Edinburgh Research Explorer

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

Isotope Depletion Mass Spectrometry (ID-MS) for accurate mass determination and improved top-down sequence coverage of intact proteins.

Running Title: ***Top-down isotope depletion mass spectrometry.***

Kelly J. Gallagher[1], Michael Palasser[1,3], Sam Hughes[1], C. Logan Mackay[1], David P. A. Kilgour[2] and David J. Clarke[*1]

[1] The EastChem School of Chemistry, University of Edinburgh, Joseph Black Building, Brewster Road, Edinburgh, EH9 3FJ, UK.

[2.] Chemistry and Forensics, Nottingham Trent University, Rosalind Franklin Building, Clifton Lane, Nottingham, NG11 8NS, U.K

[3] *Present Address:* Institut für Organische Chemie and Center for Molecular Biosciences Innsbruck (CMBI), Universität Innsbruck, Innrain, 80-82, 6020, Innsbruck, Austria.

Correspondence to: David Clarke; *e-mail:* dave.clarke@ed.ac.uk

ABSTRACT: Top-down mass spectrometry (MS) is an increasingly 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 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 twenty years ago (Marshall, Senko, Li, Li, Dillon, Guan, and Logan. *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 reveal dramatically reduced isotope distributions with the 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.

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 characterisation of post-translational and chemical protein modifications[1, 2]. More recently, the application of top-down fragmentation techniques, in conjunction 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, included: hydrogen−deuterium exchange MS [5, 6], cross-linking MS [7], and covalent labelling 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 wide-spread 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 context of a top-down fragmentation experiment, these compounding difficulties all reduce the number of fragment ions which 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 twenty 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 analysing 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

Proteins Samples.

Three well-characterised 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 palmitoyltransferase 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 ºC. This was then used to inoculate 500 ml of 2xYT media, incubated at 37 ºC until an OD_{600} 0.6-0.8 was obtained, at which point the cell culture was centrifuged at 5000 *x g* at 4 ºC for 20 minutes (200 ml per expression culture required). The resultant pellet was washed with 20 ml/ gram of pellet 5xM9 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 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 ºC for 1 hour, before protein expression was induced with 1 or 0.1 mM IPTG as required, and incubated overnight at 18 ºC [27]. After incubation cultures were centrifuged at 5000 × *g* at 4 ºC for 30 minutes, and pellets were stored at -80 ºC 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 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 elution buffer (PBS and 300 mM imidazole). 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 lysis buffer 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 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. 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 pipette tips (Agilent Technologies, Santa Clara, CA), with final elution in 50:50 water:acetonitrile with 0.1% formic acid. Protein samples were ionised by nanoelectrospray (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 12T 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 50-200 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 optimised 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 20-35V. Electron capture dissociation (ECD) was achieved by irradiating with electrons using a heated hollow dispenser cathode. Typically,

cathode conditions were: bias voltage 1.5 V, lens voltage 15 V, and a pulse length of between 5-30 ms was employed. 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, 5 (CA) transients as stated. For online top-down fragmentation protein samples were separated on an ACE 3 C4 column (75 x 0.5mm) and eluted using a 5 to 95% gradient of acetonitrile over 30 minutes. Individual charge states were isolated in the mass resolving quadrupole and fragmentation was achieved using ECD in the ICR cell. The cathode conditions were optimised to achieve maximum fragment ion yield.

Top-Down Data Analysis

FT-ICR MS data was post-processed, in absorption mode, using a specially adapted version of AutoVectis (Spectroswiss Sàrl, Lausanne, Switzerland); for detailed description see Supporting Information. For natural isotopic abundance proteins, top-down data was also analysed 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 pre-formulated 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 which 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 100-1000 μg of ^{13}C and ^{15}N depleted protein per 100 ml 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 *ca.* 20 kDa in molecular weight [17–20, 29]. Here we extend this technique for the production of proteins up to *ca.* 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 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 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 *ca.* 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 characterisation 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 analysed natural isotopic abundance and isotope 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 reduces 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 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 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 (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 and 2C). Thus, ECD of the isotopically depleted protein allowed accurate assignment of sidechain 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 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+} (Fig 2B), c_{15}^{2+} (Fig 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 characterise multiple top-down fragmentations technologies and MS platforms [15, 32, 43–45].

Individual charge state of CA was isolated and subject to ECD (Figure 3). ECD of the $[M+22]^{22+}$ (*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 dataset, 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 S7). 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 – 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 complimentary 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 ions of high mass, is a direct consequence of the inherent increase in the S/N which 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 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 (Figure 5. 5A, 5B, 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 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 un-uniform 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 characterisation of this larger proteins. Indeed, several recent reports have highlighted that top-down analysis of proteoforms >30 kDa benefit 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 increasing top-down sequence coverage in large proteins, and will be complementary to these other strategies in future studies.

The benefits of ID-MS for online top-own 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 workflows [49–51]. However, the time constraints inherent in this experimental setup precludes 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 non-linear 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 seconds respectively. The resulting spectra were analysed and fragment ions assigned (Supporting Information, Figure S12 and S13) and compared to the longer spectral acquisition time, described above (Figure 6). As expected, for natural isotopic abundance 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 timeframe of ~20 seconds for EncFtn and ~60 seconds 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 seconds, which allowed ~13 spectral scans to be summed and processed. The ECD fragmentation conditions were optimised 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 timescales.

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 analysing higher molecular mass proteins and in time-limited top down MS experiments, such as online LC-MS/MS applications.

Future Outlook

The ID-MS strategy is analogous to the isotope enrichment techniques which 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 labelling, and other structural MS techniques will all benefit from the advantages of increased sensitivity and S/N and reduced spectral complexity which 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 in 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 of biosynthesize amino acids from simple precursors and consequently do not grow efficiently on defined carbon sources. However, isotope labelling has been demonstrated in both organisms by supplementation of growth media with individually labelled amino acids. For uniform isotope labelling of recombinant proteins, the cost for supplying isotope labelled forms of all 20 amino acids is often prohibitive. To overcome this, a more economical alternative is to supplement the host with a labelled 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 labelling 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-timescale; particularly in the mass range >20 kDa.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website. Supporting information includes: 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. The MS datasets described in this study are available to download in their original format (.baf files; Bruker Daltonics) from the at Edinburgh DataShare (https://datashare.is.ed.ac.uk/handle/10283/760) using the following link: http://dx.doi.org/10.7488/ds/2446.

AUTHOR INFORMATION

Corresponding Author

* dave.clarke@ed.ac.uk

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.

REFERENCES

1.   Lanucara, F., Eyers, C.E.: Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. Mass Spectrom. Rev. 32, 27–42 (2013). doi:10.1002/mas.21348

2.   Catherman, A.D., Skinner, O.S., Kelleher, N.L.: Top Down proteomics: facts and perspectives. Biochem. Biophys. Res. Commun. 445, 683–93 (2014). doi:10.1016/j.bbrc.2014.02.041

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. 21, 221 (2015). doi:10.1255/ejms.1340

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. 1–10 (2018). doi:10.1038/nchem.2908

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. 131, 12801–12808 (2009). doi:10.1021/ja904379w

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. 90, 3079–3082 (2018). doi:10.1021/acs.analchem.7b03655

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. 17, 155–162 (2003). doi:10.1002/rcm.885

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. 23, 1306–1318 (2012). doi:10.1007/s13361-012-0403-1

9. Limpikirati, P., Liu, T., Vachet, R.W.: Covalent labeling-mass spectrometry with non-specific reagents for studying protein structure and interactions. Methods. 144, 79–93 (2018). doi:10.1016/j.ymeth.2018.04.002

10. Zubarev, R., Kelleher, N.L., McLafferty, F.W.: Electron capture dissociation of multiply charged protein cations. A nonergodic process. J. Am. Chem. Soc. 120, 3265–3266 (1998)

11. Zubarev, R.A.: Electron-capture dissociation tandem mass spectrometry. Curr. Opin. Biotechnol. 15, 12–6 (2004). doi:10.1016/j.copbio.2003.12.002

12. Good, D.M., Wirtala, M., McAlister, G.C., Coon, J.J.: Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. Mol. Cell. Proteomics. 6, 1942–1951 (2007). doi:10.1074/mcp.M700073-MCP200

13. Compton, P.D., Zamdborg, L., Thomas, P.M., Kelleher, N.L.: On the scalability and requirements of whole protein mass spectrometry. Anal. Chem. 83, 6868–74 (2011). doi:10.1021/ac2010795

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. 27, 520–531 (2015). doi:10.1007/s13361-015-1306-8

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. 29, 140–149 (2018). doi:10.1007/s13361-017-1808-7

16. Valkenborg, D., Mertens, I., Lemière, F., Witters, E., Burzykowski, T.: The isotopic distribution conundrum. Mass Spectrom. Rev. 31, 96–109 (2012). doi:10.1002/mas.20339

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. 119, 433–434 (1997)

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. 82, 3293–3299 (2010). doi:10.1021/ac100079z

19. 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. 70, 3333–3336 (1998). doi:10.1021/ac980215f

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. 75, 3263–3266 (2003). doi:10.1021/ac020690k

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. 6, 602–610 (2007). doi:10.1021/pr060354a

22. Zubarev, R.A., Demirev, P.A.: Isotope depletion of large biomolecules: Implications for molecular mass measurements. J. Am. Soc. Mass Spectrom. 9, 149–156 (1998). doi:10.1016/S1044-0305(97)00232-8

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. 71, 2076–2084 (1999). doi:10.1021/ac990196p

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. 5, e18972 (2016). doi:10.7554/eLife.18972

25. Vijay M. Krishnamurthy, George K. Kaufman, Adam R. Urbach, Irina Gitlin, Katherine L. Gudiksen, Douglas B. Weibel, A., Whitesides*, G.M.: Carbonic Anhydrase as a Model for Biophysical and Physical-Organic Studies of Proteins and Protein−Ligand Binding. Chem. Rev. 108, 946–1051 (2008). doi:10.1021/CR050262P

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. 135, 14276–14285 (2013). doi:10.1021/ja4059876

27. Marley, J., Lu, M., Bracken, C.: A method for efficient isotopic labeling of recombinant proteins. J. Biomol. NMR. 20, 71–75 (2001)

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. 15, 1235–1238 (2015). doi:10.1002/pmic.201400313

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: Experimental Resolution of Isotopic Fine Structure in Proteins. Proc. Natl. Acad. Sci. USA. 95, 11532–11537 (1998)

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. 6, 229–233 (1995). doi:10.1016/1044-0305(95)00017-8

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. 11, 320–332 (2000). doi:10.1016/S1044-0305(99)00157-9

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 Tesla FT-ICR Mass Spectrometer for Intact Protein Sequence Analysis. J. Am. Soc. Mass Spectrom. 28, 1787–1795 (2017). doi:10.1007/s13361-017-1702-3

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. 88, 2727–2733 (2016)

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. 4, 1257–1269 (2013)

35. Radestock, S., Gohlke, H.: Protein rigidity and thermophilic adaptation. Proteins Struct. Funct. Bioinforma. 79, 1089–1108 (2011). doi:10.1002/prot.22946

36. Hao, P., Adav, S.S., Gallart-Palau, X., Sze, S.K.: Recent advances in mass spectrometric analysis of protein deamidation. Mass Spectrom. Rev. 36, 677–692 (2017). doi:10.1002/mas.21491

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. 28, 253–262 (2017). doi:10.1007/s13361-016-1549-z

38. Kilgour, D.P.A., Orden, S.L. Van, Tran, B.Q., Goo, Y.A., Goodlett, D.R.: Producing Isotopic Distribution Models for Fully Apodized Absorption Mode FT-MS. Anal. Chem. 87, 5797–5801 (2015). doi:10.1021/acs.analchem.5b01032

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. 1, 1–22 (2018). doi:10.1002/mas.21560

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. 72, 563–73 (2000)

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. 101, 9528–9533 (2004). doi:10.1073/pnas.0402700101

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. 42, 5014 (2013). doi:10.1039/c3cs35477f

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. 30, 88–93 (1995). doi:10.1002/jms.1190300114

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. 99, 1774–9 (2002). doi:10.1073/pnas.251691898

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. 135, 12646–12651 (2013). doi:10.1021/ja4029654

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.
        90, 4032–4038 (2018). doi:10.1021/acs.analchem.7b05234

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.
        acs.analchem.9b04026 (2019). doi:10.1021/acs.analchem.9b04026

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. 90, 8553–8560 (2018).
        doi:10.1021/acs.analchem.8b01638

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. 10–11 (2017). doi:10.1016/j.jprot.2017.02.002

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. 17, 787–794 (2013). doi:10.1016/j.cbpa.2013.07.028

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 Tesla
        FT-ICR Mass Spectrometry. J. Proteome Res. 16, 1087–1096 (2017).
        doi:10.1021/acs.jproteome.6b00696

52.   Marshall, A.G., Comisarow, M.B.: Fourier transform methods in spectroscopy. J.

Chem. Educ. 52, 638 (1975). doi:10.1021/ed052p638

53. Ohki, S.-Y., Kainosho, M.: Stable isotope labeling methods for protein NMR

spectroscopy. Prog Nucl Magn Reson Spectrosc. 53, 208–226 (2008).

doi:10.1016/j.pnmrs.2008.01.003

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. 76, 1804–1809

(2004). doi:10.1021/ac030299e

55. Dutta, A., Saxena, K., Schwalbe, H., Klein-Seetharaman, J.: Isotope labeling in

mammalian cells. Methods Mol. Biol. 831, 55–69 (2012). doi:10.1007/978-1-61779-

480-3_4

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. 71,

173–184 (2018). doi:10.1007/s10858-018-0172-7

Figure Legends.

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 isotope 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 (*)

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 labelled, with the ions found exclusively in the isotopically depleted spectrum highlighted in green. (***C***) Two 4 *m/z* range highlighting the simplified isotopic distributions, reduced ion distribution overlap, and increased S/N increase achieved using isotopically depleted ECD (*bottom*) over natural isotopic abundance ECD (*top*). Monoisotopic signals of fragment ions are highlighted with coloured asterisks (*) and the S/N for each fragment ion is highlighted. Further examples 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.

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 labelled, 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 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.

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*•-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.

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 labelled, with the ions assigned 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 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.

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**) or 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 dataset.

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*, 5ms 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*).
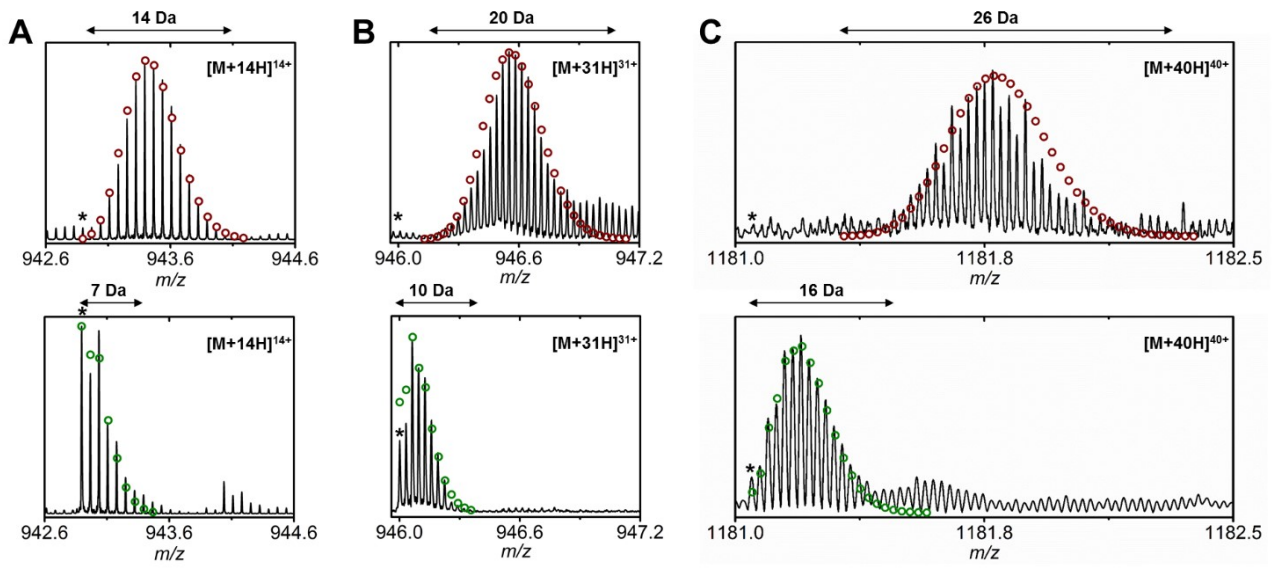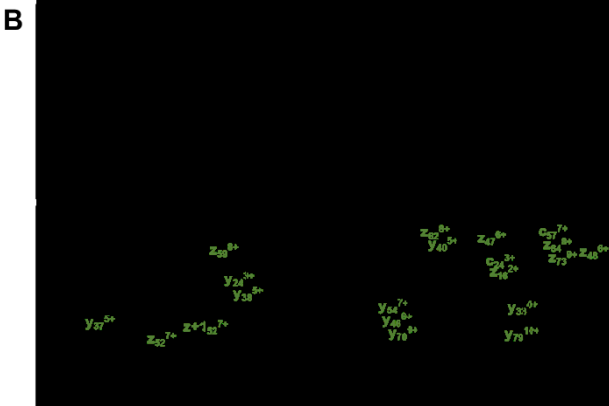
Figure 1.
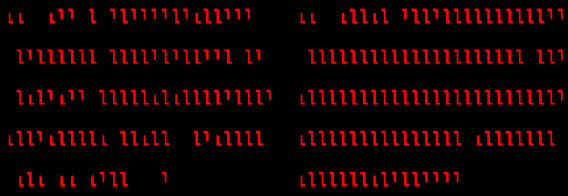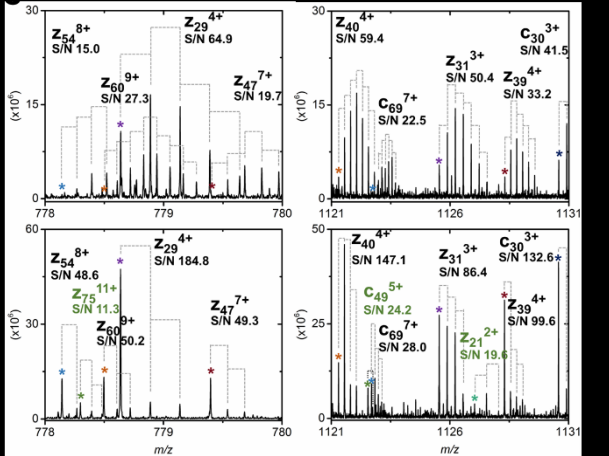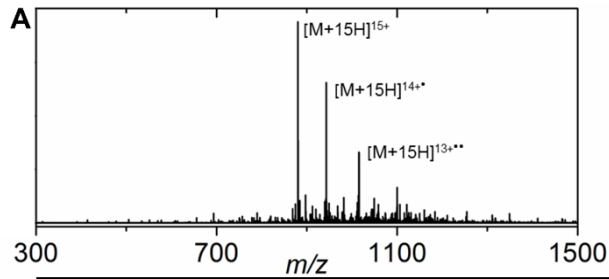
Figure 2
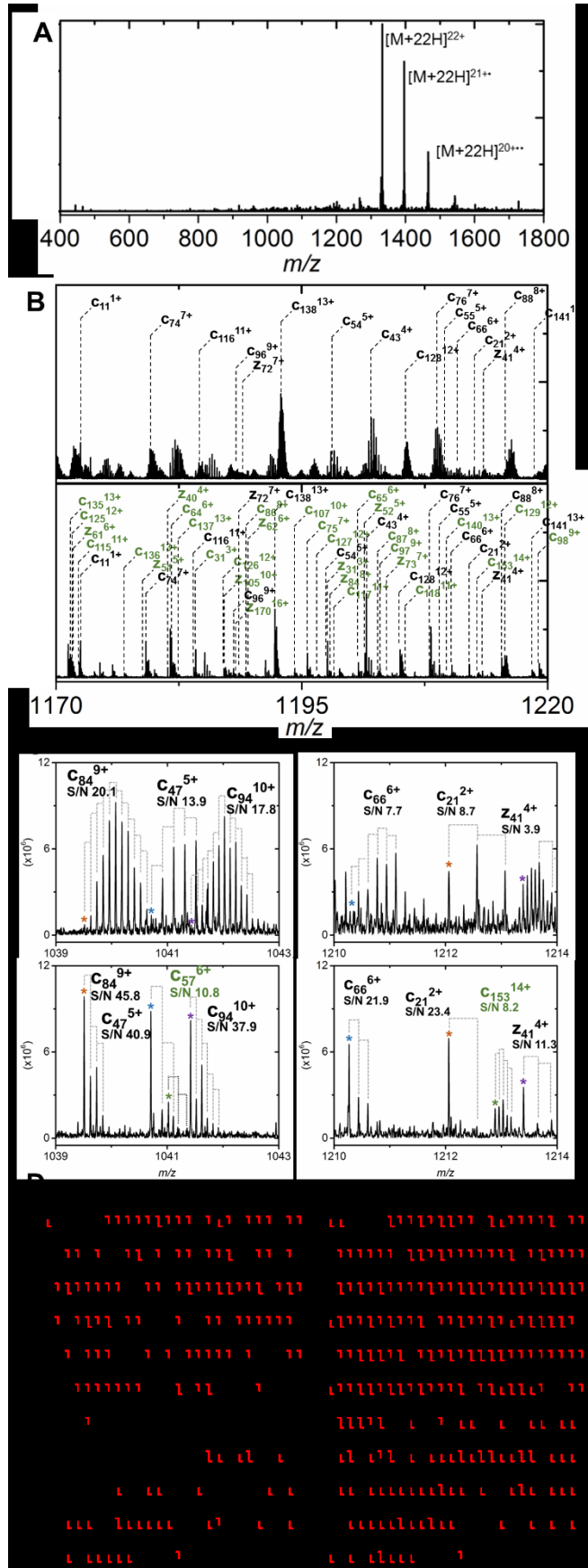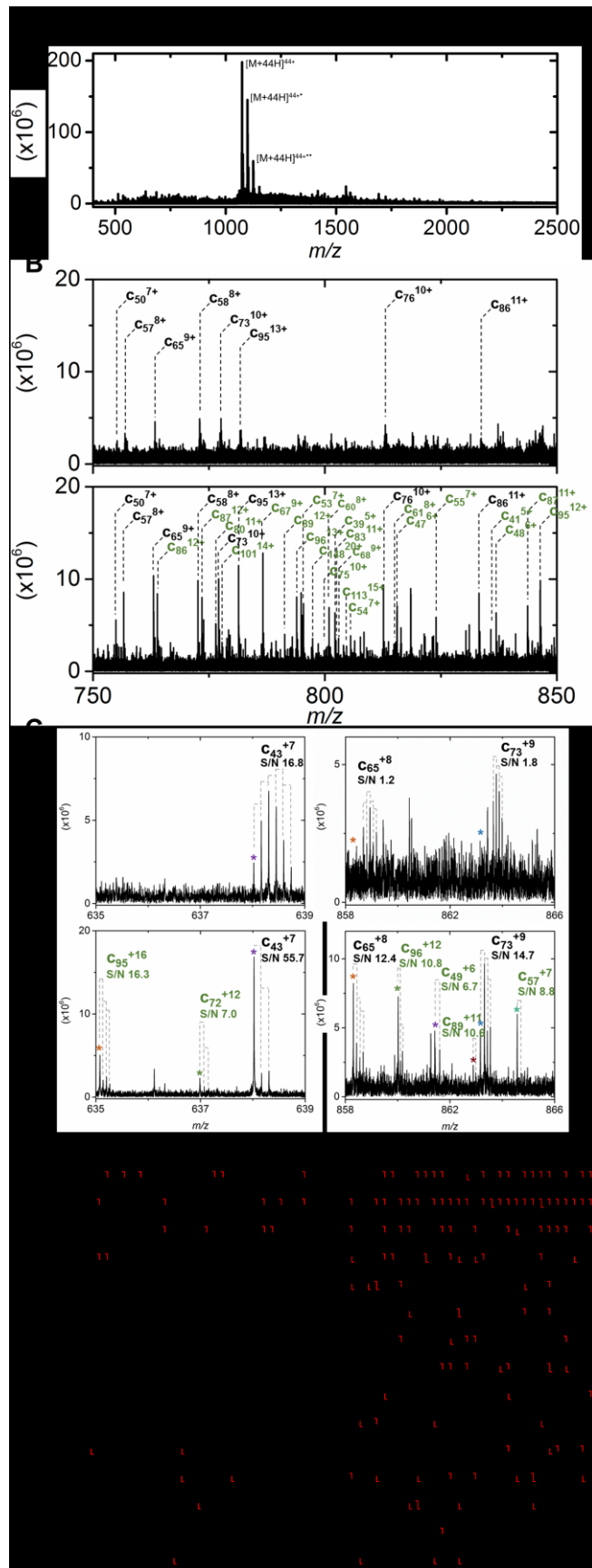
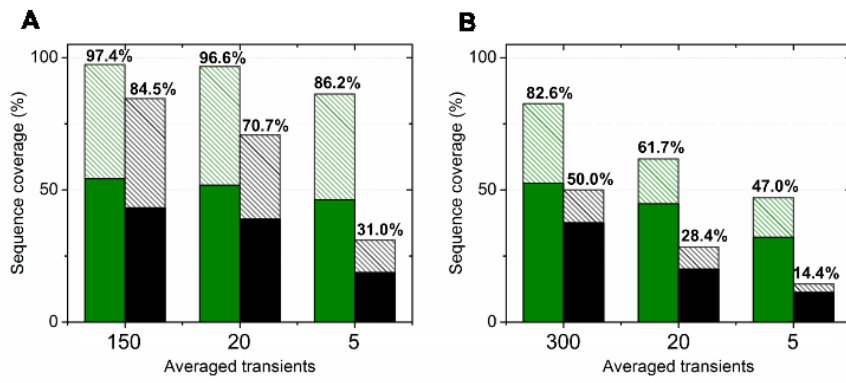**Figure 3**
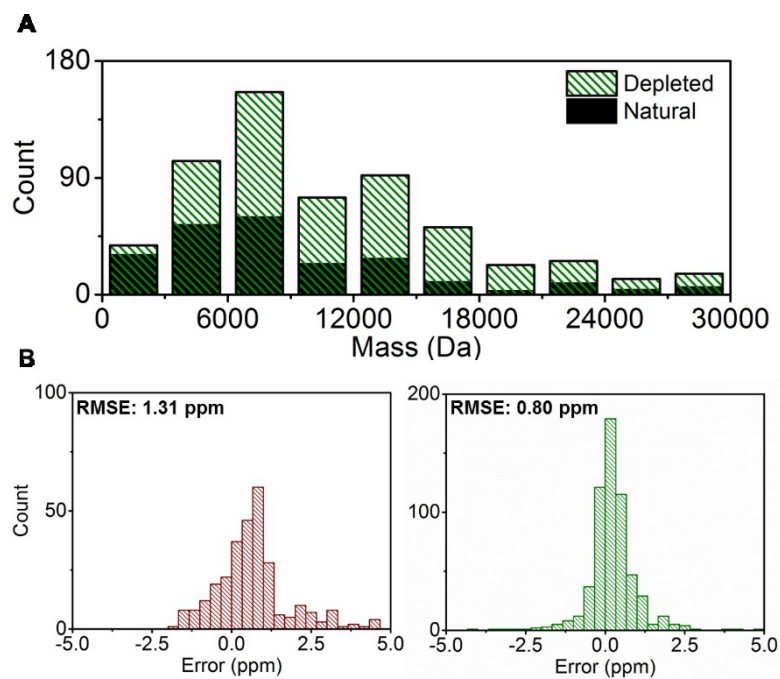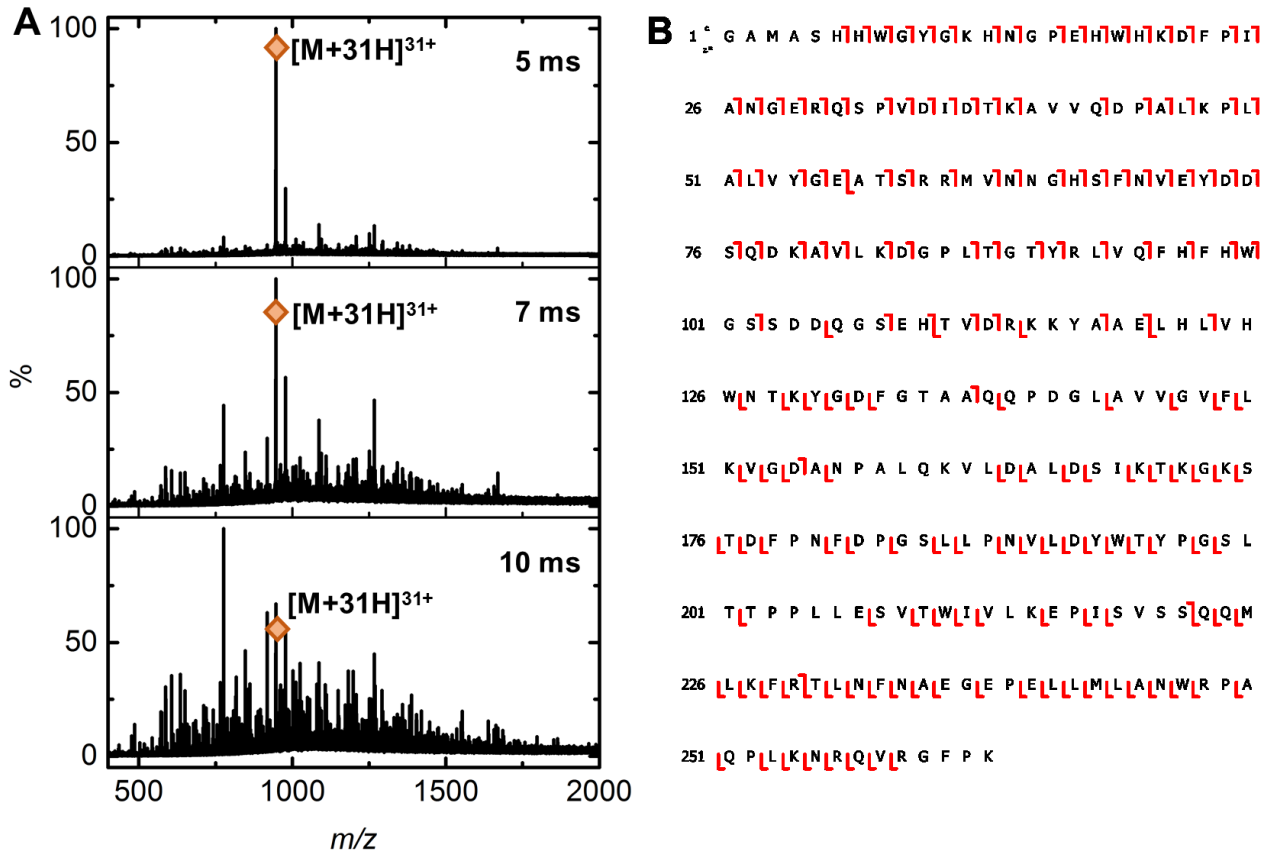
Figure 4

Figure 5

Figure 6

Figure 7

For Table of Content Use only

Isotope Depletion Mass Spectrometry (ID-MS) for accurate mass determination and improved top-down sequence coverage of intact proteins.

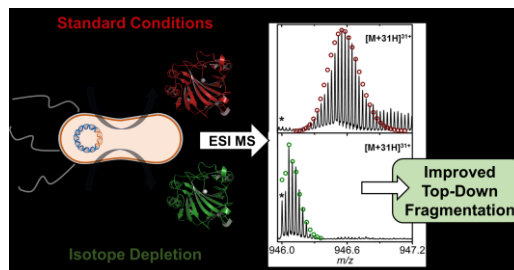Kelly J. Gallagher, Michael Palasser, Sam Hughes, C. Logan Mackay, David P. A. Kilgour and David J. Clarke

Table of Contents Graphic