

Characterization of protein structure with ion mobility mass spectrometry, multiplexed fragmentation strategies and data directed analysis

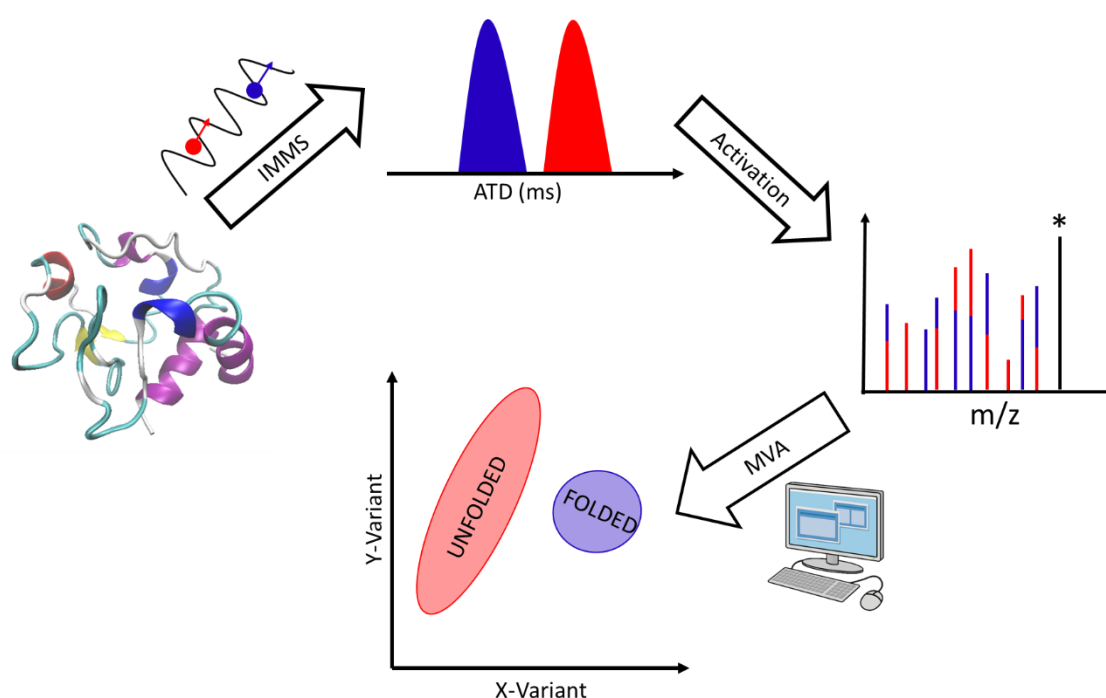
Rachelle Black¹, Alexey Barkhanskiy¹, Lennart A. I. Ramakers¹, Alina Theisen¹, Jeffery M. Brown²,
Bruno Bellina¹, Drupad Trivedi¹, Perdita E. Barran¹

¹Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute of Biotechnology,
131 Princess Street, Manchester, M1 7DN, United Kingdom

²Waters Corporation, Stamford Avenue, Altrincham Road, Wilmslow, SK9 4AX, United Kingdom

Corresponding author: perdita.barran@manchester.ac.uk

Abstract



Activated ion mobility measurements provide Insights to the stability of tertiary and quaternary structures of proteins and pairing such approaches with fragmentation can delineate which part(s) of the primary sequence are disrupted from a folded structure. In this work we use 213 nm

photodissociation coupled with ion mobility mass spectrometry and collisional activation to determine the conformational landscape of model proteins. UVPD experiments are performed on proteins following in source activation as well as on collisionally activated photoproducts post ion mobility separation. For cytochrome c, there is a significant increase in the fragmentation yield with collisional activation post mobility, for all conformational states. Similar strategies are deployed with the model multimeric proteins, concanavalin a, and haemoglobin. For these complexes' CID leads to classic asymmetric charge distribution in subunit products, which when preceded by UV irradiation yields fragments from within the sub-unit that can be mapped to the quaternary fold. Data driven, multivariate analysis (MVA) was used to determine the significant differences in UVPD and CID fragmentation pattern following in source activation. This data driven approach reveals diagnostic fragments without *a priori* assignments limited to predicated backbone cleavage and provides a new approach to map conformation landscapes that may have wider utility.

Introduction

Ion mobility mass spectrometry (IM-MS) is widely used to investigate protein structure and dynamics. Native mass spectrometry, performed under gentle ionizing conditions and using salty aqueous solutions, provides information in the form of charge state distributions, and ion mobility contributes size, charge, and shape separation. Together these data map the conformational landscape a gas-phase protein occupies under given experimental conditions¹⁻⁴. Drift times can be converted into rotationally averaged collision cross-sections, a buffer-gas dependent global size measurement which in turn can be compared to values obtained computationally from crystal structures or NMR experiments⁵. A large body of work has advanced our understanding of how solution structures and gas-phase structure might relate, and it is generally accepted that aspects of the native fold and topology can be preserved upon transfer into the gas-phase⁵⁻¹³.

A number of fragmentation strategies have been developed that are complementary to IM-MS, informing on structure, sequence and potential interactions^{10,14-16}. Collision-induced dissociation, the most widely available and commonly used technique, involves slow heating of the molecule via collisions with a neutral buffer gas until a dissociation threshold is reached, cleaving the most labile bonds first¹⁷. Due to the statistical redistribution of energy prior to dissociation, it is difficult to obtain topological information as proteins will unfold before fragmentation occurs.

Over recent years, ultraviolet photodissociation (UVPD) has become more widely used for many biomolecules from lipids and sugars to proteins and non-covalent complexes¹⁸⁻²². While different wavelengths enable different experimental strategies, most so called 'top-down' experiments are carried out at 213 or 193 nm, providing good sequence coverage throughout the entire protein^{23,24}. UVPD is sensitive to the rigidity of protein secondary structural elements and proline isomerization, inviting exploitation of UV photons for structural characterization²⁵⁻²⁷. At lower UV wavelength <220 nm, deposition of a single photon can be sufficient to cause electronic excitation and dissociation of protein molecules as the amide backbone becomes an efficient chromophore. However, the

fragmentation observed cannot be accounted for by direct dissociation only and it has therefore been proposed that a mix of direct dissociation and dissociation via internal conversion occurs²⁸. Fragments from peptides and proteins often arise due to backbone cleavages and are assigned using the Roepstorff and Fohlman nomenclature²⁹. In many workflows fragments are usually analysed either *via de novo* sequencing or comparing to a database of known proteins and their fragmentation patterns but restricted to backbone cleavages that contain either the N and C termini. These approaches are well utilised throughout academia and industry, although it has been predicted that they ignore up to 90% of the fragment ion intensity, which can arise from several sources including side chain losses, the presence of PTMs including non canonical, internal fragments and adductation.³⁰ When applied to top down sequencing and especially that for proteins in a native MS workflow, these exceptions become more prevalent, often necessitating manual assignment.³¹ Further to this, no automated assignment workflows take into account the conformation of the precursor ion. This is despite widespread evidence that the fragment spectra, and in particular the intensities of the fragment ions can be diagnostic of secondary and tertiary structure.^{10,32–34}

It is well known that the charge of an ion will vary according to the solvated conformation³⁵ and also IM-MS measurements have shown that a single precursor ion can adopt numerous conformations even within a single charge state.^{3,36} Fragmentation approaches tend to focus only on understanding the primary sequence of the protein and may go on to relate this to a 3D structure determined from X-ray crystallography,³⁷ however few approaches determine how the fragmentation yield alters depending on the original conformation or charge state. As previously reported, 213 nm UVPD can probe protein conformation, yielding different fragment spectra depending on the initial conformations according to IMMS measurements³⁸. The charge state of an ion typically results in a conformational difference when monitored *via* IM-MS,³⁹ when these different charge states are fragmented *via* CID there is an observable difference.⁴⁰

Although CID is the most common approach for inducing fragmentation, compaction and asymmetric fragmentation of subunits,^{41,42} it can limit the insights to multimeric proteins and alternative fragmentation methods may have more merit for structural studies.^{20,43–46}

Non-covalently bound fragments in ECD show as a reduced molecular ion and can, therefore, be readily distinguished from the precursor by m/z alone¹²; in UVPD however, they may be m/z coincident with the precursor ion and therefore not easily identifiable in the mass spectrum. While no studies known to us have yet investigated whether these ‘sticky’ fragments exist directly, Halim *et al.* used 213 nm UVPD on ubiquitin, either followed by or coincident with IR irradiation. For the +13 precursor ion, they reported an improvement in sequence coverage, although the yield of a-type fragments did not increase, indicating that UVPD did not result in ‘masked cleavage’ due to fragments remaining bound by non-covalent interactions.^{28,47} For Ubiquitin, the 13+ ion has been shown to exist in a fully extended conformation, and likely does not possess the tight network of non-covalent bonds present in compact, native-like structures and therefore these types of fragments should not be expected.^{48,49}

Principal component analysis (PCA) is commonly used as a data mining tool and forms the basis for multivariate data analysis. The main aim of such a method is to extract statistically relevant features from rich multidimensional datasets.⁵⁰ It is not limited by analytical technique and is widely used in metabolomics, however for proteomics, perhaps due to the success of algorithms that only consider backbone cleavages, it is rarely used. This is despite a significant portion of the mass spectra from top-down and bottom up proteomics remaining ‘dark’. PatternLab 4.0 has started to solve this problem; the software clusters similar proteomic profiles together and with a scoring system.⁵¹ Picotti and Reiter also recently introduced ‘chemoproteomics’ wherein they utilise machine learning to discern features indicative of drug binding and score them to identify protein targets for small molecules.⁵²

PatternLab 4.0 has been extended with the DiagnoProt module, which generates a PCA plot for each MS analysis. This also utilises machine learning to assess the spectral-peak intensities of tandem

mass spectrometry data that originates from HCD on precursor ion.⁵³ DiagnoProt can provide a quick view as to how similar two biological conditions are to one another via clustering of the fragmentation yield.^{54,55} These MS PCA workflows are independent of precursor conformations although PCA has been utilised for monitoring protein folding dynamics albeit with molecular dynamics and NMR.⁵⁰

In this work we show the utility of applying multiple activation strategies to probe the structure of proteins. Since such datasets are highly complex, we explore the use of multivariate analysis (MVA) for such analyses with model investigations that reveal that the fragment spectra robustly identify the difference between the ions from adjacent charge states and also between conformations adopted by a single charge state.

Experimental Section

Materials

Methanol was obtained from Sigma Aldrich (UK) with a purity of >99.9%. Ammonium acetate was purchased from Fisher Scientific (Loughborough, UK). Ultrapure water was produced by a Milli-Q Advantage A10 ultrapure water filtration system (Merck Millipore, Darmstadt, Germany). Bovine ubiquitin and bovine cytochrome c were purchased from Sigma Aldrich (UK) as lyophilized powders with purities of ≥98% and 95% respectively. The multimeric proteins, Jack Bean Concanavalin A and Human Haemoglobin were purchased from Sigma Aldrich (UK) and were dissolved in 200 mM ammonium acetate to a final concentration of 10 μM for UVPD and 20 μM. for CID experiments. Ubiquitin and cytochrome c were dissolved in 200 mM ammonium acetate and diluted to a final protein concentration of 10 μM for UVPD and 20 μM for CID experiments. Micro Bio-Spin 6 chromatography columns (Bio-Rad Laboratories, Hercules, CA, US) were used to desalt the Haemoglobin.

UVPD-IM-CID Experiments

The experimental setup for UVPD-IM-MS has been described in detail previously,³⁸ a schematic of the instrument can be found in SI Figure 1. Briefly, a laser beam is introduced into a Synapt G2-S through a CaF₂ window in the source block and directed onto the same pathway as the ion beam. Mass-selected ions are accumulated in the trap cell region of the instrument before ion mobility separation using a series of DC potential gates;⁵⁶ a mechanical shutter is then opened allowing the laser beam to interact with the trapped ions for a specified amount of time and all photo products are subsequently released into the IMS cell for ion mobility separation. Optionally, a collision voltage can be applied in the transfer cell post-IM to activate the mobility-separated photoproducts.

UVPD Workflow

All samples were ionized using nanoESI in positive ion mode. Monomeric proteins were sprayed from borosilicate glass capillaries (World Precision Instruments, Stevenage, UK) pulled in-house on a Flaming/Brown P-1000 micropipette puller (Sutter Instrument Company, Novato, CA, US) into which a platinum wire was inserted to apply a voltage to the solution. Multimeric proteins were sprayed from quartz glass capillaries on a Sutter Micropipette pulled (P-2000; Sutter Instrument Co.).

All source voltages were tuned for gentle ionization with a typical capillary voltage of 1.1-1.3 kV. Source temperature was altered for the species investigated, for monomeric proteins a source temperature of 40 °C for the native conditions, additionally, the trap bias setting was reduced to 30 V – 35 V to minimise protein activation. For multimeric the source temperature remained at 40 °C and trap gas was increased to improve the extent of multimeric species observed.

After ionization, the charge state of interest was m/z selected in the quadrupole and accumulated in the trap cell for 2 seconds to achieve an ion count of $\sim 2e3$. Monomeric protein ions were then photoactivated for 2 seconds by a 213 nm laser (CryLas, UK) operating at 1 kHz with an average pulse energy of $\sim 1 \mu J$. Photoproducts were then released into the IMS cell for mobility separation. Post-IM, ions were optionally activated by increasing the transfer collision energy (CE) voltage before proceeding to the ToF. Multimeric proteins were not trapped prior to irradiation, ubiquitin

data shows identical ATD for fragments produced via trapping and no trapping protocols (SI Figure 2), the multimeric protein was irradiated by a 213 nm laser (CryLas, Germany) operates at 1 kHz and 25 Hz and delivers 213 nm photons with an average pulse energy of 27 μ J. Multimeric proteins were investigated under a cone voltage of 10 V to yield the most folded tetrameric structure. The fragmentation difference was monitored over the course a series of different activation methods. UVPD-IM-CID was repeated at different cone voltages ranging from 10 V to 150 V for each monomeric protein. Experiments were performed in duplicate to confirm any observed trends. Fragmentation yield was calculated according to the relationship:

$$\text{Normalised fragmentation yield} = \frac{\sum \text{fragment ions}}{\sum \text{fragment ions} + \text{precursor}}$$

Data was analysed using MassLynx v4.1 (Waters Corporation, US), OriginPro 9.1 (OriginLab Corporation, US) and Microsoft Excel 2010 (Microsoft, US).

UVPD-MVA experiments were repeated at different cone voltages ranging from 10 V to 130 V to yield the folded and extended conformation. These were then processed via a MVA workflow.

CID Workflow

All samples were ionized using nanoESI in positive ion mode. These were sprayed from quartz glass capillaries on a Sutter Micropipette pulled (P-2000; Sutter Instrument Co.) into which a platinum wire was inserted to apply a voltage to the solution.

All samples were infused directly into a Waters Synapt G2-S mass spectrometer *via* nanoESI. All source voltages were tuned to maintain native-like conformations, capillary voltage was maintained at 1.1-1.3 kV with a source temperature of 60 °C and the sampling cone at 40 V. Trap gas was maintained at 2 mL/min for both protein samples. The charge state of interest was selected, for ubiquitin this was the 5+ and 6+ ions, these were collisionally activated at 60 V and 50 V respectively. For concanavalin a, dimer and tetramer ions were selected and collisionally activated at 60 V and 80 V respectively.

Multivariate analysis workflow

Raw data were processed using R ⁵⁷ and figures were produced using the package ggplot2.⁵⁸ Data points were calculated by combining scans and creating a single spectrum for each injection. To compare masses across different replicates, masses were binned to two decimal accuracies and combined into a single matrix. The final matrix was $m \times n$ structure where each row was a binned m/z and each column was corresponding ion count for each replicate. Partial least square – discriminant analysis was performed using package MixOmics ⁵⁹. Scores plot, ROC curve and box plots were used to visualise classification of data and to validate this classification, 3-fold cross validation with 100 repeats was performed.

Results and Discussion

1. Collisional activation post UVPD to probe restructuring of monomeric and multimeric proteins

1.1 Cytochrome c

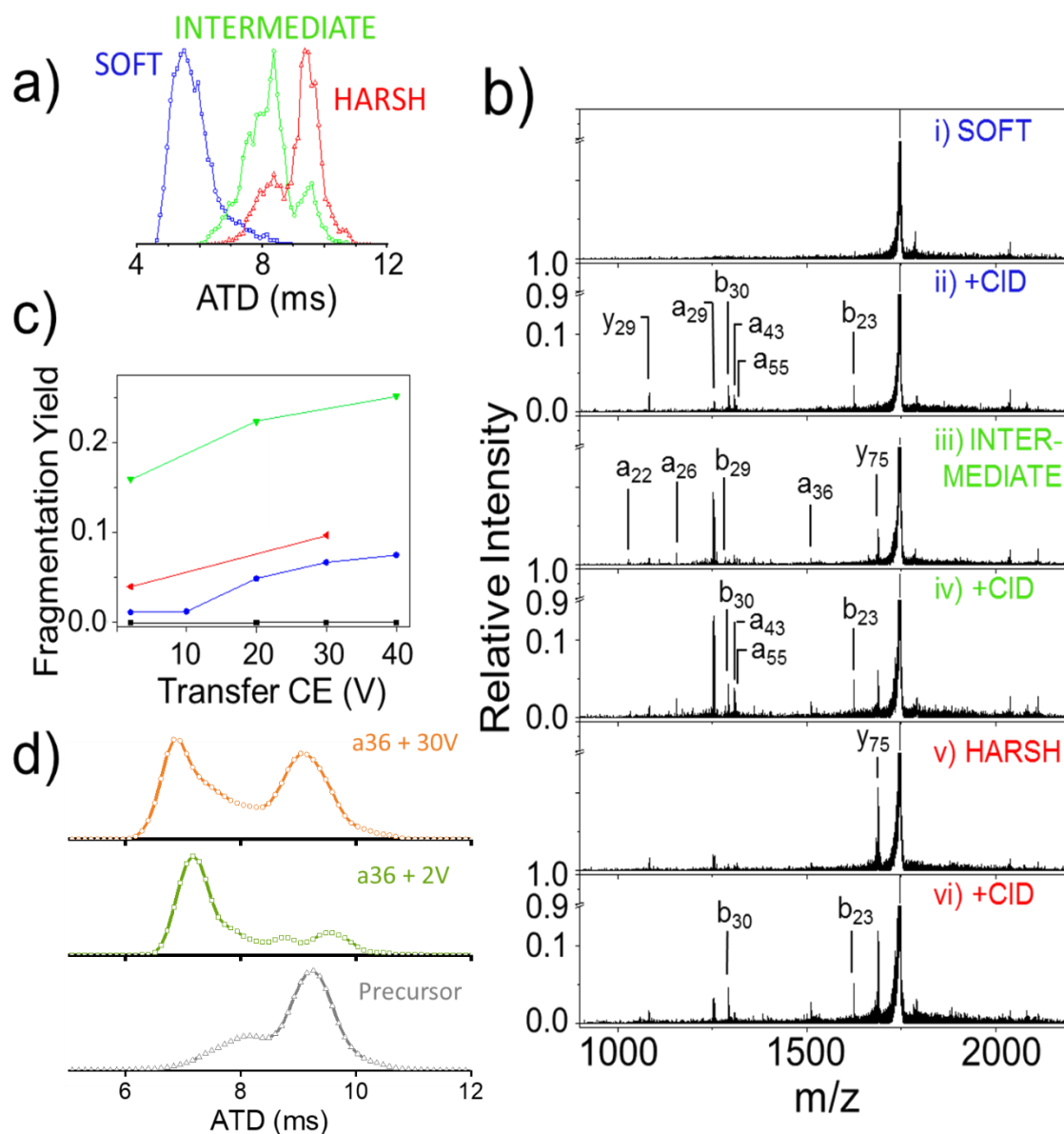


Figure 1: UVPD-IM-CID experiments performed on three conformational arrangements of $[M+7H]^{7+}$ of Cytochrome C. a) Arrival time distribution at soft (cone 10 V), intermediate (cone 85 V) and harsh (cone 120 V) source conditions. b) i) UVPD spectrum at soft conditions without and ii) with additional transfer CE of 30 V. iii) UVPD spectrum at intermediate conditions iv) with additional transfer CE of 40 V. v) UVPD spectrum at harsh conditions vi) with additional transfer CE of 30 V. c) Fragmentation

yield of a-type fragments as a function of transfer collision energy in soft (blue), intermediate (green), harsh (red) conditions and the CID-only control (black). d) Arrival time distribution of fragment a₃₆ from harsh conditions compared between transfer CE 2 V (green) and transfer CE 30 V (orange), the precursor is shown in grey.

Cytochrome c presents predominantly as ions $[M+6H]^{6+}$ and $[M+7H]^{7+}$ when sprayed from 200 mM ammonium acetate with most intensity in the latter. $[M+7H]^{7+}$ was mass selected for analysis and arrival time distributions were recorded for a range of cone voltages as shown in Figure 1a. Under soft source conditions, $[M+7H]^{7+}$ exists as a single, compact conformer (5.5 ms). Increasing the cone voltage to 85 V ('intermediate') resulted in a shift away from the compact conformer to more extended forms (8.4 and 9.4 ms). Further increase of the cone voltage to 120 V ('harsh') moves the intensity from the feature at 8.4 ms to 9.4 ms. At each discrete cone voltage, $[M+7H]^{7+}$ was subjected to UVPD as shown in Figure 1b producing a, b, c and y-type fragments. A full list of all fragments observed can be found in SI Table 1. When collisional activation was applied to photoproducts post ion mobility, an increase in the intensity of all fragments occurred along with the appearance of seven additional a- and b-type fragments from cleavage in residues 36 to 65, corresponding to a loop and two short alpha-helical regions deep in the protein core. We suggest collisional activation caused a partial unravelling of these secondary structural elements, breaking stabilising non-covalent interactions and the release of cleaved fragments. When the intermediate conformer was subjected to UVPD, a drastic increase in the intensity and the number of fragments occurs (Figure 1biii). UVPD-IM-CID increased the total number of fragments to 27 with better coverage between residues 22 to 55 and the appearance of b₆₅ (Figure 1biv). These results indicate that the helix between residues 50 and 56 is disrupted prior to the helix containing residue 65. Increasing the cone voltage to harsh conditions with 120 V resulted in a slight decrease in overall fragment intensity and a small increase in identified fragments from 20 in intermediate to 21 (Figure 1bv). In contrast to soft and intermediate, the b₆₅ ion is already present in the UVPD only data taken under harsh conditions albeit at low intensity, suggesting the increase in cone voltage was sufficient

to disrupt non-covalent bonds around this region. When collisional activation was applied post-IM, two additional fragments, b29 and c26, were identified, however, due to the low intensity in both the mass spectrum and the ion mobility we cannot exclude the possibility that these were produced from sequential fragmentation. Interestingly, no fragments from either the C- or N-terminal helices were found even after additional post-IM activation, supporting findings that these are the most stable structural elements.⁶⁰

The overall fragmentation yield increases for all source conditions with increasing transfer CE (SI Figure 3) as does the yield of only a-type fragments (Figure 1c). Since the ATD of the intermediate and the harsh source conditions both contain the same conformer but in different ratios, we used ion mobility to determine whether the increase in fragmentation yield in the harsh condition is a result of the remaining intermediate family or originates from the most extended one. Fragment a43, which exists in the UVPD spectra of both conditions, exhibits an increase in the portion of the ATD overlapping predominantly with the intermediate conformer, although a very slight increase is also found overlapping with the extended form (SI Figure 4). The ATD of fragment a36 shows a prominent increase in the arrival times overlapping with the most extended conformer as shown in Figure 1d. Overall, while we observe a fragment-dependent bias towards either intermediate or extended family, our results show that in contrast to ubiquitin (SI Figure 5), UVPD-IM-CID of the most extended conformer of cytochrome c after in-source activation can still increase fragmentation yield presumably by disrupting the remaining non-covalent interactions.

1.2 Concanavalin A

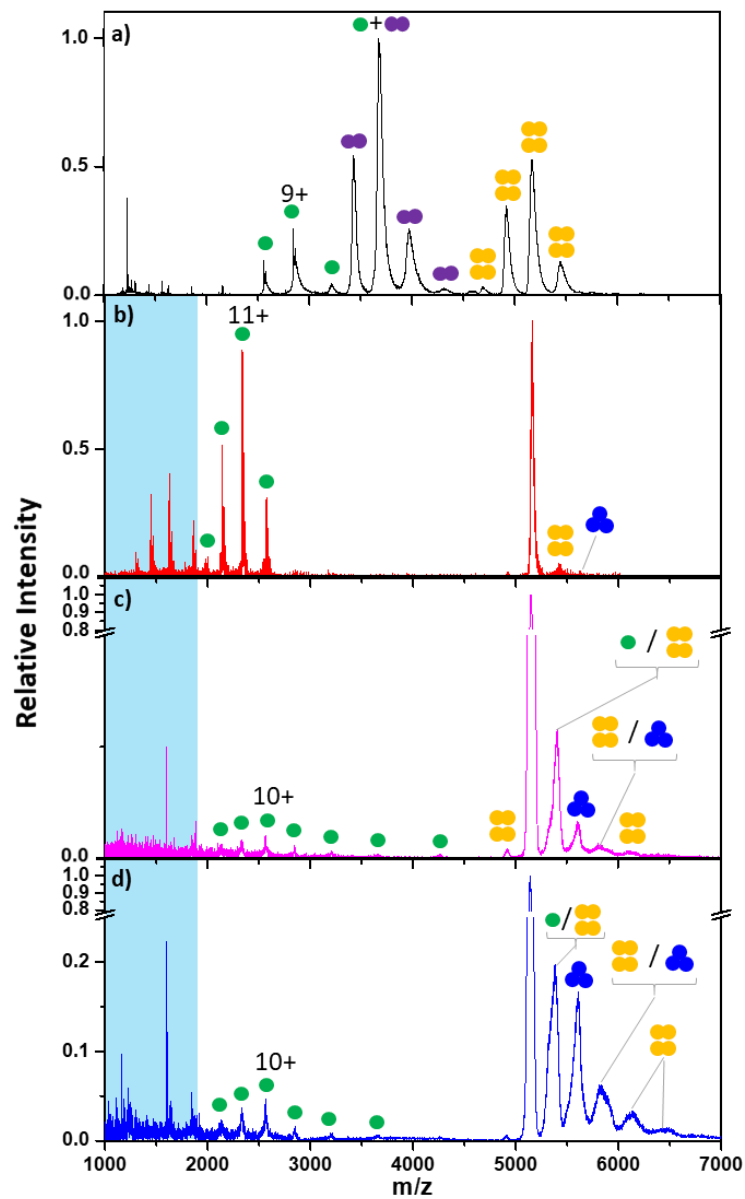


Figure 2: a) The full MS spectrum of the Concanavalin A (black), labelled to show the subunits along with the final tetrameric structure. b) CID spectra (red), at a collisional energy of 1.6 keV, of the $[4M+20H]^{20+}$ tetramer in the trap region of the mass spectrometer. UVPD spectra at a repetition rate of c) 250 Hz (pink) and d) 1 kHz (blue), of the $[4M+20H]^{20+}$ tetramer in the trap region of the mass spectrometer. Monomer, dimer, trimer and tetramer species indicated by green, purple, blue and yellow circles, respectively.

When sprayed from 200 mM ammonium acetate, jack bean concanavalin A gives a native-like mass spectrum consisting of monomer, homo-dimer and homo-tetramer peaks (**Error! Reference source not found.a**). Collisional activation was performed on both the dimeric (SI Figure 6) and tetramer (Figure 2b) species. Dimer activation yields mainly monomer species over a wide CSD, some charge stripping with additional fragments due to backbone cleavage also observed (SI Figure 6). Tetramer activation (Figure 2b) also shows significant production of monomer peaks, as previously described for native complexes. Alongside these monomeric species, CID the $[4M+20H]^{20+}$ yields low mass fragments in the mass range 1000-1800 m/z, due to fragmentation within the ejected monomer units (blue, **Error! Reference source not found.b**). Product ions at higher m/z than the $[4M+20H]^{20+}$ precursor are also observed, associated with charge stripped 19+ tetrameric ion species, consistent with a previous work on concanavalin A,⁶¹.

The $[4M+20H]^{20+}$ ion was isolated and fragmented *via* UVPD at 250 Hz (Figure 2c) and 1 kHz (Figure 2d) to monitor the effect of the repetition rate on the fragmentation spectrum. Similar to CID, the UVPD fragmentation spectra shows monomer subunits, trimer ions, charge stripped tetramer ions and a range of backbone cleavage ions. There are some evident differences between CID and UVPD spectra. The distributions of ejected monomers whilst centred on $[M+10H]^{10+}$ (UVPD) contain much lower charge states than those generated by CID consistent with more native-like monomeric forms (**Error! Reference source not found.a**). The abundance of these low charge species is greater when UVPD is carried out at 1 kHz (Figure 2d) and accompanied with a greater abundance of the backbone cleavage ions. The higher photon flux increases the probability of the precursor ion undergoing an ejection event and ejected subunit undergoing further UVPD fragmentation events. UVPD also produces charged stripped tetrameric species and trimeric species, and in addition to species observed in CID, the photo-fragmentation of the $[4M+20H]^{20+}$ concanavalin A ion yields a 5+ monomer subunit. With higher temporal photon flux (250 Hz vs. 1 kHz), the intensities of the 5+ monomer subunit, 14+ trimer ion and the 19+ tetramer ion increase by factors of ~8.7, ~8 and ~3.25, respectively and the average intensity of the other ejected monomer subunits and backbone cleavage ions are ~3-fold higher

(**Error! Reference source not found.**b). This is indicative of these processes proceeding *via* a multiphoton pathway.

Similar findings are obtained with haemoglobin (SI Figure 7). Here CID leads to the production of α and β globin monomer units, a portion of which are thought to be formed by the secondary ejection of the heme group after monomer ejection as well as a range of holo- and apo-trimeric species (SI Figure 7b). Experiments on Haemoglobin showed that the yield of the apo- and holo- forms of α - and β -globin monomers as well as sequence fragments, were highly dependent on the dissociation methodology and parameters (SI Figure 7). UVPD data, from different photon fluxes and repetition rates, along with collisionally released fragments seems to yield insights into the mechanism of dissociation within such complex systems, possibly allowing the strength of the inter-unit interactions to be explored.

2. MVA Strategies applied to fragmentation experiments in native mass spectrometry
- 2.1 Ubiquitin following CID comparing native charge states

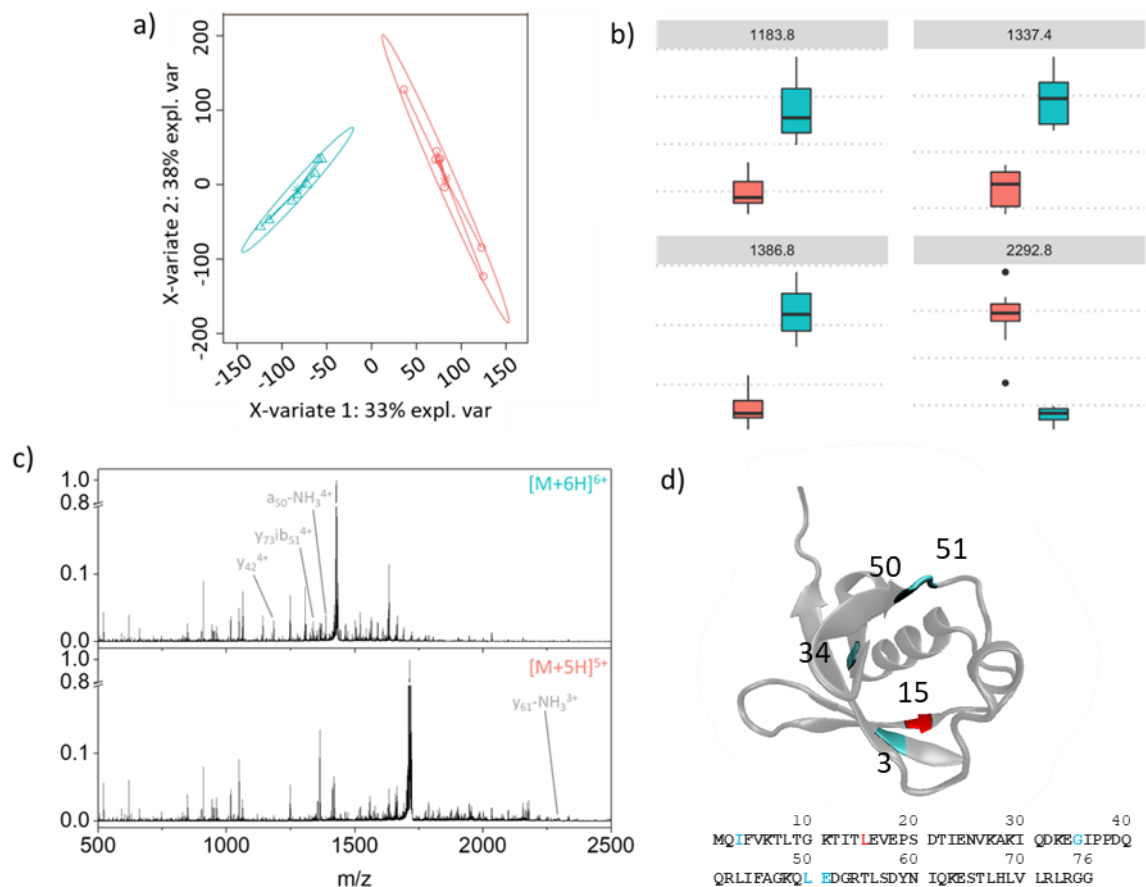


Figure 3: nESI-MS/MS-MVA of ubiquitin ions $[M+6H]^{6+}$ and $[M+5H]^{5+}$. a) Score plot for the tandem MS of the 5+ (red) and 6+ (cyan) charge states following CID. b) Boxplots for discriminative ions identified via MVA. c) CID spectra of $[M+6H]^{6+}$ (top) and $[M+5H]^{5+}$ (bottom) with identity of discriminant ions labelled. d) 3D structure of ubiquitin (PDB structure: 4Z9S), residues that undergo discriminative fragmentation between charge states from the MVA workflow are highlighted on the structure and the primary sequence shown below.

To develop a method to mine the complex data from native top down experiments, an MVA workflow was trained on Ubiquitin. The dominant native charge states $[M+6H]^{6+}$ and $[M+5H]^{5+}$ from gentle source conditions, were each selected and exposed to collisional activation (SI Figure 8). Experiments were performed in 8 replicates. MVA analysis was performed on these precursor ions to monitor for significant changes in the fragmentation pattern between the two charge states under different source activation conditions.

The activated charge states have different fragmentation patterns, this is evident from manual inspection as well as the MVA results, Figure 3a. The PLS-DA plot shows good separation and clustering for each charge state, with a significant variance within each charge states attributable to the slight differences between the repeats perhaps due to conformational differences. This separation reveals that even for adjacent charge states of a protein ion produced under gentle ionisation conditions, the precursor charge state restructures differently which in turn influences the fragmentation behaviour.

An advantage to an MVA method is that the total fragment ion spectral intensity can be monitored for significance. Features that are significant can be displayed via box plots that represent the difference in intensities when comparing the two precursor charge states, Figure 3b. There is a larger variance in these for the $[M+6H]^{6+}$ precursor ion, although this is always distinguishable from the $[M+5H]^{5+}$ precursor, inferring that particular fragments can be linked to precursor ion charge states a priori. These determined features are then annotated on the relevant MS spectrum, Figure 3c. Unlike the fully annotated MS spectrum (SI Figure 8), we here highlight fragments where there are statistically significant different intensities between the data from $[M+6H]^{6+}$ cf. $[M+5H]^{5+}$. These identified fragments were not evident by manual inspection. Whilst the example of ubiquitin is trivial, this approach could find utility in top down measurement of more exotic proteins, single point mutants, or *apo* and *holo* protein complexes.

Where the diagnostic features fragments can be mapped back to the primary sequence it can be overlay this onto the 3D crystal structure of the protein, Figure 3d, it is notable that the MVA approach applied here for CID data reveals statistically different fragments than previously observed with manual inspection of UVPD data.³⁸

2.2 Cytochrome c following in source activation and UVPD

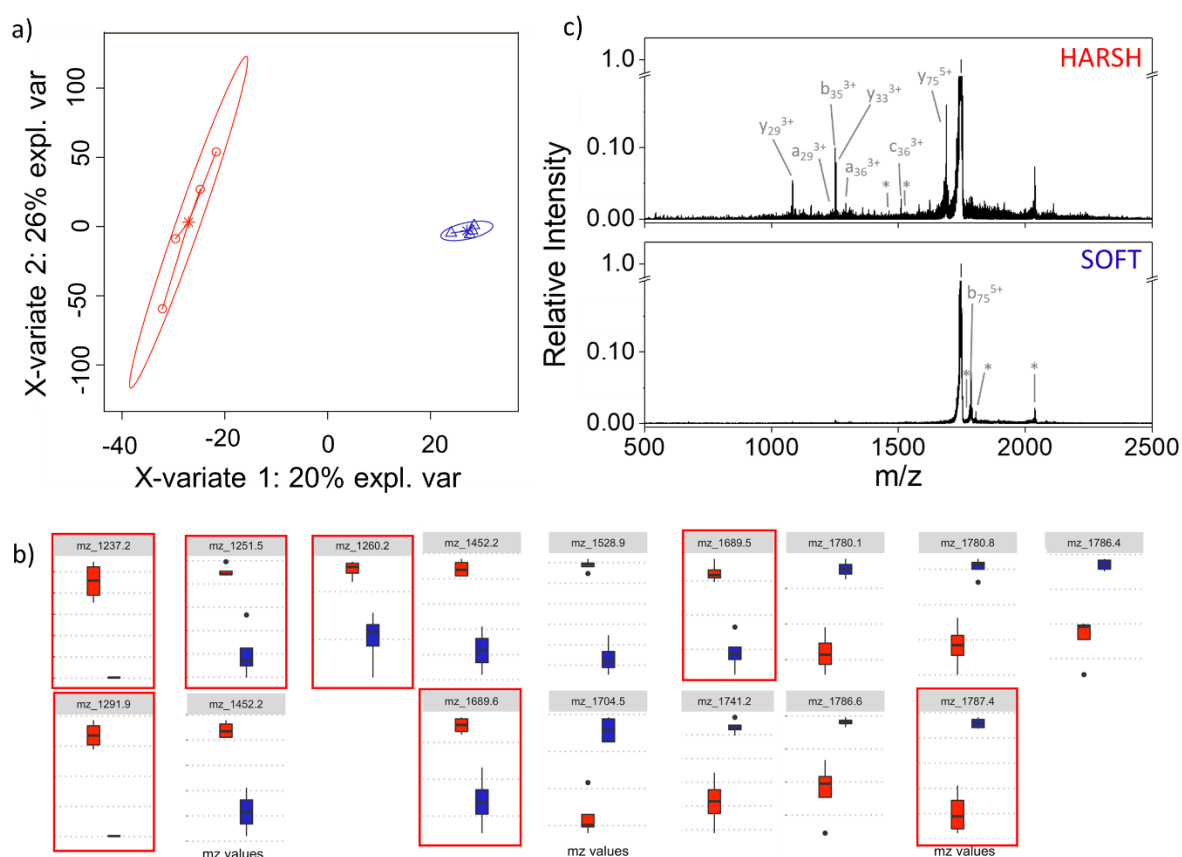


Figure 4: UVPD-IM-MVA of two conformations of $[M+7H]^{7+}$ Cytochrome c. a) Score plot for the soft (red) and harsh (blue) conditions exposed to 213 nm UVPD, b) Boxplots for the top 15 peaks identified via MVA, circled m/z represents the peaks that have been successfully identified, the remaining could not be isotopically resolved. c) UVPD spectrum at soft (cone 10 V) and harsh (cone 120 V) conditions without any additional transfer CE, labelled are the most prominent changes according to the MVA results.

To further explore the rationale for an MVA strategy we now turn to a comparison for the multi fragmentation workflow described above for cytochrome c, with a focus on variance between conformations found for a single charge state. Following soft and harsh in source activation the difference in the fragments is readily obtained with the MVA approach (Figure 4a). Data obtained under soft conditions clusters closely between experimental replicates, indicating little fluctuation in the precursor ion structure. Conversely the data obtained following harsher in source activation shows significant variance between the repeats. This indicates that the relative population of

structures in the activated precursor ions vary from replicate to replicate, although we took care to reproduce experimental conditions. The MVA protocol allows the fragments identified to be compared between these two conditions, they are then scored depending on the difference in intensities between the two conditions. The top 15 identified fragments are shown in Figure 4b, where circled peaks represent the ions that can be successfully identified in this data, the remaining were limited by the resolution of the instrument. It should be noted that there are some duplicates within the top 15 however the MVA picks up isotopically different peaks. These significant fragments all sit within regions where the tertiary structure is a less defined loop (SI Figure 9), in agreement with the previous data³⁸ and we can speculate that the unidentified fragments also sit within the less defined structural elements.

The relationship between fragments and conformation has been exploited previously to show the capability of UVPD to directly probe the conformation,³⁸ this MVA procedure enables a data directed approach. The fragments that show the greatest fluctuation between the two conditions can act as the identifier features to indicate when and where structural perturbation of a protein has occurred. Fragments identified *via* MVA that cannot be assigned are just as important and provide new reproducible features that signify the conformational state. The top 15 fragments are labelled on the MS data in Figure 4c, and have significantly different intensity dependent on activation status, which is also shown by precursor ATDs (SI Figure 10). Unassignable fragments are also labelled and combine with charge state, ATD and assigned fragments to provide a fingerprint of the conformational state.

Conclusions

We have shown that multiple fragmentation strategies coupled with ion mobility mass spectrometry can provide insights to protein structure and stability for model monomeric and tetrameric proteins. This is further revealed by a data driven analysis which is not restricted by fragments from backbone cleavage nor by the need to possess an N or C terminal amino acid. Collisional activation of

photoproducts post ion mobility separation releases non-covalently linked fragments even from compact (native) conformers. Multiple protein conformers co-exist in intermediate and harsh source conditions, and these strategies combined with the ability to map fragments to the ion mobility data facilitate a conformer directed data analysis approach. This can yield insights to the stability of the fold; for Cytochrome c even the most extended conformer released a-type (UVPD) fragments following UVPD-CID.

MVA was applied to fragmentation data from the two charge states of ubiquitin obtained under classical native conditions and clearly shows the differences between adjacent charge states implying that even these un-activated structural ensembles are different. MVA analysis was also performed on UVPD data from source activated proteins, and this indicates how the precursor ion conformation dictates which fragments are observed.

MVA analysis has provided new insights to conformer dependent fragmentation. We have shown distinct separation in trivial examples (charge state and in source activation), but given the sensitivity of modern mass spectrometers, such an approach could readily be applied as a screening tool, identifying features that alter due to PTMS, cofactor binding, or following the directed evolution of enzymes.

Acknowledgements

This work was supported by BBSRC grants BB/L002655/1, BB/L016486/1 and BB/M01108/1 as well as studentship awards to AT, RB with additional financial support from Waters Corp who have also part supported AB and PB.

PB and LAIR are grateful to the MS SPIDOC project funded by the European Union's Horizon 2020 FET-OPEN research and innovation programme (Grant agreement No. 801406) for support of ongoing photo activation IM-MS activities

AB acknowledges the EPSRC, BBSRC and AstraZeneca plc. for funding under the Prosperity Partnership EP/S005226/1 and

References

- 1 D. E. Clemmer, R. R. Hudgins and M. F. Jarrold, *J. Am. Chem. Soc.*, 1995, **117**, 10141–10142.
- 2 T. Wyttenbach, G. Von Helden and M. T. Bowers, *J. Am. Chem. Soc.*, 1996, **118**, 8355–8364.
- 3 R. Beveridge, L. G. Migas, K. A. P. Payne, N. S. Scrutton, D. Leys and P. E. Barran, *Nat. Commun.*, 2016, **7**, 1–9.
- 4 R. Beveridge, S. Covill, K. J. Pacholarz, J. M. D. Kalapothakis, C. E. Macphée and P. E. Barran, *Anal. Chem.*, 2014, **86**, 10979–10991.
- 5 E. Jurneczko and P. E. Barran, *Analyst*, 2011, **136**, 20–28.
- 6 Z. Hall, A. Politis and C. V Robinson, *Structure*, 2012, **20**, 1596–609.
- 7 J. A. Loo, J. X. He and W. L. Cody, *J. Am. Chem. Soc.*, 1998, **120**, 4542–4543.
- 8 Z. Deng, N. Thontasen, N. Malinowski, G. Rinke, L. Harnau, S. Rauschenbach and K. Kern, *Nano Lett.*, 2012, **12**, 2452–2458.
- 9 D. Hewitt, E. Marklund, D. J. Scott, C. V Robinson and A. J. Borysik, *J. Phys. Chem. B*, 2014, **118**, 8489–95.
- 10 N. C. Polfer, K. F. Haselmann, P. R. R. Langridge-smithy and P. E. Barran, *Mol. Phys.*, 2005, **103**, 1481–1489.
- 11 S. R. Harvey, M. Porrini, A. Konijnenberg, D. J. Clarke, R. C. Tyler, P. R. R. Langridge-Smith, C. E. MacPhee, B. F. Volkman and P. E. Barran, *J. Phys. Chem. B*, 2014, **118**, 12348–12359.
- 12 K. Breuker, H. Oh, D. M. Horn, B. A. Cerda and F. W. McLafferty, *J. Am. Chem. Soc.*, 2002, **124**, 6407–6420.
- 13 T. Wyttenbach and M. T. Bowers, *Annu. Rev. Phys. Chem.*, 2007, **58**, 511–33.
- 14 J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz and D. F. Hunt, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9528–33.
- 15 R. A. Zubarev, D. M. Horn, E. K. Fridriksson, N. L. Kelleher, N. A. Kruger, M. A. Lewis, B. K. Carpenter and F. W. McLafferty, *Anal. Chem.*, 2000, **72**, 563–73.
- 16 A. S. Galhena, S. Dagan, C. M. Jones, R. L. Beardsley and V. H. Wysocki, *Anal. Chem.*, 2008, **80**,

- 1425–36.
- 17 N. L. Kelleher, *Anal. Chem.*, 2004, **76**, 196 A–203 A.
- 18 M. R. Robinson, J. M. Taliaferro, K. N. Dalby and J. S. Brodbelt, *J. Proteome Res.*, 2016, **15**, 2739–2748.
- 19 A. Racaud, R. Antoine, L. Joly, N. Mesplet, P. Dugourd and J. Lemoine, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 1645–1651.
- 20 L. J. Morrison and J. S. Brodbelt, *J. Am. Chem. Soc.*, 2016, **138**, 10849–10859.
- 21 T. Ly and R. R. Julian, *J. Am. Chem. Soc.*, 2010, **132**, 8602–8609.
- 22 M. A. Halim, L. MacAleese, J. Lemoine, R. Antoine, P. Dugourd and M. Girod, *J. Am. Soc. Mass Spectrom.*, 2017, 1–14.
- 23 J. B. Shaw, W. Li, D. D. Holden, Y. Zhang, J. Griep-Raming, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher and J. S. Brodbelt, *J. Am. Chem. Soc.*, 2013, **135**, 12646–12651.
- 24 J. R. Cannon, M. B. Cammarata, S. A. Robotham, V. C. Cotham, J. B. Shaw, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher and J. S. Brodbelt, *Anal. Chem.*, 2014, **86**, 2185–2192.
- 25 T.-Y. Kim, S. J. Valentine, D. E. Clemmer and J. P. Reilly, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 1455–65.
- 26 S. Warnke, C. Baldauf, M. T. Bowers, K. Pagel and G. Von Helden, *J. Am. Chem. Soc.*, 2014, **136**, 10308–10314.
- 27 M. B. Cammarata and J. S. Brodbelt, *Chem. Sci.*, 2015, **6**, 1324–1333.
- 28 R. R. Julian, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1823–1826.
- 29 P. Roepstorff and J. Fohlman, *Biol. Mass Spectrom.*, 1984, **11**, 601–601.
- 30 J. A. Taylor and R. S. Johnson, *Anal. Chem.*, 2001, **73**, 2594–2604.
- 31 A. S. Phillips, A. F. Gomes, J. M. D. Kalapothakis, J. E. Gillam, J. Gasparavicius, F. C. Gozzo, T. Kunath, C. MacPhee and P. E. Barran, *Analyst*, 2015, **140**, 3070–3081.
- 32 J. M. D. Kalapothakis, Y. Berezovskaya, C. G. Zampronio, P. A. Faull, P. E. Barran and H. J. Cooper, *Chem. Commun.*, 2014, **50**, 198–200.

- 33 B. T. Ruotolo, S. J. Hyung, P. M. Robinson, K. Giles, R. H. Bateman and C. V. Robinson, *Angew. Chemie - Int. Ed.*, 2007, **46**, 8001–8004.
- 34 J. T. Seffernick, S. R. Harvey, V. H. Wysocki and S. Lindert, *ACS Cent. Sci.*, 2019, **5**, 1330–1341.
- 35 L. Konermann, E. Ahadi, A. D. Rodriguez and S. Vahidi, *Anal. Chem.*, 2012, **85**, 2–9.
- 36 D. Stuchfield and P. Barran, *Curr. Opin. Chem. Biol.*, 2018, **42**, 177–185.
- 37 M. Schennach and K. Breuker, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 1059–1067.
- 38 A. Theisen, R. Black, D. Corinti, J. M. Brown, B. Bellina and P. E. Barran, *J. Am. Soc. Mass Spectrom.*, 2018, **30**, 24–33.
- 39 M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson and B. T. Ruotolo, *Anal. Chem.*, 2010, **82**, 1–9.
- 40 G. E. Reid, J. Wu, P. A. Chrisman, J. M. Wells and S. A. McLuckey, *Anal. Chem.*, 2001, **73**, 3274–3281.
- 41 J. C. Jurchen and E. R. Williams, *J. Am. Chem. Soc.*, 2003, **125**, 2817–2826.
- 42 J. C. Jurchen, D. E. Garcia and E. R. Williams, *J. Am. Soc. Mass Spectrom.*, 2004, **15**, 1408–1415.
- 43 S. Tamara, A. Dyachenko, K. L. Fort, A. A. Makarov, R. A. Scheltema and A. J. R. Heck, *J. Am. Chem. Soc.*, 2016, **138**, 10860–10868.
- 44 S. N. Sipe and J. S. Brodbelt, *Phys. Chem. Chem. Phys.*, 2019, **21**, 9265–9276.
- 45 M. Sharon, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 487–500.
- 46 I. Sinelnikov, E. N. Kitova and J. S. Klassen, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 617–631.
- 47 M. A. Halim, M. Girod, L. MacAleese, J. Lemoine, R. Antoine and P. Dugourd, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 1435–1442.
- 48 J. Li, J. A. Taraszka, A. E. Counterman and D. E. Clemmer, *Int. J. Mass Spectrom.*, 1999, **185–187**, 37–47.
- 49 H. Oh, K. Breuker, S. K. Sze, Y. Ge, B. K. Carpenter and F. W. McLafferty, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 15863–8.

- 50 G. G. Maisuradze, A. Liwo and H. A. Scheraga, *J. Mol. Biol.*, 2009, **385**, 312–329.
- 51 P. C. Carvalho, D. B. Lima, F. V. Leprevost, M. D. M. Santos, J. S. G. Fischer, P. F. Aquino, J. J. Moresco, J. R. Yates and V. C. Barbosa, *Nat. Protoc.*, 2016, **11**, 102–117.
- 52 I. Piazza, N. Beaton, R. Bruderer, T. Knobloch, C. Barbisan, L. Chandat, A. Sudau, I. Siepe, O. Rinner, N. de Souza, P. Picotti and L. Reiter, *Nat. Commun.*, 2020, **11**, 1–13.
- 53 D. B. Lima, A. R. F. Silva, M. Dupré, M. D. M. Santos, M. A. Clasen, L. U. Kurt, P. F. Aquino, V. C. Barbosa, P. C. Carvalho, J. Chamot-Rooke and J. Hancock, *Bioinformatics*, 2019, **35**, 3489–3490.
- 54 A. R. F. Silva, D. B. Lima, A. Leyva, R. Duran, C. Batthyany, P. F. Aquino, J. C. Leal, J. E. Rodriguez, G. B. Domont, M. D. M. Santos, J. Chamot-Rooke, V. C. Barbosa and P. C. Carvalho, *Bioinformatics*, 2017, **33**, 1883–1885.
- 55 J. M. Silva, H. H. Wippel, M. D. M. Santos, D. C. A. Verissimo, R. M. Santos, F. C. S. Nogueira, G. A. R. Passos, S. L. Sprengel, L. A. B. Borba, P. C. Carvalho and J. de S. d. G. Fischer, *Sci. Rep.*, 2020, **10**, 1–11.
- 56 B. Bellina, J. M. Brown, J. Ujma, P. Murray, K. Giles, M. Morris, I. Compagnon and P. E. Barran, *Analyst*, 2014, **139**, 6348–6351.
- 57 R. C. Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, <http://www.r-project.org/>, (accessed 13 August 2020).
- 58 H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer US, New York, 2009.
- 59 K. A. Lê Cao, I. González and S. Déjean, *Bioinformatics*, 2009, **25**, 2855–2856.
- 60 M. M. G. Krishna, H. Maity, J. N. Rumbley and S. W. Englander, *Protein Sci.*, 2007, **16**, 1946–56.
- 61 M. Zhou, S. Dagan and V. H. Wysocki, *Analyst*, 2013, **138**, 1353–1362.