Comprehensive Characterization of Methicillin-Resistant *Staphylococcus aureus* subsp. *aureus* COL Secretome by Two-Dimensional Liquid Chromatography and Mass Spectrometry

Shobha Ravipaty and James, P. Reilly*

Department of Chemistry, Indiana University, 800E Kirkwood Avenue, Bloomington, IN 47405

*To whom correspondence should be addressed. E-mail: reilly@indiana.edu.

Running Tittle: 2D-LC and mass spectrometric characterization of *S. aureus* COL secretome

Abbreviations

Abbreviations	Defination
S. aureus	*Staphylococcus aureus*
MRSA	Methicillin resistant *Staphylococcus aureus*
SPase I	Type I signal peptidase
SPPase	Signal peptide peptidase
1DE	One-dimensional gel electrophoresis
2DE	Two-dimensional gel electrophoresis
ACN	Acetonitrile
FA	Formic acid
TCA	Trichloroacetic acid
BHI	Brain heart infusion
2D-LC	Two-dimensional liquid chromatography
SCX	Strong cation exchange
QTOF	Quadrupole time-of-flight mass spectrometer
TFA	Trifluoroacetic acid
H_2O	Water

Abstract

Two-dimensional liquid chromatography (2D-LC) combined with whole protein and peptide mass spectrometry are used to characterize proteins secreted by methicillin-resistant *S. aureus* COL. Protein identifications were accomplished via offline protein fractionation followed by digestion and subsequent peptide analysis by RPLC-ESI-LTQ-FT-MS/MS. Peptide MS/MS analysis identified 127 proteins comprising 59 secreted proteins, 7 cell wall anchored proteins, 4 lipoproteins, 4 membrane proteins and 53 cytoplasmic proteins. The identified secreted proteins included various virulence factors of known functions (cytotoxins, enterotoxins, proteases, lipolytic enzymes, peptidoglycan hydrolases etc.). Accurate whole protein mass measurement (± 1.5 Da) of the secreted proteins combined with peptide analysis enabled identification of signal peptide cleavage sites, and various post-translational modifications. In addition, new observations were possible using the present approach. Although signal peptide cleavage is highly specific, signal peptide processing can occur at more than one site. Surprisingly, cleaved signal peptides and their fragments can be observed in the extracellular medium. The prediction accuracies of several signal peptide prediction programs were also evaluated.

Introduction

Staphylococcus aureus (*S. aureus*), a gram positive human pathogen, is the leading cause of nosocomial infections, imposing tremendous economic burden on patients and hospitals throughout the world.[1-3] The spectrum of staphylococcal infections is very wide, ranging from minor skin lesions to life threatening conditions such as bacteremia, pneumonia, endocarditis, osteomyelitis, toxic shock syndrome and septicemia.[2-4] The treatment of staphylococcal infections has become extremely challenging due to its propensity to rapidly evolve antibiotic resistant strains. The methicillin resistant staphylococci (MRSA) is the most notorious, in that it

3

causes an estimated 94,000 life-threatening infections and 19,000 deaths a year in the United States.[4, 5] Hospitalization costs associated with MRSA infections are also significant with a mean attributable cost of $35,000 per infection.[6] Furthermore, the recent emergence of strains resistant to vancomycin, a glycopeptide antibiotic that is often considered as the last resort drug in treating MRSA infections, has compounded the problem.[7-9] Needless to say, it is of paramount importance to discover effective vaccines and to develop new strategies to treat *S. aureus* infections. This urgency motivated the scientific community to direct significant research effort towards whole genome sequencing of *S. aureus* strains in the past few years. The wealth of information available from the nine fully annotated and sequenced genomes of *S. aureus* has provided us with an excellent opportunity to apply powerful technologies including proteomics to gain a comprehensive understanding of the biology of this organism.[10-13]

Pathogenesis of *S. aureus* is complex and involves the synthesis of an array of virulence factors followed by their transport across the cytoplasmic membrane to destinations outside the cell. A majority of the exported proteins in *S. aureus* are predicted to be secreted via secretory (Sec) pathway, which requires an N-terminal signal peptide[14] at the N-terminus of the protein and a cleavage site that is recognized by type I signal peptidases (SPase I).[15] During translocation, or shortly thereafter, the signal peptide of the preprotein is removed by SPase I resulting in the release of the mature protein from the membrane.[16, 17] The mature protein may be further modified and is either retained on the cell surface or secreted into the extracellular host milieu. The secreted proteins of *S. aureus* are postulated to play a prominent role in host infection and are believed to be engaged in tissue damage, invasion and evasion of host immune responses. Therefore, a comprehensive description of secretory proteins (the secretome) of different *S. aureus* strains is vital to gain insights into its pathogenesis. This information will be valuable in

4

identifying novel virulence factors and should ultimately help in the development of new diagnostic tools and vaccines.

To this end, *S. aureus* secretory proteins have been identified using a variety of classical techniques including Western blot, ELISA, and one- and two-dimensional gel electrophoresis (1DE/2DE) with N-terminal sequencing.[18-20] Although gel-based techniques are well established for separating proteins mixtures, they have several drawbacks including poor reproducibility, and sensitivity and limited dynamic range and they are tedious, labor intensive and technically challenging.[21-23] Recent proteomic strategies have coupled gel electrophoresis with mass spectrometry.[24-27] In these approaches, the peptides resulting from in-gel digestion of excised spots were analyzed by MALDI/TOF-MS or by LC-ESI-MS/MS. A critical drawback of peptide analysis by mass spectrometry is that it provides very limited molecular information about intact secreted proteins. Valuable information regarding loss of signal peptides, signal peptide cleavage sites, post-translational modifications and protein degradation is usually completely lost. This is particularly true when peptide mass fingerprinting is employed. To overcome the shortcomings of the current techniques, a sophisticated gel-free approach combining whole protein and peptide mass spectrometric approaches was employed in the present work. To the best of our knowledge there is only one report in the literature pertaining to whole protein mass analysis of *S.aureus* secretory proteins. Kawano et al.[18] attempted to identify *S. aureus* secreted proteins via one-dimensional reverse phase liquid chromatography (1D-RPLC)-ESI-MS and N-terminal Edman degradation. Using this approach they were able to tentatively identify 3 and 4 secreted proteins in NCTC 8325 and MRSA 3543 strains, respectively.

In the present study the secretome of methicillin resistant *Staphylococcus aureus* subsp. *aureus* COL (*S. aureus* COL) was more comprehensively characterized. Secreted proteins were

separated with an in-house constructed automated 2D-LC system[28] with on-line fractionation followed by whole protein mass measurement by ESI-QTOF-MS. Definitive protein identifications were accomplished via offline collection of protein fractions followed by protein digestion and subsequent peptide analysis by RPLC-ESI-LTQ-FT-MS/MS. Genome-based signal peptide algorithms predict 71 secretory proteins for *S. aureus* COL.[29] Our peptide analysis successfully identified 59 of these secreted proteins from the culture supernatants of *S. aureus* COL with average sequence coverage of 79%. In addition, combined information from the two mass spectrometric approaches allowed detailed characterization of 53 of these secreted proteins. The accurate whole protein mass measurement of the secreted proteins allowed verification of signal peptide cleavage sites. Also, the current study provided us with an opportunity to compare the accuracy of several computational tools available for predicting signal peptide cleavage sites. Additionally, we report surprising findings on the presence of cleaved signal peptides and signal peptide fragments in the extracellular medium.

Materials and Methods

Materials. HPLC grade acetonitrile (ACN) and urea were purchased from EMD Chemicals (Gibbstown, NJ). Water was purified using a Barnstead/Thermolyne E-Pure system (Barnstead/Thermolyne, Dubuqe, IA). 40% aqueous methylamine was obtained from Aldrich (Aldrich Chemical, Milwaukee, WI). Sodium chloride, formic acid (FA), trifluoroacetic acid (TFA), and trichloroacetic acid (TCA) were procured from Mallinckrodt Baker, Inc. (Phillipsburg, NJ). Ammonium bicarbonate was obtained from Mallinckrodt Baker, Inc. (Paris, KY). Proteomics grade trypsin (T-6567) was purchased from Sigma Chemical Co. (St. Louis, MO) and endoproteinase Glu-C was obatained from New England Biolabs (Beverly, MA).

Bacterial Strain and Growth Conditions. The methicillin-resistant *S. aureus* COL was obtained through the NARSA program supported under NIAID/NIH Contract No. N01-AI-95359. To prepare stock cultures, a single colony of *S. aureus* COL was inoculated into 50 mL of sterile brain heart infusion broth (BHI; Difco/BD Biosciences, Sparks, MD) and incubated overnight at 37 ºC with rotary aeration (200 rpm). 1 mL of the overnight culture was used to inoculate 200 mL of fresh sterile BHI broth and incubated for 10 hours. The resulting culture was aliquoted into 2 mL stocks with 20% by volume glycerol as the cryoprotective agent and stored at -80 ºC until required. In a typical experiment, the stock culture was inoculated into 100 mL of sterile BHI broth and grown overnight at 37 ºC with rotary aeration (200 rpm). A 2 mL portion of this overnight preculture was then diluted 1:100 in fresh sterile BHI broth and incubated at 37 ºC with shaking at 200 rpm. The growth was monitored spectrophotometrically by measuring optical density at 600nm (OD_{600}). Bacteria were cultured to stationary growth phase ($OD_{600} = 4$) that was generally attained in about 8hrs.

Precipitation and Preparation of Extracellular Protein Fraction. Bacterial cells were separated from the stationary phase culture by centrifugation at 8500 x g (Sorvall RC-5B centrifuge, Du Pont, Wilmington, DE) for 30 minutes at 4 ºC. In order to remove residual bacteria, the supernatant was filtered using a Stericup®/Streritop™ filtration device with 0.22 µm pore-size PES membrane (Millipore Corp., Billerica, MA). Soluble proteins in the filtered supernatant were precipitated overnight with 10% w/v TCA at 4 ºC. The resulting precipitate was pelleted by centrifugation at 8500 x g for 70 minutes at 4 ºC, washed several times with ice cold acetone, and dried in a SpeedVac vacuum centrifuge (Jouan/Thermo Electron). The protein extract was dissolved in an appropriate amount of 8 M urea, 2 M thiourea solution and centrifuged at 14100 x g (minispin plus, Eppendorf, Westbury, NY) for 2 minutes to remove

insoluble materials, and was immediately used for proteomic analysis. Protein concentration was determined by Bradford assay[30] using bovine serum albumin as a standard.

Live-dead Staining. 500 µL of *S. aureus* COL stationary phase culture was centrifuged at 16000 x g for two minutes and the resulting cell pellet was stained with LIVE/DEAD® BacLight™ bacterial viability kit (Molecular Probes, Carlsbad, CA) following the protocol provided by the manufacturer, and visualized by fluorescence microscopy to detect cell lysis.

Isolation and Preparation of Extracellular Peptide Fraction. To isolate cleaved signal peptides present in the extracellular medium, *S. aureus* COL was cultured to stationary phase as described above. After removal of bacterial cells by centrifugation, the supernatant was filtered through Microcon Ultracel YM-10 to remove all the high molecular weight proteins. (Millipore Corp., Billerica, MA). The resulting filtrate was lyophilized, reconstituted in water, and the peptides were isolated, concentrated and desalted using PepClean C18 spin columns (Pierce, Milwaukee, WI). The eluted peptides were dried in a SpeedVac vacuum centrifuge, resuspended in 40 µL of 0.1% TFA in H_2O and analyzed directly without any enzyme pretreatment by nano-RPLC-nanoESI-LTQ-FT-MS/MS as described below.

Protein Sequence Data. *S. aureus* COL proteome was obtained from The J. Craig Venter Institute (www.jcvi.org, Genbank accession number CP000046.1).[31] Theoretical protein masses were calculated from protein sequences using Protein Analysis Worksheet Software program (PAWS: freeware edition for Windows 95/98/NT/2000, ProteoMetrics, LLC, New York, NY; available at http://bioinformatics.genomicsolutions.com/PawsDL.html). Additional sequence information was obtained from the Swiss-Prot database at ExPASY (www.expasy.org).[32]

Whole Protein Two-Dimensional Liquid Chromatography and Mass Spectrometry.

S. aureus COL secreted proteins were separated using an automated two-dimensional liquid chromatography (2D-LC) system that has been previously described.[28] In a typical analysis, *S. aureus* COL extracellular protein extract was first separated by strong cation exchange chromatography and fractionated on-line using 20 trapping columns. Contents of trap were then separated by reverse phase chromatography followed by measurement of protein masses by quadrupole time-of-flight mass spectrometer. More detailed description of the method is provided in supplemental data. Mobile phases along with the gradients used are shown in Supplemental Table 1 and Supplemental Table 2.

Peptide Nano -Liquid Chromatography - Tandem Mass Spectrometry and Data Analysis.
A detailed description of peptide analysis and information on MS/MS database search parameters are provided in supplemental data. Briefly, proteins trapped on each trapping column were eluted with organic mobile phase and collected off-line for trypsin or endoprotease Glu-C digestion. Each trap digest was subsequently analysed by nano-RPLC-nanoESI-LTQ-FT-MS/MS.

Signal Peptide and Protein Localization Predictions. To predict the presence of signal peptides and signal peptidase I cleavage sites for the proteins identified in the present work, the following prediction tools were used: SignalP-NN (neural network model) and SignalP-HMM (hidden markov model) version 2.0[33] (http://www.cbs.dtu.dk / services/SignalP-2.0/) and version 3.0[34] (http://www.cbs.dtu.dk/services/SignalP/), PrediSi[35] (position weight matrix method; http://www.predisi.de/) and SigCleave[36, 37] (weight matrix method; http://sbcr.bii.a-star.edu.sg/cgi-bin/emboss/menu/sigcleave). When required, LipoP, version 1.0[38]

(http://www.cbs.dtu.dk/services/LipoP/), and TMHMM program, version 2.0[39]

(http://www.cbs.dtu.dk/services/TMHMM/) were used to predict lipoproteins and membrane

proteins, respectively. PSORTb version 2.0[40] was used to predict protein subcellular localization.

Results

Protein Identification Strategy. Sibbald *et al.* used a rigorous approach employing a

combination of computational tools and an optimized type I signal peptidase (SpsB) recognition

search pattern to estimate that 71 extracellular proteins are produced by *S. aureus* COL.[29]

In order to identify proteins present in *S. aureus* COL extracellular medium, proteins were

extracted by TCA precipitation from stationary growth phase cultures, a phase during which

extracellular proteins are preferentially expressed.[27, 41] The first step in the analysis was to

identify the *S. aureus* COL secretome, through peptide analysis. The proteins from each of the

C4 trapping columns were digested and analyzed as discussed above. From these peptide data a

total of 127 proteins (Supplemental Table 3) were identified in the extracellular medium, and

using bioinformatics tools we classified these proteins into five categories based on their

predicted subcellular localization. We classified 59 of the identified proteins as secreted proteins

(Table 1) and all of these proteins except two contained potential Sec-type signal peptides with

Spase I cleavage sites, and lacked any cell wall or membrane retention signals. The remaining 68

proteins included seven cell wall anchored proteins, four lipoproteins, four membrane proteins

(Supplemental Table 4) and 53 cytoplasmic proteins (Supplemental Table 5); these proteins were

not predicted to be secreted into the growth medium.

The second step in the analysis of extracellular proteins involved mass measurements of

whole proteins captured on each trap by ESI-QTOF-MS. Whole protein masses from a particular

trap were assigned to molecules that had been identified by peptide analysis of the same trap.

The general experimental approach employed to identify *S. aureus* COL extracellular proteins is demonstrated below by using Trap 7 as an example. Peptides corresponding to thermonuclease (Nuc) (Figure 1A), beta-hemolysin (Hlb) SACOL0755, SACOL0859, serine protease SplA (SplA), serine protease SplE (SplE), penicillin binding protein 2 (MecA) and ribosomal protein L9 (RplI) were heavily populated in Trap 7. Since Nuc, Hlb, SACOL0755, SACOL0859, SplA , and SplE are predicted to undergo post-translational cleavage of the signal peptide, predicted masses for the mature secreted proteins were calculated by obtaining predicted cleavage site information from 6 signal peptide prediction programs; the resulting list of predicted masses was then matched against the experimental masses derived from whole protein MS analysis of Trap 7 proteins. For example, signal peptide cleavage site prediction of Nuc (theoretical mass of 259111.9 Da) yielded 5 different predictions (cleavage at residue 23, 25, 30, 57 and 60) resulting in 5 possible predicted masses for the mature Nuc (Figure 1B). It is important to point out here that the computational methods used to predict signal peptide cleavage sites frequently provide conflicting predictions as exemplified by this case. Figure 1C displays the total ion chromatogram of Trap 7 fraction. Deconvolution of the raw spectrum (Figure 1D) yields protein peak at retention time of 47 min with a mass of 18782.0 Da (Figure 1E). This observed mass matches very closely with only one of the possible predicted masses of Nuc (18782.3 Da) and confirms the signal peptide cleavage site position as Ala[60]. The cleavage site position of Nuc was further corroborated by identification of the N-terminal peptide of the mature protein by MS/MS analysis as shown in Figure 1F. Whole protein identification and confirmation of signal peptide cleavage site positions of the other Trap 7 secreted proteins was accomplished similarly. Of the six programs used in the present study, only two (SignalP V3-NN, Predisi) yielded the correct cleavage site for Nuc. Capitalizing on the ability to accurately determine the whole protein mass

and hence the cleavage site position of the secreted proteins, we took this opportunity to evaluate the prediction accuracy of different signal peptide prediction programs, and the results are discussed below.

Whole protein identification of membrane proteins and cytoplasmic proteins was achieved by directly matching the observed masses with theoretical masses calculated from the protein sequences. In the case of cell wall anchored proteins and lipoproteins that contain cleavable signal peptides the strategy used for secreted proteins was applied. For any protein that could not be identified by matching theoretical or predicted masses (after loss of signal peptide) to the observed masses, additional modifications including methylation (+ 14.03 Da) acetylation (+42.04 Da), oxidation (+15.99 Da), formylation (+27.99 Da) and protein truncation were considered. The PAWS program was used to assign unmatched protein masses derived from whole protein MS analysis to the corresponding truncated secreted proteins. In this approach, an unassigned protein mass from a particular trap was searched against the entire sequence of the suspected protein to determine if any part of the sequence has a mass that matches the input mass within the experimental error (±1.5 Da).

Post-Translational Modifications of Secreted Proteins. Whole protein mass measurements of 20 trap fractions successfully confirmed 53 of the 59 secreted proteins that had been identified by peptide analysis (Table 1). Of these, 39 proteins were identified directly by matching the observed masses to the predicted masses calculated by removing a signal peptide. Remaining proteins were identified by considering additional post-translational modifications. Six proteins could not be identified by whole protein analysis probably due to their low abundance or extensive degradation during culture or sample preparation by secreted proteases.[20, 26, 42] Only

12

notable secreted proteins exhibiting modifications other than routine signal peptide loss will next be discussed in detail.

Cytotoxins.   Delta-toxin is a 45 residue (Hld-45; 5009.1 Da) protein that does not contain a classical N-terminal signal peptide. However, the mature form is 26 residues (2978.5 Da) in length indicating that the first 19 residues constitute the signal peptide.[43] Since Hld was identified predominantly as a N-terminal methionine formylated species, some studies have suggested that the translational start codon has been misassigned such that the native form of Hld is only 26 (Hld-26) residues long and is secreted without a signal peptide.[44, 45] In the present study two forms of Hld (Figure 2) were identified by peptide MS/MS analysis of extracellular peptide fraction (see methods section); Hld[1] with a monoisotopic mass of 3020.6 Da corresponds to Hld-26 with oxidized N-terminal formylated methionine, and Hld[2] (monoisotopic mass of 2976.6 Da) corresponds to Hld-26 with unformylated N-terminal methionine. It is not totally clear if Hld[2] is formed as a result of post-translational deformylation of N-terminal formylated Hld-26 or if it results from Hld-45 by signal peptide cleavage. The ambiguity arises from the fact that deformylation by peptide deformylase is usually followed by removal of N-terminal methionine by methionine aminopeptidase if the succeeding residue has a small side chain; although Ala is the second residue in Hld-26, we did not find any evidence of methionine removal.

The theoretical mass of leukotoxin LukE (LukE) is 34819.1 Da. Signal peptide cleavage at Ala[27] was predicted by all of the programs leading to an expected mass of 31864.5 Da. Several peptides corresponding to LukE (sequence coverage, 68%) including the N-terminal peptide of the mature protein (NTNIENIGDGAEVIKR, residues 28-43) were identified in the Trap 8 tryptic digest. Whole protein mass spectra from Trap 8 did not contain any peak that matched the

predicted mass. However, as seen in Figure 3, a mass of 31751.5 Da that is 114 Da (± 1.5 Da) lower than the predicted mass of LukE was observed in Trap 8. This was identified as LukE with a single C-terminal residue (Asn) truncation. It is most likely that the truncated form resulted from C-terminal degradation by extracellular proteases.

Gamma-hemolysin, component B (HlgB) is a 325 residue protein with a theoretical mass of 36711.0 Da. The predicted signal peptide cleavage site is Ala^{26} yielding an expected mass of 34048.7 Da. Peptides corresponding to HlgB were observed in Trap 4 (sequence coverage, 69%), however, whole protein MS analysis of the same trap did not show a mass that matched with the expected mass of HlgB. Instead, three co-eluting protein masses were observed that differed in mass by less than150 Da suggesting modified forms of the same protein. The PAWS program revealed that these three protein masses matched well with the three forms of mature HlgB ($HlgB^1$, $HlgB^2$ and $HlgB^3$). As shown in Figure 4 the observed mass of 33392.6 Da corresponds to Pro^{32}- Asn^{324} ($HlgB^1$), the observed mass of 33463.6 Da corresponds to Glu^{27}- Glu^{320} ($HlgB^2$), and the observed mass of 33507.4 Da corresponds to Lys^{29}- Glu^{322} ($HlgB^3$). The most plausible explanation for the presence of three HlgB forms is truncation of both N-terminal and C-terminal residues due to proteolytic degradation as suggested above for LukE.

Aerolysin/leukocidin family protein (SACOL2006) has a theoretical mass of 40434.0 Da and signal peptide cleavage at Ser^{29} and Ala^{27} were predicted for the protein. Peptide analysis of Trap 10 digest identified SACOL2006 with a sequence coverage of 73%. Interestingly, Trap 10 whole protein MS analysis identified two co-eluting forms of the protein, $SACOL2006^1$ (37419.3 Da) and $SACOL2006^2$ (37620.8 Da) that appear to have formed as a result of signal peptide processing at both Ser^{29} and Ala^{27}, respectively (Figure 5). This observation of signal peptide processing at two different sites is quite unusual, since signal peptidases generally cleave signal

peptides with high fidelity.[46] This observation along with another example will be discussed below.

Superantigenic Toxins. Staphylococcal enterotoxin (Sek) has a theoretical mass of 27721.1 Da. Signal peptide cleavage at Ala^{23} was predicted by all the programs leading to an expected mass of 25333.1 Da. Peptide analysis of Trap 12 digest identified Staphylococcal enterotoxin (Sek) with a sequence coverage of 82%. Trap 12 whole protein MS analysis identified Sek as a C-terminal truncated protein since the observed mass of 24699.6 Da coincided with the mass of the mature Sek protein formed by signal peptide cleavage at position Ala^{23} and removal of residues YKTI from the C-terminus (24698.4 Da). The signal peptide cleavage at Ala^{23} was further confirmed by observation of the N-terminal peptide QGDIGIDNLR (Res 24-33). The presence of truncated Sek is most probably due to the degradation of the protein.

The theoretical mass of Staphylococcal enterotoxin type I (Sei) is 28184.6 Da. Signal peptide cleavage at Ala^{26} was predicted by all the programs leading to an expected mass of 25076.9 Da. Sei was identified from Trap 20 digest with a sequence coverage of 77%. Trap 20 whole protein MS analysis identified Sei as a C-terminal truncated protein with an observed mass of 24848.0 Da corresponding to signal peptide cleavage at position Ala^{26} and removal of C-terminal residues TE. The signal peptide cleavage at Ala^{26} was further confirmed by observation of N-terminal peptide DVGVINLRNFYANYEPE (Res 27-43). C-terminal truncated Sei may be formed by proteolytic degradation.

Proteases. Serine proteases SplA, SplB, SplC, SplE, SplF and cysteine protease precursor SspB were identified in *S. aureus* COL extracellular medium by peptide MS/MS as well as whole protein MS analysis and signal peptide cleavage was the only modification observed. Serine protease SplD was somewhat more interesting. Its theoretical mass is 25669.1 Da. Signal peptide

cleavage at Ala[26] was predicted by the prediction programs yielding an expected mass of 22001.8 Da. Peptide analysis of Trap 5 protein digest identified SplD with a high sequence coverage of 77%. Nevertheless Trap 5 did not yield a whole protein mass that matched the predicted mass. Instead a recurring mass peak at 22012.2 Da that is 10.4 Da higher than the predicted mass was observed co-eluting with SplF (Figure 6a). Since SplD shares 96% sequence identity with SplF,[47] we expected it to co-elute with SplF. The observed 10.4 Da mass difference was suspected to be a sequencing error. Examination of single nucleotide substitutions that could account for the observed mass discrepancy yielded two potential candidates for a sequencing error; Ser $\rightarrow$ Pro with a mass difference of 10 Da and Gln$\rightarrow$His with a mass difference of 9 Da. Phylogenetic analysis combined with mass spectrometry has been previously used in our laboratory to identify a large number of sequencing errors in *B. subtilis* strain 168.[48] A similar approach was utilized in the present study. Multiple amino acid sequence alignment of SplD from different *S. aureus* strains using ClustalW program indicated high sequence homology between *S. aureus* COL and other strains (99.6% sequence identity) and revealed Gln[68]$\rightarrow$His[68] as the plausible sequencing error (Figure 6b). This was in fact confirmed by MS/MS analysis of the peptide (residue 42-74) LITNTNVAPYSGVTWMGAGTGFVVGNHTIITNK (Figure 6c). A comparison of SplD nucleotide sequences from various *S. aureus* strains revealed that CAT codon at position 68 is highly conserved in all strains suggesting that the observed discrepancy is not a single nucleotide polymorphism. Only Spld of *S. aureus* COL has CAA codon at position 68 in the reference genome and we believe that A$\rightarrow$T nucleotide sequencing error at this position has resulted in the observed Gln[68]$\rightarrow$His[68] sequencing error.

Lipolytic Enzymes. Rollof et al. have reported that lipase (76 KDa) in *S. aureus* strain TEN 5 is secreted into the culture medium as a pro-lipase (82 KDa) after cleavage of the signal peptide;

subsequent processing (removal of propeptide) of the pro-lipase resulted in a mature lipase (44

Kda to 45KDa).[49] It is important to point out that the molecular mass of the observed pro-lipase

is significantly higher than the predicted mass (73Kda); the reason for the observed mass

difference, however, was not explained. In the present study, Lipase1 (Lip1) was identified by

peptide analysis of Trap 8 digest (96% sequence coverage), and Lipase2 (Lip2) was identified

from Trap 9 digest with a sequence coverage of 71%. Lip1 is a 680 residue protein (76675.3 Da)

and is predicted to contain a signal peptide domain (1-34 residues) followed by propeptide

domain (35-290 residues) and a mature lipase domain consisting of 390 residues (44345.3 Da).

Similarly, Lip2 (71276.8 Da) is predicted to contain a signal peptide (37 amino acids), a

propeptide (258 amino acids) and a mature lipase with a predicted mass of 44071.6 Da. The

observed mass for Lip1 (73078.0 Da) obtained from Trap 8, and the observed mass of Lip2

(67152.8 Da) obtained from Trap 9 matched with the predicted masses corresponding to the

signal peptide cleavage at positions 34 (73077.1 Da) and 37 (67152.0 Da), respectively, and

revealed that after 8 hrs of growth both Lip1 and Lip2 were present in the extracellular medium

as unprocessed pro-enzymes (Figure 7). We did not find any evidence of mature lipase forms in

the extracellular medium. In contrast to previous studies based on SDS-PAGE and Western blot

analysis where the observed masses of staphylococcal pro-lipases were inexplicably higher than

predicted masses,[49-51] our study provided accurate mass determination of pro-lipases allowing

confident identification of the proteins.

Peptidoglycan Hydrolases. Bifunctional autolysin (Atl; 138 KDa) is a bacteriolytic enzyme

capable of causing cell lysis. It consists of two functionally distinct domains. Several studies

have reported that Atl undergoes proteolytic processing to generate 62 KDa (amidase) and 51

KDa (glucosaminidase) extracellular lytic enzymes.[52, 53] In *S. aureus* COL Atl is a 1256 amino

acid protein (137334.9 Da) and is predicted to contain a signal peptide domain (1-29 residues) a propeptide domain (30-198 residues), an N-acetylmuramoyl-L-alanine amidase domain (AM;199-775 residues), and an Endo-β-N-acetylglucosaminidase domain (GL; 776-1256 residues). The predicted masses of mature AM and GL are 63008.7 Da and 53479.5 Da, respectively. Atl was identified in Traps 8 and 9 and peptide analysis yielded a high sequence coverage of 99%. Whole protein MS analysis of the same traps resulted in the identification of three gene products of Atl (Figure 8). The observed mass of 134249.5 Da corresponding to pro-Atl form (Ala^{30}-Lys^{1256}, Atl^1) matched closely with the predicted mass (134248.2 Da) confirming the signal peptide cleavage at Ala^{29}. Other two gene products, Atl^2 and Atl^3 appear to have formed as a result of proteolytic processing of Atl^1. An intermediate form of AM with a mass of 80787.3 Da (Atl^2) corresponding to sequence Ala^{30}-Lys^{775} was identified. The observed mass of 53479.7 Da corresponding to the mature GL (Ala^{776}-Lys^{1256}, Atl^3) confirmed proteolytic processing at Lys^{775} - Ala^{776}. The mature AM form, however, was not detected in the extracellular medium of *S. aureus* COL. This suggests that either the intermediate AM form requires more time (>8hrs) to undergo proteolytic processing or the enzyme responsible for its processing is not present.

Lysm domain protein (SACOL0723) has a theoretical mass of 28186.8 Da (residue 1-265). Signal peptide cleavage at Ala^{23} and Ala^{25} was predicted by the programs yielding expected masses of 25831.0 Da and 25631.8 Da, respectively. Peptide analysis of Trap 2 digest identified SACOL0723 with a high sequence coverage of 80%. Whole protein MS analysis of Trap 2, however, did not uncover any mass that matched either of the expected masses. Instead, we observed two masses that suggested degradation of SACOL0723 (cleavage between Gly^{138} - Gly^{139}) into two fragments, SACOL0723^1 and SACOL0723^2. The observed mass of

SACOL0723[1] (11849.0 Da) corresponds to N-terminal protein fragment (Ser^{26}- Gly^{138}) and the observed mass of SACOL0723[2] (13801.0 Da) corresponds to C-terminal protein fragment (Gly^{139}- His^{265}). The cleavage between Gly^{138} - Gly^{139} was further corroborated by identification of C-terminal non-tryptic peptide GYLIMPNQTLQIPNGGSG (residue 121-138) by peptide MS/MS analysis.

The theoretical mass of N-acetylmuramoyl-L-alanine amidase domain protein (SACOL2666) is 69253.2 Da (609 amino acid). Signal peptide cleavage at Ala^{27} was predicted by all the programs leading to the expected mass of 66309.7 Da. Peptide analysis of Trap 6 digest identified SACOL2666 with a high sequence coverage of 89%. Whole protein MS analysis of Trap 6 did not provide any mass that was close to the expected mass of SACOL2666. Instead masses 15296.4 Da, 42252.1 Da, and 7810.9 Da corresponding to fragments Thr^{29}-Thr^{163}, Asp^{172}-Asp^{546} and Tyr^{547}-Lys^{619} respectively were observed indicating degradation of the protein.

Miscellaneous Enzymes. For SACOL1071, signal peptide processing was observed at multiple sites (Figure 9a). The theoretical mass of SACOL1071 is 11344.8 Da, and signal peptide cleavage at Ala^{24} and Ala^{26} was predicted leading to expected masses of 8907.1 Da and 8721.0 Da, respectively. Peptide analysis of Trap 3 digest identified SACOL1071 with a high sequence coverage of 91%. Trap 3 whole protein MS analysis identified two co-eluting forms of the protein that represented signal peptide processing at Ala^{26} as well as Ala^{24}: SACOL1071[1] with an observed mass of 8720.1 Da (Thr^{27}- Lys^{105}) and SACOL1071[2] with an observed mass of 8907.1 Da (Asp^{25}-Lys^{105}). A comparison of the protein peak intensities suggested that SACOL1071[1] is the major form present in the extracellular medium. Furthermore, a whole protein mass of 9901.6 Da (SACOL1071[3]) was also observed co-eluting with SACOL1071[1]. This matched the predicted

mass of SACOL1071 (9902.0 Da) after removal of N-terminal residues 1-14. This observation

implicated another signal peptide processing site for this protein. Identification of the N-terminal

non-tryptic peptide, ATLVTPNLNADATTNTTPQIK (residue 15-35) that represents the N-

terminal peptide of the mature protein via peptide MS/MS analysis, confirmed signal peptide

processing at position 14 (Figure 9b). The general attributes of the14 residue signal sequence of

SACOL1071do not conform to those typical for a Sec-type signal peptide. The implications of

observing multiple cleavage sites in SACOL1071 will be discussed later.

Surface Adhesins . The theoretical mass of secretory extracellular matrix and plasma binding

protein (Empbp) is 38484.9 Da (residue 1-340). Signal peptide cleavage is predicted at Ala[24] and

Ala[26] yielding expected masses of 35781.6 Da and 35582.4 Da, respectively. Peptide analysis of

Trap 15 digest identified Empbp with a sequence coverage of 65%. Whole protein MS analysis

of trap 15 did not show any protein mass close to the predicted values of mature Empbp. Instead,

we observed two masses that suggested degradation of Empbp into two fragments. An observed

mass of 12891.8 Da matched that of an N-terminal protein fragment encompassing residues

Ser[27]-Thr[143] (Empbp[1]), and a peak at 22708.9 Da matched the mass of a C-terminal protein

fragment encompassing residues Gln[144]-Val[340] (Empbp[2]). The sum of the two fragments

(35600.7 Da) matched closely with the expected mass (35582.4 Da +18 Da) corresponding to

signal peptide cleavage at Ala[26].

SdrH protein (SdrH) is a 419 residue protein (46630.4 Da) with a predicted mass of

43094.9 Da after signal peptide cleavage at position Ala[32]. SdrH was identified with high

sequence coverage (81%) in the digest of Trap 3, but whole protein MS analysis of the same trap

did not show any mass that was close to the predicted mass. Instead, we detected a C-terminally

truncated protein with an observed mass of 38085.1 Da (Lys[33]- Lys[376]) formed by the removal

of residues 377-419. The signal peptide cleavage at position Ala^{32} was further confirmed via peptide MS/MS analysis by identification of N-terminal peptide KDNLNGEKPTT NLNH NITSPSVNSEMNNNETGTPHESNQTGNEGTGSNSR (Residue 33-82). Peptide analysis also indicated that the C-terminal peptides corresponding to the last 44 residues were missing.

Proteins with Unknown Functions. Staphylococcal secretory antigen ssaA (SsaA2) has a theoretical mass of 29327.1 Da (267 residues) and a predicted mass of 26715.1 Da after signal peptide cleavage at Ala^{27}. Peptides corresponding to SsaA2 were found in Trap 3. However, whole protein MS analysis of the same trap did not provide a mass that corresponded to full length mature protein. Instead SsaA2 was identified as two protein fragments; an N-terminal protein fragment with an observed mass of 12210.3 Da (Ser^{28}-Gly^{127}, SsaA2^{1}) and a C-terminal fragment with an observed mass of 14294.9 Da (Ala^{131}-His^{267}, SsaA2^{2}). Furthermore, observation of the peptide ASYSTSSNNVQVTTTMAPSSNGR (residue 131-153) that resulted from non-tryptic cleavage at the N-terminus of the peptide confirmed degradation of the protein into two fragments.

Staphyloxanthin biosynthesis protein (SACOL2295) has a theoretical mass of 17424.8 Da. Signal peptide cleavage is predicted at Ala^{22} and Ala^{27} yielding expected masses of 15297.2 Da and 14739.7 Da, respectively. Peptide analysis of Trap 3 digest identified this protein with a high sequence coverage of 71%. The observed mass (14756.7 Da) of SACOL2295 obtained from Trap 3 whole protein MS analysis was 17 Da higher than the predicted mass (14739.7 Da). This led us to propose two modifications for SACOL2295: loss of signal peptide and oxidation of the mature protein. Thiol groups of cysteine residues are known to be sensitive toward oxidation; Wolf et al. have already reported the oxidation of Cys^{69} residue in SACOL2295.[54] In this study,

it is highly possible that Cys[69] is the site of the proposed modification, however, we could not confirm the modification site by peptide MS/MS experiments as the residue was not mapped.

Virulence factor EsxA (EsxA) has a theoretical mass of 11036.2 Da and no Sec-type signal sequence is predicted at its N-terminus. In the present study, peptide analysis identified EsxA in Trap 3 digest with high sequence coverage (99%) and there was no evidence of loss of signal peptide from the protein. Burts et al. identified EsxA along with EsxB in *S. aureus* strain Newman and have shown that they are exported via an ESAT-6 secretion pathway (type VII pathway).[55] Since EsxA of *S. aureus* COL shares 100% sequence identity with that of *S. aureus* Newman, we expect that it is similarly exported. Whole protein MS analysis of trap 3, did not uncover any mass that was close to the theoretical mass (11036.2 Da). Instead, we observed a very intense peak at 10905.2 Da that was 131 Da lower than the theoretical mass suggesting the removal of N-terminal methionine by methionine aminopeptidase. Identification of the N-terminal methionine truncated peptide AMIKMSPEEIRAKSQSYGQGSDQIRQILS DLT RAQGE (residue 2- 38) corroborated this hypothesis. The whole protein MS analysis in combination with peptide analysis definitively confirmed the absence of signal peptide processing in EsxA. In mycobacterium tuberculosis, proteins EsxA and EsxB form a tight 1:1 dimer[56] that is required for stability of the proteins and this interaction is thought to take place in the cytosol prior to protein export. Burts et al.[55] reported that in *S. aureus* Newman EsxB is required for the synthesis and secretion of EsxA and vice versa. This led them to believe that EsxA and EsxB also form a heterodimer in *S. aureus*. However, in the present study EsxB was not identified despite the fact that EsxA yielded rather intense signal, which suggests that EsxB may not be required for secretion of EsxA in *S. aureus* COL. Sundaramoorthy et al. also did not

observe heterodimer formation following incubation of *S. aureus* EsxA and EsxB proteins; instead EsxA crystallized as a homodimer.[57]

The predicted mass of SACOL0270 is 30421.1 Da following signal peptide cleavage at Ala[24]. Peptides corresponding to this protein were found in Trap 12, however, whole protein MS analysis of the same trap did not show any mass that matched the predicted value within the experimental error. Instead an unmatched mass of 30377.9 Da was observed, which we believe is SACOL0270. The observed mass discrepancy of -42 Da is most probably due to arginine modification. Hydrolysis of arginine to form ornithine is a well known modification that results in a mass shift of -42 Da.

Post-Translational Modifications of Non-Secretory Proteins. Similar to secreted proteins, post-translational modifications of cell wall anchored proteins, membrane proteins, lipoproteins and cytoplasmic proteins were characterized and a detailed discussion on the observed modifications is provided in the supplemental data.

Stable Cleaved Signal Peptides and Signal Peptide Fragments. Several reports have suggested that after cleavage of a signal peptide from a pre-protein, rapid removal and degradation of the signal peptide is important for proper functioning of the export machinery.[58, 59] Nevertheless, peptide analysis of some trap fractions indicated the presence of stable cleaved signal peptides and signal peptide fragments derived from a few secreted proteins. Since TCA does not precipitate peptides efficiently[60], we suspected that there may be more peptides in the extracellular medium than those identified in the TCA protein extract. Using the procedure outlined in the methods section we attempted to isolate the peptides present in the *S. aureus* COL stationary phase culture. Indeed, RPLC-ESI-LTQ-FT-MS/MS of the peptide extract revealed the

presence of several peptides including stable cleaved signal peptides of five proteins Sle1, SACOL0723, SceD, IsaA and SACOL2295(Supplemental Table 6) and signal peptide fragments of 18 secreted proteins (Table 2). As an example, MS/MS spectra of cleaved signal peptide of IsaA and signal peptide fragment of SACOL1164 identified in the present study are shown in Figure 10. It is noteworthy that all of the observed signal peptide fragments are from C-terminal portions of respective signal peptides containing SPaseI cleaved sites, and appear to have formed by cleavage in the hydrophobic region of the signal peptide. To account for the signal peptide fragments observed in the present study, we have proposed cleavage sites in the SPase I processed signal peptides as shown in Table 2. The implications of these observations will be discussed later.

Signal Peptide Prediction Accuracy. The prediction accuracy of computational programs commonly used to predict signal peptides and cleavage site position has been debated. Few studies have evaluated the performance of signal peptide prediction programs using experimentally verified signal peptide data from different organisms.[61-63] To the best of our knowledge, the suitability of the commonly used prediction programs to predict secretory proteins and signal peptide cleavage sites in *S. aureus* has not been reported. Results from the present study are shown in Figure 11. It is evident from the figure that SignalP 3.0-HMM (92%), SignalP 2.0-HMM (90%) and SignalP 2.0-NN (87%) are superior in predicting the correct cleavage sites. In contrast, the prediction accuracies of SignalP 2.0-NN (75%), Predisi (74%), and Sigcleave (70%) are substantially lower. Although we observed only a slight improvement in the performance of SignalP 3.0-HMM compared to SignalP 2.0-HMM (older version), there is a major performance improvement in SignalP 3.0-NN in comparison with SignalP 2.0-NN. False negatives were observed for Predisi (7%), SigCleave (5%) and SignalP-NN (2%), while there

were none from SignalP-HMM programs. In addition, we tested the prediction programs on a

negative set of 58 non-secretory proteins. False positive predictions were observed for all the

programs: SignalP 3.0-HMM (5%), SignalP 2.0-HMM (7%), SignalP 3.0-NN (10%), SignalP

2.0-NN (12%), Predisi (4%), and Sigcleave (28%). It is noteworthy that false positive predictions

from Sigcleave were particularly high. Based on our results, SignalP 3.0-HMM appears to be the

best program in predicting the signal peptide cleavage sites accurately. This observation is

different from that reported by Zhang et al. [63] who found that SignalP 2.0-NN gave the best

result when tested on experimentally (Edman analysis) verified data set consisting of 270

recombinant human proteins. Recently, Gupta et al.[62] identified signal peptide cleavage sites for

94 proteins in a comprehensive proteomic analysis of Gram-negative bacterium *Shewanella*

oneidensis MR-1 via LC-MS/MS. They tested two programs, Predisi and SignalP. From their

results it appears that Predisi performed better than SignalP. The varying performances of

prediction programs may be associated with the different organisms studied: eukaryotes, gram

negative prokaryotes, gram positive prokaryotes. These observations strongly indicate the need

for experimental data on signal peptides as it will help fine-tune the existing programs.

Discussion

Secreted Proteins of *S. aureus* COL and their Post-Translational Modifications. Secretory

proteins of *S. aureus* are of particular importance to virulence and pathogenesis. Depending on

the *S. aureus* strain, approximately 70 to 90 proteins can be expected to be secreted into the

extracellular milieu.[29] Several proteomic studies have investigated the identification of *S. aureus*

extracellular proteins produced by different strains using a variety of gel-based techniques.

Bernardo et al.[19] identified 13 extracellular proteins produced by *S. aureus* ATCC 29213 and

ATCC 43300 using 1DE and MALDI-TOF-MS. Using 2DE and N-terminal sequencing, Nakano

et al.[42] identified 29 proteins in MRSA strains. Ziebandt et al.[27] identified 18 and 19 proteins from culture supernatants of *S. aureus* COL and RN6390, respectively, employing N-terminal sequencing or MALDI-TOF-MS. Using a combination of techniques (MALDI-TOF-MS/N-terminal sequencing/LC-MS-MS) Ziebandt et al.[26] also studied the influence of accessory gene regulator (*agr*) and alternative sigma factor σ^B (sigB) on the expression of extracellular proteins in RN6390 and RN6911 and identified a total of 43 proteins including secreted, cell wall associated and cytoplasmic proteins. Using 1DE and 2DE with LC-MS/MS, MALDI-TOF-MS and SEDI-MS/MS, Pocsfalvi et al. identified 119 proteins in *S. aureus* ATCC 14458 that included 22 secreted proteins containing potential signal peptides.[64] Burlak et al. have reported the identification of 256 proteins from extracellular extracts of *S. aureus* MW2 and LAC using 2DE and automated direct infusion-MS/MS.[24] However, only 38 of these were predicted to be secreted based on the presence of putative signal peptides. Similarly, Jones et al.[25] reported the identification of 541 proteins from culture supernatants of *S. aureus* UAMS-1 using 1DE and nano-LC-MS/MS, of which only 41 proteins have predicted export signals for secretion into the extracellular milieu. Although previous studies confirmed the presence of a given protein in *S. aureus* extracellular milieu, they failed to provide a detailed characterization of the proteins. This is a particular weakness of peptide-based analyses, since only a fraction of the total theoretical peptide population of a given protein may be identified. The present study provides a comprehensive picture of the secretome of *S. aureus* COL by identification of the proteins and characterization of their post-translational modifications.

All but two of the 59 secreted proteins identified in the present study, were predicted to possess Sec-type signal peptides and we were able to verify the signal peptide loss in these proteins; this confirmed that they were exported via a Sec-dependant pathway. Also, we have

confirmed that EsxA, which is known to be exported via ESAT-6 pathway does not contain a cleavable signal peptide and the only modification observed is the removal of N-terminal methionine.

In a majority of proteins signal peptide loss was the only modification observed. Other observed modifications included proteolytic processing, N-terminal fomylation, methionine removal, oxidation, formation of ornithine and protein truncation. Degradation of few secreted proteins observed in the present study indicated proteolytic activity in the culture supernatants. This observation has been reported by several investigators and it has been suggested that the secreted proteins are degraded by the action of their secreted proteases during culture and sample preparation [20, 26, 42, 65-67] Degradation of proteins by extracellular proteases has also been reported in other microorganisms; in *Bacillus subtilis* it has been demonstrated that mutants lacking proteases exhibit a substantial increase in the abundance of various extracellular proteins compared to the wild type.[65-67] Degradation of extracellular proteins may be due to slow or incorrect post-translational folding of the proteins, or to the presence of exposed protease recognition sequences in the folded protein.[52, 54] It may also be a means of nutrient recycling for survival.[20, 66, 68] However, in the present study a majority of the proteins were refractory to non-specific protease activity since they were identified as intact proteins.

Predicted Versus Observed Secreted Proteins of *S. aureus* COL. 52 *S. aureus* COL secreted proteins out of the 71 predicted proteins[29] were unambiguously identified by LTQ-FT-MS/MS analysis from three or more peptides with a Mascot score above the threshold of significance. Seven predicted secretory proteins, putative uncharacterized protein (SACOL0129), exotoxin 3 (SACOL0468), exotoxin (SACOL0470), exotoxin 3 (SACOL0478), surface protein (SACOL0479), cell wall hydrolase (SACOL1264), and hypothetical protein (SACOL1870) were

identified from one or two peptide sequences (Mascot score, $p < 0.05$) and were not included in the list of identified secreted proteins due to the stringent protein identification criteria used in the present study. These proteins are apparently present in the extracellular medium, in low abundance. The remaining 12 predicted proteins not detected in the extracellular medium are probably not secreted by *S. aureus* COL under the conditions studied or are present in trace amounts. There is no evidence from the published literature to indicate that these proteins are indeed secreted by *S. aureus* COL. Recent proteomic data on membrane, cell wall and extracellular proteins of *Bacillus subtilis* (*B. subtilis*) revealed that a good number of proteins that are predicted to contain cleavable Sec-type signal peptide and Spase I recognition site are not secreted into the medium but are in fact retained in the membrane;[69] this could be the case for the predicted proteins not identified in the present study. Furthermore, five proteins that were not predicted to be secreted due to the presence of transmembrane domains (Nuc and SACOL0442) or the presence of Thr in -1 position (SACOL2179) or +1 position (SACOL1071) or the presence of Tyr in +1 position (SACOL0270) relative to the cleavage site,[29] have been identified in the present study by whole protein and peptide MS analysis as being secreted proteins released into the medium by removal of Sec-type signal peptides. This indicates that Thr in -1 and +1 position and Tyr in +1 position are accepted by *S. aureus* SpaseI; a discussion of the amino acid residues accepted by SPaseI at positions -3 to +1 relative to the signal peptide cleavage site and the frequency of their occurrence is presented below.

A comparison of secreted proteins identified in the present study with those identified in various *S. aureus* strains[29] showed an overlap between identified proteins and revealed potential vaccine and drug candidates. Of the 56 secreted proteins that have been identified in other *S. aureus* strains,[29] 48 are encoded in *S. aureus* COL. 43 of these (Table 1) were identified in the

present study by three or more peptides and three proteins (SACOL0478, SACOL0479 and SACOL1870) were identified by one or two peptides. Two proteins (SACOL0209 and SACOL2691), however, were not identified

Signal Peptides and Cleavage Sites of *S. aureus* COL Secreted Proteins. Table 3 lists the signal sequences of 59 *S. aureus* COL proteins (secretory and cell wall associated proteins) identified in the present study. In gram positive bacteria, Spase I recognizes residues at positions -3 and -1 with respect to the cleavage site, and Ala-X-Ala is the most common sequence preceding the signal peptide cleavage site.[70] It is evident from Table 3 that the signal sequences of *S. aureus* COL proteins identified in the present study all contain the N, H and C domains of a typical Sec-type signal peptide. The length of the signal peptides varies from 23 to 60 amino acids with an average of 31 residues. Table 4 lists the residues accepted at and around the verified SPase I cleavage sites of *S. aureus* COL proteins identified in the present study. In a majority of the proteins, Ala is predominantly preferred at -3 (77%) and -1 (97%) positions. Residues, Val (10%) and Ser (10%) are also accepted at -3 position and occur with a higher frequency than Thr, Leu, and Ile (2%). With respect to -1 position, residues Ser and Thr are also accepted, however, with a markedly lower frequency (2%) than Ala. The residues found in -3 and -1 positions of *S. aureus* COL signal sequences are small and uncharged; this is in agreement with the assumption that side chains of residues at the -1 and -3 positions are bound in two shallow hydrophobic substrate-binding pockets (S1 and S3) of the active site of Spase I.[71] In contrast, the side chain of the residue at position -2 is thought to be pointing outwards from the enzyme. As a consequence, a variety of residues appear to be tolerated at the -2 position including Lys, Asn, Gln, His, Asp, Ser, Glu, Leu, Phe, Tyr, Gly, Arg with a preference for Lys (23%). There appears to be a preference for Ala (31%) at +1 position. Other residues including

Ser, Glu, Lys, Asp, Gln, Thr, Phe, Asn, Tyr, and Leu were also accepted in the +1 position. Approximately 78% of *S. aureus* COL secreted proteins identified in the present study possess signal sequences that contain a helix-breaking residue (mostly glycine) in the middle of the H-domain, and about 50% contain a helix-breaking residue (proline or glycine) at position -7 to -4 relative to the predicted processing site for SPase I. Helix-breaking residues found at the end of the H-domain are thought to facilitate cleavage by SPase I.[71]

Signal Peptide Processing at Two Cleavage Sites by SPase I. Whole protein MS analysis of SACOL1071 and SACOL2006 revealed processing of the signal peptide at more than one site. This observation is very interesting as signal peptide cleavage by SPase I is generally considered to be highly specific[72, 73] and reports on the observation of signal peptide cleavage at multiple sites in wild-type proteins are very rare. The reasons for the high fidelity of SPase I are not clearly understood. An important requirement for cleavage by SPase I is the presence of amino acids with small neutral side chains at positions -1 and -3 in the C-region; the -3 position is less restrictive than -1.[33] This is also evident from Table 4. SACOL1071 and SACOL2006 each have more than one potential cleavage site that is in compliance with the substrate specificity of SPase I and may compete for recognition. The N-terminal signal sequence MNKLLQSLSALGVS AT LVTPN L-N-A$^{24\downarrow}$-D-A$^{26\downarrow}$ of SACOL1071 contains the ubiquitous A-X-A motif (-3 to -1 position) as well as an L-X-A motif (-5 to -3 position) with observed cleavages ($^\downarrow$) at the -3 and -1 positions. The weighted average signal intensities in the whole protein mass spectrometry experiments suggest that signal peptide processing at the -1 position (82%) is preferred over the -3 position (18%). Similarly, the N-terminal sequence of SACOL2006, MKNKKRVLIASSLSCAILLLSAATT-Q-A$^{27\downarrow}$-N-S$^{29\downarrow}$ also contains two potential sites for signal peptide cleavage, the A-X-S sequence (-3 to -1 position) and the T-X-A sequence (-5 to -

3 position) with observed cleavages ($\downarrow$) at the -1 and -3 position, respectively. The weighted average of signal intensities of the two mature proteins derived from whole protein MS experiments suggests that processing at the -3 position (64%) is preferred over the -1 position (36%). From our results it appears that SPase I is capable of processing the signal peptide at more than one site depending on the availability of an alternate potential site close by (-3 and -1 position in the two examples). The differences in the nature of residues at -1, -3 and -5 positions, the proximity of the cleavage site to the hydrophobic core, the β-turn structure, and the residues that are capable of breaking α-helix or β-strand structure probably play a role in the efficiency of processing at different sites.[74] A close inspection of the signal sequences of the proteins identified in this study (Table 3) indicates that there are 11 additional sequences that have SPase I compatible residues in the -5 to -3 positions. However, signal peptide cleavage at only one site was detected in these cases. The observed cleavages at two sites for SACOL1071 and SACOL2006 may be due to some unique characteristics of their signal sequences. In general, prokaryotic signal peptides possess redundant information, in that they contain more than one potential cleavage site, more than one basic residue in the N-region and a longer H-region than required.[75, 76] It is not clear why alternate sites for signal peptide cleavage exist or if there is some biological significance to this.

Possible Secretion of SACOL1071 via Alternate Secretory Pathway. Chitin, a homopolymer of β-1,4-*N*-acetyl-D-glucosamine (GlcNAc), is one of the most abundant natural polymers. Several bacterial species secrete chitinolytic enzymes and chitin-binding proteins that are thought to degrade chitin.[77, 78] SACOL1071, a putative chitinase B protein has a predicted Sec-type signal peptide and has been identified in this study as a mature protein formed by removal of this peptide. In addition, another mature form of the protein, SACOL1071[3] has also been

identified by whole protein and peptide analysis. It appears to have been secreted by cleavage of a 14 residue N-terminal segment (Figure 9B). This 14 residue signal peptide, MNKLLQSLSALGVS, does not resemble any known signal peptide suggesting secretion of SACOL1071[3] by an alternate pathway. In gram negative *Pseudomonas aeruginosa* bacteria, Folders et al. identified a chitinase C (ChiC) protein in the extracellular medium following cleavage of an 11 residue signal peptide (MIRIDFSQLHQ).[79] ChiC also does not contain a typical N-terminal sequence. Since SACOL1071 has a functional Sec-dependent export route, it would be intriguing if it were secreted by an alternate pathway.

Non-Secretory Proteins in the Extracellular Medium. The presence of non-secretory proteins in the extracellular medium has been reported by several investigators for a number of organisms including *S. aureus*.[24, 65, 80] The possible explanation for the presence of cytoplasmic proteins in the extracellular medium is cell lysis during growth; this has been visually confirmed in the present study by fluorescence microscopy using live/dead staining method. Although only 3.0 (± 0.5%) of the cells in the stationary phase culture were lysed, that is sufficient to detect the very abundant proteins listed in Supplemental Table 5. The rest of the non-secretory proteins were most probably released into the extracellular medium by proteolytic processing, shedding or cell wall turnover as suggested by other investigators.[81, 82]

Stable Cleaved Signal Peptides and Signal Peptide Fragments. We have reported our unexpected findings on the presence of cleaved stable signal peptides and signal peptide fragments in the extracellular medium. In *Escherichia coli* (*E. coli*) two signal peptide peptidases (SPPases), a membrane bound protease IV and cytoplasmic oligopeptidase A have been identified.[83] It has been proposed that protease IV initially cleaves the SPase I processed signal peptides by making endoproteolytic cuts and the products of the initial cleavage may be diffused

or transported back into the cytoplasm and further degraded into amino acids by oligopeptidase A and other cytoplasmic enzymes. Similar proteins have also been identified in *B. subtilis*[84]. However, homologous proteins have not been found in *S. aureus*.[29] Since we have identified fragments of signal peptides, degradation of cleaved signal peptides must also happen in *S. aureus* by some unknown proteases. All the signal peptide fragments observed in the present study appear to have formed by endoproteolytic cleavage at hydrophobic residues including Leu, Thr, Ala, Val, Phe and Met (Table 2), suggesting that the protease responsible for processing signal peptide in *S. aureus* may have a preference for hydrophobic residues and its substrate specificity is similar to protease IV of *E. coli*.[83] As mentioned earlier, only C-terminal signal peptide fragments and not N-terminal signal peptide fragments have been identified in the extracellular medium in the present study. This suggests that an endoproteolytic cut in the hydrophobic region of the signal peptides releases the C-terminal signal peptide fragments into the extracellular medium, while the N-terminal fragments are retained in the membrane or released into the cytoplasm or completely degraded.

The observation of cleaved stable signal peptides in the extracellular medium raises some interesting questions. Since various reports have suggested that after cleavage from the pre-protein, rapid degradation of signal peptide is important for proper functioning of the export machinery,[58, 59] why are the observed cleaved signal peptides stable ? Also, because their long hydrophobic cores should lead to peptide retention in the membrane, how are they released into the medium? What is the biological role of the released signal peptide and signal peptide fragments? The fact that the signal peptides of Sle1, SACOL0723, SceD, IsaA, and SACOL2295 were detected in the culture supernatant indicates that they have high solubility in water. Also the Zyggregator algorithm[85] (that analyzes aggregation propensities of proteins) predicts that the

cleaved signal peptides identified in the present study should not have the propensity to aggregate. It is possible that certain unique biophysical properties allow the observed cleaved signal peptides to be readily released from the membrane into the extracellular medium after cleavage from the pre-protein.

It is important to note here that the fate of cleaved signal peptides and their fragments is not clearly understood, particularly for prokaryotes. However, recent studies have suggested that cleaved signal sequences and their fragments may have important biological functions including signaling. For instance, bacteriocin release protein (BRP), an *E. Coli* lipoprotein, is processed very slowly by signal peptidase II to yield a mature BRP and a stable signal peptide that mediate the translocation of cloacin DF13.[86] In eukaryotes, the N-terminal signal peptide fragments of preprolactin and human immunodeficiency virus-1 p-gp160 were found to be released into the cytoplasm and bound to calmodulin/Ca^{2+} suggesting that the liberated signal peptide fragments may influence signal transduction pathways in the cell.[87] As stated by Martoglio et al.,[88] signal sequences are more than just simple greasy peptides possessing a wealth of functional information. It can be conjectured that cleaved signal peptides and signal peptide fragments of *S. aureus* COL released into the extracellular medium also have the potential to perform important roles. Tristan et al., have proposed that the signal peptide of panton-valentine leukocidin LUKS (LukS-PV), a component of hetero-oligomeric pore-forming toxin (PLV) in *S. aureus* PVL positive strains, mediates increased adhesion to extracellular matrix components.[89] According to their hypothesis, LukS-PV signal peptide is released from the membrane after cleavage by SPase I and attaches to the cell wall through its unique C-terminal (TSFHESKA) sequence thus exposing the positively charged N-terminus to the extracellular matrix molecules. *S. aureus* COL does not encode PLV gene, therefore we could not confirm the release of LukS-PV signal

peptide. Also, none of the signal peptides identified in the present study share homology with LukS-PV signal peptide.

Conclusions. We have presented a new approach to study *S. aureus* extracellular proteins utilizing the combination of whole protein MS analysis and peptide MS/MS analysis that enabled us to provide the most comprehensive view of *S. aureus* extracellular proteins reported to date. We have identified 59 secreted proteins in *S .aureus* COL, and characterized their post-translational modifications. Accurate determination of signal peptide cleavage sites of *S. aureus* COL secreted proteins allowed us evaluate the prediction accuracies of several programs. Since *S. aureus* COL secreted proteins are potential drug targets or vaccine candidates, the signal peptide cleavage site information provided by this study will be useful in making constructs, recombinant proteins and in protein engineering. This information will also extend our current knowledge base on experimentally verified signal peptide cleavage sites and may assist in improving prediction accuracies of the programs. We have also detected signal peptide processing at multiple sites and the release of cleaved signal peptides and signal peptide fragments into the extracellular medium. These observations are unusual and their biological significance is yet to be understood. The current approach should be useful in secretome analysis of other *S. aureus* strains and other pathogenic organisms.

References

1. Diekema, D. J.; Pfaller, M. A.; Schmitz, F. J.; Smayevsky, J.; Bell, J.; Jones, R. N.; Beach, M., *Clinical Infectious Diseases* **2001,** *32* (Suppl. 2), S114-S132.
2. Lowy, F. D., *The New England journal of medicine* **1998,** *339* (8), 520-32.
3. Noskin Gary, A.; Rubin Robert, J.; Schentag Jerome, J.; Kluytmans, J.; Hedblom Edwin, C.; Jacobson, C.; Smulders, M.; Gemmen, E.; Bharmal, M., *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2007,** *45* (9), 1132-40.
4. Klevens, R. M.; Morrison Melissa, A.; Nadle, J.; Petit, S.; Gershman, K.; Ray, S.; Harrison Lee, H.; Lynfield, R.; Dumyati, G.; Townes John, M.; Craig Allen, S.; Zell Elizabeth, R.; Fosheim Gregory, E.; McDougal Linda, K.; Carey Roberta, B.; Fridkin Scott, K., *JAMA : the journal of the American Medical Association* **2007,** *298* (15), 1763-71.
5. Klein, E.; Smith David, L.; Laxminarayan, R., *Emerging infectious diseases* **2007,** *13* (12), 1840-6.
6. Stetson, M. G. V., R. J., *A.J.I.C* **2006,** *34* (5), E40-E41.
7. Chang, S.; Sievert Dawn, M.; Hageman Jeffrey, C.; Boulton Matthew, L.; Tenover Fred, C.; Downes Frances, P.; Shah, S.; Rudrik James, T.; Pupp Guy, R.; Brown William, J.; Cardo, D.; Fridkin Scott, K. *Infection with vancomycin-resistant Staphylococcus aureus containing the vanA resistance gene*; Epidemic Intelligence Service, Division of Applied Public Health Training, Epidemiology Program Office, Centers for Disease Control and Prevention, Atlanta 30333, USA. (Vancomycin-Resistant Staphylococcus aureus Investigative Team): United States, 2003; pp 1342-7.
8. Srinivasan, A.; Dick, J. D.; Perl, T. M., *Clinical Microbiology Reviews* **2002,** *15* (3), 430-438.
9. Whitener Cynthia, J.; Park Sarah, Y.; Browne Fred, A.; Parent Leslie, J.; Julian, K.; Bozdogan, B.; Appelbaum Peter, C.; Chaitram, J.; Weigel Linda, M.; Jernigan, J.; McDougal Linda, K.; Tenover Fred, C.; Fridkin Scott, K. *Vancomycin-resistant Staphylococcus aureus in the absence of vancomycin exposure*; The Penn State Milton S. Hershey Medical Center, Hershey, Pennsylvania 17033, USA.: United States, 2004; pp 1049-55.
10. Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Aoki, K.; Nagai, Y.; Lian, J.; Ito, T.; Kanamori, M.; Matsumaru, H.; Maruyama, A.; Murakami, H.; Hosoyama, A.; Mizutani-Ui, Y.; Takahashi, N. K.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Hirakawa, H.; Kuhara, S.; Goto, S.; Yabuzaki, J.; Kanehisa, M.; Yamashita, A.; Oshima, K.; Furuya, K.; Yoshino, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K., *Lancet* **2001,** *357* (9264), 1225-40.
11. Baba, T.; Takeuchi, F.; Kuroda, M.; Yuzawa, H.; Aoki, K.-i.; Oguchi, A.; Nagai, Y.; Iwama, N.; Asano, K.; Naimi, T.; Kuroda, H.; Cui, L.; Yamamoto, K.; Hiramatsu, K., *Lancet* **2002,** *359* (9320), 1819-1827.
12. Holden Matthew, T. G.; Feil Edward, J.; Lindsay Jodi, A.; Peacock Sharon, J.; Day Nicholas, P. J.; Enright Mark, C.; Foster Tim, J.; Moore Catrin, E.; Hurst, L.; Atkin, R.; Barron, A.; Bason, N.; Bentley Stephen, D.; Chillingworth, C.; Chillingworth, T.; Churcher, C.; Clark, L.; Corton, C.; Cronin, A.; Doggett, J.; Dowd, L.; Feltwell, T.; Hance, Z.; Harris, B.; Hauser, H.; Holroyd, S.; Jagels, K.; James Keith, D.; Lennard, N.; Line, A.; Mayes, R.; Moule, S.; Mungall, K.; Ormond, D.; Quail Michael, A.; Rabbinowitsch, E.; Rutherford, K.; Sanders,

M.; Sharp, S.; Simmonds, M.; Stevens, K.; Whitehead, S.; Barrell Bart, G.; Spratt Brian, G.; Parkhill, J., *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (26), 9786-91.

13. Lindsay, J. A.; Holden, M. T. G., *Trends in Microbiology* **2004,** *12* (8), 378-385.

14. Blobel, G., *Proceedings of the National Academy of Sciences of the United States of America* **1980,** *77* (3), 1496-500.

15. Dalbey, R. E.; Lively, M. O.; Bron, S.; Van Dijl, J. M., *Protein Science* **1997,** *6* (6), 1129-1138.

16. McCormick, J. K.; Yarwood, J. M.; Schlievert, P. M., *Annual Review of Microbiology* **2001,** *55*, 77-104.

17. Novak, P.; Ray, P. H.; Dev, I. K., *The Journal of biological chemistry* **1986,** *261* (1), 420-7.

18. Kawano, Y.; Ito, Y.; Yamakawa, Y.; Yamashino, T.; Horii, T.; Hasegawa, T.; Ohta, M., *FEMS Microbiology Letters* **2000,** *189* (1), 103-108.

19. Bernardo, K.; Fleer, S.; Pakulat, N.; Krut, O.; Hunger, F.; Kronke, M., *Proteomics* **2002,** *2* (6), 740-746.

20. Kawano, Y.; Kawagishi, M.; Nakano, M.; Mase, K.; Yamashino, T.; Hasegawa, T.; Ohta, M., *Microbiology and Immunology* **2001,** *45* (4), 285-290.

21. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J.-C., *Electrophoresis* **2000,** *21* (6), 1104-1115.

22. Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R., *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (17), 9390-9395.

23. Issaq, H. J.; Conrads, T. P.; Janini, G. M.; Veenstra, T. D., *Electrophoresis* **2002,** *23* (17), 3048-3061.

24. Burlak, C.; Hammer, C. H.; Robinson, M.-A.; Whitney, A. R.; McGavin, M. J.; Kreiswirth, B. N.; DeLeo, F. R., *Cellular Microbiology* **2007,** *9* (5), 1172-1190.

25. Jones, R. C.; Deck, J.; Edmondson, R. D.; Hart, M. E., *Journal of Bacteriology* **2008,** *190* (15), 5265-5278.

26. Ziebandt, A.-K.; Becher, D.; Ohlsen, K.; Hacker, J.; Hecker, M.; Engelmann, S., *Proteomics* **2004,** *4* (10), 3034-3047.

27. Ziebandt, A.-K.; Weber, H.; Rudolph, J.; Schmid, R.; Hoper, D.; Engelmann, S.; Hecker, M., *Proteomics* **2001,** *1* (4), 480-493.

28. Karty, J. A.; Running, W. E.; Reilly, J. P., *Journal of Chromatography, B: Analytical Technologies in the Biomedical and Life Sciences* **2007,** *847* (2), 103-113.

29. Sibbald, M. J. J. B.; Ziebandt, A. K.; Engelmann, S.; Hecker, M.; de Jong, A.; Harmsen, H. J. M.; Raangs, G. C.; Stokroos, I.; Arends, J. P.; Dubois, J. Y. F.; van Dijl, J. M., *Microbiology and molecular biology reviews : MMBR* **2006,** *70* (3), 755-88.

30. Read, S. M.; Northcote, D. H., *Analytical Biochemistry* **1981,** *116* (1), 53-64.

31. Gill, S. R.; Fouts, D. E.; Archer, G. L.; Mongodin, E. F.; DeBoy, R. T.; Ravel, J.; Paulsen, I. T.; Kolonay, J. F.; Brinkac, L.; Beanan, M.; Dodson, R. J.; Daugherty, S. C.; Madupu, R.; Angiuoli, S. V.; Durkin, A. S.; Haft, D. H.; Vamathevan, J.; Khouri, H.; Utterback, T.; Lee, C.; Dimitrov, G.; Jiang, L.; Qin, H.; Weidman, J.; Tran, K.; Kang, K.; Hance, I. R.; Nelson, K. E.; Fraser, C. M., *Journal of Bacteriology* **2005,** *187* (7), 2426-2438.

32. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.-C.; Estreicher, A.; Gasteiger, E.; Martin, M. J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M., *Nucleic Acids Research* **2003,** *31* (1), 365-370.

33. Nielsen, H.; Engelbrecht, J.; Brunak, S.; Von Heijne, G., *Protein Engineering* **1997,** *10* (1), 1-6.

34. Dyrlov Bendtsen, J.; Nielsen, H.; von Heijne, G.; Brunak, S., *J. Mol. Biol.* **2004,** *340* (4), 783-795.

35. Hiller, K.; Grote, A.; Scheer, M.; Munch, R.; Jahn, D., *Nucleic Acids Res* **2004,** *32* (Web Server issue), W375-9.

36. Von Heijne, G., *Nucleic Acids Research* **1986,** *14* (11), 4683-90.

37. Von Heijne, G., *Sequence Analysis in Molecular Biology. Treasure Trove on Trivial Pursuit.* 1987; p 188 pp.

38. Juncker, A. S.; Willenbrock, H.; Von Heijne, G.; Brunak, S.; Nielsen, H.; Krogh, A., *Protein Science* **2003,** *12* (8), 1652-1662.

39. Cserzo, M.; Wallin, E.; Simon, I.; von Heijne, G.; Elofsson, A., *Protein Eng* **1997,** *10* (6), 673-6.

40. Gardy, J. L.; Laird, M. R.; Chen, F.; Rey, S.; Walsh, C. J.; Ester, M.; Brinkman, F. S. L., *Bioinformatics* **2005,** *21* (5), 617-623.

41. Vandenesch, F.; Kornblum, J.; Novick, R. P., *Journal of Bacteriology* **1991,** *173* (20), 6313-20.

42. Nakano, M.; Kawano, Y.; Kawagishi, M.; Hasegawa, T.; Iinuma, Y.; Ohta, M., *Microbiology and Immunology* **2002,** *46* (1), 11-22.

43. Jeljaszewicz, J.; Mlynarczyk, A.; Mlynarczyk, G., *Bulletin of the Polish Academy of Sciences: Biological Sciences* **2000,** *48* (2), 77-98.

44. Fitton, J. E.; Dell, A.; Shaw, W. V., *FEBS Letters* **1980,** *115* (2), 209-12.

45. Janzon, L.; Lofdahl, S.; Arvidson, S., *Molecular & general genetics : MGG* **1989,** *219* (3), 480-5.

46. Carlos, J. L.; Paetzel, M.; Brubaker, G.; Karla, A.; Ashwell, C. M.; Lively, M. O.; Cao, G.; Bullinger, P.; Dalbey, R. E., *Journal of Biological Chemistry* **2000,** *275* (49), 38813-38822.

47. Reed, S. B.; Wesson, C. A.; Liou, L. E.; Trumble, W. R.; Schlievert, P. M.; Bohach, G. A.; Bayles, K. W., *Infect. Immun.* **2001,** *69* (3), 1521-1527.

48. Lauber, M. A.; Running, W. E.; Reilly, J. P., *Journal of Proteome Research* **2009,** *8* (9), 4193-4206.

49. Rollof, J.; Normark, S., *Journal of Bacteriology* **1992,** *174* (6), 1844-7.

50. Gotz, F.; Verheij, H. M.; Rosenstein, R., *Chemistry and Physics of Lipids* **1998,** *93* (1-2), 15-25.

51. Nikoleit, K.; Rosenstein, R.; Verheij, H. M.; Goetz, F., *European Journal of Biochemistry* **1995,** *228* (3), 732-8.

52. Foster, S. J., *J Bacteriol* **1995,** *177* (19), 5723-5.

53. Oshida, T.; Sugai, M.; Komatsuzawa, H.; Hong, Y. M.; Suginaka, H.; Tomasz, A., *Proceedings of the National Academy of Sciences of the United States of America* **1995,** *92* (1), 285-9.

54. Wolf, C.; Hochgraefe, F.; Kusch, H.; Albrecht, D.; Hecker, M.; Engelmann, S., *Proteomics* **2008,** *8* (15), 3139-3153.

55. Burts Monica, L.; Williams Wade, A.; DeBord, K.; Missiakas Dominique, M., *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (4), 1169-74.

56. Renshaw, P. S.; Lightbody, K. L.; Veverka, V.; Muskett, F. W.; Kelly, G.; Frenkiel, T. A.; Gordon, S. V.; Hewinson, R. G.; Burke, B.; Norman, J.; Williamson, R. A.; Carr, M. D., *EMBO Journal* **2005,** *24* (14), 2491-2498.

57. Sundaramoorthy, R.; Fyfe, P. K.; Hunter, W. N., *J. Mol. Biol.* **2008,** *383* (3), 603-614.

58. Koren, R.; Burstein, Y.; Soreq, H., *Proceedings of the National Academy of Sciences of the United States of America* **1983,** *80* (23), 7205-9.

59. Chen, L.; Tai, P. C.; Briggs, M. S.; Gierasch, L. M., *The Journal of biological chemistry* **1987,** *262* (4), 1427-9.

60. Yvon, M.; Chabanet, C.; Pelissier, J. P., *International Journal of Peptide & Protein Research* **1989,** *34* (3), 166-76.

61. Leversen, N. A.; de Souza, G. A.; Maalen, H.; Prasad, S.; Jonassen, I.; Wiker, H. G., *Microbiology (Reading, United Kingdom)* **2009,** *155* (7), 2375-2383.

62. Gupta, N.; Tanner, S.; Jaitly, N.; Adkins Joshua, N.; Lipton, M.; Edwards, R.; Romine, M.; Osterman, A.; Bafna, V.; Smith Richard, D.; Pevzner Pavel, A., *Genome Res* **2007,** *17* (9), 1362-77.

63. Zhang, Z.; Henzel, W. J., *Protein Science* **2004,** *13* (10), 2819-2824.

64. Pocsfalvi, G.; Cacace, G.; Cuccurullo, M.; Serluca, G.; Sorrentino, A.; Schlosser, G.; Blaiotta, G.; Malorni, A., *Proteomics* **2008,** *8* (12), 2462-2476.

65. Antelmann, H.; Tjalsma, H.; Voigt, B.; Ohlmeier, S.; Bron, S.; Van Dijl, J. M.; Hecker, M., *Genome Research* **2001,** *11* (9), 1484-1502.

66. Westers, L.; Westers, H.; Zanen, G.; Antelmann, H.; Hecker, M.; Noone, D.; Devine, K. M.; van Dijl, J. M.; Quax, W. J., *Proteomics* **2008,** *8* (13), 2704-2713.

67. Aziz, R. K.; Pabst, M. J.; Jeng, A.; Kansal, R.; Low, D. E.; Nizet, V.; Kotb, M., *Molecular Microbiology* **2003,** *51* (1), 123-134.

68. Sarvas, M.; Harwood, C. R.; Bron, S.; van Dijl, J. M., *Biochimica et Biophysica Acta, Molecular Cell Research* **2004,** *1694* (1-3), 311-327.

69. Tjalsma, H.; van Dijl Jan, M., *Proteomics* **2005,** *5* (17), 4472-82.

70. Schneewind, O.; Missiakas, D., *Bacterial Pathogenomics* **2007**, 301-326.

71. Paetzel, M.; Dalbey, R. E.; Strynadka, N. C. J., *Nature (London)* **1998,** *396* (6712), 707.

72. Sekizaki, T.; Takai, S.; Egawa, Y.; Ikeda, T.; Ito, H.; Tsubaki; Shiro, *Gene* **1995,** *155* (1), 135-6.

73. Folz, R. J.; Nothwehr, S. F.; Gordon, J. I., *Journal of Biological Chemistry* **1988,** *263* (4), 2070-8.

74. Fikes, J. D.; Barkocy-Gallagher, G. A.; Klapper, D. G.; Bassford, P. J., Jr., *Journal of Biological Chemistry* **1990,** *265* (6), 3417-23.

75. Puziss, J. W.; Fikes, J. D.; Bassford, P. J., Jr., *J Bacteriol* **1989,** *171* (5), 2303-11.

76. Fikes, J. D.; Bankaitis, V. A.; Ryan, J. P.; Bassford, P. J., Jr., *J Bacteriol* **1987,** *169* (6), 2345-51.

77. Bassler, B. L.; Yu, C.; Lee, Y. C.; Roseman, S., *The Journal of biological chemistry* **1991,** *266* (36), 24276-86.

78. Watanabe, T.; Kimura, K.; Sumiya, T.; Nikaidou, N.; Suzuki, K.; Suzuki, M.; Taiyoji, M.; Ferrer, S.; Regue, M., *J Bacteriol* **1997,** *179* (22), 7111-7.

79. Folders, J.; Algra, J.; Roelofs, M. S.; Van Loon, L. C.; Tommassen, J.; Bitter, W., *Journal of Bacteriology* **2001,** *183* (24), 7044-7052.

80. Maalen, H.; Berven, F. S.; Fladmark, K. E.; Wiker, H. G., *Proteomics* **2007,** *7* (10), 1702-1718.

81. Tjalsma, H.; Antelmann, H.; Jongbloed, J. D. H.; Braun, P. G.; Darmon, E.; Dorenbos, R.; Dubois, J.-Y. F.; Westers, H.; Zanen, G.; Quax, W. J.; Kuipers, O. P.; Bron, S.; Hecker, M.; van Dijl, J. M., *Microbiology and Molecular Biology Reviews* **2004,** *68* (2), 207-233.

82. Tjalsma, H.; van Diji, J. M., *Proteomics* **2005,** *5* (17), 4472-4482.

83. Novak, P.; Dev, I. K., *J Bacteriol* **1988,** *170* (11), 5067-75.

84. Bolhuis, A.; Matzen, A.; Hyyrylainen, H. L.; Kontinen, V. P.; Meima, R.; Chapuis, J.; Venema, G.; Bron, S.; Freudl, R.; van Dijl, J. M., *The Journal of biological chemistry* **1999,** *274* (35), 24585-92.

85. Tartaglia, G. G.; Vendruscolo, M., *Chemical Society Reviews* **2008,** *37* (7), 1395-1401.

86. Van der Wal, F. J.; Oudega, B.; Kater, M. M.; Ten Hagen-Jongman, C. M.; De Graaf, F. K.; Luirink, J., *Molecular Microbiology* **1992,** *6* (16), 2309-18.

87. Martoglio, B.; Graf, R.; Dobberstein, B., *EMBO Journal* **1997,** *16* (22), 6636-6645.

88. Martoglio, B.; Dobberstein, B., *Trends in Cell Biology* **1998,** *8* (10), 410-415.

89. Tristan, A.; Benito, Y.; Montserret, R.; Boisset, S.; Dusserre, E.; Penin, F.; Ruggiero, F.; Etienne, J.; Lortat-Jacob, H.; Lina, G.; Bowden, M. G.; Vandenesch, F., *PLoS One* **2009,** *4* (4), e5042.

Figure Legends

Figure 1. Experimental strategy for identification of Nuc in Trap 7. (A) Combined sequence coverage map of Nuc from trypsin and Glu-C digestion, the underlined amino acids were identified. Shaded region corresponds to the signal sequence. (B) Signal peptide cleavage site predictions and corresponding predicted masses for Nuc. (C) Total ion chromatogram of Trap 7 containing a peak corresponding to Nuc; (D) Raw spectrum of Nuc showing the charge state distribution. Asterisks show charge state distribution of Hlb; (E) Deconvoluted mass spectrum; (F) MS/MS spectrum of N-terminal peptide SQTDNGVNRSGSEDPTVYSATSTK of mature Nuc.

Figure 2. A) LTQ-FT-MS/MS of Hld[1], the asterisked residue (M1) is the site of modification (formylation and oxidation). B) LTQ-FT-MS/MS of Hld[2]

Figure 3. Deconvoluted spectrum showing C-terminal truncated form of LukE, corresponding sequence is shown in parenthesis.

Figure 4. Deconvoluted spectrum of HlgB showing three truncated forms, HlgB[1], HlgB[2], HlgB[3] corresponding sequences are shown in parenthesis.

Figure 5. Deconvoluted mass spectrum of SACOL2006 suggesting signal peptide processing at two sites ; SACOL2006 [1] resulted from signal peptide cleavage at position 29 and SACOL2006 [2] resulted from signal peptide cleavage at position 27.

Figure 6. Identification of sequencing error in SplD. (A) Whole protein mass spectrum of co-eluting SplF and SplD. (B) Multiple sequence alignment (residues 40-70) of SplD from *S. aureus* COL and other strains of *S. aureus*. Suspected sequencing error is shown in the box. An asterisk represents identity among the aligned residues and two dots represent strong similarity.

(C) LTQFT-MS/MS spectrum of SplD peptide consisting of residues 42-74 that confirm $Gln^{68} \rightarrow His^{68}$.

Figure 7. Deconvoluted mass spectra showing pro-enzyme forms of (A) Lip1 and (B) Lip2

Figure 8. Deconvoluted mass spectra showing the three Atl gene products identified in the present study. (A) Atl^{1}, pro-Atl form; (B) Atl^{2}, intermediate AM form; (C) Atl^{3}, mature GL form.

Figure 9. (A) Deconvoluted mass spectrum showing three forms of SACOL1071 formed as a result of signal peptide processing at different sites; $SACOL1071^{1}$, signal peptide cleavage at position 26; $SACOL1071^{2}$, signal peptide cleavage at position 26; $SACOL1071^{3}$, signal peptide cleavage at position 14. (B) LTQ-FT-MS/MS spectrum of N-terminal peptide of $SACOL1071^{3}$.

Figure 10. (A) LTQ-FT-MS/MS spectrum of stable cleaved signal peptide of Isa and (B) LTQ-FT-MS/MS spectrum of signal peptide fragment of SACOL1164.

Figure 11. Prediction accuracies of various signal peptide prediction programs.

Table 1. *S. aureus* COL secreted proteins identified in the present study.

Gene ID	Protein name	Theor. mass	Predicted mass[†]	Observed mass[§]	Mass error	Modifications	Trap Location	MS/MS % Seq. Cov.*
Cytotoxins								
SACOL1173[e]	Alpha-hemolysin precursor (HlY)	35973.3	33260.1	33260.6	0.5	-Sig P	Trap 6	84
SACOL2003[e]	Beta-hemolysin (Hlb)	37237.8	33742.6	33742.8	0.1	-Sig P	Trap 7	90
SACOL2022[1, c]	Delta-hemolysin (Hld[1])	2976.6[d]	2976.6[d]	2976.6[d]	0.0			100
SACOL2022[2, c]	Delta-hemolysin (Hld[2])	2976.6[d]	3020.6[d]	3020.6[d]	0.0	+ N-terminal formylation,+16 Da		100
SACOL2419[a]	Gamma hemolysin, component A (HlgA)	34955.7	31921.9				Trap 10	53
SACOL2422[1, e]	Gamma hemolysin, component B (HlgB[1])	36711.0	33392.0	33392.6	0.6	-Sig P, truncated protein	Trap 4	69
SACOL2422[2]	Gamma hemolysin, component B (HlgB[2])	36711.0	33462.1	33463.6	1.5	-Sig P, truncated protein	Trap 4	69
SACOL2422[3]	Gamma hemolysin, component B (HlgB[3])	36711.0	33506.2	33507.4	1.2	-Sig P, truncated protein	Trap 4	69
SACOL2421[e]	Gamma hemolysin, component C (HlgC)	35625.8	32565.2	32566.2	1.0	-Sig P	Trap 9	85
SACOL1880[e]	Leukotoxin LukD (LukD)	36888.9	34158.6	34159.1	0.5	-Sig P	Trap 6	53
SACOL1881[e]	Leukotoxin LukE (LukE)	34819.1	31750.4	31751.5	1.1	-Sig P, -C-terminal residue N	Trap 8	68
SACOL2004[e]	leukocidin subunit precursor, putative	38686.1	35573.3	35574.4	1.1	-Sig P	Trap 8	74
SACOL2006[1 b, e]	Aerolysin/Leukocidin family protein	40434.0	37619.7	37620.8	1.1	-Sig P	Trap 10	72
SACOL2006[2, b]	Aerolysin/Leukocidin family protein	40434.0	37418.4	37419.3	0.9	-Sig P	Trap 10	73
Superantigenic Toxins								
SACOL0442	Staphylococcal enterotoxin	23165.4	19343.9	19345.0	1.1	-Sig P	Trap 8	69
SACOL0886[e]	Staphylococcal enterotoxin (Sek)	27727.1	24698.4	24699.6	1.2	-Sig P, -C-terminal residues YKETI	Trap 12	82
SACOL0887[e]	Staphylococcal enterotoxin type I (Sei)	28184.6	24846.6	24848.0	1.4	-Sig P, -C-terminal residues TE	Trap 20	77
SACOL0907[e]	Staphylococcal enterotoxin B (Seb)	31435.8	28368.0	28368.5	0.5	-Sig P	Trap 8	100
Proteases								
SACOL1869[e]	Serine protease SplA (SplA)	25876.2	21853.5	21854.4	0.9	-Sig P	Trap 7	70
SACOL1868[e]	Serine protease SplB (SplB)	26096.4	22371.1	22372.0	0.9	-Sig P	Trap 6	66
SACOL1867[e]	Serine protease SplC (SplC)	26098.4	22388.0	22388.7	0.7	-Sig P	Trap 4	48
SACOL1866[e]	Serine protease SplD (SplD)	25669.1	22010.8	22012.2	1.4	-Sig P, Seq. error	Trap 5	77
SACOL1865[e]	Serine protease SplE (SplE)	25679.3	22011.9	22013.3	1.4	-Sig P	Trap 7	32
SACOL1864[e]	Serine protease SplF, (SplF)	25655.1	21941.7	21942.7	1.0	-Sig P	Trap 5	93
SACOL1057[a, e]	V8 Protease (SspA)	36312.6	33376.1				Trap 5	42
SACOL1056[e]	Cysteine protease precursor SspB (SspB1)	44519.0	40649.4	40650.5	1.1	-Sig P	Trap 5	78
SACOL1970[a, e]	Cysteine protease precursor SspB (SspB2)	44252.1	41524.8				Trap 12	56
SACOL2659[a, e]	Zinc metalloproteinase aureolysin (Aur)	56361.3	53459.8				Trap 8	48
Lipolytic enzymes								
SACOL2694[e]	Lipase1 (Lip1)	76675.3	73077.3	73078.0	0.8	-Sig P	Trap 8	96
SACOL0317[e]	Lipase 2 (Lip2)	71276.8	67152.0	67152.8	0.8	-Sig P	Trap 9	71
SACOL0078[e]	1-Phosphatidylinositol phosphodiesterase (Plc)	37086.7	34127.1	34128.0	0.9	-Sig P	Trap 6	90
Peptidoglycan Hydrolases								
SACOL0263[e]	Peptidoglycan hydrolase (LytM)	35067.4	31735.5	31736.5	1.0	-Sig P	Trap 4	66
SACOL0507[a, e]	N-acetylmuramoyl-L-alanine amidase (Sle1)	35835.7	33424.9				Trap 12	76
SACOL0723[1, e]	LysM domain protein	28186.8	11848.9	11849.0	0.1	-Sig P, N-terminal fragment	Trap 2	100
SACOL0723[2]	LysM domain protein	28186.8	13800.9	13801.8	0.9	-Sig P, C-terminal fragment	Trap 2	72
SACOL1062[1, e]	Bifunctional autolysin (Atl[1])	137334.9	134248.2	134249.5	1.3	-Sig P	Trap 9	99
SACOL1062[2]	Bifunctional autolysin (Atl[2])	137334.9	80786.9	80787.3	0.4	-Sig P, proteolytic processing	Trap 8	99
SACOL1062[3]	Bifunctional autolysin (Atl[3])	137334.9	53479.5	53479.7	0.2	-Sig P, proteolytic processing	Trap 9	99
SACOL2088[e]	SceD protein (SceD)	24096.0	21497.0	21497.7	0.7	-Sig P	Trap 3	76
SACOL2584[e]	Immunodominant antigen A (IsaA)	24203.2	21377.9	21377.5	-0.4	-Sig P	Trap 3	91
SACOL2666[e]	N-acetylmuramoyl-L-alanine amidase domain protein	69253.2	42250.8	42252.1	1.3	-Sig P, Protein degradation	Trap 6	89
Miscellaneous enzymes								
SACOL0303[e]	Acid phosphatase5-nucleotidase	33351.9	30184.3	30185.0	0.7	-Sig P	Trap 11	87
SACOL0860	Thermonuclease precursor (Nuc)	25119.9	18782.3	18782.0	-0.3	-Sig P	Trap 7	97
SACOL0962[e]	Glycerophosphoryl diester phosphodiesterase (GlpQ)	35310.7	32240.2	32241.0	0.8	-Sig P	Trap 10	100
SACOL1071[1b]	Chitinase-related protein	11344.8	8720.8	8721.0	0.2	-Sig P	Trap 3	91
SACOL1071[2b]	Chitinase-related protein	11344.8	8906.9	8907.1	0.2	-Sig P	Trap 3	89
SACOL1071[3b]	Chitinase-related protein	11344.8	9902.0	9901.6	-0.4	-Sig P	Trap 3	92
Surface adhesins								
SACOL0858[1]	Secretory extracellular matrix and plasma binding protein (Empbp[1])	38484.9	12890.5	12891.8	1.3	-Sig P, N-terminal fragment	Trap15	73

Table 1. *Contd.*

Secretory extracellular matrix and plasma ... ninal fragment Trap 15 53
SACOL0985ᵃ Surface protein, putative 15858.2 12859.7 12859.5 -0.3 -Sig P Trap 4 100

Protein	Function	Theor. mass	Predicted mass[†]	Observed mass[§]	Mass error	Modifications	Trap Location	MS/MS % Seq. Cov.*
Surface adhesins								
SACOL1164	Fibrinogen binding-related protein	12596.6	9592.1	9592.5	0.4	-Sig P	Trap 3	69
SACOL1168	Fibrinogen-binding protein (Efb)	18764.6	15850.1	15851.0	0.9	-Sig P	Trap 10	90
SACOL2002	Map protein (Map)	76945.2	73877.7				Trap 17	31
SACOL2019	SdrH protein, putative (SdrH)	46630.0	38083.8	38085.1	1.2	-Sig P, removal of C-terminal residues 377-419	Trap 3	85
SACOL2197[e]	Surface protein, putative	15447.5	12480.0	12480.1	0.1	-Sig P	Trap 3	100
SACOL2418[e]	IgG-binding protein (Sbi)	50070.2	47049.7	47049.3	-0.4	-Sig P	Trap 9	95
SACOL2660	Immunodominant antigen B (IsaB)	19370.2	15785.0	15785.3	0.3	-Sig P	Trap 3	100
Unknown Functions								
SACOL0270[e]	Staphyloxanthin biosynthesis protein, putative	33032.2	30379.1	30377.9	-1.2	-Sig P, -42 Da modification	Trap 12	59
Esxa	Virulence factor esxA (Esxa)	11306.2	10905.2	10905.2	0.2	-Met	Trap 3	100
SACOL0480	Hypothetical protein	11301.8	8380.3	8380.4	0.1	-Sig P	Trap 9	100
SACOL0669[e]	Conserved hypothetical protein	18594.2	15905.0	15905.4	0.4	-Sig P	Trap 6	100
SACOL0755[e]	Conserved hypothetical protein	16922.1	13449.9	13449.1	-0.8	-Sig P	Trap 7	79
SACOL0859	Hypothetical protein	17717.0	14812.5	14813.4	0.9	-Sig P	Trap 7	67
SACOL0908[e]	Hypothetical protein	20345.8	16271.1	16271.1	0.0	-Sig P	Trap 5	99
SACOL1166[e]	Hypothetical protein	15202.5	12304.0	12304.0	0.0	-Sig P	Trap 5	55
SACOL2179	Conserved hypothetical protein	32763.6	29134.0	29135.5	1.5	-Sig P	Trap 8	42
SACOL2291[1, e]	Staphylococcal secretory antigen ssaA2	29327.1	12211.3	12210.3	-1.0	-Sig P, N-terminal fragment	Trap 3	98
SACOL2291[2]	Staphylococcal secretory antigen ssaA2	29327.1	14294.5	14294.9	0.4	-Sig P, C-terminal fragment	Trap 3	97
SACOL2295[e]	Staphyloxanthin biosynthesis protein	17424.8	14755.7	14756.7	1.0	-Sig P, +16 Da modification	Trap 3	71
SACOL2557	Conserved domain protein	16870.3	12788.5	12789.5	1.0	-Sig P	Trap 6	61

[†] Predicted masses were calculated by subtracting the mass of signal peptide from the theoretical mass. For proteins with additional modifications recalculated predicted mass is presented in the table.

[§] Average mass of the protein from replicate analyses

* Net sequence coverage of the mature protein

[a] Proteins could not be identified by whole protein MS analysis

[b] Proteins with signal peptide cleavages at more than one site

[c] Proteins identified in the extracellular peptide extract by peptide MS/MS analysis and the reported protein mass was calculated from precursor ion mass

[d] Monoisotopic mass

[e] Proteins identified in other S. aureus strains[29]

[1-3] Numbers in the superscript indicate the different forms of the protein identified

Table 2. Observed signal peptide fragments and proposed cleavage sites to account for the fragments detected.[b]

Gene ID	Observed signal peptide fragments	Proposed cleavage sites for degradation of signal peptide
SACOL0480	VGVLATGVVGYGNQADA	MKFKKVLVATAM-↓VGVLATGVVGYGNQADA
SACOL0723[a]	FAITATSGAAAFLTHHDAQA	MKKLA-↓FAITATSGAAAFLTHHDAQA
SACOL0860	VLTLVVVSSLSSSANA	MTEYLLSAGICMAIVSILLIGMAISNVSKGQYAKRFFYFATSCLVLTL-↓VVVSSLSSSANA
SACOL0908	ALVLTTVGSGFHSSSNYNGINNVAKA	MNKKLLTRTLIAS-↓ALVLTTVGSGFHSSSNYNGINNVAKA
SACOL1062	LTLVGSAVTAHQVQA	MAKKFNYKLPSMVA-↓LTLVGSAVTAHQVQA
SACOL1164	AISLTVSTFAGESHA	MKKNFIGKSILSIA-↓AISLTVSTFAGESHA
SACOL1864	TILTSITGVGTTMVEGIQQTAKA	MNKNIIIKSIAAL-↓TILTSITGVGTTMVEGIQQTAKA
SACOL1868	TILTSVTGIGTTLVEEVQQTAKA	MNKNVVIKSLAAL-↓TILTSVTGIGTTLVEEVQQTAKA
SACOL2003	ANLLLVGALTDNSAKA	MVKKTKSNSLKKVATLAL-↓ANLLLVGALTDNSAKA
SACOL2088[a]	SLAVGLGIVAGNAGHEAHA	MKKTLLAS-↓SLAVGLGIVAGNAGHEAHA
SACOL2197	LGLLSTVGAALPSHEASA	MKLKSFVTATLA-↓LGLLSTVGAALPSHEASA
SACOL2291	AGFATIAIASGNQAHA	MKKIATATIAT-↓AGFATIAIASGNQAHA
SACOL2295[a]	ATTLTAGIGTALVGQAYHADA TAGIGTALVGQAYHADA	MKKLVT-↓ATTL-↓TAGIGTALVGQAYHADA
SACOL2418	TITLATMISNGEAKA	MKNKYISKLLVGAA-↓TITLATMISNGEAKA
SACOL2421	SVSLLAPLANPLLENAKA	MLKNKILTTTL-↓SVSLLAPLANPLLENAKA
SACOL2557	AVLFSADFTYQSVEQTHQSHA	MEYKKILIRLLIAF-↓AVLFSADFTYQSVEQTHQSHA
SACOL2584[a]	IMASSLAVALGVTGYAAGTGHQAHA SSLAVALGVTGYAAGTGHQAHA AVALGVTGYAAGTGHQAHA	MKKT-↓IMA-↓SSL-↓AVALGVTGYAAGTGHQAHA
SACOL2660	GTLIGVTVVENSAPTSKQAQA LIGVTVVENSAPTSKQAQA	MNKTSKVCVAATLAL-↓GT-↓LIGVTVVENSAPTSKQAQA

[a] Stable cleaved signal peptide also identified in the extracellular medium by LTQ-FT-MS/MS analysis
[b] Peptide Mascot scores are provided in Supplemental Table 7.

Table 3. Signal sequences, observed and predicted signal peptide cleavage sites of *S. aureus* COL proteins (secretory and cell wall associated proteins) identified in the present study.

Gene ID	Signal peptide	SPase I[b] (-3 to +1)	Obsd. Clvg. Site[c]	SignalP V3.0 NN	HMM	SignalP V2.0 NN	HMM	Predisi	SigCleave
SACOL0024	MKALLLKTSVWLVLLFSVMGLWQV **SNA**	A	27	27	27	27	29	27	27
SACOL0050	MNKNSKKKLDFLPNKLNKYSIRRFTVGTASILVGATLIFGVANDQ **AEA**	A	48	48	48	48	48	48	48
SACOL0078	MKKCIKTLFLSIILVVMSGWYHS **AHA**	S	26	26	26	26	26	26	26
SACOL0263	MEDVLYMKKLTAAAIATMGFATFTMAHQ **ADA**	A	31	31	31	29	31	31	31
SACOL0095	MKKKNIYSIRKLGVGIASVTLGTLLISGGVTPA **ANA**	A	36	36	36	36	36	36	36
SACOL0119	MKKLATVGSLIVTSTLVFSSMPFQN **AHA**	D	28	28	28	28	28	28	28
SACOL0270	MKKTILLTMTTLTLFSMSPNS **AQA**	Y	24	24	24	24	24	24	24
SACOL0303	MNKISKYIAIASLSVAVTVSAPQTTNST **AFA**	K	31	31	31	21	31	29	20
SACOL0317	MLRGQEERKYSIRKYSIGVVSVLAATMFVVSSHE **AQA**	S	37	37	37	37	37	37	37
SACOL0442	MFKKYDSKNSIVLKSILSLGIIYGGTFGIYPK **ADA**	S	35	35	35	none	35	none	none
SACOL0480	MKFKKVLVATAMVGVLATGVVGYGNQ **ADA**	K	29	29	29	29	29	none	19
SACOL0507	MQKKVIAAIIGTSAISAVAATQ **ANA**	A	25	25	25	25	25	25	25
SACOL0610	MINRDNKKAITKKGMISNRLNKFSIRKYTVGTASILVGTTLIFGLGNQE **AKA**	A	52	none	52	52	52	52	52
SACOL0669	MKKLLTASIIACSVVMGVGLVNTS **AEA**	A	27	27	27	27	27	27	27
SACOL0723	MKKLAFAITATSGAAAFLTHHD **AQA**	S	25	25	25	23	25	25	25
SACOL0755	MTVKNLFLGFVAVILTVCLIGLLILATNED **ALA**	K	33	33	33	33	33	33	33
SACOL0856	MNMKKKEKHAIRKKSIGVASVLVGTLIGFGLLSSKE **ADA**	S	39	39	39	39	39	39	39
SACOL0858	MKKKLLVLTMSTLFATQIMNSNH **AKA**	S	26	26	26	24	26	24	26
SACOL0859	MKRKVLVLTMGVICATQLWHSNH **ANA**	L	26	26	26	24	33	24	26
SACOL0860	MTEYLLSAGICMAIVSILLIGMAISNVSKGQYAKRFFYFATSCLVLTLVVVS SLSSS **ANA**	S	60	23	25	60	30	60	57
SACOL0886	MKKLISILLINIIILGVSNS **ASA**	Q	23	23	23	23	23	23	23
SACOL0887	MNKIFRILTVSLFFFTFLIKNNL **AYA**	D	26	26	26	26	26	26	26
SACOL0907	MYKRLFISHVILIFALILVISTPN **VLA**	E	27	27	27	27	27	27	27
SACOL0908	MNKKLLTRTLIASALVLTTVGSGFHSSSNYNGINNV **AKA**	S	39	24	39	24	39	30	28
SACOL0962	MTNSSKSFTKFMAASAVFTMGFLSVPT **AGA**	E	30	30	30	30	30	30	30
SACOL0985	MKLKSFITVTLALGMIATTGATVAGNE **VSA**	A	30	30	30	30	30	30	21
SACOL1056	MNSSCKSRVFNIISIIMVSMLILSLGAFANNNK **AKA**	D	36	36	36	36	36	36	29
SACOL1062	MAKKFNYKLPSMVALTLVGSAVTAHQ **VQA**	A	29	29	29	29	29	29	29
SACOL1071[a]	MNKLLQSLSALGVSATLVTPNLN **ADA**	T	26	26	24	26	24	26	24
SACOL1071[a]	MNKLLQSLSALGVSATLVTPN **LNA**	D	24	26	24	26	24	26	24
SACOL1140	MTKHYLNSKYQSEQRSSAMKKITMGTASIILGSLVYIGADSQQ **VNA**	A	46	39	46	46	46	none	46
SACOL1164	MKKNFIGKSILSIAAISLTVSTFAGE **SHA**	Q	29	29	29	29	29	29	29
SACOL1166	MKKNITKTIIASTVIAAGLLTQTND **AKA**	F	28	28	28	28	28	28	28
SACOL1168	MKNKLIAKSLLTLAAIGITTTTIAST **ADA**	S	29	29	29	29	29	29	26
SACOL1173	MKTRIVSSVTTTLLLGSILMNPV **ANA**	A	26	26	26	26	26	26	26
SACOL1864	MNKNIIIKSIAALTILTSITGVGTTMVEGIQQT **AKA**	E	36	36	36	36	36	29	24
SACOL1865	MNKNIIIKSIAALTILTSVTGVGTTVVEGIQQT **AKA**	E	36	36	36	36	36	36	none
SACOL1866	MNKNIIIKSIAALTILTSITGVGTTVVDGIQQT **AKA**	E	36	36	36	36	36	36	none
SACOL1867	MNKNIVIKSMAALAILTSVTGINAAVVEETQQI **ANA**	E	36	25	24	24	36	25	21
SACOL1868	MNKNVVIKSLAALTILTSVTGIGTTLVEEVQQT **AKA**	E	36	36	36	36	36	36	36
SACOL1869	MNKNVMVKGLTALTILTILTSLGFAENISNQPHSI **AKA**	E	38	38	38	38	38	30	23
SACOL1880	MKMKKLVKSSVASSIALLLLSNT **VDA**	A	26	26	26	26	26	26	26
SACOL1881	MFKKKMLAATLSVGLIAPLASPIQE **SRA**	N	28	28	28	28	28	28	28
SACOL2003	MVKKTKSNSLKKVATLALANLLLVGALTDNS **AKA**	E	34	34	34	34	34	34	34
SACOL2004	MIKQLCKNITICTLALSTTFTVLPAT **SFA**	K	29	29	29	29	29	29	29
SACOL2006[a]	MKNKKRVLIASSLSCAILLLSAAT **TQA**	N	27	27	29	29	29	29	29
SACOL2006[a]	MKNKKRVLIASSLSCAILLLSAATTQ **ANS**	A	29	27	29	29	29	29	29
SACOL2019	MSYHWFKKMLLSTSILILSSSSLGLATHT **VEA**	K	32	32	32	32	32	32	56
SACOL2088	MKKTLLASSLAVGLGIVAGNAGHE **AHA**	S	27	27	27	23	27	27	27
SACOL2179	MKKIFVIITTLLAVAIIIGSIIMVVFSQRQ **AQT**	F	33	31	31	31	29	31	60
SACOL2197	MKLKSFVTATLALGLLSTVGAALPSHE **ASA**	D	30	30	30	30	30	30	30
SACOL2291	MKKIATATIATAGFATIAIASGNQ **AHA**	S	27	27	27	27	27	27	27
SACOL2295	MKKLVTATTLTAGIGTALVGQAYH **ADA**	A	27	27	27	27	27	22	27
SACOL2418	MKNKYISKLLVGAATITLATMISNGE **AKA**	S	29	29	29	27	29	29	29
SACOL2419	MIKNKILTATLAVGLIAPLANPFIEI **SKA**	E	29	29	29	29	29	29	29
SACOL2421	MLKNKILTTTLSVSLLAPLANPLLEN **AKA**	A	29	29	29	29	29	29	29
SACOL2422	MKMNKLVKSSVATSMALLLLSGT **ANA**	E	26	26	26	26	26	26	26
SACOL2557	MEYKKILIRLLIAFAVLFSADFTYQSVEQTHQ **SHA**	A	35	35	35	26	35	none	19
SACOL2584	MKKTIMASSLAVALGVTGYAAGTGHQ **AHA**	A	29	29	29	29	29	29	29
SACOL2660	MNKTSKVCVAATLALGTLIGVTVVENSAPTSKQ **AQA**	A	36	36	36	27	36	27	36

Table 4. Amino acid residues around the confirmed SPase I cleavage site of *S. aureus* COL proteins. [34] [34]

R

A	47 (77%)	K	14 (23%)	A	59 (97%)	A	19 (31%)
V	6 (10%)	N	11 (18%)	S	1 (2%)	S	12 (20%)
S	6 (10%)	Q	8 (13%)	T	1 (2%)	E	11 (18%)
T	1 (2%)	H	7 (12%)			K	5 (8%)
L	1 (2%)	D	8 (13%)			D	4 (7%)
		S	3 (5%)			Q	2 (3%)
		E	3 (5%)			T	2 (3%)
		L	2 (3%)			F	2 (3%)
		F	2 (3%)			N	2 (3%)
		Y	1 (2%)			Y	1 (2%)
		G	1 (2%)			L	1 (2%)
		R	1 (2%)				

A

MTEYLLSAGICMAIVSILLIGMAISNVSKGQYAKRFFYFATSCLVLTLVV VSSLSSSANA **SQTDNGVNRSGSEDPTVYSATSTK K HKEPATLIKAIDG DTVKLMY KGQPMTFRLLLVDTPETK**HPKKGVEK**YGPEASAFTKKMVE**NAK**K IEVEFDKGQRTDKYGRGLAYIYADGKMVNEALVRQGLAKVAYVYKPNNTH**

B

	SignalP V3.0		SignalP V2.0		Predisi	SigCleave
	NN	HMM	NN	HMM		
Predicted Cleavage Site	60	30	23	25	60	57
Predicted Mass (Da)	18782.3	22038.0	22723.8	22523.6	18782.3	19038.5

C
Thermonuclease (Nuc)

D

E
Thermonuclease (Nuc)
18782.0
β-Hemolysin (Hlb)
33742.8

F
SQTDNGVNRSGSEDPTVYSATSTK
Residue 61-84

Figure 1. Experimental strategy for identification of Nuc in Trap 7. (A) Combined sequence coverage map of Nuc from trypsin and Glu-C digestion, the underlined amino acids were identified. Shaded region corresponds to the signal sequence. (B) Signal peptide cleavage site predictions and corresponding predicted masses for Nuc. (C) Total ion chromatogram of Trap 7 containing a peak corresponding to Nuc; (D) Raw spectrum of Nuc showing the charge state distribution. Asterisks show charge state distribution of Hlb; (E) Deconvoluted mass spectrum; (F) MS/MS spectrum of N-terminal peptide SQTDNGVNRSGSEDPTVYSATSTK of mature Nuc.

Figure 2. A) LTQ-FT-MS/MS of Hld[1], the asterisked residue (M1) is the site of modification (formylation and oxidation). B) LTQ-FT-MS/MS of Hld[2].

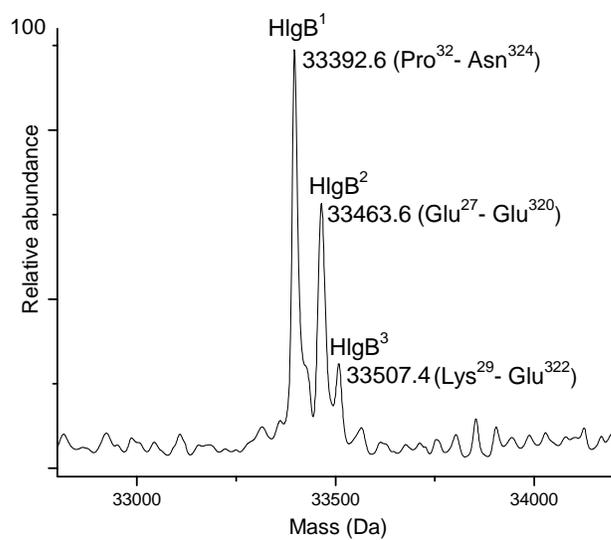**Figure 3.** Deconvoluted spectrum showing C-terminal truncated form of LukE, corresponding sequence is shown in parenthesis.

Figure 4. Deconvoluted spectrum of HlgB showing three truncated forms, HlgB[1], HlgB[2] and HlgB[3]. Corresponding sequences are shown in parenthesis.
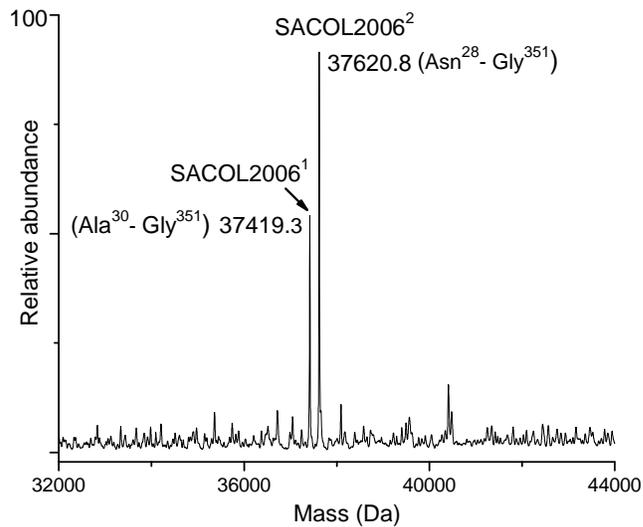
Figure 5. Deconvoluted mass spectrum of SACOL2006 suggesting signal peptide processing at two sites ; SACOL2006 [1] resulted from signal peptide cleavage at position 29 and SACOL2006 [2] resulted from signal peptide cleavage at position 27.

Figure 6. Identification of sequencing error in SplD. (A) Whole protein mass spectrum of co-eluting SplF and SplD. (B) Multiple sequence alignment (residues 40-70) of SplD from *S. aureus* COL and other strains of *S. aureus*. Suspected sequencing error is shown in the box. An asterisk represents identity among the aligned residues and two dots represent strong similarity. (C) LTQFT-MS/MS spectrum of SplD peptide consisting of residues 42-74 that confirm $Gln^{68} \rightarrow His^{68}$.
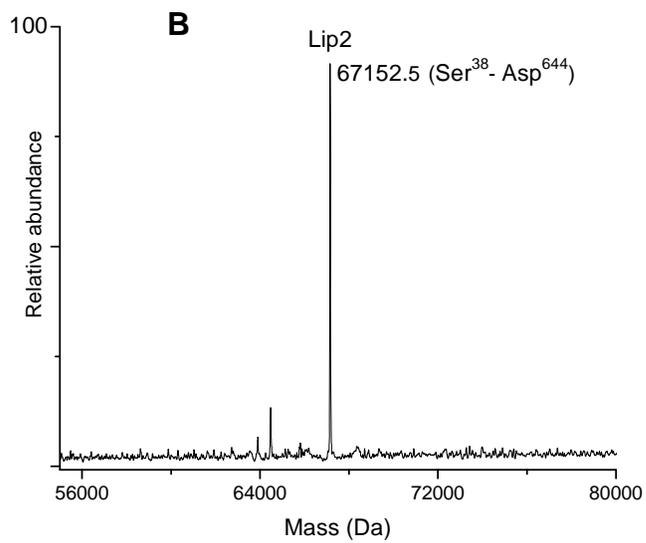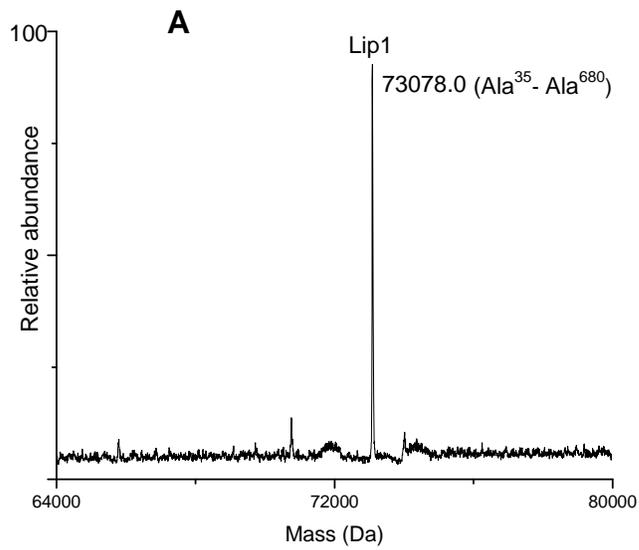
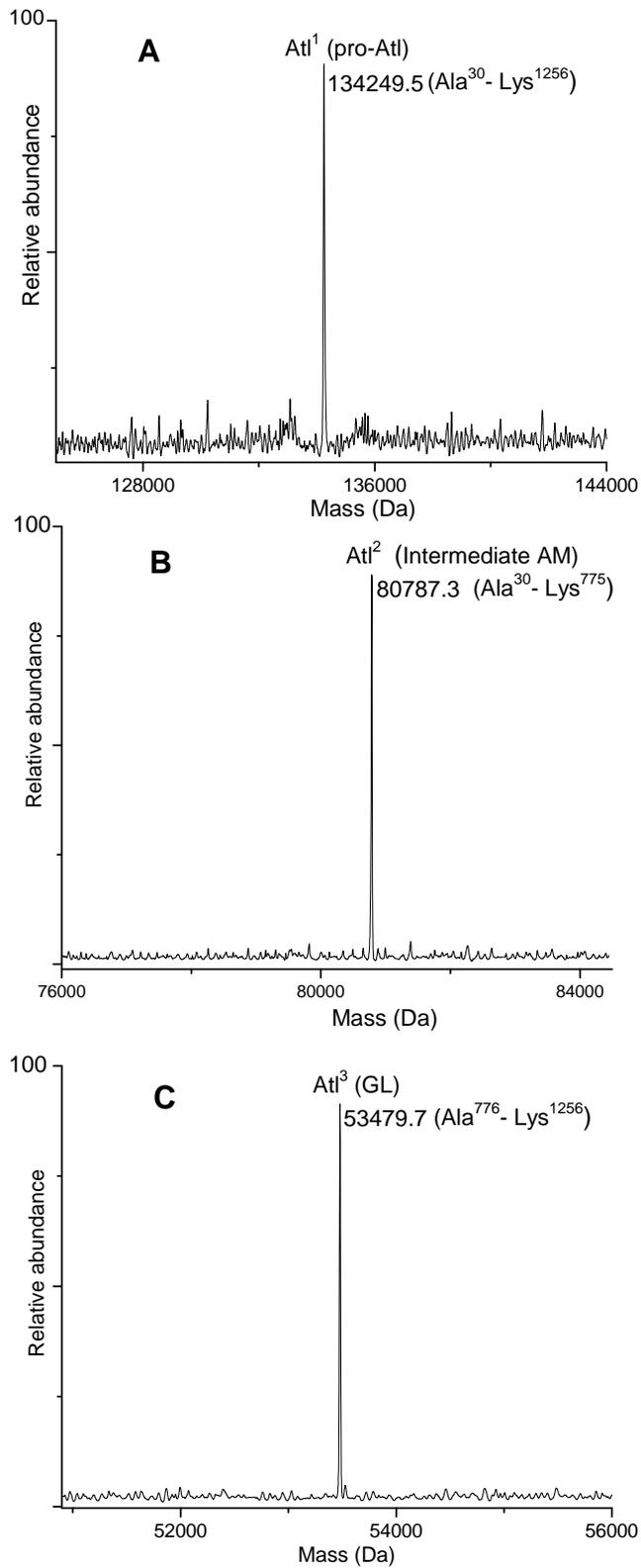**Figure 7.** Deconvoluted mass spectra showing pro-enzyme forms of (A) Lip1 and (B) Lip2

Figure 8. Deconvoluted mass spectra showing the three Atl gene products identified in the present study. (A) Atl[1], pro-Atl form; (B) Atl[2], intermediate AM form; (C) Atl[3], mature GL form.
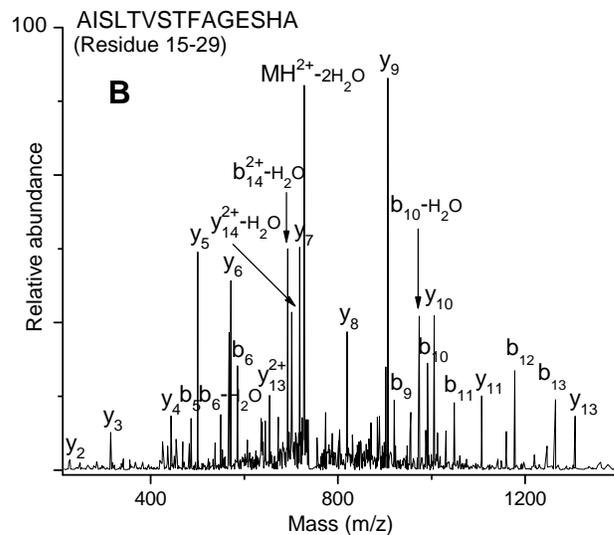
Figure 9. (A) Deconvoluted mass spectrum showing three forms of SACOL1071 formed as a result of signal peptide processing at different sites; SACOL1071[1], signal peptide cleavage at position 26; SACOL1071[2], signal peptide cleavage at position 24; SACOL1071[3], signal peptide cleavage at position 14. (B) LTQ-FT-MS/MS spectrum of N-terminal peptide of SACOL1071[3].
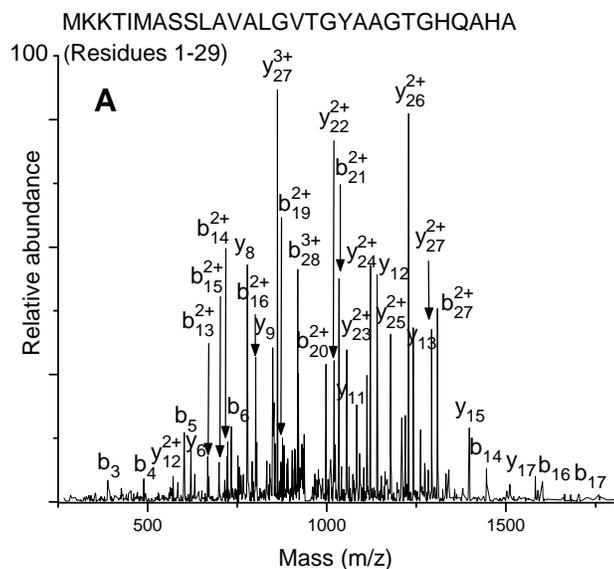
Figure 10. (A) LTQ-FT-MS/MS spectrum of stable cleaved signal peptide of Isa and (B) LTQ-FT-MS/MS spectrum of signal peptide fragment of SACOL1164.
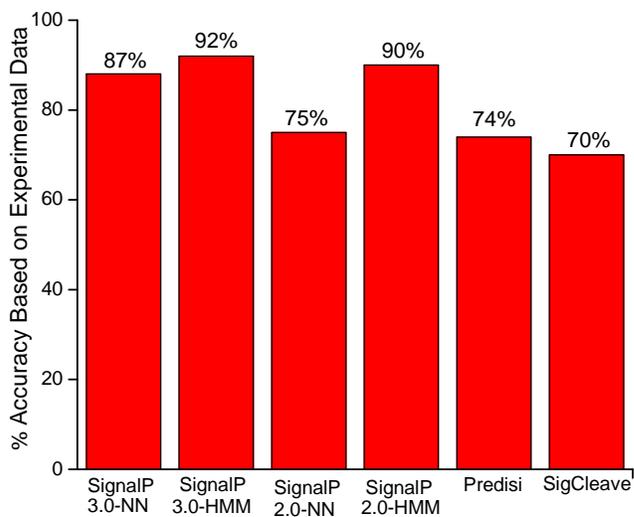

Figure 11. Prediction accuracies of various signal peptide prediction programs.