Journal of proteome research

pubs.acs.org/jpr

Article

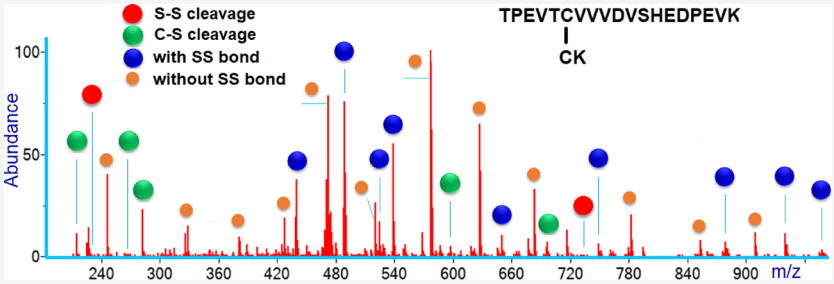
Comprehensive Analysis of Tryptic Peptides Arising from Disulfide Linkages in NISTmAb and Their Use for Developing a Mass Spectral Library

Qian Dong,* Xinjian Yan, Yuxue Liang, Sanford P. Markey, Sergey L. Sheetlin, Concepcion A. Remoroza, William E. Wallace, and Stephen E. Stein

Cite This: *J. Proteome Res.* 2021, 20, 1612−1629

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



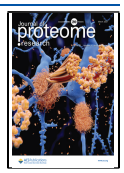
ABSTRACT: This work presents methods for identifying and then creating a mass spectral library for disulfide-linked peptides originating from the NISTmAb, a reference material of the humanized IgG1k monoclonal antibody (RM 8671). Analyses involved both partially reduced and non-reduced samples under neutral and weakly basic conditions followed by nanoflow liquid chromatography tandem mass spectrometry (LC−MS/MS). Spectra of peptides containing disulfide bonds are identified by both MS1 ion and MS2 fragment ion data in order to completely map all the disulfide linkages in the NISTmAb. This led to the detection of 383 distinct disulfide-linked peptide ions, arising from fully tryptic cleavage, missed cleavage, irregular cleavage, complex Met/Trp oxidation mixtures, and metal adducts. Fragmentation features of disulfide bonds under low-energy collision dissociation were examined. These include (1) peptide bond cleavage leaving disulfide bonds intact; (2) disulfide bond cleavage, often leading to extensive fragmentation; and (3) double cleavage products resulting from breakages of two peptide bonds or both peptide and disulfide bonds. Automated annotation of various complex MS/MS fragments enabled the identification of disulfide-linked peptides with high confidence. Peptides containing each of the nine native disulfide bonds were identified along with 86 additional disulfide linkages arising from disulfide bond shuffling. The presence of shuffled disulfides was nearly completely abrogated by refining digest conditions. A curated spectral library of 702 disulfide-linked peptide spectra was created from this analysis and is publicly available for free download. Since all IgG1 antibodies have the same constant regions, the resulting library can be used as a tool for facile identification of "hard-to-find" disulfide-bonded peptides. Moreover, we show that one may identify such peptides originating from IgG1 proteins in human serum, thereby serving as a means of monitoring the completeness of protein reduction in proteomics studies. Data are available via ProteomeXchange with identifier PXD023358.

KEYWORDS: *cysteine, disulfide, disulfide-linked peptides, NISTmAb, IgG1, MS spectral library, full MS scan, disulfide bond fragmentation*

■ INTRODUCTION

Disulfide bridges between cysteines (Cys) stabilize the three-dimensional structure of proteins, aiding with protein folding and ultimately enabling protein functionality.[1,2] Dysregulated structural disulfide bonds are characteristic of certain human diseases, such as various neurodegenerative diseases and cancers.[3] Therefore, detailed information on disulfide connectivity is critical for protein structure−function studies and biotherapeutics structure integrity assessment. Typical proteomics experiments dissociate disulfide bonds by reduction as part of the denaturing process and then cap Cys residues by alkylation to prevent the disulfide bond from re-forming. These well-established protein digestion methods can fully sequence individual proteins and identify a large fraction of low-abundance modifications in protein digests. Recently, two

Received: October 16, 2020
Published: February 8, 2021



Downloaded via 65.21.193.86 on February 24, 2022 at 17:09:38 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications

Not subject to U.S. Copyright. Published 2021 by American Chemical Society

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

detailed mass spectral library-based examinations of the NISTmAb, a reference material of the humanized IgG1κ monoclonal antibody (RM 8671),[4] and its glycosylation were reported.[5,6] In contrast, MS analysis of disulfide bonds requires a different analytical approach. Recent advances in mass spectrometry have enabled the development of customized LC−MS strategies for disulfide bond mapping in proteins.[7−9] A major requirement in sample preparation is to prevent non-native disulfide formation (i.e., disulfide bond artifacts). The disulfide shuffling occurs spontaneously in the presence of free cysteines at pH 7.5−8.5, which is typical for protein digestion in standard proteomics protocols.[10,11] These artifacts can be minimized by blocking free cysteines and using acidic to neutral pH conditions. Methods for determining disulfide patterns include liquid chromatographic ultraviolet profile comparison between reduced and non-reduced peptides, direct mass spectrometry detection of disulfide without reduction or partial reduction, chemical labeling identification, and tandem mass spectrometry sequencing of non-reduced peptides.[12−17]

Significant challenges remain in disulfide analyses due to complex intertwined disulfides, disulfide shuffling, incomplete digestion of non-reduced proteins, large peptide masses, and low ionization efficiency of disulfide-linked peptides.[18−24] Tandem mass spectra of peptides containing disulfide bonds are generally very complex because of various collisional dissociation products arising from cleavage of the component peptide backbone as well as disulfide bonds. There are no diagnostic fragments characteristic of disulfide spectra unlike other post-translational modifications that are readily detected by predominant ions resulting from residue side chain cleavage (e.g., oxonium ions of glycans, phosphoric acid (H_3PO_4) neutral loss of O-phosphorylation). Thus, most available bioinformatics tools currently are incapable of recognizing various disulfide-linked ions and interpreting their product ion spectra.

The present study is a continuation of our previous work,[5,6] in which two comprehensive spectral libraries were developed for characterizing peptides, glycopeptides, and other biologically modified peptides derived from the NISTmAb. These libraries contain a large number of low-abundance modifications occurring on a wide range of residues. Notably, many modifications are not reported by most commonly used peptide identification programs since modified peptides do not fragment primarily by peptide backbone cleavages. Instead, their side chain cleavage reactions can dominate their product ion spectra. This study extends the prior work to enable the identification of another class of hard-to-find peptide products of digestion, namely, disulfide-bonded peptides. To identify and confirm the identity of these complex peptides, datasets generated by two digestion protocols, non-reduction and partial reduction, were chosen from a large amount of data obtained by a series of common protocols used in creating the above spectral libraries.[5] The data analysis method developed in this work employs full scan MS spectra to identify potential precursor ions and validates them by a detailed fragmentation analysis of their tandem mass spectra (MS/MS spectra). Resulting spectra then serve to create a searchable mass spectral library. As a bonus, we find and demonstrate the utility of these spectra for identifying digested, but incompletely reduced IgG1-derived peptides in human serum, thereby providing a unique measure of reduction effectiveness.

■ MATERIALS AND METHODS

Materials

The primary sample 8670 NISTmAb (PS 8670, an in-house primary standard material), lot 3f1b, is an IgG1k mAb derived from a separate production lot of NISTmAb 8671 expressed in murine suspension culture.[4] It was obtained from the Bioanalytical Science Group at NIST. Digestion reagents guanidine hydrochloride, dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), *N*-ethylmaleimide (NEM), and iodoacetamide (IAM) were purchased from Sigma-Aldrich (St. Louis, MO). A revised Tris buffer (pH = 7) was prepared via adjustment of tris base solution (200 mmol) by HCl solution (1 mol/L). Sequencing-grade trypsin was purchased from Promega (Madison, WI). RapiGest, brand-name for sodium 3-[(2methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate, was purchased from Waters (Milford, MA); Zeba spin columns (7 K molecular weight cutoff (MWCO)) were purchased from Thermo Fisher Scientific (Waltham, MA). Chromatographic separations were performed on an Acclaim pepmap100 nano column (150 mm × 75 μm, C18, 3 μm particle size, 100 Å pore size, Dionex, Sunnyvale, CA).

Tryptic Digestion

To generate a wide variety of tryptic peptides that were collected in the NISTmAb peptide spectral library, parameters for a series of digestion protocol variations were used during denaturing, reduction, and alkylation of the NISTmAb. The details are described in our previous paper.[5] Among them, the two protocols of the non-reduced and partially reduced samples prepared under these general experimental conditions (pH = 8) were used in this study, therefore, disulfide bond rearrangements could occur. To prevent these shuffling reactions, two alternative methods are used for sample preparation at pH 7. All methods used in analyzing disulfides are described below.

General Experimental Conditions (pH = 8). NISTmAb (500 μg) was denatured at room temperature for 10 min using guanidine hydrochloride (6 M) in 50 μL of 100 mmol/L Tris HCI (Tris(hydroxymethyl)aminomethane hydrochloride) buffer (pH = 8). Reduction and alkylation were varied as follows: (a) the sample was neither reduced nor alkylated or (b) the sample was partially reduced by adding 5 μL of 200 mmol/L TCEP to the above denatured mixture and incubating at room temperature for 1 h, followed by alkylation with 10 μL of 200 mmol/L IAM at room temperature in the dark for 1 h. The alkylated protein solution was desalted with a Zeba spin column (7 K MWCO) prior to adding 10 μg of Promega trypsin. The resulting mixture was subjected to digestion with sequencing-grade trypsin at 37 °C for 0.25, 2, and 18 h.

Non-reduction and Alkylation with NEM (pH = 7). 500 μg of NISTmAb was denatured at room temperature for 10 min using 6 M guanidine in 50 μL of tris buffer (pH = 7). The alkylation reagent NEM was used because of its thiol specificity and stability at pH 7, contrasting to IAM with optimal reaction conditions at pH 8, which most likely results in disulfide shuffling.[8] 10 μL of 200 mmol/L NEM was added to the above denatured mixture and incubating at room temperature in the dark for 1 h. The alkylated protein solution was desalted as described above. Trypsin (10 μg) was added to digest the solution at 37 °C for 0.25, 2, and 18 h. The reaction was quenched with 5 μL of formic acid (50% v/v).

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

Table 1. Classification of the Three Major Groups of Product Ions Observed in the Spectra of Disulfide Peptide Discussed in This Section, SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPVTK[a]

group #	presence of disulfide	bond broken	number of cleavages	subgroup	example fragment	charge	assignment example
1	intact disulfide bond	amide bond	1	a	|VCLLNNFYPR | VYACEVTHQGLSSPVTK	3+	y_{10}
			2	b	|VCLL| | VYACEVTHQGLSSPVTK	2+	intVCLL
				c	|VCLLNNFYPR | VYACE|	2+	$[y_{10}:b(2)_5]$
2	disulfide bond cleavage	C-S	1	a1	VYAC(-H_2S)EVTHQGLSSPVTK	2+	$y(2)_{17}$-H_2S
		C-S & amide bond	2	a2	VYAC(+S)EVTHQGLSS|	2+	$b(2)_{13}$+S
		S-S	1	b1	VYAC(-2H)EVTHQGLSSPVTK	2+	$y(2)_{17}$-2H
		S-S & amide bond	2	b2	|AC(+0)EVTHQGLSSPVTK	2+	$y(2)_{15}$+0
3	no disulfide bond	amide bond	1	a	|SSPVTK	1+	$y(2)_6$
			2	b	|HQGLSSP|	1+	Int(2)8_14

[a]Symbols |, -H_2S, and +S, -$2H$ and +0 denote cleavage of amide, C−S, and S−S bonds, respectively.

Reduction with TCEP and Alkylation with NEM (pH = 7). 500 μg of NISTmAb was denatured using 6 M of guanidine; reduction was performed by adding 5 μL of 200 mmol/L TCEP to the above denatured mixture and incubating at room temperature for 1 h, followed by alkylation with 20 μL of 200 mmol/L NEM at room temperature in the dark for 1 h. Then, the alkylated protein solution was desalted as described above. Trypsin (10 μg) was added to digest the protein solution at 37 °C for 0.25, 2, and 18 h. The reaction was quenched with 5 μL of formic acid (50% v/v).

LC−MS/MS Analysis

The above digests (0.2 μg) were analyzed on a Dionex Ultimate 3000 RSLC Nano LC with an Acclaim pepmap100 column with a nanospray source connected to one of two mass spectrometers: a Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer or an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Waltham, MA) in the positive ion mode. Mobile phase A consisted of 0.1% formic acid in water, and mobile phase B consisted of 0.1% formic acid in ACN. The peptides were eluted by increasing mobile phase B from 1% to 90% over 120 min. Data was collected using a data-dependent mode with a dynamic exclusion of 20 s. The top 10 most abundant precursor ions were selected from a 250 m/z to 1850 m/z full scan for HCD and ion trap fragmentation (FT-CID) with normalized collisional energy (NCE) parameters (NCE of 16, 20, 24, 32, and 36). The resolution of full MS scan and MS/MS scan was set at 120,000 and 30,000 on the Fusion Lumos and at 70,000 and 17,500 on the Q Exactive, respectively.

Data Analysis

MS1 analysis of All Detectable Ion Clusters in a LC−MS Run. All detectable MS1 ions in a LC−MS run were analyzed with an in-house program, NIST ProMS.[25] In a LC−MS peptide mapping analysis, each peptide ion can be detected via a set of MS1 (LC−MS) peaks with specific isotopes, abundance, and retention times (RT). This set of peaks is denoted as an ion cluster by ProMS. ProMS identifies isotopes of each ion cluster to determine its charge state, retention time, abundance, and monoisotopic m/z in all LC−MS/MS runs. The abundance of an ion cluster is calculated using the peak area of all observable isotope peaks. The RT of an ion cluster refers to a time when the accumulated abundance is the half of the total abundances. Additionally, the RT range of a peptide is defined as the first and last eluting values of the peptide.

Estimation of the Median RT Range. The RPLC column can generate consistent peptide RT under same experimental conditions, with small variations from run to run. In this study, the median value of RT range for specific experiments is used in a discriminant filter to remove unrelated identifications. The first step is to extract the RT range of the most abundant ion of each peptide; then the median RT range is calculated from replicate runs under identical experiment conditions.

Disulfide Peptide Analysis. The IgG1 disulfide bond structure (see Figure S1) is highly conserved with a simple disulfide connectivity pattern because most disulfide bonds are constrained within an individual heavy/light domain. In all but one case, component peptides can be found to be connected by a single native disulfide bond. The exception involves the identical peptides from two heavy chains linking to each other in the hinge region by two disulfide bonds (HC229-HC229 and HC232-HC232). More complex disulfide connectivity patterns involving multiple, intertwined disulfide bonds and the proximity of Cys residues in a peptide sequence require manual investigation or alternative MS fragmentation techniques. These are outside the scope of this paper.

Spectra of Disulfide-Linked Peptides. Each disulfide-bonded peptide in this study is composed of two component peptides: peptide 1 (P1) and peptide 2 (P2). In this work, when they are in separate polypeptide chains (heavy, H, or light, L), the one from the heavy chain is shown first. When in the same chain, they are shown in order of sequence number.

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

The example given below is a peptide connected by two consecutive Cys residues in the constant region of the L chain. It is represented in two ways in this paper:

(1) SGTASVV**C**LLNNFYPR P1 (LC133)
|
VYA**C**EVTHQGLSSPVTK P2 (LC193)

(2) SGTASVVCLLNNFYPR_**SS**_VYA-CEVTHQGLSSPVTK.

For assigning product ions of disulfide-bonded peptides, each spectrum can be considered as a mixture produced from P1 modified by P2 at the Cys residue and P2 modified by P1 at the Cys residue. For example

MS/MS spectrum = b and y ion series from "SGTASVVC-(P2-2H)LLNNFYPR" + b(2) and y(2) ion series from "VYAC(P1-2H)EVTHQGLSSPVTK"

As shown in the expression, in order to distinguish fragments from different components in disulfide-containing spectra, ions from P2 are labeled by "(2)" following ion series of a, b, or y.

Classification of Product Ions. Most of the product ions of disulfide bond fragmentation observed in this work fall into three main groups, as shown in Table 1.

Group 1: Fragments with an intact disulfide bond. They are further divided into three subgroups. Ions resulting from single peptide backbone cleavage in P1 or P2 are labeled using the standard notation of a/a(2), b/b(2), and y/y(2) ion series, shown in Subgroup 1a of Table 1. Ions arising from double cleavage in P1 or P2 are denoted as internal ions, such as "intVCLL" shown in Subgroup 1b. If a fragment sequence contains more than four residues, the position range is used, such as int7_12. When ions are composed of two fragments (P1 and P2 fragments), they are denoted by both product ions, such as "$[y_{10}:b(2)_5]$" in Subgroup 1c.

Group 2: Specific fragment ions involving a cleavage of disulfide bonds, i.e., carbon−sulfur (C−S) and sulfur−sulfur (S−S). C−S bond cleavage yields fragments (Subgroup 2a) with a modification of persulfide (+S, 31.9721 Da) or dehydroalanine (-H_2S, −33.9877 Da) at Cys, while S−S bond cleavage generates fragments (Subgroup 2b) with a modification of Cys (+0 Da) and Cys thioaldehyde (-2H, −2.0156 Da).

Group 3: Ions are due to single or double peptide backbone cleavages but do not contain a disulfide bond, so a conventional annotation is used (Subgroups 3a and 3b).

Product Ion Assignment. A software tool was developed for matching experimental MS/MS fragment ions from HCD and FT-CID spectra to the in silico fragmentation of the component peptides as well as disulfide bonds. First, an analysis of fragment isotopic distributions was performed, which allows the accurate mass determination of product ion monoisotopic peaks in MS/MS spectra. The deviation threshold to accept an isotopic assignment is 20 ppm. For annotating fragments, the program makes assignments in the order of the occurrence frequency of fragment ions: (1) neutral loss of precursor ions; (2) amide bond cleavage ion series with and without a disulfide bond and their common neutral losses (H_2O, NH_3, NH_3H_2O, or $2H_2O$); (3) disulfide cleaved at C−S and S−S bonds; (4) double cleavage with and without a disulfide bond; and (5) neutral losses (C_2H_4O and $C_2H_6O_2$) from peptide bond cleavage ion series. All fragments that have reasonable isotope patterns are analyzed, and those with deviation within 20 ppm between observed and theoretical *m/z* values are annotated.

RESULTS

Identification of Disulfide Bonds

Data Analysis Strategy. Disulfide-linked peptides cannot be identified by most current database searching tools since they assume linear peptides. In recent disulfide studies,[7−9] the anticipated polypeptides containing disulfide linkages are often used for initial searching of MS2 spectra acquired from non-reduced or partially reduced digests, and the spectra matched by theoretical mass are then confirmed by product ion assignment. However, because of the complexity of disulfide bond fragmentation, many of such peptides may be undetected or misidentified. To address this issue, we used a previously developed alternative workflow for characterizing hard-to-identify peptides and modifications derived from the NISTmAb in an LC−MS/MS analysis.[5] Our method starts with searching for MS1 ion clusters to find all disulfide-containing candidates of interest based on the protein sequence and then assigns their product ion spectra to the in silico fragmentation of specific disulfide-linked peptides for validating identification results.

Figure 1 is a flowchart of the process of identifying spectra of disulfide-linked peptides, which are subsequently used for creating a spectral library. First, theoretical *m/z* values are computed for all potential disulfide-linked peptides derived from known native and potentially scrambled disulfide bond linkages based on protein sequence (step 1). Second, all detectable MS1 peaks are analyzed to determine their monoisotopic *m/z*, charge, abundance, and RT of each ion cluster (step 2). In the third step, the initial identification is conducted by matching observed *m/z* values derived from step 2 with the precursor *m/z* values in step 1. The median RT of these matched ions in the same digestion conditions is calculated and used to filter identifications of ions with inconsistent RT. After disulfide-linked peptides are tentatively identified, step 4 extracts MS2 spectra and conducts fragment ion analysis by interpreting their product ion spectra. In step 5, replicate spectra are selected based on product ion assignments. The selected spectrum should include at least 60% of peak abundances that can be assigned to fragments of disulfide-specific structure and peptide sequence. Lastly, all validated

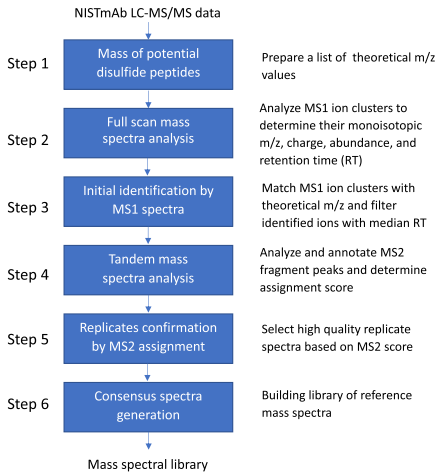


Figure 1. Building library of reference mass spectra for disulfide analysis.

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

Table 2. Disulfide-Linked Peptides Detected for Each of the Nine Native Disulfide Linkages Obtained from the Triplicate LC−MS/MS Analyses of the 18 h NISTmAb Digest Partially Reduced with TCEP and Alkylated with IAM[g,h]

bond name	position in protein	class	missed cleaved sites	peptide sequence	mass	charge	absolute abund.
1[a] V_H	HC22-HC97	**tryptic 25%**	**0**	**[e]ESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIR_SS_VTNMDPADTATYYCAR**	**5431.587**	**4,5,6**	**3.6×10^{10}**
		missed cleav.	1	[e]ESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIR_SS_VTNMDPADTATYYCARDMIFNFYFDVWGQGTTVTVSSASTK	8213.882	5,6,7	9.3×10^{10}
			2	ESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIR_SS_VTNMDPADTATYYCARDMIFNFYFDVWGQGTTVTVSSASTKGPSVFPLAPSSK	9381.511	6,7	8.6×10^{09}
			2	ESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIRQPPGK_SS_VTNMDPADTATYYCARDMIFNFYFDVWGQGTTVTVSSASTK	8721.163	5,6,7,8	3.2×10^{09}
			3	ESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIR_SS_DTSKNQVVLKVTNMDPADTATYYCARDMIFNFYFDVWGQGTTVTVSSASTK	9326.501	6,7,8	2.3×10^{09}
			3	ESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIRQPPGK_SS_VTNMDPADTATYYCARDMIFNFYFDVWGQGTTVTVSSASTKGPSVFPLAPSSK	9888.792	7,8	8.3×10^{08}
		semi- tryptic	0	ESGPALVKPTQTLTLTCTFSGF_SS_VTNMDPADTATYYCAR	4085.906	4,5	2.4×10^{09}
			0	ESGPALVKPTQTLTLTCTFSGFSL_SS_VTNMDPADTATYYCAR	4286.022	3	3.4×10^{08}
2[a] C_H1	HC147-HC203	**tryptic 94%**	**0**	**[e]STSGGTAALGCLVK_SS_DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTK**	**7916.919**	**5,6,7,8**	**1.8×10^{11}**
		missed cleav.	1	[f]STSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTK	7898.909	5,6	4.4×10^{09}
			1	STSGGTAALGCLVK_SS_DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDK	8259.11	6,7	4.8×10^{08}
		semi- tryptic	0	STSGGTAALGCLVK_SS_SWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTK	6869.428	5,6,7	4.5×10^{09}
			0	STSGGTAALGCLVK_SS_TVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTK	7069.544	6	1.8×10^{09}
			0	STSGGTAALGCLVK_SS_SLSSVVTVPSSSLGTQTYICNVNHKPSNTK	4409.221	5,6	3.3×10^{08}
			0	STSGGTAALGCLVK_SS_SVVTVPSSSLGTQTYICNVNHKPSNTK	4122.073	4,5,6	2.2×10^{08}
			0	STSGGTAALGCLVK_SS_SSVVTVPSSSLGTQTYICNVNHKPSNTK	4209.105	4,5	1.4×10^{08}
			0	STSGGTAALGCLVK_SS_TVPSSSLGTQTYICNVNHKPSNTK	3836.904	4,5	9.3×10^{07}
3[a] C_H2	HC264-HC324	**tryptic 96%**	**0**	**[e]TPEVTCVVVDVSHEDPEVK_SS_CK**	**2328.098**	**2,3,4,5**	**2.5×10^{11}**
		missed cleav.	2	TPEVTCVVVDVSHEDPEVK_SS_VVSVLTVLHQDWLNGKEYKCK	4537.287	4,5,6,7	6.6×10^{09}
			3	TPEVTCVVVDVSHEDPEVK_SS_VVSVLTVLHQDWLNGKEYKCKVSNK	4965.526	5,6	1.5×10^{09}
			1	TPEVTCVVVDVSHEDPEVK_SS_EYKCK	2748.299	4,5	9.4×10^{08}
			1	TPEVTCVVVDVSHEDPEVK_SS_CKVSNK	2756.336	4,5	2.8×10^{08}
		semi- tryptic	0	TPEVTCVVVDVSHED_SS_CK	1874.839	4	5.1×10^{08}
			0	TPEVTCVVVDVSH_SS_CK	1630.77	2	4.7×10^{08}
			0	TPEVTCVVVD_SS_CK	1307.61	2,3	3.4×10^{08}
4[a] C_H3	HC370-HC428	**tryptic 99%**	**0**	**[e]NQVSLTCLVK_SS_WQQGNVFSCSVMHEALHNHYTQK**	**3844.824**	**3,4,5,6,7**	**9.9×10^{10}**
		semi- tryptic	0	NQVSLTCLVK_SS_VFSCSVMHEALHNHYTQK	3231.563	4,5	7.6×10^{07}

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

Table 2. continued

bond name	position in protein	class	missed cleaved sites	peptide sequence	mass	charge	absolute abund.
5[b] V_L	LC23- LC87	**tryptic 92%**	**0**	**VTITCSASSR_SS_FSGSGSGTEFTLTISSLQPDDFATYYCFQGSGYPFTFGGGTK**	**5446.46**	**4,5**	**1.4×10^{10}**
		missed cleav.	1	DIQMTQSPSTLSASVGDRVTITCSASSR_SS_FSGSGSGTEFTLTISSLQPDDFATYYCFQ GSGYPFTFGGGTK	7320.344	5	1.0×10^{09}
			2	VTITCSASSR_SS_FSGSGSGTEFTLTISSLQPDDFATYYCFQGSGYPFTFGGGTKVEIKR	6071.851	5,6	2.0×10^{08}
6[b] C_L	LC133-LC193	**tryptic 66%**	**0**	**SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPVTK**	**3555.749**	**3,4,5**	**1.2×10^{11}**
		missed cleav.	1	SGTASVVCLLNNFYPR_SS_HKVYACEVTHQGLSSPVTK	3820.903	3,4,5,6,7	6.2×10^{10}
			2	SGTASVVCLLNNFYPR_SS_ADYEKHKVYACEVTHQGLSSPVTK	4427.168	5,6	6.0×10^{08}
		semi- tryptic	0	SGTASVVCLLNNFYPR_SS_VYACEVTH	2658.257	4	4.5×10^{07}
7[c] H-L	HC223-LC213	missed cleav.	1	SCDKTHTC(**CAM**)PPC(**CAM**)PAPELLGGPSVFLFPPKPK_SS_GEC	3467.639	3,4,5,6,7	2.6×10^{10}
			2	SCDKTHTC(**CAM**)PPC(**CAM**)PAPELLGGPSVFLFPPKPK_SS_SFNRGEC	3971.883	4,5,6,7	3.3×10^{09}
			3	SCDKTHTC(**CAM**)PPC(**CAM**)PAPELLGGPSVFLFPPKPKDTLMISR_SS_SFNRGEC	4788.299	4,5,6	2.7×10^{09}
			1	SCDK_SS_SFNRGEC	1260.486	2,3,4	7.8×10^{08}
8 and 9[d] Hinge	HC229-HC229 and HC232-HC232	**tryptic 82%**	**0**	**THTCPPCPAPELLGGPSVFLFPPKPK_2(SS)_THTCPPCPAPELLGGPSVFLFPPKPK**	**5456.799**	**4,5,6,7,8**	**6.7×10^{10}**
		missed cleav.	1	THTCPPCPAPELLGGPSVFLFPPKPK_2(SS)_SC(**CAM**) DKTHTCPPCPAPELLGGPSVFLFPPKPK	5889.962	5,6,7,8,9	1.2×10^{10}
			2	SC(**CAM**)DKTHTCPPCPAPELLGGPSVFLFPPKPK_2(SS)_SC(**CAM**) DKTHTCPPCPAPELLGGPSVFLFPPKPK	6323.125	4,5,6,7,8,9	2.2×10^{09}
		semi- tryptic	0	THTCPPCPAPELLGGPSVFLFPPKPK_2(SS)_THTCPPCPAPELLGGPSVF	4649.298	5	5.4×10^{08}

[a]V_H, C_H1, C_H2, and C_H3 refer to the variable and constant domains 1 to 3 in H chain, respectively. [b]V_L and C_L represent the variable and constant domains of the L chain. [c]H−L is the interchain bond linking the H and L chains. [d]Hinge region contains the two disulfide bonds connecting the two H chains. [e]Oxidized and/or metallated forms were also detected. [f]Intrapeptide form. [g]Notes: Bond: disulfide bond named by domains. Position: Cys position in each protein chain. Class: peptide class. Miss cleav.: total number of missed cleavage sites. Peptide sequence: sequence of two component peptides linked by a disulfide bond. Mass: mass of disulfide-linked peptide. z: charge state of disulfide-linked peptide. Absolute abund.: summed absolute abundance over all charge states of a peptide. [h]In Column 3, the fully tryptic peptide is shown in bold, and the relative abundance is calculated as the percentage of the tryptic peptide abundance over the total peptide abundance of the linkage site. Note that several peptides containing disulfide bonds 7, 8, and 9 were identified with carbamidomethylated (CAM) cysteine residues.

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

replicate spectra are merged to form consensus spectra for inclusion in a spectral library (step 6).

Identification of Native (Expected) NISTmAb Disulfide Linkages. Initially, we tested the above workflow with all MS and MS/MS data obtained by a wide range of common digestion protocols under alkaline pH conditions (see Table S1) that were used in developing NISTmAb spectral libraries.[5,6] We then compared different digestion profiles using both the number and ion intensities of confident identifications of all peptide classes including native disulfide-bonded peptides (see Figure S2A,B). Both non-reduction and partial reduction protocols were selected as the preferred digestion conditions under which the identification of relatively higher abundances and large numbers of disulfide-containing peptides were achieved. For developing spectral libraries, the experiments were designed to generate the widest possible variety of peptides of analytical concern. No attempts were made to optimize digestion; instead, we employed distinct but complementary protocols to achieve a broad coverage of peptides and modifications. Considering that disulfide bond shuffling can be easily induced under these common alkaline pH conditions, non-reducing and partially reducing experiments under neutral pH for protein digestions were performed in order to control shuffling induced by LC−MS analysis, as described in the Materials and Methods section.

The NISTmAb, a typical IgG1 protein, is composed of four polypeptide chains including two identical heavy (H) and two identical light (L) chains (Figure S1). Each H chain has a variable domain (V_H) and three constant domains (C_H1, C_H2, and C_H3), while each L chain has a variable domain (V_L) and a constant domain (C_L). In total, the protein has 12 individual structural domains. Each domain contains a single disulfide bond or an intrachain disulfide bond. The two H chains are linked to each other in the hinge region (Hinge) by two interchain disulfide bonds, and each H chain is connected to an L chain by an interchain disulfide bond (H−L). Thus, the NISTmAb has a total of 12 intrachain disulfide bonds and four interchain disulfide bonds. In this peptide mapping study, there are nine distinct native disulfide bonds (Table 2) that needed verification for the NISTmAb due to the symmetrical molecular structure formed by these antibodies.

Disulfide analyses were done by identifying all potential disulfide-linked peptide classes from the tryptic digests of the NISTmAb, including fully tryptic cleavage, expected missed cleavages, irregular cleavage, ESI-induced adducts, and other modifications. Most disulfide-containing peptides are connected by a single disulfide bond. The exception is the peptides in the hinge region that contain two disulfide bonds located in the proximity of each other to connect two H chains, as described in the Materials and Methods section. Table 2 shows the nine distinct native disulfide bonds mapped in the NISTmAb through the identification of the disulfide-linked peptides corresponding to each linkage. The table presents MS1 peptide data acquired from the triplicate LC−MS analyses of the 18 h solution tryptic digest of the NISTmAb partially reduced with TCEP and alkylated with IAM at pH 8. These identified disulfide-linked peptides range in mass from 1260 to 9888 Da with charge states from +2 to +9. A large majority of fully tryptic and miscleaved disulfide peptides are detected with multiple charges. Furthermore, more highly charged ions of tryptic peptides with missed cleavage are found, providing complementary information for the linkage analysis. Disulfide peptides containing multiple Met/Trp residues are detected in the composite spectrum generated from multiple mono-/dioxidized products. Disulfide peptides originated from irregular cleavage (semitryptic) were identified. Metallated ions formed by abundant tryptic disulfide peptides through electrospray ionization (ESI) were also found and verified by their precursor mass and coelution. The identification of these disulfide peptide variants facilitates reliable disulfide bond analysis and provides additional fragmentation information as shown in the Fragmentation section. Table 2 summarizes the LC−MS analyses of disulfide-linked peptides for each linkage and provides the name of disulfide bonds, the linkage location in the protein, class, the number of missed-cleaved sites, sequence of peptides, mass, observed charge states, and absolute ion signal intensities.

Major Disulfide Tryptic Products. Disulfide-bonded peptide pairs are classified as disulfide tryptic peptides when all termini are consistent with simple tryptic cleavage. Other peptide varieties such as oxidized and metallated disulfide peptides can originate from modification to these peptides. Table 2 shows that major disulfide tryptic peptides of seven (bonds 2 to 6 and 8 to 9) out of nine linkages produced 66 to 99% of the total disulfide peptide abundance (column 3 in bold of Table 2). In contrast, the percent abundance of tryptic peptides of bond 1 was low (25%) and tryptic peptides of bond 7 was not detectable, indicating the need for alternative peptide classes (e.g., missed cleavage and/or irregular cleavage) for a complete analysis of disulfides in antibodies.

Missed Cleavage Products. Disulfide bonds generally hinder proteolytic digestion by common enzymes;[22] thus, peptides with up to four missed-cleavage sites were sought. This led to the identification of 152 disulfide-bonded peptides with missed-cleavage sites, which accounted for 40% of the total identified peptides. As seen from Table 2, the number of missed-cleavage sites are shown in Column 4 and they were observed in all disulfide bonds except the bond C_H3. Varying levels of tryptic peptides with missed cleavage are identifiable with multiple charge states. The detection of such peptides is an important supplementary disulfide determination, especially when their tryptic counterparts are too small to be detected (H−L, bond 7) or low abundance species (V_H, bond 1). Four alternative disulfide peptides containing one to three missed cleavages were observed for the bond H−L at the multiple charge states shown in Table 2, ensuring confidence in the assignment of this disulfide linkage. Furthermore, the presence of reduced and alkylated Cys residues with the disulfide-bonded peptides (Bonds H−L and Hinge) can increase their ion signals, and thus their abundances presented in Table 2 are significant even with overnight digestion. Missed-cleavage products produced additional fragmentation patterns and their spectra are part of the spectral library.

Irregular Cleavage Products. Irregular cleavage products, involving a single semitryptic cleavage, are sought during the MS1 ion cluster search. In total, 27 such peptides, arising from major tryptic peptide cleavage, were identified, with most of them generated from protein digestion, whereas a few were produced from electrospray ionization (ESI) in-source fragmentation. Although these are minor products resulting from larger disulfide species, they can provide, as short peptides, useful information to aid in fragmentation analysis. For example, the semitryptic peptide ions 5 to 7 in Table 4 are selected for characterizing disulfide bond fragmentation because their MS/MS spectral data contain useful structural information.

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

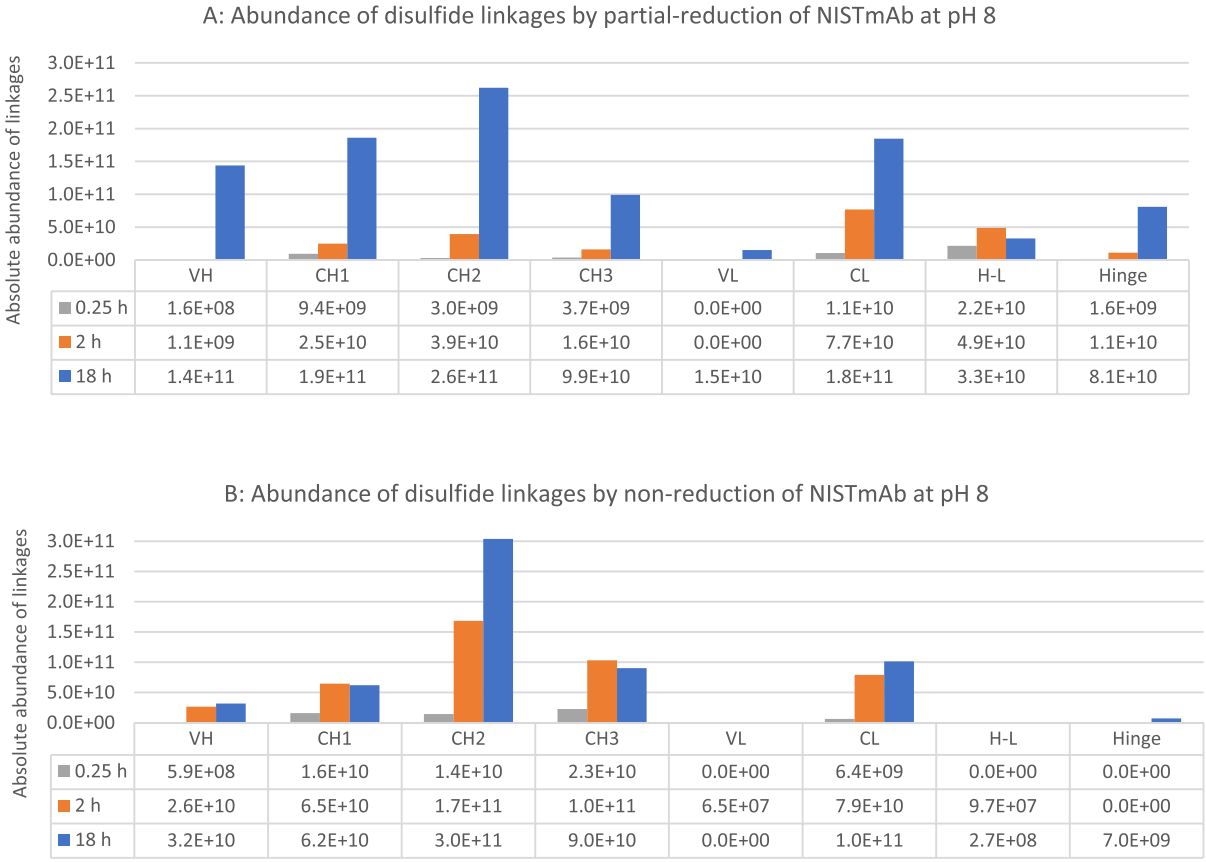


Figure 2. Summed abundances for each of the nine native disulfide bonds in the NISTmAb obtained by median results of triplicate analyses from three separate tryptic digests at 0.25, 2, and 18 h using partially reducing and non-reducing conditions. (A) Partially reduced peptide mapping using low concentration of TCEP at pH 8; (B) non-reduced peptide mapping at pH 8.

Complex Met/Trp Oxidation Mixtures of Disulfide Species. Several Met/Trp (methionine/tryptophan) oxidation products are observed with two disulfide linkages of V_H and C_H3 (bonds 1 and 4, Table 2) where multiple Met/Trp are present, in the partial reducing experiments. They are likely generated from two different sources: in-sample and in-source, which is consistent with previous observations.[5,26] Oxidized species with more than one Met/Trp residue often produce a mixture of disulfide-linked peptide spectra. Structural differences of the oxidation mixtures are due to position variations of Met or Trp or both in the sequence, making them difficult to recognize using conventional software tools. Selected ion chromatograms of oxidized disulfide-bonded peptides containing methionine/tryptophan are shown in Figure S3. Their MS2 mixture spectra are manually validated and collected in the library.

Comparison of Tryptic Digestion Profiles of Partially/Non-reduced Native Disulfide Bonds at pH 8. We conducted a comparative evaluation of the above two protocols to estimate their digestion efficiency for disulfide mapping using MS1 ion signal intensities of the unique identifications of disulfide-containing peptides. The digestion of partially reduced and alkylated NISTmAb produces three groups of Cys-containing peptides: major alkylated peptides, relatively abundant disulfide-linked peptides, and minor reduced peptides, whereas the non-reduced and non-alkylated digest almost exclusively yields peptides containing disulfide bonds. This comparison was conducted using disulfide-linked peptides. The summed abundances for each linkage site displayed in the bar graph from Figure 2A,B were obtained from two LC−MS analyses at pH 8 for partially reduced and non-reduced proteins digested for 0.25, 2, or 18 h. The site-specific quantitation combined the ion intensities of multiply charged disulfide tryptic peptides and peptides with mis-cleavages at each disulfide bond. It is well-known that LC−MS abundance measurement of low-intensity ions, such as disulfide-linked ions, is very challenging given the intrinsic bias against low abundance peptides, their lower ionization efficiency in LC − ESI − MS, and possible run-to-run variation in ion suppression and others. We used the site-specific quantitation as a measure of the relative detectability of the nine native disulfide bonds by the LC−MS analyses with the selected two protocols. At short digestion times (0.25 and 2 h), mostly low-abundance species were observed and no disulfide-containing peptides were detected from the L chain variable regions as well as from the hinge region. At 18 h, most of the major disulfide-linked peptides were observed with significantly higher signal intensities shown in Figure 2A but not in non-reduced datasets (Figure 2B, except the C_H2 linkage), which may suggest that a longer time is needed to digest non-reduced proteins.

In the digestion time course experiments for 0.25, 2, and 18 h at alkaline pH, a total of 129 native disulfide-bonded peptides were identified by the partial reduction condition with the low concentration of TCEP, whereas 85 of such peptides were detected using non-reducing and non-alkylating conditions. The spectra of these identifications are included in the spectral library. A major advantage of the pH 8 partial reduction approach is that partially reduced disulfides and

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

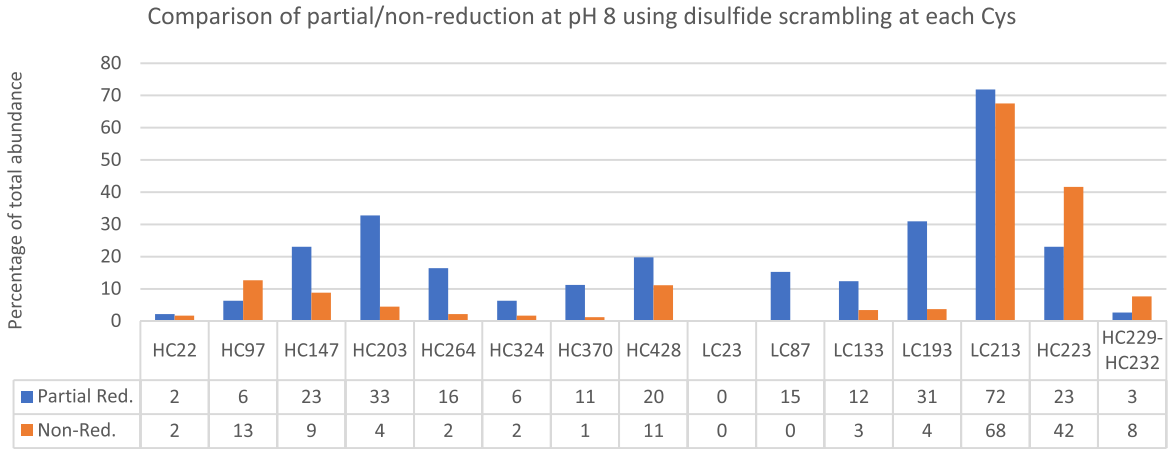


Figure 3. Levels of scrambling occurring at pH 8 for each of Cys sites are observed from non-reduction and partial reduction experiments. Partial Red.: reduction by TCEP and alkylation by IAM. Non-Red.: no reduction and alkylation for antibody digestion. Cysteines are labeled by their residue number in L or H chain, e.g., HC264 denotes Cys at residue 264 of the H chain.

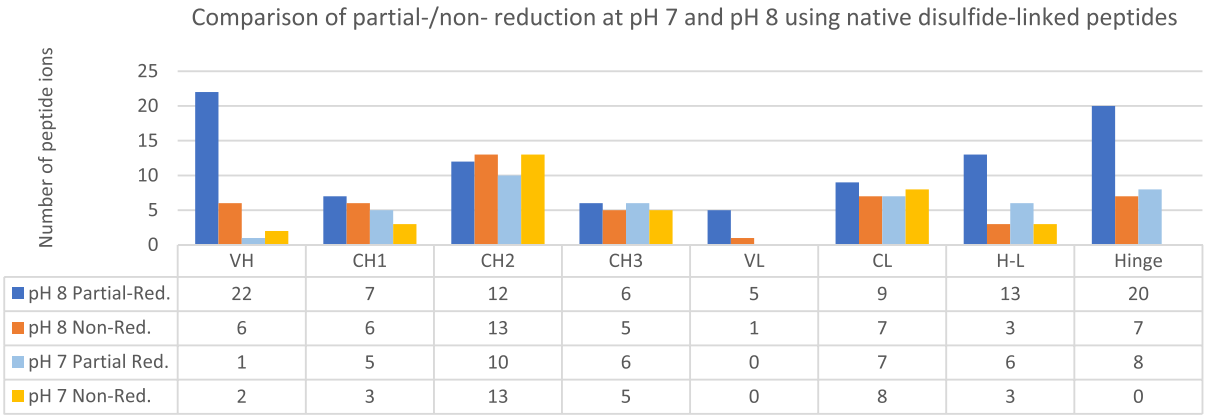


Figure 4. Peptide identifications containing native disulfide bonds obtained from four triplicate analyses of the 18 h tryptic digest using different parameters for four protocol variations: non-reducing and partially reducing conditions at pH 7 and pH 8, respectively.

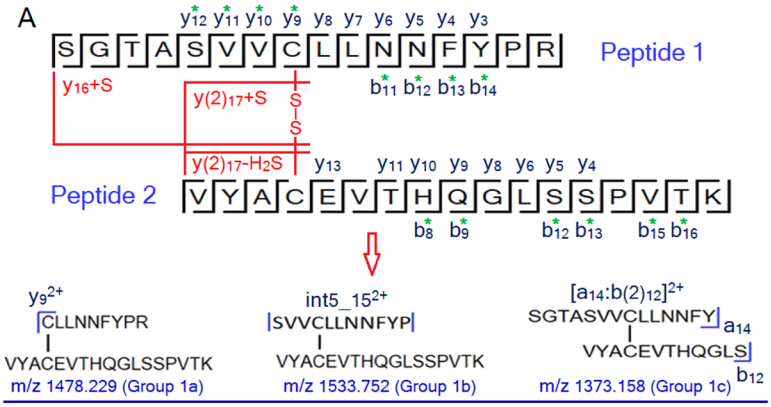
alkylated cysteines greatly improve protein digestion efficiency for linear and alkylated peptides[5] as well as for disulfide-containing peptides. This also allows the identification of disulfide-bonded peptides along with any non-disulfide-containing peptides in the same experiment.

Detection of Scrambled (Unexpected) Disulfide Bonds at pH 8 in the LC−MS/MS Analysis. After the native NISTmAb disulfide linkages were verified using the non−/partially reducing LC−MS/MS analyses, we applied the same workflow to find disulfides with unexpected linkages. All 136 possible linkages that can be formed from the 16 Cys residues were sought. Only unmodified disulfide peptides are sought. The results reveal that these reactions readily occur at nearly all Cys residues under alkaline conditions (pH 8), with the exception of LC87. In addition to the nine expected native disulfide linkages, 86 non-native disulfide linkages were observed from 109 disulfide-scrambled tryptic peptides.

Figure 3 displays the level of scrambling occurring in the 18 h tryptic digest at pH 8 for every Cys residue of the NISTmAb as the percent abundance of the total intensities of all disulfide bonds from each Cys including native and scrambled. In the partial reduction of the protein digest, 8 out of 16 Cys residues exhibit over 15% disulfide scrambling. In particular, HC203, LC193, and LC213 comprise 33, 31, and 72% of all disulfide abundances, respectively. The non-reduction condition generated much less scrambling as compared to the partial reduction, although HC223 and LC213 show higher level of scrambling at 42 and 68%, respectively. It is evident that when using pH 8 for protein digestion, whether non-reduced or partially reduced samples, wide spread disulfide rearrangement could occur. LC213 was detected as a hotspot for both conditions.

To minimize disulfide bond rearrangement at basic pH, two sets of control experiments (non-reduction and partial reduction) were performed by adjusting the buffer pH to 7 and using NEM as the alkylating reagent due to its thiol specificity and stability at or below pH 7, as described in the Experimental Section. Over 90% of the disulfide bond artifacts detected at pH 8 disappeared or were only found at the MS1 trace levels when neutral pH and NEM were used. Seven peptides with scrambled disulfides identified from MS/MS spectra are shown in Table S2A. Only four low-abundance non-native disulfide bonds were detected in the non-reduced digest, and just one was detected in the partial reduction digest. Seven Cys residues (HC22, HC147, HC264, HC324, HC370, LC23, and LC213) are involved in the formation of scrambled disulfide bonds detected at trace levels. This result illustrates that sample preparation using neutral or lower pH significantly decreases shuffling.

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629



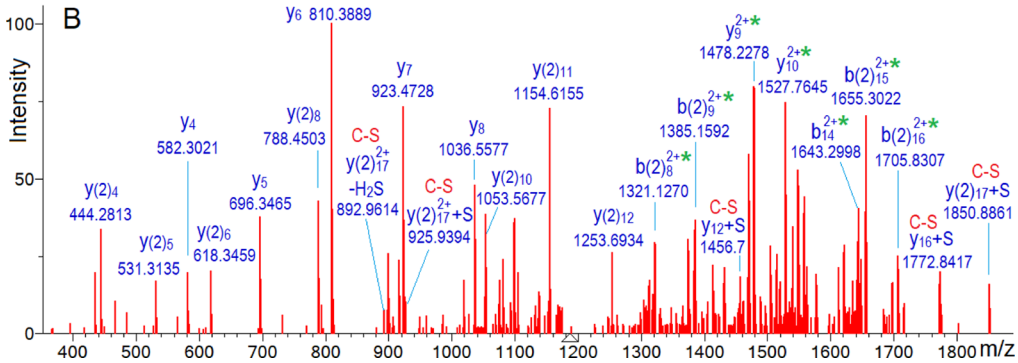


Figure 5. Different types of product ions in the MS/MS spectrum of a triply charged disulfide-linked peptide, SGTASVVCLLNNFYPR_**SS**_-YACEVTHQGLSSPVTK, at m/z 1186.257 with less than 5 ppm deviation from its theoretical value, from FT-CID fragmentation at NCE of 35%. (A) Sequencing disulfide linkage and (B) fully annotated MS/MS spectrum. The * symbol in green color represents fragments from amide bond cleavages and contains the intact disulfide bond. The annotation in red color denotes fragments from C−S bond cleavage.

To gain a better understanding of the pH 7 data, the identification of native disulfide-containing peptides at pH 7 is also compared to that at pH 8. Figure 4 shows the number of disulfide-bonded peptides detected in each of four triplicate analyses of the 18 h tryptic digest under non-reduction and partial-reduction at pH 7 and pH 8, respectively. The notable variation in the number of identified ions is observed at all linkages by different digestion protocols, although more agreeable results are seen at four linkages (C_H1, C_H2, C_H3, and C_L). In all cases, the partial reduction at pH 8 yields more disulfide-containing peptides than others. At pH 8 conditions, a total of 114 different ions containing native disulfide bonds were identified, while 55 of such ions were identified under pH 7 conditions (see Table S2B). It is evident that at pH 7, each linkage yields fewer ions than those at pH 8; specifically, no detectable ions are found at two out of the nine disulfide linkages. These results suggest that pH is a critical parameter for effective disulfide analysis, and the selection of the preferable digestion conditions should consider the detectability of both native and scrambled disulfide bonded peptides.

Fragmentation Analysis of Disulfide-Linked Peptides

A central objective of this work was to completely map all the disulfide linkages in the NISTmAb by identifying as many disulfide-bonded peptides and their product ions as possible. This was done in the following manner:

(1) Manually annotate MS/MS spectra of disulfide-linked peptides.

(2) Develop an identification tool for automatic fragment assignment using information from the previous step.

(3) Apply this tool to all spectra containing disulfide bonds that were identified by MS1 ion analysis.

(4) Develop an evaluation tool for assessing the thoroughness and accuracy of annotations by examining the mass accuracy and abundance ratio of assigned/unassigned product ions.

(5) If step 4 finds incorrect or missed annotations, go back to step 1 to obtain new knowledge of unknown fragments, and then work through steps 1 to 5 until no more significant ions are unassigned.

Overview of Different Types of Fragments Observed. In this section, we introduce different types of product ions containing a disulfide bond with an example spectrum. Then, we describe the general fragmentation features of HCD and CID spectra obtained in FT MS/MS analysis of disulfide-bonded peptides under low collisional energy.

Figure 5 shows an annotated MS/MS spectrum of a peptide containing an intrachain disulfide bond. A total of 100 unique product ions were assigned, identified with relative abundance greater than 1% of the base peak, which constituted 92% of the total ion intensity. Table 3 summarizes the 55 fragments having a relative abundance above 5% with a mass accuracy within 5 ppm. These fragments are organized into three groups (see column 1 of Table 3 and Table 1 in the Materials and Methods section for the detailed notation).

In this example, group 1 contains 35 peptide backbone fragments all with an intact disulfide linkage. For example, the dominant peak at m/z 1478.229 of group 1a, as shown in Figure 5A, corresponds to the doubly charged ion y_9 of P1

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

Table 3. Fragment Ion Annotation in an FT-CID MS/MS Spectrum of the Charge 3+ Disulfide-Linked Peptide, SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPVTK[a]

class	group	cleavage site	m/z	mass error in ppm	assignment of fragment	charge	sequence	percent abundance
intact disulfide bond fragment	1a	single C-N cleavage in either peptide	985.8223	1.1	y_9	3	CLLNNFYPR_SS_VYACEVTHQGLSSPVTK	5.5
			1075.204	1.9	y_{12}-NH_3	3	SVVCLLNNFYPR_SS_VYACEVTHQGLSSPVTK	18.5
			1080.879	1.4	y_{12}	3	SVVCLLNNFYPR_SS_VYACEVTHQGLSSPVTK	25.4
			1131.5491	2.2	$b(2)_{16}$-H_2O	3	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPVT	10.2
			1308.6421	1	b_9-H_2O	2	SGTASVVCL_SS_VYACEVTHQGLSSPVTK	6.4
			1312.1265	1	$b(2)_8$-H_2O	2	SGTASVVCLLNNFYPR_SS_VYACEVTH	19.5
			1321.1309	0.3	$b(2)_8$	2	SGTASVVCLLNNFYPR_SS_ YACEVTH	32.8
			1385.1595	0.2	$b(2)_9$	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQ	47
			1413.6871	4.6	b_{11}-H_5NO	2	SGTASVVCLLN_SS_VYACEVTHQGLSSPVTK	19.9
			1431.2117	0.4	b_{11}	2	SGTASVVCLLN_SS_VYACEVTHQGLSSPVTK	26.5
			1478.229	0.5	y9	2	CLLNNFYPR_SS_VYACEVTHQGLSSPVTK	100
			1488.2318	1.3	b_{12}	2	SGTASVVCLLNN_SS_VYACEVTHQGLSSPVTK	17.7
			1504.7283	3.3	$b(2)_{12}$-H_2O	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLS	30
			1513.7325	2.6	$b(2)_{12}$	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLS	27.7
			1527.7645	1.4	y_{10}	2	VCLLNNFYPR_SS_VYACEVTHQGLSSPVTK	98.5
			1539.2399	3.8	$b(2)_{13}$-$2H_2O$	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSS	38.4
			1553.2588	2.7	b_{13}-NH_3	2	SGTASVVCLLNNF_SS_VYACEVTHQGLSSPVTK	7.2
			1557.2498	3.3	$b(2)_{13}$	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSS	48.7
			1561.7668	0.7	b_{13}	2	SGTASVVCLLNNF_SS_VYACEVTHQGLSSPVTK	26
			1577.3018	3.3	y_{11}	2	VVCLLNNFYPR_SS_VYACEVTHQGLSSPVTK	22.8
			1612.3059	4.1	y_{12}-NH_3	2	SVVCLLNNFYPR_SS_VYACEVTHQGLSSPVTK	18.7
			1625.7781	1.8	b_{14}-H_5NO	2	SGTASVVCLLNNFY_SS_VYACEVTHQGLSSPVTK	6.9
			1629.3011	0.6	a_{14}	2	SGTASVVCLLNNFY_SS_VYACEVTHQGLSSPVTK	6.6
			1634.7865	0.1	b_{14}-NH_3	2	SGTASVVCLLNNFY_SS_VYACEVTHQGLSSPVTK	15.6
			1641.3032	2.8	$a(2)_{15}$	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPV	18.2
			1643.2998	0.1	b_{14}	2	SGTASVVCLLNNFY_SS_VYACEVTHQGLSSPVTK	55.1
			1646.7877	2.6	$b(2)_{15}$-NH_3	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPV	30.1
			1655.3022	1.8	$b(2)_{15}$	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPV	86.6
			1696.8263	1.5	$b(2)_{16}$-H_2O	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPVT	22.1
			1705.8307	1	$b(2)_{16}$	2	SGTASVVCLLNNFYPR_SS_VYACEVTHQGLSSPVT	31.8
	1b	double C-N cleavage in one peptide	1533.7521	3.1	int5-15	2	SVVCLLNNFYP_SS_VYACEVTHQGLSSPVTK	13.1
	1c	double C-N cleavage in both peptides	1306.1199	4.1	$[b(2)_{13}$-$y_{10}]$	2	VYACEVTHQGLSS_SS_VCLLNNFYPR	8.1
			1329.6398	1.2	$[a_{14}$:$b(2)_{11}]$	2	SGTASVVCLLNNFY_SS_VYACEVTHQGL	7
			1373.158	2.8	$[a_{14}$:$b(2)_{12}]$	2	SGTASVVCLLNNFY_SS_VYACEVTHQGLS	44.1
			1404.179	4.9	$[b(2)_{14}$:$y_{11}]$	2	VYACEVTHQGLSSP_SS_VVCLLNNFYPR	11
cleaved disulfide bond fragments	2a	C-S cleavage	892.9614	1.2	$y(2)_{17}$-H_2S	2	VYACEVTHQGLSSPVTK	3.5
			925.9394	3.3	$y(2)_{17}$ + S	2	VYACEVTHQGLSSPVTK	11.2
			1456.7136	4.1	y_{12} + S	1	SVVCLLNNFYPR	8.2
			1772.8417	2.3	y_{16} + S	1	SGTASVVCLLNNFYPR	17.1
			1850.8861	4.6	$y(2)_{17}$ + S	1	VYACEVTHQGLSSPVTK	10.1
fragments without disulfide bond	3a	single C-N cleavage in both peptides	435.2348	0.6	y_3	1	YPR	7.5
			444.2815	0.4	$y(2)_4$	1	PVTK	12.8

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

Table 3. continued

class	group	cleavage site	m/z	mass error in ppm	assignment of fragment	charge	sequence	percent abundance
			531.3135	0.4	$y(2)_5$	1	SPVTK	7.3
			582.3021	2.3	y_4	1	FYPR	10
			618.3459	0.3	$y(2)_6$	1	SSPVTK	8.8
			696.3465	0.2	y_5	1	NFYPR	18.5
			788.4503	1.2	$y(2)_8$	1	GLSSPVTK	21.2
			810.3889	0.5	y_6	1	NNFYPR	51.8
			899.4845	1.4	$y(2)_9$-NH_3	1	QGLSSPVTK	13.5
			916.511	1.3	$y(2)_9$	1	QGLSSPVTK	12
			923.4728	0.6	y_7	1	LNNFYPR	40.9
			1036.5577	0.3	y_8	1	LLNNFYPR	29.9
			1053.5677	1	$y(2)_{10}$	1	HQGLSSPVTK	23
			1154.6155	0.8	$y(2)_{11}$	1	THQGLSSPVTK	44.5
			1382.7231	3.1	$y(2)_{13}$	1	EVTHQGLSSPVTK	11.5

[a]Notes: Class and group: fragment class and group. Cleav. site: peptide bond or disulfide bond. m/z: the mass/charge ratio for fragment ion. Error ppm: mass deviation between observed and theoretical m/z values. z: charge state of fragment ion. Sequence: sequence of fragment ion. Abund. % relative abundance of specific product ions with respect to the base peak in a MS2 spectrum.

linked to P2 via a disulfide bond. Aside from these ions generated by single peptide bond cleavage, more complex products arise from double cleavage such as the ions in the groups 1b and 1c. For instance, the peak at m/z 1533.752 formed by double cleavage of amide bonds was identified as the doubly charged ion involving the P1 residues from 5 to 15 linked to P2 via a disulfide bond, labeled Int5_15 in Group 1b of Figure 5A. Another example is the doubly charged internal ion at m/z 1373.158 consisting of the fragment a_{14} of P1 and the fragment $b(2)_{12}$ of P2 connected by a disulfide bond (Group 1c of Figure 5A). Overall, group 1 comprises a majority (71%) of total product ion intensity and provides detailed peptide sequence.

This example spectrum also contains five product ions arising from disulfide cleavage as shown in red color in Figure 5B and Group 2a of Table 3. They were formed through the loss of H_2S (e.g., charge 2+ $y(2)_{17}$-H_2S) or the addition of S (e.g., the singly and doubly charged ion $y(2)_{17}$ + S) in the C−S bond cleavage products. Group 3 includes 15 standard peptide fragments without a disulfide bond. Some of them are highly abundant, such as the singly charged y_6 of P1 and $y(2)_{11}$ of P2. This group represents common backbone fragments, useful for sequencing both P1 and P2.

Key Features of Disulfide Linkage Fragmentation Patterns. Low-energy collision dissociation of disulfide-containing peptides generates product ions, which contain intact, cleaved, and no disulfide bonds. Cleavage of an S−S bond leads to the formation of fragments containing free Cys or Cys thioaldehyde (−2 Da) while cleavage of the C−S bond yields fragments with Cys persulfide (+32 Da) or dehydroalanine (−34 Da). The dissociation of amide bonds can produce typical b/y ions, with or without intact and cleaved disulfide bonds.

We conducted a comprehensive analysis of disulfide bonds under fragmentation by low energy collision dissociation for 15 multiply charged disulfide-linked peptide ions (Table 4). They include ions of tryptic, semitryptic, and miscleaved peptides. Different types of fragment ions arising from the disulfide bond were examined using spectra acquired at NCE values of 16%, 20%, and 24% for HCD as well as 35% for FT-CID. The last column shows the number of replicate spectra used in the calculation of median values. Three types of fragmentation patterns were observed, (1) amide bond cleavage products containing an intact disulfide bond are predominant fragments; (2) disulfide bond cleavages leading to minor yet extensive fragmentation; and (3) double cleavage products are common and significant fragments. These key features are more fully described below.

Key Feature 1: Fragments Containing an Intact Disulfide Bond Predominate under HCD and FT-CID. For the purpose of finding major fragments in disulfide spectra, we examined a variety of product ions and divided them into three categories, as shown in Figure 6. One involves fragments containing an intact disulfide bond, labeled DSB fragments in this analysis, the second are fragments not containing Cys, labeled non-DSB fragments, and the last set includes all unassigned peaks.

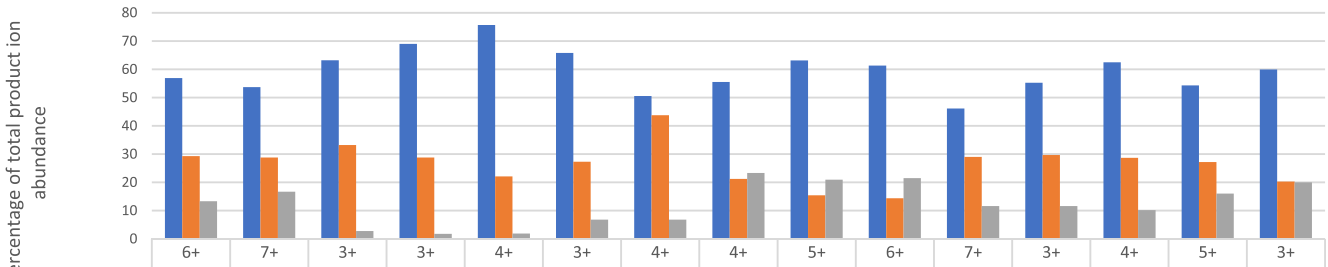
DSB fragments make the largest relative contribution in the spectra for all peptides studied, with each yielding from 46 to 76% of total product ion intensity. Non-DSB fragments from peptide bond cleavage accounted for 14 to 44% of total ion intensity in the spectra, providing complementary sequencing information to DSB fragments. The unassigned peaks appeared

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

Table 4. Selected Disulfide-Linked Peptide Ions with Varied Amino Acid Compositions and Charge States that Were Studied for Fragmentation Patterns under HCD and FT-CID

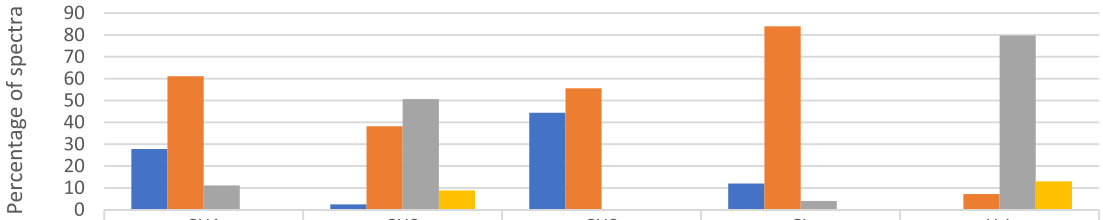
# bond	bond name	Cys position in protein	disulfide-linked peptide sequence	peptide class	# ion	charge	precursor *m/z*	replicate spectra
2	C_H1	147−203	STSGGTAALGCLVK_**SS**_DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYS	tryptic	1	6	1320.494	42
			LSSVVTVPSSSLGTQTYICNVNHKPSNTK		2	7	1131.996	157
3	C_H2	264−324	TPEVTCVVVDVSHEDPEVK_**SS**_CK	tryptic	3	3	777.04	424
					4	4	583.032	955
			*EVTCVVVDVSHEDPEVK_**SS**_CK	semi-tryptic	5	3	711.006	18
			*TPEVTCVVVDVSHED_**SS**_CK		6	3	625.854	20
					7	4	469.717	14
4	C_H3	370−428	NQVSLTCLVK_**SS**_WQQGNVFSCSVMHEALHNHYTQK	tryptic	8	4	962.213	149
					9	5	769.972	582
					10	6	641.811	41
					11	7	550.268	29
6	C_L	133−193	SGTASVVCLLNNFYPR_**SS**_VYACEVTHQGLSSPVTK	tryptic	12	3	1186.257	17
					13	4	889.945	277
					14	5	712.157	475
7	H-L	223−213	SCDK_**SS**_SFNRGEC	missed cleavage	15	3	421.169	30

[a]Notes: The symbol * indicates semitryptic peptides. # bond and bond name: sequential number and name of bonds given by this work. Cys position in protein: Cys residue in protein sequence. Peptide class: tryptic digest products. # ion: sequential number of selected ions. Charge: charge states. Precursor *m/z:* the mass/charge for peptide ions in MS1. *Replicate spectra:* repeat spectra identifications.



	6+ 1320.494 (1)	7+ 1131.996 (2)	3+ 711.006 (3)	3+ 777.04 (4)	4+ 583.032 (5)	3+ 625.954 (6)	4+ 469.717 (7)	4+ 962.213 (8)	5+ 769.972 (9)	6+ 641.811 (10)	7+ 550.268 (11)	3+ 1186.257 (12)	4+ 889.945 (13)	5+ 712.157 (14)	3+ 421.169 (15)
DSB Fragment	57	54	63	69	76	66	51	56	63	61	46	55	62	54	60
Non-DSB fragment	29	29	33	29	22	27	44	21	15	14	29	30	29	27	20
Unassigned	13	17	3	2	2	7	7	23	21	22	12	12	10	16	20

Figure 6. Comparison of three ion categories from 15 disulfide-linked peptide ions (Table 4) by percent ion intensity. DSB fragments: ions linked by a disulfide bond. Non-DSB fragments: ions not containing a disulfide bond. Unassigned: unassigned ions.



	CH1	CH2	CH3	CL	H-L
pk=0	28	2	44	12	0
pks<=3	61	38	56	84	7
pks BTW 4 and 7	11	51	0	4	80
pks up to 15	0	9	0	0	13

Figure 7. Distribution of disulfide-bond cleavages among five disulfide linkages obtained from 3200 high-quality MS2 spectra. Pks refers to fragment ions arising from disulfide cleavage.

to primarily arise from low-abundance fragments that cannot be recognized due to the lack of isotopic data.

Key Feature 2: Extensive Disulfide-Bond Cleavage Observed at Low Levels. Disulfide-bond fragmentation involving Cys side chain breakage (i.e., C−S and S−S bonds) complements disulfide connectivity information. Such product ions have been reported as rare compared to amide bond breakage via the gas phase fragmentation of disulfide-linked peptides by low-energy collision dissociation.[27,28] In contrast, we observed extensive cleavage products, although at

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

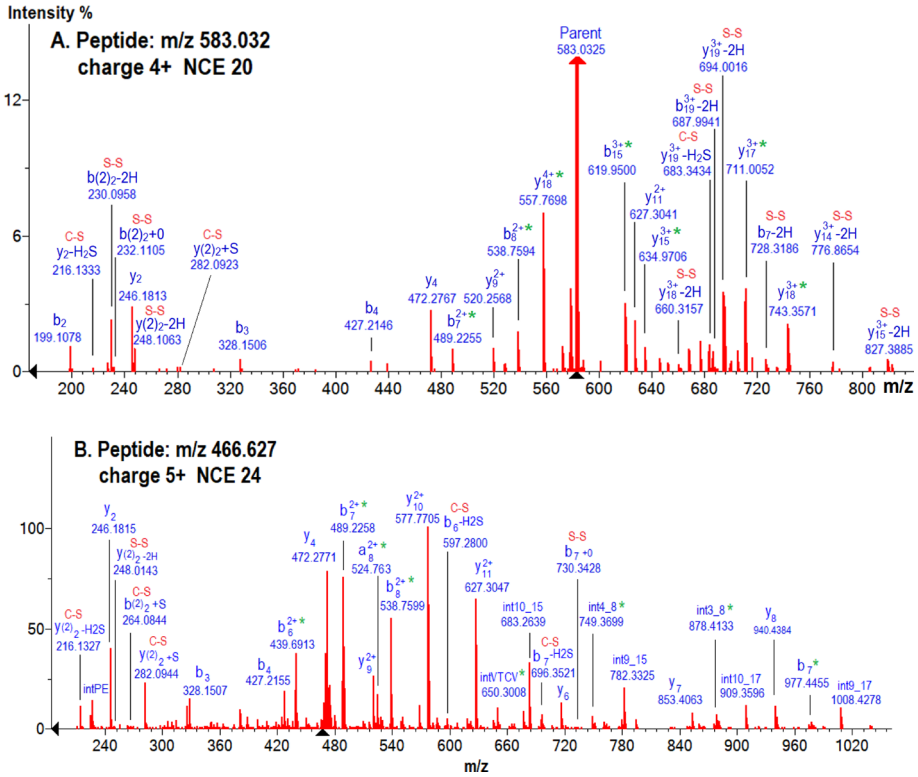


Figure 8. MS2 analysis of the charge/energy-dependence changes in HCD fragmentation of S−S and C−S bonds of disulfide-linked peptide, TPEVTCVVVDVSHEDPEVK_**SS**_CK at (A) *m*/*z* 583.032 of the charge 4+ ion at an NCE of 20 and (B) *m*/*z* 466.627 of the charge 5+ ion at an NCE of 24. The intensity axis is scaled down to an intensity range from 0 to 12% relative to the base peak to show low levels of disulfide bond cleavage in the HCD spectrum. * fragments contain an intact disulfide bond. C−S and S−S indicate cleavage from disulfide bonds.

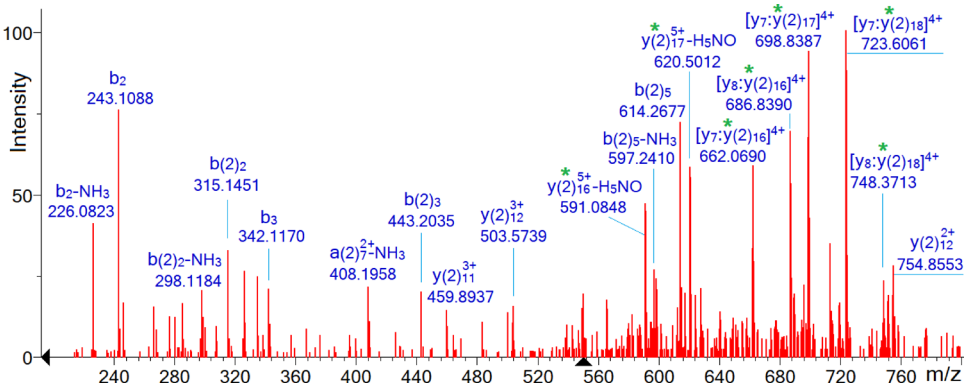


Figure 9. HCD spectrum of the charge 7+ disulfide-linked peptide ion, NQVSLTCLVK_SS_WQQGNVFSCSVMHEALHNHYTQK, at NCE = 24 shows dominant double cleavage that generated a number of major internal fragments containing a disulfide bond (in square brackets []). * fragments contain an intact disulfide bond.

low levels, arising from both S−S and C−S bonds, including free Cys, Cys thioaldehyde (−2 Da), Cys persulfide (+32 Da), and dehydroalanine (−34 Da).

Figure 7 shows the distribution of observed disulfide cleavage products resulting from over 3200 good-quality MS2 spectra of the five disulfide bonds studied in Table 4. The spectra were divided into four groups in terms of the number of disulfide cleavage products observed (e.g., peaks = 0 for no peaks from disulfide cleavage; ≤3; between 4 to 7; and >7). The percentages of spectra in each group over total replicate spectra (Column 9 of Table 4) were calculated for each bond linkage. These data indicate that most of the linkage sites produced up to three fragments resulting from disulfide fragmentation, while two disulfide linkages (C_H2 and H−L) yield prevalent disulfide bond cleavage. Over 93% (H−L) and 60% (C_H2) of their total spectra contain 4 to 15 cleavage products arising from C−S bonds and S−S bonds. This is because the two sites are made up of smaller peptides: SCDK_**SS**_SFNRGEC and TPEVTCVVVDVSHED-PEVK_**SS**_CK and its semitryptic peptides, which are more prone to undergoing S−S and C−S bond cleavages. In our data, fragments resulting from disulfide cleavage were typically observed in low abundance; for example, the summed abundance of the C−S and S−S bond cleavage is from 1 to 5% of the total product ion intensity.

We observed more ions arising from S−S bond cleavage at lower charge states, whereas more ions from C−S bond cleavage were observed at higher charge states. Figure 8

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

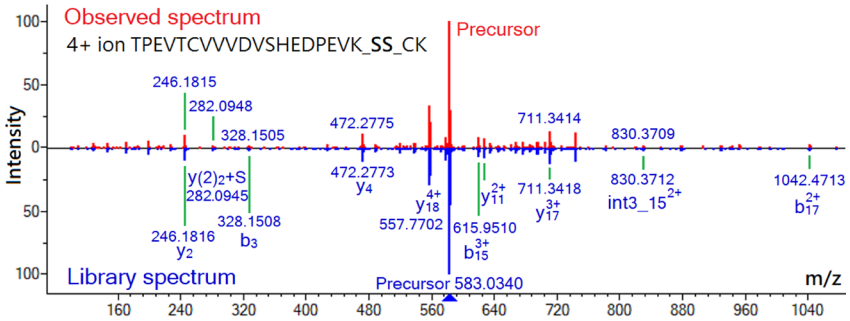


Figure 10. Head-to-tail plot of experimental and reference HCD tandem mass spectra of an ion containing a native disulfide bond C_H2. It shows an experimental ESI-MS/MS spectrum acquired from 2D LC−MS/MS analysis of reduced human plasma reference material NIST SRM 1950 matching to the charge 4+ disulfide-linked peptide spectrum in the library, "TPEVTCVVVDVSHEDPEVK_**SS**_CK". The library spectral similarity score is 930 (1000 as the highest).

demonstrates the charge/energy-dependent changes in the HCD fragmentation of S−S and C−S bonds in the charges 4+ and 5+ TPEVTCVVVDVSHEDPEVK_**SS**_CK. Their fully annotated MS/MS peak lists can be found in Table S3A,B. More intense S−S bond cleavage is evident in charge 4+ at HCD NCE = 20 in Figure 8A, while more C−S bond cleavage products are seen in the charge 5+ at HCD NCE = 24 in Figure 8B. This observation suggests that (a) the number of mobile protons that are required to induce dissociation[29] appears to have a significant effect on C−S bond cleavage and (b) with fewer mobile protons and lower NCE values, the S−S bond may tend to yield more cleavage as compared to the C−S bond.

Key Feature 3: Double Cleavage of Disulfide-Linked Peptide Ions Is Common. They are frequently observed for all disulfide peptides investigated in this study. The percent abundances of double cleavage ions range from 10 to 35% of total product ion intensity. The formation of disulfide-bonded double cleavage ions can be quite complex. They can be grouped into two categories. One involves either double cleavage of one peptide (internal ions) or a cleavage from each of two component peptides, and the other results from the cleavage of both peptide backbone and disulfide bonds (see Group 2 of Table 1). Following backbone fragmentation, further fragmentation of the resultant fragments may be derived by the breakage of the disulfide bond. The number and abundance level of double cleavage products are energy- and charge-dependent. As expected, higher collision energy and charge values appear to increase the portion of double cleavages. As illustrated in Figure 9, dominant double cleavage products were observed in the HCD spectrum of the charge 7+ C_H3 disulfide peptide. The large proportion of multiply charged major products (shown in green color) observed in the high *m/z* region can be assigned as arising from the sequential cleavage of both component peptides. The product ions were analyzed for the peptide in 4+, 5+, 6+, and 7+ charge states, showing that the relative abundances of double cleavages at each charge are 1, 4, 10, and 22%, respectively.

Consensus Mass Spectral Library. We compiled a library of consensus mass spectra of all disulfide-linked peptide ions acquired in this study from the tryptic digest of the NISTmAb. Each of them arises from multiple spectra of the same peptide ion at a given collision energy. They were derived from the LC−MS/MS analysis of 82 runs under non-reducing and partial reducing conditions. Spectra were included only if at least 60% of total peak abundance could be confidently annotated. Over 10,000 replicate spectra underlie these consensus mass spectra. The spectral library contains 702 consensus spectra for 149 native and 234 scrambled disulfide-bonded peptide ions (e.g., sequence/modification/charge). Each entry in the library provides information on protein, Cys position in sequence, disulfide bonds, precursor *m/z*, normalized collision energy, identification score, percentages of assigned/unassigned abundances, and number of replicate spectra used for creating the consensus spectrum. Peaks in each spectrum are thoroughly annotated. This library is freely available at https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:disulfidepeptides.

Use of the Library of Disulfide-Linked Spectra in General Antibody Analysis. One potential utility of the mass spectral library is to evaluate the effectiveness of reduction in peptide mapping sample preparation. We demonstrate this using LC−MS/MS results from three different datasets, all of which were performed under conventional reducing conditions: Humira (a commercially available mAb drug),[30] human plasma (a NIST reference material, SRM 1950),[31,32] and NISTmAb prepared without second reduction.[5] Note that the former two samples were prepared with a second reduction step. This analysis was carried out using the NIST MSQC Pipeline.[33] It identified 5, 11, and 29 disulfide-linked peptide spectra in Humira, NIST SRM 1950, and NISTmAb, respectively, using a match score threshold of 450 and mass error of less than 10 ppm. All identifications were manually confirmed and are shown in Table S4. Figure 10 shows an example of a charge 4+ peptide ion containing a native disulfide linkage, C_H2 detected in the 2D LC−MS/MS analysis of reduced human plasma reference material. In the reduced Humira experiments, we found two native disulfide bonds. In the NIST 1950 plasma reference material, two native and two scrambled linkages were found. Furthermore, in the NISTmAb analysis, which did not include a second reduction step, all five native disulfide bonds of constant regions as well as four scrambled disulfide bonds were detected. These observations suggest that (1) low levels of disulfide-linked ions can be identified even after reduction in LC−MS analysis and (2) the second reduction step often implemented in standard peptide mapping experiments[34] is justified as fewer disulfide bonds were detected when this extra reduction step was done in the Humira and NIST SRM 1950 analyses.

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629

CONCLUSIONS

In this study, we performed a detailed analysis of disulfide-linked peptides generated in the tryptic digestion of a reference monoclonal antibody, the NISTmAb. This was done both for understanding the fragmentation of these complex disulfide-linked peptides and for creating a reference mass spectral library for helping others to readily identify these peptides.

For this purpose, a data analysis workflow of combining MS and MS/MS was developed, using MS data to assign all possible disulfide-containing peptide spectra and then MS/MS data to resolve any ambiguous identifications. This approach helps circumvent the tedious manual inspection of MS/MS spectra necessary for the protein disulfide bond profile with high sensitivity and accuracy. Because the method searches the complete combination of all possible disulfide bonds in proteins based on their sequence, all native as well as non-native disulfide-containing peptides are covered in searches. Hence, this approach could be useful for routine disulfide mapping in general proteomics applications. In our study, this led to the identification and annotation of 428 disulfide-linked peptide ions with a variety of origins: fully tryptic, missed cleavages, irregular cleavages, containing Met/Trp oxidation, and metallated adducts. Based on these results, we effectively characterized the relative level of all the nine native disulfide bonds and detected 86 additional disulfide linkages from shuffling in the LC−MS/MS analysis of the NISTmAb. As expected, the S−S bond shuffling at pH 8 was greatly minimized under neutral condition pH 7.

This analysis highlights the unique difficulties in a comprehensive analysis of disulfide-linked peptides especially where reduction is incomplete, either by design or not, and where fragmentation lacks readily identifiable diagnostic fragment ions. Transforming these results into a spectral library will enable others to efficiently identify both native and scrambled disulfide peptides in routine proteomics analyses containing IgG1 antibodies. We showed that with a reference library of confidently assigned spectra, the unambiguous assignment of tandem mass spectra (MS/MS) to disulfide-linked peptide sequences can be easily done. We also demonstrated that this library can identify disulfide bonds of IgG1 proteins in reduced human serum digests, thereby providing a unique measure of the effectiveness of reduction methods in proteomics experiments.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00823.

Schematic diagram of the NISTmAb structure as a humanized IgG1κ protein (Figure S1); comparison of 15 tryptic digestion protocols using peptide class analysis (Figure S2); selected ion chromatograms of oxidized disulfide-bonded peptides (Figure S3); parameters used in 15 protocol variations during denaturing, reduction, and alkylation of 1 mg of the NISTmAb (Table S1); list of scrambled and native disulfide-bonded peptides identified from controlled experiments of the 18 h digestion under partial/non-reduction at pH 7 (Table S2); annotated peak list for the HCD tandem mass spectrum of the charge 4+ and 5+ disulfide-linked peptide, TPEVTCVVVDVSHEDPEVK_**SS**_CK (Table S3); library search of native and scrambled disulfide-bonded peptides from three LC−MS/MS analyses under reduced conditions (Table S4) (xlsx)

AUTHOR INFORMATION

Corresponding Author

Qian Dong − *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States;* orcid.org/0000-0001-6466-6045; Phone: 301-975-2569; Email: qian.dong@nist.gov

Authors

Xinjian Yan − *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States;* orcid.org/0000-0003-3204-7420

Yuxue Liang − *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States*

Sanford P. Markey − *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States*

Sergey L. Sheetlin − *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States*

Concepcion A. Remoroza − *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States;* orcid.org/0000-0003-1540-1635

William E. Wallace − *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States*

Stephen E. Stein − *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.0c00823

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported solely with the NIST funds. Certain commercial instruments or materials are identified in this paper to specify the experimental procedure adequately. Such an identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or instruments identified are necessarily the best available for the purpose. The mass spectrometry proteomics data have been deposited to the ProteomeXchange[35] Consortium via the PRIDE[36] partner repository with the dataset identifier PXD023358.

ABBREVIATIONS PAGE

C−S, carbon−sulfur bond; Cys, cysteine; DTT, dithiothreitol; FT, Fourier transform mass spectrometry; HCI, hydrochloric acid; HCD, higher-energy collision dissociation; IAM, iodoacetamide; IgG1, subclass 1 of immunoglobulin G (IgG1, 2, 3, and 4); LC−MS, liquid chromatography−mass spectrometry; LC−MS/MS, liquid chromatography−tandem

mass spectrometry; mAb, monoclonal antibody; Met, methionine; MS1, full MS scan; MS2, tandem MS scan; MWCO, molecular weight cut-off; NCE, normalized collisional energies; NEM, *N*-ethylmaleimide; NIST, National Institute of Standards and Technology; NISTmAb, a reference material of humanized IgG1k monoclonal antibody (RM 8671); NIST MSQC, an in-house software platform for analysis of shotgun proteomics; RPLC, reversed phase liquid chromatography; RSLC, rapid separation liquid chromatography; S−S, sulfur−sulfur bond; TCEP, tris(2-carboxyethyl)phosphine; TRIS, tris-hydroxymethyl-aminomethane; Trp, tryptophan; UPLC, ultra-high performance liquid chromatography

■ REFERENCES

(1) Sevier, C. S.; Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 836−847.

(2) Liu, H.; May, K. Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *mAbs* **2012**, *4*, 17−23.

(3) Bechtel, T. J.; Weerapana, E. From structure to redox: The diverse functional roles of disulfides and implications in disease. *Proteomics* **2017**, *17*, 1600391.

(4) National Institute of Standards and Technology. *RM 8671 − NISTmAb, Humanized IgG1k Monoclonal Antibody*. 2016, https://www-s.nist.gov/srmors/view_detail.cfm?srmD8671.

(5) Dong, Q.; Liang, Y.; Yan, X.; Markey, S. P.; Mirokhin, Y. A.; Tchekhovskoi, D. V.; Bukhari, T. H.; Stein, S. E. The NISTmAb tryptic peptide spectral library for monoclonal antibody characterization. *mAbs* **2018**, *10*, 354−369.

(6) Dong, Q.; Yan, X.; Liang, Y.; Stein, S. E. In-Depth Characterization and Spectral Library Building of Glycopeptides in the Tryptic Digest of a Monoclonal Antibody Using 1D and 2D LC-MS/MS. *J. Proteome Res.* **2016**, *15*, 1472−1486.

(7) Gorman, J. J.; Wallis, T. P.; Pitt, J. J. Protein disulfide bond determination by mass spectrometry. *Mass Spectrom. Rev.* **2002**, *21*, 183−216.

(8) Tsai, P. L.; Chen, S. F.; Huang, S. Y. Mass spectrometry-based strategies for protein disulfide bond identification. *Rev. Anal. Chem.* **2013**, *32*, 257−268.

(9) Lakbub, J. C.; Shipman, J. T.; Desaire, H. Recent mass spectrometry-based techniques and considerations for disulfide bond characterization in proteins. *Anal. Bioanal. Chem.* **2018**, *410*, 2467−2484.

(10) Monahan, F. J.; German, J. B.; Kinsella, J. E. Effect of pH and temperature on protein unfolding and thiol/disulfide interchange reactions during heat-induced gelation of whey proteins. *J. Agric. Food Chem.* **1995**, *43*, 46−52.

(11) Kerr, J.; Schlosser, J. L.; Griffin, D. R.; Wong, D. Y.; Kasko, A. M. Steric effects in peptide and protein exchange with activated disulfides. *Biomacromolecules* **2013**, *14*, 2822−2829.

(12) Cui, C.; Liu, T.; Chen, T.; Lu, J.; Casaren, I.; Lima, D. B.; Carvalho, P. C.; Beuve, A.; Li, H. Comprehensive identification of protein disulfide bonds with pepsin/trypsin digestion, Orbitrap HCD and Spectrum Identification Machine. *J. Proteomics* **2019**, *198*, 78−86.

(13) Cramer, C. N.; Kelstrup, C. D.; Olsen, J. V.; Haselmann, K. F.; Nielsen, P. K. Generic Workflow for Mapping of Complex Disulfide Bonds Using In-Source Reduction and Extracted Ion Chromatograms from Data-Dependent Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 8202−8210.

(14) Wiesner, J.; Resemann, A.; Evans, C.; Suckau, D.; Jabs, W. Advanced mass spectrometry workflows for analyzing disulfide bonds in biologics. *Expert Rev. Proteomics* **2015**, *12*, 115−123.

(15) Reynolds, K. J.; Yao, X.; Fenselau, C. Proteolytic ^{18}O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *J. Proteome Res.* **2002**, *1*, 27−33.

(16) Zhang, W.; Marzilli, L. A.; Rouse, J. C.; Czupryn, M. J. Complete disulfide bond assignment of a recombinant immunoglobulin G4 monoclonal antibody. *Anal. Biochem.* **2002**, *311*, 1−9.

(17) Formolo, T.; Ly, M.; Levy, M.; Kilpatrick, L.; Lute, S.; Phinney, K.; Marzilli, L.; Brorson, K.; Boyne, M.; Davis, D.; Schiel, J. Determination of the NISTmAb Primary Structure. In *State-of-the-art and emerging technologies for therapeutic monoclonal antibody characterization volume 2. Biopharmaceutical characterization: the NISTmAb case study*; ACS symposium series, vol 1201 , American Chemical Society: 2015; 1−62.

(18) Jones, M. D.; Patterson, S. D.; Lu, H. S. Determination of disulfide bonds in highly bridged disulfide-linked peptides by matrix-assisted laser desorption/ionization mass spectrometry with post-source decay. *Anal. Chem.* **1998**, *70*, 136−143.

(19) Cramer, C. N.; Kelstrup, C. D.; Olsen, J. V.; Haselmann, K. F.; Nielsen, P. K. Complete Mapping of Complex Disulfide Patterns with Closely-Spaced Cysteines by In-Source Reduction and Data-Dependent Mass Spectrometry. *Anal. Chem.* **2017**, *89*, 5949−5957.

(20) Zhang, W.; Czupryn, M. J. Free sulfhydryl in recombinant monoclonal antibodies. *Biotechnol. Prog.* **2002**, *18*, 509−513.

(21) Shen, Y.; Zeng, L.; Zhu, A.; Blanc, T.; Patel, D.; Pennello, A.; Bari, A.; Ng, S.; Persaud, K.; Kang, Y. K.; Balderes, P.; Surguladze, D.; Hindi, S.; Zhou, Q.; Ludwig, D. L.; Snavely, M. Removal of a C-terminal serine residue proximal to the inter-chain disulfide bond of a human IgG1 lambda light chain mediates enhanced antibody stability and antibody dependent cell-mediated cytotoxicity. *mAbs* **2013**, *5*, 418−431.

(22) Switzar, L.; Nicolardi, S.; Rutten, J. W.; Oberstein, S. A. J. L.; Aartsma-Rus, A.; van der Burgt, Y. E. M. In-Depth Characterization of Protein Disulfide Bonds by Online Liquid Chromatography-Electrochemistry-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 50−58.

(23) Zhang, H.-M.; McLoughlin, S. M.; Frausto, S. D.; Tang, H.; Emmett, M. R.; Marshall, A. G. Simultaneous reduction and digestion of proteins with disulfide bonds for hydrogen/deuterium exchange monitored by mass spectrometry. *Anal. Chem.* **2010**, *82*, 1450−1454.

(24) Liu, H.; Chumsae, C.; Gaza-Bulseco, G.; Hurkmans, K.; Radziejewski, C. H. Ranking the susceptibility of disulfide bonds in human IgG1 antibodies by reduction, differential alkylation, and LC-MS analysis. *Anal. Chem.* **2010**, *82*, 5219−5226.

(25) Yan, X.; Markey, S. P.; Marupaka, R.; Dong, Q.; Cooper, B. T.; Mirokhin, Y. A.; Wallace, W. E.; Stein, S. E. Mass Spectral Library of Acylcarnitines Derived from Human Urine. *Anal. Chem.* **2020**, *92*, 6521−6528.

(26) Dong, Q.; Yan, X.; Kilpatrick, L. E.; Liang, Y.; Mirokhin, Y. A.; Roth, J. S.; Rudnick, P. A.; Stein, S. E. Tandem mass spectral libraries of peptides in digests of individual proteins: Human Serum Albumin (HSA). *Mol. Cell. Proteomics* **2014**, *13*, 2435−2449.

(27) Lioe, H.; OrsHair, R. A. J. A novel salt bridge mechanism highlights the need for nonmobile proton conditions to promote disulfide bond cleavage in protonated peptides under low-energy collisional activation. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1109−1123.

(28) Clark, D. F.; Go, E. P.; Toumi, M. L.; Desaire, H. Collision induced dissociation products of disulfide-bonded peptides: ions result from the cleavage of more than one bond. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 492−498.

(29) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399−1406.

(30) Morsa, D.; Baiwir, D.; La Rocca, R.; Zimmerman, T. A.; Hanozin, E.; Grifnée, E.; Longuespée, R.; Meuwis, M.-A.; Smargiasso, N.; De Pauw, E.; Mazzucchelli, G. Multi-Enzymatic Limited Digestion: The Next-Generation Sequencing for Proteomics? *J. Proteome Res.* **2019**, *18*, 2501−2513.

(31) Phinney, K. W.; Ballihaut, G.; Bedner, M.; Benford, B. S.; Camara, J. E.; Christopher, S. J.; Davis, W. C.; Dodder, N. G.; Eppe, G.; Lang, B. E.; Long, S. E.; Lowenthal, M. S.; McGaw, E. A.; Murphy, K. E.; Nelson, B. C.; Prendergast, J. L.; Reiner, J. L.; Rimmer, C. A.; Sander, L. C.; Schantz, M. M.; Sharpless, K. E.; Sniegoski, L. T.; Tai, S. S.-C.; Thomas, J. B.; Vetter, T. W.; Welch, M. J.; Wise, S. A.; Wood, L. J.; Guthrie, W. F.; Hagwood, C. R.; Leigh, S. D.; Yen, J. H.; Zhang, N.-F.; Chaudhary-Webb, M.; Chen, H.; Fazili, Z.; LaVoie, D. J.;

McCoy, L. F.; Momin, S. S.; Paladugula, N.; Pendergrast, E. C.; Pfeiffer, C. M.; Powers, C. D.; Rabinowitz, D.; Rybak, M. E.; Schleicher, R. L.; Toombs, B. M.; Xu, M.; Zhang, M.; Castle, A. L. Development of a Standard Reference Material for metabolomics research. *Anal. Chem.* **2013**, *85*, 11732−11738.

(32) Simón-Manso, Y.; Lowenthal, M. S.; Kilpatrick, L. E.; Sampson, M. L.; Telu, K. H.; Rudnick, P. A.; Mallard, W. G.; Bearden, D. W.; Schock, T. B.; Tchekhovskoi, D. V.; Blonder, N.; Yan, X.; Liang, Y.; Zheng, Y.; Wallace, W. E.; Neta, P.; Phinney, K. W.; Remaley, A. T.; Stein, S. E. Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. *Anal. Chem.* **2013**, *85*, 11725−11731.

(33) National Institute of Standards and Technology. *The NIST libraries of peptide tandem mass spectra and available windows software for peptide mass spectral libraries*. Washington (DC): United States Department of Commerce; 2016, http://chemdata.nist.gov/dokuwiki/doku.php?idDpeptidew:cdownload, and http://chemdata.nist.gov/dokuwiki/doku.php?idDpeptidew:msqcpipeline. Aug 8.

(34) Rebecchi, K. R.; Go, E. P.; Xu, L.; Woodin, C. L.; Mure, M.; Desaire, H. A general protease digestion procedure for optimal protein sequence coverage and post-translational modifications analysis of recombinant glycoproteins: application to the characterization of human lysyl oxidase-like 2 glycosylation. *Anal. Chem.* **2011**, *83*, 8484−8491.

(35) Deutsch, E. W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J. J.; Kundu, D. J.; García-Seisdedos, D.; Jarnuczak, A. F.; Hewapathirana, S.; Pullman, B. S.; Wertz, J.; Sun, Z.; Kawano, S.; Okuda, S.; Watanabe, Y.; Hermjakob, H.; MacLean, B.; MacCoss, M. J.; Zhu, Y.; Ishihama, Y.; Vizcaíno, J. A. The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **2020**, *48*, D1145−D1152.

(36) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442−D450.

https://dx.doi.org/10.1021/acs.jproteome.0c00823
J. Proteome Res. 2021, 20, 1612−1629