High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer

Timothy P. Cleland,† Caroline J. DeHart,‡ Ryan T. Fellers,‡ Alexandra J. VanNispen,‡ Joseph B. Greer,‡ Richard D. LeDuc,‡ W. Ryan Parker,† Paul M. Thomas,§ Neil L. Kelleher,§ and Jennifer S. Brodbelt*†

†Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States
‡National Resource for Translational and Developmental Proteomics, Northwestern University, Evanston, Illinois 60208, United States
§Departments of Chemistry, Molecular Biosciences, and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, United States

Supporting Information

ABSTRACT: The analysis of intact proteins (top-down strategy) by mass spectrometry has great potential to elucidate proteoform variation, including patterns of post-translational modifications (PTMs), which may not be discernible by analysis of peptides alone (bottom-up approach). To maximize sequence coverage and localization of PTMs, various fragmentation modes have been developed to produce fragment ions from deep within intact proteins. Ultraviolet photodissociation (UVPD) has recently been shown to produce high sequence coverage and PTM retention on a variety of proteins, with increasing evidence of efficacy on a chromatographic time scale. However, utilization of UVPD for high-throughput top-down analysis to date has been limited by bioinformatics. Here we detected 153 proteins and 489 proteoforms using UVPD and 271 proteins and 982 proteoforms using higher-energy collisional dissociation (HCD) in a comparative analysis of HeLa whole-cell lysate by qualitative top-down proteomics. Of the total detected proteoforms, 286 overlapped between the UVPD and HCD data sets, with 68% of proteoforms having C scores greater than 40 for UVPD and 63% for HCD. The average sequence coverage (28 ± 20% for UVPD versus 17 ± 8% for HCD, p < 0.0001) was found to be higher for UVPD than HCD and with a trend toward improvement in q value for the UVPD data set. This study demonstrates the complementarity of UVPD and HCD for more extensive protein profiling and proteoform characterization.

KEYWORDS: proteomics, protein, top-down, ultraviolet photodissociation, higher-energy collisional dissociation, Orbitrap mass spectrometer, proteoform, HeLa

INTRODUCTION

Despite the enormous popularity and tremendous success of bottom-up mass spectrometry strategies,1,2 top-down approaches aimed at the analysis of intact proteins rather than mixtures of proteolytic peptides are a compelling alternative for large-scale proteomics studies.3,4 Top-down methods offer the potential to pinpoint all modifications and mutations of a protein without loss of key features, which might be overlooked owing to peptides that are not sampled in bottom-up shotgun workflows.5 Performance gains and technological advances in mass spectrometers, separation methods, and bioinformatics have accelerated the arena of top-down mass spectrometry in recent years. The number of both proteins and proteoforms identified by high-throughput top-down mass spectrometry has increased dramatically6−13 because of the use of proteome fractionation via molecular-weight separation5 or through orthogonal improvements in HPLC methods.11,12 The use of these separations has extended the number of intact proteins and proteoforms detected from the tens or hundreds14−18 to the thousands,6−9,19 while at the same time increasing both the depth and confidence of assignments of post-translational modifications (PTMs) for a diverse range of proteins. Concurrently with the development of premass spectrometry techniques, the development of new scoring metrics for top-down data (e.g., C score20) has facilitated a more complete understanding of both the confidence of protein identification and how well PTMs are localized or constrained by fragmentation data. Many high-throughput top-down analyses use higher energy collisional dissociation (HCD) for...
GELFrEE Protein Fractionation. The acetone-precipitated pellets were subsequently resuspended in 100 μL of 1% SDS by vigorous pipetting. After adding 8 μL of 1 M dithiothreitol, 12 μL of Optima-grade water (ThermoFisher Scientific), and 30 μL of 5X tris-acetate sample buffer (Expedeon) for a final volume of 150 μL, the proteins were denatured for 10 min at 95 °C. Following denaturation, all samples were pelleted at 13 200 rpm for 10 min at 4 °C. Each sample was separated into 12 MW-based fractions on a 10%T gel-eluited liquid-fraction entrapment electrophoresis (GELFrEE) cartridge, following manufacturer’s instructions (GELFrEE 8100 Fractionation System, Expedeon). For each fraction, a 10 μL sample was taken to visualize protein content and resolution by SDS-PAGE and subsequent silver nitrate stain.38 Immediately prior to LC—MS analysis, fractions were precipitated using the chloroform/methanol/water method.39 Precipitated pellets were washed with an additional four volumes of methanol to maximize SDS removal. After a short drying period (1–5 min), all pellets were resuspended in 5% ACN/0.1% formic acid using repeated pipetting aspirations. Resuspended pellets were further diluted 1:3 to 1:8 times depending on initial protein amount (as determined from Figure S1).

LC—MS

Fractions 1–7 were separated on in-house packed PLRP-S (5 μm particle size, 1000 Å pore size; Agilent Technologies) columns (trap column: 3 cm × 100 μm i.d.; analytical column: 40 cm × 75 μm i.d.) using a Dionex Ultimate 3000 UHPLC system (ThermoFisher Scientific). Mobile phases consisted of A: 0.1% formic acid and B: 99% acetonitrile, 0.1% formic acid. An initial gradient was applied as follows at a flow rate of 300 nL/min: 0–8 min at 2% B; 8–10 min increasing from 2 to 15% B; 10–37 min increasing from 15 to 55% B; 37–40 min rapidly increasing from 55 to 95% B; 40–43 min at 95% B; and 43–45 min decreasing from 95 to 2% B; 45–60 min at 2% B. 100–500 ng of protein was injected on-column from each fraction. After this initial injection, the gradient was optimized using the gradient optimization and analysis tool (GOAT; https://proteomics.swmed.edu/goat/).40 All GOAT gradients are provided in Table S1. The nanoLC system was coupled to a Fusion Lumos Orbitrap mass spectrometer (Thermo Scientific Instruments) modified for 193 nm UVVPD, as previously described for a Fusion Orbitrap mass spectrometer.41 UVPD was performed in the high-pressure trap using a single pulse (1.4 mJ, 5 ns) from a 193 nm excimer laser (Coherent Exstar XS). HCD was performed in the ion routing multipole at 15% NCE during a 0.1 ms period. The Orbitrap mass spectrometer was run using the following parameters regardless of fragmentation type: MS1 at 120 000 FTRP, 4 μscans averaged per spectrum, 1 × 105 AGC target, 15 V source fragmentation, intact protein monoisotopic precursor selection; and MS2 at 120 000 FTRP with top speed mode enabled for 7 s, 6 μscans averaged per spectrum, and 1 × 105 AGC target. (“Top speed mode” affords a variable number of MS/MS spectra per scan cycle, typically 3–4 MS2 spectra in the present study). For fractions 1–4, charge-state targets were set to 10–25+, and for fractions 5–7, charge state targets were set to 10–25+ and undetermined charge states.

Data Analysis

Raw data were uploaded to the National Resource for Translational and Developmental Proteomics (NRTDP, Northwestern University, Evanston, IL) TDPortal.35 high-performance computing environment for analysis of high-throughput
top-down proteomics data (available for academic collaborators at: http://ntdp.northwestern.edu/tdportal-request/). Intact protein MS1 spectra were first averaged using the cRAWler algorithm (developed in-house), followed by deconvolution to monoisotopic masses by means of the Xtract algorithm (Thermo Fisher Scientific). Processed data were then searched against a database generated from the SwissProt 2016_04 release of the human proteome comprising 1 × 10⁷ total candidate proteoforms. All searches entailed a three-pronged strategy, each mode of which was first defined for ProSight PTM 2.0. The first stage entailed a narrow absolute mass search (with a 2.2 Da tolerance for MS1 and a 10 ppm tolerance for MS2) to confidently identify well-matching previously detected proteoforms. An MS1 tolerance of 2.2 Da is used to be tolerant of isoping errors in the Xtract deconvolution algorithm. The second stage involved a biomarker search (equivalent to a no-enzyme search in peptide analysis, with a 10 ppm tolerance for both MS1 and MS2), which enabled the detection of previously unknown truncations. The strategy concluded with a wide absolute mass search using a 200 Da tolerance for MS1, a 10 ppm tolerance for MS2, and Δm mode enabled to accommodate unexpected post-translational modifications. The 200 Da search tolerance allows inclusion of unexpected modifications, which requires a large precursor tolerance. The choice of 200 Da is meant to encompass the possible incorporation of up to two phosphorylation modifications. Data derived from each fragmentation type (HCD or UVPD) were analyzed separately.

False discovery rate (FDR) and instantaneous q-value estimation at the protein and proteoform level were accomplished by means of a recently developed in-house algorithm (developed in-house), followed by deconvolution to monoisotopic masses by means of the Xtract algorithm (Thermo Fisher Scientific). Processed data were then searched against a database generated from the SwissProt 2016_04 release of the human proteome comprising 1 × 10⁷ total candidate proteoforms. All searches entailed a three-pronged strategy, each mode of which was first defined for ProSight PTM 2.0. The first stage entailed a narrow absolute mass search (with a 2.2 Da tolerance for MS1 and a 10 ppm tolerance for MS2) to confidently identify well-matching previously detected proteoforms. An MS1 tolerance of 2.2 Da is used to be tolerant of isoping errors in the Xtract deconvolution algorithm. The second stage involved a biomarker search (equivalent to a no-enzyme search in peptide analysis, with a 10 ppm tolerance for both MS1 and MS2), which enabled the detection of previously unknown truncations. The strategy concluded with a wide absolute mass search using a 200 Da tolerance for MS1, a 10 ppm tolerance for MS2, and Δm mode enabled to accommodate unexpected post-translational modifications. The 200 Da search tolerance allows inclusion of unexpected modifications, which requires a large precursor tolerance. The choice of 200 Da is meant to encompass the possible incorporation of up to two phosphorylation modifications. Data derived from each fragmentation type (HCD or UVPD) were analyzed separately.

For a systematic comparison of UVPD and HCD for large-scale high-throughput top-down proteomics, HeLa whole-cell lysate proteins were fractionated by molecular weight via gel-eluted liquid-fraction entrapment electrophoresis (GELFEE) prior to LC–MS/MS analysis on a high-performance Orbitrap mass spectrometer. Each of seven GELFEE fractions, ranging in MW from ∼5 to 45 kDa (Figure S1), was analyzed in triplicate using UVPD or HCD in back-to-back series to allow direct comparison of MS/MS modes, for a total of 42 LC–MS runs. Raw data files were analyzed via the TDP-Portal high-performance computing environment at Northwestern University, with an average processing time of 16 wall-clock hours, and specific metrics were used to evaluate the performance of HCD and UVPD, as summarized in the following sections. Specifically, key metrics included the number of proteins and proteoforms identified by each method as well as the sequence coverages, q values, and C scores within each set of identified proteoforms. Sequence coverage is a standard parameter used to evaluate the percentage of backbone sites cleaved in a protein. The q value is the instantaneous false discovery rate associated with the proteoform identification and represents the false discovery rate the study would have if the proteoform in question was taken as correct. The q value is derived from the distribution of Poisson scores for the forward hits compared with a distribution of scrambled results. In the present study, the q values range from 9.8 × 10⁻³ to 4.6 × 10⁻¹⁰, with lower scores being more favorable. The characterization score (C score) is a more recently introduced concept used to provide a metric for estimating the ability to confidently differentiate and assign proteoforms. C scores range from zero to greater than 500. Proteoforms assigned C scores below 3 have been neither confidently identified nor characterized, while proteoforms assigned C scores between 3 and 40 have been confidently identified but not fully characterized, and those proteoforms assigned C scores above 40 have been confidently identified and extensively characterized.

UVPD versus HCD Metrics

While high-throughput UVPD-MS has primarily been limited by bioinformatic capabilities, the recently developed TDP-Portal platform has allowed the first direct comparison of high-throughput UVPD and HCD top-down proteomics data acquired on an Orbitrap Lumos mass spectrometer. Pooling results from seven HeLa fractions analyzed in triplicate resulted in the identification of 153 proteins (Table S2; defined as UniProt accession numbers) and 489 proteoforms (Table S3; defined as a Proteoform Record, PFR; Consortium for Top-Down Proteomics Proteoform Repository http://repository.topdownproteomics.org/) at 1% FDR using UVPD and 271 proteins (Table S4) and 982 proteoforms (Table S5) at 1% FDR using HCD. The overlapping and unique proteins and proteoforms identified by UVPD and HCD are summarized in Venn diagram format in Figure 1. While HCD resulted in the identification of a greater number of proteins and proteoforms overall, the UVPD and HCD data sets still had 143 proteins (~51%) and 286 proteoforms (~24%) in common (Figure 1), which was consistent with the higher sequence coverage achieved by UVPD (Table S2).

Results and Discussion

For a systematic comparison of UVPD and HCD for large-scale high-throughput top-down proteomics, HeLa whole-cell lysate proteins were fractionated by molecular weight via gel-eluted liquid-fraction entrapment electrophoresis (GELFEE) prior to LC–MS/MS analysis on a high-performance Orbitrap mass spectrometer. Each of seven GELFEE fractions, ranging in MW from ∼5 to 45 kDa (Figure S1), was analyzed in triplicate using UVPD or HCD in back-to-back series to allow direct comparison of MS/MS modes, for a total of 42 LC–MS runs. Raw data files were analyzed via the TDP-Portal high-performance computing environment at Northwestern University, with an average processing time of 16 wall-clock hours, and specific metrics were used to evaluate the performance of HCD and UVPD, as summarized in the following sections. Specifically, key metrics included the number of proteins and proteoforms identified by each method as well as the sequence coverages, q values, and C scores within each set of identified proteoforms. Sequence coverage is a standard parameter used to evaluate the percentage of backbone sites cleaved in a protein. The q value is the instantaneous false discovery rate associated with the proteoform identification and represents the false discovery rate the study would have if the proteoform in question was taken as correct. The q value is derived from the distribution of Poisson scores for the forward hits compared with a distribution of scrambled results. In the present study, the q values range from 9.8 × 10⁻³ to 4.6 × 10⁻¹⁰, with lower scores being more favorable. The characterization score (C score) is a more recently introduced concept used to provide a metric for estimating the ability to confidently differentiate and assign proteoforms. C scores range from zero to greater than 500. Proteoforms assigned C scores below 3 have been neither confidently identified nor characterized, while proteoforms assigned C scores between 3 and 40 have been confidently identified but not fully characterized, and those proteoforms assigned C scores above 40 have been confidently identified and extensively characterized.

UVPD versus HCD Metrics

While high-throughput UVPD-MS has primarily been limited by bioinformatic capabilities, the recently developed TDP-Portal platform has allowed the first direct comparison of high-throughput UVPD and HCD top-down proteomics data acquired on an Orbitrap Lumos mass spectrometer. Pooling results from seven HeLa fractions analyzed in triplicate resulted in the identification of 153 proteins (Table S2; defined as UniProt accession numbers) and 489 proteoforms (Table S3; defined as a Proteoform Record, PFR; Consortium for Top-Down Proteomics Proteoform Repository http://repository.topdownproteomics.org/) at 1% FDR using UVPD and 271 proteins (Table S4) and 982 proteoforms (Table S5) at 1% FDR using HCD. The overlapping and unique proteins and proteoforms identified by UVPD and HCD are summarized in Venn diagram format in Figure 1. While HCD resulted in the identification of a greater number of proteins and proteoforms overall, the UVPD and HCD data sets still had 143 proteins (~51%) and 286 proteoforms (~24%) in common (Figure 1), which was consistent with the higher sequence coverage achieved by UVPD (Table S2).
thus giving ample overlap to facilitate a more detailed comparison of the MS/MS performance metrics for specific proteoforms, as described below. Solely on the basis of the number of identifications, HCD outperforms UVPD; however, the confidence in the identifications varies considerably for UVPD compared to HCD, as shown in Figures 2−4. Interestingly, 203 proteoforms (17.1%) were uniquely identified by UVPD, suggesting a high level of complementarity of UVPD and HCD.

Figure 2 displays the number of proteoforms identified for HCD and UVPD in histogram format according to q value, molecular weight, and C scores to show the range of values and clustering of proteoform parameters. The total distribution of q values (Figure 2A), molecular weights (Figure 2B), and C scores (Figure 2C) was similar for the UVPD and HCD data sets. The average molecular weight of proteins identified by HCD was slightly higher (UVPD: 13.4 ± 3.0 kDa; HCD: 14.4 ± 3.4 kDa; p < 0.001) than that identified by UVPD for all detected proteoforms (Figure 2B). UVPD afforded a significantly higher average C score (UVPD: 215 ± 260; HCD: 126 ± 187; p < 0.001) and trend toward higher average −log(q value) (UVPD: 18 ± 22; HCD: 16 ± 13; p = 0.07) than HCD. This outcome is rationalized by an increase in the average degree of sequence coverage (UVPD: 25 ± 18%; HCD: 13 ± 7%; p < 0.001) as well as the greater number of diagnostic fragment types associated with UVPD (Figure 2A,C). Furthermore, UVPD resulted in 59% of identified proteoforms being assigned a C score of 40 or higher, whereas 49% of the proteoforms identified by HCD fell into the same highest C-score category (Figure 2C). The results from Figure 2A (q values) and Figure 2C (C scores) recapitulate that the richer MS/MS spectra afforded by UVPD provides a boost in scoring metrics for proteoform identification compared with HCD.

The 286 proteoforms that overlapped for UVPD and HCD were regrouped in histogram format according to q-value and molecular weight in Figure 3. UVPD afforded a significantly higher average C score (mean UVPD gain: 89.9, 95% CI 64.0% to 115.6%, t(285 d.f.) = 6.86; p < 0.001) and higher average percent of sequence coverage (mean UVPD increase in fragment coverage: 10.6%, 95% CI 8.38% to 12.8%, t(285 d.f.) = 9.45; p < 0.001). This increase in C score is gained with no significant change in q value (t(285 d.f.) = −0.7, p = 0.4847; Figure 3A). The average molecular weight for these 286 proteoforms (13.4 kDa; Figure 3B) was representative of the molecular weight range of all of the UVPD proteoforms (p = 0.98); however, this average molecular weight was smaller than that of all proteoforms identified by HCD (p < 0.001). This discrepancy in molecular size arises from the greater number of

Figure 1. Proteins (left) and proteoforms (right) detected by high-throughput top-down proteomic analysis of seven HeLa GELFrEE fractions using HCD or UVPD fragmentation.

Figure 2. (A) −log(q value), (B) molecular weight, and (C) C-score distributions for all detected proteoforms within the HCD (dark green) and UVPD (light green) data sets.

Figure 3. (A) −log(q-value) and (B) molecular weight distributions of the 286 overlapping proteoforms between the HCD (dark green) and UVPD (light green) data sets.
proteoforms falling in the 25−30 kDa range identified by HCD; however, the 286 overlapping proteoforms provide a direct means of evaluating differences in C score (Figure 4A) and sequence coverage (Figure 4B) distributions between the UVPD and HCD data sets. The comparisons of C scores and sequence coverages obtained by UVPD and HCD of individual proteoforms are illustrated as difference plots in Figure S2, in which the 286 proteoforms are displayed across the x-axis, and the y-axis is used to convey the increase or decrease in the C score or sequence coverage for each proteoform (UVPD relative to HCD). The benchmark HCD values are tracked on the green line, and the bars that extend above the green line signify those proteoforms for which UVPD outperformed HCD.

As illustrated in Figure 4A, C scores generated from the UVPD mass spectra were ~9% higher overall than those C scores generated from HCD mass spectra (UVPD: 68% greater than C score 40; HCD: 63% greater than C score 40) for the 286 overlapping proteoforms, with 62% of overlapping proteoforms having a higher UVPD C score than HCD C score. UVPD also resulted in a significant increase in average sequence coverage (28 ± 20%, p < 0.0001) compared with HCD (17 ± 8%), with 59% of the overlapping proteoforms having greater UVPD sequence coverage (Figure 4B). On average, UVPD increased sequence coverage by 11% compared with HCD; however, five of the proteoforms displayed a sequence coverage that increased by over 60% compared with HCD. For example, there was a 74% increase in sequence coverage for small EDRK-rich factor 2 (P84101-1; PFR21392) for UVPD compared with HCD (Figure 5A). These combined results suggest that UVPD provided increased confidence in PTM localization and proteoform characterization compared with HCD, owing to the larger number of fragment ions produced, and consistent with a previous report of the promising merits of UVPD for top-down proteomics.

Localization of Post-Translational Modifications

On the basis of the increase in sequence coverage and C score observed for UVPD, the MS/MS results for 6 of the 286 overlapping proteoforms were examined in greater detail to evaluate differences in PTM localization reflected by the changes in C scores (Figure S2A, blue circles) and sequence coverage variation (Figure S2B, blue circles). Among the group of 286 proteoforms, up to 5 PTMs, including N-terminal acetylation, lysine acetylation, and arginine methylation, were localized on a single protein. For the high mobility group protein HMG-I (P17096-1; 106 residues; Figure 5B and Figure S3), at least two overlapping proteoforms possessing three (PFR17157) or four (PFR13815) PTMs were characterized. For PFR17157 (Figure 5B), UVPD localized acetylated Lys6 and yielded limited coverage of phosphorylation of Ser101 and Ser102 (C-score: 368). A similar level of localization was observed for the two phosphorylations by HCD; however, acetylation of the N-terminus or Lys6 could not be confirmed, reducing confidence in characterization (C score: 256) (Figure 5B). For PFR 13815, the methylation of Arg25 was localized by UVPD (Figure S3), and there was limited localization of the acetylation of Lys14. The phosphorylation of Ser101 and Ser102 was pinpointed by appropriate fragment ions to these two positions by UVPD (Figure S3). For HCD of this same proteoform, the phosphorylation sites were not localized, and only limited localization of the acetylation at Lys14 and methylation at Arg25 was obtained, causing the C-score to plummet to 0.1 for HCD (from 149 for UVPD). For singly

Figure 4. Distribution of gains in UVPD compared to HCD for (A) C scores and (B) sequence coverage for characterization of proteoforms. Overlays of normal distributions and kernel distributions are given for each metric.

Figure 5. Fragmentation maps of (A) EDRK-rich factor 2 (PFR21392) for HCD (12+) and UVPD (12+) and (B) high mobility group protein HMG-I with three post-translational modifications (PFR17157) for HCD (14+) and UVPD (17+). Specific residues or sites are shaded as follows: blue box, phosphorylation, red box, acetylation; green box, methylation. Backbone fragmentation markers are shown along the sequence as colored flags: a,x: green; b,y: blue; c,z: red.
modified proteoforms, a large number of N-terminal acetylations (e.g., small EDRK-rich factor 2 (Figure 5A); heat shock protein beta-1 (Figure 6A)) were mapped in addition to lysine acetylations that UVPD successfully differentiated from the alternative N-terminal forms, as exemplified by ATP synthase subunit g (PFR16756) in Figure 6B.

In some cases, HCD outperformed UVPD, as illustrated for ATP synthase delta (P30049; 146 residues; PFR1028), a proteoform that contained no PTMs. HCD returned greater sequence coverage and C-score than UVPD (Figure S4). This outcome reinforces the complementarity of HCD and UVPD, as not all proteoforms will fragment as extensively with UVPD. Some proteins are clearly better suited for HCD fragmentation than UVPD fragmentation, possibly owing to the favorable distribution of mobile protons or low frequency of amino acids susceptible to directing preferential cleavages, two factors that might otherwise suppress appropriate fragmentation upon collisional activation.

**CONCLUSIONS**

The use of UVPD fragmentation for the analysis of human proteins by high-throughput top-down proteomics resulted in both increased overall confidence in proteoform identification and improved degree of PTM localization in comparison with the more typical HCD fragmentation, as evaluated by scoring metrics provided by a new and more advanced bioinformatics platform. From these comparative analyses of HeLa GELFrEE fractions, UVPD was observed to increase sequence coverage up to 74% and the level of proteoform characterization by up to 9% compared with HCD, reflected in a particular increase in confident PTM localization. The comprehensive results obtained by the present study recapitulate the complementary nature of HCD and UVPD fragmentation for more extensive protein profiling and proteoform characterization in high-throughput top-down proteomics studies on an Orbitrap Fusion Lumos platform.

**ASSOCIATED CONTENT**

* Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.7b00043.

Figure S1. SDS-PAGE gel of HeLa GELFrEE fractions. Figure S2. Rank order C-score and sequence coverage plots of the 286 overlapping proteoforms between the HCD and UVPD data sets. Figure S3. Fragmentation maps of high mobility group protein HMG-I with four post-translational modifications (PFR13815) showing differential localization of post-translational modifications between HCD (14+) and UVPD (17+). Figure S4. Fragmentation maps of ATP synthase delta (PFR1028) showing HCD (11+) fragmentation and C-score exceeding those of UVPD (11+). Table S1. Gradient optimization and analysis tool LC %B values for fractions 1–7. (PDF)

Table S2: Summary of 153 proteins identified by pooling results from UVPD of seven HeLa fractions analyzed in triplicate. (XLSX)

Table S3: Summary of 489 proteoforms identified by pooling results from UVPD of seven HeLa fractions analyzed in triplicate. (XLSX)

Table S4: Summary of 271 proteins identified by pooling results from HCD of seven HeLa fractions analyzed in triplicate. (XLSX)

Table S5: Summary of 982 proteoforms identified by pooling results from HCD of seven HeLa fractions analyzed in triplicate. (XLSX)

**AUTHOR INFORMATION**

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

**ORCID**
Paul M. Thomas: 0000-0003-2887-4765
Neil L. Kelleher: 0000-0002-8815-3372
Jennifer S. Brodbelt: 0000-0003-3207-0217

**Notes**
The authors declare no competing financial interest. The .raw files analyzed in the searches, the .txt file used for search database creation, and all resulting .tdReport files, which
show all identified protein entries and proteoforms, as well as other statistics such as observed sequence coverage or calculated q values and C scores, are available for download here: ftp://massive.ucsd.edu/MSV000080432.

ACKNOWLEDGMENTS

Funding from the NSF (CHE-1402753), the Welch Foundation (F-1155), and NIH 1K12GM102745 (fellowship to T.P.C.) is acknowledged. Funding from the UT System for support of the UT System Proteomics Core Facility Network is gratefully acknowledged. This work was supported in part by the National Institute of General Medical Sciences P41GM108569 for the National Resource for Translational and Developmental Proteomics (NRTDP) based at Northwestern University (to N.L.K.). Further support was provided by the computational resources and staff for the Quest high performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology.

REFERENCES


