LEBANESE AMERICAN UNIVERSITY

Towards *S. aureus* MALDI-typing:
Genotypic Correlation and Shotgun Proteomic Identification

By

Dania Bassam Awad

A thesis

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Molecular Biology

School of Arts and Sciences
April 2014

© 2014

Dania B. Awad

All Rights Reserved

Lebanese American University

School of Arts and Sciences ; Byblos Campus

THESIS APPROVAL FORM

Student Name: Dania B. Awad I.D. #: 200902633

Thesis Title : Towards *S. aureus* MALDI-typing: Genotypic Correlation and Shotgun Proteomic Identification

Program: M.S. Molecular Biology

Department: Natural Sciences

School: Arts and Sciences

The undersigned certify that they have examined the final electronic copy of this thesis and approved it in Partial Fulfillment of the requirements for the degree of:

Master of Science in the major of Molecular Biology

Thesis Advisor's Name Dr. Brigitte Wex Signature Date 24.04.2014

Committee Member's Name Dr. Sima Tokajian Signature Date 25.04.2014

Committee Member's Name Dr. Roy Khalaf Signature Date 28/4/14

THESIS COPYRIGHT RELEASE FORM

LEBANESE AMERICAN UNIVERSITY NON-EXCLUSIVE DISTRIBUTION LICENSE

By signing and submitting this license, you (the author(s) or copyright owner) grants to Lebanese American University (LAU) the non-exclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video. You agree that LAU may, without changing the content, translate the submission to any medium or format for the purpose of preservation. You also agree that LAU may keep more than one copy of this submission for purposes of security, backup and preservation. You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright. If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant LAU the rights required by this license, and that such third-party owned material is clearly identified and acknowledged within the text or content of the submission. IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN LAU, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT. LAU will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

Name: Dania AlAwad

Signature:

Date: 29/4/2014

PLAGIARISM POLICY COMPLIANCE STATEMENT

I certify that:

- I have read and understood LAU's Plagiarism Policy.
- I understand that failure to comply with this Policy can lead to academic and disciplinary actions against me.
- This work is substantially my own, and to the extent that any part of this work is not my own I have indicated that by acknowledging its sources.

Name: Dania Awad

Signature:

Date: 29/4/2014

To My Loving Parents, I dedicate this work.

ACKNOWLEDGMENTS

“There is not a more pleasing exercise of the mind than gratitude. It is accompanied with such an inward satisfaction that the duty is sufficiently rewarded by the performance” Joseph Addison. None of us got to where we are alone and I couldn’t have complete this work on my own. There are a number of people to whom I am greatly indebted in the journey of obtaining my Master’s degree. First and foremost, I thank God for his uncounted blessings; for giving me strength, patience, health and the love and support of dear family and friends.

“The dream begins with a teacher who believes in you, who tugs and pushes and leads you to the next plateau, sometimes poking you with a sharp stick called truth” Dan Rather. Thank you doesn’t seem sufficient but it is said with appreciation and respect to my thesis advisor, Dr. Brigitte Wex for her guidance, support, encouragement, constant motivation, understanding and patience through reading my numerous drafts.

I warmly thank Dr. Sima Tokajian, for her valuable advice, constructive criticism, extensive discussions around my work, and her participation in my thesis committee. I extend my thanks to Dr. Roy Khalaf for his valued friendship, providing a stimulating and fun filled environment, spending his precious times reading my thesis and giving his valuable suggestions and taking part in my thesis committee.

This thesis has been kept on track and been seen through to completion with the support and encouragement of numerous people including my well-wishers, my friends,

and colleagues. My Special thanks are due to Mr. El-Hage, Ms. Harastani, Mr. Kassab, Ms. Kasbo, Mr. Younes,

I take this opportunity to sincerely acknowledge Ms. Maya Farah, our Lab coordinator, who contributed in many ways to the success of this study and made it an unforgettable experience for me. This list is incomplete without acknowledging Mrs. Rana Najjar, I can see the good shape of my thesis because of her help and suggestions in formatting my thesis.

“We can only be said to be alive in those moments when our hearts are conscious of our treasures” Thornton Wilder. My biggest treasure is my family, without whom I cease to exist. I offer my utmost form of gratitude to my father, Bassam Awad for his unconditional love and support and for being there for me in my utmost time of need. Words are short to express my deep sense of gratitude towards my dear mother, Jamal Hourri, for her continuous support, love and ardent devotion. During the inevitable ups and downs of conducting my research, I owe my gratitude and appreciation to my beloved brother Hadi Awad, for being my best friend. Last but not least, I would like to pay high regards to Ms. Hania Kassabiyyeh who willingly devoted so much time in giving guidance to me and mostly for being my second mother.

At the end of my thesis, it is a pleasant task to express my thanks to all those who made my journey harder for “In the middle of difficulty lies opportunity” Albert Einstein.

I would like to end my acknowledgments with a famous quote from B.B. King which was a source of inspiration to me in perusing my higher education, “The beautiful thing about learning is that nobody can take it away from you”.

Towards *S. aureus* MALDI-typing:

Genotypic Correlation and Shotgun Proteomic Identification

Dania Awad

Abstract

Staphylococcus aureus is possibly the most common human pathogen causing approximately one third of all bacteria-related diseases. *S. aureus*' adaptable characteristics and increasing antibiotic resistance allows it to thrive in both nosocomial and community settings. Identification of *S. aureus* is mainly performed via phenotypic methods and costly genotypic methods. Recently, Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) has been introduced into clinical microbiology as a means of bacterial identification to the species level. In this study, we compared MALDI protein fingerprints to genomic data of 25 *S. aureus* isolates, an *S. epidermidis* strain and an *E. coli* strain finding only geographic separation as visualized in a dendrogram. The absence of correlation between genomic and proteomic data to the strain level is owed to genomic mutations leading to protein isomers being falsely identified as different proteins at the MALDI level. Tryptic digests of protein extracts from one isolate were further characterized by NanoFlow Liquid Chromatography coupled with MALDI-TOF MS/MS and 94 proteins were successfully identified. Three candidate proteins are proposed for MALDI-TOF MS typing, which marks a step towards fast identification of *S. aureus* to the strain level via MALDI-TOF MS.

Keywords: Proteomics, *Staphylococcus aureus*, Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS), NanoFlow Liquid Chromatography, typing.

TABLE OF CONTENTS

Chapter	Page
I – Introduction	1 - 7
1.1 – <i>Staphylococcus aureus</i>	1
1.1.1 – <i>S. aureus</i> Virulence	2
1.2 – Bacterial Identification	3
1.2.1 – Typing of <i>S. aureus</i>	3
1.3– Matrix Assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry	4
1.3.1 – Bacterial Identification via MALDI-TOF MS	4
1.3.2 – <i>S. aureus</i> Typing via MALDI-TOF MS	5
1.3.3 – Typing via MALDI-TOF MS and Bruker Biotyper	5 - 6
1.4 – Protein Identification	6
1.4.1 – NanoFlow LC MS/MS	7
II – Materials and Methods	8 - 19
2.1 – Clinical Isolate	8 - 9
2.2 – Protein Extraction	10
2.3 – Protein Quantification	10
2.4 – In-Solution Tryptic Digestion	11
2.5 – Nano-Flow Liquid Chromatography	11
2.5.1 – NanoFlow LC Set Up	11 – 12
2.5.2 – Sample Preparation	12
2.5.3 – NanoFlow LC Run Conditions	12 - 13
2.6 – MALDI-TOF MS Analysis	13
2.6.1 – Plate Preparation	13

2.6.2 – Calibration and Error Calculation	13 - 16
2.6.3 – Protein Analysis	17
2.6.4 – Shotgun Proteomics – Peptides Analysis	17 - 18
2.7 – Bioinformatics Analysis	18
2.7.1 – Protein Analysis	18
2.7.2 – Peptides Analysis and Protein Identification	18 - 19
 III – Results	 20 - 45
 3.1 – Protein Quantification	 20 - 21
3.2 –Proteomic Fingerprinting	22 - 28
3.3 - Peak Analysis	29 - 30
3.4 - Correlating Proteomic Fingerprints and Genotypic Characteristics	31
3.5 – NanoFlow LC MS/MS Optimization	32 - 34
3.6 – Identification of <i>S. aureus</i> Proteins	35 - 45
 IV – Discussion	 46 - 53
 4.1 – Assessment of Proteomic Profiles	 47 - 49
4.2 – Evaluation of Protein Identification Method	49 - 51
4.3 – Protein Identification Audit	51 - 53
4.4 – Evaluation of Protein Extraction Method	53
 V – Conclusion	 54 - 55
 VI – References	 56 - 60
 Appendix	 61 – 71

LIST OF TABLES

Table	Page
Table 2.1. A detailed view of characteristics of <i>S. aureus</i> isolates utilized in this study	9
Table 2.2. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: JS 03, JS08, JS 10, JS 17 and JS 22.	14
Table 2.3. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: JS 50, JS 52, JS 54, JS 56 and JS 60.	14
Table 2.4. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: JS 61, JS 65 and JS 85.	15
Table 2.5. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: LS 16, LS 24, LS 34, and LS 36.	15
Table 2.6. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: LS 79, LS 89, LS 94, LS 98 and LS 103.	15
Table 2.7. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: LS 57, LS 125 and LS 126.	15
Table2.8. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for strains <i>S. epidermidis</i> and <i>E. coli</i> .	16
Table2.9. 4700 Cal Mix standards used for MALDI TOF MS calibration in reflector mode run for LS 16 peptide analysis.	16
Table 2.10. Fragmentation pattern of 4700 Cal Mix peptide with mass of 1296.4681.	16
Table 3.1. The frequency of the highest-occurring protein peaks detected by MALDI-TOF MS in linear mode of formic acid protein extracts of <i>S. aureus</i> isolates.	30
Table 3.2. Proteins identified from trypsin-digested LS 16 formic acid extract by MASCOT search engine.	41
Table 3.3. Proteins identified to species other than <i>S. aureus</i> and their corresponding	

<i>S. aureus</i> homologues. Homologues are obtained from non-redundant (nr) NCBI	
<i>S. aureus</i> database blast search.	42
Table 3.4. Proteins identified by non-redundant (nr) NCBI <i>S. aureus</i> database search for MASCOT unassigned peptide sequences (Part 1).	43
Table 3.5. Functions and Molecular weights of 59 identified proteins, obtained from Uniprot or NCBI nr databases (Part 1).	44 - 45
Table 7.1. Trypsin-digested BSA Standard and trypsin-digested LS 16 inclusion lists utilized in MALDI-TOF MS/MS Interpretation method.	63
Table 7.2. CHCA matrix exclusion list utilized in MALDI-TOF MS/MS Interpretation method.	64
Table 7.3. Fractionation pattern of trypsin-digested BSA Standard. Column 1 show BSA peptide peaks obtained from BioEngland Labs Manual of trypsin-digested BSA Standard. Columns A1 to B10 refer to one hour Nanoflow LC fractions analyzed by MALDI-TOF MS in reflector mode in mass range of 400-4,000 m/z, prior to MALDI-TOF MS/MS analysis.	65
Table 7.4. Proteins identified from trypsin-digested LS 16 formic acid extract by MASCOT search engine.	67 - 69
Table 7.5. Proteins identified in this study that were not assigned a molecular mass and a function.	70

LIST OF FIGURES

Figure	Page
Figure 3.1. BCA standard curved from which protein concentrations of isolates LS 89, LS 94, LS 98, LS 103, LS 125 and LS 126 were extrapolated from.	20
Figure 3.2. Protein Concentrations in µg/mL for each of the 25 <i>S. aureus</i> isolates in this study.	21
Figure 3.3. MALDI-TOF MS spectra for formic acid protein extracts of isolates JS 03 (top), JS 08 (middle) and JS 10 (bottom).	22
Figure 3.4. MALDI-TOF MS spectra for formic acid protein extracts of isolates JS 17 (top), JS 22 (middle-top), JS 50 (middle-bottom) and JS 52 (bottom).	23
Figure 3.5. MALDI-TOF MS spectra for formic acid protein extracts of isolates JS 54 (top), JS 56 (middle-top), JS 60 (middle-bottom) and JS 61 (bottom).	24
Figure 3.6. MALDI-TOF MS spectra for formic acid protein extracts of isolates JS 65 (top), JS 85 (middle-top), LS 16 (middle-bottom) and LS 24 (bottom).	25
Figure 3.7. MALDI-TOF MS spectra for formic acid protein extracts of isolates LS 34 (top), LS 36 (middle-top), LS 57 (middle-bottom) and LS 79 (bottom).	26
Figure 3.8. MALDI-TOF MS spectra for formic acid protein extracts of isolates LS 89 (top), LS 94 (middle-top), LS 98 (middle-bottom) and LS 103 (bottom).	27
Figure 3.9. MALDI-TOF MS spectra for formic acid protein extracts of isolates LS 125 (top), LS 126 (middle-top), <i>S. epidermidis</i> strain (middle-bottom) and <i>E. coli</i> strain (bottom).	28
Figure 3.10. Dendrogram representation of hierarchical cluster analysis of <i>S. aureus</i> isolates (LS# and JS#), <i>S. epidermidis</i> (SE) and <i>E. coli</i> (EC) strains along with their corresponding genotypic characteristics.	31
Figure 3.11. MALDI-TOF MS spectrum acquired in reflector mode (mass range: 400-4000 m/z) for Trysin-digested BSA MS Standard.	32
Figure 3.12. Nanoflow LC chromatogram of trypsin-digested BSA standard. Injection	

volume: 8 μ L. Linear gradient of 2-45% solvent B. Detection wavelength: 210nm.	
Time-based fractionation: 3min.	33
Figure 3.13. MALDI-TOF MS/MS spectrum for peptide with m/z value of 1,480.07 from Trysin-digested BSA Standard.	33
Figure 3.14. Mascot score histogram showing the result of MS/MS ion search in SwissProt database for Trysin-digested BSA MS Standard after fractionation by NanoFlow LC. Individual ions scores > 37 indicate identity or extensive homology ($p < 0.05$). Protein scores are the sum of individual ion scores.	34
Figure 3.15. MALDI-TOF MS spectra acquired in reflector mode for trypsin-digested LS 16 peptides in a mass range of 400-4,000 m/z (top) and 400-1,000 m/z (bottom).	35
Figure 3.16. MALDI-TOF MS spectra acquired in reflector mode for trypsin-digested LS 16 peptides in a mass range of 1,000-2,000 m/z (top) and 2,000-3,000 m/z (bottom).	36
Figure 3.17. Nanodrop Plot from which LS 16 peptides concentration was estimated to nearly 90 mg/mL. 1 mm Absorbance = 1 mg/ mL.	36
Figure 3.18. Nanoflow LC chromatogram of trypsin-digested LS 16. Injection volume: 8 μ L. Linear gradient of 2-45% solvent B. Detection wavelength: 210nm. Time-based fractionation: 3min.	37
Figure 3.19. MALDI-TOF MS/MS spectrum for peptide with m/z value of 1,378.72 from Trysin-digested LS 16 protein extract.	38
Figure 3.20. Mascot score histogram showing the result of MS/MS ion search in SwissProt database for Trysin-digested LS 16 protein extract after fractionation by NanoFlow LC. Individual ions scores > 31 indicate identity or extensive homology ($p < 0.05$). Protein scores are the sum of individual ion scores.	38
Figure 3.21. Mascot score histogram showing the result of MS/MS ion search in NCBI nr database for Trysin-digested LS 16 protein extract after fractionation by NanoFlow LC. Individual ions scores > 50 indicate identity or extensive homology ($p < 0.05$). Protein scores are the sum of individual ion scores.	39

Figure 7.1. BCA standard curved from which protein concentrations of isolates JS 03, JS 08, JS 10, JS 17, JS 22, JS 50 and JS 52 were extrapolated from.	61
Figure 7.2. BCA standard curved from which protein concentrations of isolates JS 54, JS 56, JS 60, JS 61, JS 65 and JS 85 were extrapolated from.	61
Figure 7.3. BCA standard curved from which protein concentrations of isolates LS 16, LS 24, LS 34, LS 36, LS 57 and LS 79 were extrapolated from.	62
Figure 7.4. BCA standard curved from which protein concentrations of <i>S. epidermidis</i> , <i>E. coli</i> and <i>S. aureus</i> isolate LS 16 were extrapolated.	62
Figure 7.5. Nanoflow LC chromatogram of trypsin-digested LS 16. Injection volume: 8 µL. Five hour linear gradient of 2-45% solvent B. Detection wavelength: 210 nm. Time-based fractionation: 3min.	66
Figure 7.6. Two dimensional electrophoresis (2-DE) representative gel of <i>S. aureus</i> isolate LS-16 performed by El Hage (2013).	73

GLOSSARY

ACN: Acetonitrile

BLAST: Basic Local Alignment Search Tool

CHCA: α -Cyano-4-Hydroxy-Cinnamic Acid

CID: Collision Induced Dissociation

HPLC: *High Performance Liquid Chromatography*

MALDI-TOF TOF: Matrix Assisted Laser Desorption/Ionization Time of Flight Time of Flight

MGF: Mascot Generic Format

MLST: Multi-locus sequence typing

MRSA: Methicillin-resistant *Staphylococcus aureus*

MSSA: Methicillin-sensitive *Staphylococcus aureus*

M/z: mass to charge ratio

PFGE: Pulsed-field Gel Electrophoresis

PVL: Panton-Valentine leukocidin

SCC: Staphylococcal Chromosomal Cassette

ST: Sequence Type

TFA: Trifluoroacetic Acid

TPCK: L -1-Tosylamide-2-phenylethyl chloromethyl ketone

TSA: Tryptone Soy Agar

Chapter One

Introduction

1.1 *Staphylococcus aureus*

Staphylococcus aureus is a gram-positive, facultative anaerobe bacterium. All members of the genus *Staphylococcus* are non-motile, non-spore-forming cocci, which form grape-like clusters upon division. The species' name *aureus* refers to the golden colonies that appear on solid agar. Nearly all strains of *S. aureus* are catalase-positive and coagulase-positive, important characteristics exploited in *S. aureus* identification (Harris, Foster, & Richards, 2002).

S. aureus is common in both community and nosocomial (hospital-related) settings. The ability of *S. aureus* to thrive in a community setting is related to its adaptable characteristics, since *S. aureus* is part of the normal human flora. On the other hand, the antibiotic resistance mechanisms have opened the path for *S. aureus* to enter and survive in nosocomial settings (Zetola, Francis, Nuermberger, & Bishai, 2005). *S. aureus* causes one of the most common nosocomial infections; its complications have high mortality rates due to bacteremia (Lark et al., 2001). Due to the increasing use of antibiotics, multiple drug resistant *S. aureus* strains have emerged; making it a major pathogen of interest (Lowy, 2000). Also, due to different selective pressures between nosocomial and community settings, isolates of each differ notably (Zetola, Francis et al. 2005).

1.1.1 *S. aureus* Virulence

S. aureus is possibly the most common human pathogen; it causes around one third of all bacterial diseases, which range from mild pus-forming superficial lesions (boils, styes and furuncles) to more serious colonization of internal organs (mastitis, phlebitis, meningitis, urinary tract infections, osteomyelitis and endocarditis).

Pathogenicity is due to the many potential virulence factors of *S. aureus*, which include: (1) surface proteins that promote colonization of host tissues; (2) invasins that promote bacterial spread in tissues (leukocidin, kinases, hyaluronidase); (3) surface factors that inhibit phagocytic engulfment (capsule, Protein A); (4) biochemical properties that enhance their survival in phagocytes (carotenoids, catalase production); (5) immunological membranes (hemolysins, leukotoxin, leukocidin); (6) exotoxins that damage host tissues or provoke symptoms of disease (SEA-G, TSST, ET); and (7) inherent and acquired resistance to antimicrobial agents (Todar 2009).

Secreted virulence factors play a major role in host pathogenesis. Type VII secretion system (T7SS) in *S. aureus* secretes virulence factors EsxA and EsxB to the extracellular surroundings. While EsxB is required for EsxA synthesis, EsxA (Ess extracellular A) is involved in abscess formation process of pathogenesis (Burts, Williams, DeBord, & Missiakas, 2005); (Sundaramoorthy, Fyfe, & Hunter, 2008). Additionally, basic *S. aureus* cell physiology is linked to its pathogenicity; it was recently shown that tricarboxylic acid cycle activity contributes to its virulence (Meyer, Liebeke, & Lalk, 2010).

1.2 Bacterial Identification

Successful elimination of a suspected bacterial infection in a patient requires the identification of the pathogen at hand and employing the best antibiotic treatment available, accordingly (Carbonnelle et al., 2011). In addition to pathogen identification, microbial typing methods offer us a better understanding of microbial population epidemiology and diversity (Boers, van der Reijden, & Jansen, 2012). The choice of typing method is usually dependent on the aim of investigation, and other factors including: discriminatory power, reproducibility, technical difficulty, time, throughput and cost (Boers et al., 2012).

In general, bacterial species are identified using conventional methods, based on physiology as well as biochemistry. These methods rely on cell structure, cellular metabolism, and pigments for identification. However, these methods lack both specificity and sensitivity, and are time-consuming. One limitation of these methods is caused by differential expression of phenotypic features among different isolates of the same species (Yang et al., 2009). Moreover, human error associated with the interpretation of some tests may compromise objectiveness (Hussain et al., 2008).

1.2.1 Typing of *S. aureus*

Strain identification of *S. aureus* is mainly performed via pulse field gel electrophoresis (PFGE), *spa*, Multilocus sequence typing (MLST), multilocus variable number of tandem repeat analysis (MLVA), and repeat-element PCR typing (Tang, 2009). These genetic typing methods are distinguished by their high discriminatory power, however, these methods remain cost intensive, time consuming, and tedious (Yang et al. 2009). Accordingly, an increasing need arises for methods with high sensitivity, specificity, speed, reproducibility and reliability for *S. aureus* identification (Yang et al. 2009).

1.3 Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry is a gentle ionization method capable of detecting and characterizing biopolymers such as peptides, proteins, oligosaccharides and nucleotides (Kaufmann, 1995).

1.3.1 Bacterial Identification via MALDI TOF MS

Bacterial identification via MALDI-TOF MS is based on the examination of sample protein contents, identification and detection of protein biomarkers that are unique to a specific bacterium (Cherkaoui et al., 2010). Mass spectrometry is a promising technique for routine bacterial identification in clinical microbiology (Anhalt & Fenselau, 1975); (Carbonnelle et al., 2007); (Degand et al., 2008); (Seng et al., 2009).

MALDI-TOF MS has a rapid turnaround time, low sample volume requirements and modest reagent costs. Samples can be prepared in a few minutes while data acquisition can be acquired in a few seconds. Identification of different bacterial species by MALDI-TOF MS has been proven to be successful, rapid, and accurate (S. Y. Hsieh et al., 2008) (Rajakaruna et al., 2009) (Sauer & Kliem, 2010) (Seng et al., 2009). By comparing MALDI-TOF MS results to multilocus variable-number tandem repeat analysis (MLVA) data, Lista and Reubsaet were able to correctly identify *Brucella* species in 99.3% of 152 *Brucella* isolates using MALDI-TOF MS (Lista et al., 2011).

1.3.2 *S. aureus* Typing via MALDI-TOF MS

Liu and Du developed a universal method for *S. aureus* identification by MALDI-TOF MS that is rapid, simple, and highly reproducible. This identification method was able to discriminate between *E. coli* and *S. aureus*, based on the unique spectral profiles generated by each, in a mass range of 1,000 to 20,000 Da (Liu, Du, Wang, & Yang, 2007). Furthermore, Du et al. was capable of differentiating between methicillin-resistant *Staphylococcus aureus* (MSSA) and methicillin-susceptible *S. aureus* (MRSA) with the use of MALDI-TOF MS (Du, Yang, Guo, Song, & Wang, 2002).

To take appropriate measures during an epidemiological outbreak, it is of great importance to differentiate among infections resulting from strains of the same clonal complex and strains of different clonal origin. For this reason, clinical isolates of *S. aureus* were collected during an epidemiological outbreak from a single hospital ward. Upon comparison of the isolates' spectra, a common peak pattern was observed. Moreover, additional *S. aureus* isolates were collected from different hospital wards and upon the comparison of the isolates' spectra, significant variations were apparent between spectra of isolates collected from different hospital wards (Bernardo et al., 2002).

1.3.3 Typing via MALDI-TOF MS and Bruker Biotyper

When MALDI-TOF MS and Bruker biotyper are coupled, the result is rapid identification of bacterial isolates. While MALDI-TOF MS generates spectral profiles of bacterial proteins, Bruker biotyper references these spectra to a compiled database, thus allowing bacterial identification to the genus and species level. Bruker Biotyper work is based on pattern matching of MALDI-generated spectra to library spectra

(Saffert et al., 2011). Additionally, Bruker Biotyper is capable of constructing a hierarchical dendrogram, based on spectral similarities and differences among submitted isolates (Tani et al., 2012).

In 2010, Szabados and Woloszyn employed Bruker Biotyper in the identification of 602 molecularly defined *S. aureus* isolates and 412 isolates that belong to different *Staphylococci* species other than *S. aureus*. Identification accuracy was 100% for both pools of isolates. Continual growth, by user supplementation and software updates of Bruker Biotyper library, continue to better MALDI-TOF MS and Biotyper-based bacterial identification (Theel, Hall, Mandrekar, & Wengenack, 2011).

1.4 Protein Identification

Identification of *S. aureus* proteome allows for better understanding of MALDI-TOF MS spectra, thus improving MALDI-TOF MS-typing efficiency. The need to characterize complex proteomic mixtures led to the coupling of mass spectrometry with liquid chromatography techniques (Moon, Myung, Plasencia, Hilderbrand, & Clemmer, 2003). This bottom up approach consists of two steps: peptides fractionation via LC followed by peptides fragmentation via MS/MS (Kocher, Pichler, Swart, & Mechtler, 2012). Protein identities are then obtained by matching resulting MS/MS spectra with proteomic databases, which contain in-silico digests and in-silico fragments (Cottrell, 2011).

1.4.1 NanoFlow LC MS/MS

The use of narrow-bore LC capillaries operated at ultralow flow rates coupled with tandem mass spectrometry results with protein identification, which is characterized with in-depth protein sequence coverage and high rate of data production (F. Zhou, Lu, Ficarro, Webber, & Marto, 2012). Reverse-phase (RP) LC is most commonly used in shotgun proteomics; whereby mobile phase polarity is gradually decreased by the addition of organic solvent (E. J. Hsieh, Bereman, Durand, Valaskovic, & MacCoss, 2013).

Collision Induced Dissociation of peptides, achieved by dual time of flight MALDI MS, results with MS/MS spectra composed of product ion patterns. Ion products are dependent on both, the peptide sequence and instrument used (Khatun, Ramkissoon, & Giddings, 2007). A database search of un-interpreted MS/MS spectral raw data culminates the process of protein identification.

In this study, we attempt to correlate proteomic profiles of *S. aureus* isolates with their corresponding genotypic characteristics. In addition, we carried out shotgun proteomics on formic acid extracts of one *S. aureus* isolate to identify *S. aureus* proteins of typable quality.

Chapter Two

Materials and Methods

2.1 Clinical Isolates:

This study was conducted on a total of twenty-five isolates of *S. aureus* collected from clinical specimens and provided by Dr. Sima Tokajian. Thirteen of the twenty-five isolates originated from the Jordan University Hospital (JUH) in Amman, accordingly isolates were designated JS ##. The other twelve isolates originated from the American University in Beirut-Medical Center (AUB-MC), accordingly designated LS ##. SCC*mec* types, MLST types, *spa* types, and PVL virulence were previously characterized for all isolates, by senior colleagues. A detailed view of the isolates and their characteristics are listed in Table 2.1. Additionally, one strain of *Staphylococcus epidermidis* (*S. epidermidis*) ATCC 12228 and another of *Escherichia coli* (*E. coli*) ATCC 25922 were included in this study. Isolate LS 16 was chosen for further peptide analysis and protein identification. *S. aureus* isolates involved in this study have been previously thoroughly studied by Harastani et al. (2013) and Al-Bakri et al. (2013).

Table 2.1. A detailed view of characteristics of *S. aureus* isolates utilized in this study (Harastani et al. (2013) and Al-Bakri et al. (2013)).

Sample ID	<i>mecA</i> type	PVL	MLST-ST type	MLST-CC	<i>spa</i> type	<i>spa</i> -CC	Site of Infection
JS 003	MSSA	-	ST-046	CC045	t 0065	CC0065	Nose
JS 008	MRSA-IV	-	ST-025	CC005	t 0922	Singletons	Nose
JS 010	MSSA	-	ST-005	CC005	t 0002	CC0214	Nose
JS 017	MSSA	-	ST-006	CC005	t 0701	CC0024/CC0701	Stool
JS 022	MSSA	-	ST-022	CC022	t 0005	CC0223	Nose
JS 050	MRSA-IV	-	ST-395	CC395	t 1651	No Founder	Nose
JS 052	MSSA	-	ST-30	CC030	t 0012	CC0012	Nose
JS 054	MSSA	-	ST-015	CC015	t 0084	CC0346/CC2398/CC0084	Nose
JS 056	MSSA	-	ST-008	CC008	t 0024	CC0024/CC0701	Stool
JS 060	MRSA-IV	+	-	-	t 0044	CC0044	Stool
JS 061	MSSA	-	ST-001	CC001	t 0267	Singletons	Nose
JS 065	MSSA	-	ST-097	CC097	t 0166	Singletons	Nose
JS 085	MSSA	-	ST-034	CC030	t 5842	No Founder	Stool
LS 016	MRSA-IV	+	ST-030	CC030	t 0318	CC0021	Wound
LS 024	MRSA-III	-	ST-239	CC239	t 0037	CC0021	Brain
LS 034	MRSA-V	-	ST-097	CC097	t 0267	Singletons	Neck mucus
LS 036	MRSA-III	-	ST-239	CC239	t 0030	CC0021	Wound
LS 057	MRSA-IV	+	ST-080	CC080	t 0131	CC0044	Eye
LS 079	MSSA	-	ST-005	CC005	t 0002	CCb	Blood
LS 089	MSSA	-	ST-080	CC080	t 0937	CCd	Wound
LS 094	MSSA	-	ST-030	CC030	t 0012	CC0021	Blood
LS 098	MSSA	+	ST-121	CC121	t 0159	Singletons	Bronchi
LS 103	MRSA-IV	-	ST-022	CC022	t 0032	CC0005	Sputum
LS 125	MRSA-IV	+	ST-005	CC005	t 0311	CCb	Sputum
LS 126	MRSA-IV	-	ST-008	CC008	t 0008	CC0008	Semen

2.2 Protein Extraction:

Formic acid extraction of *S. aureus* proteins was performed according to Wolters et. al (2011) with few modifications. Accordingly, *S. aureus* isolates were cultured on TSA plates and incubated at a temperature of 37 °C for 22-24 hrs. A full loop of colonies was suspended in 300 µL of HPLC-grade water. Vigorous vortexing followed the addition of 900 µL of ethanol. Cell lysate was then centrifuged at 12,000×g for 2 min at room temperature. After discarding the supernatant, centrifugation was repeated to remove any residual ethanol. Then, the pellet was resuspended in 50 µL of 70% formic acid. This step was followed by the addition of 50 µL of ACN. A 10 min vigorous vortexing followed the addition of each, formic acid and ACN. Centrifugation was applied at 20,000×g for 2 min. From the supernatant, 80 µL were collected and stored at –20 °C. Three biological replicates were collected for each *S. aureus* isolate.

2.3 Protein Quantification:

Prior to protein quantification, protein extracts were dried using a speedvac and reconstituted in HPLC-grade water. Protein quantification was carried out using a Pierce BCA Protein Assay Kit (p/n: 23337, Thermo Scientific). Absorbance measurements were recorded at a wavelength of 595 nm using a Multi-scan microplate reader (Thermo Scientific). A standard curve was generated based on absorbance measurements of serial dilutions of Bovine Serum Albumin (250, 125, 50, 25, 5, and 0 µg/mL). After linear regression analysis, protein concentrations were computed. Readings for standard and biological extracts were taken in triplicates. Concentrations of 25 *S. aureus*, *S. epidermidis*, *E. coli* and LS 16 protein extracts were determined accordingly.

2.4 In-Solution Tryptic Digestion

In-solution tryptic digestion protocol was adopted from the mass spectrometry department of University of California in San Francisco (UCSF). A total of 1,000 μL of LS 16 protein extract was dried via Speed Vac and reconstituted in 50 mM NH_4HCO_3 . Reduction was attained by the addition of 100 μL of 100 mM DTT and sequential incubations at 95 $^\circ\text{C}$ for 5 min and 60 $^\circ\text{C}$ for 30 min. Then, alkylation proceeded with the addition of 200 μL of 55 mM iodoacetamide and dark incubation at room temperature for 20 min. Protein extract was then treated with 100 μL (0.1 $\mu\text{g}/\mu\text{L}$) of fresh TCPK-treated trypsin solution (p/n: t1426, Sigma-Aldrich). Digestion took place at 37 $^\circ\text{C}$ for 3 hrs and was terminated by placing peptides on ice.

2.5 Nano-Flow Liquid Chromatography

2.5.1 NanoFlow LC Set Up

Peptide separation was carried out on an Agilent 1260 Infinity Nanoflow Liquid Chromatography System. NanoLC system consisting of 8 modules: Solvent Cabinet, Micro Vacuum Degasser, Binary Nanoflow Pump, High Performance Micro Autosampler, Autosampler Thermostat, Thermostatted Column Compartment, Multiple Wavelength Detector, and Micro-scale Fraction Collector/Spotter. The NanoLC system was operated by ChemStation software version B.04.03-SP1 (Agilent Technologies, USA).

The NanoLC system configuration was set to deliver peptides directly on a nanobore trapping column (Agilent Technologies, Zorbax SB300-C18 0.75 μm x

150 mm, 3.5 μ m). The total dead volume of the NanoLC system was measured using acetone to be 2.5 μ L.

Needle washes preceded every NanoLC run for 30 sec using 15% methanol 84% HPLC-grade water and 0.1% formic acid solution. Solvent A bottle contained HPLC-grade water and solvent B bottle contained ACN. Both solvents were freshly prepared and filtered using 0.45 μ m filters (Sartorius). Additionally, 0.1% formic acid was added to both solvents to ensure that the mobile phase pH does not exceed 7.

2.5.2 Sample Preparation

Given the complexity of sample at hand and taking the dilution factor introduced by Nanoflow LC into consideration, LS 16 in-solution digests were concentrated from 12 mL to 50 μ L using a speedvac. Due to the absence of guard column in Nanoflow LC setup, LS 16 peptides were cleaned and further concentrated using C18 ZipTips (Millipore) according to manufacturer's instructions. Prior to Nanoflow LC separation, concentration of LS 16 peptides was estimated using Nanodrop.

2.5.3 NanoFlow LC Run Conditions

NanoFlow LC protocol is adopted from Kocher et al. (2012). Following an injection volume of 8 μ L, NanoFlow LC system ran for 10 min with 2% solvent B. Peptides were then eluted at a stable flow rate of 275 nL/min with an increasing organic solvent of 2-45% solvent B for 1 hr. The gradient was followed by a 5 min steep increase to 90% solvent B, which was maintained for another 5 min. This step allows for column washing and was followed by an equilibration phase, by which the system returned to its original settings of 2% Solvent B. Detection wavelength and response time was set to 210 nm with 500 nm as a reference and 0.05 min, respectively.

Fractionation trigger mode was time-based; individual fraction duration was limited to 3 min for a total volume of 0.8 μ L/well. Throughout the NanoFlow LC run, the Autosampler Thermostat temperature was set to 4 °C and the Thermostatted Column Compartment temperature was not controlled.

The NanoFlow LC system was first optimized using a Trysin-digested BSA MS Standard (New England Biolabs, P8108S). Following optimization process, LS 16 peptides were subjected to the same protocol as BSA-digest. Additionally, LS 16 peptides were subjected to a 5 hrs NanoFlow LC separation by extension of linear gradient time.

2.6 MALDI-TOF MS Analysis

2.6.1 Plate Preparation

Prior to MALDI-TOF MS analysis, samples were overlaid with CHCA matrix solution (10 mg of α -cyano-4-hydroxy-cinnamic acid in 50% ACN with 0.1% TFA) upon their spotting on a stainless steel target plate (Opti-TOF TM 384 Well Insert, 123×81 mm RevA, Applied Biosystems, USA). Sample spots were air-dried before and after matrix addition. In addition, 13 calibration spots were spotted on each stainless steel target plate.

2.6.2 Calibration and Error Calculation

All MALDI-TOF MS analysis were conducted on a 4800 MALDI TOF/TOF Analyzer instrument operated by the 4000 Series Explorer software version 3.5.1 (Applied Biosystems, USA). Instrument calibration was achieved in reflector mode with the use of 4700 Proteomics Analyzer Calibration mixture (p/n: 4333604, Applied

Biosystems). Des-Arg-Bradykinin, Angiotensin, 1 Glu-Fibrinopeptide B, ACTH (clip 1-17), ACTH (clip 18-39) and ACTH (clip 7-38) constituted the 4700 Cal Mix peptide peaks with theoretical m/z values of 904.4681, 1,296.6853, 1,570.6774, 2,093.0867, 2,465.1989 and 3,657.9294, respectively. Error calculation was preceded by the recording and averaging of 13 MALDI-TOF MS calibration spots readings. The values are recorded in Tables 2.2-2.9 as “Detected Mass of 13 cal. spots”. Errors in ppm were computed using the following formula (Tables 2.2-2.9):

$$\frac{\text{Experimental } m/z - \text{Theoretical } \frac{m}{z}}{\text{Theoretical } \frac{m}{z}} \times 10^6$$

Instrument was also calibrated in MS/MS mode using 4700 Cal Mix peptide peak at 1296.4681. The fragmentation pattern of peak 1296.4681 is listed in Table 2.10.

Table 2.2. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: JS 03, JS08, JS 10, JS 17 and JS 22.

Standard	Monoisotopic Mass (Da)	Detected Mass of 13 cal. spots (Da)	Error (ppm)
des-Arg-Bradykinin	904.4681	904.4027	72.3077
Angiotensin I	1,296.6853	1,296.7717	66.6475
Glu-Fibrinopeptide B	1,570.6774	1,571.0015	206.3157
ACTH(clip1-17)	2,093.0867	2,092.7261	172.2756
ACTH(clip18-39)	2,465.1989	2,465.3929	78.7098
ACTH(clip 7-38)	3,657.9294	3,656.4228	411.8841
Average Error			168.0234

Table 2.3. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: JS 50, JS 52, JS 54, JS 56 and JS 60.

Standard	Monoisotopic mass (Da)	Detected Mass of 13 cal. spots (Da)	Error (ppm)
des-Arg-Bradykinin	904.4681	904.5708	113.5959
Angiotensin I	1,296.6853	1,297.2807	459.1387
Glu-Fibrinopeptide B	1,570.6774	1,571.7740	698.1402
ACTH(clip1-17)	2,093.0867	2,094.2680	564.3862
ACTH(clip18-39)	2,465.1989	2,467.2067	814.4623
ACTH(clip 7-38)	3,657.9294	3,660.9397	822.9565
Average Error			578.7800

Table 2.4. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: JS 61, JS 65 and JS 85.

Standard	Monoisotopic Mass (Da)	Detected Mass of 13 cal. spots (Da)	Error (ppm)
des-Arg-Bradykinin	904.4681	904.9730	558.2023
Angiotensin I	1,296.6853	1,296.9701	219.6180
Glu-Fibrinopeptide B	1,570.6774	1,570.8216	91.8065
ACTH(clip1-17)	2,093.0867	2,092.5460	258.3126
ACTH(clip18-39)	2,465.1989	2,464.3443	346.6676
ACTH(clip 7-38)	3,657.9294	3,653.9397	1,090.6896
Average Error			427.5494

Table 2.5. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: LS 16, LS 24, LS 34, and LS 36.

Standard	Monoisotopic Mass (Da)	Detected Mass of 13 cal. spots (Da)	Error (ppm)
des-Arg-Bradykinin	904.4681	904.7987	365.5000
Angiotensin I	1,296.6853	1,297.1851	385.4556
Glu-Fibrinopeptide B	1,570.6774	1,570.5702	68.2337
ACTH(clip1-17)	2,093.0867	2,093.7726	327.6849
ACTH(clip18-39)	2,465.1989	2,467.2394	827.7332
ACTH(clip 7-38)	3,657.9294	3,662.0363	1,122.7289
Average Error			516.2227

Table 2.6. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: LS 79, LS 89, LS 94, LS 98 and LS 103.

Standard	Monoisotopic Mass (Da)	Detected Mass of 13 cal. spots (Da)	Error (ppm)
des-Arg-Bradykinin	904.4681	904.5371	76.2616
Angiotensin I	1,296.6853	1,296.6878	1.9001
Glu-Fibrinopeptide B	1,570.6774	1,570.1998	304.0658
ACTH(clip1-17)	2,093.0867	2,093.7460	314.9893
ACTH(clip18-39)	2,465.1989	2,468.3850	1,292.4509
ACTH(clip 7-38)	3,657.9294	3,661.3653	939.3042
Average Error			456.8708

Table 2.7. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for isolates: LS 57, LS 125 and LS 126.

Standard	Monoisotopic Mass (Da)	Detected Mass of 13 cal. spots (Da)	Error (ppm)
des-Arg-Bradykinin	904.4681	904.9841	570.4568
Angiotensin I	1,296.6853	1,297.0376	271.7052
Glu-Fibrinopeptide B	1,570.6774	1,570.1998	304.0658
ACTH(clip1-17)	2,093.0867	2,093.7460	314.9893
ACTH(clip18-39)	2,465.1989	2,467.2832	845.5061
ACTH(clip 7-38)	3,657.9294	3,661.3653	939.3042
Average Error			541.0045

Table2.8. 4700 Cal Mix standards used for MALDI TOF MS calibration in linear mode run for strains *S. epidermidis* and *E. coli*.

Standard	Monoisotopic Mass (Da)	Detected Mass of 13 cal. spots (Da)	Error (ppm)
des-Arg-Bradykinin	904.4681	904.2227	271.2797
Angiotensin I	1,296.6853	1,296.4977	144.6487
Glu-Fibrinopeptide B	1,570.6774	1,570.3512	207.6953
ACTH(clip1-17)	2,093.0867	2,092.0023	518.0733
ACTH(clip18-39)	2,465.1989	2,464.6131	237.6460
ACTH(clip 7-38)	3,657.9294	3,656.5749	370.2931
Average Error			291.6060

Table2.9. 4700 Cal Mix standards used for MALDI TOF MS calibration in reflector mode run for LS 16 peptide analysis.

Standard	Monoisotopic Mass (Da)	Detected Mass of 13 cal. spots (Da)	Error (ppm)
des-Arg-Bradykinin	904.4681	904.4648	3.6426
Angiotensin I	1,296.6853	1,296.6942	6.8678
Glu-Fibrinopeptide B	1,570.6774	1,570.6844	4.4993
ACTH(clip1-17)	2,093.0867	2,093.0938	3.3778
ACTH(clip18-39)	2,465.1989	2,465.2043	2.2083
ACTH(clip 7-38)	3,657.9294	3,657.9345	1.3877
Average Error			3.6639

Table 2.10. Fragmentation pattern of 4700 Cal Mix peptide with mass of 1296.4681 obtained from 4000 Series Explorer software.

Fragment Mass	Fragment Ion Type
110.072	H
269.161	y2(+1)
272.136	b2(+1)
416.23	y3(+1)
513.283	y4(+1)
517.241	b4-17(+1)
650.341	y5(+1)
767.384	b6-17(+1)
784.411	b6(+1)
1,046.542	b8 + H ₂ O(+1)
1,165.591	b9(+1)
1,181.658	y9(+1)
1,183.601	b9 + H ₂ O(+1)

2.6.3 Protein Analysis

MALDI-TOF MS spectra were acquired in linear positive mode at a laser intensity of 5000 within a mass range of 800 to 12,000 Da. Focus mass was set to 6,000 Da. Acceleration voltage was 20 kV, IS2 voltage was maintained at 19.3 kV, extraction delay time was 160 ns and laser frequency was set to 500 MHz. For each sample spot, one sum spectrum was accumulated from 1,000 measurements and minimum signal-to-noise ratio was set to 20. Eleven replicate readings were recorded for every biological replicate. A total of 924 spectra were obtained in this study. Automated preprocessing (smoothing, baseline subtraction, and peak picking) was done using 4000 Series Explorer Software version 3.5.3 (Applied Biosystems, USA).

2.6.4 Shotgun proteomics - Peptides Analysis

MALDI-TOF MS spectra were acquired in reflector positive mode at a laser intensity of 4600 within a mass range from 400 to 4,000 Da, with a focus mass of 1,200 Da. Acceleration voltage was 20 kV, IS2 voltage was maintained at 16.1 kV, the extraction delay time was 300 ns, and laser frequency was set to 500 MHz. For each sample spot, one sum spectrum was accumulated from 800 measurements and minimum signal-to-noise ratio was set to 20.

An interpretation method was utilized to collect MS/MS spectra in batch mode for Nanoflow LC fragments. This method allows for the dissociation of the strongest 20 precursors/spot. An adduct exclusion list as well as peptides exclusion and inclusion lists is also supported in this method. An inclusion list was created for each of BSA-digest and LS 16 digest based on a mass list extracted from MS Spectra in reflector mode. An exclusion list for CHCA matrix was also created in the same manner, in addition to an adduct exclusion for sodium and potassium at 21.982 and 37.956 m/z,

respectively. Inclusion and exclusion lists are found in the Appendix (Tables 7.1 and 7.2).

In addition to MALDI TOF MS/MS analysis of NanoFlow LC fragments, complete in-solution digest of LS 16 acquired in reflector mode in the ranges of 400-1,000, 1,000-2,000, 2,000-3,000, 3,000-4,000 and 400-4,000 m/z were subjected to MS/MS analysis.

MS/MS Spectra were acquired in a positive mode using CID by air (medium dissociation mode). For each sample spot, one sum spectrum was accumulated from 500 measurements and minimum signal-to-noise ratio was set to 3.

2.7 Bioinformatics

2.7.1 Protein Analysis

Subsequent statistical analysis of *S. aureus* proteomic profiles was performed using BioNumerics software version 7.1, developed by Applied Maths. The criteria for considering a protein peak present in an isolate was its occurrence in 6 out of 11 MALDI replicates and 2 out of 3 biological replicate. Accordingly and taking experimental error (600 ppm) into consideration, one mass list was generated for every *S. aureus* isolate, *S. epidermidis* and *E. coli* strain. A dendrogram was created based upon a comparison of the *S. epidermidis*, *E. coli* strains' and 25 *S. aureus* isolates' peak lists.

2.7.2 Peptides Analysis and Protein Identification

Raw MS/MS spectra were searched against SwissProt and NCBIInr databases using licensed MASCOT software, version 2.2.04. Other mammalia and other Firmicutes were chosen for BSA-digest and LS 16 taxonomy, respectively. A

maximum of 1 missed cleavage was permitted for tryptic digestion. Carbamidomethylation of cysteine residues and oxidation of methionine residues were set as fixed and variable modifications, respectively. MALDI TOF-TOF was chosen to be the instrument of this project, consequently the peptide charge was set to “1+”. As recommended by MASCOT, peptide and fragment tolerance was inferred using a standard, in this case, BSA digest. Accordingly, peptide and fragment tolerance were set to 0.6/0.2 and 0.1, respectively. Submitted data were in MGF and error tolerance was allowed. Unassigned protein sequences were blasted on non-redundant NCBI protein database with strict parameters; expect threshold was set to 10^{-5} , gap cost of 7 for existence and 2 for extension and database was limited to *S. aureus* proteins. Additionally, parameters were adjusted for short input sequences. The biological functions of identified proteins were obtained from NCBItr and UniProt databases.

Chapter Three

Results

3.1 Protein Quantification

Protein concentrations were measured for the 25 *S. aureus* isolates, *S. epidermidis*, *E. coli* strains and isolate LS 16. A total of 246 measurements were recorded for all isolates and strains, rising from 3 absorbance readings for each biological replicate and 3 biological replicates for each isolate/strain. Given the fact that a 96-well plate cannot accommodate this big number of samples plus the BSA standards, measurements were determined in 5 plates. Therefore 5 BCA-standard curves were plotted, one for each plate. Figure 3.1 illustrates the BCA standard curve used in the measurement of protein concentrations of isolates LS 89, LS 98, LS 103, LS 125 and LS 126. BCA standard curves used in the measurements of protein concentrations of remaining isolates can be viewed in the Appendix (Figures 7.1-7.4)

Figure 3.1. BCA standard curved from which protein concentrations of isolates LS 89, LS 94, LS 98, LS 103, LS 125 and LS 126 were extrapolated from.

Figure 3.2. Protein Concentrations in µg/mL for each of the 25 *S. aureus* isolates in this study.

Concentrations of *S. aureus* protein extracts ranged from 184 to 1,708 µg/mL, with an average concentration of 793 µg/mL. Figure 3.2 illustrates the variation in protein concentrations of the 25 *S. aureus* isolates involved in this study. Additionally, the concentrations of protein extracts of *S. epidermidis* and *E. coli* strains were found to be 823 and 188 µg/mL, respectively. LS 16, the *S. aureus* isolate chosen for further MS analysis in the process of protein identification, had an average protein concentration of 427 µg/mL.

3.2. Proteomic Fingerprinting

MALDI-TOF MS analysis verified that every isolate has a unique proteomic fingerprint. Figures 3.3 to 3.9 represent MALDI-TOF MS spectra (linear mode with mass range 800-1,200 m/z) for formic acid protein extracts of JS isolates, LS isolates, *S. epidermidis* and *E. coli* strains.

Figure 3.3. MALDI-TOF MS spectra for formic acid protein extracts of isolates JS 03 (top), JS 08 (middle) and JS 10 (bottom).

Figure 3.4. MALDI-TOF MS spectra for formic acid protein extracts of isolates JS 17 (top), JS 22 (middle-top), JS 50 (middle-bottom) and JS 52 (bottom).

Figure 3.5. MALDI-TOF MS spectra for formic acid protein extracts of isolates JS 54 (top), JS 56 (middle-top), JS 60 (middle-bottom) and JS 61 (bottom).

Figure 3.6. MALDI-TOF MS spectra for formic acid protein extracts of isolates JS 65 (top), JS 85 (middle-top), LS 16 (middle-bottom) and LS 24 (bottom).

Figure 3.7. MALDI-TOF MS spectra for formic acid protein extracts of isolates LS 34 (top), LS 36 (middle-top), LS 57 (middle-bottom) and LS 79 (bottom).

Figure 3.8. MALDI-TOF MS spectra for formic acid protein extracts of isolates LS 89 (top), LS 94 (middle-top), LS 98 (middle-bottom) and LS 103 (bottom).

Figure 3.9. MALDI-TOF MS spectra for formic acid protein extracts of isolates LS 125 (top), LS 126 (middle-top), *S. epidermidis* strain (middle-bottom) and *E. coli* strain (bottom).

3.3. Peak analysis

Upon data analysis, not a single peak was found common between all *S. aureus* isolates. Protein peaks, which fall shortly outside the considered error of 600 ppm in this study, were averaged for protein peaks that are considered discriminatory for *S. aureus* by Abdallah (2011) and Wolters et al. These protein peaks appear to be the highest-occurring protein peaks and are listed in Table 3.1.

Table 3.1. The frequency of the highest-occurring protein peaks detected by MALDI-TOF MS in linear mode of formic acid protein extracts of *S. aureus* isolates.

Peak Mass (Da)	JS 03	JS 08	JS 10	JS 17	JS 22	JS 50	JS 52	JS 54	JS 56	JS 60	JS 61	JS 65	JS 85	LS 16	LS 24	LS 34	LS 36	LS 57	LS 79	LS 89	LS 94	LS 98	LS 103	LS 125	LS 126	Peak Count	% Occurrence
4813.17														x	x	x	x									4	16%
4819.16																			x	x			x	x		4	16%
4825.1			x		x																				x	3	12%
4829.33	x	x									x	x	x													5	20%
4833.67						x	x	x																		3	12%
4824.09★																										19	76%
6411.1														x	x	x	x		x	x		x		x	x	9	36%
6424.73	x	x	x	x	x						x	x	x													8	32%
6436.12						x	x	x	x	x																5	20%
6423.98★																										22	88%
6586.29															x	x			x	x		x	x			6	24%
6600.09			x		x							x		x			x							x	x	7	28%
6613.29		x		x						x	x															4	16%
6635.59						x		x																		2	8%
6608.81★																										19	76%
8903.83		x	x		x																					3	12%
8923.55							x	x	x	x																4	16%
8875.75														x	x	x	x									4	16%
8887.78																							x		x	2	8%
8897.73★																										13	52%
9621.97														x	x	x	x		x	x		x	x	x	x	10	40%
9645.88	x	x	x	x	x						x	x	x													8	32%
9666.97						x	x	x	x	x																5	20%
9644.94★																										23	92%

★ Average masses of protein peaks that are assumed to be isomers of the same protein.

3.4. Correlating Proteomic Fingerprints and Genotypic

Characteristics

To correlate proteomic data generated by MALDI-TO MS to genotypic typing characteristics (MLST-cc, spa-cc, mecA types, PVL and site of location), MALDI-TOF MS raw data acquired for formic acid protein extracts of *S. aureus* isolates, *S. epidermidis* and *E. coli* strains were compared using BioNumerics software and a dendrogram was generated, Figure 3.10. The dendrogram was based on similarities of protein masses among different isolates. Genotypic characteristics are listed on the right of the dendrogram. While no relationship was spotted between proteomic fingerprints and genotypic characteristics, a geographic separation of isolates was apparent in the dendrogram. *S. epidermidis* and *E. coli* strains were not clustered separately in the dendrogram, hence Cluster Analysis based on peak similarity is not enough to differentiate between *S. aureus* isolates and *S. epidermidis* and *E. coli* strains.

Figure 3.10. Dendrogram representation of hierarchical cluster analysis of *S. aureus* isolates (LS# and JS#), *S. epidermidis* (SE) and *E. coli* (EC) strains along with their corresponding genotypic characteristics.

3.5. NanoFlow LC MS/MS Optimization

MALDI-TOF MS spectrum (reflector mode) of trypsin-digested BSA standard prior to Nanoflow LC separation is shown in Figure 3.11. The spectrum allows the visualization of the most prominent peaks of trypsin-digested BSA standard.

Figure 3.11. MALDI-TOF MS spectrum acquired in reflector mode (mass range: 400-4000 m/z) for Trypsin-digested BSA MS Standard.

Trypsin-digested BSA standard was used in the process of optimizing Nanoflow LC MS/MS; the resulting chromatogram is shown in Figure 3.12. Peptides fractionation was optimized by visualizing peptides in MALDI-TOF MS reflector mode-generated raw data. Only prominent peaks in MALDI-TOF MS spectrum of trypsin digested standard (Figure 3.11) appear in MALDI-TOF MS spectra of fractionated trypsin-digested BSA standard. Peptide fractionation profile can be viewed in the Appendix (Table 7.3).

Figure 3.12. Nanoflow LC chromatogram of trypsin-digested BSA standard. Injection volume: 8 μ L. Linear gradient of 2-45% solvent B. Detection wavelength: 210nm. Time-based fractionation: 3min.

MS/MS spectrum for the peptide with mass of 1,480.07 m/z from trypsin-digested BSA standard is shown in Figure 3.13. MS/MS ion search of trypsin-digested BSA standard was successfully performed using MASCOT, following one hour Nanoflow separation coupled with MALDI-TOF MS/MS analysis for the 20 highest peaks per well (Figure 3.14).

Figure 3.13. MALDI-TOF MS/MS spectrum for peptide with m/z value of 1,480.07 from Trypsin-digested BSA Standard.

Mascot Score Histogram

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 37 indicate identity or extensive homology ($p < 0.05$). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

Peptide Summary Report

Format As: Peptide Summary
[Help](#)

Significance threshold $p < 0.05$
Max. number of hits: AUTO
Hide error tolerant matches: ☐

Standard scoring: ☐ MudPIT scoring: ☐ Ions score or expect cut-off: 0
Show sub-sets: 0

Show pop-ups: ☐ Suppress pop-ups: ☐ Sort unassigned: Decreasing Score
Require bold red: ☐

Preferred taxonomy: All entries

Select All Select None Search Selected ☐ Error tolerant

- ALBU_BOVIN** Mass: 71244 Score: 175 Matches: 20(9) Sequences: 16(8) emPAI: 0.62
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
458	927.4647	926.4574	926.4861	-0.0287	0	36	0.12	2		K.YLYEIAR.R
463	974.4298	973.4225	973.4505	-0.0280	0	35	0.18	1	U	K.DLGEHHFK.G
468	1034.4471	1033.4398	1033.4651	-0.0253	0	34	0.17	1	U	R.NECFLSHK.D
508	1107.4873	1106.4800	1106.5066	-0.0266	0	39	0.061	1	U	K.EACFAVEGPK.L
511	1163.6000	1162.5927	1162.6234	-0.0306	0	52	0.0026	1	U	K.LVNELTEFAK.T
582	1283.6819	1282.6746	1282.7034	-0.0287	0	45	0.011	1		R.HPEYAVSVLLR.L
583	1291.5748	1290.5675	1289.6656	0.9020	0	(38)		1		R.HPEYAVSVLLR.L + [+6.9622 at C-term R]
584	1305.6860	1304.6787	1304.7088	-0.0301	0	56	0.0011	1		K.HLVDEPQNLIK.Q
585	1399.6592	1398.6519	1398.6853	-0.0334	0	71	2.9e-05	1		K.TVMENFVAFVDK.C
586	1415.6541	1414.6468	1414.6803	-0.0334	0	(21)	3.3	3		K.TVMENFVAFVDK.C + Oxidation (M)
665	1479.7646	1478.7573	1478.7881	-0.0308	0	30	0.39	2		K.LGEYGFQNALIVR.Y
699	1502.5828	1501.5755	1501.6065	-0.0309	0	27	0.64	1	U	K.EYEATLEECCK.D
702	1554.6278	1553.6205	1553.6457	-0.0252	0	63	0.00015	1	U	K.DDPHACYSTVEDK.L
703	1567.7109	1566.7036	1566.7354	-0.0318	0	(29)	0.45	3	U	K.DAFLGSLYEYSR.R
704	1576.7263	1575.7190	1576.7627	-1.0436	0	50		8	U	K.DAFLGSLYEYSR.R + [+10.0272 at F7]
787	1724.8030	1723.7957	1723.8273	-0.0316	0	48	0.0047	1		R.MPCTEDYLSLIILNR.L
788	1740.7915	1739.7842	1739.8222	-0.0380	0	(43)	0.015	1		R.MPCTEDYLSLIILNR.L + Oxidation (M)
789	1749.6364	1748.6291	1748.6553	-0.0261	0	57	0.00054	1		K.ECCHGDLLECAADR.A
798	1907.8750	1906.8677	1906.9135	-0.0458	0	27	0.52	5	U	K.LFTFHADICTLPDTEK.Q
799	1955.9067	1954.8994	1954.9524	-0.0529	0	40	0.028	3	U	K.DAIPENLPPLTADFADK.D
- ALBU_SHEEP** Mass: 71139 Score: 133 Matches: 10(6) Sequences: 7(5) emPAI: 0.38
Serum albumin OS=Ovis aries GN=ALB PE=2 SV=1
☐ Check to include this hit in error tolerant search

Figure 3.14. MASCOT score histogram showing the result of MS/MS ion search in SwissProt database for Trysin-digested BSA MS Standard after fractionation by NanoFlow LC. Individual ions scores > 37 indicate identity or extensive homology ($p < 0.05$). Protein scores are the sum of individual ion scores.

3.6. Identification of *S. aureus* Proteins

A trypsin-digested LS 16 protein sample was analyzed by MALDI-TOF MS prior to Nanoflow LC separation. Spectra of trypsin-digested LS 16 peptides acquired in reflector mode in the ranges of 400-4,000, 400-1,000, 1,000-2,000 and 2,000-3,000 m/z are illustrated in Figures 3.15 and 3.16, respectively. The spectra allow the visualization of the most prominent peaks of trypsin-digested LS 16 peptides in different mass windows. MALDI-TOF MS/MS analysis of the spectra mentioned was performed and followed by MASCOT database search.

Figure 3.15. MALDI-TOF MS spectra acquired in reflector mode for trypsin-digested LS 16 peptides in a mass range of 400-4,000 m/z (top) and 400-1,000 m/z (bottom).

Figure 3.16. MALDI-TOF MS spectra acquired in reflector mode for trypsin-digested LS 16 peptides in a mass range of 1,000-2,000 m/z (top) and 2,000-3,000 m/z (bottom).

Figure 3.17. Nanodrop Plot from which LS 16 peptides concentration was estimated to nearly 90 mg/mL. 1 mm Absorbance = 1 mg/ mL.

Following speedvac concentration and ZipTip clean-up, LS 16 peptides concentration was estimated to be 90 mg/mL using Nanodrop (Figure3.17). Protein identification for *S. aureus* was performed in the same manner as Trysin-digested BSA

MS standard. Accordingly, an hour NanoFlow LC separation method was set up to separate LS 16 peptides over a gradient of 2-45 % Solvent B (Figure 3.18).

Figure 3.18. Nanoflow LC chromatogram of trypsin-digested LS 16. Injection volume: 8 μ L. Linear gradient of 2-45% solvent B. Detection wavelength: 210nm. Time-based fractionation: 3min.

MS/MS spectrum for the peptide with mass of 1,378.72 m/z from trypsin-digested and Nanoflow LC fractionated LS 16 is shown in Figure 3.19. MS/MS ion search of trypsin-digested LS 16 peptides was successfully performed using MASCOT, following Nanoflow LC separation coupled with MALDI-TOF MS/MS analysis for the 20 highest peaks per well/fraction. Figures 3.20 and 3.21 illustrates successful protein identification by MASCOT for SwissProt and NCBItr databases, respectively.

Figure 3.19. MALDI-TOF MS/MS spectrum for peptide with m/z value of 1,378.72 from Trysin-digested LS 16 protein extract.

MATRIX SCIENCE MASCOT Search Results

User : Dania Awad
E-mail : dania.awad@lau.edu
Search title : Project: DANIA, Spot Set: DANIA\CHCA_18022014 2
MS data file : pps_CHCA_18022014 2_1392794079 (1hour).txt
Database : SwissProt 2014_01 (542,258 sequences; 192,776,118 residues)
Taxonomy : Other Firmicutes (60,373 sequences)
Timestamp : 20 Feb 2014 at 18:45:51 GMT

☒ All ☐ Non-significant ☐ Unassigned [\[help\]](#) As

Not what you expected? Try [the peptide summary](#).

Search parameters

Score distribution

Peptide score distribution. Ions score is $-10\log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 31 indicate **identity** or **extensive homology** ($p < 0.05$).

[Deprecated] Score distribution for family members in the first 50 proteins. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein families.

Legend

Protein Family Summary

Significance threshold p< Max. number of families [\[help\]](#)
 Ions score or expect cut-off Dendrograms cut at
 Error tolerant matches:
 Preferred taxonomy

Proteins (4) [Report Builder](#) [Unassigned \(383\)](#)

Protein families 1-4 (out of 4)

per page

Accession

1 **DBH_STAAC** 1016 DNA-binding protein HU OS=Staphylococcus aureus (strain COL) GN=hup PE=3 SV=1
2 **DBH1_BACHD** 101 DNA-binding protein HU-1 OS=Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 /

Figure 3.20. MASCOT score histogram showing the result of MS/MS ion search in SwissProt database for Trysin-digested LS 16 protein extract after fractionation by NanoFlow LC. Individual ions scores > 31 indicate identity or extensive homology ($p < 0.05$). Protein scores are the sum of individual ion scores.

MASCOT Search Results

User : Dania Awad
E-mail : dania.awad@lau.edu
Search title : Project: DANIA, Spot Set: DANIA\CHCA_18022014 2
MS data file : pps_CHCA_18022014_2_1392794079 (1hour).txt
Database : NCBIInr 20140122 (35,149,712 sequences; 12,374,887,350 residues)
Taxonomy : Other Firmicutes (5,013,137 sequences)
Timestamp : 20 Feb 2014 at 20:38:15 GMT

☒ All
 ☐ Non-significant
 ☐ Unassigned
 [\[help\]](#)
 As

Not what you expected? Try [the peptide summary](#).

Search parameters

Score distribution

Peptide score distribution. Ions score is $-10\log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 50 indicate **identity** or **extensive homology** ($p < 0.05$).

[Deprecated] Score distribution for family members in the first 50 proteins. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein families.

Legend

Protein Family Summary

Significance threshold $p <$
 Max. number of families
[\[help\]](#)
 Ions score or expect cut-off
 Dendrograms cut at
 Error tolerant matches:
 Preferred taxonomy

Proteins (2)
 [Report Builder](#)
[Unassigned \(438\)](#)

Protein families 1-2 (out of 2)

per page

 Accession

1 [gi|15924463](#) 700 DNA-binding protein II [Staphylococcus aureus subsp. aureus Mu50]
3 [gi|328957353](#) 100 non-specific DNA-binding protein Hbsu [Carnobacterium sp. 17-4]

Figure 3.21. MASCOT score histogram showing the result of MS/MS ion search in NCBIInr database for Trysin-digested LS 16 protein extract after fractionation by NanoFlow LC. Individual ions scores > 50 indicate identity or extensive homology ($p < 0.05$). Protein scores are the sum of individual ion scores.

Proteins identified in this study for isolate LS 16 are listed in Tables 3.2. Proteins numbered 1-22 in Tables 3.2 are identified from MS/MS spectra of one hour Nanoflow LC fractions. Protein numbered 23 in Table 3.2 is identified from MS/MS spectra of five hour Nanoflow LC fractions. Proteins numbered 24-32 in Table 3.2 are identified from MS/MS spectra of complete and unfractionated in-solution tryptic-digest of LS 16 collected in reflector mode with mass range of 400-1,000 m/z. Proteins numbered 32-36 in Table 3.2 are identified from MS/MS spectra of complete and

unfractionated in-solution tryptic-digest of LS 16 collected in dissected reflector mode of mass ranges of 400-1,000, 1,000-2,000, 2,000-3,000 and 3,000-4,000 m/z.

Proteins identified to species other than *S. aureus* were blasted on non-redundant (nr) NCBI *S. aureus* database. Resulting *S. aureus* homologues are listed in Table 3.3. *S. aureus* homologues were not found for 2 identified proteins with molecular weight of 25,571 and 40,960 Da. Table 3.4 lists protein blast results of MASCOT unassigned peptides in non-redundant (nr) NCBI *S. aureus* database. Highest scoring protein is reported for blastp searches. Tables 3.5 and 3.6 list functions of 59 identified proteins. Identified proteins that were not assigned a molecular weight and a function are listed in Table 7.5 in the Appendix.

Table 3.2. Proteins identified from trypsin-digested LS 16 formic acid extract by MASCOT search engine.

No.	Database MW (Da)	ID	Protein	Species
1	9620	DBH_STAAC	DNA-binding protein HU	<i>Staphylococcus aureus</i>
2	13348	Y1117_STAHI	UPF0342 protein SH1117	<i>Staphylococcus haemolyticus</i>
3	9651	DBH1_BACHD	DNA-binding protein HU-1	<i>Bacillus halodurans</i>
4	29580	Y235_STAEQ	UPF0447 protein SERP0235	<i>Staphylococcus epidermidis</i>
5	28960	Y3362_GEOTN	UPF0447 protein GTNG_3362	<i>Geobacillus thermodenitrificans</i>
6	88841	PBPA_CLOAB	Penicillin-binding protein 1A	<i>Clostridium acetobutylicum</i>
7	107792	SYI_HALOH	Isoleucine--tRNA ligase	<i>Halothermothrix orenii</i>
8	40023	GCST_GEOKA	Aminomethyltransferase	<i>Geobacillus kaustophilus</i>
9	9619	gi488404375	transcriptional regulator	<i>Staphylococcus epidermidis</i>
10	9646	gi446966605	transcriptional regulator	<i>Staphylococcus aureus</i>
11	9650	gi488380409	transcriptional regulator	<i>Staphylococcus</i>
12	13302	gi15924835	hypothetical protein SAV1845	<i>Staphylococcus aureus</i>
13	12256	gi532359921	hypothetical protein SAAG_01746	<i>Staphylococcus aureus</i>
14	13332	gi446212447	hypothetical protein	<i>Staphylococcus aureus</i>
15	13287	gi379796200	hypothetical protein SAMSHR1132_16910	<i>Staphylococcus aureus</i>
16	9805	gi90962747	DNA-binding protein HU	<i>Lactobacillus salivarius</i>
17	9678	gi497173090	DNA-binding protein HU	<i>Weissella ceti</i>
18	14040	gi222151808	hypothetical protein MCCL_1563	<i>Macroccoccus caseolyticus</i>
19	25571	gi553957006	hypothetical protein	<i>Eubacterium brachy</i>
20	9617	gi319892463	DNA-binding protein HBSu	<i>Staphylococcus pseudintermedius</i>
21	9604	gi446966606	transcriptional regulator	<i>Staphylococcus aureus</i>
22	9634	gi488364546	transcriptional regulator	<i>Staphylococcus</i>
23	44043	TGT_BACLD	Queuine tRNA-ribosyltransferase	<i>Bacillus licheniformis</i>
24	43882	BRPA_STRMU	Biofilm regulatory protein A	<i>Streptococcus mutans serotype c</i>
25	8085	RL29_STAA2	50S ribosomal protein L29	<i>Staphylococcus aureus</i>
26	10279	Y1445_STAAN	UPF0297 protein	<i>Staphylococcus aureus</i>
27	11029	ESXA_STAAC	Virulence factor EsxA	<i>Staphylococcus aureus</i>
28	10229	RS16_STAA1	30S ribosomal protein S16	<i>Staphylococcus aureus</i>
29	14607	RS9_STAA1	30S ribosomal protein S9	<i>Staphylococcus aureus</i>
30	31122	gi557828691	2-dehydropantoate 2-reductase	<i>Exiguobacterium sp. MH3</i>
31	11271	SP5G_STAA2	Putative septation protein SpoVG	<i>Staphylococcus aureus</i>
32	13150	gi492418262	Fic family protein	<i>Eubacterium saphenum</i>
33	40960	gi495696491	hydrogenase expression/formation protein HypD	<i>Clostridium sp. Maddingley MBC34-26</i>
34	28880	gi493632938	DNA repair protein RecO	<i>Lactobacillus jensenii</i>
35	34894	HPRK_CLOPE	HPr kinase/phosphorylase	<i>Clostridium perfringens</i>
36	7770	Y1696_STAAN	UPF0435 protein	<i>Staphylococcus aureus</i>

Table 3.3. Proteins identified to species other than *S. aureus* and their corresponding *S. aureus* homologues. Homologues are obtained from non-redundant (nr) NCBI *S. aureus* database blast search.

No.	Database MW (Da)	ID	Protein	Species	<i>S. aureus</i> Homologue ID	Homologue Protein
1	13348	Y1117_STAHI	UPF0342 protein SH1117	<i>Staphylococcus haemolyticus</i>	Y1663_STAAN	UPF0342 protein SA1663
2	9651	DBH1_BACHD	DNA-binding protein HU-1	<i>Bacillus halodurans</i>	gi 580255613	DNA-binding protein HU
3	29580	Y235_STAEQ	UPF0447 protein SERP0235	<i>Staphylococcus epidermidis</i>	Y537_STAAB	UPF0447 protein SAB0537c
4	28960	Y3362_GEOTN	UPF0447 protein GTNG_3362	<i>Geobacillus thermodenitrificans</i>	gi 379795059	hypothetical protein SAMSHR1132_05310
5	88841	PBPA_CLOAB	Penicillin-binding protein 1A	<i>Clostridium acetobutylicum</i>	gi 49483639	penicillin-binding protein 2
6	107792	SYI_HALOH	Isoleucine--tRNA ligase	<i>Haloethermothrix orenii</i>	gi 446306826	isoleucyl-tRNA synthetase
7	40023	GCST_GEOKA	Aminomethyltransferase	<i>Geobacillus kaustophilus</i>	gi 446015501	glycine cleavage system protein T
8	9619	gi 488404375	transcriptional regulator	<i>Staphylococcus epidermidis</i>	gi 15924463	DNA-binding protein II
9	9805	gi 90962747	DNA-binding protein HU	<i>Lactobacillus salivarius</i>	gi 580255613	DNA-binding protein HU
10	9678	gi 497173090	DNA-binding protein HU	<i>Weissella ceti</i>	gi 580255613	DNA-binding protein HU
11	14040	gi 222151808	hypothetical protein MCCL_1563	<i>Macrococcus caseolyticus</i>	gi 446212447	hypothetical protein
12	25571	gi 553957006	hypothetical protein	<i>Eubacterium brachy</i>		No Homologue found
13	9617	gi 319892463	DNA-binding protein HBsu	<i>Staphylococcus pseudintermedius</i>	gi 446966605	transcriptional regulator
14	44043	TGT_BACLD	Queuine tRNA-ribosyltransferase	<i>Bacillus licheniformis</i>	gi 579740496	queuine tRNA-ribosyltransferase
15	43882	BRPA_STRMU	Biofilm regulatory protein A	<i>Streptococcus mutans serotype c</i>	gi 82751903	transcriptional regulator
16	31122	gi 557828691	2-dehydropantoate 2-reductase	<i>Exiguobacterium</i> sp. MH3	gi 379796910	putative ketopantoate reductase
17	13150	gi 492418262	Fic family protein	<i>Eubacterium saphenum</i>	gi 507417608	PfkB family carbohydrate kinase
18	40960	gi 495696491	hydrogenase expression/formation protein HypD	<i>Clostridium</i> sp. Maddingley MBC34-26		No Homologue found
19	28880	gi 493632938	DNA repair protein RecO	<i>Lactobacillus jensenii</i>	gi 514066500	DNA repair protein RecO
20	34894	HPRK_CLOPE	HPr kinase/phosphorylase	<i>Clostridium perfringens</i>	gi 486396223	HPr kinase/phosphorylase

Table 3.4. Proteins identified by non-redundant (nr) NCBI *S. aureus* database search for MASCOT unassigned peptide sequences.

No.	Observed Mass (m/z)	Peptide Sequence	ID	Protein
1	1729.8831	IEDIDVFEATAIQK + [+138.1045 at C-term K]	gi 579935718	acetyl-CoA acetyltransferase, partial
2	1378.7352	SPETFGMLEVR	gi 583105701	methionine sulfoxide reductase B, partial
3	841.5296	VSPGQLIK	gi 379796404	sensor kinase protein
4	1881.8877	QIAANA GLEGSVIVERVR	gi 577289267	chaperonin, partial
5	1414.7722	IRAALPTLQYLR	gi 580848717	oxidoreductase
6	1307.6460	YLSAVNGLLETK	gi 584847631	hypothetical protein U620_02608
7	1480.8762	IVLTGHSLTIAEAR	gi 580689402	PTS system, ascorbate-specific IIC
8	1601.7997	IDRLSMTLMNTMK + 3 Oxidation (M)	gi 580965588	coenzyme A disulfide reductase, partia
9	1911.0636	HSQATDVLIMLSIQNK	gi 447003213	membrane protein
10	1917.9103	LLGFAHQEMDNSRMR	gi 581796105	protein translocase subunit secA 2
11	2063.9983	DLNAADIEAAMRMIEGTAR + Oxidation (M)	gi 584655548	50S ribosomal protein L11
12	2104.0530	IMDINGLDKISTEVETAR	gi 487738499	putative ATP-binding domain protein, partial
13	1464.7075	LNAKSVEDAFNTR	gi 580927722	hypothetical protein U042_02830
14	954.4770	NILVLGLGR	gi 487722202	amino acid ABC transporter substrate-binding protein
15	1679.8086	IQNGASPEDFAVLYR	gi 582606875	hypothetical protein U343_02428, partial
16	1730.9032	VGMHCAVASPKGYEPK	gi 446045965	acetyl-CoA synthetase
17	1918.8976	NDIMLTQQTISRQQR + Oxidation (M)	gi 386729591	Sodium/proline symporter
18	1361.6914	VTTLMGDKVEPR + Oxidation (M)	gi 582616066	DNA topoisomerase 4 subunit B, partial
19	1935.8772	TAHGSWADSQMEPPEHR	gi 487730301	membrane protein
20	1036.5518	FGVDTRSVR	gi 582450771	hypothetical protein U271_01845
21	852.4320	LNA YPFK	gi 487730899	30S ribosomal protein S2
22	1736.8127	LDEMEQDGTDFVLPK	gi 582657440	DNA topoisomerase 3, partial
23	928.5194	ILEAPEEK	gi 17227188	hypothetical protein pETB_p11
24	1306.6812	LADMVTPEMRK + Oxidation (M)	gi 580256465	50S ribosomal protein L17
25	553.3952	GGYTR	gi 487368615	peptide methionine sulfoxide reductase msrB
26	2225.1006	EEGLYVDIVSGKPLFTSKDK	gi 514309011	chloramphenicol/florfenicol resistance protein, partial
27	2698.2009	LMPIGNKYSIEILDADYDFK + Oxidation (M)	gi 578670717	PTS system EIIBC component, partial
28	1008.5385	VVGPGTLIPR	gi 581372899	30S ribosomal protein S11, partial
29	1882.9065	VAGGKHSFIEHEEDISK	gi 581372899	30S ribosomal protein S11, partial
30	1400.7163	EVFVPIEQGKVR	gi 579648211	cysteinyl-tRNA synthetase, partial
31	934.4845	EKAAFIQK	gi 510813435	hypothetical protein
32	1735.8058	KTVGILTEMQAEEDR + Oxidation (M)	gi 487721653	GTP diphosphokinase
33	920.4692	DVDSDTLR	gi 487754004	hypothetical protein
34	1934.8690	FMFADAVQCDFDKQGR	gi 521212208	cell division protein MraZ
35	1957.9141	GWEAAVSAIEMANVLRGLA	gi 580654011	6,7-dimethyl-8-ribityllumazine synthase
36	2274.0833	AGLTEIMVAMNYLVHNPQIK + 2 Oxidation (M)	gi 379795214	peptidase T
37	2059.9998	EVEEHLHVAKADPWMR	gi 487741723	DNA polymerase III subunit alpha
38	1740.8979	LDELPGGALVGTSSLRR	gi 447152960	hypothetical protein, partial
39	1463.7273	GSFLSNISKNLQR	gi 580840478	hypothetical protein T991_02716
40	1956.8464	IAESRTYPQFPYTDR	gi 447004115	hypothetical protein
41	912.4804	IYNAYLR	gi 582796996	hypothetical protein U433_02585
42	1900.9148	ASMALNMDIATNSKFDR + Oxidation (M)	gi 580175801	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
43	2064.9751	ESTIGIDAGGTLTKIAYLNK	gi 446612392	pantothenate kinase
44	1381.6901	AGINMDGINFSAR + Oxidation (M)	gi 584237569	cation:proton antiporter
45	1678.8104	EEQDSEMLEVKTTPK + Oxidation (M)	gi 446056903	GTPase Era
46	1912.0410	GMWAMPDLMAEMLKQK + 2 Oxidation (M)	gi 487739866	TelA
47	2282.1580	AMETVGVKPLIKPIRGDTGSR	gi 487709438	peptidase T
48	2058.9949	IIEQIGTYNPTSANAEIK	gi 15924228	30S ribosomal protein S16
49	977.6641	TGGHFIAMK + Oxidation (M)	gi 584237570	cation:proton antiporter
50	2331.1619	HVLLSHHGKGEWGSPKPLVR	gi 470191645	Cmp-binding-factor=1
51	531.4152	VPSTK	gi 580754041	transposase for insertion sequence-like element IS431mec
52	2275.0618	IPLILEVTSNQYESHTQFR	gi 379795714	Nuclease sbcCD subunit C
53	2040.9270	TPAANKMDIFEDGSIVGLF + Oxidation (M)	gi 446923043	serine protease
54	1974.8595	LDGKELPPVEMMQTLNKK + 2 Oxidation (M)	gi 379796531	DJ-1/PfpI family protein
55	1975.8680	LIEMGAEEVEENGDSVLVR + Oxidation (M)	gi 584763348	kinase-associated protein B
56	2103.0493	VGGLFSKEIQEAGNVVISQK	gi 10956153	recombination protein
57	2090.9849	ISGDEALEIISKIFRPFR	gi 446986725	UDP-N-acetyl-D-galactosamine dehydrogenase
58	1023.7339	ISYPESSLK	gi 379796521	tagatose-6-phosphate kinase

Table 3.5. Functions and Molecular weights of 59 identified proteins, obtained from Uniprot or NCBI databases.

No.	ID	Protein	Molecular Weight (Da)	Database	Function
1	gi 379796404	sensor kinase protein	101,506	uniprot	membrane protein involved in kinase transferase
2	gi 447003213	membrane protein	15,526	uniprot	membrane protein with uncharacterized function
3	gi 487738499	putative ATP-binding domain protein, partial	5,026	uniprot	ATP-binding protein
4	gi 487722202	amino acid ABC transporter substrate-binding protein	9174	NCBI (protein)	Protein involved in amino acid transport and metabolism and Signal transduction mechanisms
5	gi 446045965	acetyl-CoA synthetase	12255	NCBI (protein)	Acetyl-CoA synthetase containing an Adenylate forming domain and a CoA binding site
6	gi 386729591	Sodium/proline symporter	56,409	uniprot	Transmembrane ion sodium ion transporter protein
7	gi 487730301	membrane protein	12,465	uniprot	Membrane protein with uncharacterized function
8	gi 487730899	30S ribosomal protein S2	15128	NCBI (protein)	Ribosomal protein involved in translation
9	gi 17227188	hypothetical protein pETB_p11	5,864	uniprot	Uncharacterized protein
10	gi 487368615	peptide methionine sulfoxide reductase msrB	16,226	uniprot	Repair protein with peptide-methionine (R)-S-oxide reductase activity that functions in response to oxidative stress
11	gi 514309011	chloramphenicol/florfenicol resistance protein, partial	25,395	uniprot	Cytoplasm protein involved in methylation and rRNA methylation
12	gi 510813435	hypothetical protein	4061	NCBI (protein)	Uncharacterized protein
13	gi 487754004	hypothetical protein	40,205	uniprot	Guanosine tetraphosphate that belongs to relA/spoT family. A mediator of the stringent response that coordinates a variety of cellular activities in response to changes in nutritional abundance
14	gi 521212208	cell division protein MraZ	16,959	uniprot	Cell cycle protein involved in cell division
15	gi 379795214	peptidase T	45,930	uniprot	Cytoplasm proteolytic enzyme that Cleaves the N-terminal amino acid of tripeptides
16	gi 487741723	DNA polymerase III subunit alpha	89,622	uniprot	Protein with 3'-5' exonuclease activity that functions in DNA replication
17	gi 447152960	hypothetical protein, partial	15331	NCBI (protein)	An intermediate enzyme in the biosynthetic pathway of tetrapyrrolic ring systems, such as heme, chlorophylls, and vitamin B12
18	gi 447004115	hypothetical protein	16,235	uniprot	Uncharacterized function
19	gi 446612392	pantothenate kinase	29,165	uniprot	Cytoplasm kinase involved in Coenzyme A biosynthesis
20	gi 446056903	GTPase Era	34,330	uniprot	Essential GTPase that binds GDP and GTP, with rapid nucleotide exchange. Plays a role in 16S rRNA processing and 30S ribosomal subunit biogenesis and possibly in cell cycle regulation and energy metabolism
21	gi 487739866	TelA	22302	uniprot	Toxic anion resistance protein (TelA)
22	gi 487709438	peptidase T	8,674	uniprot	Protein with hydrolase activity
23	gi 15924228	30S ribosomal protein S16	10,235	uniprot	Ribosomal protein involved in translation
24	gi 470191645	Cmp-binding-factor=1	33,711	uniprot	Metal-ion binding protein with phosphoric diester hydrolase activity
25	gi 379795714	Nuclease sbcCD subunit C	117,130	uniprot	DNA repair protein with nuclease activity
26	gi 446923043	serine protease	84,510	uniprot	Serine-type endopeptidase activity
27	gi 379796531	DI-1/PfpI family protein	23445	NCBI (protein)	Type 1 glutamine amidotransferase with (GATase I)-like domain
28	gi 10956153	recombination protein	48,496	uniprot	DNA binding protein
29	gi 446986725	UDP-N-acetyl-D-galactosamine dehydrogenase	48014	NCBI (protein)	Outer membrane protein that belongs to the UDP-glucose/GDP-mannose dehydrogenase family with possible function in cell envelope biogenesis
30	gi 379796521	tagatose-6-phosphate kinase	33,950	uniprot	ATP-binding protein involved in lactose metabolism

No.	ID	Protein	Molecular Weight (Da)	Database	Function
31	Y1663_STAAN	UPF0342 protein SA1663	13,310	uniprot	Uncharacterized function
32	Y537_STAAB	UPF0447 protein SAB0537c	29,328	uniprot	Uncharacterized function
33	gi 379795059	hypothetical protein SAMSHR1132_05310	29,252	uniprot	Uncharacterized function
34	gi 49483639	penicillin-binding protein 2	79,295	uniprot	Membrane protein that binds penicillin
35	gi 446306826	isoleucyl-tRNA synthetase	104,843	uniprot	Cytoplasm protein involved in protein synthesis by Catalyzing the attachment of isoleucine to tRNA
36	gi 15924463	DNA-binding protein II	9,626	uniprot	DNA-binding protein involved in chromosome condensation
37	gi 580255613	DNA-binding protein HU	40,504	uniprot	Cytoplasm protein with aminomethyltransferase activity leading to glycine degradation
38	gi 446212447	hypothetical protein	13,340	uniprot	Uncharacterized function
39	gi 446966605	transcriptional regulator	9,652	uniprot	DNA-binding protein involved in chromosome condensation
40	gi 82751903	transcriptional regulator	34,768	uniprot	Protein involved in transcription regulation
41	gi 379796910	putative ketopantoate reductase	32,451	uniprot	Cytoplasm protein that binds NADP and is involved in pantothenate biosynthetic process
42	gi 514066500	DNA repair protein RecO	28,207	uniprot	Protein involved in DNA repair and RecF pathway recombination
43	gi 486396223	HP γ kinase/phosphorylase	6022	NCBI (protein)	Kinase protein with specific and non-specific DNA binding sites. Helix-turn-helix XRE-family like protein
44	DBH_STAAC	DNA-binding protein HU	9620	uniprot	Histone-like DNA-binding protein which is capable of wrapping DNA to stabilize it, and thus prevents its denaturation under extreme environmental conditions
45	gi 446966605	transcriptional regulator	9646	uniprot	DNA-binding protein involved in chromosome condensation
46	gi 488380409	transcriptional regulator	9650	uniprot	DNA-binding protein involved in chromosome condensation
47	gi 15924835	hypothetical protein SAV1845	13302	uniprot	Uncharacterized function
48	gi 532359921	hypothetical protein SAAG_01746	12256	uniprot	Uncharacterized function
49	gi 446212447	hypothetical protein	13332	uniprot	Uncharacterized function
50	gi 379796200	hypothetical protein SAMSHR1132_16910	13287	uniprot	Uncharacterized function
51	gi 446966606	transcriptional regulator	9604	uniprot	DNA-binding protein involved in chromosome condensation
52	gi 488364546	transcriptional regulator	9634	uniprot	Non-specific DNA-binding protein HBSu signal recognition particle-like (SRP) component
53	RL29_STAAC	50S ribosomal protein L29	8085	uniprot	Ribosomal protein involved in translation
54	Y1445_STAAN	UPF0297 protein	10279	uniprot	Uncharacterized function
55	ESXA_STAAC	Virulence factor EsxA	11029	uniprot	Secreted virulence factor involved in pathogenesis by interacting with EsxB. Involved in abscess formation
56	RS16_STAAC	30S ribosomal protein S16	10229	uniprot	Ribosomal protein involved in translation
57	RS9_STAAC	30S ribosomal protein S9	14607	uniprot	Ribosomal protein involved in translation
58	SP5G_STAAC	Putative septation protein SpoVG	11271	uniprot	Protein with possible function in septation
59	Y1696_STAAN	UPF0435 protein	7770	uniprot	Uncharacterized function

Chapter Four

Discussion

MALDI-TOF MS analysis of bacterial isolates is renowned as an efficient identification method in clinical laboratories (Gagnaire et al., 2012). On the other hand, culture conditions and sample preparation methods introduce many variables to the identification process. Those variables affecting spectrum quality and reproducibility of *S. aureus* profiling were first thoroughly studied by Goldstein et al. in August 2013. Cultures grown in broth and processed following protein extraction method appears to offer most promising results for *S. aureus* characterization (Goldstein, Zhang, Borrer, Rago, & Sandrin, 2013).

Protein concentration introduces an additional variable to MALDI-TOF MS analysis; concentrations of the 25 *S. aureus* protein extracts were measured to determine the effect of protein concentration on proteomic analysis. Protein concentrations of different isolates range 10 folds; variation of the protein concentration was not corrected for in this study. It was also observed that protein concentrations of the same isolate differ drastically among biological replicates. The concentration of the one biological replicate was almost 4 times higher than another biological replicate for the case of isolate LS 34. A major difficulty lies in adjusting this variation, given that it will lead to dilution of protein extracts. One solution to this challenge lies in adjusting cell density prior to protein extraction method (Sitepu et al., 2012).

In a previous study in our lab, El Hage (2012) showed the absence of correlation between protein concentrations of *S. aureus* formic acid extracts and protein peak counts in MALDI-TOF MS spectra (linear mode). This result was confirmed in this study, whereby protein extract concentration of isolate LS 34 was 184 µg/mL and the protein peak count 34, whereas extracts of isolate JS 17 had a protein concentration of 878 µg/mL and a protein peak count of 9.

4.1 Assessment of Proteomic Profiles

MALDI-TOF MS spectra, acquired in linear mode with a mass range of 800-12,000 m/z for formic acid protein extracts of 25 *S. aureus* isolates, *S. epidermidis* isolate and *E. coli* isolate, demonstrates a unique fingerprint for every isolate based on visual inspection of Figures 3.3-3.9 and spectral analysis using BioNumerics software. It must be noted that isolates are expected to exhibit differential growth rates, accordingly extracting proteins at the same time (22-24 hrs) for all isolates introduces a variation in the type and quantity of the extracted proteins. In this study, growth curves were not prepared for *S. aureus* isolates as to mimic the identification process in clinical settings.

Upon comparison of MALDI-TOF MS spectra of 25 *S. aureus* isolates, *S. epidermidis* isolate and *E. coli* isolate, a cluster analysis, based on Jeffrey's X similarity coefficient was generated. Our cluster analysis reveals geographic separation of *S. aureus* isolates, e.g. into isolates originating from the Lebanese vs. Jodanian hospital. No relationship between proteomic fingerprints and *spa*-cc, MLST-cc, *mecA*, PVL or site of location was detected. Furthermore, clustering was not species-dependent; *S. epidermidis* and *E. coli* were not clustered separately in the dendrogram.

Previously, MALDI-typing was achieved to the species-level, yet a large database of spectra, e.g. Bruker Biotyper was implemented in the process.

Although identification of *S. aureus* to the species level based on MALDI-TOF MS spectra in mass range of 800-12,000 m/z is applied in clinical laboratories, protein signatures in this range were not evaluated until January 2013 by Josten et al. Asymmetry of spectra among isolates can be owed to peak intensity variation, peak loss and peak shift. While the former two conditions can be directly correlated with culture conditions, the latter can also be an indication of a gene mutation of the detected peptides. Altered gene product results in molecular weight difference due to an amino acid exchange leading to peak shifts (Josten et al., 2013).

Considering an experimental error of 600 ppm, any amino acid exchange occurring in a specific protein of one of the isolates will cause a peak shift, hence the protein becomes easily mistakable as a different protein. For example, if a protein with a mass of 6,000 m/z undergoes an amino acid exchange of glycine with tryptophan, its mass becomes 6,129 m/z. A mass shift of 129 Da exceeds the observed error range of ± 3.6 Da at 6,000 m/z in this study, hence the protein at 6,129 m/z would be falsely reported as a different protein. Upon data analysis considering a low error range of 600 ppm, none of the protein peaks in the mass range of 800-12,000 m/z were found unique to a specific MLST or *spa* clonal complex.

Moreover, none of the protein peaks were found common to all *S. aureus* isolates. Abdallah (2011) confirmed the presence of 5 protein peaks with m/z of 4,828, 6,612, 6,428, 8,903 and 9,644 in 13 *S. aureus* isolates considering an error of 1500 ppm, which were considered *S. aureus* discriminatory peaks by Wolters et al. 2011. Protein peaks listed in Table 3.1 refer to protein peaks, which are considered as different proteins, although they only differ by few Daltons considering an error of 600

ppm. We suggest that these protein peaks correspond to isomers of the same protein, which differ in mass due to amino acid exchange. To confirm this theory, we aligned proteins identified in this study with the molecular weights of 9,620, 9,651, 9,650, 9,617 and 9,634, via NCBI. Although, these proteins have different accession numbers and were identified as DNA-binding proteins and transcriptional regulators, sequence similarity was 99, 98, 97 and 78%. Considering their high-occurrence rates, we suggest the use of these proteins in MALDI-typing of *S. aureus*.

In 2013, Josten et al. attempted to characterize clonal complexes of *S. aureus* using MALDI-TOF MS, by identifying proteins in mass range of 800-12,000 m/z. Clonal lineages were correlated to peak shifts based on the nature of mutations. Accordingly, different protein isoforms resulting from peak shifts were assigned to *spa* clonal complexes and MLST clonal complexes. For example, ST80 CA-MRSA strains are characterized by a gene mutation of *S. aureus* SAR1012 ORF shift, leading to a peak at 4,497 m/z, whereas ST88 CA-MRSA strains are characterized by a peak at 4,511 m/z corresponding to the same SAR1012 ORF.

4.2 Evaluation of Protein Identification Method

Protein identification is mainly performed in two workflows: 2-D gel electrophoresis followed by PMF for each gel spot or LC MS/MS. In a previous study in our lab, El Hage (2012) identified 9 *S. aureus* proteins using the former method. According to the 2-D electrophoresis gel, the formic acid extraction yielded approximately 50 spots corresponding to one or multiple proteins. 2-D electrophoresis gel of El Hage (2012) can be viewed in the Appendix (Figure 7.6).

Protein identification by coupling Nanoflow LC and MALDI-TOF MS/MS method is distinguished by its fast and efficient identification power of proteins; this

is due to the batch interpretation mode utilized in MS/MS acquisition method. Peptide fragments generated by MALDI-TOF MS/MS were successfully assigned to 36 proteins by MASCOT, 16 of which belong to *S. aureus*. Homologues of the remaining proteins were found by NCBI nr protein blasts. Two identified proteins with molecular weight of 25,571 and 40,960 Da were not matched to *S. aureus* homologues; we conclude that these 2 proteins are not yet characterized in *S. aureus*. Additionally, unassigned peptide fragments were also blasted on NCBI nr.

Twenty-three proteins identified in this study are based on MALDI-TOF MS/MS analysis of Nanoflow LC fractions of in-solution digest of isolate LS 16. Additionally, 13 proteins were identified based on MALDI-TOF MS/MS analysis of complete in-solution digests of isolate LS 16 analyzed by MALDI-TOF MS in reflector mode in an array of mass ranges.

Only one protein, virulence factor EsxA protein, with the mass of 11,029 m/z was identified by both methods of identification; 2-D gel electrophoresis followed by PMF for gel spot 6 (Appendix, Figure 7.6) and Nanoflow LC MS/MS. Additionally, UPF0342 protein SH1117 (MW: 13348 m/z), a homologue of UPF0342 protein SA1663 (MW: 13301), which was identified by El Hage (2012), was identified in this study.

In addition to the hour NanoFlow LC separation, tryptic-digest of isolate LS 16 was subjected to a 5 hour Nanoflow LC separation, by increasing gradient time. Chromatogram of 5 hour Nanoflow LC run is shown in Appendix (Figure 7.5). As a result, the number of MASCOT assigned peptides decreased leading to less identified proteins. This phenomena can be justified by the effect of gradient length on peak capacity. For a complex peptide mixture with low concentration, longer gradient lengths produce increased peak widths, resulting in dilution of eluted peptides in

several MALDI target plate spots, thus resulting with a decreased MALDI-TOF MS signal intensity (Kocher et al., 2012). A more concentrated protein sample is recommended for longer NanoFlow LC separation, whereby longer separation gradient length was shown to share a linear relationship with number of proteins identified (Kocher et al., 2012).

4.3 Protein Identification Audit

Ninety-four proteins were identified in this study from formic acid extract of *S. aureus* LS 16. Proteins identified in this study range in mass from 4,061 to 117,130 Da. The functions of proteins identified in this study were obtained from Uniprot and NCBI protein databases and are listed in Tables 3.5 and 3.6. Thirty-three proteins that were identified in this study were not assigned a molecular weight and function (Appendix, Table 7.5). The existence of these proteins in protein databases is due to automated computational analysis using gene prediction method or conceptual translation.

The functions of the identified proteins varied vastly from DNA transcription regulation, DNA-binding, DNA-repair, kinases, proteases and few uncharacterized proteins. The majority of the identified proteins can be divided into four groups. The first group is comprised of hypothetical proteins with molecular weights of 5,864, 4,061, 40,205, 15,331, 16,235, 29,252, 13,340, 13,302, 12,256, 13,332, 13,287 m/z and 6 proteins with unassigned molecular weights. The second group consists of 6 transcriptional regulators with molecular weight of 9,652, 3,4678, 9,646, 9,650, 9,604 and 9,634 m/z. Ribosomal proteins, which constitute almost 50% of cellular proteins, comprise the third group. Six 30S ribosomal protein subunits with molecular weight of 15,128, 10,235, 10,229, 14,607 m/z and 2 proteins with unassigned

molecular weights were identified, in addition to 3 50S ribosomal protein subunits with molecular weight of 8,085 and 2 proteins with unassigned molecular weights. The forth group contains 4 DNA-binding protein HU with molecular weights of 9,620, 9,626, 40,504 m/z and one protein with unassigned molecular weight. It must be noted that isomers of the same protein, which may be present in different isolates of *S. aureus*, could have been simultaneously identified in this study by homology. For example, identified transcriptional proteins with molecular weights of 9,652, 9,646, 9,650 and 9,604 are assumed to be isomers of the same protein present in different *S. aureus* isolates; their simultaneous identification in one isolate, LS 16, is due to homology.

Josten et al. (2013) proposed that hypothetical proteins in the mass range of 800-12,000 m/z are useful for typing due to their occurrence in multiple isoforms and showed a correlation of their peak shift patterns and corresponding genotypic clonal lineages. In this study, 19 proteins were identified in this mass range. Accordingly, hypothetical protein pETB_p11 (MW: 5,864) and hypothetical protein (MW: 4,061) are considered candidates for MALDI-TOF MS typing. While ribosomal proteins are prominent in MALDI-TOF MS mass range 800-12000 m/z, they are not useful for typing, since any knock-down introduced to ribosomal genes in studying peak shift patterns leads to growth arrest and alteration of protein levels (Josten et al., 2013).

The majority of proteins identified in this study with assigned molecular weights and functions lie outside the MALDI-TOF MS range (800-12,000 m/z), e.g. 39 proteins out of 59. Hence, identified proteins were examined for virulence characteristic, given that virulent proteins make good candidates for MALDI-TOF MS typing. One virulence factor identified in this study, EsxA (MW: 11,029 Da), appears to be a promising typable protein especially that it lies in the MALDI-TOF MS range

(800-12,000 m/z). Virulence factor EsxA existence in multiple drug-resistant strains has been proven, especially for MRSA strains. Pathogenesis of this protein is related to its ability to induce antibodies in *S. aureus* infected patients. Moreover, Virulence factor EsxA makes a good target antigen for new *S. aureus* control and prevention vaccines (H. Zhou et al., 2013).

4.4 Evaluation of Protein Extraction Method

Results of correlating proteomic fingerprints to genotypic data as well as the identification process in this study are limited by the fact that formic acid does not extract the whole proteome of *S. aureus*. Moreover, the concentration of proteins extracted by formic acid is too low for analysis by NanoFlow LC. This led to absence of resolved peptide peaks on NanoFlow LC chromatogram. Hence, the chromatograms generated by ChemStation software serve to monitor the gradient only, and fractionation was time-based. Loading peptides on a trapping column before eluting them on an analytical column can overcome the low concentration challenge, since it allows the injection of a bigger sample volume at an elevated flowrate of 20 $\mu\text{L}/\text{min}$ (Kocher et al., 2012). The detection of peptide peaks on NanoFlow LC chromatogram allows for peak-based fractionation, which will result in better MS/MS spectral quality and thus identification of more proteins.

Chapter Five

Conclusion

In this study, MALDI-TOF MS spectral analysis of *S. aureus* formic acid protein extracts revealed unique proteomic profiles of 25 *S. aureus* isolates. On the other hand, no correlation was possible between proteomic profiles and genotypic data. This is possibly due to genomic mutations leading to protein peak shifts, causing protein isoforms to be falsely reported as a different protein, considering an error of 600 ppm.

Protein identification by NanoFlow LC coupled with MALDI-TOF MS/MS is a powerful, fast and efficient tool. Proteins identified in this study are involved in a wide-range of cellular functions, including transcription regulation, DNA-binding, DNA-repair, kinases, proteases, and few hypothetical proteins. We propose 3 candidate proteins for MALDI-TOF MS typing, present in MALDI-TOF MS detectable range of 800-12,000 m/z; 2 hypothetical proteins with molecular weight of 5,864 and 4,061 Da and Virulence factor Esx A with molecular weight of 11,029 Da. The former were chosen for their existence in several isoforms and the latter was chosen for its virulence function.

Further analysis of Virulence factor Esx A in large pool of *S. aureus* isolates coupled with comprehensive study of possible genomic mutations affecting its peak shift patterns is needed in future studies. Additionally, future work is recommended to proceed with a more suitable extraction protocol that allows for acquisition of concentrated whole protein content of *S. aureus*, followed by 2-Dimensional Nanoflow

LC MS/MS. 2-D LC allows for better peptides separation leading to an increased number of identified proteins.

References

- Anhalt, J., & Fenselau, C. (1975). Identification of bacteria using mass spectrometry. *Anal Chem*, 47(2), 219-225
- Abdallah, K. (2011). *A proteomic approach for fast identification of Staphylococcus aureus isolates by Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) based protein mass fingerprinting* (Unpublished master's thesis). Lebanese American University, Byblos, LB
- Bernardo, K., Pakulat, N., Macht, M., Krut, O., Seifert, H., Fler, S., . . . Kronke, M. (2002). Identification and discrimination of staphylococcus aureus strains using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Proteomics*, 2(6), 747-753. doi: 10.1002/1615-9861(200206)2:6<747::AID-PROT747>3.0.CO;2-V
- Boers, S. A., van der Reijden, W. A., & Jansen, R. (2012). High-throughput multilocus sequence typing: Bringing molecular typing to the next level. *PLoS One*, 7(7), e39630. doi: 10.1371/journal.pone.0039630 PONE-D-12-09114 [pii]
- Burts, M. L., Williams, W. A., DeBord, K., & Missiakas, D. M. (2005). Esxa and esxb are secreted by an esat-6-like system that is required for the pathogenesis of staphylococcus aureus infections. *Proc Natl Acad Sci U S A*, 102(4), 1169-1174. doi: 10.1073/pnas.0405620102
- Carbonnelle, E., Beretti, J. L., Cottyn, S., Quesne, G., Berche, P., Nassif, X., & Ferroni, A. (2007). Rapid identification of staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*, 45(7), 2156-2161. doi: JCM.02405-06 [pii] 10.1128/JCM.02405-06
- Carbonnelle, E., Mesquita, C., Bille, E., Day, N., Dauphin, B., Beretti, J. L., . . . Nassif, X. (2011). Maldi-tof mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clin Biochem*, 44(1), 104-109. doi: S0009-9120(10)00305-X [pii] 10.1016/j.clinbiochem.2010.06.017
- Cherkaoui, A., Hibbs, J., Emonet, S., Tangomo, M., Girard, M., Francois, P., & Schrenzel, J. (2010). Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *J Clin Microbiol*, 48(4), 1169-1175. doi: JCM.01881-09 [pii] 10.1128/JCM.01881-09
- Cottrell, J. S. (2011). Protein identification using ms/ms data. *J Proteomics*, 74(10), 1842-1851. doi: 10.1016/j.jprot.2011.05.014

- Degand, N., Carbonnelle, E., Dauphin, B., Beretti, J. L., Le Bourgeois, M., Sermet-Gaudelus, I., . . . Ferroni, A. (2008). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of nonfermenting gram-negative bacilli isolated from cystic fibrosis patients. *J Clin Microbiol*, 46(10), 3361-3367. doi: JCM.00569-08 [pii] 10.1128/JCM.00569-08
- Du, Z., Yang, R., Guo, Z., Song, Y., & Wang, J. (2002). Identification of staphylococcus aureus and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Chem*, 74(21), 5487-5491.
- El Hage, H. (2013). *Optimization of sample preparation method, and evaluation of formic acid protein extracts for MALDI-typing of Staphylococcus aureus* (Unpublished master's thesis). Lebanese American University, Byblos, LB.
- Gagnaire, J., Dauwalder, O., Boisset, S., Khau, D., Freydiere, A. M., Ader, F., . . . Vandenesch, F. (2012). Detection of staphylococcus aureus delta-toxin production by whole-cell maldi-tof mass spectrometry. *PLoS One*, 7(7), e40660. doi: 10.1371/journal.pone.0040660
- Goldstein, J. E., Zhang, L., Borrer, C. M., Rago, J. V., & Sandrin, T. R. (2013). Culture conditions and sample preparation methods affect spectrum quality and reproducibility during profiling of staphylococcus aureus with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Lett Appl Microbiol*, 57(2), 144-150. doi: 10.1111/lam.12092
- Harris, L. G., Foster, S. J., & Richards, R. G. (2002). An introduction to staphylococcus aureus, and techniques for identifying and quantifying s. Aureus adhesins in relation to adhesion to biomaterials: Review. *Eur Cell Mater*, 4, 39-60. doi: vol004a04 [pii]
- Hsieh, E. J., Bereman, M. S., Durand, S., Valaskovic, G. A., & MacCoss, M. J. (2013). Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow lc-ms/ms of complex proteomic samples. *J Am Soc Mass Spectrom*, 24(1), 148-153. doi: 10.1007/s13361-012-0508-6
- Hsieh, S. Y., Tseng, C. L., Lee, Y. S., Kuo, A. J., Sun, C. F., Lin, Y. H., & Chen, J. K. (2008). Highly efficient classification and identification of human pathogenic bacteria by maldi-tof ms. *Mol Cell Proteomics*, 7(2), 448-456. doi: M700339-MCP200 [pii] 10.1074/mcp.M700339-MCP200
- Hussain, M., von Eiff, C., Sinha, B., Joost, I., Herrmann, M., Peters, G., & Becker, K. (2008). Eap gene as novel target for specific identification of staphylococcus aureus. *J Clin Microbiol*, 46(2), 470-476. doi: JCM.01425-07 [pii] 10.1128/JCM.01425-07
- Josten, M., Reif, M., Szekat, C., Al-Sabti, N., Roemer, T., Sparbier, K., . . . Bierbaum, G. (2013). Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of staphylococcus aureus identifies mutations that allow

- differentiation of the main clonal lineages. *J Clin Microbiol*, 51(6), 1809-1817. doi: 10.1128/JCM.00518-13
- Kaufmann, R. (1995). Matrix-assisted laser desorption ionization (maldi) mass spectrometry: A novel analytical tool in molecular biology and biotechnology. *J Biotechnol*, 41(2-3), 155-175.
- Khatun, J., Ramkissoon, K., & Giddings, M. C. (2007). Fragmentation characteristics of collision-induced dissociation in maldi tof/tof mass spectrometry. *Anal Chem*, 79(8), 3032-3040. doi: 10.1021/ac061455v
- Kocher, T., Pichler, P., Swart, R., & Mechtler, K. (2012). Analysis of protein mixtures from whole-cell extracts by single-run nanolc-ms/ms using ultralong gradients. *Nat Protoc*, 7(5), 882-890. doi: 10.1038/nprot.2012.036
- Lark, R. L., Saint, S., Chenoweth, C., Zemencuk, J. K., Lipsky, B. A., & Plorde, J. J. (2001). Four-year prospective evaluation of community-acquired bacteremia: Epidemiology, microbiology, and patient outcome. *Diagn Microbiol Infect Dis*, 41(1-2), 15-22. doi: S0732-8893(01)00284-X [pii]
- Lista, F., Reubsaet, F. A., De Santis, R., Parchen, R. R., de Jong, A. L., Kieboom, J., . . . Paauw, A. (2011). Reliable identification at the species level of brucella isolates with maldi-tof-ms. *BMC Microbiol*, 11, 267. doi: 1471-2180-11-267 [pii] 10.1186/1471-2180-11-267
- Liu, H., Du, Z., Wang, J., & Yang, R. (2007). Universal sample preparation method for characterization of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl Environ Microbiol*, 73(6), 1899-1907. doi: AEM.02391-06 [pii] 10.1128/AEM.02391-06
- Lowy, F. D. (2000). Is staphylococcus aureus an intracellular pathogen? *Trends Microbiol*, 8(8), 341-343. doi: S0966-842X(00)01803-5 [pii]
- Meyer, H., Liebeke, M., & Lalk, M. (2010). A protocol for the investigation of the intracellular staphylococcus aureus metabolome. *Anal Biochem*, 401(2), 250-259. doi: 10.1016/j.ab.2010.03.003
- Moon, M. H., Myung, S., Plasencia, M., Hilderbrand, A. E., & Clemmer, D. E. (2003). Nanoflow lc/ion mobility/cid/tof for proteomics: Analysis of a human urinary proteome. *J Proteome Res*, 2(6), 589-597.
- Rajakaruna, L., Hallas, G., Molenaar, L., Dare, D., Sutton, H., Encheva, V., . . . Shah, H. N. (2009). High throughput identification of clinical isolates of staphylococcus aureus using maldi-tof-ms of intact cells. *Infect Genet Evol*, 9(4), 507-513. doi: S1567-1348(09)00013-6 [pii] 10.1016/j.meegid.2009.01.012
- Saffert, R. T., Cunningham, S. A., Ihde, S. M., Jobe, K. E., Mandrekar, J., & Patel, R. (2011). Comparison of bruker biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometer to bd phoenix automated

- microbiology system for identification of gram-negative bacilli. *J Clin Microbiol*, 49(3), 887-892. doi: JCM.01890-10 [pii] 10.1128/JCM.01890-10
- Sauer, S., & Kliem, M. (2010). Mass spectrometry tools for the classification and identification of bacteria. *Nat Rev Microbiol*, 8(1), 74-82. doi: nrmicro2243 [pii] 10.1038/nrmicro2243
- Seng, P., Drancourt, M., Gourié, F., La Scola, B., Fournier, P. E., Rolain, J. M., & Raoult, D. (2009). Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis*, 49(4), 543-551. doi: 10.1086/600885
- Sitepu, I. R., Ignatia, L., Franz, A. K., Wong, D. M., Faulina, S. A., Tsui, M., . . . Boundy-Mills, K. (2012). An improved high-throughput Nile red fluorescence assay for estimating intracellular lipids in a variety of yeast species. *J Microbiol Methods*, 91(2), 321-328. doi: 10.1016/j.mimet.2012.09.001
- Sundaramoorthy, R., Fyfe, P. K., & Hunter, W. N. (2008). Structure of staphylococcus aureus esxa suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein. *J Mol Biol*, 383(3), 603-614. doi: 10.1016/j.jmb.2008.08.047
- Szabados, F., Woloszyn, J., Richter, C., Kaase, M., & Gatermann, S. (2010). Identification of molecularly defined staphylococcus aureus strains using matrix-assisted laser desorption/ionization time of flight mass spectrometry and the biotyper 2.0 database. *J Med Microbiol*, 59(Pt 7), 787-790. doi: 10.1099/jmm.0.016733-0
- Tang, Y. W. (2009). Progress toward rapid and accurate staphylococcus aureus strain typing. *Clin Chem*, 55(12), 2074-2076. doi: clinchem.2009.135707 [pii] 10.1373/clinchem.2009.135707
- Tani, A., Sahin, N., Matsuyama, Y., Enomoto, T., Nishimura, N., Yokota, A., & Kimbara, K. (2012). High-throughput identification and screening of novel methylobacterium species using whole-cell MALDI-TOF/MS analysis. *PLoS One*, 7(7), e40784. doi: 10.1371/journal.pone.0040784 PONE-D-12-11302 [pii]
- Theel, E. S., Hall, L., Mandrekar, J., & Wengenack, N. L. (2011). Dermatophyte identification using matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*, 49(12), 4067-4071. doi: JCM.01280-11 [pii] 10.1128/JCM.01280-11
- Todar, K. (2009). Todar's online textbook of bacteriology. Retrieved from <http://textbookofbacteriology.net/staph.html>
- Wolters, M., Rohde, H., Maier, T., Belmar-Campos, C., Franke, G., Scherpe, S., . . . Christner, M. (2011). MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant staphylococcus aureus lineages. *Int J Med Microbiol*, 301(1), 64-68. doi: 10.1016/j.ijmm.2010.06.002

- Yang, Y. C., Yu, H., Xiao, D. W., Liu, H., Hu, Q., Huang, B., . . . Huang, W. F. (2009). Rapid identification of staphylococcus aureus by surface enhanced laser desorption and ionization time of flight mass spectrometry. *J Microbiol Methods*, 77(2), 202-206. doi: S0167-7012(09)00041-4 [pii] 10.1016/j.mimet.2009.02.004
- Zetola, N., Francis, J. S., Nuermberger, E. L., & Bishai, W. R. (2005). Community-acquired meticillin-resistant staphylococcus aureus: An emerging threat. *Lancet Infect Dis*, 5(5), 275-286. doi: S1473-3099(05)70112-2 [pii] 10.1016/S1473-3099(05)70112-2
- Zhou, F., Lu, Y., Ficarro, S. B., Webber, J. T., & Marto, J. A. (2012). Nanoflow low pressure high peak capacity single dimension lc-ms/ms platform for high-throughput, in-depth analysis of mammalian proteomes. *Anal Chem*, 84(11), 5133-5139. doi: 10.1021/ac2031404
- Zhou, H., Du, H., Zhang, H., Shen, H., Yan, R., He, Y., . . . Zhu, X. (2013). Esxa might as a virulence factor induce antibodies in patients with staphylococcus aureus infection. *Braz J Microbiol*, 44(1), 267-271. doi: 10.1590/S1517-83822013005000019

Appendix

Figure 7.1. BCA standard curved from which protein concentrations of isolates JS 03, JS 08, JS 10, JS 17, JS 22, JS 50 and JS 52 were extrapolated from.

Figure 7.2. BCA standard curved from which protein concentrations of isolates JS 54, JS 56, JS 60, JS 61, JS 65 and JS 85 were extrapolated from.

Figure 7.3. BCA standard curved from which protein concentrations of isolates LS 16, LS 24, LS 34, LS 36, LS 57 and LS 79 were extrapolated from.

Figure 7.4. BCA standard curved from which protein concentrations of *S. epidermidis*, *E. coli* and *S. aureus* isolate LS 16 were extrapolated.

Table 7.1. Trypsin-digested BSA Standard and trypsin-digested LS 16 inclusion lists utilized in MALDI-TOF MS/MS Interpretation method.

Trypsin-Digested BSA Standards	Trysin-Digested LS 16	
400.3659	400.9892	1658.6572
415.7304	409.2798	1692.9045
432.0742	415.1955	1729.9358
432.9656	437.1907	1735.8422
463.0389	440.2151	1740.7375
497.6936	441.0574	1880.8461
499.1149	441.2589	1881.9633
523.3412	459.2380	1911.9417
541.6649	469.2853	1917.8364
525.1186	476.8908	1918.6981
598.0327	477.2423	1935.8468
604.0181	482.2712	1957.7061
642.8904	516.2645	1968.8343
643.6831	564.1211	1973.6842
650.7220	569.1252	1974.7076
800.2753	671.9273	1992.7404
816.6121	687.9029	1995.7043
820.2831	841.3866	1996.6979
783.8229	852.4586	2060.0527
793.6814	854.2908	2065.0491
821.2779	868.4467	2104.1035
892.6742	883.3168	2224.8964
901.2573	895.4429	2274.1482
951.9488	912.3694	2282.5722
962.2470	920.4929	2330.2117
973.9276	921.3868	2363.5114
992.3522	928.4517	2446.9229
1081.0747	934.5127	2545.0062
1089.4435	954.4488	2557.1691
1108.2006	1011.3958	2558.1724
1171.1273	1095.4201	2560.6564
1186.8218	1306.7030	2594.1483
1189.8515	1320.4968	2602.0212
1214.8206	1357.5681	2603.0171
1231.2137	1361.7376	2612.2791
1244.5976	1378.6877	2617.0286
1260.5962	1380.7633	2635.9960
1268.8035	1400.5566	2636.9885
1275.2923	1416.5301	2697.2922
1294.2341	1435.5941	2714.0079
1309.1736	1463.7559	2758.3713
1197.1156	1480.7234	2797.2835
1327.5022	1601.7405	2845.2769
1370.8749		3372.5111
1281.8655		
1434.5519		
1442.2743		
1447.2046		
1451.3566		
1466.6840		
1472.9178		
1477.8373		
1481.8613		
1544.1682		
1596.1889		
1612.2456		
1624.6964		

Table 7.2. CHCA matrix exclusion list utilized in MALDI-TOF MS/MS Interpretation method.

CHCA Matrix Exclusion	
401.0392	669.3860
402.3249	685.3578
413.2334	713.4137
416.3138	714.2794
417.0104	729.3738
423.0239	730.4315
428.3367	757.4382
429.2236	758.4529
430.3585	773.4007
432.9860	774.4645
438.9992	798.5811
442.3173	801.4617
442.9858	802.4811
444.3380	817.4243
445.0007	818.4971
445.9857	826.6236
446.9829	829.2689
447.9823	833.0397
448.9814	839.0522
450.9806	842.6113
452.9804	845.4854
454.9711	846.5159
460.9903	855.0192
480.1221	861.0352
493.2947	861.4551
504.2826	862.5279
506.1058	870.6411
507.1683	870.9956
522.0987	877.0112
523.0917	881.2251
524.1136	886.6346
526.1271	889.5095
537.3210	890.5449
550.0951	906.5539
552.1107	914.6673
566.0856	930.6591
567.0877	933.5286
568.1056	934.5731
570.0988	950.5784
572.0860	958.6900
572.4180	974.6819
578.6308	978.5985
581.3305	994.6036
590.0927	1002.7141
596.0506	1018.7061
598.0504	1022.6254
599.0502	1038.6269
600.0490	1044.0508
601.0500	1046.7365
602.0502	1050.0660
603.0366	1060.0129
604.0487	1062.7255
605.0485	1066.0282
606.0553	1066.6468
607.0593	1082.0056
618.2561	1082.6546
622.0271	1090.7610
625.3666	1106.7473
628.0410	1110.6747
634.0290	1126.6728
635.0242	1134.7675
636.0256	1150.7644
637.0262	1154.6993
638.0206	1170.6979
639.0265	1249.0397
640.0216	1255.0361
641.0311	1271.0158
644.0040	1277.0366
644.9932	1337.7323
650.0180	1475.9990
656.0395	1482.0018
665.9916	2161.7256
	2162.7834

Table 7.3. Fractionation pattern of trypsin-digested BSA Standard. Column 1 show BSA peptide peaks obtained from BioEngland Labs Manual of trypsin-digested BSA Standard. Columns A1 to B10 refer to one hour Nanoflow LC fractions analyzed by MALDI-TOF MS in reflector mode in mass range of 400-4,000 m/z, prior to MALDI-TOF MS/MS analysis.

[illegible]

Figure 7.5. Nanoflow LC chromatogram of trypsin-digested LS 16. Injection volume: 8 μ L. Five hour linear gradient of 2-45% solvent B. Detection wavelength: 210 nm. Time-based fractionation: 3min.

Table 7.4. Proteins identified from trypsin-digested LS 16 formic acid extract by MASCOT search engine.

No.	Database MW (Da)	ID	Protein	Species	Database	Protein Score	% Sequence Coverage	Peptide mass	Peptide score	Peptide sequence
1	9620	DBH_STAAC	DNA-binding protein HU	<i>Staphylococcus aureus</i>	SwissProt	1002	66	1378.7079	(57)	K.VQLIGFGNFEVR.E
								1381.7021	(36)	K.VQLIGFGNFEVR.E + [+2.9883 at Q2]
								1601.8016	(33)	K.TDLINA VAEQADLT.K
								1601.8126	(12)	K.TDLINA VAEQADLT.K
								1601.8127	(20)	K.TDLINA VAEQADLT.K
								1601.8132	43	K.TDLINA VAEQADLT.K
								1729.8931	61	K.TDLINA VAEQADLT.K.E
								1735.7988	(42)	K.TDLINA VAEQADLT.K.E + [+6.0201 at L3]
								1736.8186	(33)	K.TDLINA VAEQADLT.K.E + [+7.0172 at L13]
								1917.9169	107	K.EAGSA VDA VFESIQNSLA.K.G + [-18.0106 at N-term E]
								1918.9213	(34)	K.EAGSA VDA VFESIQNSLA.K.G + [-16.9902 at N15]
								1934.8983	(48)	K.EAGSA VDA VFESIQNSLA.K.G + [-0.9476 at E11]
								1935.8981	(92)	K.EAGSA VDA VFESIQNSLA.K.G
								1957.8936	(50)	K.EAGSA VDA VFESIQNSLA.K.G + [+21.9819 at E11]
								2063.9854	13	K.KEAGSA VDA VFESIQNSLA.K.G
2	13348	Y1117_STAHI	UPF0342 protein SH1117	<i>Staphylococcus haemolyticus</i>	SwissProt	154	20	954.4895	52	K.KLFDEF.R.E
								1880.8973	(53)	M.AVNLYDYANQLEQALR.D
								1881.9097	56	M.AVNLYDYANQLEQALR.D + [+0.9970 at E12]
3	9651	DBH1_BACHD	DNA-binding protein HU-1	<i>Bacillus halodurans</i>	SwissProt	101	17	1601.8132	43	K.TDLINA VAEQAELSK.K
								1729.8931	43	K.TDLINA VAEQAELSK.K.D
4	29580	Y235_STAEQ	UPF0447 protein SERP0235	<i>Staphylococcus epidermidis</i>	SwissProt	35	2	954.4895	35	K.KIVYEMR.F + Oxidation (M)
5	28960	Y3362_GEOTN	UPF0447 protein GTNG_3362	<i>Geobacillus thermodenitrificans</i>	SwissProt	35	2	954.4895	35	K.KLVYEMR.F + Oxidation (M)
6	88841	PBPA_CLOAB	Penicillin-binding protein 1A	<i>Clostridium acetobutylicum</i>	SwissProt	25	1	1729.9060	25	K.QNSIQGASTITQQLIK.N
7	107792	SYL_HALOH	Isoleucine--tRNA ligase	<i>Halothermothrix orenii</i>	SwissProt	15	1	1217.6532	15	K.LVMARELVDR.V + Oxidation (M)
8	40023	GCST_GEOKA	Aminomethyltransferase	<i>Geobacillus kaustophilus</i>	SwissProt	15	2	1217.6532	15	R.RLVGEMIDR.G + Oxidation (M)
9	9619	gi488404375	transcriptional regulator	<i>Staphylococcus epidermidis</i>	NCBI nr	230	66	1378.7394	61	K.VQLIGFGNFEVR.E
								1414.7581	17	K.EIDIPASKVPFAK.A
								1601.8132	43	K.TDLINA VAEQADLT.K
								1729.8931	61	K.TDLINA VAEQADLT.K.E
								1934.8983	(48)	K.EAGSA VDA VFESIQNSLA.K.G
								1935.8967	86	K.EAGSA VDA VFESIQNSLA.K.G + [+0.9970 at S12]

No.	Database MW (Da)	ID	Protein	Species	Database	Protein Score	% Sequence Coverage	Peptide mass	Peptide score	Peptide sequence
10	9646	gi446966605	transcriptional regulator	<i>Staphylococcus aureus</i>	NCBI	182	66	1378.7394	61	K.VQLGFGNFEVR.E
								1414.7581	17	K.EIDP AS KVP AFK.A
								1601.8132	43	K.TDLNAVAEQADLT.K
								1729.8931	61	K.TDLNAVAEQADLT.K
								1917.9169	61	K.EAGLAVDAVFESQNSLAK.G + [-43.9898 at D7]
11	9650	gi488380409	transcriptional regulator	<i>Staphylococcus</i>	NCBI	166	52	1378.7394	61	K.VQLGFGNFEVR.E
								1601.8132	43	K.TDLNAVAEQADLT.K
								1729.8931	61	K.TDLNAVAEQADLT.K
								1935.9030	73	K.EAGS AVDAVFESQNSLSK.G + [-16.0313 at L17]
12	13302	gi15924835	hypothetical protein SAV1845	<i>Staphylococcus aureus</i>	NCBI	104	20	954.4895	52	K.KLFDEF.R
								1880.8973	(53)	MAVNLYDYANQLEQAL.R
								1881.9097	56	MAVNLYDYANQLEQAL.R + [+0.9970 at E12]
13	12256	gi532359921	hypothetical protein SAAG_01746	<i>Staphylococcus aureus</i>	NCBI	104	21	954.4895	52	K.KLFDEF.R
								1880.8973	(53)	MAVNLYDYANQLEQAL.R
								1881.9097	56	MAVNLYDYANQLEQAL.R + [+0.9970 at E12]
14	13332	gi446212447	hypothetical protein	<i>Staphylococcus aureus</i>	NCBI	104	20	954.4895	52	K.KLFDEF.R
								1880.8973	(53)	MAVNLYDYANQLEQAL.R
								1881.9097	56	MAVNLYDYANQLEQAL.R + [+0.9970 at E12]
15	13287	gi379796200	hypothetical protein SAMSHR1132_16910	<i>Staphylococcus aureus</i>	NCBI	104	20	954.4895	52	K.KLFDEF.R
								1880.8973	(53)	MAVNLYDYANQLEQAL.R
								1881.9097	56	MAVNLYDYANQLEQAL.R + [+0.9970 at E12]
16	9805	gi90962747	DNA-binding protein HU	<i>Lactobacillus salivarius</i>	NCBI	61	13	1378.7394	61	K.VQLGFGNFEVR.E
17	9678	gi497173090	DNA-binding protein HU	<i>Weissella ceti</i>	NCBI	61	34	1378.7394	61	K.VQLGFGNFEVR.S
								1935.9169	76	K.DAAS AVDAVFSSQDSLAK.G + [+41.0265 at S11]
18	14040	gi222151808	hypothetical protein MCCL_1563	<i>Macrococcus caseolyticus</i>	NCBI	52	5	954.4895	52	K.KLFDEF.R
19	25571	gi553957006	hypothetical protein	<i>Eubacterium brachy</i>	NCBI	52	3	954.4895	52	K.KLFDEF.R
20	9617	gi319892463	DNA-binding protein HBsu	<i>Staphylococcus pseudintermedius</i>	NCBI	206	52	1378.7394	80	K.VQLGFGNFEVR.E
								1381.7084	(53)	K.VQLGFGNFEVR.E + [+1.9941 at Q2]
								1601.7924	51	K.TDLNAVAEQADLT.K
								1729.8931	80	K.TDLNAVAEQADLT.K + [+0.9589 at H]
								1730.8715	(63)	K.TDLNAVAEQADLT.K
								1917.8853	56	K.EAGLAVDAVFIAQTS LAK.G + [-16.0313 at L4]

No.	Database MW (Da)	ID	Protein	Species	Database	Protein Score	% Sequence Coverage	Peptide mass	Peptide score	Peptide sequence
21	9604	gi446966606	transcriptional regulator	<i>Staphylococcus aureus</i>	NCBItr	206	52	1378.7394	80	K.VQLGFGNFEVR.E
								1381.7084	(53)	K.VQLGFGNFEVR.E + [+19941 at Q2]
								1601.7924	51	K.TDLNVAEQADLT.K
								1729.8931	80	K.TDLNVAEQADLT.KK.E + [+0.9589 at H]
								1730.8580	(34)	K.TDLNVAEQADLT.KK.E
								1730.8715	(63)	K.TDLNVAEQADLT.KK.E
								1957.8710	62	K.EAGS AVDAVFEAKNSLAK.G + [+37.9559 at E11]
22	9634	gi488364546	transcriptional regulator	<i>Staphylococcus</i>	NCBItr	206	52	1378.7394	80	K.VQLGFGNFEVR.E
								1381.7084	(53)	K.VQLGFGNFEVR.E + [+19941 at Q2]
								1601.7924	51	K.TDLNVAEQADLT.K
								1729.8931	80	K.TDLNVAEQADLT.KK.E + [+0.9589 at H]
								1730.8715	(63)	K.TDLNVAEQADLT.KK.E
								1918.8909	(54)	K.EAGS AVDAVFESQKSLAK.G + [-31.0422 at K15]
								1957.8608	61	K.EAGS AVDAVFESQKSLAK.G + [+8.0142 at K15]
23	44043	TGT_BACLD	Queuine tRNA-ribosyltransferase	<i>Bacillus licheniformis</i>	SwissProt	14	3	1511.6504	14	R.LTS YHNLFLK.L
24	43882	BRPA_STRMU	Biofilm regulatory protein A	<i>Streptococcus mutans serotype c</i>	SwissProt	33	1	954.6473	33	K.TMTSLER.D + Oxidation (M)
25	8085	RL29_STAA2	50S ribosomal protein L29	<i>Staphylococcus aureus</i>	SwissProt	25	28	920.6347	8	K.EELFNLR.F
								1464.0115	25	R.FQLATGQLEETAR.I
26	10279	Y1445_STAAN	UPF0297 protein	<i>Staphylococcus aureus</i>	SwissProt	23	13	1511.9556	23	K.FDYELPTQDVR.D
27	11029	ESXA_STAAC	Virulence factor EsxA	<i>Staphylococcus aureus</i>	SwissProt	23	27	1320.8972	4	MAMKMSPEER.A + Oxidation (M)
								1735.0984	23	R.AQGEIAANWEGQAFS.R.F
28	10229	RS16_STAA1	30S ribosomal protein S16	<i>Staphylococcus aureus</i>	SwissProt	23	27	852.6048	10	K.RNPFYR.I
								2059.4185	23	R.IEIQGTYNPTS ANAPEIK.V
29	14607	RS9_STAA1	30S ribosomal protein S9	<i>Staphylococcus aureus</i>	SwissProt	14	16	2273.5215	14	K.GNYDVLNVHGGGFTGQAQIR.H
30	31122	gi557828691	2-dehydropantoate 2-reductase	<i>Exiguobacterium sp. MH3</i>	NCBItr	56	2	954.6473	56	R.QLFDEMR.T + Oxidation (M)
31	11271	SP5G_STAA2	Putative septation protein SpoVG	<i>Staphylococcus aureus</i>	SwissProt	39	17	1912.244	39	K.ALVSITLDEAFVHDLR.V
32	13150	gi492418262	Fic family protein	<i>Eubacterium saphenum</i>	NCBItr	59	6	954.594	59	K.KMDEAR.N + 2 Oxidation (M)
33	40960	gi495696491	hydrogenase expression/formation protein HypD	<i>Clostridium sp. Maddingley MBC34-26</i>	NCBItr	59	1	826.2953	34	K.FIDEFR.N
								954.3608	24	M.KFIDEFR.N
34	28880	gi493632938	DNA repair protein RecO	<i>Lactobacillus jensenii</i>	NCBItr	56	2	826.2862	31	K.FLDEMR.R + Oxidation (M)
								954.3587	25	K.KFLDEMR.R + Oxidation (M)
35	34894	HPRK_CLOPE	HPr kinase/phosphorylase	<i>Clostridium perfringens</i>	SwissProt	38	3	1378.5558	38	R.SPEITFGMLEVR.G
36	7770	Y1696_STAAN	UPF0435 protein	<i>Staphylococcus aureus</i>	SwissProt	33	25	1740.7404	34	R.LSPSEVTAADALGQLR.H

Table 7.5. Proteins identified in this study that were not assigned a molecular mass and a function.

No.	ID	Protein
1	gi 583105701	methionine sulfoxide reductase B, partial
2	gi 579935718	acetyl-CoA acetyltransferase, partial
3	gi 577289267	chaperonin, partial
4	gi 580848717	Oxidoreductase
5	gi 584847631	hypothetical protein U620_02608
6	gi 580689402	PTS system, ascorbate-specific IIC
7	gi 580965588	coenzyme A disulfide reductase, partial
8	gi 581796105	protein translocase subunit secA 2
9	gi 584655548	50S ribosomal protein L11
10	gi 580927722	hypothetical protein U042_02830
11	gi 582606875	hypothetical protein U343_02428, partial
12	gi 582616066	DNA topoisomerase 4 subunit B, partial
13	gi 582450771	hypothetical protein U271_01845
14	gi 582657440	DNA topoisomerase 3, partial
15	gi 580256465	50S ribosomal protein L17
16	gi 578670717	PTS system EIIBC component, partial
17	gi 581372899	30S ribosomal protein S11, partial
18	gi 581372899	30S ribosomal protein S11, partial
19	gi 579648211	cysteinyI-tRNA synthetase, partial
20	gi 487721653	GTP diphosphokinase
21	gi 580654011	6,7-dimethyl-8-ribityllumazine synthase
22	gi 580840478	hypothetical protein T991_02716
23	gi 582796996	hypothetical protein U433_02585
24	gi 580175801	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
25	gi 584237569	cation:proton antiporter
26	gi 584237570	cation:proton antiporter
27	gi 580754041	transposase for insertion sequence-like element IS431mec
28	gi 584763348	kinase-associated protein B
29	gi 580255613	DNA-binding protein HU
30	gi 446015501	glycine cleavage system protein T
31	gi 580255613	DNA-binding protein HU
32	gi 579740496	queuine tRNA-ribosyltransferase
33	gi 507417608	PfkB family carbohydrate kinase

Figure 7.6. Two dimensional electrophoresis (2-DE) representative gel of *S. aureus* isolate LS-16 performed by El Hage (2013).