N-Terminal Amino Acid Sequence Determination of Proteins by N-Terminal Dimethyl Labeling: Pitfalls and Advantages When Compared with Edman Degradation Sequence Analysis

Elizabeth Chang,[1] Sergei Pourmal,[2] Chun Zhou,[2] Rupesh Kumar,[3] Marianna Teplova,[2] Nikola P. Pavletich,[2] Kenneth J. Marians,[3] and Hediye Erdjument-Bromage[1,3,]*

[1]*Microchemistry and Proteomics Core Laboratory,* [2]*Structural Biology and* [3]*Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA*

In recent history, alternative approaches to Edman sequencing have been investigated, and to this end, the Association of Biomolecular Resource Facilities (ABRF) Protein Sequencing Research Group (PSRG) initiated studies in 2014 and 2015, looking into bottom-up and top-down N-terminal (Nt) dimethyl derivatization of standard quantities of intact proteins with the aim to determine Nt sequence information. We have expanded this initiative and used low picomole amounts of myoglobin to determine the efficiency of Nt-dimethylation. Application of this approach on protein domains, generated by limited proteolysis of overexpressed proteins, confirms that it is a universal labeling technique and is very sensitive when compared with Edman sequencing. Finally, we compared Edman sequencing and Nt-dimethylation of the same polypeptide fragments; results confirm that there is agreement in the identity of the Nt amino acid sequence between these 2 methods.

KEY WORDS: Nt protein dimethylation, Nt sequencing, bottom-up MS, in-gel digestion

INTRODUCTION

Limited proteolysis coupled to Edman degradation remains the routine approach to mapping of protein domains by biochemists and structural biologists.[1] Edman degradation yields sequential coupling and cleavage of amino acids from the N terminus of proteins and polypeptides and with picomoles of starting material, can conclusively identify 20–30 N-terminal (Nt) amino acids of a protein. Hsu *et al.*[2] published a simple method for dimethyl labeling of proteins, where N terminus and ε-amino groups of lysines are derivatized by formaldehyde through reductive amination. Recent work by Deng *et al.*[3] has outlined and compared in detail Nt labeling of proteins using N-succinimidyloxycarbonyl-methyl-tris(2,4,6-trimethyoxyphenyl) phosphonium bromide and dimethyl labeling. This year's ABRF PSRG study[4] (https://www.abrf.org/sites/default/files/temp/RGs/PSRG/abrf_psrg_poster2015_final_mcgrath.pdf and https://www.abrf.org/sites/default/files/temp/RGs/PSRG/psrg2015presentation_final.pdf) investigated the limits of detection of the Nt-dimethyl labeling method with low picomole and femtomole amounts of starting materials. We used low picomole amounts of starting materials to demonstrate the sensitivity of the chemistry. We discuss the advantages of this approach compared with Edman degradation sequencing and pitfalls in the bottom-up approach, which relies on in-gel proteolysis, purification, and identification of the dimethyl-derivatized Nt-most peptide. Here, we report applications where Nt domains from large proteins were identified and mapped using in-gel Nt-dimethyl derivatization, followed by a bottom-up mass spectrometry (MS) approach, using in-gel digestion and nano liquid chromatography (LC)-tandem MS (MS/MS) of resolved peptides.

MATERIALS AND METHODS
Materials

Quantitated horse myoglobin was donated by ABRFPSRG. It can also be purchased from Sigma-Aldrich (St. Louis, MO, USA).

Formaldehyde (36.5–38% v/v) was purchased from Sigma-Aldrich; working solution was 4% (v/v). Sodium cyanoborohydride was purchased from Sigma-Aldrich; working solution was 260 mM. Ammonium hydroxide solution (28–30%) was purchased from Sigma-Aldrich; working solution was 4% (v/v). Sodium acetate buffer solution, pH 5 (3 M), was purchased from Sigma-Aldrich; working solution was 100 mM.

FIGURE 1

SDS-PAGE analysis of sample 710SP. Band-annotated 710SP_A1 was excised and used for the Nt-dimethylation/identification experiment. Molecular weight (Mw) markers (in kilodaltons, kDa) are indicated on left side of the gel.

Protein Fragment "710SP"

This fragment of Bloom Syndrome Protein (BLM) was expressed as a glutathione-S-transferase (GST) fusion protein in insect cells, with a GST tag that was subsequently removed by tobacco etch virus protease. The resulting fragment was purified by ion exchange and gel filtration chromatography (20 mM Tris HCl, pH 8.0, 300 mM NaCl, 2 mM DTT). Its calculated molecular weight is ~40 kDa, but on a 4–12% Bis-Tris SDS-PAGE gel, it migrates ~50 kDa. Band "710SP_A1" (**Fig. 1**) was excised and subjected to Nt-dimethylation study.

Protein Fragment "826RK"

Previous work with MukB suggests a role of conformational transitions in its function.[5] Therefore, limited proteolysis of the protein was used to probe changes in the MukB conformation during various steps. The *Escherichia coli* MukB dimer (250 nM; 1.7 μg) was subjected to limited proteolysis by 625 nM trypsin (280 ng) for 15 min at 37°C in a 20 μl reaction. The digestion was carried out in a buffer containing 50 mM HEPES KOH (pH 7.5), 20 mM KCl, 10 mM DTT, 7.5% glycerol, 0.5 mM Mg(OAc)$_2$. The reaction was terminated by the addition of SDS-PAGE sample loading buffer, followed by heating at 95°C for 10 min. The reaction products were resolved on a 5% SDS polyacrylamide gel and visualized by staining in colloidal SimplyBlue SafeStain (Thermo Fisher Scientific, Grand Island, NY, USA). One of the products from the limited proteolysis of MukB, "826RK_A2," was excised and subjected to Nt identification by Nt-dimethylation (**Fig. 2**).

Protein Fragment "831MT"

The human Fragile X Mental Retardation Protein (FMRP) fragment (aa 1-444, excluding aa 331-396) was cloned between *Bam*HI and *Xho*I restriction sites into a modified pRSFDuet vector with a His$_6$-Sumo tag at the N terminus. The extra Nt amino acid, serine at position 1, results from the proteolytic removal of the Nt His$_6$-Sumo tag with ubiquitin-like-specific protease 1. During gel electrophoresis of the cloned protein, 2 bands were detected: 1 at predicted molecular weight (~43.7 kDa) of the cloned, full-length protein, annotated as "831MT_F1," and a faster migrating band at an apparent M_r ~37 kDa, annotated as "831MT_T1" (**Fig. 3**). Both bands were excised and subjected to Nt-dimethylation study. To carry out a direct comparison of the result obtained from dimethylation of the N terminus, the same sample was transferred onto a PVDF membrane (Bio-Rad Laboratories, Hercules, CA, USA) in 25 mM Tris-glycine/20% ethanol buffer using a GENIE blotter (Idea Scientific, Minneapolis, MN, USA), according to the manufacturer's instructions for 40 min. After transfer, while the PVDF membrane was still wet, it was stained briefly with Coomassie Blue (~2 min), and the unspecifically stained background was removed with 50% (v/v) methanol. Stained membrane was imaged after drying (**Fig. 4**), and the 2 PVDF-bound protein bands, "831MT_F_PVDF" (full-length) and "831MT_T_PVDF" (truncated), were excised and subjected to Edman chemical sequencing.
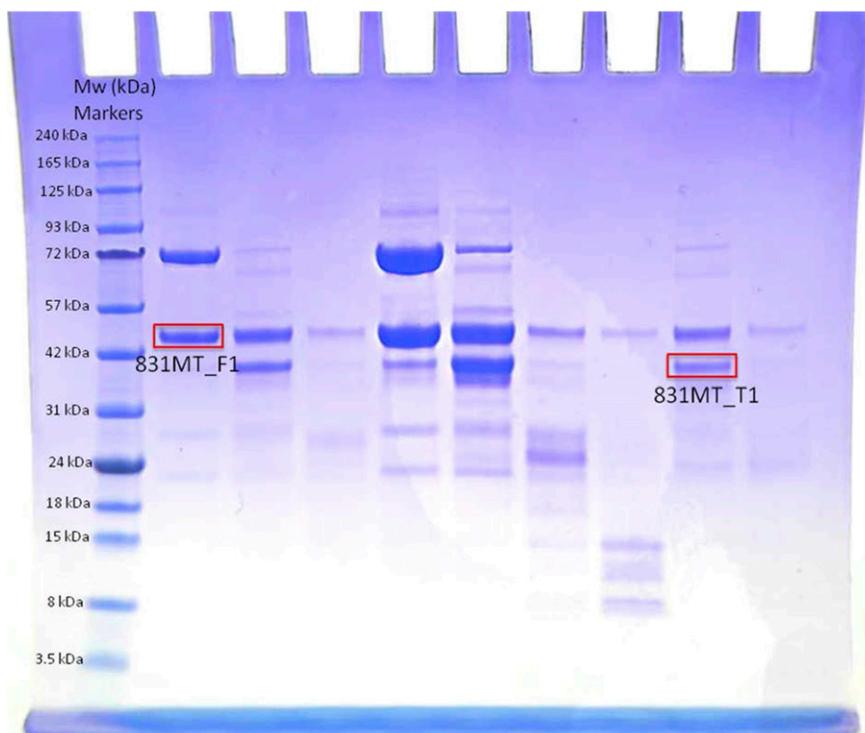
In-Gel Dimethylation

Post SDS-PAGE, gel-bound protein of interest was excised and destained and SDS removed by washes with 50% methanol (v/v) and if necessary, 200 mM ammonium bicarbonate/acetonitrile,

FIGURE 2

SDS-PAGE analysis of sample 826RK. Band-annotated 826RK_A2 was excised in 3 lanes, as indicated, and used for the Nt-dimethylation/identification experiment. Mw markers (in kDa) are also indicated on left side of the gel.

50% (v/v). Reduction and alkylation of the cysteinylated protein were done as follows: to dried, gel-bound protein, freshly made 10 mM DTT in 25 mM ammonium bicarbonate was added until the pieces were covered. Reduction of disulfide bonds was allowed to proceed for 1 h at 56°C. Following, reducing agent was removed, and freshly made 55 mM iodoacetamide in 25 mM ammonium bicarbonate was added to gel-bound protein. Alkylation took place in the dark at room temperature for 45 min, after which

time, the supernatant was removed and gel pieces rinsed with 100 mM ammonium bicarbonate, 3 times. Finally, gel pieces were dried using a SpeedVac. For the dimethylation reaction, gel pieces were reconstituted in 25 μl 100 mM sodium acetate, pH 5. In-gel dimethylation is carried out following a protocol by Hsu et al.,[2] as optimized by the PSRG Study Group.[4] In summary, to the gel-bound protein in 100 mM sodium acetate, a mixture of 1 μl 4% formaldehyde/1 μl 260 mM sodium cyanoborohydride was added, and the

FIGURE 3

SDS-PAGE analysis of sample 831MT; Bands annotated 831MT_F1 and 831MT_T1, as indicated, and were used for the Nt-dimethylation/identification experiment. Mw markers (in kDa) are also indicated on left side of the gel.

FIGURE 4

Coomassie Blue staining of electroblotted 831MT on PVDF; bands annotated 831MT_F_PVDF and 831MT_T_PVDF were excised, as indicated, and placed in the reaction cartridge of an automated sequencer for the Nt Edman degradation/identification experiment. Mw markers (in kDa) are also indicated on left side of the gel.

reaction was allowed to proceed for 20 min. Dimethylation was stopped by removing all of the solution above the gel piece and by rinsing the sample with 50 μl 4% ammonium hydroxide; the latter served as a quenching agent to ensure that no additional dimethylation was taking place by any remaining dimethylation reactants. Gel pieces were rinsed further with 200 μl 100 mM ammonium bicarbonate, a total of 5 times, removing liquid after each wash. Finally, they were lyophilized to dryness and stored until the in-gel digestion step.

In-Gel Digestion

To dried, gel-bound protein, 200 ng trypsin (modified sequencing grade; Promega, Madison, WI, USA) in 10 μl 100 mM ammonium bicarbonate was added. An additional volume of ammonium bicarbonate was added until the gel piece was covered with the digestion solution; this volume is usually 25–50 μl, depending on the acrylamide volume (surface area) of the original gel piece that was excised for analysis. Tryptic digest was allowed to proceed at 37°C for 2 h. Generated peptides were removed and "cleaned up"

FIGURE 5

LC-MS/MS analysis of dimethyl-labeled myoglobin; ~1 pmol, gel-bound sample was analyzed by nano LC-MS/MS. A) Base peak over 50 min acetonitrile gradient. B) Extracted ion chromatogram of underivatized Nt peptide at m/z: 908.455. C) Nt-dimethylated peptide at m/z: 922.471. D) Nt-dimethylated and deamidated peptide at m/z: 922.968. E) Nt-dimethylated peptide with 2 deamidations at m/z: 923.459.
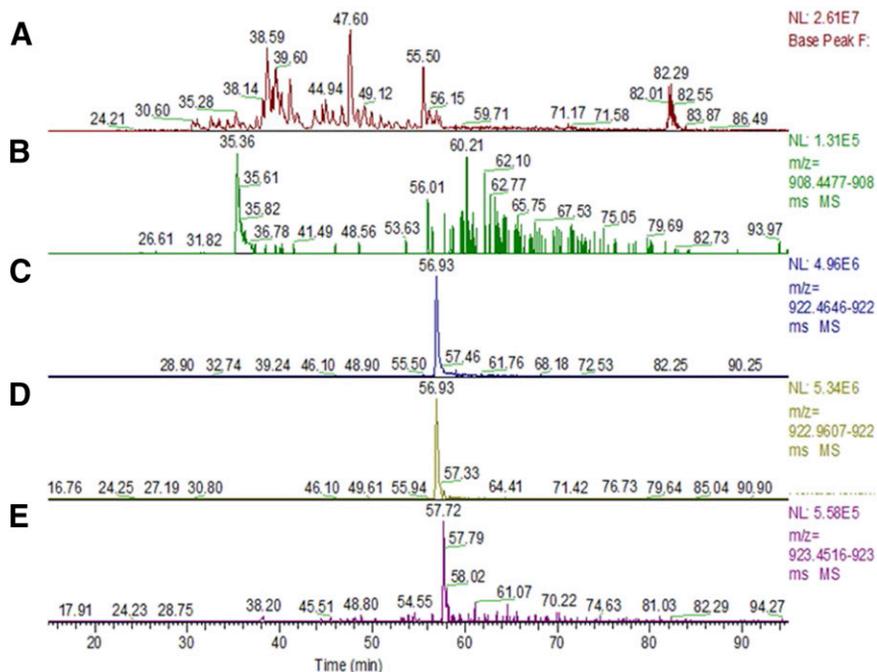
TABLE 1

Extracted Ions for Nt-Dimethylated and -Underivatized GLSDGEWQQVLNVWGK

Amino acid sequence	M_r (calc)	Total measured AUCs					
		2.5 pmol			0.5 pmol		
	m/z (Obsrv), elution time	A1	A2	A3	B1	B2	B3
[1]GLSDGEWQQ VLNVWGK[16]	1814.895 908.455 (2+), 57 min	2.64E + 07	3.75E + 07	2.53E + 07	2.83E + 06	1.67E + 05	3.56E + 06
Dimethyl GLSDGEWQQ VLNVWGK[16]	1842.926 922.471 (2+), 58 min	1.03E + 09	1.16E + 09	1.25E + 09	7.06E + 07	1.02E + 07	1.17E + 08
Dimethyl GLSDGEWQQ VLN$_{Deam}$VWGK[16]	1843.910 922.9681 (2+), 58 min	1.18E + 09	1.46E + 09	1.48E + 09	6.44E + 07	7.85E + 06	1.13E + 08
Dimethyl GLSDGEWQ$_{Deam}$Q VLN$_{Deam}$VWGK[16]	1844.895 923.459 (2+), 59 min	1.70E + 08	3.18E + 08	3.06E + 08	8.79E + 06	9.27E + 05	1.72E + 07

Extracted ion current measurements, expressed as AUC for the myoglobin Nt-dimethylated and -underivatized peptide GLSDGEWQQVLNVWGK. Xcalibur analysis (version 2.1.0) was used to isolate experimental masses (m/z; charge: 2+) for the following ions: underivatized, Nt-dimethylated, and Nt-dimethylated peptide with 1 deamidated (Deam) asparagine and with 2 deamidated residues—asparagine and glutamine. For these quantitative measurements, MS1 window with mass accuracy of 5 ppm was used. A1, A2, and A3, Gel-bound myoglobin samples, where half of 5 pmol starting material digest was analyzed by LC-MS/MS. B1, B2, and B3, analysis half of 1.0 pmol myoglobin digests. calc, calculated; obsrv, observed.

using Poros R2 beads, as the standard protocol described.[6] Other peptide clean-up chromatographic tools, such as Empore C18 Extraction Disks (3M, St. Paul, MN, USA) or Zip Tips (EMD Millipore, Billerica, MA, USA), can be substituted successfully at this step.

LC Coupled to MS/MS

Peptides were eluted with 3 µl 40% (v/v) acetonitrile containing 0.1% (v/v) formic acid and diluted to 20 µl with 0.1% formic acid for nano LC-MS coupled to MS/MS. If not analyzed immediately, the peptide pools are stable when kept in organic buffer (40% acetonitrile/0.1% formic acid) and stored at −80°C. Peptide mixtures were loaded onto a trapping guard column (0.3 × 5 mm Acclaim PepMap 100 C18 cartridge; LC Packings, Sunnyvale, CA, USA) using the Nano MDLC system (Eksigent Technologies, Dublin, CA, USA) at a flow rate 20 µl/min. After washing, the flow was reversed through the guard column and the peptides eluted with a 5–45% acetonitrile gradient over 85 min at a flow rate of 200 nl/min, onto and over a 75 µm × 15 cm-fused silica capillary PepMap 100 C18 column (LC Packings). The eluent was directed to a 75 µm (with 10 µ orifice)-fused silica nano-electrospray needle (New Objective, Woburn, MA, USA). The electrospray ionization needle was set at 1800 V. A linear ion trap quadrupole (LTQ) Orbitrap hybrid analyzer (Thermo Fisher Scientific, San Jose, CA, USA) was operated in automatic, data-dependent MS/MS acquisition mode with 1 MS full scan [m/z 450–2000] in the Orbitrap analyzer at 60,000 mass resolution and up to 10 concurrent MS/MS scans in the LTQ for the 10 most intense peaks selected from each survey scan. Survey scans were acquired in profile mode, and MS/MS scans were acquired in centroid mode. The collision energy was automatically adjusted in accordance with the experimental mass (m/z) value of the precursor ions selected for MS/MS. Minimum ion intensity of 2000 counts was required to trigger an MS/MS spectrum; dynamic exclusion duration was set at 60 s.

After LC-MS/MS analysis, database searches in SwissProt 2014 containing 546,000 protein entries were carried out using Mascot (Matrix Science, Boston, MA, USA), version 2.5.00,[7] with the following search parameters: 1) 2 missed cleavage tryptic sites allowed, 2) precursor ion mass tolerance = 10 ppm, and 3) fragment ion mass tolerance = 0.8 Da; variable peptide modifications were allowed for methionine oxidation, deamidation of asparagine to glutamine, protein Nt acetylation, and protein and peptide Nt and lysine dimethylation. We carried out additional searches with N-formyl as a modification, as dimethylation and formylation modifications are close in resulting mass difference (δ mass). If reduction and alkylation were done, then carbamidomethylation of cysteine was used as a fixed modification. A decoy database search was always activated, and in general, with $P < 0.01$, a false discovery rate (FDR) averaged <1%.

T A B L E 2

Estimation of Nt-Dimethyl Labeling

	Myoglobin amount derivatized					
	2.5 pmol			0.5 pmol		
Summed AUCs of Nt peptides	A1	A2	A3	B1	B2	B3
Dimethylated	2.38E + 09	2.94E + 09	3.04E + 09	1.44E + 08	1.90E + 07	2.47E + 08
Nondimethylated	2.64E + 07	3.75E + 07	2.53E + 07	2.83E + 06	1.67E + 05	3.56E + 06
% Derivatized Nt peptides	98.9%	98.7%	99.2%	98.1%	99.1%	98.6%

Relative Nt-dimethylation of myoglobin was estimated for 2.5 and 0.5 pmol protein by summing extracted ion for unmodified peptide GLSDGEWQQVLNVWGK and Nt-labeled GLSDGEWQQVLNVWGK with deamidated species detected at times of elution (min) indicated. Therefore, the final ratio accounts for all of the unmodified, as well as dimethyl-labeled, peptides.

Scaffold (Proteome Software, Portland, OR, USA), version 4.4.1, was used to validate and cross-tabulate further the MS/MS-based peptide and protein identifications. Protein and peptide probability was set at 95%, with a minimum peptide requirement of 1.

We also subjected experimental data to de novo sequencing analysis using "PEAKS" (version 7.5; Bioinformatics Solutions, ON, Canada) to see if we could identify the same Nt peptides identified by the Mascot search algorithm without using a sequence database.[8] The same search parameters used for the Mascot search were used for the PEAKS de novo sequence analysis.

Nt Sequence Analysis

PVDF-bound protein fragments, at an apparent M_r ~44 kDa for the full-length-expressed protein 831MT_F_PVDF and apparent M_r ~38 kDa for the truncated expressed protein 831MT_T_PVDF were excised and subjected to 8 cycles of Nt sequence analysis using a Procise 494 protein sequencer, as described in Tempst et al.[9]

FIGURE 6

Representative MS/MS of unmodified myoglobin tryptic peptide (A) and N-terminally dimethylated tryptic peptide (B). C) PEAKS de novo sequencing result confirming Nt-dimethylated peptide GLSDGEWQQVLNVWGK for myoglobin sample. D) Representation of the experimental data that identified horse myoglobin. All green residues have been confirmed with >95% probability by the Scaffold FDR algorithm. There is only 1 tryptic peptide that identifies the Nt-dimethylation, peptide-spanning aa 2–17. Myoglobin consists of 19 lysines; experimental data map, 15 lysines; 9/15 lysines are found to be dimethylated (see mapped, green regions with "Dimethyl K" boxes); nearly all dimethylated lysines are found to be internal (miscleaved) lysines.

C

D
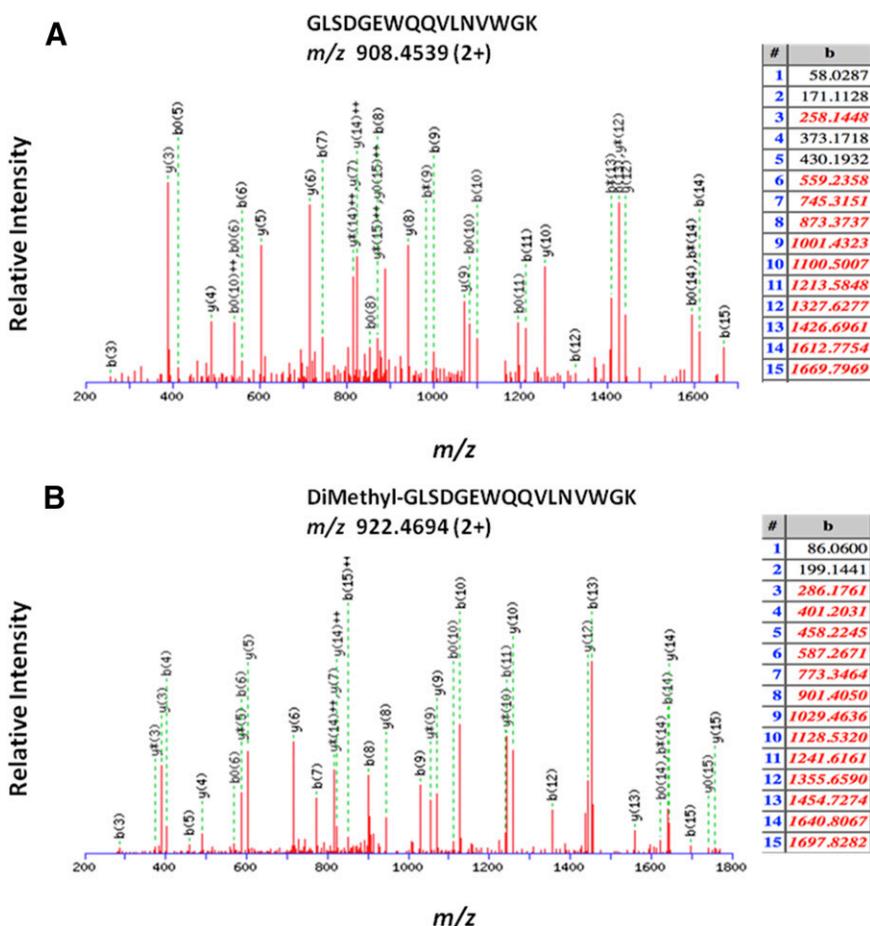
FIGURE 6–(*continued*)
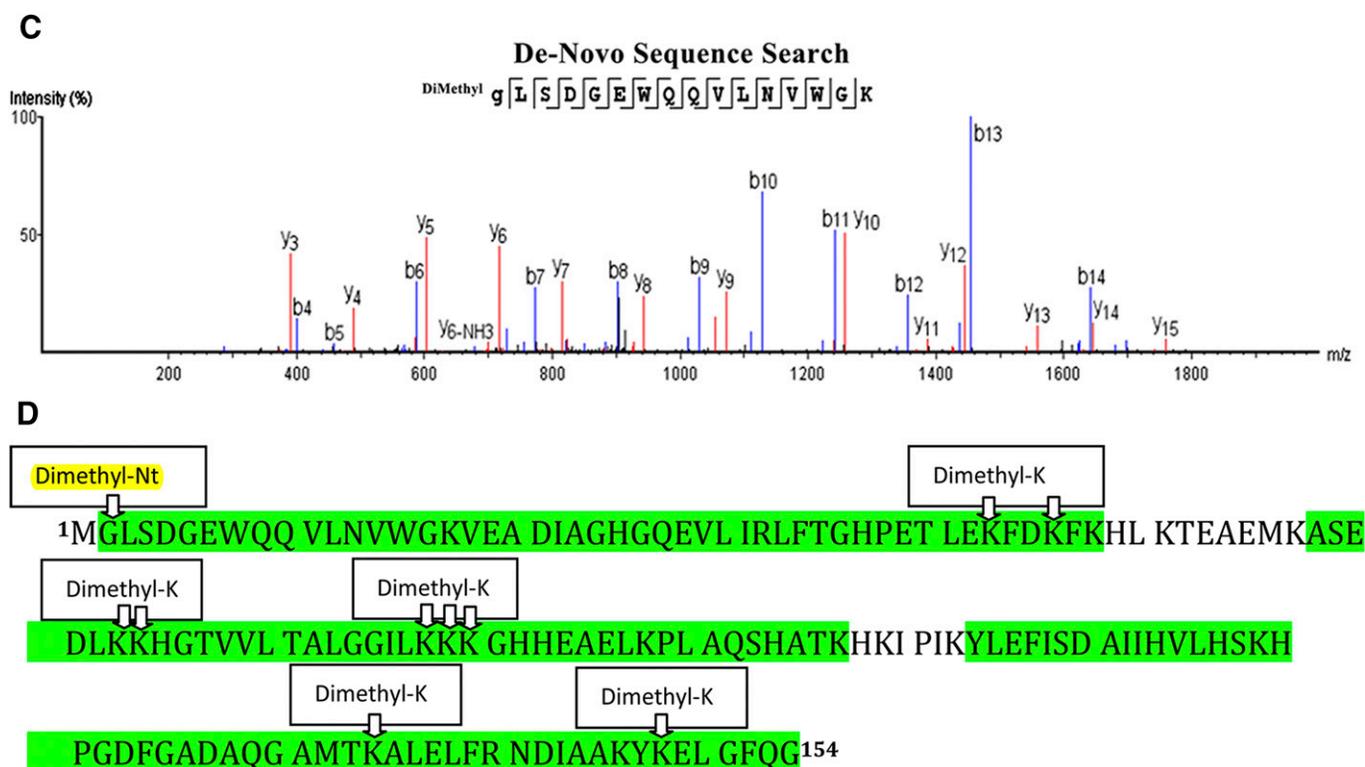
RESULTS AND DISCUSSIONS

Standard Amounts of Myoglobin Dimethylation

Amounts (5 and 1 pmol) of myoglobin were gel purified and subjected to in-gel dimethylation, followed by in-gel proteolysis. One-half of the generated peptide amounts was LC-MS/MS analyzed in triplicate, to assess the efficiency Nt dimethylation. Relative quantities of the underivatized, as well as dimethylated Nt tryptic peptide "GLSDGEWQQVLNVWGK," identified and found in each set, were assessed by integrating the area under the curve (AUC) for all the derivatized, as well as the underivatized, peptide populations (**Fig. 5** and **Table 1**). Extracted ions for triplicate experiments were consistent as shown; estimation of Nt-dimethylation averaged at 98.8% and 98.4% for 2.5 and 500 fmole triplicate analyses sets, respectively (**Table 2**).

Nt-dimethylation did not cause an adverse effect on the detection and final identification of the generated peptide, as can be confirmed by the consistent ion series obtained for MS/MS of the underivatized and Nt-dimethylated myoglobin tryptic peptide GLSDGEWQQVLNVWGK (**Fig. 6A**, **B**, respectively). Nt-dimethylation of myoglobin was confirmed by visual inspection of the MS/MS data, where nearly all of the Nt-originating b-ion series[10] were found (Fig. 6A, B). We also carried out additional de novo sequence searches of the raw MS data, which resulted in the same Nt-dimethylated amino acid sequence of the myoglobin peptide (Fig. 6C and Supplemental Table 1). Experimental data obtained on the gel-bound, in-gel

dimethylated myoglobin, mapping to the published myoglobin sequence, is shown in Fig. 6D; there were many peptides that were identified to have internal dimethylated lysine residues, as documented in Supplemental Table 2).

710SP: Nt Dimethylation Result

We detected and identified 2 N-terminally dimethylated peptides; "dimethyl MAAVPQNNLQEQLER" (**Fig. 7A**) and "dimethyl AAVPQNNLQEQLER" (Fig. 7B). Both of these peptides correspond to the published sequence of human Bloom syndrome protein, BLM_HUMAN (M_r ~158,901 Da), spanning aa 1–15 and 2–15, respectively. Experimental data obtained on the gel-bound, in-gel, dimethylated 710SP fragment indicate that this fragment is an N-terminal-most fragment of published Bloom syndrome protein (BLM; Fig. 7C). When we inspect all of the dimethylated peptides found in this experiment, we find 4 peptide counts for dimethyl MAAVPQNNLQEQLER and 5 peptide counts for dimethyl AAVPQNNLQEQLER. There were many peptides that were identified to have dimethylated lysine residues (see Supplemental Table 3). Nt-dimethylation and Nt identification of the protein fragment are confirmed by visual inspection of the MS/MS spectrum.

826RK: Nt Dimethylation Result

We identified tryptic peptide with Nt-dimethylation: "dimethyl ELAEHNGAEGDLEADYQAASDHLNLVQTALR"
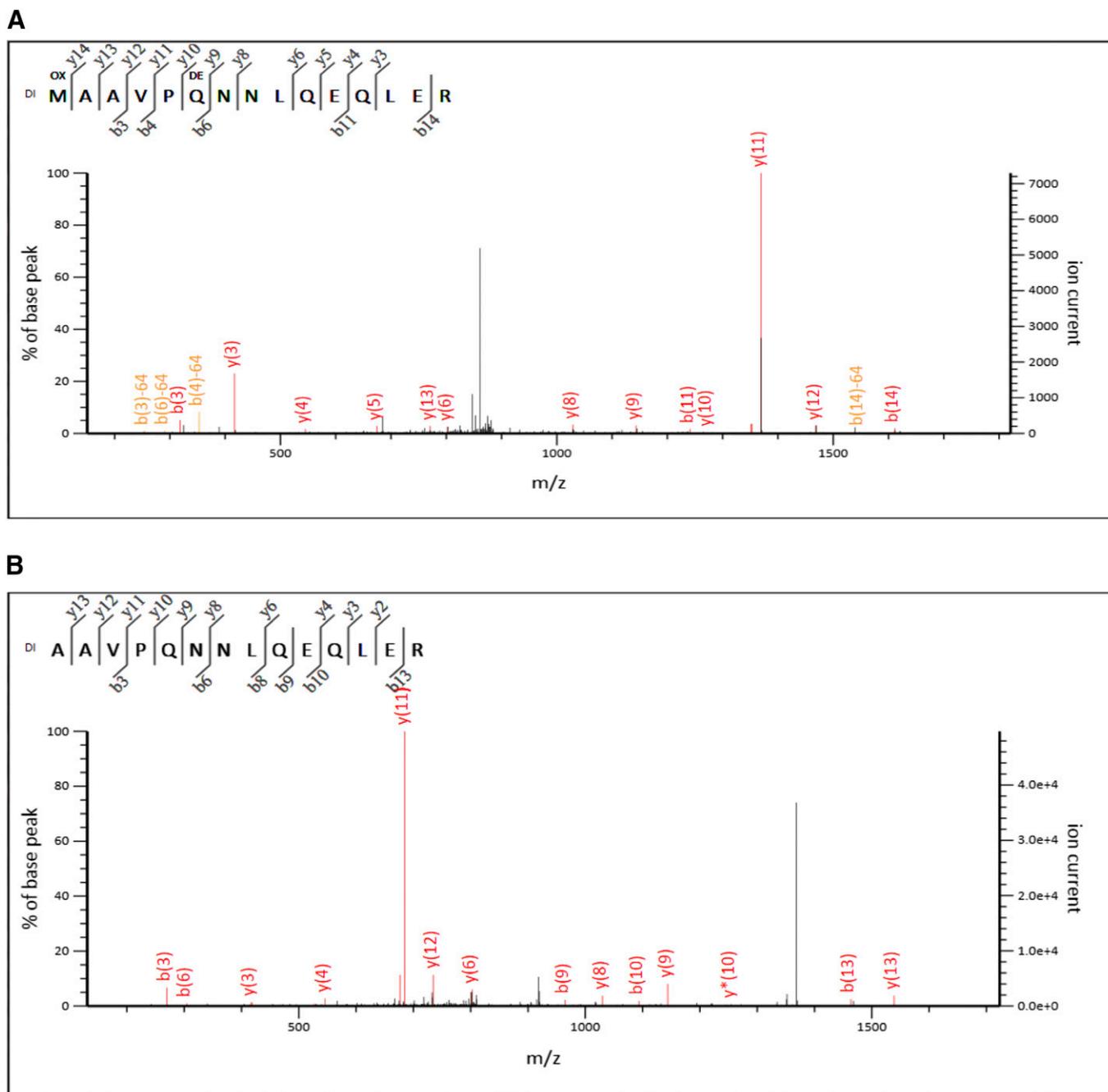
FIGURE 7

MS/MS analysis and identification of the Nt tryptic peptide a dimethyl MAAVPQNNLQEQLER (*A*) and dimethyl AAVPQNNLQEQLER (*B*) in sample 710SP_A1. *C*) Representation of the 710SP_A1 experimental data that identified the fragment of BLM. All green residues have been confirmed with >95% probability by the Scaffold FDR algorithm. There are 2 tryptic peptides that identify the Nt-dimethylation: peptides spanning aa 1–15 and 2–15. BLM protein has 123 lysine residues. Experimental data map 27 lysines (out of 35 that are present in the proteolytic fragment of BLM); 15/27 lysines are found to be dimethylated (see mapped, green regions with Dimethyl K); and nearly all are internal (miscleaved) lysines.

(**Fig. 8***A*), which is 100% identical to the published sequence of *E. coli* chromosome partition protein, MukB (Uniprot Accession MUKB_ECOBW) (M_r ~170,696 Da), spanning aa 315–345 of the published sequence. This result suggests that 826RK_A2 probably contains a fragment of mukB, with the first amino acid at 315; experimental data obtained on the gel-bound 826RK_A2, mapping to the published MukB sequence, are shown in Fig. 8*B*. We identified an overwhelming number of 87 peptides with "dimethyl **E**LAEHNGAEGDLEADYQAASDHLNLVQTALR." There were also internal-dimethylated, lysine-containing peptides that were identified, as documented in Supplemental Table 4.

C

Dimethyl-Nt

Dimethyl-Nt   Dimethyl K   Dimethyl K   Dimethyl K

¹MAAVPQNNLQ EQLERHSART LNNKLSLSKP KFSGFTFKKK TSSDNNVSVT NVSVAKTPVL

Dimethyl K   Dimethyl K   Dimethyl K

RNKDVNVTED FSFSEPLPNT TNQQRVKDFF KNAPAGQETQ RGGSKSLLPD FLQTPKEVVC

Dimethyl K   Dimethyl K

TTQNTPTVKK SRDTALKKLE FSSSPDSLST INDWDDMDDF DTSETSKSFV TPPQSHFVRV

Dimethyl K

STAQKSKKGK RNFFKAQLYT TNTVKTDLPP PSSESEQIDL TEEQKDDSEW LSSDVICIDD

Dimethyl K   Dimethyl K

GPIAEVHINE DAQESDSLKT HLEDERDNSE KKKNLEEAEL HSTERVPCIE FDDDDYDTDF

Dimethyl K   Dimethyl K

VPPSPEEIIS ASSSSSKCLS TLKDLDTSDR REDVLSTSKD LLSKPEK³⁴⁷MSM QELNPETSTD

CDARQISLQQ QLIHVMEHIC KLIDTIPDDK LKLLDCGNEL LQQRNIRRKL LTEVDFNKSD ASLLGSLWRY RPDSLDGPME GDSCPTGNSM
KELNFSHLPS NSVSPGDCLL TTTLGKTGFS ATRKNLFERP LFNTHLQKSF VSSNWAETPR LGKKNESSYF PGNVLTSTAV KDQNKHTASI
NDLERETQPS YDIDNFDIDD FDDDDDWEDI MHNLAASKSS TAAYQPIKEG RPIKSVSERL SSAKTDCLPV SSTAQNINFS ESIQNYTDKS
AQNLASRNLK HERFQSLSFP HTKEMMKIFH KKFGLHNFRT NQLEAINAAL LGEDCFILMP TGGGKSLCYQ LPACVSPGVT VVISPLRSLI
VDQVQKLTSL DIPATYLTGD KTDSEATNIY LQLSKKDPII KLLYVTPEKI CASNRLISTL ENLYERKLLA RFVIDEAHCV SQWGHDFRQD
YKRMNMLRQK FPSVPVMALT ATANPRVQKD ILTQLKILRP QVFSMSFNRH NLKYYVLPKK PKKVAFDCLE WIRKHHPYDS GIIYCLSRRE
CDTMADTLQR DGLAALAYHA GLSDSARDEV QQKWINQDGC QVICATIAFG MGIDKPDVRF VIHASLPKSV EGYYQESGRA GRDGEISHCL
LFYTYHDVTR LKRLIMMEKD GNHHTRETHF NNLYSMVHYC ENITECRRIQ LLAYFGENGF NPDFCKKHPD VSCDNCCKTK DYKTRDVTDD
VKSIVRFVQE HSSSQGMRNI KHVGPSGRFT MNMLVDIFLG SKSAKIQSGI FGKGSAYSRH NAERLFKKLI LDKILDEDLY INANDQAIAY
VMLGNKAQTV LNGNLKVDFM ETENSSSVKK QKALVAKVSQ REEMVKKCLG ELTEVCKSLG KVFGVHYFNI FNTVTLKKLA ESLSSDPEVL
LQIDGVTEDK LEKYGAEVIS VLQKYSEWTS PAEDSSPGIS LSSSRGPGRS AAEELDEEIP VSSHYFASKT RNERKRKKMP ASQRSKRRKT
ASSGSKAKGG SATCRKISSK TKSSSIIGSS SASHTSQATS GANSKLGIMA PPKPINRPFL KPSYAFS¹⁴¹⁷

FIGURE 7–(*continued*)

However, the evidence for glutamic acid 315 as the main Nt amino acid of the protein fragment is very good. We also confirmed Nt-dimethylation and, therefore, Nt identification of this proteolytically generated fragment of MukB by inspection of the MS/MS spectrum that contained consistent series of b ions (Fig. 8*A*).

831MT_F and 831MT_T: Nt Dimethylation Result

For 831MT_F full length and 831MT_T truncated fragments, we identified "dimethyl **S**MEELVVEVR" (**Fig. 9***A*, ***B***, respectively) peptides that correspond to the N-terminal amino acids of the cloned FMRP (Uniprot Accession FMR1_HUMAN) (M_r ~71,174 Da), spanning amino acids 1–10 (Fig. 9*C*). For the full-length fragment, we identified 120 peptides as dimethyl **S**MEELVVEVR, and for the truncated 831MT, 87 peptide counts mapped to this result. There were other internal lysine-dimethylated peptides that were identified (see Supplemental Tables 5 and 6). Visual inspection of the MS/MS spectrum did not show presence of a b ion series that confirms Nt-dimethylated fragment ions; therefore, we carried out a secondary independent search by PEAKS de novo sequencing (Fig. 9*D* for 831MT_F and Fig. 9*E* for 831MT_T), which confirmed the database search result and that the serine at aa 1 is the main Nt amino acid of both the full-length and the truncated protein fragments (see Supplemental Tables 7 and 8).

831MT_F_PVDF and 831MT_T_PVDF: Nt Sequence Analysis (Edman Degradation) Results

Both proteins have the same Nt sequence result, "**SMEELVVE**." This corresponds to the Nt amino acid sequence of the cloned protein sequence **SMEELVVE** and

is identical to FMRP sequence after the first amino acid serine.

Nt-dimethyl labeling of proteins is a universal labeling method and does not depend on the nature of the Nt amino acid. For the 3 applications above, the dimethylated Nt amino acids were found to be methionine and alanine for sample-annotated 710SP, glutamic acid for samples 826RK, and finally, serine for sample 831MT. In cases where multiple Nt dimethylation was confirmed, this was a result of the Nt heterogeneity of the protein fragment. For example, for sample 710SP, we identified an Nt peptide of the protein with and without the initiator methionine, as dimethyl MAAVPQNNLQEQLER and as dimethyl AAVPQNNL-QEQLER (Fig. 7*A*, *B*).

Database searches with N formyl as a variable modification did not result in any significant peptide identifications, as documented in the supplemental tables, for myoglobin, BLM, MukB, and FMRP. This result further strengthens the specificity and use of the Nt derivatization study presented here. The monoisotopic mass difference between Nt-dimethyl and Nt-formyl is 0.0365 Da. Before the use of high mass accuracy and resolution instrumentation, this difference could have been considered to be small enough, causing N-formylated and N-dimethylated peptide identifications to be inseparable. We advocate, whenever possible, use of improved mass spectrometers with high accuracy and resolution to make unequivocal identification of protein fragments.

Sequencing searches (de novo), carried out for standard myoglobin, as well as the full-length and truncated proteins 831MT_F and 831MT_T, confirmed our initial finding by the Mascot search engine. For PEAKS de novo searches,

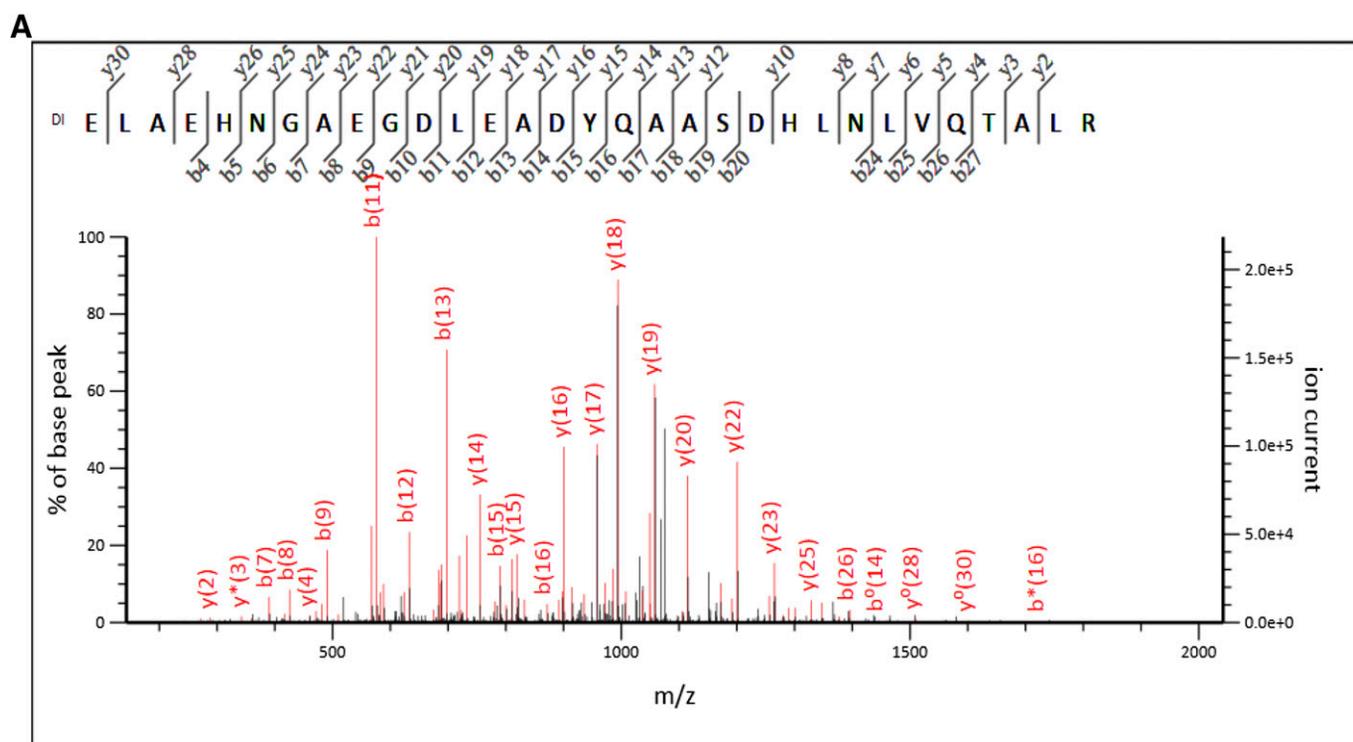**FIGURE 8**

A) MS/MS analysis and identification of the Nt tryptic peptide-dimethyl ELAEHNGAEGDLEADYQAASDHLNLVQTALR in sample 826RK. *B*) We show the representation of the 826RK_A2 experimental data that identified protein fragment in chromosome partitioning protein, MukB. All green residues have been confirmed with >95% probability by the Scaffold FDR algorithm. There is only 1 tryptic peptide that identifies the Nt-dimethylation: the peptide spanning aa 315–345. MukB protein has 49 lysine residues. Experimental data map 17 lysines (out of 19 that are present in the proteolytic fragment of MukB); 11/19 lysines are found to be dimethylated (see mapped, green regions with Dimethyl K); and nearly all are internal (miscleaved) lysines.

where we used Nt-dimethylation, asparagine, and glutamine deamidation, methionine oxidation and trypsin specificity in the absence of a protein database, multiple high-scoring (average local confidence scores >90%) matches were obtained for "Nt-dimethyl GLSDGEWQQVLNVWGK" for myoglobin raw data (Fig. 6*D*). It was encouraging to see similar, multiple, high-scoring (average local confidence scores >90%) matches for "Nt-dimethyl **S**MEELVVEVR" for 831MT_F and 831MT_T samples (Fig. 9*D*, *E*).

There are 2 approaches to Nt-dimethylation labeling of proteins: 1) in-gel Nt-dimethyl derivatization, followed by a bottom-up approach, where the decision to first resolve and purify protein(s) by SDS-PAGE aids in removal of dimethylation chemicals and offers selectivity of domain of interest, and 2) Nt-dimethylation on protein mixtures followed by SDS-PAGE and in-gel digestion, which provides comparable results to the first approach, as long as care is taken to neutralize the acidic pH of the dimethylation reaction before SDS-PAGE. This study, as well as this year's PSRG study,[4] confirm that Nt-dimethylation of gel-bound proteins is a sensitive method. An apparent molecular weight of 826RK_A2, which was

identified as MukB, as estimated by SDS-PAGE (Fig. 2), is ~110 kDa. To carry out Edman sequence analysis successfully, we would require ~2 μg of this fragment (this roughly corresponds to 20 pmol starting materials, assuming a single Nt fragment at stained protein band 826RK_A2). As can be estimated from the SimplyBlue SafeStain image in Fig. 2, the amounts used for Nt identification of the fragment 826RK_A2 in a single lane, using the Nt demethylation approach, was 25–50 ng. This illustrates that it is possible to obtain Nt identification of a protein fragment by the Nt-dimethylation approach, using starting materials that are a much less than what is required for Edman sequencing approach.

We have observed that under the dimethylation conditions that we have followed, the internal lysine dimethylation was present, as confirmed by inspection of the MS/MS spectra. In most cases, lysine dimethylations were common in tryptic peptides with miscleaved internal lysines. It was encouraging to see de novo sequencing softwares confirming database search results. However, visual confirmation of Nt peptides remains a very important practice that was used throughout this documented research.

B

[1]MIERGKFRSL TLINWNGFFA RTFDLDELVT TLSGGNGAGK STTMAAFVTA LIPDLTLLHF RNTTEAGATS GSRDKGLHGK LKAGVCYSML

DTINSRHQRV VVGVRLQQVA GRDRKVDIKP FAIQGLPMSV QPTQLVTETL NERQARVLPL NELKDKLEAM EGVQFKQFNS ITDYHSLMFD

LGIIARRLRS ASDRSKFYRL IEASLYGGIS SAITRSLRDY LLPENSGVRK AFQDMEAALR ENRMTLEAIR VTQSDRDLFK HLISEATNYV

AADYMRHANE RRVHLDKALE FRRELHTSRQ QLAAEQYKHV DMAR

Dimethyl-Nt →[315]ELAEHN GAEGDLEADY QAASDHLNLV QTALRQQEKI ERYEADLDEL QIRLEEQNEV

Dimethyl K

VAEAIERQEE NEARAEAAEL EVDELKSQLA DYQQALDVQQ TRAIQYNQAI AALNRAKELC

Dimethyl K

HLPDLTADCA AEWLETFQAK ELEATEKMLS LEQKMSMAQT AHSQFEQAYQ LVVAINGPLA

Dimethyl K

RNEAWDVARE LLREGVDQRH LAEQVQPLRM RLSELEQRLR EQQEAERLLA DFCKRQGKNF

DIDELEALHQ ELEARIASLS DSVSNAREER MALRQEQEQL QSRIQSLMQR APVWLAAQNS

LNQLSEQCGE EFTSSQDVTE YLQQLLERER EAIVERDEVG ARKNAVDEEI ERLSQPGGSE

DQRLNALAER FGGVLLSEIY DDVSLEDAPY FSALYGPSRH AIVVPDLSQV TEHLEGLTDC PEDLYLIEGD

Dimethyl K   Dimethyl K

PQSFDDSVFS VDELEKAVVV KIADRQWRYS RFPEVPLFGR AARESRIESL HAEREVLSER FATLSFDVQK

Dimethyl K

TQRLHQAFSR FIGSHLAVAF ESDPEAEIRQLNSRRVELER ALSNHENDNQ QQRIQFEQAK EGVTALNRIL

Dimethyl K

PRLNLLADDS LADRVDEIRE RLDEAQEAAR FVQQFGNQLA KLEPIVSVLQ SDPEQFEQLK

Dimethyl K

EDYAYSQQMQ RDARQQAFAL TEVVQRRAHF SYSDSAEMLS GNSDLNEKLR ERLEQAEAERTRAR

Dimethyl K

EALRGH AAQLSQYNQV LASLKSSYDT KKELLNDLQR[1040] ELQDIGVRAD SGAEERARIR RDELHAQLSN NRSRRNQLEK

ALTFCEAEMD NLTRKLRKLE RDYFEMREQV VTAKAGWCAV MRMVKDNGVE RRLHRRELAY LSADDLRSMS DKALGALRLA VADNEHLRDV

LRMSEDPKRP ERKIQFFVAV YQHLRERIRQ DIIRTDDPVE AIEQMEIELS RLTEELTSRE QKLAISSRSV ANIIRKTIQR EQNRIRMLNQ

GLQNVSFGQV NSVRLNVNVR ETHAMLLDVL SEQHEQHQDL FNSNRLTFSE ALAKLYQRLN PQIDMGQRTP QTIGEELLDY RNYLEMEVEV

NRGSDGWLRA ESGALSTGEA IGTGMSILVM VVQSWEDESR RLRGKDISPC RLLFLDEAAR LDARSIATLF ELCERLQMQL IIAAPENISP

EKGTTYKLVR KVFQNTEHVH VVGLRGFAPQ LPETLPGTDE APSQAS[1486]

FIGURE 8 –(*continued*)

CONCLUSION

Edman degradation sequence analysis has provided scientists with definitive amino acid sequence information, Nt or internal, for decades. To obtain information on ~20 aa that are at the N terminus of a single protein, we need at least 20 pmol starting materials. This is a result of losses of amino acid yields inherent in the chemistry and derivatized amino acids on automated sequencers. Furthermore, Edman degradation will not work on N-terminally blocked proteins, which may occur in proteins expressed in insect and mammalian cells. With the advent and easy use of biologic mass spectrometers, alternative approaches have
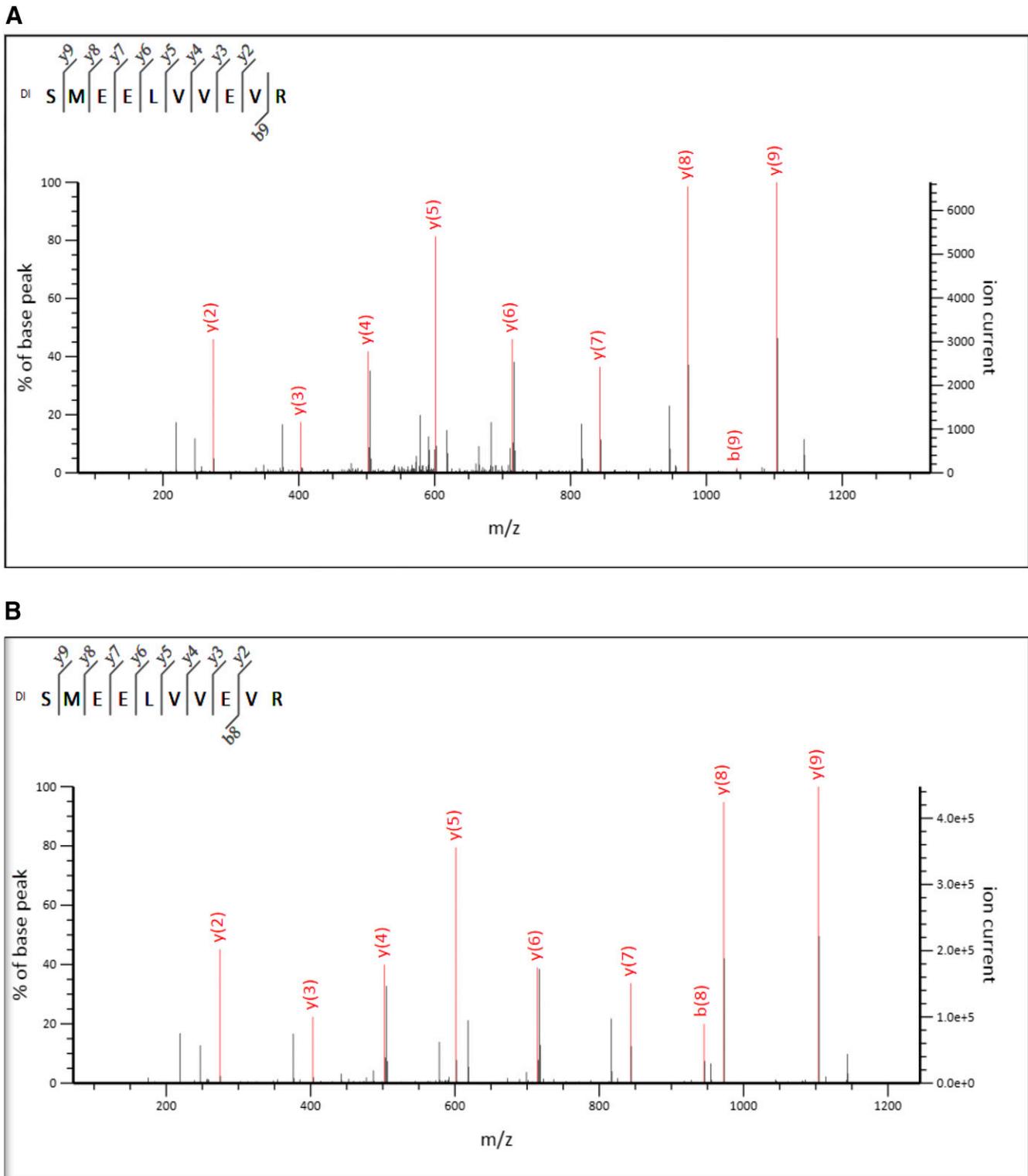
A

B

FIGURE 9

MS/MS analysis and identification of the Nt tryptic peptide-dimethyl SMEELEVR in sample 831MT_F (*A*) and sample 831MT_T (*B*). *C*) We show the representation of the experimental data that identified the fragment of FMRP in samples 831MT_T and 831MT_F. All green residues have been confirmed with >95% probability by the Scaffold FDR algorithm. There was only 1 tryptic peptides that identifies the Nt-dimethylation: peptide spanning aa 1–10 of the expressed protein. Expressed fragment of the FMR_1 protein has 33 lysine residues. Experimental data map 26 lysines; 11/26 lysines are found to be dimethylated (see mapped, green regions with "dimethyl lysines"); and nearly all are internal (miscleaved) lysines. *D* and *E*) Sequencing result (de novo), confirming Nt-dimethylated peptides SMEELVVEVR for samples 831MT_T and 831MT_F, respectively, that identified FMRP.

C

Dimethyl-Nt → **1**SMEELVVEVRG SNGAFYKAFV KDVHEDSITV AFENNWQPDR QIPFHDVRFP PPVGYNKDIN

ESDEVEVYSR ANEKEPCCWW LAKVRMIKGE FYVIEYAACD ATYNEIVTIE RLRSVNPNKP

ATKDTFHKIK LDVPEDLRQM CAKEAAHKDF KKAVGAFSVT YDPENYQLVI LSINEVTSKR

AHMLIDMHFR SLRTKLSLIM RNEEASKQLE SSRQLASRFH EQFIVREDLM GLAIGTHGAN

IQQARKVPGV TAIDLDEDTC TFHIYGEDQD AVKKARSFLE FAEDVIQVPR NLVGKVIGKN

GKLIQEIVDK SGVVRVRIEA ENEKNVPQEE EIMPPNSLPS NNSRVGPNAP EEKKHLDIKE

NSTHFSQPNS TKVQRVLVAS SVVAGESQKP ELKAWQGMVP FVFVGTKDSI ANATVLLDYH

LNYLKEVDQL RLERLQIDEQ LRQI**379**GASSRP PPNRTDKEKS YVTDDGQGMG RGSRPYRNRG HGRRGPGYTS GTNSEASNAS ETESDHRDEL SDWSLAPTEE ERESFLRRGD GRRRGGGGRG QGGRGRGGGF KGNDDHSRTD NRPRNPREAK GRTTDGSLQI RVDCNNERSV HTKTLQNTSS EGSRLRTGKD RNQKKEKPDS VDGQQPLVNG VP**632**

D

De-Novo Sequence Search

E

De-Novo Sequence Search

FIGURE 9–(*continued*)

been developed to obtain similar definitive information to what is classically provided by the Edman degradation sequencing method. Nt-dimethylation, as shown in this work, is simple, robust, and reproducible, and it is also near quantitative. Hsu et al.[2] has already shown that dimethyl labeling results in enhanced detection of the generated peptides. It should be emphasized that Nt-demethylation will not work with N-terminally "blocked" proteins. Another major limitation of the use of Nt-dimethylation to identify the Nt amino acid of a protein is when the proteolytically generated peptide is too short or too long. For example, after proteolysis and dimethyl derivatization, we may have a peptide that consists of 2 or 3 aa (Nt-dimethyl amino acid X-R/K or Nt-dimethyl amino acid XY-R/K). There is a good probability that this di- or tripeptide may not bind to the subsequent HPLC column and therefore, may not be recovered and identified. If the generated, Nt-dimethylated tryptic peptide is long and hydrophobic, it may not elute from the reverse-phase HPLC column, and just like the very short and hydrophilic di- or tripeptide, it might never be recovered and identified at the end of the subsequent LC-MS/MS experiment. However, in these situations, in-gel proteolysis may be customized or optimized to help recovery of Nt candidate peptides that are protein sequence dependent.

REFERENCES

1. Koth CM, Orlicky SM, Larson SM, Edwards AM. Use of limited proteolysis to identify protein domains suitable for structural analysis. *Methods Enzymol* 2003;368:77–84.
2. Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 2003; 75:6843–6852.
3. Deng J, Zhang G, Huang FK, Neubert TA. Identification of protein N-termini using TMPP or dimethyl labeling and mass spectrometry. *Methods Mol Biol* 2015; 1295:249–258.
4. McGrath SC, Erdjument-Bromage H, Cavey GS, English RD, Luo X, Remmer HA, Phinney BS. N-Terminal identification of a standard protein at low picomole levels by dimethyl labeling and bottom-up mass spectrometry. *ABRF 2015 Annual Meeting*, PSRG Poster #260, 2015.
5. Petrushenko ZM, Lai CH, Rai R, Rybenkov VV. DNA reshaping by MukB. Right-handed knotting, left-handed supercoiling. *J Biol Chem* 2006;281:4606–4615.
6. Erdjument-Bromage H, Lui M, Lacomis L, Grewal A, Annan RS, McNulty DE, Carr SA, Tempst P. Examination of micro-tip reversed-phase liquid chromatographic extraction of peptide pools for mass spectrometric analysis. *J Chromatogr A* 1998; 826:167–181.
7. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence data-bases using mass spectrometry data. *Electrophoresis* 1999;20: 3551–3567.
8. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics 2012;11:M111.010587. 1074/mcp.M111.010587.
9. Tempst P, Geromanos S, Elicone C, Erdjument-Bromage H. Improvements in microsequencer performance for low pico-mole sequence analysis. *Methods Comp Meth Enzymol* 1994;6: 248–261.
10. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 1984;11:601.