analytical chemistry

pubs.acs.org/ac Article

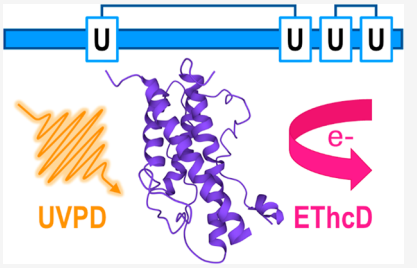
Integrated Top-Down and Bottom-Up Mass Spectrometry for Characterization of Diselenide Bridging Patterns of Synthetic Selenoproteins

Eleanor Watts, Ross Thyer, Andrew D. Ellington, and Jennifer S. Brodbelt*

Cite This: *Anal. Chem.* 2022, 94, 11175−11184 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: With the rapid acceleration in the design and development of new biotherapeutics, ensuring consistent quality and understanding degradation pathways remain paramount, requiring an array of analytical methods including mass spectrometry. The incorporation of non-canonical amino acids, such as for synthetic selenoproteins, creates additional challenges. A comprehensive strategy to characterize selenoproteins should serve dual purposes of providing sequence confirmation and mapping of selenocysteine bridge locations and the identification of unanticipated side products. In the present study, a combined approach exploiting the benefits of both top-down and bottom-up mass spectrometry was developed. Both electron-transfer/higher-energy collision dissociation and 213 nm ultraviolet photodissociation were utilized to provide complementary information, allowing high quality characterization, localization of diselenide bridges for complex proteins, and the identification of previously unreported selenoprotein dimers.



INTRODUCTION

Disulfide bridges, formed by the oxidation and linkage of two cysteine residues, contribute significantly to the stabilities and structures of proteins.[1,2] The escalating interest in the assembly, degradation, and characterization of disulfide bridges is unsurprising given their ubiquitous presence in antibody-based biotherapeutics.[3,4] Incorporation of selenocysteine instead of cysteine in proteins has recently been explored as an innovative strategy to advance the development of biotherapeutics.[5,6] Selenocysteine has a significantly lower pK_a than cysteine, resulting in the facile deprotonation of selenocysteine at physiological pH.[7] Moreover, selenocysteine has a much lower redox potential than cysteine, leading to favorable generation of diselenide bridges with increased stability relative to conventional disulfide bonds.[8] As a result of this unique chemistry, the incorporation of selenocysteine in the place of cysteine has the potential to improve the production and stability of biotherapeutic proteins. However, the unusual biosynthetic pathway involving the enzymatic conversion of serine to selenocysteine on its tRNA and constraints within the coding sequence of the target protein imposed by the native protein and RNA factors made this previously unachievable in recombinant protein systems.[9,10] Breakthroughs in synthetic biology have led to the site-specific incorporation of selenocysteine and formation of diselenide bridges in recombinant proteins generated in *Escherichia coli*.[5,6]

Given the great potential for enhancing design and development of novel biotherapeutics via incorporation of selenocysteines and/or other non-canonical amino acids, it is critical to develop versatile analytical methods for their characterization.[2,3] The characterization of proteins containing disulfide bridges and their diselenide analogues presents a challenge, and an array of approaches ranging from bottom-up to top-down tandem mass spectrometry (MS/MS) methods have been employed.[2,11,12] For bottom-up methods, proteins are typically reduced prior to enzymatic digestion, thus facilitating characterization of the primary protein sequences without the hindrance of disulfide bridges. Structural analysis of disulfide bridges has also been achieved through strategic digestion without reduction.[2] Identification of peptides resulting from enzymatic digestion is traditionally accomplished by utilizing collision induced dissociation (CID); however, in the case of peptides containing disulfide bridges, little success is achieved with CID alone.[2] As a means to reduce disulfide bonds and streamline the mapping of disulfide bridges, innovative ionization methods, including corona discharge[13] and solvent assisted photoionization,[14] as well as online electrochemical reduction,[15] have been implemented. Improved characterization has also been reported for peptides containing intramolecular disulfide bonds by examining internal ions as well as their disulfide bridge-containing complements, or "external ions" produced by CID.[16] Other MS/MS methods using alternative ion activation techniques

Received: April 1, 2022
Accepted: July 20, 2022
Published: August 5, 2022

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184

have been used to augment the characterization of disulfide-bridged peptides.[2] Both MALDI in-source decay[17−20] and electron transfer dissociation (ETD)[21−28] have been shown to yield fragment ions resulting from the cleavage of the disulfide bond. Hybrid ETD methods, including electron-transfer/CID and electron-transfer/higher energy collisional dissociation (EThcD), boost the generation of sequence-related fragment ions.[23,25] The capabilities of ultraviolet photodissociation (UVPD) for analysis of disulfide-bridged peptides have been explored.[29−31] 266 nm UVPD results in site-specific cleavage of disulfide bonds, offering higher confidence in peptide pairing and better characterization when combined with electron capture dissociation (ECD).[30,31] 193 nm UVPD has also achieved complete characterization of peptides containing complex disulfide bridges by integrating informative fragment ions resulting from cleavage of disulfide bridges with fragment ions retaining disulfide linkages.[29] The ability of 193 nm UVPD to achieve high quality characterization without requiring MS^3 approaches represented a promising benchmark for analysis of disulfide- and diselenide-bridged peptides.

While bottom-up mass spectrometry has been ideal for identifying individual disulfide bridges, top-down mass spectrometry offers the potential for greater confidence in protein characterization and allows the opportunity for comprehensive differentiation of complex proteoforms.[32,33] As also witnessed in bottom-up workflows, disulfide bridges and their analogues present a major challenge for top-down methods because the disulfide bonds often remain intact upon protein fragmentation, limiting the sequence coverage in sections spanned by the bridged regions. ETD and ECD of intact proteins result in limited cleavage of disulfide bridges, with most identified fragments resulting from backbone cleavage outside of constrained regions.[11,34−37] While the lack of fragment ions over the disulfide-bridged sections limits sequence coverage, it can help confirm the presence and location of disulfide bonds, and the reduced complexity of the resulting spectra can increase confidence in the fragment ion identifications.[11,38] Hybrid electron-capture/higher energy collisional dissociation and 157 nm UVPD promote cleavage of disulfide-bridged subunits, as demonstrated for intact antibodies; however, characterization remains limited owing to survival of some intrapeptide disulfide bonds.[12] Another hybrid method, activated ion ETD (AI-ETD), has exhibited the greatest success in characterizing the constrained regions of disulfide bridged proteins, with enhanced cleavage of intra-peptide disulfide bonds resulting in sequence coverage comparable to that of reduced proteins.[39,40] A recent study of insulin dimers has demonstrated the ability of 213 nm UVPD to successfully analyze large peptides resulting from complex linkages.[41] Although the insulin dimers examined in this seminal study were smaller than the proteins considered in most top-down studies, the results represented a promising adaptation of 213 nm UVPD for disulfide characterization.[41]

While the characterization of therapeutic proteins containing disulfide bridges has been explored extensively with tandem mass spectrometry, there has been little emphasis on proteins containing diselenide linkages.[42] A few bottom-up studies have successfully identified cysteine−selenocysteine linkages in naturally occurring selenoproteins by using CID methods.[43,44] Recent advances in development of custom recombinant selenoproteins have reinforced the need for improved analytical methods for characterization of diselenide bonds.[5,6] Initial studies of synthetic selenoproteins utilized top-down mass spectrometry, facilitated by 193 nm UVPD.[5,6] While the methodology was successful in confirming the mass of the proteins and localizing the diselenide linkages, the sequence coverage was constrained by the diselenide bridges, hindering complete characterization of the proteins.[5,6] As described in the present study, 213 nm UVPD, EThcD, ETD, and higher energy collisional dissociation (HCD) are integrated in a combined top-down/bottom-up workflow to characterize synthetic selenoproteins with the aim of characterizing the entire sequence of the proteins as well as interrogating diselenide linkages.

EXPERIMENTAL SECTION

Samples and Reagents. Recombinant selenoproteins samples were prepared as previously reported.[5] All Millipore OmniSolv LC−MS grade solvents were obtained from VWR. Lysyl-endopeptidase (Lys-C) was purchased from Santa Cruz Biotechnology. All other reagents were purchased from Thermo-Fisher Scientific unless otherwise noted.

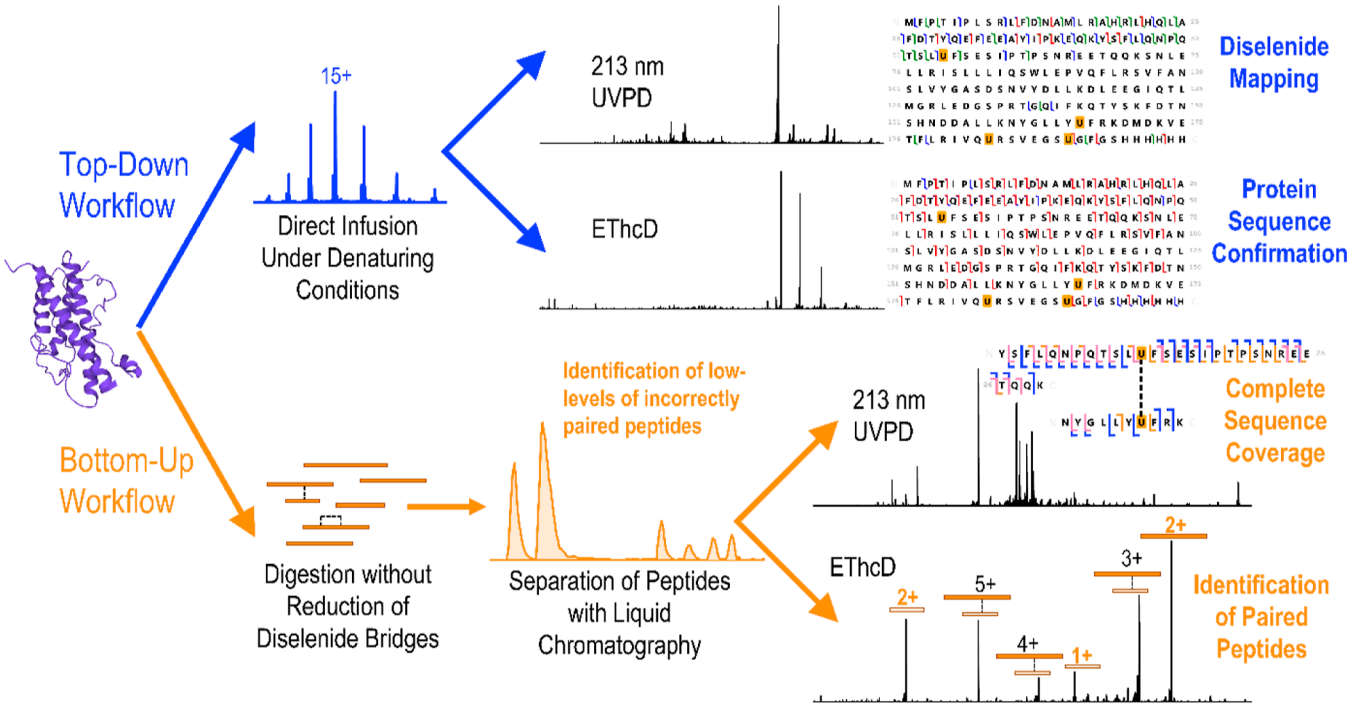
Top-Down Mass Spectrometry. Each protein sample was buffer exchanged into water with Micro Bio-Spin 6 columns (BioRad) and then diluted to approximately 10 μM in 50:50 water/acetonitrile with 1% formic acid. Samples were then loaded into house-made borosilicate emitters coated in Au/Pd for nano-electrospray infusion with an applied voltage of 0.9−1.3 kV. All experiments were performed with an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) equipped with 213 nm UVPD. For all MS1 spectra, 100 microscans were collected with a resolving power of 120,000 at m/z 200. MS2 experiments combined 500 microscans with a resolving power of 240,000 at m/z 200. Except when otherwise noted, top-down activation parameters were as follows: 15% normalized collisional energy (NCE) for HCD, 5 ms activation time for ETD and EThcD with 15% NCE supplemental activation for EThcD, and 50 ms activation period for 213 nm UVPD, equivalent to 125 pulses (approximately 3 μJ/pulse) at 2500 Hz. Top-down mass spectra were deconvoluted with the Xtract algorithm in Thermo Scientific Freestyle using a signal to noise (S/N) threshold of 3, fit factor of 80%, and remainder threshold of 25%. Fragment ions were identified using ProSight Lite with a 10 ppm tolerance for fragment ions. TDValidator was used to generate the fragment ion distributions included in Figure S1 to demonstrate the quality of the isotope patterns generated for representative fragment ions with high S/N and fit factors (Figure S1A) as well as lower S/N (Figure S1B) and fit factors close to the cutoff value (Figure S1B,C). The unusual isotope pattern of selenium resulted in deconvoluted masses shifted by approximately −1.0 Da. As a compensatory measure, 2.01 Da was subtracted from each selenocysteine reside to account for the isotope pattern and the loss of one hydrogen to form a diselenide bridge.

Bottom-Up Liquid Chromatography−Mass Spectrometry. For digestion, 50 μg of the sample was prepared in Tris−HCL at pH 6.0. Lys-C was then added at a 1:25 enzyme-to-protein ratio and samples were incubated at 37 °C for 18 h. Digests were buffer-exchanged into water with Amicon Ultra 3 kDa molecular weight cutoff centrifugal filters. Samples were stored at −80 °C until analysis. Samples were diluted to 0.1−0.3 mg/mL for a 1 μl injection.

Reversed phase liquid chromatography (LC) was performed using a Dionex Ultimate nano LC system configured to trap and elute. Trap and analytical columns were packed in-house to a length of 3 and 20 cm, respectively, using 1000 Å and 50

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184

Scheme 1. Workflow Diagram Displaying Key Information Obtained Using a Combination of Top-Down and Bottom-Up Mass Spectrometry and a Combination of 213 nm UVPD and EThcD as Illustrated for HGH (Crystal Structure PDB ID: 1HGU)



μm bulk PLRP media from Agilent. Samples were loaded onto the trap column with starting conditions of 2% acetonitrile and 0.1% formic acid in water at 5 μL/min. After 5 min of loading, a valve switch placed the trap column in-line with the analytical column. Analytical mobile phases comprising water (A) and acetonitrile (B) with 0.1% formic acid (A and B) were applied at an initial gradient of 2 to 6% B over 2 min followed by a gradient up to 35% B over 33 min at a flow rate of 300 nL/min. Introduction into the Orbitrap Fusion Lumos mass spectrometer occurred with an applied voltage of 2−2.4 kV. MS1 scans with a resolving power of 60,000 at m/z 200 and an AGC target of 4×10^5, using an m/z range of 400−2000. Top 10 data dependent MS2 scans were selected with an intensity filter of 1×10^5. All MS2 scans were collected with an AGC target of 5×10^5, within an m/z range of 300−2000. Except for where otherwise noted, three microscans were averaged for each MS^2 spectrum with a resolving power of 120,000 at m/z 200. HCD scans were collected with 30% NCE. ETD and EThcD scans were acquired with calibrated charge-dependent ETD parameters and 15% NCE supplemental activation in the case of EThcD. 213 nm UVPD activation occurred over a 100 ms activation period, equivalent to 250 pulses (approximately 3 μJ/pulses) at 2500 Hz, unless otherwise indicated.

Bottom-up data was analyzed with Byonic (Protein Metrics version 4.2.10) and was searched against a single protein database with the appropriate protein sequence. The cleavage site was set to lysine, fully specific, with two missed cleavages allowed, and both precursor and fragment mass tolerances were set to 20 ppm. The conversion of selenocysteine to serine was set as a custom modification with a mass loss of 63.92 Da at selenocysteine. Diselenide, disulfide, and Se−S bonds were searched as a custom cross-link with the loss of two hydrogens. When calculating total protein sequence coverage, only peptides that were identified in two out of three replicates were included. All peptides identified containing selenocysteine, including those with serine incorporation in the place of selenocysteine, were manually verified by examining and deconvoluting the associated 213 nm UVPD mass spectra and identifying matching fragments with ProSight Lite. EThcD mass spectra were also manually verified for specific high abundance fragments related to cleavage across the diselenide bond to confirm the presence or absence of interpeptide diselenide bridges.

DISCUSSION

As illustrated in the workflow overview in Scheme 1, the combination of top-down and bottom-up strategies provides complementary information that allows the most comprehensive characterization of selenoproteins. While top-down mass spectrometry affords high-level characterization of sequences and locations of diselenide bridges, bottom-up methods offer confident identification of incorrectly paired peptides that contribute to the protein heterogeneity. Additionally, while ETD and HCD were also explored to evaluate their attributes for mapping diselenide bridges, 213 nm UVPD and EThcD provided the most detailed information and were highly complementary in both top-down and bottom-up approaches. The benefits and limitations of each method, as well as the advantages in combining the workflows, are described in more detail in the following sections.

Top-Down Analysis. High-resolution (120,000 at m/z 200) MS1 spectra acquired for each intact selenoprotein, as exemplified in Figure 1, allows the first level of assessment of the three selenoprotein constructs. The first construct examined is green fluorescent protein (GFP), which contains only one selenocysteine pair. The only notable masses in the deconvoluted spectrum of GFP (Figure 1A) correspond to the intact protein (28,011 Da) and the intact protein after methionine loss (27,880 Da), a common truncation caused by the enzyme methionine aminopeptidase, which cleaves

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184

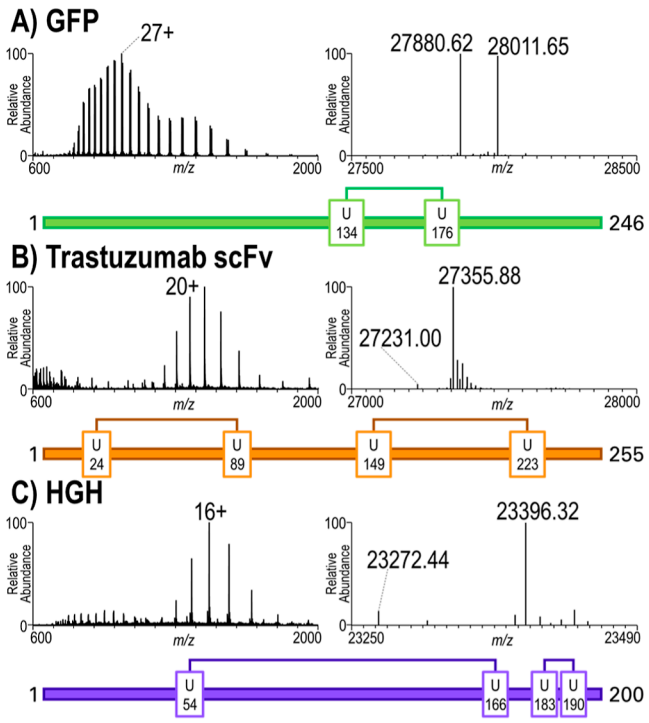


Figure 1. Electrospray ionization (left) and deconvoluted (right) mass spectra for three modified selenocysteine-containing proteins including (A) GFP, (B) trastuzumab scFv, and (C) HGH. All spectra were collected with a resolution of 120,000 at m/z 200. Schematic depictions of the pairings of diselenides for each protein are shown below each spectrum.

methionine co-translationally.[45] The other two proteins, trastuzumab single chain variable fragment (scFv) and human growth hormone (HGH), both contain two pairs of selenocysteine residues. While methionine loss was not observed for either protein, another truncated product corresponding to a mass shift of 124.9 Da was noted for each protein in the deconvoluted mass spectra (Figure 1B,C). The net mass shift of 124.9 Da is explained by the incorporation of two serine residues in place of two selenocysteine residues. The incorporation of serine instead of selenocysteine is a known side reaction that occurs because serylated $tRNA^{Sec}$ is an intermediate in the generation of selenylated $tRNA^{Sec}$.[5] Interestingly, the incorporation of serine in place of selenocysteine was observed only in pairs, a phenomenon that will be examined in greater detail in the last section of the discussion.

After initial mass analysis of the intact proteins, MS/MS spectra were collected utilizing HCD, ETD, EThcD, and UVPD. The optimization of activation parameters (e.g., collision energy, activation time) for each activation method is demonstrated for two representative charge states (27+, 30+) of GFP in Figure S2, and examples of optimized MS/MS spectra for GFP (27+) are displayed in Figure S3. Little difference in sequence coverage was observed based on charge state, with the lower and more abundant charge state (27+) typically resulting in slightly higher sequence coverage than the higher and less abundant charge state (30+). Considering these results, high abundance charge states were selected for each protein for MS/MS analysis. For trastuzumab, the 20+ and 19+ charge states had similar abundances, and the 20+ charge state was selected to maximize fragmentation for charge-dependent methods, whereas for GFP, the 16+ charge state was significantly more abundant than adjacent charge states and was therefore utilized to maximize fragment ion abundances. Based on the outcomes in Figure S2 for GFP and similar assessments for the other two proteins (data not shown), the optimal activation parameters were 15% NCE for HCD, 5 ms activation time for ETD and EThcD with 15% NCE supplemental activation for EThcD, and 50 ms activation period for 213 nm UVPD. A series of representative sequence maps and MS/MS spectra are shown in Figures 2 and S4, respectively, for trastuzumab scFv (20+). While the highest sequence coverage was obtained using EThcD (Figure 2C), a significant number of fragment ions resulted from cleavage of the diselenide bridges linking U24 to U89 and U129 to U233, creating ambiguity in bridge locations. Bracketing of the diselenide bonds was more cleanly accomplished by HCD and ETD (Figure 2A,B), as indicated by the large stretches of the sequence that reveal no or few backbone cleavages. However, HCD and ETD resulted in low sequence coverage, particularly in the N-terminal regions. 213 nm UVPD (Figure 2D) resulted in higher sequence coverage than HCD and ETD and better bracketing of diselenide bonds than EThcD. Similar results are illustrated for GFP in Figure S3. EThcD and 213 nm UVPD proved to be highly complementary methods and the utilization of both led to confident identification of diselenide bridge locations as well as characterization of the protein sequence.

The final protein examined with top-down mass spectrometry was HGH, and examples of sequence maps and MS/MS spectra are displayed in Figures 3 and S5. The two diselenide bridges in HGH constrain most of the protein sequence, and three (U166, U183, and U190) of the selenocysteines are nearly adjacent, curbing sequence coverage and complicating the identification of diselenide bridges. This limitation of the top-down approach is most evident for HCD and ETD (Figure 3A,B) for which very low sequence coverage is achieved and the locations of the diselenide bridges are ambiguous. For EThcD (Figure 3C) the sequence coverage remains high due to significant cleavage of diselenide bridges, but the diselenide bridge locations cannot be identified. 213 nm UVPD (Figure 3D) offers the greatest success in bracketing the diselenide bridges of HGH, but given the proximity of U166, U183, and U190, as well as backbone cleavages between U54 and U166, the confidence in pinpointing the location of the diselenide bridge is low. An additional impediment observed in the series of MS/MS spectra for HGH is that although EThcD offered consistently high sequence coverage, patchy fragmentation between bridged selenocysteine residues restricted coverage. To adequately characterize complex patterns of diselenide bridges, such as those observed for HGH, and decipher entire protein sequences constrained by diselenide bridges, a bottom-up method is required to fill in the gaps.

Bottom-Up Characterization of Predicted Peptides. Analysis of HGH using a bottom-up (LysC digestion) LC−MS/MS approach eliminated the shortcomings imposed by diselenide-constrained regions in the top-down strategies. High sequence coverages were observed for each MS/MS method, including 98% for HCD, 76% for ETD, 98% for EthcD, and 96% for 213 nm UVPD based on identification of 8, 6, 8, and 7 unique peptides, respectively, as listed in Table S1. A representative chromatogram is included in Figure S6. The peptides of primary interest for HGH are the two diselenide-bridged species, one containing two peptides with 29 and 10

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184

Figure 2. Sequence coverage maps for trastuzumab scFv (20+) obtained using (A) HCD (15% NCE), (B) ETD (5 ms activation time), (C) EThcD (5 ms activation time and 15% NCE supplemental activation), and (D) 213 nm UVPD (50 ms activation period).

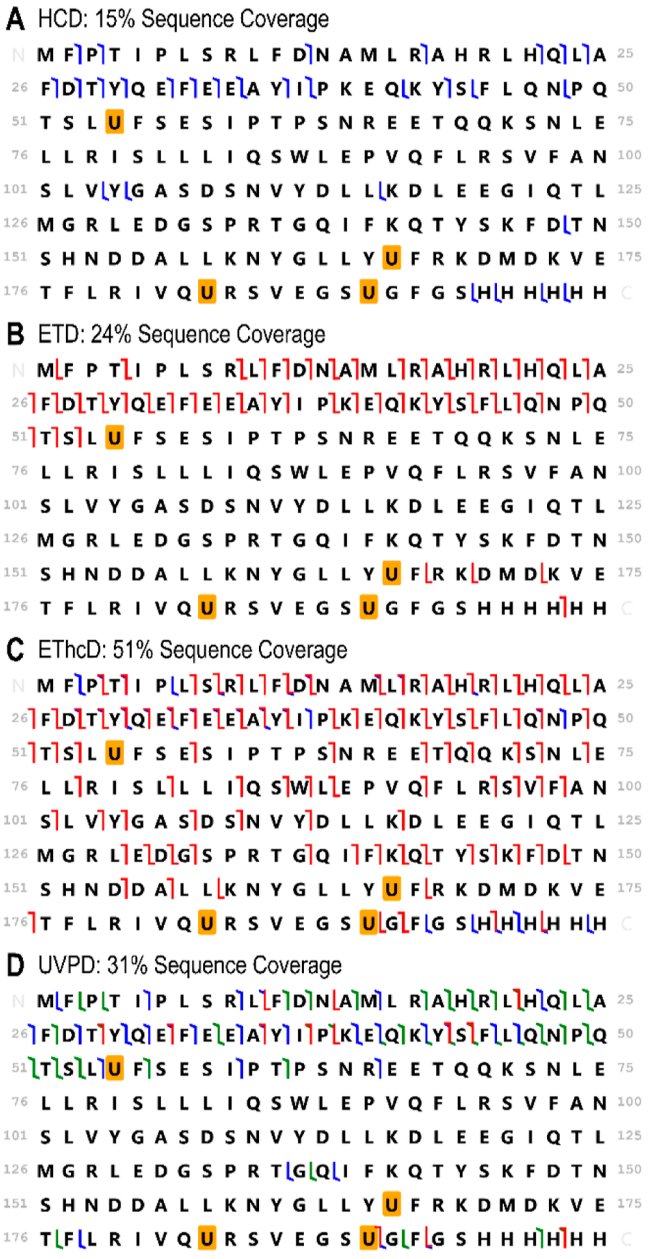


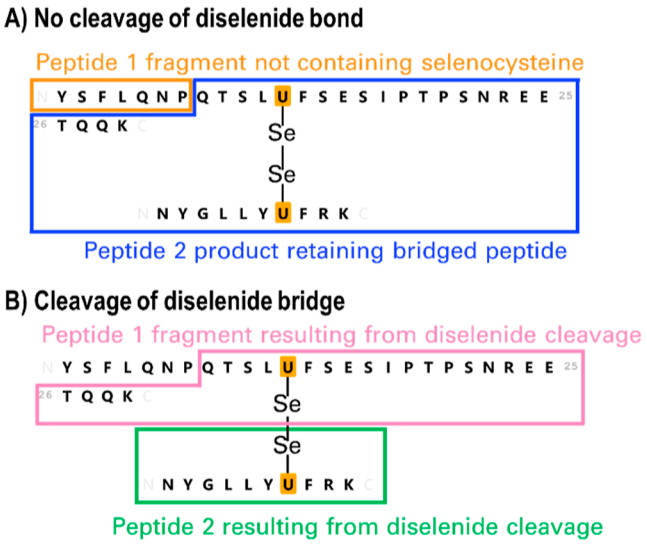
Figure 3. Sequence coverage maps for HGH (16+) obtained using (A) HCD (15% NCE), (B) ETD (5 ms activation time), (C) EThcD (5 ms activation time and 15% NCE supplemental activation), and (D) 213 nm UVPD (50 ms activation period).

residues each and the U54−U166 linkage and the other containing 28 residues and the U183−U190 linkage, highlighted in Scheme S1. Both peptides were identified in the LC−MS data and were selected to evaluate MS/MS parameters. Standard parameters of 30 and 15% NCE were used for HCD and EThcD, respectively, and calibrated charge-dependent activation periods were employed for ETD and EThcD. For 213 nm UVPD, ideal parameters were selected based on mass resolution, number of microscans, and activation period (e.g., number of laser pulses) (Figure S7).

For MS/MS analysis of diselenide-linked peptides produced after digestion of the proteins, there are a variety of possible fragment ions that can be used for sequencing. Each possibility is illustrated in Scheme 2. In theory, peptide fragment ions can be produced without cleavage of the diselenide bond (Scheme

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184

Scheme 2. Illustration of Possible Fragments Resulting from MS/MS of a Pair of Peptides (1 and 2) Bridged by a Diselenide Bond[a]



[a](A) One possibility is cleavage of a peptide bond without cleavage of the diselenide bonds, resulting in a terminal fragment ion from peptide 1 and not containing selenocysteine, outlined in orange, and a fragment retaining all of peptide 2 and part of peptide 1, outlined in blue. (B) Another possibility is cleavage across the diselenide bond, which can occur either in addition to cleavage of the peptide backbone, exemplified by the product outlined in pink, or without cleavage of the peptide backbone, resulting in the product outlined in green.

2A) or with cleavage of the diselenide bridge (Scheme 2B). In the case of diselenide cleavage, the resulting products can either undergo additional backbone cleavage to result in a sequence ion (pink fragment in Scheme 2B) or release a partner peptide without backbone fragmentation (green fragment in Scheme 2B). When cleavage of the diselenide bridge occurs, it may entail cleavage of the Se−Se bond (homolytic) or may occur at either of the C−Se bonds in an asymmetric manner that results in products containing two selenium atoms or none. Very few peptide fragment ions were identified that resulted from asymmetric cleavage, and it was therefore not considered further in the identification of sequence ions.

As illustrated in the MS/MS spectra shown in Figures S8 and 4, HCD, EThcD, and 213 nm UVPD all resulted in high quality characterization of the U54−U166 diselenide-bridged peptide pair from the HGH protein digest. While all activation methods have merits, significant differences were observed in the resulting MS/MS spectra. For HCD (Figure S8A) high quality characterization of the sequence was achieved; however, limited production of complementary bi-directional *b*/*y* ion pairs and the lack of cleavage across the diselenide bond resulted in lower confidence compared to the other methods. The performance of ETD for the same diselenide-bridged peptide was less impressive (Figure S8B). ETD has been reported to result in peptide partners after cleavage across disulfide bonds, confirming the presence of the disulfide bridges between peptides.[21,22,26,27] This unique cleavage was observed in low abundance upon ETD of the selenocysteine-bridged peptide in the 5+ charge state (Figure S8B), and the production of other fragment ions was sparse. Cleavage of the diselenide bond was observed in much greater abundance for the same peptide in the higher 6+ charge state (Figure S9B), and a larger array of sequence ions was produced. While the charge dependence of ETD could be mitigated by analyzing multiple charge states of each peptide in a run, the same limitation was not observed for EThcD (Figures S8C and S9C) or 213 nm UVPD (Figures 4 and S9D). EThcD promoted homolytic cleavage of the diselenide bond, resulting in the two complementary peptide partner products as the highest abundance fragment ions, regardless of charge state of the precursor ion. EThcD also resulted in improved sequence coverage compared to ETD, although HCD still surpassed EThcD in total sequence coverage. Only 213 nm UVPD resulted in both the generation of the peptide partner products from homolytic diselenide bond cleavage and complete sequence coverage for both peptide partners (Figure 4). In addition to homolytic Se−Se cleavage, some peptide products originated from alternative C−Se cleavage upon 213 nm UVPD, resulting in fragment ions that contained two selenium atoms (or none). Although the observation of these C−Se cleavage products was interesting, they were identified only in the peptide partners, not in the sequence ions, and therefore did not contribute to the peptide characterization. The high sequence coverage obtained by 213 nm UVPD was enabled by the generation of a significant number of high abundance fragment ions arising from cleavage of the diselenide bond (Figure 4). While the overall best characterization and

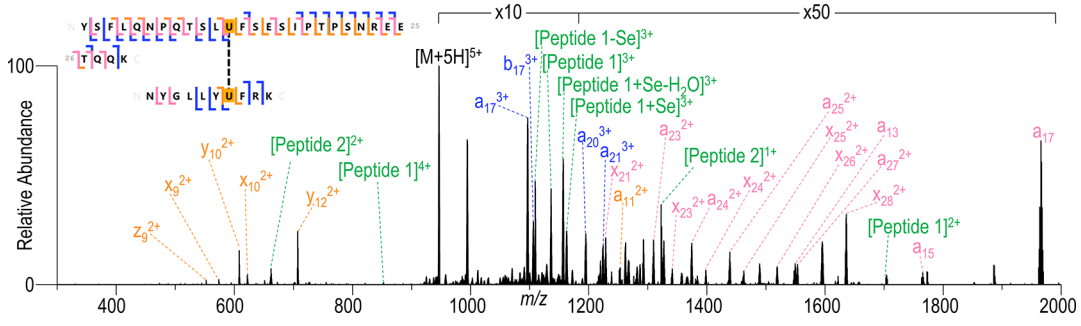


Figure 4. MS/MS spectrum and sequence coverage map for one selenocysteine-bridged peptide pair from HGH (5+) obtained using 213 nm UVPD (100 ms activation period). All spectra were collected with 120k resolution and 3 μscans, and 8 spectra were averaged. Only the most abundant fragment ions in each region are labeled. Fragment ions related to free peptides after cleavage of diselenide bonds are marked as peptide 1 (larger peptide) and peptide 2 (smaller peptide). Any fragment ion labeled with a star originated from peptide 2. The fragment ion coloring is defined in Scheme 2.

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184

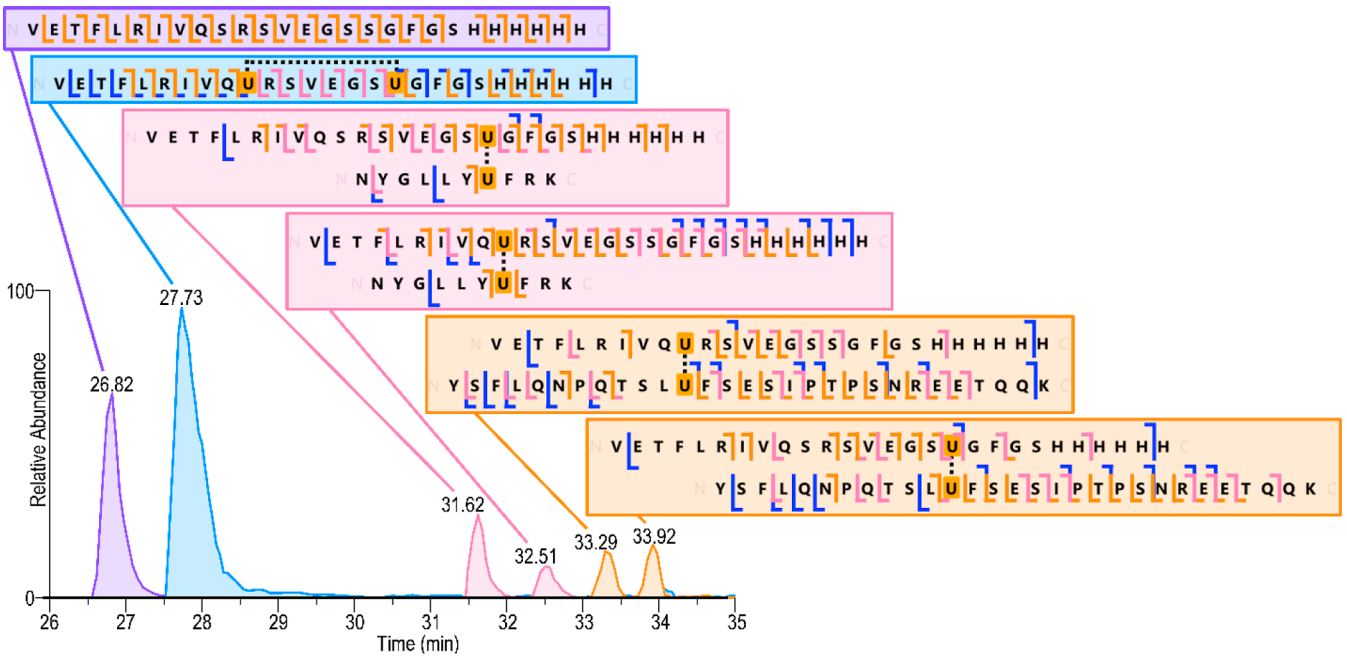


Figure 5. Extracted ion chromatograms for all identified peptides containing U183 and U190 with no missed cleavages from the HGH digest. The most abundant charge state from the MS1 spectrum acquired for each chromatographic peak is extracted for each peptide, and a sequence coverage map for each peptide is included based on the results from 213 nm UVPD. The peptides highlighted in pink and orange are pairs of isomers that elute separately and share extracted ion chromatograms.

sequence coverage was afforded by 213 nm UVPD, EThcD yielded the highest abundance fragment ions. The peptide partners from homolytic diselenide cleavage are critical indicators for unambiguous confirmation of the presence of diselenide bridges in tryptic peptides and are uniquely observed in high abundance with EThcD. The combination of EThcD and 213 nm UVPD is thus ideal for the identification and characterization of these bridged peptides.

The second predicted diselenide-bridged peptide from HGH, with an intrapeptide U183−U190 diselenide bridge, was also analyzed in the 5+ charge state (Figure S10). For this type of species, fragment ions originating from backbone cleavages between the two selenocysteine residues are observed only if the constraining diselenide bond is also cleaved, allowing separation of the two peptide partners. These homolytic cleavage fragment ions are crucial for obtaining complete characterization of the constrained peptide. Sequence ions resulting from concurrent homolytic cleavage and backbone cleavage were observed for ETD, EThcD, and 213 nm UVPD (but not HCD), and this second diselenide-bridged peptide is well-characterized as illustrated in the maps in Figure S10.

Characterization of Diselenide Bridge Heterogeneity. Assessing the best way to decipher diselenide-bridged peptides utilizing the predicted digestion products also allows more facile analysis of peptides containing unanticipated diselenide bridges. During the proteolytic digestion process used for bottom-up analysis, the possibility of diselenide scrambling must be considered because disulfide scrambling has been reported to occur during digestion.[2,46] The replacement of sulfur with diselenide for cysteine bridges, however, has repeatedly been shown to lead to improved stability.[5,47,48] To mitigate diselenide scrambling, the digestion was performed at pH 6, which has previously been shown to limit disulfide scrambling.[2,46] After establishing the conditions to limit scrambling, the identification of unanticipated diselenide bonds is possible.

In addition to the successful characterization of the correctly linked peptides described in the previous section, a significant number of peptides containing serine in the place of selenocysteine and/or containing an incorrect diselenide bond were identified in the HGH digest based on the database search. To confirm the diselenide linkages, EThcD and 213 nm UVPD mass spectra were examined for each match. For a match to be validated for a pair of peptides with interpeptide diselenide bonds, the EThcD spectra must display high abundance peptide partner ions resulting from homolytic cleavage of diselenide bonds. For a peptide containing an intrapeptide diselenide bond, the highest abundance ions in the EThcD spectra were expected to be consistent with charge reduction products (no change in mass). Because 213 nm UVPD tended to result in the greatest number of fragment ions and the highest sequence coverage, 213 nm UVPD mass spectra were also examined for each peptide match to confirm sequence identities. All confirmed peptides from selenocysteine-containing HGH are listed in Table S1. Most of the incorrectly paired diselenide bridges are clearly the result of the mis-incorporation of serine in the place of selenocysteine, as showcased in the extracted ion chromatograms (EIC) in Figure 5 generated for all HGH peptides containing U183 and U190, which are expected to participate in an intrapeptide diselenide bridge. The highest abundance peptide, eluting at 27.73 min (highlighted in blue), corresponds to the predicted sequence and correct diselenide bridge. The next highest abundance peptide eluting at 26.82 min (highlighted in purple) originates from the incorporation of two serine residues in the place of each selenocysteine and thus has no diselenide bond. The presence of a peptide containing two serine residues is not surprising given the identification of a protein containing two serine replacements in the top-down

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184

MS1 spectrum for HGH (Figure 1C). The remaining four peaks (highlighted in pink and orange) in the EIC trace in Figure 5 result from the replacement of either U183 or U190 with serine, allowing the other selenocysteine to pair with a different selenocysteine residue within the protein. No peptides were identified that indicated participation of U183 and U190 in an incorrect pairing unless either U183 or U190 was replaced with serine. These results convey that incorrect selenocysteine pairing was due to the unintended incorporation of serine.

In addition to the diselenide bridges illustrated in Figure 5, which could easily be explained by the misincorporation of serine within the observed peptides, there were a few other peptides identified that were consistent with more surprising diselenide bridges. Namely, the peptides containing U166−U166 and U54−U54 diselenide bridges listed in Table S1 could only have resulted from the pairing of two separate HGH proteins. This was unexpected because an intact HGH dimer was not observed in the high-resolution top-down MS1 spectrum (Figure 1C). While one explanation is that these linkages resulted from diselenide scrambling during digestion, the low pH utilized during digestion should have limited that possibility. An alternative explanation is that the dimer is present in the sample but was simply not detected in the high-resolution mass spectrum. In an Orbitrap analyzer, a higher mass ion, such as a dimer, may experience accelerated signal decay relative to a higher abundance monomer, a phenomenon resulting from more frequent gas collisions for the species with the larger size (i.e., greater collision cross-section).[49,50] This accelerated signal decay of the higher mass dimer may result in an artificially low signal abundance. This signal decay discrepancy is substantiated by collection of MS1 spectra using lower resolution by curtailing the time that the transient is acquired and truncating the signal decay. The lower resolution (R 15,000 at m/z 200) spectrum displayed in Figure S11A reveals the presence of an HGH dimer not observed in the higher resolution (R 120,000 at m/z 200) spectrum in Figure S11B. While the low-resolution data could not be deconvoluted with traditional software, deconvolution with Unidec was applied to reveal additional information about the dimer (Figure S12). The monomer population in the low-resolution mass spectrum was deconvoluted as a single species corresponding to the complete incorporation of four selenocysteine residues (4xU). The dimer population, however, comprised several species consistent with the incorporation of either two serine residues (6xU 2xS, nominally 46,700 Da) or four serine residues (4xU 4xS, nominally 46,570 Da) in place of selenocysteine. The fact that no HGH dimers were identified containing complete incorporation of selenocysteine (8xU 0xS) supports the idea that dimer formation is uniquely caused by the incorporation of at least one serine residue per protein. HGH dimers can occur naturally as a consequence of a genetic mutation resulting in the incorporation of serine in the place of C54, leading to bridging of unpaired C166 residues of two HGH molecules.[51] Given the greater electronegativity of selenium relative to sulfur,[8] it is unsurprising that a similar phenomenon would be observed in the selenocysteine system. It can be reasoned that the HGH dimer containing two serine residues resulted from a partnership of a pair of HGH monomers, each with one serine replacement, leaving one free selenocysteine of each protein to form a dimer. Similarly, the dimer with four serine misincorporations likely formed from interaction of a protein containing one serine incorporation with a protein containing three serine incorporations, resulting in one free selenocysteine on each HGH molecule. The production of dimers owing to free (unpaired) selenocysteines also explains a phenomenon first observed in the high-resolution mass spectra in Figure 1B,C. The monomers were always observed either with complete selenocysteine incorporation, as expected, or with the misincorporation of two serine residues (low abundance species), but never with an odd number of serine residues. The spontaneous generation of dimers whenever a free selenocysteine residue is present would eliminate any monomeric proteins containing uneven numbers of selenocysteine residues.

Given evidence of dimer formation for HGH, the MS1 spectra of trastuzumab scFv and GFP proteins were reacquired at lower resolution. Unlike HGH, dimers were not observed in the lower resolution data for GFP or trastuzumab scFv Figure S13). To further explore the possible incorporation of serine residues, the LysC digest of GFP was examined in greater detail (Figure S14 and Table S2), revealing the presence of additional peptides consistent with the misincorporation of Ser (Figure S15). In addition to the correctly bridged pair of peptides, peptides resulting from the misincorporation of serine at both U134 and U176 were observed (retention times 23.88, 17.22, and 24.41 min). Additionally, peptides resulting from the pairing of U134−U134 and U176−U176, which could be explained only by the presence of a dimer, were identified (retention times 16.35 and 27.60 min). The fact that peptides originating from the GFP dimers were identified in the tryptic digest, but intact dimers could not be observed in the top-down MS1 spectra, suggests that detection of these low abundance intact mixed S/U proteins was suppressed in top-down analysis and supports the value of the bottom-up approach for identification of low abundance side-products.

CONCLUSIONS

Given the substantial potential for selenoproteins to aid in the development of biotherapeutics, it is critical to establish adequate characterization techniques. The combination of top-down and bottom-up mass spectrometry, presented herein, led to the complete characterization of customized selenoprotein constructs. While the bottom-up approach is critical for the identification of peptide pairings not identified with top-down analysis, the context of combinatorial modifications is lost. Additionally, when correctly paired peptides are identified, it is not possible to confirm that they did not originate from a dimer owing to the contextual information lost in the digestion process. With the top-down approach, it is possible to individually isolate the proteins with complete selenocysteine incorporation, allowing the corroboration of diselenide bridges in the proteins without serine misincorporation. The link between incorrect diselenide bridge formation and serine misincorporation suggests that improving the enzymatic incorporation of selenocysteine would eliminate the generation of incorrectly bridged proteins. In addition, it was gratifying to not observe any apparent mis-folded proteins with a full complement of selenocysteine residues. These results indicate that many complex selenoproteins can potentially be made by simple replacement of cysteine with selenocysteine. Furthermore, the presence of the dimeric species upon incorporation of a single serine residue presents the opportunity to purify selenoproteins with size-based separation techniques.

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184

The combination of EThcD and 213 nm UVPD proved beneficial for comprehensive top-down characterization. EThcD cleaved diselenide bonds and yielded the highest sequence coverage for diselenide-bridged proteins. In a complementary manner, 213 nm UVPD resulted in many sequence ions without cleavage of diselenide bonds, allowing localization of diselenide bridges. In the bottom-up strategy, a combination of 213 nm UVPD and EThcD also offered the highest quality characterization of diselenide-bridged peptides given the unique ions generated by EThcD and the high sequence coverage achieved with 213 nm UVPD. Using these methods, the identification of both monomers with incomplete diselenide incorporation and dimers resulting from unpaired selenocysteine residues also sheds light on factors mediating the production of synthetic selenoproteins that could have key implications on protein function. The integrated approach presented here offers a compelling strategy to advance the development of selenoproteins for biotherapeutic applications.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c01433.

Optimization of MS/MS conditions, top-down MS/MS spectra for all proteins, additional bottom-up MS/MS spectra, chromatographic traces, low resolution mass spectra of proteins, schemes showing diselenide-bridge patterns, tables of peptide identifications from bottom-up analysis, and a master table listing all fragment ions matched from bottom-up data (PDF)

(XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Jennifer S. Brodbelt − *Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0003-3207-0217; Email: jbrodbelt@cm.utexas.edu

Authors

Eleanor Watts − *Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States*

Ross Thyer − *Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, United States*

Andrew D. Ellington − *Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0001-6246-5338

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c01433

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from NSF (CHE-2203602), the Robert A. Welch Foundation (F-1155), and the UT System for support of the UT System Proteomics Core Facility Network.

■ REFERENCES

(1) Feige, M. J.; Braakman, I.; Hendershot, L. M. CHAPTER 1.1. Disulfide Bonds in Protein Folding and Stability. *Oxidative Folding of Proteins*; Royal Society of Chemistry, 2018; pp 1−33.

(2) Lakbub, J. C.; Shipman, J. T.; Desaire, H. *Anal. Bioanal. Chem.* **2018**, *410*, 2467−2484.

(3) Campuzano, I. D. G.; Sandoval, W. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1861−1885.

(4) Weinfurtner, D. CHAPTER 1.4. Analysis of Disulfide Bond Formation in Therapeutic Proteins. *Oxidative Folding of Proteins*; Royal Society of Chemistry; 2018; pp 81−98.

(5) Thyer, R.; Shroff, R.; Klein, D. R.; d'Oelsnitz, S.; Cotham, V. C.; Byrom, M.; Brodbelt, J. S.; Ellington, A. D. *Nat. Biotechnol.* **2018**, *36*, 624−631.

(6) Thyer, R.; Robotham, S. A.; Brodbelt, J. S.; Ellington, A. D. *J. Am. Chem. Soc.* **2015**, *137*, 46−49.

(7) Mousa, R.; Notis Dardashti, R.; Metanis, N. *Angew. Chem., Int. Ed.* **2017**, *56*, 15818−15827.

(8) Arnér, E. S. J. *Exp. Cell Res.* **2010**, *316*, 1296−1303.

(9) Cardoso, B. R.; Ganio, K.; Roberts, B. R. *Metallomics* **2019**, *11*, 1974−1983.

(10) Thyer, R.; Filipovska, A.; Rackham, O. *J. Am. Chem. Soc.* **2013**, *135*, 2−5.

(11) Srzentić, K.; Nagornov, K. O.; Fornelli, L.; Lobas, A. A.; Ayoub, D.; Kozhinov, A. N.; Gasilova, N.; Menin, L.; Beck, A.; Gorshkov, M. V.; Aizikov, K.; Tsybin, Y. O. *Anal. Chem.* **2018**, *90*, 12527−12535.

(12) Shaw, J. B.; Liu, W.; Vasil'ev, Y. V.; Bracken, C. C.; Malhan, N.; Guthals, A.; Beckman, J. S.; Voinov, V. G. *Anal. Chem.* **2020**, *92*, 766−773.

(13) Li, G.; Pei, J.; Yin, Y.; Huang, G. *Analyst* **2015**, *140*, 2623−2627.

(14) Adhikari, S.; Yang, X.; Xia, Y. *Anal. Chem.* **2018**, *90*, 13036−13043.

(15) Cramer, C. N.; Haselmann, K. F.; Olsen, J. V.; Nielsen, P. K. *Anal. Chem.* **2016**, *88*, 1585−1592.

(16) Chin, S.; Chen, T.; Hannoush, R. N.; Crittenden, C. M. *J. Pharm. Biomed. Anal.* **2021**, *195*, 113893.

(17) Schnaible, V.; Wefing, S.; Resemann, A.; Suckau, D.; Bücker, A.; Wolf-Kümmeth, S.; Hoffmann, D. *Anal. Chem.* **2002**, *74*, 4980−4988.

(18) Massonnet, P.; Haler, J. R. N.; Upert, G.; Smargiasso, N.; Mourier, G.; Gilles, N.; Quinton, L.; De Pauw, E. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1995−2002.

(19) Huwiler, K. G.; Mosher, D. F.; Vestling, M. M. *J. Biomol. Tech.* **2003**, *14*, 289−297.

(20) van der Burgt, Y. E. M.; Kilgour, D. P. A.; Tsybin, Y. O.; Srzentić, K.; Fornelli, L.; Beck, A.; Wuhrer, M.; Nicolardi, S. *Anal. Chem.* **2019**, *91*, 2079−2085.

(21) Riley, N. M.; Coon, J. J. *Anal. Chem.* **2018**, *90*, 40−64.

(22) Rombouts, I.; Lagrain, B.; Scherf, K. A.; Lambrecht, M. A.; Koehler, P.; Delcour, J. A. *Sci. Rep.* **2015**, *5*, 12210.

(23) Bishop, A.; Brodbelt, J. S. *Int. J. Mass Spectrom.* **2015**, *378*, 127−133.

(24) Liu, F.; van Breukelen, B.; Heck, A. J. R. *Mol. Cell. Proteomics* **2014**, *13*, 2776−2786.

(25) Cole, S. R.; Ma, X.; Zhang, X.; Xia, Y. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 310−320.

(26) Clark, D. F.; Go, E. P.; Desaire, H. *Anal. Chem.* **2013**, *85*, 1192−1199.

(27) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 1811−1822.

(28) Zhang, L.; English, A. M.; Bai, D. L.; Ugrin, S. A.; Shabanowitz, J.; Ross, M. M.; Hunt, D. F.; Wang, W.-H. *Mol. Cell. Proteomics* **2016**, *15*, 1479−1488.

(29) Quick, M. M.; Crittenden, C. M.; Rosenberg, J. A.; Brodbelt, J. S. *Anal. Chem.* **2018**, *90*, 8523−8530.

(30) Agarwal, A.; Diedrich, J. K.; Julian, R. R. *Anal. Chem.* **2011**, *83*, 6455−6458.

(31) Wongkongkathep, P.; Li, H.; Zhang, X.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A. *Int. J. Mass Spectrom.* **2015**, *390*, 137−145.
(32) Siuti, N.; Kelleher, N. L. *Nat. Methods* **2007**, *4*, 817−821.
(33) Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y. *Anal. Chem.* **2018**, *90*, 110−127.
(34) Ganisl, B.; Breuker, K. *ChemistryOpen* **2012**, *1*, 260−268.
(35) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2013**, *85*, 4239−4246.
(36) Fornelli, L.; Parra, J.; Hartmer, R.; Stoermer, C.; Lubeck, M.; Tsybin, Y. O. *Anal. Bioanal. Chem.* **2013**, *405*, 8505−8514.
(37) Zhang, J.; Loo, R. R. O.; Loo, J. A. *Int. J. Mass Spectrom.* **2015**, *377*, 546−556.
(38) Greisch, J.-F.; den Boer, M. A.; Schuurman, F.; Tamara, J.; Bondt, S.; Heck, A.; Heck, A. J. R. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1326−1335.
(39) Lodge, J. M.; Schauer, K. L.; Brademan, D. R.; Riley, N. M.; Shishkova, E.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2020**, *92*, 10246−10251.
(40) Rush, M. J. P.; Riley, N. M.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2018**, *90*, 8946−8953.
(41) Gammelgaard, S. K.; Petersen, S. B.; Haselmann, K. F.; Nielsen, P. K. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1910−1918.
(42) Lamarche, J.; Ronga, L.; Szpunar, J.; Lobinski, R. *Int. J. Mol. Sci.* **2021**, *22*, 6283.
(43) Ma, S.; Hill, K. E.; Burk, R. F.; Caprioli, R. M. *Biochemistry* **2003**, *42*, 9703−9711.
(44) Ma, S.; Hill, K. E.; Burk, R. F.; Caprioli, R. M. *J. Mass Spectrom.* **2005**, *40*, 400−404.
(45) Wingfield, P. *Curr. Protoc. Protein Sci.* **2017**, *88*, 6.14.1.
(46) Sung, W.-C.; Chang, C.-W.; Huang, S.-Y.; Wei, T.-Y.; Huang, Y.-L.; Lin, Y.-H.; Chen, H.-M.; Chen, S.-F. *Biochim. Biophys. Acta, Proteins Proteomics* **2016**, *1864*, 1188−1194.
(47) Mousa, R.; Hidmi, T.; Pomyalov, S.; Lansky, S.; Khouri, L.; Shalev, D. E.; Shoham, G.; Metanis, N. *Commun. Chem.* **2021**, *4*, 30.
(48) Shimodaira, S.; Asano, Y.; Arai, K.; Iwaoka, M. *Biochemistry* **2017**, *56*, 5644−5653.
(49) Makarov, A.; Denisov, E. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1486−1495.
(50) Sanders, J. D.; Grinfeld, D.; Aizikov, K.; Makarov, A.; Holden, D. D.; Brodbelt, J. S. *Anal. Chem.* **2018**, *90*, 5896−5902.
(51) Sander, M.; Wu, Z.; Strasburger, C. J. *J. Biol. Chem.* **2020**, *295*, 4893−4901.

https://doi.org/10.1021/acs.analchem.2c01433
Anal. Chem. 2022, 94, 11175−11184