Isotope Depletion Mass Spectrometry (ID-MS) for Enhanced Top-Down Protein Fragmentation.

Kelly J. Gallagher¹, Michael Palasser¹,2, Sam Hughes¹, C. Logan Mackay¹, David P. A. Kilgour³ and David J. Clarke¹*

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

² Present Address: Institut für Organische Chemie and Center for Molecular Biosciences Innsbruck (CMBI), Universität Innsbruck Innrain 80-82, 6020 Innsbruck, Austria.

³ Chemistry and Forensics, Rosalind Franklin Building, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, U.K.

*Corresponding author: dave.clarke@ed.ac.uk
Abstract.

Top-down mass spectrometry has become an important technique for the identification of proteins and characterisation of chemical and posttranslational modifications. However, as the molecular mass of proteins increases intact mass determination and top-down fragmentation efficiency become more challenging due to the partitioning of the mass spectral signal into many isotopic peaks. In large proteins, this results in reduced sensitivity and increased spectral complexity and signal overlap. This phenomenon is a consequence of the natural isotopic heterogeneity of the elements which comprise proteins (notably $^{13}$C). Here we present a bacterial recombinant expression system for the production of proteins depleted in $^{13}$C and $^{15}$N and use this strategy to prepare a range of isotopically depleted proteins. High resolution MS of isotope depleted proteins reveal dramatically reduced isotope distributions, which results in increases in sensitivity and deceased spectral complexity. We demonstrate that the monoisotopic signal is observed in mass spectra of proteins up to ~50 kDa. This allows confident assignment of accurate molecular mass, and facile detection of low mass modifications (such as deamidation).

We outline the benefits of this isotope depletion strategy for top-down fragmentation. The reduced spectral complexity alleviates problems of signal overlap; the presence of monoisotopic signals allow more accurate assignment of fragment ions; and the dramatic increase in single-to-noise ratio (up to 7-fold increases) permits vastly reduced data acquisition times. Together, these compounding benefits allow the assignment of ca. 3-fold more fragment ions than analysis of proteins with natural isotopic abundances. Thus, more comprehensive sequence coverage can be achieved; we demonstrate near single amino-acid resolution of the 29 kDa protein carbonic anhydrase from a single top-down MS experiment. Finally, we demonstrate that the ID-MS strategy allows far greater sequence coverage to be obtained in time limited top-down data acquisitions – highlighting potential advantages for top-down LC-MS/MS workflows and top-down proteomics.
Introduction

Top-down mass spectrometry (MS) has emerged as a powerful technique for the analysis of protein sequence and the detailed characterisation of chemical modifications to protein side-chains.[1,2] Consequently, top-down MS is a powerful strategy for the comprehensive identification and characterisation of all proteoforms arising from genetic variation, alternative splicing, and post-translational modifications (PTMs). The technique consists of first measuring the intact molecular mass of a protein, followed by gas phase fragmentation of a selected proteoform ion by tandem mass spectrometry. The resulting fragment ions are assigned based on their observed accurate mass. If sufficient numbers of fragments can be assigned, top-down MS can provide a complete description of protein sequence and PTM state.[3]

Over the last two decades several fragmentation techniques have been employed for top-down studies. However, electron-based fragmentation techniques such as electron capture dissociation (ECD)[4,5] and electron transfer dissociation (ETD)[6] have been the most widely applied and they offer improved diversity of backbone site cleaved when compared techniques which rely on vibrational excitation, such as collision induced dissociation (CID) and infrared multiphoton dissociation (IRMPD).[7,8] Thus top-down ECD and ETD can provide comprehensive sequence coverage for the analysis of small proteins (<20 kDa). However, as protein mass increases, top-down fragmentation efficacy notably diminishes; spectra become increasingly complex, and a series of other compounding factors result in reduced sequence coverage (the challenges of top-down MS have been discussed in depth in several recent publications).[9–11]

One fundamental factor which proves detrimental in top-down analysis is the increasing breadth of the isotopic distribution that accompanies increasing molecular mass. For proteins, the isotopic heterogeneity of the organic elements (particularly the ~1.1% natural abundance of $^{13}$C) results in the ion signal being spread over a distribution of discrete isotopologues (the isotope distribution); with each isotopologue differing in composition by a neutron. As protein mass increases, this isotope distribution widens, and so the overall signal gets more disperse. For example, a 10 kDa protein the isotope distribution will consist of ~12 isotopologue signals; whereas for a 50 kDa protein, the number of isotopologues observed can be over 40.[12] This phenomenon reduces the signal to noise ratio (S/N) and can lead to the overlapping of signals for species which are close in mass (e.g. proteoforms of the same protein with similar masses). Furthermore, 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 as protein mass increases.
One solution to this problem is the production of isotopically enriched/depleted proteins, by recombinant production in hosts grown on carbon/nitrogen sources that are enriched or depleted in specific naturally occurring heavy isotopes. The feasibility of this strategy has been demonstrated by Marshall et al., in 1997.[13] They reported the production and intact mass analysis of the 12 kDa FK506-binding protein, in media depleted in $^{13}$C and $^{15}$N. Using this approach,[13] the authors demonstrated an increase in the mass spectral sensitivity and the detection limit. Despite the publication of this seminal report over twenty years ago, the use of this strategy has been limited; despite theoretical studies highlighting the potential benefits of the approach.[9,14] The reason for this may be the technical difficulty in producing isotopically depleted proteins, and currently the strategy has only been applied to produce a handful of small proteins (<15 kDa).[15–17]

Herein we detail a robust method for the recombinant production of isotopically depleted protein in *E.coli* and demonstrate its benefits to top-down protein analysis by producing and characterizing 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 strategy, termed *isotope depletion mass spectrometry* (ID-MS), we show that the monoisotopic mass signals can be observed in isotopically depleted proteins up to ~50 kDa. This allows direct and accurate determination of molecular mass for large proteins and protein fragment ions, for the first time. Finally, we perform top-down fragmentation of isotopically depleted proteins and demonstrate that the reduced spectral complexity and increased S/N allow assignment of fragment ions with increased confidence, and results in dramatically improved sequence coverage.
Results and Discussion.

We chose three well-characterised proteins as model systems for this study – encapsulated ferritin (EncFtn, 13.2 kDa),[18] carbonic anhydrase (CA, 29.3 kDa),[19] and serine palmitoyltransferase (SPT, 47.3 kDa).[20] These proteins were recombinantly expressed in E. coli using M9 minimal growth media, containing glucose and ammonium sulfate as the sole carbon and nitrogen sources. This allowed isotopically doubly-depleted protein samples to be prepared by using isotopically-depleted glucose (99.9% ¹²C₆) and ammonium sulfate (99.99% ¹⁴N₂) in the cell culture preparation. Full details of the expression protocol can be found in the (Supporting Information, Figure S1).

ID-MS allows direct determination of monoisotopic mass of intact proteins up to 50 kDa.

After protein expression and purification, MS analysis of the intact proteins was performed using high resolution electrospray (ESI) Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). ESI mass spectra 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 when compared to natural isotopic abundance proteins (Figure 1).

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% ¹²C, 99.63% ¹⁴N; isotopically depleted abundances 99.90% ¹²C, 99.99% ¹⁴N). In each spectrum, the monoisotopic species is highlighted with an asterisk (*).
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, in the corresponding mass spectra obtained from isotopically depleted proteins, the monoisotopic peak was easily identifiable. For isotopically depleted CA, the monoisotopic peak accounted for ~13% of the total signal, and for the isotopically depleted SPT protein, the monoisotopic peak still accounts for ~2% of the total signal. Similar to EncFtn, the isotopically depleted variants of CA and SPT exhibited a reduction in isotopic distribution width of 20 Da to 10 Da and 26 Da to 16 Da respectively. In all case, the S/N also improved as the same number of proteins ions were partitioned between fewer isotopologue peaks.

As first proposed by Marshall in 1997,[13] we have demonstrated that $^{13}$C and $^{15}$N double depletion increases the proportion of the monoisotopic isotopologue and results in an observable monoisotopic signals for protein up to ~50 kDa. This allows the unambiguous and immediate assignment of the accurate molecular mass of 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.[21,22] 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, regardless of the resolution achieved in the data acquisition.[12,23] Consequently, this makes confident detection of low molecular mass PTMs, such as disulfide bond formation or deamidation, at the protein level, particularly challenging.[24–27] 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 in top-down fragmentation. Typically, top-down fragmentation generates many hundreds of fragment ions, with each ion appearing in multiple charge states, and exhibiting its own isotopic distribution. 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 forms of both EncFtn and CA, using both CID and ECD fragmentation. Initially, CID was performed on the [M+16H]^{16+} precursor ion of natural isotopic abundance and isotopically depleted 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_{374}^{+} and y_{105}^{+} 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 every isotopically depleted fragment ion’s isotope distribution. 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. 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 natural isotopic abundance spectrum (Supporting Information, Figure S5). This increase in the observed fragment ion number is similar to that demonstrated by Akashi et al., who reported an 63% increase in the number of assigned fragment ions when performing CID of an isotopically-depleted version of the 10 kDa protein cystatin. However, 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 (-H₂O, -CO, -NH₃). 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 S5).

The lack of product ion specificity, and the biased nature of fragmentation with CID has been well documented,[8] and this limits the utility of the technique for top-down studies of proteins over 10-15 kDa. 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.[29–31] 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.
Figure 2: ECD fragmentation of EncFtn (13 kDa). (A) 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. (B) A 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 figure S7. (C) The fragmentation maps (protein sequence coverage) achieved after ECD of the [M+16H]16+ charge state of: natural isotopic abundance EncFtn (left; 84.5%) and isotopically depleted EncFtn (right; 97.4%).

Figure 2 shows the spectrum obtained after ECD of EncFtn (spectral averaging of 150 acquired transients; magnitude mode). Post-acquisition, Autovectis was used to process the data in absorption mode and assign fragment ions (for details see Supporting Information, Figure S6). ECD of a single charge state of natural isotopic abundance EncFtn yielded 131 c-
ions and 125 z-ions. This included fragment ions from throughout the protein sequence and represented a total sequence coverage of 84.5% (Figure 2C, left); in our hands, this result is entirely typical for top-down ECD fragmentation of a 13 kDa protein. In comparison to ECD of natural isotopic abundance EncFtn, fragment ions obtained from ECD of isotopically depleted EncFtn displayed reduced isotopic distribution widths with a dominant monoisotopic signal and increased signal abundance (typically ~2-fold to 7-fold S/N increases were observed; dependant on fragment ion molecular mass). These factors allow assignment of many more low abundance ECD fragment ions, and accurate assignment of fragment ions which overlap in the natural isotopic abundance spectrum (Figure 2B and further examples in Supporting Information, Figure S7). In addition, 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). ECD of the isotopically depleted EncFtn yielded 276 c-ions and 220 z-ions fragment ions from this single experimental condition; a total sequence coverage of 97.4% (Figure 2C, right). Cleavages N-terminal to proline are not generally observed in ECD.[32] Remarkably, if this is taken into account, 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.

It is clear from our analysis of EncFtn that three characteristics of the isotope depletion MS strategy lead to dramatic improvements for top-down fragmentation - namely,

(i) improved overall S/N,
(ii) increased monoisotopic signal abundance, and
(iii) decreased isotope distribution width.

These compounding benefits should be more evident as the precursor protein mass increases over 20 kDa. Therefore, we tested the utility of top-down isotopically depleted MS at higher mass, analysing bovine CA (29 kDa) by top-down ECD. CA has been used extensively to characterise top-down fragmentations technologies by multiple research groups and on multiple MS platforms;[11,23,33–35] thus it constitutes an ideal model study.

Either the [M+32H]^{32+} (m/z 916) or the [M+22]^{22+} (m/z 1332) charge state of CA was isolated and subject to ECD (Figure 3). ECD of the [M+22H]^{22+} charge state of natural isotopic abundance CA produced highly complex spectra (20,000 peaks with S/N > 2.5), which exhibit overlapping fragment ion isotope distributions throughout the spectrum (Figure 3A,3B, top). In addition, following substantial spectral averaging (300 averaged transients), even more fragment ions were observed with low S/N, this was especially evident as fragment ion mass increased. In total, from this single dataset, 229 c- and z- fragment ions could be assigned, representing 50.0% sequence coverage (Figure 3C, left). Low sequence coverage was especially evident in the central region of the protein.
As expected, ECD of the isotopically depleted CA resulted in significantly reduced spectral complexity and fragment ion distribution overlap. Fragment ions were observed with increased S/N (~2- to 8-fold increase – similar to previous results, vide infra; Figure 3B 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 should be 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. approximately a three-fold increase in the number of fragment ions assigned from the natural isotopic abundance CA sample. These fragment ions yielded a sequence coverage of 82.6% for the isotopically depleted protein (Figure 3C, right). Comparable assignment rate increases were possible when analysing the ECD spectra of the [M+32H]^{32+} charge state of isotopically depleted CA (Supporting information, Figure S10). If the sequence coverage observed for both charge states are combined, the overall sequence coverage obtained for isotopically depleted CA was 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 S10). To our knowledge, represents the most comprehensive sequence coverage of CA observed to date, irrespective of fragmentation technique or MS platform.
Figure 3: ECD fragmentation of CA. (A) A region of the fragmentation spectra of IN (top) and isotopically depleted protein (bottom). The fragment ions assigned from each spectrum are labelled, with the ions found exclusively in the isotopically depleted spectrum highlighted in green. (B) A 4 m/z region allowing a 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 coloured asterisk (*) and the S/N for each ion is highlighted. Further examples in figure S9. (C) The fragmentation maps (protein sequence coverage) achieved after ECD of the [M+22H]^22+ charge state of natural isotopic abundance CA (left; 50%) and isotopically depleted CA (right; 82.6%).
One striking characteristic of the ECD of isotopically depleted proteins is the ability to assign extended stretches of complimentary c- and z- ions, even in central regions of larger proteins. In effect, allowing comprehensive fragment ion sequence coverage ‘deeper’ into the protein sequence. Comparison of the mass distributions of the fragment ions assigned after ECD of natural isotopic abundance and isotopically depleted CA show that more fragment ions are assigned in the isotopically depleted ECD spectrum from across all molecular mass ranges (Figure 4A). However, these histograms highlight that the ID MS strategy has the greatest benefit for the assignment of fragment ions of higher masses; where ECD of isotopically depleted CA consistently affords 3- to 8-fold more fragment ions than ECD of natural isotopic abundance CA. For example, in ECD of isotopically depleted CA resulted in 52 fragment ions in the mass range 15-18 kDa; whereas only 10 fragment ions of similar mass were assigned from natural isotopic abundance CA.

Not only is sequence coverage improved in isotopically depleted ECD, but fragment ions are also assigned with lower error in the isotopically depleted ECD spectrum. Figure 4B shows the distribution of errors for the assigned ECD spectra of natural isotopic abundance and isotopically depleted CA. Using a ‘poly-averagine’-based approach for deconvolution of the ECD spectrum of natural isotopic abundance CA, the resulting fragment ions were assigned with a RMS error of 1.306 ppm. While AutoVectis analysis of the ECD spectrum of the same charge state of isotopically depleted CA allowed assignment of fragment ions with a RMS error of 0.800 ppm.

The ability to assign dramatically more fragment ions, especially fragment ions of mass >10 kDa, is a direct consequence of the inherent increase in the S/N which accompanies isotopically depleted MS. In addition, the ability to directly observe the monoisotopic signals in isotopically depleted fragment ions is also highly advantageous, as it removes the requirement to obtain isotopic distributions with sufficient S/N for precise poly-averagine based deconvolution methods. Furthermore, the mass error introduced using the poly-averagine approximation during deconvolution is removed; leading to assignment of fragment ions with lower overall mass error. Therefore, higher confidence in fragment ion assignment can be achieved, which is particularly important for the interpretation of highly complex spectra, such as top-down analysis of large proteins or assigning branched protein ions.\[36\]
**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.

**ID-MS improves top-down ECD on an LC Timescale.**

One of the overarching goals of top-down mass spectrometry is to achieve comprehensive protein sequence coverage using spectral acquisition times that are compatible with front-end chromatography; potentially allowing top-down protein analysis to be used in an LC-MS/MS proteomics workflow. Significant advances have been made in this field of top-down proteomics in recent years;\(^{37–39}\) however, it is particularly challenging to achieve extensive sequence coverage as protein molecular mass increases, and the sequence coverage achieved in top-down LC-MS/MS experiments is often restricted to limited regions at the N- and C-termini of the protein. Although this is often sufficient to provide a ‘sequence-tag’ and allow protein identification, low sequence coverage is insufficient for confidently mapping protein modifications and full characterisation at the proteoform level. These limitations are due to the time-constraints of the experiment and the inability to perform the extensive spectral
averaging required to obtain fragment ion signals of sufficient abundance. In effect, there is a compromise that exists between the signal-to-noise 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.\[40\] Because of this non-linear relationship, the increased S/N inherent in our isotopically depleted MS approach should be particularly effective for increasing the fragment ion sequence coverage obtainable with limited spectral averaging. In order to investigate this, ECD spectra were acquired in the same fashion for both the natural isotopic abundance and isotopically depleted forms of EncFtn (13 kDa) and CA (29 kDa) using both 20 or 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 S11 and S12) and compared to the longer spectral acquisition time, described above (Figure 5). As expected, for natural isotopic abundance EncFtn reduction in the spectral averaging reduces the obtained protein sequence coverage significantly and 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, and 86.2% protein sequence coverage was achieved with only 5 averaged spectra. For the larger protein, CA (29 kDa), it is clear that 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. As discussed above, this phenomenon is well-documented in larger proteins, and is a current bottleneck in top-down proteomics. 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. These initial findings demonstrate the potential benefit of applying isotopically depleted strategies in top-down proteomic workflows and highlight the possibility of achieving comprehensive sequence coverage of larger proteins on chromatographic timescales.
Figure 5. 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.

Conclusion
We have produced several isotopically depleted 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. In addition, direct observation of the monoisotopic signal of isotopically depleted proteins is possible; allowing accurate molecular mass to be directly determined, even in large proteins. Applying ID-MS in conjunction with top-down fragmentation affords reduced spectral complexity, increased S/N and increased mass accuracy; together this allows assignment of dramatically more fragment ions (typically 2- to 3-fold) and consequently increased protein sequence coverage. We also highlight the potential of applying ID-MS for performing top-down fragmentation on a chromatographic timescale for top-down proteomic applications.

Finally, we note that this isotope depletion strategy is analogous to the isotope enrichment techniques which have become integral to biomolecular nuclear magnetic resonance (NMR) spectroscopy.\[^{[41]}\] Similarly, it is clear that ID-MS has huge promise for many biomolecular MS applications, particularly for proteins (or other biomolecules\[^{[42]}\]) of high molecular mass. Techniques such as hydrogen/deuterium exchange MS, native protein MS, and structural MS will all benefit greatly from the advantages which accompany isotopic depletion.
Experimental Section
Isotopically-depleted proteins were produced by recombinant expression in *E. coli* using minimal media supplemented with $^{12}\text{C}(99.9\%)$–glucose and $^{14}\text{N}(99.99\%)$-ammonium sulfate as the sole carbon and nitrogen sources; see Supporting Information for detailed protocols. MS experiments were performed on a 12T SolariX Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer equipped with an Infinity ICR cell (Bruker Daltonics, Bremen, Germany). Ionisation was achieved with a TriVersa NanoMate nanoelectrospray robot (Advion Bioscience, Ithaca, NY). Data was processed in magnitude mode, using Data Analysis (Bruker Daltonics, Bremen, Germany); and, in the absorption mode, using an in-house developed, enhanced version of AutoVectis (Nottingham Trent University and Spectroswiss Sàrl, Lausanne, Switzerland). Full details are available in the Supporting Information.

Datasets
All mass spectrometry datasets used in this study are available to download, in their original data formats, at Edinburgh DataShare ([www.http://datashare.is.ed.ac.uk/handle/10283/760](http://datashare.is.ed.ac.uk/handle/10283/760)), using the following link: [http://dx.doi.org/10.7488/ds/2446](http://dx.doi.org/10.7488/ds/2446).

Acknowledgements
This work was funded by the University of Edinburgh and a Royal Society Research Grant awarded to D.J.C. 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 the kind gift of bacterial expression plasmids.

Keywords (5)
FT-ICR MS, protein mass spectrometry, top-down mass spectrometry, electron capture dissociation, isotope depletion, proteomics.
References


