Received: 21 December 2018 | Revised: 1 March 2019 | Accepted: 9 March 2019

DOI: 10.1002/rcm.8435

RESEARCH ARTICLE

WILEY Rapid Communications in Mass Spectrometry

Identification of N-terminal protein processing sites by chemical labeling mass spectrometry

Santosh A. Misal[1] | Sujun Li[2] | Haixu Tang[2] | Predrag Radivojac[2] | James P. Reilly[1]

[1] Department of Chemistry, Indiana University, Bloomington, Indiana, USA

[2] School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana, USA

Correspondence
J. P. Reilly, Department of Chemistry, Indiana University, Bloomington, Indiana, USA.
Email: reilly@indiana.edu

Funding information
National Institutes of Health, Grant/Award Numbers: U54 GM074807 and R01 GM103725

Rationale: Proteins undergo post-translational modifications and proteolytic processing that can affect their biological function. Processing often involves the loss of single residues. Cleavage of signal peptides from the N-terminus is commonly associated with translocation. Recent reports have suggested that other processing sites also exist.

Methods: The secreted proteins from *S. aureus* N315 were precipitated with trichloroacetic acid (TCA) and amidinated with *S*-methyl thioacetimidate (SMTA). Amidinated proteins were digested with trypsin and analyzed with a high-resolution orbitrap mass spectrometer.

Results: Sixteen examples of *Staphylococcus aureus* secretory proteins that lose an N-terminal signal peptide during their export were identified using this amidination approach. The N-termini of proteins with and without methionine were identified. Unanticipated protein cleavages due to sortase and an unknown protease were also uncovered.

Conclusions: A simple N-terminal amidination based mass spectrometry approach is described that facilitates identification of the N-terminus of a mature protein and the discovery of unexpected processing sites.

1 | INTRODUCTION

The N-terminus of a mature protein is often not defined by the genome sequence from which it is translated. Rather, the occurrence of post-translational modifications (PTMs) and signal peptide cleavages can affect the N-terminal sequence and determine the eventual fate of the functional protein. The significance of these processes has motivated the growth of the new field of N-terminomics.[1] Mass spectrometric methods have been commonly applied to identify N-terminal residues and their modifications.[2] They often rely on chemical derivatizations using reagents such as heavy/light TMPP (trimethoxyphenyl phosphonium), biotin *N*-hydroxysuccinimide ester (biotin-heavy/light-NHS) and SATA (*N*-succinimidyl *S*-acetylthioacetate).[3-5] SATA and NHS biotin modified N-terminally labeled peptides can be enriched from complex mixtures. TMPP applies a fixed charge tag to a peptide, but it is very bulky and can be lost during fragmentation.[6] *N*-Acetoxy-D3/H3-succinamides and D2/H2 formaldehyde have been utilized to label N-termini in quantitative proteomics studies.[7] Unfortunately, for all these cases, coverage of low-abundance proteins can sometimes be poor.[3,8] An additional complication is that succinimidyl esters can undergo undesirable reactions with other nucleophiles such as OH groups, as demonstrated in chemical labeling[9-11] and cross-linking experiments.[12-14] The advantages of *S*-methyl thioacetimidate (SMTA) in this application are that it is small and so it exhibits good solvent accessibility, it increases the basicity of peptides, and it does not undergo competing reactions with hydroxyl groups. Finally, the amidino tag that is created by SMTA reaction facilitates the formation of b_1 ions that definitively locate the labeling site. It is of interest to develop novel and efficient approaches to study proteolytic events in the cell, and in particular to identify N-termini that result from proteolytic processing of the secreted proteins of pathogenic bacteria.

Staphylococcus aureus is an important community-acquired human pathogen for which multi-drug resistance represents a major worldwide health threat. Secretion of proteins such as cytotoxins, enterotoxins, proteases, lipolytic enzymes and peptidoglycan hydrolases is one of the essential cell functions that makes *S. aureus* pathogenic.[15,16] These proteins are synthesized in the cytoplasm and translocated outside of the cell by specific secretory pathways. Distinct N-terminal leader (signal) peptide sequences serve as determinants for one protein secretory pathway.[17-20] However, there is growing evidence of other secretory pathways that involve novel processing sites typically at or near protein N-termini.[21,22] Covalent chemical labeling experiments exploit the accessibility of these sites to identify their locations. We report here the combination of protein N-terminal amidination by SMTA with high-resolution mass spectrometry analysis to confidently identify proteolytic processing sites. The present work describes methodology for identifying N-termini of *S. aureus* N315 proteins that occur in both predictable and unexpected sequence locations.

2 | EXPERIMENTAL

2.1 | Bacterial growth and protein extraction

S. aureus N315 cells were stored as a glycerol stock culture at −80°C in a freezer. The bacterial cells were streaked on a Brain Heart Infusion (BHI) agar plate and grown overnight at 37°C. Single colonies were selected, inoculated in 5 mL of sterile BHI broth (37 g/L) and incubated overnight at 37°C in a 250 rpm orbital shaker. Then 1 mL of this overnight culture was further inoculated in 100 mL of fresh BHI or Tryptic Soy Broth (TSB) and grown for 6 h at 37°C. This culture was aliquoted in 1 mL of 20% (v/v) glycerol stocks and stored at −80°C until required. The working culture was prepared by inoculating 100 mL of fresh sterile BHI or TSB broth with 1 mL of glycerol stock and incubating overnight at 37°C in an orbital shaker at 250 rpm. A volume of 5 mL of this overnight culture was diluted to 500 mL with fresh BHI or TSB broth and allowed to grow for 6 h at 37°C. This yielded an optical density of between 0.8 and 1.0 at 660 nm as measured by UV–visible spectrophotometry. Cells were pelleted by centrifugation at 8000 rpm for 10 min using a JA10 Beckman Coulter rotor. The cell pellet was discarded, and the supernatant was filtered through a 0.22 µM Millipore Stericup filter to remove any residual bacteria. 10% TCA (w/v) was added to the filtered solution and incubated overnight with continuous stirring at 4°C in order to precipitate soluble proteins. This solution was then centrifuged at 15,000 rpm for 45 min and the resulting protein pellet was washed with ice-cold acetone at least three times. The pellet was dried in a vacuum evaporator for 10 min and then resuspended in 400 µL of 100 mM ammonium bicarbonate, pH 8, for chemical labeling. This solution was clarified by centrifugation at 15,000 rpm to remove insoluble precipitates and the protein concentration was determined by Bradford assay using BSA as a standard.

2.2 | Chemical labeling of proteins with SMTA

SMTA was prepared by reaction of methyl iodide with thioacetamide in anhydrous diethyl ether as described by Thumm et al.[23] SMTA was added to a 500 mM concentration in a 600 µg protein solution and incubated at room temperature for 2 h with continuous slow vortex. The pH of the reaction mixture was monitored and adjusted to 7.5–8 every 20 min with 100 mM ammonium bicarbonate and 250 mM KOH. Following the labeling reaction, the protein sample was loaded on a strong cation-exchange (SCX) chromatography column that had been pre-equilibrated with mobile phase A consisting of 6 M urea, 20 mM glacial acetic acid, pH 5.0. Following protein loading, excess unreacted SMTA was washed away with mobile phase A for 20 min. Proteins were eluted from the column using a 120 min gradient of 0 to 90% mobile phase B (6 M urea, 20 mM glacial acetic acid, 0.5 M NaCl, pH 5.0). Eluting proteins were captured on 20 reversed-phase C4 trapping columns. The C4 columns were desalted with 90% 0.1% formic acid (FA) in water and the trapped proteins were then eluted with 90% ACN, 0.1% FA. Proteins directed to a Waters fraction collector and the 20 samples were lyophilized in a vacuum evaporator. Dried samples were re-dissolved in 25 mM ammonium bicarbonate, pH 7.8. Proteins were reduced with 5 mM DTT and alkylated with 5 mM iodoacetamide. Trypsin was added to each sample at 1:50 ratio (trypsin/protein) and incubated for 18 h at 37°C. The digestion was terminated by adding 0.1% FA. Tryptic digests were centrifuged at 14000 rpm before loading on a reversed-phase C18 column.

2.3 | Orbitrap mass spectrometry and database search

Peptides were analyzed using a Thermo Orbitrap Fusion Lumos mass spectrometer coupled with a nanoAcquity liquid chromatography (LC) system. Approximately 1 µg of the peptide was loaded onto a nanoACQUITY UPLC Symmetry C18 trap column (Waters) in 95% solvent A (0.1% FA in water (HPLC grade)) and 5% solvent B (0.1% FA in acetonitrile (HPLC grade)). Peptides were eluted and separated using a 60-min gradient from 3 to 48% of solvent B at a flow rate of 300 nL/min on a C18 ACQUITY UPLC HSS T3 column (Waters). The eluent from the C18 column was electrosprayed in positive ion mode using 1.8 kV voltage. Peptides with the precursor mass in the *m/z* range 300 to 2000 with 2 to 7 charges were selected for HCD activation. The resolution for MS1 was set at 120,000 and HCD fragment ions were analyzed with a resolution of 30,000. Mass spectrometry raw data were converted into Mascot generic format (mgf) using MS Convert software. The mgf files were submitted to Mascot and searched against the *S. aureus* N315 proteome database that was downloaded from Uniprot (https://www.uniprot.org/proteomes/UP000000751). Peptides were searched using semi-tryptic enzyme cleavage (non-specific cleavage at the N-terminus of the peptide) with a maximum of three missed cleavages. Mass tolerances for precursor and fragment ions were set to 5 ppm and

0.02 Da, respectively. Variable modifications included methionine oxidation, and lysine (+41.0265) and N-terminal amidination (+41.0265). Peptides having a score of greater than 30 with 1% false discovery rate (FDR) were considered as confident identifications. Each spectrum that was identified as arising from an N-terminal amidinated peptide was manually checked to verify that it contained a b_1 ion with an intensity of at least 10% of the base peak.

3 | RESULTS AND DISCUSSION

3.1 | N-terminal amidination of proteins

In order to identify N-termini, protein samples are derivatized using the amidinating reagent *S*-methyl thioacetimidate (SMTA). This reagent reacts with amine groups at each N-terminus and lysine side chain.[24] Subsequent to the reaction, proteins are enzymatically digested and peptides that contain an N-terminal amidination label are identified. A major advantage of using SMTA over NHS esters as the tagging reagent is that it does not exhibit secondary reactivity toward protein OH groups.[23] It is also relatively small, and reacts fast and efficiently. Most importantly, N-terminally amidinated peptides undergo a unique and highly reproducible fragmentation pattern upon collisional activation. The amidine group at the N-terminus of a peptide promotes cleavage of the N-terminal peptide bond to produce a distinctive and intense b_1 fragment ion.[24-26] This is striking because b_1 ions of unlabeled peptides are generally unstable and not observed.[27] Thus, observation of a b_1 ion in conjunction with other fragment ions in the tandem mass (MS/MS) spectrum of an amidinated peptide provides clear evidence that the N-terminus of this peptide was labeled *before* the enzymatic digestion step. That indicates that this was an accessible protein N-terminus.

The experimental workflow is depicted in Figure 1. *S. aureus* N315 cells were grown to mid-log phase and pelleted out of the culture medium. Secreted proteins were precipitated from the supernatant solution by the addition of trichloroacetic acid (TCA), then resuspended in ammonium bicarbonate buffer and amidinated with SMTA. Following amidination, intact proteins were separated by strong cation-exchange (SCX) chromatography and fractionated. Individual protein fractions were digested with trypsin and the resulting peptides were analyzed using a Thermo Orbitrap mass spectrometer. Ions were fragmented with high-energy collisional dissociation (HCD). N-terminally amidinated peptides yield stable b_1 fragment ions that appear as intense features in the spectrum.[26]

Initially, proof of concept was demonstrated by identifying N-terminal peptides from proteins that had undergone either no or trivial N-terminal processing. Many of these are proteins known to be highly abundant in the cytoplasm that are not expected to be exported. Their detection indicates that a small fraction of cells undergo inadvertent lysis, releasing their contents into the extracellular medium. Over 35 N-terminal peptides derived from proteins of this sort were identified and are listed in Table 1. Amidination was typically found either at the initial methionine or at the next residue; the latter indicates that excision by methionine aminopeptidase had occurred. As a representative example, Figure 2A displays the MS/MS spectrum of the N-terminal peptide of ribosomal protein L22. The peptide MEAKAVAR is amidinated at both the lysine residue and at the N-terminus as demonstrated by the observed peptide mass, the masses of all fragments and, in particular, the presence of an intense b_1 ion peak. While the presence of ribosomal proteins in the extracellular milieu may seem a little surprising, they are small and abundant and are probably products of inadvertent cell lysis. Previously, we found that ~3% of the *S. aureus* COL cells were lysed during the stationary phase and that is enough to detect the abundant cytoplasmic proteins in the extracellular fraction.[16]

Figure 2B displays the MS/MS spectrum of the N-terminal peptide PIITDVYAR derived from enolase protein, a glycolytic enzyme that catalyzes the dehydration of 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP). This protein has been reported to be in the cytoplasm and on the cell surface. Its high abundance in the

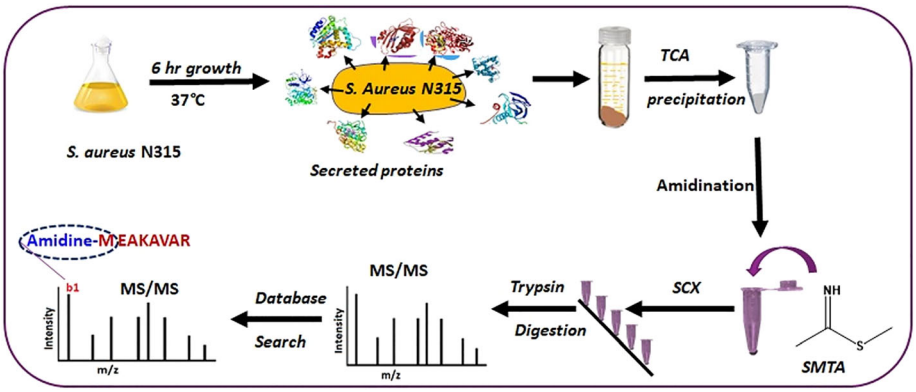

FIGURE 1 Chemical labeling/mass spectrometry work flow. Secreted proteins were isolated, precipitated with 10% TCA, amidinated with 500 mM SMTA, separated with ion-exchange chromatography, fractionated and digested with trypsin. Mass spectrometry data were acquired using a Thermo Orbitrap fusion Lumos with HCD fragmentation and a resolution of 120,000 for full MS scans. Raw data were analyzed using Mascot [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Amidinated peptides from the N-termini of proteins

No.	Pre	Peptide	Post	Accession	Protein	Subcellular location	Spectral counts	Amidinated
1	M	PIITDVYAR	E	P99088	Enolase	Extracellular/cell surface/cytoplasm	130	104
2	M	TKSALVTGASR	G	P99093	3-oxoacyl-[acyl-carrier-protein] reductase FabG	Cytoplasm/extracellular/cell surface	3	3
3	M	SNNFKDDFEKNR	Q	Q7A5I6	Elastin-binding protein EbpS	Membrane	2	2
4	M	PTEDTTMFDQVAEVIER	L	A0A0H3JM38	SA0797 protein	Predicted cytoplasm/membrane	2	2
5	M	AVNLYDYANQLEQALR	E	Q7A4V3	UPF0342 protein SA1663	Cytoplasm/membrane	12	12
6	M	AVFKVFYQHNR	D	Q99V08	UPF0356 protein SA0941	Cytoplasm/membrane	2	2
7	M	AVPNELKY	S	P64214	Glycine cleavage system H protein	Cytoplasm/membrane	7	7
8	M	KKVVNLIMTKER	R	A0A0H3JKT7	Probable transposase	Unknown	2	2
9	M	TVDNNKAKQAYDNQTGVNEKER	E	P99157	Alkaline shock protein 23, ASP23	Unknown	13	13
10	M	PIVNVKLLEGR	S	P99132	Probable tautomerase SA1195.1	Cytoplasm	16	16
11	-	MEQNSYVIIDETGIHAR	P	P99143	Phosphocarrier protein HPr	Cytoplasm	121	9
12	M	EQNSYVIIDETGIHAR	P	P99143	Phosphocarrier protein HPr	Cytoplasm	19	12
13	M	ATISAKLVKELR	K	P99171	Elongation factor Ts	Cytoplasm	2	2
14	M	ENFDKVKDIIVDR	L	P0A002	Acyl carrier protein	Cytoplasm	10	9
15	-	MEAKAVAR	T	Q7A460	Ribosomal protein L22	Cytoplasm	2	2
16	M	EAKAVAR	T	Q7A460	Ribosomal protein L22	Cytoplasm	2	2
17	M	SAIIEAKKQLVDEIAEVLSNSVSTVIVDYR	G	P99155	Ribosomal protein L10	Cytoplasm	27	25
18	M	PTINQLVR	K	P0A0G8	Ribosomal protein S12	Cytoplasm	10	8
19	M	GQKINPIGLR	V	P66553	Ribosomal protein S3	Cytoplasm	24	24
20	-	MKLHELKPAEGSR	K	P0A0F6	Ribosomal protein L15	Cytoplasm	2	2
21	M	KLHELKPAEGSR	K	P0A0F6	Ribosomal protein L15	Cytoplasm	29	2
22	M	TNHKLIEAVTKSQLR	T	P66083	Ribosomal protein L19	Cytoplasm	2	2
23	M	AISQER	K	Q7A5X8	Ribosomal protein S15	Cytoplasm	4	4
24	M	AIKKYKPITNGR	R	P60432	Ribosomal protein L2	Cytoplasm	21	19
25	M	PKMKTHR	G	P66276	Ribosomal protein L35	Cytoplasm	5	5
26	M	TKGILGR	K	P60449	Ribosomal protein L3	Cytoplasm	12	12
27	M	TMTDPIADMLTR	V	P66630	Ribosomal protein S8	Cytoplasm	16	16
28	M	AQVEYR	G	P66646	Ribosomal protein S9	Cytoplasm	5	5
29	M	ANHEQIIEAIK	E	P99154	Ribosomal protein L7/L12	Cytoplasm	3	3
30	M	AVKVAINGFGR	I	P99136	Glyceraldehyde-3-phosphate dehydrogenase 1	Cytoplasm	11	8

(Continues)

TABLE 1 (Continued)

No.	Pre	Peptide	Post	Accession	Protein	Subcellular location	Spectral counts	Amidinated
31	MA	VKVAINGFGR	I	P99136	Glyceraldehyde-3-phosphate dehydrogenase 1	Cytoplasm	6	2
32	M	AVISMKQLLEAGVHFGHQTR	R	P66544	Ribosomal protein S2	Cytoplasm	19	19
33	MA	VISMKQLLEAGVHFGHQTR	R	P66544	Ribosomal protein S2	Cytoplasm	2	2
34	M	AKLQITLTR	S	P0A0G0	Ribosomal protein L30	Cytoplasm	16	10
35	MA	KLQITLTR	S	P0A0G0	Ribosomal protein L30	Cytoplasm	71	27
36	MA	SLKSIIR	Q	Q7A7B3	Ribosomal protein L25	Cytoplasm	3	3

extracellular matrix is associated with its role in the staphylokinase activation of plasminogen and pathogen invasion.[28] Nevertheless, the protein does not lose a signal peptide. The MS/MS spectrum of PIITDVYAR is assigned to the peptide amidinated at the N-terminal proline following methionine excision. All ion peaks, and especially the appearance of an intense b_1 fragment ion, are consistent with this interpretation. We observed N-terminal peptides from three different proteins that retained methionine and 30 proteins that had lost this residue. Some proteins such as phosphocarrier protein HPr, and ribosomal proteins L15 and L22, appeared in both forms, with and without their initial methionine. It is known that the *E. coli* methionine aminopeptidase (MetAP1) can tolerate a broad range of amino acids at the second and third position and that cleavage adjacent to small residues is preferred.[29,30] A few proteins listed at the bottom of Table 1 also lose an alanine residue adjacent to the methionine before their N-termini are amidinated. This has not been previously reported. It may result from methionine aminopeptidase activity. Alternatively, we have detected two other aminopeptidases, cytosol aminopeptidase (ampA) and SA2244 protein, in our extracellular medium and one of them may be involved. It is worth pointing out that the extent of N-terminal amidination is not always 100%. This makes sense since the N-terminus is not necessarily solvent accessible. For example, the N-terminal peptides of phosphocarrier protein HPr and ribosomal protein L15 are observed numerous times but are only rarely amidinated. For one, loss of methionine appears to increase the labeling rate while for the other protein the opposite is true.

3.2 | SPase I cleavage sites confirmed by N-terminal amidination

In a previous study we identified signal peptidase cleavage sites by observing the total masses of exported proteins that remained after these cleavages had occurred.[16] In general, our observations were very consistent with results from the SignalP 4.1 program that predicts the presence and location of signal peptide cleavage sites. In the present work, we were able to observe these signal peptidase cleavage sites without measuring the masses of whole proteins, but simply by observing N-terminally amidinated peptides. We were able to detect 16 known secretory proteins in the extracellular medium and identify their signal peptidase cleavage sites using this methodology. These proteins are listed in Table 2. The most abundant of these as judged by spectral counts include immunoglobulin G binding protein A precursor, putative surface protein SA2285, extracellular matrix-binding protein EbhB, immunodominant antigen A, and uncharacterized proteins. As expected, their SPase cleavage sites adjacent to standard AXA motifs were confirmed in these experiments. The amidination rate of the N-termini created by SPase I cleavage is almost 100% indicating that these sites become very solvent accessible. Immunoglobulin G binding protein A (SPa) is a cell wall associated protein that enables bacteria to evade the host immune response. It is anchored by its C-terminal region and has an

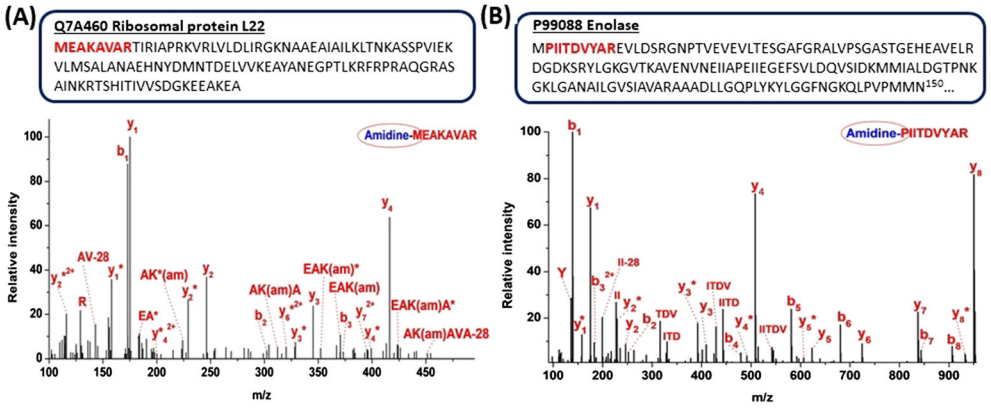

FIGURE 2 A, MS/MS spectrum of N-terminally amidinated peptide of ribosomal protein L22. B, MS/MS spectrum of N-terminally amidinated peptide of enolase enzyme whose N-terminal methionine had previously been cleaved by methionine aminopeptidase. (the b_1 ion peak is off-scale). The intense b_1 ions clearly indicate that the proteins are N-terminally amidinated, demonstrating the accessibility of these sites and exhibiting a key advantage of SMTA as a tagging reagent [Color figure can be viewed at wileyonlinelibrary.com]

LPXTG sortase motif. Signal peptidase cleavage adjacent to a typical AXA (ANA↓AQH) site removes a 36-residue signal peptide as previously reported.[16,21] In the current work we observed the succeeding peptide AQHDEAQQNAFYQVLNMPNLNADQR 28 times and it was amidinated 89% of the time. The MS/MS spectrum in Figure 3A displays an excellent match to this peptide; observation of the intense b_1 ion confirms that the peptide is amidinated at its N-terminal amine. The C-terminal portion of this protein provided a

TABLE 2 Amidinated peptides detected after signal peptide cleavage

No.	Accession	Protein	Pre	Peptide	Post	Cleavage pattern	Spectral counts	Amidinated
1	P99134	Immunoglobulin G binding protein A	A	AQHDEAQQNAFYQVLNMPNLNADQR	N	ANA↓AQH	28	25
2	P61598	Putative surface protein SA2285	A	AENNIENPTTLKDNVQSK	E	AEA↓AEN	12	10
3	Q7A5M1	Extracellular matrix-binding protein EbhB	A	AETNQPASVVKQKQQSNNEQTENR	E	AHA↓AET	8	8
4	P99160	Immunodominant antigen A	A	AEVNVDQAHLVDLAHNHQDQLNAAPIK	D	AHA↓AEV	4	4
5	A0A0H3JPH2	Uncharacterized protein	A	ATVHVAGGVWSHGIGKHYVWSYYSHNKR	N	AEA↓ATV	16	16
6	A0A0H3JNR9	Uncharacterized protein	A	SETNQKVSTNQESK	A	AEA↓SET	2	2
7	A0A0H3JTK7	Glycerophosphoryl diester phosphodiesterase	A	EQTNQIANKPQAIQWHTNLTNER	F	AGA↓EQT	6	5
8	A0A0H3JPQ1	SA1000 protein	A	QTKNVEAAKK	Y	SHA↓QTK	6	6
9	A0A0H3JNG8	Staphylocoagulase	A	IVTKDYSKESR	V	ADA↓IVT	4	4
10	A0A0H3JT93	SA0620 protein	A	STQHTVQSGESLWSIAQKYNTSVESIK	Q	AQA↓STQ	2	2
11	P68800	Fibrinogen-binding protein	A	SEGYGPR	E	ADA↓SEG	7	6
12	Q99SU9	Staphylococcal complement inhibitor	A	STSLPTSNEYQNEKLANELK	S	AQA↓STS	35	30
13	A0A0H3JK15	Uncharacterized protein	A	ASGNSIDTVK	Q	AEA↓ASG	2	2
14	A0A0H3JPG2	Uncharacterized protein	A	NEEQKSSLLENQK	E	AYA↓NEE	2	2
15	P65289	Lipase 1	A	AEKQVNMGNSQEDTVTAQSIGDQQTR	E	AQA↓AEK	2	2
16	Q7A6P2	Thermonuclease	A	SQTDNGVNR	E	ANA↓SQT	12	12

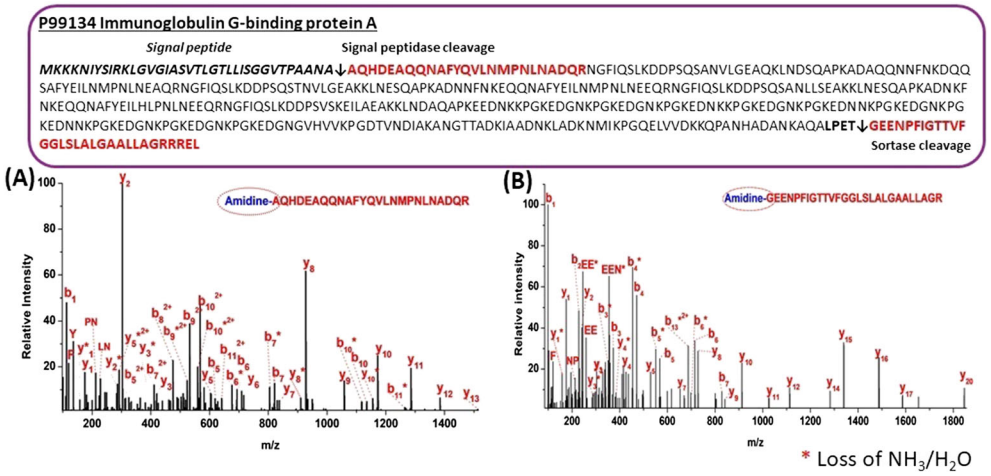

FIGURE 3 Mass spectra demonstrating the N-terminal amidination of *S. aureus* N315 secreted proteins. A, MS/MS spectrum of the N-terminally amidinated peptide from immunoglobulin G-binding protein A. Its N-terminus became accessible for labeling following a normal signal peptidase cleavage. B, MS/MS spectrum of an N-terminally amidinated peptide from the same protein. Its N-terminus became accessible for labeling following cleavage by sortase at an LPXTG site [Color figure can be viewed at wileyonlinelibrary.com]

more interesting observation. The N-terminally amidinated peptide GEENPFIGTTVFGGLSLALGAALLAGRRREL was identified from the spectrum in Figure 3B. Although this LPXTG cleavage by sortase has already been reported, this particular peptide has not been previously observed.[31] Amidination may facilitate its detection. The mass of the peptide, and the appearance of multiple b-type ions including the prominent b_1 fragment ion peak, confirm that its N-terminus is labeled. This spectrum was recorded 123 times over the course of eight experiments reflecting the abundance of this peptide in the extracellular medium. It was amidinated in every case. Sortase reportedly cleaves and anchors other surface or cell wall associated proteins including fnbA, fnbB, clfA, clfB, sdrC, sdrD, sdrE and pls. Unamidinated peptides from these proteins were observed but we did not detect C-terminal peptides that characterize the sortase cleavage site.[32] This may not be surprising since Mazmanian et al[32] did indicate that peptides C-terminal to this cleavage site are degraded.

Probable transglycosylase (IsaA) is a housekeeping protein that is non-covalently bound to the cell wall. Its high immunogenic properties make it a potential target for immuno-therapy development.[33] A mature form of the protein is secreted after removal of the 29-residue signal peptide MKKTIMASSLAVAL GVTGYAAGTGHQAHA. As displayed in Figure 4A, the resulting peptide AEVNVDQAHLVDLAHNHQDQLNAAPIK was observed always N-terminally amidinated. The spectral match was excellent and the plethora of b- and y-type ions including the intense b_1 ion confirmed this assignment. Observation of this signal peptide cleavage

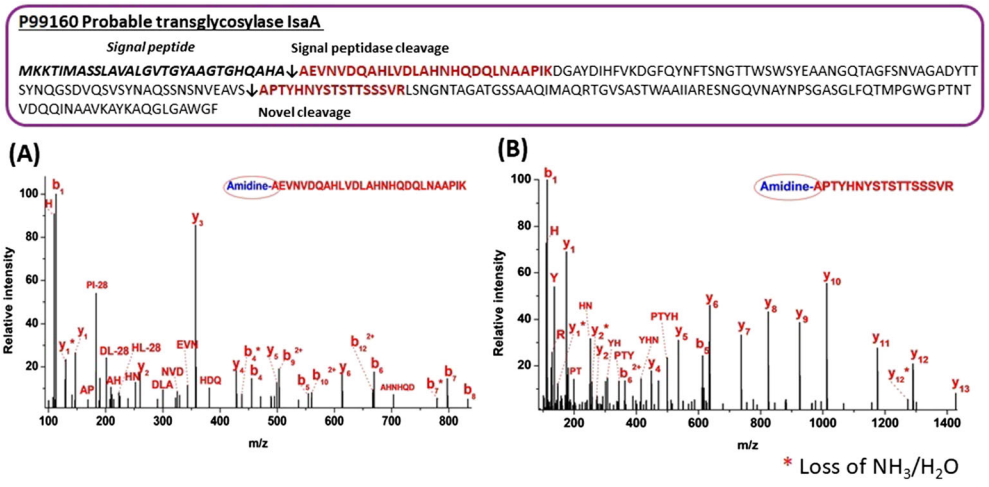

FIGURE 4 A, MS/MS spectrum of an N-terminally amidinated peptide of IsaA protein confirming another normal signal peptidase cleavage site. B, MS/MS spectrum of an N-terminally amidinated peptide from the same protein demonstrating an unexpected internal cleavage site. (note that the b_1 ion peak is off-scale) [Color figure can be viewed at wileyonlinelibrary.com]

is consistent with previous reported results.[16,34] Much more interestingly, we observed another cleavage in the middle of this protein following residue 130 (AVS^{130}↓APT). As shown in Figure 4B, the peptide APTYHNYSTSTTSSSVR was detected with an N-terminally amidinated alanine. The MS/MS spectral match was again excellent, and numerous b- and y-type fragment ions including an intense b_1 ion peak substantiate this peptide assignment. This cleavage, which has not been previously reported, was observed in multiple replicates and probably occurs during translocation of the protein into the extracellular matrix. Although it is not near the protein N-terminus, the cleavage site (AVS↓) somewhat resembles a standard AXA↓ signal peptidase motif, suggesting that this protein may be processed twice for different biological functions. While less than 5% of the peptides identified in our sample were found to be N-terminally amidinated, this particular peptide was amidinated 28 of the 47 times that we observed it in eight separate experiments. As noted above, we are amidinating folded proteins or protein fragments and labeling sites are not always completely accessible. Consequently, we often detect peptides whose N-termini are in both labeled and unlabeled states. Nevertheless, definitive observation of an N-terminally amidinated peptide implies that cleavage had occurred before the labeling reaction. In eight separate experiments, we observed 36725 total PSMs and 365 proteins. Out of 365 proteins, 116 proteins were labeled at N-termini including the unexpected middle proteolytic cleavages. However, many peptides from the N-termini of exported proteins were not observed. This may be because of their inaccessibility or because near the N-termini of proteins there are often multiple lysine and arginine basic residues that could lead to very small peptides or an array of partially labeled and partially unlabeled peptides.

Approximately 30 proteins found in the extracellular medium contained unexpected proteolytic cleavages of this sort, presumably associated with their extracellular transport. In all cases these internal cleavages were identified through the observation of N-terminally amidinated peptides whose MS/MS spectra contained b_1 ions. Corroborating peptide and top-down protein mass spectrometric evidence of all of these observations will be presented in a subsequent publication. Future work will be aimed at investigating whether any cytosolic proteins also display these unexpected internal cleavages or whether they are confined to exported proteins. We also aim to identify cleavage mechanisms and determine the effects that this processing has on *Staphylococcus aureus* protein function.

4 | CONCLUSIONS

Pathogenicity of *S. aureus* is expedited by the secretion of a variety of toxic proteins and enzymes into the external environment. This intrinsic protein secretion is associated with known and unknown secretory pathways including those that produce N-terminal signal peptides. The chemical labeling/mass spectrometry methodology presented in this work facilitates the identification of secreted mature proteins and the determination of their cleavage sites. When N-terminally amidinated peptide ions fragment they generally yield distinctive b_1 ions that assist in their identification. N-terminal amidination also probes the solvent accessibility of the pre- and post-processed protein N-terminus. Besides identifying secreted protein cleavage sites that had been expected, a number of unforeseen cleavage sites have also been observed. Future experiments of this type will investigate the prevalence of these unexpected processes with the eventual goal of determining the mechanisms and reasons for their occurrence.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (Grant Nos. R01 GM103725 and U54 GM074807. It was partially funded by the Indiana University Vice Provost for Research through the Faculty Research Support Program and the Indiana University Precision Health Initiative. The authors wish to thank the Floyd Romesberg group at Scripps Research Institute, San Diego, California, for providing a stock culture of *S. aureus* N315 cells.

ORCID

Santosh A. Misal https://orcid.org/0000-0002-8835-0130
James P. Reilly https://orcid.org/0000-0002-9784-1804

REFERENCES

1. Hartmann EM, Armengaud J. N-terminomics and proteogenomics, getting off to a good start. *Proteomics*. 2014;14(23–24):2637-2646.
2. Thompson CR, Champion MM, Champion PA. Quantitative n-terminal footprinting of pathogenic mycobacteria reveals differential protein acetylation. *J Proteome Res*. 2018;17(9):3246-3258. https://doi.org/10.1021/acs.jproteome.8b00373
3. Kim JS, Dai Z, Aryal UK, et al. Resin-assisted enrichment of N-terminal peptides for characterizing proteolytic processing. *Anal Chem*. 2013;85(14):6826-6832.
4. Timmer JC, Enoksson M, Wildfang E, et al. Profiling constitutive proteolytic events in vivo. *Biochem J*. 2007;407(1):41-48.
5. Westermann B, Jacome ASV, Rompais M, Carapito C, Schaeffer-Reiss C. Doublet N-terminal oriented proteomics for N-terminomics and proteolytic processing identification. *Methods Mol Biol*. 2017;1574:77-90.
6. He Y, Parthasarathi R, Raghavachari K, Reilly JP. Photodissociation of charge tagged peptides. *J Am Soc Mass Spectrom*. 2012;23(7):1182-1190. https://doi.org/10.1007/s13361-012-0379-x
7. Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem*. 2003;75(24):6843-6852.
8. Bland C, Hartmann EM, Christie-Oleza JA, Fernandez B, Armengaud J. N-Terminal-oriented proteogenomics of the marine bacterium *Roseobacter denitrificans* och114 using (*N*-succinimidyloxycarbonylmethyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP) labeling and diagonal chromatography. *Mol Cell Proteomics*. 2014;13(5):1369-1381.
9. Miller BT, Collins TJ, Nagle GT, Kurosky A. The occurrence of O-acylation during biotinylation of gonadotropin-releasing hormone and analogs. Evidence for a reactive serine. *J Biol Chem*. 1992;267(8):5060-5069.

10. Miller BT, Kurosky A. Elevated intrinsic reactivity of seryl hydroxyl-groups within the linear peptide triads His-Xaa-Ser or Ser-Xaa-His. *Biochem Biophys Res Commun*. 1993;196(1):461-467.
11. Miller BT, Collins TJ, Rogers ME, Kurosky A. Peptide biotinylation with amine-reactive esters: Differential side chain reactivity. *Peptides*. 1997;18(10):1585-1595.
12. Swaim CL, Smith JB, Smith DL. Unexpected products from the reaction of the synthetic cross-linker 3,3′-dithiobis (sulfosuccinimidyl propionate), DTSSP with peptides. *J Am Soc Mass Spectrom*. 2004;15(5):736-749.
13. Kalkhof S, Sinz A. Chances and pitfalls of chemical cross-linking with amine-reactive *N*-hydroxysuccinimide esters. *Anal Bioanal Chem*. 2008;392(1–2):305-312.
14. Madler S, Bich C, Touboul D, Zenobi R. Chemical cross-linking with NHS esters: A systematic study on amino acid reactivities. *J Mass Spectrom*. 2009;44(5):694-706.
15. Green ER, Mecsas J. Bacterial secretion systems: An overview. *Microbiol Spectr*. 2016;4(1). https://doi.org/10.1128/microbiolspec.VMBF-0012-2015
16. Ravipaty S, Reilly JP. Comprehensive characterization of methicillin-resistant *Staphylococcus aureus* subsp. aureus col secretome by two-dimensional liquid chromatography and mass spectrometry. *Mol Cell Proteomics*. 2010;9(9):1898-1919. https://doi.org/10.1074/mcp.M900494-MCP200
17. de Keyzer J, van der Does C, Driessen AJ. The bacterial translocase: A dynamic protein channel complex. *Cell Mol Life Sci*. 2003;60(10):2034-2052.
18. Sibbald MJ, Ziebandt AK, Engelmann S, et al. Mapping the pathways to staphylococcal pathogenesis by comparative secretomics. *Microbiol Mol Biol Rev*. 2006;70(3):755-788.
19. Tjalsma H, Antelmann H, Jongbloed JD, et al. Proteomics of protein secretion by *Bacillus subtilis*: Separating the "secrets" of the secretome. *Microbiol Mol Biol Rev*. 2004;68(2):207-233.
20. Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM. Signal peptide-dependent protein transport in bacillus subtilis: A genome-based survey of the secretome. *Microbiol Mol Biol Rev*. 2000;64(3):515-547.
21. Craney A, Dix MM, Adhikary R, Cravatt BF, Romesberg FE. An alternative terminal step of the general secretory pathway in *Staphylococcus aureus*. *MBio*. 2015;6(4). https://doi.org/10.1128/mBio.01178-15
22. Morisaki JH, Smith PA, Date SV, et al. A putative bacterial abc transporter circumvents the essentiality of signal peptidase. *MBio*. 2016;7(5). https://doi.org/10.1128/mBio.00412-16
23. Thumm M, Hoenes J, Pfleiderer G. S-methylthioacetimidate is a new reagent for the amidination of proteins at low ph. *Biochim Biophys Acta*. 1987;923(2):263-267.
24. Beardsley RL, Reilly JP. Fragmentation of amidinated peptide ions. *J Am Soc Mass Spectrom*. 2004;15(2):158-167. https://doi.org/10.1016/j.jasms.2003.10.007
25. Beardsley RL, Sharon LA, Reilly JP. Peptide de novo sequencing facilitated by a dual-labeling strategy. *Anal Chem*. 2005;77(19):6300-6309. https://doi.org/10.1021/ac050540k
26. Li S, Dabir A, Misal SA, Tang H, Radivojac P, Reilly JP. Impact of amidination on peptide fragmentation and identification in shotgun proteomics. *J Proteome Res*. 2016;15(10):3656-3665. https://doi.org/10.1021/acs.jproteome.6b00468
27. Harrison AG. To b or not to b: The ongoing saga of peptide b ions. *Mass Spectrom Rev*. 2009;28(4):640-654.
28. Carneiro CR, Postol E, Nomizo R, Reis LF, Brentani RR. Identification of enolase as a laminin-binding protein on the surface of staphylococcus aureus. *Microbes Infect*. 2004;6(6):604-608.
29. Frottin F, Martinez A, Peynot P, et al. The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics*. 2006;5(12):2336-2349.
30. Xiao Q, Zhang F, Nacev BA, Liu JO, Pei D. Protein N-terminal processing: Substrate specificity of *Escherichia coli* and human methionine aminopeptidases. *Biochemistry*. 2010;49(26):5588-5599.
31. Becker S, Frankel MB, Schneewind O, Missiakas D. Release of protein a from the cell wall of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A*. 2014;111(4):1574-1579.
32. Mazmanian SK, Ton-That H, Schneewind O. Sortase-catalysed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. *Mol Microbiol*. 2001;40(5):1049-1057.
33. Stapleton MR, Horsburgh MJ, Hayhurst EJ, et al. Characterization of IsaA and SceD, two putative lytic transglycosylases of *Staphylococcus aureus*. *J Bacteriol*. 2007;189(20):7316-7325.
34. Craney A, Romesberg FE. Stable signal peptides and the response to secretion stress in *Staphylococcus aureus*. *MBio*. 2017;8(6). https://doi.org/10.1128/mBio.01507-17

How to cite this article: Misal SA, Li S, Tang H, Radivojac P, Reilly JP. Identification of N-terminal protein processing sites by chemical labeling mass spectrometry. *Rapid Commun Mass Spectrom*. 2019;33:1015–1023. https://doi.org/10.1002/rcm.8435