Published in final edited form as:

Proteomics. 2014 May ; 14(10): 1165–1173. doi:10.1002/pmic.201300364.

Characterization of Green Fluorescent Proteins by 193 nm Ultraviolet Photodissociation Mass Spectrometry

Joe R. Cannon¹, Christien Kluwe², Andrew Ellington², and Jennifer S. Brodbelt¹

¹Department of Chemistry, University of Texas at Austin, Austin, Texas

²Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas

Abstract

We investigate the utility of 193 nm ultraviolet photodissociation (UVPD) in comparison to collision induced dissociation (CID), higher energy CID (HCD), and electron transfer dissociation (ETD) for top down fragmentation of highly homologous green fluorescent proteins (GFP) in the gas phase. Several GFP variants were constructed via mutation of surface residues to charged moieties, demonstrating different isoelectric points and presenting a challenge for identification by mass spectroscopy. Presented is a comparison of fragmentation techniques utilized for top down characterization of four variants with varying levels of surface charge. UVPD consistently resulted in identification of more fragment ions relative to other tandem mass spectrometry (MS/MS) methods, allowing higher confidence identification. In addition to the high number of fragment ions, the sites of fragmentation were more evenly spread throughout the protein backbone, which proved key for localizing the point mutations.

Keywords

Green Fluorescent Protein; Orbitrap; Top-down; Ultraviolet Photodissociation

1 Introduction

Mass spectrometric-based proteomic methodologies have diverged into three orthogonal strategies, termed, (in order of increasing polypeptide mass) bottom up, middle down, and top down. Bottom up, which is by far the most popular method, exploits the analysis of peptides after proteolysis of the proteins of interest and can routinely identify hundreds to thousands of peptides (and consequently hundreds or thousands of proteins) in a single experiment, albeit with low sequence coverage [1,2]. The versatility of implementation and robustness of enzymatic digestion using trypsin make this an ideal method for large-scale discovery type experiments. Additionally, bioinformatic search engines optimized for tryptic digestion-produced peptides and CID have simplified data processing [3]. This experimental simplicity is not without cost, however. Bottom up experiments often suffer from identifications that display poor sequence coverage, and it is commonplace to identify

Corresponding author: Jennifer S. Brodbelt 1 University Station A5300 University of Texas at Austin Department of Chemistry and Biochemistry Austin, TX 78712-0165 jbrodbelt@cm.utexas.edu Phone: (512)-471-0028 FAX: (512)-471-8696.

The authors have declared no conflict of interest.

proteins based on the observation of only two peptides or more recently a single peptide based on high mass accuracy measurement of the peptide and product ion m/z values that allow a reduction in candidate peptide search space [4]. This experimental avenue particularly falls short in identification of post translational modifications (PTMs), especially those that are labile during fragmentation. Moreover, PTMs cannot be mapped in regions of the proteins that are not sequenced in the bottom up approach. Middle down methods alleviate in part the aforementioned problems but at the expense of throughput and sensitivity. Peptides produced by chemical or enzymatic methods that are selective for a single amino acid or a small, but less frequently observed, primary sequence motif are longer on average than their bottom up counterparts [5–7]. The increase in peptide size, which is the basis of the middle down approach [8], inherently increases sequence coverage and the likelihood of observing single or cohabitating PTMs [9], but requires higher resolution detection for accurate charge state determination of both precursor and product ions [6,10]. In addition to the increased time required for higher resolution detection on Fourier transform (FT) based instruments, the larger peptides often occupy more than one charge state which dilutes the total ion abundance while simultaneously crowding the m/z landscape [11].

Top down mass spectrometry employs no proteolysis and thus entails the analysis of intact proteins and as such provides 100% sequence coverage [12] but at a still greater cost of throughput and sensitivity [11]. As polypeptide length increases, the aforementioned issues with charge state based signal dilution and sensitivity are compounded, even before factoring in the greater technical complexities associated with intact protein separations. Aside from potential sensitivity and separation issues, obtaining nearly complete fragmentation remains a formidable challenge for intact proteins. Methods like CID, which are well adapted for small tryptic peptides, fall short due to strong preferences for cleavage at particularly labile Xxx-Pro, Asp-Xxx, and Glu-Xxx bonds [13,14]. Higher energy CID (HCD) often produces a greater number of diagnostic ions [15], but suffers from the same cleavage preferences, and both methods result in the loss of chemically labile PTMs. Electron-based methods, such as electron transfer and electron capture dissociation (ETD and ECD) [16,17], typically maintain chemically labile PTMs and provide broad sequence coverage, but these methods demonstrate a strong charge state dependency [18]. The use of complementary MS/MS methods have improved the sequence coverage obtained for many proteins, but there are often still large patches of sequences devoid of fragment ions, a particular shortcoming for analysis of proteins with sequence mutations or PTMs that fall in those patchy regions. Observation of product ions arising from cleavage between every amino acid at sufficiently high resolution and mass accuracy would provide unambiguous sequence confirmation (aside from Leu and Ile), including assignment of potential single amino acid substitutions) [19], and in the case of proteins that are post translationally modified, confident assignment of modified sites at single amino acid resolution. We recently reported the implementation of ultraviolet photodissociation (UVPD) in the HCD cell of an Orbitrap mass spectrometer for characterization of intact proteins [20]. Our interest in UV photoactivation for top down protein characterization was motivated by our past work and that of others which demonstrated the fast high energy deposition of UVPD for analysis of peptides [21–25], as well as a few initial reports suggesting the potential

promise of UVPD for intact proteins [20,23,26,27]. As we have recently reported, 193 nm UVPD using a single 5 ns pulse creates an array of diagnostic *a*, *b*, *c*, *x*, *y*, and *z*-type fragment ions that provide unprecedented intact protein characterization, seemingly limited only by the downstream detector. Here we examine the capabilities of UVPD for characterization of a series of green fluorescent protein (GFP) variants [28].

To further explore the efficacy of UVPD, several GFP variants possessing varying isoelectric points were analyzed using CID, HCD, ETD, and UVPD. GFP has been used extensively in biochemistry immunofluorescence imaging and resonant energy transfer experiments [29]. These protein sequence variants provide an important model for top down analysis as a method for confirming the sequence of chemically engineered proteins, especially those with minor sequence variations that are easily missed by bottom-up methods. Additionally, they serve as a nearly ideal model for determining the effects of charge on gas phase fragmentation of a conserved solution structure. The fluorescent properties of GFPs are the direct result of the aromaticity of an intrinsic chromophore formed by covalent cross-linking of three buried residues when the protein is properly folded. Although these proteins were not analyzed under native conditions, an easily identifiable green color implied a folded state in solution.

2 Materials and Methods

Green Fluorescent Protein Sequence Variants

GFP variants based on the superfolder GFP backbone were designed as previously described [28]. Selected variants were chosen by calculated isoelectric point, and named according to their predicted net charge at neutral pH excluding the effect of the hexahistidine tag: RscG -11 (G3), RscG -5 (G2), RscG +7 (WT), RscG +11 (G12), and RscG +15 (G13). GFP genes were fabricated in the gene synthesis facility in house and cloned into pET21 *E. coli* expression vectors (Novagen). Plasmids encoding the GFP variants were transformed into strain BL21(DE3) cells (NEB) and grown to an OD of 0.6 to 0.8, after which cultures were chilled to 18°C, induced with 1 mM IPTG (Sigma) and further grown for 18 hours. Cells were isolated by pelleting at 3,000×g for 20 minutes at 4°C, then resuspended in 1/40th volume IMAC Buffer (20 mM MOPS, 20 mM imidazole, 500 mM NaCl, pH 7.5) supplemented with 250 U Benzonase (Novagen). Cells were lysed with a Model 500 Sonic Dismembrator (Fisher Scientific) using a 50% duty cycle at 40% amplitude for 3 minutes and centrifuged at 40,000×g for 30 minutes at 4°C to pellet insoluble proteins and debris. The soluble fraction was then applied to a chelating sepharose fast-flow column (GE Life Sciences) for immobilized metal ion affinity chromatography (IMAC) using the hexahistidine tag. Non-specific proteins were removed by washing with IMAC buffer supplemented to 60 mM imidazole. Purified GFPs were eluted by IMAC buffer supplemented to 400 mM imidazole and dialyzed twice overnight into fresh 10 mM Tris, 1 mM EDTA (pH 8.0). Protein purity was confirmed by automated electrophoresis on an Agilent 2200 tape station. Each GFP variant (~28 kDa) was buffer exchanged into LC-MS grade water three times prior to analysis using 10 kDa molecular weight cutoff filters (Millipore, Billerica, MA). The sequences of each protein are given in **Supplemental Figure 1**

Electrospray Ionization Mass Spectrometry (ESI-MS)

GFP variants were diluted to 25 μ M in 50/49/1 acetonitrile/water/formic acid. The variants were then infused at 5 microliters per minute directly into an Orbitrap Elite MS (ThermoFisher, San Jose, CA). UVPD was implemented in a manner similar to what we have previously described for infrared photodissociation in which UVPD was undertaken in the HCD cell[30]. The in source voltage was set to 35 V to remove loosely covalent adducts. For all analyses, the nominal gas pressure in the HCD cell was reduced to $\sim$ 2 mTorr relative to the normal operating pressure of $\sim$ 10 mTorr. This decrease in pressure facilitated a higher sensitivity analysis via a reduction in dephasing collisions between the intact proteins and excess bath gas[31], and was used for all CID, ETD, HCD, and UVPD data collection. For each GFP, the 25+, 29+, and 33+ charge states were isolated and fragmented using CID at normalized collision energy of 35 and using HCD at normalized collision energy of 15. Electron transfer dissociation was performed only on the 33+ charge state due to the well-known relationship between increased charge state and ETD efficiency[17,18]. For all scans, the charge states of interest were isolated using a 25 m/z isolation window. For UVPD the isolated protein was activated via a single 5 ns 1.5 mJ laser pulse at 193 nm from an ArF excimer laser (Coherent Excistar XS). For all MS/MS methods, five hundred scans (10 averaged scans were combined with 50 microscans each) were averaged prior to bioinformatic analysis of the resulting spectra.

Fragment Ion Matching

Tandem mass spectra were deconvolved using the THRASH algorithm [32] (with a signal to noise ratio of 5:1) and searched against their known sequences using a beta version of ProSightPC 3.0 that was modified on an *ad hoc* basis for analysis of the most abundant ion types produced from UVPD[20]. The algorithm used for searching UVPD data included the canonical ion types produced from both collision based and electron based methods, so it was used to search all spectra. Deconvolved neutral masses were matched to within 5 ppm of their predicted masses. The mass shift associated with the GFP chromophore molecular rearrangement to 4-(phydroxybenzylidene)-imidazolidin-5-one was applied to Tyr 66 ($m = -20.0256$ Da).

3 Results and Discussion

193 nm UVPD for Sequence Confirmation

Despite the success of bottom up methods for large-scale discovery type proteomic experiments, they are generally unsuitable for high confidence characterization of proteins that have minor sequence variations, such as those that are chemically engineered to have point mutations. Mutations that result in charge inversion at a specific site (Lys/Arg to Asp/Glu substitutions) create particular problems for bottom-up methods that rely on tryptic digestion as a result of the concomitant change in the peptide mass fingerprint that occurs due to the larger or lower number of Lys/Arg residues. For example, rather than observing a specific mass shift associated with a modification of a tryptic peptide, a single mutation resulting in charge inversion in either direction (Glu/Asp to Lys/Arg or Lys/Arg to Asp/Glu) would create a new set of peptides equivalent to a missed cleavage with an additional mass shift. This outcome would complicate or prohibit the successful analysis of digests

containing proteins with multiple or even single mutations. An identical problem would occur using less common proteases specific for acidic residues (GluC/AspN). The previously mentioned complications with bottom up methods are ancillary to the main problem – the unlikely observation of every peptide. Primary sequence confirmation is well suited to top down methods provided product ion spectra are sufficiently informative. In the present study, the CID, HCD, ETD and UVPD fragmentation patterns of three charge states (25+, 29+, 33+) of five GFPs were analyzed with respect to the number and types of fragment ions and N-terminal versus C-terminal coverage. Among the set of five GFPs, two protein variants with isoelectric points higher and two protein variants with isoelectric points lower than that of the wild type (WT) sequence were used to evaluate the impact of charge sites on gas-phase fragmentation. For each GFP variant, there were from 4 to 11 residues mutated (with the WT variant as a reference) across 16 sites along the 238 residue sequence that contained amino acids with either positively or negatively charged side chains. One series of deconvolved MS/MS spectra for WT GFP (33+) are shown in **Supplemental Figure 2**. As expected, the CID and HCD spectra display mostly *a*, *b* and *y* ions, and the ETD spectrum shows mostly *c/z* ions. The UVPD mass spectrum exhibits a far greater array of ions, including *a*, *b*, *c*, *x*, *y* and *z* ions.

To evaluate the efficacy of top down mass spectrometry using 193 nm UVPD in comparison to other fragmentation methods, the resulting ions were searched against the known sequence of all prepared GFP variants. The nine most frequently observed ion types (*a*, *a*[•], *b*, *c*, *x*, *x*[•], *y*, *y*-*I*, and *z*[•]) were used for matching based on prior algorithm optimization with commonly used model proteins. Unlike collision- and electron-based methods, UVPD spectra contain a large proportion of *a*-type ions, specifically *a* and *a*[•]. The canonical *b*- and *y*-type fragments are observed less frequently. *X*, *x*[•], *y*, and *z*[•] ions are the dominant C-terminal containing ion series. The *c*-type ions are not as frequently observed in UVPD as in ETD, and the *y*-*I*-type ions (the latter which differ from the more commonly observed *y*-type ion by a decrease in mass equivalent to single hydrogen atom (-1.0078 Da). A version of ProSightPC customized on an *ad hoc* basis to accommodate all of the ion types produced following UVPD[20] allowed all ion types to be searched and tabulated in a manner consistent for all of the MS/MS spectra acquired for each of the four activation methods, as summarized in **Figure 1**. In general, the bar graphs show that the distributions of the six ion types varied depending on the ion activation method. For the purposes of practical protein identification and characterization, even and odd electron species were combined into collective bars.

The two collision-based methods (CID and HCD) and ETD displayed fragment ions that were characteristic of a pairwise fragmentation mechanism, producing primarily *b* and *y* ions for CID[14] and *c/z* ions for ETD [16,17], as observed in **Figure 1**. UVPD on the other hand, which is proposed to be the result of both vibrational energy redistribution as well as direct dissociation from excited electronic states [14,25], produced predominantly *a*-type ions originating from the N terminus and a nearly equivalent distribution of *x*, *y*, and *z* type ions arising from the C terminus. In terms of the total number of fragment ions produced, the averages across all charge states and variants ranged from 98 ions for CID, 109 ions for

HCD, 147 ions for ETD, and 252 ions for UVPD, the latter highlighting the possibility of exceptional sequence coverage upon UVPD.

Comparison of Metrics for Ion Activation Methods

UVPD consistently outperformed all other MS/MS methods for each charge state and variant in terms of the total number of fragment ions produced. After excluding redundant peaks (different charge states of the same ion type), the total number of observed ions for each variant in each charge state using CID, HCD, and UVPD are shown in **Figure 2** and re-grouped based on charge state in **Supplemental Figure 3**. Removal of redundant peaks in this case simplified analysis. Situations in which product ion sequence coverage is less extensive or mass accuracy is lower could utilize single fragment ion charge state redundancy as an added confidence metric.

The increased number of fragment ions observed for UVPD relative to CID and HCD affords higher sequence coverage and deeper analysis of protein sequence. The collision-based methods typically provide better sequence coverage at the termini of proteins rather than at their mid-sections. In contrast, UVPD produces fragments throughout the sequence including well into the interior from either direction. To illustrate this result, the 238 residue sequence of each GFP variant was divided into thirds (where the first two thirds each contains 80 residues (1-80 and 81-160), and the final third contains 78 residues (161-238)). **Figure 3** demonstrates the trends in the depth of coverage, as defined by the total amount of N-terminally derived fragment ions that represent the first (shorter *a*, *b*, *c* fragment ions), middle (mid-sized *a*, *b*, *c* fragment ions), and last third (longer *a*, *b*, *c* fragment ions) of the protein sequence. The 25+ charge state for each variant, which was influenced the least by spectral processing limitations, was selected for comparison (with the exception of ETD, which was only acquired for the 33+ charge state). The production of a much greater array of mid-sized and longer *a*, *b*, and *c* ions upon UVPD compared to CID and HCD is readily apparent in **Figure 3**.

The fraction of fragment ions from the interior of the protein sequence (corresponding to fragment ions terminating at residues 81-160 or 161-238 from the N terminus) did not fall below 40% of the total ions observed for UVPD. For CID in comparison, only 15% of the resulting fragment ions represented either of the latter two thirds of the protein sequence. In terms of practical outcome, this means that fragmentation using UVPD results in an increased likelihood of observing ions that are diagnostic for changes in the middle of the protein. **Supplemental Figure 4** depicts a comparison of ion maps for the $z = 25$ charge state of the WT variant using CID, HCD, and UVPD. Of the 16 mutated sites in this set of GFPs, CID produced an ion map that is devoid of fragment ions originating from backbone cleavages between four consecutive modification sites and HCD lacked fragment ions covering five modification sites. In contrast, UVPD resulted in fragment ions between every site of modification, and after normalizing the number of fragments arising from backbone cleavages between sites to the total number of amino acids, UVPD exhibited 0.80 fragments/residue in the regions between consecutive sites of modification, while HCD and CID achieved 0.42 and 0.34, respectively. The same comparison using the higher $z = 33$ charge state resulted in 0.57 fragments/residue for UVPD and 0.46 for ETD. Using the resulting

fragments from the same spectra used in the comparison depicted in **Figure 3**, the achieved sequence coverage (as defined by the total number of non-redundant inter-residue cleavages divided by the total number possible) was 23.7%, 27.1%, and 38.1% for CID, HCD, and ETD, respectively, while UVPD achieved 61.4%.

To evaluate the impact of the substitution of basic residues for neutral or acidic residues on the MS/MS patterns, the UVPD mass spectra of each variant were compared. Shown in **Figure 4** are the *a* type ion series (both odd and even electron) produced for each GFP variant displayed as yellow bars across the entire sequence. The 25+ charge state precursor was selected to minimize the complications arising from deconvolution of the more isotopically crowded higher charge states. Examination of the location of identified cleavages across the primary sequence of each variant (depicted by yellow bars at each amino acid position) served to highlight the effects of mutations that either inverted the side chain charge from acidic to basic or simply added basic sites where there previously were uncharged residues. For the five GFPs, the mutations are most prevalent at the N-termini of the proteins, and there appears to be an increase in the number of *a* ions produced that cover this N-terminal region as the frequency of Lys increases. A gap in sequence coverage (as highlighted by the black rectangle) is observed at the site in the primary sequence where molecular rearrangement occurs to form 4-(phydroxybenzylidene)-imidazolidin-5-one, the GFP chromophore, shown in the inset. Specific mutation of residues #8 (Thr in the G2, G3, and WT variants), #18 (Asp in the G3, G2, and WT variants), and #27 (Ser in the G2, G3, and WT variants) to Lys served to close an eleven residue gap in sequence coverage, leading to identification of a nearly complete ladder type sequence of the first fifty residues (**Figure 4**). This outcome can be attributed to the sequestration of protons at the very basic side-chains of Lys that are not present for Thr, Asp, or Ser, thus facilitating cleavages adjacent to those sites. A general increase in the total number of ions using CID and HCD as charge state increased was observed, but the ions were still localized to the two termini.

Changes in Fragmentation as a Function of Charge State

It has been observed using collisional activation methods that the number of identified fragment ions often increases as the charge state increases (up to the point of limiting proton mobility via Coulombic repulsion)[15]. Previous work from this lab demonstrated a relatively uniform extent of dissociation regardless of precursor charge state using 193 nm UVPD for ubiquitin (8.5 kDa) and myoglobin (16.9 kDa)[20]. For the larger GFPs in the present study (28 kDa), the number of fragments identified upon CID, HCD, and UVPD of the highest charge state ($z = 33$) precursors was lower than that identified in the $z = 29$ charge state, irrespective of sequence variant (**Figure 2A**). While this trend in a reduction in the number of identified fragment ions as a function of increasing precursor charge state was observed for all MS/MS methods, it was especially pronounced for the UVPD analyses (although the total number of fragment ions produced upon UVPD still remained almost double that generated by CID and HCD for all charge states). Using the $z = 29$ charge state of the G12 variant as an example, both CID and HCD methods yielded less than 120 identified fragment ions, whereas 193 nm UVPD generated two-fold more. The sheer number of fragment ions produced upon UVPD caused greater crowding of the m/z landscape with overlapping isotope clusters. As demonstrated in **Figure 2B**, from all three

interrogated precursor charge states analyzed with UVPD the proportion of identified fragment ions in higher charge states decreases as precursor charge state increases. This is consistent with increasing difficulty in deconvolution due to closer isotopic spacing.

The more notable decrease in the number of identified fragment ions for the $z = 33$ precursor upon UVPD compared to CID or HCD is attributed to be due to a reduction in the accuracy of the deconvolution algorithm. With an increase in charge comes a proportional decrease in the isotopic spacing of each isotopic cluster. Intuitively, one can imagine that the fragment ion charge state distribution should reflect the precursor charge state to some extent. For example, if a $z = 10$ precursor ion is dissociated, there will be fragments observed with charge states spanning from one up to the precursor ion charge state ($10+$), and as the precursor charge state increase, both the range of fragment ion charge state and the portion in higher charge states should increase accordingly. The distribution of fragment ion charge states upon UVPD does not reflect this expected trend, as shown for the WT GFP in **Figure 2B**. Despite the increase in precursor charge state (going from $25+$ to $29+$ to $33+$), the more highly charged fragment ions are observed less frequently. A further analysis of the deconvolution process undertaken on the same spectra using two different algorithms (Xtract (ThermoFisher, San Jose, CA) and THRASH [32]) demonstrated a similar decrease in the number of deconvolved fragment ions as precursor charge state increased (**Figure 2C**), irrespective of the algorithm used. This result is consistent with the increase in spectral complexity as isotope clusters become closer together and exhibit greater overlap. Visual inspection of identical segments of the m/z landscape across all three charge states of the same variant clearly showed a large increase in spectral complexity for the $z = 33$ charge state that was not reflected in the total number of deconvolved fragment ions (**Figure 5**). In short, the apparent decrease in the number of identified fragment ions is explained by the complexity of the UVPD spectra and the limitations of currently available deconvolution algorithms. The main implication of this shortcoming with current high resolution detection options is a requirement for auxiliary methods to either increase the available resolution, spread out the total ion current across the m/z landscape, or generally decrease the charge states interrogated for MS/MS analysis. Ion-ion reactions such as proton transfer reactions have been shown to effectively decrease product ion m/z for analysis of tandem mass spectra resulting from highly charged precursors[33]. For single protein infusion type experiments, an increase in spectral resolution can be achieved by increasing the total number of averaged scans. Future advancements in both hybridized fragmentation methods (ETDUVPD or UVPD-PTR) and MS detectors and analyzers will be undoubtedly be required to advance the field toward 100% top down sequence coverage.

4 Conclusions

This comparative study of fragmentation of GFP variants demonstrated that 193 nm UVPD provided a much greater number of diagnostic fragment ions that facilitated more confident identification of primary sequence for proteins as large as 28 kDa. An increase in the number of fragment ions was observed localized to sites of mutagenesis where a basic residue was introduced. The benefit of UVPD does not result solely from an increase in the *number* of identified ions or ion types, but rather the widespread distribution of diagnostic ions well into the interior of the primary sequence. For all variants analyzed, the total

number of identified fragments decreased from the 29+ to the 33+ charge state for UVPD. Based on the observed decrease in the proportion of more highly charged fragments and visual inspection of the resulting spectra, the results suggested limitations arising from isotopic overlap prior to deconvolution of the UVPD spectra. The successful analysis of sequences of the GFPs containing a few key point mutations demonstrates the higher performance of UVPD for intact protein analysis and its applicability to primary sequence confirmation. Identification of closely related variants from one another is particularly relevant in applications where precise quantitative analysis of protein states is necessary. With this UVPD technology we foresee the possibility to examine important protein modifications, which might be ‘invisible’ from other methods.

Acknowledgments

Funding from the NIH (R21GM099028 JSB) and the Welch Foundation (F1155 to JSB and F1654 to AE) is acknowledged. Christien Kluwe and the production of the GFPs was supported by the National Security Science and Engineering Faculty Fellowship (FA9550-10-1-0169), Defense Advanced Research Projects Agency (HR-0011-10-0052 and 5-55068), the Defense Threat Reduction Agency (HDTRA1-12-C-0007). We would also like to acknowledge assistance from Neil Kelleher, Paul Thomas, Ryan Fellers, and Bryan Early for providing a customized version of ProSightPC.

Abbreviations

CID	collision induced dissociation
ECD	electron capture dissociation
ETD	electron transfer dissociation
FT	Fourier transform
GFP	green fluorescent protein
HCD	higher energy collision induced dissociation
MS	mass spectrometry
MS/MS	tandem mass spectrometry
PTM	post translational modification
UVPD	ultraviolet photodissociation
WT	wild type

References

1. Geiger T, Velic A, Macek B, Lundberg E, et al. Initial Quantitative Proteomic Map of 28 Mouse Tissues Using the SILAC Mouse. *Mol. Cell. Proteomics*. 2013; 12:1709–1722. [PubMed: 23436904]
2. Kim W, Bennett EJ, Huttlin EL, Guo A, et al. Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Mol. Cell*. 2011; 44:325–340. [PubMed: 21906983]
3. Li W, Ji L, Goya J, Tan G, Wysocki VH. SQUID: An Intensity-Incorporated Protein Identification Algorithm for Tandem Mass Spectrometry. *J. Proteome Res*. 2011; 10:1593–1602. [PubMed: 21204564]

4. Hansen TA, Kryuchkov F, Kjeldsen F. Reduction in Database Search Space by Utilization of Amino Acid Composition Information from Electron Transfer Dissociation and Higher-Energy Collisional Dissociation Mass Spectra. *Anal. Chem.* 2012; 84:6638–6645. [PubMed: 22799558]
5. Wu S-L, Kim J, Hancock WS, Karger B. Extended Range Proteomic Analysis (ERPA): A New and Sensitive LC–MS Platform for High Sequence Coverage of Complex Proteins with Extensive Post-translational Modifications. *Comprehensive Analysis of Beta-Casein and Epidermal Growth Factor Receptor (EGFR)*. *J. Proteome Res.* 2005; 4:1155–1170. [PubMed: 16083266]
6. Cannon J, Lohnes K, Wynne C, Wang Y, et al. High-Throughput Middle-Down Analysis Using an Orbitrap. *J. Proteome Res.* 2010; 9:3886–3890. [PubMed: 20557100]
7. Wu C, Tran JC, Zamborg L, Durbin KR, et al. A protease for “middle-down” proteomics. *Nat Meth.* 2012; 9:822–824.
8. Yergey JA, Cotter RJ, Heller D, Fenselau C. Resolution requirements for middle molecule mass spectrometry. *Anal. Chem.* 1984; 56:2262–2263.
9. Young NL, Plazas-Mayorca MD, DiMaggio PA, Flaniken IZ, et al. Collective mass spectrometry approaches reveal broad and combinatorial modification of high mobility group protein A1a. *J. Am. Soc. Mass Spectrom.* 2010; 21:960–970. [PubMed: 20202861]
10. Boyne MT, Garcia BA, Li M, Zamborg L, et al. Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *J. Proteome Res.* 2008; 8:374–379. [PubMed: 19053528]
11. Compton PD, Zamborg L, Thomas PM, Kelleher NL. On the Scalability and Requirements of Whole Protein Mass Spectrometry. *Anal. Chem.* 2011; 83:6868–6874. [PubMed: 21744800]
12. Mørtz E, O'Connor PB, Roepstorff P, Kelleher NL, et al. Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence data bases. *Proc. Natl. Acad. Sci.* 1996; 93:8264–8267. [PubMed: 8710858]
13. Liu Z, Schey KL. Fragmentation of multiply-charged intact protein ions using MALDI TOF-TOF mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2008; 19:231–238. [PubMed: 17693096]
14. Wysocki VH, Tsaprailis G, Smith LL, Brei LA. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* 2000; 35:1399–1406. [PubMed: 11180630]
15. Chanthamontri C, Liu J, McLuckey SA. Charge state dependent fragmentation of gaseous α -synuclein cations via ion trap and beam-type collisional activation. *Int. J. Mass Spectrom.* 2009; 283:9–16. [PubMed: 20160958]
16. Zubarev RA, Kelleher NL, McLafferty FW. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* 1998; 120:3265–3266.
17. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:9528–9533. [PubMed: 15210983]
18. Good DM, Wirtala M, McAlister GC, Coon JJ. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics.* 2007; 6:1942–1951. [PubMed: 17673454]
19. Wynne C, Edwards NJ, Fenselau C. Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC-MS/MS on an Orbitrap. *PROTEOMICS.* 2010; 10:3631–3643. [PubMed: 20845332]
20. Shaw JB, Li W, Holden DD, Zhang Y, et al. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* 2013; 135:12646–12651. [PubMed: 23697802]
21. Thompson MS, Cui W, Reilly JP. Factors that impact the vacuum ultraviolet photofragmentation of peptide ions. *J. Am. Soc. Mass Spectrom.* 2007; 18:1439–1452. [PubMed: 17543535]
22. Madsen JA, Boutz DR, Brodbelt JS. Ultrafast Ultraviolet Photodissociation at 193 nm and its Applicability to Proteomic Workflows. *J. Proteome Res.* 2010; 9:4205–4214. [PubMed: 20578723]
23. Moon JH, Shin YS, Cha HJ, Kim MS. Photodissociation at 193 nm of some singly protonated peptides and proteins with m/z 2000–9000 using a tandem time-of-flight mass spectrometer equipped with a second source for delayed extraction/post-acceleration of product ions. *Rapid Commun. Mass Spectrom.* 2007; 21:359–368. [PubMed: 17206742]

24. Reilly JP. Ultraviolet photofragmentation of biomolecular ions. *Mass Spectrom. Rev.* 2009; 28:425–447. [PubMed: 19241462]
25. Parthasarathi R, He Y, Reilly JP, Raghavachari K. New Insights into the Vacuum UV Photodissociation of Peptides. *J. Am. Chem. Soc.* 2010; 132:1606–1610. [PubMed: 20078132]
26. Guan Z, Kelleher NL, O'Connor PB, Aaserud DJ, et al. 193 nm photodissociation of larger multiply-charged biomolecules. *Int. J. Mass Spectrom. Ion Process.* 1996; 157:357–364.
27. Kim T-Y, Thompson MS, Reilly JP. Peptide photodissociation at 157 nm in a linear ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* 2005; 19:1657–1665. [PubMed: 15915476]
28. Der BS, Kluwe C, Miklos AE, Jacak R, et al. Alternative Computational Protocols for Supercharging Protein Surfaces for Reversible Unfolding and Retention of Stability. *PLoS ONE.* 2013; 8:e64363. [PubMed: 23741319]
29. Tsien RY. The green fluorescent protein. *Annu. Rev. Biochem.* 1998; 67:509–544. [PubMed: 9759496]
30. Vasicek LA, Ledvina AR, Shaw J, Griep-Raming J, et al. Implementing Photodissociation in an Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* 2011; 22:1105–1108. [PubMed: 21953052]
31. Makarov A, Denisov E. Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* 2009; 20:1486–1495. [PubMed: 19427230]
32. Horn DM, Zubarev RA, McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* 2000; 11:320–332. [PubMed: 10757168]
33. Chi A, Bai DL, Geer LY, Shabanowitz J, Hunt DF. Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. *Donald F Hunt Honour Issue.* 2007; 259:197–203.

Figure 1.

Histogram of the percentage of ion types across all charge states for each GFP variant (25+, 29+, 33+). Electron transfer dissociation spectra were only acquired for the 33+ charge state. The large number of *y-1* ions observed in the ETD spectra can be attributed to the nearly identical mass of *z* ions after ammonia loss.

Figure 2.

(A) Histograms showing the total number of fragment ions identified as a function of charge state for each GFP variant using CID (top), HCD (middle), and UVPD (bottom). Note that the y-axis scale is different for the UVPD histogram. (B) Distribution of observed fragment ion charge states resulting from UVPD of the WT variant for three precursor charge states. (C) Bar graph displaying the total number of deconvolved fragment ions (normalized to the largest number of fragments).

Figure 3.

(A) Histograms illustrating the proportion of N terminal ions (*a*, *b*, and *c*) terminating at the first eighty (1-80) (left), middle eighty (81-160) (middle), and last 78 (161-238) (right) residues of the 25+ charge state of each GFP variant using CID, HCD, UVPD, and ETD (note that ETD was only performed on the 33+ charge state). All histograms are normalized to the same y-axis.

Figure 4.

(A) Each residue of the GFP variant primary sequence is represented by a bar with the N terminus on the left and the C terminus on the right. Sites of identified *a*-type ions from UVPD analysis of the $z = 25$ charge state from each sequence variant are highlighted in yellow. Shown in the inset is the small structural region devoid of fragment ions, which corresponds to the intrinsic hydroxybenzylidene imidazole chromophore that prevents confident assignment of sequence ions due to its inconsistent mass shift during fragmentation. (B) Relationship between the total number of fragment ions and the isoelectric point of the GFP protein (related to substitution of Lys/Arg sites for negative or uncharged residues).

Figure 5.

UVPD spectra from different charge states of the same WT variant. (A) precursor charge state 33+, (B) precursor charge state 29+, and (C) precursor charge state 25+.