analytical chemistry

Article

pubs.acs.org/ac

Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale

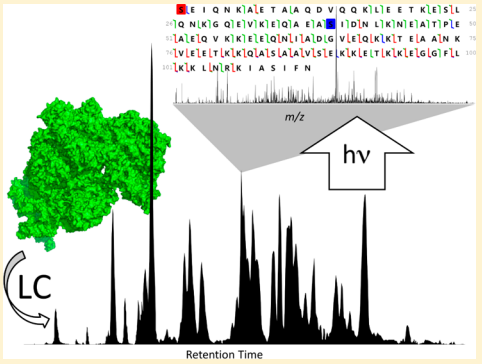
Joe R. Cannon,[†] Michael B. Cammarata,[†] Scott A. Robotham,[†] Victoria C. Cotham,[†] Jared B. Shaw,[†] Ryan T. Fellers,[‡] Bryan P. Early,[‡] Paul M. Thomas,[‡] Neil L. Kelleher,[‡] and Jennifer S. Brodbelt*[,†]

[†]Department of Chemistry, University of Texas at Austin, 1 University Station A5300, Austin, Texas 78712, United States

[‡]Departments of Chemistry and Molecular Biosciences and the Proteomics Center of Excellence, Northwestern University, Evanston, Illinois 60208, United States

Supporting Information

ABSTRACT: Intact protein characterization using mass spectrometry thus far has been achieved at the cost of throughput. Presented here is the application of 193 nm ultraviolet photodissociation (UVPD) for top down identification and characterization of proteins in complex mixtures in an online fashion. Liquid chromatographic separation at the intact protein level coupled with fast UVPD and high-resolution detection resulted in confident identification of 46 unique sequences compared to 44 using HCD from prepared *Escherichia coli* ribosomes. Importantly, nearly all proteins identified in both the UVPD and optimized HCD analyses demonstrated a substantial increase in confidence in identification (as defined by an average decrease in *E* value of ∼40 orders of magnitude) due to the higher number of matched fragment ions. Also shown is the potential for high-throughput characterization of intact proteins via liquid chromatography (LC)−UVPD-MS of molecular weight-based fractions of a *Saccharomyces cerevisiae* lysate. In total, protein products from 215 genes were identified and found in 292 distinct proteoforms, 168 of which contained some type of post-translational modification.



Proteomic analysis using mass spectrometry (MS) can be divided into three distinct approaches, termed (in order of increasing polypeptide mass) bottom up, middle down, and top down. The vast majority of biological samples are interrogated using bottom up methods, which use robust collision-based fragmentation methods to sequence the small peptides that result from tryptic digestion.[1] Middle down methods exploit more restricted proteases or chemical methods that are specific for a single amino acid or less commonly observed primary sequence motif to create peptides that are generally larger than those produced using bottom up methods.[2−5] The interest in middle down methods is motivated by the fact that as polypeptide mass increases, so does the resulting sequence coverage from each identification. Finally, top down methods lack a proteolytic step and are able to correlate observed deviations from the theoretical intact mass for a more accurate picture of the biologically relevant proteoform.[6] Each of these approaches has its own drawbacks and benefits. Because of the ease of separation, ionization, and detection of small peptides, bottom up methods provide unparalleled throughput in terms of identification, but the relative fraction of characterized protein sequence from each identification is typically far lower than the other two methods. Middle down methods result in higher sequence coverage but require higher resolution detection of both the precursor and product ions for accurate charge state deconvolution.[4]

Top down methods lack a proteolytic step and exploit high accuracy precursor and product ion masses for comparison to the expected translated sequences.[7] This measurement provides immediate feedback on post translational modifications (PTMs) that may or may not be present in the analyzed sample. Agreement in precursor mass between theoretical and observed measurements constitutes a major leap forward toward characterizing the identified protein, as opposed to just confirming its presence. Several groups have demonstrated nearly complete characterization of intact proteins, but the analyses are typically targeted, single protein infusion type experiments.[8−11] Thus far, high-throughput top down analysis of complex mixtures has rarely been reported, with a few exceptions.[12,13] While impressive results in terms of the total number of identified proteins have been achieved, fully characterizing each identified protein remains a substantial challenge that has not been surmounted by collision induced dissociation (CID), electron capture dissociation (ECD), or electron transfer dissociation (ETD) methods. The high mass accuracy product ion measurements achieved in top down experiments provide an impressive level of specificity, requiring relatively few matching fragments for a positive identification.[14] Top down search algorithms have capitalized on this by

Received: November 27, 2013
Accepted: January 21, 2014
Published: January 21, 2014

   dx.doi.org/10.1021/ac403859a | *Anal. Chem.* 2014, 86, 2185−2192

allowing expanded precursor mass tolerances that can accommodate unforeseen PTMs and mass shifts (including subtractive modifications like sequence truncations or incorrect start sites).[15,16] This search strategy can readily identify proteins whose masses differ significantly from the translated sequence. While confirmation of a PTM's presence on a given protein is an important achievement by itself, its localization and relative quantitation on the matched sequence is an ultimate goal. Although top down proteomic methods provide a large amount of information, the analytical challenges associated with their implementation (i.e., requiring satisfactory chromatography and high-resolution MS measurements on a liquid chromatography−mass spectrometry (LC−MS) time scale) are significant impediments to its widespread adoption in the field. This requirement limits practitioners to Fourier transform ion cyclotron resonance (FTICR), Orbitrap, or high resolution time-of-flight (TOF) instruments. For the former two platforms, resolution is proportional to transient acquisition time, and since high-resolution measurements are required for both precursor and their highly charged product ions, longer duty cycles result in decreased throughput. In addition to the fundamental difficulty in high-resolution detection, proteins and large biomolecules produce charge state distributions that clutter the spectral landscape while simultaneously diluting the total ion current. Indeed, computational studies have demonstrated that this is the single largest detriment to sensitivity in intact protein MS.[17] All of these challenges are encountered even before facing the difficulties in activation and fragmentation. Currently, the majority of top down proteomics identifications are the result of some variation of threshold fragmentation (either collision induced dissociation (CID), higher energy CID (HCD), or nozzle skimmer[18]/prefolding[19] dissociation (NSD)). These robust methods are widely applicable across a large range of polypeptide sizes but suffer from several shortcomings. Generally, as polypeptide length increases, collision-based methods result in fewer and fewer fragments that are representative of the middle region of the protein sequence.[20] Additionally, these threshold fragmentation methods cleave the weakest bonds first, resulting in a bias toward Xxx-Pro and Glu-Xxx/Asp-Xxx bonds[21] and the preferential cleavage of labile PTMs (such as phosphorylation and sulfation).

We recently reported the performance metrics of ultraviolet photodissociation (UVPD) for individual proteins up to 29 kDa by direct infusion,[22] thus motivating our effort to evaluate UVPD for LC−MS workflows. Presented here is a comparison of 193 nm ultraviolet photodissociation to collision-based methods for fragmentation of protein ions in a high-throughput format using the *Escherichia coli* ribosome and the fractionated *Saccharomyces cerevisiae* proteome as models. Ribosomes are the molecular machines responsible for translation of mRNA into functional proteins and their constituent protein complement serves as an ideal yardstick for method comparison due to its relative simplicity (~56 proteins in *E. coli*), generally small size amenable to top down MS analysis, high average isoelectric point, and stoichiometric abundance. Previous methods analyzing ribosomes from various organisms have achieved success using a variety of approaches.[4,23,24] Ribosomal proteins constitute a large proportion of the proteins identified in top down experiments due to both their high abundance relative to the basal proteome during normal growth conditions and high ionization efficiency in positive polarity (stemming from the high average isoelectric point).[25] For the present comparative study, ribosomes isolated from *E. coli* for *in vitro* translation of mRNA transcripts were analyzed using both collision based (HCD) and UV photodissociation methods.

MATERIALS AND METHODS

High-Throughput UVPD Optimization. Horse myoglobin, β-lactoglobulin, and β- casein were acquired from Sigma Aldrich (St. Louis, MO). Model proteins (10 μM) were prepared in 50/49/1 acetonitrile/water/formic acid (v/v/v). Proteins were infused at 5 μL/min. In order to determine the optimal laser conditions for UVPD, either one or two laser pulses was used and the laser power was varied from 0.5 to 3.0 mJ/pulse. Product ion spectra were acquired with 1, 3, and 5 averaged scans at three different resolution settings. This optimization resulted in the selection of a single 2.5 mJ pulse with three scans averaged for the LC−UVPD-MS experiments.

Ribosomal Preparation. *E. coli* ribosomes were purchased from New England Biolabs (Ipswich, MA). Nucleic acids were precipitated from solution following the method of Hardy et al.[26] Briefly, 0.25 volumes of $Mg(OAc)_2$ and one volume of glacial acetic acid were added to the ribosome suspension. The solution was incubated at 4 °C for 1 h prior to centrifugation to pellet the rRNA using a benchtop centrifuge. The samples were then neutralized via buffer exchange into 25 mM NH_4HCO_3 using 3 kDa molecular weight cutoff filters (Millipore, Billerica, MA). The ribosomal proteins were then reduced by addition of dithiothreitol to 5 mM and incubation at 55 °C for 30 min. Alkylation was performed in the dark for 30 min with 5 mM iodoacetamide. Proteins were injected directly without further manipulation.

Yeast Lysate Preparation and GELFREE Fractionation. *S. cerevisiae* BY4742 were grown at 30 °C and harvested during log-phase growth (OD_{600} 0.7). Cells were lysed by boiling in a buffer comprised of 5% sodium dodecyl sulfate (SDS), 5% glycerol, 50 mM dithiothreitol (DTT), and 50 mM Tris (pH 7.5) supplemented with protease and phosphatase inhibitors (Halt inhibitor cocktail, ThermoPierce, Rockford, IL). Lysates were cleared by centrifugation (7000*g*, 5 min) prior to protein quantification using the bicinchoninic acid assay (BCA, ThermoPierce, Rockford, IL). Protein (400 μg) was prepared for GELFrEE by acetone precipitation followed by resuspension in GELFrEE loading buffer, DTT, and water per manufacturer's instructions (Expedeon, Cambridgeshire, U.K., GELFREE 8100). After separation, fractions were precipitated[27] from their 0.1% SDS containing buffer and resuspended in aqueous solvent containing formic acid.

LC−UVPD-MS/MS. Reduced and alkylated ribosomes (3.8 pmol) were loaded onto a C4 trap column and subsequently eluted onto a 40 cm C4 analytical column packed in house with 5 μm particles containing 300 Å pores (Bruker-Michrom, Auburn, CA). Mobile phase A was 0.1% aqueous formic acid, and mobile phase B was 0.1% formic acid (v/v) in acetonitrile. Proteins were eluted using a 90 min gradient with a linear increase to 15% eluent B in the first 10 min followed by an increase to 60% eluent B over the next 80 min at 300 nL/min. Proteins eluted directly off of the column into a Thermo Scientific Orbitrap Elite mass spectrometer (Bremen, Germany) customized to accommodate photodissociation (previously described elsewhere).[22,28] Briefly, a flange containing a CaF_2 window was installed on the back of the Orbitrap in line with the center of the HCD collision cell. A 193 nm ArF excimer laser (Coherent ExciStar XS) was synchronized to initiate a single 2.5 mJ, and a 5 ns pulse at 500 Hz as the

dx.doi.org/10.1021/ac403859a | *Anal. Chem.* 2014, 86, 2185−2192

targeted ion cloud was transferred into the HCD cell. For ribosomal analyses, MS1 spectra were acquired at 240k resolving power (at m/z 400) and product ion spectra after UVPD were acquired for the top three most abundant precursors by averaging three scans at 240k resolving power. For all analyses, the HCD cell pressure was reduced to ~2 mTorr (relative to the standard HCD cell operating pressure of 10 mTorr) which decreased the pressure in the Orbitrap mass analyzer and thus enhanced the detection of low abundance and larger fragment ions.

Data Processing. RAW files were deconvoluted using the Xtract algorithm embedded in a custom version of ProSightPC 3.0, modified ad hoc to accommodate all of the product ion types encountered in UVPD.[22] The Poisson-based P-Score model was adjusted, as needed, to accommodate the additional ion types. Ribosomal analyses were searched against both a forward and decoy custom database (170 candidate sequences) of ribosomal proteins using a 1000 Da precursor mass tolerance and a 10 ppm product ion tolerance in delta M mode. Yeast fractions were searched against forward and decoy versions of a preannotated yeast database (1 196 890 candidate sequences) with a 2.2 Da precursor mass tolerance and a 10 ppm product ion tolerance without using delta M mode. All searches were conducted to accommodate the nine major UVPD ion types (a, $a + 1$, b, c, x, $x + 1$, y, $y - 1$, and z)[29] unless otherwise noted. Fragment ions were matched within a single spectrum for each precursor and its associated set of product ions. False positive rates were determined empirically and all reported identifications were based on E values that correlated to less than a 1% false positive rate across all protein spectral matches in a given analysis. For all analyses described, an E value cutoff of 1×10^{-3} correlated to a false positive rate of less than 1%.

RESULTS AND DISCUSSION

As stated previously, top down protein characterization thus far has largely been limited to single proteins in infusion type experiments for which high purity and copious amounts of sample allow extensive signal averaging to enhance the resolution of crowded product ion spectra. Using this approach, complete or nearly complete characterization of several proteins has been demonstrated, in which fragment ions representing cleavage at every inter-residue bond were observed.[20,22] To attain this level of characterization on an LC time scale requires informative product ion spectra to be obtained in each scan without excessive signal averaging. To confirm that product ion spectra were acquired at ideal conditions for UVPD, a series of model proteins was infused to allow assessment and optimization of experimental parameters to be used in an online LC−MS method. Specifically, the number of averaged scans and the user defined nominal resolution (which correlates to the resolving power observed at m/z 400) were interrogated. Supplemental Figure 1A,B in the Supporting Information shows the outcomes for one protein (myoglobin) with respect to the changes in the number of matched fragment ions as each parameter was varied. At resolving powers of 60k and 120k, the number of matched ions was 64 on average per spectrum, whereas it increased to an average of 104 at a resolving power of 240k. In terms of the impact of the number of scans, a single scan resulted in 34 fragment ions identified, whereas averaging 3 or 5 scans led to the identification of 96 or 104 fragment ions per spectrum, respectively, for myoglobin.

To isolate each variable from another when examining the effects of the change in resolving power on the number of fragments observed, the maximum number of averaged scans (five) was used, and for evaluating the effects of the number of averaged scans, the maximum resolving power was used (240k at m/z 400). From examination of these two variables individually, the benefit of using the highest possible resolution is clearly supported (Supplemental Figure 1A in the Supporting Information), whereas the gain in matched fragment ions for averaging more than three scans was marginal. From this analysis for myoglobin as well as other model proteins of varying sizes (data not shown), the conditions were set for the ribosomal analysis to use a single pulse at 2.5 mJ for UVPD and to acquire the resulting spectra with three averaged scans at the maximum resolution (240k).

Intact protein separation is a notoriously challenging part of high-throughput top down analysis, and is widely considered to be the bottleneck preventing top down proteomics from being more widely adopted. This problem has been overcome previously by maximizing peak capacity via multiple dimensions of both isoelectric point and size-based fractionation prior to reversed phase LC−MS.[13] The ribosome represents a mixture of low complexity in comparison to whole cell lysates, but even with low complexity samples attaining adequate separation is crucial for top down fragmentation. This requirement stems from the wide precursor ion isolation windows (15−100 m/z) used to increase transmission efficiency and coisolate multiple proteoforms (with a single phosphorylation, for example) that are likely to share a high proportion of fragment ions.[30] As proteins become larger, they occupy more charge states and as such are more likely to be observed throughout the m/z landscape examined. Combining the wide isolation width requirement with the increased likelihood of charge state envelopes occupying more of the m/z landscape means that without adequate resolution separations, there will be a considerable degree of precursor overlap during ion isolation, complicating the database search and increasing the likelihood of false positives. For more complex mixtures, such as the yeast whole cell lysate also examined, proteins were separated first using gel eluted liquid fraction entrapment electrophoresis (GELFrEE) prior to online fractionation by nanoLC with elution directly into the mass spectrometer. Chromatographic separation was highly reproducible using a C4 300 Å pore size reversed phase column (Supplemental Figure 2 in the Supporting Information).

To confirm that optimal conditions were achieved for the HCD analyses under the reduced pressures required for online intact protein and product ion detection, the samples were run using three different normalized collision energy settings (NCE 15, 25, and 35) and the results compared (Supplemental Figure 3 in the Supporting Information). In general, the optimal HCD conditions were dependent on the protein analyzed, more so than any variations in the corresponding UVPD performance obtained for different proteins. While the three HCD settings tested by no means represent an exhaustive optimization procedure, they are an adequate parameter range for determining methods that will optimally dissociate the largest proportion of proteins encountered in a top down high-throughput analysis. To most accurately portray the amount of information obtained for each individual protein, the number of fragment ions observed in each product ion spectrum are divided by the total number of residues for the matched protein as shown in Figure 1. As reflected by the uniformly higher

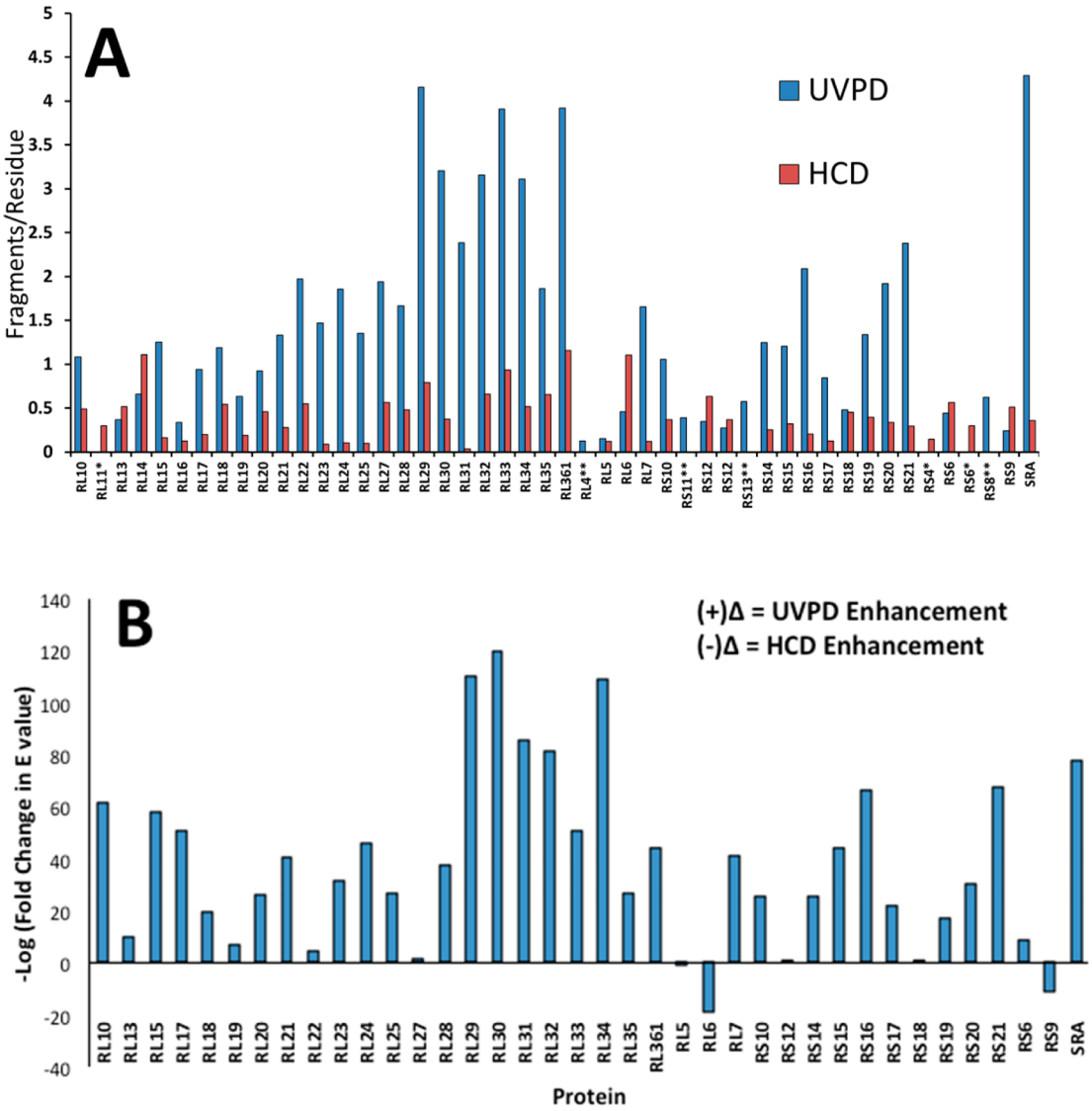


Figure 1. (A) Histogram depicting the number of fragments matched for ribosomal proteins normalized to the protein length for UVPD (1 pulse, 3 scans averaged) in comparison to the optimal HCD conditions (25 NCE and 3 scans averaged). (B) Histogram of the log of the magnitude change in confidence associated with UVPD compared to HCD. In part A, proteins followed by a single asterisk (∗) were observed only in the HCD analysis, while proteins followed by two asterisks (∗∗) were observed only in the UVPD analysis.

number of fragment ions observed for UVPD in Figure 1, UVPD vastly outperformed HCD in terms of both characterization and identification for the majority of proteins observed. All identified proteoforms and their sequence coverages are summarized in Supplemental Table 1 in the Supporting Information.

All analyses were performed using the same ad hoc modified search algorithm in order to match each spectrum with the most inclusive array of possible fragment ions. The search algorithm included a, a + 1, b, c, x, x + 1, y, y − 1, and z-type ions. The fragments produced from HCD were predominantly b- and y-type, as expected, and primarily a- and x-type for UVPD along with secondary contributions from b, c, y, and z ions. The relative portion of a/x, b/y, and c/z product ions generated upon UVPD-MS analysis of the ribosomal proteins is shown in Figure 4. The increase in protein characterization by UVPD is not due solely to an increase in the number of fragmentation channels. If a small peptide dissociates into just b- and y-type ions, it is understandable that a lower number of fragment ions would be observed than if fragmentation channels resulting in a, b, c, x, y, and z-type ions were accessed; however, UVPD does not simply result in additional *redundant* ions. UVPD produced both a greater number of fragment ions than HCD as well as many additional ones arising from unique stretches of the protein sequence. Several examples are illustrative of this outcome from the ribosomal analysis. In an extreme case, the protein that returned the highest number of fragment ions per residue, the stationary-phase-induced ribosome associated protein (P68191, M_r = 5 092.77 Da, abbreviated as "SRA" in Figure 1) resulted in identification of 193 fragment ions using UVPD and 81 fragment ions under the best HCD conditions examined (Figure 1). The protein was completely characterized by UVPD (i.e., fragments were observed representing every inter-residue position), while seven inter-residue positions were missed using HCD. Of the 193 ions observed for UVPD, 83 contained the N terminus and 51% of those were positionally unique (no redundancy in charge state or ion type relative to a specific interresidue site),

dx.doi.org/10.1021/ac403859a | *Anal. Chem.* 2014, 86, 2185−2192

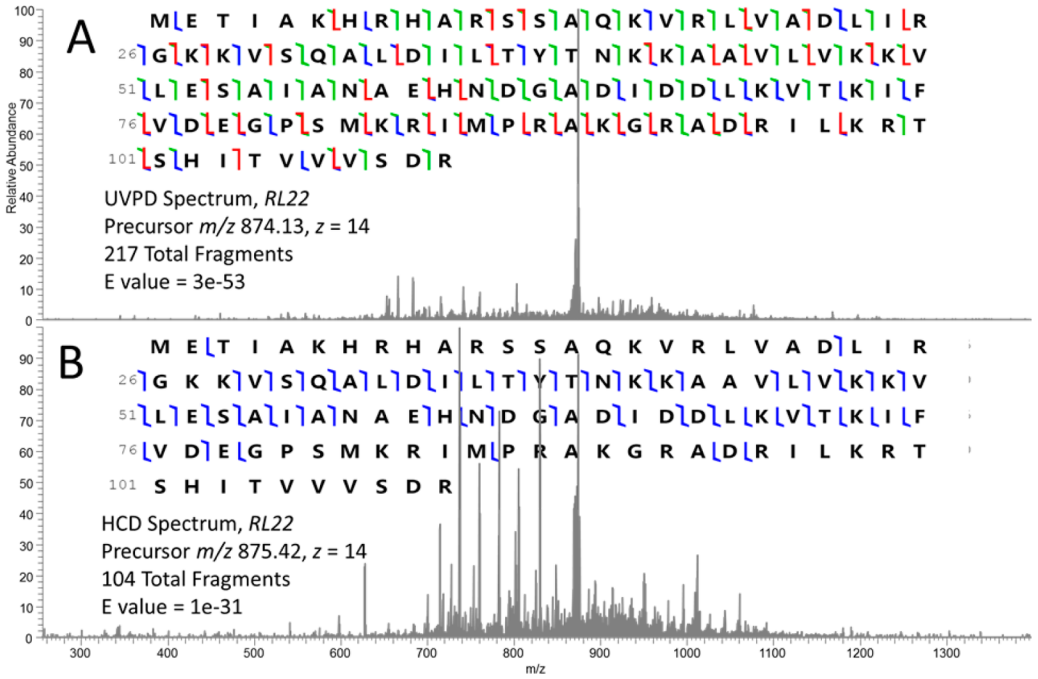


Figure 2. MS/MS spectra and associated ion maps for ribosomal protein L22 (C from *E. coli* identified using (A) UVPD and using (B) HCD. In each ion map, the cleavages resulting in b- and y-type ions are blue, the cleavages resulting in a- and x-type ions are green, and the cleavages resulting in c- and z-type ions are red.

while 110 were C terminally derived and 46% were positionally unique. The corresponding best HCD spectrum (NCE 25) was interrogated using the same candidate ion types as UVPD, and of the 81 fragments observed 40 were N-terminally derived (with 68% positionally unique) and 41 were observed from the C terminus (with 73% positionally unique). For the protein that yielded the highest number of fragment ions using HCD, RL22 (M_r = 12 218.77 Da) (Figure 2) resulted in 104 total fragment ions. Of those, from the N terminus 39 out of 45 (87%) were positionally unique, and 51 of 59 (86%) from the C terminus were positionally unique. UVPD resulted in a greater than 2-fold increase in the total number of ions (217 in total), with positionally unique fragmentation resulting from 89 of 124 (72%) N terminal ions and 54 of 93 (58%) C terminal ions.

On the basis of these results, UVPD exhibits some informational redundancy due to the multiplicity of ion types produced but there are also a significantly greater number of ions covering regions of the protein sequence (typically the middle region) that are not generated by HCD.[20] In short, the richer fragmentation patterns produced by UVPD outweigh the cost of increasing the number of searched ion types and offset the potential drawbacks of redundant sequence ions, especially with respect to online *characterization* of proteins. To examine the changes in sensitivity for identification associated with increasing the number of searchable ion types from two (b, y) to nine (a, a + 1, b, c, x, x + 1, y, y − 1, z), the HCD data set which provided the best results (NCE25) was used for parallel bioinformatic analyses. The same deconvoluted .PUF file was searched against the same database using the two different algorithms (two ion types versus nine ion types). As expected, the total number of false positives (as defined by hits matched to a randomized decoy database) increased with the number of searchable ion types for the HCD data (Figure 3).

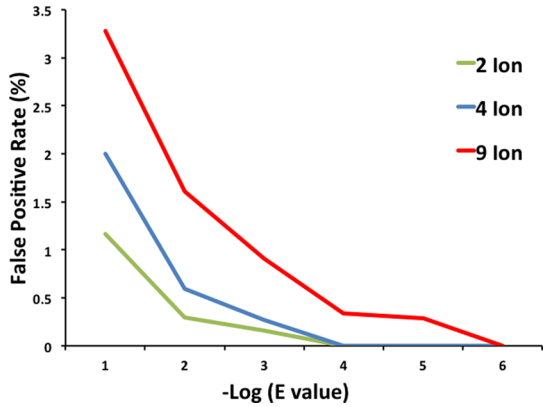


Figure 3. Graphical display of the increase in false positive rate as a function of the number of ion types interrogated as shown by the relationship between false positive rate and *E* value (displayed here as the inverse log for convenience) for the ribosomal analysis performed with HCD using the standard algorithm for b- and y-type ions (green line) and for UVPD using both the conservative four ion algorithm for identification (blue line) and the more aggressive nine ion algorithm for characterization (red line).

For more complex mixtures, a two-phase approach was evaluated as a method for increasing the potential for protein *identification* by UVPD. On the basis of the relationship depicted in Figure 3, an algorithm to match only the most abundant ion types produced upon UVPD (a, a + 1, x, and x + 1, see Figure 4) was created to decrease the amount of false positives as an initial step to narrow down the search space for increased confidence and sensitivity in identification. The three curves depicted in Figure 3 demonstrate the decrease in confidence associated with the total number of fragment ion types interrogated. Comparing the green (HCD analysis searching two ion types) and the red (UVPD analysis searching nine ion types) analyses results in a 2.8-fold increase (from 1.2

dx.doi.org/10.1021/ac403859a | *Anal. Chem.* 2014, 86, 2185−2192

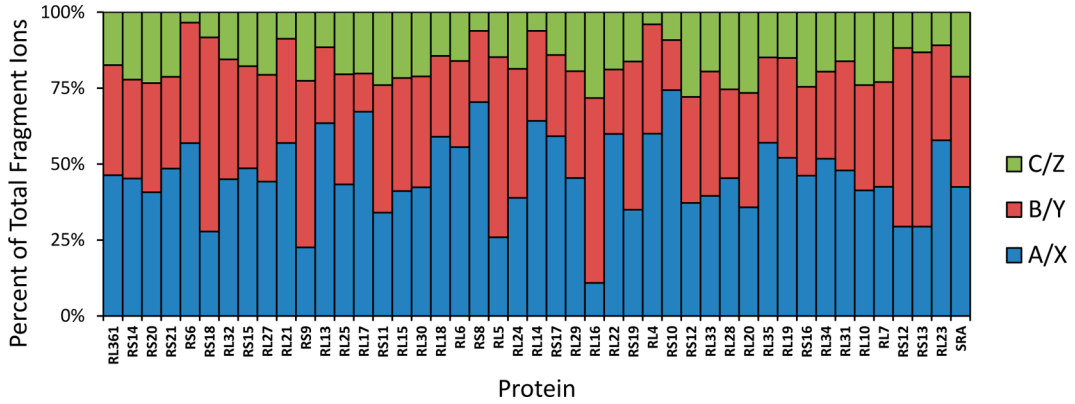


Figure 4. Distribution of fragment ions for identified ribosomal proteins using UVPD. The high proportion of AX ions (46% on average) allowed for creation of a faster and more sensitive custom algorithm to be used for identification purposes only. Postidentification, other UVPD fragment ion types can then be included for characterization of the entire protein primary structure.

to 3.3%) in false positive rate at the highest (least confident) *E* values examined. To alleviate this deficiency in identification that would have adverse effects on lower abundance proteins with suboptimal quality product ion spectra, the algorithm represented by the blue curve represents a practical compromise. In an optimized two-step bioinformatic workflow, the UVPD algorithms can be used in sequence starting with the more sensitive 4-ion version to effectively narrow down the search space. As stated previously, the 4-ion algorithm searches for a-, a + 1-, x-, and x + 1-type ions, which represent an average of 46% of all fragment ions identified using UVPD (Figure 4).

This high proportion ensures a low likelihood of proteins being identified with the 9-ion algorithm but not the 4-ion (these theoretical "missed hits" would represent false negatives in the first more sensitive search). Once the search space has been reduced to include only candidate sequences with *E* values lower than that which correlates to a threshold empirically determined false positive rate, the 9-ion algorithm can be used for more complete characterization. While this two step searching approach will have little effect on the *E* values of proteins identified with, for example, 150 matched fragment ions, the combined effects of a drastic reduction in the number candidate sequences and an increase in the number of searched ion types will benefit lower abundance proteins identified from suboptimal quality product ion spectra. Using both the 4- and 9-ion algorithms on a single data set will allow both identification- and characterization-centric bioinformatic analyses to be performed with a minimal loss in sensitivity when compared to fragmentation methods that produce only two ion types.

Following analysis of the *E. coli* ribosome, a significantly more complex mixture, an *S. cerevisiae* whole cell lysate, was fractionated based on molecular weight and the resulting fractions were analyzed by LC−MS/MS to demonstrate the utility of UVPD for high-throughput characterization. Spectra were searched against both forward and decoy versions of a preannotated yeast proteome database using the more inclusive nine ion algorithm for characterization. From this analysis, 292 proteoforms were identified representing 215 canonical sequences and of those 168 proteoforms contained some sort of PTM. All identified proteoforms, their sequence coverages, and a histogram depicting the mass distribution of identified proteins are included in Supplemental Table 2 in the Supporting Information. An example of a charge deconvoluted UVPD mass spectrum acquired for a protein containing a PTM is shown in Figure 5.

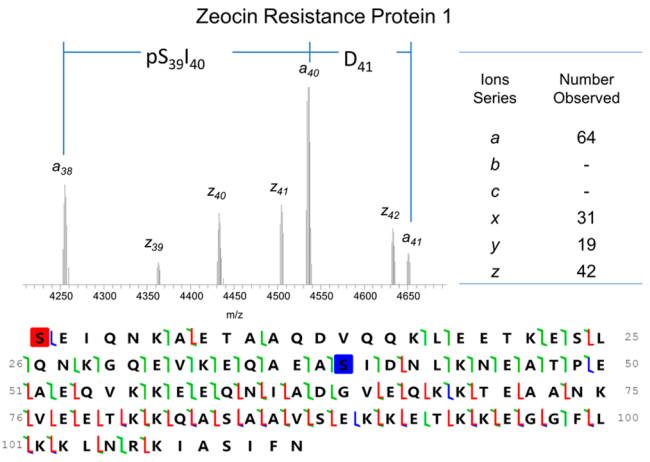


Figure 5. Shown at top left is a zoomed charge deconvoluted region of the spectrum assigned to Zeocin resistance protein (Q08245, M_r = 12 450.46 Da), which displayed nearly complete interresidue fragmentation upon UVPD. In the ion map, the cleavages resulting in b- and y-type ions are blue, the cleavages resulting in a- and x-type ions are green, and the cleavages resulting in c- and z-type ions are red. This allowed localization of the single phosphorylated moiety to residue Ser39 (residue highlighted in a blue box in the sequence). This proteoform was also acetylated at the N terminus (coded in a box at the first residue) following demethionylation. The fragment ion type distribution is shown on the right.

Zeocin resistance protein has been identified as phosphorylated on several different sites in multiple large scale global phosphoproteomic analyses.[31−34] Despite the fact that the sites of phosphorylation have been identified previously, top down analysis provides unprecedented information by confirmation of phosphorylated and nonphosphorylated sites simultaneously. Single residue specific localization of the phosphorylation was enabled not only by the extensive inter-residue fragmentation throughout the primary sequence but also by the fact that UVPD maintains the CID- and HCD-labile modification on the serine side chain. Previous analysis using peptides (in both positive[35] and negative[36,37] polarity) and proteins has demonstrated that modifications are maintained similarly to electron-based fragmentation methods. After UV photon

dx.doi.org/10.1021/ac403859a | *Anal. Chem.* 2014, 86, 2185−2192

absorption, ions in excited electronic states may exhibit nonergodic dissociation behavior that permits access to many fragmentation pathways,[38] including ones that do not disrupt the labile bonds associated with PTMs. This level of characterization identifies a specific proteoform that can be correlated to specific biological conditions. The bottom-up analyses were able to identify that there were multiple phosphorylated sites on the protein,[31−34] but they were all present on different tryptic peptides, making observation of all sites at the same time impossible using bottom-up methods. With top down UVPD, the intact mass measurement provided information concerning the number of phosphorylations (just a single phosphorylation was observed at detectable levels), and the extensive fragmentation upon UVPD allowed unambiguous localization to a single residue, while simultaneously ruling out other previously observed sites. This same or higher level of characterization was observed for many proteins, thus facilitating confident identification of multiple proteins that differed by a single amino acid. ATPase proteolipid 1 and 2 differ by a single substitution of serine to alanine in the middle of the protein sequence. Compositionally, the chemical change associated with this substitution is identical to artifactual oxidation, which is frequently observed in mass spectrometry. The high level of product ion sequence coverage achieved on either side of the changed residue using UVPD allowed unambiguous matching to the correct product ion sequence.

CONCLUSION

We have demonstrated the utility of 193 nm UVPD for intact protein identification and characterization in a high-throughput online fashion. Method development work performed using prepared *E. coli* ribosomes confirmed the superior performance of UVPD when compared to HCD under optimal conditions. From a single analysis of the *E. coli* ribosome, UVPD identified 45 unique sequences compared to 44 identified by HCD. Of greater importance, the average number of fragment ions observed was approximately 2-fold greater in the matched UVPD spectra resulting in a substantial increase in confidence (in the form of decreased *E* values). Despite the fact that there are more fragmentation channels accessed by UVPD, the higher number of fragment ions does not result entirely from a duplicative multiplicity, and the benefit of increased UVPD fragmentation outweighs the decrease in confidence from accommodating all produced ion types. Additionally, an increase in confidence is achieved by tailoring the database search to match only the most abundant ion types produced following UVPD. Finally, a more complex fractionated yeast whole cell lysate was examined, resulting in identification of 292 proteoforms from 215 canonical sequences. In conclusion, UVPD has been shown to outperform HCD for high-throughput protein characterization and similar comparisons to electron-based fragmentation methods are underway.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jbrodbelt@cm.utexas.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank John Savaryn and Emma Doud for assistance in creating the yeast sample. Funding from the NIH (Grant R21 GM099028 to J.S.B., Grant R01 GM067193 to N.L.K., Grant P30 DA018310 to N.L.K.) and the Welch Foundation (Grant F1155 to J.S.B.) and assistance provided by Thermo Fisher Scientific in implementing UVPD on an Orbitrap Elite mass spectrometer is acknowledged.

REFERENCES

(1) Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233−6237.
(2) Wu, S.-L.; Kim, J.; Hancock, W. S.; Karger, B. *J. Proteome Res.* **2005**, *4*, 1155−1170.
(3) Boyne, M. T.; Garcia, B. A.; Li, M.; Zamdborg, L.; Wenger, C. D.; Babai, S.; Kelleher, N. L. *J. Proteome Res.* **2008**, *8*, 374−379.
(4) Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. *J. Proteome Res.* **2010**, *9*, 3886−3890.
(5) Wu, C.; Tran, J. C.; Zamdborg, L.; Durbin, K. R.; Li, M.; Ahlf, D. R.; Early, B. P.; Thomas, P. M.; Sweedler, J. V.; Kelleher, N. L. *Nat. Methods* **2012**, *9*, 822−824.
(6) Smith, L. M.; Kelleher, N. L. *Nat. Methods* **2013**, *10*, 186−187.
(7) Roth, M. J.; Forbes, A. J.; Boyne, M. T.; Kim, Y.-B.; Robinson, D. E.; Kelleher, N. L. *Mol. Cell. Proteomics* **2005**, *4*, 1002−1008.
(8) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1774−1779.
(9) Thomas, C. E.; Kelleher, N. L.; Mizzen, C. A. *J. Proteome Res.* **2006**, *5*, 240−247.
(10) Li, X.; Yu, X.; Costello, C. E.; Lin, C.; O'Connor, P. B. *Anal. Chem.* **2012**, *84*, 6150−6157.
(11) Peng, Y.; Chen, X.; Zhang, H.; Xu, Q.; Hacker, T. A.; Ge, Y. *J. Proteome Res.* **2012**, *12*, 187−198.
(12) Wu, S.; Brown, R. N.; Payne, S. H.; Meng, D.; Zhao, R.; Tolić, N.; Cao, L.; Shukla, A.; Monroe, M. E.; Moore, R. J.; Lipton, M. S.; Paša-Tolić, L. *Int. J. Proteomics* **2013**, *2013*, 1−10.
(13) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254−258.
(14) Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. *Nat. Biotechnol.* **2001**, *19*, 952−957.
(15) LeDuc, R. D.; Taylor, G. K.; Kim, Y.-B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. *Nucleic Acids Res.* **2004**, *32*, W340−W345.
(16) Liu, X.; Sirotkin, Y.; Shen, Y.; Anderson, G.; Tsai, Y. S.; Ting, Y. S.; Goodlett, D. R.; Smith, R. D.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2012**, *11*, M111.008524.
(17) Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2011**, *83*, 6868−6874.
(18) Loo, J.; Edmonds, C.; Smith, R. *Science* **1990**, *248*, 201−204.
(19) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. *Science* **2006**, *314*, 109−112.
(20) Cannon, J. R.; Kluwe, C.; Ellington, A.; Brodbelt, J. S. *Proteomics*, submitted for publication.
(21) Chanthamontri, C.; Liu, J.; McLuckey, S. A. *Int. J. Mass Spectrom.* **2009**, *283*, 9−16.
(22) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135*, 12646−12651.
(23) Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P. *J. Proteome Res.* **2006**, *6*, 337−347.
(24) Swatkoski, S.; Gutierrez, P.; Wynne, C.; Petrov, A.; Dinman, J. D.; Edwards, N.; Fenselau, C. *J. Proteome Res.* **2008**, *7*, 579−586.

(25) Wynne, C.; Edwards, N. J.; Fenselau, C. *Proteomics* **2010**, *10*, 3631−3643.
(26) Hardy, S. J. S.; Kurland, C. G.; Voynow, P.; Mora, G. *Biochemistry (Mosc.)* **1969**, *8*, 2897−2905.
(27) Wessel, D.; Flügge, U.-I. *Anal. Biochem.* **1984**, *138*, 141−143.
(28) Vasicek, L. A.; Ledvina, A. R.; Shaw, J.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1105−1108.
(29) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135*, 12646−12651.
(30) Cannon, J. R.; Edwards, N. J.; Fenselau, C. *J. Mass Spectrom.* **2013**, *48*, 340−343.
(31) Albuquerque, C. P.; Smolka, M. B.; Payne, S. H.; Bafna, V.; Eng, J.; Zhou, H. *Mol. Cell. Proteomics* **2008**, *7*, 1389−1396.
(32) Holt, L. J.; Tuch, B. B.; Villén, J.; Johnson, A. D.; Gygi, S. P.; Morgan, D. O. *Science* **2009**, *325*, 1682−1686.
(33) Li, X.; Gerber, S. A.; Rudner, A. D.; Beausoleil, S. A.; Haas, W.; Villén, J.; Elias, J. E.; Gygi, S. P. *J. Proteome Res.* **2007**, *6*, 1190−1197.
(34) Reinders, J.; Wagner, K.; Zahedi, R. P.; Stojanovski, D.; Eyrich, B.; van der Laan, M.; Rehling, P.; Sickmann, A.; Pfanner, N.; Meisinger, C. *Mol. Cell. Proteomics* **2007**, *6*, 1896−1906.
(35) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *J. Proteome Res.* **2010**, *9*, 4205−4214.
(36) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. *Proteomics* **2011**, *11*, 1329−1334.
(37) Luo, Y.; Yogesha, S. D.; Cannon, J. R.; Yan, W.; Ellington, A. D.; Brodbelt, J. S.; Zhang, Y. *ACS Chem. Biol.* **2013**, *8*, 2042−2052.
(38) Cui, W.; Thompson, M. S.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1384−1398.