0.01	2.014102	1.00628
Carbon	**^{12}C**	**98.93**	**12.00000**	
	^{13}C	**1.07**	**13.00336**	**1.00336**
Nitrogen	**^{14}N**	**99.63**	**14.00307**	
	^{15}N	**0.37**	**15.00011**	**0.99704**
Oxygen	^{16}O	99.76	15.99492	
	^{17}O	0.04	16.99913	1.00422
	^{18}O	0.21	17.99916	2.00424
Sulfur	^{32}S	94.93	31.97207	
	^{33}S	0.76	32.97146	0.999387
	^{34}S	4.29	33.96787	1.995796

Peptides and proteins contain ten to thousands of atoms, and the probability of heavy stable isotope incorporation is the same at each occurrence of that atom. [16,18] Although the percentage abundance of the heavier isotopes shown in Table 1.1 is relatively low, the probability of heavier isotope incorporation into a polypeptide chain increases dramatically with the molecular weight. Figure 1.4 is a simulated model showing the 'spreading' effect of increasing molecular weight on the isotopologue distribution. Based on the repeating theoretical averagine amino unit [19,20] (n[$C_{4.9384}H_{7.7583}O_{1.4773}N_{1.3577}S_{0.0417}$]) which is the average of each element observed across the twenty natural amino acids.

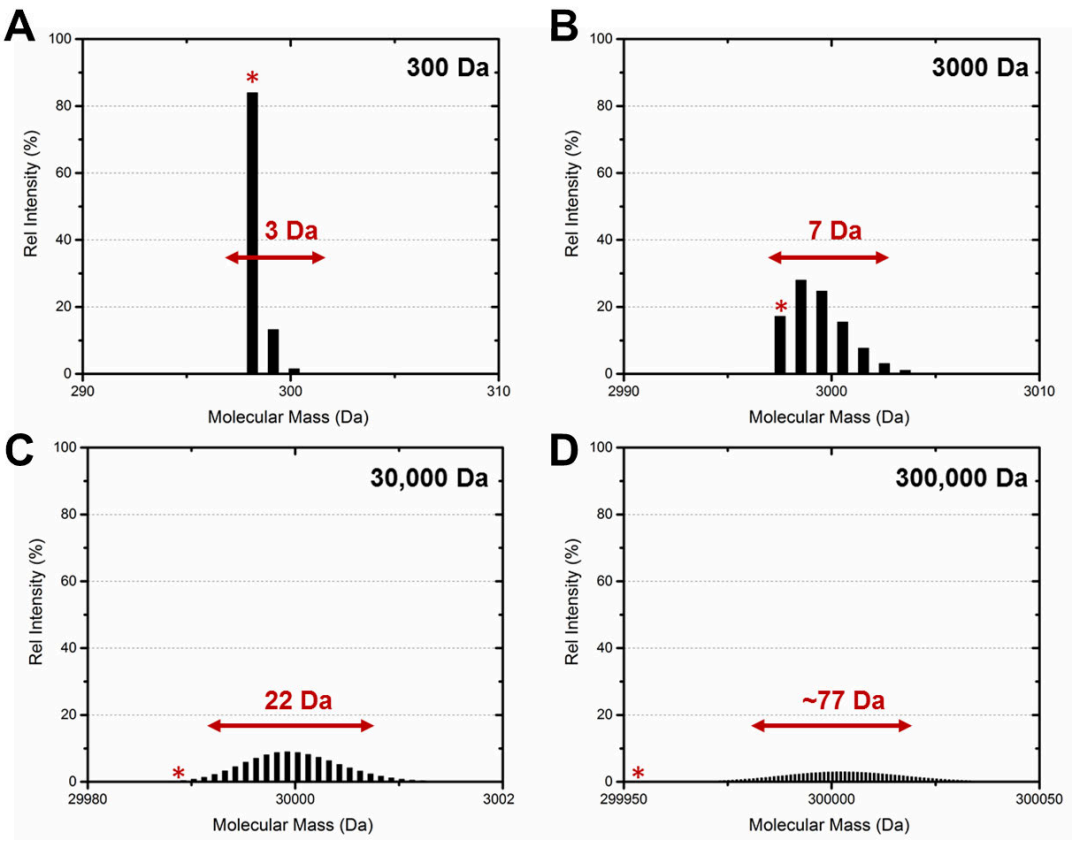


Figure 1.4: Isotopologue distribution increases with size

Increasing molecular weight of a protein from 300 Da (A), in stages (3000 Da (B) and 30,000 (C)) to 300,000 Da (D) results in an increased atomic complexity of the molecule. This increases the probability for incorporation of heavy isotopes. Observed as an increased number of isotopologue peaks within the isotopologue distribution spread over a greater mass range. As a result, the percentage intensity of the most abundant peak in the distribution becomes lower as the total signal of the distribution is spread amongst each of the isotopologue peaks in the distribution.

At low mass (300 Da), the isotopologue distribution consists of three peaks and is therefore 3 Da wide (Figure 1.4a); however at high mass the distribution will consist of many isotopologue peaks, for example 300,000 Da the width is approximately 77 Da (Figure 1.4d). As a consequence, there is an observed reduction in S/N as the total signal of the ion is distributed between an ever-greater number of isotopologue peaks. The monoisotopic peak (*) is the most abundant isotopologue peak for Figure 1.4a, is observable but not the most abundant peak in Figure 1.4b and is not observable in the isotopologue distribution of either Figure 1.4c or Figure 1.4d. The monoisotopic peak for polypeptides around ~13 kDa (Figure 1.6) is still statistically likely at low abundance, but is often not observed above the noise floor in the acquired spectra. As the protein continues to increase in molecular weight (Figure 1.4c or Figure 1.4d) the predicted monoisotopic mass isotopologue is out-with the range of the observed distribution (Figure 1.4d). Resulting in a measured average mass of the protein which can be significantly larger than the accurate (monoisotopic) mass.

1.1.5 The isotopic fine structure of protein

Mass defect is defined as the difference in the monoisotopic mass from nominal mass. This is as a result of the binding energy which is released as the constituent parts (protons and neutrons) formed the atom. [21] For each element the mass defect is a result of different required binding energy, and becomes particularly relevant for this study in the removal of heavier stable isotopes. As shown in Table 1.1, the difference in mass between the heavier stable isotope and the most abundance isotope does not produce the same increase in mass at a rounded integer (e.g., 1 or 2 Da). The result of this is displayed in Figure 1.5 and is known as the isotopic fine structure of the isotopic distribution. The modelled small peptide ($C_{69}H_{114}N_{20}O_{26}S_1$, 1670.79 Da) (Figure 1.5a) display 6 isotopologue peaks, the most abundant of which are M, M+1 and the M+2 peaks. As previously described, the M+1 peak is formed as a result of a

single heavy isotope contributing an extra ~1 Da to the peptide mass. This mass can contributed by ^{15}N (+0.99704 Da), ^{33}S (+0.99939 Da), ^{13}C (+1.00336 Da) or ^{2}H (+1.00628 Da) [15,22,23] as shown in Figure 1.5b. The separate abundance contributed by each isotope all contribute to the total abundance of the M+1 peak in Figure 1.5a. Similarly, the M+2 peak is the result of a single isotope contributing ~2 Da, or two isotopes which both contribute ~1 Da. Due to the number of carbon atoms in the peptide and the natural abundance of ^{13}C (1.07%) therefore it has the greatest contribution in both Figure 1.5b and Figure 1.5c.

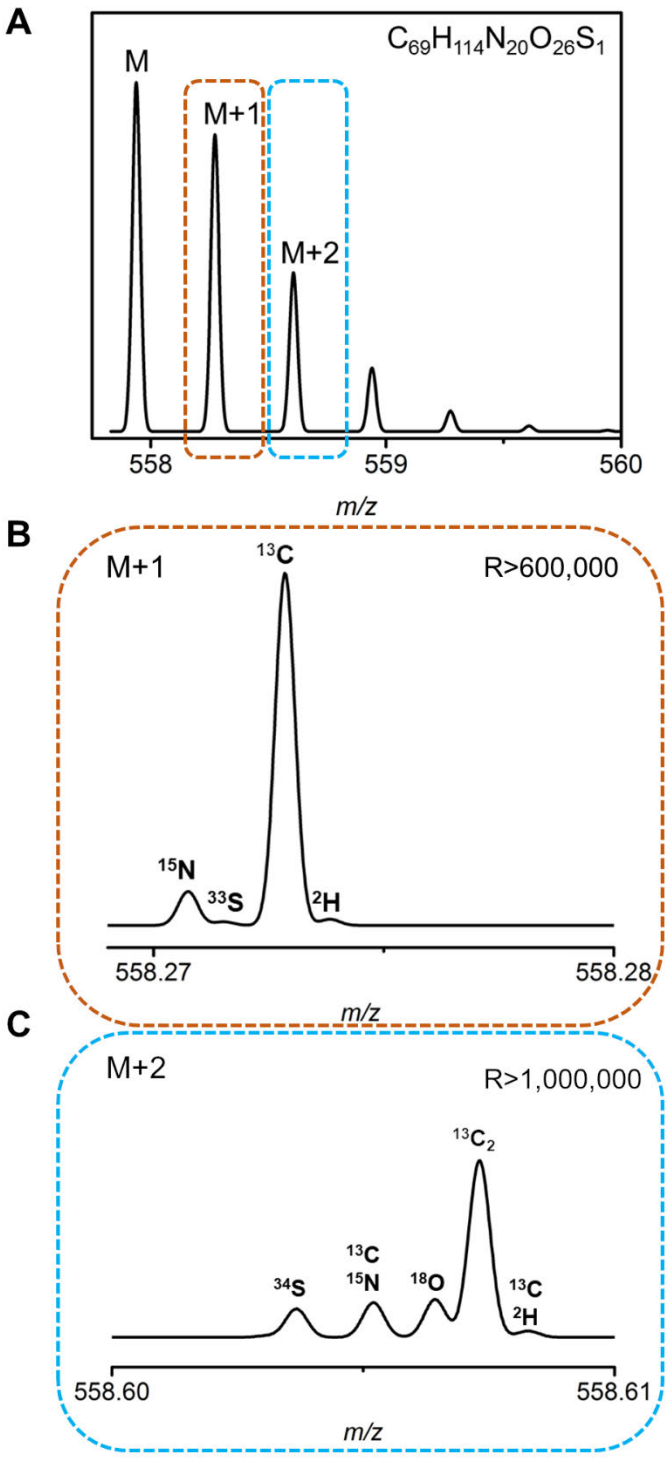


Figure 1.5: Isotopic fine structure of a small peptide

The theoretical model of a small peptide $[C_{69}H_{114}N_{20}O_{26}S_1]^{3+}$ (A) and a closer view of both the M+1 (B) and the M+2 peak (C) to observe the isotopic fine structure of the peptide.

In order to be able to observe the isotopic fine structure of this small peptide displayed in Figure 1.5, a high-resolution acquisition is required (typically >600,000). Predictably, as the mass of the peptide increases it becomes less likely that isotopic fine resolution will be sufficiently resolved, and combination of heavier stable isotope incorporation will increase in complexity [24]. Producing the statistical array of isotopologue peaks we observe in the isotopologue distribution (Figure 1.4).

1.1.6 Mass Resolution and Resolving Power

The mass resolution of a mass spectrometer is the ability to differentiate between two neighbouring peaks on a mass spectrum. More commonly measured, the resolving power of a mass spectrometer is defined as $m/\Delta m$, the full width half maximum (FWHM). [25] The narrower the width of the distinct peaks, the greater the distinction which can be made. This is demonstrated in Figure 1.6, modelled on the chemical structure of a small 13 kDa polypeptide chain $(C_{572}H_{884}N_{172}O_{185}S_2)^{10+}$. As the resolution increases the spectra modifies from a single peak Gaussian distribution (blue line) to the isotopologue distribution (black line), but both represent the same protein distribution. At the lowest resolution (blue line) only the average mass of the peak can be measured and has a larger FWHM meaning that if two peaks at similar *m/z* and overlapping may appear as a single extended or humped distribution making assignment more difficult. As the resolution increases to the highest resolution (black line) the monoisotopic peak can be identified allowing for accurate measurement of the polypeptide. Each separate isotopologue peak of the isotopic distribution are regularly spaced peaks and can provide information on the charge state or separate multiple distributions which partially overlap, so can be distinguished and assigned separately.

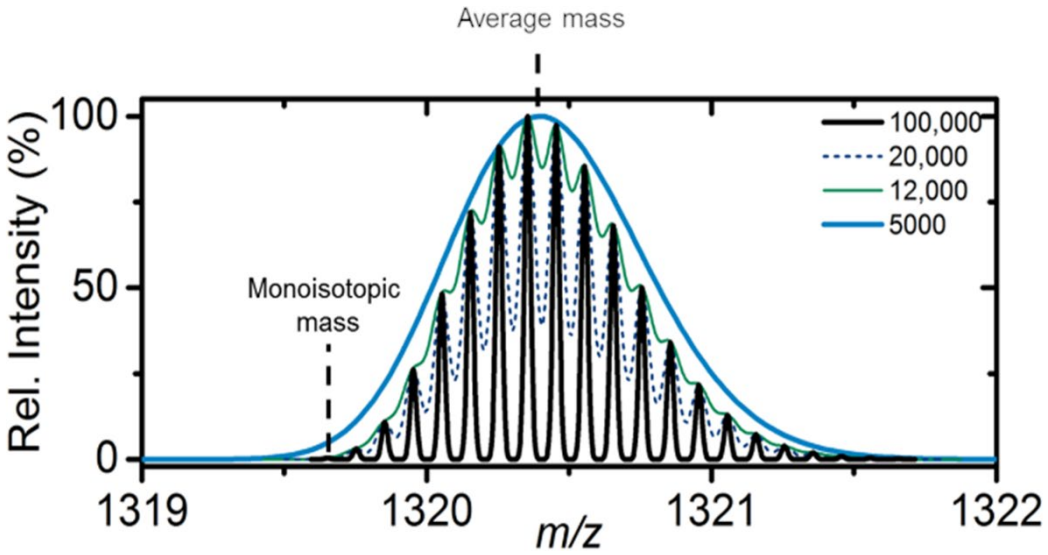


Figure 1.6: Varying resolution of a polypeptide

At varying resolution, the information which can be extracted from a spectrum differs. Lower resolution (5000, blue line) produces wider peaks from which an average mass can be calculated. As the resolution increases (12,000, green line; 20,000, dashed line) the isotopologue peaks become observable. Further increase in resolution (100,000, black line) is sufficient to fully resolve the isotopologue distribution of the polypeptide.

The resolving power achieved by an instrument can be limited by a number of different factors, such as the type of mass analyser used, the mass of the analyte, the number of charges an analyte carries, and the presence of attached salt adducts. Adducts like sodium and potassium ions are the result of trace salt concentrations in the ESI buffer. It is present in the droplets and associate with the protein molecules as they enter the gas state (Figure 1.2). They cause an increase to the total mass of the protein and create multiple smaller secondary distributions which will significantly overlap with the original protein distribution. This is a particular issue for larger mass species, as the distribution for both the protein and the adduct may be observed as a single, overlapping distributions in the spectrum. This can result in a calculated average mass

that would not be an accurate reflection of the mass of the protein species and will ultimately limit the analysis of the protein if undetected.

1.1.7 Time of Flight Mass Analysers

Mass spectrometers have developed greatly in achievable resolution since first developed by J.J Thomson. [26–28] The highest resolution commonly used for protein analysis is achieved by time-of-flight (TOF), FT-ICR and orbitrap analysis [29] TOF analysers first involves acceleration of the ions across a potential difference into a region without an electric field (the drift region). This acceleration results in the transfer of kinetic energy (K.E. = $\frac{mv2}{2}$) to the ions. The velocity (v) of the ion once it reaches the drift region is controlled by the *m/z* of the ion. If multiple ions of the same *m/z* have varying kinetic energy as they move from the accelerated region to the drift region, they will have different velocities and so reach the detector at different time. This

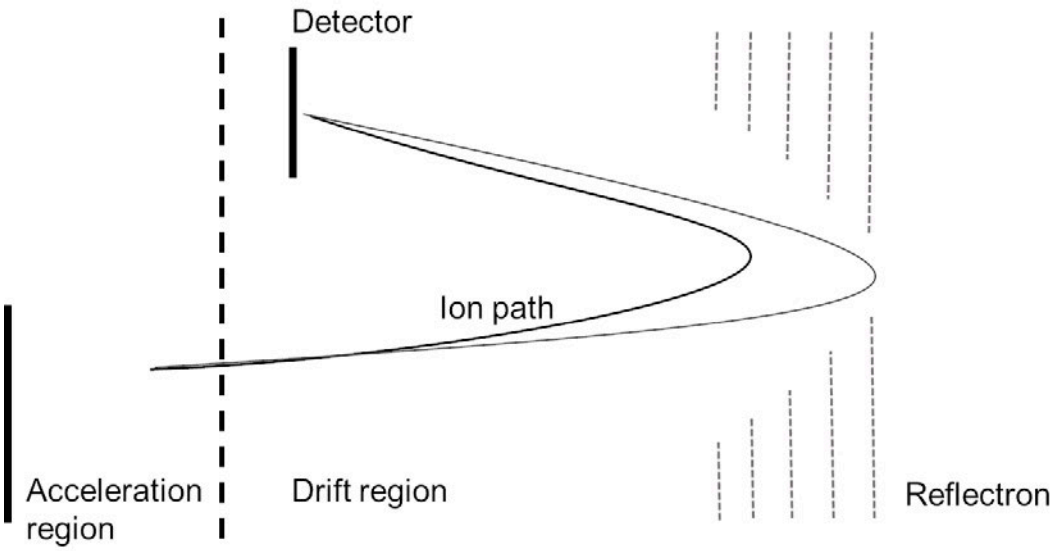


Figure 1.7: Time-of-flight mass analyser with a reflectron.

Charged ions are accelerated into a drift region with a constant kinetic energy. The same ion with different kinetic energies will travel through the drift region with different velocities. The reflectron will deaccelerate the ions and reflect their trajectory towards the detector. Ions of greater velocity will penetrate further into the reflectron electrostatic field but travels faster and so will reach the detector at the same time.

would reduce the resolution of the MS detector. [29] This issue is resolved using an electrostatic ion mirror, shown in Figure 1.7 and referred to as a reflectron. [30]

As the ions of the same *m/z* but different KE move from the acceleration region to the drift region they travel towards the reflectron, which deaccelerates the ion and directs it towards the detector. The ions with greater velocity travel further into the reflectron, but all ions with the same *m/z* will ultimately reach the detector at the same time. [29] The reflectron provides a second benefit for the resolution of TOF analysers, which is dependent on the length of the ion flight time (t_f). The reflectron increases the length of the ion path, without requiring an impractically large mass spectrometer. The *m/z* of a given ion after acceleration through the acceleration region (Figure 1.7) of length s by a voltage E is calculated as a measure of the ion flight time (t_f) through a drift region of distance x. This is described by Equation 1, however for a given TOF instrument E, s and x are typically constant for a given instrument. Therefore Equation 1 can be simplified to Equation 2, using k as a value for an encapsulating constant.

$$m/z = \frac{2t_f^2 Es}{(2s + x)^2}$$

Equation 1

$$m/z = kt_f^2$$

Equation 2

For larger biomacromolecule analysis TOF analysers are powerful tools, but often are insufficient to observe fully resolved isotope peaks. For this Fourier transform ion cyclotron resonance (FT-ICR) or orbitrap analysers are required.

1.1.8 Fourier transform ion cyclotron resonance mass analysers

FT-ICR mass spectrometers were first described by Comisarow and Marshall in 1974. [31,32] In FT-ICR, the *m/z* of an ion is calculated using the measured cyclotron frequency (ω) imparted on the ion by a strong magnetic field. Typically for protein samples, (n)ESI is the used technique to produce gaseous ions and these are accelerated through stages of increasing vacuum, until contained within a Penning trap detection cell [33] held within a helium cooled magnet. The movement of the ions at this point are considered to move within the spatially uniform magnetic field (**B**), inside the voltage gated cell which contains the ions. Newton's second law of motion (Equation 3) in an electromagnetic field can be rewritten to produce the left section of Equation 4, also known as the Lorentz force. In which q and v are charge and velocity respectively. [34] As ions are directed in the *xy* plane, they are perpendicular to the magnetic component of the Lorentz force, causing the ion to have angular acceleration, which is described in the right section of Equation 4 (highlighted in **bold**).

$$F = mass\ x\ acceleration$$

Equation 3

$$F = qv_{xy}B = m\frac{\boldsymbol{v_{xy}^2}}{\boldsymbol{r}}$$

Equation 4

Considered as two-dimensional, angular velocity ($(\omega = \frac{v_{xy}}{r}$), Figure 1.8b) can be used to simplify Equation 4, and produce the angular cyclotron frequency (ω_c) equation (Equation 5).

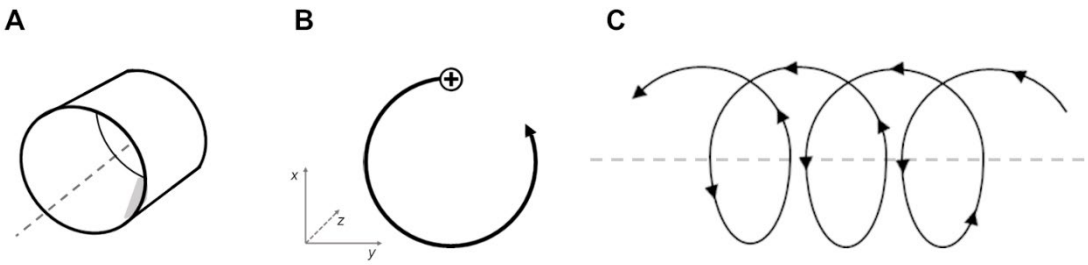


Figure 1.8: The movement of ions in an electromagnetic field.

*As a result of the magnetic component of the Lorentz force an ion trapped within an ICR cell (A), moving in the xy plane, will also display angular velocity (B), resulting in circular motion around the xy plane (C) with a defined radius (**r**).*

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

Equation 5

An important aspect of Equation 5, is that in a constant magnetic field all ions of a given *m/z* will display the same angular cyclotron frequency. [32,34] Regardless of the ion velocity obtained during ion acceleration in the ICR cell. It must be noted that this (Equation 5) is a simplified outlook of the ion movement and does not account for the trapping voltages at either end of the ICR which maintains the oscillation of the ions inside the ICR cell (Figure 1.8c). The trapping voltages on the ICR cell combats the Lorentz force and results in axial movement in the z plane (ω_z, Equation 6). In which V_{trap} is the trapping potential, α is a constant of the ICR cell geometry and *a* is the separation of the trapping electrodes.[35]

$$\omega_z = \sqrt{\frac{2qV_{trap}\alpha}{ma^2}}$$

Equation 6

As the contribution of ω_z acts to reduce the cyclotron motion, what is actually measured is the reduced cyclotron motion (ω_+). Equation 7 displays the relationship between the cyclotron motion (ω_c) and the axial movement (ω_z) to calculate ω_+. It can also show the relationship between ω_c and the magnetron frequency (ω_-), which is significantly smaller than ω_+ and rarely measured.

$$\omega_{\pm} = \frac{\omega_c}{2} \pm \sqrt{(\frac{\omega_c}{2})^2 - \frac{\omega_z^2}{2}}$$

Equation 7

Ions that enter into and are trapped in the ICR cell display a reduced cyclotron motion (ω_+). It is important to note that ω_+ will not produce a detectable signal within the ICR cell. In order for the ions to be detected, they are required to be excited to a greater cyclotron radius through application of an oscillating electric field ($E_{(t)}$, Equation 8).

$$E(t) = E_0 cos\ \omega_c\ tj$$

Equation 8

$$E_0 = \frac{2V_0}{d} = \frac{V_{p-p}}{d}$$

Equation 9

$$E(t) = E_R(t) + E_L(t)$$

Equation 10

Assuming E_0 with the application of $+V_0$ and $-V_0$ across parallel conductive plates separated by a distance of *d* meters (Equation 9), the peak-to-peak voltage difference between the excitation plates (V_{p-p}) pushes the ions to greater cyclotron radius. The

oscillating radiofrequency is composed of two alternating components (Equation 10) $E_R(t)$ rotating in the same sense as the positive ions, 'in resonance' so pushes the ion. The opposite component $E_L(t)$, rotates in the opposite direction and is off-resonance, and so has a negligible impact on the motion of the ion. The post-excitation radius (r) is calculated using Equation 11, and is independent of the ion *m/z.* Therefore, all ions can be excited to the same post-excitation radius during the radiofrequency sweep.

$$r = \frac{V_{p-p} T_{excite}}{2dB_0}$$

Equation 11

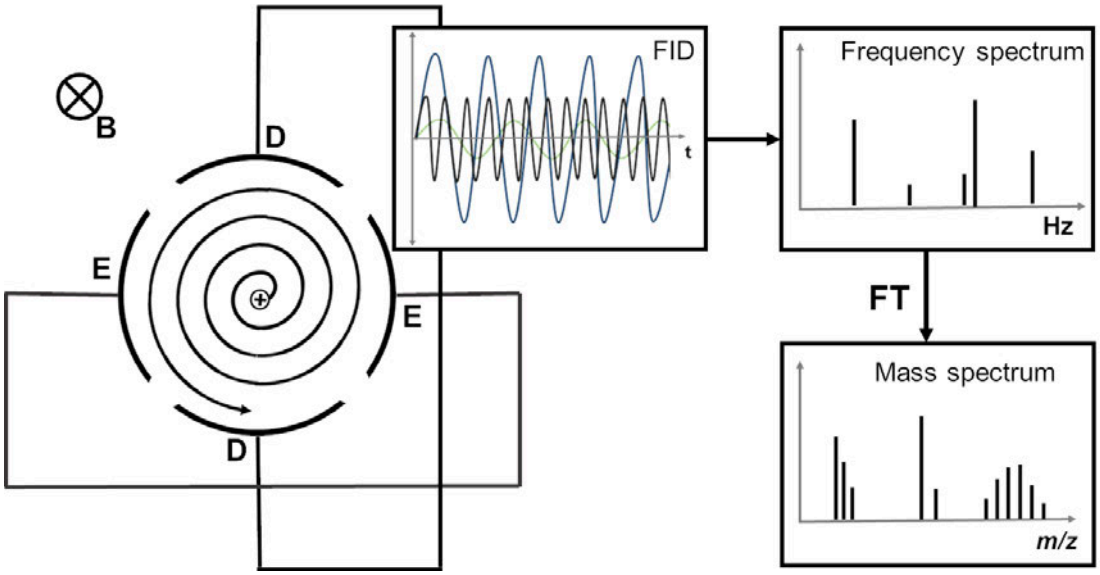


Figure 1.9: Detection of ions within an ICR cell

Ions are excited with a radiofrequency using the excitation plates (E) to a greater radius. At this point the ions will induce a current across the detection plates (D) to produce a complex overlap of waveforms for each m/z ion present, known as a free induction decay (FID). The Fourier transform function (FT) can separate each waveform in the FID and can calculate its frequency, which can then be converted into the m/z of the ion.

As well as the parallel plates which apply the oscillating radiofrequency and excite (E, Figure 1.9) the ions, the cell contains also contains two detector plates (D, Figure 1.9). Ions move around the cell with the same post-excitation radius (Equation 11) but with different velocity. Detector plates have an induced current for each separate *m/z*, all of which overlap to produce a complicated overlap of signal known as the free induction decay (FID, Figure 1.9). The Fourier transform can separate each waveform of the FID and calculate the frequency of each set of ions. Which can be related to the *m/z* of the ion using Equation 5. Due to the accuracy with which the current can be measured, FT-ICR provides very high-resolution mass spectra. This is further improved by the distance the ion travels during the acquisition. As in time of flight, increased travel time increases resolution and during a 1 second acquisition an ion can travel several kilometres. [34]

1.1.9 Orbitrap mass spectrometry

Orbitrap analysis, like FT-ICR, is a high resolution technique which is a development of the Kingdon ion trap.[36] A cylindrical ion trap with a central wire, and a voltage applied to both promotes elliptical oscillation of ions accelerated into the trap. Similarly, the orbitrap (Figure 1.10) is composed of a central spindle electrode encompassed by a larger, barrel electrode. Ions are introduced parallel and off-centre to the *z* axis and are trapped between voltages applied across the electrodes. Ensuring the ions move away from the outer electrode and rotate the central spindle. [37]

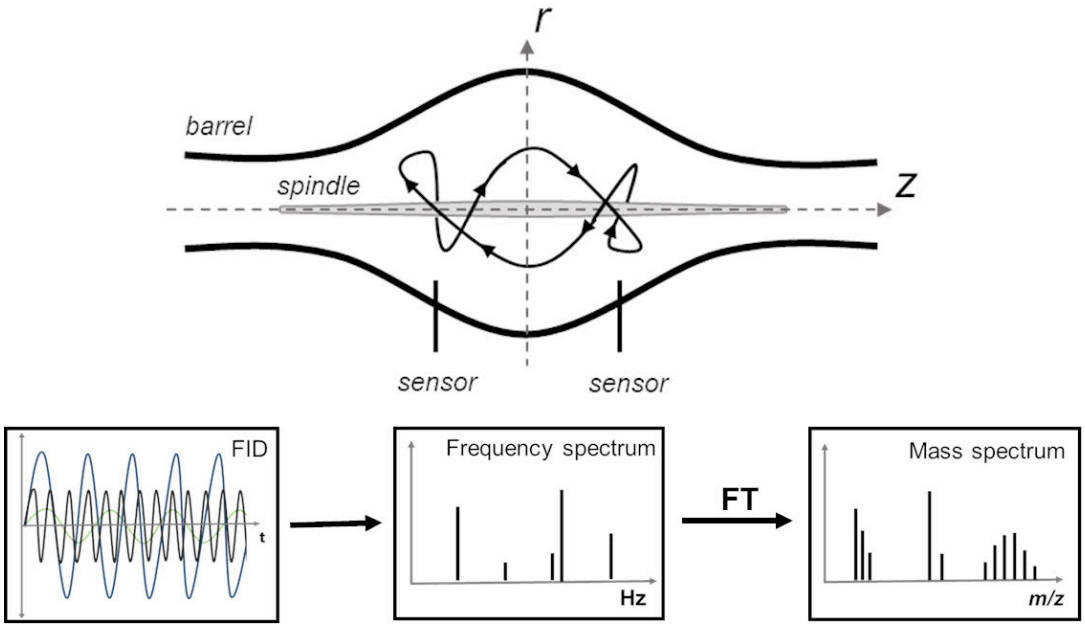


Figure 1.10: Oscillation of ions within the orbitrap detector

Ions are introduced to the orbitrap parallel to the z-axis and held between the voltages applied to the central spindle and outer barrel. The axial frequency (ω) of the ion along the z-axis is dependent on the m/z of the ion and detected as a free induction decay (FID) and the axial frequency identified using a Fourier transform. From this the m/z ratio of the present ions can be calculated and shown within a mass spectrum.

$$\omega = \sqrt{k \Big/ \left(\frac{m}{z}\right)}$$

Equation 12

$$z(t) = z_0 \cos(\omega t) + \sqrt{\left(\frac{2E_z}{k}\right)} \sin(\omega t)$$

Equation 13

The ions display a complicated axial rotation around the spindle electrode (Equation 13), but fortunately the measured axial frequency (ω) is primarily the result of the *m/z* of the ion (Equation 12). [37] Similarly to FT-ICR, the movement of the ions are detected as they approach the sensors (Figure 1.10) generating a complex FID, identifying the axial frequency (ω) using the Fourier transform and using Equation 12 can calculate the *m/z* ratio of the ions within the detector.

As previously mentioned, FT-ICR and orbitraps are the highest resolution techniques currently available for MS analysis. Both involve the trapping of ions, held within electrostatic fields at high vacuum and analysis through ion current detection to produce an FID and converted to a mass spectrum. FT-ICR was developed decades prior to the orbitrap, [32,37] and the achievable resolution is dependent on the mass and charge of the ion and the strength of the magnetic field (Equation 5). Therefore, resolution is inversely related to the *m/z,* but can be improved with a greater acquisition time or increasing the ion cyclotron motion (ω_c) by having a larger field strength magnet. The largest of which worldwide is currently 21T. [38,39] In an orbitrap as there is only an electromagnetic field the resolution does not decrease as rapidly with ion *m/z* (Equation 12). Thus analysis of the same sample for the same length of time (FID) below a critical *m/z* value, FT-ICR will provide the greatest resolution and above which the orbitrap will be superior. [40]

Using high-resolution mass analysers when carrying out analysis of a biomacromolecule permits visualisation of the separate isotopologue peaks (Figure 1.6), and is normally coupled to a range of MS^2 and MS^3 analysis techniques.

1.2 Bottom-up vs Top-down Protein Mass Spectrometry

Fragmentation techniques are frequently used within MS to gain greater information on the composition of the protein analysed. This could be to identify a specific unknown protein, to identify the location of a post-translational modification (PTM),

protein or small molecule interactions sites, or the protein active site. It can also be used to gain proteomic scale analysis of cell culture. [41,42] Fragmentation analysis traditionally follows either 'bottom-up' or 'top-down' analysis workflows.

Bottom-up fragmentation analysis of a single protein, first involves enzymatic digestion of the polypeptide, preferably by an enzyme like trypsin which has a predictable digest pattern (after lysine, K or arginine, R). [43] The resultant mixture of digested peptides can then be analysed on the MS spectra, and individual digest ions can be isolated for further fragmentation (MS^2) to the individual amino acids to determine the amino acid sequence (Figure 1.11). In bottom-up proteomic studies the most common strategy involves an entire cellular lysate being enzymatically digested and separated using a C18 reverse phase column; the resulting peptide ions are analysed by MS and MS^2 and compared to protein databases. In effect each peptide acts as a 'proxy', to allow identification of the parent protein ion. [44] There are limitations to bottom-up analysis. Due to the multi-step sample preparation and the enzymatic digestion can result in loss in sensitivity in analysis. This includes the loss of peptides and therefore achievable sequence and detail coverage. This also includes the separation and identification of potential proteoforms. [45] Produced by alternative splicing events, single nucleotides polymorphisms and PTMs, all of which can have an impact on protein activity or have combinatorial regulatory functions in the cell. [42,46–48]

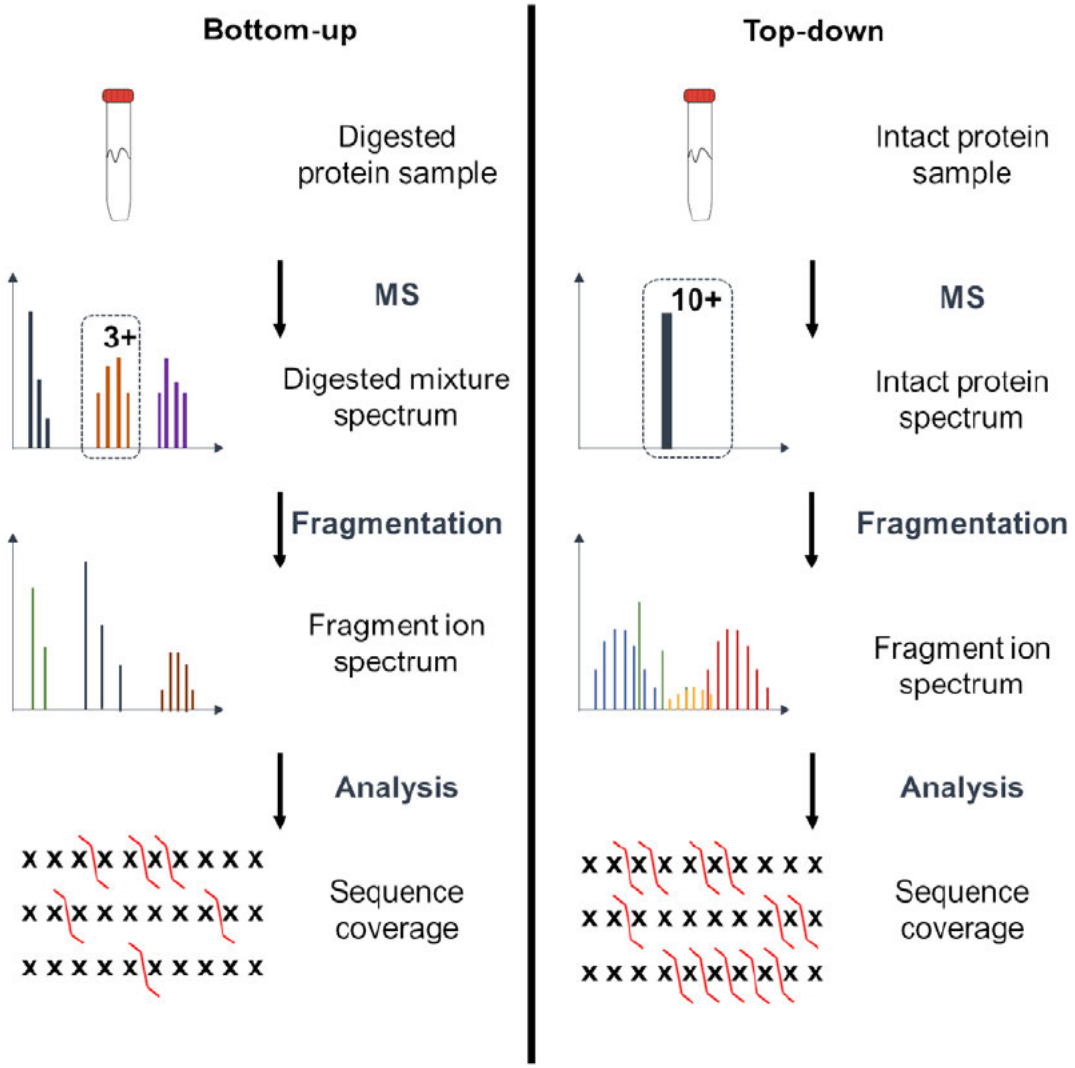


Figure 1.11: Comparison of the bottom-up and top-down protein MS workflows

Bottom-up analysis involves enzymatic digestion of the protein sample, while top-down analysis focuses on an initially intact protein sample. Once the sample is introduced into the mass spectrometer both styles can follow similar general fragmentation workflows.

In contrast, a top-down approach does not employ the initial enzymatic digestion. Instead, the single protein analyte or proteomic sample are first analysed as intact protein ions (MS), before an individual charge state of a single proteoform is isolated inside the instrument and fragmented (MS^2) to produce fragment ions (Figure 1.11). [49] This provides many advantages over the bottom-up workflow described above. As the inlet separation step takes place on the intact protein, it is potentially possible to

separate different proteoforms and isolate a single proteoform on which to perform fragmentation. Theoretically allowing full characterisation of the single proteoform. Thereby creating a detailed analysis on the number of different proteoforms, and characterisation of their differences.

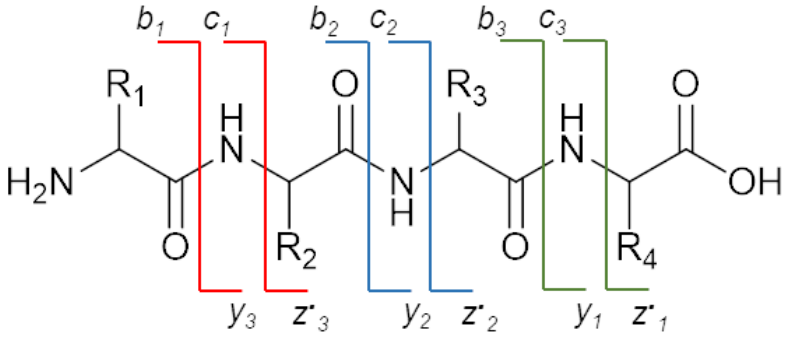


Figure 1.12: Fragmentation labelling nomenclature

Fragment ions are labelled in relation to the covalent bond which is broken. For b-type and y- type ions, the amide bond between amino acids is fragmented. For c-type and z˙ -type ions the N-Cα bond becomes fragmented. The b-type and c-type fragment ions are numbered sequentially and include the N-terminus, while the y-type and z˙ -type fragments are sequentially numbered and include the C-terminus.

Over the last two decades a range of gas phase fragmentation techniques have been developed. The nomenclature of the produced protein fragment ions is dependent on which bond in the polypeptide chain has been fragmented. As of product of specific fragmentation techniques. The N-terminal fragment ions typically produce *b/c*-type ions and are numbered sequentially from the N-terminus of the polypeptide, while the C-terminal fragment ions produce *y/z˙* -type ions and are numbered sequentially from the C-terminus. This is displayed in Figure 1.12, showing the potential bond cleavages of three standard amino acids. Numbered in the appropriate direction for N-terminal and C-terminal fragments. There is a wide variety in the available fragmentation techniques which are utilised over the many different MS platforms.

1.2.1 Collision induced dissociation

Collision induced dissociation (CID) or collisionally activated dissociation (CAD) is the most commonly used fragmentation technique. It uses high voltages to accelerate ions into many collisions with an inert gas, such as helium, nitrogen or argon. In FT-ICR, the protein ions contained in the ICR cell are excited at a frequency slightly off-resonant (sustained off-resonant irradiation, SORI) to the ω_c of the protein ion. [50] The protein ions can undergo successive rounds of excitation and de-excitation permitting multiple collisions with gas molecules. In other mass spectrometers, a separate collision cell filled with the inert gas is used. Through which the protein ions undergo multiple collisions with the gas molecules. The advantage to this is that there is no lower mass cut-off caused by the off-resonant irradiation. In the different forms of CID, multiple collisions increase the internal vibrational energy of the protein, which can disrupt non-covalent interactions and once a sufficient energy threshold is reached will also cause covalent bond breakages in the amide backbone. Resulting in the formation of *b*- and *y*- type ions (Figure 1.12) which can undergo further fragmentation or neutral loss ($-H_2O$ or $-NH_3$). [51]

The 'slow-heating' fragmentation techniques, like CID, are best described by the mobile proton model; [43] in which protons are transferred from basic side chains (e.g., arginine, R) to the nitrogen atom in the amide backbone which causes the structural rearrangement resulting in fragmentation. The rates of fragmentation are dictated by the ratio of carried charge to charge carrying site. [43] So having a larger number of basic residues in the polypeptide sequence will increase the rate of fragmentation observed. Similarly a range of other 'slow-heating' fragmentation techniques including infrared multi-photon dissociation (IRMPD) [52,53] and blackbody infrared radiative dissociation (BIRD) [51] can also be applied to cause fragmentation. In IRMPD, typically and infra-red laser supplies the required energy to, like CID, increase the

vibrational energy of the protein and therefore cause fragmentation. Again, in BIRD fragmentation the energy is supplied as a result of the thermal electromagnetic radiation of the ESI process and (typically) the ICR trap. As it is a low energy fragmentation technique BIRD fragmentation requires long reactions times, ranging from seconds to minutes. [54] Each of these fragmentation techniques redistribute the accumulated vibrational energy across the entire peptide and results in production of *b*- and *y*- type fragmentation products. Displaying a bias towards fragmentation of regions with greater flexibility, around the termini as the molecular weight of the protein increases. It is therefore ideally suited as a fragmentation technique for bottom-up analysis, and fragment smaller enzymatic digest ions. [43,55] Another detrimental factor for using these slow-heating techniques, is that the energy redistribution can result in the neutral loss of PTMs. [56] Further reducing the applicability of these techniques for thorough top-down investigations.

1.2.2 Electron capture dissociation and Electron transfer dissociation

The collective electron dissociation techniques (ExD), which encompass electron capture dissociation (ECD) or electron transfer dissociation (ETD) involves using low energy electrons (~1 eV) [55] to produce protein ion fragments. The ExD techniques unlike the multiple collisions in the slow-heating fragmentation methods, have the energy supplied for fragmentation with one electron capture/transfer event, resulting in fragmentation close to the site of electron capture. [55] maintaining non-covalent interactions and PTMs.

ECD is normally carried out on FT-ICR platforms (page 18), as protein ions can be trapped within the ICR cell which enter from the front end of the mass spectrometer.[57] The electrons are produced with heated tungsten hollow ring cathode [58,59] which can enter into the ICR cell through the opposite end of the ICR cell. Importantly, there is significant overlap of the cyclotron motion (ω_c) of the protein ions and the electrons to allow significant electron capture, and ultimately fragmentation before excitation and detection. This is represented in Figure 1.13, showing a cartoon schematic of the electron path through the ICR cell, and producing *c*-type and *z*·-type ions.

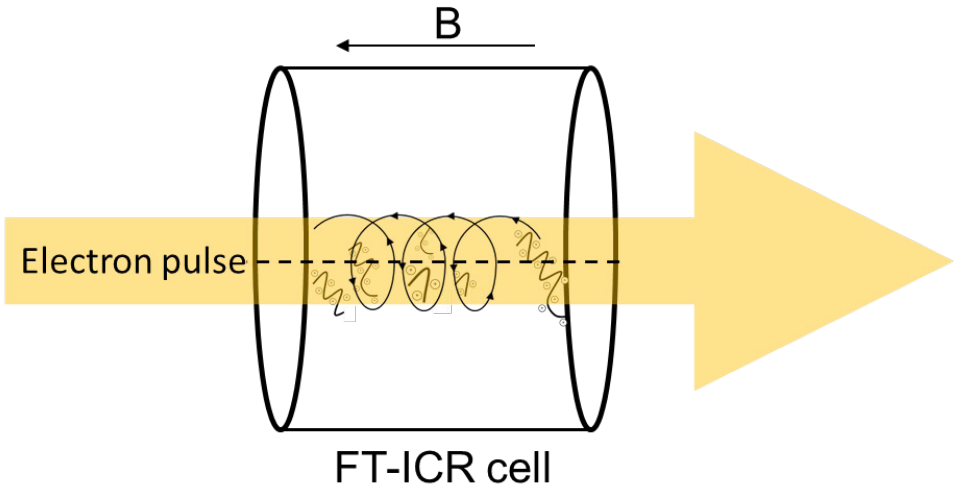


Figure 1.13: Cartoon diagram depiction of ECD fragmentation

Prior to electron capture, ions are trapped within the ICR cell, which can be bombarded with the generated electron pulse allowing electron capture which can lead to fragmentation of the multiply charged protein ions into c-type, containing the N-terminus or z· -type ions containing the C-terminus.

In ECD, the electron capture step frequently does not result in protein ion fragmentation. The electron capture event without fragmentation produces 'charged reduced species' or 'electron capture without dissociation' (ECnoD) which are observed at a lower charge value, plus the mass of an electron. [60,61] These ECnoD fragments could also have the potential to be composed of *c*-type and *z*· -type product

ions which still interact non-covalently, therefore is observed at the combined mass of both fragments. The highly specialised instrumentation required for ECD makes it a very successful protein fragmentation technique on FT-ICR platforms. However it is not an optimal electron delivery system for the majority of MS platforms. [62]

If the energy of the supplied electron increases it is described as *hot* ECD (~3 -10 eV) [57,63,64] which produces fragmentation in a similar manner to ECD, but is also capable of producing secondary fragmentation events. If the electron energy is further increased (> 10 eV) it is described as electron detachment dissociation (EDD). Which has shown to be a powerful tool for the fragmentation of negatively charged peptides. [65,66]

ETD was developed to allow ExD fragmentation to be applied initially in a linear ion trap mass spectrometer, [62] which are often coupled to orbitrap detectors for high resolution analysis. When first developed, an anthracene electron donor was added from the back end of the mass spectrometer and overlapped with the protein ions inside a linear ion trap with differential voltage control to optimise fragmentation. [62] The technique has further developed allowing front end fragmentation in an installed multi-dissociation reaction cell which can be applied to a wide range of mass spectrometers. [38,67,68] In modern ThermoScientific instruments, a fluoranthene (202.26 Da) electron donor is applied at the front end of the instrument near the nESI inlet [67,68]

The mechanism of the ExD techniques is still disputed but best known are the Cornell mechanism, which postulates that the electron is taken up primarily by the NH_3^+ side chain of lysine and the now hypervalent residue will donor a proton to nearby carbonyl groups on peptide backbones and cause the structural rearrangement and fragmentation. The Utah-Washington mechanism suggests that the electron is taken

up by the unoccupied pi orbitals ultimately causing fragmentation of the N-Cα bond prior to the proton transfer step. [61,69]

Regardless of the mechanism it is clear that ExD techniques primarily result in fragmentation of the N-Cα bond throughout the polypeptide backbone. Similarly, ultraviolet photodissociation (UVPD) is another fragmentation technique. It applies a laser (typically 193 nm, 6.4 eV) to cause fragmentation. UVPD is novel to the other fragmentation techniques previously described. In that it can cause vibrational activation and fragmentation, producing *b/y*- type ions and rapid direct dissociation producing *c/z*- type ions. It can also produce *a/x*- type ions (C-C_α bond cleavage) and internal fragments via secondary fragmentation events. [70–73] So there are several different mechanisms through which polypeptide fragmentation can occur. With the ExD and UVPD higher energy fragmentation techniques being more applicable for maintaining PTM's and permitting greater sequence coverage.

1.2.3 Proton Transfer Charge Reduction

Proton transfer charge reduction (PTCR) is another form of ion/ion reaction which can reduce the charge carried by multiply charged ions, producing $[M+(n-1)H]^{(n-1)+}$ species. [74–76] Proton transfer, much like electron transfer, uses radically charged anions which readily accepts a proton from multiply charged protein. The multiply charged species can react many times with the supplied anions producing an array of lower charged ions. Economically PTCR can be applied using the same anion as in ETD, but with a nitrogen adduct introduced during reagent inlet (216 Da). [77] However, perfluoroperhydrophenanthrene (624 Da) is now used in the most recent commercial instruments (ThermoFisher). [78,79] This an extremely versatile technique, particularly in a top-down fragmentation workflow, as it can be employed to simplify complex regions of the fragmentation spectrum, as shown in Figure 1.14. [79]

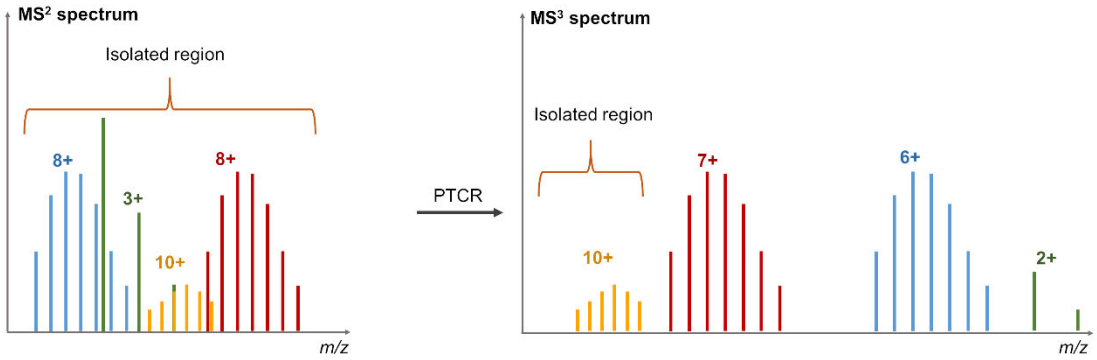


Figure 1.14: Impact on a mass spectrum with the application of PTCR

A standard workflow in top-down analysis applying PTCR to a region of dense, overlapping ion isotopologue distributions (MS^2 spectrum). As a result of applying the PTCR reagent (MS^3 spectrum) fragment ions which were previously contained within the isolated region of the mass spectrum to appearing at a greater m/z value as a result in the reduction of charge on the fragment ion.

PTCR can be applied stepwise to small isolated windows across the MS^2 fragmentation spectra, reducing the fragment ion charge. Resulting in the complicated, overlapping fragment ion isotopologue distributions at lower charge states across higher *m/z* range in the MS^3 spectrum (Figure 1.14). This provides the opportunity to assign fragment ions which can be left unassigned or hidden by the isotopologue distribution of abundant ions. It can also be extremely useful for isolation of a specific proteoform that cannot be sufficiently separated using a quadrupole prior to fragmentation. PTCR application can create a pseudo charge state distribution. The correct one of which can then be isolated and subjected to the fragmentation method. [80] The combination of the many ion manipulation techniques allows for complex and detailed analysis of many different biological sample types.

1.3 Limitation of top-down analysis

Typical top-down fragmentation experiment produces a highly complex spectrum of fragment ions at multiple charge states. The issue with analysis occurs due to the fragment ions appearing over a comparatively small range of *m/z* (typically ~400-2500) (Figure 1.15). Thus, overlapping fragment ion signals are commonplace. This complexity further increases with increasing molecular weight of the protein analyte. As previously discussed, larger MW protein are composed of a greater number of amino acids, which contribute to a greater incorporation of heavier isotopes. Therefore, with fragmentation, a larger MW protein will produce more fragment ions,

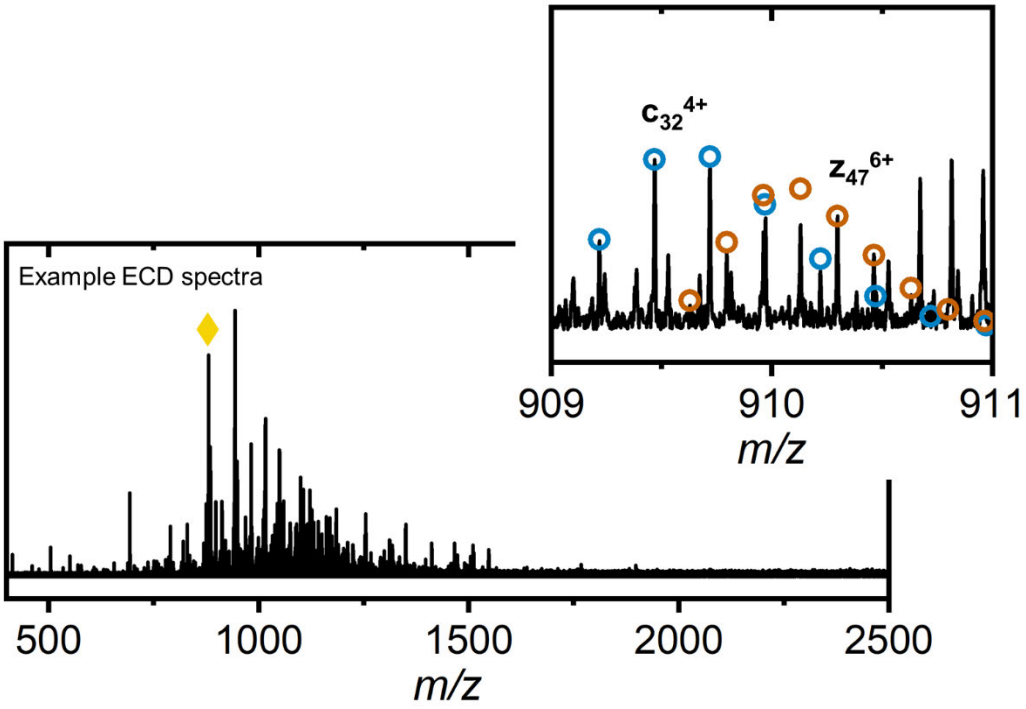


Figure 1.15: Example of overlapping fragment ions.

An example of the spectral complexity as a result of top-down (ECD) fragmentation reaction. An example spectrum of a 13 kDa protein with isolated precursor ion highlighted (♦). The inset panel displays the overlapping distributions of the c_{32}^{4+} and z_{47}^{6+} ions and the separate ion distributions are highlighted by the modelled overlaid theoretical distributions

many of which will also be of larger MW. These, in turn, will display a greater abundance of isotopologue peaks.

As shown in the inset of Figure 1.15, being able to differentiate the isotopologue peaks of separate ions can be challenging as many exist in very close proximity or may overlap. It normally requires deconvolution software, [81] calculating the monoisotopic peak for unknown isotopologue distributions based on the central building block of averagine. Algorithms such as THRASH or SNAP [82,83] can deconvolute the charge of isotopologue distribution and identify the ions like c_{32}^{4+} and z_{47}^{6+} (Figure 1.15) and calculate the *m/z* of the monoisotopic peak. While this model can be an efficient system to calculate mass and assign protein peaks it can also introduce error, if for example the protein contains a higher sulfur content than typical globular protein. However it is currently one of the best analysis methods to facilitate assignment of unknown proteins through distribution deconvolution, assigning charge [19,84] and accurately estimating monoisotopic masses. With larger MW protein the fragmentation spectrum will eventually be overly complex, with reduced fragment ion signal. Ultimately limiting the detectable ions which can be assigned.

Recently, several alternative analysis workflows have been introduced. AutoVectis is a different spectral analysis software which does not use charge deconvolution to identify separate isotope distributions. Instead it uses the produced FID to identify all the peaks within the mass spectrum and calculate regional baseline noise values. [85–87] It can also autocorrect for the 'phase-shift' in FT-ICR analysis and produce the absorption mass spectrum, as opposed to the normally produced magnitude spectrum. This serves to produce the same mass spectrum with increased resolution. AutoVectis can also produce comprehensive fragment ion databases based on the amino acid sequence of the protein being analysed. Therefore, the software can also compare the detected peaks to the predicted databases for the assignment of ions

and so effectively carry out the full analysis of particular spectra types. However, there are still further advances to be made, in order to improve our ability to analyse larger MW protein and achieve ever greater sequence coverage.

1.4 The history of isotope depletion

Instrument developments are crucial for further development of top-down fragmentation. To become implemented within research these instrument upgrades are often expensive. A complementary strategy to could be to simplify the sample. An initial study into the potential benefit of this simplification was carried out in 1997 by Marshall *et al*. [88] In this study they were the first to display the observed isotopologue distribution change of a small protein (11.8 kDa) as a result of *isotope depletion* resulting in reduced isotope complexity. Produced with an expression media composed of controlled ion abundance (^{12}C 99.95%, ^{14}N 99.99%) and different from the natural isotope abundance (^{12}C 98.93%, ^{14}N 99.63; Table 1.1). They postulated that the isotope depletion strategy would have multiple benefits even at a large molecular weight (~60 kDa).

Following this initial study, a computational analysis of the theoretical impact and usefulness of isotope depletion when applied to lower resolution mass spectrometers was carried out by Zubarev and Demirev in 1998. [17] The authors discussed the benefits of the technique and concluded that the resultant shift in the centroid *m/z* of an isotopically depleted protein can be used to calculate the average mass of the isotopically natural protein in low-resolution acquisitions, only when the approximate level of isotope depletion is known. Experimentally, the field was further developed by Akashi *et al*. also in 1998 [89] who carried out CID fragmentation on a 11 kDa protein using a 7T FT-ICR analyser. They reported an increase in the number of assignable fragments, and greater S/N of fragments within the spectra. Similarly, Jensen *et al* in 1999 and Sharma *et al* in 2007, [90,91] both demonstrated the proof of principle that an

isotopically depleted sample can be applied to proteomic analysis discussing the impact on the isotopologue distribution. However, neither study focus on in-depth analysis of the isotopically depleted samples, nor described the difference, if any, in the identified proteins between isotopically natural and isotopically depleted samples.

Another isotopically depleted study was carried out by Stump *et al* in 2003, [92] which analysed the ribosomal fraction of an *E. coli* strain grown within isotopically depleted media using MALDI-TOF MS. Again, this study also reported the simplification of the isotopic distribution, and S/N increase of the isotopically depleted ribosomal protein. Facilitating the identification of a greater number of ribosomal proteins with reduced data acquisition. This trend is also observed by Bou-Assaf *et al* in 2010, [93] in which isotopically depleted protein samples were applied to hydrogen-deuterium exchange (HDX) who reported the simplification of peptide isotopologue distributions. Therefore, observed less peptide overlap once the deuterium exchange causes an increase in their isotopic complexity. As the depletion of the isotopes occurs within the cellular expression media, the entire cellular content therefore is composed of atypical isotope abundance. Isotope depletion can facilitate improved MS analysis of more the just protein, particularly ribonucleotide as was reported by Xiong *et al* in 2004, [94] demonstrated that isotopically depleted ribonucleotides display a similar trend to protein. Opening another potential avenue for isotopic depletion investigation.

Finally, in 2011 Compton *et al* [95] carried out a theoretical analysis of the impact of different isotope depletion ratios on protein S/N with an increase of MW. They concluded that isotope depletion of carbon and nitrogen will eventually reach an upper mass limit at which the benefit of isotope depletion on the observed S/N of the protein distributions (>80 kDa).

Variations of the isotope depletion strategy has been applied to a handful of varied analysis of protein and ribonucleotides over the past couple of decades. However,

isotopic depletion has not been widely adopted by the MS community in that time. This may be in part due to the focus on the development of the instrumentation but is more likely due to the added effort or cost of producing isotopically depleted samples. In all the above investigations, isotopically depleted samples were expressed using commercial media (Cambridge Isotope Laboratories) which is no longer available. Or in M9 minimal media without detailed discussion of the effectiveness in their isotopically depleted protein expression.

A direct comparison of isotopically depleted protein to the natural equivalent, in order to experimentally determine the benefit of the isotopic depletion strategy has not been carried out. In order for the isotopic depletion strategy to become more widely adopted, the impact of its application needs to be thoroughly investigated.

1.5 Aims of the project

The aim of this investigation is to continue from this handful of initial studies and produce an efficient method to allow the consistent and controlled incorporation of depleted isotopes into protein which does not rely on a preformulated medium. Thereby regulating the simplifying of their isotopologue distributions. The development of an efficient (and cost effective) expression protocol, isotope depletion and applied to multiple protein of varying molecular weight facilitates the assessment of the strategy. Therefore, highlighting the potential benefits. As isotope depletion applied to larger molecular weight should partially counter the isotopologue spreading and described by Compton *et al* [95] as impacting the S/N of natural protein ions beyond 50 kDa. The impact of isotope depletion on various MS analysis needs to be assessed through the comparative study of the expressed model protein at various molecular weight.

Application of top-down fragmentation techniques to isotopically natural and depleted forms of the same protein will highlight any differences in their respective spectra and

benchmark improvement in fragment ion analysis. As the isotope depletion strategy results in a fundamentally altered protein sample, the isotopologue simplification has the potential to positively impact a wide range of biological MS techniques. As further proof of principle of the benefit of applying isotope depletion to protein MS, carrying out comparison, and applied to native analysis.

In simpler terms, the aim of the project is to develop the methodology and investigate the many advantages *isotope depletion* can have on protein MS.

2. Materials and Methods

2.1 Materials

Glucose (^{12}C; 99.90 %) was purchased from Cambridge Isotope Laboratories, Inc. Ammonium sulphate (^{14}N; 99.99 %) was purchased from ISOTEC™

R. rubrum ($A0973_{sH}$) in pET28a was kindly gifted from Dr Jon Marles-Wright (Newcastle University). BCA2 in pETHISTEV was constructed within the Clarke group by Dr Sam Hughes (University of Edinburgh). *S. paucimobilis* (AmSPT) in pET28a was kindly gifted from Prof Dominic Campopiano (University of Edinburgh).

2.2 Methods

2.2.1 Molecular biology

2.2.1.1 Site-directed mutagenesis

The expression pET28a plasmid containing the gene sequence for the encapsulated ferritin from *R. rubrum* with a C-terminal truncation and hexahistadine tag (Rru_A0973_{sH}) [96] underwent site-directed mutagenesis carried out using Quikchange Lightning (Agilent UK). Both a M30L and M68L mutation were sequentially introduced. Mutations were confirmed with Sanger Sequencing carried out at GATC Biotech UK

A M30L, M68L variant of Rru_A0973_{sH} was created with site-directed mutagenesis carried out using Quikchange Lightning (Agilent UK), M30L and M68L mutations were sequentially introduced. Mutations were confirmed with Sanger Sequencing carried out at GATC Biotech UK.

Primer sequences:

M30L Forward- 5'-CTTCCAGTTCCTCCAAGACCGACACGATCGC -3'

M30L Reverse- 5'-GCGATCGTGTCGGTCTTGGAGGAACTGGAAG-3'

M68L Forward- 5'-CCATTCCAGGGTCAACGCCGCGTGTTCTT-3'

M68L Reverse- 5'-AAGAACACGCGGCGTTGACCCTGGAATGG-3'

2.2.2 Agarose Gel Electrophoresis

A 0.8% $^{w}/_{v}$ agarose gel was prepared using agarose (Lonza Bioscience) in TBE buffer (New England Biolabs) and stained with SYBR Safe DNA gel stain (Invitrogen). DNA was mixed with 6x gel loading dye (New England Biolabs). The gel was run at 100 V

for approximately 45 minutes using 1 kb ladder (New England Biolabs) and visualised with a UV light.

2.3 Protein expression

2.3.1 Plasmid transformation

The required plasmids (50 ng/µl) were added to a 30 µL aliquot of competent cells for plasmid DNA amplification (DH5α) or protein expression (DE3). The cells were incubated on ice for 30 minutes before 45 seconds of heat shock at 42 ^{0}C. The cells were again incubated on ice for 2 minutes before the addition of 200 µL of SOC outgrowth media to the cell culture. The culture was incubated at 37 ^{0}C for 45 minutes, then 100 µL were spread onto a solid agar plate containing the appropriate antibiotic and incubated overnight at 37 ^{0}C.

2.3.2 Expression of protein in LB media

A single colony of transformed BL21 (DE3) cells containing the desired expression plasmid was used to inoculate 10 ml of LB media supplemented with ampicillin (100 µg/mL), kanamycin (50 µg/mL) or chloramphenicol (25 µg/mL) as required by the expression plasmid. Starter cultures were incubated overnight at 37 ^{0}C, which was then used to inoculate 500 ml of LB media which was grown at 37 ^{0}C until an OD_{600} 0.6-0.8 was obtained. Protein expression was induced with β-D-thiogalactoside (IPTG 0.1- 1mM) and the culture was incubated for 4 hours at 37 ^{0}C or for 18 hours at 18 ^{0}C. Cultures were centrifuged at 5000 *x g* at 4 ^{0}C for 30 minutes, and the pellets were stored at -20 ^{0}C until further required.

2.3.3 Expression of protein in M9 minimal media

A single colony of transformed BL21 (DE3) cells containing the desired expression plasmid was used to inoculate 10 ml of LB media supplemented with ampicillin (100 µg/mL), kanamycin (50 µg/mL) or chloramphenicol (25 µg/mL) as required by the expression plasmid. Starter cultures were incubated overnight at 37 ^{0}C, then used to inoculate 500 ml of LB media which was grown at 37 ^{0}C until an OD_{600} 0.6-0.8 was obtained. The 200 ml of the LB culture was centrifuged at 5000 *x g* at 4 ^{0}C for 20 minutes, and the resulting cell pellet was washed twice in 5X M9 Salts solution (33.9 g/L Na_2HPO_4, 15 g/L KH_2PO_4, 2.5 g/L NaCl) without a nitrogen source [97] and pelleted. The washed pellet was suspended in 100 ml of isotopically standard M9 Minimal Media or isotopically depleted M9 Minimal Media, (Table 2.1) the culture was incubated at 37 ^{0}C for 1 hour. Protein expression was induced with IPTG (0.1- 1 mM)

and the minimal media cultures were incubated for 18 hours at 18 ^{0}C, adapted from [96,97]. Cultures were centrifuged at 5000 *x g* at 4 ^{0}C for 30 minutes, and the pellets were stored at -20 ^{0}C until further required.

Table 2.1 M9 media and 5x M9 salts recipes

M9 growth media recipe (1L)		5 x M9 Salts recipe	
		g/L	
10 ml	40% glucose	33.9	Na_2HPO_4
2 ml	1 M MgSO4	15	KH_2PO_4
100 µl	1 M CaCl2	2.5	NaCl
200 ml	5 x M9 salts	12.4	$(NH_4)_2SO_4$

2.3.4 Protein purification

To prevent cross contamination between the isotopically natural and isotopically doubly depleted protein the purification process kept separate, using independent buffers and ion- affinity columns.

2.3.4.1 Encapsulated Ferritin

Cell pellets containing EncFtn (Rru_A0973$_{sH}$) were suspended in HisA buffer (50 mM Tris, 50 mM imidazole, 500 mM NaCl, pH 8) and the cells disrupted with sonication at 10 µm amplitude in 15 second bursts over 8 minutes. The lysate was clarified by centrifugation at 20,000 *x g* at 4 ^{0}C for 30 minutes and filtered through a 0.22 µm syringe filter (Millipore UK).

Filtered lysate was loaded onto a 1 ml HisTrap FF column (GE Healthcare) pre-equilibrated with HisA buffer. Unbound proteins were washed out with 5 column volumes (CV) HisA buffer and EncFtn was eluted using a linear gradient between 40-80% HisB (50 mM Tris, 500 mM imidazole, 500 mM NaCl, pH 8.0). Fractions were visualised on an 18% acrylamide gel and selected fractions were pooled and concentrated with a Vivaspin spin centrifugal concentrator (5,000 MWCO). Concentrated protein sample was buffer exchanged into the storage buffer (50 mM Tris, 500 mM NaCl, 10% $^v/_v$ Glycerol, pH 8.0), and aliquoted at a standard concentration of 10 µM which was stored at -80 ^{0}C, adapted from He, D. *et al*. 2016. [96].

2.3.4.2 Carbonic Anhydrase

Cell pellets containing BCA (BCA II) were resuspended in buffer A_{BCA} (PBS with 10 mM imidazole, pH 7.4) and the cells were disrupted with sonication at 10 µm amplitude in 15 second bursts over 8 minutes. The cell lysate was clarified by centrifugation at 12000 × *g* for 1 hour at 4 °C and filtered using a 0.22 µm syringe filter (Millipore, UK).

The filtered cell lysate was loaded onto a 1 ml HisTrap column (GE Healthcare, UK), pre-equilibrated with lysis buffer. Unbound protein was washed out with 5 CV lysis buffer and protein was eluted using a linear gradient between 0-100 % of buffer B_{BCA} (PBS and 300 mM imidazole, pH 7.4). Hexahisitdine tag cleavage was carried out by incubating protein with TEV protease overnight at 4 °C. Non-cleaved protein and cleaved tags were removed by diluting the protein sample 1:2 with buffer A_{BCA} and reloading onto 1 ml HisTrap column and the flowthrough collected. The collected flowthrough was concentrated to a standard concentration of 10 µM using a Vivaspin centrifugal concentrator (5000 Da MWCO) and exchanged into a storage buffer (50 mM Tris-HCl, 10% glycerol, pH 7.4) using a Desalt column (GE Healthcare UK) and stored at -80 ºC.

2.3.4.3 Serine Palmitoyltransferase

Cell pellets containing SPT (AmSPT) [98] were resuspended in Buffer A_{SPT} (20 mM potassium phosphate, 150 mM NaCl, 25 µM PLP, 10 mM imidazole, pH 7.5) and disrupted with sonication at 10 µm amplitude in 15 second bursts over 8 minutes. The cell lysate was clarified by centrifugation at 12000 × *g* for 1 hour at 4 °C and filtered using a 0.22 µm syringe filter (Millipore, UK).

The filtered cell lysate was loaded onto a 1 ml HisTrap column (GE Healthcare, UK), pre-equilibrated with Buffer A_{SPT}. Unbound protein was washed with 10 CV Buffer A_{SPT} and the bound protein was eluted using a linear gradient between 40-100 % Buffer B_{SPT} (20 mM potassium phosphate, 150 mM NaCl, 25 µM PLP, 10 mM imidazole, pH 7.5. The fractions containing eluted SPT were concentrated (10,000 MWCO) and loaded onto a superdex 300/600 column (GE Healthcare, UK) equilibrated with SPT_{SE} buffer (20 mM Tris, pH 7.5) for purification via size exclusion chromatography.

2.3.4.4 Polyacrylamide gel electrophoresis

Purified protein was initially visualised using an in-house 12% (Table 2.2) sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or a 4-12% NuPage

bis tris gels (Invitrogen). The gels were run with MES buffer at 180-200 V for 35 minutes or as needed.

Table 2.2 Recipe for 12% SDS-PAGE for protein.

	Stacking gel	Separating gel
Water	4.08 mL	7.92 mL
Polyacrylamide (30% $^{v}/_{v}$) (Bio-Rad)	1.02 mL	9.6 mL
Tris-HCl (1 mol dm-3, pH 6.8)	0.75 mL	-
Tris-HCl (1.5 mol dm-3, pH 8.8)	-	6 mL
Sodium dodecyl sulfate (10% $^{w}/_{v}$)	60 µL	240 µL
Ammonium persulfate (10% $^{w}/_{v}$)	60 µL	240 µL
Tetramethylethylenediamine	6 µL	9.6 µL

2.4 Mass Spectrometry

Experiments involving Q-TOF were carried out on a Synapt G2 mass spectrometer (Waters). Higher resolution analyses were carried out on a 12T SolariX FT-ICR (Bruker Daltonics), or an Eclipse orbitrap (ThermoFisher) where mentioned.

2.4.1 Liquid Chromatography mass spectrometry

The Acquity UPLC was coupled to a Synapt G2 TOF mass spectrometer (Waters). Reversed-phase LC was performed using a 50 x 2.1 mm C4 column (Phenomenex) without temperature control. The flow rate was maintained at 0.2 mL min^{-1} and typically, approximately 25 µmol of protein was loaded onto the column. After a three-minute wash step with 95% mobile phase A (water, 0.1 % $^{v}/_{v}$ formic acid), the protein was eluted over a 5-minute linear gradient from 5 – 95% mobile phase B (acetonitrile, 0.1 % $^{v}/_{v}$ formic acid). On-line MS was enabled by ESI, with a capillary voltage of 4 kV, source temperature of 120 °C and sampling cone voltage of 30 V. Mass spectrum was acquired continuously after the three-minute wash step.

The dionex HPLC was coupled to a 12T SolariX FT-ICR (Bruker Daltonics). Reverse phase LC was performed using ACE 3 C4 (75 x 0.5 mm) typically without temperature control. The flow rate was maintained at 0.3 mL min^{-1} with a 10:1 flow splitter to control pressure. Typically, 2-5 µM of protein was loaded onto the column depending on if MS^1 or MS^2 analysis was taking place. Typically, the protein was eluted over a 15 to 30-minute linear gradient from 5 – 95% mobile phase B (acetonitrile, 0.1 % $^{v}/_{v}$ formic

acid). On-line MS was enabled by ESI, with a capillary voltage of 4 kV and source temperature of 200 °C. Acquisition of mass spectrum continued throughout the LC elution times and MS acquisition method was varied to match 'ideal' settings for MS^1 and MS^2 spectra as developed using direct infusion ESI analysis.

2.4.2 Direct infusion nanoelectrospray ionisation

Typically, a 5-10 µM protein sample was buffer exchanged using Bond Elut OMIX C4 Ziptip (Agilent Technologies) and eluted 30µl in acetonitrile (50% $^{v}/_{v}$, 0.1% Formic acid). Alternatively, for native analysis, between 1-25 µM protein was exchanged into 10-100 mM ammonium acetate protein sample using P6 micro-spin columns (Bio-rad).

Mass spectrometry was carried out in an Infinity or Paracell 12T SolariX FT-ICR (Bruker Daltonics) with nESI ionisation was either performed using a TriVersa nanomate (Advion Bioscience) or borosilicate tips (OD 1.2 mm, ID 0.9 mm, World Precision Instruments) pulled to a tip in-house using a model P-1000 micropipette puller using a 2.5 mm-wide box filament (Sutter instruments). Sample is loaded into the pulled borosilicate tips and a platinum wire inserted into the protein solution to conduct approximately 1 kV voltage. The angle and distance of the tip was manually altered relative to the sample-inlet to optimise the spectra.

2.4.3 Top-Down fragmentation

In preparation for MS^2 experiments, selected MS^1 protein ion were accumulated to the greatest possible ion abundance, typically ~$1x10^8$ within the 12T FT-ICR. Automatic gain control (AGC) was minimally set to $2x10^5$ to control ion accumulation of the MS^1 protein ion. Ion isolation and accumulation for further fragmentation was controlled with continuous accumulation of selected ions (CASI).

2.4.3.1 Collision induced dissociation (CID)

In the 12T FT-ICR, CASI isolated ions were accelerated with a collision energy set between 9-15V for the fragmentation of multiple quadrupole isolated charge state peaks. Scan transients were set are as required by the molecular weight of the protein ranging from 5 to 300 averaged transients as described.

2.4.3.2 Electron capture dissociation (ECD)

In the 12T FT-ICR, typical conditions for ECD involved an ECD current of 1.6 A, lens voltage of 12 V, bias of 1.2 V with a pulse rate ranging from 2-20 ms. CASI ion isolation

and accumulation was altered as necessary. Spectra were acquired with 5-300 averaged ICR transients as required. Analysis and modelled isotopologue distributions were carried out using DataAnalysis software (Bruker, Version 4.4) or AutoVectis (Pro version 8.8).

2.4.3.3 Electron transfer dissociation (ETD)

In the orbitrap Eclipse, isolated precursor was subjected to ETD fragmentation carried out on the orbitrap Eclipse. Isolated MS^1 precursor was isolated with AGC set to 1 x 10^6. The precursor ion was subjected to ETD fluoranthene reagent for 0.5-3 ms as required to cause fragmentation. Spectra were typically collected as 100 scan averaged transients.

2.4.3.4 Proton transfer chain reaction (PTCR)

Isolated ETD MS^2 spectral regions in the orbitrap Eclipse were isolated and AGC accumulated at $5x10^5$ and subjected to the PTCR perfluoroperhydrophenanthrene reagent for 0.1-15 ms as required for efficient proton transfer. Spectra were typically collected as 100 scan averaged transients.

2.4.4 Data analysis

LC-MS data collected on synapt G2 Q-TOF (Waters) were analysed using MassLynx (Version 4.1, Waters) using the Transform algorithm (Waters) to calculate the average mass of the protein in the charge state distribution.

FT-ICR acquired spectra were initially visualised and assessed using DataAnalysis (Bruker) which was also used to generate theoretical model distributions for spectral overly. The charge state distribution spectra of encapsulin was smoothed with 1.2 *m/z* width and baseline subtracted to produce the processed charge state distribution image. Isotopically natural protein spectra were initially deconvoluted using SNAP analysis (Bruker Daltonics), with the S/N cut-off was set to 3, and the quality factor was set to 0.3. Typically, a single point internal calibration on DataAnalysis was performed with a consistent fragment ion c_{20} for EncFtn and c_{34} for BCA. Theoretical ion distributions were created using the Simulate Pattern function of DataAnalysis. To which the required chemical formula and either natural or depleted isotope abundance was supplied.

MS^2 fragmentation spectra for comparative analysis were processed into sequence coverage and generate fragmentation maps using an adapted version of AutoVectis.

Fragment ion peaks were typically identified between 400-2000 *m/z* in absorption mode and compared to theoretical fragment peak lists generated by AutoPiquer based on the amino acid sequence of the studied protein. Typically, the S/N cut-off was set at 3, as was the SNR value within AutoPiquer. The spectra were calibrated using a linear trend line through the main cluster of fragment ions. Using a plot of error by *m/z* to determine the linear trend line of the spectra.

The coherency of each fragment ion was assessed by visually inspecting each ion distribution and compared to a theoretical distribution. Autopiquer assigned ions which were deemed to not be a correct assignment were manually excluded from the assigned ion list at this point. AutoVectis provides the error and S/N value for each assigned isotopologue peaks. The appropriate peak for isotopically natural protein was based on the full isotopologue distribution and for isotopically depleted fragments ions the most abundant peak was typically the monoisotopic peak, or M+1 peak for larger fragments. The error and S/N values were always taken from the same isotopologue peak.

On-line fragmentation files acquired on 12T FT-ICR were first observed on DataAnalysis, accumulating the spectra in the appropriate area of the LC chromatograph. The selected chromatograph area was analysed using AutoVectis to assign fragment ions as described above.

ThermoFisher RAW. files produced by the orbitrap Eclipse were converted to a mzML files using MS convert software (http://proteowizard.sourceforge.net/tools.shtml) and analysed using a specialised version of AutoVectis (AutoPiquer mzML) and analysed as described above.

3. Development of an efficient method to produce isotopically depleted protein in M9 minimal media.

3.1 Isotopic heterogeneity in protein

The biosynthesis of amino acids is an essential process and intrinsically connected to the central metabolism pathways, such as the Citric acid cycle and the shikimate pathway. [99] These pathways incorporate element sources, like glucose for carbon to keep the bacterial and eukaryotic cells resources of these amino acids sufficient to meet the transcriptional demand of the cell. As a consequence of this, heavier isotopes such as ^{13}C are incorporated into amino acids at the same ratio to the naturally occurring abundance (^{13}C, 1.07%, Table 1.1). Therefore, the translation of the amino acid into polypeptides produce protein which also contain natural abundances of heavier isotopes.

The production of an organism which contains an unnatural isotopic composition is therefore fundamentally controlled by the nutrient source upon which an organism is grown. This as a concept is well understood and has been utilised, extensively in NMR studies, to enrich peptide and protein samples for NMR-active heavier isotopes such as ^{13}C and ^{15}N. As there is the demand, there is also a supply of commercially produced growth media which is enriched in whichever heavy stable isotope required. For isotope *depletion*, the ultimate goal is to remove heavier isotopes such as ^{13}C and ^{15}N, instead enriching for the lighter isotopes ^{12}C and ^{14}N. While there has been commercial media enriched for carbon (^{12}C, ~99.95%) and nitrogen (^{14}N, ~99.97%) in the past [90,91], it is no longer commercially available and its use would ultimately limit isotopic depletion as a technique.

3.2 Isotopically depleted culture media

A defined growth media, such as M9 minimal media, supplies BL21 cells for protein expression with a controlled carbon source, nitrogen source, some inorganic salts ($MgSO_4$ and $CaCl_2$) and water. Thus, allowing strict control of each component within the media and minimal media (used to enrich for heavier isotopes for NMR analysis

for decades). [100] This is in contrast to more commonly used rich media broths, such as LB media, which facilitate fast growth and high cellular densities for bacterial cultures. but supply undefined yeast extract and tryptone to the BL21 cells. [101]

However, as minimal media has a basic level of nutrient availability, it is often difficult to grow the expression cells (BL21) to an appropriate density in order to recombinantly express suitable levels of protein. Thus, the resulting yield of recombinant protein can often be significantly lower than that achieved by culture in rich media. The protocol developed in this study, to express recombinant proteins enriched protein with lighter isotopes, was developed based on a findings by Marley *et al*. 2001. [100] To produce the required cellular density, the BL21 cultures are initially grown in an isotopically natural rich yeast-based media, such as LB or 2xYT. The cells would then be washed and exchanged into the isotopically controlled media before protein expression is induced.

This is highlighted in Figure 3.1, which for stages 1-3 follows the basic workflow for the growth of protein expression bacteria cultures. [102] The BL21 *E. coli* cells are only placed in the M9 minimal media after the washing step at stage 4. Once placed in the final minimal media (stage 5) the BL21 cells require at least 30 minutes at 37^0C to rest and acclimatise to the new nutrient availability. The M9 minimal media can be further supplemented with metal ligands or co-factors as required by the protein which is to be recombinantly expressed. It is important to note that the success of this strategy relies on the strict control of protein expression during the initial growth in rich media, this is achieved using the T7/lacO inducible expression system. [102] The Isopropyl β-D-1-thiogalactopyranosid (IPTG) induction is best carried out at low temperature with a longer expression time to facilitate lighter isotope incorporation into the polypeptide chain.

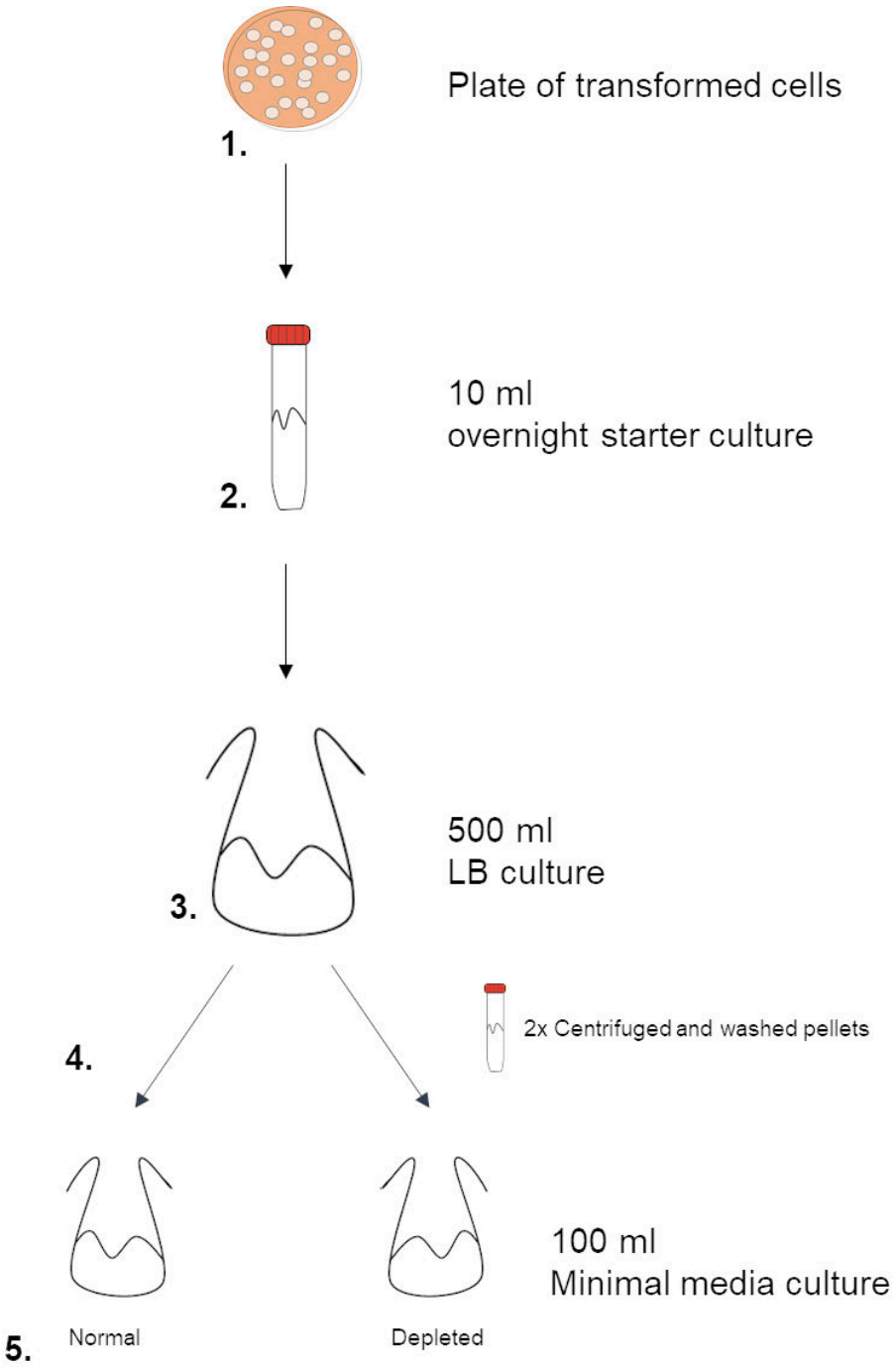


Figure 3.1: Workflow diagram of E. coli culture

The schematic workflow for effective transfer of E. coli cells from a rich media into isotopically controlled media. This workflow initially follows normal expression protocols, transformation of an expression plasmid into an expression cell line and grown on an antibiotic selective plate (1). A single colony of transformed BL21 cells is selected, and a starter culture is grown (2) and used to inoculate the rich expression media (3). To then transfer into isotopically controlled media, the cells grown in stage 3 are centrifuged to form pellets and washed with M9 salt solution (4). The final stage of the process is to resuspend the washed pellet in isotopically controlled M9 minimal media (5) where IPTG induction of protein expression can be carried out.

3.3 Optimising expression of purification of model proteins in minimal media.

Before isotope depletion in minimal media was attempted, it was first necessary to optimise expression and purification protocols for the three model proteins used in this study. Outlined below are the typical purification results for each protein expressed from a 100 ml isotopically natural M9 minimal media culture.

3.3.1 Encapsulated ferritin (EncFtn)

The EncFtn (EncFtn_sHis) from *Rhodospirillum Rubrum* was expressed [96] in rich (LB) and M9 minimal media as described previously. The protein was purified from the partially clarified cell lysate using nickel-affinity chromatography to selectively bind to the N-terminal hexahistidine tag. The purification trace from the ÄKTA start is shown in Figure 3.2. In Figure 3.2a, we can see the purification when carried out on an *E. coli* culture grown and induced in a 100 ml culture of LB media. The sample loading stage refers to the mass loading of filtered cellular lysate after sonication to disrupt the *E. coli* cell pellets and partial clarification to remove cellular membranes and organelles. This is composed of endogenous *E. coli* protein which has a limited capability to bind to the affinity column. The imidazole elution refers to the gradual increase of the elution buffer, which in nickel affinity purification requires a high concentration of imidazole. At an imidazole concentration of approximately 275 mM provided in the elution buffer, the tagged EncFtn elutes, with a peak maximum at around 1955 mAu.

The culture of LB media acts as a direct comparison to minimal media cultures, which are normally 100 ml in order to limit the requirement for the depleted substrates within the media. M9 minimal media has an overall limited nutrition availability, and therefore the short recovery period in stage 5 of Figure 3.1 will not support the BL21 cells to allow further cellular division events before expression was induced with IPTG. Figure 3.2b is the purification of recombinant EncFtn in 100 ml M9 minimal media. His-tagged EncFtn elutes at a similar concentration of imidazole as previously seen for the

expression of the culture grown in LB. However, there is a significant reduction in the elution peak maximum at 421 mAu. This is a 4.6-fold reduction in the total

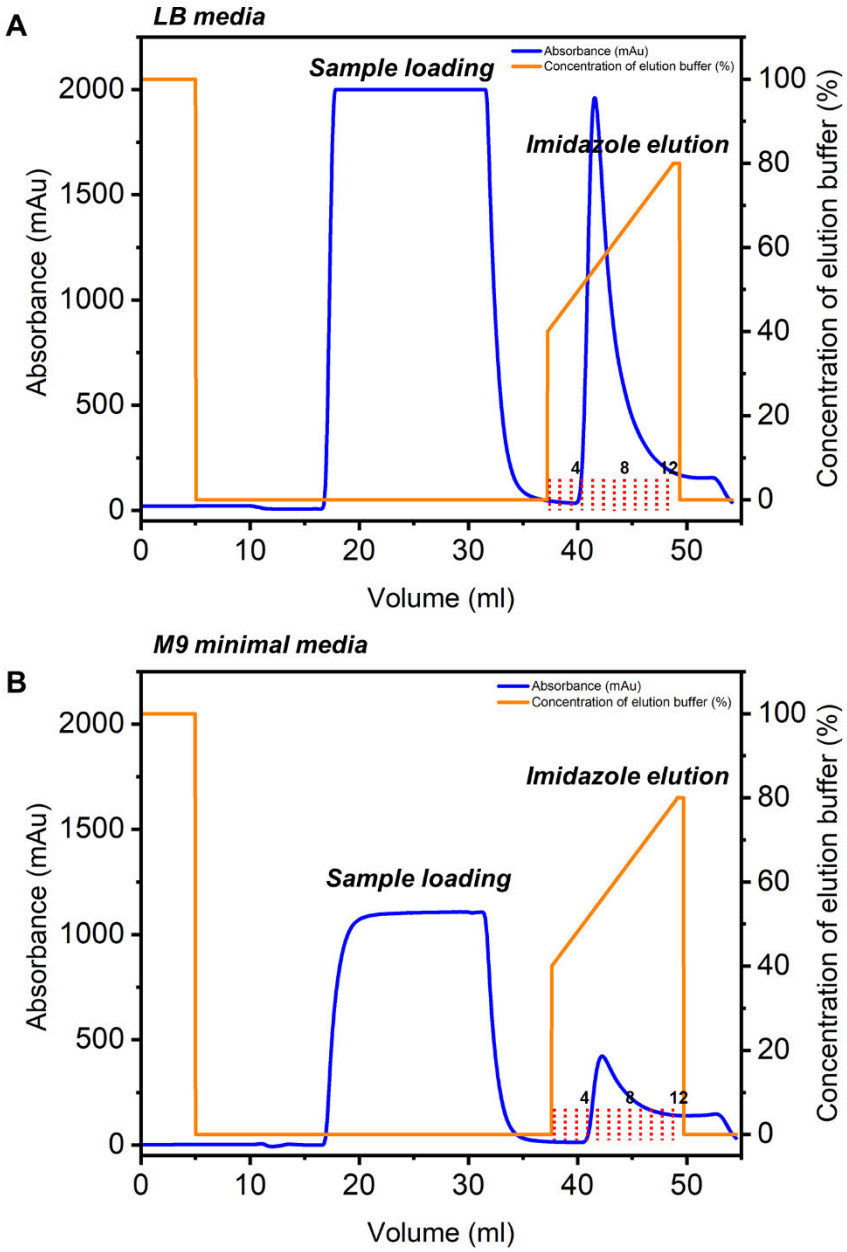


Figure 3.2: ÄKTA chromatograms for nickel affinity chromatography for EncFtn

Purification of recombinantly expressed EncFtn in (A) LB media or (B) M9 minimal media. The absorbance at 280 nm (blue line) tracked protein concentration through the purification and the concentration of elution buffer (orange line) was strictly controlled by the ÄKTA system to ensure reproducibility. The separation of each IMAC fraction indicated by the dotted line (---)

absorbance, which corresponds to similar fold reduction in the total protein yield.

While the large decrease in potential expressed protein concentration is not ideal, most mass spectrometric techniques only require micromolar concentrations of protein so ultimately even with the reduction in concentration, it should still be possible obtain a sufficient concentration of recombinantly expressed protein to allow efficient MS analysis. The content of the elution peak from the M9 minimal media culture was visualised on a polyacrylamide gel (Figure 3.3).

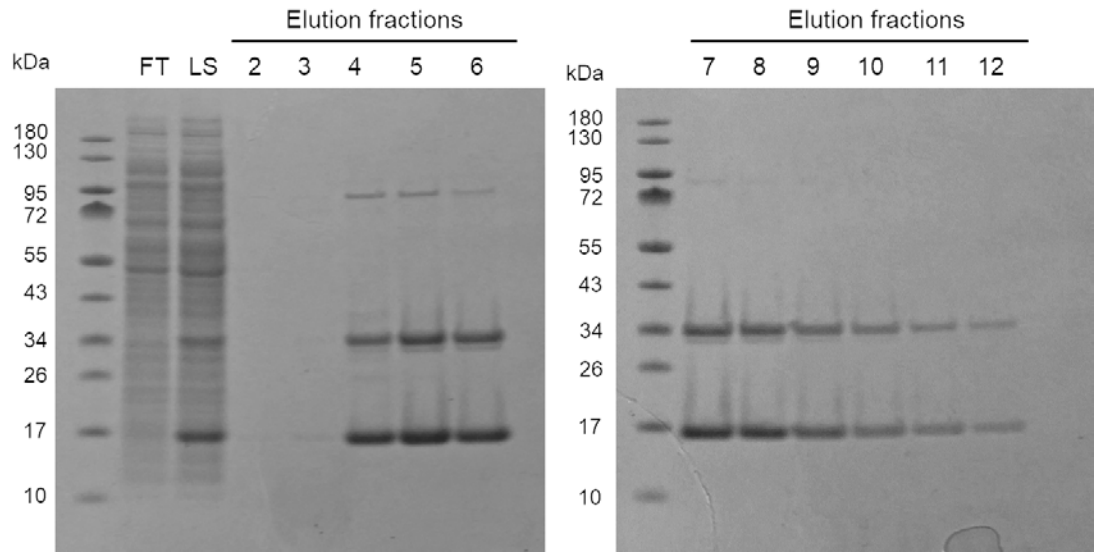


Figure 3.3: Polyacrylamide gels of eluted EncFtn IMAC fractions.

Elution fractions of EncFtn after nickel chromatography elution. Cellular lysate (LS) is loaded onto the column, flowthrough (FT) contains endogenous protein which does not bind to the resin. Bound protein was selectively eluted using an imidazole concentration over twelve elution fractions.

EncFtn can assemble into a protein homodecamer in the wild-type protein and with available free iron, however is initially expressed as a monomer (this construct is approximately 13 kDa). [96,103] Figure 3.3 shows the expressed protein content of the in M9 minimal fractions collected during the imidazole elution stage of Figure 3.2b. EncFtn monomer is seen in the eluted fractions 4-12 near the 17 kDa marker and is

not seen in the purification flowthrough (FT) which contains protein within the cellular lysate. This typically does not bind to the column during the sample loading stage (Figure 3.2). Due to the strength of the dimer interface, this is still seen on SDS-PAGE gels at around 34 kDa [96]. A band with less intensity seen at 95 kDa which is thought to be either a low intensity contaminant or the EncFtn forming small amounts of higher order structures.

Measurement of the accurate mass of the protein visualised on the polyacrylamide gel (Figure 3.3) was performed by mass spectrometry. LCMS is a fast efficient method for protein mass determination, using a C4 reverse phase column (50 x 2.1 mm) to exchange the protein sample into the required solvent [water: acetonitrile: formic acid (49.95%:49.95%:0.1%)].The produced charge state distribution in Figure 3.4 contains the labelled monomer charge distribution acquired on a Q-TOF, and a lower abundance of the dimer distribution is also observable.

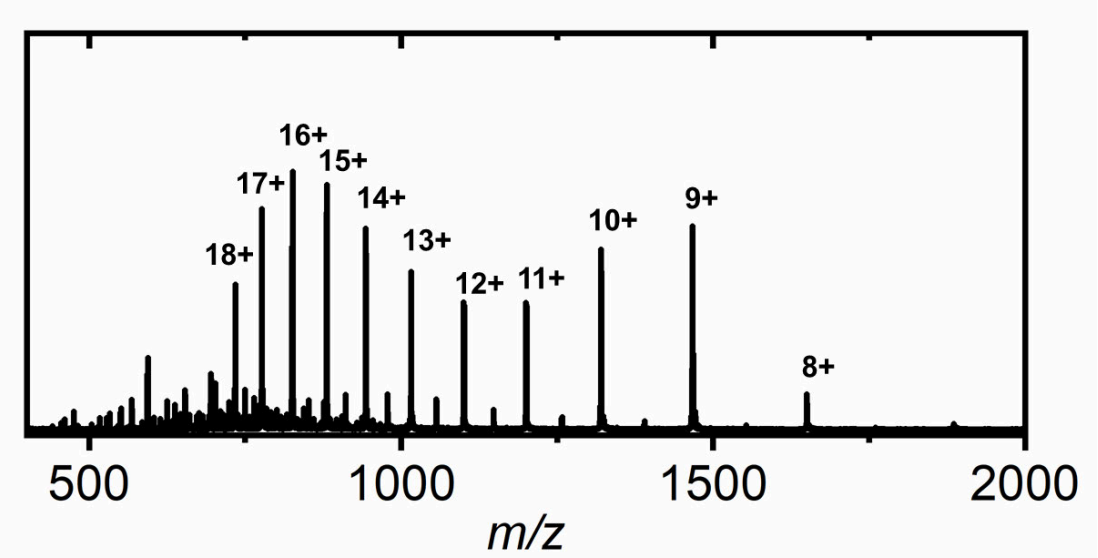


Figure 3.4: Mass confirmation of EncFtn

LC-Q-TOF charge state distribution of purified EncFtn expressed in M9 minimal media. The monomer peaks are labelled with the appropriate charge state. Dimer peaks seen between the monomer peaks.

The deconvoluted average mass of EncFtn in Figure 3.4 is 13193.78 Da, which is consistent with the theoretical average mass 13194.33 Da, providing confidence that the expressed protein matches the predicted amino acid sequence.

3.3.2 Carbonic anhydrase (BCA)

Bovine carbonic anhydrase (BCA_HisTEV) was expressed in 100 ml minimal media and loaded onto a nickel-affinity chromatography column to selectively purify the hexahistidine tag linked to BCA via a TEV protease cleavage site. [104] Figure 3.5a shows a typical chromatogram for the two-step purification of BCA. For the initially expressed TEV-tagged BCA protein, there is a large imidazole elution peak which saturated the detector of the ÄKTA start system (MAX. 2000 mAu). The eluted protein was collected and subjected to overnight cleavage with a recombinant His-tagged TEV protease to remove the hexahistidine tag. [104] The protein was then reloaded onto the nickel affinity column, to which the cleaved protein will have limited binding affinity to the nickel resin. This is the result observed in Figure 3.5b, in which the large sample loading stage is the loaded carbonic anhydrase without the hexahisitidine tag. By attempting to elute any bound non-cleaved protein, we can quality assess the TEV cleavage, as any protein which did not get cleaved by the protease will as in Figure 3.5a bind to the affinity resin.

BCA purification was initially observed using SDS-PAGE, showing each stage of the purification process (Figure 3.5), shown in Figure 3.6.

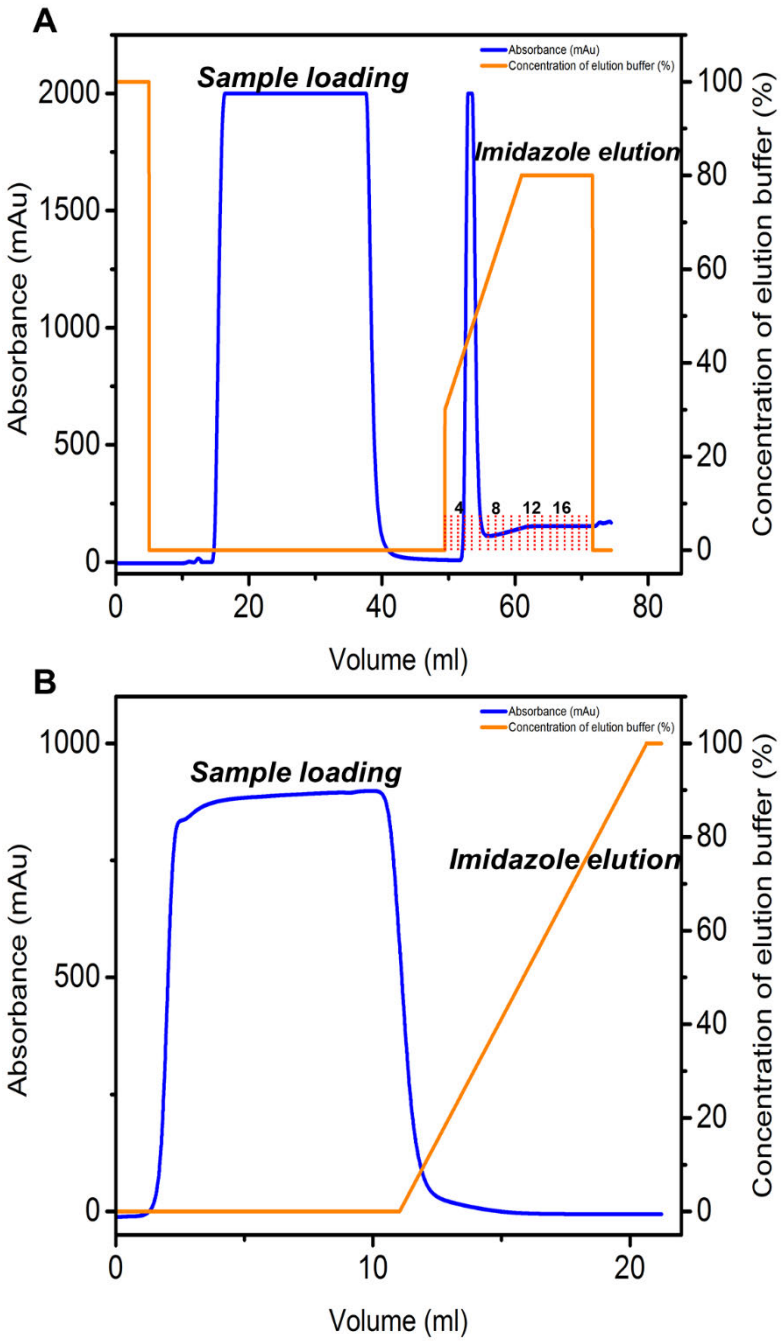


Figure 3.5: Purification of BCA

Purification of recombinantly expressed BCA using nickel-affinity chromatography both prior to (A) and subsequent cleavage (B) of the hexahistidine tag. The absorbance at 280 nm (blue line) tracked total protein concentration through the purification and the concentration of elution buffer (orange line) indicates the elution profile. The initial elution fractions are indicated by the dotted line (---).

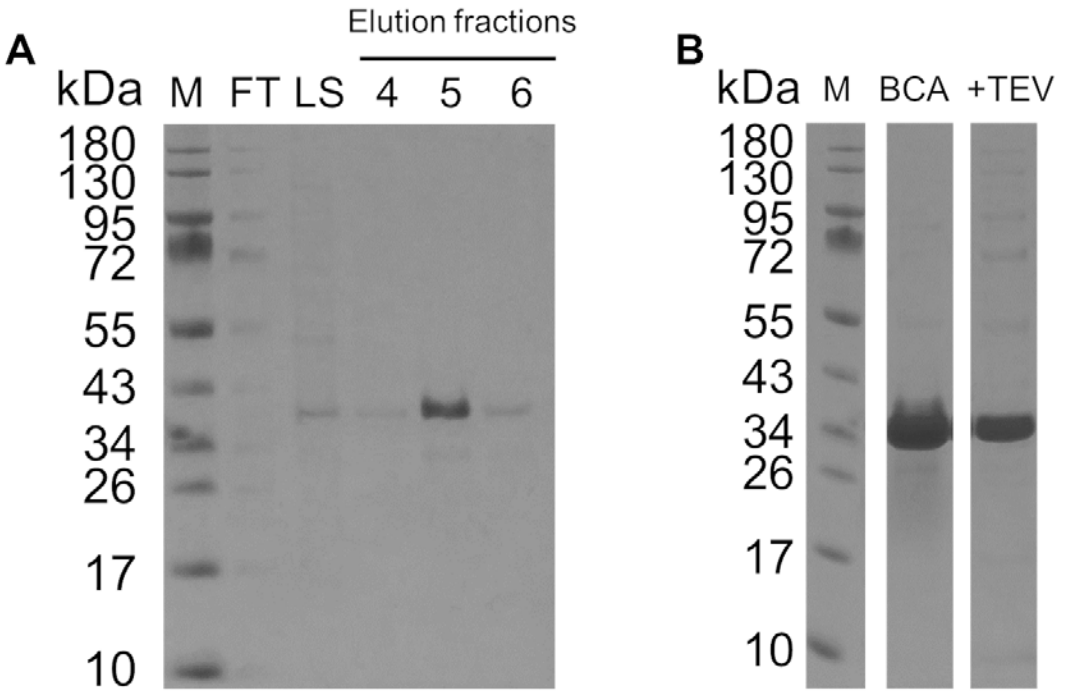


Figure 3.6: Polyacrylamide of BCA purification

Polyacrylamide gel of initial IMAC purification of BCA (A), loaded cellular lysate (LS) is loaded onto the column, flowthrough (FT) of the cellular lysate was highly diluted and contained contamination from the MW ladder (M). Bound protein was selectively eluted using an imidazole gradient is almost exclusively present in elution fraction 5. The concentrated BCA was subjected to (B) was subjected to an overnight TEV cleavage to produce the final protein.

The initial cellular lysate was purified as shown in Figure 3.5a and contained a single high absorbance elution peak. This is almost entirely contained within elution fraction 5. This fraction was used in subsequent TEV cleavage of the hexahistidine tag, changing the average mass of the protein from 32281 Da to 29313 Da. This is not clearly observable in Figure 3.6b so the MW of the purified protein after TEV cleavage was verified via LC-MS. The charge state distribution in Figure 3.7 was obtained on a Q-TOF, to measure the MW of the protein. Which was calculated with an average mass of 29311.59 Da, which is consistent with the predicted average mass of BCA at 29312.67 Da.

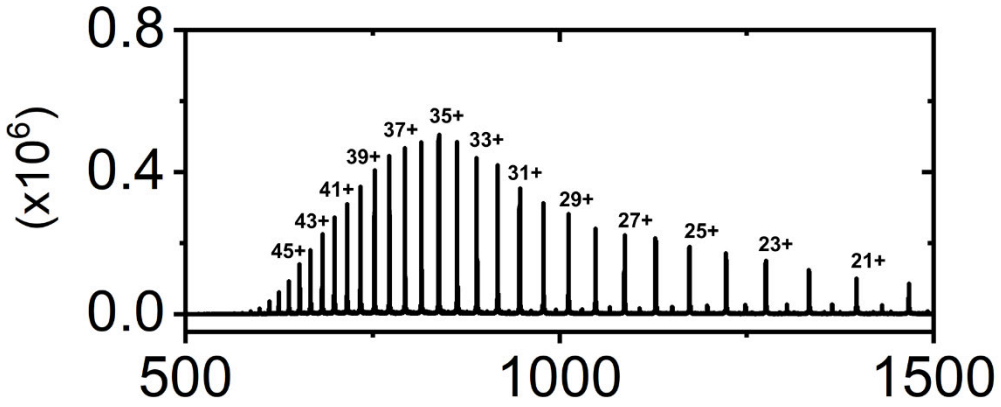


Figure 3.7: LC-TOF MS of BCA

LC-Q-TOF charge state distribution of recombinantly purified BCA after TEV cleavage of the hexahistidine tag. The odd numbered charge states are labelled.

3.3.3 Serine palmitoyltransferase (SPT)

Serine palmitoyltransferase (AmSPT) from *Sphingomonas paucimobilis* was expressed in 100 ml minimal media further supplemented with pyridoxal phosphate (PLP) which is an integral co-factor for enzymatic activity and protein stability. As with EncFtn (page 53) and BCA (page 57), the expression plasmid contains a hexahistidine tag allowing it to be purified via selective elution from a nickel-affinity column (Figure 3.8a). Due to the PLP ligand bound to the protein the fractions which contained SPT protein displayed a yellow colour. In Figure 3.8a there are two elution peaks, the first peak at 53% (159 mM imidazole) of the elution buffer and a second peak at 83% (249 mM imidazole). Both peaks were collected in fractions which visually appeared yellow. Which when analysed on a polyacrylamide gel (Figure 3.8b) showed the overexpressed SPT protein (47, 232 Da) in all the elution fractions underneath both elution peaks (elution fractions 4-10).

This is likely because SPT is known to exist as a homodimer, but when expressed in minimal media results in a mixture of protein expressed as monomer or dimer. The dimer will result in a complex which contains two hexahistidine tags, so will require a

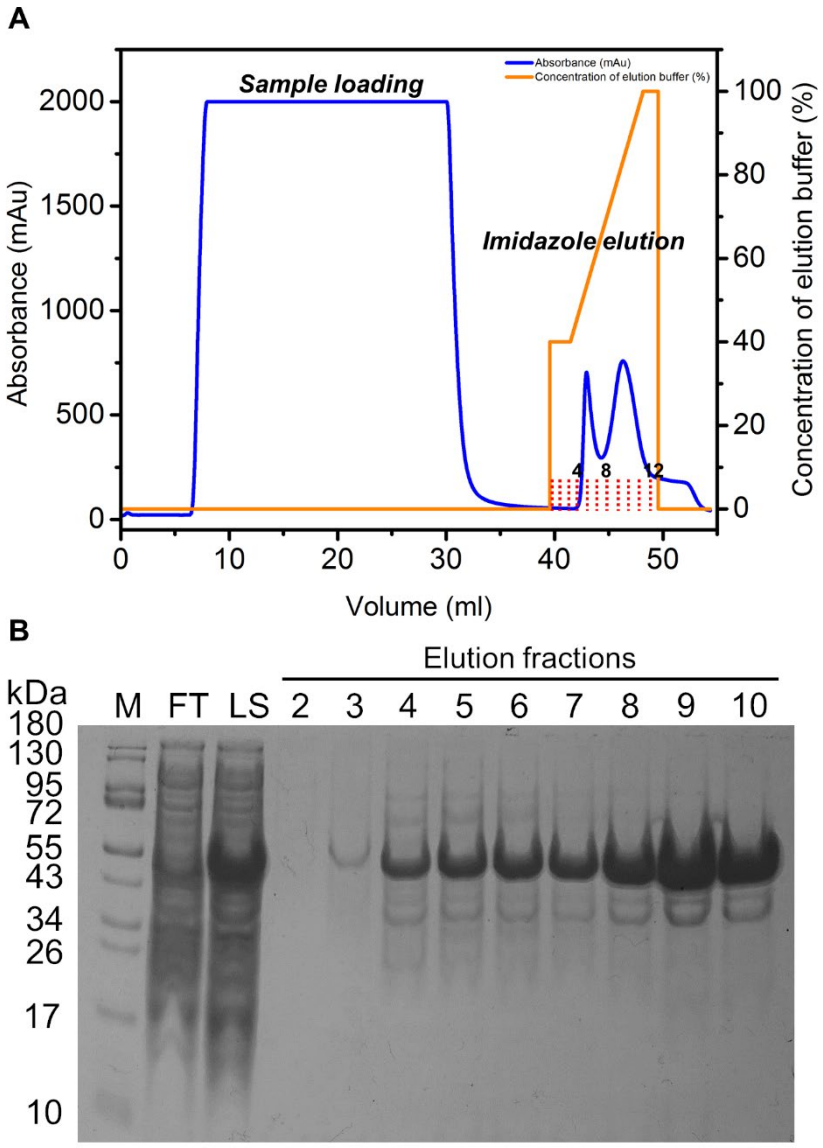


Figure 3.8: Nickel- affinity purification of SPT

Purification of recombinantly expressed SPT (47,232 Da) was carried out using nickel-affinity chromatography (A). The absorbance at 280 nm (blue line) tracked total protein concentration through the purification and the concentration of elution buffer (orange line) indicates the elution profile. The resultant elution fractions (---) were visualised on a polyacrylamide gel (B).

greater concentration of imidazole to elute from the affinity resin. Therefore, elution fractions 4-10 were concentrated in a centrifugal concentrator to be further purified by size exclusion chromatography (SEC, Figure 3.9).

As with the imidazole elution in Figure 3.8a, the protein eluted from SEC as two peaks, presumably again as a consequence of monomer and dimer forms. Both elution peaks were yellow is colour suggesting that both fractions contain the PLP-bound enzyme. LC-Q-TOF MS analysis of the eluted protein highlights the complexity of the sample.

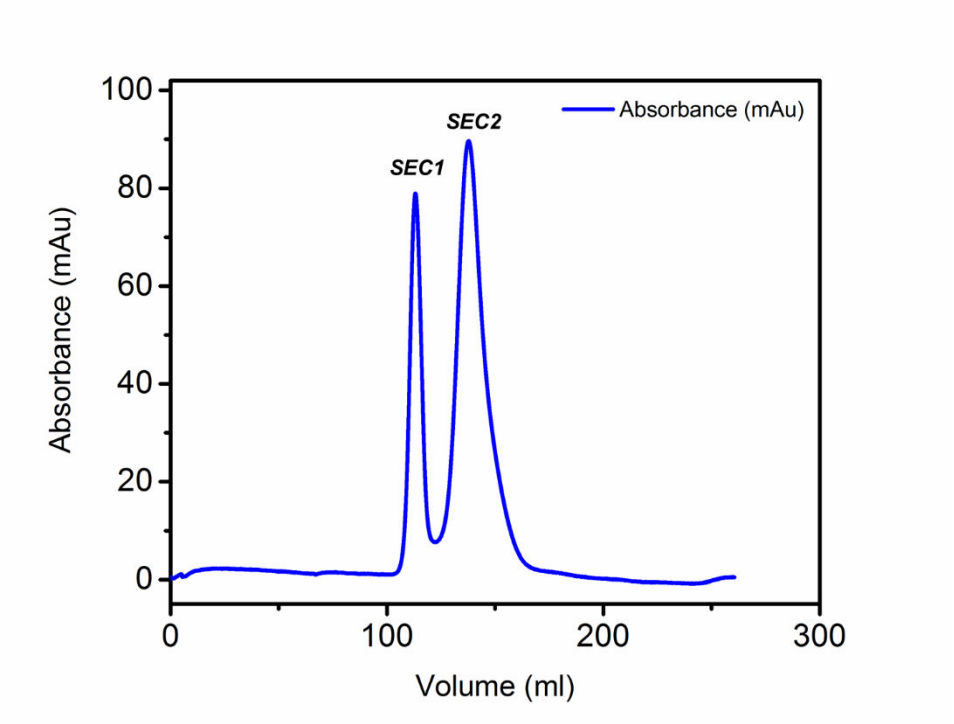


Figure 3.9: Size- exclusion chromatography of SPT

Size exclusion of collected nickel- affinity elution fractions containing SPT protein. The absorbance at 280 nm (blue line) tracked total protein concentration through the purification. Two major elution peaks are labelled sequentially as they eluted from the column (SEC_1 and SEC_2).

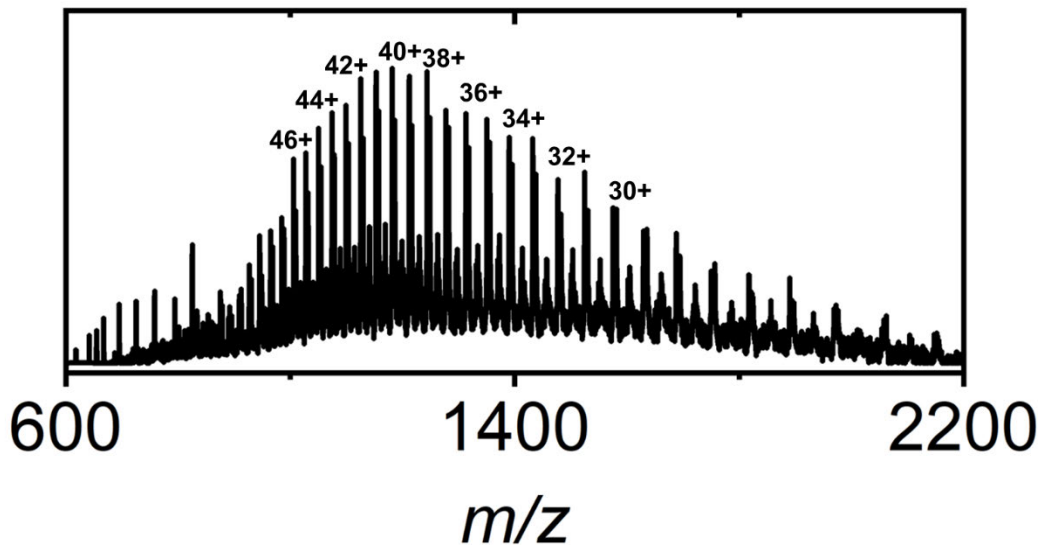


Figure 3.10: Charge state distribution of purified SPT

LC-Q-TOF charge state distribution of recombinantly purified SPT containing combined SEC1 and SEC2.The positive charge states of monomeric SPT are labelled

Figure 3.10 displays the charge state distribution of SPT, the most abundant displayed a calculated average mass of 47229.9 Da, consistent with the expected average mass of SPT at 47231.6 Da. In addition, it is clear that there is still dimer present within the spectra even with LC conditions which contains solvent and 0.1% formic acid. Closer inspection of the charge state distribution also reveals partial modification on the SPT protein (Figure 3.11).

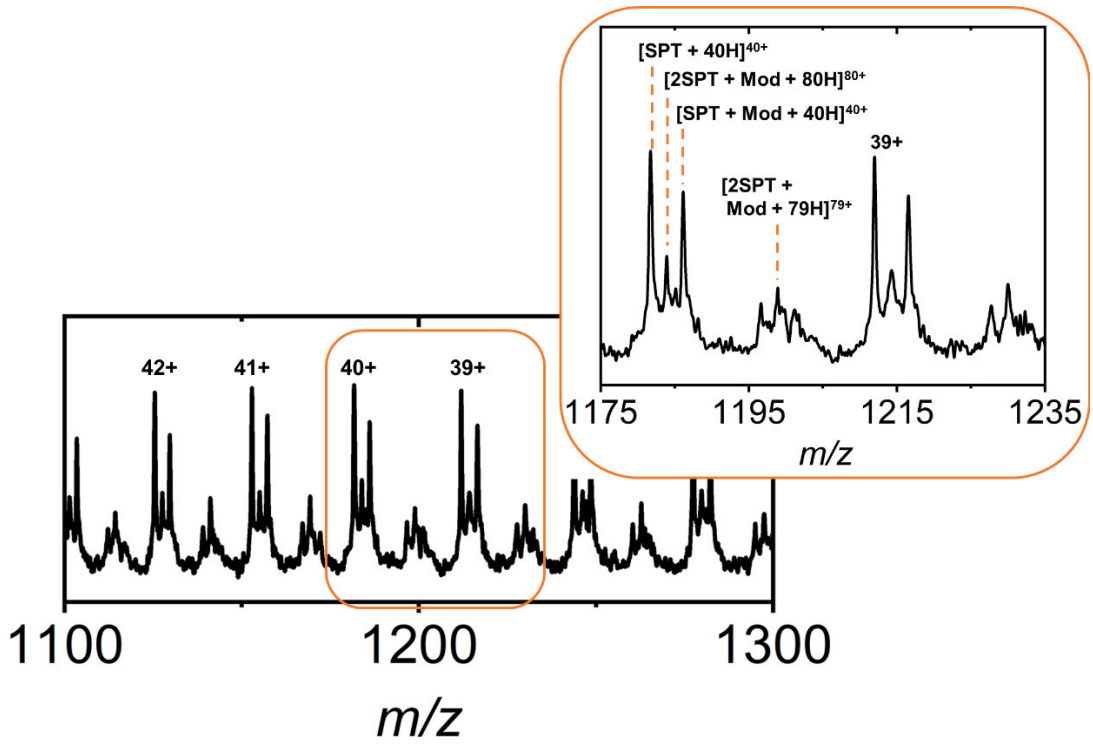


Figure 3.11: Closer view of charge state distribution

Closer view of LC-Q-TOF charge state distribution, highlighting the modification and dimer species of SPT.

The calculated MW of SPT containing the modification is 47407.9 Da, which results the increase in MW of 178 Da, which is suspected to be an α-N-6-gluconoylation ($C_6H_{10}O_6$) [105] known to affect the hexahistidine tag of protein with a known sequence (S-S-$[H]_6$-) expressed from pET28a from which this construct of SPT was expressed. This observation may explain the two-peak elution of SPT observed during nickel affinity chromatography. It maybe that the modification interferes with efficient binding of the hexahistidine tag to the immobilised nickel on the column, allowing the protein containing the tag to be eluted with a lower imidazole concentration. All downstream studies on SPT were carried out on the protein mixed with the modification.

3.4 Using established expression and purification protocols to express isotopically depleted protein.

The workflow in Figure 3.1 highlights that isotopically depleted *E. coli* cultures are produced alongside the isotopically natural minimal media culture, prepared from the same starter culture in LB media. Producing protein which are chemically identical, only differing in the ratio of heavier isotope availability. They can be purified in the same manner as described above. This strategy, while enriching for lighter isotopes is referred to as isotopic *depletion* to differentiate and prevent confusion. Due to the relative isotopic abundance found in nature (Table 1.1) and prevalence within a protein polypeptide chain [20], the two main elements to target for the process of isotopic depletion is carbon and nitrogen [88,91]. Controlled by the addition of *depleted* isotope sources, ^{12}C (99.9%)–glucose (Cambridge Isotope Laboratories) and ^{14}N (99.99%)-ammonium sulfate.

After the expression and purification procedures had been established, isotopically depleted protein samples were produced by substituting 99.9% ^{12}C-Glucose and 99.99% ^{14}N-ammonium sulfate into the workflow at stage 5 of the workflow (Figure 3.1). SDS-PAGE analysis of the purified proteins from both standard and isotope depleted preparations are shown in Figure 3.12.

Purified EncFtn (lanes 1 and 2), BCA (lanes 3 and 4) and SPT (lanes 5 and 6) are, in relation to protein purification, identical when produced in minimal media, irrespective of the isotope content (Figure 3.12). This results in purified isotopically natural protein which are the ideal control for the analysis of isotopically controlled protein. In which the only difference between the two samples are the relative abundance of heavier isotopes within the isotope sources supplied to the *E. coli* cells within the M9 minimal media.

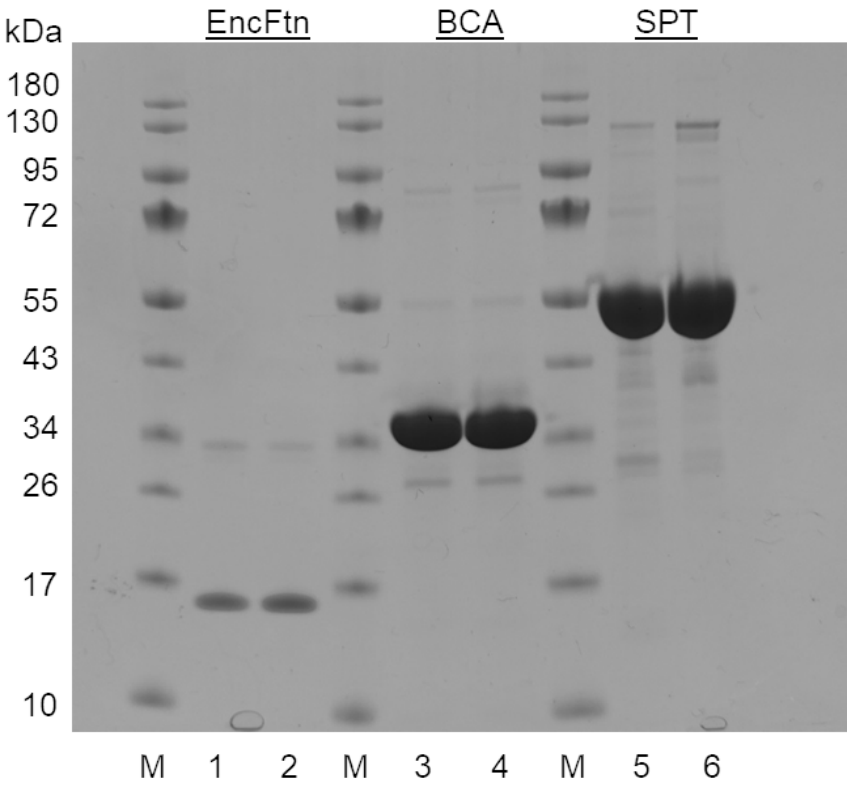


Figure 3.12: Polyacrylamide gel comparison of isotopic and depleted protein.

Purified protein from minimal media. Lane 1 natural isotopic abundance EncFtn, Lane 2 isotope depleted EncFtn (13.2 kDa), Lane 3 natural CA, Lane 4 depleted CA (29.3 kDa) Lane 5 natural SPT and Lane 6 depleted SPT (47.2 kDa). Both natural and depleted protein were expressed in M9 minimal media and were purified in an identical manner.

3.4.1 Single isotope depletion

The resultant change in the isotopologue profile caused by isotopic depletion can be observed within a mass spectrum. As described in section 1.1.4 (page 8) the shape of the peak is dictated by the chemical formula and the availability for heavier atom to incorporate into a protein. The charge state distribution was measured on a Q-TOF (Waters) which does not provide sufficient resolution to allow visualisation of the isotopic peaks. Therefore, the studies into the impact of isotopic depletion on the isotopologue distribution required analysis using a 12T FT-ICR (Bruker). In

comparison to the charge state distribution in Figure 3.4, the charge state distribution in Figure 3.13 was collected using nESI on the FT-ICR. The calculated average mass from the monomer distribution is the same as that calculated in Figure 3.4.

One of the advantages of acquiring spectra with greater resolution is measurement of accurate mass, calculated on the identified monoisotopic peak. The 10^+ charge state in Figure 3.13 (outlined in orange) has an average *m/z* value of 1320.6 and is predicted by the computation deconvolution software as having a monoisotopic molecular mass of 13,187.45 Da (Bruker's DataAnalysis). This is similar to the EncFtn theoretical monoisotopic of 13,186.45 Da (error 76 ppm) and can be further validated as the correct protein using alignment of a simulated model of the chemical formula. The associated error is larger than expected, but this is largely due to bias of the SNAP algorithm and misidentification of the monoisotopic peak within the isotope

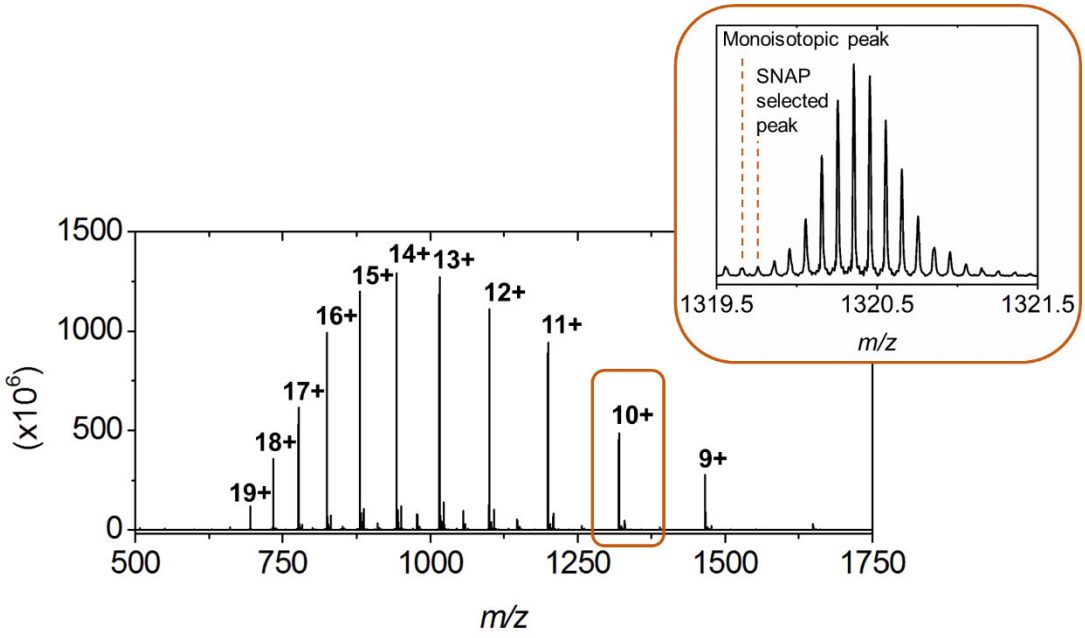


Figure 3.13: ESI FT-ICR MS of EncFtn

Protein sample was infused into FT-ICR to verify the recombinantly expressed EncFtn (13.2 kDa). The 10^+ charge state (highlighted in orange) and shown in greater detail in the panel. The monoisotopic peak and the peak calculated as the monoisotopic using deconvolution software are both highlighted.

distribution by one isotopologue. This assignment error may also be explained due to the introduction of a deamidation PTM (+0.98402 Da), [106] occurring primarily on asparagine and glutamine residues, modifying the total mass of the protein. This can be identified with top-down fragmentation and assigning fragment ions containing the mass shift (+0.98402 Da). With further downstream analysis of the protein (i.e., various fragmentation techniques applied to EncFtn) producing fragments consistent with the unmodified form of the protein.

For larger peptides and proteins the misidentification of the SNAP monoisotopic peak becomes a common issue with this type of analysis, and can result in the introduction of error. [95] Using EncFtn as an example we can investigate the impact of isotopic depletion on the protein spectra. As discussed previously, incorporation of isotopically depleted isotopes into the M9 culture media (Figure 3.1) will result in incorporation of depleted isotopes into the polypeptide chain. Therefore, we can test the incorporation of single isotopes and observe the change (Figure 3.14).

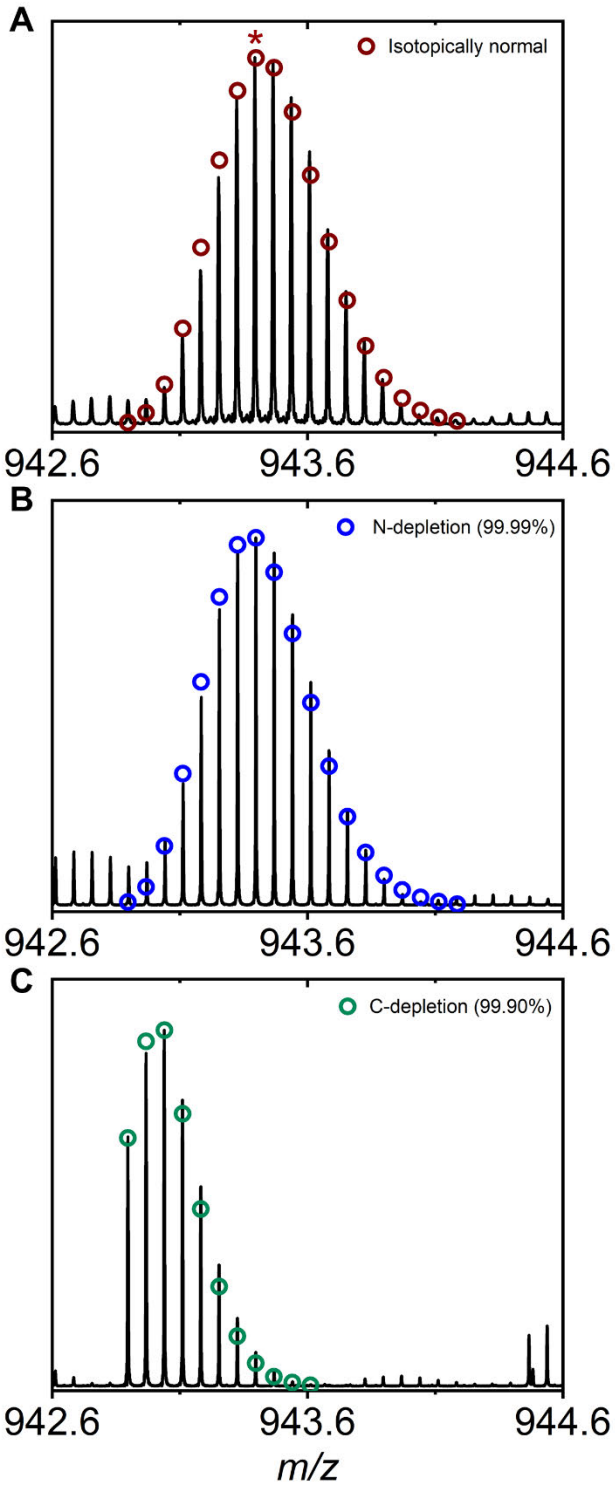


Figure 3.14: A comparison of the 14+ charge state isotope peak of EncFtn after recombinant expression in isotopically depleted media.

EncFtn expressed within isotopically natural media (A) displays a predictable Gaussian profile similar to the theoretical model (o). Expression of EncFtn with depleted isotope sources; nitrogen (B) and carbon (C) produce a different isotopologue distribution which also can be compared to a computational model (o - 14N 99.99%, o - 12C 99.90%).

Figure 3.14 shows the impact of isotopic depletion of a single element on the isotopologue structure. EncFtn ($C_{572}H_{884}N_{172}O_{185}S_2$) was expressed in M9 minimal media, as described previously in figure 3.1. The isotopically natural sample (Figure 3.14a) exhibits a near Gaussian distibution, typical of intact protein. This distribution contains several isotopologue peaks, which span an observable range in mass of 14 Da. Each the result of statistical incorporation of heavier isotopes, [20] which is closely matched by the predicted averagine model (represented by the overlaid scatter plot, o). The most intense peak (*, 943.398 *m/z*) contributes 13.6% of the total signal of the charge state, and is 1.1 times greater in intensity than the isotopologue 1 Da lighter (943.326 *m/z)* which contributes 12.4% to the total charge state signal. By providing an isotopically depleted nitrogen source (99.99% ^{14}N) to the minimal media (Figure 3.14b) we can observe a very subtle change in the isotopologue distribution. It continues to span a 14 Da range, and the most abundant peak is the reciprocal to that in the isotopically natural spectrum (*, 943.398 *m/z*) which similarly contribtes 14.1% of the total signal. The difference becomes apparent in the lighter isotopologue peak (943.326 *m/z)* which contributes 13.8%, and so is very similar to the most intense peak which is only 0.08 times greater in intensity. Likewise the lighter isotopologue all contribute a greater percentage to the total than the reciprocal isotopologue in the isotopically natural sample.

A more dramatic change is observed within the isotopologue distribution as a result of carbon depletion (Figure 3.14c), the distribution is asymmetic and more narrow than that of the isotopically natural, decreasing from 14 Da to 8 Da, as a result of there being almost half the number of isotoplogue peaks. The most intense peak is the M+2 peak which contributes 22.5% of the total signal. With the first five isotopologue peaks contributing almost 90% of the entire charge state signal. The reciprocal five peaks contribute 12% and 16% in isotopically normal and nitrogen depleted samples.

Another clear advantage of the carbon depletion sample is the significant increase in the monoisotopic peak signal. As the intensity of the monoisotopic peak in isotopically normal samples is typically very low (0.06% in EncFtn) it can be misidentified within the background noise (Figure 3.13). In Figure 3.14c the monoisotopic peak is clearly observable, and allows accurate determination of the monoisotopic mass of the protein (13186.45 Da, error -0.12 ppm). The abundance of the monoisotopic peak also shows that in Figure 3.14c the EncFtn sample has not undergone significant deamidation, without the need for further fragmentation.

Figure 3.14c displays a isotopologue distribution which matches closely to the predicted model (^{12}C, 99.90%), we have observed that the level of observed isotopic depletion is batch variable, and in most samples we observe carbon depletion greater to that seen in the figure. This is particularly evident in our samples which have been recombinantly expressed in minimal media which is isotopically depleted in both carbon and nitrogen (^{12}C, 99.90%; ^{14}N, 99.99%), as the technique to exchange BL21 cells into minimal media became further refined. Within our doubly depleted protein, as seen in Figure 3.15b, displays an isotopologue more consistant with the theoretical model with a greater level of carbon depltion than expected (o - ^{12}C 99.95%, ^{14}N 99.99%).

3.4.2 Double isotope depletion

Carbon and nitrogen double depletion has a more dramatic impact on the isotopologue distribution than single depletion. The double depletion EncFtn protein displays an isotopologue distribution which only spans a range of 7 Da, while carbon depletion alone (Figure 3.14c), has a distribution which spans 8 Da. Within the isotopologue distribution in Figure 3.15b the monoisotopic peak which is the most abundant peak contributes 38.4% of the entire signal, and 80% of the entire signal contained within the first three isotopologue peaks. As explained in Figure 1.5, the

M+1 and M+2 peak are a result of heavier isotope incorporation. But are largely a result of the contribution of one or two ^{13}C isotopes. Which ultimately means, within a perfect system that the majority of the signal in double depleted (~80%) is produced

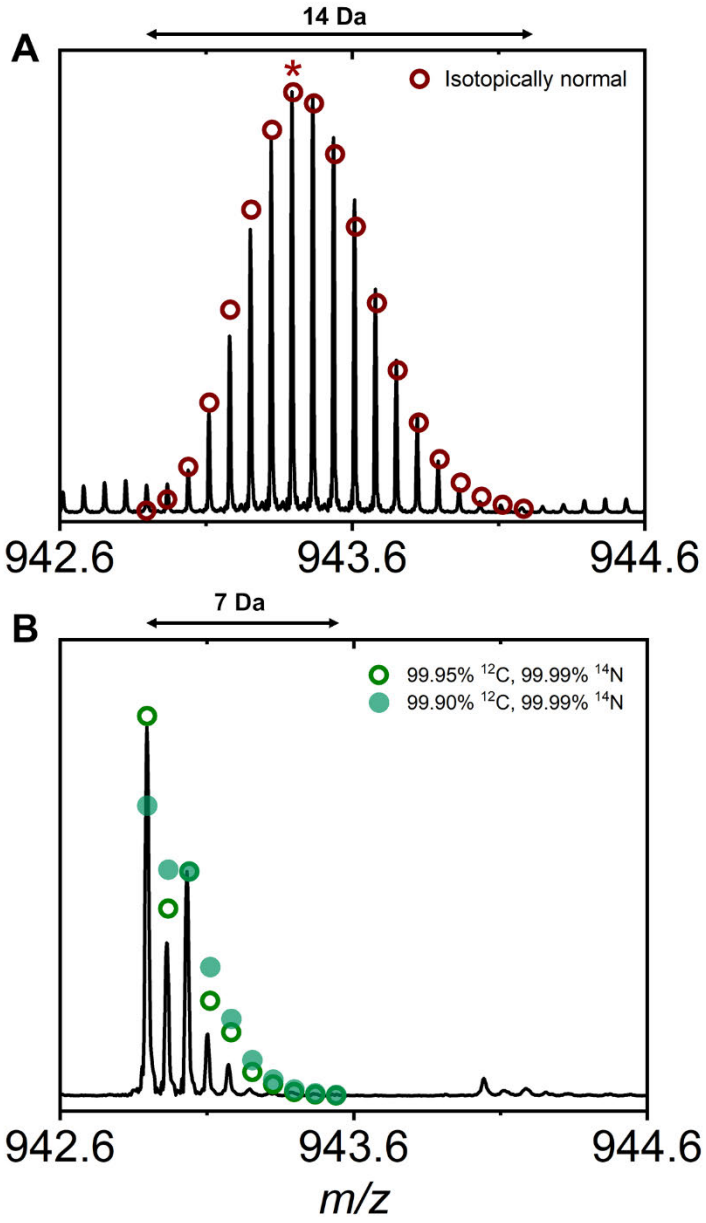


Figure 3.15: Comparison of the 14+ charge state isotopes peak of EncFtn after recombinant expression in isotopically double depleted media.

EncFtn expressed within isotopically natural media (A) displays a predictable Gaussian profile similar to the theoretical model (o). Expression of EncFtn with two depleted isotope sources; nitrogen and carbon (B) produce a different isotopologue distribution which can be compared to theoretical models with varying levels of isotopic depletion (● - 12C 99.90%, 14N 99.99%; o - 12C 99.95%, 14N 99.99%).

by a polypeptide chain with the addition of no ^{13}C (38.4%), one ^{13}C (18.9%) or two ^{13}C (22.7%) out of a possible 572 carbon atoms within the chemical formula of EncFtn.

3.4.3 Sulfur depletion

While during the scope of this investigation isotopic depletion of only carbon and nitrogen was attempted. The effect of sulfur depletion was partially investigated through the mutation of methionine (-$C_5H_9NO_2S$-) residues 30 and 68 to leucine (-$C_6H_{11}NO_2$-) residues, therefore removing the opportunity of any heavy stable sulfur isotope incorporation (^{33}S, 0.76%; ^{34}S, 4.29%).

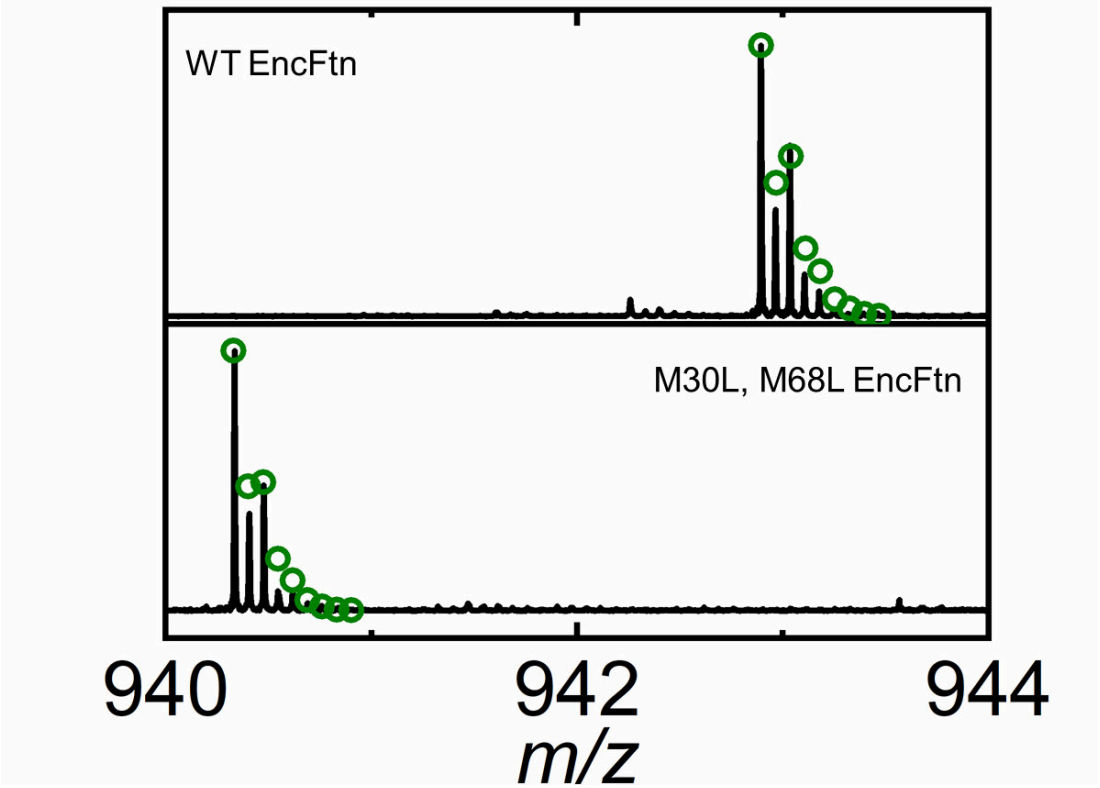


Figure 3.16: Comparison of isotopologue distribution of double mutant EncFtn to WT EncFtn.

The 14+ charge state of both isotopically depleted WT EncFtn ($C_{572}H_{884}N_{172}O_{185}S_2$) and double mutant EncFtn ($C_{574}H_{888}N_{172}O_{185}$). Both proteins were expressed isotopically depleted M9 minimal media. The theoretical model of the protein is overlaid (o - ^{12}C 99.95%, ^{14}N 99.99%).

The M30L, M68L EncFtn has a reduced mass of approximately 36 Da compared to the WT EncFtn, causing the left shift of the isotope distribution. As WT EncFtn only contains two sulfur atoms, we can only expect a minimal change in the overall distribution. The most abundant heavy isotope for sulfur is ^{34}S, present in the environment at 4.29%, will have the greatest impact on the the M+2 peak of the distribution. Therefore as a result of sulfur removal the M+2 peak in Figure 3.16 reduces in height from 23% in the WT isotoplogue distribution to 21% in the double mutant distribution. The relative contribution of the monoisotopic peak to the total signal increases from 38.4% in the WT EncFtn to 42.4% in the mutant EncFtn. With sulphur depletion the M+2 peak is mainly composed of ^{18}O, meaning that oxygen depletion would also need to be required to reduce the M+2 peak further.

With each successive element required to be depleted increases the cost and the technical difficulty to produce depleted polypeptide. The main sulfur source is within the ammonium sulfate, which would also need to contain depleted nitrogen. Oxygen depletion is potentially more tricky, and would require isotopically depleted water but ^{18}O could still be taken up from the air, making is difficult to have suffient levels of depletion without having a controlled air supply. This would require investment in infrastructure, making oxygen depletion potentially very expensive.

3.5 Isotopic fine structure analysis of double depleted protein

The elemental composition of the doubly depleted WT EncFtn was further investigated by probing the isotopic fine structure, explained in detail in section 1.1.8 (page 18). To achieve the fine structure of a polypeptide requires a high-resolution mass spectrum. This is more achievable with lower mass, so prior to analysis EncFtn was subjected to an enzymatic digest. An abundant digest ion $[EHAAMTLEWLR]^{3+}$ ($C_{60}H_{93}N_{17}O_{17}S$) was analysed using 8 megaword transient size (FID recorded for 3.2s

[107]) to facilitate ultra-high resolution. As this peptide contains a sulfur atom, it gives us an opportunity to investigate the role of sulfur in the fine isotopic structure.

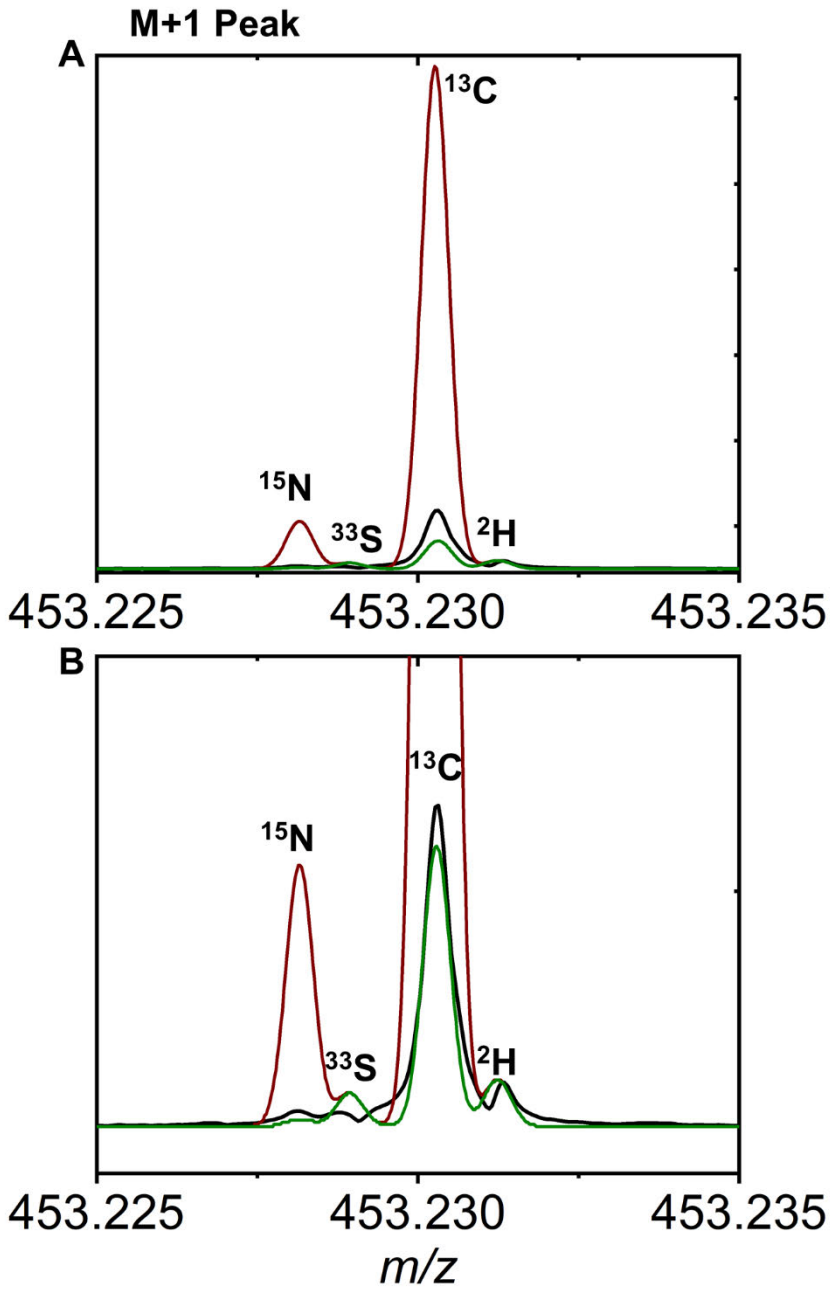


Figure 3.17: Fine structure of M+1 peak of digest ion from EncFtn.

The M+1 isotopologue peak of a tryptic peptide from EncFtn acquired at high resolution (black line), overlaid with isotopically natural (red line) and isotopically depleted (green line, 12C 99.90%, 14N 99.99%) produced using the chemical formula of the ion (C60H93N17O17S). The isotope responsible for each observable fine structure peak is labelled. The full spectra view (A) highlights the difference in 13C abundance between models, and a closer view of the lower abundance fine structure peaks (B) within the same spectra allows greater detail to be seen.

Within the M+1 peak of the isotopically double depleted digest ion we can compare the acquired data of the digest ion (black line) to the theoretical model, as above, with isotopically natural (red line) and isotopically double depleted (green line). From Figure 3.17a, it is clear that within the isotopically depleted sample, the abundance of ^{13}C is far below that of the tall theoretical peak expected in the isotopically natural sample. The difference in height of the ^{15}N peak is more obvious within Figure 3.17b which the change in y-axis allows greater clarity of the lower signal region of the same mass spectrum. The acquired spectrum does not exactly match either the isotopically natural or isotopically double depleted theoretical model lines. But shows a far greater similarity to the isotopically depleted model. Interestingly, the digest was carried out on EncFtn from the same batch as that shown in Figure 3.15b, which is shown to be more consistent with the greater carbon depletion model (^{12}C 99.95%, ^{14}N 99.99%). The model in Figure 3.17 was more consistent with the lower carbon depletion ^{12}C 99.90%, ^{14}N 99.99%), but this highlights the general variability of protein samples. For instance, in Figure 3.17b the ^{2}H peak in the isotopically natural model, double depleted model and the acquired spectra all overlay and have the same peak height. However, in the ^{33}S peak, the acquired spectra is lower than both models. But this is likely due to the resolution achieved in the acquired tryptic peptide spectra ~900,000 not being sufficient to fully resolve the ^{33}S peak from the ^{13}C peak.

As we progress through the isotopologue peaks, they are composed of a more complex array of heavier isotope additions, either a single isotope adding two mass units, or a mixture each adding one. This is observed within the M+2 peak (Figure 3.18) of the same ion ($C_{60}H_{93}N_{17}O_{17}S$). Figure 3.18, as in Figure 3.17, shows the overlaid models on the acquired spectrum data from isotopically double depleted sample of digested EncFtn. In this case the largest peak in the isotopically natural model is due to the addition of two ^{13}C isotopes which more than double the height of any of the other fine structure peaks (Figure 3.18a).

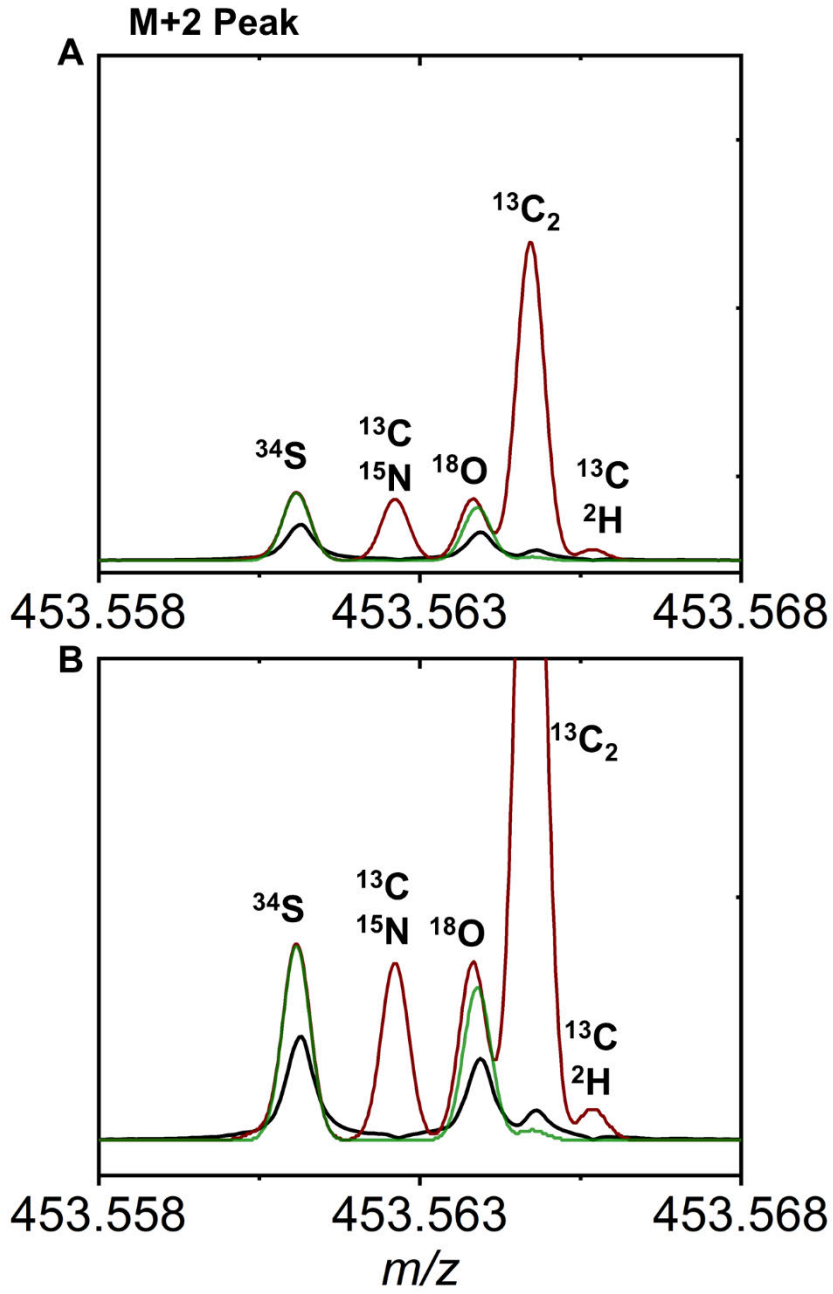


Figure 3.18: Fine structure of M+2 peak of digest ion from EncFtn.

The M+1 isotopologue peak of a digest ion from EncFtn acquired at high resolution (black line), overlaid with isotopically natural (red line) and isotopically depleted (green line ^{12}C 99.90%, ^{14}N 99.99%) produced using the chemical formula of the ion ($C_{60}H_{93}N_{17}O_{17}S$). The isotope responsible for each fine structure peak is labelled. The full spectra view (A) highlights the difference in ^{13}C abundance between models, and a closer view of the lower abundance fine structure peaks (B) within the same spectra allows greater detail to be seen.

In Figure 3.18b, we can clearly the peak formed as a result of the addition of $^{13}C/^{15}N$ within the isotopically natural model is non-present in the acquired spectrum (black line). Similarly, the peak caused by the addition of $^{13}C/^{2}H$ is reduced as a result of active carbon depletion. Further reduced by the probability of the digested ion with ^{2}H incorporation (Figure 3.18) will also contain a ^{13}C isotope within the same digest ion amino acids. The height of the M+2 peak, as mentioned before, in the isotopically depleted sample and model is dictated by the ^{34}S and ^{18}O peaks as there is such low abundance of the $^{13}C_2$ peak (green and black line). So therefore, in order to further reduce the contribution of the M+1 peak greater levels of depletion of ^{13}C would be required and for the M+2 peak, to provide isotopically depleted sulfur and oxygen sources.

Again interestingly, the acquired spectra show that elements which have not had isotope availability altered also do not match the theoretical model. This is likely, as mentioned before, as this is the same digest ion, due to numbers of sulfur and oxygen within the peptide, and the statistical probability of heavier isotope incorporation. It is also potentially a result of the separation achieved by the theoretical models. It would be an interesting experiment to achieve the required resolution, R> 6,000,000 which would be required to resolve the ^{13}C peak from the ^{15}N fine structure peak in the M+1 peak of intact EncFtn (13.2 kDa) protein. Acquisition of the intact protein spectrum with R> 14,000,000 would be required to differentiate the ^{13}C from the ^{2}H fine structure peak, which so far has not been accomplished. [108,109]

Figure 3.15, shows the effect of double isotope depletion on EncFtn which is a relatively small protein. As protein mass increases, there are a greater total of carbon atoms. Meaning that even with the 99.95% ^{12}C carbon source there will still be a significant number of ^{13}C atoms incorporated. This number will continue to increase with larger MW protein, which become more difficult to analyse. Isotope depletion

should theoretically have a greater impact on mass spectrometry of larger protein. As seen in Figure 3.12, the protein expressed in isotopically depleted M9 minimal media can be purified in an identical manner to that expressed in isotopically natural media. The comparison of isotopically natural to isotopically double depleted BCA and SPT are shown in Figure 3.19.

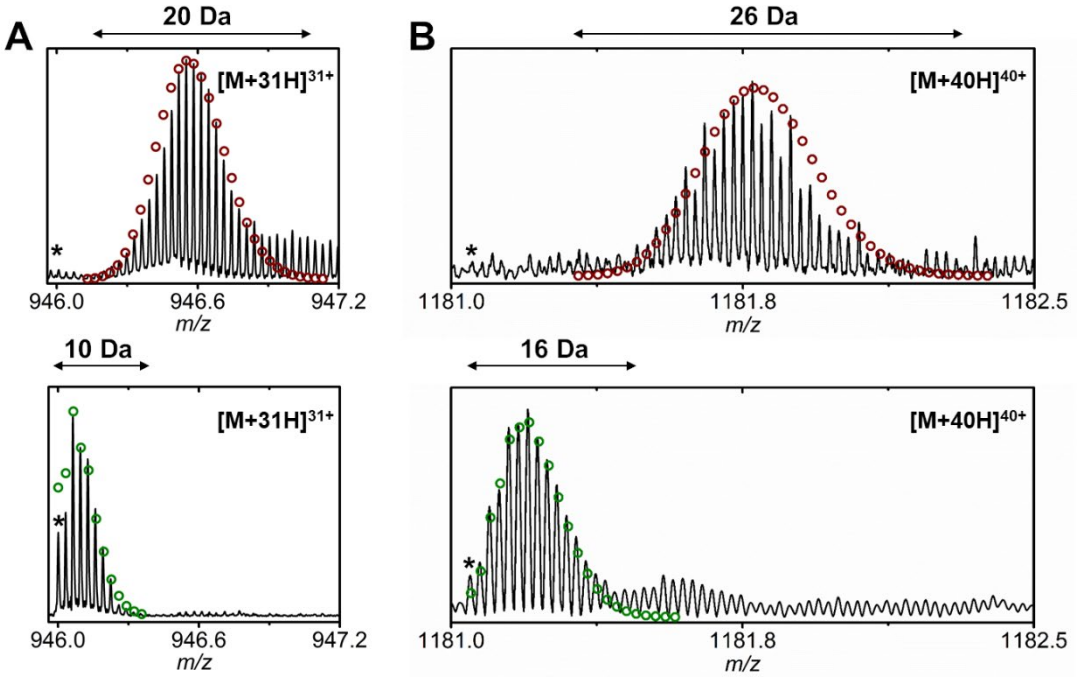


Figure 3.19: Comparison of larger protein after recombinant expression in isotopically double depleted media.

Comparison of BCA (A) and SPT (B) expressed with isotopically natural and isotopically depleted conditions. The isotopologue distributions are overlaid with the theoretical models (o $^{-12}$C 98.89%, ^{14}N 99.63%; o $^{-12}$C 99.95%, ^{14}N 99.99%) and the monoisotopic peak highlighted with an asterisk (*).

The distribution of isotopically depleted BCA (Figure 3.19a) and SPT (Figure 3.19b) again shows, like in EncFtn (Figure 3.15), a reduction in the width of the isotopic profile from 20 Da to 10 Da and 26 Da to 16 Da respectively. In the isotopically natural species of each protein the monoisotopic peak (*) is not observed due to the statistical improbability of a monoisotopic species at higher MW. In both isotopically depleted

BCA and SPT the monoisotopic peak is observable and accounts for ~13% and ~2% of the total isotope envelope. Overall, we can instinctively see the simplification of the isotopic profile, the leftward shift and overall simplification of the isotopically depleted profiles should simplify and improve our ability to study larger protein via mass spectrometry-based techniques.

3.6 Conclusion

Using the recombinant expression system outlined in Figure 3.1 we can quickly and effectively produce protein with controlled isotope abundance. Which can be purified identically to isotopically natural protein counterparts, and the only difference is reduced availability of ^{13}C and ^{15}N to 0.05% and 0.01% respectively. This limited incorporation of heavy isotopes simplifies the isotopic distribution observed within their mass spectrum. The difference in heavier isotope abundance has been demonstrated at the fine structure level of a short tryptic digest peptide from EncFtn. This recombinant expression method is a robust method and has been successfully applied to a range of protein.

While there is a visual difference in the produced isotopologue distribution between the isotopically natural and isotopically depleted protein. Predominantly observed in the increase in the abundance of the monoisotopic and lighter isotopologue peaks. Its contribution to MS analysis needs to be further assessed.

4. Isotopic depletion and the improvement for top-down fragmentation analysis.[1]

[1] The data in this chapter has previously been compiled into a manuscript published in JASMS (Appendix 4).

4.1 Top down fragmentation of depleted protein

With the expression and purification of isotopically depleted protein readily available, the usefulness of isotopic depletion as a technique in mass spectrometry can be assessed. Top-down fragmentation analysis is the common name for the application of one or more of several fragmentation techniques on an individual isolated intact protein species (proteoform) from the MS charge state distribution. It is applied to protein sample in a variety MS-based experiments including structural and proteomic scale investigations. [60,110–113] The multiple fragmentation technologies available make it an extremely versatile technique.

In this chapter, the direct comparison of the fragmentation spectra and assignable ions of isotopically natural protein in comparison to the isotopically depleted equivalent protein has been performed. The results highlight the potential usefulness of isotope depletion for improving top-down fragmentation analysis.

4.2 CID fragmentation of EncFtn

CID, as described previously in section 1.2.1 (page 28), is a fragmentation technique which results in the production of *b*- type and *y*- type ions of varying molecular weight. It is an easily incorporated technique and is the most commonly applied fragmentation technique. [114–117] The technique is also commonly applied at low voltage to improve fragmentation with any of the other available fragmentation techniques – for example in concert with ECD, in a process known as ion activation (IA-ECD). [118]

The $[M+16H]^{16+}$ charge state of isotopically natural and isotopically depleted EncFtn (13 kDa) was subjected to CID fragmentation. In Figure 4.1, the fragmentation pattern of isotopically natural (Figure 4.1a) and isotopically depleted (Figure 4.1b) have remarkably similar fragmentation spectra. With many of the same major ions

appearing between 800-1200 *m/z*. The isotopically depleted ions only differing by being present with a greater S/N.

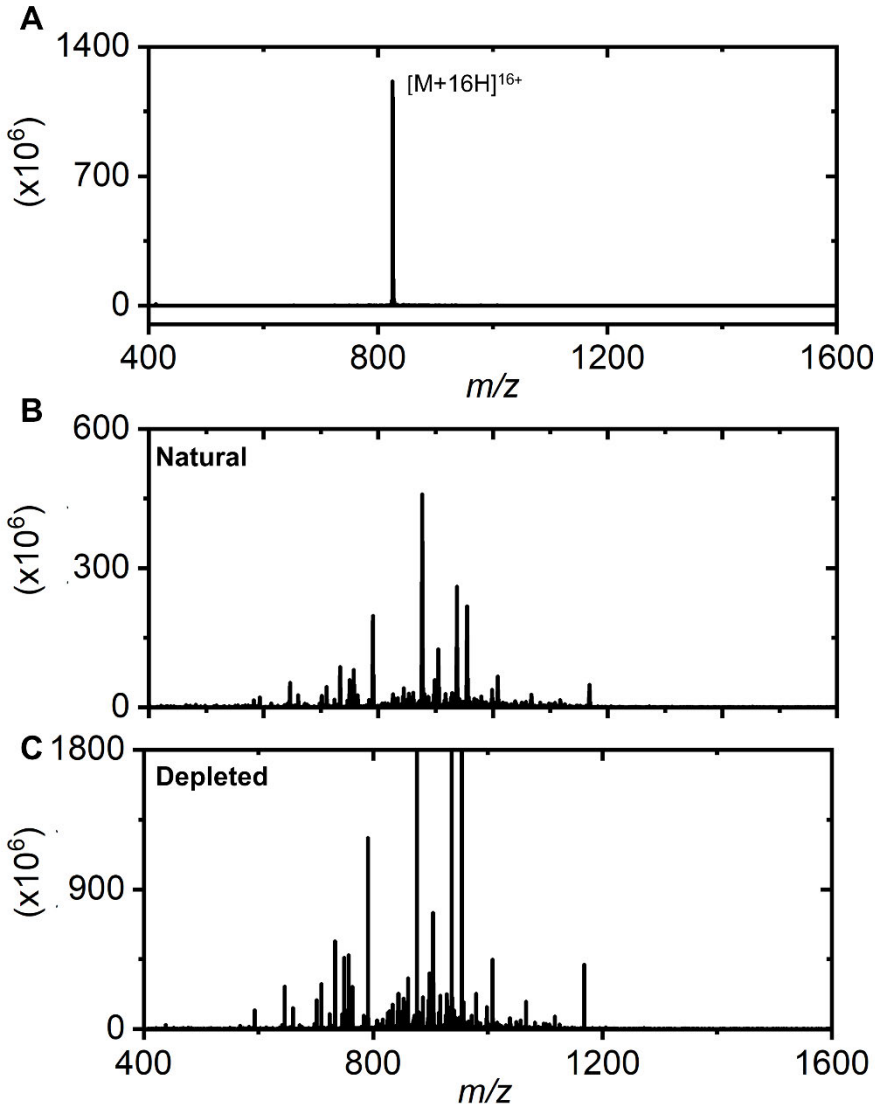


Figure 4.1: CID fragmentation spectra of the [M+16H]$^{16+}$ charge state of EncFtn

CID fragmentation spectra of the 16+ charge state of isotopically natural (A) and isotopically depleted (B) EncFtn. These spectra were the result of a single acquisition with 200 averaged transients.

In both spectra the major ions are clearly distinguishable, but many of the produced fragment ions are acquired with low S/N. Often in densely packed regions of fragment ions making them difficult to identify and assign. This is demonstrated in Figure 4.2,

displaying a closer view of the region 868-873 *m/z* from the full CID fragmentation spectra of Figure 4.1. Isotopically natural EncFtn fragments (Figure 4.2a) display a complex arrangement of overlapping fragment ion distributions expected in a fragmentation spectrum. This region of the spectra contains six assigned fragment ions, each overlaid with the theoretical model of the assigned ion.

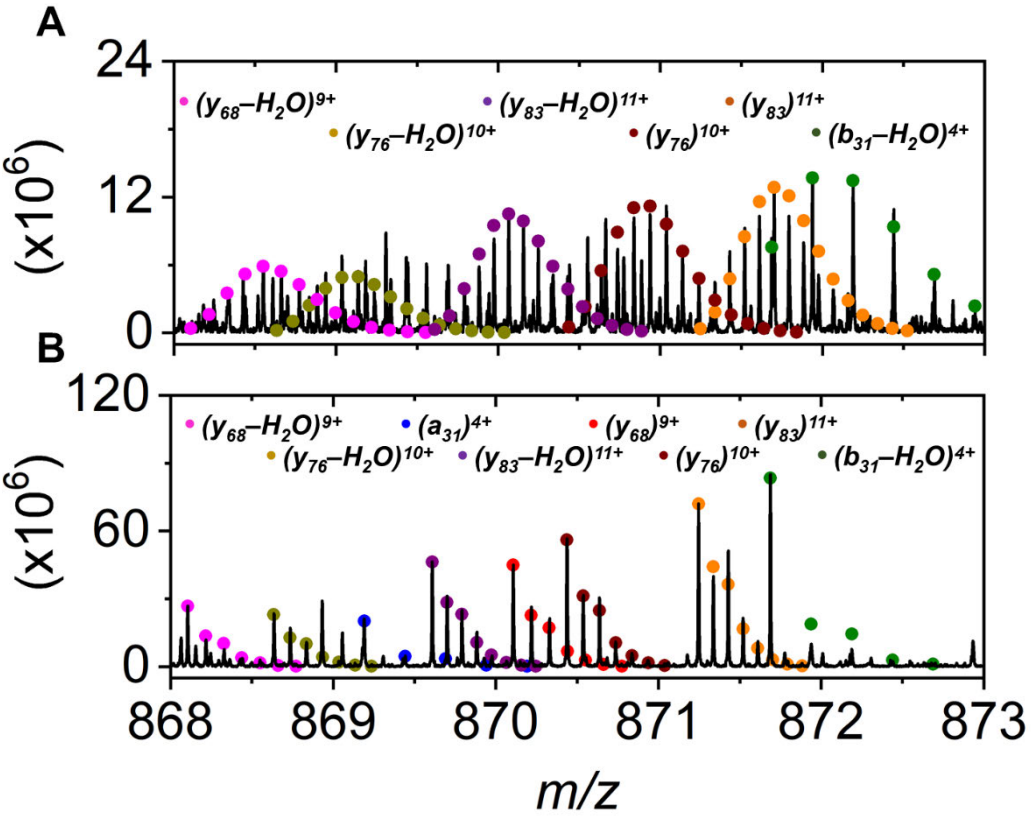


Figure 4.2: Region of 868-873 m/z of EncFtn CID fragmentation spectra

Region of assigned ions within the CID fragmentation spectra of isotopically natural (A) and isotopically depleted (B) EncFtn. The theoretical models of assigned ions are overlaid as dots on the spectra, y_{68}-H_2O^{9+} (●), y_{76}-H_2O^{10+} (●), a_{31}^{4+} (●), y_{83}-H_2O^{11+} (●), y_{68}^{9+} (●), y_{76}^{10+} (●), y_{83}^{11+} (●), b_{31}-H_2O^{4+} (●).

The signal in the isotopically depleted fragments (Figure 4.2b), like with the major ions in Figure 4.1 are consistently higher than the same fragment ions in the corresponding isotopically natural spectra (Figure 4.2a). Crucially however, the isotopically depleted

fragment ions display spectral simplification in the number of isotopologue peaks present for each separate ion. Allowing the assignment of two extra fragment ions.

This region of 868-873 *m/z* is interesting as we can observe the spectral complexity which arises in the isotopically natural EncFtn (Figure 4.2a) causing many of the ion distributions to overlap. For many of the ions there is also a loss of water which further complicates the spectra as the same fragment ion (minus water) will produce another isotope distribution close to the original fragment ion. Even within this small section of spectra it is possible to see that there is isotopologue peaks which are not part of the isotope distributions of the assigned ions. Which means that there is a wealth of information hidden underneath the assigned ions which cannot be resolved. Thereby limiting the full analysis isotopically natural EncFtn. Deconvolution of these regions can be particularly challenging for the averagine based detection algorithms. [20]

Using the y_{83}^{11+} (●) fragment ion as an example, in Figure 4.2a contains 12 isotopologue peaks. The most abundant isotopologue peak of the ion contributes ~16% to the total signal. Conversely the depleted ion (Figure 4.2b) contains eight isotopologue peaks, the most abundant of which is the monoisotopic peak contributing ~40%, and is ~5.6 times more abundant than that in Figure 4.2a. Similarly, for the complementary fragment ion, b_{32}^{5+}(which contains the rest of the fragmented protein ion), the isotopically standard ion has 9 isotopologue peaks, the most abundant displaying ~25% of the total signal; while the isotopically depleted ion contains only 5 isotopologue peaks, with the monoisotopic peak contributing ~75% of the total signal. As before, the isotopologue with the greatest abundance in b_{32}^{5+} (Figure 4.2b) is ~5.7 times greater than the isotopologue with the greatest abundance in the isotopically natural spectra.

The simplification of the isotopically depleted CID fragmentation spectrum is consistent for all the produced fragment ions. In Figure 4.2b, it facilitates the assignment of two extra ions. As this is repeated throughout the fragmentation

spectra, it ultimately increases the total number of assigned fragment ions, and the achievable sequence coverage (Figure 4.3).

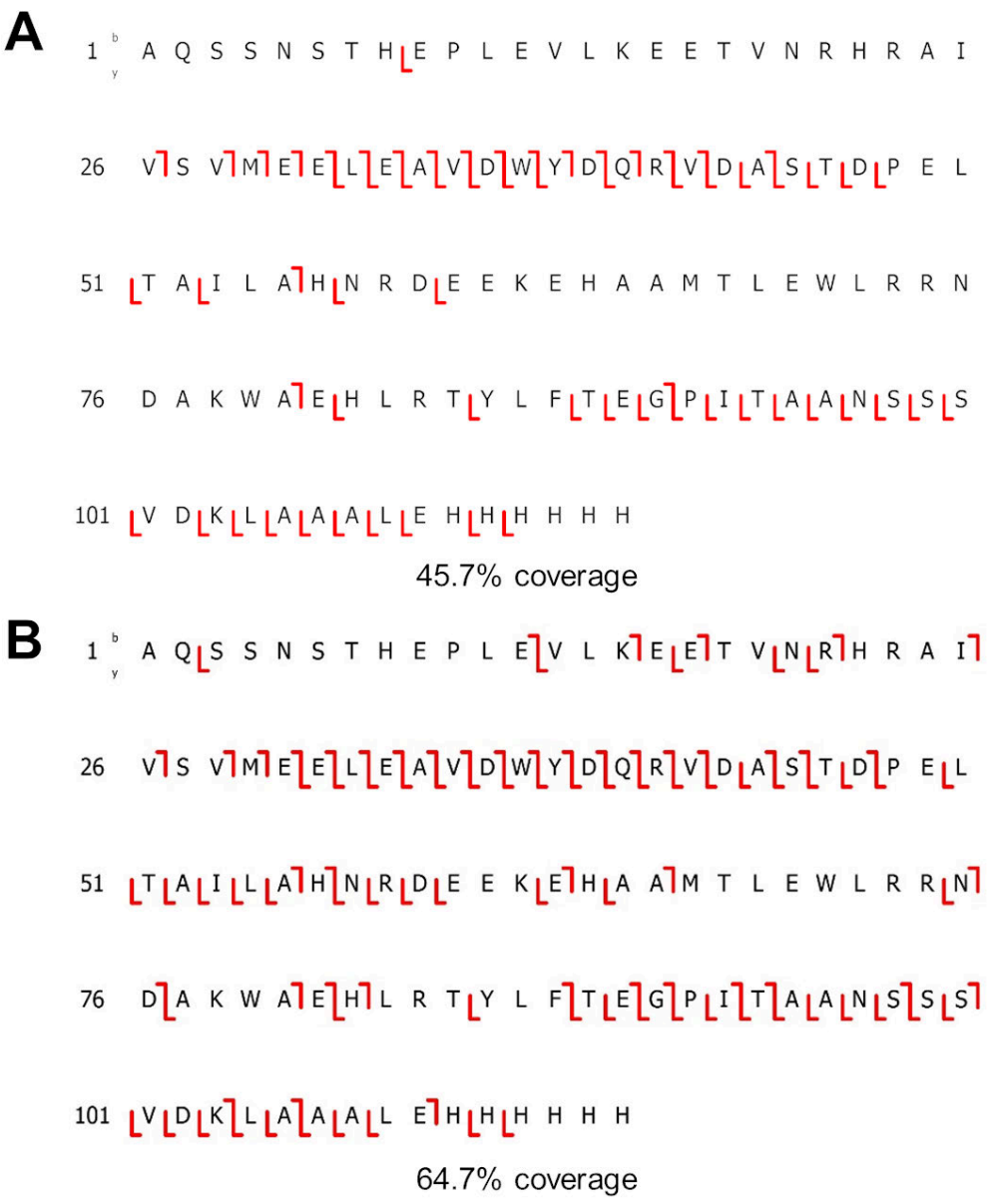


Figure 4.3: Sequence coverage of EncFtn after CID fragmentation

CID fragmentation of isotopically natural (A) and isotopically depleted (B) EncFtn results in production of fragments ions. Indicated by the red lines in between the amino acids. The direction of the line indicates if the fragment is a N-terminal fragment ion (b-type ion, ┐) or a C-terminal ion (y-type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

Analysis and assignment of the fragmentation spectra was carried out using Autovectis (section 1.3, page 34) focusing on assignment of N-terminal *b*- type and C-terminal *y*- type ions. Figure 4.3a shows the amino acid coverage after CID of the $[M+16H]^{16+}$ of EncFtn, 110 *b* and *y* fragment ions were assigned in the natural isotopic abundance spectrum (39 *b*-ions, 71 *y*-ions; 45.7% total sequence coverage); in comparison, 217 *b* and *y* fragment ions (84 *b*-ions, 133 *y*-ions; 64.7% total sequence coverage) were assigned in the depleted isotopic abundance spectrum (Figure 4.3b). It is clear from the spectra (Figure 4.2) that there are a greater variety of ions produced in CID fragmentation than *b*-ions and *y*-ions. Further analysis of the unassigned fragment ions in both CID spectra revealed a substantial number of internal fragments, and widespread neutral loss during fragmentation ($-H_2O$, $-CO$, $-NH_3$). Taking these multiple fragmentation channels into considerations allowed assignment of a total of 448 product ions (*a*, *b*, *x*, *y*, and y_{-H2O} ions; 82% total sequence coverage) in the CID spectrum of isotopically depleted EncFtn. This increase is consistent with previous studies by Akashi, *et al*., [89] who reported a 63% increase in the number of assigned fragment ions when performing CID of an isotopically-depleted version of the 11 kDa protein cystatin.

While isotopically depleted protein does allow a greater number of ions to be assigned, it is clear from the fragmentation maps in Figure 4.3 that there are regions of the protein from which fragments are not observed in the spectra. This ultimately limits CID as a technique for this style of top-down intact protein analysis, as the 'ideal result' would be to achieve as close to 100% sequence coverage (i.e. single amino acid resolution) as possible. This is largely as a result of the mechanism of fragmentation where the vibrational energy of the molecule increases with each collision until the weakest bond is broken. As a result, the utility of CID fragmentation is limited to small proteins, and is well documented that CID is unlikely to cause

fragmentation in the central amino acid region of larger protein (> ca.15 kDa).[114,119] Therefore, even with the increase in sequence coverage seen as a result of isotopic depletion (Figure 4.3), CID would be a limiting technique if applied as a fragmentation technique on the larger proteins in this study.

4.3 ECD fragmentation of EncFtn

ExD fragmentation techniques (section 1.2.2, page 29) require a capture of electrons onto the positive protein ions resulting in fragmentation of the N-Cα bond generating

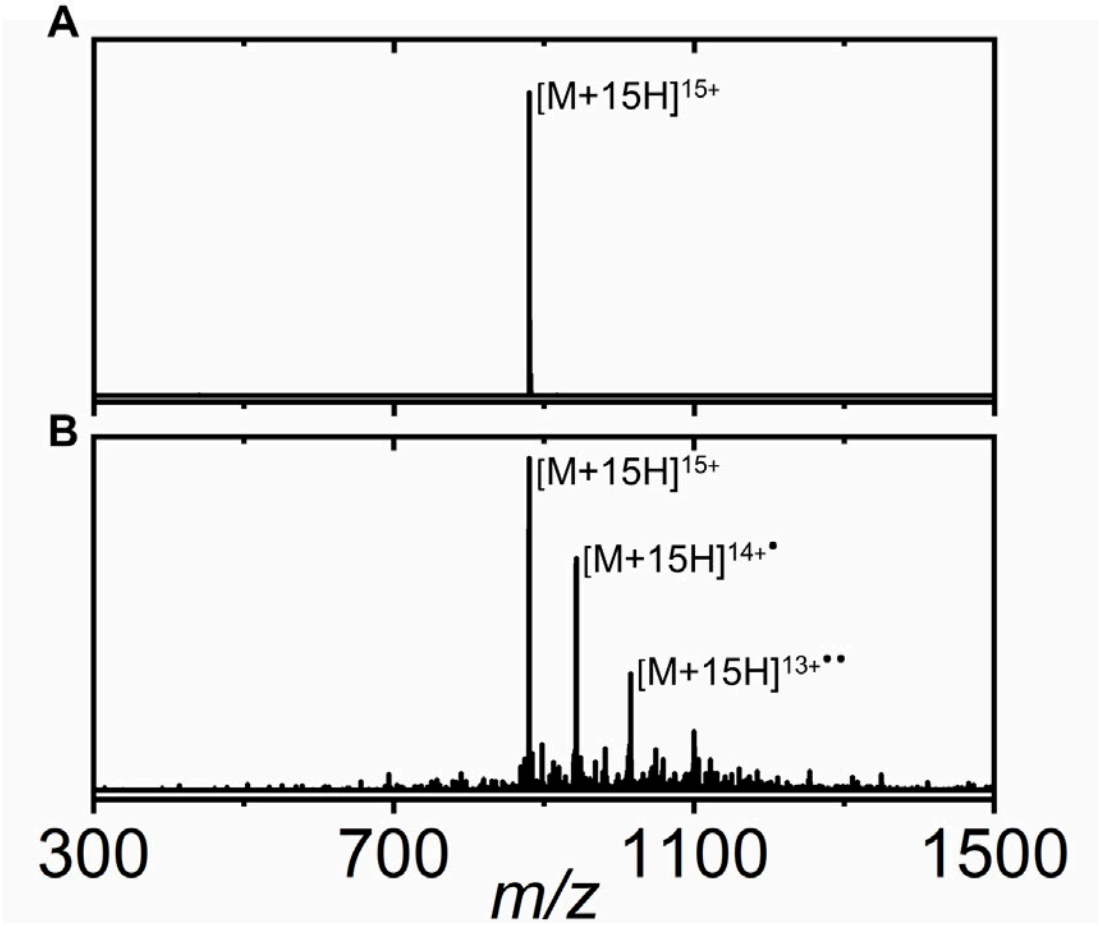


Figure 4.4: ECD fragmentation of the 15+ charge state of isotopically depleted EncFtn

The charge state of interest, the 15+ charge state, was isolated from the ESI charge state distribution (A). With the application of electrons into the FT-ICR detection cell the protein will capture an electron which will result in charge reduced species or the production of c-type or z· -type fragment ions (B). These spectra were the result of a single acquisition with 150 averaged transients.

c-type ions and $z^{\cdot}$ -type ions (Figure 1.12). Importantly it is described as a 'non-ergodic' technique, [61,120,121] which conversely to the 'slow-heating' increase of energy with each subsequent collision in CID, results in the fragmentation of the protein within a relative proximity to where the electron capture event took place.[120] The majority of the analysis in this report was carried out using a 12T FT-ICR SolariX (Bruker), equipped with ECD. The electrons are produced using a heated filament source [58,59] situated behind the ICR cell. Electrons are introduced to the protein ions within the ICR cell, so all resultant fragments ions, theoretically should be detectable. [120] The change to the protein spectra by applying ECD fragmentation to isotopically depleted EncFtn is observed in Figure 4.4. The isolated $[M+15H]^{15+}$ EncFtn precursor (Figure 4.4a) was subjected to electron bombardment to cause fragmentation (Figure 4.4b). The highest abundance peaks within Figure 4.4b are precursor $[M+15H]^{15+}$ ions, which have not undergone capture/fragmentation events. The next tallest peaks $[M+15H]^{14+\bullet}$ and $[M+15H]^{13+\bullet\bullet}$ are precursor ions which have captured an electron ($^{\bullet}$) but have not undergone fragmentation (ECnoD) or undergone some fragmentation but remain bound due to non-covalent interactions. These species are known as charge reduced species, where they are observed at a charge state lower than the precursor ion. The charge is neutralised by the capture of the negative electron, so gain the mass of an electron onto the ion. These peaks are present at a far greater abundance than the fragment ions which populate the bottom of the spectra. As with the CID fragments (Figure 4.2) we can compare the produced fragments in Figure 4.4, observed on a closer axis in Figure 4.5.

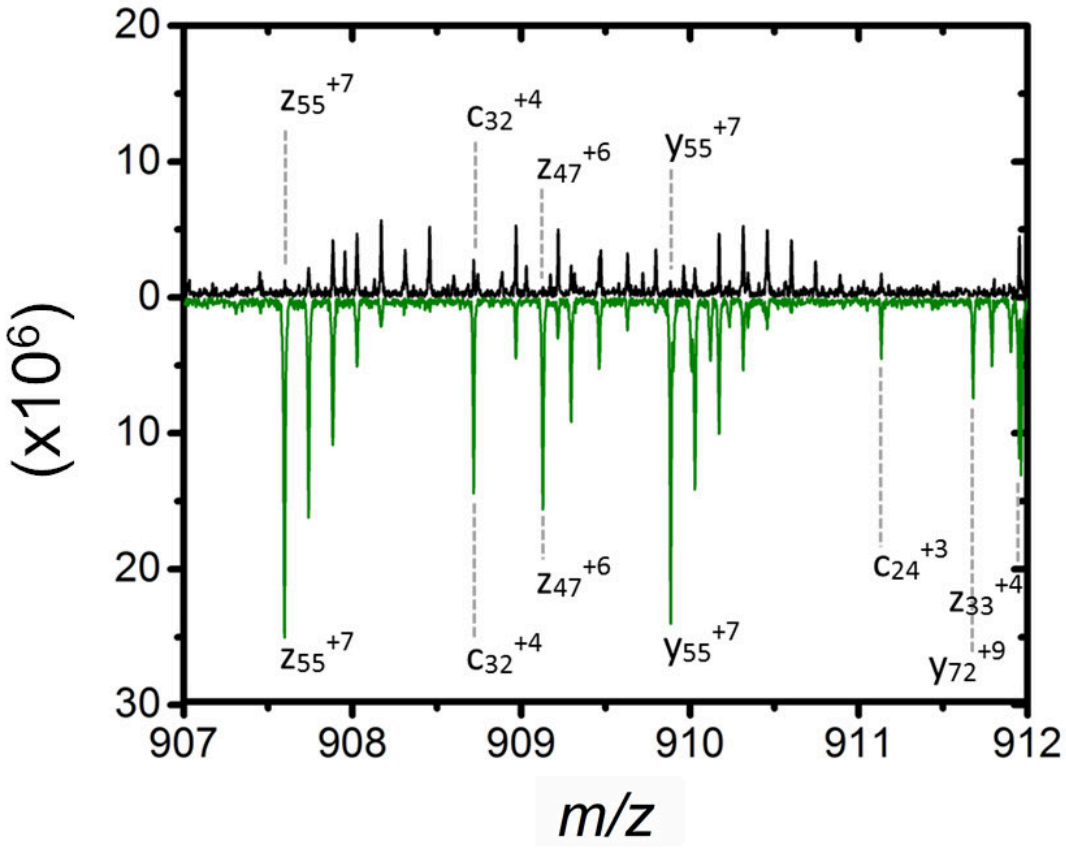


Figure 4.5: Mirror view of 907-912 m/z of the EncFtn ECD fragmentation spectra

Comparison of isotopically natural (black line) and isotopically depleted (green line). The isotopically depleted spectrum is displayed on the negative axis using the same signal intensity scale. The assigned ions are labelled at the monoisotopic peak.

The reflective axis in Figure 4.5 again highlights the advantage of isotopic depletion on the isotopologue distribution of the protein fragments. The large increase in the abundance of the monoisotopic peak allows for rapid visual identification of ions, the reduced number of isotopologue peaks results in an increase in the S/N of each ion. Therefore, ultimately results in almost double the number of assignable ions within this small section of the entire fragmentation spectra. Using the z_{55}^{7+} ion (6346.13 Da) as an example for comparison, within the isotopically natural spectra is composed of approximately 8 isotopologue peaks. While the inverse spectra for the isotopically

depleted protein displays approximately 5 isotopologue peaks. As is also observable in Figure 4.5 the most abundant isotopologue peak (monoisotopic) in the isotopically depleted spectra is 4X greater in signal than the most abundant isotopologue in the isotopically natural sample (M+3). Interestingly, fragments ions smaller than ~2000 Da within the isotopically depleted spectra exist almost exclusively as a monoisotopic peak, which must be assigned solely on the accurate mass.

This is observable in Figure 4.6, which is another small section of the same ECD fragmentation spectra of EncFtn, highlighting the different isotopologue peaks of each fragment ion. The monoisotopic peak of each ion highlighted by the asterisk (*). The c_{15}^{2+} fragment ion (*, 1637.84 Da) in Figure 4.6b exists as a large monoisotopic peak, S/N 106.5. The second isotopologue peak for the ion exists close to the noise baseline, contributing little to the overall abundance of the ion signal. The isotopically natural sample shows the monoisotopic c_{15}^{2+} ion, S/N 42.4, the second isotopologue peak almost achieves a similar level of signal, so the population is split between the fragment ions which are fully monoisotopic and those which contain a single ^{13}C, and also a smaller peak which contain two ^{13}C out with the range of Figure 4.6. So further reducing the signal achievable by the monoisotopic peak of the isotopically natural c_{15}^{2+} ion.

This region of the ECD spectra is a further excellent example of the impact for isotope depletion, as Figure 4.6a, is a complex overlap of ions. In particular, the *m/z* of the monoisotopic peak for z_{14}^{2+} and z_{43}^{6+} ions are very close and could result in misassignment of the ion or monoisotopic peak which would result in a greater associated error. As discussed before, the isotopically depleted fragmentation spectra contain fewer isotope peaks, so there is far less overlap, this and the increased abundance of the monoisotopic peak contributes to the identification of two extra fragment ions, labelled in green.

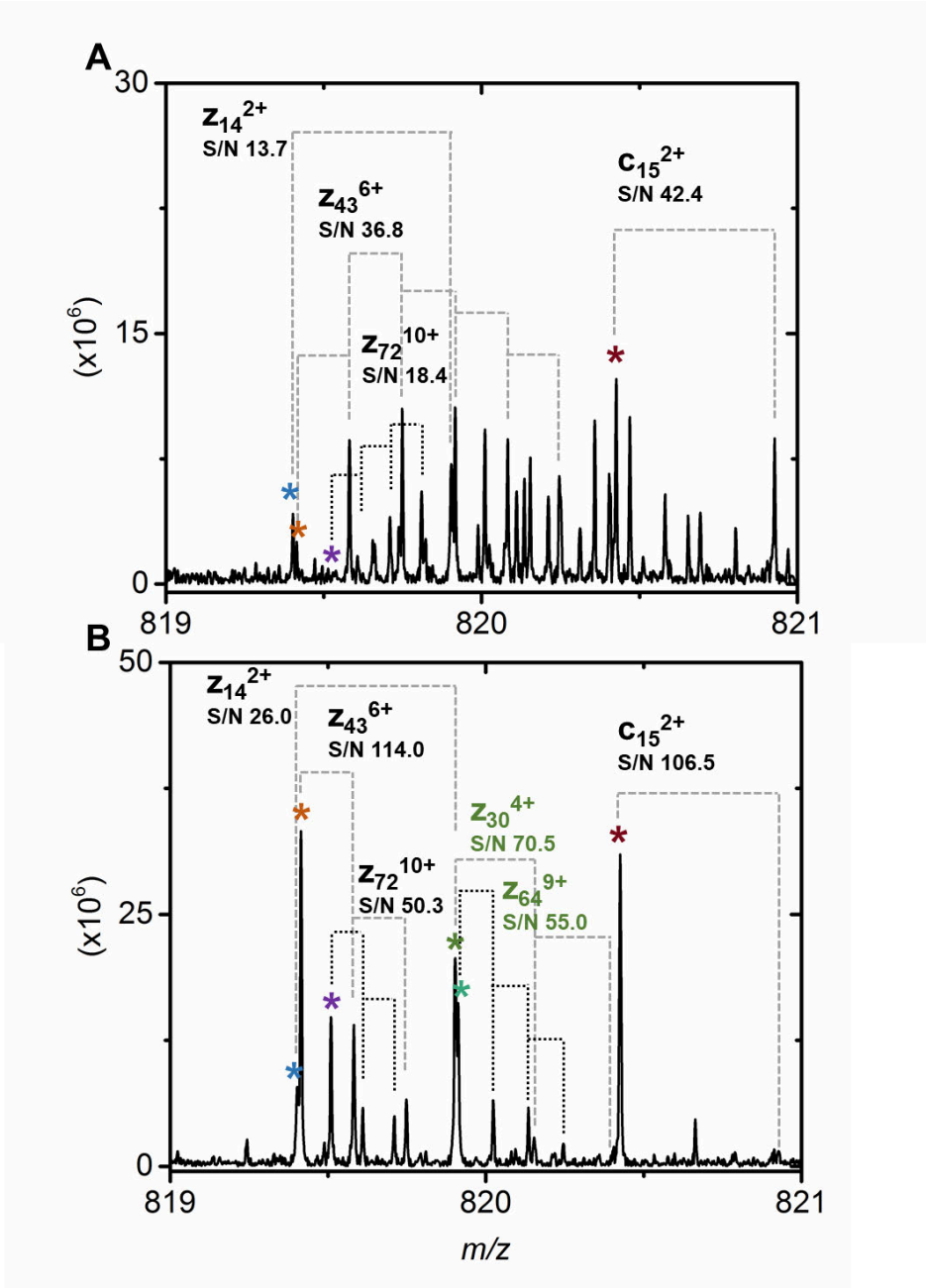


Figure 4.6: Isotopologue view of ECD fragmentation of the $[M+15H]^{15+}$ charge state of EncFtn.

In complex sections of the fragmentation spectra it can be difficult to identify the isotopologue peaks contributing to a single fragment ion. The dashed lines originate at a labelled monoisotopic peak (*) and direct the reader to each isotopologue peak. Fragments found in the isotopically depleted spectrum (B) and not in the natural equivalent (A) are assigned in green.

The S/N value of each assigned peak is displayed underneath the ion assignment.

As with the CID fragmentation acquisition, the ECD fragments can be compiled into fragmentations maps. Which in this case the N-terminus ions are *c*-type ions (┐) and the C-terminus ions (└) are *z*·-type ions.

Figure 4.7: Sequence coverage of EncFtn after ECD fragmentation

ECD fragmentation of isotopically natural (A, 74.1% coverage) and isotopically depleted (B, 96.6% coverage) EncFtn results in production of fragments ions. Indicated by the red lines between the amino acids. The direction of the line indicates if the fragment is a N-terminal ion (c-type ion, ┐) or a C-terminal ion (z· -type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

Overall ECD produces a more comprehensive fragmentation pattern than that achieved during CID fragmentation (Figure 4.3). But similarly, the isotopically depleted EncFtn (Figure 4.7b), due to the change in the isotopologue distributions allow for a greater number of fragment ions to be assigned. As EncFtn as a monomer is relatively small protein (~13 kDa), it is readily fragmented throughout the polypeptide chain. The isotopically natural EncFtn (Figure 4.7a) achieved a high sequence coverage of 74.1%, with approximately 31% displaying identification of complementary ions. Which is the assignment of both the C-terminal and N-terminal ions from a single bond fragmentation. In the isotopically depleted EncFtn fragmentation map, as we might predict due to the extra ion assignments seen in both Figure 4.5 and Figure 4.6b, there is a greater sequence coverage observed. Out of a total of 114 possible bonds, only 6 bonds were not assigned within this fragmentation spectra. But if we consider that ECD fragmentation N-terminal to proline residues is rarely observed, [42,57] this effectively means that only 3 fragmentation events were not observed. It is impressive that more than 77% of the amino acid sequence is covered by complementary ion pairs, which is more than the total sequence coverage observed for the isotopically natural EncFtn protein (Figure 4.7a).

Effective fragmentation of smaller protein has been well documented using many methods of fragmentation, [122] our ability to use these same approaches to analyse quickly decreases with the increasing MW of the protein of interest.

4.4 ECD fragmentation of BCA

BCA is a commonly used MS standard at a molecular weight ~30 kDa, has frequently been used to report the progression of our current capabilities for top-down fragmentation.[38,120,123–125] For example, using UVPD, Shaw *et al*. 2013 managed to achieve 87% coverage with the accumulation of the spectra over 500 data scans on a modified ThermoScientific Orbitrap Elite mass spectrometer.[126] A similar sequence

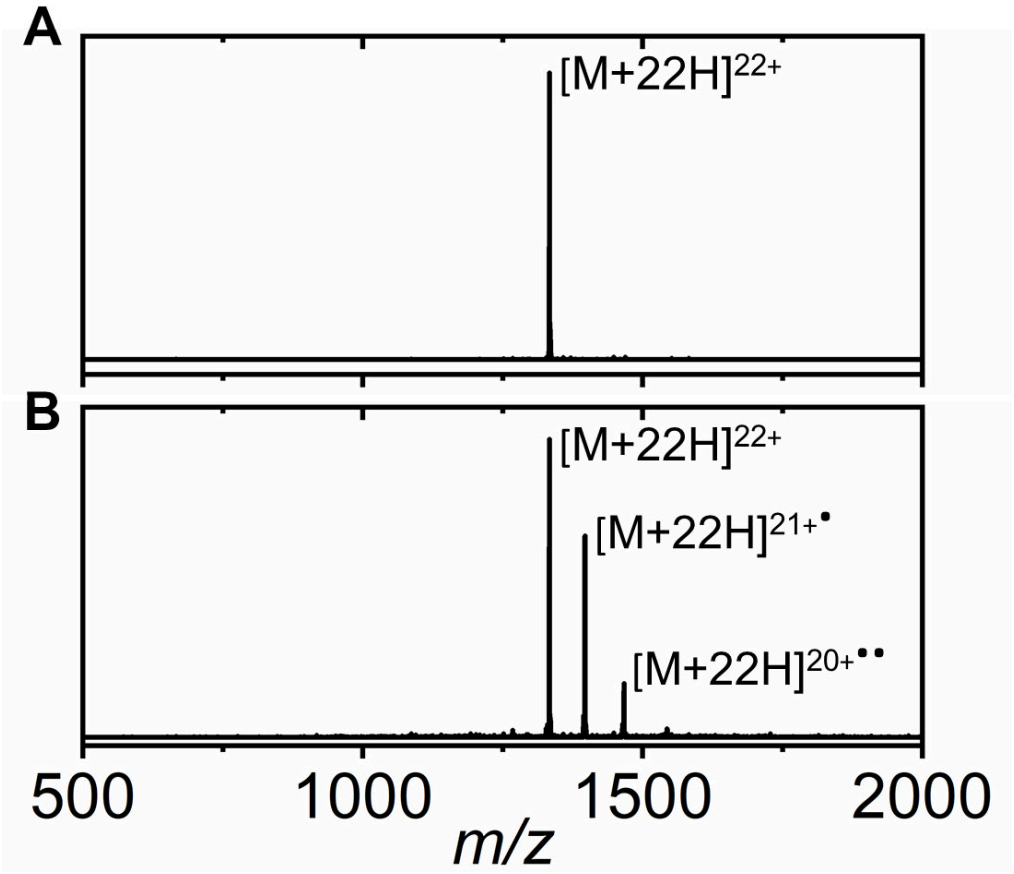


Figure 4.8: ECD fragmentation of the $[M+22H]^{22+}$ charge state of isotopically natural BCA

ECD fragmentation of the isolated $[M+22H]^{22+}$ charge state from the ESI charge state distribution (A). With the application of electrons into the FT-ICR detection cell the protein will capture an electron which will result in charge reduced species or the production of c-type or z·.-type fragment ions (B). These spectra were the result of a single acquisition with 300 averaged transients.

coverage (87%) was obtained by Weisbrod, *et al.* 2017 on a 21T FT-ICR over 1500 averaged transients. [38]

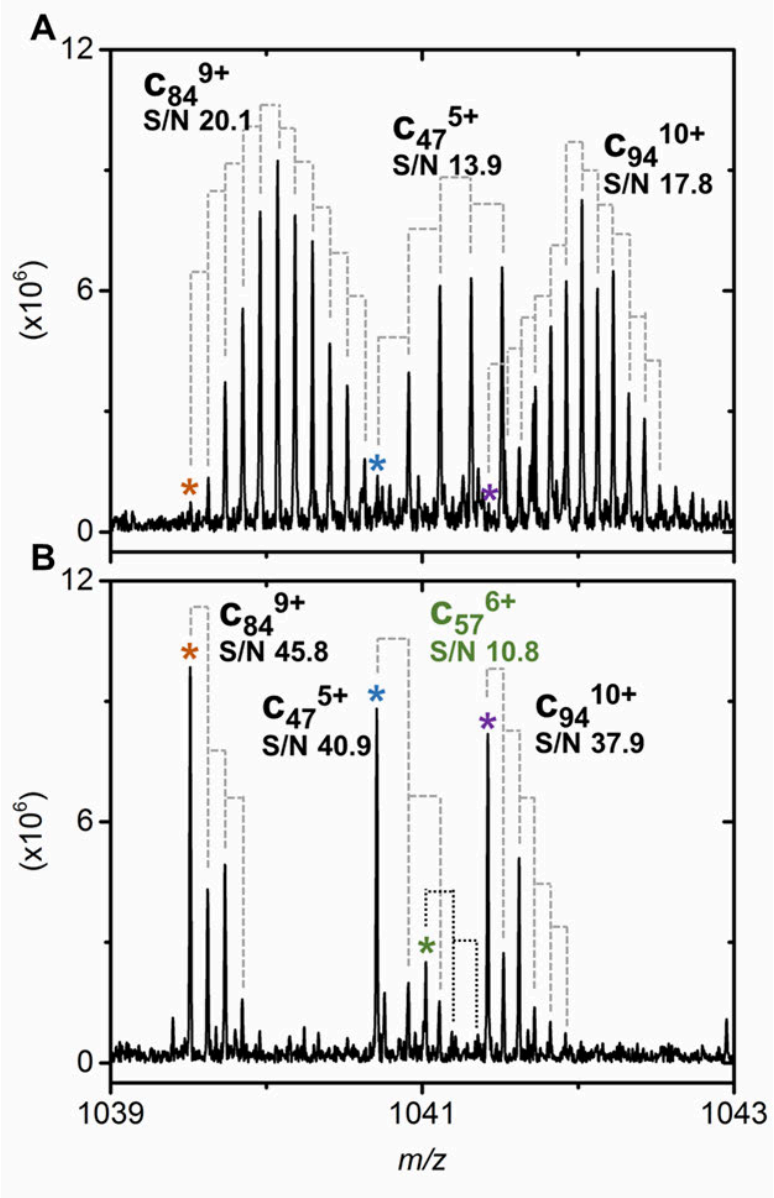


Figure 4.9: Isotopologue view of ECD fragmentation of the $[M+22H]^{22+}$ charge state of BCA

In complex sections of the fragmentation spectra it can be difficult to identify the isotopologue peaks contributing to a single fragment ion. The dashed lines originate at a labelled monoisotopic peak (*) and direct the reader to each isotopologue peak. Fragments found in the isotopically depleted spectrum (B) and not in the natural equivalent (A) are assigned in green.

We applied ECD fragmentation to our recombinantly expressed BCA to test if the increase in sequence coverage observed in EncFtn (Figure 4.7) as a result of isotopic depletion would remain as the protein MW increased.

As observed during ECD of EncFtn (Figure 4.4b), bombardment of electrons during ECD fragmentation of BCA protein (Figure 4.8b) displays the isolated precursor and charge reduced species as the most abundant ion species in the fragmentation spectra.

Therefore, we can again compare the isotopically natural ECD spectra to that of the isotopically depleted protein sample. Figure 4.9, much like Figure 4.6 displays a small section of the top-down fragmentation spectra of BCA. Again, it is apparent that the number of isotopologue peaks for each isotopically depleted ion have reduced, the monoisotopic peak has become the most abundant isotopologue in the distribution and there is at least a two-fold increase in the signal of each ion. As the molecular weight has increased, the range of size and charge in the resultant fragments is larger than that observed with EncFtn. The fragments c_{84}^{9+} (9346.55 Da) and c_{94}^{10+} (10404.14 Da) both have similar masses to a small protein and the resultant change in their isotopologue distribution is not dissimilar to the isotopologue distribution seen for intact EncFtn (Figure 3.15). This ultimately reduces the overall complexity observed within this small region, allowing assignment of the c_{57}^{6+} ion in Figure 4.9b. This trend, as with EncFtn, is observed throughout the entire spectra.

An example of this is apparent in Figure 4.10, showing a 50 *m/z* window of the ECD fragmentation spectra of BCA. While there is less observed detail of the separate isotopologue distributions it can highlight the increase in ions assignment observed within the isotopically depleted BCA fragmentation spectra. The fragment ions labelled in green, are ions assigned only in the isotopically depleted protein sample (Figure 4.10b), while also continuing to have all the assigned fragments from the

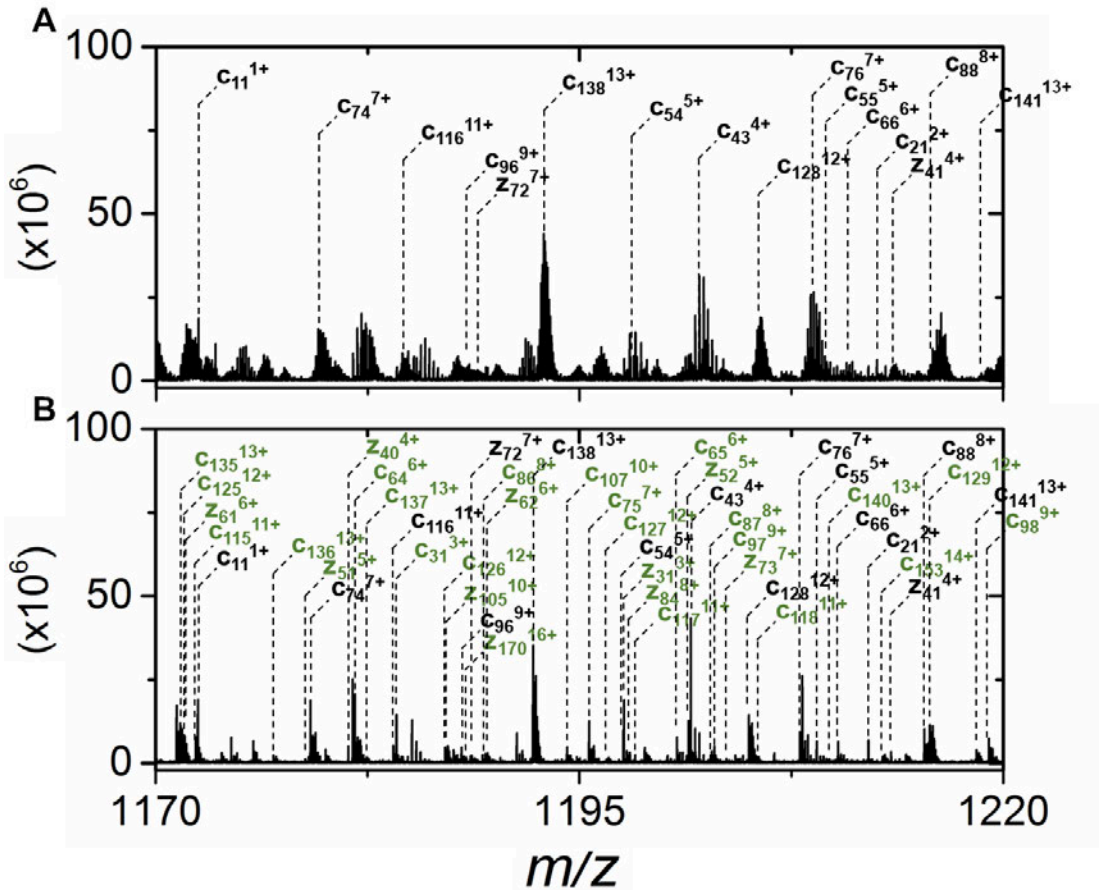


Figure 4.10: A 50 m/z window of ECD fragmentation with ion assignments

Comparison of the number of assignable ions between isotopically natural (A) and isotopically depleted (B) BCA. Each ion assignment is labelled, with ions found only in the isotopically depleted spectrum (B) highlighted in green and connected to the relative m/z of the distribution with a dashed line (---).

isotopically natural sample (Figure 4.10a). This results in an approximately 3 times a greater number of total ions being assigned. This is particularly useful as many of the assigned ions in Figure 4.10b are higher mass and charge state, which would normally be a complicated overlap of isotopologue peaks within in the background noise. Fragment ions like c_{137}^{13+} (15358.44 Da), z_{170}^{16+} (18990.75 Da), c_{129}^{12+} (14576.12 Da) and c_{235}^{17+} (26038.08 Da; not shown) are all large fragments ions mostly generated from fragmentation events within the middle of the polypeptide sequence. Each fragment is larger than the intact EncFtn (13 kDa) and are all low

abundance ions. Therefore without the benefits conferred from isotopic depletion, i.e., the isotopically natural spectrum (Figure 4.10a), it fundamentally very difficult to achieve enough signal to allow for confident assignment of these ions. Compounding this problem, the monoisotopic peak of these large fragments is not observed above the baseline noise, this contributes to an increased mass error when assigning fragment identities, as monoisotopic mass must be inferred computationally. Comparison of the number of assigned ions is summarised in Figure 4.11.

Figure 4.11: Sequence coverage of the $[M+22H]^{22+}$ charge state of BCA after ECD fragmentation

ECD fragmentation of isotopically natural (A, 50% coverage) and isotopically depleted (B, 82.6% coverage) BCA results in production of fragments ions. Indicated by the red lines between the amino acids. The direction of the line indicates if the fragment is a N-terminal ion (c-type ion, ┐) or a C-terminal ion (z· -type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

The obtained sequence coverage of isotopically natural (Figure 4.11a) and isotopically depleted (Figure 4.11b) BCA further highlights the impressive increase in sequence coverage possible with isotope depletion. There is a substantial increase from 50% to 82.6% total coverage. Nearly equivalent to the 87% achieved by Shaw and Weisbrod. [38,126] This increase is particularly evident in the region of 151-200 amino acids of the polypeptide chain, which in Figure 4.11a, has very small number of ion assignments (i.e. higher mass, 12-20 kDa, fragment ions). The same area in Figure 4.11b has far greater number of fragment ions, and a greater number of complementary ions at the beginning of the polypeptide chain (1-151 amino acids).

Assigned fragments in Figure 4.11a (172 *c*- ions, 57 *z·* - ions) and Figure 4.11b (377 *c*- ions, 216 *z·* - ions) both show preferential fragmentation at the N-terminus of the protein. With both spectra displaying a greater number of *c*-type ions overall. Many of the larger *z·* -type ions observed, particularly in Figure 4.11b, are also focused at the N-terminus of the protein. This produces a large number of complementary ions in this area, which thanks to the increased S/N in the depleted spectra can be identified and assigned. The larger number of fragments around the N-terminus is likely as a result of the primary sequence of BCA, [127] containing a greater number of positively charged residues which promotes electron capture.

The typical process for ion assignment in isotopically natural protein mass spectrum involves the determination of charge on the ion and deconvolution of mass [83]. As previously described (Figure 3.13) this process can lead to the introduction of error. The process requires the calculation of the monoisotopic peak, which on the whole is relatively accurate and deconvolution software is used in many commercial software packages. An example of this can be seen in Figure 4.12, in which the z_{40}^{5+} fragment was correctly assigned in the isotopically depleted fragmentation spectra (Figure 4.12b), but in the isotopically natural spectra (Figure 4.12a) the deconvolution software incorrectly calculated the monoisotopic peak, assigning it to the M+1 peak.

This results in the z_{40}^{5+} fragment not being included in the fragmentation list. The z_{40}^{5+} fragment is an extreme example of the error which can be introduced via sophisticated numerical annotation procedure (SNAP) analysis. [82,83]

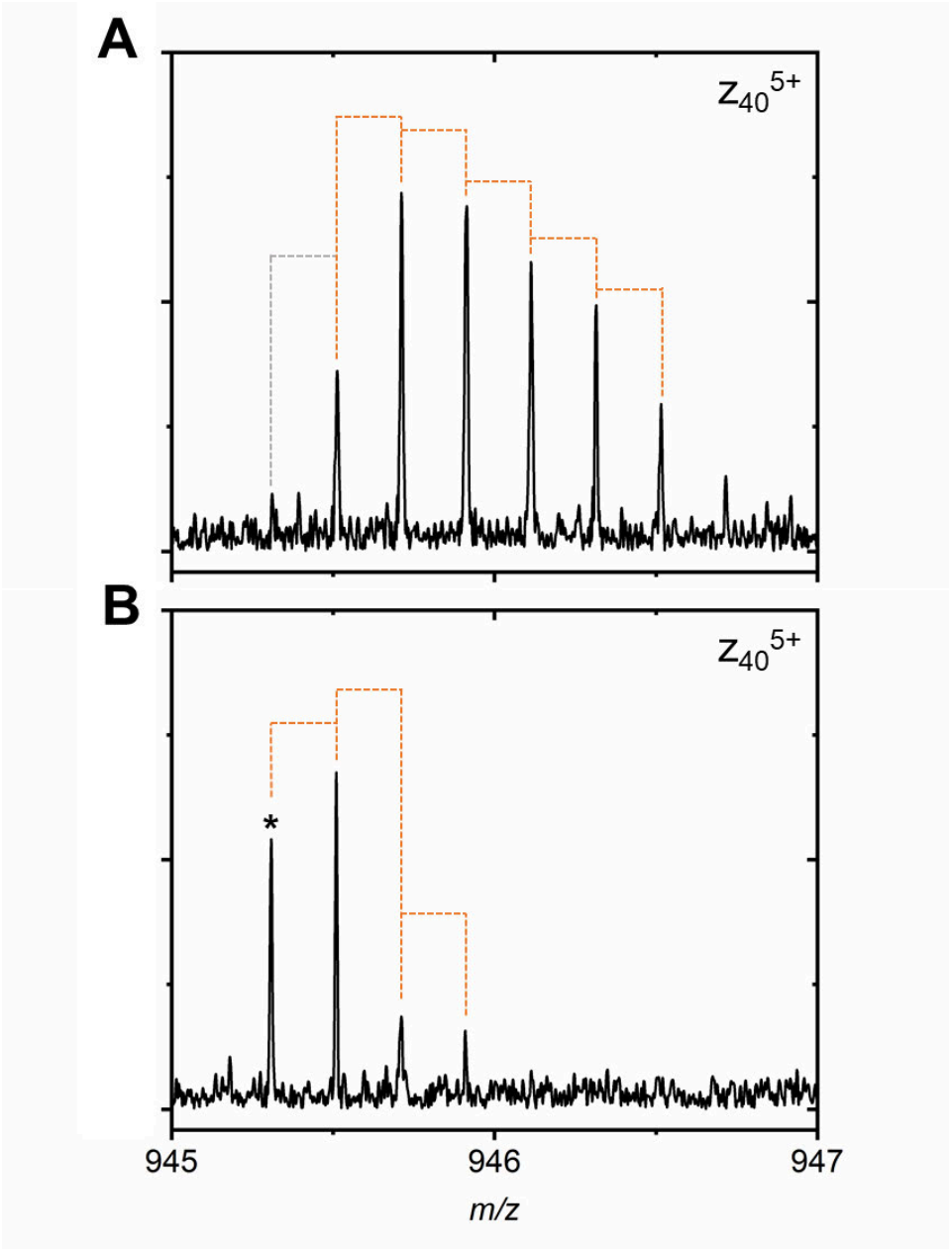


Figure 4.12: SNAP deconvolution of the monoisotopic peak

Isotopologue distribution of the z_{40}^{5+} for isotopically natural (A) and isotopically depleted (B) BCA. The uniform m/z difference between identified isotopologue peaks is highlighted with the orange dash line (---) and the monoisotopic peak is indicated by an asterisk (*). The isotopologue which were not identified are connected to the next isotopologue with a grey dash line (---).

The presence of a prominent monoisotopic peak in the fragment ions produced during fragmentation of isotopically depleted protein greatly simplifies this process. Particularly as the monoisotopic peak often contributes to a significant proportion of the overall abundance of the ion, it can be easily identified and remove the requirement for computational calculation and inference of the monoisotopic *m/z*. Therefore, improving the error for each ion assignment. This has been visualised in Figure 4.13, which is a histogram of the collected errors in fragment assignment observed from both isotopically natural and isotopically depleted BCA. The error of the isotopically natural BCA was calculated using the SNAP deconvolution algorithm within the DataAnalysis software (Bruker). [82,128]

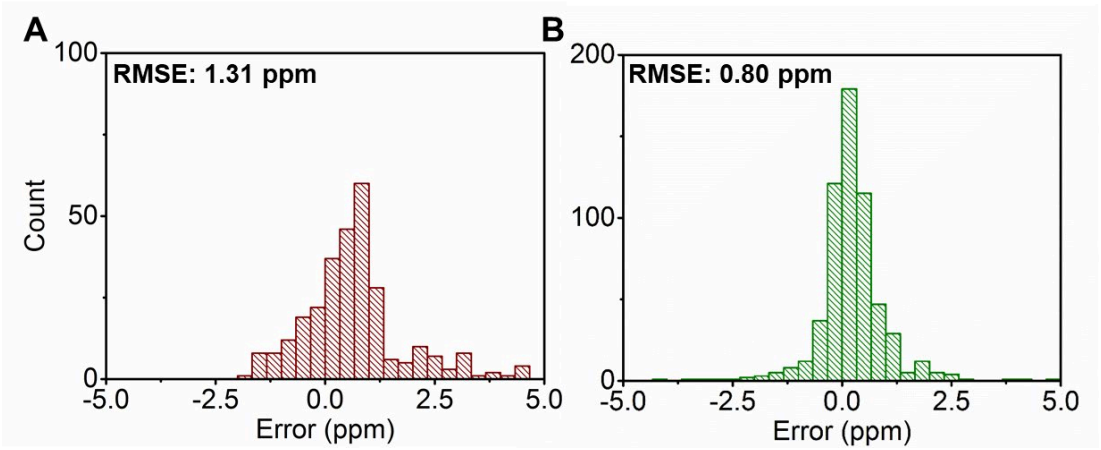


Figure 4.13: Error distributions for the fragment ions assigned after ECD of the $[M+22H]^{22+}$ charge state of BCA.

Histogram displaying the distribution of the mass-error (ppm) for the observed fragment ions after ECD of natural isotopic abundance CA (A, red) and isotopically depleted CA (B, green); bin size = 0.333 ppm. The root-mean-square-error (RMSE) for each distribution is shown.

The mass errors from both spectra are classed in 0.333 ppm bins. Isotopically natural BCA (Figure 4.13a), displays a skewed distribution which is likely as a result of the calibration of the data file within DataAnalysis. However, fewer total assigned ions in the fragmentation spectra of Figure 4.13a will also contribute to this observation. As

result of this, an obvious statistical distribution is less likely to appear, and the root-mean-square-error (RMSE) for the dataset is 1.31 ppm. In the isotopically depleted dataset (Figure 4.13b), as there are a far greater number of ions, there is an obvious distribution centred around 0 ppm. The RMSE at 0.80 ppm, is impressively low for the complex fragmentation spectra. Which is made less complex due to the simplified isotopic profiles caused by isotopic depletion.

It should also be noted that Figure 4.13 is perhaps a biased view of the improvement to error which accompanies isotopic depletion. As it is the combined improvement of both isotopic depletion and removing reliance on deconvolution to calculate the monoisotopic mass. The ions assigned in Figure 4.13a were identified using the

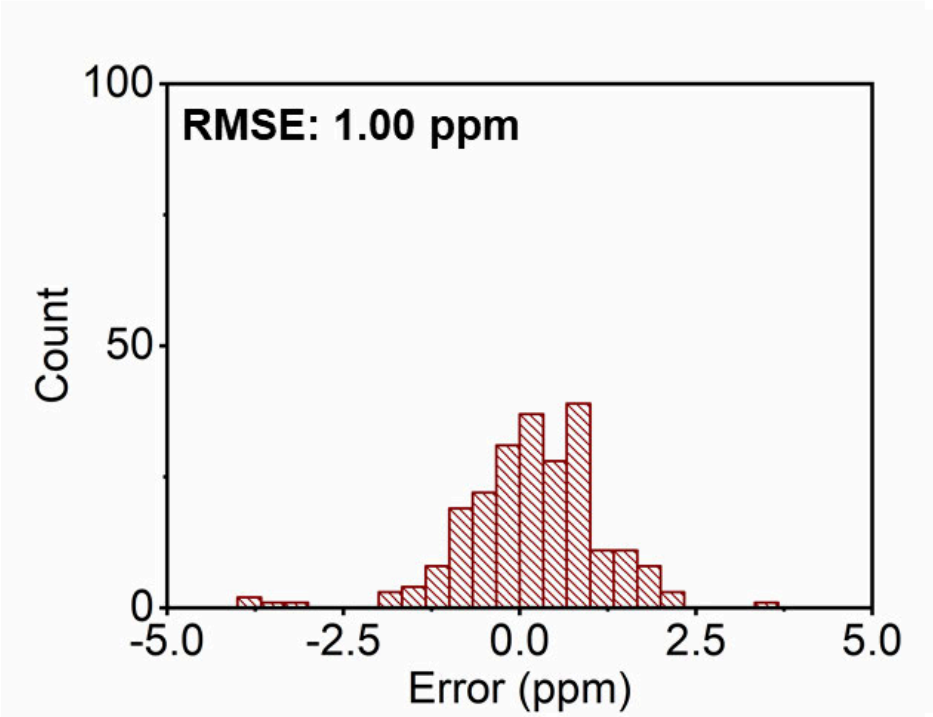


Figure 4.14: Error distribution of assigned fragment ions after ECD of the $[M+22H]^{22+}$ charge state of isotopically natural BCA.

Histogram displaying the distribution of the mass-error (ppm) for the observed fragment ions after ECD of natural isotopic abundance CA analysed using AutoVectis software; bin size = 0.333 ppm. The root-mean-square-error (RMSE) for the distribution is shown.

charge deconvolution SNAP algorithm [82,83] in DataAnalysis and assigned as a particular fragment ion using PrositeLite [81]. Which at the time of writing cannot be applied to isotopically depleted spectra. All spectra, excepting that in Figure 4.13a, were analysed using the peak-picking software AutoVectis. [87] Which when applied to the fragmentation spectra in Figure 4.13a, results in the error histogram shown in Figure 4.14.

Although Figure 4.13a and Figure 4.14 are analysis of the same fragmentation spectra, and either method of analysis have a similar number of assigned ions. But there are clear differences between each histogram. Firstly, the shape in Figure 4.14 better follow a normal distribution. Which is more closely centred around 0 ppm, with an overall RMSE of 1 ppm. This is significantly lower than the 1.31 ppm in Figure 4.13a, which relies on charge deconvolution to calculate the mass of ions. However, is still greater than that of the isotopically depleted histogram in Figure 4.13b further highlighting the improvement to the analysis of top-down fragmentation using isotopically depleted protein can allow.

4.5 Maintaining sequence coverage with reduced spectral averaging

Using FT-ICR for the attainment of relatively high top-down fragmentation sequence coverage of a single isolated protein, (Figure 4.7 and Figure 4.11) requires the accumulation of ions within the ICR cell over time, up to several seconds. The time required for the data acquisition is dependent on multiple settings within the acquisition method.[129] Which ultimately varies the transient length, therefore varies the time required to collect a single data scan. However, for many experimental workflows time is a limiting factor in data acquisition. An obvious example of this is in experiments requiring liquid chromatography. In an ideal situation, the analyte of interest either peptide or protein, elutes in a single sharp peak. This allows the

greatest concentration of the analyte to enter into the MS for efficient detection. But reduces the available time for accumulation and acquisition, particularly if fragmentation of the sample is required. In each case, there is a limit to the number of scans that can be collected. Consequently, in the context of a top-down experiment, the signal to noise ratio of observed fragments is often low and the total sequence coverage obtained is limited.

This effect was demonstrated by reducing the number of accumulated data scans, i.e. changing the number of acquired transients in a top-down ECD experiment. This is observed in Figure 4.15, which displays the decrease in sequence coverage obtained after ECD fragmentation of the 13 kDa EncFtn protein (Figure 4.15a) and the 29 kDa BCA protein (Figure 4.15b) as the number of collected and averaged scans is reduced. For EncFtn (Figure 4.15a), the sequence coverage of the isotopically natural sample (black) reduces from 84.5% to 70.7% as the number of averaged transients

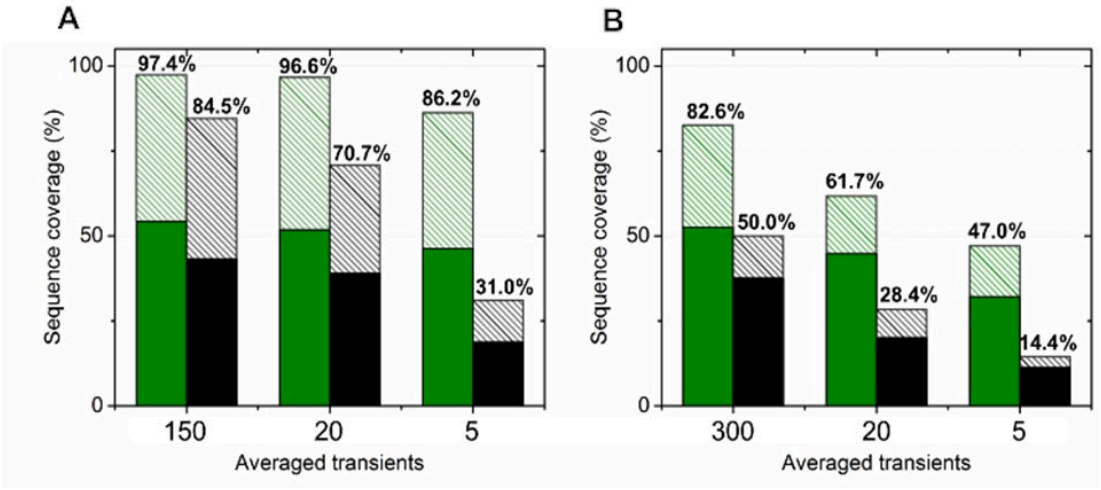


Figure 4.15: Sequence coverage reduction as a result of reduced spectral averaging.

Sequence coverage after ECD fragmentations of the $[M+16H]^{16+}$ charge state of EncFtn (A) and the $[M+22H]^{22+}$ of BCA (B) with reduced spectral averaging. For both proteins sequence coverage for isotopically depleted protein (green) and isotopically normal (black) are divided as a ratio between N-terminal, c-type ions (solid colour) and C-terminal, z· -type ions (hatched bars).

reduced from 150 transients to 20 transients. This is not a substantial decrease, around 14%, considering there was approximately 7X decrease in the averaged accumulated data. When we further decrease the averaged transients from 20 to 5, the ECD sequence coverage of the isotopically natural sample more than halves to 31% coverage. At this level of averaged transient only the most abundant ions are assigned and identified. This is very different to the sequence coverage decrease displayed by isotopically depleted EncFtn (green). Reducing from 150 to 20 averaged transients; the sequence coverage only decreases from 97.4% to 96.6%. Crucially, the sequence coverage as the number of averaged transients' decreases further to 5 still remains high at 86.2% coverage. It is interesting to note that this observed sequence coverage is greater than achieved by the isotopically natural protein acquired with 150 summed transients.

The different rates of sequence coverage decrease are dictated by the total number of assigned ions in each fragmentation spectra. With the acquisition of 150 summed transients for ECD of EncFtn, both isotopically natural and isotopically depleted proteins achieve a high overall sequence coverage (84.5%; 256 ions and 97.4%; 496 ions respectively). Both spectra contain a large number of assigned ions, many of which are redundant assignments, *i.e.* observing the same fragment ion at multiple charge states. As we reduce the available summed transients, we observe the reduction of assigned fragment ions (70.7%; 138 ions and 96.6%; 349 ions). The isotopically depleted protein still generates a large number of fragment ions, maintaining the redundancy of the assignments. While the isotopically natural protein with 138 assignable ions still has a relatively high sequence coverage, but with significantly less redundancy. As a result, further ion assignment loss has a direct impact on the obtained sequence coverage, which is what we observe in the isotopically natural EncFtn with 5 averaged transients (31%; 48 ions). For the isotopically depleted EncFtn, again a large number of ions are still assigned. The

redundancy is maintained even at 5 averaged transients (86.2%, 227 ions) thereby facilitating the large sequence coverage.

A similar situation is observed for top-down fragment assignment of the larger protein, BCA (Figure 4.15b). The decrease from 300 to 20 averaged transients in the isotopically natural BCA (black) the sequence coverage drops from 50% to 28.4%. With a further decrease to 14.4% coverage with 5 averaged transients. The isotopically depleted BCA (green) with the reduction of 300 to 20 averaged transients displays a reduction in sequence coverage from 82.6% to 61.7%. Again a further reduction to 47% is obtained with 5 averaged transients, which is comparable to the isotopically normal BCA at 300 averaged transients. This again is controlled by the number of ions assigned in each spectra, and the redundancy of ion assignments. In isotopically natural BCA (50%, 229 ions; 28.4%, 102 ions and 14.4%, 46 ions) and isotopically depleted BCA (82.6%, 593 ions; 61.7%, 331 ions and 47%, 186 ions) we observe the same trend as before in EncFtn. BCA is more than twice the molecular mass of EncFtn, with ~244 possible cleavages, *i.e.* ~488 potential fragment ions to identify. So understandably displays a greater decrease in sequence coverage than EncFtn.

Analysis of the ECD spectrum of isotopically natural BCA acquired with 300 averaged transients (Figure 4.11) allows 229 fragment ions to be assigned constituting 50% of the primary sequence. As in the case of ECD of EncFtn, many of the fragment ions are assigned, in multiple charge states, adding to the complexity of the fragmentation spectra. This is also the case for ECD of isotopically depleted BCA where the 593 assigned ions encompass 82.6% of the primary sequence. Upon reduction of the combined averaged transients collected to 20, the sequence coverage of both proteins decreases by ~20% of the total protein (*i.e.* cleavage between 50 amino

acids), with the loss of $z^{\cdot}$-ions located in the C-terminus of the protein (hatched bars). However visually, there is an obvious reduction in the quality of the spectra (Figure 4.16), which may ultimately limit the differentiation of low abundance fragments from background noise. Resulting in a small level of sequence coverage reduction.

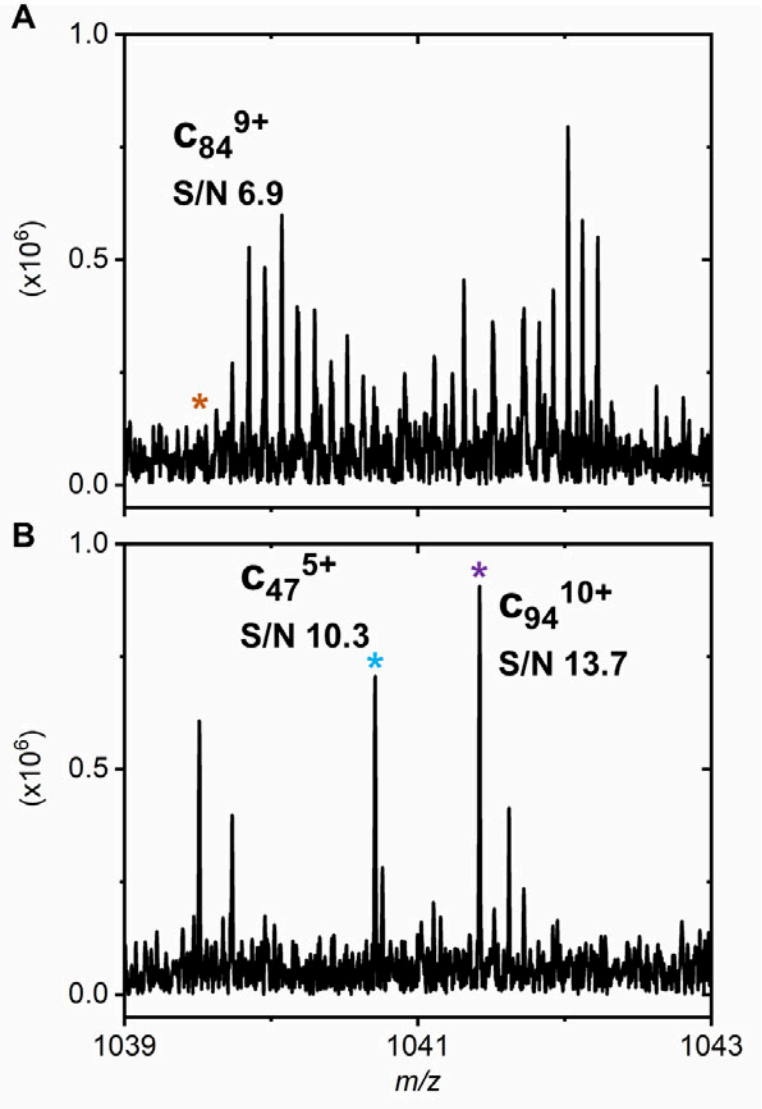


Figure 4.16: Isotopologue view of ECD fragmentation of the $[M+22H]^{22+}$ charge state of BCA with 20 accumulated data scans.

Reduced averaged transients in the data accumulation results in less cohesive isotopologue distributions. The assigned ions for isotopically natural (A) and isotopically depleted BCA (B) are labelled at the monoisotopic peak (*).

The S/N value of each assigned peak is displayed underneath the ion assignment.

This is demonstrated in Figure 4.16, displaying 1039-1043 *m/z* acquired with 20 averaged transients. This spectral region can be compared to that previously displayed in Figure 4.9, which were acquired with the sum of 300 averaged transients. The first thing to note with reduced numbers of averaged transients is the decrease in the total signal, which is approximately 6X lower. This ultimately reduces the S/N of the ions more than 4X to the point where only c_{84}^{9+} can be assigned in the isotopically natural BCA (Figure 4.16a), the same ion cannot be assigned in the isotopically depleted BCA fragmentation spectra (Figure 4.16b). However, the S/N of the monoisotopic peak of the $c_{47}^{5+}$and c_{94}^{10+} in the isotopically depleted BCA spectrum to permit the assignment of the fragment ions. In both spectra of Figure 4.16, it is possible to observe isotope peaks of unassigned ions.

As mentioned above, fragmentation in the C-terminus is low efficiency and the $z^{\cdot}$ - ions produced are characterised by low S/N. Therefore, at the lower number of accumulated scans it is not possible resolve these fragment ions from the background noise. The decrease for both isotopically natural and isotopically depleted protein to 5 averaged transients results in a further decrease in sequence coverage ~15%. However isotopically depleted BCA, while having a similar decrease in overall sequence coverage, at this point has roughly 4x greater number of assigned ions. Allowing for the far greater observed sequence coverage.

Figure 4.15 also highlights the fundamental impact of increasing molecular weight on the accessibility of analysis via top-down MS. The increase from EncFtn (13 kDa) to BCA (30 kDa) increases the number of fragmentation channels that the protein has the potential to undergo. Therefore, in order to obtain comprehensive sequence coverage, an increased number of averaged transients are required to improve the signal of the many different potential ions to a level above the S/N threshold. As a consequence, when analysing proteins >20 kDa, a far greater level of decline in sequence coverage is observed when the number of transients are significantly

reduced. This is precisely the reason that isotope depletion is a powerful tool for protein MS – the technique allows a greater number of fragment ion assignments with fewer averaged transients. Therefore, it is ideally suited to online analysis.

4.6 Online fragmentation of EncFtn

Due to the result observed in Figure 4.15, we hypothesise that our isotope depletion strategy can significantly increase the sequence coverage obtained during online-top-down workflows. The number of transients available to be accumulated during an LC workflow is significantly less than that normally acquired in top-down workflows such as those used above, 150 and 300 averaged transients. This is controlled by the flow rates and column types and chromatography used. [130]

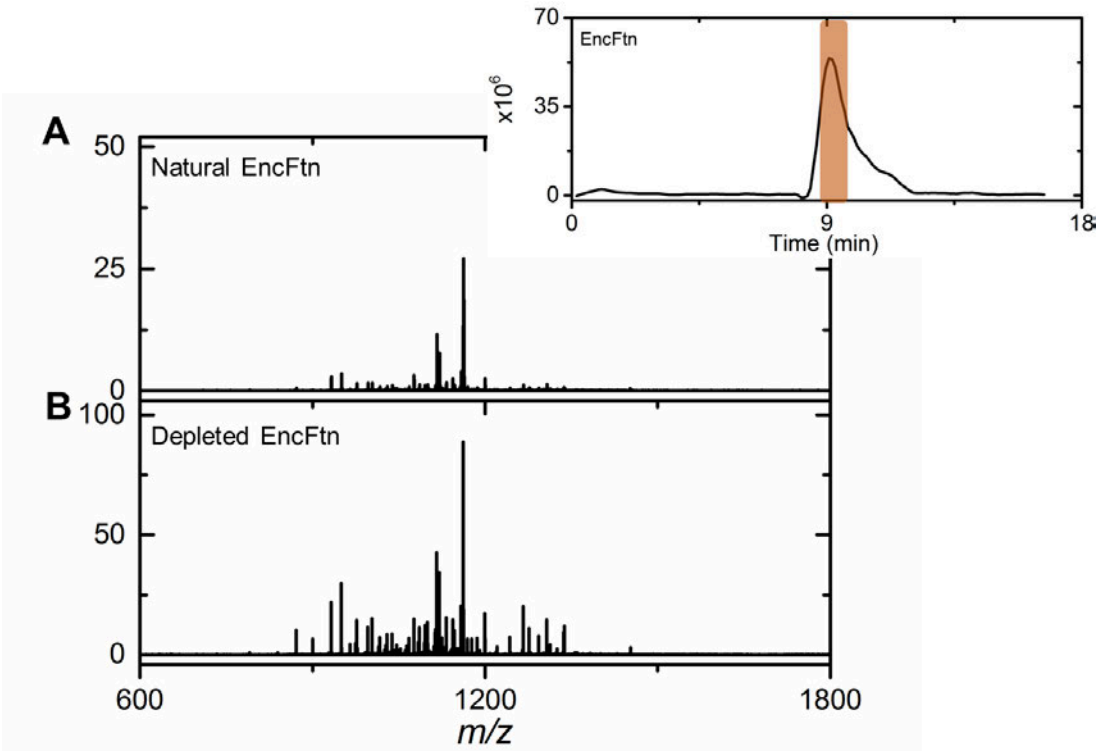


Figure 4.17: LC- top down CID fragmentation of EncFtn.

CID fragmentation of the $[M+14H]^{14+}$ charge state of isotopically natural (A) and isotopically depleted (B) EncFtn eluted through a C4 column (75 x 0.5 mm). Main chromatography elution peak for EncFtn eluted over approximately one minute. Area highlighted (orange) encapsulates ~5 transients over which data accumulation was averaged.

The spectra produced using CID fragmentation on EncFtn after elution from a C4 column is seen in Figure 4.17. The protein chromatograph shows a wide elution peak, with the area of greatest signal intensity used as the region over which the spectra is acquired. As in Figure 4.3, isotopically natural (Figure 4.17a) and isotopically depleted (Figure 4.17b) have similar fragmentation patterns, with the depleted fragments appearing to have a greater intensity due to the reduced number of isotopologue peaks present in each fragment ion. Due to the greater ion abundance of the isotopically depleted spectra in Figure 4.17b, the assignment of a greater number of ions is achieved. In Figure 4.18b, the 99 assigned *b* -type and *y* -type ions coverage 50.8% coverage. In contrast in isotopically natural EncFtn (Figure 4.18a) 15.5% of the primary sequence is covered by 33 fragment ions. As previously discussed, CID fragmentation is less likely to cause significant fragmentation in the central region of the polypeptide. Therefore the sequence coverage obtained with LC-CID fragmentation (Figure 4.18) is lower than that of the ECD fragmentation spectra with 5 transient accumulation in Figure 4.15a.

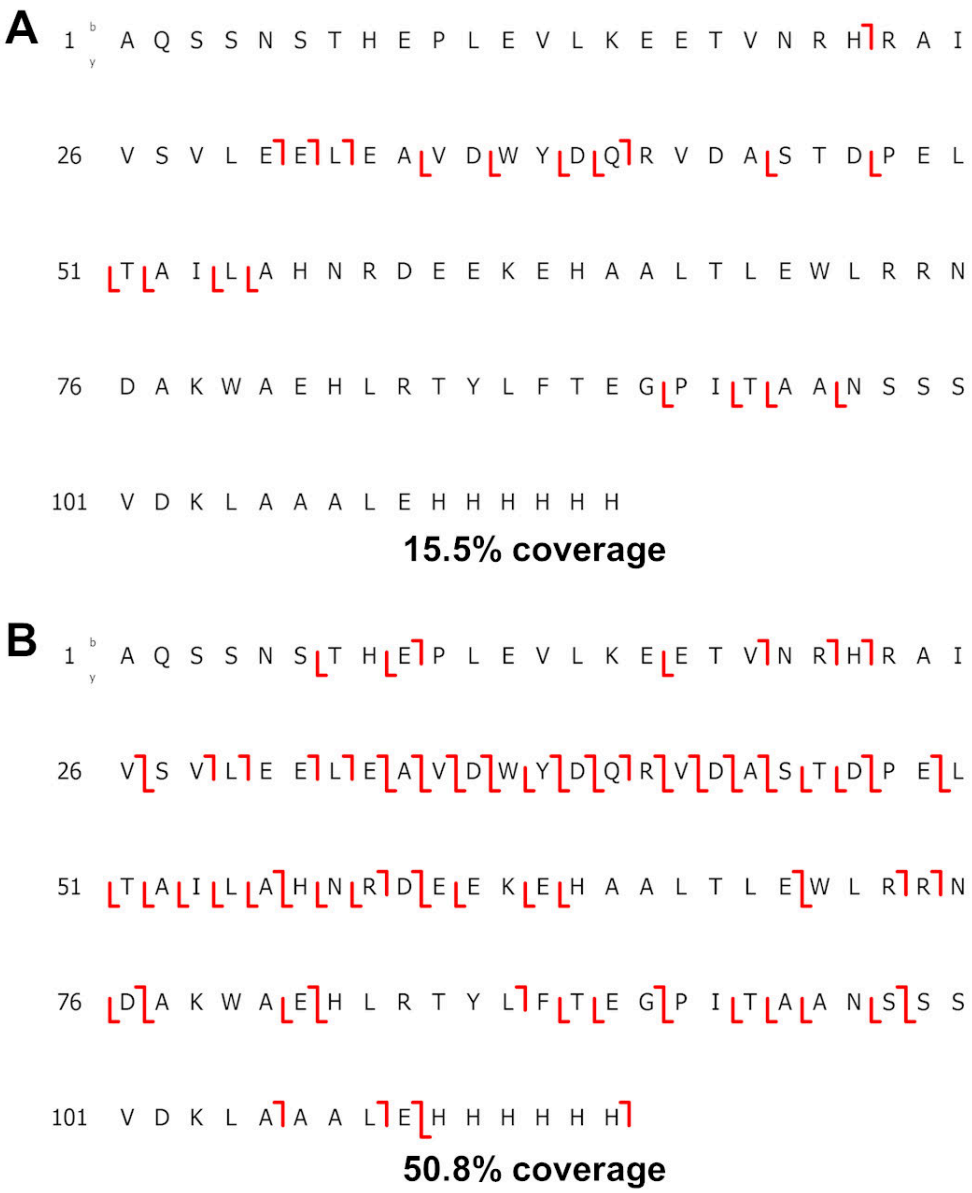


Figure 4.18: Sequence coverage of LC-CID fragmentation of EncFtn.

Sequence coverage of the [M+14H]$^{14+}$charge state of isotopically natural (A) and isotopically depleted (B) EncFtn after LC-CID fragmentation. The direction of the line indicates if the fragment is a N-terminal ion (b-type ion, ┐) or a C-terminal ion (y-type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

The same LC- MS/MS analysis can be carried out using ECD as the fragmentation technique.

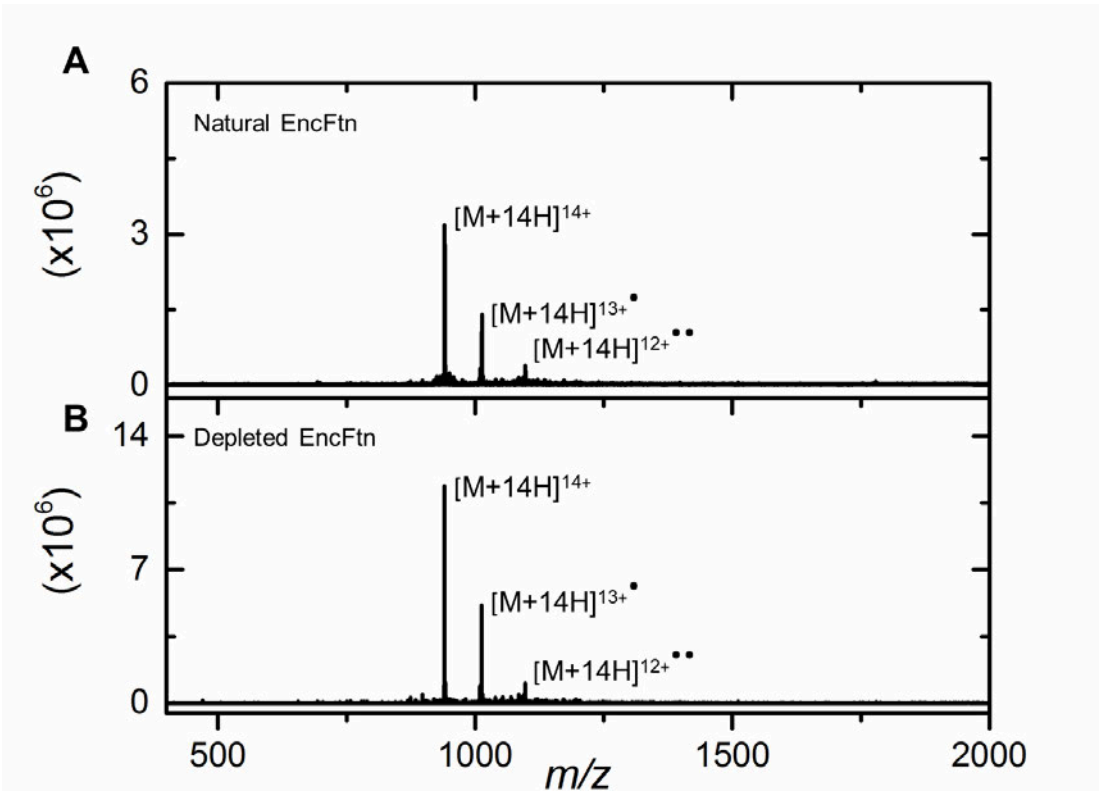


Figure 4.19: LC-ECD fragmentation spectra of EncFtn.

LC-ECD fragmentation of the [M + 14H]$^{14+}$ charge state of isotopically natural (A) and isotopically depleted (B) EncFtn. Main chromatography elution peak for EncFtn eluted over approximately one-minute allowing ~5 transients over which data accumulation was averaged.

The transient accumulation of EncFtn elution through the C4 column (75 x 0.5mm) was kept constant, producing consistent chromatography for the ECD spectra. In Figure 4.19 the charge reduced species of the [M+14H]$^{14+}$ charge state, which indicate electron capture, are clearly visible. Both fragmentation spectra display a low intensity due to the lack of spectral averaging. However low-level fragmentation can be particularly observed between the [M+14H]$^{13+\cdot}$ and the [M+14H]$^{12+\cdot\cdot}$. This is visually similar to the spectra obtained for the 5 transients averaged in Figure 4.15a, against which the sequence coverage can also be directly compared. The sequence coverage for the LC-ECD spectra is seen in Figure 4.20. As previously seen in the LC-CID fragmentation of EncFtn (Figure 4.18) the isotopically depleted EncFtn maintains a

high sequence coverage even with the low transient accumulation number. For the LC-ECD of EncFtn, the sequence coverage is maintained at 81%, greater than the 50.8% sequence coverage obtained during LC-CID (Figure 4.18b). The sequence coverage for isotopically depleted LC-ECD (81%, 199 ions, Figure 4.20b) is roughly similar to the 86.2% sequence coverage obtained in the 5 transient accumulations using direct infusion nESI (227 ions, Figure 4.15a). Interestingly, the isotopically natural sequence coverage after LC-ECD is only 17% (30 ions, Figure 4.20a), while the sequence coverage obtained using nESI is nearly double at 31% (48 ions, Figure 4.15a). The reduced number of ions and sequence coverage is due to the reduced efficiency of the ECD used to cause fragmentation. The number of ions are similar for both direct infusion nESI- ECD and LC-ECD, however isotopically depleted EncFtn consistently allows 2-4 X greater number of ions and sequence coverage compared to the isotopically natural EncFtn under identical circumstances.

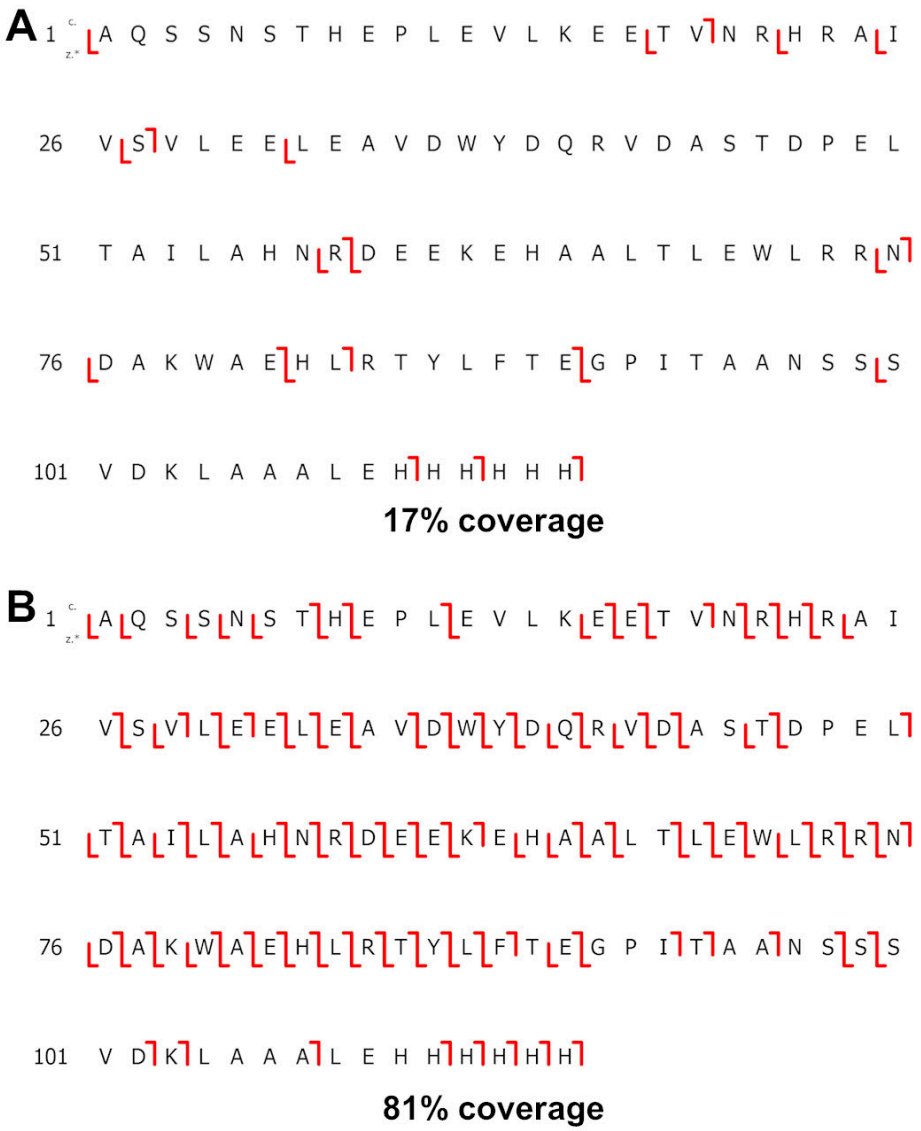


Figure 4.20: Fragmentation map of LC-ECD of EncFtn.

Sequence coverage of the $[M+14H]^{14+}$ charge state of isotopically natural (A) and isotopically depleted (B) EncFtn after LC-ECD fragmentation. The direction of the line indicates if the fragment is a N-terminal ion (c-type ion, ┐) or a C-terminal ion (z· -type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

4.7 Online fragmentation of BCA

Fragmentation of CA after elution from a C4 column (75 x 0.5 mm) was also analysed to determine if the sequence coverage obtained with lower transient averaging (Figure

4.15b) is reproducible. As carbonic anhydrase is a larger protein than EncFtn, the pulse length for ECD used for fragmentation will have a greater impact on the number of fragments produced and ultimately sequence coverage. This is observed in Figure 4.21 for isotopically natural CA. As the pulse length increases the number of

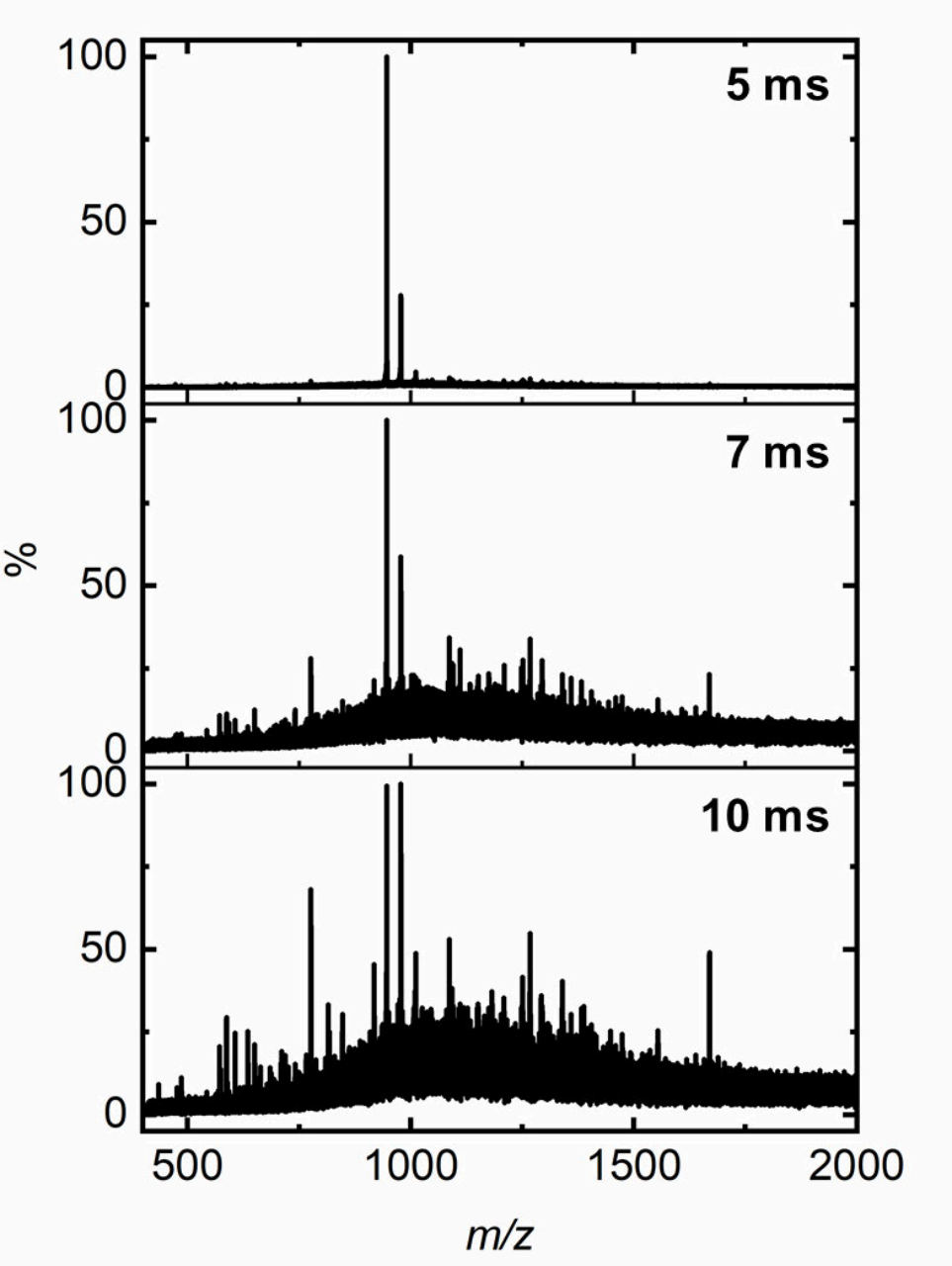


Figure 4.21: LC-ECD fragmentation spectra for isotopically natural BCA

ECD fragmentation spectra for isotopically natural spectra after elution from a C4 column (75 x 0.5 mm) at various ECD pulse length times (5 ms, 7 ms and 10 ms).

fragments in the spectra can be seen to be more numerous. The background noise of the spectra also increases, potentially as a result of more fragmentation channels being utilised to produce a greater number of fragments. Conversely will also result in the reduction of the total fragment ion abundance.

The precursor ion and charge reduced species which are very prominent within the 5 ms spectra, become greatly diminished as we increase first to 7ms then to 10 ms. Noticeably the 10 ms spectra in Figure 4.21 while displaying low S/N throughout the spectrum, contains the greatest number of fragment ions. When compared to the isotopically depleted BCA protein under identical conditions (Figure 4.22) there is an obvious visual increase of the fragment ions within the spectra. ECD fragmentation with a 5 ms pulse length in the isotopically natural BCA (Figure 4.21) particularly has few visible ions. In the isotopically depleted 5 ms spectrum (Figure 4.22) there are several ions that are easily distinguished from the background noise. So much so that several information rich areas are present across the spectra which directly contributes to the sequence coverage obtained. Likewise there is a further increase in the number of identifiable ions at 7 ms and 10 ms using isotopically depleted BCA. Both fragmentation spectra have clear fragment ions of varying intensity which have not become dispersed and lost in the background noise, as observed for some regions of the 10 ms of the isotopically natural BCA (Figure 4.21). The sequence coverage of the 10 ms isotopically natural and isotopically depleted fragmentation spectra after 10 ms ECD pulse length can be compared (Figure 4.23).

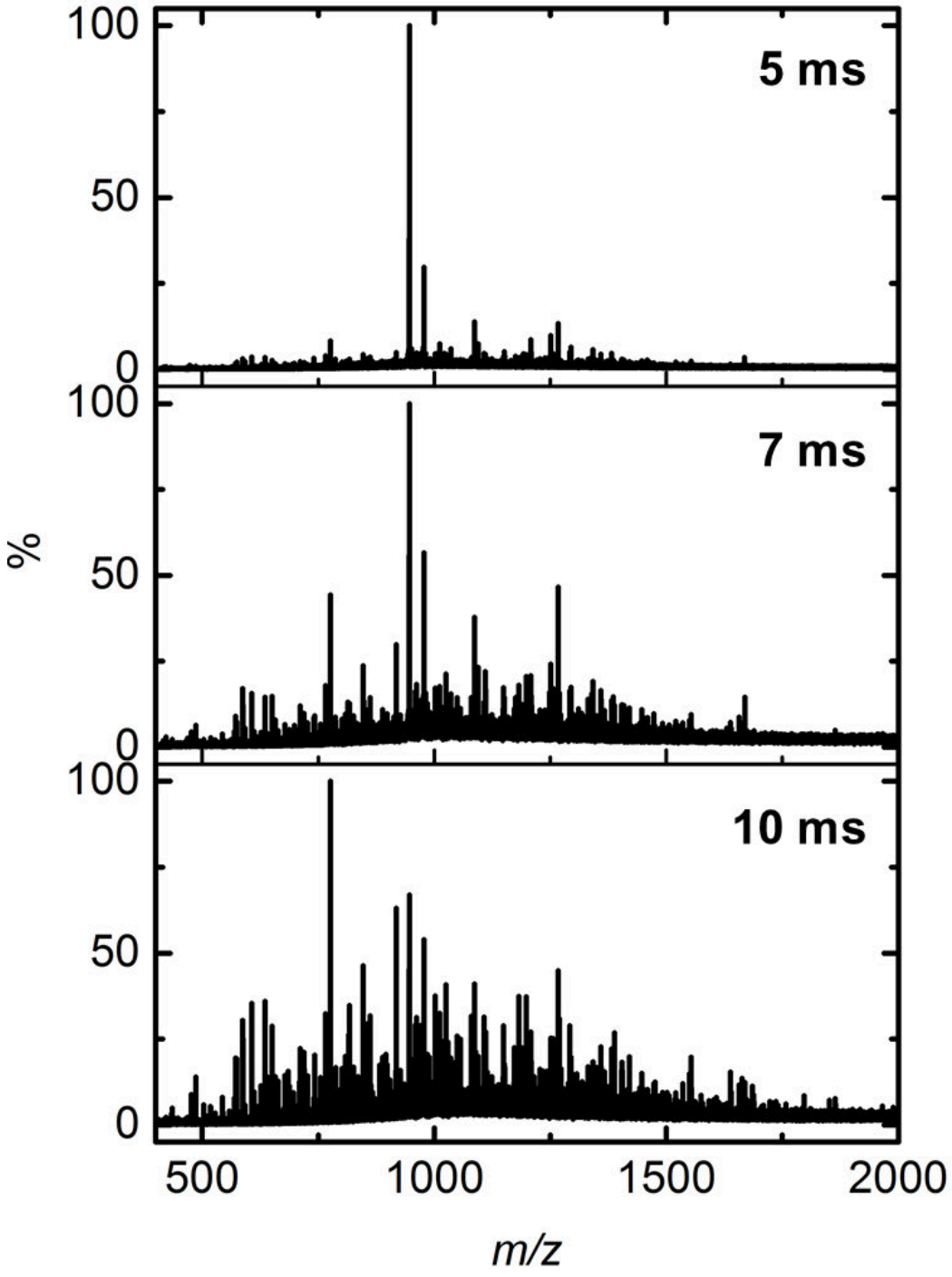


Figure 4.22: LC-ECD fragmentation for isotopically depleted BCA

ECD fragmentation spectra for isotopically depleted spectra after elution from a C4 column (75 x 0.5 mm) at various ECD pulse length times (5 ms, 7 ms and 10 ms).

It is clear from Figure 4.23 that the isotopically depleted BCA allows a significantly greater number of ions to be assigned, allowing a high level of sequence coverage. Due to the elution of BCA from the C4 column, the spectra and sequence maps are acquired over 13 averaged scans. When compared to the 20 averaged transients collected with direct infusion nESI (Figure 4.15), that 10 ms of ECD fragmentation of isotopically depleted BCA (Figure 4.23b) has a similar observed sequence coverage (64.3%, 345 ions) to that observed with 20 averaged transients (Figure 4.15). The fragmentation spectra of isotopically natural BCA (Figure 4.23a), also acquired over 13 averaged transients has an assigned sequence coverage (12.1%, 36 ions) a lower

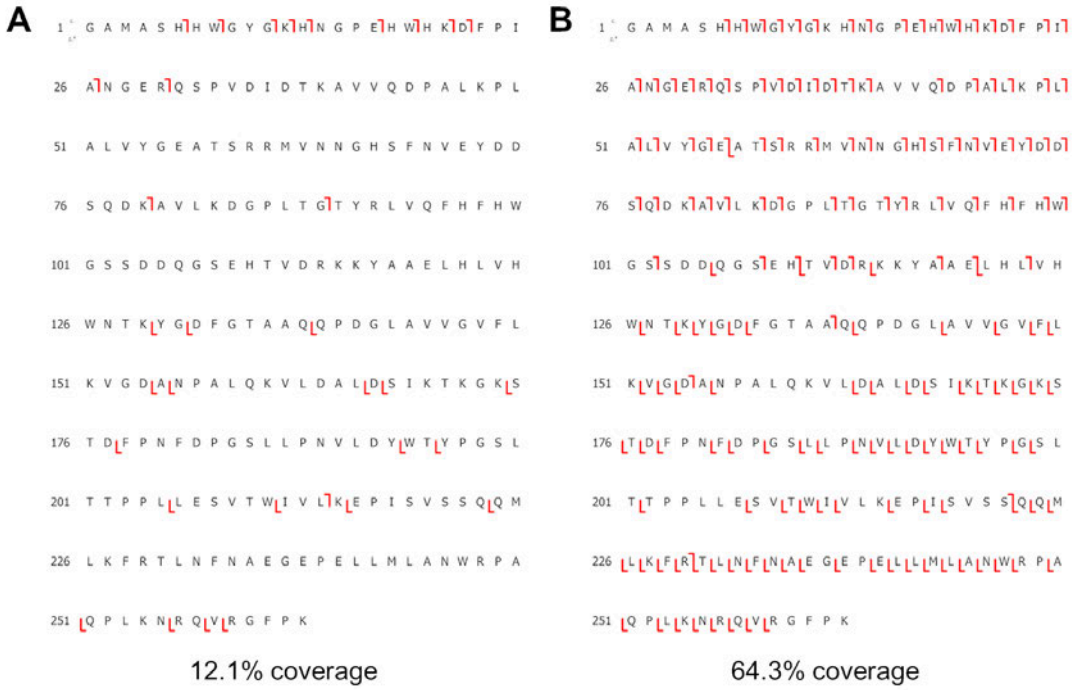


Figure 4.23: Fragmentation map of LC-ECD of BCA.

Sequence coverage of the $[M+31H]^{31+}$ charge state of isotopically natural (A) and isotopically depleted (B) BCA after LC-ECD fragmentation with a pulse length of 10 ms. The direction of the line indicates if the fragment is a N-terminal ion (c-type ion, ┐) or a C-terminal ion (z· -type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

than the nESI (Figure 4.15b) with 5 averaged transients, is within the region of sequence coverage we would predict.

Crucially, the isotopically depleted protein for EncFtn (Figure 4.20b) and BCA (Figure 4.23b) both display a significantly greater sequence coverage than their isotopically natural counterparts. As many on-line experiments are complex biological mixtures with a proteomics outlook, the sequence coverage achieved would be sufficient to obtain unique sequence tags allowing confident identification of the protein if the sample was unknown. [111] The same cannot be said of the isotopically normal protein under theses experimental conditions, which do not generate regions of heavy fragmentation to allow de novo sequencing required to allow identification of the an unknown protein. [131]

4.8 Top down fragmentation of SPT

From the increase in molecular weight from EncFtn (13 kDa) to BCA (30 kDa) we can see a significant decrease in the achievable sequence coverage. With a further increase in MW, we can predict this trend to continue. The increasing MW results in the total signal of the protein decreasing, as the signal is split between greater numbers of isotopologue peaks (Figure 1.4) [132,133]. Direct infusion ECD analysis of SPT (47 kDa) on the 12T FT-ICR is observed in Figure 4.24.

SPT as well as being a larger molecular weight protein, poses an extra challenge due to the post translational modification ($+C_6H_{10}O_6$, +178 Da) on the hexahistidine tag (Figure 3.10), which further diminishes the available concentration of the unmodified precursor ion. [105]

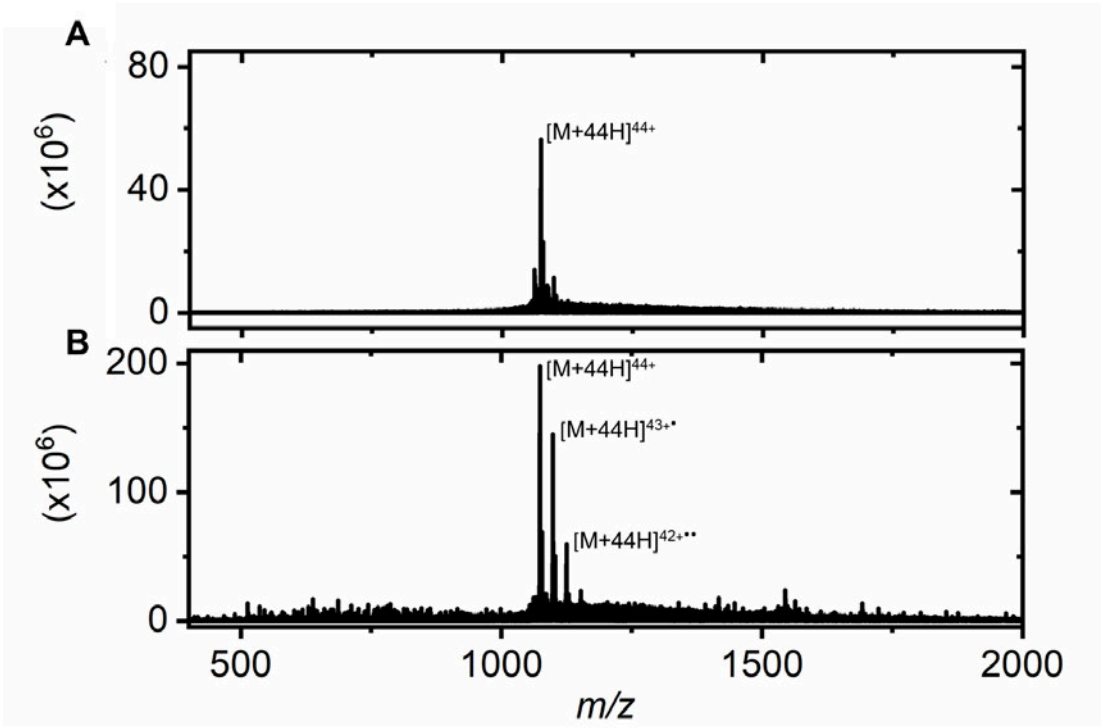


Figure 4.24: ECD fragmentation spectra of $[M+44]^{44+}$ charge state of SPT

ECD fragmentation spectra of the $[M+44]^{44+}$ charge state of isotopically depleted SPT. Isolation of the $[M+44]^{44+}$ charge state was centred on unmodified species for fragmentation. The spectrum was the result of a single acquisition with 300 averaged transients.

In Figure 4.24 we can see the characteristic charge reduced species and the produced low level fragment ions which can be seen in greater detail over the *m/z* region of 750-850 in Figure 4.25. Highlighting how few ions can be assigned within this region of the mass spectrum of the isotopically natural SPT spectra (Figure 4.25a), only 8 ions over a 100 *m/z* region could be assigned confidently above the background noise. Over the same region in the isotopically depleted SPT fragmentation spectrum (Figure 4.25b) contains 29 fragment ions.

Looking at both spectra of Figure 4.25 there are a number of peaks which are partially resolved but have not been assigned. This in part is due the low ion abundance of the $[M+44]^{44+}$ charge state containing the modification that was not efficiently removed by

quadrupole isolation of the unmodified $[M+44]^{44+}$ charge state of SPT. Secondly, many of the fragment ions are not sufficiently resolved from the background noise for the ions to be assigned with sufficient confidence. However, comparison of the ECD spectra of SPT are an excellent example of the potential improvement isotopic depletion can facilitate.

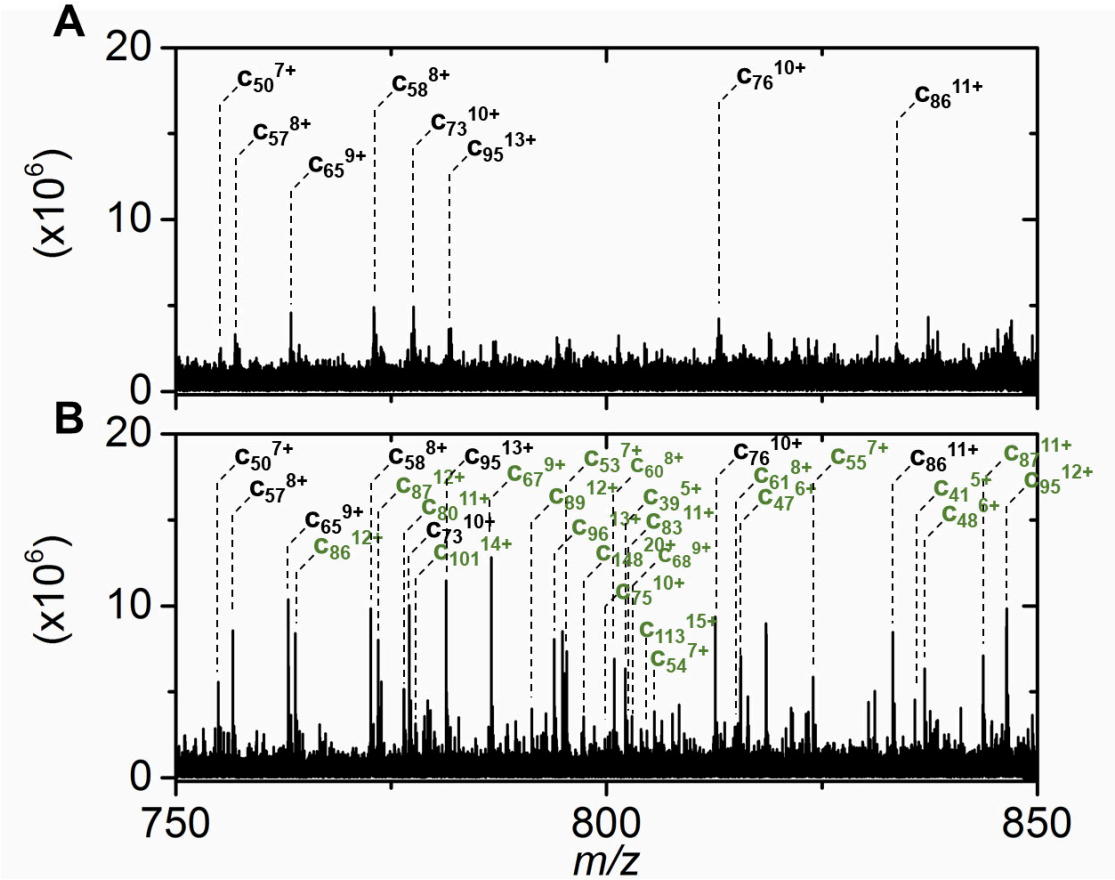


Figure 4.25: ECD fragment ion assignments of SPT.

View of the ECD fragmentation spectra, over the m/z region of 750-850 of the isotopically natural SPT (A) and isotopically depleted SPT (B). Assigned ions are labelled with the ions only found in the isotopically depleted spectra labelled in green.

The difference in the quality of the obtained spectra can be more closely seen in the smaller 2 m/z (Figure 4.26a) and 8 m/z (Figure 4.26b) ranges. Figure 4.26a covers the 776-778 m/z range, which is present within the range of Figure 4.25 , in which the

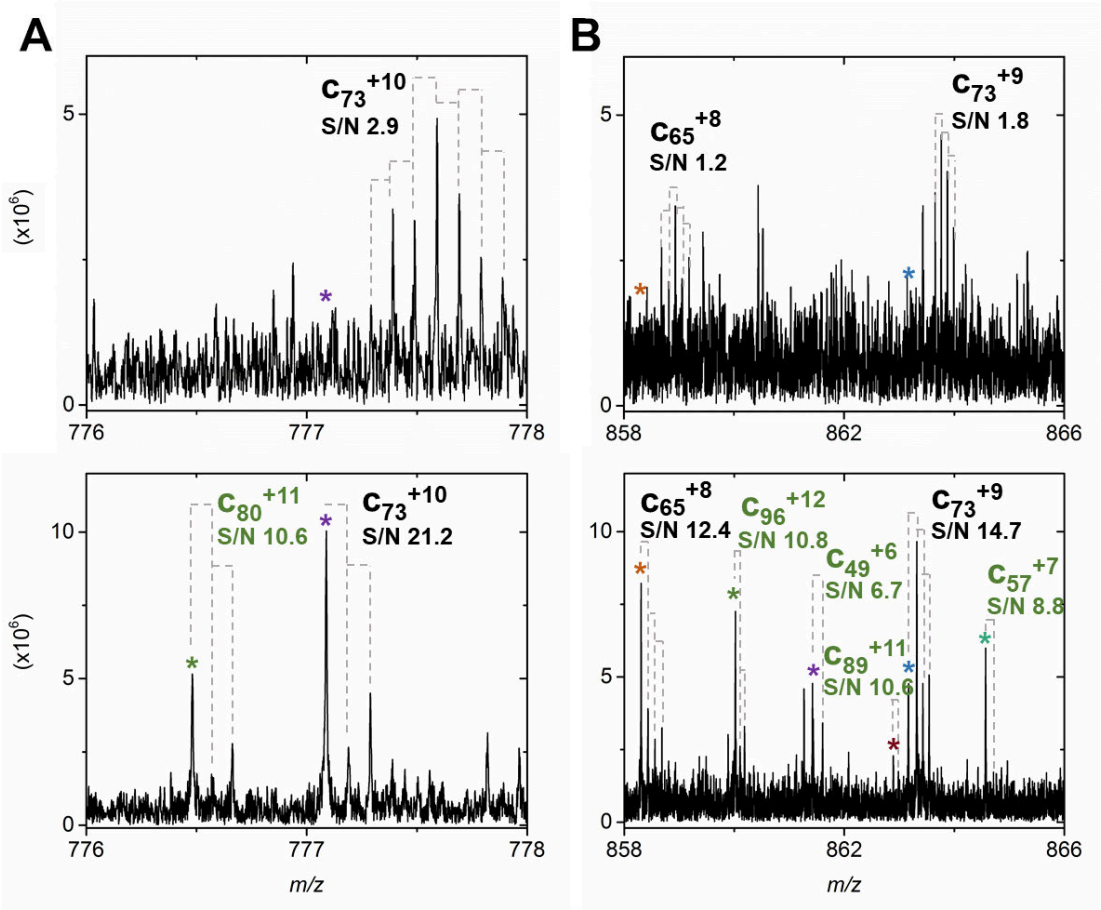


Figure 4.26: Close view at 2 m/z and 8 m/z of SPT fragmentation spectra.

Small section of the ECD fragmentation spectra comparing isotopically natural (top) and isotopically depleted (bottom) protein of SPT. The region of 776-778 (A) and 858-866 (B) highlight the difference in the number of ions which can be assigned between each spectrum.

c_{73}^{10+} ion was assigned in both spectra. The fragment ion has an average mass of 7765.56 Da. So is a large enough molecular weight for the monoisotopic peak (*, 777.09 m/z) to not contribute significant intensity to the isotopically natural isotopologue distribution (~1%). Due to the low signal of the assigned c_{73}^{10+} ion the monoisotopic and many of the lower abundance isotopologue peaks cannot be resolved from the background noise.

The monoisotopic peak of the isotopically depleted c_{73}^{10+} ion contributes ~55% of the total fragment abundance and possesses half the number of total isotopologue peaks;

ultimately this ion is detected with 7X greater signal than in the isotopically natural spectrum. The same trend can also be seen for the c_{73} ion at the 9^{+} charge state

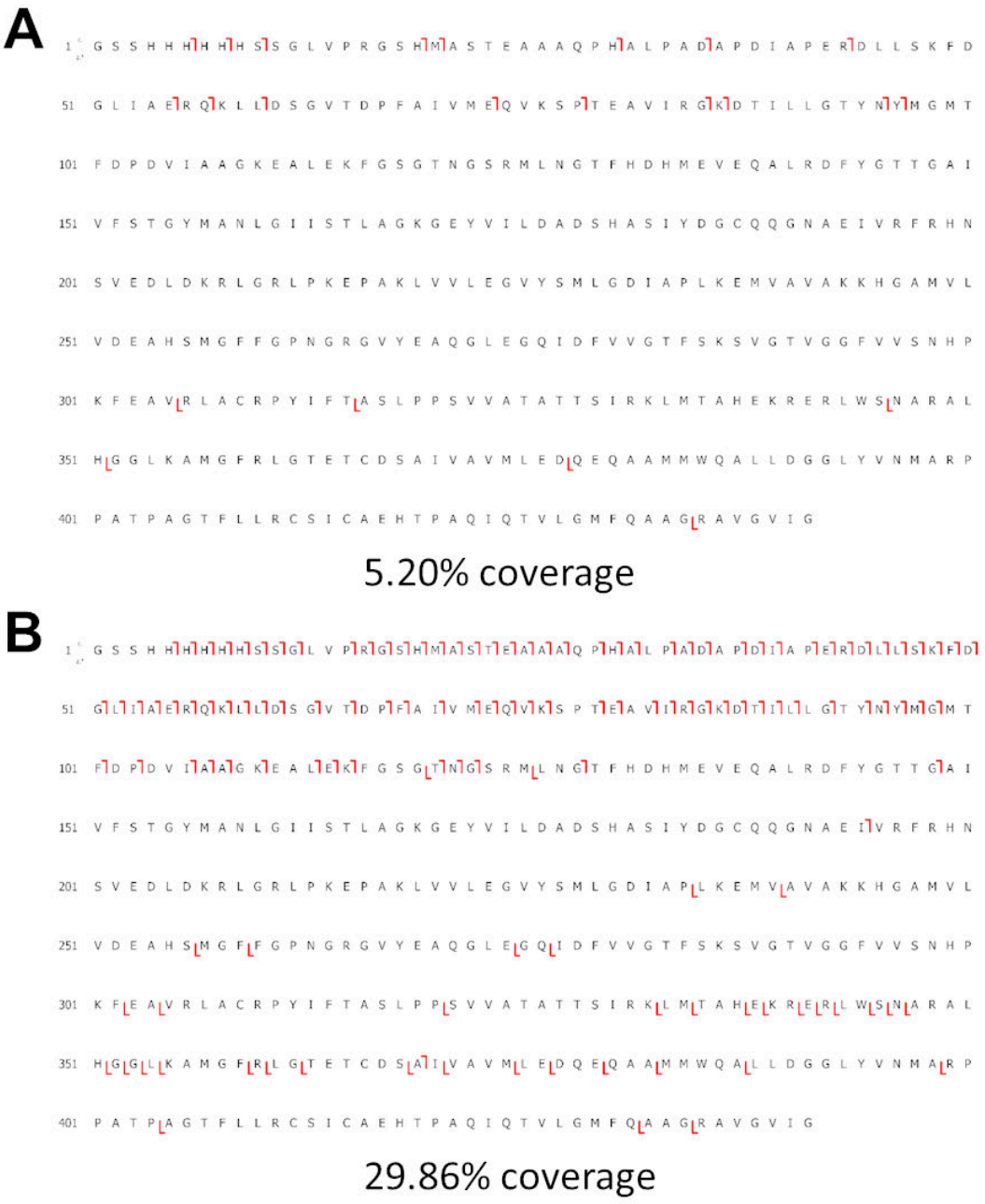


Figure 4.27: Sequence coverage of SPT after ECD fragmentation spectra.

Sequence coverage of SPT after ECD fragmentation. The direction of the line indicates if the fragment is a N-terminal ion (c-type ion, ┐) or a C-terminal ion (z· - type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

(Figure 4.26b). For both the isotopically natural and isotopically depleted forms of c_{73}^{9+} exist with just over half the signal observed in the c_{73}^{10+} ion (Figure 4.26a). However, the isotopically depleted c_{73}^{10+} and c_{73}^{9+} are detected with approximately 8X greater signal than the isotopically natural fragment ions. Both Figure 4.25 and Figure 4.26 highlight the improvement in ion assignment density observed within the isotopically depleted SPT ECD fragmentation spectra. This is seen at the amino acid sequence level in Figure 4.27.

As a result of the low number of fragments observed in Figure 4.25a and Figure 4.26a, it is predictable that there is a very low sequence coverage in Figure 4.27a. The fragmentation events primarily focus at the N-terminus of the protein, with a few fragmentation events present at the C-terminus. [51,90,134,135] These ions are small MW ions and can be resolved from the background noise. Due to the increase in potential fragment ions which can be produced, larger fragments from the central regions of the protein are unlikely to have sufficient intensity to be resolved from the background noise. Therefore the sequence coverage is very low, with only 23 assigned fragments, accounting for 5% of the sequence coverage. Such limited sequence coverage makes downstream analysis of any form difficult. Under the same challenging conditions, the isotopically depleted protein in Figure 4.27b is just under 6X greater in total sequence coverage (~30%) and 121 fragment ions. Like in the isotopically natural SPT (Figure 4.27a) the fragmentation of isotopically depleted SPT (Figure 4.27b) occurred almost exclusively at the N-terminus of the amino acid sequence, but has the advantage of assignment of clusters of fragmentation events along the polypeptide sequence at the N-terminus. Having several sequential fragmentation events (sequence tags) is ideal for de novo sequencing if the protein was an unknown and can allow identification of the protein from databases. [84,136] While it is clear that isotope depletion allows a substantial increase in the achievable sequence coverage in a challenging sample

like SPT, there is also an obvious lack of fragmentation around the C-terminus (Figure 4.27b). Containing few positively charged residues (lysine and arginine) which promote electron capture on the polypeptide chain. [127] So would likely have limited assignable fragmentation at the C-terminus even with superior quality of the fragmentation spectra.

4.9 Conclusion

It is clear from the data shown in this chapter that recombinantly expressed protein from isotopically depleted M9 minimal media, can provide a benefit to top-down fragmentation analysis. The observed simplification of the protein isotopologue distribution confers many benefits to protein MS analysis. This was tested with the application of ECD fragmentation on a 12T FT-ICR platform over a range of molecular weight protein standards, ranging from ~13 kDa to ~47 kDa. Displaying a varying improvement in the achievable protein sequence coverage. The total achievable sequence coverage decreased with each increase in molecular weight. For each of the MS^2 fragmentation spectra shown in this chapter are of single spectral acquisitions of various transient number. These chosen spectra are often the best example of protein fragmentation and will display small variations between different fragmentation acquisitions and charge states. These small variations are likely the result of interactions between the polypeptide chains or maintained tertiary structure. The chosen charge state for fragmentation will also have an impact, as larger number of charges will have a greater fragmentation efficiency. This ultimately means that every sample and fragmentation acquisition will produce slight differences in the assignable ions.

The isotopically depleted protein at each molecular weight consistently permitted the assignment of a greater number of fragment ions, regardless of fragmentation technique *i.e.*, CID or ECD, therefore facilitated a greater sequence coverage. This is

directly the result of the simplification of the isotopologue distribution. With the fragmentation of the intact protein precursor, many fragment ions of a MW less than that of the precursor are produced. In isotopically depleted protein, the fragment ions also display simplified isotopologue distributions and more prominent monoisotopic peaks. Allowing confident assignment on the fragment ion accurate mass. As the ion signal gets split between fewer isotopologue peaks, there is often at least a 2X greater S/N for the observed fragment ions. However, it is important to note that due to the sample variation, fragment ions were assigned in the isotopically natural fragmentation spectra which were not observed in the isotopically depleted comparison. This only occurred for very few fragment ions and was always vastly outnumbered by the large number of ions found in the isotopically depleted spectra, compared to the isotopically natural.

For both EncFtn and BCA isotopic depletion it has been shown to provide greater sequence coverage resilience when the number of transients averaged during data collection substantially decreases. This has the potential to vastly improve analysis within the rapidly developing '-omics' fields which aims to achieve a comprehensive view of the global biomolecule content within a cell line [130,137–139] or recent developments within single cell proteomics. [140] In which simplified isotopologue profiles and a prominent monoisotopic peak could easily improve current analysis. SPT at 47 kDa is approaching the upper limit of our current capacity to facilitate effective fragmentation of our protein ion precursor. [110,122] Many studies have focused on technological improvements to MS instruments to extend this limit. [38,122] This chapter demonstrates that isotope depletion is a cost effective method to obtain significant improvement in ion assignment, fragmentation products or intact protein.

5. Isotope Depletion for improved top-down sequence coverage of larger molecular weight proteins – using ETD and Proton Transfer Charge Reduction.

5.1 FT-ICR vs Orbitrap

As described in the introduction (section 1.1.8, page 18 and section 1.1.9, page 22) FT-ICR and orbitrap analysers both produce FID data outputs, which is converted into high resolution mass spectrum. Top-down ECD fragmentation carried out on a 12T FT-ICR (chapter 4) highlighted the increase in the number of assignable fragment ions within isotopically depleted fragmentation spectrum. The orbitrap as an analogous analyser to FT-ICR supports ETD fragmentation which will also result in the fragmentation of the N-Cα bond and produce *c*-type and *z·* -type fragment ions. Applying the same top-down fragmentation analysis on the orbitrap Eclipse demonstrates that isotope depletion can improve protein MS analysis on multiple platforms and consistently allow greater sequence coverage of the model protein in top-down fragmentation analysis.

The orbitrap Eclipse can also facilitate proton transfer (PTCR), which has been demonstrated to help with top-down analysis of larger MW protein. [78,79] In this chapter, PTCR is used in combination with isotope depletion to improve the top-down analysis of large proteins.

5.2 ETD fragmentation of EncFtn and BCA

Top-down ETD of EncFtn and BCA was performed on an orbitrap Eclipse (Thermo Fisher Scientific). As in chapter 4, a comparison of the top-down fragmentation spectra of the isotopically natural and depleted forms of the protein to ETD fragmentation again displays the simplified isotopologue distribution in isotopically depleted protein spectra. Applying ETD analysis on both EncFtn and BCA we can observe a trend similar to that observed in the ECD fragmentation (chapter 4).

Figure 5.1a displays the ETD fragmentation spectrum of the $[M+14H]^{14+}$ charge state of EncFtn. As in the previous ECD analysis, direct comparison of fragment ions

produced with ETD fragmentation of isotopically natural (Figure 5.1b) and isotopically depleted EncFtn (Figure 5.1c). Again, we observe the simplification of the depleted isotopologue distribution. This is repeated with the ETD fragmentation of BCA (Figure 5.2).

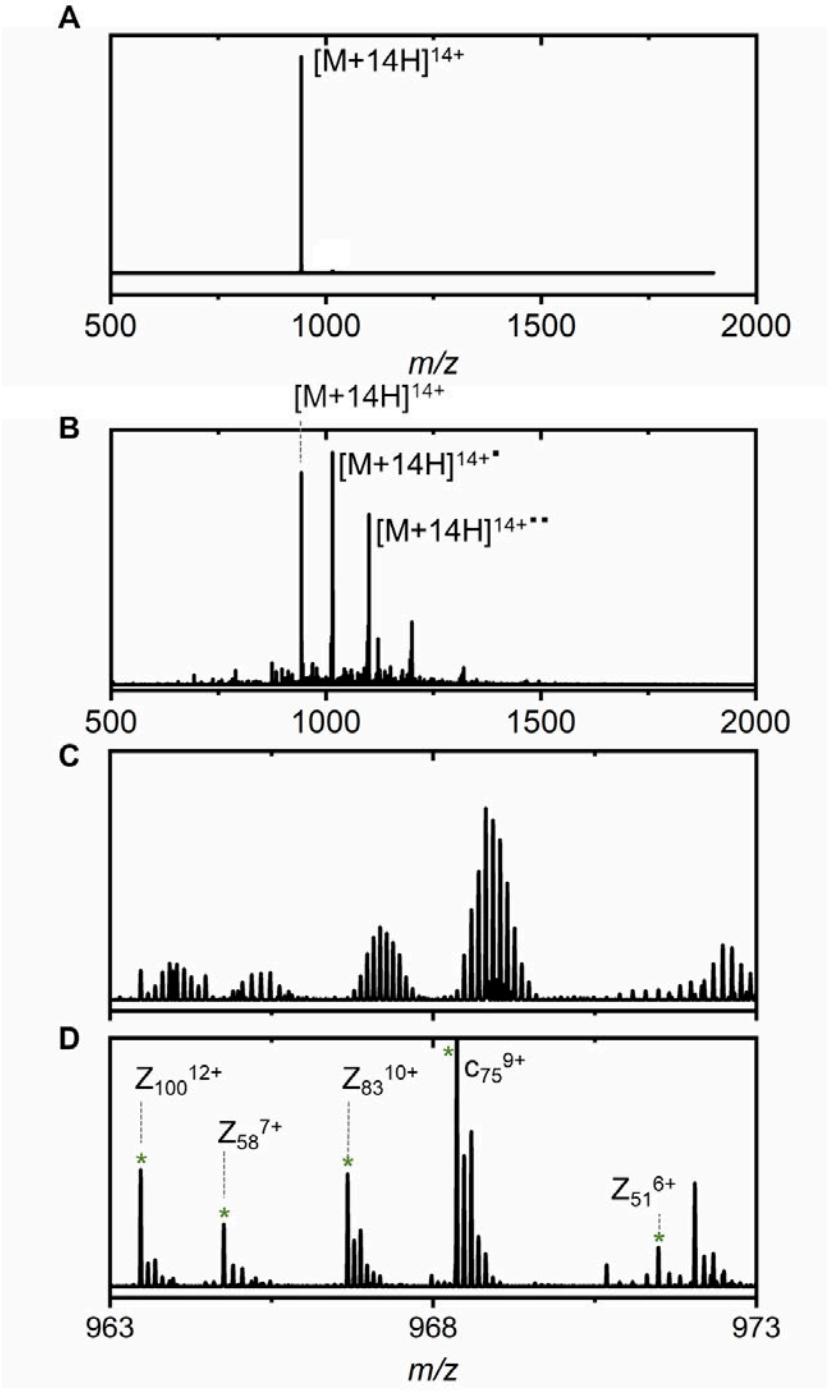


Figure 5.1: ETD fragmentation of $[M+14H]^{14+}$charge state of EncFtn

Comparison of the produced fragment ions of the $[M+14H]^{14+}$charge state of EncFtn after 3 ms of ETD fragmentation (A). Examples of the produced isotopologue distribution of isotopically natural (B) and isotopically depleted (C) fragments within the 963-973 m/z region. The asterisk (*) indicates the monoisotopic peak of the isotopically depleted distribution and ion assignment labelled above the depleted ion can also be observed in both fragmentation spectra.

Isotopically depleted BCA (Figure 5.2c) similarly is simplified and the monoisotopic peak the most prominent within the isotopic envelope. The isotopically depleted forms of both proteins allow greater fragment ion assignment, and therefore greater sequence coverage. The collected spectra for both EncFtn and BCA are similar to that achieved during ECD fragmentation. As a result of this, the obtained sequence coverage is similar to that previously achieved. Increasing from 52% in isotopically natural EncFtn (Figure 5.1b) to 95% in isotopically depleted EncFtn (Figure 5.1c). Likewise isotopically natural BCA (Figure 5.2b) has 32% sequence covered, while the isotopically depleted BCA 72% of sequence coverage were obtained (Figure 5.2c). Much like in chapter 4, we can conclude that the simplification of the isotopologue distribution will consistently increase the information we can extract from a single fragmentation spectrum.

As the ETD fragmentation of EncFtn and BCA are very similar to those previously discussed in chapter 4 and the benefit of analysing the isotopically depleted protein was so thoroughly discussed in chapter 4 a full repetitive analysis is not required for this ETD fragmentation. Both Figure 5.1 and Figure 5.2 display examples of large fragments, such as z_{100}^{12+} ($C_{503}H_{771}N_{151}O_{160}S_2$, 11550.6 Da, Figure 5.1) and c_{190}^{24+} ($C_{934}H_{1414}N_{263}O_{281}S_2$, 20874.4 Da, Figure 5.2) both encompass over 70% of the total of their respective protein sequence. In the isotopically depleted protein spectra of both protein (Figure 5.1c and Figure 5.2c), simplified isotope distributions facilitate the assignment of these large mass fragment ions using the accurate mass. The usefulness of this increases with protein MW. As larger MW protein have a greater number of fragmentation products, they ultimately will produce more densely populated fragmentation spectra and larger MW fragment ions. Each fragment ion will ultimately be present in the spectra at lower signal. This in part contributes to the challenge of applying top-down fragmentation techniques to larger MW protein.

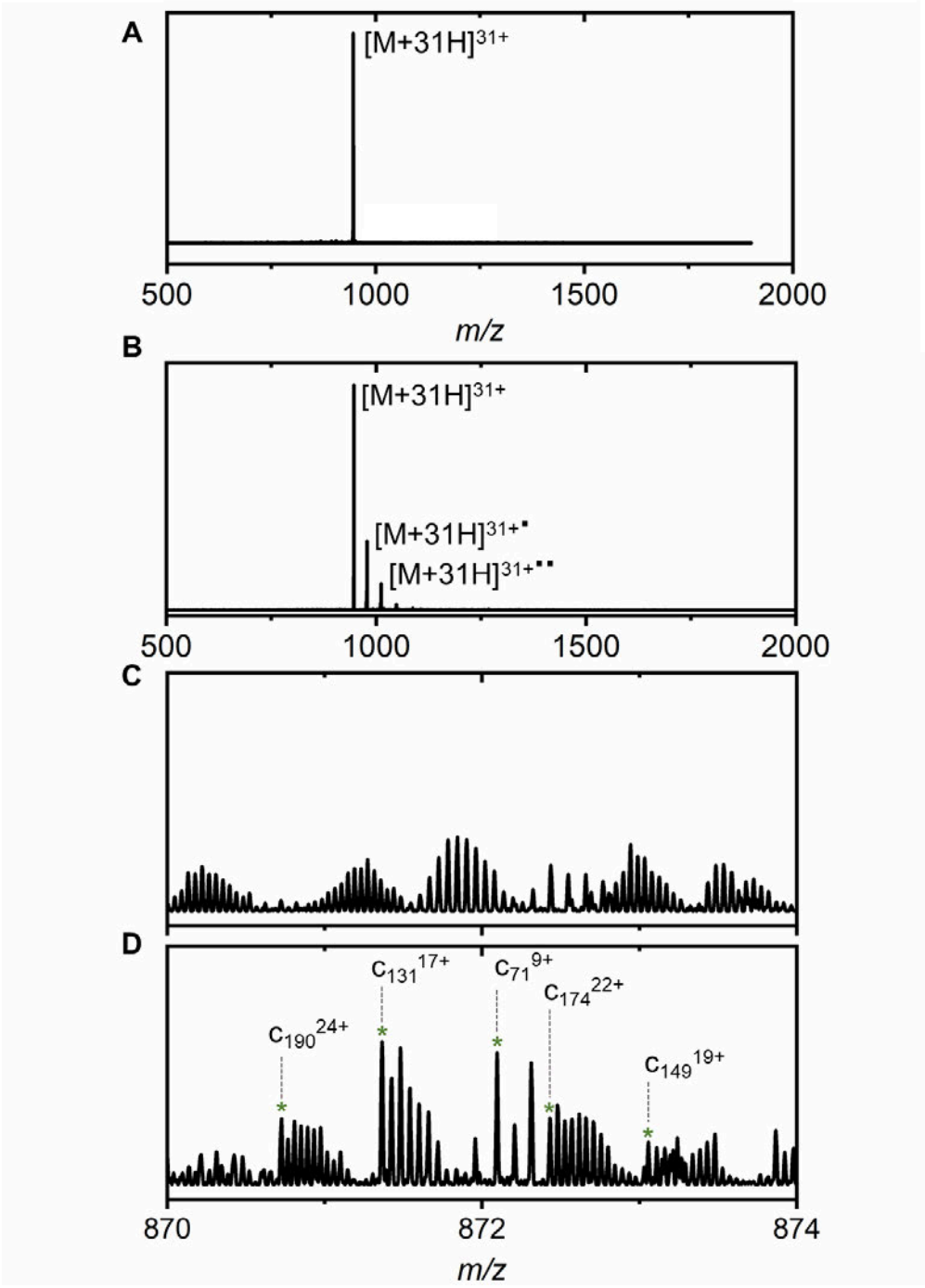


Figure 5.2: ETD fragmentation of the $[M+31H]^{31+}$BCA

Comparison of the produced fragment ions of the $[M+31H]^{31+}$charge state of EncFtn after 1 ms of ETD fragmentation (A). Examples of the produced isotopologue distribution of isotopically natural (B) and isotopically depleted (C) fragments within the 870-874 m/z region. The asterisk (*) indicates the monoisotopic peak of the isotopically depleted distribution and ion assignment labelled above the depleted ion can also be observed in both fragmentation spectra.

Pushing the drive for technologies which improve top-down down fragmentation performance (sensitivity, fragmentation efficiency, sequence coverage) at higher mass.

5.3 ETD fragmentation of SPT

The top-down ECD analysis of SPT described in chapter 4 attempted to demonstrate improvement in achievable sequence coverage with isotope depletion. However, the highest achieved sequence coverage on the 12T FT-ICR was ~5% for the isotopically natural SPT and ~30% for isotopically depleted SPT (Figure 4.27). SPT protein was also subjected to ETD fragmentation on the orbitrap Eclipse, which is demonstrated by the fragmentation spectra of isotopically depleted SPT in Figure 5.3. For this analysis, the electron exposure time was varied, and it was evident that increasing the exposure time, led to an increased number of fragments ions. At 0.75 ms of ETD fragmentation, there is very little electron capture and fragmentation and the isolated $[M+37H]^{37+}$ charge state remains the most abundant ion. Increasing the fragmentation time to 1 ms or further to 1.5 ms, produces a range of ETD fragment ions the lower m/z range (600-1000 m/z) and significantly decreases the abundance of the of the remaining isolated precursor and charge reduced species. The 1.5 ms produced the largest number of fragment ions and displays significant fragmentation within the central region of the protein. Shown in the fragmentation map of Figure 5.4b.

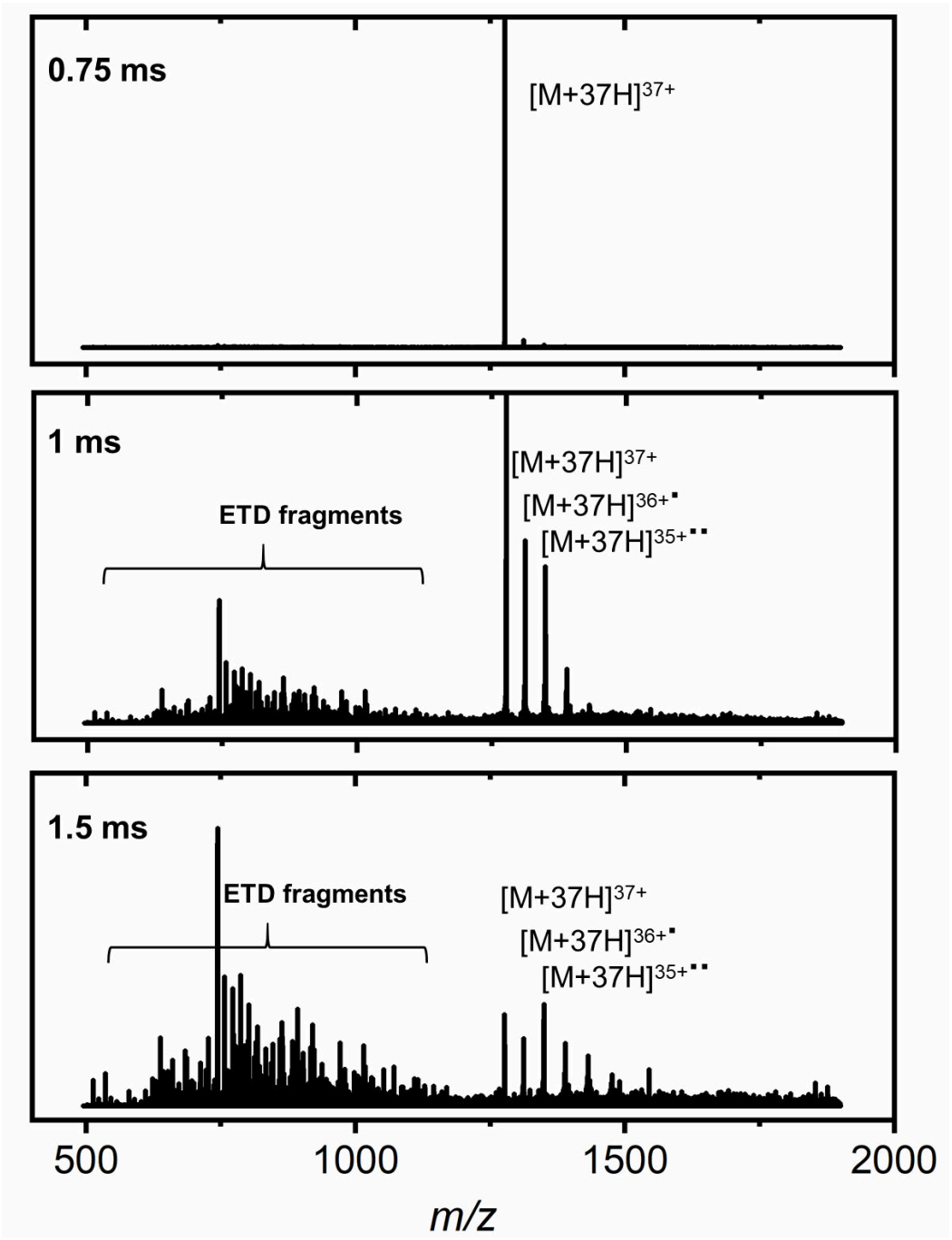


Figure 5.3: ETD fragmentation of SPT with increasing fragmentation time.

ETD fragmentation spectra of the $[M+37H]^{37+}$ charge state of isotopically depleted SPT after 0.75, 1, or 1.5 ms exposure to the electron donor. The precursor and the first two charged reduced species are labelled and the lower m/z fragment ions are indicated.

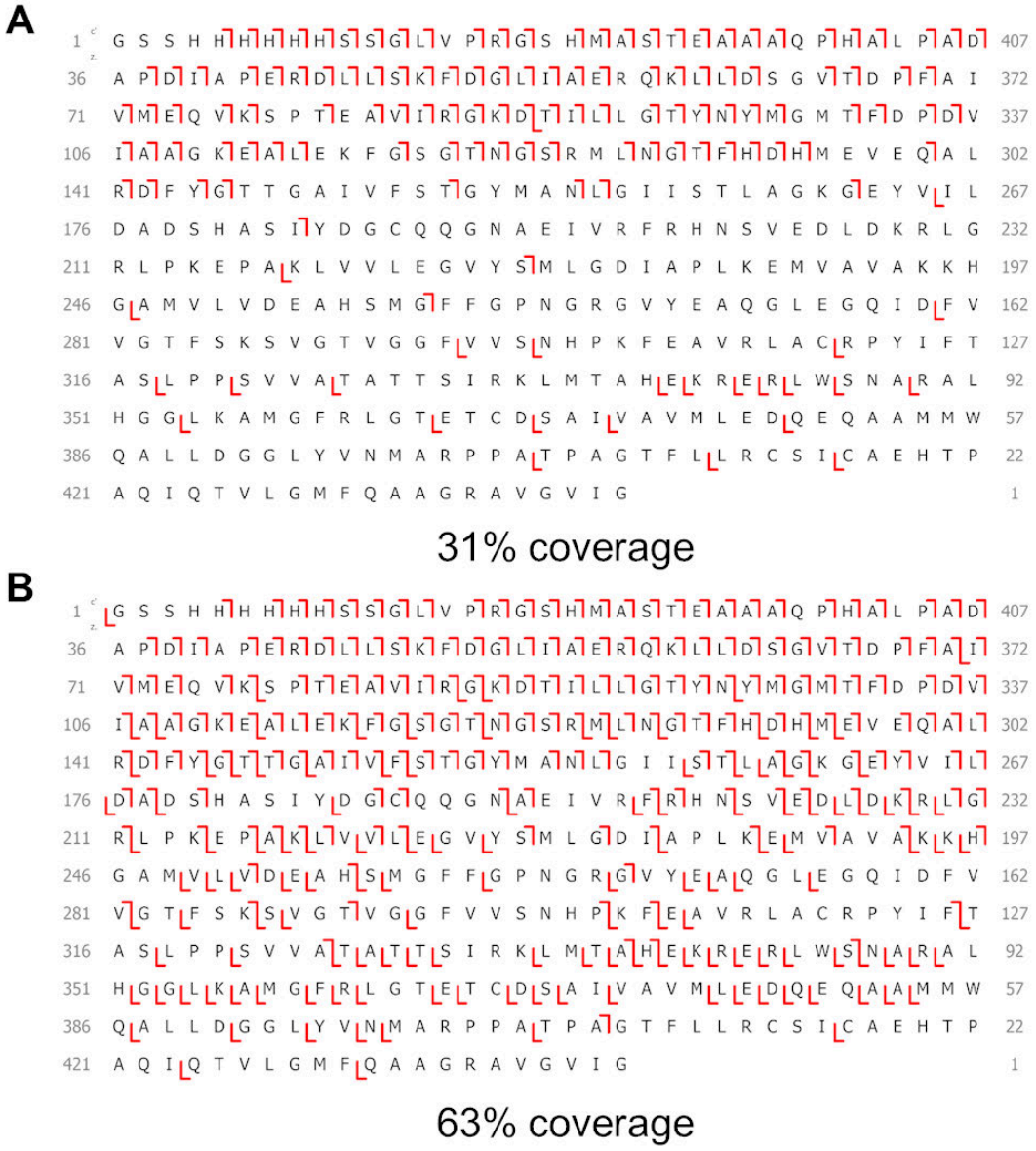


Figure 5.4: Sequence coverage of the $[M+37H]^{37+}$ charge state of SPT after ETD fragmentation.

Sequence coverage of isotopically natural (A, 31% coverage) and isotopically depleted (B, 62% coverage) SPT after 1.5 ms of ETD fragmentation. The direction of the line indicates if the fragment is a N-terminal ion (c-type ion, ┐) or a C-terminal ion (z· -type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

Both acquisitions are the sum of 100 averaged data scans.

The isotopically natural protein (Figure 5.4a) had sufficient ion assignment to cover 31% of the protein sequence. Which is comparable to that achieved with isotopically depleted SPT with ECD fragmentation (Figure 4.27b). Similarly ETD fragmentation of the isotopically depleted SPT (Figure 5.4b) displays a large increase in sequence coverage, which is more than double that previously achieved with ECD (Figure 4.27b) to 63% coverage. These initial results suggest that the orbitrap platform maybe more suitable for top-down fragmentation of larger proteins.

The (orbitrap) ETD spectra in Figure 5.4 display similarities to the isotopically depleted (ICR) ECD fragmentation spectra. Notably, fragmentation is again primarily focused around the N-terminus. With almost full coverage over the first 100 amino acids for both isotopically natural (>200 *c*-type fragment ions) and depleted SPT (>500 *c*-type fragment ions; Figure 5.4). In contrast, very little fragmentation is observed at the C-terminus. As this is similar to the ECD fragmentation of SPT on the FT-ICR platform and we propose this consistency in fragmentation is maybe due to the limited number of positive residues in this region of the protein.

The difference in sequence coverage obtained for SPT between the 12T FT-ICR and orbitrap Eclipse is surprising. For both EncFtn and BCA the sequence coverage obtained was comparable. As discussed in the introduction (section 1.1.8, page 18) FT-ICR spectra resolution, as a result of the magnetic field, is inversely related to the *m/z* of the ion. Whereas the orbitrap is inversely related to the square root of the *m/z.* However, the difference in sequence coverage is more likely as a result in the different fragmentation efficiencies of the ExD techniques. Reaction times for ECD can range from ~5 ms to ~70 ms and are typically at the lower end of the time range for larger MW protein. ETD reaction times tend to vary between ~50 ms to ~200 ms, [60] and produces less electron capture products (ETnoD) than ECD. [61] Therefore provide the range of fragments ions better signal. Allowing for a greater total sequence coverage.

Attainment of 63% sequence coverage of an isotopically depleted 47 kDa protein is an impressive achievement. However, new technologies aimed at improving current capabilities of top-down analysis have also been developed to maximise the sequence coverage.

5.4 PTCR of SPT fragmentation spectra

One method is to combine the sequence coverage obtained from multiple fragmentation spectra, using the same or multiple fragmentation techniques [79,122,126] that can have a combined analysis. Another developed method to extract as much information as possible is using PTCR, (section 1.2.3, page 32) [74,75] . This technique uses a radical anion to accept a proton from the multiply charge ion resulting in charge reduction ($[M+H]^{n+}$ to $[M+(n-1)H]^{(n-1)+}$) shifting ions to higher *m/z* values. When applied stepwise to multiple isolated windows containing regions of the complex MS^2 ETD fragments of Figure 5.3 can help simplify the fragmentation spectra and increase sequence coverage. There is a disadvantage to this style of analysis, in that it requires a large quantity of protein sample, a mass spectrometer which can facilitate the use of multiple fragmentation techniques and is not readily applicable for on-line analysis. However, for complicated samples which are not time-limited or sample-limited it can help improve top-down analysis.

5.4.1 PTCR analysis of isotopically natural SPT

PTCR can be applied to the top-down fragmentation of SPT, in stepwise sections which are isolated and mixed with the radical anion reagent to cause proton transfer. This is shown in Figure 5.5, as the multiple coloured brackets display separate regions of dense fragment ions which were sequentially isolated. The separate coloured brackets, each cover swathes of the most densely populated regions of the MS^2 spectrum. For different protein fragmentation spectra other regions can also be selected for further analysis, As for isotopically natural SPT, the regions between the

charge reduced peaks and the higher m/z range (1450-2000 m/z) were not as densely populated as the lower m/z region (600-1260 m/z), so PTCR was not applied. For protein of varying molecular weight may require application of PTCR within these regions as required.

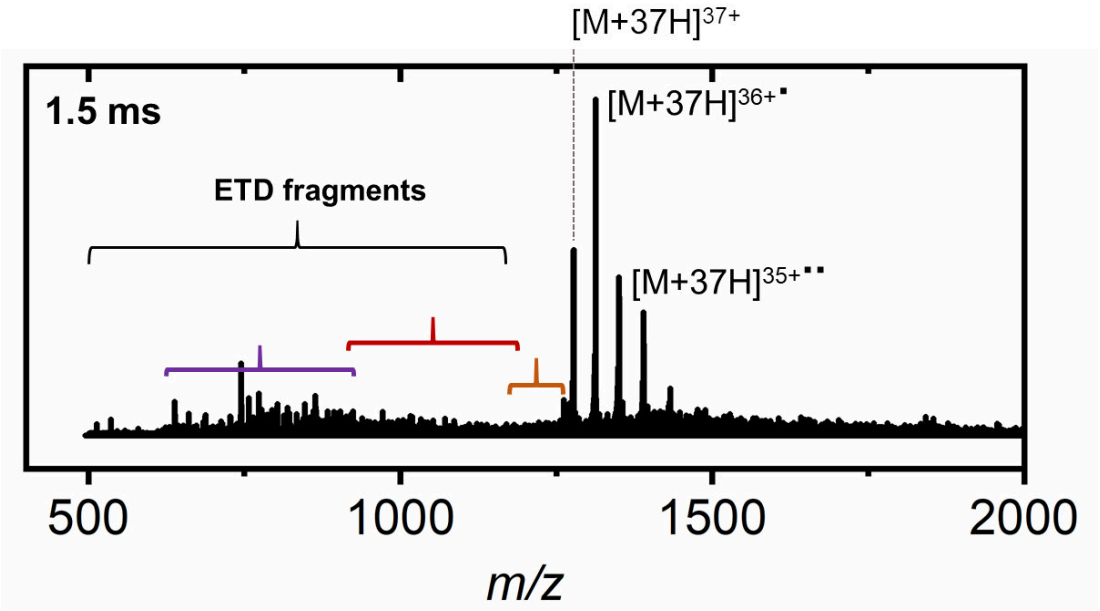


Figure 5.5: ETD fragmentation spectra of the [M+37H]37+ charge state of isotopically natural SPT

Isotopically natural SPT undergoing ETD fragmentation for 1.5 ms results in a complicated spectra of overlapping fragment ion distributions. Particularly in the region of 700-1300 m/z. To apply the MS3 PTCR reaction first requires the isolation of regions within the dense overlapping ions. Highlighted by different coloured brackets.

The length of time for which the ion-ion PTCR reaction occurs is also important to consider. As a single, multiply charged fragment ion can react with a succession of radical anions, an appropriate reaction time for the isolated fragment ions must be experimentally determined.

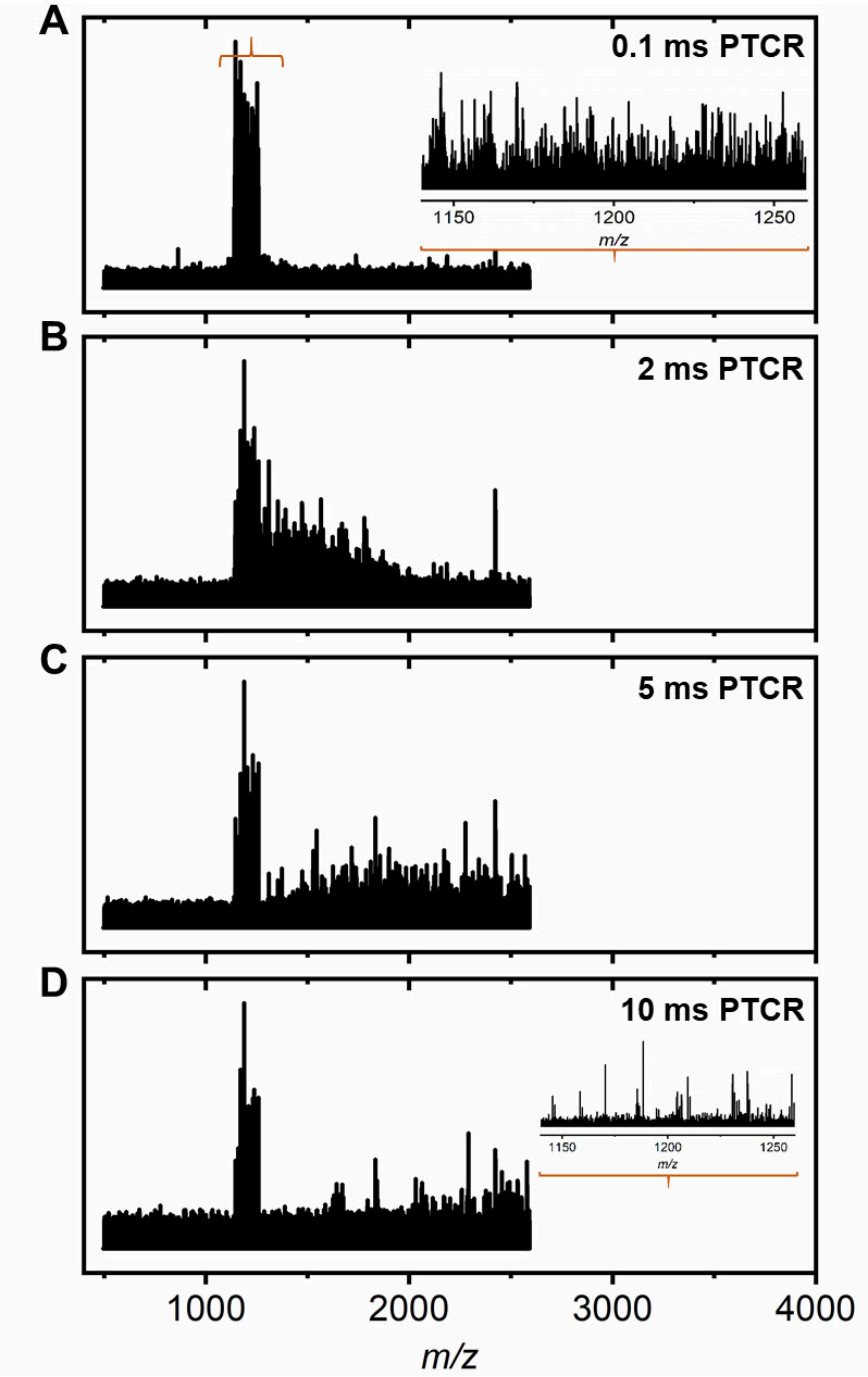


Figure 5.6: PTCR of isolated fragment ions centred at 1200 m/z from MS2 ETD fragmentation spectra

Produced ETD fragments of isotopically natural SPT were further analysed using PTCR. A selected region of fragments ions appearing at 1140-1260 m/z were isolated and subjected to proton transfer for 0.1 ms (A), 2 ms (B), 5 ms (C) and 10 ms (D). Proton transfer causes the fragment ions to appear at a lower charge state, so fragments appear within a higher m/z range.

This demonstrated in Figure 5.6, showing the isolated window of 1140-1260 *m/z* of the MS^2 ETD spectrum with different applied PTCR reaction time. In Figure 5.6a, the PTCR reaction time is 0.1 ms to act as a control for the isolated fragment ions as little to no proton transfer can occur in the small time-frame. The isolated regions of fragment ions are shown in greater detail in the inset spectra, clearly displaying the spectral complexity of the initial isolated region. The reaction time was increased from 0.1 ms to 2 ms (Figure 5.6b), 5 ms (Figure 5.6c) or 10 ms (Figure 5.6d) to observe increasing proton transfer. At 2 ms PTCR (Figure 5.6b) we can observe the rightward shift of the fragment ions away from the isolated region. But at this reaction time there is still a large region of ion overlap and does not utilise the full range of the recorded mass spectrum. Increasing to 5 ms reaction time improves this, as the ions achieve a greater distribution across the spectrum and don't have regions which are excessively populated by ions. With a further increase in reaction time to 10 ms many of the fragment ions have lost a sufficient number of protons that they achieve a greater *m/z* than that within the recorded range and so do not appear on the mass spectrum. The spectral insert also shows the original isolated range of 1140-1260 *m/z* and the reduced number of fragment ions which have not undergone significant proton transfer and can still be observed within this region. Conveniently, these ions are more likely to be assigned due to the migration of the other fragment ions away from this isolated region.

Within the range of 1140-1260 *m/z* within the ETD MS^2 spectra it was only possible to assign 25 fragment ions (14 *c*-ions, 11 $z^{\cdot}$ -ions). With PTCR applied for 5 ms (Figure 5.6c) the total number of fragment ions assigned increase to 55 fragment ions (39 *c*-ions; 16 $z^{\cdot}$ -ions), with the further increase in PTCR reaction time to 10 ms the number of assigned fragments decreases again to 24 fragment ions (14 *c*-ions; 10 $z^{\bullet}$ -ions).

As this increases the number of new fragment ions if the peak lists are combined it ultimately causes an increase in the achievable sequence coverage.

In the ETD MS^2 fragmentation spectra, only 31% of the total sequence coverage can be assigned. Addition of the assigned MS^3 ions from the regions outlined by brackets in Figure 5.5 allows the assignment of ~600 fragment ions. Many of which will be redundant ions previously identified in the MS^2 spectrum. Some however will allow characterisation of new fragmentation events. The result of which is shown in Figure 5.7b as the superposition of the fragmentation maps from the MS^2 and several MS^3 SPT spectra. It is still mainly populated by the red fragments, meaning that a large proportion of the 600 identified fragment ions are redundant ion assignments. In that they were previously assigned in the MS^2 spectra. Interestingly, the contribution from the PTCR spectra can be observed at the C-terminus which had very few observed fragment ions in the MS^2 spectra. In the combined fragmentation map (Figure 5.7b) are displayed in a range of colours from separate PTCR acquisitions indicating that every spectrum contributes to the total combined sequence coverage. Increasing it to 45% sequence coverage. The increase observed with PTCR could also likely be further increased by reducing the *m/z* ranges over which PTCR is applied. Generating a larger number of spectra to attempt to gain as much detail from the fragmentation spectra as possible.

The increase from 31% to 45% sequence coverage of isotopically natural SPT highlights the positive impact PTCR can have on top-down analysis. When used in conjecture with isotope depletion it can potentially facilitate significant sequence coverage of larger molecular weight protein like SPT.

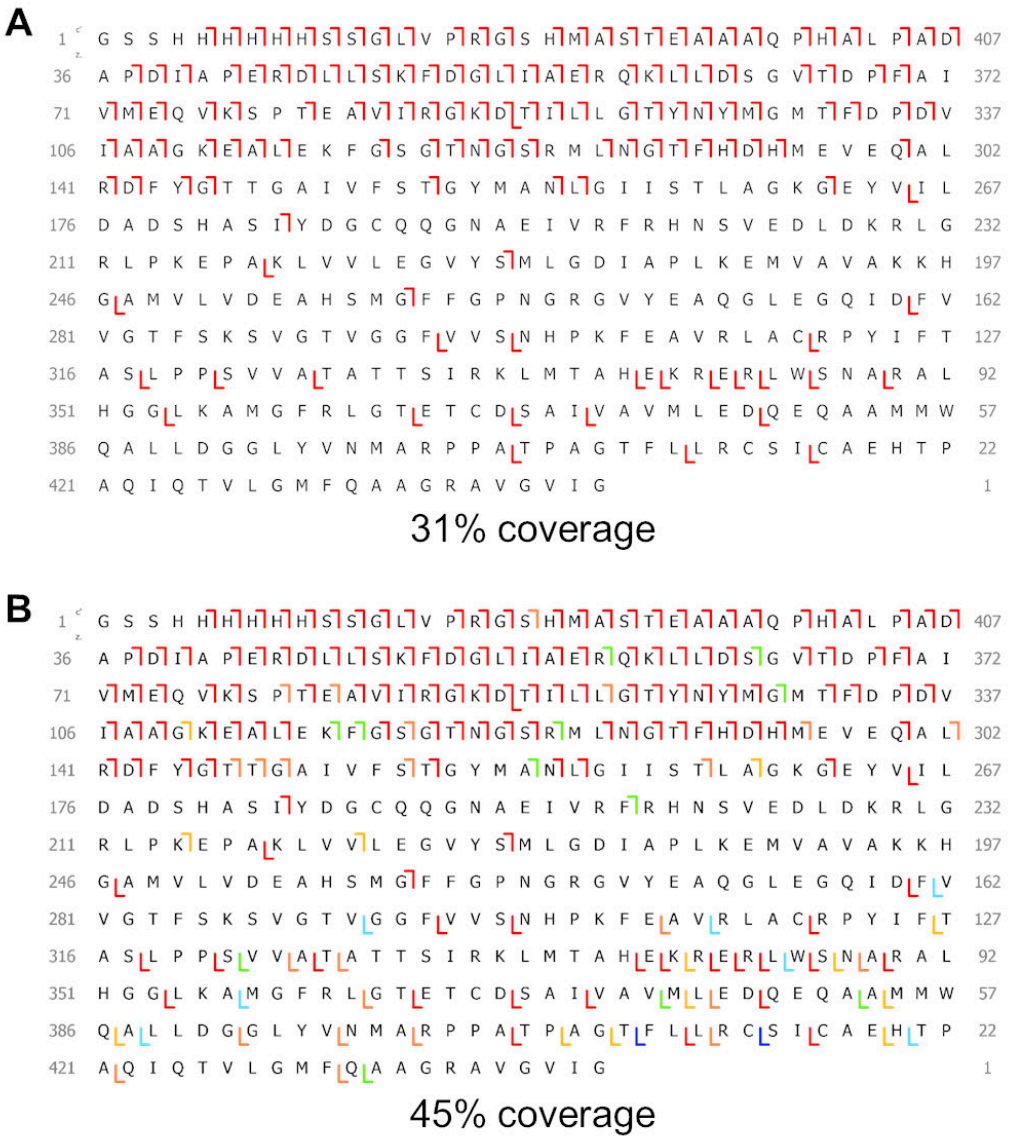


Figure 5.7: Sequence coverage increase for isotopically natural SPT after ETD fragmentation with PTCR applied.

Sequence coverage of isotopically natural SPT after 1.5 ms of ETD fragmentation (A, 31% coverage; red fragments). The combined sequence coverage from the PTCR assigned fragments (B, 45% coverage), 300 m/z isolation centred at 800 m/z, 5 ms (green fragments), 300 m/z isolation centred at 1050 m/z, 5 ms (orange fragments) and 10 ms (light blue fragments). A 120 m/z isolation centred at 1200 m/z, 5 ms (yellow fragments) and 10 ms (dark blue fragments). The direction of the line indicates if the fragment is a N-terminal ion (c-type ion, ┐) or a C-terminal ion (z· -type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

5.4.2 PTCR analysis of isotopically depleted SPT

An identical analysis on the impact of PTCR can be applied to isotopically depleted SPT protein. Much like in the isotopically natural SPT MS^2 spectra, regions of overlapping can be isolated to be subjected to PTCR.

This is demonstrated in Figure 5.8, in which 1140-1260 *m/z* of the ETD fragments ions were isolated and PTCR applied for a range of ion-ion reaction times. Again 0.1 ms demonstrates the spectra with effectively no proton transfer *(*Figure 5.8a), and the spectral insert demonstrates the density of the isolated region. PTCR was applied for 2 ms (Figure 5.8b), 5 ms (Figure 5.8c) and 10 ms (Figure 5.8d) and like for the isotopically natural SPT creates a rightward shift of the overlapping fragments which appear at greater *m/z* values. At 10 ms, many fragment ions appeared at a greater *m/z* value, therefore extending the recorded *m/z* value was required.

Isotopically depleted SPT displays an identical trend to the isotopically natural protein with the application of PTCR at varying reaction times (Figure 5.8).

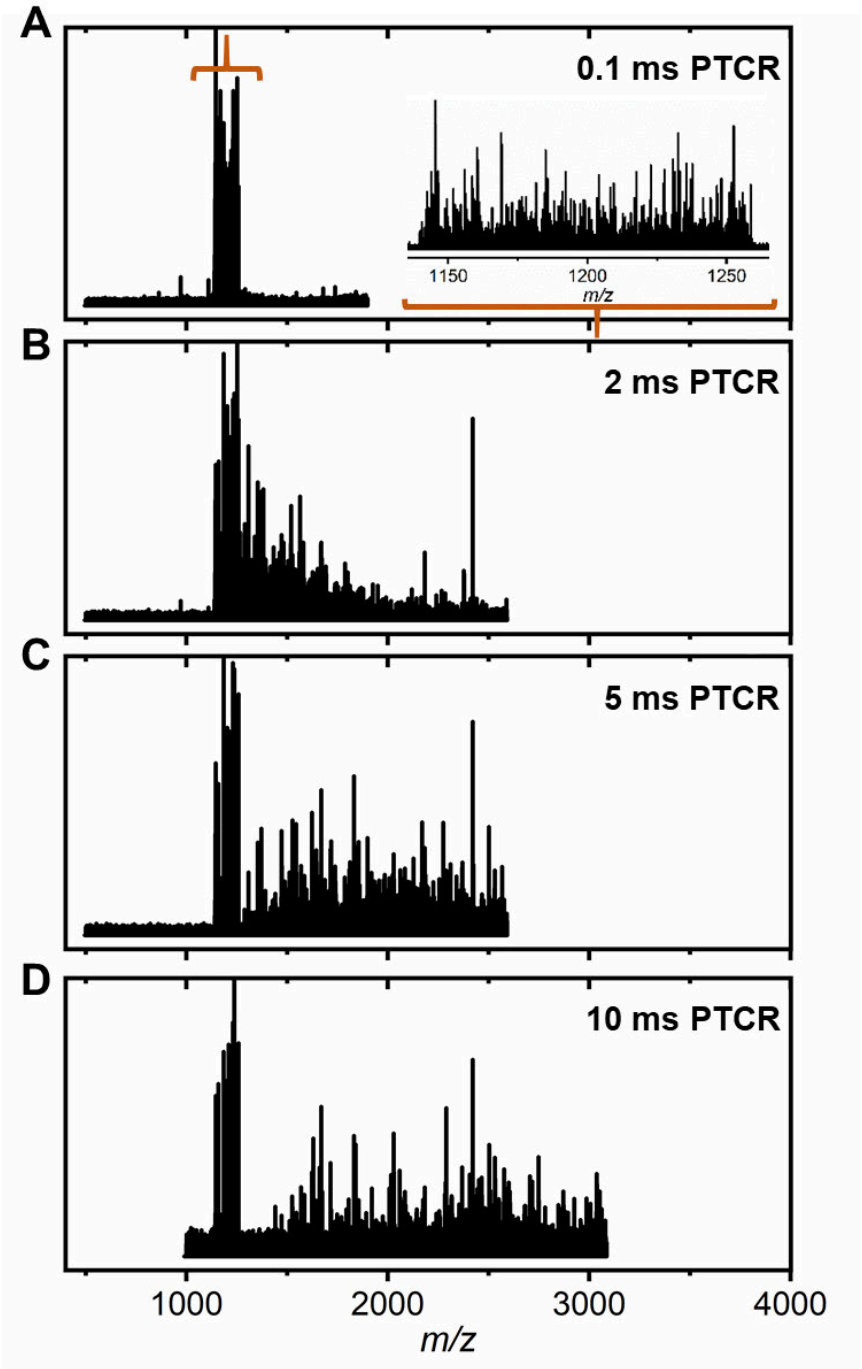


Figure 5.8: PTCR of isolated fragment ions centred at 1200 m/z from MS^2 ETD fragmentation spectra

Produced ETD fragments of isotopically depleted SPT were further analysed using PTCR. A selected region of fragments ions appearing at 1140-1260 m/z were isolated and subjected to proton transfer for 0.1 ms (A), 2 ms (B), 5 ms (C) and 10 ms (D). Proton transfer causes the fragment ions to appear at a lower charge state, so fragments appear within a higher m/z range.

In the MS^2 fragmentation map (Figure 4.4b), a total of 63% sequence coverage with isotopically depleted SPT was achieved. Within the isolated region of Figure 5.8 a total of 111 fragment ions (56 *c*-ions; 55 $z^{\cdot}$-ions) were assigned, over 4X greater than the 25 ions in the isotopically natural MS^2 spectra. Similarly with the application of PTCR to the isolated region, the assignable ions increases to 357 ions (274 *c*-ions; 83 $z^{\cdot}$-ions) in the 5 ms reaction time spectra (Figure 5.8c) and 156 ions (112 *c*-ions; 44 $z^{\cdot}$-ions) when increased to a 10 ms reaction time. The application of PTCR to the same highlighted regions outlined by brackets in Figure 5.5 contribute an extra 2470 assigned fragment ions, again roughly 4X greater than that of the isotopically natural protein.

An example of the fragments which can be assigned in the spectrum of Figure 5.8c are shown in Figure 5.9. The assigned ions identified in the isotopically depleted SPT PTCR spectrum (Figure 5.9b) are labelled and the monoisotopic peak identified by the dashed line. These same fragment ions could not be assigned in the natural SPT 5 ms PTCR spectrum (Figure 5.6c and Figure 5.9a). The c_{246}^{11+} fragment ion (26474.29 Da) assigned on the isotopically depleted spectrum is a particularly large fragment ion which comprises approximately half of the SPT protein. As a result of isotopic depletion, the monoisotopic peak of the fragment ion is present which facilitates our assignment on the accurate monoisotopic mass. This is particularly impressive as the fragment ion has achieved and maintained a sufficient abundance throughout the ETD fragmentation, quadrupole isolation and proton charge transfer to still display a coherent isotopologue distribution.

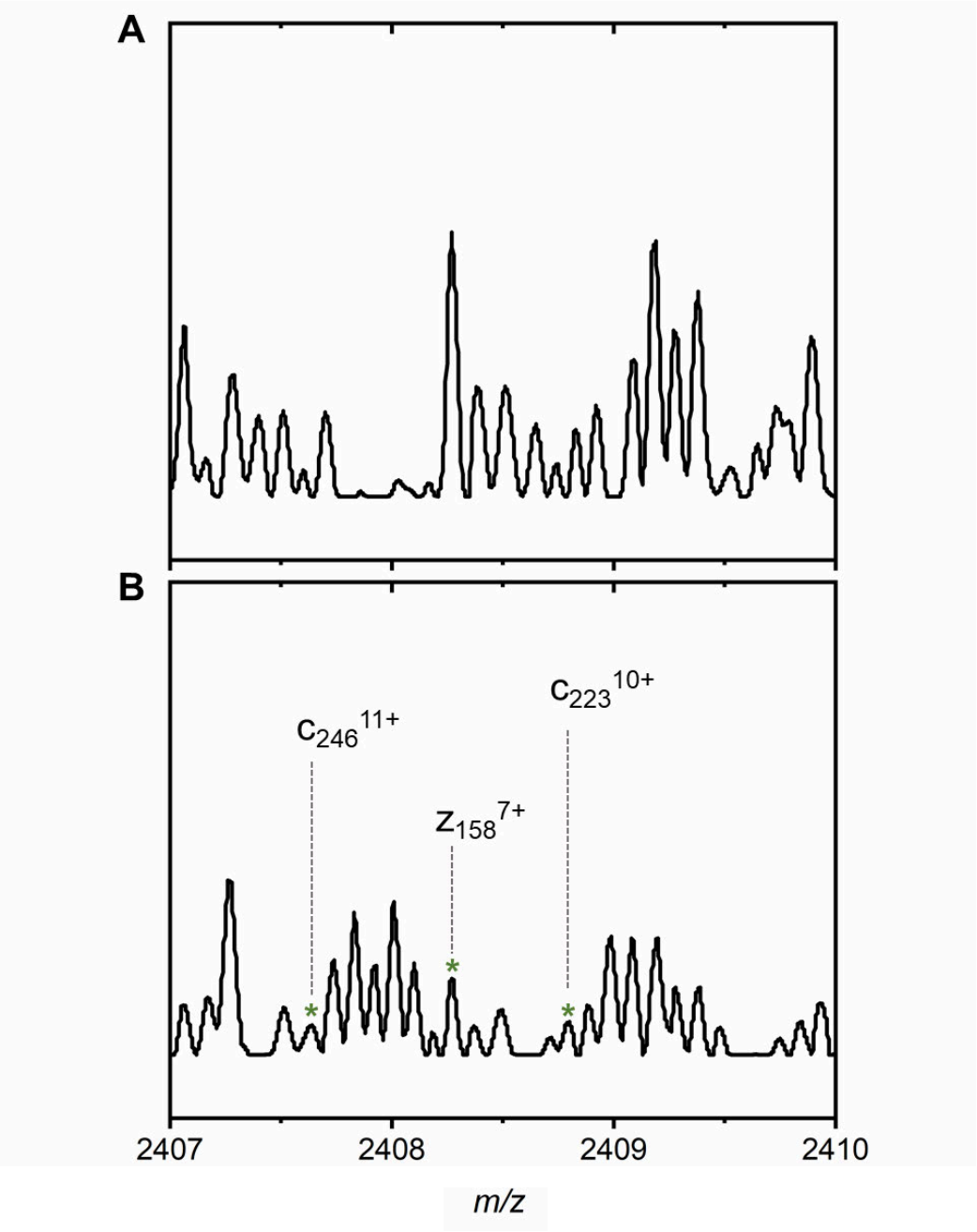


Figure 5.9: MS^3 fragments SPT after 5 ms PTCR

Comparison of the assigned ions produced after MS^3 5 ms of PTCR of isolated MS^2 fragments centred around 1200 m/z of isotopically natural (A) and isotopically depleted (B) SPT. The dashed line indicates the monoisotopic peak and the fragment ion assignment above.

The multiple PTCR MS^3 sequence coverage maps can be combined with that of the ETD MS^2 sequence coverage maps, shown in Figure 5.10. The combined PTCR sequence coverage maps of Figure 5.10b, again show an increase in sequence coverage in comparison to the isotopically natural coverage maps in Figure 5.8. The isotopically natural fragmentation increased in total sequence coverage from 31% to 45%. Likewise, the isotopically depleted SPT also displays an increase from 63% in the MS^2 to 78% with the combined MS^2 and MS^3 assigned fragments.

As discussed for Figure 5.7, many of the PTCR assigned fragment ions are redundant, as they cover fragments which were already assigned in the MS^2 spectrum. However, in the combined isotopically depleted fragmentation map Figure 5.10b there are examples of the PTCR fragment ions increasing the number of complementary fragment ion pairs within the densely assigned N-terminus. In a manner similar to that observed in the fragmentation of the isotopically natural protein, the greatest impact of combining the PTCR fragment ions is observed at the C-terminus, very little of which is assignable in the MS^2 spectrum. This is indicative of limited fragmentation due to a lack of positive residues, [127] a low abundance fragment ion which will become lost within the complex overlapping fragments of the MS^2 spectrum, as demonstrated in the insert of Figure 5.8a. Using PTCR, the low abundance fragment ions can emerge and be assigned contributing to the overall sequence coverage demonstrated in Figure 5.9b, compared to the same region of the isotopically natural SPT spectra (Figure 5.9a). Therefore, even the analysis of isotopically depleted protein can be improved with a combination of top-down techniques.

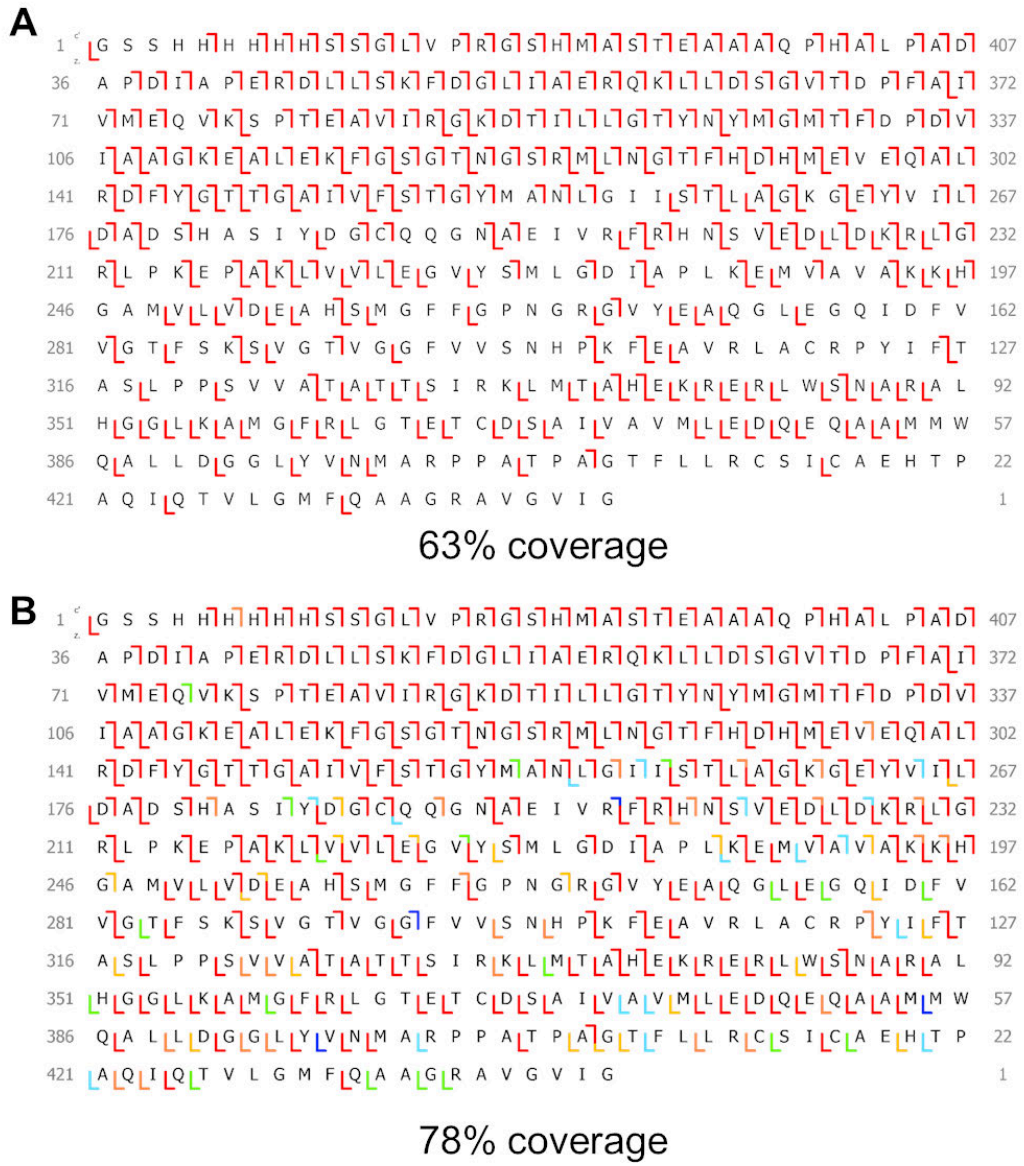


Figure 5.10: Sequence coverage increase forisotopically depleted SPT after ETD fragmentation with PTCR applied.

Sequence coverage of isotopically depleted SPT after 1.5 ms ETD fragmentation applied (A, 63% coverage, red fragments). The combined sequence coverage from the PTCR assigned fragments (B, 78% coverage), 300 m/z isolation centred at 800 m/z, 5 ms (green fragments), 300 m/z isolation centred at 1050 m/z, 5 ms (orange fragments) and 10 ms (light blue fragments). A 120 m/z isolation centred at 1200 m/z, 5 ms (yellow fragments) and 10 ms (dark blue fragments). The direction of the line indicates if the fragment is a N-terminal ion (c-type ion, ┐) or a C-terminal ion (z˙-type ion, └), and the total sequence coverage is represented as a percentage of the total number of bonds which can undergo fragmentation.

One of the benefits of top-down analysis regardless of the fragmentation technique used, is that the produced fragment ions can be directly related to a known amino acid sequence which was purposely overexpressed. While this is an important avenue for study, much of protein MS analysis takes place on samples on which the present protein is not known. To carry out identification and assignment of an unknown sample, often a probability-based 'scoring' algorithm is used. Algorithms like Mascot, Andromeda and prosight [81,141,142] and many others match peptide sequence from databases to the spectra from MS^2 and MS^3 spectra deconvolute and assign the present protein.

The software of these algorithms, as previously discussed, do not readily accept isotopically depleted spectra; in automated software, multiple isotopologue peaks are essential for the fragment ion identification. As the isotope depletion strategy becomes more widely adopted updates to the software will allow deconvolution of the isotopically depleted spectra. Once the software gets updated the simplified isotopic distribution, increased monoisotopic peak and increased signal to noise will facilitate an increased number of fragment ions which can be assigned in the isotopically depleted spectra. Likely present with greater signal than that in the isotopically natural spectra, as has been previously shown in this thesis. It would be expected that as with the comparative analysis performed above the change to the isotopologue distribution with the isotope depletion strategy that the software will identify and assign a greater number of fragment ions, thereby producing greater number of sequence tags. [90,91] It can be expected that this will allow the identification of proteins with a greater 'confidence score' than isotopically natural protein samples and potentially allow the identification of more fragment ions.

5.5 Conclusions

The data in this chapter has shown that the application of top-down analysis of isotopically depleted protein is analogous across the currently available MS platforms, namely high-resolution 12T FT-ICR and orbitrap Eclipse. It has also demonstrated that comprehensive top-down sequence coverage is feasible for larger MW protein. Applying ETD fragmentation to SPT on the orbitrap Eclipse permitted assignment of significantly greater sequence coverage than that previously achieved on the FT-ICR. The fundamental change in the isotopically depleted distribution of larger MW continues to result in simplified and improved analysis of protein via top-down analysis. High-resolution analysers allow distinction of separate isotopologue peaks, and the identification of the monoisotopic peak is essential for accurate assignment of fragment ions. Particularly when the distributions cannot be accurately assigned in the isotopically natural protein spectra.

While isotope depletion is shown as a powerful tool for larger protein top-down analysis. It can achieve even greater protein sequence coverage in conjunction with instrumental improvements and the other supplemental fragmentation techniques which are being developed. The upper mass limit for reasonable sequence coverage needs to be experimentally determined. Therefore, increases in the MW of the protein analyte is required to continue to develop isotope depletion of ^{13}C and ^{15}N as a tool for improved top-down analysis. Further isotope depletion, instrumental improvements and the application of further processing, like PTCR, will be ultimately continuing to extend this limit further in the future.

6. Isotope Depletion for Improved Native Protein Mass Spectrometry.

6.1 Native mass spectrometry

Native mass spectrometry is a technique in which the electrospray process takes place from a solution mimicking physiological conditions (e.g. aqueous, neutral pH) in an effort to maintain the non-covalent interactions and protein higher order structure during ionisation and transmission in the gaseous phase (section 1.1.2, Figure 1.2b).[143] It has developed into a versatile structural biology technique which is used to investigate protein conformation, protein complex assembly, protein-protein interactions, and protein-ligand binding. [144–147] Native mass spectrometry is often coupled to other MS techniques such as ion mobility MS, collisional activation, HDX-MS or fragmentation analysis to further analyse the gas phase 'folded' protein. As such, it has become an important orthogonal technique to traditionally used structural biology techniques. Such as protein crystallography, NMR, dynamic light scattering and electron microscopy.[143,148–150]

Native MS commonly depends on recombinant expression of large quantities of protein. Making the use of isotope depletion expression media, is therefore would require only a small change to the typical native MS workflow to adopt isotope depletion to these native MS techniques. The change in native protein spectra and the potential improvements for analysis as an initial proof of principle are discussed in this chapter.

6.2 Native Mass Spectrometry analysis of isotopically depleted BCA

Carbonic anhydrase is a well characterised model protein (29 kDa) for many MS analysis and other characterisation focused techniques, due to its known stability and well understood catalytic mechanism. [52,151–153] It has a defined catalytic pocket, which uses a triad of histidine residues to coordinate a zinc ion, required for the protein to be catalytically active. [154] To produce isotopically depleted protein, BCA was recombinantly expressed in M9 minimal media as described in chapter 3. BCA protein

purified from isotopically natural and isotopically depleted M9 minimal media were analysed under native conditions from 20 mM ammonium acetate, using nESI on a 12T FT-ICR instrument.

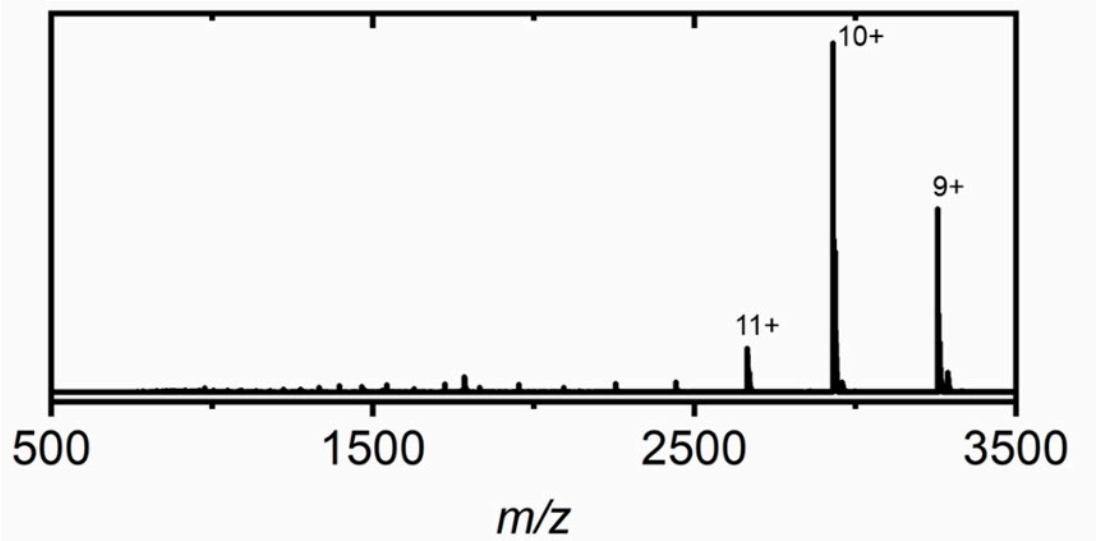


Figure 6.1: Native spectrum of isotopically depleted BCA

Charge state distribution of native isotopically depleted BCA, displaying the 9^+, 10^+ and 11^+ charge states.

The observed charge state distribution is the most striking difference between a natively electrospray protein and an unfolded electrospray protein. In comparison to the spectrum obtained for unfolded BCA (Figure 3.7), which displays the 20+ to 50+ charge states spread across the 500-1500 *m/z* region, the charge state distribution in the case of native is predominantly observed as the 9+ to 11+ charge states across the 2500-3500 *m/z* region (Figure 6.1). This is a consequence of the retained higher order protein structure, which is maintained during the native electrospray process, leading to fewer available surface residues to carry charge. As the folded protein consistently produces a predictable secondary structure so has less variation in the number of acquired charges. As with previous analysis, isotope depletion results in a

simplification of the isotopologue distribution. The 10^{+} charge state of the native charge state distribution in Figure 6.1, is shown in greater detail in Figure 6.2.

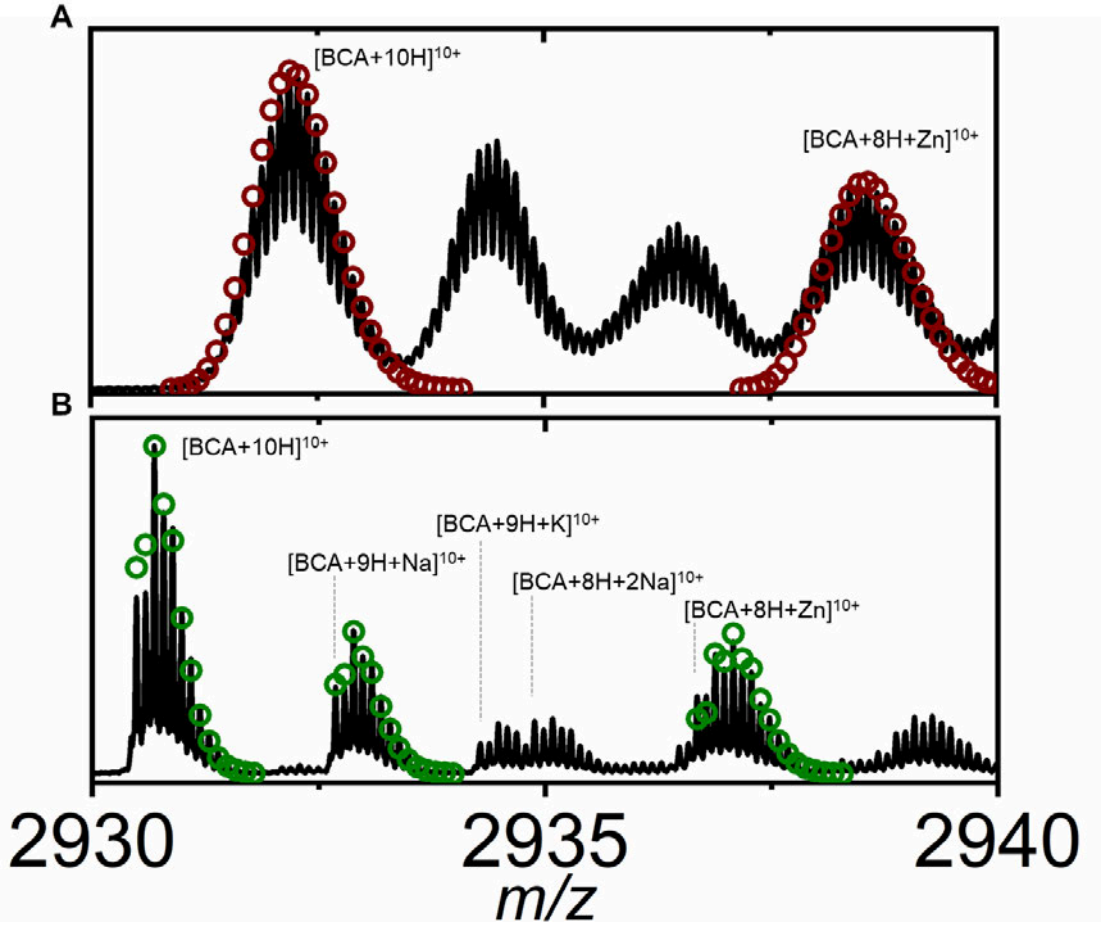


Figure 6.2: Native spectra of the 10+ charge state of BCA

Native mass spectra of isotopically natural (A) and isotopically depleted (B) BCA. Recombinant expression of BCA in M9 minimal media does not provide sufficient concentration of zinc for all of the expressed BCA to be zinc bound. Salt adducts, a common detrimental feature of native protein mass spectrometry are labelled and the theoretical distribution is overlaid (O).

Figure 6.2 displays the difference in isotopologue distribution of the $[BCA+10H]^{10+}$ and a range of associated salt adducts as a result of isotope depletion (Figure 6.2b). As seen in previous chapters, isotope depletion results in simplified distributions. At 29 kDa, isotopically depleted BCA also has a significant abundance of the monoisotopic peak. Facilitating the accurate assignment of BCA plus many of the observed cationic

salt adducts are observed ($[M+9H+Na]^{10+}$, $[M+9H+K]^{10+}$ etc), a phenomenon commonly encountered during native MS experiments. For the 10^+ charge state of isotopically natural BCA (Figure 6.2a), the different salt adduct peaks overlap to a degree that they are difficult to assign. The increased variety assigned in the isotopically depleted spectrum is likely hidden in the isotopically natural spectrum due to the high level of isotopologue distribution overlap. Interesting for both spectra, the most abundant distribution is the protein in the apo-form, without a zinc (Zn^{2+}) ion bound. This is likely as a result of the lack of available free zinc within the culture expression in the M9 minimal media. As the protein is largely over-expressed, there

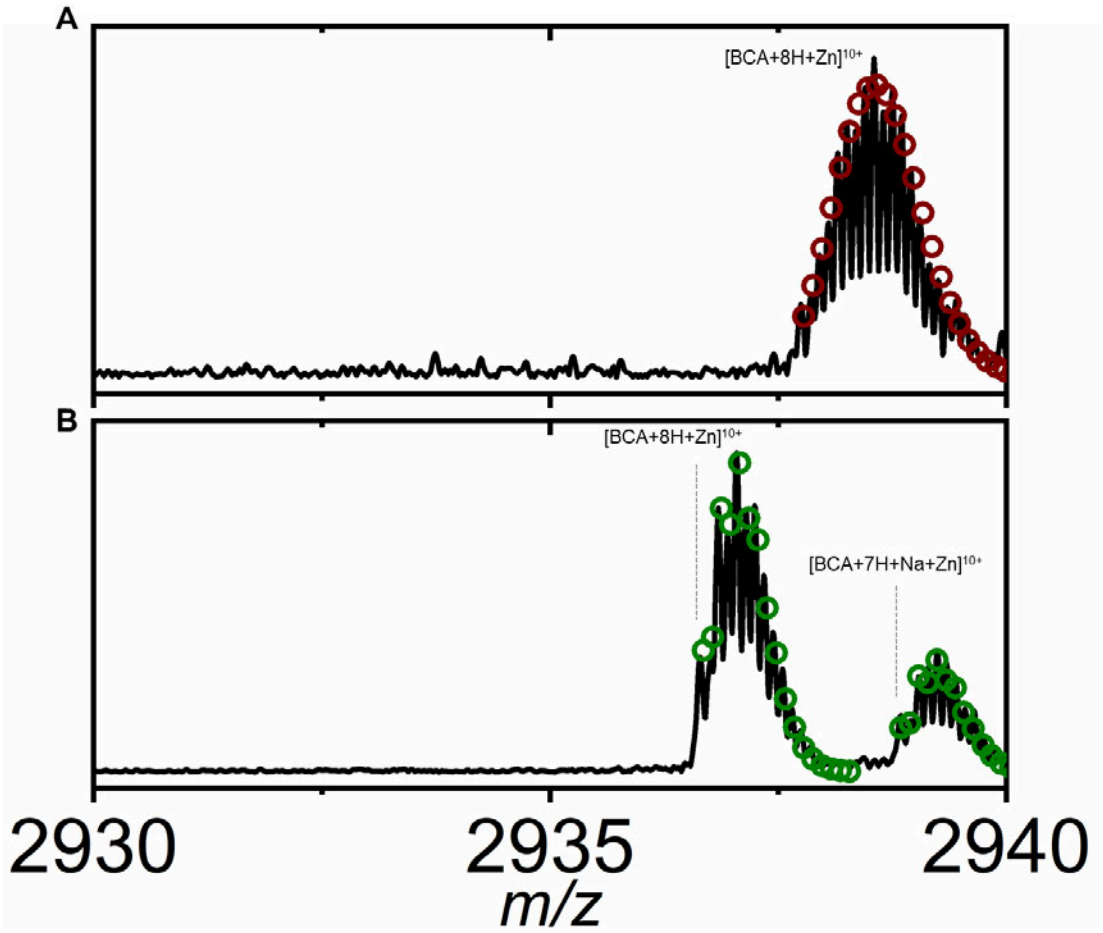


Figure 6.3: Native spectrum of the 10^+ charge state of BCA with zinc loaded

Native MS of isotopically natural (A) and isotopically depleted (B) with the addition of zinc to the electrospray buffer which is sequestered by the present BCA protein. Theoretical model for each protein distribution is overlaid (O).

is insufficient zinc for all the protein. The process can therefore be repeated with zinc supplemented into the ammonium acetate electrospray buffer. This is shown in Figure 6.3, in which for both protein spectra the $[BCA+8H+Zn]^{10+}$ is the only isotopologue distribution present within the 2930-2940 *m/z* range.

The isotopologue distribution in Figure 6.3 is not as clearly formed as that in Figure 6.2, however the depleted $[BCA+8H+Zn]^{10+}$ isotopologue distribution from both Figure 6.3b and Figure 6.2b both display a varied isotopologue distribution when compared to the isotopically depleted $[BCA+10H]^{10+}$ distribution in Figure 6.2b. Therefore, the change in isotopologue distribution is a result of the zinc ion bound to the protein.

6.3 Isotope depletion and the impact of metal binding

Metalloprotein like BCA are common, the metal ions are essential for function or structure [147,155] and the variety of metal ions can modulate a diverse action of enzymes. So are key regulators of cellular biology. [156] Being able to characterise and identify which metals are binding to a protein is an important step in native analysis. [108,147,157]

The depleted isotopologue distribution $[BCA+8H+Zn]^{10+}$ (Figure 6.2b and Figure 6.3b) displays an altered isotopologue distribution to the protein without an associated zinc ion, $[BCA+10H]^{10+}$. As this is a result of the isotopic contribution of the zinc ion, it maybe be an unexpected benefit of the isotope depletion strategy that facilitate metal ion binding assignment based on the asymmetric distribution of protein plus the addition of metal ions. This was investigated computationally, using predictive software to produce the theoretical isotopologue distribution of isotopically depleted BCA with the addition of a transition metal, this is shown in Figure 6.4.

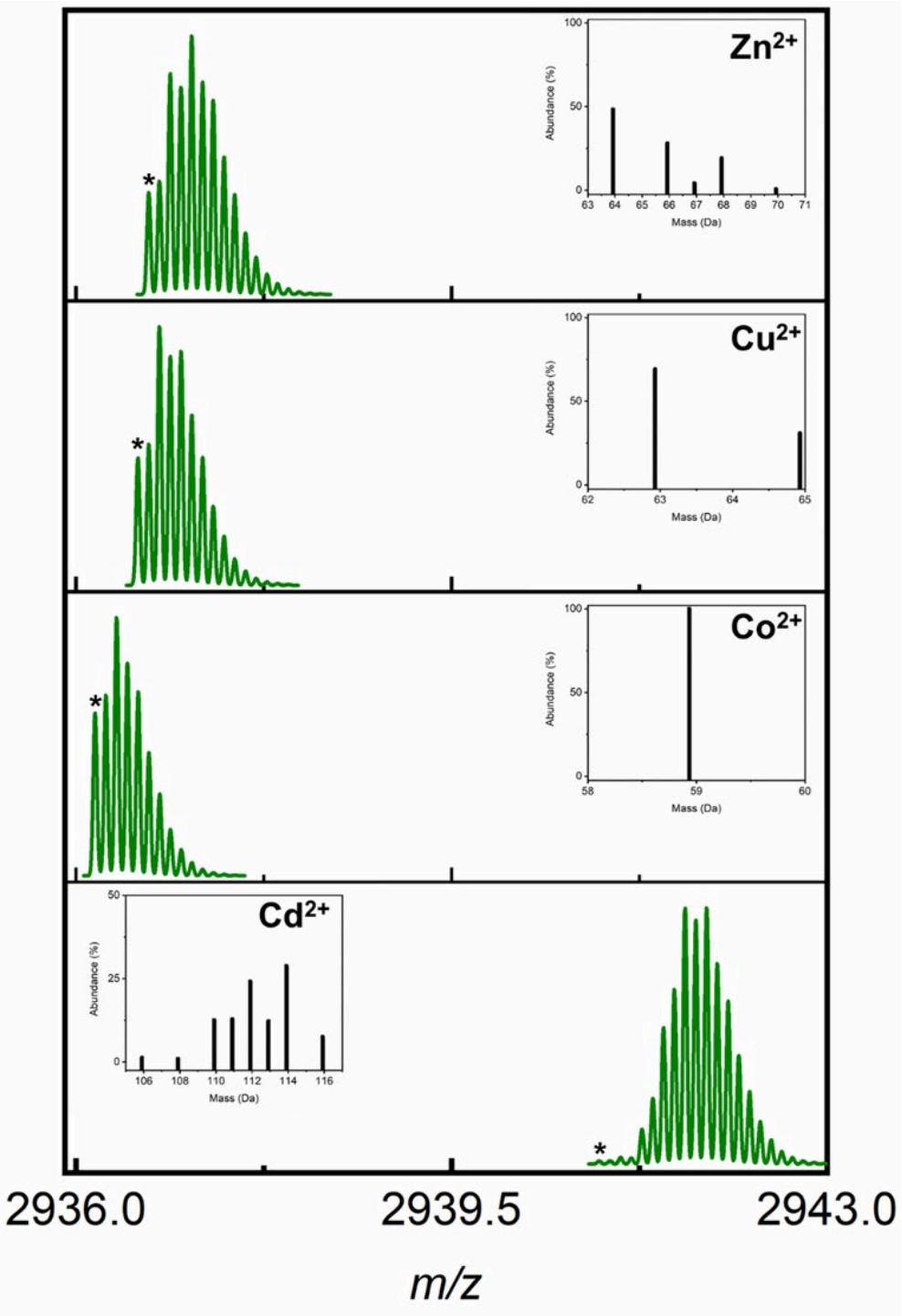


Figure 6.4: Isotopologue profile of BCA (10^{+}) with a transition metal ligand

Theoretical models of isotopically depleted BCA, with a bound metal co-factor. The isotopologue distribution of protein plus zinc (Zn^{2+}), copper (Cu^{2+}), cobalt (Co^{2+}) and cadmium (Cd^{2+}) were all investigated. The isotope distributions for each metal ion is included as a spectral insert. The monoisotopic peak (M) of each distribution is highlighted with an asterisk (*)

The theoretical model of isotopically depleted BCA with a zinc ion (Figure 6.4), is largely similar to the acquired spectral isotopologue distributions of $[BCA+8H+Zn]^{10+}$ in Figure 6.2b and Figure 6.3b. Particularly the decrease in abundance of the M+3 peak, i.e. the fourth isotopologue peak, making the M+2 and M+4 the two most abundant isotopologue peaks in the distribution. Copper (Cu^{2+}) and cobalt (Co^{2+}) have limited isotopic variation containing two and one isotopologue peak respectively. Cobalt therefore will have no impact on the isotopologue distribution, only causing an increase in mass of 58.9332 Da. Copper, due to its second isotope with have an impact on the protein isotopologue on the M+2 peak and beyond, and so will produce a protein isotopologue distribution different to that produced with the addition of a zinc ion. The theoretical addition of a cadmium (Cd^{2+}) ion produces a very different protein isotopologue distribution. Firstly, it adds a greater mass than the other metals, resulting in a greater mass shift on the mass spectrum. Secondly, the most abundant isotope peaks of the metal ion are not also the lightest isotopes peaks as is the case for all the elements considered previously in this study. It will therefore have the greatest impact on the M+4 isotopologue peak and beyond. Interestingly, the theoretical distribution in Figure 6.4 resembles a Gaussian distribution of an isotopically natural distribution.

While the isotopologue distributions of Figure 6.4 are produced computationally, comparison of the theoretical zinc-bound BCA closely resembles that of the acquired spectra of isotopically depleted BCA protein bound to zinc (Figure 6.2b and Figure 6.3b). Suggesting that the other theoretical models will closely match the acquired spectra if acquired.

As the histidine triad within the catalytic pocket of BCA can complex can complex a range of divalent cations [158] the resultant change in the isotopologue distribution was further assessed by loading copper ions (Cu^{2+}) into the active site of BCA. This is shown in Figure 6.5.

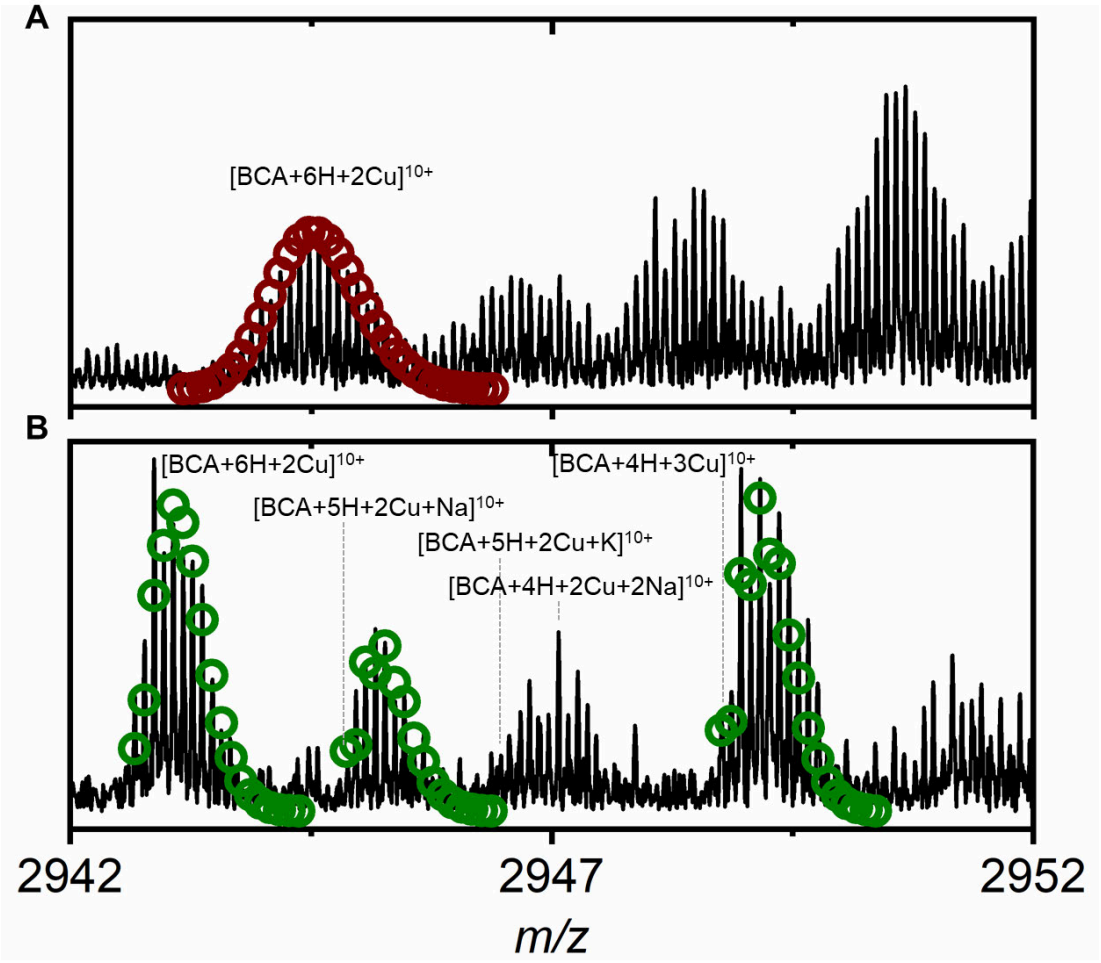


Figure 6.5: Native spectrum of copper loaded BCA

Native MS of isotopically natural (A) and isotopically depleted (B) with the addition of copper to the electrospray buffer which is sequestered by the present BCA protein.

Unfortunately two copper ions (Cu^{2+}) have been demonstrated to interact with folded BCA protein within the catalytic pocket, as well as at a secondary site [159]. This is what was observed within the acquired spectra of Figure 6.5. Which is an added level of complexity to the analysis of the protein-metal bound isotopologue distribution in comparison to the theoretical distributions in Figure 6.4. The isotopologue distribution

will be further altered by the addition of two Cu^{2+} ions. This can also be modelled for comparison, shown in Figure 6.6.

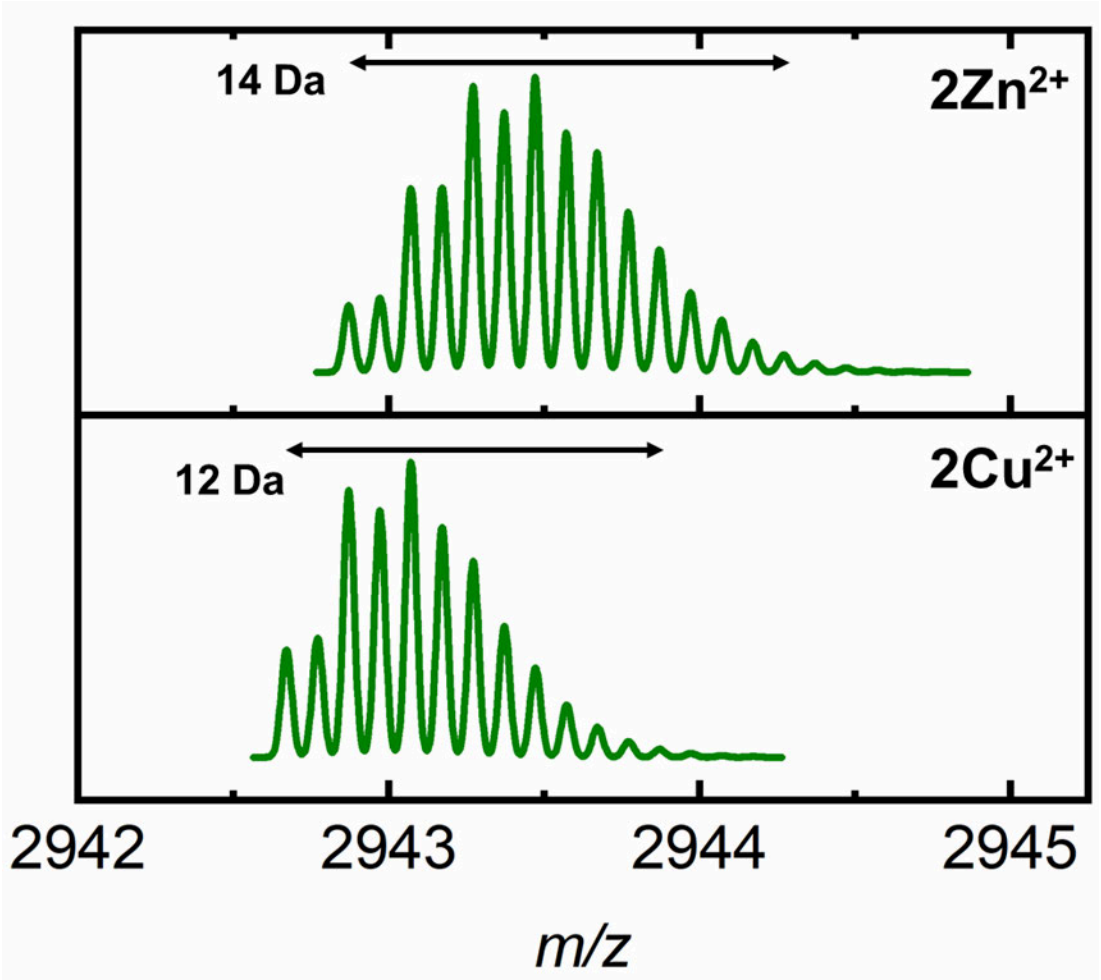


Figure 6.6: Theoretical distribution of BCA (10+) with two bound divalent ions.

Simulated distribution of isotopically depleted BCA with the added mass of two divalent metal cations. The addition of protein plus two zinc (Zn^{2+}) and two copper (Cu^{2+}) ions.

This displays the similarity of the isotopologue distribution with the addition of two zinc (Zn^{2+}) or two copper (Cu^{2+}) ions. Therefore, accurate assignment of the monoisotopic peak is crucial for the assignment of the protein-metal ion. The most abundant isotope of copper (62.9296 Da) has a difference in mass of only 0.9995 Da to zinc (63.9291 Da). However, another key difference is the width of the isotopologue distribution, as

the addition of zinc ions cause a wider isotopologue distribution than the addition of copper. This ultimately means that the average mass of the protein-metal also plays a role in the final assignment of the protein-metal ion. The addition of two zinc ions results in a mass increase of 1.9991 Da greater than that of two copper ions. The average mass of BCA bound to two zinc ions however will be 3.7254 Da greater than that of BCA bound to two copper ions.

As a result of can be assigned efficiently in the isotopically depleted BCA spectrum in Figure 6.5b due to the accurate mass assignment of the monoisotopic peak, and with confidence that the ion is bound to copper ions due to the characteristics of the metal bound isotopologue distribution. This also applies to the range of salt adducts and the BCA protein bound to a third copper (Cu^{2+}) ion. These cannot be assigned with such confidence in the isotopically natural spectrum (Figure 6.5a).

A further example of using the asymmetric isotopic distribution was demonstrated by Piergentili et al., [160] investigating the assembly pathway of EncFtn as it forms decamer structures and further elucidate the catalytic mechanism. [96,103] In Figure 6.7 the theoretical models of an EncFtn dimer (-,Figure 6.7a), EncFtn dimer plus two iron ions (-,Figure 6.7b) and EncFtn dimer plus two zinc ions (O- overlaid, Figure 6.7c) were compared to acquired data shown in Figure 6.7c with both iron and zinc supplied in the aqueous buffer. While EncFtn is catalytically active with iron (Fe^{2+}) ions, the model of zinc (Zn^{2+}) bound EncFtn dimer matches the metal bound dimer distribution. The isotopologue distribution also closely matches the theoretical asymmetric distribution as a result of zinc binding to the protein, further confirming the zinc bound protein structure. From Figure 6.7, the analysis of isotopically depleted EncFtn dimer with iron or zinc bound isotope distributions do not overlap. This analysis carried out with isotopically natural protein would have wider isotopologue distributions, which could result in the isotope distribution being spread over the predicted *m/z* for the

EncFtn dimer bound to either iron or zinc, which could complicate the analysis. Similarly, in the isotopically natural distributions of metal bound BCA in Figure 6.3a and Figure 6.5a, the monoisotopic peak cannot be observed and display multiple overlapping distributions which are more difficult to distinguish or assign. This section continues to highlight the advantages of isotope depletion as a method to benefit this type of analysis; accurate assignment, distinction between protein-adduct distributions and isotopologue pattern changes caused by transition metal binding all become simplified and improve our native MS analysis.

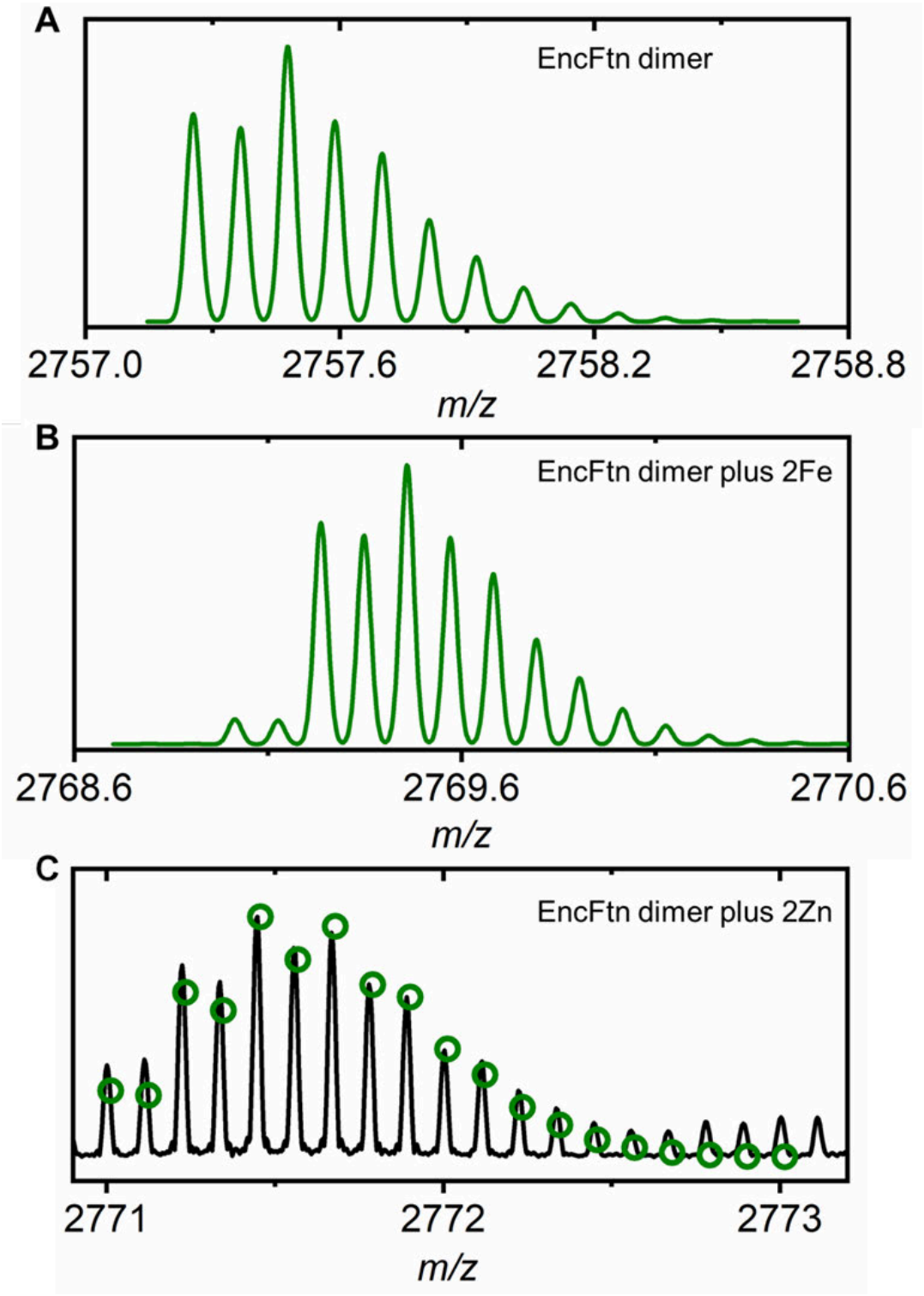


Figure 6.7: Comparison of predicted metal bound EncFtn dimer isotopologue distributions.

Direct comparison of the isotopologue distribution of EncFtn dimer (-, A) with either Fe^{2+} (-, B) or Zn^{2+} (O, C), with acquired EncFtn spectrum distribution (-, C) with access to both Fe^{2+} and Zn^{2+} in the aqueous buffer.

6.4 Isotope depletion and native MS of protein higher order structure.

MS analysis of larger protein poses a greater challenge than analysis of lower molecular weight protein (BCA, 30 kDa). As the molecular weight of analyte protein increases, previously described (Figure 1.4), the number of isotopologue peaks correspondingly increase. The greater number of isotopologue peaks 'spread' over a greater *m/z* range, so can increase overlap of the analyte protein isotopologue distribution to that of the protein plus adduct distributions, or small molecule binding complicating analysis.

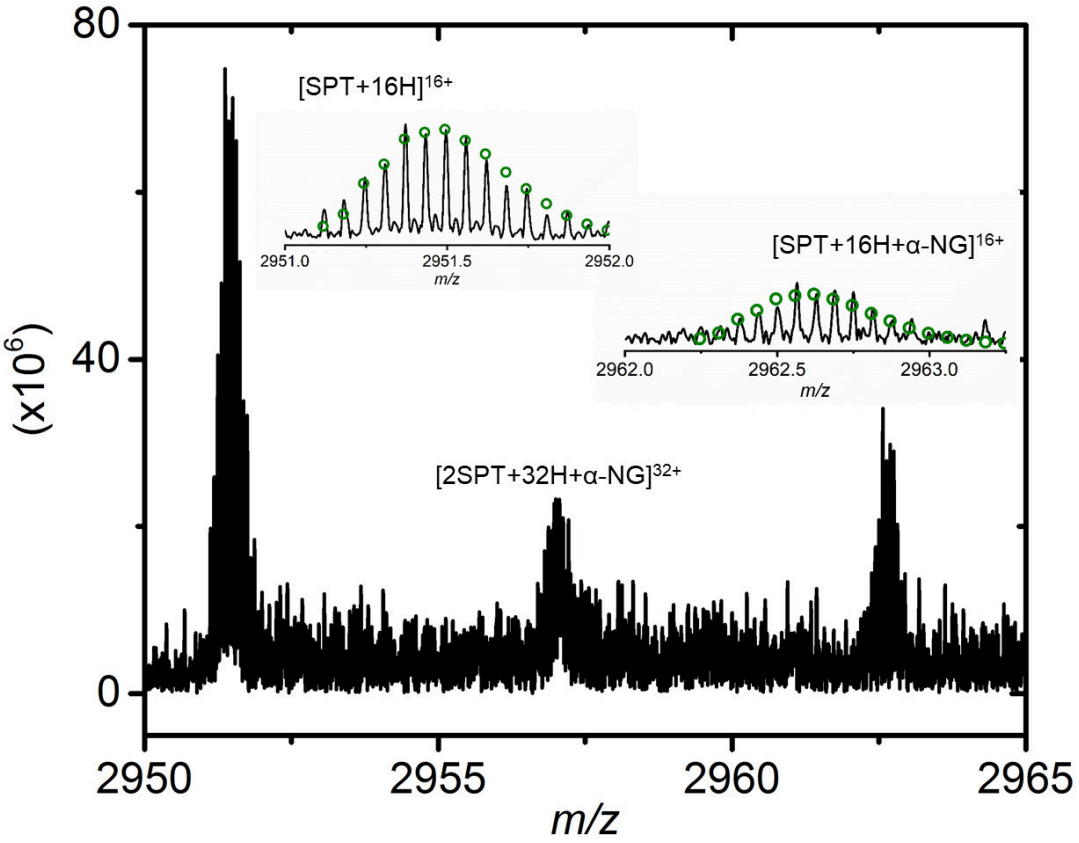


Figure 6.8: Native spectrum of isotopically depleted SPT monomer

Native nESI of isotopically depleted SPT displaying the [SPT+16H]$^{16+}$, the [SPT+16H+α-NG]$^{16+}$ and [2SPT+32H+αNG]$^{32+}$ isotope distributions. The monomer distributions are shown as inserts with the theoretical distribution is overlaid (O).

MS analysis of the SPT construct used in this study (47 kDa) poses an extra challenge, hampered by the addition the modification of the alpha-N-gluconolylation (α-NG) modification ($C_6H_{10}O_6$, +178 Da). [105] This creates a dual distribution of the protein (Figure 6.8) at approximately 2:1 ratio of signal abundance. This has the potential to complicate and obscure information which may overlap the $[SPT+16H+α\text{-}NG]^{16+}$ distribution and limit the abundance of $[SPT+16H]^{16+}$ in the spectrum which can be investigated and analysed. Due to a high collisional activation voltage (30V) in the experimental tuning the SPT protein in Figure 6.8 is primarily as a monomer (* monoisotopic peak, m/z 2951.1). Whereas it is catalytically active and exists as a

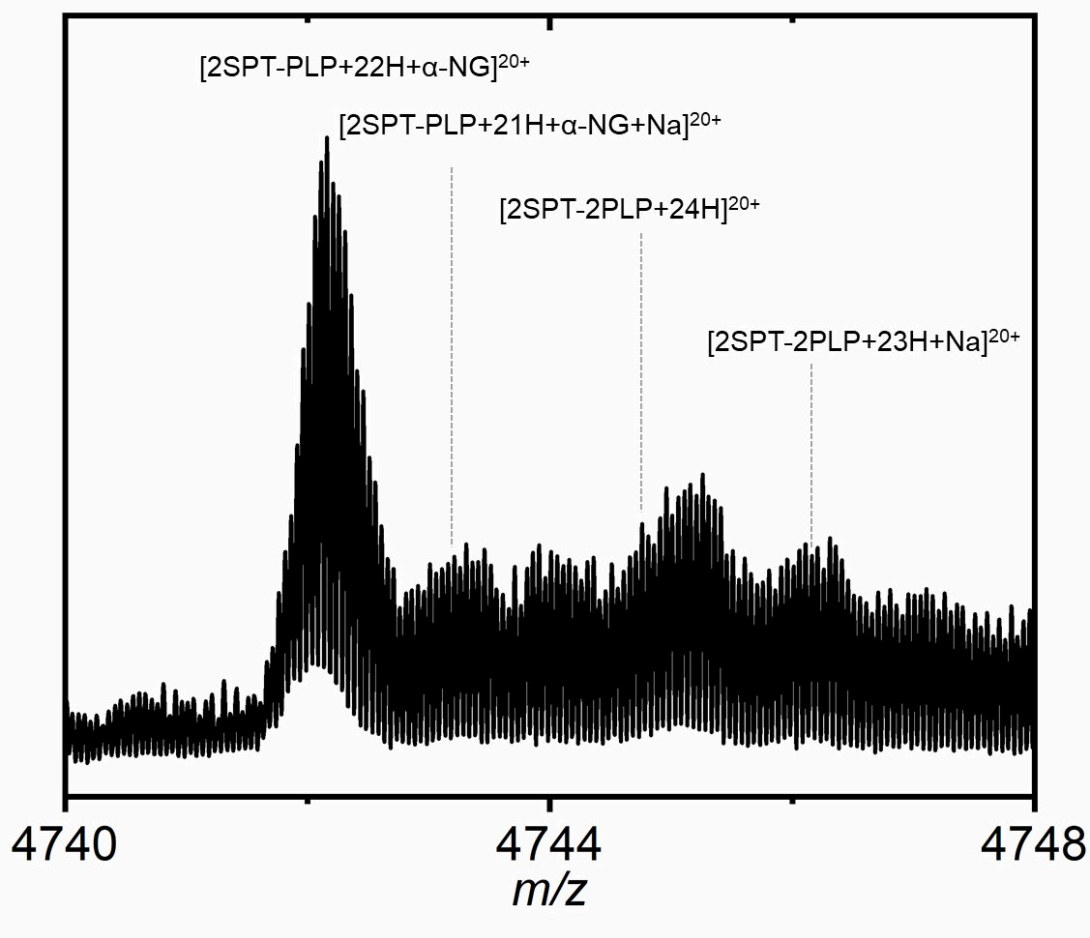


Figure 6.9: Native spectrum of isotopically depleted SPT dimer (20^+) with PLP ligand bound

Native nESI of the 20^+ isotopically depleted SPT dimer bound to catalytic co-factor, PLP (~95 kDa) and associated salt ion adducts.

homodimer species. This is displayed in Figure 6.9, showing the 20^+ charge state of the SPT dimer bound to the pyridoxal phosphate (PLP) co-factor in the active site.

The observed isotopically depleted $[2SPT\text{-}PLP+22H+\alpha\text{-}NG]^{20+}$ species in Figure 6.9 has an average mass of ~95 kDa but is composed of multiple protein overlapping distribution. Including $[2SPT\text{-}2PLP+24H]^{20+}$ and associated sodium adducts, all of which are observed within the 6 *m/z* range, making them difficult to fully resolve and assign.

Our ability to observe a native isotopically depleted 95 kDa protein clearly shows the advantage this as a technique which will improve protein analysis via MS. As the isotopically natural SPT protein distribution under the same native conditions was not sufficient to resolve any isotopologue peaks. The equivalent spectra for isotopically natural SPT dimer was not acquired to provide a direct comparison.

From previous examples and the predictive theoretical models, we know can assume the isotopically natural $[2SPT\text{-}PLP+22H+\alpha\text{-}NG]^{20+}$ species will be composed of a greater number of isotopologue peaks spread over the greater *m/z* range. Theoretically ranging over 3 *m/z* (60 Da) as opposed to the isotopically depleted distribution at 1.7 *m/z* (33 Da) more clearly displayed in Figure 6.10. The signal of the isotopically natural ion species can therefore be predicted to be significantly reduced than that of the isotopically depleted distribution. Which ultimately would reduce the likelihood of differentiation of the ion from the associated adducts.

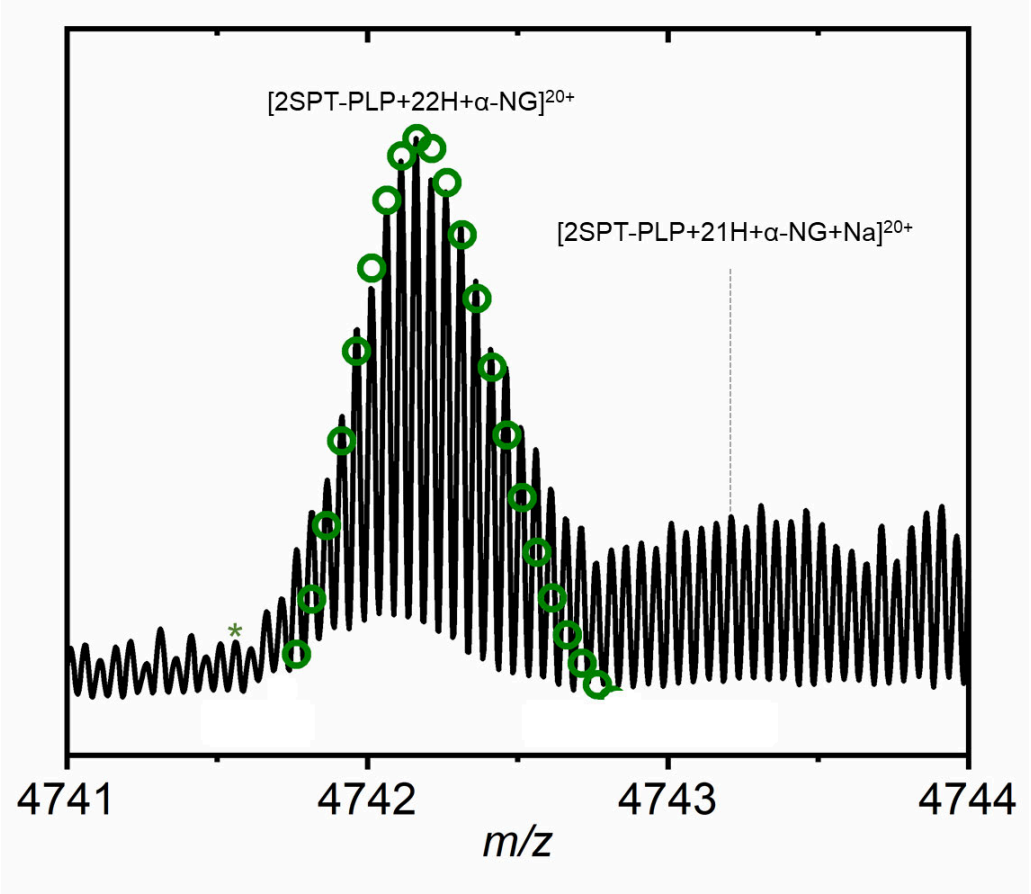


Figure 6.10: Isotopic distribution of isotopically depleted [2SPT-PLP+22H+ α-NPG]$^{20+}$

Closer view of native nESI isotopically depleted SPT dimer displaying the isotopologue distribution in detail. The monoisotopic peak of [2SPT-PLP+22H+ α-NG]$^{20+}$ is indicated by the asterisk (*). The monoisotopic peak of [2SPT-PLP+21H+ α-NG+Na]$^{20+}$ is indicated by the dashed line (---).

However, using SPT as a dimer (~95 kDa) is a clear advantage of double isotope depletion. The reduced width of the depleted isotopologue distribution facilitating the assignment and as the MW of the analysed protein continue to increase, the upper mass limit for feasible high-resolution protein analysis continues to be extended experimentally. If further element depletion was to take place, analysis of larger MW protein would only improve due to further simplification of the isotopologue distribution.

6.5 The contribution of isotope depletion for native MS analysis of megadalton protein assemblies.

Native MS analysis have focused on ever larger protein complexes and it has directed the development of specialised mass spectrometers, such as high mass range Q-TOF, orbitrap and FT-ICR with high magnetic field strength. [25,137] Facilitating the analysis of a range of protein complexes, such as viral capsids, ribosomes or membrane proteins. [161–163] Another example of a larger protein assemblies is encapsulin (Enc, depleted monomer average mass 28954.01 Da) which can assemble into large 60-mer nanocompartments (depleted average mass 1737240.5 Da) and encapsulate the decamers of EncFtn. [14,96,164–167] Decamer EncFtn are

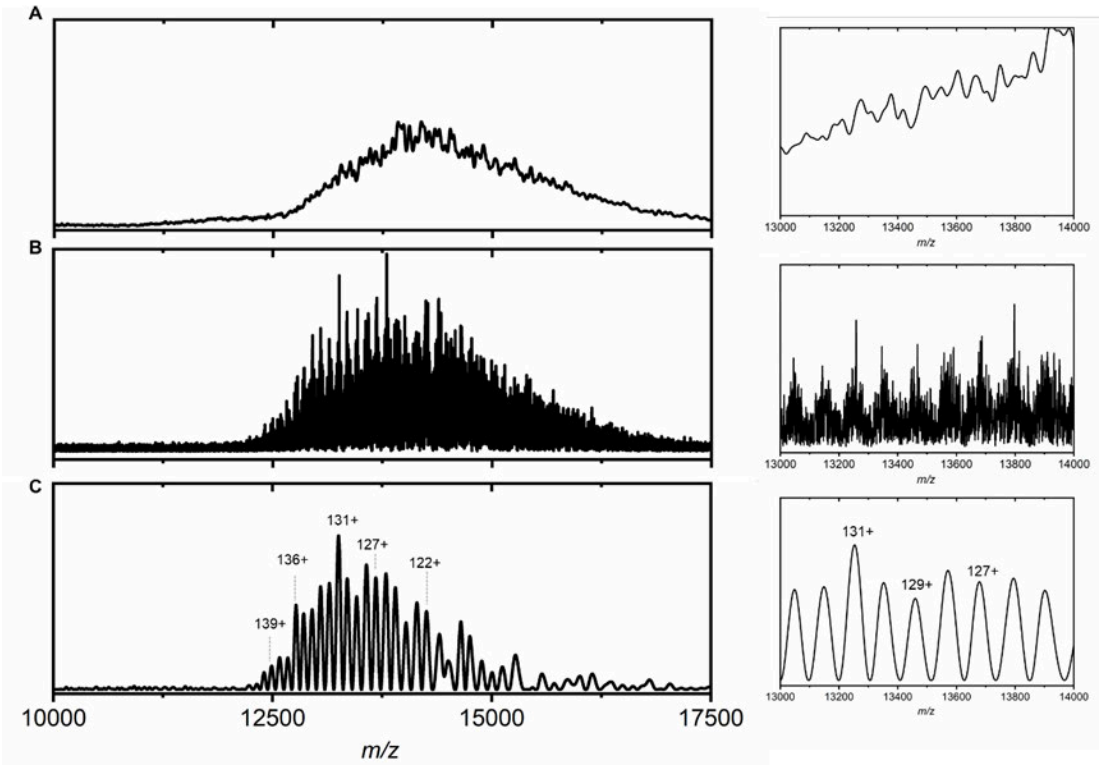


Figure 6.11: Native nESI of assembled Enc 60-mer

Native nESI charge state distribution of unprocessed acquired isotopically natural (A) and the unprocessed acquired isotopically depleted (B) 60-mer Enc complex. The charge state for the isotopically natural (A) Enc could not be sufficiently resolved. Partially resolved isotopically depleted (B) underwent post-acquisition processing resulting in a 'smoothed' spectrum (C) from which an average mass can be measured, and charge assigned.

catalytically active and use the Enc 60-mers as a store for the oxidised iron. Keeping it away from the bacterial cytosol [103] acting as a bacterial organelle.

The isotopically natural and isotopically depleted assembled 60-mer without any EncFtn present was analysed natively on a 12T FT-ICR, shown in Figure 6.11.

In the isotopically natural Enc spectrum (Figure 6.11a), the charge state distribution cannot be clearly resolved and has the visual appearance of a single large humped peak, so no mass information can be obtained. However in the isotopically depleted Enc spectrum (Figure 6.11b) we can observe the large dense charge state distribution of the Enc 60-mer complex. The isotopologue distribution pre-processing do not form a cohesive isotopologue distribution. Therefore, processing and 'smoothing' of the spectra is required to measure the average mass of the protein charge state (Figure 6.11c). This is seen more clearly at closer range in the panels of 13000-14000 *m/z* regions of the spectra. For the isotopically natural Enc (Figure 6.11a) little-to-no information can be resolved from the spectra. Comparatively, the isotopologue distribution can be partially resolved in the closer view of isotopically depleted Enc (Figure 6.11b). However due to the large MW and number of associated charges it cannot be fully resolved in to separate isotopologue peaks, as previously shown in lower MW protein analysis. To improve analysis the isotopically depleted spectrum was smoothed with a 1.5 *m/z* window. This effectively groups the isotopologue peaks of a single charge state together to create a single lower resolution peak from which an average mass can be measured.

As an example of the improvement of isotope depletion in megadalton protein native MS analysis, the Enc 60-mer comparison is visually striking. Enc 60-mer is sufficiently large, it can acquire many charges in the native electrospray process and displays dense charge state distribution. So much so that the isotopically natural protein (Figure 6.11a) cannot be analysed in this manner. The isotopically depleted Enc

assigned charge states (Figure 6.11c) can be used to calculate the molecular mass of the 60-mer with 0.028% error (Appendix 2).

This is an impressive application of the isotope depletion strategy. As it moves larger megadalton molecular weight protein out-with the feasible experimental range of FT-ICR into the working range, without any change in MS equipment. The difference between the Enc 60-mer isotopologue distributions were further investigated in Figure 6.12. The isotopically natural distribution in Figure 6.12a epitomises the challenges faced when studying larger protein via MS. The large number of isotopically natural isotopologue peaks make it difficult to distinguish between them (**O**). The predicted FWHM of the isotopically natural Enc distribution is approximately double that of the isotopically depleted Enc, and consequently to both these observations, the isotopically natural Enc has a reduced observed distribution intensity (Figure 6.12a). This therefore directly relates to what we observe in Figure 6.11a. Likewise it is possible to observe closer to isotopic resolution of isotopically depleted in Figure 6.11b, which facilitated protein analysis with further processing (Figure 6.11c). This is a direct result of the fewer total number of isotopologue peaks, therefore resulting in in an observed greater signal intensity. Reduced chemical complexity results in a reduced distribution width, allowing observation of isotopically depleted Enc in Figure 6.11b.

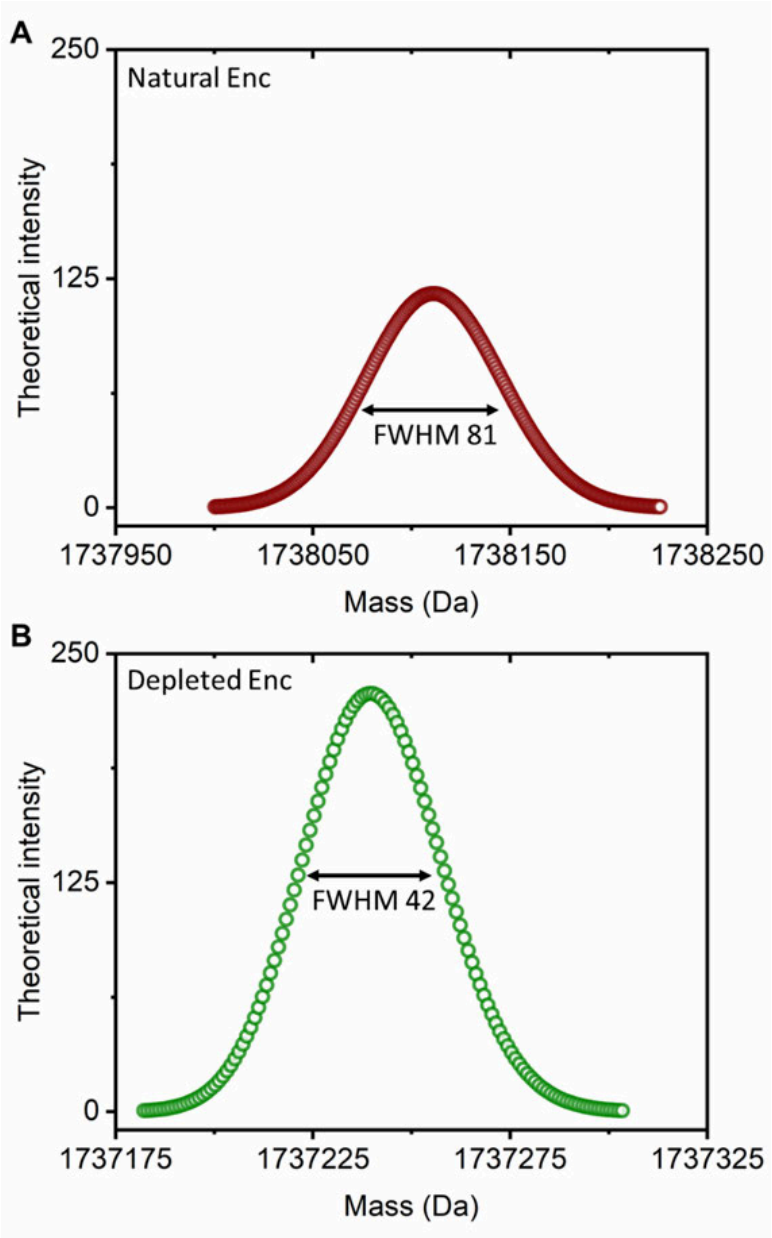


Figure 6.12: Width of distribution of encapsulin 60-mer

*Theoretical isotopologue distribution of Enc 60-mer as isotopically natural (A, **O**, 1738111.72 Da) and isotopically depleted (B, **O**, 1737240.52 Da) protein complexes. Each isotopologue of the protein distributions of each protein is indicated by a coloured circle (**O**) and the calculated FWHM value is indicated.*

These underlying changes to the isotopologue distribution with isotope depletion are constant throughout this investigation. It is impressive that even in very large protein complexes like Enc 60-mers ($C_{77760}H_{121920}N_{21360}O_{23280}S_{300}$) it can still result in a

significant improvement, in what can be observed within the spectrum and facilitating improved protein analysis.

6.6 Conclusion

This chapter highlights the role that isotopic depletion can contribute to native MS and improve analysis. It has demonstrated the benefit of simplified isotopologue profiles contributing to the greater abundance of the monoisotopic peak and reduced width of the distribution. This has been shown to be useful in the identification of different metal ions which can bind to metalloprotein, like BCA. As well as the contribution of the observed asymmetric distribution, which can provide a distinctive change to the isotopologue, aiding in confident assignment. Similarly, as the MW of the protein increased to ~47 kDa in the case of SPT, it allowed acquisition of the protein as a monomer and a dimer with and without an added alpha-N-gluconolylation post-translational modification. That was not observed experimentally with isotopically natural SPT. This is also observed for larger MW protein complexes like Enc, unobservable as an isotopically natural sample can be isotopically simplified into the working range of the 12T FT-ICR.

Native protein MS is a powerful analytical technique which maintains and investigates the structural and conformational traits as a result of maintaining the secondary and higher order structures. Isotope depletion and the simplification of the protein distribution allows a greater distinction between separate protein distributions, and if applied has the potential to drastically improve the analysis of protein and extend the current feasible working range for larger MW protein.

7. Conclusions

The main premise of this investigation was to develop the methodology for the efficient production of isotopically depleted protein and investigate the many advantages isotope depletion can have on protein MS.

The efficient protein expression was accomplished using M9 minimal expression media supplying ^{12}C (99.90%) and ^{14}N (99.99%) and was applied to a range of proteins with differing molecular mass and higher order structure. Depletion of either carbon or nitrogen resulted in a change to the isotopologue distribution, but consistently producing protein with greater levels of carbon depletion (^{12}C, 99.95%) in the double depleted protein. Importantly, the double depletion of these model protein simplified and reduced the widths of the isotopologue distributions, an increased abundance of the monoisotopic peak and as a consequence of this, ions are present at greater S/N.

One MS technique which can potentially benefit from this change in the isotopologue distribution is top-down fragmentation analysis. ECD fragmentation on a 12T FT-ICR was applied to the three-model protein. The comparison of the number of assignable ions found that across the multiple MW the isotopically depleted samples consistently allowed the assignment of 2-4X the number of ions. Directly resulting in an increased sequence coverage for all the model protein. The benefit of the isotope depletion strategy can be observed for top-down fragmentation across multiple platforms. A repeat analysis was carried out using a complementary high-resolution mass analyser, the orbitrap Eclipse. Using the orbitrap analyser, it was possible to observe the same trend as that recorded on the 12T FT-ICR. Complementary to the improvement in sequence coverage caused by isotope depletion, other instrumental developments designed to also improve top-down analysis, like PTCR, was also shown to have a significant impact on analysis. By applying a secondary fragment separation and acquisition, therefore providing another opportunity to identify

fragment ions. The combination of isotope depletion and PTCR allowed the massive increase in sequence coverage from 31% in the isotopically natural SPT (47 kDa) to 78% with the application of PTCR to the isotopically depleted SPT protein.

The fundamental change in the isotopic composition of depleted samples is entirely the cause of the observed improvement in top-down analysis. It also facilitates efficient sequence coverage even at lower numbers of transient accumulation. Primarily due to the abundant monoisotopic peak and greater S/N. With the use of isotopic depletion this study has shown that significantly improved sequence coverage can be obtained with limited transient averaging with protein up to 30 kDa, and likely higher, however the upper mass limit of top down analysis using isotopically depleted protein is still to be experimentally determined. However, the abundance of top-down fragmentation spectra in this thesis are strong proof that the application of isotope depletion does have a positive impact on our ability to improve the analysis of larger MW protein.

As the simplification of the isotopologue distribution and increase in S/N of the ion ultimately can benefit more than just top-down fragmentation analysis. Applying the isotope depletion strategy to native MS analysis has been shown to provide a range of benefits. For lower MW protein, it can facilitate the separation of the isotopologue distribution for the naturally observed overlapping distributions. Similarly, at the same mass range, being able to visualise atypical isotopologue contribution from metal ions can provide a subtle contribution to analysis, such as that shown in the work by Piergentili et al., to identify the binding or outcompeting of metal ions within an assembled higher order complex.

It has also been shown to be largely beneficial in the improvement for analysis of protein complexes up to 1.7 MDa. The resulting simplification to the isotopologue

distributions even at this molecular weight is likely to be similarly beneficial at even greater MW.

7.1 Future outlook

This thesis has built on the initial studies of isotope depletion and developed a simple and robust protocol for the recombinant expression of ^{13}C and ^{15}N depleted protein in *E. coli*. The benefit to top-down fragmentation up to 50 kDa has been investigated. Therefore, the obvious development within this area of research is to continue to push the boundary of the MW analysed. This can lead to many interesting developments, including the depletion of greater number of elements like sulfur and oxygen, and further application of the technique for online fragmentation analysis.

The isotope depletion strategy can and should be applied to structural MS techniques for biomacromolecules of larger MW. Specialised techniques like HDX protein analysis, native MS; including top-down fragmentation and protein interactions (small molecules, peptides, other proteins etc). All of these techniques will be directly improved by the simplified isotope profile, increased S/N and a prevalent monoisotopic peak. Increasing the achievable sensitivity and increase the working mass range. The application of isotope depletion to HDX in particular would be extremely beneficial, as HDX increases the width and complexity of the isotopologue peaks, simplifying the initial ion prior to the addition of deuterium (^{2}H, +1.00628). Thereby increasing the working mass range of the peptide ion which can be analysed via HDX and improve identification of the number of deuterium ions which have exchanged onto the peptide ion.

Ultimately this thesis has been limited by only showing recombinant protein, expressed in the workflow described at the start of Chapter 3 (page 52) using *E. coli* BL21 cells. As this workflow incorporated the depleted isotopes as glucose and

ammonium sulphate, the depleted isotope sources will be incorporated throughout the entire cellular proteome. This opens the potential of the isotope depletion strategy to the investigation of proteomic scale analysis. An area which would vastly improve the isotope depletion strategy would be to modify the described workflow of cell culture to facilitate the culture of eukaryotic cells. Such as yeast cells, which can be cultured on defined carbon sources, and will facilitate the recombinant expression of complex folding and the addition of PTMs. This will also increase the proteomic potential of the isotope depletion strategy.

The analysis of all complex protein samples has the potential to become simplified using the isotope depletion strategy. The initial groundwork covered in this study will hopefully form the basis of reference to allow the rapid expansion of isotope depletion throughout the mass spectrometry community and become a widely exploited tool for biomolecular analysis.

8. References

(1) Fenn, J. B. Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chemie - Int. Ed.* **2003**, *42* (33), 3871–3894. https://doi.org/10.1002/anie.200300605.

(2) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. *Anal. Chem.* **1985**, *57* (3), 675–679. https://doi.org/10.1021/ac00280a023.

(3) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151–153.

(4) Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal. Chem.* **1988**, *60* (20), 2299–2301. https://doi.org/10.1021/ac00171a028.

(5) Taylor, G. Disintegration of Water Drops in an Electric Field. *R. Soc.* **1964**, *280* (1382), 383–397. https://doi.org/10.1098/rspa.1964.0151.

(6) 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**, 1–40. https://doi.org/10.1155/2012/282574.

(7) 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.

(8) Konermann, L.; Rodriguez, A. D.; Liu, J. On the Formation of Highly Charged Gaseous Ions from Unfolded Proteins by Electrospray Ionization. *Anal. Chem.* **2012**, *84* (15), 6798–6804. https://doi.org/10.1021/ac301298g.

(9) 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.

(10) Wilm, M.; Mann, M. Analytical Properties of the Nanoelectrospray Ion Source. *Anal. Chem.* **1996**, *68* (1), 1–8. https://doi.org/10.1021/ac9509519.

(11) Juraschek, R.; Dülcks, T.; Karas, M. Nanoelectrospray - More than Just a Minimized-Flow Electrospray Ionization Source. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (4), 300–308. https://doi.org/10.1016/S1044-0305(98)00157-3.

(12) Schmidt, A.; Karas, M.; Dülcks, T. Effect of Different Solution Flow Rates on Analyte Ion Signals in Nano-ESI MS, or: When Does ESI Turn into Nano-ESI? *J. Am. Soc. Mass Spectrom.* **2003**, *14* (5), 492–500. https://doi.org/10.1016/S1044-0305(03)00128-4.

(13) Cassou, C. A.; Williams, E. R. Desalting Protein Ions in Native Mass Spectrometry Using Supercharging Reagents. *Analyst* **2014**. https://doi.org/10.1039/c4an01085j.

(14) Lössl, P.; Snijder, J.; Heck, A. J. R. Boundaries of Mass Resolution in Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2014**. https://doi.org/10.1007/s13361-014-0874-3.

(15) Rosman, K, J, R.; Taylor, P, D, P. ISOTOPIC COMPOSITIONS OF THE ELEMENTS (IUPAC Technical Report). *Pure Appl. Chem.* **1998**, *70* (1), 217–235.

(16) Rockwood, A. L.; Van Orden, S. L.; Smith, R. D. Rapid Calculation of Isotope Distributions. *Anal. Chem. Rockwood. A. L. Rapid Commun. Mass Spectrom. J Chem. Educ. J. A. Int. J. Mass Spectrom. Ion Phys* **1995**, *67* (52), 2699–

2704.

(17) Zubarev, R. A.; Demirev, P. A. Isotope Depletion of Large Biomolecules: Implications for Molecular Mass Measurements. *J. Am. Soc. Mass Spectrom.* **1998**. https://doi.org/10.1016/S1044-0305(97)00232-8.

(18) Rockwood, A. L.; Van Orden, S. L. Ultrahigh-Speed Calculation of Isotope Distributions. *Anal. Chem.* **1996**. https://doi.org/10.1021/ac951158i.

(19) Senko, M. W.; Beu, S. C.; McLaffertycor, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* **1995**. https://doi.org/10.1016/1044-0305(95)00017-8.

(20) Valkenborg, D.; Jansen, I.; Burzykowski, T. A Model-Based Method for the Prediction of the Isotopic Distribution of Peptides. *J. Am. Soc. Mass Spectrom.* **2008**. https://doi.org/10.1016/j.jasms.2008.01.009.

(21) Pourshahian, S. Mass Defect from Nuclear Physics to Mass Spectral Analysis. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (9), 1836–1843. https://doi.org/10.1007/s13361-017-1741-9.

(22) G. Marshall, A.; T. Blakney, G.; Chen, T.; K. Kaiser, N.; M. McKenna, A.; P. Rodgers, R.; M. Ruddy, B.; Xian, F. Mass Resolution and Mass Accuracy: How Much Is Enough? *Mass Spectrom.* **2013**, *2* (Special_Issue), S0009. https://doi.org/10.5702/massspectrometry.s0009.

(23) Liu, Q.; Easterling, M. L.; Agar, J. N. Resolving Isotopic Fine Structure to Detect and Quantify Natural Abundance-and Hydrogen/Deuterium Exchange-Derived Isotopomers. *Anal. Chem.* **2014**, *86* (1), 820–825. https://doi.org/10.1021/ac403365g.

(24) Shi, S. D. H.; Hendrickson, C. L.; Marshall, A. G. Counting Individual Sulfur Atoms in a Protein by Ultrahigh Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Experimental Resolution of Isotopic Fine Structure in Proteins. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11532–11537. https://doi.org/10.1073/pnas.95.20.11532.

(25) Snijder, J.; Heck, A. J. R. Analytical Approaches for Size and Mass Analysis of Large Protein Assemblies. *Annu. Rev. Anal. Chem* **2014**, *7*, 43–64. https://doi.org/10.1146/annurev-anchem-071213-020015.

(26) Thomson, J. J. Charge Carried by the Cathode Rays . *Philos. Mag.* **1897**, *44* (293), 1–24.

(27) Thomson, J. J. Some Further Applications of the Method of Positive Rays. *Nature* **1913**, *91* (28), 333–337.

(28) Thomson, J. J. The Occurrence of Helium and Neon in Vacuum Tubes [3]. *Nature* **1927**, *118* (2980), 838–839. https://doi.org/10.1038/118838a0.

(29) Xian, F.; Hendrickson, C. L.; Marshall, A. G. High Resolution Mass Spectrometry. *Anal. Chem.* **2012**, *84* (2), 708–719. https://doi.org/10.1021/ac203191t.

(30) Mamyrin, B. A. Time-of-Flight Mass Spectrometry (Concepts, Achievements, and Prospects). *Int. J. Mass Spectrom.* **2001**, *206*, 251–266.

(31) Comisarow, M. B.; Marshall, A. G. Resolution-Enhanced Fourier Transform Ion Cyclotron Resonance Spectroscopy. *J. Chem. Phys.* **1975**, *62* (1), 293–295. https://doi.org/10.1063/1.430222.

(32) Marshall, A. G.; Comisarow, M. B. Fourier Transform Methods in Spectroscopy. *J. Chem. Educ.* **1975**, *52* (10), 638. https://doi.org/10.1021/ed052p638.

(33) Guan, S.; Marshall, A. G. *Mass Spectrometry and Ion Processes Ion Traps for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Principles and Design of Geometric and Electric Configurations*; 1995; Vol. 146.

(34) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev.* **1998**, *17*, 1–35. https://doi.org/10.1002/(SICI)1098-2787(1998)17:1<1::AID-MAS1>3.0.CO;2-K.

(35) Scigelova, M.; Hornshaw, M.; Giannakopulos, A.; Makarov, A. Fourier Transform Mass Spectrometry. *Mol. Cell. Proteomics* **2011**, *10* (7). https://doi.org/10.1074/mcp.M111.009431.

(36) Perry, R. H.; Cooks, R. G.; Noll, R. J. Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications. *Mass Spectrom. Rev.* **2008**, *27* (6), 661–699. https://doi.org/10.1002/mas.20186.

(37) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72* (6), 1156–1162. https://doi.org/10.1021/ac991131p.

(38) Weisbrod, C. R.; Kaiser, N. K.; Syka, J. E. P.; Early, L.; Mullen, C.; Dunyach, J. J.; English, A. M.; Anderson, L. C.; Blakney, G. T.; Shabanowitz, J.; Hendrickson, C. L.; Marshall, A. G.; Hunt, D. F. Front-End Electron Transfer Dissociation Coupled to a 21 Tesla FT-ICR Mass Spectrometer for Intact Protein Sequence Analysis. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (9), 1787–1795. https://doi.org/10.1007/s13361-017-1702-3.

(39) Anderson, L. C.; DeHart, C. J.; Kaiser, N. K.; Fellers, R. T.; Smith, D. F.; Greer, J. B.; LeDuc, R. D.; Blakney, G. T.; Thomas, P. M.; Kelleher, N. L.; Hendrickson, C. L. Identification and Characterization of Human Proteoforms

by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry. *J. Proteome Res.* **2017**, *16* (2), 1087–1096. https://doi.org/10.1021/acs.jproteome.6b00696.

(40) Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85* (11), 5288–5296. https://doi.org/10.1021/ac4001223.

(41) Sidoli, S.; Cheng, L.; Jensen, O. N. Proteomics in Chromatin Biology and Epigenetics: Elucidation of Post-Translational Modifications of Histone Proteins by Mass Spectrometry. *J. Proteomics* **2012**, *75*, 3419–3433. https://doi.org/10.1016/j.jprot.2011.12.029.

(42) Lanucara, F.; Eyers, C. E. Top-down Mass Spectrometry for the Analysis of Combinatorial Post-Translational Modifications. *Mass Spectrometry Reviews*. 2013. https://doi.org/10.1002/mas.21348.

(43) Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G. Mass Spectrometry of Peptides and Proteins. *Methods*. 2005. https://doi.org/10.1016/j.ymeth.2004.08.013.

(44) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422*, 198–207. https://doi.org/10.1038/nature01511.

(45) Smith, L. M.; Kelleher, N. L.; Proteomics, T. C. for T.-D. Proteoform: A Single Term Describing Protein Complexity. *Nat Methods* **2013**, *10* (3), 186–187. https://doi.org/10.1038/nmeth.2369.Proteoform.

(46) Wiśniewski, J. R.; Zougman, A.; Krüger, S.; Mann, M. Mass Spectrometric Mapping of Linker Histone H1 Variants Reveals Multiple Acetylations, Methylations, and Phosphorylation as Well as Differences between Cell Culture and Tissue. *Mol. Cell. Proteomics* **2007**, *6* (1), 72–87. https://doi.org/10.1074/mcp.M600255-MCP200.

(47) Tweedie-Cullen, R. Y.; Brunner, A. M.; Grossmann, J.; Mohanna, S.; Sichau, D.; Nanni, P.; Panse, C.; Mansuy, I. M. Identification of Combinatorial Patterns of Post-Translational Modifications on Individual Histones in the Mouse Brain. *PLoS One* **2012**, *7* (5). https://doi.org/10.1371/journal.pone.0036980.

(48) Roth, M. J.; Forbes, A. J.; Boyne, M. T.; Kim, Y. Bin; Robinson, D. E.; Kelleher, N. L. Precise and Parallel Characterization of Coding Polymorphisms, Alternative Splicing, and Modifications in Human Proteins by Mass Spectrometry. *Mol. Cell. Proteomics* **2005**, *4* (7), 1002–1008. https://doi.org/10.1074/mcp.M500064-MCP200.

(49) Waldera-Lupa, D. M.; Stefanski, A.; Meyer, H. E.; Stühler, K. The Fate of B-Ions in the Two Worlds of Collision-Induced Dissociation. *Biochim. Biophys. Acta - Proteins Proteomics* **2013**. https://doi.org/10.1016/j.bbapap.2013.08.007.

(50) Mortensen, D. N.; Jones, C. A.; Dearden, D. V. Appropriate Choice of Event Length in Sustained Off-Resonance Irradiation Collision-Induced Dissociation (SORI-CID) Experiments: Activated Ion Collision-Induced Dissociation. *Int. J. Mass Spectrom.* **2012**, *330–332*, 241–245. https://doi.org/10.1016/j.ijms.2012.08.026.

(51) Ge, Y.; Horn, D. M.; Mclafferty, F. W. *Blackbody Infrared Radiative Dissociation of Larger (42 KDa) Multiply Charged Proteins*; 2001.

(52) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994**. https://doi.org/10.1021/ac00090a004.

(53) Crowe, M. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation (IRMPD) and

Collisionally Activated Dissociation of Peptides in a Quadrupole Ion Trap with Selective IRMPD of Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2004**. https://doi.org/10.1016/j.jasms.2004.07.016.

(54) Price, W. D.; Schnier, P. D.; Jockusch, R. A.; Strittmatter, E. F.; Williams, E. R. Unimolecular Reaction Kinetics in the High-Pressure Limit without Collisions. *J. Am. Chem. Soc.* **1996**, *118* (43), 10640–10644. https://doi.org/10.1021/ja961812r.

(55) Elviri, L. *Tandem Mass Spectrometry - Applications and Principles*; Jeevan, P., Ed.; INTECH EUROPE, JANEZA TRDINE9, RIJEKA, 51000, CROATIA, 2012. https://doi.org/10.5772/1327.

(56) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88* (1), 30–51. https://doi.org/10.1021/acs.analchem.5b04563.

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

(58) Tsybin, Y. O.; Witt, M.; Baykut, G.; Kjeldsen, F.; Håkansson, P. Combined Infrared Multiphoton Dissociation and Electron Capture Dissociation with a Hollow Electron Beam in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (15), 1759–1768. https://doi.org/10.1002/rcm.1118.

(59) Tsybin, Y. O.; Quinn, J. P.; Tsybin, O. Y.; Hendrickson, C. L.; Marshall, A. G. Electron Capture Dissociation Implementation Progress in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (6), 762–771. https://doi.org/10.1016/j.jasms.2008.02.007.

(60) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, A.; Tsybin, Y. O. Principles of Electron Capture and Transfer Dissociation Mass Spectrometry Applied to Peptide and Protein Structure Analysis. *Chem. Soc. Rev. Chem. Soc. Rev* **2013**, *42* (42), 5014–5030. https://doi.org/10.1039/c3cs35477f.

(61) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical Solutions: Principles and Application of Electron-Based Dissociation in Mass Spectrometry-Based Analysis of Protein Structure. *Mass Spectrom. Rev.* **2018**, *37* (6), 750–771. https://doi.org/10.1002/mas.21560.

(62) 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.* **2004**, *101* (26), 9528–9533. https://doi.org/10.1073/pnas.0402700101.

(63) Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Jensen, F.; Zubarev, R. A. Dissociative Capture of Hot (3-13 EV) Electrons by Polypeptide Polycations: An Efficient Process Accompanied by Secondary Fragmentation. *Chem. Phys. Lett.* **2002**, *356* (3–4), 201–206. https://doi.org/10.1016/S0009-2614(02)00149-5.

(64) Williams, J. P.; Creese, A. J.; Roper, D. R.; Green, B. N.; Cooper, H. J. Hot Electron Capture Dissociation Distinguishes Leucine from Isoleucine in a Novel Hemoglobin Variant, Hb Askew, Beta-54(D5) Val-Ile. **2009**, *20*, 1707–1713. https://doi.org/10.1016/j.jasms.2009.05.002.

(65) Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A. Electron Detachment Dissociation of Peptide Di-Anions: An Electron-Hole Recombination Phenomenon. *Chem. Phys. Lett.* **2001**, *342* (3–4), 299–302. https://doi.org/10.1016/S0009-2614(01)00501-2.

(66) Ganisl, B.; Valovka, T.; Hartl, M.; Taucher, M.; Bister, K.; Breuker, K. Electron Detachment Dissociation for Top-down Mass Spectrometry of Acidic Proteins. *Chem. - A Eur. J.* **2011**, *17* (16), 4460–4469. https://doi.org/10.1002/chem.201003709.

(67) Earley, L.; Anderson, L. C.; Bai, D. L.; Mullen, C.; Syka, J. E. P.; English, A. M.; Dunyach, J. J.; Stafford, G. C.; Shabanowitz, J.; Hunt, D. F.; Compton, P. D. Front-End Electron Transfer Dissociation: A New Ionization Source. *Anal. Chem.* **2013**, *85* (17), 8385–8390. https://doi.org/10.1021/ac401783f.

(68) Rose, C. M.; Russell, J. D.; Ledvina, A. R.; McAlister, G. C.; Westphall, M. S.; Griep-Raming, J.; Schwartz, J. C.; Coon, J. J.; Syka, J. E. P. Multipurpose Dissociation Cell for Enhanced ETD of Intact Protein Species. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (6), 816–827. https://doi.org/10.1007/s13361-013-0622-0.

(69) Syrstad, E. A.; Turecek, 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.

(70) Bowers, W. D.; Delbert, S. S.; Hunter, R. L.; Mclver, R. T. Fragmentation of Oligopeptide Ions Using Ultraviolet Laser Radiation and Fourier Transform Mass Spectrometry. *J. Am. Chem. Soc.* **1984**, *106* (23), 7288–7289. https://doi.org/10.1021/ja00335a094.

(71) Cui, W.; Thompson, M. S.; Reilly, J. P. Pathways of Peptide Ion Fragmentation Induced by Vacuum Ultraviolet Light. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1384–1398. https://doi.org/10.1016/j.jasms.2005.03.050.

(72) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein

Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651. https://doi.org/10.1021/ja4029654.

(73) Morrison, L. J.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation Mass Spectrometry of Tetrameric Protein Complexes Provides Insight into Quaternary and Secondary Protein Topology. *J. Am. Chem. Soc.* **2016**, *138*, 10849–10859. https://doi.org/10.1021/jacs.6b03905.

(74) Stephenson, J. L.; McLuckey, S. A. *Simplification of Product Ion Spectra Derived from Multiply Charged Parent Ions via Ion/Ion Chemistry*; 1999; Vol. 70. https://doi.org/10.1021/ac9802832.

(75) Stephenson, J. L.; McLuckey, S. A. Ion/Ion Reactions in the Gas Phase: Proton Transfer Reactions Involving Multiply-Charged Proteins. *J. Am. Chem. Soc.* **1996**, *118* (31), 7390–7397. https://doi.org/10.1021/ja9611755.

(76) Gunawardena, H. P.; He, M.; Chrisman, P. A.; Pitteri, S. J.; Hogan, J. M.; Hodges, B. D. M.; McLuckey, S. A. Electron Transfer versus Proton Transfer in Gas-Phase Ion/Ion Reactions of Polyprotonated Peptides. *J. Am. Chem. Soc.* **2005**, *127* (36), 12627–12639. https://doi.org/10.1021/ja0526057.

(77) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. *Anal. Chem.* **2016**, *88* (1), 1008–1016. https://doi.org/10.1021/acs.analchem.5b03911.

(78) Huguet, R.; Mullen, C.; Srzentić, K.; Greer, J. B.; Fellers, R. T.; Zabrouskov, V.; Syka, J. E. P.; Kelleher, N. L.; Fornelli, L. Proton Transfer Charge Reduction Enables High-Throughput Top-Down Analysis of Large Proteoforms. *Anal. Chem.* **2019**, *91* (24), 15732. https://doi.org/10.1021/acs.analchem.9b03925.

(79) Sanders, J.; Mullen, C.; Watts, E.; Holden, D.; Syka, J. E.; Schwartz, J. C.; Brodbelt, J. S. Enhanced Sequence Coverage of Large Proteins by Combining Ultraviolet Photodissociation with Proton Transfer Reactions. *Anal. Chem.* **2019**, acs.analchem.9b04026. https://doi.org/10.1021/acs.analchem.9b04026.

(80) Wenger, C. D.; Lee, M. V.; Hebert, A. S.; Mcalister, G. C.; Phanstiel, D. H.; Westphall, M. S.; Coon, J. J. Gas-Phase Purification Enables Accurate, Large-Scale, Multiplexed Proteome Quantification with Isobaric Tagging. *Nat Methods* **2012**, *8* (11), 933–935. https://doi.org/10.1038/nmeth.1716.

(81) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical Software to Analyze Top-down Mass Spectrometry Data. *Proteomics* **2015**, *15*, 1235–1238. https://doi.org/10.1002/pmic.201400313.

(82) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (4), 320–332. https://doi.org/10.1016/S1044-0305(99)00157-9.

(83) Köster, C.; Holle, A. A New Intelligent Annotation Procedure: SNAP. **1999**.

(84) Mann, M.; Wilm, M. *Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags*; Academic Press, 1994; Vol. 66.

(85) 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.

(86) Hussong, R.; Tholey, A.; Hildebrandt, A. Efficient Analysis of Mass Spectrometry Data Using the Isotope Wavelet. *AIP Conf. Proc.* **2007**, *940*

(February), 139–149. https://doi.org/10.1063/1.2793396.

(87) 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**. https://doi.org/10.1007/s13361-016-1549-z.

(88) Marshall, A. G.; Senko, M. W.; Li, W. Q.; Li, M.; Dillon, S.; Guan, S. H.; Logan, T. M. Protein Molecular Mass to 1 Da By C-13, N-15 Double-Depletion and Ft-Icr Mass Spectrometry. *J. Am. Chem. Soc.* **1997**, *119* (2), 433–434.

(89) 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 13 C, 15 N-Doubly Depleted Protein. *Anal. Chem.* **1998**, *70* (15), 3333–3336. https://doi.org/10.1021/ac980215f.

(90) Jensen, P. K.; Paša-Tolić, L.; Anderson, G. A.; Horner, J. A.; Lipton, M. S.; Bruce, J. E.; Smith, R. D. Probing Proteomes Using Capillary Isoelectric Focusing-Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **1999**, *71* (11), 2076–2084. https://doi.org/10.1021/ac990196p.

(91) Sharma, S.; Simpson, D. C.; Tolić, N.; Jaitly, N.; Mayampurath, A. M.; Smith, R. D.; Paša-Tolić, L. Proteomic Profiling of Intact Proteins Using WAX-RPLC 2-D Separations and FTICR Mass Spectrometry. *J. Proteome Res.* **2007**, *6* (2), 602–610. https://doi.org/10.1021/pr060354a.

(92) Stump, M. J.; Jones, J. J.; Fleming, R. C.; Lay, J. O.; Wilkins, C. L. Use of Double-Depleted 13C and 15N Culture Media for Analysis of Whole Cell Bacteria by MALDI Time-of-Flight and Fourier Transform Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (11), 1306–1314.

https://doi.org/10.1016/S1044-0305(03)00577-4.

(93) 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.

(94) Xiong, Y.; Schroeder, K.; Greenbaum, N. L.; Hendrickson, C. L.; Marshall, A. G. Improved Mass Analysis of Oligoribonucleotides by 13C, 15N Double Depletion and Electrospray Ionization FT-ICR Mass Spectrometry. *Anal. Chem.* **2004**. https://doi.org/10.1021/ac030299e.

(95) Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L. On the Scalability and Requirements of Whole Protein Mass Spectrometry. *Anal. Chem.* **2011**. https://doi.org/10.1021/ac2010795.

(96) He, D.; Vanden-Hehir, S.; Georgiev, A.; Altenbach, K.; Tarrant, E.; Mackay, C. L.; Waldron, K. J.; Clarke, D. J.; Marles-Wright, J. Structural Characterisation of an Encapsulated Decameric Ferritin Provides Insight into Iron Storage within Bacterial Nanocompartments. *Nat Struct Mol Biol* **2016**. https://doi.org/10.7554/eLife.18972.

(97) Marley, J.; Lu, M.; Bracken, C. A Method for Efficient Isotopic Labeling of Recombinant Proteins. *J. Biomol. NMR* **2001**, *20* (1), 71–75.

(98) Wadsworth, J. M.; Clarke, D. J.; Mcmahon, S. A.; Lowther, J. P.; Beattie, A. E.; Langridge-Smith, P. R. R.; Broughton, H. B.; Dunn, T. M.; Naismith, J. H.; Campopiano, D. J. The Chemical Basis of Serine Palmitoyltransferase Inhibition by Myriocin. *J. Am. Chem. Soc.* **2013**, *135*, 14276−14285. https://doi.org/10.1021/ja4059876.

(99) Jiang, M.; Zhang, H. Engineering the Shikimate Pathway for Biosynthesis of Molecules with Pharmaceutical Activities in E. Coli. *Current Opinion in Biotechnology*. 2016. https://doi.org/10.1016/j.copbio.2016.01.016.

(100) Marley, J.; Lu, M.; Bracken, C. A Method for Efficient Isotopic Labeling of Recombinant Proteins. *J. Biomol. NMR* **2001**, *20* (1), 71–75. https://doi.org/10.1023/A:1011254402785.

(101) Bertani, G. Studies on Lysogenesis. I. The Mode of Phage Liberation by Lysogenic Escherichia Coli. *J. Bacteriol.* **1951**, *62* (3), 293–300. https://doi.org/10.1128/JB.62.3.293-300.1951.

(102) Green, M. R.; Sambrook, J. *Molecular Cloning : A Laboratory Manual*, 4th ed.; Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, [2012] Copyright ©2012.: N.Y, 2012.

(103) 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.

(104) Liu, H.; Naismith, J. H. A Simple and Efficient Expression and Purification System Using Two Newly Constructed Vectors. *Protein Expr. Purif.* **2009**. https://doi.org/10.1016/j.pep.2008.09.008.

(105) Geoghegan, K. F.; Dixon, H. B. F.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; Lemotte, P. K.; McColl, A. S.; Kamath, A. V.; Stroh, J. G. Spontaneous α-N-6-Phosphogluconoylation of a “His Tag” in Escherichia Coli: The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. *Anal. Biochem.* **1999**, *267* (1), 169–

184. https://doi.org/10.1006/abio.1998.2990.

(106) Robinson, N. E.; Zabrouskov, V.; Zhang, J.; Lampi, K. J.; Robinson, A. B. Measurement of Deamidation of Intact Proteins by Isotopic Envelope and Mass Defect with Ion Cyclotron Resonance Fourier Transform Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2006**, *20* (23), 3535–3541. https://doi.org/10.1002/rcm.2767.

(107) Kanawati, B.; Bader, T. M.; Wanczek, K. P.; Li, Y.; Schmitt-Kopplin, P. Fourier Transform (FT)-Artifacts and Power-Function Resolution Filter in Fourier Transform Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2017**, *31* (19), 1607–1615. https://doi.org/10.1002/rcm.7940.

(108) Nagao, T.; Yukihira, D.; Fujimura, Y.; Saito, K.; Takahashi, K.; Miura, D.; Wariishi, H. Power of Isotopic Fine Structure for Unambiguous Determination of Metabolite Elemental Compositions: In Silico Evaluation and Metabolomic Application. *Anal. Chim. Acta* **2014**, *813*, 70–76. https://doi.org/10.1016/J.ACA.2014.01.032.

(109) Dittwald, P.; Valkenborg, D.; Claesen, J.; Rockwood, A. L.; Gambin, A. On the Fine Isotopic Distribution and Limits to Resolution in Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2015**. https://doi.org/10.1007/s13361-015-1180-4.

(110) Horn, D. M.; Ge, Y.; McLafferty, F. W. Activated Ion Electron Capture Dissociation for Mass Spectral Sequencing of Larger (42 KDa) Proteins. *Anal. Chem.* **2000**, *72* (20), 4778–4784. https://doi.org/10.1021/ac000494i.

(111) Shaw, J. B.; Malhan, N.; Vasil'Ev, Y. V.; Lopez, N. I.; Makarov, A.; Beckman, J. S.; Voinov, V. G. Sequencing Grade Tandem Mass Spectrometry for Top-Down Proteomics Using Hybrid Electron Capture Dissociation Methods in a Benchtop Orbitrap Mass Spectrometer. *Anal. Chem.* **2018**, *90* (18), 10819–

10827. https://doi.org/10.1021/acs.analchem.8b01901.

(112) Chalmers, M. J.; Håkansson, K.; Johnson, R.; Smith, R.; Shen, J.; Emmett, M. R.; Marshall, A. G. Protein Kinase A Phosphorylation Characterized by Tandem Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Proteomics* **2004**, *4*, 970–981. https://doi.org/10.1002/pmic.200300650.

(113) Cui, W.; Zhang, H.; Blankenship, R. E.; Gross, M. L. Electron-Capture Dissociation and Ion Mobility Mass Spectrometry for Characterization of the Hemoglobin Protein Assembly. *Protein Sci.* **2015**, *24* (8), 1325–1332. https://doi.org/10.1002/pro.2712.

(114) Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **2011**. https://doi.org/10.1021/pr1011729.

(115) Pacholarz, K. J.; Barran, P. E. Distinguishing Loss of Structure from Subunit Dissociation for Protein Complexes with Variable Temperature Ion Mobility Mass Spectrometry. *Anal. Chem.* **2015**, *87* (12), 6271–6279. https://doi.org/10.1021/acs.analchem.5b01063.

(116) Hall, Z.; Hernández, H.; Marsh, J. A.; Teichmann, S. A.; Robinson, C. V. The Role of Salt Bridges, Charge Density, and Subunit Flexibility in Determining Disassembly Routes of Protein Complexes. *Structure* **2013**, *21* (8), 1325–1337. https://doi.org/10.1016/j.str.2013.06.004.

(117) Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. *J. Am. Chem. Soc.* **2012**, *134* (7), 3429–3438. https://doi.org/10.1021/ja2096859.

(118) Quintyn, R. S.; Zhou, M.; Yan, J.; Wysocki, V. H. Surface-Induced Dissociation Mass Spectra as a Tool for Distinguishing Different Structural Forms of Gas-Phase Multimeric Protein Complexes. *Anal. Chem.* **2015**, *87* (23), 11879–11886. https://doi.org/10.1021/acs.analchem.5b03441.

(119) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical Solutions: Principles and Application of Electron-Based Dissociation in Mass Spectrometry-Based Analysis of Protein Structure. *Mass Spectrom. Rev.* **2018**, *37* (6), 750–771. https://doi.org/10.1002/mas.21560.

(120) 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.

(121) Qi, Y.; Volmer, D. A. Electron-Based Fragmentation Methods in Mass Spectrometry: An Overview. *Mass Spectrometry Reviews*. 2017. https://doi.org/10.1002/mas.21482.

(122) Riley, N. M.; Westphall, M. S.; Coon, J. J. Sequencing Larger Intact Proteins (30-70 KDa) with Activated Ion Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (1), 140–149. https://doi.org/10.1007/s13361-017-1808-7.

(123) Riley, N. M.; Sikora, J. W.; Seckler, H. S.; Greer, J. B.; Fellers, R. T.; LeDuc, R. D.; Westphall, M. S.; Thomas, P. M.; Kelleher, N. L.; Coon, J. J. The Value of Activated Ion Electron Transfer Dissociation for High-Throughput Top-Down Characterization of Intact Proteins. *Anal. Chem.* **2018**, *90* (14), 8553–8560. https://doi.org/10.1021/acs.analchem.8b01638.

(124) O'Connor, P. B.; Speir, J. P.; Senko, M. W.; Little, D. P.; McLafferty, F. W. Tandem Mass Specatrometry of Carbonic Anhydrase (29 KDa). *J. Mass*

Spectrom. **1995**, *30* (1), 88–93. https://doi.org/10.1002/jms.1190300114.

(125) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Top-down Mass Spectrometry of a 29-KDa Protein for Characterization of Any Posttranslational Modification to within One Residue. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (4), 1774–1779. https://doi.org/10.1073/pnas.251691898.

(126) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651. https://doi.org/10.1021/ja4029654.

(127) Hughes, S. Investigating Protein Structure by Means of Mass Spectrometry, University of Edinburgh, 2019.

(128) Köster, C. Mass Spectrometrty Method for Accurate Mass Determination of Unknown Ions. *US Patients 6,188,064 B1* **2001**.

(129) Smith, D. F.; Kharchenko, A.; Konijnenburg, M.; Klinkert, I.; Paša-Tolić, L.; Heeren, R. M. A. Advanced Mass Calibration and Visualization for FT-ICR Mass Spectrometry Imaging. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (11), 1865–1872. https://doi.org/10.1007/s13361-012-0464-1.

(130) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down Proteomics: Facts and Perspectives. *Biochemical and Biophysical Research Communications*. 2014. https://doi.org/10.1016/j.bbrc.2014.02.041.

(131) Cao, X.; Nesvizhskii, A. I. Improved Sequence Tag Generation Method for Peptide Identification in Tandem Mass Spectrometry. *J. Proteome Res.* **2008**, *7* (10), 4422–4434. https://doi.org/10.1021/pr800400q.

(132) Valkenborg, D.; Mertens, I.; Lemière, F.; Witters, E.; Burzykowski, T. The Isotopic Distribution Conundrum. *Mass Spectrom. Rev.* **2012**, *31* (1), 96–109. https://doi.org/10.1002/mas.20339.

(133) Breuker, K.; Jin, M.; Han, X.; Jiang, H.; McLafferty, F. W. Top-Down Identification and Characterization of Biomolecules by Mass Spectrometry. *J Am Soc Mass Spectrom* **2008**, *19* (8), 1045–1053. https://doi.org/10.1016/j.jasms.2008.05.013.

(134) Cai, W.; Tucholski, T.; Chen, B.; Alpert, A. J.; Mcilwain, S.; Kohmoto, T.; Jin, S.; Ge, Y. Top–down Proteomics of Large Proteins up to 223 KDa Enabled by Serial Size Exclusion Chromatography Strategy. **2018**, *89* (10), 5467–5475. https://doi.org/10.1021/acs.analchem.7b00380.Top.

(135) Kellie, J. F.; Tran, J. C.; Lee, J. E.; Ahlf, D. R.; Thomas, H. M.; Ntai, I.; Catherman, A. D.; Durbin, K. R.; Zamdborg, L.; Vellaichamy, A.; Thomas, P. M.; Kelleher, N. L. The Emerging Process of Top Down Mass Spectrometry for Protein Analysis: Biomarkers, Protein-Therapeutics, and Achieving High Throughputw. *R. Soc. Chem.* **2010**, *6*, 1532–1539. https://doi.org/10.1039/c000896f.

(136) Vyatkina, K.; Wu, S.; Dekker, L. J. M.; VanDuijn, M. M.; Liu, X.; Tolić, N.; Luider, T. M.; Paša-Tolić, L.; Pevzner, P. A. Top-down Analysis of Protein Samples by de Novo Sequencing Techniques. *Bioinformatics* **2016**, *32* (18), 2753–2759. https://doi.org/10.1093/bioinformatics/btw307.

(137) Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A. An Integrated Native Mass Spectrometry and Top-down Proteomics Method That Connects Sequence to Structure and Function of Macromolecular Complexes. *Nat. Chem.* **2018**, *10*, 139–148.

https://doi.org/10.1038/nchem.2908.

(138) Köcher, T.; Superti-Furga, G. Mass Spectrometry-Based Functional Proteomics: From Molecular Machines to Protein Networks. *Nat. Methods* **2007**. https://doi.org/10.1038/nmeth1093.

(139) Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science (80-.).* **2006**, *312*, 212–217. https://doi.org/10.1126/science.1124619.

(140) Specht, H.; Slavov, N. Transformative Opportunities for Single-Cell Proteomics. *Journal of Proteome Research*. American Chemical Society August 3, 2018, pp 2565–2571. https://doi.org/10.1021/acs.jproteome.8b00257.

(141) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

(142) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10* (4), 1794–1805. https://doi.org/10.1021/pr101065j.

(143) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 5–13. https://doi.org/10.1007/s13361-016-1545-3.

(144) Breuker, K.; Mclafferty, F. W. The Thermal Unfolding of Native Cytochrome c in the Transition from Solution to Gas Phase Probed by Native Electron

Capture Dissociation. **2005**, 4911–4914. https://doi.org/10.1002/anie.200500668.

(145) Zhang, J.; Reza Malmirchegini, G.; Clubb, R. T. C.; Loo, J. Native Top-down Mass Spectrometry for the Structural Characterization of Human Hemoglobin. *Eur. J. Mass Spectrom.* **2015**, *21* (3), 221. https://doi.org/10.1255/ejms.1340.

(146) Jones, C. M.; Beardsley, R. L.; Galhena, A. S.; Dagan, S.; Cheng, G.; Wysocki, V. H. Symmetrical Gas-Phase Dissociation of Noncovalent Protein Complexes via Surface Collisions. *J. Am. Chem. Soc.* **2006**, *128* (47), 15044–15045. https://doi.org/10.1021/ja064586m.

(147) Wongkongkathep, P.; Han, J. Y.; Choi, T. S.; Yin, S.; Kim, H. I.; Loo, J. A. Native Top-Down Mass Spectrometry and Ion Mobility MS for Characterizing the Cobalt and Manganese Metal Binding of α -Synuclein Protein. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1870–1880. https://doi.org/10.1007/s13361-018-2002-2.

(148) Chapman, H. N.; Fromme, P.; Barty, A.; White, T. A.; Kirian, R. A.; Aquila, A.; Hunter, M. S.; Schulz, J.; Deponte, D. P.; Weierstall, U.; Doak, R. B.; Maia, F. R. N. C.; Martin, A. V; Schlichting, I.; Lomb, L.; Coppola, N.; Shoeman, R. L.; Epp, S. W.; Hartmann, R.; Rolles, D.; Reich, C.; Pietschner, D.; Timneanu, N.; Seibert, M. M.; Andreasson, J.; Rocker, A.; Jo, O.; Krasniqi, F.; Bott, M.; Schmidt, K. E.; Wang, X.; Grotjohann, I.; Holton, J. M.; Barends, T. R. M.; Neutze, R.; Marchesini, S.; Fromme, R. Femtosecond X-Ray Protein Nanocrystallography. *Nature* **2011**, *470*, 73–78. https://doi.org/10.1038/nature09750.

(149) Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; Wan,

W.; Stubbs, G.; Schwieters, C. D.; Lee, V. M. Y.; George, J. M.; Rienstra, C. M. Solid-State NMR Structure of a Pathogenic Fibril of Full-Length Human α -Synuclein. *Nat. Struct. Mol. Biol.* **2016**, *23* (5), 409–417. https://doi.org/10.1038/nsmb.3194.

(150) Milne, J. L. S.; Borgnia, M. J.; Bartesaghi, A.; Tran, E. E. H.; Earl, L. A.; Schauder, D. M.; Lengyel, J.; Pierson, J.; Patwardhan, A.; Subramaniam, S. Cryo-Electron Microscopy - A Primer for the Non-Microscopist. *FEBS J.* **2013**, *280* (1), 28–45. https://doi.org/10.1111/febs.12078.

(151) Saito, R.; Sato, T.; Ikai, A.; Tanaka, N. Structure of Bovine Carbonic Anhydrase II at 1.95 Å Resolution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (4), 792–795. https://doi.org/10.1107/S0907444904003166.

(152) Woods, L. A.; Dolezal, O.; Ren, B.; Ryan, J. H.; Peat, T. S.; Poulsen, S.-A. Native State Mass Spectrometry, Surface Plasmon Resonance, and X-Ray Crystallography Correlate Strongly as a Fragment Screening Combination. *J. Med. Chem.* **2016**, *59* (5), 2192. https://doi.org/10.1021/acs.jmedchem.5b01940.

(153) Sanders, J. D.; Mullen, C.; Watts, E.; Holden, D. D.; Syka, J. E. P.; Schwartz, J. C.; Brodbelt, J. S. Enhanced Sequence Coverage of Large Proteins by Combining Ultraviolet Photodissociation with Proton Transfer Reactions. *Anal. Chem.* **2019**. https://doi.org/10.1021/acs.analchem.9b04026.

(154) Lindskog, S. Structure and Mechanism of Carbonic Anhydrase. *Pharmacol. Ther.* **1997**, *74* (1), 1–20. https://doi.org/10.1016/S0163-7258(96)00198-2.

(155) Bradley, J. M.; Moore, G. R.; Le Brun, N. E.; Brun, L. Diversity of Fe 2+ Entry and Oxidation in Ferritins. *Curr. Opin. Chem. Biol.* **2017**, *37*, 122–128. https://doi.org/10.1016/j.cbpa.2017.02.027.

(156) Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M. Design of Functional Metalloproteins. *Nature*. August 13, 2009, pp 855–862. https://doi.org/10.1038/nature08304.

(157) Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L. Native Electrospray and Electron-Capture Dissociation FTICR Mass Spectrometry for Top-down Studies of Protein Assemblies. *Anal. Chem.* **2011**, *83*, 5598–5606. https://doi.org/10.1021/ac200695d.

(158) Lindskog, S.; Malmstrom, B. G. Metal Binding and Catalytic Activity in Bovine Carbonic Anhydrase. *J. Biol. Chem.* **1962**, *237* (4), 1129–1137.

(159) Tabbì, G.; Magrì, A.; Rizzarelli, E. The Copper(II) Binding Centres of Carbonic Anhydrase Are Differently Affected by Reductants That Ensure the Redox Intracellular Environment. *J. Inorg. Biochem.* **2019**. https://doi.org/10.1016/j.jinorgbio.2019.110759.

(160) Piergentili, C.; Ross, J.; He, D.; Gallagher, K. J.; Stanley, W. A.; Adam, L.; Mackay, C. L.; 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**, *44* (0), jbc.RA120.014502. https://doi.org/10.1074/jbc.ra120.014502.

(161) Johnson, J. E. Virus Particle Maturation: Insights into Elegantly Programmed Nanomachines. *Curr. Opin. Struct. Biol.* **2010**, *20* (2), 210–216. https://doi.org/10.1016/j.sbi.2010.01.004.

(162) Ilag, L. L.; Videler, H.; McKay, A. R.; Sobott, F.; Fucini, P.; Nierhaus, K. H.; Robinson, C. V. Heptameric (L12)6/L10 Rather than Canonical Pentameric Complexes Are Found by Tandem MS of Intact Ribosomes from Thermophilic Bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (23), 8192–8197.

https://doi.org/10.1073/pnas.0502193102.

(163) Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V. Micelles Protect Membrane Complexes from Solution to Vacuum. *Science (80-.).* **2008**, *321* (5886), 243–246. https://doi.org/10.1126/science.1159292.

(164) Snijder, J.; Rose, R. J.; Veesler, D.; Johnson, J. E.; Heck, A. J. R. Studying 18 MDa Virus Assemblies with Native Mass Spectrometry. *Angew. Chemie - Int. Ed.* **2013**. https://doi.org/10.1002/anie.201210197.

(165) Snijder, J.; Van De Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.; Agbandje-Mckenna, M.; Makarov, A.; Heck, A. J. R. Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J. Am. Chem. Soc.* **2014**. https://doi.org/10.1021/ja502616y.

(166) Snijder, J.; Kononova, O.; Barbu, I. M.; Uetrecht, C.; Rurup, W. F.; Burnley, R. J.; Koay, M. S. T.; Cornelissen, J. J. L. M.; Roos, W. H.; Barsegov, V.; Wuite, G. J. L.; Heck, A. J. R. Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded Bacterial Nanocompartment Encapsulin. *Biomacromolecules* **2016**. https://doi.org/10.1021/acs.biomac.6b00469.

(167) Giessen, T. W. Encapsulins: Microbial Nanocompartments with Applications in Biomedicine, Nanobiotechnology and Materials Science. *Current Opinion in Chemical Biology*. 2016. https://doi.org/10.1016/j.cbpa.2016.05.013.

Appendices

Appendix 1 Protein Samples

Encapsulated Ferritin (EncFtn) Amino Acid Sequence
AQSSNSTHEPLEVLKEETVNRHRAIVSVMEELEAVDWYDQRVDASTDPELTAILAH
NRDEEKEH
AAMTLEWLRRNDAKWAEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH

Encapsulin (Enc) Amino Acid Sequence
MDLLKRHLAPIVPDAWSAIDEEAKEIFQGHLAGRKLVDFRGPFGWEYAAVNTGELR
PIDDTPEDVDMKLRQVQPLAEVRVPFTLDVTELDSVARGATNPDLDDVARAAERM
VEAEDSAIFHGWAQAGIKGIVDSTPHEALAVASVSDFPRAVLSAADTLRKAGVTGP
YALVLGPKAYDDLFAATQDGYPVAKQVQRLVVDGPLVRANALAGALVMSMRGGD
YELTVGQDLSIGYAFHDRSKVELFVAESFTFRVLEPGAAVHLRYAR

Bovine Carbonic Anhydrase II (CA) Amino Acid Sequence
GAMASHHWGYGKHNGPEHWHKDFPIANGERQSPVDIDTKAVVQDPALKPLALVY
GEATSRRMVNNGHSFNVEYDDSQDKAVLKDGPLTGTYRLVQFHFHWGSSDDQG
SEHTVDRKKYAAELHLVHWNTKYGDFGTAAQQPDGLAVVGVFLKVGDANPALQK
VLDALDSIKTKGKSTDFPNFDPGSLLPNVLDYWTYPGSLTTPPLLESVTWIVLKEPIS
VSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK

Serine Palmitoyltransferase (SPT) Amino Acid Sequence
GSSHHHHHHSSGLVPRGSHMASTEAAAQPHALPADAPDIAPERDLLSKFDGLIAE
RQKLLDSGVTDPFAIVMEQVKSPTEAVIRGKDTILLGTYNYMGMTFDPDVIAAGKEA
LEKFGSGTNGSRMLNGTFHDHMEVEQALRDFYGTTGAIVFSTGYMANLGIISTLAG
KGEYVILDADSHASIYDGCQQGNAEIVRFRHNSVEDLDKRLGRLPKEPAKLVVLEG
VYSMLGDIAPLKEMVAVAKKHGAMVLVDEAHSMGFFGPNGRGVYEAQGLEGQID
FVVGTFSKSVGTVGGFVVSNHPKFEAVRLACRPYIFTASLPPSVVATATTSIRKLMT
AHEKRERLWSNARALHGGLKAMGFRLGTETCDSAIVAVMLEDQEQAAMMWQALL
DGGLYVNMARPPATPAGTFLLRCSICAEHTPAQIQTVLGMFQAAGRAVGVIG

Appendix 2. Charge and error calculations of Enc

The charge state determination of the average mass of Enc was carried out using the theoretical average mass of the isotopically depleted Enc 60-mer at 1737240.52 Da ($(C_{1296}H_{2032}N_{356}O_{388}S_5)_{60}$). The *m/z* of each 'smoothed' and processed charged states were taken from the centroid *m/z* in DataAnalysis. The standard deviation was calculated using the standard deviation formula in Microsoft excel for all charge states and focusing on charge state 126 to 136. The average mass was calculated using all charge states and compared to the calculated theoretical depleted average to calculate percentage error.

m/z	z		Mass	FWHM
14256.92	122		1739222.77	46.69
14147.19	123		1739981.22	49.10
14023.09	124		1738739.63	40.20
13901.66	125		1737582.07	45.61
13794.67	126		1738002.03	47.86
13678.35	127		1737023.58	43.70
13571.15	128		1736979.12	47.03
13459.76	129		1736180.37	41.64
13352.32	130		1735670.99	42.20
13253.22	131		1736040.98	46.39
13148.76	132		1735503.76	43.25
13048.52	133		1735320.55	39.82
12951.72	134		1735397.08	39.22
12854.96	135		1735284.94	35.57
12768.54	136		1736386.06	36.32
12674.56	137		1736277.46	25.46
12580.51	138		1735972.13	29.37
12489.50	139		1735900.83	21.65
		Average mass	1736748.09	40.06
		SD	1377.35	
		SD_z_126 to136	831.08	42.09
		% Error	-0.028	

Appendix 3. Datastore location

All datasets used in this thesis are available on the University of Edinburgh Datastore.

Each dataset and its location are listed below.

Chapter 3

Development of an efficient method to produce isotopically depleted protein in M9 minimal media.

– Datastore: https://doi.org/10.7488/ds/2902.

-Q-TOF LCMS of isotopically natural EncFtn
-Q-TOF LCMS of isotopically natural BCA
-Q-TOF LCMS of isotopically natural SPT

-EncFtn with NITROGEN single depletion expression
-EncFtn with CARBON single depletion expression

-EncFtn M30L M68L mutant (FTMS LCMS file)
-EncFtn digest ion isolated at 453 for fine structure

Intact Mass- Isolated charge states - Datastore: https://doi.org/10.7488/ds/2446

- EncFtn (depleted) [M+14H]14+ isolated charge spectrum
- EncFtn (natural) [M+14H]14+ isolated charge spectrum
- CA (depleted) [M+31H]31+ isolated charge spectrum
- CA (natural) [M+31H]31+ isolated charge spectrum
- SPT (depleted) [M+40H]40+ isolated charge spectrum
- SPT (natural) [M+40H]40+ isolated charge spectrum (LCMS data)

Chapter 4

Isotopic depletion and the improvement for top-down fragmentation analysis.

ECD fragmentation of EncFtn - Datastore: https://doi.org/10.7488/ds/2446

- EncFtn (depleted) [M+16H]16+ ECD fragmentation 150 scans
- EncFtn (depleted) [M+16H]16+ ECD fragmentation 20 scans
- EncFtn (depleted) [M+16H]16+ ECD fragmentation 5 scans
- EncFtn (natural) [M+16H]16+ ECD fragmentation 150 scans
- EncFtn (natural) [M+16H]16+ ECD fragmentation 20 scans
- EncFtn (natural) [M+16H]16+ ECD fragmentation 5 scans
- EncFtn (depleted) [M+14H]14+ ECD fragmentation 200 scans
- EncFtn (natural) [M+14H]14+ ECD fragmentation 200 scans

ECD fragmentation of CA- Datastore: https://doi.org/10.7488/ds/2446

- CA (depleted) [M+22H]22+ ECD fragmentation 300 scans
- CA (depleted) [M+22H]22+ ECD fragmentation 20 scans
- CA (depleted) [M+22H]22+ ECD fragmentation 5 scans
- CA (natural) [M+22H]22+ ECD fragmentation 300 scans
- CA (natural) [M+22H]22+ ECD fragmentation 20 scans
- CA (natural) [M+22H]22+ ECD fragmentation 5 scans
- CA (depleted) [M+32H]32+ ECD fragmentation 200 scans
- CA (natural) [M+32H]32+ ECD fragmentation 200 scans

Protein charge state distributions- Datastore: https://doi.org/10.7488/ds/2446

- EncFtn (depleted) charge state distribution
- EncFtn (natural) charge state distribution
- CA (depleted) charge state distribution
- CA (natural) charge state distribution

CID fragmentation of EncFtn- Datastore: https://doi.org/10.7488/ds/2446

- EncFtn (depleted) [M+16H]16+ isolated charge spectrum
- EncFtn (natural) [M+16H]16+ isolated charge spectrum

Chapter 5

Isotope Depletion for improved top-down sequence coverage of larger molecular weight proteins.

MS^2 ETD fragmentation of EncFtn, BCA and SPT

–Datastore: https://doi.org/10.7488/ds/2902.

-EncFtn (depleted) [M+16H]16+ ETD fragmentation 100 scans (with FAIMS -30V)
-EncFtn (natural) [M+16H]16+ ETD fragmentation 100 scans (with FAIMS -30V)
-BCA (depleted) [M+31H]31+ ETD fragmentation 100 scans
-BCA (natural) [M+31H]31+ ETD fragmentation 100 scans
-SPT (depleted) [M+37H]37+ ETD fragmentation 100 scans
-SPT (natural) [M+37H]37+ ETD fragmentation 100 scans

MS^3 using ETD of SPT – Datastore: https://doi.org/10.7488/ds/2902.

-SPT (depleted) [M+37H]37+ ETD fragmentation plus 0p1ms PTCR 1200 centred w 120 100 scans
-SPT (depleted) [M+37H]37+ ETD fragmentation plus 5ms PTCR 800 centred w 300 100 scans
-SPT (depleted) [M+37H]37+ ETD fragmentation plus 5ms PTCR 1050 centred w 300 100 scans
-SPT (depleted) [M+37H]37+ ETD fragmentation plus 5ms PTCR 1200 centred w 120 100 scans

-SPT (depleted) [M+37H]37+ ETD fragmentation plus 10ms PTCR 1050 centred w 300 100 scans
-SPT (depleted) [M+37H]37+ ETD fragmentation plus 10ms PTCR 1200 centred w 120 100 scans

-SPT (natural) [M+37H]37+ ETD fragmentation plus 0p1ms PTCR 1200 centred w 120 100 scans
-SPT (natural) [M+37H]37+ ETD fragmentation plus 5ms PTCR 800 centred w 300 100 scans
-SPT (natural) [M+37H]37+ ETD fragmentation plus 5ms PTCR 1050 centred w 300 100 scans
-SPT (natural) [M+37H]37+ ETD fragmentation plus 5ms PTCR 1200 centred w 120 100 scans
-SPT (natural) [M+37H]37+ ETD fragmentation plus 10ms PTCR 1050 centred w 300 100 scans
-SPT (natural) [M+37H]37+ ETD fragmentation plus 10ms PTCR 1200 centred w 120 100 scans

Chapter 6

Isotope Depletion for Improved Native Protein Mass Spectrometry.

– Datastore: https://doi.org/10.7488/ds/2902.

-BCA (depleted) Native without metal ion
-BCA (natural) Native without metal ion
-BCA (depleted) Native with zinc ion bound
-BCA (natural) Native with zinc ion bound
-BCA (depleted) Native with copper ion bound
-BCA (natural) Native with copper ion bound
-SPT monomer (depleted) [M+16H]16+ plus alpha-N-phosphogluconolylation modification
-SPT dimer (depleted) [M+20H]20+ plus alpha-N-phosphogluconolylation modification
-Enc (depleted) 60-mer charge state distribution
-Enc (natural) 60-mer charge state distribution

Appendix 4. Isotope Depletion Mass Spectrometry (ID-MS) for Accurate Mass Determination and Improved Top-Down Sequence Coverage of Intact Proteins.

Almost all the data included in chapter 3 and all of chapter 4 have previously been included into a manuscript which was published in JASMS in January 2020 (10.1021/jasms.9b00119), manuscript included on next page.

ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Research Article

Isotope Depletion Mass Spectrometry (ID-MS) for Accurate Mass Determination and Improved Top-Down Sequence Coverage of Intact Proteins

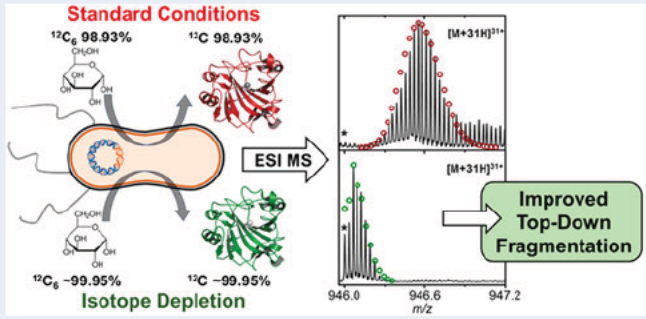
Kelly J. Gallagher, Michael Palasser, Sam Hughes, C. Logan Mackay, David P. A. Kilgour, and David J. Clarke*

Cite This: *J. Am. Soc. Mass Spectrom.* 2020, 31, 700−710

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Top down mass spectrometry (MS) is an increas ingly important technique for protein characterization. However, in many biological MS experiments, the practicality of applying top down methodologies is still limited at higher molecular mass. In large part, this is due to the detrimental effect resulting from the partitioning of the mass spectral signal into an increasing number of isotopic peaks as molecular mass increases. Reducing the isotopologue distribution of proteins via depletion of heavy stable isotopes was first reported over 20 years ago (Marshall, A. G.; Senko, M. W.; Li, W.; Li, M.; Dillon, S., Guan, S.; Logan, T. M.. Protein Molecular Mass to 1 Da by ^{13}C, ^{15}N Double Depletion and FT ICR Mass Spectrometry. *J. Am. Chem. Soc.* **1997**, *119*, 433−434.) and has been demonstrated for several small proteins. Here we extend this approach, introducing a new highly efficient method for the production of recombinant proteins depleted in ^{13}C and ^{15}N and demonstrating its advantages for top down analysis of larger proteins (up to ∼50 kDa). FT ICR MS of isotopically depleted proteins reveals dramatically reduced isotope distributions with monoisotopic signal observed up to 50 kDa. In top down fragmentation experiments, the reduced spectral complexity alleviates fragment ion signal overlap, the presence of monoisotopic signals allows assignment with higher mass accuracy, and the dramatic increase in signal to noise ratio (up to 7 fold) permits vastly reduced acquisition times. These compounding benefits allow the assignment of ∼3 fold more fragment ions than comparable analyses of proteins with natural isotopic abundances. Finally, we demonstrate greatly increased sequence coverage in time limited top down experiments—highlighting advantages for top down LC−MS/MS workflows and top down proteomics.



KEYWORDS: *isotope depletion, top down fragmentation, intact protein, FT ICR MS, electron capture dissociation*

Downloaded via UNIV OF EDINBURGH on June 25, 2020 at 08:51:05 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

■ INTRODUCTION

In the last two decades, top down mass spectrometry (MS) has become an established technique for the analysis of protein sequence and the detailed characterization of post translational and chemical protein modifications.[1,2] More recently, the application of top down fragmentation techniques, in con junction with native mass spectrometry, has also emerged as a potentially valuable tool for studying the structure of proteins and protein complexes.[3,4] Indeed, the application of the top down approach has many potential benefits for a wide range of structural MS workflows, including hydrogen−deuterium exchange MS,[5,6] cross linking MS,[7] and covalent labeling MS.[8,9] Although, recent advances in instrumentation and fragmentation techniques now allow comprehensive sequence coverage to be obtained for small proteins (<20 kDa),[10−12] top down MS suffers from a rapid drop off in fragmentation efficacy as protein mass increases beyond this mass.[13−15] Therefore, strategies for increasing the efficiency of top down techniques are required to facilitate the widespread adoption of the top down philosophy in all areas of biological MS.

One fundamental factor, which proves severely detrimental in this regard, is the increasing breadth of the isotopic distribution that accompanies increasing molecular mass.[16] The dispersal of the ion signal over more isotopologues reduces the signal to noise ratio (S/N), increases spectral complexity, and results in the overlapping of signals for species which are close in mass (e.g., proteoforms of the same protein with similar masses). In addition, proteins, or protein fragment ions, over ∼10 kDa, commonly do not display a monoisotopic signal of sufficient ion abundance to accurately assign. In the

Received: December 18, 2019
Revised: January 31, 2020
Accepted: January 31, 2020
Published: January 31, 2020

ACS Publications

© 2020 American Society for Mass Spectrometry. Published by the American Chemical Society. All rights reserved.

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

context of a top down fragmentation experiment, these compounding difficulties all reduce the number of fragment ions that can be confidently assigned. One solution to this problem is to produce isotopically depleted proteins, which display less disperse isotope distributions. The feasibility of this strategy was demonstrated by Marshall et al. in 1997 who reported the production and intact mass analysis of a 12 kDa protein depleted in ^{13}C and ^{15}N isotopes.[17] Since the publication of this report over 20 years ago, the isotope depletion strategy has been applied to a handful of studies,[18−21] all of which highlight the associated benefits of isotopic depletion to protein MS.[13,22] It is surprising therefore that this this strategy has not been widely adopted; this is perhaps due to the technical challenge in producing isotopically depleted protein samples.

Herein we detail a new method for the production of isotopically depleted proteins, by recombinant expression in *E. coli*, using isotopically controlled culture media. Our new methodology allows depletion of ^{13}C and ^{15}N stable isotopes by restricting the C and N sources to ^{12}C (∼99.9%) glucose and ^{14}N (∼99.99%) ammonium sulfate during protein expression. We demonstrate the benefits of this approach to top down protein analysis by producing and analyzing a series of proteins up to ∼50 kDa. All isotopically depleted proteins displayed dramatically simplified isotope distributions, and as a consequence, we report a reduction in mass spectral complexity and dramatic S/N increases. Using this isotope depletion mass spectrometry (ID MS) strategy, in combination with top down electron capture dissociation (ECD), allows assignment of fragment ions with increased confidence, results in dramatically improved sequence coverage, and allows shorter acquisition times. Thus, we demonstrate that ID MS is a powerful strategy for increasing fragmentation coverage in online top down proteomic workflows.[21,23] Furthermore, incorporating the ID MS strategy into other structural MS workflows, such as native top down and hydrogen/deuterium exchange MS, will extend the feasibility of implementing top down methodologies in biological MS studies.

EXPERIMENTAL SECTION

Protein Samples. Three well characterized recombinant proteins were chosen for this study—encapsulated ferritin from *Rhodospirillum rubrum* (EncFtn, 13.2 kDa),[24] bovine carbonic anhydrase (CA, 29.3 kDa),[25] and serine palmitoyl transferase from *Sphingomonas paucimobilis* (SPT, 47.3 kDa).[26]

Molecular Biology. Details of expression plasmids and preparation of single point mutants can be found in Supporting Information.

Protein Expression. Protein expression was performed using *E. coli* BL21 (DE3) cells transformed with the required expression plasmid. A single colony was used to inoculate 10 mL of LB media supplemented with the appropriate antibiotic, before overnight incubation at 37 °C. This was then used to inoculate 500 mL of 2× YT media, incubated at 37 °C until an OD_{600} of 0.6−0.8 was obtained, at which point the cell culture was centrifuged at 5000*g* at 4 °C for 20 min (200 mL per expression culture required). The resultant pellet was washed with 20 mL/gram of pellet 5× M9 salts solution without ammonium sulfate nitrogen source (33.9 g/L Na_2HPO_4, 15 g/L KH_2PO_4, 2.5 g/L NaCl). After washing, the pellet was resuspended in 100 mL of M9 minimal media. Isotopically depleted M9 minimal media was supplemented with ^{12}C (99.9%) glucose (Cambridge Isotope Laboratories) and ^{14}N (99.99%) ammonium sulfate, as the sole carbon and nitrogen sources. The M9 minimal media cultures were further incubated at 37 °C for 1 h, before protein expression was induced with 1 or 0.1 mM IPTG as required and incubated overnight at 18 °C.[27] After incubation, cultures were centrifuged at 5000*g* at 4 °C for 30 min, and pellets were stored at −80 °C until required.

Protein Purification. The purification for both EncFtn and SPT have been modified from previously published protocols.[24,26] Carbonic anhydrase cell pellets were resuspended in lysis buffer (PBS with 10 mM imidazole, pH 7.4), and cells were disrupted via sonication for 18 cycles of 10 s bursts at 10 μm amplitude. The cell lysate was clarified by centrifugation at 12 000*g* for 1 h at 4 °C and filtered using a 0.22 μm syringe filter (Millipore, UK).

The filtered cell lysate was loaded onto a 1 mL HisTrap column (GE Healthcare, UK), pre equilibrated with lysis buffer. Unbound protein was washed out with 5 CV of lysis buffer, and protein was eluted using a linear gradient between 0 and 100% of elution buffer (PBS and 300 mM imidazole). Hexahisitdine tag cleavage was carried out by incubating protein with TEV protease overnight at 4 °C. Noncleaved protein and cleaved tags were removed by diluting the protein sample 1:2 with lysis buffer, reloading it onto a 1 mL HisTrap column, and collecting the flowthrough. The collected flowthrough was concentrated to a standard concentration of 10 μM using a Vivaspin centrifugal concentrator (5000 MWCO) and exchanged into a storage buffer (50 mM Tris HCl, 10% glycerol, pH 7.4) using a desalting column (GE Healthcare UK) and stored at −80 °C. The purification of isotopically double depleted protein was kept separate from protein with natural isotopic abundance to prevent cross contamination.

FT-ICR Mass Spectrometry. Prior to mass spectrometry, intact protein samples were desalted using C4 reverse phase Bond Elut OMIX pipet tips (Agilent Technologies, Santa Clara, CA), with final elution in 50:50 water/acetonitrile with 0.1% formic acid. Protein samples were ionized by nano electrospray (nESI) using a Nanomate infusion robot (Advion Biosciences, Ithaca, NY) at a typical concentration of 5 μM. MS analyses were performed on a SolariX FT ICR instrument equipped with an Infinity ICR cell and a 12 T magnet (Bruker Daltonics, Bremen, Germany). For intact mass analysis, spectra were acquired between m/z 500 and 5000, to yield a broadband 1 or 2 megaword (MW) time domain transient. Ion accumulation was set to between 50 and 200 ms, and typically, each spectrum was the sum of 50 acquisitions. CASI (continuous accumulation of selected ions) was used to improve the isotope distribution profile obtained for SPT (quadrupole width: 100 m/z).

For top down mass spectrometry, a specific protein charge state was isolated using the mass resolving quadrupole, and ion accumulation time was optimized to accumulate ∼10^8 ions in the FT ICR cell per fill. Collision induced dissociation (CID) was performed external to the ICR cell with a typical collision voltage of between 20 and 35 V. Electron capture dissociation (ECD) was achieved by irradiating with electrons using a heated hollow dispenser cathode. Typically, cathode conditions were a bias voltage of 1.5 V, lens voltage of 15 V, and pulse length of between 5 and 30 ms. Top down spectra were recorded between m/z 300 to 5000 and were the sum of 150, 20, or 5 1 MW time domain transients (EncFtn) or 300, 20, or 5 (CA) transients as stated. For online top down fragmenta

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

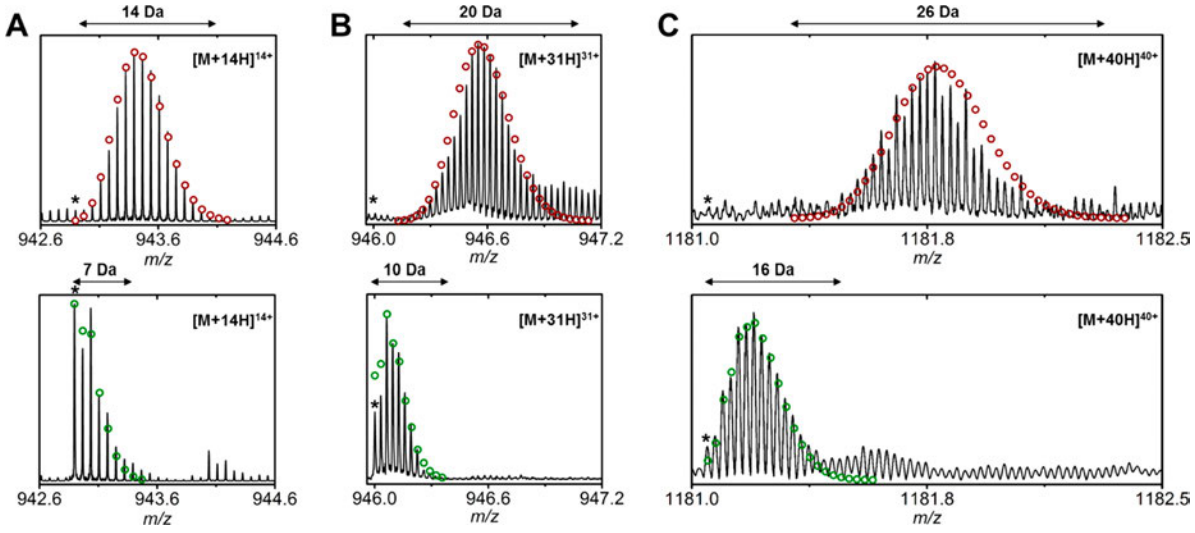


Figure 1. High resolution ESI FT ICR MS analysis of the isotope distributions of proteins used in this study (magnitude mode data shown). (A) EncFtn (13.2 kDa); (B) CA (29.3 kDa); (C) SPT (47.3 kDa). (Top) The observed isotope distribution for samples prepared from natural isotope abundance cell culture, and (bottom) the observed isotopic distribution for samples prepared from isotopically depleted cell culture. The theoretical isotopic distributions are overlaid on the spectra as scatter plots (natural abundance: 98.89% ^{12}C, 99.63% ^{14}N; isotopically depleted abundances: 99.90% ^{12}C, 99.99% ^{14}N). In each spectrum, the monoisotopic species is highlighted with an asterisk (*).

tion, protein samples were separated on an ACE 3 C4 column (75 × 0.5 mm) and eluted using a 5 to 95% gradient of acetonitrile over 30 min. Individual charge states were isolated in the mass resolving quadrupole, and fragmentation was achieved using ECD in the ICR cell. The cathode conditions were optimized to achieve the maximum fragment ion yield.

Top-Down Data Analysis. FT ICR MS data was postprocessed, in absorption mode, using a specially adapted version of AutoVectis (Spectroswiss Sàrl, Lausanne, Switzerland); for a detailed description, see the Supporting Information. For natural isotopic abundance proteins, top down data was also analyzed using a "poly averagine" method (see Results and Discussion). The sophisticated numerical annotation procedure (SNAP; Bruker Daltonics) was used to deconvolute the monoisotopic mass of fragment ions. For deconvolution, the S/N cut off was set to 3, and the quality factor threshold was set to 0.3. A single point internal calibration was performed using the theoretical m/z of a consistent fragment ion (typically c_{20} for EncFtn data and c_{34} for CA data). The resulting calibrated monoisotopic peak list was search against the protein sequence using Prosight Lite.[28]

RESULTS AND DISCUSSION

Efficient Production of Isotopically Depleted Proteins in *E. coli*. All proteins were recombinantly expressed in *E. coli* using M9 minimal growth media, containing isotopically depleted glucose (99.9% $^{12}C_6$) and ammonium sulfate (99.99% $^{14}N_2$) as the sole carbon and nitrogen sources (see Supporting Information and Figure S1). The expression protocol developed here does not rely on commercial preformulated expression media. Therefore, individual elemental isotope composition can be controlled by selection of the required feedstock. In this report, we have employed isotope depletion of carbon and nitrogen; however, this protocol could also be easily adapted for depletion of other elements, e.g., sulfur depletion in sulfur rich proteins by employing ^{33}S and ^{34}S depleted sulfate. In addition, we have developed a protocol that employs a starter culture in rich (LB) media. This allows the production of sufficient cell densities before the cells are transferred into defined (isotopically depleted) media and recombinant protein expression is induced. This strategy allows the efficient and economically viable production of isotopically depleted proteins. Using this method, the achieved yields ranged from 100 to 1000 μg of ^{13}C and ^{15}N depleted protein per 100 mL of minimal media culture at a cost of ∼ $100.

ID-MS Allows Direct Determination of Monoisotopic Mass of Intact Proteins up to 50 kDa. To date, isotope depletion techniques have been used to produce recombinant protein samples up to ∼20 kDa in molecular weight.[17−20,29] Here we extend this technique for the production of proteins up to ∼50 kDa using our efficient expression system outlined above. MS analysis of the purified proteins was performed using high resolution electrospray (ESI) Fourier transform ion cyclotron resonance mass spectrometry (FT ICR MS). ESI MS of samples prepared in natural abundance cell culture and double depleted cell culture exhibited identical charge state distributions (Figure S2). However, dramatically simplified isotope distributions were observed in the mass spectra of proteins produced in isotopically depleted media (Figure 1). For EncFtn (monoisotopic molecular mass 13 186.4 Da, Figure 1A), the width of the isotopic distribution decreased from 14 Da in the natural isotopic abundance protein to 7 Da in the isotopically depleted protein (isotopologues with abundance greater than 0.5% of the base peak), and the monoisotopic peak increased from ∼0.06% of the total signal (i.e., below the noise) to ∼30% of the total signal and was the highest peak in the distribution. For CA (29 294.8 Da) and SPT (47 201.8 Da), the monoisotopic peak was not visible in the natural isotopic abundance protein distribution. In contrast, the isotopically depleted proteins displayed an easily identifiable monoisotopic peak, accounting for ∼13 and ∼2% of the total signal for CA and SPT, respectively. A dramatic reduction in the widths of the isotopic distribution was also evident in CA and SPT. In all cases, the S/N also improved as the same number of ions were partitioned between fewer isotopologue peaks. The obtained isotope distributions of all

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

proteins studied was consistent with a reduction of ^{13}C to 0.05−0.10% and ^{15}N to <0.01%.

The ability to directly observe the monoisotopic isotopologue allows the unambiguous and immediate assignment of the accurate molecular mass of large intact proteins. In contrast, for accurate mass assignment of natural isotopic abundance proteins above ∼10 kDa, it is necessary to infer the monoisotopic mass by matching the observed isotope distribution with the calculated theoretical isotope distribution of an "average" protein, i.e., a repeating polymer of the model amino acid averagine.[30,31] This "poly averagine approximation" method relies on obtaining a statistically reliable experimental isotope distribution and often results in significant mass error (up to ∼3 ppm) and/or the misassignment of the monoisotopic mass by ±1 Da (or 2 Da), regardless of the resolution achieved in the data acquisition.[16,32]

As a consequence, confident detection of changes to protein primary structure that induce relatively small changes in mass (e.g., amidation/deamidation, disulfide bridge formation, single amino acid substitutions, etc.) is particularly challenging.[33−36] Therefore, the ID MS strategy may be particularity powerful for the detection and characterization of these low molecular mass PTMs. In order to demonstrate this, we produced an EncFtn "deamidated" single point variant (N58D), in an isotopically depleted form. MS analysis of isotopically depleted N58D EncFtn allowed confident detection of the deamidation at the protein level. Direct detection of protein deamidation was also possible from mixtures of WT and deamidated proteoforms (Supporting Information, Figure S3).

ID-MS Dramatically Improves Protein Sequence Coverage Achieved Using Top-Down Fragmentation. We then investigated the benefits of using the ID MS strategy in top down fragmentation experiments. Typically, top down fragmentation generates many hundreds of fragment ions, with each ion appearing in multiple charge states and exhibiting its own isotopic distribution.[37] Thus, the resulting spectra are highly complex and consist of many thousands/tens of thousands of individual peaks, over a wide dynamic range of ion abundance. As the observed fragment ions fall in a comparatively narrow m/z range (typically m/z 500−2000), fragment ion isotope distributions often overlap, and even with high resolving power instrumentation, superposition of peaks is common. Therefore, fragment ions can be overlooked or misassigned due to low signal and/or signal overlap.

In order to investigate the benefit of the isotope depletion strategy for top down mass spectrometry, we analyzed natural isotopic abundance and isotopically depleted proteins using both CID and ECD fragmentation. Initially, CID was performed on the $[M + 16H]^{16+}$ precursor ion of natural isotopic abundance and isotopically depleted 13 kDa EncFtn. Both fragmentation spectra were remarkably similar on initial inspection, displaying identical high abundance fragment ions at similar m/z (Figure S4A). However, all fragment ions derived from the isotopically depleted EncFtn exhibited reduced isotope distribution widths, which greatly reduce signal overlap of individual fragments. In addition, the S/N ratio displayed by isotopically depleted fragment ions was dramatically increased (for example, the complementary ion pairs b_{37}^{4+} and y_{78}^{10+} exhibit S/N gains of 7.0 fold and 4.7 fold in the isotopically depleted spectrum when compared to the natural isotopic abundance spectrum). It was also apparent that, for this 13 kDa isotopically depleted protein, the monoisotopic signal was the base peak (i.e., the highest signal) in the isotope distribution of every isotopically depleted fragment ion. This allowed direct determination of the accurate monoisotopic mass of every fragment ion (Figure S4C). Taken together, these three advantages led to confident assignment of substantially more CID product ions in the isotopically depleted EncFtn CID spectrum. Data analysis was performed in absorption mode, using AutoVectis (Supporting Information, Figure S5).[37,38] For CID of the $[M + 16H]^{16+}$ of EncFtn, 110 b and y fragment ions were assigned in the natural isotopic abundance spectrum (39 b ions, 71 y ions; 45.7% total sequence coverage); in comparison, 217 b and y fragment ions (84 b ions, 133 y ions; 64.7% total sequence coverage) were assigned in the depleted isotopic abundance spectrum (Supporting Information, Figure S6). This increase in the observed fragment ion number is similar to that demonstrated by Akashi et al.,[19] who reported a 63% increase in the number of assigned fragment ions when performing CID of an isotopically depleted version of the 10 kDa protein cystatin. However, we noted that for CID of both natural isotopic abundance and isotopically depleted EncFtn, the assigned b and y ions only constitute only around 20−30% of the total number of observed fragments, and even when employing an isotopically depleted strategy with top down CID, it is clear that there are regions of the protein with limited sequence coverage. Further analysis of the unassigned fragment ions in both CID spectra revealed a substantial number of internal fragments and widespread neutral loss during fragmentation ($-H_2O$, $-CO$, $-NH_3$). Taking these fragmentation channels into consideration allowed assignment of a total of 448 product ions (a, b, x, y, and y_{-H2O} ions; 82% total sequence coverage) in the CID spectrum of isotopically depleted EncFtn (Supporting Information, Figure S6).

The lack of product ion specificity and the biased nature of fragmentation with CID has been well documented, and this limits the utility of the technique for top down studies of proteins over 10−15 kDa.[39] In contrast to CID, electron driven dissociation techniques (such as ECD and ETD, together termed "ExD") are thought to result in relatively unbiased fragmentation throughout the protein sequence.[40−42] Thus, potentially higher sequence coverage has been reported (especially in larger proteins), and ExD fragmentation is a far more attractive technique for top down fragmentation as protein mass increases. However, one drawback of the ExD approach is its relatively inefficient precursor to product ion conversion, and so, ExD characteristically results in c and z type fragment ions of low ion abundance. Therefore, we reasoned that the substantial increased S/N evident in top down ID MS may potentially be of more benefit when used in conjunction with ExD studies.

Electron capture dissociation (ECD) was performed on isotopically natural and isotopically depleted EncFtn (Figure 2). The resulting fragmentation spectra were very similar—displaying identical high abundance ions at a similar m/z. However, all fragment ions resulting from isotopically depleted EncFtn displayed (i) dramatically increased monoisotopic signal, (ii) increased overall S/N, and (iii) reduced isotopic distribution widths, resulting in reduced ion distribution overlap (Figure 2B,C). Thus, ECD of the isotopically depleted protein allowed accurate assignment of side chain losses and revealed low abundance ions in "congested" regions of the spectrum (see Supporting Information, Figure S8). After peak assignment using Autovectis, 496 fragment ions were assigned

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

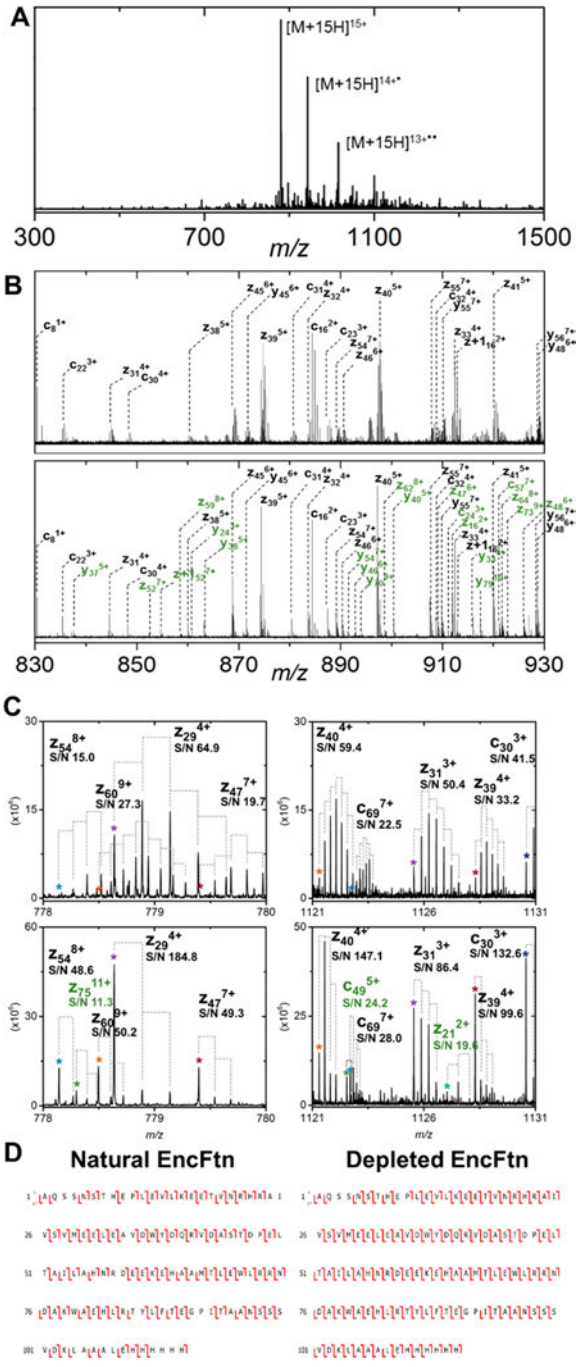


Figure 2. ECD fragmentation of EncFtn (13 kDa). (A) ECD spectrum of the $[M + 15H]^{15+}$ charge state of EncFtn. (B) Comparison of a region of the ECD fragmentation spectra of natural isotopic abundance EncFtn (top) and isotopically depleted EncFtn (bottom). The fragment ions assigned from each spectrum are labeled, with the ions found exclusively in the isotopically depleted spectrum highlighted in green. (C) A 2 and a 10 m/z range highlighting the simplified isotopic distributions, reduced ion distribution overlap, and increased S/N achieved using isotopically depleted ECD (bottom) over natural isotopic abundance ECD (top). Monoisotopic signals of fragment ions are highlighted with colored asterisks (*), and the S/N for each fragment ion is highlighted. Further examples are in Figure S7. (D) The fragmentation maps (protein sequence coverage) achieved after ECD of natural isotopic abundance EncFtn (left; 84.5%) and isotopically depleted EncFtn (right; 97.4%). Both spectra are the sum of 150 acquired transients.

in the isotopically depleted EncFtn ECD spectrum (~2 fold more than in the natural abundance spectrum), and over 97% sequence coverage was obtained (Figure 2D). Cleavages N terminal to proline are not generally observed in ECD. Remarkably, if this is considered, of 114 peptide bonds in EncFtn, only 2 possible cleavages were not observed. In addition, complementary c and z ion pairs cover over 85% of the protein sequence.

We also note that isotopically depleted low mass fragment ions (<2000 Da) exist almost exclusively as monoisotopic signals, displaying low abundance heavy isotope peaks (for examples, see c_8^{1+} (Figure 2B) and c_{15}^{2+} (Figure S7)). In fragment ions below ~20 amino acids in length, the isotopic signature can often not be obtained with high enough S/N to assign charge based on the spacing between isotopologues. Consequently, determination of the neutral mass for such ions is not possible. In targeted top down analyses, where protein sequence is known, fragment ion assignment can still be confidently made based on the accurate monoisotopic m/z signal. However, this hinders assignment of small fragment ions in untargeted top down experiments and *de novo* top down sequencing workflows. In such cases, protein identification must be first proposed based on larger fragment ion sequence tags before smaller fragment ions, displaying no isotopic signal, can be confidently assigned.

The compounding benefits of ID MS should be more evident as the precursor protein mass increases over 20 kDa. Therefore, we tested the utility of top down ID MS using bovine CA (29 kDa). CA constitutes an ideal model study, as it has been used extensively to characterize multiple top down fragmentations technologies and MS platforms.[15,32,43−45]

The individual charge state of CA was isolated and subject to ECD (Figure 3). ECD of the $[M + 22H]^{22+}$ (m/z 1332) charge state of natural isotopic abundance CA produced highly complex spectra (over 20 000 peaks with $S/N > 2.5$), which exhibited overlapping fragment ion isotope distributions throughout the spectra (Figure 3B, 3C, top). In total, from a single data set, 229 c and z fragment ions could be assigned, representing 50% sequence coverage (Figure 3D, left). The S/N of assigned CA fragment ions was low, especially for higher mass fragment ions; consequently, low sequence coverage was evident in the central region of the protein.

As expected, ECD of the same charge state of isotopically depleted CA resulted in significantly reduced spectral complexity and fragment ion distribution overlap. Fragment ions were observed with increased S/N (~2 to ~7 fold increase; Figure 3C and further examples in Supporting Information, Figure S9). Interestingly, compared to the equivalent natural isotopic abundance spectrum, a similar number of individual peaks were observed in the ECD spectrum of isotopically depleted CA, suggesting that substantially more fragmentation channels were evident. As a consequence, from the ECD spectrum of the $[M + 22H]^{22+}$ of isotopically depleted CA, 593 fragment ions (377 c ions, 216 z ions) were assigned, i.e., a ~3 fold increase. These fragment ions yielded a sequence coverage of 83% for the isotopically depleted protein (Figure 3D, right). Comparable assignment rate increases were apparent after ECD of the $[M + 32H]^{32+}$ charge state (Supporting Information, Figure S10). Combining the sequence coverage for both charge states of isotopically depleted CA resulted in over 90% (95.2% if bonds with adjacent proline residues were discounted), i.e., only 12 cleavages were not observed in this 263 amino acid protein—

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

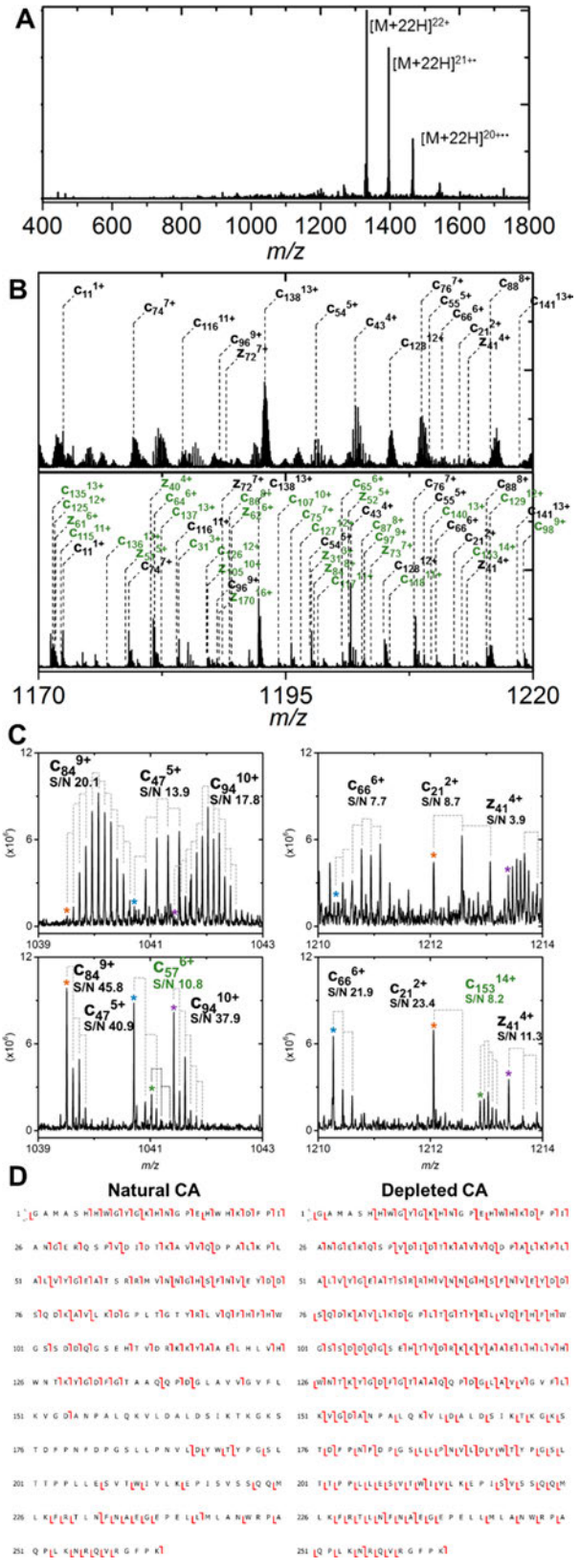


Figure 3. ECD fragmentation of CA (29 kDa). (A) ECD of the $[M + 22H]^{22+}$ charge state of CA. (B) A region of the ECD spectra of standard (top) and isotopically depleted protein (bottom). Assigned fragment ions are labeled, with the ions found exclusively in the isotopically depleted spectrum highlighted in green. (C) Two 4 m/z regions allowing comparison of the isotopic distribution of fragment ions from natural isotopic abundance CA (top) and isotopically depleted CA (bottom). Monoisotopic signal of each assigned fragment ion is highlighted with a colored asterisk (*), and the S/N for each ion is highlighted. Further examples are in Figure S9. (D) Fragmentation maps achieved after ECD of the $[M + 22H]^{22+}$ charge state of natural isotopic abundance CA (left; 50%) and isotopically depleted CA (right; 82.6%). Both spectra are the sum of 300 acquired transients.

very close to the "ideal" of single amino acid level resolution throughout the protein sequence (Supporting Information, Figure S11).

One striking characteristic of the ECD spectra of isotopically depleted proteins is the ability to assign extended stretches of complementary c and z ions, even in central regions of larger proteins. Comparison of the mass distributions of the assigned ECD fragment ions of natural isotopic abundance and isotopically depleted CA highlights that the ID MS strategy has the greatest benefit for the assignment of fragment ions of higher masses (>10 kDa), where ECD of isotopically depleted protein consistently affords 4 to 8 fold more fragment ions (Figure 4A). This ability to assign dramatically more fragment

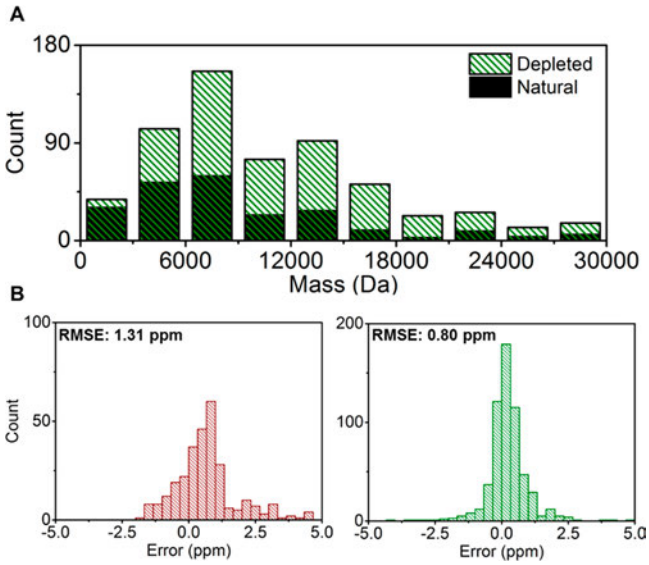


Figure 4. Mass and error distributions for the fragment ions assigned after ECD of the $[M + 22H]^{22+}$ charge state of natural isotopic abundance and isotopically depleted CA (c' and $z^{\bullet}$ ions only). (A) Histograms displaying the distribution of mass (Da) for the observed fragment ions after ECD of natural isotopic abundance CA (filled black bars) and isotopically depleted CA (green hatched bars); bin size = 3000 Da. (B) Histogram displaying the distribution of the mass error (ppm) for the observed fragment ions after ECD of natural isotopic abundance CA (left, red) and isotopically depleted CA (right, green); bin size = 0.333 ppm. The root mean square error (RMSE) for each distribution is shown. The top down isotopically depleted strategy allows assignment of a far greater number of fragment ions with lower mass error.

ions of high mass is a direct consequence of the inherent increase in the S/N that accompanies ID MS. In addition, the ability to directly observe the monoisotopic signals in isotopically depleted fragment ions is highly advantageous, as it removes the requirement to obtain isotopic distributions with sufficient S/N for precise poly averagine based

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

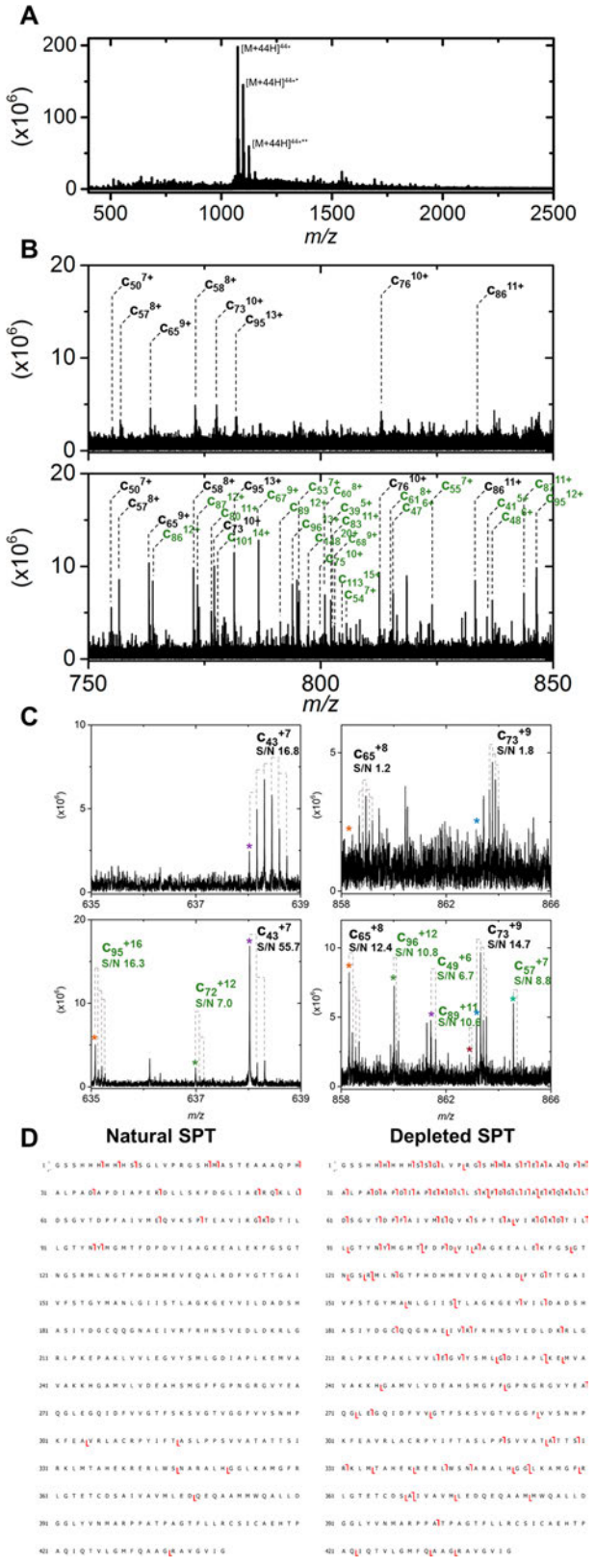


Figure 5. ECD fragmentation of SPT (47 kDa). (A) ECD of the $[M + 44H]^{44+}$ charge state of SPT. (B) A region of the ECD spectra of standard (top) and isotopically depleted protein (bottom). Assigned fragment ions are labeled, with the ions assigned exclusively in the isotopically depleted spectrum highlighted in green. (C) A 4 and an 8 m/z region allowing comparison of the isotopic distribution of fragment ions from natural isotopic abundance SPT (top) and isotopically depleted SPT (bottom). Monoisotopic signal of each assigned fragment ion is highlighted with a colored asterisk (*), and the S/N for each ion is highlighted. (D) Fragmentation maps achieved after ECD of the $[M + 44H]^{44+}$ charge state of natural isotopic abundance SPT (left; 20 assigned ions, 5.2%) and isotopically depleted SPT (right; 143 assigned ions, 29.9%). Both spectra are the sum of 300 acquired transients.

deconvolution methods. An additional benefit is that the mass error introduced using the poly averagine approximation during deconvolution is removed, leading to assignment of fragment ions with lower overall mass error (Figure 4B). As a result, higher confidence in fragment ion assignment can be achieved using the ID MS approach; this is particularly important for the interpretation of highly complex spectra, such as top down analysis of large proteins or assigning branched protein ions.[46]

Top-Down ID-MS of Large Proteins. Finally, we subjected natural isotope abundance and isotopically depleted SPT (47 kDa) to top down ECD fragmentation. Fragmentation of proteins of this molecular mass is highly challenging—not only due to the increase in isotope heterogeneity described above but also the compounding factors of (i) reduction in the precursor ion number, (ii) an increase in the number of possible fragmentation channels, and (iii) competing non dissociative channels. Together, this typically results in very low fragment ions yields and poor sequence coverage, even when employing extended acquisition times. This is demonstrated to good effect in the ECD fragmentation of the $[M + 44H]^{44+}$ charge state of natural isotope abundance SPT (5A, 5B, and 5C, top). In this analysis, extensive electron capture without dissociation was observed, resulting in dominant "EC no D" signals. Fragment ions signals are present at low S/N, and it is clear that many fragment signals are not sufficiently above the noise band to assign, even with extensive spectral averaging. It is also apparent that signal overlap also hampers assignment when fragments are observed at such low S/N. Thus, only 23 fragment ions can be assigned, the majority of low molecular mass c ions are assigned, constituting a sequence coverage of only ~5% (Figure 5D, left). Again, employing isotope depletion has a dramatic effect on the number of assigned fragment ions. An increase in the S/N of fragment ions of up to ~10 fold is observed in the resulting ECD fragments from this 47 kDa precursor ion (Figure 5B, 5C, bottom). This allows assignment of a total of 121 fragment ions from this single data acquisition, a sequence coverage of 30% (Figure 5D, right).

Although this sequence coverage achieved is not comprehensive, sufficient information is obtained for confident protein identification from a Uniprot database search. In contrast, this was not possible using the fragmentation data obtained from natural isotope abundance SPT. It is also interesting to note the nonuniform distribution of assigned fragments—the majority arising from N terminal c ions. This was observed in ECD analysis of several charge states of the protein, suggesting ECD alone is not sufficient for complete top down characterization of these larger proteins. Indeed, several recent reports have highlighted that top down analysis of proteoforms >30 kDa benefits from implementing several orthogonal fragmentation techniques and/or employing gas phase ion–ion proton transfer reactions (PTR).[47,48] It is clear that isotope depletion adds to the growing number of strategies for

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

increasing top down sequence coverage in large proteins and will be complementary to these other strategies in future studies.

Benefits of ID-MS for Online Top-Down MS. One of the principal goals of top down mass spectrometry is to achieve comprehensive protein sequence coverage using spectral acquisition times that are compatible with front end chromatography, thereby allowing LC−MS/MS work flows.[49−51] However, the time constraints inherent in this experimental setup preclude extensive spectral averaging. Consequently, as fragment ion signals are not acquired with sufficient abundance to assign, protein sequence coverages are often limited (this is especially true for proteins >20 kDa).

In effect, there is a compromise that exists between the S/N level achieved and the spectral acquisition time. Spectral averaging produces a gain in the S/N ratio that is approximately proportional to the square root of the number of scans averaged.[52] Because of this nonlinear relationship, the increased S/N inherent in our ID MS approach should be particularly effective for increasing the fragment ion sequence coverage obtainable with limited spectral averaging.

In an effort to demonstrate this, we began by acquiring ECD spectra for both the natural isotopic abundance and isotopically depleted forms of EncFtn (13 kDa) and CA (29 kDa), using both 20 and 5 spectral averages, which constituted total data collection times of ∼25 and ∼6 s, respectively. The resulting spectra were analyzed, and fragment ions were assigned (Supporting Information, Figures S12 and S13) and compared to the longer spectral acquisition time, described above (Figure 6). As expected, for natural isotopic abundance

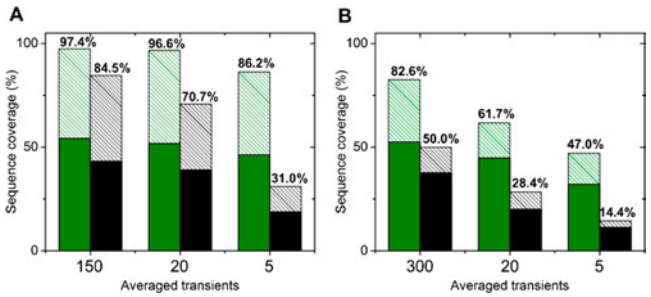


Figure 6. Fragment ion sequence coverage obtained after ECD of natural isotopic abundance and isotopically depleted forms of (A) EncFtn (13 kDa; $[M + 15H]^{15+}$) and (B) CA (29 kDa; $[M + 22H]^{22+}$ charge state) with varying degrees of spectral averaging. For all data: green and black bars represent sequence coverage obtained for isotopically depleted protein and natural isotopic abundance protein, respectively; *c* ions are represented with solid bars, and *z* ions are represented with hatched bars. The total sequence coverage (*c* and *z* ions) is stated for each data set.

EncFtn, reduction in the spectral averaging reduces the protein sequence coverage significantly. With spectral averaging limited to 5 transients, only 48 ions could be assigned, constituting 31% total sequence coverage. In contrast, for isotopically depleted EncFtn, the reliance on extensive spectral averaging to obtain high sequence coverage is far less pronounced; here, 86.2% protein sequence coverage was achieved with only 5 averaged spectra. For the larger protein, CA (29 kDa), without extensive spectral averaging, the sequence coverage obtained after ECD of the natural isotopic abundance protein is severely limited—28.4% sequence coverage is obtained with 20 averaged spectra, and 14.4% sequence coverage is obtained upon averaging only 5 spectra. Dramatic improvements are observed using the isotopically depleted strategy, and in depth sequence coverage can still be assigned under time limited data acquisitions. ECD of isotopically depleted CA using 20 and 5 spectral averages affords sequence coverage of 61.7 and 47% respectively.

Following these initial results, we went on to perform online LC−MS/MS of the isotopically natural and depleted forms of the two proteins using reverse phase chromatography separation with the FT ICR configured for precursor ion selection in the quadrupole and ECD fragmentation in the ICR cell. Under these conditions, the elution of both proteins occurs over a time frame of ∼20 s for EncFtn and ∼60 s for CA, and the isotopically normal and depleted forms of each protein displayed identical elution profiles. For LC−MS/MS of EncFtn, 5 scans were collected over the time course of the protein elution. After spectral summing, for natural isotopic abundance EncFtn, 30 fragment ions could be assigned above the noise band, which constitute a sequence coverage of 17%. In contrast, the increased S/N obtained with the isotope depletion strategy led to the assignment of 199 fragment ions and a sequence coverage of 81% (Figure S14) under identical LC−MS/MS conditions.

The superior sequence coverage achieved using isotope depletion was also evident upon LC−MS/MS analysis of the 29 kDa CA protein. Under the experimental conditions used, CA eluted over a retention period of ∼60 s, which allowed ∼13 spectral scans to be summed and processed. The ECD fragmentation conditions were optimized to increase fragment ion yield; however, only 36 fragment ions could be assigned above the noise band for isotopically natural CA (representing a sequence coverage of 12%). An average increase in fragment ion S/N level of over 4 fold was achieved in the analysis of isotopically depleted CA. This allowed the confident assignment of 345 fragment ions (61% sequence coverage) under identical experimental conditions (Figure 7). These initial studies clearly demonstrate the potential benefit of applying isotope depletion strategies in top down proteomic workflows and highlight the ability to achieve more comprehensive sequence coverage of larger proteins on chromatographic time scales.

■ CONCLUSION

We have developed an efficient method for the production of milligram quantities of isotopically depleted recombinant proteins and demonstrated the production of several proteins with molecular masses up to ∼50 kDa. We demonstrate that mass spectra of intact isotopically depleted proteins display decreased isotope distribution widths and increased S/N. The observable monoisotopic signal of isotopically depleted proteins allows accurate molecular mass to be directly determined, even in large proteins. Applying ID MS in conjunction with top down fragmentation results in reduced spectral complexity, increased S/N, and increased mass accuracy; together, these allow assignment of dramatically more fragment ions (typically 2 to 4 fold) and consequently increased protein sequence coverage. We have demonstrated that the benefits of top down ID MS are particularly evident when analyzing higher molecular mass proteins and in time limited top down MS experiments, such as online LC−MS/MS applications.

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

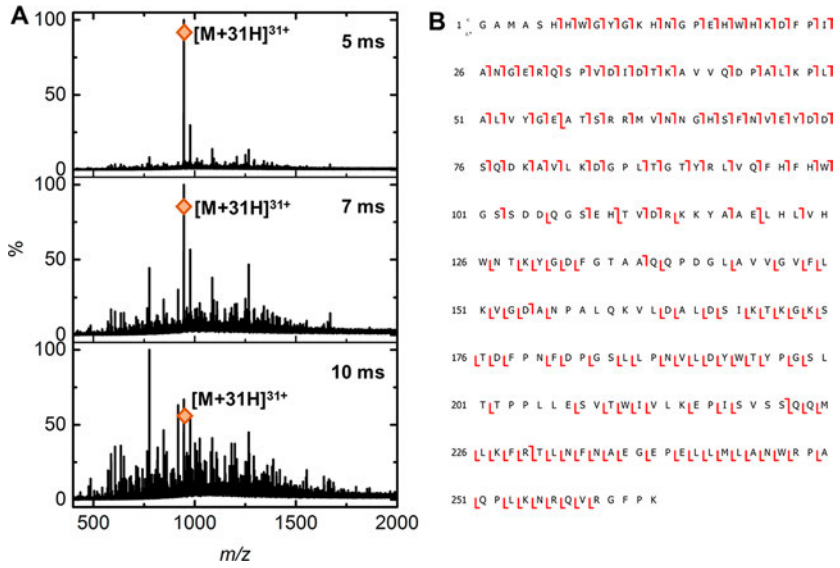


Figure 7. Online top down ECD (LC−MS/MS) of bovine CA (29 kDa). (A) Summed ECD spectra of the $[M + 31H]^{31+}$ charge state obtained during chromatographic elution. The ECD pulse length was varied to optimize fragment ion yield (top, 5 ms ECD pulse length; middle, 7 ms pulse length; bottom, 10 ms pulse length). Each spectrum is the sum of ∼15 data scans collected over the time course of the protein elution. (B) Fragmentation maps achieved after online ECD of the $[M + 31H]^{31+}$ charge state of isotopically depleted CA under optimized ECD conditions (64%; 345 assigned fragment ions).

FUTURE OUTLOOK

The ID MS strategy is analogous to the isotope enrichment techniques that have become integral to biomolecular NMR spectroscopy.[53] Similarly, it is clear that ID MS has huge promise for many structural biomolecular MS applications, particularly for proteins (or other biomolecules, e.g., oligonucleotides[54]) of high molecular mass. Techniques such as hydrogen/deuterium exchange MS, (top down) native protein MS, protein footprints via covalent labeling, and other structural MS techniques will all benefit from the advantages of increased sensitivity and S/N and reduced spectral complexity that accompany isotopic depletion. For this technique to become established in the community, the production of isotopically depleted recombinant proteins on a milligram scale must be reliably achieved without a prohibitive economic barrier. The method described here demonstrates that this is possible using an *E. coli* recombinant expression system.

Moreover, this strategy should not be limited to protein production in bacterial systems, and future adaption for the production of recombinant proteins in eukaryotic hosts should be possible. Yeast species (e.g., *Pichia pastoris, Saccharomyces cerevisiae*, and *Saccharomyces pombi*) are routinely used for recombinant protein production and can be successfully cultured using defined minimal media conditions (e.g., Basal Salts with a defined carbon source, such as glycose, glycerol, or methanol). Adaption of these existing culture protocols utilizing ^{13}C depletion will allow access to the extensive protein folding and PTM cellular machinery in higher organisms and afford isotopically depleted forms of the more complex proteoforms characteristic of eukaryotic organisms. More recently, strategies have been developed for the production of isotopically defined proteins in insect cells and mammalian cell lines.[55,56] These hosts have a limited capacity to biosynthesize amino acids from simple precursors and consequently do not grow efficiently on defined carbon sources. However, isotope labeling has been demonstrated in both organisms by supplementation of growth media with individually labeled amino acids. For uniform isotope labeling of recombinant proteins, the cost for supplying isotope labeled forms of all 20 amino acids is often prohibitive. To overcome this, a more economical alternative is to supplement the host with a labeled algal or yeast extract as a comprehensive amino acid source. This strategy has been successfully used to produce (^{13}C, ^{2}H, ^{15}N) isotopically enriched proteins for biomolecular NMR studies. It remains to be seen if this strategy can be performed with the required incorporation efficiency to allow effective isotope depletion for protein mass spectrometry. However, these existing reports suggest that isotope depletion in higher organisms is technical feasible in the near future.

Finally, it should be stressed that isotope labeling results in isotope depletion of the entire proteome of the host organism. Thus, the technique has potential benefits for proteomic applications, as was first demonstrated in 1999.[23] The results presented here clearly demonstrate that isotope depletion is particularly well suited for applying in combination with top down proteomic studies, which will benefit from the improved sensitivity and accuracy for intact mass determination and improved sequence coverage obtained in time limited top down fragmentation experiments. These benefits will help improve the number of proteoforms that can be identified on an LC time scale; particularly in the mass range >20 kDa.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.9b00119.

Further details of protein used in this study; further experimental detail; SDS PAGE analysis of proteins used in this study; MS analysis of N58D EncFtn variant; further top down fragmentation data (PDF)

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710

AUTHOR INFORMATION

Corresponding Author

David J. Clarke − *The EastChem School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, U.K.;* orcid.org/0000 0002 3741 2952; Email: dave.clarke@ed.ac.uk

Authors

Kelly J. Gallagher − *The EastChem School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, U.K.*

Michael Palasser − *The EastChem School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, U.K.*

Sam Hughes − *The EastChem School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, U.K.*

C. Logan Mackay − *The EastChem School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, U.K.;* orcid.org/0000 0003 1018 8353

David P. A. Kilgour − *Chemistry and Forensics, Nottingham Trent University, Nottingham NG11 8NS, U.K.*

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.9b00119

Author Contributions

D.J.C. conceived the study and designed the experiments. K.J.G. developed the protein expression protocols. K.J.G, M.P., and S.H. performed the experiments. MS data collection was performed under the guidance of D.J.C. and C.L.M. D.P.A.K. designed and coded the data analysis software and discussed the manuscript. The manuscript was written by K.J.G and D.J.C. in discussions with all coauthors.

Notes

The authors declare no competing financial interest.

The MS data sets described in this study are available to download in their original format (.baf files; Bruker Daltonics) from the Edinburgh DataShare (https://datashare.is.ed.ac.uk/handle/10283/760) using the following link: http://dx.doi.org/10.7488/ds/2446.

ACKNOWLEDGMENTS

This work was funded by a Royal Society Research Grant awarded to D.J.C. (RG160814). K.J.G. is supported by an Engineering and Physical Sciences Research Council (EPSRC) PhD studentship. M.P. was supported by the Erasmus programme. We thank Prof. Dominic Campopiano and Dr. John Marles Wright for expression plasmids.

REFERENCES

(1) Lanucara, F.; Eyers, C. E. Top down mass spectrometry for the analysis of combinatorial post translational modifications. *Mass Spectrom. Rev.* **2013**, *32*, 27−42.

(2) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* **2014**, *445*, 683−93.

(3) Zhang, J.; Reza Malmirchegini, G.; Clubb, R. T. C.; Loo, J. Native top down mass spectrometry for the structural characterization of human hemoglobin. *Eur. J. Mass Spectrom.* **2015**, *21*, 221.

(4) Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A. An integrated native mass spectrometry and top down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **2018**, *10*, 139−148.

(5) 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*, 12801−12808.

(6) 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*, 3079−3082.

(7) Kruppa, G. H.; Schoeniger, J.; Young, M. M. A top down approach to protein structural studies using chemical cross linking and Fourier transform mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 155−162.

(8) Chen, J.; Cui, W.; Giblin, D.; Gross, M. L. New Protein Footprinting: Fast Photochemical Iodination Combined with Top Down and Bottom Up Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1306−1318.

(9) Limpikirati, P.; Liu, T.; Vachet, R. W. Covalent labeling mass spectrometry with non specific reagents for studying protein structure and interactions. *Methods* **2018**, *144*, 79−93.

(10) Zubarev, R.; Kelleher, N. L.; McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, *120*, 3265−3266.

(11) Zubarev, R. A. Electron capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **2004**, *15*, 12−6.

(12) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics* **2007**, *6*, 1942−1951.

(13) Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L. On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* **2011**, *83*, 6868−74.

(14) Riley, N. M.; Mullen, C.; Weisbrod, C. R.; Sharma, S.; Senko, M. W.; Zabrouskov, V.; Westphall, M. S.; Syka, J. E. P.; Coon, J. J. Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 520−531.

(15) Riley, N. M.; Westphall, M. S.; Coon, J. J. Sequencing Larger Intact Proteins (30−70 kDa) with Activated Ion Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 140−149.

(16) Valkenborg, D.; Mertens, I.; Lemière, F.; Witters, E.; Burzykowski, T. The isotopic distribution conundrum. *Mass Spectrom. Rev.* **2012**, *31*, 96−109.

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

(18) 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*, 3293−3299.

(19) Akashi, S.; Takio, K.; Matsui, H.; Tate, S.; Kainosho, M. Collision Induced Dissociation Spectra Obtained by Fourier Trans form Ion Cyclotron Resonance Mass Spectrometry Using a 13 C, 15 N Doubly Depleted Protein. *Anal. Chem.* **1998**, *70*, 3333−3336.

(20) Charlebois, J. P.; Patrie, S. M.; Kelleher, N. L. Electron capture dissociation and 13C,15N depletion for deuterium localization in intact proteins after solution phase exchange. *Anal. Chem.* **2003**, *75*, 3263−3266.

(21) Sharma, S.; Simpson, D. C.; Tolić, N.; Jaitly, N.; Mayampurath, A. M.; Smith, R. D.; Paša Tolić, L. Proteomic profiling of intact proteins using WAX RPLC 2 D separations and FTICR mass spectrometry. *J. Proteome Res.* **2007**, *6*, 602−610.

(22) Zubarev, R. A.; Demirev, P. A. Isotope depletion of large biomolecules: Implications for molecular mass measurements. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 149−156.

(23) Jensen, P. K.; Paša Tolić, L.; Anderson, G. A.; Horner, J. A.; Lipton, M. S.; Bruce, J. E.; Smith, R. D. Probing proteomes using capillary isoelectric focusing electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **1999**, *71*, 2076−2084.

(24) 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*, e18972.

(25) Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. Carbonic Anhydrase as a Model for Biophysical and Physical Organic Studies of Proteins and Protein−Ligand Binding. *Chem. Rev.* **2008**, *108*, 946−1051.
(26) Wadsworth, J. M.; Clarke, D. J.; McMahon, S. A.; Lowther, J. P.; Beattie, A. E.; Langridge Smith, P. R. R.; Broughton, H. B.; Dunn, T. M.; Naismith, J. H.; Campopiano, D. J. The chemical basis of serine palmitoyltransferase inhibition by myriocin. *J. Am. Chem. Soc.* **2013**, *135*, 14276−14285.
(27) Marley, J.; Lu, M.; Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **2001**, *20*, 71−75.
(28) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical software to analyze top down mass spectrometry data. *Proteomics* **2015**, *15*, 1235−1238.
(29) Shi, S. D. H.; Hendrickson, C. L.; Marshall, A. G. Counting Individual Sulfur Atoms in a Protein by Ultrahigh Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Exper imental Resolution of Isotopic Fine Structure in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 11532−11537.
(30) Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229−233.
(31) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320−332.
(32) Weisbrod, C. R.; Kaiser, N. K.; Syka, J. E. P.; Early, L.; Mullen, C.; Dunyach, J. J.; English, A. M.; Anderson, L. C.; Blakney, G. T.; Shabanowitz, J.; Hendrickson, C. L.; Marshall, A. G.; Hunt, D. F. Front End Electron Transfer Dissociation Coupled to a 21 T FT ICR Mass Spectrometer for Intact Protein Sequence Analysis. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1787−1795.
(33) Thurlow, S. E.; Kilgour, D. P.; Campopiano, D. J.; Mackay, C. L.; Langridge Smith, P. R. R.; Campbell, C. J.; Clarke, D. J. Determination of protein thiol reduction potential by isotope label ing and intact mass measurement. *Anal. Chem.* **2016**, *88*, 2727−2733.
(34) Scotcher, J.; Clarke, D. J.; Mackay, C. L.; Hupp, T.; Sadler, P. J.; Langridge Smith, P. R. R. Redox regulation of tumour suppressor protein p53: identification of the sites of hydrogen peroxide oxidation and glutathionylation. *Chem. Sci.* **2013**, *4*, 1257−1269.
(35) Radestock, S.; Gohlke, H. Protein rigidity and thermophilic adaptation. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 1089−1108.
(36) Hao, P.; Adav, S. S.; Gallart Palau, X.; Sze, S. K. Recent advances in mass spectrometric analysis of protein deamidation. *Mass Spectrom. Rev.* **2017**, *36*, 677−692.
(37) 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*, 253−262.
(38) Kilgour, D. P. A.; Van Orden, S. L.; Tran, B. Q.; Goo, Y. A.; Goodlett, D. R. Producing Isotopic Distribution Models for Fully Apodized Absorption Mode FT MS. *Anal. Chem.* **2015**, *87*, 5797−5801.
(39) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical solutions: Principles and application of electron based dissociation in mass spectrometry based analysis of protein structure. *Mass Spectrom. Rev.* **2018**, *37*, 750−771.
(40) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **2000**, *72*, 563−73.
(41) 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*, 9528−9533.
(42) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü.A.; Tsybin, Y. O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem. Soc. Rev.* **2013**, *42*, 5014.
(43) O'Connor, P. B.; Speir, J. P.; Senko, M. W.; Little, D. P.; McLafferty, F. W. Tandem mass specatrometry of carbonic anhydrase (29 kDa). *J. Mass Spectrom.* **1995**, *30*, 88−93.
(44) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Top down mass spectrometry of a 29 kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1774−9.
(45) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135*, 12646−12651.
(46) Chen, D.; Gomes, F.; Abeykoon, D.; Lemma, B.; Wang, Y.; Fushman, D.; Fenselau, C. Top Down Analysis of Branched Proteins Using Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 4032−4038.
(47) Sanders, J.; Mullen, C.; Watts, E.; Holden, D.; Syka, J. E.; Schwartz, J. C.; Brodbelt, J. S. Enhanced Sequence Coverage of Large Proteins by Combining Ultraviolet Photodissociation with Proton Transfer Reactions. *Anal. Chem.* **2020**, *92*, 1041.
(48) Riley, N. M.; Sikora, J. W.; Seckler, H. S.; Greer, J. B.; Fellers, R. T.; LeDuc, R. D.; Westphall, M. S.; Thomas, P. M.; Kelleher, N. L.; Coon, J. J. The Value of Activated Ion Electron Transfer Dissociation for High Throughput Top Down Characterization of Intact Proteins. *Anal. Chem.* **2018**, *90*, 8553−8560.
(49) Fornelli, L.; Toby, T. K.; Schachner, L. F.; Doubleday, P. F.; Srzentić, K.; DeHart, C. J.; Kelleher, N. L. Top down proteomics: Where we are, where we are going? *J. Proteomics* **2018**, *175*, 3−4.
(50) Ahlf, D. R.; Thomas, P. M.; Kelleher, N. L. Developing top down proteomics to maximize proteome and sequence coverage from cells and tissues. *Curr. Opin. Chem. Biol.* **2013**, *17*, 787−794.
(51) Anderson, L. C.; DeHart, C. J.; Kaiser, N. K.; Fellers, R. T.; Smith, D. F.; Greer, J. B.; LeDuc, R. D.; Blakney, G. T.; Thomas, P. M.; Kelleher, N. L.; Hendrickson, C. L. Identification and Characterization of Human Proteoforms by Top Down LC 21 T FT ICR Mass Spectrometry. *J. Proteome Res.* **2017**, *16*, 1087−1096.
(52) Marshall, A. G.; Comisarow, M. B. Fourier transform methods in spectroscopy. *J. Chem. Educ.* **1975**, *52*, 638.
(53) Ohki, S. Y.; Kainosho, M. Stable isotope labeling methods for protein NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53*, 208−226.
(54) Xiong, Y.; Schroeder, K.; Greenbaum, N. L.; Hendrickson, C. L.; Marshall, A. G. Improved mass analysis of oligoribonucleotides by C 13, N 15 double depletion and electrospray ionization FT ICR mass spectrometry. *Anal. Chem.* **2004**, *76*, 1804−1809.
(55) Dutta, A.; Saxena, K.; Schwalbe, H.; Klein Seetharaman, J. Isotope labeling in mammalian cells. *Methods Mol. Biol.* **2012**, *831*, 55−69.
(56) Franke, B.; Opitz, C.; Isogai, S.; Grahl, A.; Delgado, L.; Gossert, A. D.; Grzesiek, S. Production of isotope labeled proteins in insect cells for NMR. *J. Biomol. NMR* **2018**, *71*, 173−184.

https://dx.doi.org/10.1021/jasms.9b00119
J. Am. Soc. Mass Spectrom. 2020, 31, 700−710