analytical chemistry

Article

Cite This: Anal. Chem. 2019, 91, 3263−3269

pubs.acs.org/ac

Classification of Plasma Cell Disorders by 21 Tesla Fourier Transform Ion Cyclotron Resonance Top-Down and Middle-Down MS/MS Analysis of Monoclonal Immunoglobulin Light Chains in Human Serum

Lidong He,[†] Lissa C. Anderson,[‡] David R. Barnidge,[§] David L. Murray,[∥] Surendra Dasari,[⊥] Angela Dispenzieri,[#] Christopher L. Hendrickson,[†,‡] and Alan G. Marshall*[,†,‡]

[†]Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32310, United States

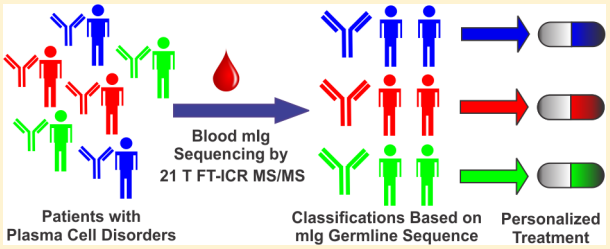
[‡]National High Magnetic Field Laboratory, Florida State University, 1800 East Paul Dirac Dr., Tallahassee, Florida 32310, United States

[§]The Binding Site, Rochester, Minnesota 55901, United States

[∥]Department of Laboratory Medicine and Pathology, [⊥]Department of Health Science Research, [#]Division of Hematology, Department of Medicine, Mayo Clinic, Rochester, Minnesota 55905, United States

Supporting Information

ABSTRACT: The current five-year survival rate for systemic AL amyloidosis or multiple myeloma is ∼51%, indicating the urgent need for better diagnosis methods and treatment plans. Here, we describe highly specific and sensitive top-down and middle-down MS/MS methods owning the advantages of fast sample preparation, ultrahigh mass accuracy, and extensive residue cleavages with 21 telsa FT−ICR MS/MS. Unlike genomic testing, which requires bone marrow aspiration and may fail to identify all monoclonal immunoglobulins produced by the body, the present method requires only a blood draw. In addition, circulating monoclonal immunoglobulins spanning the entire population are analyzed and reflect the selection of germline sequence by B cells. The monoclonal immunoglobulin light chain FR2-CDR2-FR3 was sequenced by database-aided de novo MS/MS and 100% matched the gene sequencing result, except for two amino acids with isomeric counterparts, enabling accurate germline sequence classification. The monoclonal immunoglobulin heavy chains were also classified into specific germline sequences based on the present method. This work represents the first application of top/middle-down MS/MS sequencing of endogenous human monoclonal immunoglobulins with polyclonal immunoglobulins background.



Plasma cell disorders include monoclonal gammopathy of undetermined significance (MGUS, asymptomatic, can progress to multiple myeloma), multiple myeloma (symptoms including hypercalcemia, renal failure, anemia, and bone lesion), AL amyloidosis (deposition of free immunoglobulin light chains locally or systemically which may lead to organ failure), POEMS syndrome (symptoms including polyneuropathy, organomegaly, and endocrinopathy), and Waldenstrom's macroglobulinemia (WM, a disorder of B-lymphocytes with plasmacytic differentiation characterized by high concentrations of immunoglobulin M), all of which are characterized by the expansion of plasma cell clones. If there is clinical suspicion for one of these disorders, blood serum is tested for the presence of an elevated level of a monoclonal immunoglobulin (mIg or M protein, composed of two identical ∼25 kDa light chains and two identical ∼50 kDa heavy chains linked together by disulfide bonds) secreted by clonal plasma cells.[1,2] Despite recent research advances, the systemic AL amyloidosis and multiple myeloma five-year survival rates are still ∼51%,[3,4] indicating the need for improved diagnosis and treatment plans.

A predilection of mIgs lambda light chain LV1 gene usage (>95% of the time) was found in POEMS syndrome.[5] To elucidate the mechanisms of AL amyloidosis and multiple myeloma, immunoglobulin gene usage preference has been studied by gene sequencing.[6−10] However, only a few patients were involved in those studies due to the use of time-consuming (∼2 days) polymerase chain reaction (PCR)-based sequencing to characterize the mIg germline sequence. In addition, the specificity of a designed primer to the target gene sequence needs to be assured. Otherwise inaccurate sequence identifications may occur if the primer binds to off-target germline variable region sequences with somatic mutations. Moreover, the genomic technique requires invasive and costly

Received: July 23, 2018
Accepted: January 28, 2019
Published: February 25, 2019

DOI: 10.1021/acs.analchem.8b03294
Anal. Chem. 2019, 91, 3263−3269

bone marrow aspiration. Finally, the sequenced mIg gene cannot represent the identities and abundances of all mIgs in circulation because it is based only on the B cells isolated from the aspirated bone marrow. In contrast, mass spectrometry-based approaches analyze immunoglobulins directly from blood serum, and thus, represent the entire circulating immunoglobulin population.[11] Furthermore, immunoglobulin post-translational modifications (PTMs) are well characterized, including N-/C-terminal truncation and glycosylation, which play an important role in immune response.[12]

Conventional bottom-up liquid chromatography−tandem mass spectrometry (LC−MS/MS) is widely used to characterize the mIg sequence in serum and tissue.[13−16] Recently, a large clinical study with bottom-up de novo MS/MS correlated mIg light chain gene usage and different levels of organ damage in systemic or localized AL amyloidosis.[17] However, that method requires tissue biopsy, which is invasive. In addition, it is much more difficult to apply bottom-up de novo MS/MS to endogenous mIg in serum or plasma matrix, because digestion of the polyclonal immunoglobulins (pIgs) background yields highly abundant constant region tryptic peptides, which can suppress the signals from mIg variable region peptides of lower abundance. In the generic bottom-up approach, to gain maximum sequence coverage, laborious protein alkylation and enzymatic digestions are required, and the resulting peptide mixture is complex and can confound proteome analysis. For example, it is difficult to assign the origin of peptides if more than one protein is present.[18]

Top-down (mIg light chain, heavy chain) and middle-down (mIg light chain, heavy chain Fc/2, and heavy chain Fd subunits after protease digestion) LC−MS/MS offer the advantages of fast and simple sample preparation with minimal artifacts introduction,[19,20] and it can unambiguously assign sequence and proteoforms to each online LC-eluted protein.[21,22] We recently reported extensive characterization of monoclonal antibody (mAb, which is a therapeutic drug and is differentiated from endogenous mIg) light chain and heavy chain sequences by top-down (requiring only prior disulfide reduction) and middle-down (requiring IdeS enzyme digestion and disulfide reduction) MS/MS.[23]

Here, we report the development of a specific (mIg germline sequence-classified), low limit of detection (∼33 fmol mIg), and rapid (1 h sample preparation and two 50 min LC−MS/MS analyses) approach for typing and characterizing endogenous mIgs in human serum matrix. In contrast to bottom-up MS/MS in which serum pIgs-digested peptides obscure signals of mIg enzymatic peptides, top-down/middle-down MS/MS selects a specific m/z corresponding to mIg precursor ions (thus eliminating most of the pIg interference) for fragmentation. Two AL amyloidosis patient samples (one containing mIgs with kappa light chain and the other containing mIgs with lambda light chain) were used for proof-of-concept experiments. The mIg light chain sequences were extensively characterized in a blind analysis, and the determined sequences accurately matched the gene sequencing result. In addition, the middle-down approach provided detailed glycoform profiling of the mIg heavy chain in an AL amyloidosis sample and was found to be different from the normal human polyclonal immunoglobulins glycoform profile. The nano-LC 21 tesla (T) FT−ICR MS and top/middle-down MS/MS provide ultrahigh mass accuracy and extensive residue cleavages for accurate germline sequence identification and PTM characterization. This method has the potential for future precision diagnosis of mIg-involved diseases, such as MGUS, multiple myeloma, AL amyloidosis, POEMS syndrome, and WM.

EXPERIMENTAL METHODS

Materials. AL amyloidosis samples were provided by Mayo Clinic in compliance with the Mayo Clinic and Florida State University Institutional Review Board (HSC No. 2017.21179). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl), myoglobin, acetic acid (≥99.99%), and isopropyl alcohol (LCMS grade) were acquired from Sigma−Aldrich (St. Louis, MO). Water (LCMS grade) and acetonitrile (ACN, LCMS grade) were obtained from Honeywell Burdick & Jackson (Muskegon, MI).

Immunoglobulin Purification. For the patient serum sample (the first AL amyloidosis sample), Melon Gel (Thermo Fisher Scientific, Waltham, MA) was used to purify immunoglobulins from human serum according to the manufacturer's protocol.[24] An Amicon Ultra-0.5 Centrifugal Filter Unit with an Ultracel-10k (10 kDa molecular weight cutoff) membrane (EMD Millipore, Darmstadt, Germany) was used for further desalting. The final volume was ∼20 μL. For the patient plasma sample (the second AL amyloidosis sample), immunoglobulins were purified from camelid-derived nanobodies (Thermo Fisher Scientific) directed against the lambda light chain constant domain (prior mIg light chain immunofixation established that the isotype was lambda). Future use of the camelid-derived nanobodies mixture directed against kappa/lambda light chain constant domains does not require prior mIg light chain isotyping as previously reported.[25] The final volume was ∼20 μL.

Immunoglobulin Reduction. The 1 μL purified sample was added to 11 μL of 10 mM TCEP in 0.1% (v/v) acetic acid in water followed by incubation at 75 °C for 15 min to produce intact antibody light (∼24 kDa) and heavy (∼50 kDa) chains. Then, 1% (v/v) acetic acid was added to samples to make the pH ∼2 prior to nano-LC ESI MS/MS analysis. The final mIg concentration was ∼0.25 μM (immunoglobulins initial concentration in clinical samples was determined by nephelometry).

Immunoglobulin IdeS Digestion and Reduction. One unit of IdeS (immunoglobulin G-degrading protease from FabRICATOR, Genovis, Cambridge, MA) was added per microgram of IgG, followed by incubation at 10 μg/μL at 37 °C for 30 min. After digestion, TCEP reduction yielded intact antibody light (∼24 kDa), heavy chain Fd (∼25 kDa), and heavy chain Fc/2 (∼25 kDa) subunits followed by acidification as described above.

Liquid Chromatography and Mass Spectrometry. A Waters Acquity UPLC M-Class System (Milford, MA) was used for separation. A custom-fabricated trap column (3 cm × 150 μm, packed with 5 μm spherical diameter Poroshell 300 SB-C_8 stationary phase, Agilent Technologies, Santa Clara, CA) was used for sample desalting, and an analytical column (15 cm × 75 μm, packed with 5 μm Poroshell 300 SB-C_8 stationary phase) was used for online separation of immunoglobulins prior to ionization. Mobile phase A consisted of 0.3% formic acid in 4.7% ACN and 95% water. Mobile phase B was 0.3% formic acid in 4.7% water, 47.5% ACN, and 47.5% isopropyl alcohol (all % expressed as v/v). Samples (∼0.5 pmol mIg) were loaded onto a trap column with 5% B at a flow rate of 2.5 μL/min for 10 min. Separation was then performed with a steep linear gradient from 5% B to 30% B

DOI: 10.1021/acs.analchem.8b03294
Anal. Chem. 2019, 91, 3263−3269

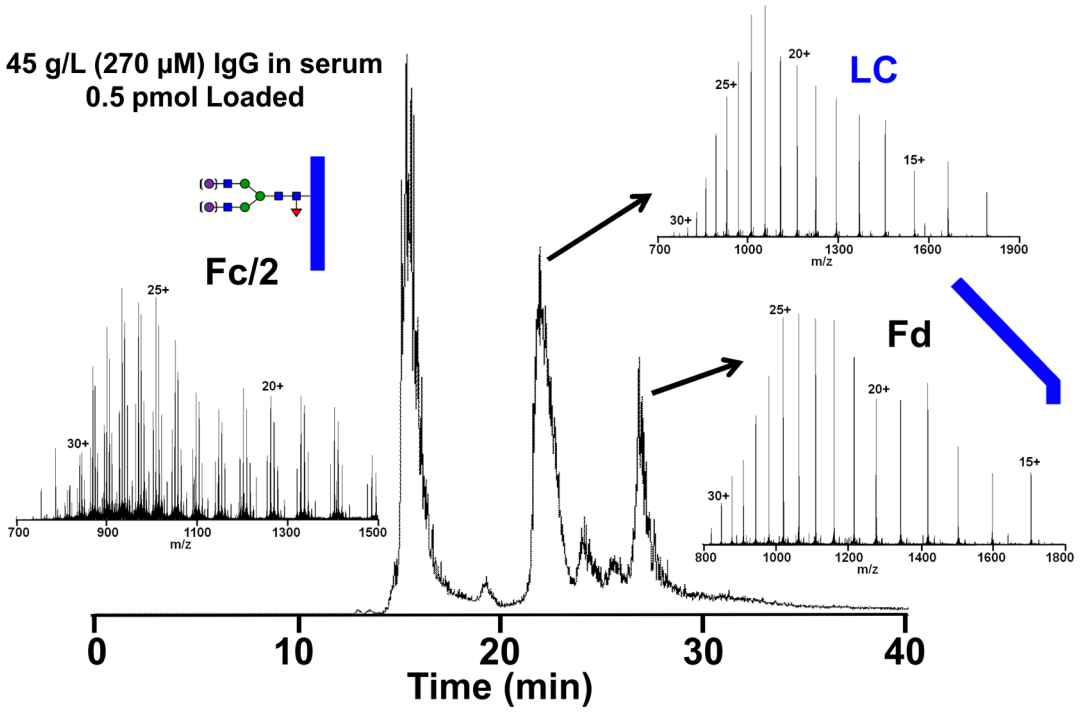


Figure 1. Total ion chromatogram for AL amyloidosis patient sample 1 subunits after IdeS digestion and TCEP reduction. Broadband mass spectra of mIg light chain, heavy chain Fc/2, and heavy chain Fd charge state distributions are shown as insets. Glycan notation is as in Figure 5. The minor peak at ∼19.3 min corresponds to a light chain with one disulfide bond reformation (isotopic distribution shifted to lower m/z by two isotopic peaks). Minor peaks between the light chain and heavy chain Fd correspond to the IdeS enzyme and slight carryover of the light chain.

over 5 min followed by a shallow linear ramp from 30% B to 60% B over 30 min at a flow rate of 0.3 μL/min.

Mass spectra were acquired with our custom-built 21 T FT−ICR mass spectrometer.[26,27] The electrospray ionization (ESI) source voltage was biased at 3 kV, and the heated capillary temperature was 325 °C. Fragmentation was performed by either electron transfer dissociation (ETD) or collision-induced dissociation (CID) in a Velos Pro linear ion trap (Thermo Fisher Scientific, San Jose, CA). CID and ETD were performed as separate LC−MS/MS experiments (isolation window width 15 Th). The automatic gain control (AGC) target was set at 1 million charges (1E6) for MS^1, 0.75 million charges (7.5E5) for CID MS/MS, and 0.2 million charges (2E5) for ETD MS/MS. AGC for ETD reagent (fluoranthene radical anions) was set to 0.4 million ions (4E5) with an ∼5 ms injection period prior to reaction with analyte at a ratio of 2:1. An external quadrupole ion storage device was used to accumulate larger ion populations (3 million charges for product ions) with multiple fills from the Velos Pro MS prior to delivery to the ICR cell. Broadband mass spectra were acquired from m/z 650−2000 with a 0.76 s time-domain acquisition period and 1 microscan, and product ion spectra were acquired from m/z 300−2000 with a 0.38 s acquisition period and 4 microscans (magnitude mode achieved baseline resolution for mIg light chain, and therefore, absorption mode with higher resolving power was not used). All data were stored as .raw files (reduced profile mode). Mass calibration was performed with myoglobin (most abundant isotopologues of the 13+ to 24+ charge states).

Data Analysis. Data were manually interpreted by use of Xcalibur 2.1 software (Thermo Fisher Scientific). The blind analysis of endogenous mIgs (from AL amyloidosis patient) sequences was performed with our recently developed mIgs sequencing algorithm.[28] Briefly, a database containing one, two, three, and four amino acid fragment exact masses was first constructed in R 3.3.2, after which mIgs light chain kappa/lambda classes were identified by MS/MS fragments matched to the putative constant region sequence in ProSight Lite with a 10 ppm fragment mass tolerance after Xtract deconvolution (Xtract deconvolution yields monoisotopic mass from averagine formula approximation, which leads to higher mass error for MS2 fragments compared to manual validation without deconvolution).[23,29] Next, framework regions (FR2 and FR3) and complementarity-determining region (CDR2) adjacent fragments (1−4 AA segment) mass differences were searched against the constructed 1−4 AA residue accurate mass database within a 10 ppm threshold, for which ppm error is defined by (adjacent fragments mass differences − amino acids candidate residue mass)/(110 Da*AA positions), approximated as 110 Da/AA. The assigned amino acids were matched to the Ig light chain germline sequence (amino acid mutations allowed) for germline sequence classification. Xtract parameters were neutral monoisotopic masses generated from deconvolution, fit factor 44%, remainder 25%, resolving power (at m/z 400) 150 000 (corresponding to the experimental time-domain data acquisition period of 0.38 s), S/N threshold 3, and max charge 25.

RESULTS AND DISCUSSION

We recently employed human serum with spiked adalimumab as a model for a plasma cell disorder to evaluate the biomarker assay, based on the exceptional mass measurement accuracy and sensitivity of nano-LC 21 T FT−ICR MS/MS.[23] The mass spectrum from a normal serum sample exhibits a broad range of charge-defined but otherwise unresolved peaks representing the normal polyclonal light chain background. In contrast,

DOI: 10.1021/acs.analchem.8b03294
Anal. Chem. 2019, 91, 3263−3269

serum spiked with adalimumab exhibits a clearly defined therapeutic monoclonal antibody (adalimumab) signal above the polyclonal background (for mAb concentration in serum as low as 0.2 μM or 0.03 g/L). After the mAb was detected, we further analyzed its sequence by nano-LC FT−ICR MS/MS experiments. The light chain fragmentation map was obtained with 72% amino acid sequence coverage, based on combined results from two targeted nano-LC FT-ICR MS/MS experiments.[30] Fragment ions generated by MS/MS of the light chain clearly match portions of both the variable region (The N-terminal 107 AA, 71% sequence coverage) and the constant region of adalimumab. Therefore, the FT−ICR top/middle-down MS/MS-based method provides in-depth mIgs marker light chain sequence information.

Classification and Characterization of Endogenous mIgs in Plasma Cell Disorder Clinical Samples. Analysis of the plasma cell disorder model shows that >85% sequence coverage was achieved in the adalimumab light chain FR2-CDR2-FR3, indicating the potential of de novo sequencing for that region. To demonstrate the feasibility of the top/middle-down de novo sequencing of endogenous mIgs, we analyzed two AL amyloidosis samples from the Mayo Clinic. For both patients, the mIg light chain gene had previously been sequenced after isolation from neoplastic plasma cells in bone marrow aspirates. The patients' mIgs gene sequences were not provided prior to MS/MS analysis.

For sample 1, LC−MS revealed the elution order and charge state distribution for the mIg light chain, heavy chain Fc/2 subunit, and heavy chain Fd subunit (Figure 1). For sample 2, LC−MS gave similar information for intact light and heavy chains. For both samples, the light chain 22+ charge state was chosen (isolation window width 15 Th) for CID (normalized collision energy 35, q-activation 0.25) and ETD fragmentation (activation period 10 ms).

Analysis of AL Amyloidosis Patient Sample 1. mIg light chain ETD and CID fragments were first searched against kappa and lambda constant region amino acid germline sequences (converted from DNA germline sequences) and matched well to the kappa germline sequence (Figure 2). The IMGT kappa germline sequence database includes 39 germline sequences, and each germline sequence differs from the others in the FR2-CDR2-FR3 region (http://www.imgt.org/3Dstructure-DB/cgi/DomainDisplay.cgi).[30] The 43rd and 44th (based on International ImMunoGeneTics (IMGT) unique numbering) amino acids are highly conserved to be "QQ" (in 33 germline sequences) or "LQ" (in 6 germline sequences), and the corresponding fragment ions were observed in high abundance. Therefore, mass differences between consecutive ETD fragments (128.05858 and 128.05858 Da) for "QQ" and (113.08406 and 128.05858 Da) for "LQ" were searched from the deconvolved mass spectrum at ~4730 Da (43 AA*110 Da/AA). After "QQ" was located for the mIg light chain from this sample, the amino acids N-/C-terminal to the initially characterized amino acids "QQ" were then assigned by matching the amino acid candidates to the adjacent fragments (separated by 1 AA) mass difference for germline sequence assignment. When sequence gaps include more than one amino acid, the gap widths (mass differences) were searched against a database containing exact masses for all combinations of 1−4 amino acid residues. If the gap mass matched multiple gap sequence candidates, the candidate sequence with the highest germline sequence similarity was selected.[28] For example, KV1-33 was

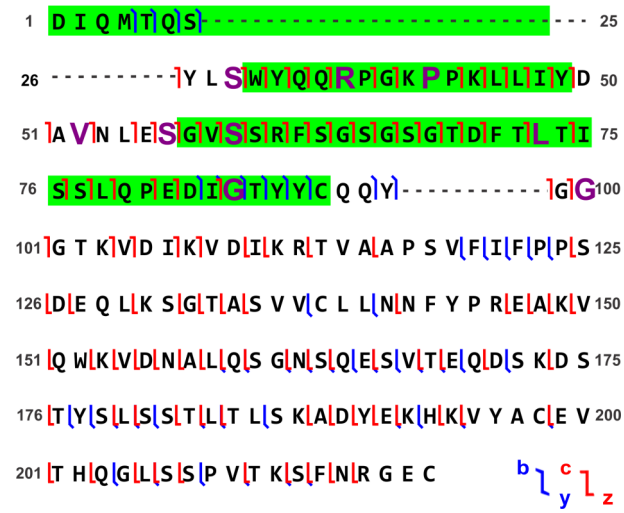


Figure 2. Top-down MS/MS sequencing of the mIg light chain of the first AL amyloidosis patient sample. The constant region matches well (70% sequence coverage) to the kappa isotype, and the variable region germline sequence is classified as KV1-33. Three framework regions are highlighted in green. Mutations are shown in purple.

assigned based on amino acids matched to adjacent fragments (gap = 1 AA), and three candidates "PQP", "KGH", and "KPP" fitted the exact mass requirement (10 ppm threshold) at positions 42−44 (from N-terminus). Because "KPP" matched 67% of the KV1-33 germline sequence ("KAP" at that position), whereas "PQP/KGH" matched only 33% of the sequence, "KPP" was selected as the sequence between c41 and c44, in the absence of additional information.

The sequence up to the N-terminal 83rd amino acid was assigned (starting from the FR2 region) by ETD c ions alone, and complementary CID fragmentation was used to further extend the amino acid assignment; the assigned c83 fragment mass was first converted to the corresponding b83 fragment mass ($\text{mass}_{\text{b-ion}} = \text{mass}_{\text{c-ion}} - 17.02655$ Da), and de novo sequencing was continued for the deconvolved CID mass spectrum. This approach yielded eight more amino acid assignments (up to N-terminal 91st amino acid). The complete sequence of FR2-CDR2-FR3 was assigned, subject to seven amino acid mutations (Figure 2), thereby confidently stratifying the M protein light chain to be the KV1-33 germline sequence.

The sequence characterized for the N-terminal portion of FR1 further confirmed the KV1-33 germline sequence assignment, and the joint region sequence was de novo characterized as the JK3 germline sequence with the P to G amino acid mutation identified (the joint region de novo sequencing started from the N-terminal of the assigned constant region sequences by converting assigned z-ion mass to its complementary c-ion mass, because the sum of the mass of c_x and the mass of z_{n-x} equals the precursor mass of a protein with n amino acids). The assigned amino acids were compared to the gene sequencing result, and 100% match for all assigned AA residues was achieved (Figure 3, top). In addition, the light chain MS1 broadband RMS error of 0.3 ppm (Figure 4) and 67% sequence coverage (Figure S1) further confirm the gene sequencing result.

The top/middle-down MS/MS can provide additional information not obtainable by gene sequencing. Figure 5 shows the heavy chain Fc/2 27+ charge state broadband

DOI: 10.1021/acs.analchem.8b03294
Anal. Chem. 2019, 91, 3263−3269

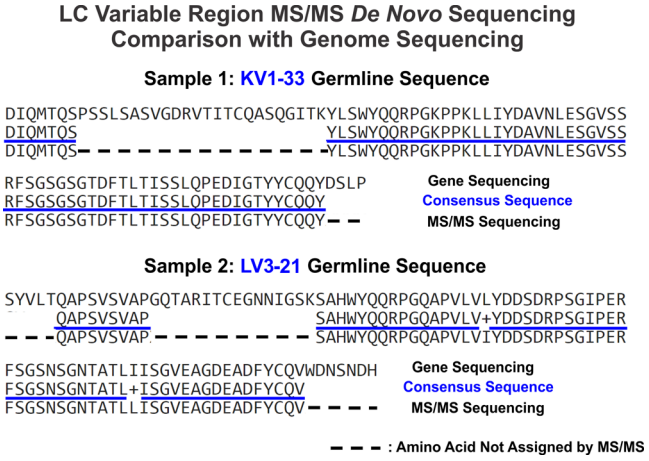


Figure 3. Samples from two AL-amyloidosis patient light chain variable regions MS/MS de novo sequences compared to the genome sequence. The consensus sequences are underlined in blue. The "+" symbol indicates that leucine, and isoleucine cannot be distinguished by mass alone.

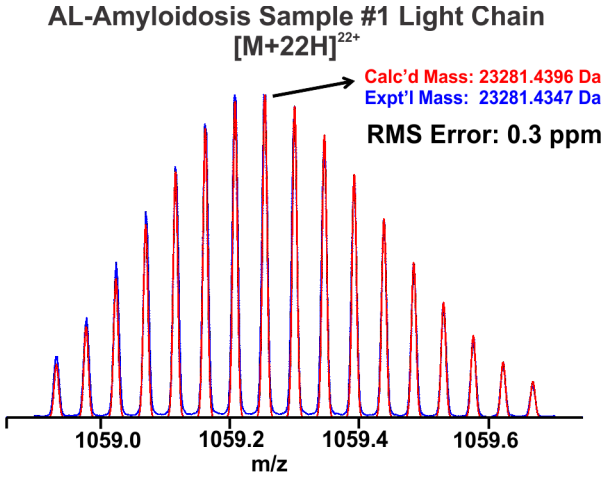


Figure 4. Mass scale-expanded segment for the mIg light chain 22+ charge state (in blue), fitted to a mass spectrum (in red) simulated from the known elemental composition from gene sequencing of the AL amyloidosis patient sample 1. The 17 highest magnitude peaks were assigned with an RMS error of 0.3 ppm.

spectrum from the first sample, and nine glycoforms are characterized with extensive loss of ammonia (either endogenous or induced by LC−MS conditions). The heavy chain was identified as human IgG2 after the G0F glycoform mass scale-expanded segment (blue) was fitted to a mass spectrum (red) simulated from the known elemental composition (human IgG2, also confirmed by a heavy chain Fd subunit C-terminal constant region MS/MS, data not shown) with a 17 peak RMS error of 0.2 ppm (Figure S2). The most abundant glycoform is G0F, followed by G1F, i.e., different from the glycoform profile for normal human IgG2 polyclonal immunoglobulins,[31,32] and that difference could potentially serve as an additional biomarker for diagnosis of plasma cell disorders.

Analysis of AL Amyloidosis Patient Sample 2. mIg light chain ETD and CID fragments matched well to the lambda constant region germline sequence LC3 (Figure S3). The IMGT lambda germline sequence database contains 33 germline sequences, and each germline sequence differs in the

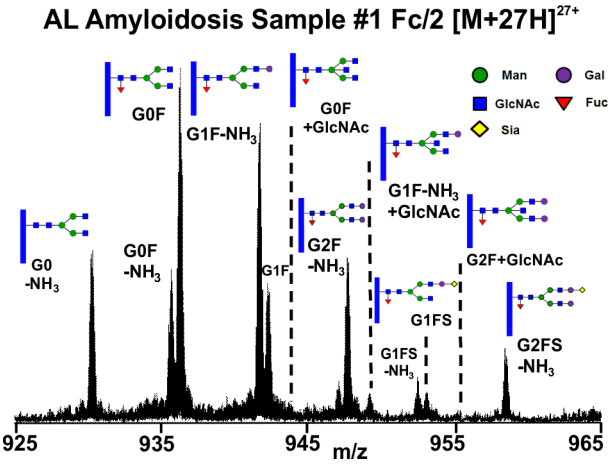


Figure 5. The mIg heavy chain glycoform profile from AL amyloidosis patient sample 1 (mass scale-expanded segment for the heavy chain Fc/2 subunit 27+ charge state). Nine glycoforms were detected, with relative abundances significantly different from the same nine glycoforms from the normal human serum.

FR2-CDR2-FR3 region.[30] The 43rd and 44th (based on IMGT unique numbering) amino acids are highly conserved to be "QQ" in all 33 germline sequences, and high abundances were observed for those fragments. Therefore, de novo sequencing began by searching mass differences between consecutive fragments as (128.05858 Da, 128.05858 Da) for "QQ" at ∼4730 Da (43 AA*110 Da/AA) with a 10 ppm threshold. After the "QQ" segment was initially identified, the fragments N-/C-terminal to the "QQ" segment were then assigned by matching the amino acid candidates to the adjacent fragment mass difference.

The sequence up to the N-terminal 83rd amino acid was assigned (starting from the FR2 region) by ETD c ions alone. Combining ETD c ions with CID b ions yielded five more amino acid assignments (up to N-terminal 88th amino acid). The complete sequence of FR2-CDR2-FR3 could then be assigned, as well as five amino acid mutations (Figure 6), thereby stratifying the M protein light chain as the LV3-21 germline sequence.

The sequence assignment in the middle of FR1 further confirmed the LV3-21 germline sequence identification, and the joint region sequence was de novo characterized as the JL2 germline sequence with K to T amino acid mutation identified. The de novo assigned amino acids were compared to the gene sequencing result, and all MS/MS-assigned amino acids matched, except for two amino acids with isomeric counterparts (L at position 46 and I at position 72), which cannot be differentiated by exact mass alone (Figure 3, bottom).

The monoisotopic mass from the light chain gene sequencing and the experimental light chain mass spectrum differs by ∼87 Da. The LV3-21 germline sequence has an N-terminal serine with a residue accurate mass of 87.03203 Da, and we hypothesize that N-terminal serine clipping occurred to the light chain as a post-translational modification. This interpretation is confirmed by the light chain MS1 broadband RMS error of 0.4 ppm (Figure S4) and 67% sequence coverage (Figure S5) after removal of the N-terminal serine. These results demonstrate that the ultrahigh mass accuracy and extensive residue cleavages combined with the high resolving power mass spectrometry yield information complementary to that from the genomic test.

DOI: 10.1021/acs.analchem.8b03294
Anal. Chem. 2019, 91, 3263−3269

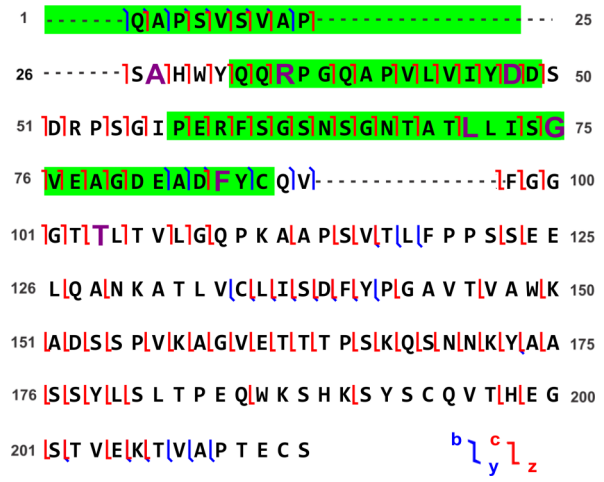


Figure 6. Top-down MS/MS de novo sequencing of the mIg light chain from the second AL amyloidosis sample. The constant region matches well (61% sequence coverage) to the lambda isotype, and the variable region germline sequence is classified as LV3-21. Three framework regions are highlighted in green. Mutations are shown in purple.

CONCLUSIONS

The present results demonstrate that nano-LC 21 T FT−ICR MS/MS provides fast and accurate classification of endogenous mIgs light chains with ultrahigh mass measurement accuracy and broad sequence coverage. The mIg light chains from two AL amyloidosis patient samples were blindly analyzed, and the MS/MS assigned sequence matched 100% to the gene sequencing result, except for two amino acids with isomeric counterparts. In addition, MS and MS/MS identify PTMs, which cannot be detected by genomic tests alone (common PTMs can be characterized by database searching or de novo sequencing). We are also able to classify the heavy chain germline sequences (HV 3-33 for the first sample and HV 1-3 for the second sample). The middle-down MS/MS approach (analyzing heavy chain subunits of ∼25 kDa) yields deeper heavy chain sequence information than top-down MS/MS (analyzing heavy chain of ∼50 kDa).[23] Because no heavy chain gene sequencing data was available for comparison (as a result of more sequence variability in heavy chain which hinders accurate primer design), database-aided de novo sequencing of heavy chain is not discussed in this article. Therefore, the mIg light chain and heavy chain germline sequence can be determined from LC−MS/MS run simultaneously.

Currently, the top-down or middle-down characterization of protein sequences is mostly based on database searching. However, immunoglobulins possess variable regions whose sequences are heterogeneous on an organism level, and a database search approach cannot yield correct identifications if the immunoglobulin sequences are not in the database, or genomic information is not available for the corresponding immunoglobulins. In addition, novel PTMs cannot be identified from the database-based approach with a limited number of recorded proteoforms. In these cases, our database-aided de novo sequencing method (∼20 min sequencing time per mIg light chain) represents the only choice.

The propensity of certain germline sequences to be more amyloidogenic and their association with organ tropism have been studied extensively.[17,33−35] For example, a monoclonal gammopathy patient with the LV6-57 germline sequence should be comprehensively evaluated for systemic immunoglobulin light-chain amyloidosis (AL_S, with a worse prognosis than for multiple myeloma), because it occurs more often in AL_S than in the normal B-cell repertoire. The KV1-33 germline sequence identified in the first sample is the most common kappa germline sequence identified in AL amyloidosis and is more likely to be associated with liver involvement. Moreover, the LV3-21 germline sequence identified in the second sample is less commonly involved in renal problems.

The gene sequencing approach has the drawback of being invasive and does not always accurately reflect the circulating immunoglobulin population, and laborious bottom-up MS/MS is limited by incomplete peptide sequence coverage and ambiguity of the origin of the assigned peptide. Therefore, the nano-LC 21 T FT-ICR top-down and middle-down MS/MS sets a new benchmark for less invasive and more accurate mIg amino acid sequencing, and we anticipate further utility of this unique instrumentation in clinical studies. This method reveals the diversity of mIgs light chain phenotypes in circulation and reflects the selection of the germline sequence by B cells as well as PTMs affected following gene transcription. Although we focused here on AL amyloidosis, the assay can be further applied to other mIg-related diseases, such as multiple myeloma, MGUS, POEMS syndrome, and Waldenstrom's macroglobulinemia (WM). The mIg concentration is below 30 g/L in MGUS and above 30 g/L in multiple myeloma.[36] The median concentration of the clonal-free light chain is 0.2 g/L from AL amyloidosis patients, where ∼50% of the patients also secrete intact mIgs (median concentration 10.8 g/L) with the same light chain sequence.[37] The median M protein concentration in WM is 4.4 g/L for African-Americans and 12 g/L for Caucasians.[38] Finally, the median M protein concentration in POEMS syndrome is 11 g/L.[39] We previously demonstrated that extensive light chain variable region cleavages were obtained at concentrations as low as 0.2 g/L monoclonal antibody adalimumab in serum (33 fmol mAb loaded, a model for clinical monoclonal gammopathy),[23] and therefore, the present approach promises to enhance our ability to type diseases based on mIg gene usage and could potentially guide future personalized therapy.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b03294.

Amino acid sequence coverage (67%) for AL amyloidosis sample 1 mIg light chain, mass scale-expanded segment for mIg heavy chain Fc/2 27+ charge state (in blue) fitted to a mass spectrum (in red) simulated from the known elemental composition (human IgG2) of AL amyloidosis sample 1, human lambda light chain constant region germline sequence database with best match to experimental MS/MS data circled in blue, mass scale-expanded segment for mIg light chain 23+ charge state (in blue) overlaid with a mass spectrum (in red) simulated from the known elemental composition from gene sequencing of AL amyloidosis sample 2, and amino acid sequence coverage (67%) for AL amyloidosis sample 2 mIg light chain (PDF)

DOI: 10.1021/acs.analchem.8b03294
Anal. Chem. 2019, 91, 3263−3269

AUTHOR INFORMATION

Corresponding Author

*Phone: +1 850 644 0529; Fax: +1 850 644 1366; E-mail: marshall@magnet.fsu.edu.

ORCID

Alan G. Marshall: 0000-0001-9375-2532

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported the National Science Foundation Cooperative Agreements DMR-11-57490 and DMR-1644779 and the State of Florida.

REFERENCES

(1) Kyle, R. A.; Rajkumar, S. V. *Best Practice & Research Clinical Haematology* **2007**, *20*, 637−664.

(2) Rajkumar, S. V.; Dispenzieri, A.; Kyle, R. A. *Mayo Clin. Proc.* **2006**, *81*, 693−703.

(3) Kumar, S. K.; Gertz, M. A.; Lacy, M. Q.; Dingli, D.; Hayman, S. R.; Buadi, F. K.; Short-Detweiler, K.; Zeldenrust, S. R.; Leung, N.; Greipp, P. R.; Lust, J. A.; Russell, S. J.; Kyle, R. A.; Rajkumar, S. V.; Dispenzieri, A. *Mayo Clin. Proc.* **2011**, *86*, 12−18.

(4) Cancer Stat Facts: Myeloma; https://seer.cancer.gov/statfacts/html/mulmy.html.

(5) Dispenzieri, A. *Am. J. Hematol.* **2017**, *92*, 814−829.

(6) Kosmas, C.; Stamatopoulos, K.; Stavroyianni, N.; Belessi, C.; Viniou, N.; Yataganas, X. *Leuk. Lymphoma* **1999**, *33*, 253−265.

(7) Kiyoi, H.; Naito, K.; Ohno, R.; Saito, H.; Naoe, T. *Leukemia* **1998**, *12*, 601−609.

(8) Rettig, M. B.; Vescio, R. A.; Cao, J.; Wu, C. H.; Lee, J. C.; Han, E.; Der Danielian, M.; Newman, R.; Hong, C.; Lichtenstein, A. K.; Berenson, J. R. *Blood* **1996**, *87*, 2846−2852.

(9) Comenzo, R. L.; Wally, J.; Kica, G.; Murray, J.; Ericsson, T.; Skinner, M.; Zhang, Y. N. *Br. J. Haematol.* **1999**, *106*, 744−751.

(10) Ramirez-Alvarado, M. *Curr. Top. Med. Chem.* **2013**, *12*, 2523−2533.

(11) Barnidge, D. R.; Lundstrom, S. L.; Zhang, B.; Dasari, S.; Murray, D. L.; Zubarev, R. A. *J. Proteome Res.* **2015**, *14*, 5283−5290.

(12) Murray, D.; Barnidge, D. *Crit. Rev. Clin. Lab. Sci.* **2013**, *50*, 91−102.

(13) Barnidge, D. R.; Tschumper, R. C.; Theis, J. D.; Snyder, M. R.; Jelinek, D. F.; Katzmann, J. A.; Dispenzieri, A.; Murray, D. L. *J. Proteome Res.* **2014**, *13*, 1905−1910.

(14) Vrana, J. A.; Gamez, J. D.; Madden, B. J.; Theis, J. D.; Bergen, H. R.; Dogan, A. *Blood* **2009**, *114*, 4957−4959.

(15) Bergen, H. R.; Dasari, S.; Dispenzieri, A.; Mills, J. R.; Ramirez-Alvarado, M.; Tschumper, R. C.; Jelinek, D. F.; Barnidge, D. R.; Murray, D. L. *Clin. Chem.* **2016**, *62*, 243−251.

(16) Dasari, S.; Theis, J. D.; Vrana, J. A.; Meureta, O. M.; Quint, P. S.; Muppa, P.; Zenka, R. M.; Tschumper, R. C.; Jelinek, D. F.; Davila, J. I.; Sarangi, V.; Kurtin, P. J.; Dogan, A. *J. Proteome Res.* **2015**, *14*, 1957−1967.

(17) Kourelis, T. V.; Dasari, S.; Theis, J. D.; Ramirez-Alvarado, M.; Kurtin, P. J.; Gertz, M. A.; Zeldenrust, S. R.; Zenka, R. M.; Dogan, A.; Dispenzieri, A. *Blood* **2017**, *129*, 299−306.

(18) Mann, M.; Kelleher, N. L. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18132−18138.

(19) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2013**, *85*, 4239−4246.

(20) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. *Anal. Chem.* **2014**, *86*, 3005−3012.

(21) Toby, T. K.; Fornelli, L.; Kelleher, N. L. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499−519.

(22) Anderson, L. C.; DeHart, C. J.; Kaiser, N. K.; Fellers, R. T.; Smith, D. F.; Greer, J. B.; LeDuc, R. D.; Blakney, G. T.; Thomas, P. M.; Kelleher, N. L.; Hendrickson, C. L. *J. Proteome Res.* **2017**, *16*, 1087−1096.

(23) He, L.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 827−838.

(24) Barnidge, D. R.; Dasari, S.; Botz, C. M.; Murray, D. H.; Snyder, M. R.; Katzmann, J. A.; Dispenzieri, A.; Murray, D. L. *J. Proteome Res.* **2014**, *13*, 1419−1427.

(25) Mills, J. R.; Kohlhagen, M. C.; Dasari, S.; Vanderboom, P. M.; Kyle, R. A.; Katzmann, J. A.; Willrich, M. A. V.; Barnidge, D. R.; Dispenzieri, A.; Murray, D. L. *Clin. Chem.* **2016**, *62*, 1334−1344.

(26) Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1626−1632.

(27) Weisbrod, C. R.; Kaiser, N. K.; Syka, J. E. P.; Early, L.; Mullen, C.; Dunyach, J. J.; English, A. M.; Anderson, L. C.; Blakney, G. T.; Shabanowitz, J.; Hendrickson, C. L.; Marshall, A. G.; Hunt, D. F. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1787−1795.

(28) He, L.; Weisbrod, C. R.; Marshall, A. G. *Int. J. Mass Spectrom.* **2018**, *427*, 107−113.

(29) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. *Proteomics* **2015**, *15*, 1235−1238.

(30) Lefranc, M. P.; Giudicelli, V.; Duroux, P.; Jabado-Michaloud, J.; Folch, G.; Aouinti, S.; Carillon, E.; Duvergey, H.; Houles, A.; Paysan-Lafosse, T.; Hadi-Saljoqi, S.; Sasorith, S.; Lefranc, G.; Kossida, S. *Nucleic Acids Res.* **2015**, *43*, D413−D422.

(31) Wuhrer, M.; Stam, J. C.; van de Geijn, F. E.; Koeleman, C. A. M.; Verrips, C. T.; Dolhain, R. J. E. M.; Hokke, C. H.; Deelder, A. M. *Proteomics* **2007**, *7*, 4070−4081.

(32) Goetze, A. M.; Zhang, Z. Q.; Liu, L.; Jacobsen, F. W.; Flynn, G. C. *Mol. Immunol.* **2011**, *49*, 338−352.

(33) Vrana, J. A.; Theis, J. D.; Dasari, S.; Mereuta, O. M.; Dispenzieri, A.; Zeldenrust, S. R.; Gertz, M. A.; Kurtin, P. J.; Grogg, K. L.; Dogan, A. *Haematologica* **2014**, *99*, 1239−1247.

(34) Baden, E. M.; Sikkink, L. A.; Ramirez-Alvarado, M. *Curr. Protein Pept. Sci.* **2009**, *10*, 500−508.

(35) del Pozo Yauner, L.; Ortiz, E.; Sanchez, R.; Sanchez-Lopez, R.; Guereca, L.; Murphy, C. L.; Allen, A.; Wall, J. S.; Fernandez-Velasco, D. A.; Solomon, A.; Becerril, B. *Proteins: Struct., Funct., Genet.* **2008**, *72*, 684−692.

(36) Durie, B. G.; Salmon, S. E. *Cancer* **1975**, *36*, 842−854.

(37) Palladini, G.; Dispenzieri, A.; Gertz, M. A.; Kumar, S.; Wechalekar, A.; Hawkins, P. N.; Schonland, S.; Hegenbart, U.; Comenzo, R.; Kastritis, E.; Dimopoulos, M. A.; Jaccard, A.; Klersy, C.; Merlini, G. *J. Clin. Oncol.* **2012**, *30*, 4541−4549.

(38) Gertz, M. A. *Am. J. Hematol.* **2017**, *92*, 209−217.

(39) Dispenzieri, A.; Kyle, R. A.; Lacy, M. Q.; Rajkumar, S. V.; Therneau, T. M.; Larson, D. R.; Greipp, P. R.; Witzig, T. E.; Basu, R.; Suarez, G. A.; Fonseca, R.; Lust, J. A.; Gertz, M. A. *Blood* **2003**, *101*, 2496−2506.

DOI: 10.1021/acs.analchem.8b03294
Anal. Chem. 2019, 91, 3263−3269