Published in final edited form as:

Biochim Biophys Acta. 2010 June ; 1804(6): 1394–1404. doi:10.1016/j.bbapap.2010.02.006.

Using chemical derivatization and mass spectrometric analysis to characterize the post-translationally modified *Staphylococcus aureus* surface protein G

Moo-Jin Suh^{*}, David J. Clark, Prashanth P. Parmer, Robert D. Fleischmann, Scott N. Peterson, and Rembert Pieper

Pathogen Functional Genomics Resource Center, J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, U.S.A.

Abstract

The *Staphylococcus aureus* surface protein G (SasG) is an important mediator of biofilm formation in virulent *S. aureus* strains. A detailed analysis of its primary sequence has not been reported to date. SasG is highly abundant in the cell wall of the vancomycin-intermediate *S. aureus* strain HIP5827, and was purified and subjected to sequence analysis by MS. Data from MALDI-TOF and LC-MS/MS experiments confirmed the predicted N-terminal signal peptide cleavage site at residue A₅₁ and the C-terminal cell wall anchor site at residue T₁₀₈₆. The protein was also derivatized with N-succinimidylloxycarbonyl-methyl-tris(2,4,6-trimethoxyphenyl) phosphonium bromide (TMPP-Ac-OSu) to assess the presence of additional N-terminal sites of mature SasG. TMPP-derivatized SasG peptides featured *m/z* peaks with a 572 Da mass increase over the equivalent underivatized peptides. Multiple N-terminal peptides, all of which were observed in the 150 amino acid segment following the signal peptide cleavage at the residue A₅₁, were characterized from MS and MS/MS data, suggesting a series of successive N-terminal truncations of SasG. A strategy combining TMPP derivatization, multiple enzyme digestions to generate overlapping peptides and detailed MS analysis will be useful to determine and understand functional implications of PTMs in bacterial cell wall-anchored proteins, which are frequently involved in the modulation of virulence-associated bacterial surface properties.

Keywords

N-terminal truncation; TMPP labeling; multiple enzyme digestion; SasG; mass spectrometry; post-translational modifications

1. Introduction

Secreted proteins play an important role in the pathogenicity of *Staphylococcus aureus* and are the primary source of vaccine and diagnostic candidates. A majority of these proteins are exported via the signal peptidase I-dependent pathway, and have a signal peptide that is cleaved off during the secretion process. The signal peptide is normally between 15 and 50 amino acids

© 2009 Elsevier B.V. All rights reserved.

^{*} To whom correspondence should be addressed, Phone (301) 795-7708, Fax (301) 838-9142, msuh@jvci.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

long and is essential for protein secretion and correct localization of the target proteins [1]. *S. aureus* surface protein G (SasG), one of cell wall-associated proteins, is a multi-domain protein with a predicted N-terminal signal peptide of 50 amino acids, an A-domain, a B-domain with 128 amino acid repeat units, and a C-terminal segment [2]. The C-terminal part of SasG harbors the LPXTG motif in which a threonine residue is ultimately linked to the peptidoglycan via an ester bond. The protein was associated with adhesion to epithelial cells, when it was recombinantly expressed in strains not naturally producing SasG [3]. Thereafter, SasG was later shown to promote the formation of biofilms [4,5]. Of note, *S. aureus* strains capable of biofilm formation generally show higher resistance to antibiotics and protection from the human immune defense system [6]. While the A-domain of SasG has been associated with biofilm formation [7], the role of the B-repeats is more controversial. The latter domain has glycine-rich regions, which were associated with N-acetyl-D-glucosamine binding. Its role as a molecular spring, allowing presentation of the A-domain at the cell surface was also proposed [7]. Finally, its involvement in biofilm formation was suggested [5]. The strain HIP5827 derived from a clinical vancomycin-intermediate *S. aureus* (VISA) isolate expresses high quantities of SasG in the cell wall. Judged by the complexity of 2D gel spots from our previous work, this cell surface protein appeared to be extensively modified, resulting in differences of size and charge [8].

Targeted truncations at export signal cleavage and sorting sites are common PTMs. Prediction of export signal sites in N-terminal protein regions is accomplished by running one of the existing computational prediction programs such as SignalP [9] and PrediSi [10]. This can be experimentally verified by sequencing the mature protein's N-terminus by Edman analysis and MS [11]. MS has been the method of choice for the identification of proteins and their PTMs. MS often provides accurate information of cleavage sites and is sufficiently sensitive even if only small amounts of protein are available. The determination of specific terminal peptides can be compromised by physicochemical characteristics rendering these peptides unsuitable for ionization, detection and/or fragmentation (non-proteotypic peptides) in bottom-up analysis approach.

TMPP-Ac-OSu has been introduced as a reagent for N- and C-terminal sequence analysis of proteins [12,13]. *De novo* sequencing of protein termini with the TMPP labeling method provides higher fidelity [14]. The ability to identify low mass peptides with as few as four residues and the improved ionization efficiency during MS/MS fragmentation are further advantages of the TMPP labeling method. Based on the extensive 2DE spot trains formed by SasG and the apparent M_r value differences in 2DE gels [8], TMPP derivatization was employed to elucidate whether this protein is proteolytically clipped beyond its *in silico* predicted N-terminus. Key to the identification of N-terminal peptides is a comparison of mass spectra prior to and after derivatization with the TMPP reagent. To validate our findings, a multiple enzyme digestion strategy was employed. Peptides with overlapping sequence regions yielded MS and MS/MS data verifying many of the newly detected N-terminal sites. Exhaustive MS analysis also provided nearly complete sequence coverage for SasG. The strategy described here sets the stage for the comprehensive analysis of post-translational truncations of bacterial cell wall-immobilized proteins which may contribute to their activating or deactivating functions.

2. Materials and Methods

2.1. Materials

Sequencing grade modified trypsin was purchased from Promega Corporation (Madison, WI). Sequencing grade endopeptidases Glu-C and Asp-N were obtained from Roche Diagnostics (Dallas, TX). (N-succinimidylloxycarbonyl-methyl)tris(2,4,6-trimethoxyphenyl)

phosphonium bromide (TMPP-Ac-OSu), α -cyano-4-hydroxy-cinnamic acid (CHCA) and other chemicals were purchased from Sigma-Aldrich (St. Louis, MO).

2.2. *S. aureus* strains, bacterial culture conditions and isolation of SasG

The bacterial strain used in this study was a vancomycin-intermediate resistant *S. aureus* (VISA) strain derived from the HIP5827 clinical isolate [15]. The strain was usually grown to stationary phase ($OD_{600} \sim 2.0$) in brain heart infusion broth without antibiotics as previously described [8]. Cells were washed with PBS and pelleted by centrifugation at $6,000 \times g$ for 20 min prior to isolation of cell wall proteins.

2.3. Isolation of the cell wall digestion supernatant

The cell wall digestion supernatant generated with lysostaphin (1-CWDS) was prepared from freshly harvested cells of strain HIP5827 as described before [8], with a few modifications. Briefly, lysostaphin (75 $\mu\text{g/mL}$) and lysozyme (100 $\mu\text{g/mL}$) were added to cell suspensions (6 mL/g cells) followed by incubation in a hypertonic buffer at 37°C for *ca.* 6 h. The 1-CWDS fraction was separated from the periplasts and undigested cells by centrifugation at $7,000 \times g$ for 20 min and adjusted to 60% saturation with solid $(\text{NH}_4)_2\text{SO}_4$. Protein precipitation on ice was followed by centrifugation at $50,000 \times g$ for 30 min at 4°C. The protein was concentrated at $3,000 \times g$ in Ultrafree-10k centrifugal filtration devices (Fisher Scientific) at 4°C and exchanged into buffer B (25 mM NH_4HCO_3 , 1 mM benzamidinium and 1 mM EDTA) for analysis in gels or, alternatively, into PBS and 15% glycerol for storage at -80°C . All fractions were analyzed in 4–12% T SDS-PAGE gels followed by protein staining with Coomassie Blue G-250 (CBB).

2.4. Purification of SasG

As described above, SasG was enriched in a 1-CWDS fraction followed by a 60% saturated $(\text{NH}_4)_2\text{SO}_4$ precipitation step. The precipitate was resuspended in buffer C (PBS, 2mM EDTA, 1 mM BAM and 2 mM PMSF) at a concentration of *ca.* 10 mg/mL, loaded onto a SuperdexTM200 column (1 m $\times$ 1.6 cm) and eluted by size-exclusion chromatography (SEC) at a flow rate of 0.8 mL/min with buffer C. Peaks were detected at $\lambda=A_{280}$. Fractions were collected and analyzed by SDS-PAGE. Fractions apparently enriched in SasG (based on band intensities in the M_r range of *ca.* 125–160 kDa) were pooled, diluted in a solution with a final concentration of 50 mM NaCl, and adjusted to pH 6.8. The pooled sample was loaded onto a 3.8 mL POROS-HS 20 column. SCX chromatography was performed eluting with a linear 1 M NaCl gradient of buffer A at a flow rate of 2 mL/min, followed by analysis of all fractions in SDS-PAGE gels and BCA assays to determine protein quantities. Samples containing SasG were concentrated to *ca.* 1–2 mg/mL using Ultrafree 10K filters and stored in PBS and 15% glycerol at -80°C . All chromatography steps were carried out at 20°C, sample concentration steps at 4°C.

2.5. Derivatization of SasG with TMPP

A solution of 0.1 M TMPP-Ac-OSu in CH_3CN :water (2:8, v/v) was added at a molar ratio of 200:1 to 50 μg of purified SasG solubilized in 50 mM NH_4HCO_3 (pH 8.0) as previously described [14]. At this pH, the more basic lysine amino group is protected by protonation and does not react with TMPP-Ac-OSu. The reaction mixture was vortexed and incubated for 1 h at 20°C. Residual labeling reagent was quenched by incubation with a solution of 0.1 M hydroxylamine for 1 h at 20°C. To remove all traces of TMPP-Ac-OSu, SasG sample was loaded on a 4–12% SDS PAGE after completion of the reaction of TMPP-Ac-OSu with the protein.

2.6. Protein analysis in SDS-PAGE and 2D gels

Purified SasG was applied to separate in 2D gels. The gels were stained with CBB, and spots of interest were subjected to automated spot excision, in-gel tryptic digestion and mass spectrometric analysis using methods previously described [8]. Briefly, 24 cm IPG strips (GE Healthcare, pH range 4–7) were used in the 1st dimension separation, followed by reduction/alkylation steps and 2nd dimension separation in SDS-PAGE slab gels (25 × 19.5 × 0.15 cm). In addition, unmodified SasG and TMPP-derivatized SasG were separated in 4–12%T SDS-PAGE gels using standard electrophoresis conditions and stained with CBB.

2.7. In-solution and in-gel proteolytic digestion of SasG

Purified and TMPP-derivatized SasG (50 µg each) were incubated with two or three site-specific proteases in 100:1 (w/w) ratios (substrate to enzyme) for 24 h using the following conditions. Digestion with the endopeptidase Glu-C digestion was performed in 100 µL of 50 mM NH₄HCO₃ (pH 8.0) at 25°C. Digestion with the endopeptidases Asp-N and trypsin was performed in 100 µL of 50 mM Tris buffer (pH 8.0) at 37°C. The proteolytic reactions were quenched by addition of 5 µL 25% formic acid, thus decreasing the pH to 3. Each of the samples were divided into 10 µL aliquots and stored at –80°C prior to further use. In control experiments, each enzyme was also incubated without the presence of SasG to monitor enzyme autolysis products. For in-gel digestion, a method described previously was applied [8]. Briefly, plugs from either SDS-PAGE or 2D gels containing SasG in the M_r range of ca. 125–165 kDa were excised. This was followed by overnight digestion with 125 ng trypsin and, in the case of SDS-PAGE gel plugs, with 125 ng Glu-C.

2.8. Mass spectrometry analysis

MALDI-TOF MS: MALDI target plates were applied to analysis using a MALDI-TOF/TOF mass spectrometer (4700 Proteomics Analyzer, Applied Biosystems). For peptide identification experiments, 100 mM CHCA in 50% aqueous acetonitrile/ 0.1% TFA was used as the matrix. The ten most intensive ions from the MS scan were selected for MS/MS fragmentation. Spectra represented averaged signals of 1,000 laser shots for MS analysis and 3,000 shots for MS/MS analysis. A default calibration was applied using a six-component peptide standard mixture which was spotted onto six edge positions of the MALDI target plates.

LC-nESI-MS/MS: LC-nESI-MS/MS analysis was performed with a LTQ ion trap mass spectrometer (Thermo-Finnigan, San Jose, CA) equipped with a Finnigan nanoESI source. An Agilent 1100 series solvent delivery system (Agilent, Palo Alto, CA) was interfaced with the LTQ instrument to deliver samples to a trapping cartridge (CapTrap, Michrom BioResources, Auburn, CA), followed by a reversed-phase gradient to elute peptides off the trap and a subsequent PicoTip microcapillary reversed-phase column (BioBasic C₁₈, 75 µm × 10 cm, New Objective, Woburn, MA). Spectra were acquired in automated MS/MS mode, with the top five parent ions above a threshold of 10,000 selected for fragmentation. Details of the LTQ instrument setup were previously provided [8].

MS data and bioinformatic analyses: MS and MS/MS data obtained from the MALDI and LC-nESI-MS/MS experiments were searched against the latest releases of the Swiss-Prot and non-redundant NCBI protein databases (*S. aureus* Mu3 strain) using the search engine Mascot v. 2.2 (Matrix Science, London, UK). An in house-developed LIMS system was used to automate Mascot data processing, display the identified peptides with their ranks, the MS/MS scores and a protein sequence coverage summary. For the prediction of N-terminal signal peptides, data were searched with the algorithms SignalP [9] and PrediSi [10]. For the identification of N-terminal peptides, the Mascot searches were modified accounting for N-terminal modification with the TMPP tag. The mass of 572.18 Da was considered as a variable modification. In addition, semi-trypsin and semi-Glu-C peptides were used in the field specifying enzymatic

digestion conditions. The MS/MS spectra of N-terminally labeled peptides were also manually inspected. Algorithms provided in the website FindMol (<http://www.expasy.ch/sprot/findmod.html>) and Inspect [16] were used to examine the data with respect to other peptide modifications, followed by manual inspection of the mass spectra.

3. Results and Discussion

3.1. Gel display of the multi-domain protein SasG

Proteins exported from the cytoplasm to the cell envelope require proteolytic cleavage of an N-terminal signal peptide with a length of typically 15–50 amino acid residues. The reaction is catalyzed by a membrane-anchored signal peptidase as part of the main secretion (Sec) pathway [17]. Additional modifications, frequently also proteolytic cleavage, occur in proteins exported as pro-enzymes. The glycyl-glycine endopeptidase LytM is an example of a *S. aureus* pro-enzyme [18]. Modifications also occur at the C-terminus, as observed in proteins of *Staphylococci* anchored in the cell wall peptidoglycan [19]. The sorting signal near the protein's C-terminus typically is a highly conserved LPXTG motif. Cell wall anchoring prevents secretion of such proteins into extracellular media. SasG is a cell wall-anchored protein with an LPXTG motif involved in the formation of *S. aureus* biofilms. N-terminal proteolytic processing may play a role in functional activation of SasG, but this has not yet been thoroughly investigated.

Sequence analysis of the *sasG* gene from the VISA strain HIP5827 revealed about 99% amino acid sequence identity with the *sasG* gene (locus SAHV_2481) annotated in the Mu3 strain database (unpublished data). Surprisingly, *sasG* genes detected in the genomes of most *S. aureus* strains remain poorly annotated, and are often described as hypothetical or uncharacterized cell surface-anchored proteins. Assuming the absence of additional PTMs, one would predict a single SasG spot with a M_r of 122224 and a pI of 5.69 in 2DE gels. Gel display data shown in Figure 1 reveal that SasG solubilized from cell walls with lysostaphin has modifications affecting its M_r and pI values. All spots from 2DE gels were analyzed by peptide mass fingerprinting after in-gel digestion and confirmed as the SasG (data not shown). Furthermore, immunoblot analysis with polyclonal anti-SasG chicken IgY antibodies revealed an extended band area in SDS-PAGE gels and M_r variability of SasG (data not shown). In fact, the M_r values were estimated to be up to 40 kDa greater than the computed M_r of 122224, while the pI values revealed a range from 5.1 to 5.5. Upon solubilization from the cell wall, SasG clearly is a heterogeneous protein. We examined this heterogeneity on the primary sequence level with mass spectrometry methods.

3.2. Sequence coverage of SasG purified from the cell wall of strain HIP5827 using a combination of enzymatic digests and MS analysis

PCR analysis of *sasG* revealed a gene length in strain HIP5827 identical to that of *sasG* annotated in the Mu3 genome (unpublished data). To assess sequence coverage by MS, it was therefore reasonable to use the SasG protein sequence annotated as SAHV_2481 in strain Mu3 as a template. MS data from in-solution digests of purified SasG yielded 77% sequence coverage with trypsin. The combination of trypsin and Glu-C was particularly effective, increasing the sequence coverage to 86% with a redundancy of 59% between the two datasets. MS-based sequence coverage data derived from two or three enzyme combinations are summarized in Table 1. About 7% of the nascent SasG sequence corresponds to the N- and C-terminal signal peptide regions (1M–A₅₀ and 1087G–N₁₁₁₅, respectively). These segments were not included in the calculation of protein sequence coverage. Finally, only 4% of the mature SasG sequence was not covered by our MS data, possibly due to short, frequently cleaved peptide fragments or other physicochemical properties resulting in inability to detect such peptides by MS. Sequence coverage data presented in Figure S1 demonstrate that a combination

of enzymatic digests (Asp-N, Glu-C and trypsin) resulted in coverage of nearly the entire protein sequence by MS (96%).

3.3. N-deamidation results in SasG variants with different pI values

Analysis of a comprehensive set of MS and MS/MS data on SasG enabled us to systematically search for amino acid residue modifications with a potential effect on the protein's pI value. The horizontal spot trains in the 2D gel image displayed in Figure 1A were indicative of modifications resulting in regular pI shifts while maintaining the M_r . Using three endopeptidases for the generation of SasG peptides, amino acid modifications affecting the same residue in overlapping peptides were studied. The PMFs derived from high mass accuracy MALDI-TOF datasets revealed N-deamidation for four different residues (N_{204} , N_{257} , N_{401} , and N_{1051}) in SasG, all of which are part of NG motifs prone to chemical deamidation. MS/MS spectra confirmed the presence of deamidated residues, as shown in one example in Figure S2. Previous studies have shown that deamidation of N and Q residues in proteins is common, particularly when proteins are exposed to alkaline conditions [20]. Although N-deamidations can be chemical artifacts, we cannot rule out the occurrence of these deamidations as a natural protein aging process. The exposure of SasG at the bacterial cell surface makes a natural aging process plausible.

3.4. N- and C-terminal processing of SasG

2D gel data also revealed substantial differences in M_r values of the SasG spot trains. Two alternative explanations were considered: (1) variability in the truncation of SasG at the N- and/or C-terminus beyond the predicted signal peptide cleavage and sortase cleavage sites; (2) variability in the retention of peptidoglycan fragments at the C-terminus of SasG due to incomplete cleavage of pentaglycine bridges by lysostaphin during cell wall digestion. SasG was previously proposed to be N-terminally cleaved serving as a maturation step to allow *S. aureus* biofilm formation [4]. To assess whether SasG underwent proteolytic modifications, N-terminal labeling with TMPP was combined with a three-enzyme digestion strategy and comprehensive peptide analysis by MS and MS/MS. The prediction of signal peptides by computational prediction programs (PrediSi and Signal IP) suggested that the sequence of exported SasG started with either the amino acid residue A_{48} or A_{51} , respectively. Peptide data representing the N-terminal part of the SasG sequence were searched and analyzed. The signal peptidase cleavage site was shown to reside between amino acids A_{50}/A_{51} . The peptide $_{51}AENNIENPT\ TLK_{62}$ was detected at m/z 1343.75 (singly charged ion) using MALDI-TOF MS and at m/z 672.46 (doubly charged ion) using LC-MS/MS. As shown in Figure 2, MS/MS data from both MALDI-TOF/TOF and LC-nESI-MS/MS experiments confirmed the sequence identity of this peptide.

The LPXTG motif of the SasG peptide $_{1083}LPXTG_{1087}$ is generally associated with C-terminal processing and cell wall anchorage of *S. aureus* proteins. *S. aureus* sortase A cleaves most surface protein precursors bearing a C-terminal LPXTG motif sorting signal between the threonine residue and the glycine residue [21]. Thus, the predicted peptidoglycan anchor site in SasG is T_{1086} . Indeed, m/z values of PMF data were consistent with the C-terminal threonine, followed by C-terminal glycines that are derived from the peptidoglycan pentaglycine cross-bridge. This cross-bridge is the cleavage site for the enzyme lysostaphin which was used to solubilize SasG from the cell wall (Figure 3A). The m/z values matching the masses of the peptides $_{1078}KKRAELPKT_{1086}GG$ and $_{1078}KKRAELPKT_{1086}GGG$ were identified from Glu-C digests. Each peptide represents a natural fusion peptide; part protein and part peptidoglycan. A zoomed-in area of the PMF obtained by MALDI-TOF analysis is shown in Figure 3B and depicts both peptides (m/z values of 1184.70 and 1241.72). The m/z values did not match any other SasG peptides resulting from Glu-C digestion. To rule out the possibility that precursor peaks at m/z 1184.70 and 1241.72 represented something else, MS/MS data were

acquired and confirmed as the sequence identity of the peptides $_{1078}\text{KKRAELPKT}_{1086}\text{GG}$ and $_{1078}\text{KKRAELPKT}_{1086}\text{GGG}$ (Figure 3C and D), respectively. Ion scores of these peptides are summarized in Table 2. These results are consistent with previously obtained data on a lysostaphin-released anchor peptide from the staphylococcal surface protein IsdC, an iron (Fe^{2+}) ABC superfamily ATP binding cassette transporter that harbors a different sorting motif: NPQTN [19]. In summary, MS data initially confirmed the predicted PTMs at the N- and the C-terminus of cell wall-immobilized SasG. This data did not sufficiently explain the M_r differences of SasG spot trains observed in 2DE gels.

3.5. MS analysis of TMPP-derivatized peptides reveals further N-terminal truncation of SasG

Several research groups have speculated that targeted N-terminal truncation of SasG is an activation step for biofilm formation [4,7]. The M_r diversity of SasG indeed supports such truncation events. The extensive analysis of TMPP-derivatized peptides provided precise information on N-terminal residues in SasG variants. The N-terminal labeling reagent TMPP-Ac-OSu was selected for following reasons: (1) increased ionization efficiency during MS analysis; (2) known experimental conditions for N-terminal protein derivatization; (3) identification of lower mass peptides (consisting of 4–5 amino acids) by the addition of a mass tag. N-terminal α -amine coupling ($\text{pK}_a \sim 8$) without derivatization of lysine side chain ϵ -amines ($\text{pK}_a \sim 10.5$) was achieved by a controlled pH environment (pH 8.0) [22]. Following the coupling reaction, SasG was separated from excess TMPP reagent by SDS-PAGE. As shown in Figure 1B, TMPP derivatization did not interfere with the M_r -based separation of SasG variants in SDS-PAGE gels. Comparing TMPP-derivatized SasG and underivatized SasG as a control, peptide analysis of in-gel digests with Trypsin and Glu-C was carried out, and MS data were queried for the characteristic TMPP-associated mass shifts of 572 Da.

The anticipated N-terminal peptide ($_{51}\text{AENNIENPTTLK}_{62}$) labeled with TMPP (m/z 1915.86 of TMPP-Ac- $_{51}\text{AENNIENPTTLK}_{62}$) was identified from both LC-MS/MS and MALDI-TOF MS data. The MS/MS fragmentation spectra for the tryptic peptide TMPP-Ac- $_{51}\text{AENNIENPTTLK}_{62}$ are depicted in Figure 4. The spectral peak signal intensity of b-type ions shown in Figure 4 was indicative of increased ionization efficiency due to TMPP-modification, which introduced a permanent positive charge at the N-terminus. In a similar fashion, other mass spectral peaks were investigated. The PMF in Figure 5B shows an m/z segment in which a new peak (m/z 1172.52) was assigned to the peptide TMPP-Ac- $_{124}\text{EVQPK}_{128}$. This peptide peak was absent in the PMF of unmodified SasG (Figure 5A). Internal peptides remained unchanged with respect to their m/z values. Correct peptide assignment for the new peak was validated by MS/MS data as a TMPP-Ac- $_{124}\text{EVQPK}_{128}$. The TMPP method clearly permitted the identification of lower mass peptides (4–5 amino acids) which are rendered detectable as a result of the added mass tag. The absence of the b_8 ion in the MS/MS spectrum shown in Figure 4 and the low abundance of the b_4 ion in the MS/MS spectrum shown in Figure 5C can be attributed to the ‘proline effect’, a trait of MS experiments associated with proteolytic cleavage C-terminal to a P residue and resulting in lower peak intensity [23].

The TMPP labeling method permits the discernment of a peptide with a natural N-terminus from an equivalent peptide derived from protease cleavage during MS sample preparation. For example, the peptide (K) $_{89}\text{SEVTSNKDTIEHEASVK}_{105}$ was detected in tryptic digests of purified SasG (Figure 6A). Likewise, the peptide (K) $_{89}\text{SEVTSNKDTIE}_{99}$ which begins with the same N-terminus but different C-terminal end was also detected from Glu-C digests of SasG (Figure 6B). Therefore, we had tentative evidence for a naturally occurring peptide terminus at S_{89} . To gain further support for a naturally occurring peptide terminus at S_{89} , the TMPP labeling method was used followed by digestion with trypsin and Glu-C. The peptide TMPP-Ac- $_{89}\text{SEVTSNK}_{95}$ at m/z 668.96 (2+) was identified from tryptic digests of TMPP-

labeled SasG by LC-MS/MS (Figure 6C). The corresponding peptide from Glu-C digests was not detected. However, MS/MS data for the TMPP-derived peptide with an N-terminal residue at S₈₉ from tryptic digests strongly supported the notion that this residue was indeed a natural SasG N-terminus.

As shown in Table 3, additional N-terminal peptides indicative of post-translational protein processing steps were identified, and the characteristic 572 Da mass shift in TMPP-modified peptides was observed. Nearly all N-terminal peptide sites clustered in the SasG sequence range between the amino acid positions 63 to 202. Support for correct sequence assignments was obtained from MS/MS data of overlapping peptides derived from both Glu-C and trypsin digests. Analysis of TMPP-derivatized peptides in separate SDS-PAGE gel bands supported the notion that a decrease in the denatured M_r of SasG in SDS-PAGE was associated with a loss of amino acid residues at its N-terminus (Figure 1B). The tryptic peptide TMPP-Ac-₅₁AENNIENPTTLK₆₂ at *m/z* 1915.86 was only detected in the highest M_r band of three or four SDS-PAGE bands. The peptide at *m/z* 1172.51 (TMPP-Ac-₁₂₄EVQPK₁₂₈) was only observed in the lower M_r range. Two peptides, TMPP-Ac-₁₂₉SSVTHNAEAPK₁₃₉ and TMPP-Ac-₁₄₄RSVDEGSFDITR₁₅₅, were not detected in top band, but in the middle and bottom SDS-PAGE bands. These data are summarized in Table 4. In contrast, the C-terminal peptide ₁₀₇₈KKRAELPKT₁₀₈₆GG was identified by MS in all of the SasG bands.

Based on the strategy outlined above, characterization of TMPP label peptides was extended to Glu-C digests. A series of peptides featuring +572 Da TMPP mass shifts are listed in Table S1. The assignment of N-terminal peptides to gel band positions was similar to that of tryptic digests. For example, detection of the *m/z* 1391.57 peak corresponding to the peptide TMPP-Ac-₆₃DNVQSKE₆₉ was observed only in PMFs pertaining to the top gel band. Eight of ten SasG N-termini were verified by data derived from both trypsin and Glu-C digests. Identical nucleotide sequences of the *sasG* segment encoding the N-terminal protein region comparing strains HIP5827 and Mu3 verified that MS analysis based on the protein sequence in locus SAHV_2481 (strain Mu3) was appropriate. While the N-terminal region of the SasG 'A' domain was shown to be susceptible to proteolytic cleavage, it is not entirely certain that the truncation process occurs *in vivo* (prior to cell wall digestion with lysostaphin). However, the addition of several protease inhibitors such as PMSF and leupeptin during cell wall digestion suggest that SasG truncation is initiated at the intact cell surface of strain HIP5827.

The TMPP labeling method allows peptide detection and subsequence tandem MS analysis of peptides at picomol levels [22]. This method has been used for the selective isolation of α -amine reactive peptides and for *de novo* sequencing of such peptides from complex mixture after proteolytic digestion with Lys-N [12]. Furthermore, *de novo* sequencing of protein termini tagged with TMPP was shown to yield higher fidelity [14]. To our knowledge, this report is the first one applying this method to the characterization of extensive variability of N-terminal peptides derived from the same protein. There is currently no evidence for the biological significance of limited N-terminal truncation of SasG, e.g. in the context of intercellular aggregation, cell-to-cell adhesion and biofilm formation in *S. aureus*. However, Corrigan et al. [4] speculated that N-terminal truncation in recombinant SasG may be indicative of similar process in native SasG expressing strains. It was hypothesized that the A domain of SasG is proteolytically cleaved to allow cell-to-cell interactions to initiate autoaggregation. Our data do not support the notion that the entire SasG domain is degraded during cell autoaggregation and biofilm formation. Future studies will be aimed at understanding the intricate relationship between proteolytic processing of SasG and the unique morphological features of the strain HIP5827 and other SasG-expressing clinical *S. aureus* isolates.

4. Concluding Remarks

TMPP-Ac-OSu labeling and a multi-enzyme digestion strategy were combined to analyze post-translational processing steps and other modifications of the *S. aureus* surface protein SasG. First, MS data revealed extensive variation in the degree of N-terminal truncation of mature SasG. Secondly, the C-terminus of SasG was identified at the *in silico* predicted site, the LPXTG anchor motif. Third, four deamidation sites part of NG motifs were identified and confirmed by MS/MS analysis. Using multiple enzyme digestions we were able to expand the sequence coverage of SasG and validate PTMs via analysis of overlapping SasG peptides. SasG appears to have a proteolytically unstable N-terminal region encompassing roughly 150 amino acid residues beyond the A₅₁ signal peptidase cleavage site. This finding raises intriguing questions about the N-terminal processing mechanisms of this and other cell surface proteins that are anchored to the Gram-positive bacterial cell wall. N-terminal processing may modulate the proteins' functions as mediators of cell aggregation, cell adhesion, interactions with host cell surface proteins and host cell invasion. In future studies, we will examine the links between the PTMs observed for mature SasG and to its functional role at the bacterial cell surface.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was performed under the Pathogen Functional Genomics Resource Center (contract No. N01-AI15447), funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health. The authors thank Dr. Srilatha Kuntumalla for review of this manuscripts and Shih-Ting Huang and Tony Alami for technical supports.

Abbreviations used

PTMs	post-translational modifications
PMF	peptide mass fingerprinting
TMPP-Ac-OSu	N-succinimidylloxycarbonyl-methyl-tris(2,4,6-trimethoxyphenyl) phosphonium bromide
SasG	<i>Staphylococcus aureus</i> surface protein G
ESI-MS	electrospray ionization mass spectrometry
MALDI-TOF	matrix-assisted laser desorption ionization-time of flight
TFA	trifluoroacetic acid
2DE	two dimensional gel electrophoresis
CID	collision-induced dissociation
PBS	phosphate buffered saline
EDTA	ethylenediaminetetraacetic acid
BAM	benzamidinium-HCl
PMSF	phenylmethylsulfonyl fluoride

References

1. Nakai K. Protein sorting signals and prediction of subcellular localization. Adv. Protein Chem 2000;54:277–344. [PubMed: 10829231]

2. Roche FM, Meehan M, Foster TJ. The *Staphylococcus aureus* surface protein SasG and its homologues promote bacterial adherence to human desquamated nasal epithelial cells. *Microbiology* 2003;149:2759–2767. [PubMed: 14523109]
3. Resch A, Leicht S, Saric M, Pásztor L, Jakob A, Götz F, Nordheim A. Comparative proteome analysis of *Staphylococcus aureus* biofilm and planktonic cells and correlation with transcriptome profiling. *Proteomics* 2006;6:1867–1877. [PubMed: 16470655]
4. Corrigan RM, Rigby D, Handley P, Foster TJ. The role of *Staphylococcus aureus* surface protein SasG in adherence and biofilm formation. *Microbiol* 2007;153:2435–2446.
5. Conrady D, Brescia C, Horii K, Weiss A, Hassett D, Herr A. A zinc-dependent adhesion module is responsible for intercellular adhesion in staphylococcal biofilms. *PNAS. USA* 2008;105:19456–19461. [PubMed: 19047636]
6. Halliman DG, Ahearn DG. Relative Susceptibilities to Vancomycin and Quinupristin-Dalfopristin of Adhered and Planktonic Vancomycin-Resistant and Vancomycin-Susceptible Coagulase-Negative *Staphylococci*. *Curr. Microbiol* 2004;48:214–218. [PubMed: 15057468]
7. Kuroda M, Ito R, Tanaka Y, Yao M, Matoba K, Saito S, Tanaka I, Ohta T. *Staphylococcus aureus* surface protein SasG contributes to intercellular autoaggregation of *Staphylococcus aureus*. *Biochem. Biophys. Res. Commun* 2008;377:1102–1106.
8. Gatlin CL, Pieper R, Huang S-T, Mongodin E, Gebregeorgis E, Parmar PP, Clark DJ, Alami H, Papazisi L, Fleischmann RD, Gill SR, Peterson SN. Proteomic profiling of cell envelope-associated proteins from *Staphylococcus aureus*. *Proteomics* 2006;6:1530–1549. [PubMed: 16470658]
9. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved Prediction of Signal Peptides: SignalP 3.0. *J. Mol. Biol* 2004;340:783–795. [PubMed: 15223320]
10. Hiller K, Grote A, Scheer M, Munch R, Jahn D. PrediSi: prediction of signal peptides and their cleavage positions. *Nucl. Acids Res* 2004;32:W375–W379. [PubMed: 15215414]
11. Berg T, Leth-Larsen R, Holmskov U, Højrup P. Structural characterisation of human proteinosis surfactant protein A. *Biochim. Biophys. Acta* 2000;1543:159–173. [PubMed: 11087951]
12. Kuyama H, Sonomura K, Nishimura O, Tsunasawa S. A method for N-terminal de novo sequence analysis of proteins by matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Biochem* 2008;380:291–296. [PubMed: 18577371]
13. Kuyama H, Shima K, Sonomura K, Yamaguchi M, Ando E, Nishimura O, Tsunasawa S. A simple and highly successful C-terminal sequence analysis of proteins by mass spectrometry. *Proteomics* 2008;8:1539–1550. [PubMed: 18340634]
14. Gallien S, Perrodou E, Carapito C, Deshayes C, Reyat J-M, Van Dorsselaer A, Poch O, Schaeffer C, Lecompte O. Ortho-proteogenomics: Multiple proteomes investigation through orthology and a new MS-based protocol. *Genome Res* 2009;19:128–135. [PubMed: 18955433]
15. Mongodin E, Finan J, Climo MW, Rosato A, Gill S, Archer GL. Microarray Transcription Analysis of Clinical *Staphylococcus aureus* Isolates Resistant to Vancomycin. *J. Bacteriol* 2003;185:4638–4643. [PubMed: 12867477]
16. Tanner S, Shu H, Frank A, Wang L-C, Zandi E, Mumby M, Pevzner PA, Bafna V. InsPecT: Identification of Posttranslationally Modified Peptides from Tandem Mass Spectra. *Anal. Chem* 2005;77:4626–4639. [PubMed: 16013882]
17. Paterson GK, Mitchell TJ. The biology of Gram-positive sortase enzymes. *Trends Microbiol* 2004;12:89–95. [PubMed: 15036325]
18. Odintsov SG, Sabala I, Marcyjaniak M, Bochtler M. Latent LytM at 1.3 Å Resolution. *J. Mol. Biol* 2004;335:775–785. [PubMed: 14687573]
19. Marraffini LA, Schneewind O. Anchor Structure of Staphylococcal Surface Proteins. *J. Biol. Chem* 2005;280:16263–16271. [PubMed: 15718231]
20. Suh M-J, Alami H, Clark DJ, Parmar PP, Robinson JM, Huang S-T, Fleischmann RD, Peterson SN, Pieper R. Widespread Occurrence of Non-Enzymatic Deamidations of Asparagine Residues in *Yersinia pestis* Proteins Resulting from Alkaline pH Membrane Extraction Conditions. *Open Proteomics J* 2008;1:106–115.
21. Ton-That H, Liu G, Mazmanian SK, Faull KF, Schneewind O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *PNAS. USA* 1999;96:12424–12429. [PubMed: 10535938]

22. Huang Z-H, Wu J, Roth KDW, Yang Y, Gage DA, Watson JT. A Picomole-Scale Method for Charge Derivatization of Peptides for Sequence Analysis by Mass Spectrometry. *Anal. Chem* 1997;69:137–144. [PubMed: 8997893]
23. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandem mass spectrometry. *PNAS. USA* 1986;83:6233–6237. [PubMed: 3462691]

Figure 1.

Variants of SasG in 2D and SDS-PAGE gels. (A) Spot pattern in a high M_r region of a 2 D gel (M_r range ca. 125–160 kDa; pI range ca. 5.1–5.5). (B) Visualization of SasG bands by SDS-PAGE in the same M_r range. Coomassie Blue G250 was used to stain proteins. The three bands marked with arrowheads in (B) were excised for in-gel digestion and MS analysis. Lane 1: underivatized SasG; lane 2: TMPP-derivatized SasG; lane 3: M_r standards.

Figure 2.

MS/MS data for the N-terminal peptide $_{51}\text{AENNIENPTTLK}_{62}$. MS/MS data for this peptide was obtained from peptidoglycan-immobilized SasG digested in solution under denaturing conditions with Trypsin. (A) MALDI-MS/MS spectrum of a peak at m/z 1343.75; (B) LC-ESI-MS/MS spectrum of a peak at m/z 672.46 (2+).

Figure 3.

The C-terminal peptide of peptidoglycan-anchored SasG. (A) Schematic depicting the position of peptidoglycan/protein fusion peptide of SasG after lysostaphin cleavage. (B) MALDI-TOF spectrum of the m/z segment featuring the C-terminal peptides with peaks at m/z 1184.70 and 1241.72 from Glu-C digests. LC-nESI-MS/MS spectra of peptides corresponding to KKRAELPKTGG (C) and KKRAELPKTGGG (D), respectively.

Figure 4.

MS/MS data for the N-terminal peptide ₅₁AENNIENPTTLK₆₂ of SasG following derivatization with N-succinimidylloxycarbonyl-methyl)tris(2,4,6-trimethoxyphenyl) phosphonium bromide (TMPP-Ac-OSu). (A) MALDI-MS/MS spectrum of a peak at m/z 1915.83; (B) LC-nESI-MS/MS spectrum of a peak at m/z 959.5 (2+). TMPP adds a mass of 573.18 Da to the peptide.

Figure 5.

Comparison of spectra for the SasG peptide $_{124}$ EVQPK $_{128}$ from MALDI-TOF MS experiments with and without TMPP labeling. (A) m/z segment for trypsin-digested SasG; (B) the same m/z segment for trypsin-digested TMPP-derivatized SasG; the mass of the new peak at m/z 1172.52 corresponded to the predicted mass for TMPP-Ac- $_{124}$ EVQPK $_{128}$; (C) LC-nESI-MS/MS spectrum for the TMPP-Ac- $_{124}$ EVQPK $_{128}$ peptide.

Table 1

Sequence coverage of purified SasG by LC-nESI-MS/MS using three endopeptidases for protein digestion

	Sequence Coverage 1115 a.a.	Sequence Coverage [†] 1037 a.a.	Redundancy 1115 a.a.	Redundancy [†] 1037 a.a.
Asp-N	55% (616 a.a.)	59%	-	-
Glu-C	68% (754 a.a.)	73%	-	-
Trypsin	77% (861 a.a.)	83%	-	-
Asp-N+Glu-C	80% (887 a.a.)	86%	43% (481 a.a.)	46%
Asp-N+Trypsin	83% (928 a.a.)	89%	49% (549 a.a.)	53%
Glu-C+Trypsin	86% (956 a.a.)	92%	59% (656 a.a.)	63%
Overall	89% (992 a.a.)	96%	71% (789 a.a.)	76%

The sequence coverage for SasG is based on searches with the gene locus annotation SAHV_2481, which is the predicted *sasG* gene of *S.aureus* strain Mu3.

[†] calculated after loss of the signal peptide (50 amino acids) and truncation of SasG in the LPKTG motif at the C-terminus (28 amino acids).

Table 2
Sequence coverage of C-terminal peptides of peptidoglycan-anchored SasG by LC-nESI-MS/MS.

Sequence	Obs. (<i>m/z</i>), Charge	Cal. (<i>m/z</i>)	ΔM_r^a	Miss Cleavage	Ions Score	Expect ^b
LPKTGG	572.2, 1+	571.3	-0.12	0	31	4.3×10^{-5}
KKRAELPKTGG	395.7, 3+	1183.7	0.47	1	42	3.3×10^{-6}
LPKTGGG	315.4, 2+	628.3	0.44	0	17	8.4×10^{-3}
KKRAELPKTGGG	441.8, 3+	1240.7	0.85	1	35	1.2×10^{-4}

The sequence coverage for SasG is based on searches with the gene locus annotation SAHV_2481, which is the predicted *sasG* gene of *S.aureus* strain Mu3.

^a Difference between experimental and calculated monoisotopic masses

^b MASCOT expectation value

Table 3

List of N-terminal peptides detected in MALDI-MS data after TMPP derivatization of SasG. The TMPP-modified peptides were obtained following in-gel protein digestion with trypsin.

Amino acid position	Sequence	TMPP-derivatized Peptide		Free peptide	
		cal. (m/z)	obs. (m/z)	cal. (m/z)	obs. (m/z)
51–62	(A)AENNIENPTTLK(D)	1915.86	1915.86	1343.68	1343.68
–68	(A)AENNIENPTTLKDNVQSK(E)	2587.19	2587.17	2015.01	2015.00
63–68	(K)DNVQSK(E)	1262.52	1262.53	690.34	N.A.
–71	(K)DNVQSK(V)	1618.73	N.D.	1046.55	N.D.
74–78	(E)VTNK(D)	1162.49	1162.48	590.31	N.A.
–88	(E)VTNKDTAPQGVAK(S)	2158.98	N.D.	1586.80	N.D.
89–95	(K)SEVTSNK(D)	1336.56	1336.53	764.38	N.A.
–105	(K)SEVTSNKDTIEHEASVK(A)	2446.09	2446.05	1873.91	1873.91
93–95	(T)SNK(D)	920.37	920.38	348.19	N.A.
–105	(T)SNKDTIEHEASVK(A)	2029.90	N.D.	1457.72	N.D.
121–128	(A)NVAEVQPK(S)	1456.66	1456.63	884.45	884.49
–139	(A)NVAEVQPKSSVTHNAEAPK(V)	2578.21	N.D.	2006.03	N.D.
124–128	(A)EVQPK(S)	1172.52	1172.51	600.34	N.A.
–139	(A)EVQPKSSVTHNAEAPK(V)	2294.06	N.D.	1721.88	N.D.
129–139	(K)SSVTHNAEAPK(V)	1712.75	1712.74	1140.56	1140.57
–141	(K)SSVTHNAEAPKVR(K)	1967.91	N.D.	1395.73	N.D.
144–155	(A)RSVDEGSDITR(D)	1953.85	N.D.	1381.67	N.D.
–158	(A)RSVDEGSDITRDSK(N)	2284.01	N.D.	1711.83	N.D.
202–213	(N)ESNGLIGALQLK(N)	1814.89	1815.90	1242.71	N.D.
–215	(N)ESNGLIGALQLK(N)	2057.02	N.D.	1484.84	N.D.

Abbreviations: N.D., not detected; N.A., data not available due to exclusion from the detectable mass range. One missed cleavage site was allowed. Calibration of peptide mass spectra was done using a standard peptide calibration kit. Internal recalibration with well-resolved tryptic peptides of SasG that bracketed the m/z range of interest followed.

Table 4

Assignment of N-terminal peptides of SasG to distinct M_r variants, as observed in a SDS-PAGE gel (Figure 1B).

Position [*]	TMPP-Ac-SasG (in gel)		1D gel location [†]
	Glu-C	Trypsin	
51~	L	M+L	Top
63~	M+L	M+L	Top
74~	L	M+L	Top , Middle
89~		M+L	Top, Middle
93~	M+L	M+L	Top, Middle
121~	M	M+L	Middle , Bottom
124~	M+L	M+L	Middle , Bottom
129~	M+L	M+L	Middle, Bottom
144~	M+L	L	Middle, Bottom
202~		M+L	Bottom

* Amino acid sequence position according to SAHV_2481 sequence of SasG

[†] SDS-PAGE band position determined from in-gel digestion with trypsin of TMPP-derivatized SasG, as shown in Figure 1B. MALDI- TOF MS and LC-nESI-MS/MS data were obtained. Bold script indicates that peptides with the denoted amino acid start positions were particularly abundant and where in the SDS-PAGE band pattern they were abundant. M, identification by MALDI-TOF MS; L, identification by LC-nESI- MS/MS.