



HAL
open science

De Novo Sequencing of Antibody Light Chain Proteoforms from Patients with Multiple Myeloma

Mathieu Dupré, Magalie Duchateau, Rebecca Sternke-Hoffmann, Amelie Boquoi, Christian Malosse, Roland Fenk, Rainer Haas, Alexander K Buell, Martial Rey, Julia Chamot-Rooke

► **To cite this version:**

Mathieu Dupré, Magalie Duchateau, Rebecca Sternke-Hoffmann, Amelie Boquoi, Christian Malosse, et al.. De Novo Sequencing of Antibody Light Chain Proteoforms from Patients with Multiple Myeloma. *Analytical Chemistry*, 2021, 93 (30), pp.10627-10634. 10.1021/acs.analchem.1c01955 . pasteur-03395532

HAL Id: pasteur-03395532

<https://pasteur.hal.science/pasteur-03395532>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

De Novo Sequencing of Antibody Light Chain Proteoforms from Patients with Multiple Myeloma

Mathieu Dupré,[§] Magalie Duchateau,[§] Rebecca Sternke-Hoffmann, Amelie Boquoi, Christian Malosse, Roland Fenk, Rainer Haas, Alexander K. Buell, Martial Rey, and Julia Chamot-Rooke*



Cite This: *Anal. Chem.* 2021, 93, 10627–10634



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: In multiple myeloma diseases, monoclonal immunoglobulin light chains (LCs) are abundantly produced, with, as a consequence in some cases, the formation of deposits affecting various organs, such as the kidney, while in other cases remaining soluble up to concentrations of several $\text{g}\cdot\text{L}^{-1}$ in plasma. The exact factors crucial for the solubility of LCs are poorly understood, but it can be hypothesized that their amino acid sequence plays an important role. Determining the precise sequences of patient-derived LCs is therefore highly desirable. We establish here a novel *de novo* sequencing workflow for patient-derived LCs, based on the combination of bottom-up and top-down proteomics without database search. PEAKS is used for the *de novo* sequencing of peptides that are further assembled into full length LC sequences using ALPS. Top-down proteomics provides the molecular masses of proteoforms and allows the exact determination of the amino acid sequence including all posttranslational modifications. This pipeline is then used for the complete *de novo* sequencing of LCs extracted from the urine of 10 patients with multiple myeloma. We show that for the bottom-up part, digestions with trypsin and *Nepenthes* digestive fluid are sufficient to produce overlapping peptides able to generate the best sequence candidates. Top-down proteomics is absolutely required to achieve 100% final sequence coverage and characterize clinical samples containing several LCs. Our work highlights an unexpected range of modifications.

INTRODUCTION

Multiple myeloma (MM) is a malignancy of plasma cells characterized by a clonal expansion of an abnormal B-cell.^{1,2} B-cells accumulate in the bone marrow and secrete large amounts of monoclonal light chains (LCs) additionally to the complete immunoglobulin. Fifteen percent of the MM patients produce exclusively LCs.³ The 25 kDa LC proteins are usually either excreted or degraded by the kidney, but high monoclonal quantities and low renal clearance can induce deposition in the kidney's extracellular matrix.⁴ The deposits contain diverse LC aggregates, which can lead to various diseases, such as LC deposition disease (LCDD), where the formed aggregates have an amorphous nature,^{5,6} and amyloid LC (AL) amyloidosis, where aggregates consist of amyloid fibrils.⁷

Currently, the *in vivo* aggregation behavior of a particular monoclonal LC found in the blood of a patient with LC disease cannot be predicted. To better understand the factors affecting the solubility of LCs and their aggregation propensity, their biophysical properties have recently been explored.^{8–13} Unfortunately no clear-cut conclusion could be drawn. To achieve this goal, a database including the biophysical properties of LCs and their sequence is essential. Sequencing monoclonal

LCs is challenging, because they all have a unique amino acid sequence determined by somatic recombination and various mutations.^{14,15} This sequence diversity translates into a diverse clinical picture,¹⁶ and thus, the mechanisms behind a particular disease are hard to decipher. LCs consist of a N-terminal variable region (v-region), which is capable of recognizing the antigen, and a C-terminal constant region (c-region), which specifies the effector function of the molecule.¹⁷ Important for the antigen binding site are three hypervariable loops, so-called CDRs (complementarity-determining regions), present in the variable region. The diversity of sequences is created by somatic recombination of variable (V) and joining (J) gene segments (V–J combination) during the early stages of B-cell maturation.¹⁸ LCs exist in two isotypes: kappa (κ) and lambda (λ).¹⁹

Received: May 8, 2021

Accepted: July 12, 2021

Published: July 22, 2021



The sequence of a LC can be partially determined by sequencing the RNA of the producing B-cell clone.^{2,20,21} However, a bone marrow aspiration is allowed solely to improve diagnosis and treatment of the patients because of ethical reasons. This cannot be achieved for the majority of studies. Furthermore, this approach gives no insight into potential posttranslational modifications (PTMs), which can be important for their solubility.

In recent years, mass spectrometry (MS) has been extensively used for antibody characterization.^{22,23} The most common approach is bottom-up proteomics (BUP), which relies on the protein digestion and LC–MS/MS analysis of peptides. BUP can provide high sequence coverage confirming the sequence of recombinant antibodies and the presence of expected PTMs in particular when using a combination of enzymes.²⁴ However, the situation is much more difficult for unknown antibodies for which *de novo* sequencing is required.^{25–27} Several papers have shown that a combination of BUP and intact mass profiling can be of great help.^{28,29} For LCs, which are in the 25 kDa range, the use of top-down proteomics (TDP) based on the fragmentation of intact proteins is also possible.^{30,31} For instance, this approach has been used by Marshall and co-workers to classify plasma cell disorders (including amyloidosis) from the analysis of monoclonal immunoglobulin LCs in human serum.³² A 21 T FT-ICR mass spectrometer was employed in combination with a database-aided *de novo* MS/MS algorithm. However, achieving a complete sequence at a single amino acid resolution (i.e., achieving fragment ions for each peptide bond) is a difficult task.^{33–35} In Marshall's paper, only 70% sequence coverage was obtained on LCs.³² A solution is to combine BUP and TDP and use dedicated software tools such as TBnovo.³⁶ However, sequence gaps often remain because of the lack of fragment ions both in BUP and TDP.

We, therefore, developed here a complete *de novo* sequencing workflow for the characterization of patient-derived LC proteoforms based on a combination of BUP and TDP with specific data analysis. PEAKS is used for the *de novo* sequencing of peptides that are further assembled into full length LC sequences using ALPS. TDP provides the molecular masses of proteoforms and allows the exact determination of the amino acid sequence including all PTMs. The characterization of 10 different clinical samples covers the amino acid sequence with in most of the cases I/L distinction, all PTMs (including disulfide bonds), and the ratio between the monomeric and dimeric proteoforms in the clinical samples. To our knowledge such a deep characterization of LCs extracted from the urine of patients has never been achieved so far. In addition, our work highlights an unexpected range of modifications.

EXPERIMENTAL SECTION

Chemicals and Reagents. PBS (1×, Dulbecco's Phosphate-Buffered Saline, Gibco) was purchased from Thermo Fisher. Ammonium bicarbonate (AB), urea, Tris 1 M HCl pH 8.5 solution, iodoacetamide (IAA), Tris(2-carboxyethyl) phosphine 0.5 M solution (TCEP), and formic Acid (FA) were purchased from Sigma-Aldrich. Ethanol (70%), methanol (MeOH), and acetonitrile (ACN) were purchased from Carlo Erba. Trypsin, Lys-C, and chymotrypsin were purchased from Promega. Sep-Pak C₁₈ SPE cartridges were purchased from Waters. Pepsin columns (ref AP-PC-001s) were purchased from Affipro. *Nepenthes* digestive fluid was extracted from *Nepenthes* plants of the botanical garden of Lyon and prepared as described by Rey et al.³⁷

Ethical Considerations. All patients of whom samples were used in the study have signed an informed consent with the university hospital Düsseldorf (study number 5926R and registration ID 20170664320).

LC Sample Preparation. LCs were extracted from urine samples as described in Sternke-Hoffmann et al.¹² Samples are listed in Table S1, including their correspondence with the patients/samples described in Sternke-Hoffmann et al.¹²

Enzymatic Digestions. Each LC sample was digested with the following enzymes: trypsin, Lys-C, chymotrypsin, pepsin, and *Nepenthes* fluid. All experimental conditions are given in the Supporting Information (SI). Resulting peptides were desalted and concentrated on a Sep-Pak C₁₈ SPE cartridge. Peptides were eluted using 50% ACN and 0.1% FA. Purified peptides were lyophilized and kept at –80 °C.

LC–MS Analysis of Peptide Digests. For *de novo* sequencing, LC digests were analyzed in LC–MS/MS on a Q-Exactive Plus mass spectrometer using standard conditions (see the SI), except for the number of μ scans that was set to four to obtain high quality MS/MS data. For isoleucine/leucine discrimination, trypsin and chymotrypsin LC digests were mixed (1:1 ratio) and analyzed by LC–MS/MS on a Fusion Lumos mass spectrometer. Ions corresponding to peptides containing one or more leucine or isoleucine were added to an inclusion list to be fragmented by EThcD (MS²) or HCD (MS⁴) as previously described.^{38,39}

De Novo Peptide Sequencing and Concatenation. PEAKS Studio X was used for peptide *de novo* sequencing. The data were refined using precursor mass correction only and the chimera scan option activated. *De novo* searches were performed with 2 ppm error for precursor mass, 0.01 Da for the fragment ions with fixed carbamidomethylation (Cys) and variable oxidation (Met) as PTMs. Enzyme rules were specified for each sample, with no rules for the *Nepenthes* digestive fluid and pepsin digestions. *De novo* sequencing results were exported and sequences, local confidence, and area were used in ALPS²⁹ to concatenate the overlapping peptides. Although other values were tested, kmer from 7 to 9 were finally used to generate putative LC sequences. The theoretical masses of these sequences were compared to those obtained from the intact mass measurement of LCs to ensure correct concatenation and select the appropriate sequences.

BUP Data Analysis. The raw files obtained for the tryptic digests of each LC sample were searched with MaxQuant (parameters described in the SI). Draw Map from MSTools—Web applications was used for visualization of protein sequence coverage.⁴⁰ For disulfide bridge localization, data obtained from nonreduced nonalkylated tryptic digests were searched against the corresponding LC sequences with Mass Spec Studio⁴¹ using the CRIMP workflow. A loss of two hydrogens (–2.0156 Da) was used as a virtual cross-linker mass modification. The search parameters are provided in the SI. The most intense cross-linked peptides identified were used to assign disulfide bridges.

LC–MS Analysis of Intact LCs (TDP). Intact LCs were analyzed on an Orbitrap Fusion Lumos mass spectrometer with or without reduction/alkylation. Both MS and targeted MS/MS experiments were undertaken. To maximize sequence coverage, 12 different fragmentation conditions were used. All details are given in the SI.

TDP Data Analysis. Intact protein mass spectra were deconvoluted using Protein Deconvolution v3.0 software (Thermo-Scientific) either with the Xtract algorithm for isotopically resolved mass spectra or with the ReSpect algorithm

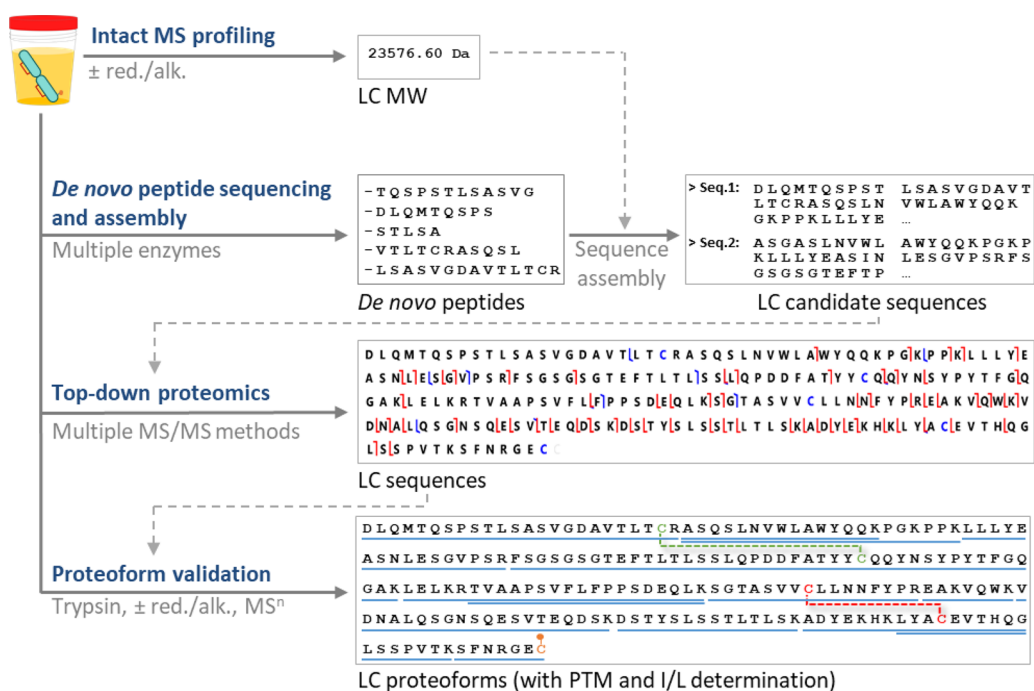


Figure 1. Combination of intact mass profiling, *de novo* peptide sequencing using multiple enzymes, TDP with multiple MS/MS method, and BUP with and without reduction/alkylation for confident identification of LC proteoforms.

for unresolved mass spectra. MS/MS spectra were deconvoluted in FreeStyle 1.6 with the Xtract algorithm. For Xtract, the following parameters were used: signal to noise ratio (S/N) of 3, fit factor of 44%, and remainder threshold of 25%. For the ReSpect algorithm, a noise rejection threshold of 95% and 20 ppm mass tolerance were used. Fragment ions produced by HCD (b, y), ETD (c, z), EThcD (b, y, c, z), or UVPD (a, a + 1, b, c, x, x + 1, y, y - 1, z) were identified using ProSight Lite v1.4 with a mass tolerance of ± 5 ppm.

RESULTS AND DISCUSSION

Development of the De Novo Sequencing Strategy.

We used the P15 sample for optimizing the four major steps of our workflow (Figure 1): intact MS profiling, *de novo* peptide sequencing using multiple enzymes, proteoform characterization using TDP, and proteoform validation using BUP, including the I/L discrimination.

Intact MS Profiling. The intact MS profiling is an important piece of information since it allows the number of proteoforms and their isotopic molecular mass to be obtained. It also provides, using a reduction/alkylation process, the information on the number of disulfide bridges. Since a slight band at 50 kDa corresponding to the potential presence of a dimer had been previously observed in the SDS-PAGE analysis of P15,¹² we used both low and high-resolution (HR) settings to allow for the mass measurement of all protein species (both monomers and dimers) present in the sample. As shown in Figure 2, two protein distributions are observed in the low-resolution MS spectrum (Figure 2A) corresponding, respectively, to molecular masses of 23,590.3 Da (major one) and 46,941.5 Da (average masses, see Table S2) with a relative intensity ratio of 10/1. At a higher resolution (120k) the signal intensity decreases and only the lowest mass is observed (Figure 2B) (23,576.60 Da, monoisotopic mass).⁴² Finally, the reduced/alkylated sample (Figure 2C) leads to a distribution shifted toward lower m/z , due to an extensive unfolding upon S–S bond reduction,

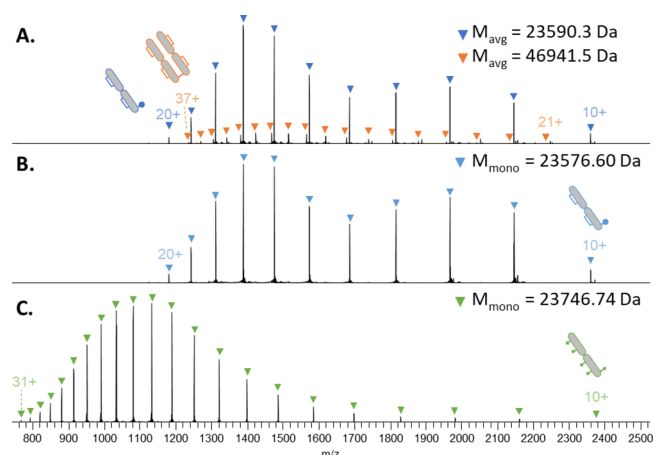


Figure 2. Intact mass spectra of (A) P15 sample using low-resolution MS (15k), (B) P15 sample using HRMS (120k), and (C) reduced/alkylated P15 sample using HRMS (120k).

corresponding to a monoisotopic mass of 23,746.74 Da. The discrepancy between the two monoisotopic measured masses highlights a Δ mass of 170.14 Da. This cannot only be explained by one or several carbamidomethylations (+57.02 Da) following the S–S bond reduction (+2.02 Da) and indicates other cysteine modifications.

De Novo Peptide Sequencing and Assembly. We first performed a digestion using four different enzymes (trypsin, Lys-C, pepsin, and *Nepenthes* digestive fluid) to maximize the generation of overlapping peptides. The LC–MS/MS method was optimized to obtain the high-quality data required for *de novo* sequencing (see the Experimental Section). For instance, the number of microscans was increased compared to a regular proteomics analysis.

The data generated were analyzed with PEAKS to obtain *de novo* sequenced peptides.⁴³ These peptides were further

assembled using ALPS, with the intact mass previously measured used as a constraint with 1 Da tolerance. The ALPS kmer parameter, which represents the number of overlapping amino acids required to concatenate two peptides, was found to be optimal between 7 and 9, depending on the LC sample. For P15, this process led to two different candidate sequences corresponding to κ LCs. They both contain five cysteines, which fits with what is expected for a LC. They differ from each other only by 1 Da and a single amino acid (N/D). One of these sequences has a theoretical molecular mass of 23,746.60 Da (monoisotopic mass for the reduced/alkylated LC) matching the experimentally measured one (23,746.74 Da) within 5.8 ppm and was thus selected as the best candidate.

This allows us to calculate the theoretical Δ mass due to reduction/alkylation, which is thus 289.14 Da. Compared to the 170.16 Da previously obtained, this leads to a residual Δ mass of 118.98 Da, which corresponds to a cysteinylolation (theoretical monoisotopic mass of 119.00 Da). This LC modification, originally described by the group of Costello⁴⁴ and confirmed a few years later by Gadgil et al.,⁴⁵ has been suggested to play an important role in the stabilization of LCs.

Top-Down Proteomics. The P15 sample was analyzed in its reduced and alkylated form in targeted LC-MS/MS using multiple activation techniques. The objective was to obtain as many complementary fragment ions as possible. This is required to increase the confidence in the assembly of *de novo* peptides and identify potential errors. As shown in Figure S1, each fragmentation method generates different fragment ions and are thus complementary. HCD and CID lead to very similar results with CID giving the best sequence coverages between both. EThcD combines the advantages of both ETD and HCD and thus provides very high sequence coverage in a single experiment (69%). Finally, UVPD leads to unique fragment ions, which makes this method useful in particular in combination with other techniques. Figure 3 shows the fragmentation map

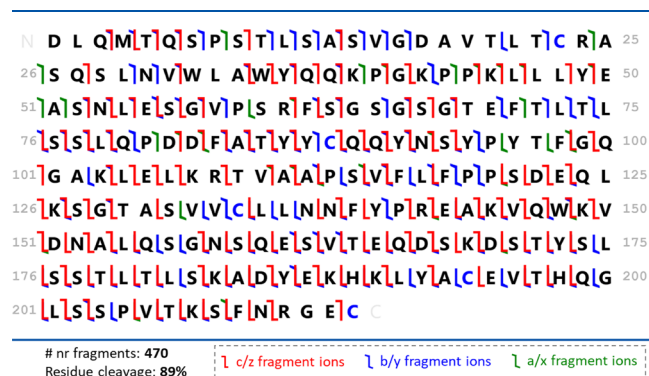


Figure 3. Fragmentation map obtained for the P15 reduced/alkylated LC sequence using a combination of 12 top-down MS/MS analyses performed with different fragmentation methods (described in the SI).

reconstructed using the 470 assigned nonredundant fragment ions obtained from all conditions. This corresponds to 89% residue cleavage (Table S3) indicating only a few remaining gaps in the peptide bond cleavages.

Proteoform Validation. As shown in Figure 3, a few amino acid stretches are not covered by the TDP data. To confirm these parts of the P15 sequence, we performed a classical BUP search using the tryptic digest data against the Uniprot *Homo sapiens* reference proteome supplemented with the P15 sequence. The MaxQuant results score P15 as the best match

with 100% sequence coverage (Table S3 and Figure S2). We thus used the information achieved on individual tryptic peptides to confirm the parts of the sequence poorly covered by the TDP data. Figure S3 shows for instance the fragmentation spectrum obtained for the [1–24] peptide allowing the DAVTITCR to be confirmed at the single amino acid level. BUP data completely confirmed the sequence depicted in Figure 3 for P15. We then evaluated the information brought in by each of the four enzymes (see Figure S4 for P15 sequence coverage map). We concluded that a combination of trypsin and *Nepenthes* digestive fluid is sufficient to achieve a 100% sequence coverage, and we, therefore, only used these enzymes for all other samples. For the assignment of the disulfide bridges and cysteinylolation, we analyzed a nonreduced/nonalkylated tryptic digest of P15. Using MassSpec Studio,⁴¹ the disulfide bonds were mainly identified between C23–C88 and C134–C194 (Figure S5). Other combinations could also be identified, but to a much lesser extent. Cysteinylolation was primarily identified on the C-terminal cysteine, which fits with the previously determined disulfide bond assignment. Note that the size of the C-terminal peptide (GEC) precluded the fragmentation of the corresponding cysteinylated form (less than 500 Da and probably singly charged) in our experiments. For the I/L attribution, we analyzed the tryptic and chymotryptic digests in LC-MSⁿ using either EThcD or HCD (MS⁴) on selected I/L-containing peptides (Table S4 and Figure S6).^{38,39} For P15, this strategy allowed us to formally attribute 20/24 residues (Figure 4). The remaining four residues were assigned by sequence homology with the LC sequences found in this work.

Application to the Other Clinical Samples. The same strategy as the one described for P15 was applied to all other clinical samples, and the corresponding results are presented in the paragraphs below and in the SI. The samples that share common characteristics and a high sequence homology are discussed together.

P6, P7, P18, and P20. These samples exhibited the same behavior as P15 and could be *de novo*-sequenced exactly in the same way (Table 1). These samples contain a single κ LC, which exists both in a monomeric and dimeric form (with various relative abundances). These LCs contain five cysteines, two disulfide bonds, and a cysteinylolation at the C-terminus. For all samples, the dimer results from the combination of two monomers that are linked through a single disulfide bond. All intact MS spectra and TDP fragmentation maps are provided in Figures S7 and S8. For these LCs, sequence coverages larger than 80% were obtained with TDP and increased to 100% when combined with BUP data (Figure S2 and Table S3). For the I/L discrimination, the vast majority could be assigned thanks to our peptide MS/MS experiments (ca. 90%), and the remaining ones were assigned by sequence homology. Finally, P6, P7, P18, and P20 exhibit LC sequences that share 80.8% homology between them and 79% when including P15 (Figure S9). Regarding the monomer/dimer ratio, two different behaviors were observed. P6 and P7 have only a small amount of dimer, while they represent around 25% of P18 and P20 samples.

P8 and P19. For these two samples, the situation is very different since the intact mass measurement indicates the presence of two proteoforms per sample. After reduction, P19 still exhibits two proteoforms although only one remains for P8. Using the workflow described above, all sequences were found to be part of the κ isotype, as the previous ones, and contain five cysteines involved in two disulfide bridges. For the two samples, the two respective proteoforms were found to share the same

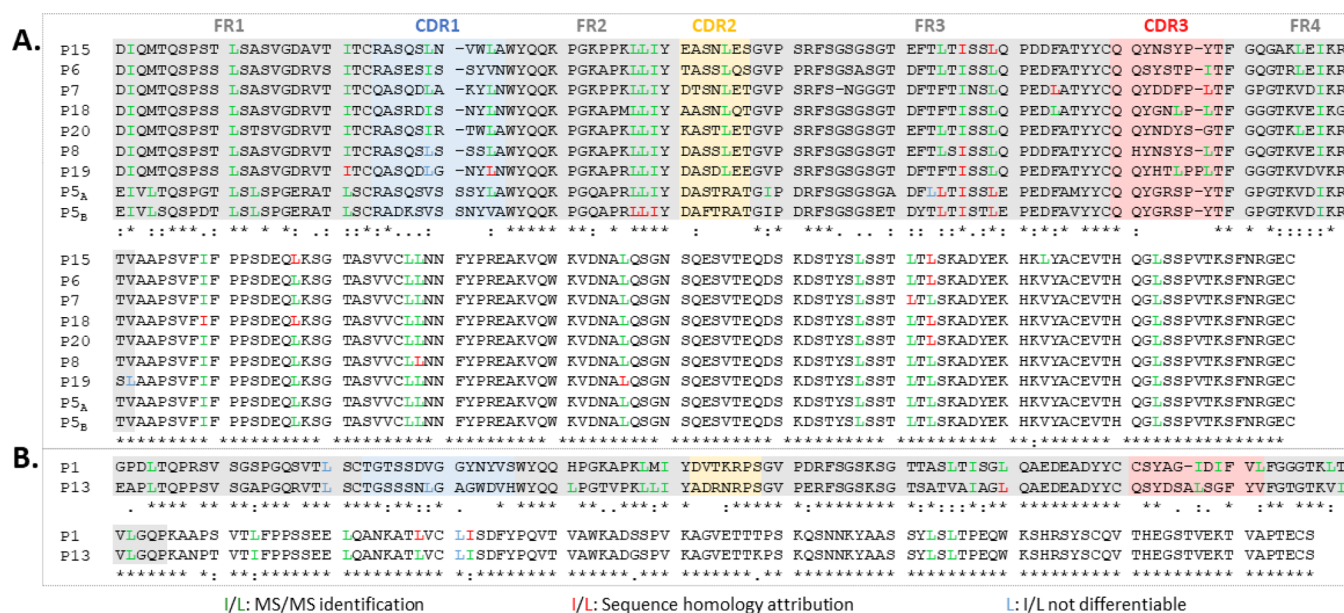


Figure 4. Multiple sequence alignment of (A) nine κ LC sequences and (B) two λ LC sequences (done with Uniprot, “*” (asterisk) indicates fully conserved residue, “:” (colon) indicates residues with strong similar properties, and “.” (period) indicates residues with weak similar properties). Framework (FR) and CDRs of the LC variable domains were numbered according to the Kabat tool (<http://www.abysis.org/abysis/index.html>).

Table 1. Summary Results Obtained for All LC Samples

sample	#proteiform	monomer		LC isotype	LC length	LC post-translational modifications (monomer)
		(relative abundance in %)				
P15	1	P15 (100)	P15–P15 (11)	κ	214	2 S–S bonds (C23–C88, C134–C194); 1 cysteinylolation ^a (C214)
P6	1	P6 (100)	P6–P6 (48)	κ	214	2 S–S bonds (C23–C85, C134–C194); 1 cysteinylolation ^a (C214)
P7	1	P7 (100)	P7–P7 (7)	κ	213	2 S–S bonds (C23–C87, C133–C193); 1 cysteinylolation ^a (C213)
P18	1	P18 (100)	P18–P18 (12)	κ	214	2 S–S bonds (C23–C88, C134–C194); 1 cysteinylolation ^a (C214)
P20	1	P20 (100)	P20–P20 (56)	κ	214	2 S–S bonds (C23–C88, C134–C194); 1 cysteinylolation ^a (C214)
P8	2	P8 _A (39), P8 _B (100)	P8–P8 (0.5)	κ (P8 _A)	214 (P8 _A)	P8 _A : 2 S–S bonds (C23–C88, C134–C194); 1 cysteinylolation ^a (C214)
				κ (P8 _B)	214 (P8 _B)	P8 _B : 2 S–S bonds (C23–C88, C134–C194); 1 coenzyme M ^b (C214)
P19	2	P19 _A (100), P19 _B (15)	P19 _A –P19 _A (42) P19 _A –P19 _B (26)	κ (P19 _A)	215 (P19 _A)	P19 _A : 2 S–S bonds (C23–C88, C135–C195); 1 cysteinylolation (C215)
				κ (P19 _B)	215 (P19 _B)	P19 _B : 2 S–S bonds (C23–C88, C135–C195); 1 cysteinylolation ^a (C215); 1 HexNAc(1)dHex(1) ^c
P5	2	P5 _B (10)	P5 _A –P5 _A (100) P5 _A –P5 _B (55) P5 _B –P5 _B (46)	κ (P5 _A)	215 (P5 _A)	P5 _A : 2 S–S bonds (C23–C89, C135–C195); 1 cysteinylolation ^a (C215)
				κ (P5 _B)	215 (P5 _B)	P5 _B : 2 S–S bonds; 1 cysteinylolation ^a
P1	1		P1–P1 (100)	λ	217	2 S–S bonds
P13	1	P13 (52)	P13–P13 (100)	λ	216	2 S–S bonds; 1 cysteinylolation ^a (C216)

^aCysteinylolation: 119.00 Da. ^bCoenzyme M: 139.96 Da. ^cHexNAc(1)dHex(1): 349.14 Da.

amino acid sequence, the difference between them arising from additional modifications. For P8, the first proteoform (P8_A) carries a cysteinylolation on the C-terminal cysteine as the ones previously described. The second proteoform (P8_B), which is the most abundant (Table 1), differs from P8_A only by another modification present on a cysteine, since it is also removed after the reduction/alkylation process. The delta mass measured is 139.96 Da, which does not correspond to any described PTM.

This mass only fits with an elemental formula of C₂H₄O₃S₂. Considering all possible thiol-based structures, we determined that P8_B was probably modified by the coenzyme M (C₂H₆O₃S₂), a small molecule often used as an adjuvant in chemotherapy. The medical details of the patient confirmed that his chemotherapy treatment contained this adjuvant confirming our identification. To our knowledge, such a modification has never been described so far on a protein and in particular on a

LC. The fact that both modifications are on a cysteine explains why after the reduction, the two proteoforms become a single one. This also explains why a single dimer, however in very low abundance (Table 1), is present in this sample, since it is created by the binding of the C-terminal cysteines of the two monomeric proteoforms. For P19, the two proteoforms are cysteinylated on the C-terminus. For the second proteoform (P19_B), an additional oligosaccharide HexNAc(1)dHex(1) (349.14 Da) was identified at S160. This location is in agreement with our TDP results, leading to a highest sequence coverage with 362 fragment ions (Figure S7, P19_B). The presence of this modification is unique to P19 among all clinical samples studied here. For this sample, the P19_A–P19_A homodimer and the only possible heterodimer are detected as indicated by the intact mass measurements. We did not observe any P19_B–P19_B homodimer, probably because of the low abundance of the glycosylated proteoform (Table 1).

At last, 80% of the I/L residues were attributed by MS, 13% by sequence homology, while three remains not identified. To facilitate the sequence alignment, we chose to assign them as leucines on Figure 4. Finally, for both P8 and P19 samples, the sequence coverage for all proteoforms were found to be between 82 and 86% with TDP and 100% when including BUP data (Figure S2 and Table S4). The sequence homology between P8 and P19 sequences is 88.4% (Figure S8).

P5. This sample is again different from the others, since it contains two different κ proteoforms that have different amino acid sequences. The proteoforms share a common pattern of five cysteines, two disulfide bonds, and a cysteinylated at the C-terminus but differ by 13 residues present in the variable part of the LC (94% of sequence homology, Figure S8). The fact that this sample contained a mixture of two sequences, close to each other drastically complicated the de novo sequencing in particular the assembly of peptides. The TDP data play a very important role here since the information obtained with this approach is specific to a proteoform, which is not the case in BUP where all peptides are mixed making the assignment to a specific proteoform barely impossible. It would have been clearly very complicated to assign the final sequences with high precision without TDP. Interestingly, for the P5 sample, all dimers are formed (both homo and hetero) and constitute the majority (95%) of the sample (Table 1).

P1 and P13. Finally, these two last samples (P1 and P13) were also found peculiar since they contain only a single LC proteoform but of the λ isotype. The P1 proteoform contains six cysteines, with only four involved in a disulfide bridge. The presence of two free cysteines probably explains the easy and exclusive formation of a dimer, since the monomeric form is absent in this sample (Table 1). This number of cysteines is very unusual since LCs generally contain only five cysteines. This particularity, combined with the fact that two cysteines are vicinal largely complicates the exact determination of the disulfide bridges. It was therefore possible to formally identify only the C139–C197 disulfide bond. Note that P1, in contrast to all previously characterized LCs, does not carry any cysteinylated. For P13, the sequence contains five cysteines, including two disulfide bonds, the remaining cysteine being cysteinylated as observed for the κ LCs present in the other samples. In this sample, the dimeric form is also predominant. Note that the S–S bond assignment was made difficult by the very low abundance of this species in the mixture. The sequence homology between P1 and P13 is 79.3% (Figure S8).

DISCUSSION

As shown from our results, the variability observed for the LCs extracted from all clinical samples is very important and much higher than expected. First, both λ and κ isotypes are identified, even if the majority is constituted of the latter. Second, some samples contain a single proteoform although others contain several, which arise either from different amino acid sequences or from the presence of various PTMs. The sequence homology between all sequences found in our work is 66.7% for the κ isotypes and 94% for the two λ ones (Figure S8). The presence of different LC sequences for the same patient suggests the existence of several plasma cell clones or point toward an unpredictable maturation process. Another difference is also the ratio between the monomeric and dimeric forms of the LCs that range from 0 to 100%. Could the exclusive presence of dimers in some sample be tightly related to an increased aggregation? For the PTMs, we find the cysteinylated of the last cysteine of the sequence to be shared by almost all samples. We also identify a N-glycosylation and a modification with coenzyme M that are either very unusual or never described so far. Our results clearly show the added value of TDP to achieve a bird's eye view of the various proteoforms present in each sample as well as for the de novo sequencing. For one of sample containing two closely related sequences, the direct analysis of proteoforms showed an extraordinary advantage over the analysis of peptides obtained by digestion. Our data also demonstrate that the combination of various activation techniques is required to improve the sequence coverage.

CONCLUSIONS

We developed here a novel workflow allowing the complete characterization of LCs extracted from the urine of patients with MM. This workflow is based on the combination of BUP and TDP approaches, as well as the use of appropriate software tools that are all commercially available. Using this pipeline, we could characterize for the first time a high variability in the LCs in terms of sequence, PTMs, and the presence of monomeric or dimeric forms. The precision of our workflow, and in particular the data obtained on intact proteoforms, allowed us to identify an unexpected modification linked to a specific medical treatment for one patient. Top-down proteomics provides straightforward information on the different proteoforms present in each sample (in particular monomer/dimer presence) and is also invaluable in deciphering closely related proteoforms present in the same clinical sample. Our results have now to be integrated to the biophysical data already obtained for all samples.⁴⁶ Indeed, determining LC sequences is only the first step to achieve the overall goal of this study, which is to identify the major factors influencing the propensity of these LCs to aggregates and leading to disease. Integrating all the data obtained is now required to understand the aggregation process and be able to prevent it in the future.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.1c01955>.

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD025884. Fragmentation maps obtained for P15 in TDP (Figure S1), sequence coverage maps obtained in BUP (tryptic

digests) (Figure S2), the HCD MS/MS spectrum of the N-terminal tryptic peptide of P15 (Figure S3), sequence coverage map obtained for P15 digested by the four enzymes (Figure S4), the HCD MS/MS spectrum of the C134–C194 disulfide-linked peptide of P15 (Figure S5), spectra showing I/L differentiation (Figure S6), mass spectra of intact proteins under various conditions (Figure S7), fragmentation maps obtained for all LCs in TDP (Figure S8), and sequence identity (%) between LCs (Figure S9) (PDF)

LC samples studied in this work (Table S1), detailed information on all LC proteoforms (Table S2), sequence coverages obtained in TDP and BUP for the tryptic digests (Table S3), and peptides and type of experiment used for the I/L discrimination (Table S4) (XLSX)

Mass Spec Studio search parameters for S–S bond assignment (TXT)

AUTHOR INFORMATION

Corresponding Author

Julia Chamot-Rooke – Mass Spectrometry for Biology Unit, CNRS USR2000, Institut Pasteur, CNRS, Paris 75015, France; orcid.org/0000-0002-9427-543X; Email: julia.chamot-rooke@pasteur.fr

Authors

Mathieu Dupré – Mass Spectrometry for Biology Unit, CNRS USR2000, Institut Pasteur, CNRS, Paris 75015, France; Present Address: DMPK Sanofi, Chilly-Mazarin 91,380, France; orcid.org/0000-0002-1845-0048

Magalie Duchateau – Mass Spectrometry for Biology Unit, CNRS USR2000, Institut Pasteur, CNRS, Paris 75015, France

Rebecca Sternke-Hoffmann – Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, Düsseldorf 40225, Germany

Amelie Boquoi – Department of Hematology, Oncology and Clinical Oncology, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany, Düsseldorf 40225, Germany

Christian Malosse – Mass Spectrometry for Biology Unit, CNRS USR2000, Institut Pasteur, CNRS, Paris 75015, France

Roland Fenk – Department of Hematology, Oncology and Clinical Oncology, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany, Düsseldorf 40225, Germany

Rainer Haas – Department of Hematology, Oncology and Clinical Oncology, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany, Düsseldorf 40225, Germany

Alexander K. Buell – Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby 2800, Denmark; orcid.org/0000-0003-1161-3622

Martial Rey – Mass Spectrometry for Biology Unit, CNRS USR2000, Institut Pasteur, CNRS, Paris 75015, France; orcid.org/0000-0002-7378-1106

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.analchem.1c01955>

Author Contributions

§M. Dupré and M. Duchateau contributed equally.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier (ongoing process). This work has been supported by EPIC-XS, project number 823839, funded by the Horizon 2020 programme of the European Union, the Institut Pasteur and CNRS. The authors are grateful to David Scherberich, responsible of the scientific pole of the botanical garden of Lyon, and to Thibault Chaze for his help on the LC glycosylation analysis. R.S.H. thanks the Manchot foundation for funding. A.K.B. thanks the Novo Nordisk Foundation for support (NNFSA170028392).

REFERENCES

- (1) Bakkus, M. H.; Heirman, C.; Van Riet, I.; Van Camp, B.; Thielemans, K. *Blood* **1992**, *80*, 2326–2335.
- (2) Chapman, M. A.; Lawrence, M. S.; Keats, J. J.; Cibulskis, K.; Sougnez, C.; Schinzel, A. C.; Harview, C. L.; Brunet, J. P.; Ahmann, G. J.; Adli, M.; Anderson, K. C.; Ardlie, K. G.; Auclair, D.; Baker, A.; Bergsagel, P. L.; Bernstein, B. E.; Drier, Y.; Fonseca, R.; Gabriel, S. B.; Hofmeister, C. C.; Jagannath, S.; Jakubowiak, A. J.; Krishnan, A.; Levy, J.; Liefeld, T.; Lonial, S.; Mahan, S.; Mfuko, B.; Monti, S.; Perkins, L. M.; Onofrio, R.; Pugh, T. J.; Rajkumar, S. V.; Ramos, A. H.; Siegel, D. S.; Sivachenko, A.; Stewart, A. K.; Trudel, S.; Vij, R.; Voet, D.; Winckler, W.; Zimmerman, T.; Carpten, J.; Trent, J.; Hahn, W. C.; Garraway, L. A.; Meyerson, M.; Lander, E. S.; Getz, G.; Golub, T. R. *Nature* **2011**, *471*, 467–472.
- (3) Magrangeas, F.; Cormier, M. L.; Descamps, G.; Gouy, N.; Lode, L.; Mellerin, M. P.; Harousseau, J. L.; Bataille, R.; Minvielle, S.; Avet-Loiseau, H. *Blood* **2004**, *103*, 3869–3875.
- (4) Herrera, G. A. *Arch. Pathol. Lab. Med.* **2009**, *133*, 249–267.
- (5) Buxbaum, J.; Gallo, G. *Hematol. Oncol. Clin. North Am.* **1999**, *13*, 1235–1248.
- (6) Sanders, P. W.; Booker, B. B. *J. Clin. Invest.* **1992**, *89*, 630–639.
- (7) Glenner, G. G.; Ein, D.; Terry, W. D. *Am. J. Med.* **1972**, *52*, 141–147.
- (8) Andrich, K.; Hegenbart, U.; Kimmich, C.; Kedia, N.; Bergen, H. R., 3rd; Schonland, S.; Wanker, E.; Bieschke, J. *J. Biol. Chem.* **2017**, *292*, 2328–2344.
- (9) Arosio, P.; Owczar, M.; Muller-Spath, T.; Rognoni, P.; Beeg, M.; Wu, H.; Salmona, M.; Morbidelli, M. *PLoS One* **2012**, *7*, No. e33372.
- (10) Blancas-Mejia, L. M.; Horn, T. J.; Marin-Argany, M.; Auton, M.; Tischer, A.; Ramirez-Alvarado, M. *Biophys. Chem.* **2015**, *207*, 13–20.
- (11) Brumshtein, B.; Esswein, S. R.; Sawaya, M. R.; Rosenberg, G.; Ly, A. T.; Landau, M.; Eisenberg, D. S. *J. Biol. Chem.* **2018**, *293*, 19659–19671.
- (12) Sternke-Hoffmann, R.; Boquoi, A.; Lopez, Y. N. D.; Platten, F.; Fenk, R.; Haas, R.; Buell, A. K. *PeerJ* **2020**, *8*, No. e8771.
- (13) Weber, B.; Hora, M.; Kazman, P.; Gobl, C.; Camilloni, C.; Reif, B.; Buchner, J. *J. Mol. Biol.* **2018**, *430*, 4925–4940.
- (14) Lefranc, M.-P.; Lefranc, G., *The immunoglobulin factsbook*. Academic Press: San Diego, 2001; 457.
- (15) Sakano, H.; Huppi, K.; Heinrich, G.; Tonegawa, S. *Nature* **1979**, *280*, 288–294.
- (16) Enqvist, S.; Sletten, K.; Stevens, F. J.; Hellman, U.; Westermark, P. *PLoS One* **2007**, *2*, No. e981.
- (17) Teng, G.; Papavasiliou, F. N. *Annu. Rev. Genet.* **2007**, *41*, 107–120.
- (18) Gearhart, P. J.; Johnson, N. D.; Douglas, R.; Hood, L. *Nature* **1981**, *291*, 29–34.

- (19) McBride, O. W.; Hieter, P. A.; Hollis, G. F.; Swan, D.; Otey, M. C.; Leder, P. *J Exp Med* **1982**, *155*, 1480–1490.
- (20) Cheung, W. C.; Beausoleil, S. A.; Zhang, X.; Sato, S.; Schieferl, S. M.; Wieler, J. S.; Beaudet, J. G.; Ramenani, R. K.; Popova, L.; Comb, M. J.; Rush, J.; Polakiewicz, R. D. *Nat. Biotechnol.* **2012**, *30*, 447–452.
- (21) Upadhyay, A. A.; Kauffman, R. C.; Wolabaugh, A. N.; Cho, A.; Patel, N. B.; Reiss, S. M.; Havenar-Daughton, C.; Dawoud, R. A.; Tharp, G. K.; Sanz, I.; Pulendran, B.; Crotty, S.; Lee, F. E.; Wrammert, J.; Bosinger, S. E. *Genome Med.* **2018**, *10*, 20.
- (22) Iwamoto, N.; Shimada, T. *Pharmacol. Ther.* **2018**, *185*, 147–154.
- (23) Ladwig, P. M.; Barnidge, D. R.; Willrich, M. A. V. *Clin. Vaccine Immunol.* **2017**, *24*, No. e00545-16.
- (24) Ruggles, K. V.; Krug, K.; Wang, X.; Clauser, K. R.; Wang, J.; Payne, S. H.; Fenyó, D.; Zhang, B.; Mani, D. R. *Mol. Cell. Proteomics* **2017**, *16*, 959–981.
- (25) Vitorino, R.; Guedes, S.; Trindade, F.; Correia, I.; Moura, G.; Carvalho, P.; Santos, M. A. S.; Amado, F. *Expert Rev. Proteomics* **2020**, *17*, 595–607.
- (26) Peng, W.; Pronker, M. F.; Snijder, J. J. *Proteome Res.* **2021**, 3559.
- (27) Sen, K. I.; Tang, W. H.; Nayak, S.; Kil, Y. J.; Bern, M.; Ozoglu, B.; Ueberheide, B.; Davis, D.; Becker, C. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 803–810.
- (28) Bandeira, N.; Pham, V.; Pevzner, P.; Arnott, D.; Lill, J. R. *Nat. Biotechnol.* **2008**, *26*, 1336–1338.
- (29) Tran, N. H.; Rahman, M. Z.; He, L.; Xin, L.; Shan, B.; Li, M. *Sci. Rep.* **2016**, *6*, 31730.
- (30) Dupre, M.; Duchateau, M.; Malosse, C.; Borges-Lima, D.; Calvaresi, V.; Podglajen, I.; Clermont, D.; Rey, M.; Chamot-Rooke, J. J. *Proteome Res.* **2021**, *20*, 202–211.
- (31) Gault, J.; Ferber, M.; Machata, S.; Imhaus, A. F.; Malosse, C.; Charles-Orszag, A.; Millien, C.; Bouvier, G.; Bardiaux, B.; Pehau-Arnaudet, G.; Klinge, K.; Podglajen, I.; Ploy, M. C.; Seifert, H. S.; Nilges, M.; Chamot-Rooke, J.; Dumenil, G. *PLoS Pathog.* **2015**, *11*, No. e1005162.
- (32) He, L.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Dasari, S.; Dispenzieri, A.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2019**, *91*, 3263–3269.
- (33) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10313–10317.
- (34) Ren, D.; Pipes, G. D.; Hambly, D.; Bondarenko, P. V.; Treuheit, M. J.; Gadgil, H. S. *Anal. Biochem.* **2009**, *384*, 42–48.
- (35) Srzentic, K.; Fornelli, L.; Tsybin, Y. O.; Loo, J. A.; Seckler, H.; Agar, J. N.; Anderson, L. C.; Bai, D. L.; Beck, A.; Brodbelt, J. S.; van der Burgt, Y. E. M.; Chamot-Rooke, J.; Chatterjee, S.; Chen, Y.; Clarke, D. J.; Danis, P. O.; Diedrich, J. K.; D'Ippolito, R. A.; Dupre, M.; Gasilova, N.; Ge, Y.; Goo, Y. A.; Goodlett, D. R.; Greer, S.; Haselmann, K. F.; He, L.; Hendrickson, C. L.; Hinkle, J. D.; Holt, M. V.; Hughes, S.; Hunt, D. F.; Kelleher, N. L.; Kozhinov, A. N.; Lin, Z.; Malosse, C.; Marshall, A. G.; Menin, L.; Millikin, R. J.; Nagornov, K. O.; Nicolardi, S.; Pasa-Tolic, L.; Pengelley, S.; Quebbemann, N. R.; Resemann, A.; Sandoval, W.; Sarin, R.; Schmitt, N. D.; Shabanowitz, J.; Shaw, J. B.; Shortreed, M. R.; Smith, L. M.; Sobott, F.; Suckau, D.; Toby, T.; Weisbrod, C. R.; Wildburger, N. C.; Yates, J. R., 3rd; Yoon, S. H.; Young, N. L.; Zhou, M. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1783–1802.
- (36) Liu, X.; Dekker, L. J.; Wu, S.; Vanduijn, M. M.; Luider, T. M.; Tolic, N.; Kou, Q.; Dvorkin, M.; Alexandrova, S.; Vyatkina, K.; Pasa-Tolic, L.; Pevzner, P. A. *J. Proteome Res.* **2014**, *13*, 3241–3248.
- (37) Rey, M.; Yang, M.; Burns, K. M.; Yu, Y.; Lees-Miller, S. P.; Schriemer, D. C. *Mol. Cell. Proteomics* **2013**, *12*, 464–472.
- (38) Lebedev, A. T.; Damoc, E.; Makarov, A. A.; Samgina, T. Y. *Anal. Chem.* **2014**, *86*, 7017–7022.
- (39) Xiao, Y.; Vecchi, M. M.; Wen, D. *Anal. Chem.* **2016**, *88*, 10757–10766.
- (40) Kavan, D.; Man, P. *Int. J. Mass Spectrom.* **2011**, *302*, 53–58.
- (41) Sarpe, V.; Rafiei, A.; Hepburn, M.; Ostan, N.; Schryvers, A. B.; Schriemer, D. C. *Mol. Cell. Proteomics* **2016**, *15*, 3071–3080.
- (42) Perry, R. H.; Cooks, R. G.; Noll, R. J. *Mass Spectrom. Rev.* **2008**, *27*, 661–699.
- (43) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342.
- (44) Lim, A.; Wally, J.; Walsh, M. T.; Skinner, M.; Costello, C. E. *Anal. Biochem.* **2001**, *295*, 45–56.
- (45) Gadgil, H. S.; Bondarenko, P. V.; Pipes, G. D.; Dillon, T. M.; Banks, D.; Abel, J.; Kleemann, G. R.; Treuheit, M. J. *Anal. Biochem.* **2006**, *355*, 165–174.
- (46) Sternke-Hoffmann, R.; Pauly, T.; Norrild, R. K.; Hansen, J.; Dupré, M.; Tucholski, F.; Duchateau, M.; Rey, M.; Metzger, S.; Boquiol, A.; Platten, F.; Egelhaaf, S. U.; Chamot-Rooke, J.; Fenk, R.; Nagel-Steger, L.; Haas, R.; Buell, A. K. *bioRxiv* **2021**, DOI: 10.1101/2021.05.12.443858.