International Journal of Mass Spectrometry 464 (2021) 116588

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Characterization of native protein structure with ion mobility mass spectrometry, multiplexed fragmentation strategies and multivariant analysis

Rachelle Black [a], Alexey Barkhanskiy [a], Lennart A.I. Ramakers [a], Alina Theisen [a], Jeffery M. Brown [b], Bruno Bellina [a], Drupad K. Trivedi [a], Perdita E. Barran [a, *]

[a] Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute of Biotechnology, 131 Princess Street, Manchester, M1 7DN, United Kingdom
[b] Waters Corporation, Stamford Avenue, Altrincham Road, Wilmslow, SK9 4AX, United Kingdom

ARTICLE INFO

Article history:
Received 8 November 2020
Received in revised form
9 February 2021
Accepted 15 March 2021
Available online 21 March 2021

Keywords:
Ion mobility mass spectrometry
Proteins
UVPD
PCA

ABSTRACT

Activated ion mobility measurements provide insights to the stability of tertiary and quaternary structures of proteins and when coupled with dissociation strategies can delineate how the fold is disrupted. In this work, we use 213 nm photodissociation coupled with ion mobility mass spectrometry and collisional activation to probe the conformational landscape of model proteins. UVPD experiments are performed on proteins following in source activation and on collisionally activated photoproducts post ion mobility separation. For all observed conformations of cytochrome c, there is a significant increase in the UVPD fragmentation yield with the addition of collisional activation post mobility. Similar strategies are deployed with the multimeric proteins, concanavalin a, and haemoglobin. For these complexes' CID results in 'classical' asymmetric charge distribution in subunit products, which when preceded by UV irradiation, yields fragments from within sub-units. Combining these strategies provides complex multidimensional datasets, rich in information, which here we mine with multivariate analysis (MVA). This approach readily determines differences in UVPD and CID fragmentation patterns as a function of conformation and reveals diagnostic information about the precursor native structure without the limitations of current methods that only assign backbone cleavage sites.

1. Introduction

Ion mobility mass spectrometry (IM-MS) is widely used to investigate protein structure and dynamics. Native mass spectrometry, performed under gentle ionizing conditions, 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 conditions [1–4]. 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 experiments [5]. 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 at least some aspects of the native fold and topology can be preserved upon transfer into the gas-phase [5–13]. Several 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 first [17]. Due to the statistical redistribution of energy before dissociation, it is difficult to obtain topological information as proteins will restructure before fragmentation occurs [18,19] and alternative fragmentation methods may have more merit for structural studies [20–24].

Over recent years, ultraviolet photodissociation (UVPD) has become more widely available for the analysis of biomolecules including lipids and sugars to proteins and non-covalent complexes [24–28]. While different wavelengths enable different experimental

* Corresponding author.
E-mail address: perdita.barran@manchester.ac.uk (P.E. Barran).

https://doi.org/10.1016/j.ijms.2021.116588
1387-3806/

strategies, most 'top-down' experiments are carried out at 213 or 193 nm. For proteins sprayed from denaturing solutions this can provide good sequence coverage throughout the entire protein [29,30]. UVPD is sensitive to the rigidity of protein secondary structural elements and to proline isomerization, inviting exploitation of UV photons for structural characterization [31–33]. 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 occurs [34].

Fragments from peptides and proteins *via* CID or UVPD often arise due to backbone cleavages and therefore conventionally assigned using the Roepstorff and Fohlman nomenclature [35]. 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, 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 adduction [36]. When applied to top-down sequencing and especially that for proteins in a native MS workflow, such exceptions become more prevalent, often necessitating manual assignment [37]. 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 intensity of the fragment ions can be diagnostic of secondary and tertiary structure [10,38–40].

It is well known that the charge state distribution of a protein can vary according to the solvated conformation(s) [41] and IM-MS measurements have shown that a single precursor ion can adopt numerous conformations even within a single charge state [3,42]. Fragmentation approaches focus on determining primary sequence and can relate this to a 3D structure determined from X-ray crystallography [43], however, there have been fewer attempts to determine how the fragment yield and pattern alters with conformation [44,45]. As previously reported, 213 nm UVPD can probe protein conformation, yielding different fragment spectra depending on the initial conformations according to IM-MS measurements [46,47].

Non-covalently bound fragments in ECD present as reduced species and can, therefore, be readily distinguished from the precursor by *m/z* alone [12]; in UVPD such species are less 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, followed by and coincident with IR irradiation and for the 3+ precursor ion, they report an improvement in sequence coverage, although the yield of a-type fragments did not increase, indicating that UVPD did not here cause 'masked' cleavage where fragments remain bound by non-covalent interactions [34,48]. 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 so perhaps these fragments should not be expected [49,50].

Principal component analysis (PCA) is commonly used as a data mining approach 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 [51]. 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 even though a significant portion of the mass spectra from top-down and bottom-up proteomics remains 'dark'. PatternLab 4.0 has started to solve this problem; the software clusters similar proteomic profiles together and with a scoring system [52]. 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 [53].

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 [54]. DiagnoProt can provide a quick view as to how similar two biological conditions are to one another via clustering of the fragmentation yield [55,56]. 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 [51].

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.

2. Experimental section

2.1. 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 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 CID experiments. Micro Bio-Spin 6 chromatography columns (Bio-Rad Laboratories, Hercules, CA, US) were used to desalt the Haemoglobin. All samples were ionized using nanoESI in positive ion mode. Proteins were sprayed from borosilicate glass or quartz capillaries (World Precision Instruments, Stevenage, UK) pulled in-house on a Flaming/Brown P-1000 or a P-2000 micropipette puller (Sutter Instrument Company, Novato, CA, US) respectively. Platinum wire (Goodfellow) was inserted to the capillary to allow a voltage to be applied to the solution.

The mass spectrometers are calibrated using 2 μg/μL NaI and CsI made up in 1:1 Water:Isopropanol, (SI Fig. 1).

2.2. UVPD-IM-CID experiments

The experimental setup for UVPD-IM-MS has been described in detail previously [46], a schematic of the instrument can be found in SI Fig. 2. Briefly, a laser beam is introduced into the ion mobility mass spectrometer, through a CaF_2 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 [57]; 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. Large multimeric proteins were not trapped however it has been shown that the trapping protocol had no effect on the ATDs for Ubiquitin (SI Fig. 3).

2.2.1. UVPD workflow

All source voltages were tuned for gentle ionization with a typical capillary voltage of 1.1–1.3 kV. The source temperature was 40 °C and the trap bias setting was reduced to 30 V–35 V to minimise protein activation. For multimeric proteins, the trap gas was increased to improve the extent of multimeric species observed. After ionization, the charge state of interest was selected in the quadrupole and accumulated in the trap cell for two seconds to achieve an ion count of ~ 2e3. Protein ions were then photo-activated with a 213 nm laser (CryLas, UK). 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. Fragmentation yields are calculated according to the relationship:

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

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

2.2.2. CID workflow

All samples were infused directly into a Waters Synapt G2-S mass spectrometer *via* nanoESI, with similar non-activating conditions as for the UVPD experiments: the capillary voltage was maintained at 1.1–1.3 kV; source temperature 60 °C; sampling cone at 40 V; trap gas was maintained at 2 mL/min for both protein samples. For ubiquitin, 5+ and 6+ ions, 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.

2.2.3. Multivariate analysis workflow

Raw data were processed using R [58] and figures were produced using the package ggplot2 [59]. 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* x *n* structure where each row was a binned *m/z* and each column was corresponding ion count for each replicate. Data were analysed in R using MetaboAnalystR package. Features that had a constant or single value across samples were removed and remaining missing values were imputed in the remainder of data. If there were more than 50% missing values these features were removed and for the remaining data, missing values were replaced by one fifth of the minimum positive value for each feature. The features were normalised by total ion count for each sample and data was then mean centred prior to MVA.

3. Results and discussion

3.1. Collisional activation post UVPD to probe protein restructuring

3.1.1. Cytochrome c

When sprayed from 200 mM ammonium acetate, Cytochrome c presents predominantly as ions $[M+6H]^{6+}$ and $[M+7H]^{7+}$ 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 (Fig. 1a). Under soft source conditions, $[M+7H]^{7+}$ is found as a single conformer, previously found to be compact by many studies [60]. Increasing the cone voltage to 85 V ('intermediate') results in a shift away from this compact conformer to more extended forms. Further increase of the cone voltage to 120 V ('harsh') moves the intensity to the most extended form.

At each discrete cone voltage, $[M+7H]^{7+}$ was subjected to UVPD producing a, b, c and y-type fragments, a full list of all fragments observed can be found in SI Table 1. The main observation from these differing conditions is the changed intensities of certain fragments as a function of activation (Fig. 1b). When collisional activation was performed post UVPD and mobility separation on the compact conformer, an increase in the intensity of all fragments is observed. This is concomitant with the appearance of additional a- and b-type fragments from cleavage within the primary sequence from Phe36 to Met65. This region in the crystal structure comprises a loop and two short alpha-helices [61], which are located deep in the protein core. It follows that collisional activation causes a partial unravelling of these secondary structural elements, breaking stabilising non-covalent interactions and allowing the release of cleaved fragments.

When the intermediate conformer was subjected to UVPD (prior to any activation), we observe an increase in the summed fragment ion intensity as well as in the number of fragments (Fig. 1biii). UVPD-IM-CID further increases the fragment yield, and releases ions containing residues Lys22 to Lys55. These results indicate that the helix between Asp50 and Gly56 is disrupted before the helix containing Met65. Increasing the cone voltage to harsh conditions decreases the overall fragment yield (Fig. 1bv). When collisional activation was applied post-IM, a b29 ion is observed, suggesting again that non-covalent interactions remain even under harsh source conditions and extra collisional energy is required to release this UVPD fragment. Interestingly, under soft, intermediate and harsh conditions, fragments that contain the C- and N-terminal stretches are produced (Asp2 to Cys14 and Lys87 to Thr103 respectively). No shorter fragments that contain these regions are ever seen even after additional post-IM activation, supporting prior observations that these are the most stable structural elements [62].

The overall fragmentation yield increases with increasing transfer CE (SI Fig. 4) for all source conditions, as does the yield of only a-type fragments (Fig. 1c). Since the ATD of the intermediate and the harsh source conditions both contain the same conformer, we used ion mobility to determine whether the increase in fragmentation yield (harsh) is a result of the remaining intermediate family or originates from the most extended one. As collisional activation occurs post ion mobility, we can use the ion mobility of fragments to determine where they dissociated and whether this was due to the initial UVPD activation or due to the collisional activation allowing the release of the fragment. The a36 ion has a distinct ATD with little overlap with the ATD of the intact protein (Fig. 1d lower and middle panel). Upon additional collisional activation post mobility, there is an increase in the fragmentation yield of this ion and its ATD now contains a distribution that resembles that of the intact protein (Fig. 1d lower and top panel). This supports the assertion that increasing the transfer CE allows the a36 ion, caused by UVPD, to depart the precursor, but of course it retains the same ATD, since it traversed the drift cell prior to this departure, the difference in ion intensity suggests that most is retained until the CE activation. This is observed for other a-ions (SI Fig. 5).

Overall, collisional activation post IM increases the fragmentation yield presumably by disrupting non-covalent interactions, that keep cleaved regions together. Ubiquitin shows similar behaviour (SI Fig. 6).

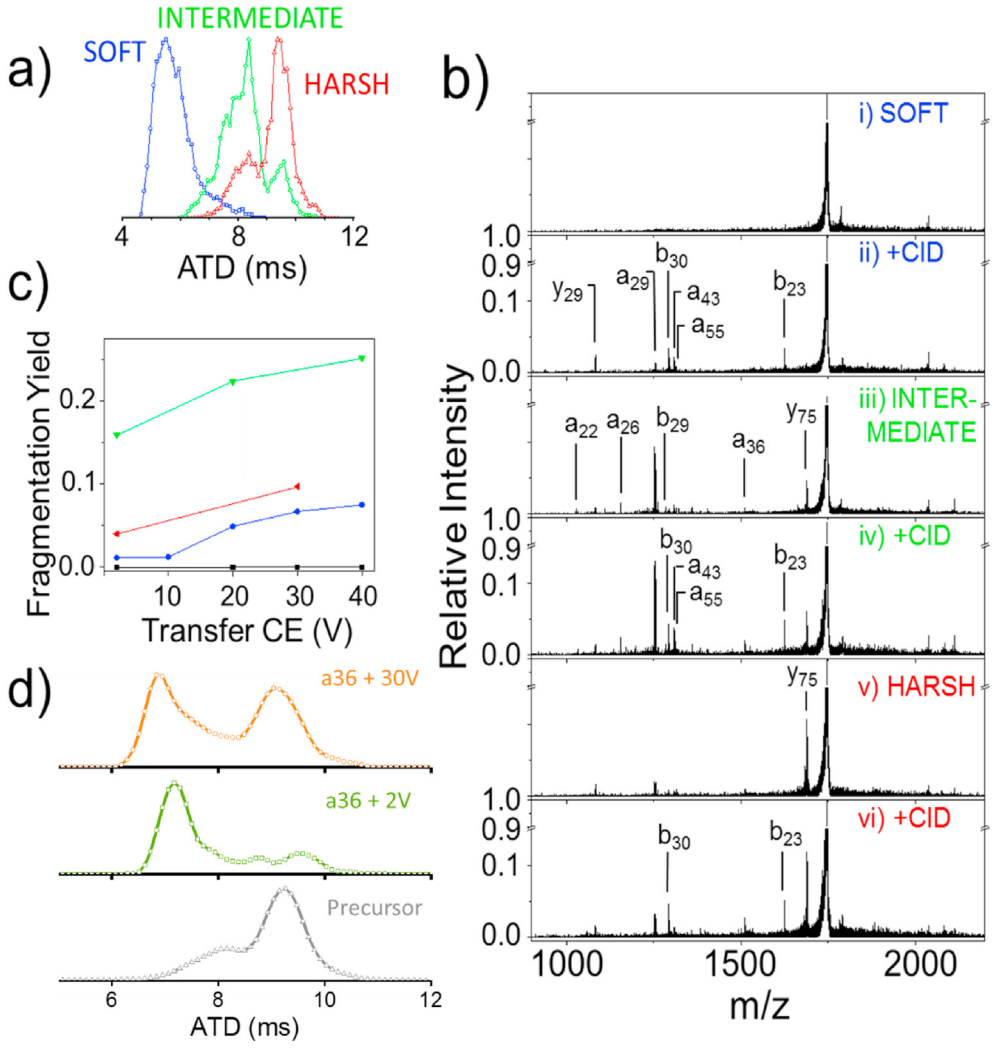

Fig. 1. UVPD-IM-CID experiments performed on the $[M+7H]^{7+}$ ion of Cytochrome C. a) Arrival time distribution at soft (cone 10 V), intermediate (cone 85 V) and harsh (cone 120 V) source conditions. b) i) UVPD spectrum at soft conditions without and ii) with additional transfer CE of 30 V. iii) UVPD 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.

3.2. UVPD analysis of large multimeric proteins

3.2.1. Concanavalin A

Concanavalin A was subjected to similar experiments as described above for cytochrome c (Fig. 2, and SI Figs. 7 and 8). Collisional activation of the tetramer (Fig. 2b) produced significant monomer, as previously described. Alongside these monomeric species, CID of $[4M + 20H]^{20+}$ yields low mass fragments (1000–1800 m/z) potentially from ejected monomer units (blue, Fig. 2b). Product ions with higher m/z than the $[4M + 20H]^{20+}$ precursor are also observed, and are assigned as charge stripped species as previously described [63]. Collisional activation was also performed on the dimeric form yielding mainly monomer species over a wide charge state distribution (CSD) and some charge stripping and low intensity backbone cleaved fragments (SI Fig. 7 and SI Table 2).

The $[4M + 20H]^{20+}$ ion was isolated and fragmented *via* UVPD at 250 Hz (Figure 2c) and 1 kHz (Fig. 2d) to monitor the effect of the repetition rate on the fragmentation spectrum. Similar to CID, the UVPD fragmentation spectra shows monomer, trimer, charge stripped tetramer and a range of backbone cleavage ions. There are some evident differences between CID and UVPD spectra, the distributions of ejected monomers in UVPD whilst centred on $[M+10H]^{10+}$ contain lower charge states than those generated by CID (Fig. 2c and SI Table 2). The abundance of these low charge species is greater when UVPD is carried out at 1 kHz (Fig. 2d) and is also accompanied by more backbone cleavage ions. The higher photon flux is acting to promptly dissociate the monomer, akin to SID, and also to generate more fragments than reported with CID [24,63,64]. Whilst the low signal to noise and the large number of possible fragments which may include side chain losses, internal fragments and even non-covalent interactions from two subunits, means these ions are hard to assign. Here we note dominant fragments from the 7+ → 10+ charge states of the $\{a_{120}\text{-}NH_3\}$, a_{120}, b_{120}, b_{121}, $\{a_{122}\text{-}NH_3\}$ and $\{b_{122}\text{-}NH_3\}$ ions (SI Fig. 8). UVPD also produces charged stripped tetrameric and trimeric species, and the photo-fragmentation of the $[4M + 20H]^{20+}$ concanavalin A ion yields a 5+ monomer subunit, not seen with CID. From this UVPD data, we can conclude that single monomers can be dissociated from the tetramer even with a lower photon flux, and the CSD infers that they are less unfolded than with CID, and may retain some of the native structure, although we do not see dimer.

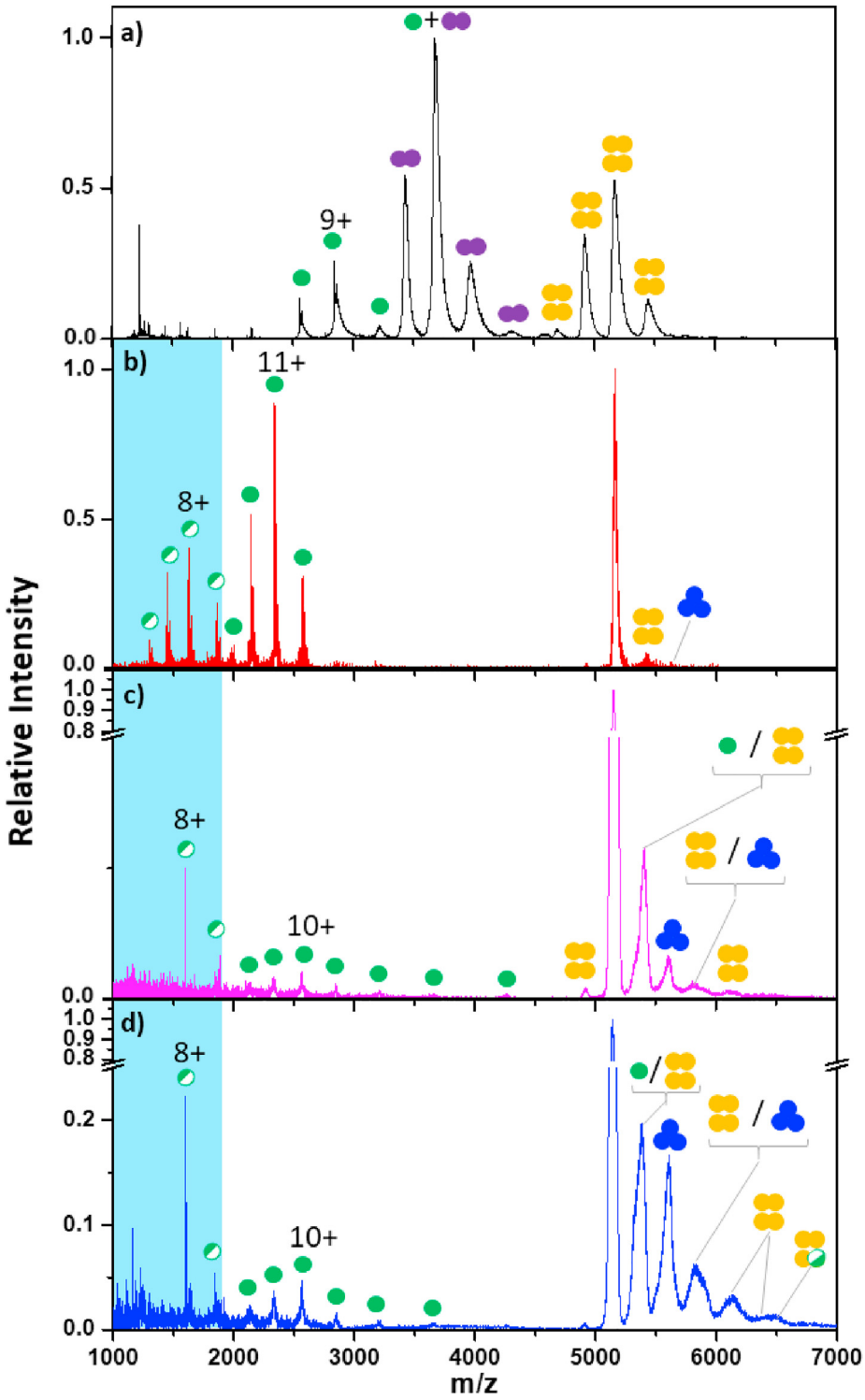

Fig. 2. a) The full MS spectrum of the Concanavalin A (black), labelled to show the subunits along with the final tetrameric structure. b) CID spectra (red), at a collisional energy of 1.6 keV, of the $[4M + 20H]^{20+}$tetramer in the trap region of the mass spectrometer. UVPD spectra at a repetition 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 sequence ions as well as monomer, dimer, trimer and tetramer species indicated by half-filled green, green, purple, blue and yellow circles, respectively.

3.2.2. *Haemoglobin*

Similar findings are obtained with haemoglobin (SI Fig. 9). 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 Fig. 9b). UVPD also leads to formation of monomer and trimer subunits along with a greater amount of primary sequence ions (SI Fig. 9c and d) confirming that large multicomponent protein complexes are amenable to UVPD within this instrument.

As with many multimeric protein complexes, the dominant products of collisional activation of tetrameric concanavalin and haemoglobin are highly charged monomeric species, and low charged trimer, where both undergo substantial structural rearrangement, as shown by IM data and the charge of the fragments prior to dissociation [24,65,66]. By contrast, UVPD facilitates the observation of fragments from the intact tetramer (Fig. 2) which can provide more information about the native fold and how it is disrupted. This approach is perhaps most informative on combining UVPD with collisional activation as shown here for cytochrome c (Fig. 1), but here even for a simple monomeric protein, the wealth of multidimensional data produced, prevents a full assessment of this related to structure. In order to guide interpretation of these multidimensional datasets, we decided to explore the use of MVA, with an initial focus on small monomeric proteins.

3.3. MVA strategies applied to fragmentation experiments in native mass spectrometry

3.3.1. Ubiquitin following CID comparing native charge states

To develop a method to mine the complex data from native top-down experiments, an MVA workflow was trained on Ubiquitin. Experiments were performed *with seven* replicates. MVA analysis was performed on these precursor ions to monitor for significant changes in the fragmentation pattern between the two charge states. CID was also performed at concentration 5 μM and 20 μM (SI Fig. 10) and showed no significant difference in the fragmentation pattern that was produced.

The activated charge states yield different fragmentation patterns when exposed to collisional activation, this is evident on manual comparison of the MS (SI Fig. 11 and SI Table 3) and *via* MVA when a sPLS-DA plot is produced (Fig. 3a), yielding distinct separation between the two charge states. The sPLS-DA plot shows good separation and clustering for each charge state, with a significant variance within each charge state, attributable to the slight differences between the repeats perhaps due to conformational differences. To account for any spray changes and fluctuations within the total ion count (TIC) the MVA data was also normalised to the sum of all the intensities (SI Fig. 12), again the separation between the two charge states remains significant. This separation implies that even for adjacent charge states of a protein ion produced under soft ionization conditions, the precursor charge state restructures differently which in turn influences the fragmentation behaviour. Principal Component Analysis (PCA) was carried out on the normalised data and scores plots were generated (SI Fig. 12).

An advantage to an MVA method is the significance of any spectral feature can be determined, upon performing a t-test the fragments with greater significance ($p < 0.05$) and a fold change greater than 1.5 or −1.5 (SI Table 4) have been chosen and identified (Fig. 3b). We find eight significant fragments that are highly dependable on the precursor protein charge, (five for $[M+6H]^{6+}$, and three for $[M+5H]^{5+}$ Fig. 3b and SI Table 4). These infer that specific fragments can be linked to precursor ion charge states *a priori*. Unlike the fully annotated MS spectrum, we highlight fragments here where there are statistically significant different intensities between the data from $[M+6H]^{6+}$ in comparison to $[M+5H]^{5+}$. Although some of these differences can be observed manually, some fragment ions hide within the "dark" regions of the MS that may not have been taken into consideration. Whilst the example of ubiquitin is trivial, this approach could find utility in the top-down measurement of more 'exotic' proteins, single point mutants, or *apo*- and *holo*-protein complexes.

3.3.2. Cytochrome c following in source activation and UVPD

To further explore the rationale for an MVA strategy we now apply this to a 213 nm UVPD fragmentation workflow for cytochrome c, with a focus on variation between conformations (SI Fig. 13) found for a single charge state.

Following this soft and harsh in-source activation the UVPD fragmentation yields were processed *via* the MVA protocol which involved monitoring the fragmentation patterns and then grouping the replicates. To overcome discrepancies caused by spray quality and nESI tips, the MVA was performed on a repeated MS to minimise these fluctuations between runs. The MS applied to the MVA was a summation of the entire 20 min' acquisition under different in-source activation, after MVA analysis was performed all identified features were confirmed to be "real" before continuing. The two in-source activation conditions remain remarkably separate from one another in the resulting score plot following this approach (Fig. 4a). Data obtained under soft conditions clusters closely between experimental replicates, indicating little fluctuation in the precursor ion structure or little fluctuation due to the background noise. Conversely, the data obtained following harsher in-source activation shows significant variance between the repeats, this may be due to the inherent structural flexibility under harsher conditions or due to the TIC variation owing to spray stability or

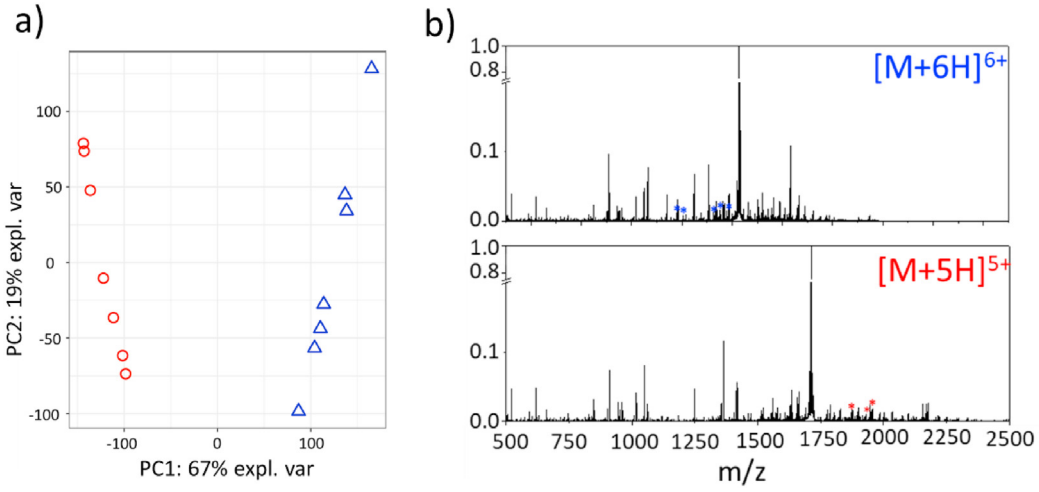

Fig. 3. nESI-MS/MS-MVA of ubiquitin ions $[M+6H]^{6+}$ and $[M+5H]^{5+}$. a) Score plot for the tandem MS of the 5+ (red) and 6+ (blue) charge states following collisional activation. b) CID spectra of $[M+6H]^{6+}$ (top) and $[M+5H]^{5+}$(bottom) with discriminant ions determined with MVA labelled on the spectra.

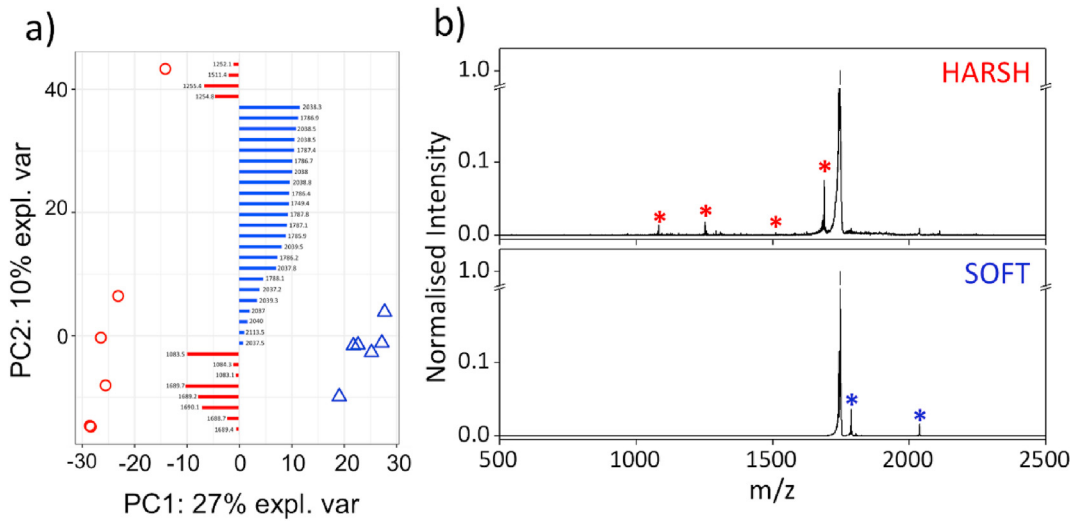

Fig. 4. UVPD-IM-MVA of two conformations of $[M+7H]^{7+}$ Cytochrome c. a) Score plot for the soft (blue) and harsh (red) conditions exposed to 213 nm UVPD with MVA determined features overlayed showing prevalence under both source conditions, b) UVPD spectrum at soft (cone 10 V) and harsh (cone 120 V) conditions without any additional transfer CE, labelled are the most prominent changes according to the MVA results.

slightly different capillary fluctuations. To account for the TIC variations, the MVA data was normalised to account for this (SI Fig. 14), whilst it appears that the harsh source conditions result in a higher population of noise, this does not affect the separation of the two conditions. Therefore, the use of MVA to identify the conformation of the protein remains meaningful however the difference between the distribution between each data point is reduced and therefore we may be limited in the assignment of features to individual conformers following activation.

This separation in the sPLS-DA plot occurs because of the differing UVPD fragments from conformers that are produced by different activation conditions. This paves the way to use MVA to highlight features of characteristic *m/z* values (Fig. 4a) that can be used to monitor the folded state of a protein. The fragments chosen have greater significance ($p < 0.05$) and a fold change greater than 1.5 or −1.5 (SI Table 5). Features identified do include isotopes and different charge states of the same ion (Fig. 4b) although this redundancy is here managed manually and in the future will be computationally. MVA applied to these top-down fragmentation data sets from native protein mass spectra finds features that are diagnostic of protein conformation. The identified features can also be compared to assignment ions (SI Table 1) to confirm their location in the protein fold.

This MVA procedure enables a data directed approach to assess the conformational landscape utilising the UVPD fragmentation information. Fragments that show the greatest change between the two conditions can act as the identifier features to indicate when and where structural perturbation of a protein has occurred. The fragments identified *via* MVA that cannot be assigned are just as important and provide new reproducible features that signify the conformational state, here it is less about the need to assign the entire sequence and more about utilising all the information available in the "dark regions" of the MS to gauge about the protein conformation from the fragmentation yield. MVA shows its clear benefits of monitoring the fragmentation yield and relating this to the conformation *via* its position on the sPLS-DA plot.

3.3.3. Combining MVA and standard assignment methods

Multivariate approaches are widely utilised in a range of analytical and physical techniques [67–70], to allow data to be extracted from congested spectra containing features which are hard to assign. Combining MVA approaches with such techniques allows tools to be developed that are capable of extracting useful information from complicated spectra, including important structural aspects of DNA binding interactions [67], development of diagnostic tests for endemic diseases [68], particle size during crystallisation [69] and allowing for rapid quality control to be performed [70]. The overall aim of the MVA applied here was to develop a novel tool to study the conformational state of proteins.

Following MVA analysis, the identified features that have the highest confidence ($p < 0.05$) for ubiquitin and cytochrome c (SI Tables 4 and 5, respectively) are then assigned *via* Roepstorff and Fohlman nomenclature and mapped onto the crystal structure of the protein (Fig. 5a). For each fragment, the corresponding part of structure is highlighted and colour coded to match the initial charge state of the precursor ion. Interestingly, fragments from the higher charge state precursor originate deeper in the tertiary fold, with a significant portion coming from the β-sheets. Whilst CID is known to be a slow heating fragmentation technique that would result in unfolding before fragmentation [72] this data indicates that the higher charge state is already more disrupted. Whereas the regions that are preferentially cleaved in $[M+5H]^{5+}$ consist of the turn (Thr9) and the end of the first β-sheet (Val6 and Lys7), and not from any part of the protein with helical or β-sheet secondary structure.

MVA determined features from UVPD, were also mapped to the structure of cytochrome c (Fig. 5b). Again, the significant features are colour coded. As expected, regions shaded in blue which are due to fragments that are found from soft conditions, are found mainly in the loop and turn regions of the protein (Pro71 and Arg91), where there are little non-covalent interactions that could prevent the dissociation. On the other hand, many of the identified features from data taken under harsh conditions arise from regions of the sequence close to, and in one case within, α-helices (Gly29, Phe36 and Ile75). As for ubiquitin, harsher activation causes unfolded conformations with fewer non-covalent interactions therefore permitting dissociation of such fragments. The identified regions are comparable to those seen previously, where we use IM-MS alone to examine the effect of conformation on the 213 nm UVPD fragmentation patterns [46] illustrating the effectiveness of the MVA approach to identify structurally significant fragments *a priori*.

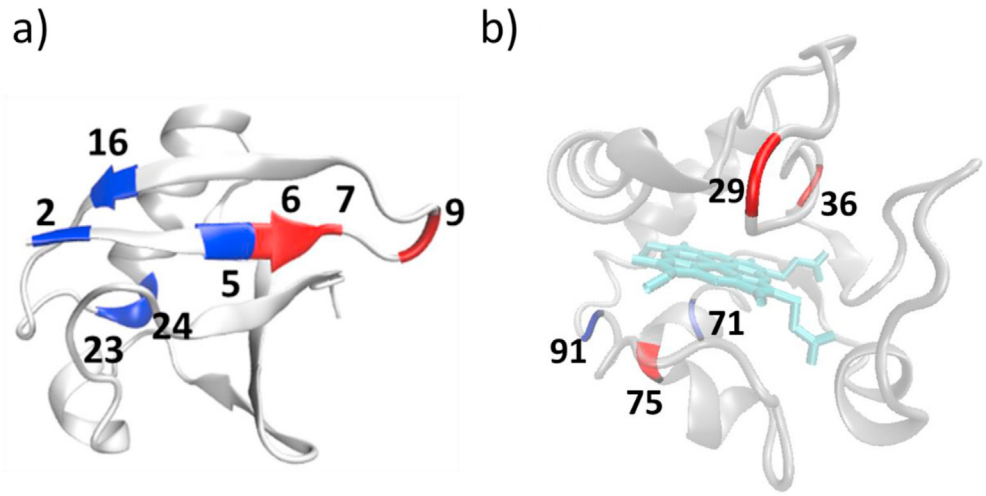

Fig. 5. The Top m/z identified through MVA are mapped onto the crystal structure of the protein, a) Ubiquitin (PDB structure: 4Z9S [71]), the regions where significant difference in fragmentation yield between the two charge states $[M+5H]^{5+}$ (red) and the $[M+6H]^{6+}$ (blue). b) Cytochrome C (PDB structure: 2B4Z [61]), the regions where there is significant difference in fragmentation yield between the two different conformations from soft (blue) and harsh (red) in-source activation.

The proteins and structural changes investigated here are well studied and understood allowing them to act as model systems for the development of MVA as an analytical tool. Our results indicate that this approach could be applied to other less common variations such as PTMs or cofactor binding, to gain a deeper understanding of their structural impact on protein systems. It is also important to note, that following standard Roepstorff and Fohlman (and without IMS knowledge), the difference between both conditions (Fig. 4) would have been overlooked. It then becomes important to consider all the data that has been obtained, even those in the "dark" unassignable region of the MS can give valuable information on the protein structure.

4. 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. 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.

Similarly, when UVPD is applied to larger multimeric proteins, we have shown here that *via* UVPD workflows we're able to get more information on the native-fold of larger multimeric proteins *via* the observed fragments from the intact tetramer.

The CID-MVA procedure has permitted the separation of the two initial precursor charge states, both of which originate from soft ionization conditions and believed to be "native-like", however, both undergo a different fragmentation pathway. We propose here that MVA has allowed more information to be taken from this data and has shown us that the precursor charge state does have an effect on the fragmentation of the ion.

MVA was also successfully applied to UVPD data where it has been previously shown that the precursor ion conformation dictates which fragments are observed. Here, we showed how to utilise MVA strategies on this data to yield sPLS-DA plots that show the direct difference between the two conformations utilising only the MS fragmentation data. It is hoped that in future work this will allow one to determine the point at which the conformation of a protein has been disrupted without the requirement of IMS.

We have shown here the benefits of utilising multiple activation methods to gain structural and conformational information on a set of model proteins. The purpose of this paper is to demonstrate that MVA can be used to determine, if a given conformation is identical or different to another, without the extra effort of fragment assignment based on limited predicted fragments. This allows for the "dark" regions of the MS to be accounted for even when assignment may not previously be possible. Given the sensitivity of modern mass spectrometers, such an approach could readily be applied as a screening tool, for example, to map the structural features that alter due to PTMs, cofactor binding, or in the directed evolution of enzymes.

Author statement

All authors agree to the submission of this document.

RB, AB, AT, LAIR and BB performed experimental work, and data analysis. DT led the MVA work. PB, RB, DT, JB and BB designed the project. All authors contributed to writing the manuscript over many long zoom sessions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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 partly 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijms.2021.116588.

References

[1] D.E. Clemmer, R.R. Hudgins, M.F. Jarrold, M.F. Jamold, J. Am. Chem. Soc. 117 (1995) 10141–10142.
[2] T. Wyttenbach, G. Von Helden, M.T. Bowers, J. Am. Chem. Soc. 118 (1996) 8355–8364.
[3] R. Beveridge, L. G. Migas, K. A. P. Payne, N. S. Scrutton, D. Leys and P. E. Barran, Nat. Commun., , DOI:10.1038/ncomms12163.
[4] R. Beveridge, S. Covill, K.J. Pacholarz, J.M.D. Kalapothakis, C.E. MacPhee, P.E. Barran, Anal. Chem. 86 (2014) 10979–10991.
[5] E. Jurneczko and P. E. Barran, 2011, 20-28.
[6] Z. Hall, A. Politis, C.V. Robinson, Structure 20 (2012) 1596–1609.
[7] J.A. Loo, J.X. He, W.L. Cody, J. Am. Chem. Soc. 120 (1998) 4542–4543.
[8] Z. Deng, N. Thontasen, N. Malinowski, G. Rinke, L. Harnau, S. Rauschenbach, K. Kern, Nano Lett. 12 (2012) 2452–2458.
[9] D. Hewitt, E. Marklund, D.J. Scott, C.V. Robinson, A.J. Borysik, J. Phys. Chem. B 118 (2014) 8489–8495.
[10] N.C. Polfer, K.F. Haselmann, P.R.R. Langridge-smithy, P.E. Barran, Mol. Phys. 103 (2005) 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, P.E. Barran, J. Phys. Chem. B 118 (2014) 12348–12359.
[12] K. Breuker, H. Oh, D.M. Horn, B.A. Cerda, F.W. McLafferty, J. Am. Chem. Soc. 124 (2002) 6407–6420.
[13] T. Wyttenbach, M.T. Bowers, Annu. Rev. Phys. Chem. 58 (2007) 511–533.
[14] J.E.P. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 9528–9533.
[15] R.A. Zubarev, D.M. Horn, E.K. Fridriksson, N.L. Kelleher, N.A. Kruger, M.A. Lewis, B.K. Carpenter, F.W. McLafferty, Anal. Chem. 72 (2000) 563–573.
[16] A.S. Galhena, S. Dagan, C.M. Jones, R.L. Beardsley, V.H. Wysocki, Anal. Chem. 80 (2008) 1425–1436.
[17] N.L. Kelleher, Anal. Chem. 76 (2004) 196 A–203 A.
[18] J.C. Jurchen, E.R. Williams, J. Am. Chem. Soc. 125 (2003) 2817–2826.
[19] J.C. Jurchen, D.E. Garcia, E.R. Williams, J. Am. Soc. Mass Spectrom. 15 (2004) 1408–1415.
[20] S. Tamara, A. Dyachenko, K.L. Fort, A.A. Makarov, R.A. Scheltema, A.J.R. Heck, J. Am. Chem. Soc. 138 (2016) 10860–10868.
[21] S.N. Sipe, J.S. Brodbelt, Phys. Chem. Chem. Phys. 21 (2019) 9265–9276.
[22] M. Sharon, JAM 21 (2010) 487–500.
[23] I. Sinelnikov, E.N. Kitova, J.S. Klassen, J. Am. Soc. Mass Spectrom. 18 (2007) 617–631.
[24] L.J. Morrison, J.S. Brodbelt, J. Am. Chem. Soc. 138 (2016) 10849–10859.
[25] M.R. Robinson, J.M. Taliaferro, K.N. Dalby, J.S. Brodbelt, J. Proteome Res. 15 (2016) 2739–2748.
[26] A. Racaud, R. Antoine, L. Joly, N. Mesplet, P. Dugourd, J. Lemoine, J. Am. Soc. Mass Spectrom. 20 (2009) 1645–1651.
[27] T. Ly, R.R. Julian, J. Am. Chem. Soc. 132 (2010) 8602–8609.
[28] M.A. Halim, L. MacAleese, J. Lemoine, R. Antoine, P. Dugourd, M. Girod, J. Am. Soc. Mass Spectrom. (2017) 1–14.
[29] 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, J.S. Brodbelt, J. Am. Chem. Soc. 135 (2013) 12646–12651.
[30] 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, J.S. Brodbelt, Anal. Chem. 86 (2014) 2185–2192.
[31] T.-Y. Kim, S.J. Valentine, D.E. Clemmer, J.P. Reilly, J. Am. Soc. Mass Spectrom. 21 (2010) 1455–1465.
[32] S. Warnke, C. Baldauf, M.T. Bowers, K. Pagel, G. von Helden, J. Am. Chem. Soc. 136 (2014) 10308–10314.
[33] M.B. Cammarata, J.S. Brodbelt, Chem. Sci. 6 (2015) 1324–1333.
[34] R.R. Julian, J. Am. Soc. Mass Spectrom. 28 (2017) 1823–1826.
[35] P. Roepstorff, J. Fohlman, Biomed. Mass Spectrom. 11 (1984) 601.
[36] J.A. Taylor, R.S. Johnson, Anal. Chem. 73 (2001) 2594–2604.
[37] A.S. Phillips, A.F. Gomes, J.M.D.D. Kalapothakis, J.E. Gillam, J. Gasparavicius, F.C. Gozzo, T. Kunath, C. MacPhee, P.E. Barran, Analyst 140 (2015) 3070–3081.
[38] J.M.D. Kalapothakis, Y.V. Berezovskaya, C.G. Zampronio, P.A. Faull, P.E. Barran, H.J. Cooper, Chem. Commun. (Camb). 50 (2014) 198–200.
[39] B.T. Ruotolo, S.J. Hyung, P.M. Robinson, K. Giles, R.H. Bateman, C.V. Robinson, Angew. Chem. Int. Ed. 46 (2007) 8001–8004.
[40] J.T. Seffernick, S.R. Harvey, V.H. Wysocki, S. Lindert, ACS Cent. Sci. 5 (2019) 1330–1341.
[41] L. Konermann, E. Ahadi, A.D. Rodriguez, S. Vahidi, Anal. Chem. 85 (2013) 2–9.
[42] D. Stuchfield, P. Barran, Curr. Opin. Chem. Biol. 42 (2018) 177–185.
[43] M. Schennach, K. Breuker, J. Am. Soc. Mass Spectrom. 26 (2015) 1059–1067.
[44] N. C. Polfer, K. F. Haselmann, P. R. R. Langridge-Smith and P. E. Barran, Mol. Phys., DOI:10.1080/00268970500095998.
[45] G.E. Reid, J. Wu, P.A. Chrisman, J.M. Wells, S.A. McLuckey, Anal. Chem. 73 (2001) 3274–3281.
[46] A. Theisen, R. Black, D. Corinti, J.M. Brown, B. Bellina, P.E. Barran, J. Am. Soc. Mass Spectrom. (2018) 1–10.
[47] M.F. Bush, Z. Hall, K. Giles, J. Hoyes, C.V. Robinson, B.T. Ruotolo, Anal. Chem. 82 (2010) 9557–9565.
[48] M.A. Halim, M. Girod, L. MacAleese, J. Lemoine, R. Antoine, P. Dugourd, J. Am. Soc. Mass Spectrom. 27 (2016) 1435–1442.
[49] J. Li, J.A. Taraszka, A.E. Counterman, D.E. Clemmer, Int. J. Mass Spectrom. 185 (1999) 37–47.
[50] H. Oh, K. Breuker, S.K. Sze, Y. Ge, B.K. Carpenter, F.W. McLafferty, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 15863–15868.
[51] G.G. Maisuradze, A. Liwo, H.A. Scheraga, J. Mol. Biol. 385 (2009) 312–329.
[52] D. B. Lima, F.V. Leprevost, M. D. M Santos, J. S. G Fischer, P. F. Aquino, J. J. Moresco, J. R. Yates, V. C. Barbosa and P. C. Carvalho, Nat. Protoc., , DOI: 10.1038/nprot.2015.133.
[53] I. Piazza, N. Beaton, R. Bruderer, T. Knobloch, C. Barbisan, L. Chandat, A. Sudau, I. Siepe, O. Rinner, N. de Souza, P. Picotti, L. Reiter, Nat. Commun. 11 (2020) 1–13.
[54] 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, J. Hancock, Bioinformatics 35 (2019) 3489–3490.
[55] 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, P.C. Carvalho, DiagnoProt: a tool for discovery of new molecules by mass spectrometry, Bioinformatics 33 (12) (15 June 2017) 1883–1885.
[56] 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, J. de S.d.G. Fischer, Sci. Rep. 10 (2020) 1–11.
[57] B. Bellina, J.M. Brown, J. Ujma, P. Murray, K. Giles, M. Morris, I. Compagnon, P.E. Barran, Analyst 139 (2014) 6348–6351.
[58] R. C. Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
[59] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer US, New York, 2009.
[60] A. P. France, L. G. Migas, E. Sinclair, B. Bellina and P. E. Barran, 1–23.
[61] N. Mirkin, J. Jaconcic, V. Stojanoff, A. Moreno, Proteins 70 (2008) 83–92.
[62] M.M.G. Krishna, H. Maity, J.N. Rumbley, S.W. Englander, Protein Sci. 16 (2007) 1946–1956.
[63] M. Zhou, S. Dagan, V.H. Wysocki, Analyst 138 (2013) 1353–1362.
[64] K. Pagel, S.J. Hyung, B.T. Ruotolo, C.V. Robinson, Anal. Chem. 82 (2010) 5363–5372.
[65] R.S. Quintyn, M. Zhou, S. Dagan, J. Finke, V.H. Wysocki, Int. J. Ion Mobil. Spectrom. 16 (2013) 133–143.
[66] Z. Hall, H. Hernández, J.A. Marsh, S.A. Teichmann, C.V. Robinson, Structure 21 (2013) 1325–1337.
[67] R. Fritzsch, P.M. Donaldson, G.M. Greetham, M. Towrie, A.W. Parker, M.J. Baker, N.T. Hunt, Anal. Chem. 90 (2018) 2732–2740.
[68] P. Heraud, P. Chatchawal, M. Wongwattanakul, P. Tippayawat, C. Doerig, P. Jearanaikoon, D. Perez-Guaita, B.R. Wood, Malar. J. 18 (2019) 1–11.
[69] C. Ferreira, J. Cardona, O. Agimelen, C. Tachtatzis, I. Andonovic, J. Sefcik, Y.C. Chen, Powder Technol. 376 (2020) 1–11.
[70] A. Alberti, T.P. Machado dos Santos, A.A. Ferreira Zielinski, C.M. Eleutério dos Santos, C.M. Braga, I.M. Demiate, A. Nogueira, LWT - Food Sci. Technol. 65 (2016) 436–443.
[71] O. Levin-Kravets, N. Shohat, G. Prag, Biochemistry 54 (2015) 4704–4710.
[72] L.G. Migas, A.P. France, B. Bellina, P.E. Barran, Int. J. Mass Spectrom. 427 (2017) 20–28.