ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Research Article

Characterization of Structural Hemoglobin Variants by Top-Down Mass Spectrometry and R Programming Tools for Rapid Identification

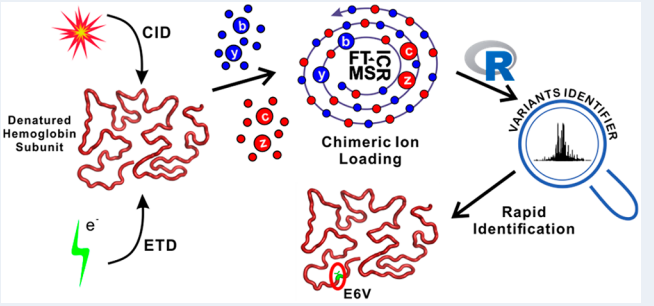
Yuan Lin, Archana M. Agarwal, Alan G. Marshall,* and Lissa C. Anderson*

Cite This: *J. Am. Soc. Mass Spectrom.* 2022, 33, 123−130

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Hemoglobinopathies are one of the most prevalent genetic disorders, affecting millions throughout the world. These are caused by pathogenic variants in genes that control the production of hemoglobin (Hb) subunits. As the number of known Hb variants has increased, it has become more challenging to obtain unambiguous results from routine chromatographic assays employed in the clinical laboratory. Top-down proteomic analysis of Hb by mass spectrometry is a definitive method to directly characterize the sequences of intact subunits. Here, we apply "chimeric ion loading" to characterize Hb β subunit variants. In this technique, product ions derived from complementary dissociation techniques are accumulated in a multipole storage device before delivery to a 21 T Fourier-transform ion cyclotron resonance mass spectrometer for simultaneous detection. To further improve the efficiency of identification of Hb variants and localization of the mutation site(s), we developed an R programming script, "Variants Identifier", to search top-down data against a database containing accurate intact mass differences and diagnostic ions from investigated Hb variants. A second R script, "PredictDiag", was developed and employed to determine relevant diagnostic ions for additional Hb variants with known sequences. These two R scripts were successfully applied to the identification of a Hb $\delta-\beta$ fusion protein and other Hb variants. The combination of chimeric ion loading and the above R scripts enables rapid and reliable interpretation of top-down mass spectrometry data, regardless of activation type, for Hb variant identification.



KEYWORDS: *top-down proteomics, Fourier-transform ion cyclotron resonance, FTMS, FT-ICR, chimeric ion loading, clinical mass spectrometry*

INTRODUCTION

Hemoglobin (Hb), the oxygen transporter, is the most abundant protein in human red blood cells (RBCs) and is composed of four globin subunits. Each subunit is bound to one heme group that reversibly binds one molecule of oxygen. Normal adult hemoglobin (Hb A) consists of two α and two β subunits. The α subunit is encoded by the two genes located on chromosome 16, and a gene on chromosome 11 encodes the β subunit.[1] Genetic disorders originating from globin gene mutations, or hemoglobinopathies, are among the most prevalent heritable diseases and affect approximately 7% of the world's population.[2] Hemoglobinopathies are mainly subdivided into two categories: (1) thalassemia syndromes, characterized by decreased production of globin subunits, and (2) structurally abnormal Hb variants.[3] For most Hb variants, the change of a single nucleotide results in one altered amino acid (AA) residue in either the α or β globin subunit. The best-known example is sickle hemoglobin (Hb S), the cause of sickle cell disease (SCD). In SCD, substitution of glutamic acid by valine at the sixth position (Glu6Val) of the β subunit results in polymerization of deoxy-Hb S, which tends to aggregate and ultimately distorts the shape of RBCs.[2,3] The sickled RBCs are more rigid and fragile, resulting in blood vessel occlusion and premature cell death leading to acute symptoms such as anemia and sickle cell crisis (pain).[2−4] Approximately 30 million people worldwide are affected by SCD (Hb S homozygotes),[5] and 300 million people carry the sickle cell trait (Hb AS heterozygotes).[6] Other clinically important Hb variants (Hb E, Hb C, Hb D, etc.) can also lead to mild to severe symptoms when combined with other inherited Hb variants or thalassemia syndromes. For instance, Hb AS and Hb AC are often asymptomatic; Hb SC, however, causes chronic hemolytic anemia and mild to moderate SCD symptoms.[2] Currently, more than 1800 Hb mutants have been

Received: September 17, 2021
Revised: December 1, 2021
Accepted: December 2, 2021
Published: December 26, 2021

https://doi.org/10.1021/jasms.1c00291
J. Am. Soc. Mass Spectrom. 2022, 33, 123−130

identified—1357 structural Hb variants and 507 thalassemia syndromes,[7] which can complicate diagnosis. Accurate diagnosis of hemoglobinopathies, even those without clinical effects, is important to control or manage potential interactions between other inherited Hb variants or thalassemia syndromes in offspring, especially in high-frequency populations.[8,9]

Of the many methods that have been applied to detect hemoglobinopathies, the most routine tests employed in clinical centers are gel electrophoresis, high-performance liquid chromatography (HPLC), and capillary electrophoresis.[1] However, these methods are not always decisive, because some Hb variants cannot be differentiated by separation alone.[10,11] For example, Hb E coelutes with 18 other variants (13 of which are also β-subunit variants) by ion-exchange HPLC.[11] Genetic testing is employed when results from standard assays are ambiguous, but this is time-consuming and expensive.

Mass spectrometry (MS) provides an alternative to gene sequencing to determine the AA sequence of Hb variants.[12] With "bottom-up" MS approaches, Hb subunits are subjected to proteolysis with sequence-specific enzymes, generating a collection of peptides. Then, peptides are typically separated by HPLC and sequence identified by tandem mass spectrometry (MS/MS).[13−16] However, not all resultant peptides are ultimately detected or successfully sequenced, and loss of sequence information is inevitable.[17,18] "Top-down" MS approaches can overcome these problems because mass and sequence information are obtained directly by MS/MS of intact proteins.[18−22]

We recently applied a top-down MS/MS approach by use of 21 tesla (T) Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS)[23] for *de novo* sequencing of Hb variants based on ultra-accurate intact mass determination and MS/MS employing collision-induced dissociation (CID) and electron-transfer dissociation (ETD).[24] CID generally cleaves the amide bonds of the backbone to generate sequence informative b-/y-type product, or fragment, ions[25,26] Conversely, ETD cleaves backbone N−C_α bonds to produce c-/$z^\bullet$-type product ions. CID and ETD MS/MS are often complementary,[27,28] and although the combined sequence coverage enabled unequivocal variant identification, it required performance of two separate MS/MS experiments.[24]

Weisbrod et al. recently introduced "chimeric ion loading", in which CID and ETD product ions derived from separate "batches" of precursor ions are accumulated in an external multipole storage device (MSD) and injected together into the ICR cell for high-resolution mass analysis.[29] In this way, both b-/y- and c-/$z^\bullet$-ions can be detected simultaneously in a single time-domain transient. Here, the "chimeric ion loading" technique is applied to diagnose Hb β variants. The ultrahigh mass measurement accuracy achieved at 21 T for both intact protein precursor and product ions along with the extensive cleavages from combined CID and ETD fragmentation enable unequivocal identification and localization of the mutated AA(s). In addition, an R programming script called "Variants Identifier" is used to identify Hb β variants by comparing fragments observed in experimental data (after m/z-to-mass deconvolution) to a manually curated database of diagnostic ions. To further expand this database, another R script, "PredictDiag", is used to predict the diagnostic product ions of Hb variants not previously analyzed by top-down MS/MS. Consequently, Hb variants, even for heterozygotes, can be rapidly and definitively characterized.

METHODS

Materials. Patient whole blood samples (Hb A and heterozygous Hb AE, AC, AD, AS, A/Riyadh, A/Abruzzo, and A/Lepore-Washington-Boston) were provided by Associated Regional and University Pathologists (ARUP) Laboratories in compliance with University of Utah (IRB_00102396) and Florida State University Institutional Review Boards. Acetic acid (99.99%) was acquired from MilliporeSigma. Water (LC-MS grade) and acetonitrile (LC-MS grade) were obtained from Honeywell Burdick & Jackson.

Sample Preparation. For Hb variant analyses, whole blood samples were centrifuged at 1000*g* for 10 min at 25 °C. The plasma was discarded, and 1 μL of the remaining aggregated RBCs was lysed by dilution with 2 mL of infusion solution (49.7% water, 50% acetonitrile, 0.3% acetic acid) and centrifuged at 13,500*g* for 5 min to pellet cell debris. The supernatant was used for direct infusion electrospray ionization (ESI) MS/MS experiments.

Mass Spectrometry. Mass spectra were acquired with our custom-built 21 T FT-ICR mass spectrometer.[23] The heated ESI probe voltage was biased at 3.7 kV, and the heated metal capillary temperature was 325 °C. Intact mass spectra (MS1) were acquired from m/z 600 to 2000 as the sum of 100 0.76 or 1.56 s time-domain transient acquisitions (corresponding to 300k and 600k resolving power at m/z 400). Product ion spectra (MS2) were acquired from m/z 300 to 2000 as the sum of 25 0.76 s time-domain acquisitions. Fragmentation was performed by ETD and CID in the high-pressure cell of a Velos Pro linear ion trap assembly (Thermo Fisher Scientific) with a 10 Th precursor isolation window. The automatic gain control target was set at 1 million charges (1E6) for MS1 and 0.2 million charges (2E5) for MS2. The automatic gain control for the ETD reagent (fluoranthene radical anions) was set to 0.4 million charges (4E5), which required an approximately 10 ms injection period prior to reacting with Hb precursor ions at a ratio of 2:1 charges for 10 ms. An external MSD was used to accumulate large populations of product ions (4 million charges for product ions) prior to high-resolution mass analysis in the ICR cell. Spectra were stored in .raw file format (Thermo Fisher Scientific) in reduced profile mode (i.e., noise baseline-subtracted). External mass calibration was performed with the Hb A α subunit (most abundant isotopologues of the 13+ to 19+ charge states) separately for the m/z ranges 300 to 2000 and 600 to 2000.

Data Analysis. Data were visualized by use of Xcalibur 3.0 software and deconvolved by Xtract (Thermo Fisher Scientific). Xtract parameters were set as follows: singly charged monoisotopic masses returned, fit factor 44%, remainder 25%, resolving power 300k (at m/z 400), signal-to-noise ratio (S/N) threshold of 3, and maximum charge of 18+ for MS2 data and 23+ for MS1 data. Fragments were matched to putative variant sequences by use of ProSight Lite[30] with a fragment mass tolerance of ±10 ppm. All assignments were manually validated.

R programming scripts were created in R3.6.1 to automate identification of Hb variant diagnostic ions and to predict diagnostic ions for untested Hb variants. The term *diagnostic ion* describes a product ion that reveals compositional information about its precursor ion.[31] More specifically, here it refers to unique product ions that distinguish Hb variant sequences and localize the mutated sites. For the R script, "Variants Identifier", a manually curated database containing

https://doi.org/10.1021/jasms.1c00291
J. Am. Soc. Mass Spectrom. 2022, 33, 123−130

intact monoisotopic mass differences (Δmass) relative to Hb A β and diagnostic ions observed in chimeric MS2 spectra of the tested ARUP Hb β variant samples (excluding Hb Lepore-Washington-Boston) was constructed. The observed Δmass value (Da) is calculated from the deconvolved MS1 spectrum and searched against the database within a user-defined mass tolerance (e.g., ±0.06 Da) to limit the search space to only those Hb variants that share a similar mass shift. Then, a table of singly charged ([MH]+), monoisotopic product ion masses and their corresponding abundances from the deconvolved MS2 spectrum is searched against the diagnostic ions corresponding to those candidates. Product ion abundances are normalized, and those present at <2% relative abundance are eliminated from consideration. For the remaining product ions, if the mass of an experimental fragment matches the mass of a diagnostic ion contained in the database within a user-defined mass tolerance (e.g., −2 to 5 ppm), then it is assigned the corresponding product ion type and AA index (e.g., y28, c26, etc.). The ppm error is defined as [(experimental deconvolved mass − monoisotopic mass of diagnostic ion)/monoisotopic mass of diagnostic ion] × 10^6. The program then returns a list of identified diagnostic ions and the Hb variants to which they correspond. A second R program, "PredictDiag", was developed to determine appropriate diagnostic ions for untested β subunit variants based on a curated reference base of manually validated product ions that are observed in deconvolved MS2 spectra and matched to the sequence of Hb A β. The accurate monoisotopic mass of each predicted diagnostic ion is then calculated for inclusion in the database employed by "Variants Identifier".

RESULTS AND DISCUSSION

Accurate Intact Mass Analysis at 21 T. Positive ESI FT-ICR MS1 spectra of intact Hb A (Figure 1) α and β subunits

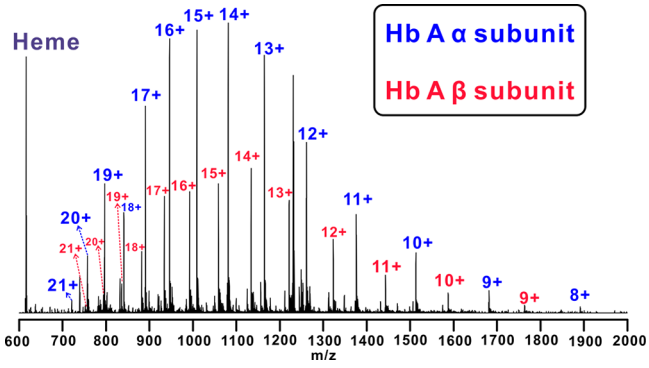


Figure 1. Positive ESI 21 T FT-ICR mass spectra of intact Hb A α and β subunits. The RMS errors in mass measurement accuracy for assigned peaks are 0.14 and 0.20 ppm, respectively.

exhibit charge states ranging from 8+ to 21+. External mass calibration employing abundant isotopologue peaks of Hb A α improved the mass measurement accuracy of assigned peaks by a factor of 2 compared to calibration with myoglobin.[24] The root-mean-square (RMS) errors for α and β subunit peak assignments were 0.14 and 0.20 ppm, based on the seven highest-magnitude isotopologue peaks for each of the 11 most abundant charge states (11+ to 21+). Ultrahigh mass measurement accuracy and resolving power afforded by 21 T FT-ICR enable unequivocal characterization of even heterozygous samples such as Hb AE, in which the monoisotopic masses of the β subunits differ by less than one Dalton (Figure 2). The partially resolved peaks distinguish the Hb A (15857.2497 Da) and Hb E (15856.3021 Da) β subunits with RMS errors of 0.29 and 1.07 ppm for definitive assignment.

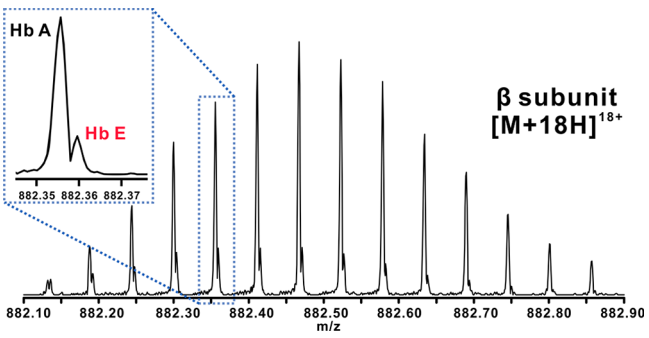


Figure 2. Isotopic distributions for the 18+ charge state of Hb AE β subunits were distinguished with a mass resolving power ($m/\Delta m_{50\%}$) of 270k and peaks assigned with RMS errors of 0.29 ppm (Hb A β) and 1.07 ppm (Hb E β).

Experimental Optimization of Precursor Charge Selection and Chimeric Ion Loading Conditions. Weisbrod et al. analyzed a complex mixture of proteins and utilized chimeric ion loading of product ions from 15 iterations of ETD and 5 iterations of CID of precursors for data-dependent acquisition of MS2 spectra.[29] Here, targeted analysis of Hb subunits afforded the opportunity to achieve optimal sequence coverage by considering appropriate selection of the precursor charge state and the ratio of iterations of ETD to CID. The influence of the precursor charge state on the degree of sequence coverage obtained from ETD of intact proteins is well understood.[27,32] However, CID of intact proteins is not as straightforward. Collision-based methods have been shown to perform better for some denatured precursors (particularly myoglobin) with lower charge.[33,34] However, experiments involving native ESI and precursors charge reduced by derivatization and proton-transfer reactions demonstrate that collisional activation of low-charged precursors results in far fewer unique backbone cleavage sites and strong residue proximity biases.[35,36] In addition, it is important to balance the number of product ion fills of the MSD derived from each activation method. Since ETD results in more residue cleavages and wider product ion charge state distributions, the spectral S/N is typically significantly lower than CID.

To optimize sequence coverage for Hb variant identification, we evaluated the 16+, 17+, and 18+ precursors of Hb A α and β and performed chimeric ion loading of ETD and CID product ions in different proportions. An untenable proportion of β subunit precursors carrying fewer than 16 charges exhibited electron-transfer without dissociation (data not shown). Moreover, the low relative abundance of precursors carrying more than 19 charges (Figure 1) caused impractically long ion accumulation periods. For all sets of conditions, sequence coverages derived from chimeric MS2 spectra following Xtract deconvolution and fragment matching with Prosight Light are given in Table S1. Minor differences in percent sequence coverage were obtained from the three precursor charge states investigated. Extensive unique backbone cleavages generated by ETD cause significant dilution of

https://doi.org/10.1021/jasms.1c00291
J. Am. Soc. Mass Spectrom. 2022, 33, 123−130

the signal for any single fragment ion. Lower S/N of c-/z$^{\bullet}$-ions necessitated the use of more fragment ion fills of the MSD to achieve a S/N comparable to CID-generated b-/y-ions. Generally, sequence coverage was poorer when fragments from one activation technique dominated the chimeric MS2 spectra (e.g., 2 iterations of CID and 18 iterations of ETD or 10 iterations of CID and ETD). For the Hb A α subunit, an optimal sequence coverage of 86% was achieved with 14 iterations of ETD and 6 iterations of CID of the 16+ precursor charge state. Sequence coverage for the β subunit was optimized to 79% with 12/8 iterations of ETD/CID of the 17+ precursor. Subsequent manual validation of this spectrum improved sequence coverage to 84%. As shown in Figure 3, complementary ETD/CID product ions confirmed nearly the complete sequence of Hb A β, so those settings were employed in subsequent experiments to identify patient Hb β variants.

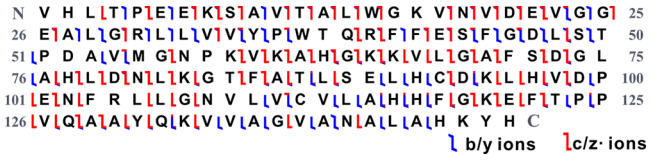


Figure 3. Amino acid sequence coverage (84% following manual validation) map for Hb A β obtained by chimeric ion loading of product ions from ETD (12 iterations) and CID (8 iterations) of the 17+ precursor charge state.

Chimeric ion loading of mixtures of ETD and CID products derived from intact proteins produces highly complex MS2 spectra. The added spectral congestion can be mitigated by high-resolution mass analysis. As shown in Figure 4a, closely spaced Hb A β fragment isotopic distributions are clearly resolved by 21 T FT-ICR MS. Overlapping isotopologue peaks for diagnostic ions (c_{26}^{2+}) from a patient Hb β variant sample (Hb AE) are partially resolved and assigned in Figure 4b. These assignments allow the substitution of a Glu to a Lys to be confidently localized to AA residue 26.

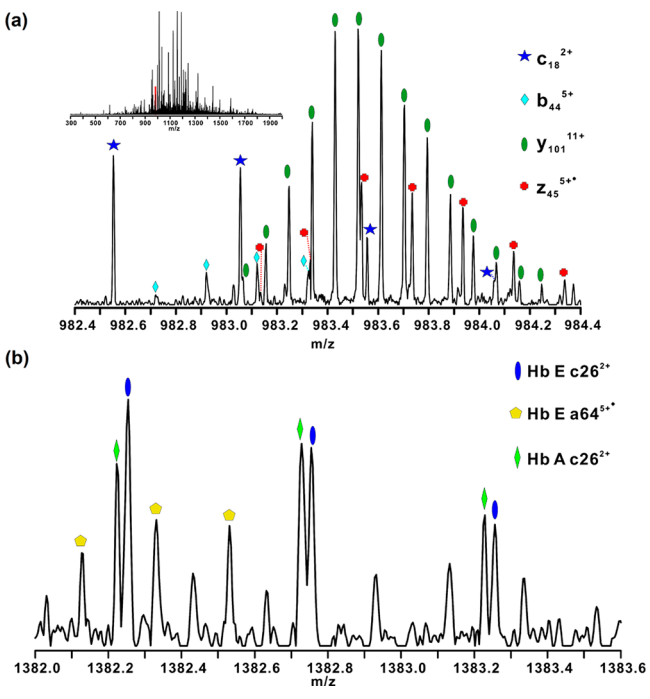


Figure 4. Scale-expanded segments of chimeric MS2 spectra for the 17+ charge state of (a) the Hb A β subunit (example shown inset with the m/z-scale-expanded segment indicated in red) and (b) the patient Hb β variant (Hb AE). Isotopic envelopes for product ions are assigned and color-coded for straightforward visualization.

Diagnostic Ion Selection for Hb Variants' Database. Previously, Lescuyer et al. developed a workflow to identify Hb variants based on selected diagnostic ions generated by top-down ETD MS/MS.[37] Reliable localization of the mutated AA residue position was achieved by visual comparison of Hb β variant spectra to Hb A β control spectra to determine the presence or absence of variant product ions matching Hb A β diagnostic ions. We took what could be considered the inverse approach—identification based on observation of diagnostic ions unique to specific Hb β variants. To simplify the process of interpreting MS2 data and ease the burden of manual validation, we extensively analyzed chimeric MS2 spectra of patient Hb β variant samples and constructed a database containing accurate monoisotopic masses ($[MH]^+$) of selected diagnostic ions. The selection criteria were: (1) diagnostic ions should contain the mutated AA residue, and (2) the AA residue index of the diagnostic ion should be less than that for the next position containing the substituted AA.

Following these criteria, potential diagnostic ion indices for substitution of each AA residue in Hb A β are listed in Table S2. For example, for substitution of the Glu at position 26 (from the N-terminus), the starting residue number of N-terminal diagnostic ions must be the same as the residue number of the substituted AA (26). The ending residue index of candidate N-terminal diagnostic ions is 42 as the next position at which a Glu occurs is residue 43. As an exception to these criteria, if two consecutive residues are the same, such as Glu in positions 6 and 7, then the ending residue number of diagnostic ions was expanded to the next position after the redundancy (i.e., Glu22). This procedure was necessary to enlarge the number of diagnostic ion candidates for identification of variants such as Hb S (Glu6Val), although technically only observation of a diagnostic ion cleaved between residues 6 and 7 can definitively site-localize that substitution. Accurate monoisotopic masses ($[MH]^+$) of diagnostic ions observed (and manually validated) in chimeric MS2 spectra of variant β subunits were compiled into a database with corresponding accurate Δmass to automate their identification. Table 1 lists examples of diagnostic ions and their corresponding accurate monoisotopic mass for each variant. All diagnostic ions in the curated database are given in Table S3.

An R programming script, "Variants Identifier", was developed to simplify and semiautomate the process of identifying Hb β variants from top-down MS/MS data. First, the user can optionally supply an observed intact Δmass value, which can be calculated from the deconvolved MS1 spectrum and lowers the probability of false positive identification. Then, the user enters a table containing monoisotopic product ion masses and abundances generated by deconvolution of the Hb variant MS2 spectrum. Variants Identifier searches product ion masses against the masses of diagnostic ions in the database corresponding to variants that share a similar intact Δmass. If the mass of an experimental product ion matches the mass of a diagnostic ion contained in the database, the diagnostic ion and Hb variant to which it corresponds are returned in a list to the user.

https://doi.org/10.1021/jasms.1c00291
J. Am. Soc. Mass Spectrom. 2022, 33, 123−130

Table 1. Example of Diagnostic Ions from a Manually Curated Database[a]

Diag. Ion	Mass (Da)	Variant	Mutation	Δmass (Da)
y8	972.5374	Hb Abruzzo	His143Arg	19.0422
y11	1199.664	Hb Abruzzo	His143Arg	19.0422
b14	1475.858	Hb C	Glu6Lys	−0.9476
b20	2159.234	Hb C	Glu6Lys	−0.9476
y28	2980.543	Hb Riyadh	Lys120Asn	−14.0520
y29	3127.611	Hb Riyadh	Lys120Asn	−14.0520
b7	776.4301	Hb S	Glu6Val	−29.9742
b8	904.5251	Hb S	Glu6Val	−29.9742
y35	3800.996	Hb D	Glu121Gln	−0.98401
z26	2792.476	Hb D	Glu121Gln	−0.98401
b26	2745.457	Hb E	Glu26Lys	−0.94763
c26	2762.484	Hb E	Glu26Lys	−0.94763

[a]Identification of Hb variants with Variants Identifier.

Figure 5a shows m/z scale-expanded segments of MS1 spectra containing α subunit 16+ charge state and β subunit

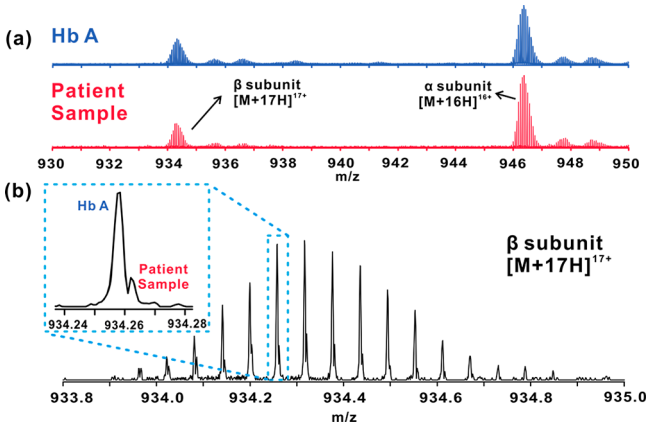


Figure 5. (a) Scale-expanded segments of MS1 spectra containing α and β subunit isotopic peak clusters for Hb A (top) and a patient variant sample (bottom). The patient α subunit peak assignments match Hb A α with 0.31 ppm RMS error (based on the seven highest-magnitude isotopic peaks for charge states 11+ to 21+). (b) Further scale-expanded segment of the patient sample MS1 spectrum shown in panel a. The isotopic distributions for β subunits at charge state 17+ are partially resolved with a mass resolving power ($m/\Delta m_{50\%}$) of 260k.

17+ charge state isotopic envelopes for Hb A (top) and a patient variant sample (bottom). The α subunit peak assignments match Hb A α. However, the presence of a β subunit variant is readily apparent upon inspection of the overlapping isotopic peaks, shown in Figure 5b. These partially resolved doublets indicate that the patient sample is heterozygous. The observed Δmass following deconvolution of the MS1 spectrum is −0.9339 Da. The subdatabase generated by Variants Identifier for a user-defined Δmass of −0.93 ± 0.06 Da is shown in Table 2. If a user does not supply a Δmass value, no subdatabase is generated, and the data are searched against all diagnostic ions in the database. The results list (Table 3) returned following analysis of MS2 data with Variants Identifier indicates that there are diagnostic ions derived from Hb C β, and the variant sample can be identified as an Hb AC heterozygote.

Prediction of Diagnostic Ions for Untested Hb Variants and Analysis of δ–β Fusion Variant. It is

Table 2. Subdatabase Produced by Variants Identifier Following Input of Experimental Δmass

Diag. Ion	Mass (Da)	Variant	Mutation	Δmass (Da)
b14	1475.8580	Hb C	Glu6Lys	−0.9476
b20	2159.2335	Hb C	Glu6Lys	−0.9476
c6	693.4406	Hb C	Glu6Lys	−0.9476
c7	822.4832	Hb C	Glu6Lys	−0.9476
c8	950.5782	Hb C	Glu6Lys	−0.9476
y35	3800.9958	Hb D	Glu121Gln	−0.9840
z26	2792.4757	Hb D	Glu121Gln	−0.9840
z28	2977.5921	Hb D	Glu121Gln	−0.9840
z29	3124.6605	Hb D	Glu121Gln	−0.9840
z30	3261.7194	Hb D	Glu121Gln	−0.9840
z40	4267.2624	Hb D	Glu121Gln	−0.9840
b26	2745.4570	Hb E	Glu26Lys	−0.9476
c26	2762.4836	Hb E	Glu26Lys	−0.9476
c29	3003.6262	Hb E	Glu26Lys	−0.9476
c39	4259.3339	Hb E	Glu26Lys	−0.9476
c40	4415.4350	Hb E	Glu26Lys	−0.9476

Table 3. List of Results from Variants Identifier

Diag. Ion	Variant	Mutation	Δmass (Da)
c7	Hb C	Glu6Lys	−0.9476
c8	Hb C	Glu6Lys	−0.9476

impossible to analyze all Hb variants by top-down MS/MS. In addition, manually curating lists of diagnostic fragments observed in experimental data for inclusion in databases like the one described here is impractical. To further automate these processes, a second R script, called "PredictDiag", was developed to predict diagnostic ions for additional Hb β variants based on a reference base containing manually validated product ions from deconvolved chimeric MS2 spectra that were matched to the sequence of Hb A β with Prosight Lite.[30] This procedure assumes that one altered AA will not impact residue cleavage sites.

The diagnostic product ions included in the PredictDiag Hb A β reference base are given in Table S4. They derive from chimeric MS2 spectra of β subunits from patient (homozygous) Hb A and from six of the seven heterozygous patient variant samples. Rapid data interpretation relies on use of deconvolution algorithms to decharge and deisotope spectra before they can be programmatically analyzed. When the monoisotopic peak of an isotopic peak cluster is of low relative abundance or isotopic peak clusters overlap, the number of heavy isotopes in a specific peak can be misinterpreted causing miscalculation of the monoisotopic mass.[38] Here, the increased complexity of mixed/heterozygous spectra can further obfuscate deconvolution and disqualify product ions that are not accurately or reproducibly matched to the sequence. To ensure the reliability of the diagnostic ions selected, only monoisotopic masses of product ions that were manually validated in raw (not deconvolved) data and matched to the sequence of Hb A β following deconvolution for a minimum of three of seven samples were included.

The reference base is used in conjunction with Table S2 as illustrated in Figure S1. First, β subunit variant sequences are supplied in the FASTA file format and compared to the sequence of Hb A β to determine the specific mutation(s), their location(s), and their intact Δmass. Then, by employing Table S2, the types (N- or C-terminal containing) and AA

https://doi.org/10.1021/jasms.1c00291
J. Am. Soc. Mass Spectrom. 2022, 33, 123−130

residue indices of potential diagnostic ions are determined for each variant sequence. If the type and index of a potential diagnostic ion (determined from the previous step) matches the type (N-terminal b-/c-type; C-terminal y-/z•-type) and index of an ion in the Hb A β reference base (Table S4), it is assumed that the backbone of the variant will similarly undergo cleavage at that position, and the resulting product ion can be used diagnostically. As shown in Figure S1a, comparison of the Hb S and A β sequences reveals that the AA substitution is Glu6Val. For substitution of Glu6, Table S2 provides indices ranging from 6 to 21 for potential N-terminal diagnostic ions and from 141 to 146 for potential C-terminal diagnostic ions (Figure S1b). These indices are searched against the Hb A β reference base, and product ions of overlapping type and index are returned as diagnostic ions. Predicted Hb S β diagnostic ions are listed in Figure S1c. Accurate monoisotopic masses of predicted diagnostic ions are calculated based on the supplied sequences, and the PredicDiag results output is formatted for direct incorporation into the Variants Identifier program.

To demonstrate the accuracy and utility of the two R scripts, PredictDiag and Variants Identifier were employed to identify the gene fusion protein Hb Lepore-Washington-Boston β subunit from top-down MS/MS data. The variant is a $\delta-\beta$ hybrid: the first 87 AA residues are identical to those of the δ subunit, and residues 116 to 146 are identical to a β subunit (residues 88−115 share 100% sequence identity). No diagnostic ions from the δ subunit were previously established, and even at ultrahigh resolving power, no Δmass could be determined from the MS1 spectrum (Figure S2), making it a difficult test case. PredictDiag was used to predict diagnostic ions for each of the variant's six substituted AA residues. The predicted diagnostic ions were added to the Variants Identifier database, which also contained the previously established diagnostic ions for the β chains of Hb C, D, E, S, Riyadh, and Abruzzo (Table S3). The deconvolved data from the chimeric MS2 spectrum of the patient variant sample were searched against the full expanded database with Variants Identifier. Eight diagnostic ions matching the sequence of Hb Lepore-Washington-Boston β were identified and are shown in Figure 6a. Although the substituted AA residues were not individually site-localized, several N-terminal fragments containing the first three substituted residues and a C-terminal fragment containing the remaining substitutions indicated the presence of the $\delta-\beta$ fusion protein in this patient variant sample. When all deconvolved product ion masses were subsequently matched to the sequences of Hb Lepore-Washington-Boston and Hb A β with Prosight Lite, as shown in Figure 6b and c, the extensive sequence coverage obtained for both forms of the β subunit enabled site-localization of 5 of the 6 substituted residues and unequivocally confirmed the diagnosis of heterozygous Hb A/Lepore-Washington-Boston.

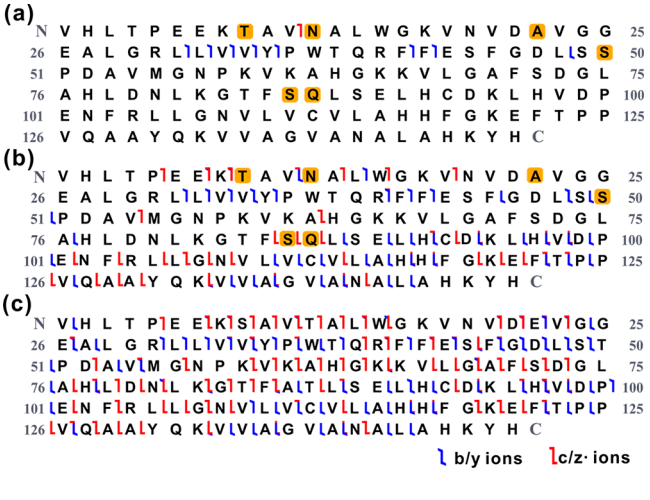


Figure 6. (a) Sequence coverage map showing predicted diagnostic ions of Hb Lepore-Washington-Boston β identified in chimeric MS2 data by Variants Identifier. AA residues that differ from Hb A β are highlighted. (b and c) Total sequence coverage maps of Hb Lepore-Washington-Boston and Hb A β subunits obtained with ProSight Lite.

CONCLUSION

Experiments were performed with a customized 21 T FT-ICR mass spectrometer, which imparts unparalleled resolving power and mass measurement accuracy to any analysis. As demonstrated, some variant β subunits require ultrahigh resolution to detect them in the presence of wildtype β (Hb AC and AE) in MS1 spectra. However, obtaining sufficient mass resolving power is not possible for all variants (as demonstrated with Hb Lepore-Washington-Boston), nor is it possible with lower-resolution mass analyzers. We also utilized a new data acquisition/ion manipulation strategy—chimeric ion loading of product ions generated by ETD and CID, which provides extensive sequence coverage of Hb variants. Although chimeric ion loading is not implemented on any commercially available mass spectrometry instrument, its use enabled consideration of a wider array of diagnostic ions for accurate variant identification and provided a rationale for development of data analysis tools that support electron- and collision-based dissociation techniques. It also reduced sample consumption, data acquisition, and analysis times by a factor of 2 relative to the performance of separate ETD and CID MS/MS experiments.[24,29]

Top-down MS/MS circumvents the significant protein inference problem associated with bottom-up MS/MS analysis of Hb variants because it reliably and directly links protein and gene sequences.[39] The time and technical skill required for sample preparation and data acquisition are also significantly reduced when Hb subunits are analyzed intact. Given the sensitivity, specificity, potential for automation, and rapid result turnaround of top-down MS/MS for diagnosis of structural Hb variants, methods that can be routinely employed in the clinical laboratory should realize significant impact. A significant hurdle to widespread adoption of top-down Hb variant analysis is a lack of available tools to aid data interpretation. Here, we developed a method combining top-down MS/MS and the R scripts "PredictDiag" and "Variants Identifier" for database construction and data interpretation to identify Hb variants rapidly and accurately from patient whole blood samples.

The method was successfully applied to diagnose patient samples containing β chain variants, including heterozygous cases in which β subunits differ in mass by less than one Dalton, and a $\delta-\beta$ fusion protein with no discernible intact mass shift detected at 21 T. Variants Identifier significantly diminished manual data interpretation and validation requirements and accurately identified Hb variant diagnostic ions. PredictDiag eliminated the need to manually curate lists of diagnostic ions specific to each Hb variant from experimental results and enabled batch processing of sequences for construction of comprehensive variant databases such as HbVar.[7] Future efforts will focus on enabling automatic

https://doi.org/10.1021/jasms.1c00291
J. Am. Soc. Mass Spectrom. 2022, 33, 123−130

detection of Hb A for heterozygous samples and reporting total sequence coverage for all variants for which diagnostic ions are identified. We also intend to package these tools into a web application to improve user-friendliness and accessibility. PredictDiag and Variants Identifier, in their current state, are open source and publicly available on GitHub (https://github.com/nhmfl-icr/VariantsID). All raw data associated with this manuscript can be found at the OSF DOI: 10.17605/OSF.IO/ZPGYU. We invite contributions from collaborators and independent investigators to employ these materials in their own research.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.1c00291.

Extended tables and figures. Sequence coverage of Hb A obtained with chimeric ion loading spectra; diagnostic ion indices for substitution of each amino acid residue in the Hb A β subunit; prediction of diagnostic ions for Hb variants; prediction of diagnostic ions for Hb variants; diagnostic fragment ions for Hb variants; diagnostic ions for the Hb A β subunit; Hb Lepore-Washington-Boston heterozygous β subunit MS1 spectrum and simulated spectrum (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Alan G. Marshall − *Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32308, United States; Ion Cyclotron Resonance Program, National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, United States;* orcid.org/0000-0001-9375-2532; Phone: +1 850-644-0529; Email: marshall@magnet.fsu.edu; Fax: +1 850-644-1366

Lissa C. Anderson − *Ion Cyclotron Resonance Program, National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, United States;* orcid.org/0000-0001-8633-0251; Phone: +1 850-644-1765; Email: anderson@magnet.fsu.edu; Fax: +1 850-644-1366

Authors

Yuan Lin − *Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32308, United States;* orcid.org/0000-0001-8653-0050

Archana M. Agarwal − *Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah 84132, United States; ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah 84108, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.1c00291

Author Contributions

All authors contributed to the conception of the experiments. Y.L. and L.C.A. performed experiments. Y.L. wrote the R scripts. A.M.A. provided clinical Hb variant samples. 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

This work was performed at the National High Magnetic Field Laboratory Ion Cyclotron Resonance User Facility, which is supported by the National Science Foundation Divisions of Chemistry and Materials Research through Cooperative Agreement No. DMR-1644779 and the State of Florida. The authors thank Drs. Christopher L. Hendrickson, Lidong He, David S. Butcher, and Chad R. Weisbrod for helpful advice and discussions.

■ REFERENCES

(1) Amitava Dasgupta, A. W. Hemoglobinopathy. In *Clinical Chemistry, Immunology and Laboratory Quality Control*; Amitava Dasgupta, A. W., Ed.; Elsevier: San Diego, 2014; Chapter 21, pp 363−390.

(2) Kohne, E. Hemoglobinopathies. *Dtsch. Arztebl.* **2011**, *108* (31−32), 532−540.

(3) Forget, B. G.; Bunn, H. F. Classification of the Disorders of Hemoglobin. *Cold Spring Harbor Perspect. Med.* **2013**, *3* (2), a011684.

(4) Thom, C. S.; Dickson, C. F.; Gell, D. A.; Weiss, M. J. Hemoglobin Variants: Biochemical Properties and Clinical Correlates. *Cold Spring Harbor Perspect. Med.* **2013**, *3* (3), a011858.

(5) Boga, C.; Ozdogu, H. Pregnancy and Sickle Cell Disease: A Review of the Current Literature. *Crit. Rev. Oncol. Hematol.* **2016**, *98*, 364−374.

(6) Grant, A. M.; Parker, C. S.; Jordan, L. B.; Hulihan, M. M.; Creary, M. S.; Lloyd-puryear, M. A.; Goldsmith, J. C.; Atrash, H. K. Public Health Implications of Sickle Cell Trait. *Am. J. Prev. Med.* **2011**, *41*, S435.

(7) Giardine, B.; Borg, J.; Viennas, E.; Pavlidis, C.; Moradkhani, K.; Joly, P.; Bartsakoulia, M.; Riemer, C.; Miller, W.; Tzimas, G.; Wajcman, H.; Hardison, R. C.; Patrinos, G. P. Updates of the HbVar Database of Human Hemoglobin Variants and Thalassemia Mutations. *Nucleic Acids Res.* **2014**, *42* (D1), 1063−1069.

(8) Wajcman, H.; Moradkhani, K. Abnormal Haemoglobins: Detection Characterization. *Indian J. Med. Res.* **2011**, *134* (10), 538−546.

(9) Weatherall, D.; Akinyanju, O.; Fucharoen, S.; Olivieri, N.; Musgrove, P. Inherited Disorders of Hemoglobin. In *Disease Control Priorities in Developing Countries*, 2nd ed.; 2006; Chapter 34, pp 663−680.

(10) Higgins, T.; Mack, M.; Khajuria, A. Comparison of Two Methods for the Quantification and Identification of Hemoglobin Variants. *Clin. Biochem.* **2009**, *42* (7−8), 701−705.

(11) Hoyer, J. D.; Scheidt, R. M. Identification of Hemoglobin Variants by HPLC. *Clin. Chem.* **2005**, *51* (7), 1303−1304.

(12) Haynes, C. A.; Guerra, S. L.; Fontana, J. C.; DeJesús, V. R. HPLC-ESI-MS/MS Analysis of Hemoglobin Peptides in Tryptic Digests of Dried-Blood Spot Extracts Detects HbS, HbC, HbD, HbE, HbO-Arab, and HbG-Philadelphia Mutations. *Clin. Chim. Acta* **2013**, *424*, 191−200.

(13) Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537* (7620), 347−355.

(14) Boersema, P. J.; Kahraman, A.; Picotti, P. Proteomics beyond Large-Scale Protein Expression Analysis. *Curr. Opin. Biotechnol.* **2015**, *34* (Aug), 162−170.

(15) Bogdanov, B.; Smith, R. D. Proteomics by FTICR Mass Spectrometry: Top down and Bottom Up. *Mass Spectrom. Rev.* **2005**, *24* (2), 168−200.

(16) Tubaon, R. M.; Haddad, P. R.; Quirino, J. P. Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016. *Proteomics* **2017**, *17* (20), 1700011.

(17) Cassidy, L.; Kaulich, P. T.; Maaß, S.; Bartel, J.; Becher, D.; Tholey, A. Bottom-up and Top-down Proteomic Approaches for the Identification, Characterization, and Quantification of the Low Molecular Weight Proteome with Focus on Short Open Reading Frame-encoded Peptides. *Proteomics* **2021**, No. e2100008.
(18) Smith, L. M.; Kelleher, N. L. Proteoform: A Single Term Describing Protein Complexity. *Nat. Methods* **2013**, *10* (3), 186−187.
(19) Cui, W.; Rohrs, H. W.; Gross, M. L. Top-down Mass Spectrometry: Recent Developments, Applications and Perspectives. *Analyst* **2011**, *136* (19), 3854−3864.
(20) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9* (9), 499−519.
(21) Mclafferty, F. W.; Bockhoff, F. M. Separation/Identification System for Complex Mixtures Using Mass Separation and Mass Spectral Characterization. *Anal. Chem.* **1978**, *50* (1), 69−76.
(22) Reid, G. E.; McLuckey, S. A. 'Top down' Protein Characterization via Tandem Mass Spectrometry. *J. Mass Spectrom.* **2002**, *37* (7), 663−675.
(23) 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. 21 T Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (9), 1626−1632.
(24) He, L.; Rockwood, A. L.; Agarwal, A. M.; Anderson, L. C.; Weisbrod, C. R.; Hendrickson, C. L.; Marshall, A. G. Diagnosis of Hemoglobinopathy and β-Thalassemia by 21 T Fourier Transform Ion Cyclotron Resonance Mass Spectrometry and Tandem Mass Spectrometry of Hemoglobin from Blood. *Clin. Chem.* **2019**, *65* (8), 986−994.
(25) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Statistical Characterization of the Charge State and Residue Dependence of Low-Energy CID Peptide Dissociation Patterns. *Anal. Chem.* **2005**, *77* (18), 5800−5813.
(26) Molina, H.; Matthiesen, R.; Kandasamy, K.; Pandey, A. Comprehensive Comparison of Collision Induced Dissociation and Electron Transfer Dissociation. *Anal. Chem.* **2008**, *80* (13), 4825−4835.
(27) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics* **2007**, *6* (11), 1942−1951.
(28) Haselmann, K. F.; Jørgensen, T. J. D.; Budnik, B. A.; Jensen, F.; Zubarev, R. A. Electron Capture Dissociation of Weakly Bound Polypeptide Polycationic Complexes. *Rapid Commun. Mass Spectrom.* **2002**, *16* (24), 2260−2265.
(29) Weisbrod, C. R.; Anderson, L. C.; Greer, J. B.; DeHart, C. J.; Hendrickson, C. L. Increased Single-Spectrum Top-Down Protein Sequence Coverage in Trapping Mass Spectrometers with Chimeric Ion Loading. *Anal. Chem.* **2020**, *92* (18), 12193−12200.
(30) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical Software to Analyze Top-down Mass Spectrometry Data. *Proteomics* **2015**, *15* (7), 1235−1238.
(31) Murray, K. K.; Boyd, R. K.; Eberlin, M. N.; Langley, G. J.; Li, L.; Naito, Y. Definitions of Terms Relating to Mass Spectrometry (IUPAC Recommendations 2013). *Pure Appl. Chem.* **2013**, *85* (7), 1515−1609.
(32) Riley, N. M.; Hebert, A. S.; Dürnberger, G.; Stanek, F.; Mechtler, K.; Westphall, M. S.; Coon, J. J. Phosphoproteomics with Activated Ion Electron Transfer Dissociation. *Anal. Chem.* **2017**, *89* (12), 6367−6376.
(33) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646−12651.
(34) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. *J. Proteome Res.* **2017**, *16* (7), 2653−2659.
(35) Haverland, N. A.; Skinner, O. S.; Fellers, R. T.; Tariq, A. A.; Early, B. P.; LeDuc, R. D.; Fornelli, L.; Compton, P. D.; Kelleher, N. L. Defining Gas-Phase Fragmentation Propensities of Intact Proteins During Native Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (6), 1203−1215.
(36) Bashyal, A.; Sanders, J. D.; Holden, D. D.; Brodbelt, J. S. Top-Down Analysis of Proteins in Low Charge States. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (4), 704−717.
(37) Coelho Graça, D.; Hartmer, R.; Jabs, W.; Beris, P.; Clerici, L.; Stoermer, C.; Samii, K.; Hochstrasser, D.; Tsybin, Y. O.; Scherl, A.; Lescuyer, P. Identification of Hemoglobin Variants by Top-down Mass Spectrometry Using Selected Diagnostic Product Ions. *Anal. Bioanal. Chem.* **2015**, *407* (10), 2837−2845.
(38) Senko, M. W.; Beu, S. C.; McLaffertycor, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6* (4), 229−233.
(39) Ntai, I.; LeDuc, R. D.; Fellers, R. T.; Erdmann-Gilmore, P.; Davies, S. R.; Rumsey, J.; Early, B. P.; Thomas, P. M.; Li, S.; Compton, P. D.; Ellis, M. J. C.; Ruggles, K. V.; Fenyö, D.; Boja, E. S.; Rodriguez, H.; Townsend, R. R.; Kelleher, N. L. Integrated Bottom-up and Top-down Proteomics of Patient-Derived Breast Tumor Xenografts. *Mol. Cell. Proteomics* **2016**, *15* (1), 45−56.

https://doi.org/10.1021/jasms.1c00291
J. Am. Soc. Mass Spectrom. 2022, 33, 123−130