Lebanese American University

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

By

Khaled A. Abdallah

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Molecular Biology

School of Arts and Sciences

Spring 2011

LEBANESE AMERICAN UNIVERSITY

Thesis approval Form (Annex III)

Student Name: Khaled Abdallah I.D. #: 200804569

Thesis Title : "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."

Program : MS in Molecular Biology

Division/Dept : Natural Sciences

School : **School of Arts and Sciences**

Approved by:

Thesis Advisor: [redacted]

Member : [redacted]

Member :

Member :

Date 06.06.2011

(This document will constitute the first page of the Thesis)

THESIS /PROJECT COPYRIGHT RELEASE FORM

LEBANESE AMERICAN UNIVERSITY

By signing and submitting this license, I (the author(s) or copyright owner) grant the Lebanese American University (LAU) the non-exclusive right to reproduce, translate (as defined below), and/or distribute my submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video. I agree that LAU may, without changing the content, translate the submission to any medium or format for the purpose of preservation. I also agree that LAU may keep more than one copy of this submission for purposes of security, backup and reservation. I represent that the submission is my original work, and that I have the right to grant the rights contained in this license. I also represent that my submission does not, to the best of my knowledge, infringe upon anyone's copyright. If the submission contains material for which I do not hold copyright, I represent that I 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, I REPRESENT THAT I HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT. LAU will clearly identify my 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 my submission.

Name: Khaled Abdallah

Signature: 

Date: 7/06/2011

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: Khaled Abdallah

Signature: 

Date: 7/06/2011

ACKNOWLEDGMENTS

This thesis would not have been accomplished without the supervision, assistance and guardianship of my professors, colleagues, family and friends.

First and foremost I like to offer my sincerest gratitude to my advisors, Dr. Brigitte Wex and Dr. Sima Tokajian whose support and constant motivation were bedrocks on which every aspect of my post-graduate study heavily relied. They have been more than advisors to me; very dear friends. Friends to whom much of this thesis would not have flourished if it were not for their encouragement and support. Genuine gratitude is also expressed to Dr. Roy Khalaf for his support and participation in the evaluation committee.

I wish to take this opportunity to earnestly thank Engineer Mazen Naamani for development of the data processing program and to thank my colleagues who gave away time and shared their knowledge to help me complete my thesis.

I am also thankful for all the faculty members at the Biology department, who have made my academic experience rich and memorable.

My deepest appreciation goes to my best friends Omar Al Hafez, Rami Al Oweini, Mohammad Habbal, Mahmoud Ghalayini and Khalid Shawwa for their patience, love and full support.

Finally, always indebted to a loving family whose nurture, support and love drove me limits afar and raised me mountains high, I dedicate this achievement and promise them ongoing successful accomplishments. Thank you Dad, Mom, Mohammad, Hanadi and Riham, your presence radiates my life and makes me a better person.

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

Khaled A. Abdallah

Abstract

Staphylococcus aureus is an important human pathogen that is capable of causing a wide range of diseases, from relatively mild skin infections to deep systemic infections such as pneumonia. Identification and typing of *S. aureus* is currently based on various genomic and phenotypic methods; however, these methods are time-consuming, expensive and sometimes lack accuracy. A total of 14 *S. aureus* isolates were previously identified by conventional methods including multiple locus sequence typing (MLST), *SCCmec* typing and *spa* typing. Our purpose is to evaluate the use of Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) for rapid identification and analysis of these *S. aureus* isolates. Formic acid extraction was used to extract proteins which were then subjected to MALDI-TOF analysis. An identification profile for each isolate was created and 2 programs were developed for further analysis of the obtained data. With this approach setting stringent parameters on reproducibility of data, we were able to identify the masses of 5

peaks (4823, 6428, 6612, 8903, and 9644 m/z) that are common to all *S. aureus* isolates. To determine whether these 5 peaks were unique to *S. aureus*, the same procedure of protein extraction and data analysis was applied on *S. epidermidis and Escherichia coli.* Compared to the m/z list of *S. aureus*, 3 peaks were discovered to be common in *S. epidermidis* but none in *E. coli.* Thus we were able to identify 3 peaks unique for the genus *Staphylococcus* and 2 peaks unique for the species *S. aureus*. Also, our results suggest that the program developed is valid and very powerful in analyzing high-through put data obtained from MALDI analysis. Moreover, MALDI-TOF MS has high potential to be used as a proteomic tool for the identification of bacteria. Future work includes identification of peaks specific at the strain level and Tandem-MS of the unique peaks to determine their nature.

Keywords: *Staphylococcus aureus*, Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF), Mass Spectrometry, Proteomics, Genomics, m/z value.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

GLOSSARY

CA-MRSA: community-acquired methicillin-resistant *Staphylococcus aureus*

E. coli: *Escherichia coli*

FAE: Formic Acid Extract

HA-MRSA: hospital-acquired methicillin-resistant *Staphylococcus aureus*

MALDI-TOF MS: Matrix Assisted Laser Desorption Ionization Method Time of Flight Mass Spectrometry

MLST: Multi-locus sequence typing

MRSA: methicillin-resistant *Staphylococcus aureus*

MSSA: methicillin-sensitive *Staphylococcus aureus*

m/z: mass to charge ratio

PCR: polymerase chain reaction

PFGE: Pulsed-field gel electrophoresis

PVL: Panton-Valentine leukocidin

S. aureus: *Staphylococcus aureus*

S. epidermidis: *Staphylococcus epidermidis*

SCC: staphylococcal chromosomal cassette

***spa* CC:** *spa* clonal complex

spa: gene for Staphylococcal Protein A

ST: Sequence type

TSA: Tryptone Soy Agar

Chapter One

INTRODUCTION

1.1 Overview of *Staphylococcus aureus*

Staphylococcus aureus is a facultative anaerobic gram-positive bacterium that has been discovered in Aberdeen, Scotland, in 1880 by the surgeon Sir Alexander Ogston. In 1884, Rosenbach isolated and cultured these bacteria from abscesses and entitled them *S. aureus* because of the yellow-orange or ‘‘gold’’ pigmented appearance of the colonies, thus the Latin word ‘‘*aureus*’’ (van Belkum *et al.*, 2009). *S. aureus* are non-spore forming cocci characterized by their non-flagellated, spherical structure and resemble grapelike structures when observed under the microscope (Seybold *et al.*, 2006). Moreover, they are known to be catalase-positive (Todar, 2008) with an optimal growth temperature between 15 to 45 °C. The majority of *S. aureus* strains are opportunistic pathogens that colonise the human skin and nares for either a short or extended periods of time without showing any symptoms. However, once the immune system of the

individual becomes compromised, infection is inevitable (Oliveira, Milheirico, & de Lencastre, 2006).

Since its discovery, *S. aureus* was found to be an important cause of serious community and health-care-associated infections (Thwaites *et al.*, 2011). Such infections can range from relatively minor infections of the skin to life-threatening conditions such as wound infections, bacteraemia, infections of the central nervous system, respiratory and urinary tracts, and infections associated with intravascular devices and foreign bodies (Oliveira *et al*., 2006).

1.2 Pathogenesis of Infection

S. aureus infections occur after several phases of pathogenesis. Initial colonization of the bacteria can lead to local sores of skin. The infection can remain local or disseminate to reach the bloodstream, facilitating metastasis into any organ. Afterwards, toxinosis or septic shock can develop. If not treated properly, these systemic infections result in high mortality rates. In addition, *S. aureus* toxins can lead to serious disorders without the necessity of bacterial invasion into the

bloodstream, such as toxic shock syndrome and scalded skin syndrome. Therefore, the control of virulent *S. aureus* infections is of major significance (Archer, 1998).

1.3 Virulence Factors

Increases in the mortality and morbidity rate in hospitals worldwide have been associated with the emergence of Methicillin-resistant *S. aureus* (MRSA) strains (Wolters *et al.*, 2011). MRSA strains have developed due to the presence of *mecA* gene, that codes for altered penicillin-binding protein (PBP2a) which in turn has a decreased affinity to β-lactam antibiotics (Makgotlho *et al.*, 2009; Decker, 2008). *mecA* is located on a mobile genetic element called staphylococcal cassette chromosome *mec* (SCC*mec*) (Makgotlho *et al.*, 2009). Among the eight identified types of SCC*mec* (I-VIII) (Chambers & Deleo, 2009), community acquired MRSA (CA-MRSA) mostly possesses the type IV or V, while hospital acquired MRSA (HA-MRSA) contain types I, II, and III (Decker, 2008).

S. aureus' capability to produce many virulence factors contribute to its high pathogenicity (Palavecino, 2007). Panton-Valentine leukocidin (PVL) is one of the virulence factors engaged in the pathogenesis of most *S. aureus* infections (Gillet *et al.*, 2002). PVL is an exotoxin that is considered a genetic indicator for CA-MRSA strains, which causes more severe diseases than HA-MRSA due to the production of this toxin (Vandenesch *et al.*, 2003). Other than CA-MRSA infections, the importance of PVL lies in its association with necrotizing pneumonia and skin infections, such as cutaneous abscesses and cellulitis (Labandeira-Rey *et al.*, 2007). PVL belongs to the synergohymenotropic toxins family where it damages neutrophils by forming pores in membranes through the synergistic process of two unrelated proteins, LukS-PV and LukF-PV (Genestier *et al.*, 2005).

An essential virulence factor found in all *S. aureus* strains is the Staphylococcal protein A (Spa). Protein A is 42-kD bound to the cell wall whereas its extracellular portion binds the Fc-region of immunoglobulin IgG of diverse mammalian species (Patel, Nowlan, Weavers, & Foster, 1987). Protein A plays important roles in *S. aureus* pathogenesis by preventing phagocytosis, stimulating B lymphocytes proliferation, and inducing inflammatory reactions in

airways and corneal epithelial cells (Gustafsson *et al.*, 2009). Other virulence factors of *S. aureus* include the toxic shock syndrome toxin-1 (TSST-1) and the enterotoxins which stimulate the release of massive amounts of cytokines, such as interleukins (IL) and tumor necrosis factors (TNF), resulting in toxic shock (McCormick, Yarwood, & Schlievert, 2001).

1.4 Molecular Typing of *S. aureus*

To perform effective control measures against the wide spread of *S. aureus* infections, it is essential to distinguish between not only sporadic and epidemic strains, but also to distinguish one epidemic clone from another (Jackson, Edwards-Jones, Sutton, & Fox, 2005). Therefore, a variety of phenotypic and genomic typing methods have been employed to differentiate among *S. aureus* strains (Feil & Enright, 2004). Selecting the appropriate technique is essential for tracking the epidemiology of MSSA and most importantly MRSA infections, investigating MRSA outbreaks, and consequently supporting control measures of *S. aureus* spread (Hallin *et al.*, 2008).

Phenotypic methods have been widely used in the clinical settings to distinguish *S. aureus* isolates. The most frequently used phenotypic means is the antimicrobial susceptibility test which provides rapid and useful data. However, this method of typing lacks the precise discrimination between isolates because MRSA strains are resistant to most classes of antibiotics. Therefore, genotypic approaches are considered more reliable for MRSA strains characterization (Palavecino, 2007).

The most common genotyping methods of *S. aureus* include single locus sequencing, such as staphylococcal protein A gene (*spa*) and coagulase gene (*coa*) typing, multilocus sequence typing (MLST), and pulsed-field gel electrophoresis (PFGE) (Turner & Feil, 2007). All of the above mentioned techniques achieve the ultimate purpose, which is subtyping of *S. aureus* strains, but each with a different discriminatory power, accuracy, and reproducibility (Weller, 2000).

1.4.1 *spa* Typing

Spa typing technique is based on sequencing the polymorphic X region of a single hyper-variable locus of protein A gene (Crisostomo *et al.*, 2001). It involves the analysis of sequence variation and repeat number that varies from 1 to 15, thus providing a binary typing scheme (van Belkum *et al.*, 2009). Because this repeat region is highly subjected to spontaneous mutations, *spa* typing is considered useful and reliable to examine temporary and local MRSA outbreaks (Mellmann *et al.*, 2006). Furthermore, *spa* typing is rapid, inexpensive, reproducible, and more resolving than *coa* typing (Harmsen *et al.*, 2003). However, the strong resolving power of *spa* typing has become a drawback because even epidemiologically related strains are categorized as different through this technique due to the continuous sequence alterations (van Belkum *et al.*, 2009).

1.4.2 Multi-Locus Sequence Typing

Unlike *spa* typing, MLST is a molecular method that targets seven stable loci of housekeeping genes, useful for studying long-standing population relatedness and assessing the evolution of *S. aureus* strains. Each isolate is characterized by a sequence type (ST), which is a set of seven numbers representing the seven alleles of each locus (Enright, Day, Davies, Peacock, & Spratt, 2000). Closely related STs are classified in a clonal cluster (CC), providing a detailed phylogenetic tree of *S. aureus* population structures (Robinson & Enright, 2004). However, because MLST targets only housekeeping genes which belong to the core genome and far from frequent genetic variation, it is not reliable to discriminate epidemiologically unrelated strains or recognize local MRSA outbreaks in hospitals (Cooper & Feil, 2004). Another disadvantage of MLST lies in its high cost and the need of a sequencing facility, which is not often available in clinical settings (Harmsen *et al.*, 2003).

1.4.3 Pulsed-Field Gel Electrophoresis

Another genomic typing technique is the PFGE that is based on digesting bacterial chromosomal DNA with a rare cutting restriction endonuclease, producing a unique pattern of bands to each *S. aureus* strain (Palavecino, 2007; van Belkum *et al.*, 2009). Compared to other typing methods, PFGE has been reported to possess a higher discriminatory power than *spa* typing and MLST to better investigate MRSA outbreaks in hospitals and study the population structure of *S. aureus* strains (Vainio *et al.*, 2008). Although considered a "gold standard", PFGE may not trace *S. aureus* long-term clones, due to the constancy of restriction sites (Hallin *et al.*, 2008). Moreover, it is cost expensive, time consuming (Jackson *et al.*, 2005), and requires high level of technical expertise (Carbonnelle *et al.*, 2011).

1.5 Characterization of Pathogens using MALDI

Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is known to be a "soft ionization" method producing ions from biological molecules such as proteins and DNA (Lay & Holland, 2000).

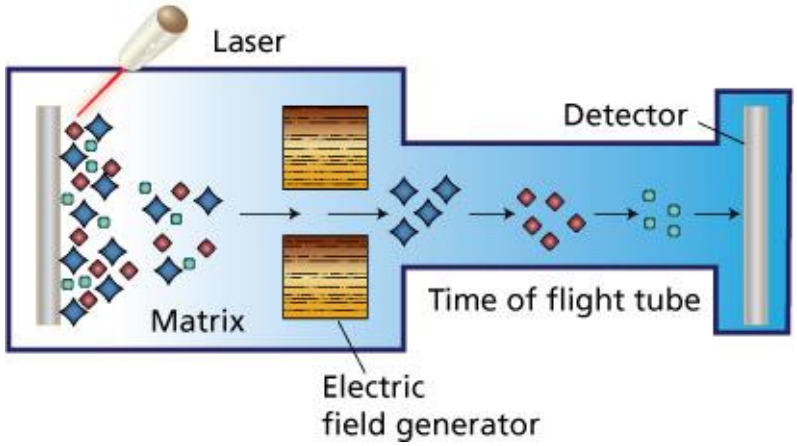

Figure 1. Schematic overview of the ionization method

The biological molecules are first crystallized with acidic organic particles known as the "matrix". A pulsed laser will then irradiate these co-crystals at a wavelength near the resonant absorption band of the matrix molecules. Therefore stimulating energy transfer and

desorption process that lead to vaporization of the analyte and matrix molecules into a highly vacuumed gas phase. Next, a proton-transfer reaction between matrix and analyte is induced leading to singly charged ions that are then accelerated by an electric field toward a mass analyzer (Sauer, 2007). Ions are then separated based on their mass-dependent velocities and are characterized by: molecular mass (m), charge (z), mass to charge ratio (m/z), and a relative intensity of the signal (Carbonnelle *et al.*, 2011).

Figure 2. 4800 MALDI TOF/TOF analyzer

The method's high sensitivity, specificity, speed and reliability has endorsed its introduction into an increasing number of diagnostic laboratories worldwide (Christner *et al.*, 2010). The increase of interest in time-of-flight mass spectrometry (TOF) has occurred in 1988, after the invention of MALDI by Karas and Hillenkamp (1988). Later, Holland *et al.* (1996) were the first to illustrate the use of MALDI for the analysis of intact microbial cells.

In recent years, several studies have shown that MALDI-TOF MS can be applied for the identification of diverse species and strains of bacteria (Yang *et al.*, 2009). For example, this technique was applied on whole cells for the identification of major human pathogens such as *Listeria* and *Pneumococci* to the species level (Carbonnelle *et al.*, 2011) and for subspecies typing of *Streptococcus pneumonia* (Williamson *et al.*, 2008). In addition, its differentiation among *L. monocytogenes* strains is also critical for proper therapy of listeriosis, a serious infection that affects pregnant women and immune-compromised patients. Barbuddhe *et al.* (2008) have performed MALDI-TOF MS on bacterial extracts as a promising method to identify, fingerprint and discriminate the clonal lineages of the pathogenic *L. monocytogenes*. *Corynebacterium* is another toxigenic organism where MALDI-TOF MS can be applied to rapidly identify

between its different species, such as *C. ulcerans*, *C. pseudotuberculosis* and *C. diphtheriae* (Konrad *et al.*, 2010). Veloo, Welling, & Degener, (2011) have reported the use of MALDI-TOF MS on whole cells for the identification of anaerobic bacteria whether being Gram-positive, such as *Clostridium spp.*, or Gram-negative anaerobes, such as *Prevotella spp.*, *Fusobacterium spp.*, and *Bacteroides spp*. Moreover, fingerprints of unknown bacterial samples obtained from MALDI-TOF MS, can be accurately identified by comparing them with reference database of known bacteria (Holland *et al.*, 1996; Stackebrandt, Pauker, & Erhard, 2005). Ferreira *et al.* (2010) have reported the creation of MALDI reference database profiles for *Brucella* species, which is also a pathogenic microorganism in many countries. This is useful for recognizing *Brucella* from culture plates and blood cultures bottles.

Other than identification purposes for practice clinical microbiology, MALDI-TOF MS may assist in infection control measures (Barbuddhe *et al.*, 2008; Carbonnelle *et al.*, 2011). Other studies showed that MALDI-TOF was a powerful technique in identifying crystal (cry) toxins from different novel *Bacillus thuringiensis* (Bt) strains. Under normal circumstances, a successful PCR technique would be sufficient to detect the cognate genes of novel Cry toxins in

novel Bt strains; however when this failed, the MALDI-TOF MS emerged as a potent remedy. The MALDI-TOF MS of in-gel trypsin digested proteins was successfully used here to match the peptide masses to their corresponding proteins. The MALDI-TOF MS technique was very effective that it even helped in differentiating and detecting multiple Cry toxins from a single protein band, that are of very similar molecular weights and isoelectric points (Ranasinghe & Akhurst, 2002).

1.6 Identification and typing of *S. aureus* using MALDI

MALDI-TOF MS has been shown to be a powerful tool for identification of *S. aureus* (Meetani & Voorhees, 2005). Edwards-Jones *et al.* (2000) showed that this technique can be applied directly on whole cells for the differentiation between methicillin-resistant and methicillin-sensitive *S. aureus* strains. They were able to identify peaks specific to MRSA, others specific to MSSA, and few specific to individual strains. Also, Du *et al.* (2002) showed similar results were whole cells of *S. aureus* strains harbouring the *mecA* were successfully detected by MALDI-TOF as MRSA, while those lacking

the resistant gene were successfully identified as MSSA. In addition, Wolters *et al.*, (2011) have shown that MALDI-TOF MS of bacterial extracts has the ability to become an important screening tool for prediction of clonal lineages and outbreak investigations. In brief, they were able to identify 13 peaks differentiating between MRSA isolates belonging to the 5 major HA-MRSA clonal complexes (CC5, CC8, CC22, CC30, CC45).

In this study, we utilized a total of 14 *S. aureus* strains already identified by regular genomic methods including PFGE, MLST, and *spa* typing. We propose a proteomic approach for fast identification of *S. aureus* isolates by MALDI-TOF MS based protein mass fingerprinting. Proteins were extracted using formic acid and they were subjected to MALDI-TOF analysis. The resulting identification profiles were analyzed using a newly developed program with strict parameters. Moreover, the same experimental procedure and data analysis were applied on two bacteria to assess the specificity of our proteomic approach. Bacteria used were *S. epidermidis*, the closest species to *S. aureus*, and *E. coli*, which belongs to a genus that is evolutionary far from *Staphylococcus.*

Chapter Two

MATERIALS AND METHODS

2.1 Clinical Isolates

A total of 14 *S. aureus* isolates (10 MRSA and 4 MSSA), obtained from clinical specimens kindly provided by Dr. George Araj from American University Hospital (AUH) in Beirut, were used in this study. The list of isolates and their characteristics are found in table 1. Five isolates were recovered from wound infections, 2 from blood, 2 from sputum, 1 from semen, 1 from eye, 1 from brain, 1 from Neck mucus, and 1 from descending thoracic aorta (DTA). The samples were streaked on Tryptone Soy Agar (TSA) and stored in Cryobanks at -80 °C. All isolates were previously characterized by *spa* typing, SCC*mec* typing and MLST. *spa* typing classified the isolates into 14 different *spa* types, which could be designated to 9 different clonal complexes based on previously determined MLST typing data. Six of 14 isolates belonged to the 4 major HA-MRSA clonal complexes CC5 (2 isolates), CC8 (1), CC22 (1), and CC30 (2). Two isolates were assigned to unrelated sporadic strain types, e.g. t044 (ST-80)

and t127 (ST-1) described as widely disseminated CA-MRSA clones. In addition, the study included *Staphylococcus epidermidis* 80 and *Escherichia coli* J96 ATCC.

Table 1: "An overview of the major MRSA and MSSA spa types and their corresponding MLST clones recovered from Lebanon" (Tokajian et al., 2010)

Isolate SA	SCC*mec* type[1]	PVL[2]	MLST ST[3]	MLST CC[4]	Allelic profile	*spa* type	*spa*-CC[5]	Pathogenic genes	Site of Infection	E-Test Susceptibility (#R)
23	IVc	+	ST-80	CC80	1-3-1-14-11-51-10	t 044	*spa*-CC044	None	Wound	OX
94	MSSA	-	ST-30	CC30	2-2-2-2-6-3-2	t 012	*spa*-CC021	SEA,SEG,SEI,TSST	blood	None
89	MSSA	-	ST-80	CC80	1-3-1-14-11-51-10	t 937	*spa*-CCd	SEI	wound	TR, TS
16	IVc	+	ST-30	CC30	2-2-2-2-6-3-2	t 318	*spa*-CC021	SED	Wound	NX, TC
126	IVc	-	ST-8	CC8	3-3-1-1-4-4-3	t 008	*spa*-CC008	SEA, SEI, eta	Semin	AZ, CH
95	MSSA	-	ST-1	CC1	1-1-1-1-1-1-1	t 127	sg #2	SEC, SEI	Wound	AZ, CM, EM, TR
79	MSSA	-	ST-5	CC5	1-4-1-4-12-1-10	t 002	*spa*-CCb	SEG,SEI	Blood	CH, LE, NX, OF, TC
57	IVc	+	ST-80	CC80	1-3-1-12-11-51-10	t 131	*spa*-CC044	SEG,SEI	Eye	AZ, CM, EM, OX, TC
24	III	-	ST-239	CC239	2-3-1-1-4-4-3	t 037	*spa*-CC021	SEA	Brain	AZ, CM, CL, EM, OX, TC
103	IVc	-	ST-22	CC22	7-6-1-5-8-4-6	t 032	*spa*-CC005	SED,SEI	Sputum	CI, EM, NX, OF, OX, TC, VA
125	IVc	+	ST-5	CC5	1-4-1-4-12-1-10	t 311	*spa*-CCb	SEA	Sputum	EM ,TC

Isolate SA	SCC*mec* type[1]	PVL[2]	MLST ST[3]	MLST CC[4]	Allelic profile	*spa* type	*spa*-CC[5]	Pathogenic genes	Site of Infection	E-Test Susceptibility (#R)
34	V	-	ST 97	CC 97	3-1-1-1-1-5-3	t 267	*sg #6*	None	Neck mucus	OX, TC
98	MSSA	+	ST 121	CC 121	6-5-6-2-7-14-5	t 159	*sg #3*	SEB,SEG,SEI	BRW	None
36	III	-	ST 239	CC239	2-3-1-1-4-4-3	t 030	*spa*-CC021	None	Wound	AZ, CH, CM, CL, EM, GM, LE, NX, OF, OX, RI, TC, TR, TS
128	IVc	-	ST 6	CC 6	12-4-1-4-12-1-3	t 304	*spa*-CC008	SEA,eta	DTA	None

2.2 MALDI-TOF MS

For sample preparation, formic acid extraction of bacteria was carried out using a modified protocol by Wolters *et al.* 2011. In brief, bacterial sample were cultured on TSA plates for 22-24 h at 37 °C. A full loop of colonies was then suspended in 300 µL of sterile water. 900 µL of ethanol were then added and the mixture is vortexed. The suspended bacteria were then centrifuged at 12,000×g for 2 min at room temperature. Discard the supernatant and centrifuge again to remove any residual ethanol. The pellet was then resuspended in 50 µL of formic acid (70%) and 50 µL of acetonitrile. Centrifugation was then applied at maximum speed for 2 minutes and 80 µL of the supernatant were collected and stored and stored at −20 ºC until use. For plate preparation, 1 µL aliquots of the supernatant was spotted onto a stainless steel target plate (Opti-TOF ^{TM} 384 Well Insert, 123×81 mm RevA, Applied Biosystems, USA) and air dried for 15 min. 1 µL of matrix solution (saturated solution of α-cyano-4-hydroxy-cinnamic acid in 50% acetonitrile with 2.5% trifluoroacetic acid) was added to each sample spot and again air dried for 15 min. MALDI-TOF analysis was applied on a 4800 MALDI TOF/TOF Analyzer instrument operated by the 4000 Series Explorer software

(Applied Biosystems, USA). Spectra were acquired in linear positive mode at a laser intensity of 7540 within a mass range from 800 to 20,000 Da. Acceleration voltage was 20 kV, IS2 voltage was maintained at 19.3 kV, and the extraction delay time was 160 ns. For each sample spot, one average spectrum was accumulated from 500 measurements. For each of the 14 isolates, 3 extracts were generated from independent subcultures. Each extract was analyzed in multiplets (11 per each), resulting in 33 mass spectra for each isolate and a total of 560 mass spectra for the whole study.

2.3 Calibration and Error Analysis

Calibration was performed using Applied Biosystems 4700 Proteomics Analyzer Calibration mixture. The internal standards used and their average theoretical mass are listed in Table 2. The percentage error was calculated by comparing measured values of 4700 Cal Mix standard (Linear mode) to their theoretical one. Also, Table 2 exemplifies how the percentage error was calculated for each experiment.

Table 2. 4700 Cal Mix Internal standards used for instrument calibration

Internal Standards	Theoretical Mass (Da)	Average Measured Mass (Da)	Error (ppm)	% Error
des-Arg-Bradykinin	904.4681	904.6936	249.32	0.02
Angiotensin 1	1296.6853	1297.7828	846.38	0.09
Glu-Fibrinopeptide B	1570.6774	1572.2834	1022.48	0.10
ACTH (clip 1-17)	2093.0867	2.09E+03	756.39	0.08
ACTH (clip 18-39)	2465.1989	2468.0046	1138.12	0.11
ACTH (clip 7-38)	3657.9294	3663.185	1436.76	0.14

2.4 Development of Program for High Throughput Data Analysis

Mass spectrum analysis was performed using two programs (Program 1 and Program 2) developed by Engineer Mazen Naamani. Strict parameters were set on the developed program to ensure minimum error during data analysis. Briefly, Program 1 is concerned with analysis of centroid mass peaks obtained from a single isolate. The input will represent the 33 peak lists obtained from each isolate (Figure I-a in appendix I). The output will represent a peak list containing centroid masses detected with their percentage of occurrence in the 33 runs (Figure I-b in appendix I). Program 2 is concerned with analysis of data between all the isolates. The input in this program is the output of all the isolates from program 1 (Figure I-

c in appendix I). At this stage, the program will collect all the centroid masses from all the peak lists and search for their occurrence in all the isolates. This will result in a new sheet that shows all the centroid masses with their percentage of occurrence in each isolate (Figure I-d in appendix I). Then an averaging process will take place where peaks are averaged based on the following criteria: a) they are within the determined percentage error, b) they follow same trend of occurrence and percentage among the 14 isolates. The last step of data analysis is concerned with selecting peaks present in all 14 *S. aureus* isolates (Figure I-e in appendix I). We set the selection process based on the follow criteria: a) Centroid mass in each isolate showing a percentage of occurrences less than 15% is set as absent. Any centroid mass showing a percentage of occurrences above 85% is set to exist. Any centroid mass that shows a percentage of occurrences between 15% and 85% is removed from the list. The unique m/z peaks identified in all *S. aureus* strains were validated if the observed phenotype (peak presence or absence) was coherent within 10 random spectra of each isolate.

Chapter Three

RESULTS

3.1 Identification of Discriminatory m/z Values of *S. aureus*

To recognize discriminatory m/z values for identification of *S. aureus*, 14 *S. aureus* isolates were analyzed by MALDI-TOF MS. Spectra were taken over a mass (m/z) range of 800 to 20,000 with a focus mass of 5000 m/z. The majority of peaks were collected between masses 800 and 12,000 m/z (corresponding to 0.8 and 12 kDa).

The figure below represents a sample of the spectrum obtained from each isolate.

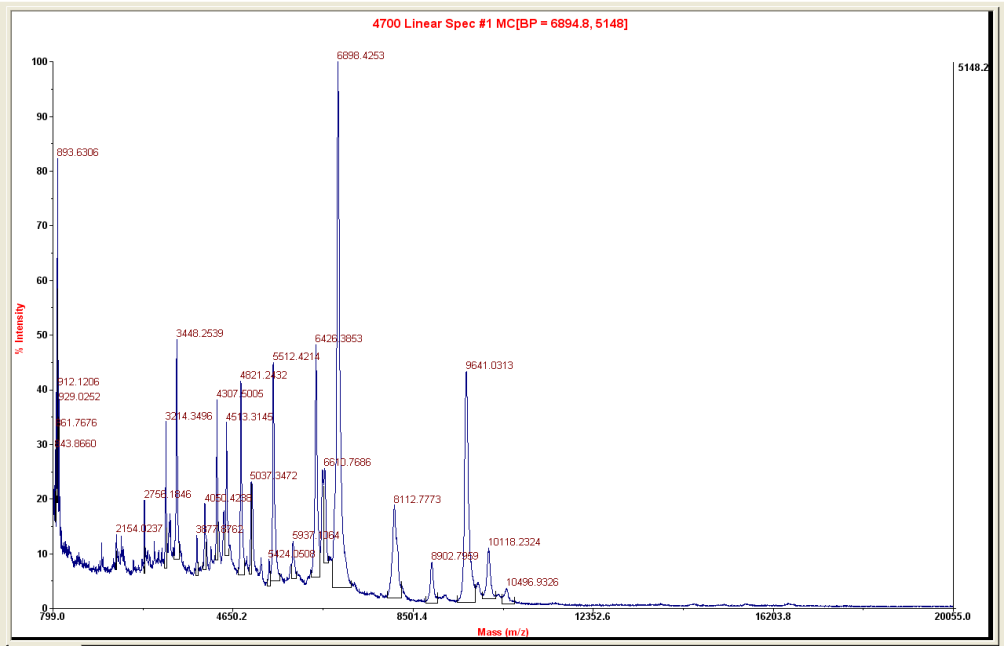

Figure 3. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 94). The m/z value shown on the *x*-axis corresponds to the molecular weight of proteins (Da) if the peak has a single positive charge. The percentage intensities of the peaks are shown on the *y* axis.

The spectrum for each isolate can be found in the appendix.

3.2 Data Analysis of Mass Spectra

As mentioned before, 33 spectra were acquired for each *S. aureus* isolate (equivalent to 33 peak lists). The peak lists are processed using two programs developed for this purpose. Table 3 shows the average peaks obtained for each isolate.

Table 3: Masses observed for each *S. aureus* isolate from MALDI-TOF-MS analysis. "1" indicates presence and "0" indicates absence of a peak at a given strain

Peak (m/z)	SA16	SA103	SA125	SA126	SA34	SA36	SA94	SA128	SA23	SA24	SA57	SA79	SA89	SA95
942.4756	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1820.2366	0	0	0	0	0	0	0	0	0	0	0	0	1	1
2239.3262	0	0	0	0	1	0	0	1	0	0	0	0	1	0
2300.2373	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2362.5288	0	0	0	0	1	0	0	0	0	0	0	0	1	0
2364.6428	0	0	0	0	1	0	0	1	0	0	0	0	1	0
2367.4386	0	0	0	0	1	0	0	1	0	0	0	0	1	0
2755.1882	1	0	0	0	0	1	1	0	0	1	0	0	0	0
2755.9189	1	0	1	0	0	1	1	0	0	1	0	0	0	0
2756.0125	1	0	1	1	0	1	1	0	0	1	0	0	0	0
3216.0315	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3220.4548	1	1	1	1	1	0	1	1	1	1	1	1	1	1
3220.8413	1	1	1	1	1	0	0	0	1	1	1	1	1	1
3221.5632	1	1	1	1	0	0	0	0	1	1	1	1	1	1
3446.6311	1	1	1	1	1	0	1	1	1	1	1	1	1	1
3450.0225	1	1	1	1	1	0	1	1	1	1	1	1	1	1
3454.8655	1	1	1	1	0	0	0	0	1	1	1	1	1	1
3464.6926	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4513.5945	0	0	0	0	0	0	1	1	0	0	0	1	1	0
4822.9492	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4829.5181	1	1	1	1	1	1	1	0	1	1	1	1	1	1
5250.5417	0	0	0	0	0	1	0	0	0	0	0	1	0	0
5303.8354	0	0	0	0	0	0	0	0	1	0	1	1	0	0
5314.8062	0	0	0	0	0	0	0	0	1	0	1	0	0	0

Peak (m/z)	SA16	SA103	SA125	SA126	SA34	SA36	SA94	SA128	SA23	SA24	SA57	SA79	SA89	SA95
5508.7493	1	0	0	0	0	0	1	0	0	0	0	0	0	0
5509.5884	1	0	0	1	0	0	1	0	0	0	0	0	0	0
5509.989	1	0	1	1	0	0	1	0	0	0	0	0	0	0
5513.198	1	1	1	1	0	0	1	0	0	0	0	0	0	0
5529.8896	1	1	1	1	1	0	0	1	1	1	1	1	1	1
5537.3621	0	1	1	1	1	0	0	1	1	1	1	1	1	1
5543.5898	0	1	1	1	1	0	0	0	1	1	1	1	1	1
5546.1924	0	1	1	1	0	0	0	0	1	1	1	1	1	1
6427.8223	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6437.812	1	1	1	1	1	1	1	0	1	1	1	1	1	1
6579.4497	1	1	1	1	0	0	1	0	1	0	1	1	1	1
6593.9648	1	1	1	1	0	0	0	0	0	0	0	0	0	1
6595.6606	1	1	1	1	0	0	0	0	1	0	0	0	0	1
6596.4404	1	1	1	1	0	0	0	0	1	0	1	0	0	1
6596.7476	1	1	1	1	0	0	0	0	1	1	1	0	0	1
6598.73	1	1	1	1	1	0	0	0	1	1	1	1	1	1
6603.3701	1	1	1	1	1	0	1	1	1	1	1	1	1	1
6612.1521	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6852.0515	0	0	0	0	0	1	0	0	0	0	0	0	0	0
6899.5774	1	1	1	1	1	0	1	1	1	1	1	1	1	1
6910.4861	1	1	1	1	1	0	1	0	1	1	1	1	1	1
8903.4185	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9643.9966	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11527.4260	0	0	0	0	0	0	0	0	0	0	0	1	0	0

3.3 Data Validation and Analysis

From the data obtained, we were able to identify 5 peaks (4823± 8, 6428 ± 10, 6612 ± 11, 8903 ± 14, and 9644 ± 15 m/z) common to all the 14 *S. aureus* isolates. Peaks were 90% reproducible and a validation test was applied to the data analysis method. Ten readings were selected randomly from each isolate and verified visually for the presence of the above mentioned peaks. All the readings (a total of 140 spectra) confirmed the presence of the 5 identified peaks. Figure 7 presents an example of the validation test.

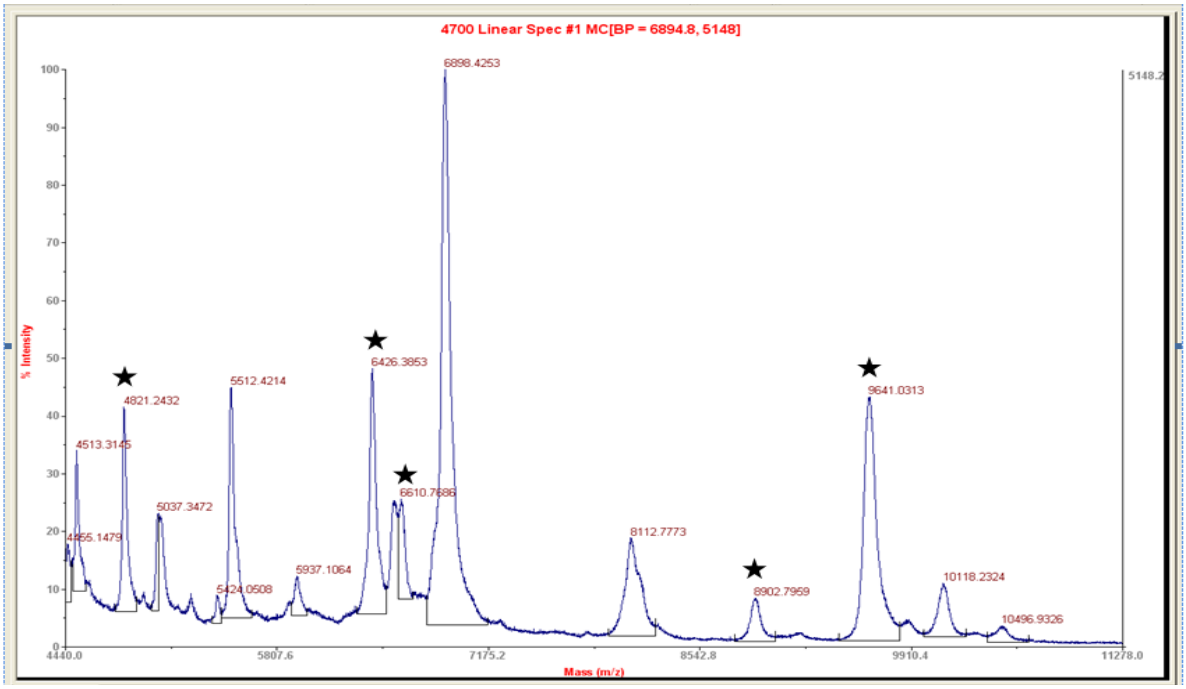

Figure 4. Data verification spectrum showing common peaks in all 14 *S. aureus* isolates. Stars represent the 5 peaks (4823± 8, 6428 ± 10, 6612 ± 11, 8903 ± 14, and 9644 ± 15 m/z) identified.

3.4 Identification of Discriminatory m/z Values of *S. epidermidis* and *E. coli*

Two additional bacteria (*S. epidermidis* 80 and *E. coli* J96 ATCC 700336) were subjected to formic acid extraction and MALDI-TOF analysis following the same procedure and analysis applied on *S. aureus.* We were able to identify 23 peaks in *E. coli* isolate and 13 peaks for *S. epidermidis* isolate. Table 4 represents the list of peaks for *S. aureus*, *E. coli* and *S. epidermidis*.

Table 4. List of m/z values common in each type of bacteria

S. aureus	*S. epidermidis*	*E. coli*
4823	803	845
6428	845	883
6612	3355	976
8903	3754	1106
9644	4308	2701
	4914	3140
	5135	3656
	5366	4203
	6607	4383
	6717	4631
	8135	4798
	8900	5119
	9659	5398
		6321
		6856
		7310
		7903
		8394
		8942
		8973
		9259
		9590
		10312

Four peaks from *E. coli* were identified to be ribosomal proteins. The list of these proteins is present in Table 5.

Table 5. Comparison between peaks derived from our study to ribosomal proteins identified by Freiwald *et al.* (2009)

***E. coli* J96 ATCC 700336**	**Peaks from *E. coli* Freiwald *et al.* (2009)**	**% Error**	**Identified Ribosomal Proteins**
845			
883			
976			
1106			
2701			
3140			
3656			
4203			
4383			
4631			
4798			
5119			
5398	5381	0.3159	RS34
6321	6316	0.0791	RL32
6856			
7310			
7903			
8394	8370	0.2867	RS21
8942			
8973			
9259			
9590			
10312	10300	0.1165	RS19

Table 6. Highlighted peaks represent the similarity identified between *S. aureus* and *S. epidermidis*

S. aureus	*S. epidermidis*	% Error
	803	
	845	
	3355	
	3754	
	4308	
	4914	
	5135	
	5366	
6612	6607	0.0841
	6717	
	8135	
8903	8900	0.0390
9644	9659	0.1531

Peaks are considered to be similar when the % error between peaks of *S. aureus* and *S. epidermidis* is within our experimental error of 0.31%.

Chapter Four

DISCUSSION

The sensitivity and robustness of MALDI-TOF have made it a very important proteomic tool for the fast identification of bacteria (Barbuddhe *et al.*, 2008). Several studies have shown the potential of this technique to differentiate bacterial species. Dieckmann *et al.* (2008) have reported identification and classification of Salmonellae at both species and subspecies level. Williamson *et al.* (2008) showed the ability of MALDI-TOF MS to differentiate *Streptococcus pneumonia* strains. Regarding *S. aureus*, previous studies have confirmed the practicability and the high discriminatory power of MALDI-TOF as observed for the diverse *S. aureus* fingerprints obtained by this technique (Edwards-Jones *et al.*, 2000). Moreover, Wolter *et al.* (2011) have attempted to discriminate major MRSA lineages by MALDI-TOF analysis of bacterial protein extracts. Some of their findings will be discussed later in this section.

In this study, bacterial extracts were prepared by formic acid extraction from *S. aureus* (4 MSSA and 10 MRSA), *S. epidermidis* 80, and *E. coli* J96 ATCC 700336) isolates. The results obtained have shown highly reproducible spectra. Therefore we can conclude that the protocol is practical for routine applications. These results are in concordance with the findings of Wolters *et al.* (2011) regarding the protocol applied and reproducibility of spectra.

Despite the 22-24 hours needed to subculture bacteria on TSA, we were able to obtain the results for MALDI-TOF analysis of all the 16 bacteria within 3 hours. This time includes 45 minutes for sample preparation, 45 minutes for spotting and drying of sample/matrix on the target plate, 60 minutes for MALDI-TOF measurement, and 30 minutes for data processing. On the other hand, genomic typing methods remain laborious and time consuming. For example, PFGE, which is known to be the "gold standard" for bacterial identification, needs 3-6 days for the results to be revealed (Murchan *et al.*, 2003). Once developed, our proteomic approach will be much faster than the regular genomic methods.

Regarding *S. aureus* isolates, 48 unique peaks were found to be distributed among the 14 isolates. Table 3 represents the peaks identified and their absence or presence in each isolate. We noticed that out of the 48 peaks, 5 (4823, 6428, 6612, 8903, and 9644 m/z) were found to be conserved among all the *S. aureus* isolates. When these peaks are compared to the list of m/z values (4813, 6888 and 9626) known to be highly conserved among *S. aureus* (Wolters *et al.*, 2011), we matched those at 4813 and 9626 m/z. Also, a validation test has confirmed the presence of these peaks in randomly selected spectra from each isolate, thus implying that our data analysis is valid. Figure 7 is an example of a spectrum showing the peaks common in all strains. Peaks are determined within a percentage error of 0.16% (1500 ppm) determined from the calibration standards.

As mentioned earlier, all our *S. aureus* isolates were previously typed using various genomic methods (MLST, SCC*mec* typing and *spa* typing). The clonal complexes: CC 5 (SA 79, SA 125), CC 22 (SA 103), CC 8 (SA 126), and CC 30 (SA 16, SA 94) used in this study have been found similar to those used by Wolters *et al.* (2011). We were able to identify 48 peaks distributed among the different clonal complexes. However, Wolter *et al.* (2011) have identified only 13 peaks. Out of the 13 peaks that they have found to be distributed in

these clonal complexes, only 5 (4511, 5508, 5524, 6591, and 6612 m/z) matched those we presented in Table 3. Although data found between the two studies show some similarity in the distribution of these peaks among the clonal complexes, variation still exists because of the different *spa* types that the same clonal complexes could have (Chambers & Deleo, 2009). Other reasons for these variations might be due to experimental variations and differences in analysis processing of data (Carbonnelle *et al.*, 2011). Furthermore, a study performed by Wunschel *et al.* (2005) showed that differences in the various commercial instruments are a source for differences in results even if all the experimental parameters were held exactly the same. The nature of each peak has not been identified until now, but additional research done by Wolters *et al.* (2011) using data from genome-sequenced reference *S. aureus* strains, revealed that the 4 peaks at m/z 3216, 3447, 4823, and 6428 in Table 3 are ribosomal proteins (via personal communication with Wolters group). However, further investigation is still needed to reveal the exact identity of the other peaks.

A study by Bittar *et. al.* (2009) revealed that *S. aureus* isolates carrying the PVL toxin can be directly identified using MALDI-TOF MS. They have identified a peak at 4448 m/z that was thought to be a

marker for the presence of the PVL toxin. However this approach failed when it was applied on four of our *S. aureus* isolates, namely SA 23, SA 16, SA 57, and SA 125, carrying the PVL toxin. Thus we conclude that there is no correlation between peak at 4448 m/z value and PVL toxin. One possible explanation is that this peak is excluded from the list of Table 3 because it failed to match the stringent criteria in the program's data analysis. However, when searching through the unprocessed data, we noticed that this peak at 4448 m/z is present in 2 strains (SA 103 and SA 34) lacking the PVL toxin. Another possible explanation for this difference is that we performed formic acid extraction of proteins and not whole cell preparation as performed by Bittar *et al.* (2009). But Szabados *et al.* (2011) performed MALDI-TOF MS on whole cells and could not find any relation between the peak at 4448 m/z and presence of the PVL toxin, therefore supporting our findings.

To determine whether the 5 common identified peaks are unique to *S. aureus*, two additional bacteria, *S. epidermidis* 80 and *E. coli* J96 ATCC 700336, were subjected to MALDI-TOF analysis following the same procedure of formic acid extraction and analysis as applied on *S. aureus.* It is important to note that *S. epidermidis* are

evolutionary the closest species to *S. aureus*, while *E. coli* belongs to a genus evolutionary far from the genus *Staphylococcus.*

Regarding MALDI-TOF analysis of *E. coli*, we have identified 23 peaks completely conserved in all the spectra obtained from the data analysis. To determine the nature of these peaks presented in Table 4, we compared our peak list to the results obtained by Freiwald & Sauer (2009) on the phylogenetic differentiation of bacteria by mass spectrometry. We were able to match 4 peaks at m/z 5398, 6320, 8394, and 10312 as *E. coli* ribosomal proteins RS 34, RL 32, RS 21, and RS 19, respectively (Table 5).

As for the MALDI-TOF analysis of *S. epidermidis*, 13 peaks were found to be present among all the spectra analyzed. But this time, none of the peaks identity was determined. One main reason is absence of a database access (such as BioTyper 2.0 and ClinProTools^{TM} 2.0) for bacterial identification.

Of the 5 peaks found to be common to all *S. aureus* isolates, 3 were found to be present in *S. epidermidis* (Table 6). However, none were found to be present in *E. coli*. Thus we can say that the 3 peaks (6612, 8903 and 9644 m/z) found in common between *S. aureus* and *S.*

epidermidis are unique for the genus *Staphylococcus* and the 2 peaks (4823 and 6428 m/z) are unique for the species *S. aureus*. Also, the developed program is designed to match an unknown sample to one of the 14 *S. aureus* isolates used in this study. So when we tested the samples related to *S. epidermidis*, the highest match obtained with one of our *S. aureus* isolate was around 39%. But when *E. coli* was tested, maximum match was close to 8%.

Furthermore, if we look at Table 3, we noticed that 5 isolates have unique peaks specific only for them. SA 34, SA 103, SA 16, SA 36, and SA 79 show one unique peak at 942, 2300, 3464, 6852, and 11527 m/z respectively. Since each of these strains belongs to a unique clonal complex and *spa* type, we were able to identify peaks specific for *S. aureus* at the strain level. However, further investigations are still necessary to identify peaks specific for all the other strains. Finally, it is important to note that due to the lack of software, we were unable to correlate the results obtained from our proteomic and genomic data.

Chapter Five

Conclusion

In conclusion, monitoring the spread of *S. aureus* is an essential step for initiation of targeted infection control measures against this pathogen. In this study, we were able to prove that MALDI-TOF can be a promising proteomic tool for fast and reliable identification of *S. aureus*. The highly reproducible data obtained and simple sample preparations have given MALDI-TOF an advantage over regular genomic methods. Moreover, the identification and validation of 3 peaks specific to the genus *Staphylococcus* and 2 peaks specific for *S. aureus* species, has proved that the program developed is powerful and effective in analyzing high-through put data obtained from MALDI-TOF analysis of bacterial protein extracts. However, program revision need to be undertaken to allow for analysis of bacterial strains of the same species. Also, Future work will address the Tandem-MS of the unique peaks to determine their nature, the creation of a database from other bacterial groups to help identify unknown bacteria and the correlation between both genomic and proteomic data obtained.

Chapter Six

BIBLIOGRAPHY

Archer, G. L. (1998). Staphylococcus aureus: A well-armed pathogen. *Clin Infect Dis, 26(5),* 1179-1181.

Barbuddhe, S. B., Maier, T., Schwarz, G., Kostrzewa, M., Hof, H., Domann, E., ...Hain, T. (2008). Rapid identification and typing of listeria species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl Environ Microbiol, 74(17),* 5402-5407.

Bittar, F., Ouchenane, Z., Smati, F., Raoult, D., & Rolain, J. M. (2009). MALDI-TOF-MS for rapid detection of staphylococcal Panton-Valentine leukocidin. *Int J Antimicrob Agents, 34(5),* 467-470.

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.

Chambers, H. F., & Deleo, F. R. (2009). Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat Rev Microbiol, 7(9),* 629-641.

Christner, M., Rohde, H., Wolters, M., Sobottka, I., Wegscheider, K., & Aepfelbacher, M. (2010). Rapid identification of bacteria from positive blood culture bottles by use of matrix-assisted laser desorption-ionization time of flight mass spectrometry fingerprinting. *J Clin Microbiol, 48(5),* 1584-1591.

Cooper, J. E., & Feil, E. J. (2004). Multilocus sequence typing: What is resolved? *Trends Microbiol, 12(8),* 373-377.

Crisostomo, M. I., Westh, H., Tomasz, A., Chung, M., Oliveira, D. C., & de Lencastre, H. (2001). The evolution of methicillin resistance in Staphylococcus aureus: Similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. *Proc Natl Acad Sci U S A, 98(17),* 9865-9870.

De Bruyne, K., Slabbinck, B., Waegeman, W., Vauterin, P., De Baets, B., & Vandamme, P. (2011). Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Syst Appl Microbiol, 34(1),* 20-29.

Decker, C. F. (2008). Pathogenesis of MRSA infections. *Dis Mon, 54(12),* 774-779.

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.

Edwards-Jones, V., Claydon, M. A., Evason, D. J., Walker, J., Fox, A. J., & Gordon, D. B. (2000). Rapid discrimination between methicillin-sensitive and methicillin-resistant Staphylococcus aureus by intact cell mass spectrometry. *J Med Microbiol, 49(3),* 295-300.

Enright, M. C., Day, N. P., Davies, C. E., Peacock, S. J., & Spratt, B. G. (2000). Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus. *J Clin Microbiol, 38(3),* 1008-1015.

Feil, E. J., & Enright, M. C. (2004). Analyses of clonality and the evolution of bacterial pathogens. *Curr Opin Microbiol, 7(3),* 308-313.

Ferreira, L., Vega Castano, S., Sanchez-Juanes, F., Gonzalez-Cabrero, S., Menegotto, F., Orduna-Domingo, A., …Munoz-Bellido, J. L. (2010). Identification of Brucella by MALDI-TOF mass spectrometry. Fast and reliable identification from agar plates and blood cultures. *PLoS One, 5(12).*

Freiwald, A., & Sauer, S. (2009). Phylogenetic classification and identification of bacteria by mass spectrometry. *Nat Protoc, 4(5),* 732-742.

Genestier, A. L., Michallet, M. C., Prevost, G., Bellot, G., Chalabreysse, L., Peyrol, S., ...Genestier, L. (2005). Staphylococcus aureus Panton-Valentine leukocidin directly targets mitochondria and induces Bax-independent apoptosis of human neutrophils. *J Clin Invest, 115(11)*, 3117-3127.

Gillet, Y., Issartel, B., Vanhems, P., Fournet, J. C., Lina, G., Bes, M., …Etienne, J. (2002). Association between Staphylococcus aureus strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. *Lancet, 359(9308)*, 753-759.

Gustafsson, E., Karlsson, S., Oscarsson, J., Sogard, P., Nilsson, P., & Arvidson, S. (2009). Mathematical modelling of the regulation of spa (protein A) transcription in Staphylococcus aureus. *Int J Med Microbiol, 299(1)*, 65-74.

Hallin, M., Denis, O., Deplano, A., De Ryck, R., Crevecoeur, S., Rottiers, S., ...Struelens, M.J. (2008). Evolutionary relationships between sporadic and epidemic strains of healthcare-associated methicillin-resistant Staphylococcus aureus. *Clin Microbiol Infect, 14(7)*, 659-669.

Harmsen, D., Claus, H., Witte, W., Rothganger, J., Turnwald, D., & Vogel, U. (2003). Typing of methicillin-resistant Staphylococcus aureus in a university hospital setting by using novel software for spa repeat determination and database management. *J Clin Microbiol, 41(12)*, 5442-5448.

Holland, R. D., Wilkes, J. G., Rafii, F., Sutherland, J. B., Persons, C. C., Voorhees, K. J., …Lay, J.O., Jr. (1996). Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom, 10(10),* 1227-1232.

Jackson, K. A., Edwards-Jones, V., Sutton, C. W., & Fox, A. J. (2005). Optimisation of intact cell MALDI method for fingerprinting of methicillin-resistant Staphylococcus aureus. *J Microbiol Methods, 62(3),* 273-284.

Karas, M., & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem, 60(20),* 2299-2301.

Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., McKenna, T., …Shah, H.N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect Genet Evol, 4(3),* 221-242.

Konrad, R., Berger, A., Huber, I., Boschert, V., Hormansdorfer, S., Busch, U., ...Sing, A. (2010). Matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry as a tool for rapid diagnosis of potentially toxigenic Corynebacterium species in the laboratory management of diphtheria-associated bacteria. *Euro Surveill, 15(43).*

Labandeira-Rey, M., Couzon, F., Boisset, S., Brown, E. L., Bes, M., Benito, Y., …Bowden, M.G. (2007). Staphylococcus aureus Panton-Valentine leukocidin causes necrotizing pneumonia. *Science, 315(5815),* 1130-1133.

Lay, J. O., Jr., & Holland, R. D. (2000). Rapid identification of bacteria based on spectral patterns using MALDI-TOFMS. *Methods Mol Biol, 146,* 461-487.

Makgotlho, P. E., Kock, M. M., Hoosen, A., Lekalakala, R., Omar, S., Dove, M., & Ehlers, M.M. (2009). Molecular identification and genotyping of MRSA isolates. *FEMS Immunol Med Microbiol, 57(2),* 104-115.

McCormick, J. K., Yarwood, J. M., & Schlievert, P. M. (2001). Toxic shock syndrome and bacterial superantigens: An update. *Annu Rev Microbiol, 55,* 77-104.

Meetani, M. A., & Voorhees, K. J. (2005). MALDI mass spectrometry analysis of high molecular weight proteins from whole bacterial cells: Pretreatment of samples with surfactants. *J Am Soc Mass Spectrom, 16(9),* 1422-1426.

Mellmann, A., Friedrich, A. W., Rosenkotter, N., Rothganger, J., Karch, H., Reintjes, R., & Harmsen, D. (2006). Automated DNA sequence-based early warning system for the detection of

methicillin-resistant Staphylococcus aureus outbreaks. *PLoS Med, 3(3)*.

Murchan, S., Kaufmann, M. E., Deplano, A., de Ryck, R., Struelens, M., Zinn, C. E., ...Cookson, B. (2003). Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant Staphylococcus aureus: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J Clin Microbiol, 41(4),* 1574-1585.

Oliveira, D. C., Milheirico, C., & de Lencastre, H. (2006). Redefining a structural variant of staphylococcal cassette chromosome mec, SCCmec type VI. *Antimicrob Agents Chemother, 50(10),* 3457-3459.

Palavecino, E. (2007). Clinical, epidemiological, and laboratory aspects of methicillin-resistant Staphylococcus aureus (MRSA) infections. *Methods Mol Biol, 391*, 1-19.

Patel, A. H., Nowlan, P., Weavers, E. D., & Foster, T. (1987). Virulence of protein A-deficient and alpha-toxin-deficient mutants of Staphylococcus aureus isolated by allele replacement. *Infect Immun, 55(12),* 3103-3110.

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.

Ranasinghe, C., & Akhurst, R. J. (2002). Matrix assisted laser desorption ionisation time of flight mass spectrometry (MALDI-TOF MS) for detecting novel Bt toxins. *J Invertebr Pathol, 79(1),* 51-58.

Robinson, D. A., & Enright, M. C. (2004). Multilocus sequence typing and the evolution of methicillin-resistant Staphylococcus aureus. *Clin Microbiol Infect, 10(2),* 92-97.

Sauer, S. (2007). The essence of DNA sample preparation for MALDI mass spectrometry. *J Biochem Biophys Methods, 70(2),* 311-318.

Seybold, U., Kourbatova, E. V., Johnson, J. G., Halvosa, S. J., Wang, Y. F., King, M. D., & Blumberg, H.M. (2006). Emergence of community-associated methicillin-resistant Staphylococcus aureus USA300 genotype as a major cause of health care-associated blood stream infections. *Clin Infect Dis, 42(5),* 647-656.

Stackebrandt, E., Pauker, O., & Erhard, M. (2005). Grouping myxococci (Corallococcus) strains by Matrix-Assisted Laser Desorption Ionization Time-of-Flight (MALDI TOF) mass spectrometry: Comparison with gene sequence phylogenies. *Curr Microbiol, 50(2),* 71-77.

Strommenger, B., Braulke, C., Heuck, D., Schmidt, C., Pasemann, B., Nubel, U., & Witte, W. (2008). *Spa* Typing of Staphylococcus aureus as a frontline tool in epidemiological typing. *J Clin Microbiol, 46(2),* 574-581.

Szabados, F., Becker, K., von Eiff, C., Kaase, M., & Gatermann, S. (2011). The matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS)-based protein peaks of 4448 and 5302 Da are not associated with the presence of Panton-Valentine leukocidin. *Int J Med Microbiol, 301(1),* 58-63.

Thwaites, G. E., Edgeworth, J. D., Gkrania-Klotsas, E., Kirby, A., Tilley, R., Torok, M. E., …Llewelyn, M.J. (2011). Clinical management of Staphylococcus aureus bacteraemia. *Lancet Infect Dis, 11(3),* 208-222.

Todar, K. (2008). *Todar's online textbook of bacteriology*. Retrieved from

http: www.textbookofbacteriology.net

Tokajian, S.T., Abou Khalil, P.A., Jabbour, D., Rizk, M., Farah, M.J., Hashwa, F.A., & Araj, G.F. (2010). Molecular characterization of *Staphylococcus aureus* in Lebanon. *Epidemiol. Infect, 138,* 707-712.

Turner, K. M., & Feil, E. J. (2007). The secret life of the multilocus sequence type. *Int J Antimicrob Agents, 29(2),* 129-135.

Vainio, A., Karden-Lilja, M., Ibrahem, S., Kerttula, A. M., Salmenlinna, S., Virolainen, A., & Vuopio-Varkila, J. (2008). Clonality of epidemic methicillin-resistant Staphylococcus aureus strains in Finland as defined by several molecular methods. *Eur J Clin Microbiol Infect Dis, 27(7),* 545-555.

Van Belkum, A., Melles, D. C., Nouwen, J., Van Leeuwen, W. B., Van Wamel, W., Vos, M. C., ...Verbrugh, H.A. (2009). Co-evolutionary aspects of human colonisation and infection by Staphylococcus aureus. *Infect Genet Evol, 9(1),* 32-47.

Vandenesch, F., Naimi, T., Enright, M. C., Lina, G., Nimmo, G. R., Heffernan, H., …Etienne, J. (2003). Community-acquired methicillin-resistant Staphylococcus aureus carrying Panton-Valentine leukocidin genes: Worldwide emergence. *Emerg Infect Dis, 9(8),* 978-984.

Veloo, A. C., Welling, G. W., & Degener, J. E. (2011). The identification of anaerobic bacteria using MALDI-TOF MS. *Anaerobe*.

Weller, T. M. (2000). Methicillin-resistant Staphylococcus aureus typing methods: Which should be the international standard? *J Hosp Infect, 44(3),* 160-172.

Williams, T. L., Andrzejewski, D., Lay, J. O., & Musser, S. M. (2003). Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. *J Am Soc Mass Spectrom, 14(4)*, 342-351.

Williamson, Y. M., Moura, H., Woolfitt, A. R., Pirkle, J. L., Barr, J. R., Carvalho Mda, G., …Sampson, J.S. (2008). Differentiation of Streptococcus pneumoniae conjunctivitis outbreak isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl Environ Microbiol, 74(19)*, 5891-5897.

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.

Wunschel, S. C., Jarman, K. H., Petersen, C. E., Valentine, N. B., Wahl, K. L., Schauki, D., ...White, E. (2005). Bacterial analysis by MALDI-TOF mass spectrometry: An inter-laboratory comparison. *J Am Soc Mass Spectrom, 16(4)*, 456-462.

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.

Chapter Seven

Appendices

Appendix I

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%	0.31%
2	A10	A11	A12	A13	A3	A4	A5	A6	A7	A8	A9	B10	B11	B12	B13	B3	B4
3	843.52	877.48	877.48	877.51	877.56	877.51	877.59	843.52	843.68	877.45	877.35	877.50	877.46	877.46	877.59	877.70	877.48
4	877.30	878.36	893.40	893.43	878.50	893.50	893.51	877.37	877.49	893.43	893.41	893.43	893.41	878.30	879.40	893.70	879.38
5	893.29	893.46	894.40	894.32	879.45	894.39	894.39	878.29	879.53	894.37	909.35	909.34	895.28	893.38	893.60	909.68	893.53
6	894.19	894.43	909.39	896.31	893.61	895.63	895.30	893.47	893.48	896.08	911.25	2,755.71	909.52	895.48	896.69	2,756.72	894.49
7	895.49	895.29	1,084.16	909.39	894.45	909.52	909.50	894.33	894.39	909.42	912.27	3,213.70	2,153.59	897.13	909.57	3,214.29	895.47
8	909.31	896.40	2,755.60	2,755.49	895.54	911.65	910.37	895.46	895.36	2,755.29	2,755.33	3,447.34	2,754.92	909.44	912.18	3,448.02	896.55
9	910.31	909.46	3,212.91	3,213.17	896.20	912.84	927.95	897.52	896.42	3,212.92	3,212.76	4,049.24	3,212.39	2,755.47	2,755.80	4,049.85	909.49
10	912.47	912.47	3,307.77	3,307.83	909.58	927.99	2,755.78	909.38	909.56	3,308.22	3,307.91	4,306.50	3,445.85	3,213.59	3,213.36	4,307.26	910.45
11	1,083.43	1,122.31	3,446.90	3,412.20	911.90	2,755.86	3,213.60	911.61	910.38	3,412.14	3,446.96	4,451.49	4,047.31	3,447.45	3,447.27	4,452.96	911.70
12	2,754.93	2,755.56	4,305.84	3,447.14	928.09	3,213.58	3,308.56	2,755.63	1,311.82	3,446.94	4,305.70	4,593.18	4,305.60	4,306.12	4,048.66	4,595.04	913.36
13	3,212.23	3,212.95	4,451.39	4,048.27	2,755.74	3,307.78	3,412.53	3,212.90	2,755.52	4,305.88	4,450.43	4,819.85	4,449.66	4,452.21	4,306.45	4,821.27	2,756.08
14	3,307.45	3,308.04	4,819.37	4,076.31	3,213.52	3,447.55	3,447.59	3,308.16	3,213.43	4,450.95	4,819.07	5,040.12	4,591.74	4,820.69	4,450.96	5,043.68	3,213.96
15	3,446.68	3,447.29	4,941.49	4,306.07	3,308.69	4,305.68	4,049.91	3,447.06	3,308.68	4,818.98	5,042.70	5,247.79	4,817.59	5,042.71	4,593.59	5,249.52	3,447.35
16	4,305.35	4,305.56	5,041.22	4,450.87	3,447.99	4,451.99	4,306.35	4,305.85	3,447.42	4,942.35	5,509.12	5,509.59	5,034.75	5,510.02	4,820.07	5,510.96	3,465.33
17	4,450.10	4,452.21	5,509.50	4,819.18	4,049.61	4,820.34	4,451.92	4,451.13	4,305.60	5,036.64	6,422.36	5,878.25	5,508.08	6,423.06	5,040.40	5,936.27	4,306.87
18	4,819.37	4,820.25	5,934.52	4,943.04	4,306.67	5,042.23	4,819.90	4,819.59	4,451.79	5,057.89	6,569.97	5,935.61	5,876.56	6,562.32	5,247.76	6,424.58	4,453.93
19	5,035.69	5,035.62	6,422.53	5,037.47	4,452.62	5,510.19	4,943.63	5,040.38	4,819.80	5,509.51	6,610.75	6,423.15	5,933.06	6,609.10	5,509.92	6,564.06	4,822.16
20	5,508.49	5,057.12	6,570.34	5,057.26	4,820.47	6,423.36	5,043.38	5,509.79	4,941.62	5,935.42	6,821.09	6,561.27	6,421.74	6,895.15	5,934.24	6,609.90	5,035.97
21	5,543.51	5,509.92	6,610.53	5,509.41	4,943.89	6,570.91	5,510.18	5,935.69	5,035.86	6,422.86	6,891.96	6,608.51	6,558.31	8,099.52	6,422.95	6,895.35	5,067.54
22	5,934.72	6,423.27	6,821.14	5,935.00	5,038.63	6,611.40	5,937.20	6,422.91	5,057.91	6,569.96	8,095.97	6,893.20	6,608.85	8,146.28	6,562.62	8,100.40	5,509.61
23	6,422.00	6,570.66	6,892.19	6,422.76	5,058.28	6,822.61	6,423.59	6,570.95	5,509.70	6,611.57	8,143.41	8,095.64	6,890.58	8,897.85	6,607.70	8,148.20	5,546.27
24	6,569.45	6,611.66	8,095.63	6,568.42	5,510.95	6,894.52	6,570.62	6,611.07	5,935.46	6,822.37	8,895.29	8,146.00	8,091.61	9,184.46	6,893.79	8,899.39	6,424.39
25	6,609.51	6,822.51	8,142.52	6,611.07	5,937.27	8,100.40	6,611.34	6,821.37	6,423.11	6,892.48	9,632.86	8,896.19	8,143.03	9,636.81	8,096.10	9,184.43	6,564.28

Input output(O)

Figure I-a. Layout of the first step to be applied in program 1. Row 1 represents the percentage error for each spot, Row 2 indicates the location of the spot on the MALDI target and the other rows refer to the m/z values.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1	Centroid Mass	A10	A11	A12	A13	A3	A4	A5	A6	A7	A8	A9	B10	B11	B12	B13	B3	B4	B5	B6	B7	B8	B9	C10	C11	C12	C13	C3	C4	C5	C6	C7	C8	C9	%	
2	843.517151	1							1	1															1	1	1								18.2%	
3	843.521729	1							1	1															1	1	1								18.2%	
4	843.619019	1							1	1															1	1	1								18.2%	
5	843.659363	1							1	1															1	1	1								18.2%	
6	843.684143	1							1	1															1	1	1								18.2%	
7	843.689819	1							1	1															1	1	1								18.2%	
8	861.44458																							1						1	1			1	12.1%	
9	861.524963																							1						1	1			1	12.1%	
10	861.607361																							1						1	1			1	12.1%	
11	861.706543																							1						1	1			1	12.1%	
12	877.105957	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
13	877.302856	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
14	877.336121	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
15	877.350525	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
16	877.37207	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
17	877.386414	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
18	877.439575	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
19	877.453979	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
20	877.45459	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
21	877.458862	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
22	877.461121	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
23	877.475708	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
24	877.478821	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	
25	877.4823	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0%	

Input | output(Q)

Figure I-b. Layout of the second step to be applied in program 1. Row 1 represents the number of the spot. Column 1 refers to the centroid mass (m/z value) values and "1" indicates presence of a peak.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	0.31%		0.31%		0.31%		0.31%		0.31%		0.14%		0.14%		0.14%		0.14%		0.17%	
2	Matrix	%	SA16	%	SA103	%	SA125	%	SA126	%	SA34	%	SA36	%	SA94	%	SA128	%	SA23	%
3	804.92865	100%	843.517151	18%	843.564453	21%	843.567993	95%	824.289368	3%	805.004517	24%	804.675232	25%	804.003357	26%	804.307739	18%	803.759949	33%
4	806.921204	100%	843.521729	18%	843.575562	21%	843.618958	95%	825.300415	3%	805.078003	28%	804.87146	33%	804.304749	41%	805.105347	33%	803.966858	35%
5	834.214172	100%	843.619019	18%	843.600647	21%	843.6203	95%	843.584351	79%	805.224182	28%	804.912598	38%	804.318481	41%	805.201111	33%	804.113159	38%
6	835.422791	100%	843.659363	18%	843.671326	21%	843.636475	95%	843.656189	79%	805.8927	34%	805.10022	45%	804.412781	41%	805.243225	33%	804.394836	38%
7	836.329651	100%	843.684143	18%	843.677917	21%	843.658936	95%	843.659607	79%	805.906799	34%	805.110229	45%	804.53833	46%	805.251221	33%	804.48584	38%
8	840.227356	100%	843.689819	18%	843.702759	21%	843.664673	95%	843.674622	79%	805.971375	34%	805.139404	45%	804.542175	46%	805.370178	33%	804.525757	38%
9	842.117676	100%	861.44458	12%	843.715576	21%	843.681091	95%	843.682678	79%	806.013489	34%	805.234009	48%	804.599304	46%	805.403564	33%	804.68634	38%
10	847.255188	100%	861.524963	12%	852.112305	3%	843.683228	95%	843.695923	79%	806.070251	34%	805.29834	48%	804.738281	46%	805.53009	30%	804.787476	38%
11	856.210571	100%	861.607361	12%	861.659241	6%	843.68457	95%	843.696777	79%	806.092712	34%	805.453186	48%	804.912781	49%	805.548767	30%	804.87677	38%
12	857.31543	100%	861.706543	12%	861.734131	6%	843.687317	95%	843.714111	79%	806.191284	31%	805.758423	50%	805.074402	51%	805.780823	33%	804.883484	38%
13	862.321289	100%	877.105957	100%	877.332947	82%	843.699951	95%	843.72229	79%	806.205872	31%	805.818481	48%	805.154419	49%	805.995117	33%	804.996399	38%
14	863.429443	100%	877.302856	100%	877.358459	82%	843.70282	95%	843.72345	79%	806.362122	24%	805.868896	48%	805.161743	49%	806.08783	33%	804.999146	38%
15	872.960388	100%	877.336121	100%	877.375977	82%	843.706299	95%	843.730957	79%	806.486694	24%	805.889282	48%	805.234192	49%	806.205811	33%	805.121277	38%
16	878.273132	100%	877.350525	100%	877.453003	82%	843.710754	95%	843.737488	79%	806.491028	24%	806.012268	45%	805.238037	49%	806.780457	13%	805.252869	35%
17	882.213501	100%	877.37207	100%	877.455566	82%	843.733459	95%	843.750244	79%	809.658325	0%	806.029297	45%	805.243286	49%	817.923279	55%	805.361511	33%
18	907.541565	100%	877.386414	100%	877.464417	82%	843.737	95%	843.763733	79%	817.78656	10%	806.073853	43%	805.292725	49%	817.93103	55%	806.14856	23%
19	1021.797485	100%	877.439575	100%	877.487549	82%	843.738098	95%	843.764832	79%	817.813843	10%	806.152893	43%	805.610596	41%	817.93689	55%	817.962891	13%
20	1046.359375	100%	877.453979	100%	877.50769	82%	843.757751	95%	843.768616	79%	817.816833	10%	806.190369	43%	805.661438	41%	817.950867	55%	818.065063	13%
21	1062.642822	100%	877.45459	100%	877.562439	82%	843.76123	95%	843.772278	79%	817.817871	10%	806.345276	38%	805.875305	31%	817.953186	55%	818.122253	13%
22	1068.436401	100%	877.458862	100%	877.574036	82%	843.767212	95%	843.800415	79%	817.834167	10%	806.448608	33%	806.047729	28%	817.959778	55%	818.164001	13%
23	1070.151733	100%	877.461121	100%	877.582947	82%	843.768921	95%	843.806946	79%	817.849487	10%	806.714661	30%	817.896973	92%	817.973022	55%	818.263794	13%
24	1084.597046	100%	877.475708	100%	877.583984	82%	843.769043	95%	843.814819	79%	818.751587	10%	817.857117	30%	817.918213	92%	817.990479	55%	825.181519	3%
25	1154.829346	100%	877.478821	100%	877.586731	82%	843.777466	95%	843.815674	79%	819.685974	0%	817.862793	30%	817.926453	92%	818.009216	55%	843.923523	85%

Input | Avr(Q) | Avr2(W) | Input Bact | Analysis(E)

Figure I-c. Layout of the first step to be applied in program 2. Row 1 represents the percentage error for each spot; Row 2 indicates each *S. aureus* strain and the other rows refer to the m/z values with their percentage of occurrence.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Centroid Mass	Matrix	SA16	SA103	SA125	SA126	SA34	SA36	SA94	SA128	SA23	SA24	SA57	SA79	SA89	SA95	Species	+-%	
2	802.2771											55%				14%		0.00%	
3	803.0066525	100%							26%		38%	78%	23%	23%	18%	36%		0.01%	
4	803.148743	100%							26%		38%	78%	23%	23%	24%	36%		0.00%	
5	803.182068	100%							41%		38%	78%	23%	23%	24%	36%		0.00%	
6	803.310608	100%							41%	18%	38%	78%	23%	23%	24%	39%		0.01%	
7	803.554504	100%						25%	46%	18%	38%	78%	33%	35%	24%	41%		0.00%	
8	803.6395875	100%						25%	46%	18%	38%	78%	35%	35%	24%	41%		0.00%	
9	803.7134705	100%						25%	46%	18%	38%	78%	35%	35%	24%	48%		0.00%	
10	803.759613	100%						33%	46%	18%	38%	78%	35%	35%	24%	48%		0.00%	
11	803.796265	100%						38%	49%	18%	38%	78%	35%	35%	24%	48%		0.00%	
12	803.9155275	100%					24%	38%	49%	18%	38%	78%	35%	35%	38%	57%		0.00%	
13	803.952209	100%					24%	38%	51%	18%	38%	78%	35%	35%	38%	57%		0.00%	
14	803.966858	100%					28%	38%	51%	18%	38%	78%	35%	35%	38%	57%		0.00%	
15	803.997986	100%					28%	45%	51%	33%	38%	78%	35%	35%	38%	61%		0.00%	
16	804.094177	100%					28%	45%	51%	33%	38%	78%	35%	38%	38%	61%		0.00%	
17	804.1824035	100%					28%	48%	51%	33%	38%	78%	35%	38%	38%	61%		0.01%	
18	804.443512	100%					28%	48%	51%	33%	38%	78%	35%	38%	38%	68%		0.02%	
19	804.7067565	100%					28%	50%	51%	33%	38%	78%	35%	38%	38%	68%		0.00%	
20	805.1897585	100%					34%	50%	51%	33%	38%	78%	35%	38%	38%	68%		0.05%	
21	805.7264095	100%					34%	50%	51%	33%	38%	75%	35%	38%	38%	68%		0.01%	
22	806.0176695	100%					34%	50%	51%	33%	38%	70%	35%	38%	38%	68%		0.02%	
23	806.1953125	100%					34%	50%	51%	33%	38%	68%	35%	38%	38%	68%		0.00%	
24	806.256012	100%					34%	50%	49%	33%	38%	68%	35%	38%	38%	68%		0.01%	
25	806.345276	100%					34%	50%	49%	33%	38%	65%	35%	38%	38%	68%		0.00%	

Input | Avr(O) | Avr2(W) | Input Bact | Analysis(E)

Figure I-d. Layout of the second step to be applied in program 2. Row 1 indicates each *S. aureus* strain and the other rows refer to the percentage of occurrence of each centroid mass indicated in column A.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Centroid Mass	Matrix	SA16	SA103	SA125	SA126	SA34	SA36	SA94	SA128	SA23	SA24	SA57	SA79	SA89	SA95	Species	+-%	
2	942.475647						1											0.1617%	
3	1820.236572														1	1		0.0099%	
4	2239.326172	1					1			1					1			0.1276%	
5	2300.237305	1		1														0.0000%	
6	2362.528809	1					1								1			0.0000%	
7	2364.642823	1					1			1					1			0.0745%	
8	2367.438599						1			1					1			0.0421%	
9	2755.188233		1					1	1			1						0.0246%	
10	2755.918946		1		1			1	1			1						0.0019%	
11	2756.012452		1		1	1		1	1			1						0.0010%	
12	3216.031494	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.1182%	
13	3220.454834	1	1	1	1	1	1		1	1	1	1	1	1	1	1		0.0000%	
14	3220.841309	1	1	1	1	1	1				1	1	1	1	1	1		0.0079%	
15	3221.563232	1	1	1	1	1					1	1	1	1	1	1		0.0000%	
16	3446.631104	1	1	1	1	1	1		1	1	1	1	1	1	1	1		0.0249%	
17	3450.022461		1	1	1	1	1		1	1	1	1	1	1	1	1		0.0732%	
18	3454.865479		1	1	1	1					1	1	1	1	1	1		0.0000%	
19	3464.692627		1															0.0313%	
20	4513.594482								1	1				1	1			0.1028%	
21	4822.949219		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.1233%	
22	4829.518066		1	1	1	1	1	1	1		1	1	1	1	1	1		0.0000%	
23	5250.541748							1						1				0.0749%	
24	5303.835449										1		1	1				0.0780%	
25	5314.806152										1		1					0.0000%	

Input | Avr(Q) | Avr2(W) | Input Bact | Analysis(E)

Figure I-e. Layout of the third step to be applied in program 2. Row 1 indicates each *S. aureus* strain and the other rows indicates presence or absence of each centroid mass indicated in column A. Highlighted rows are the values to be present in all strains including matrix.

Appendix II

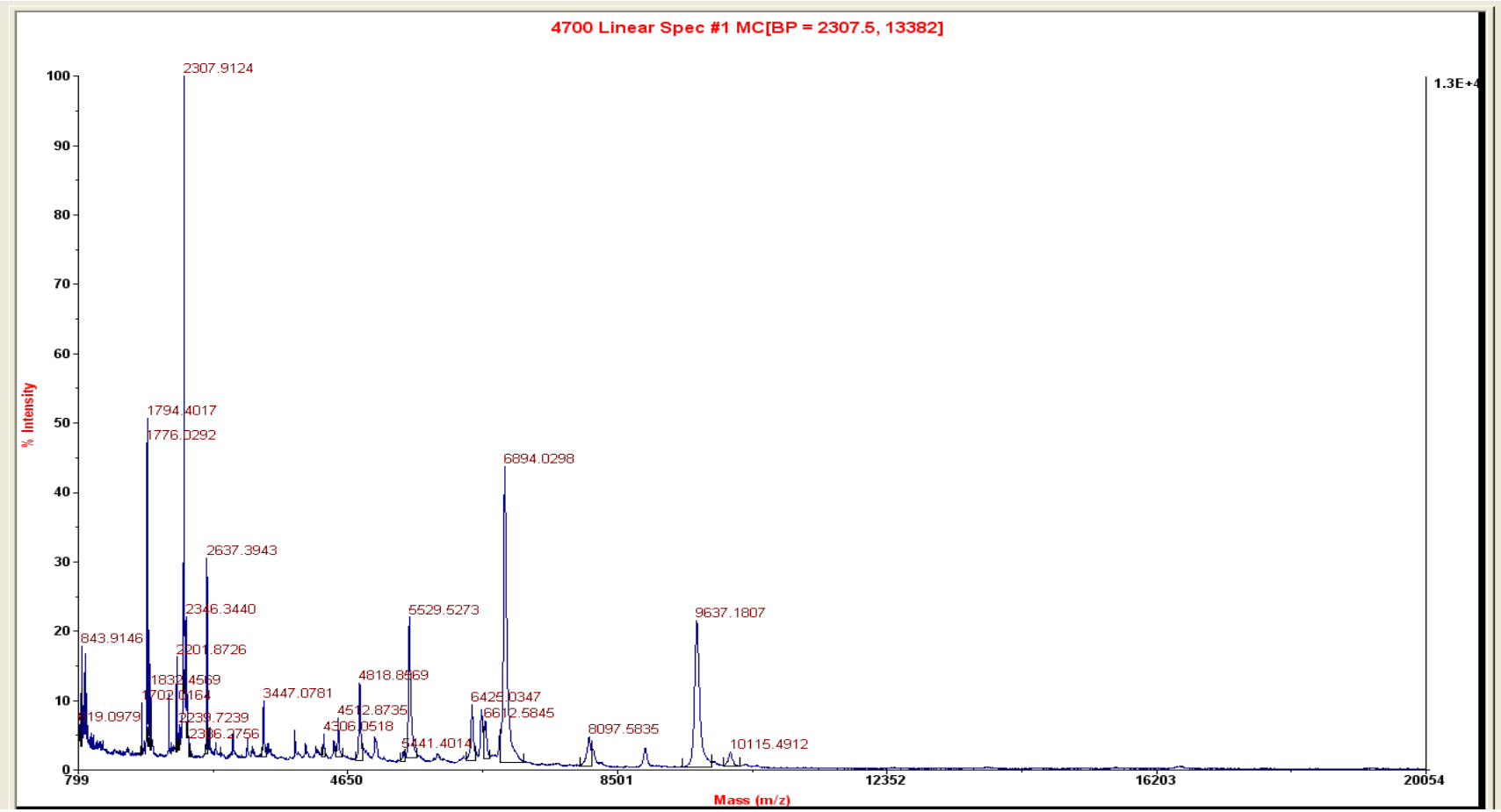

Figure II-a. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 128). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

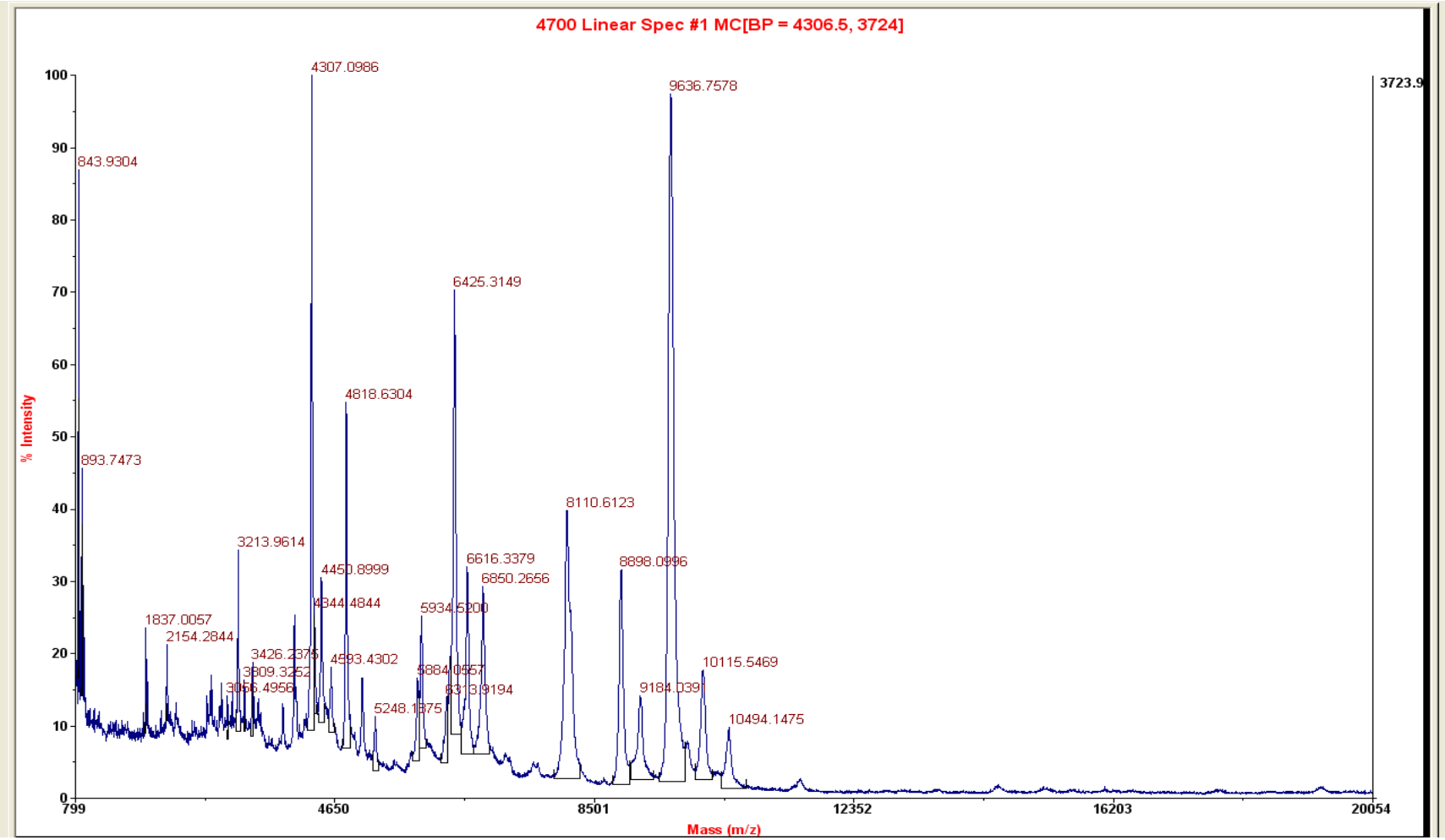

Figure II-b. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 36). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

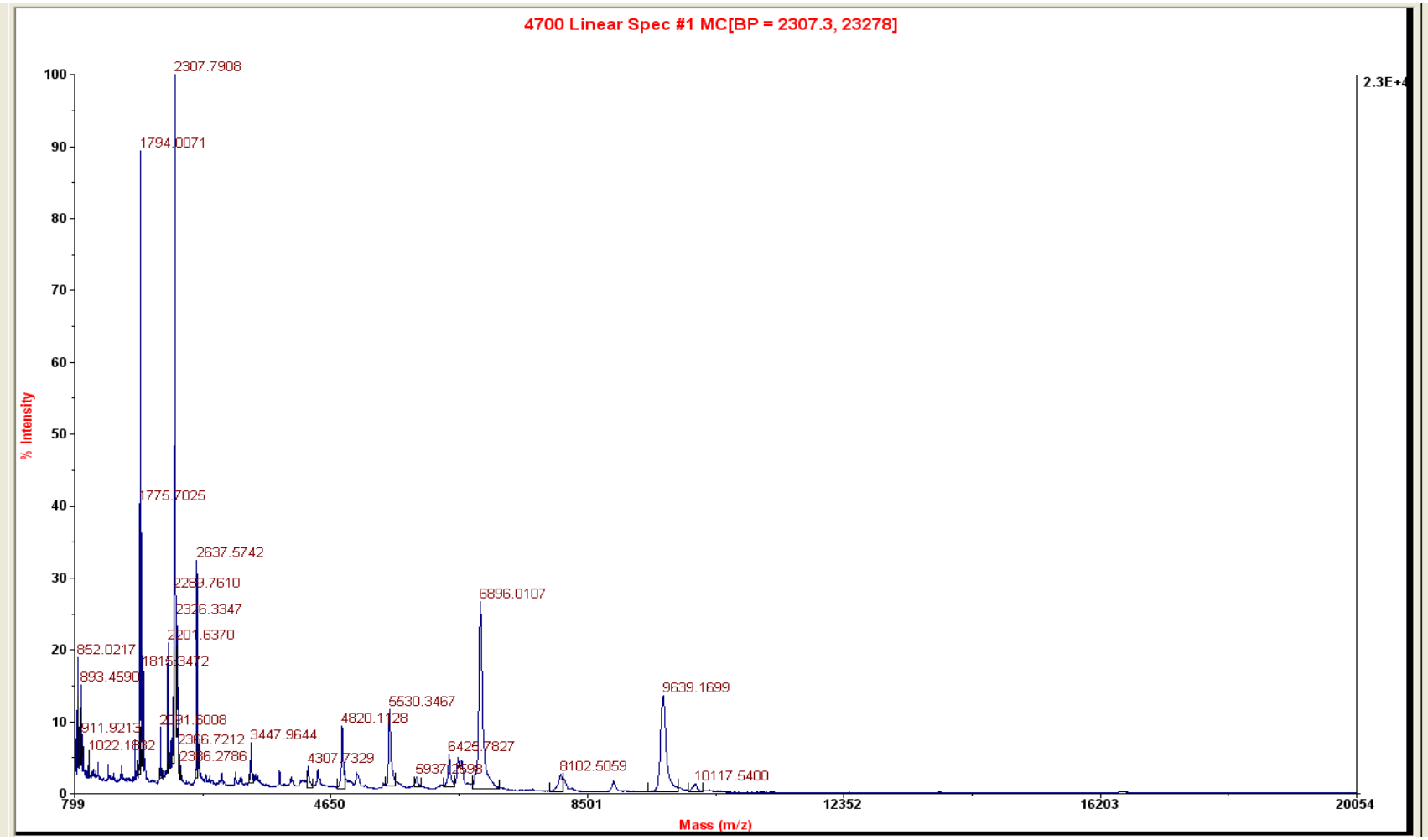

Figure II-c. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 34). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

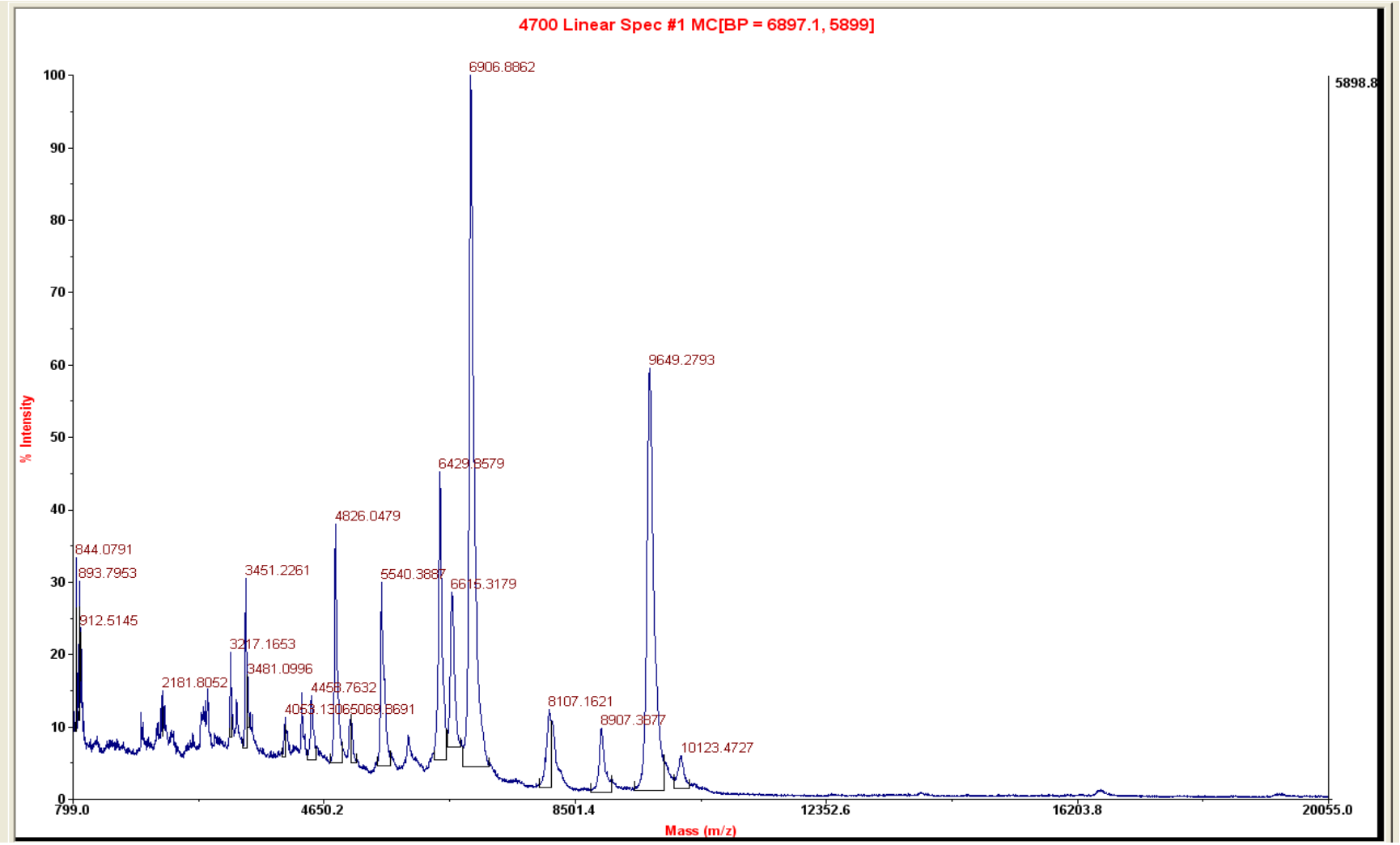

Figure II-d. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 24). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

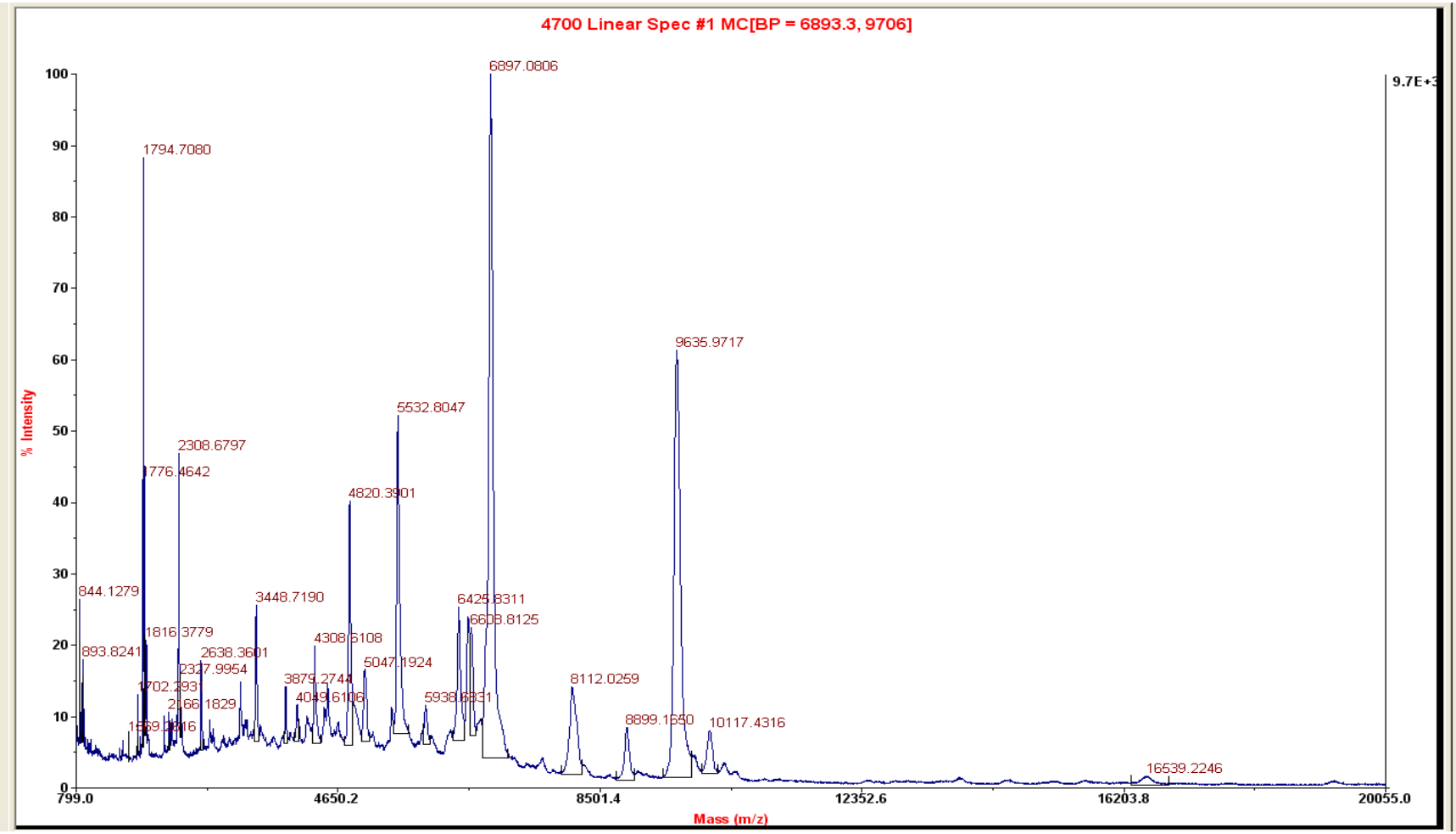

Figure II-e. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 95). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

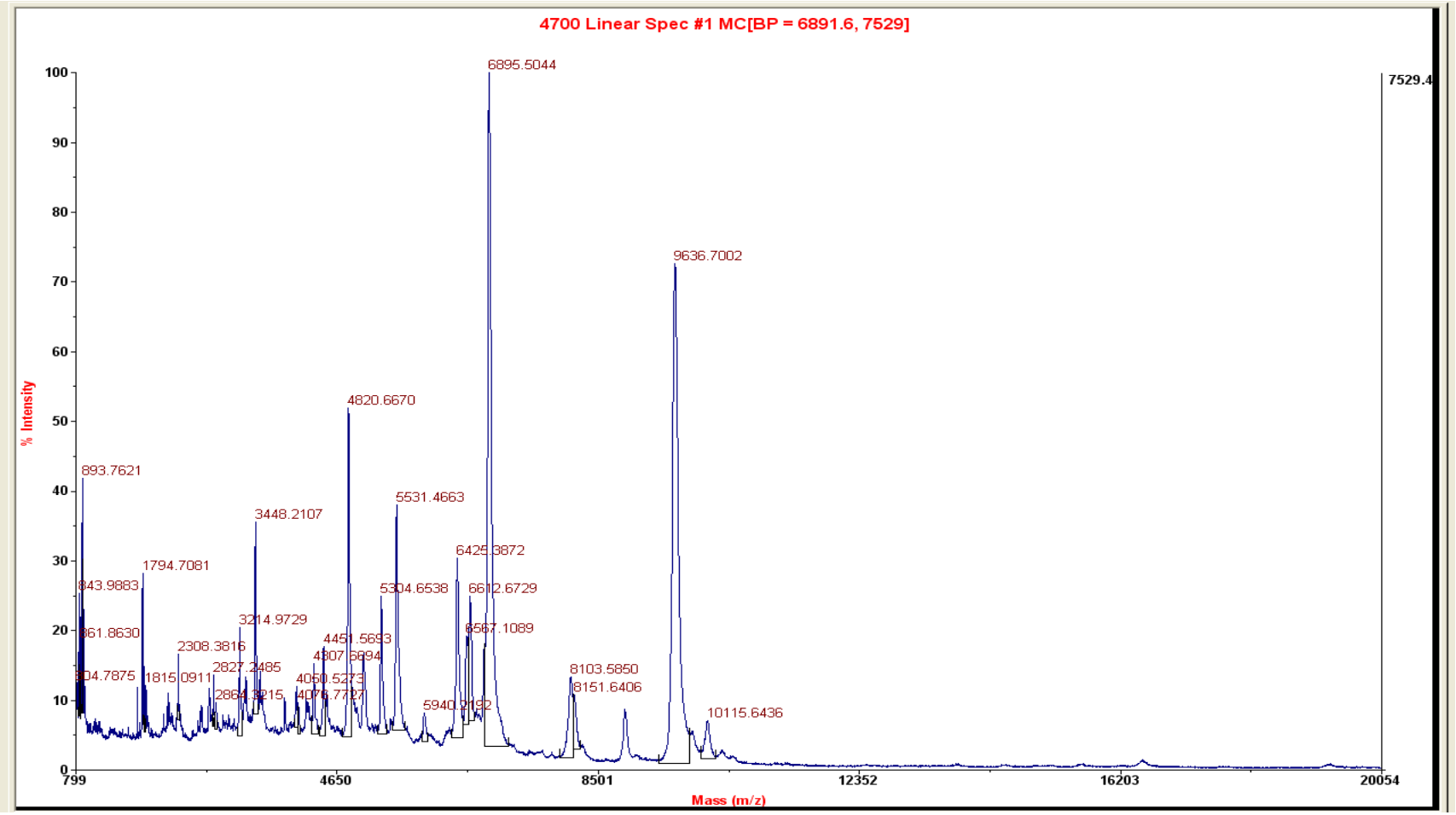

Figure II-f. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 23). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

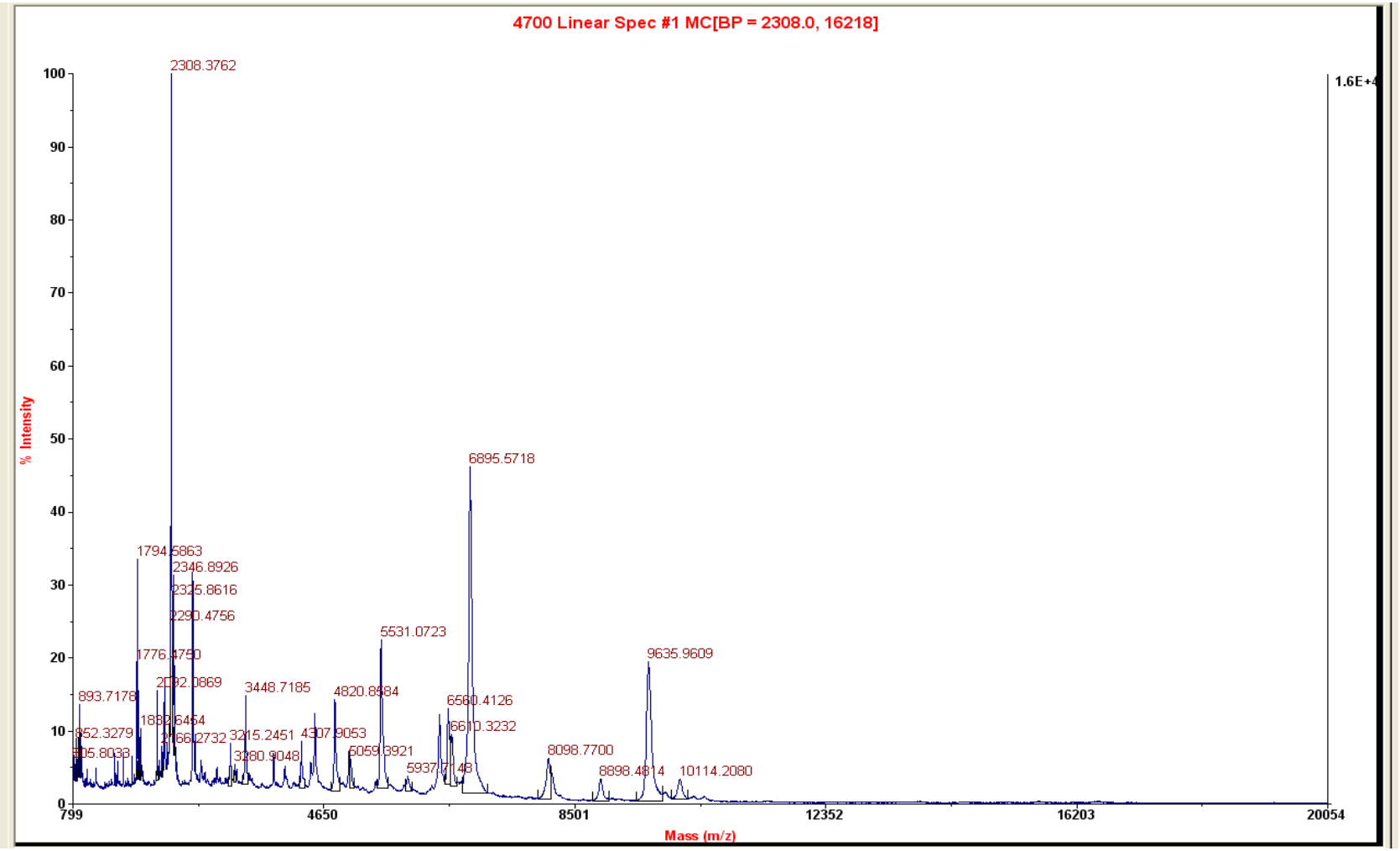

Figure II-g. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 89). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

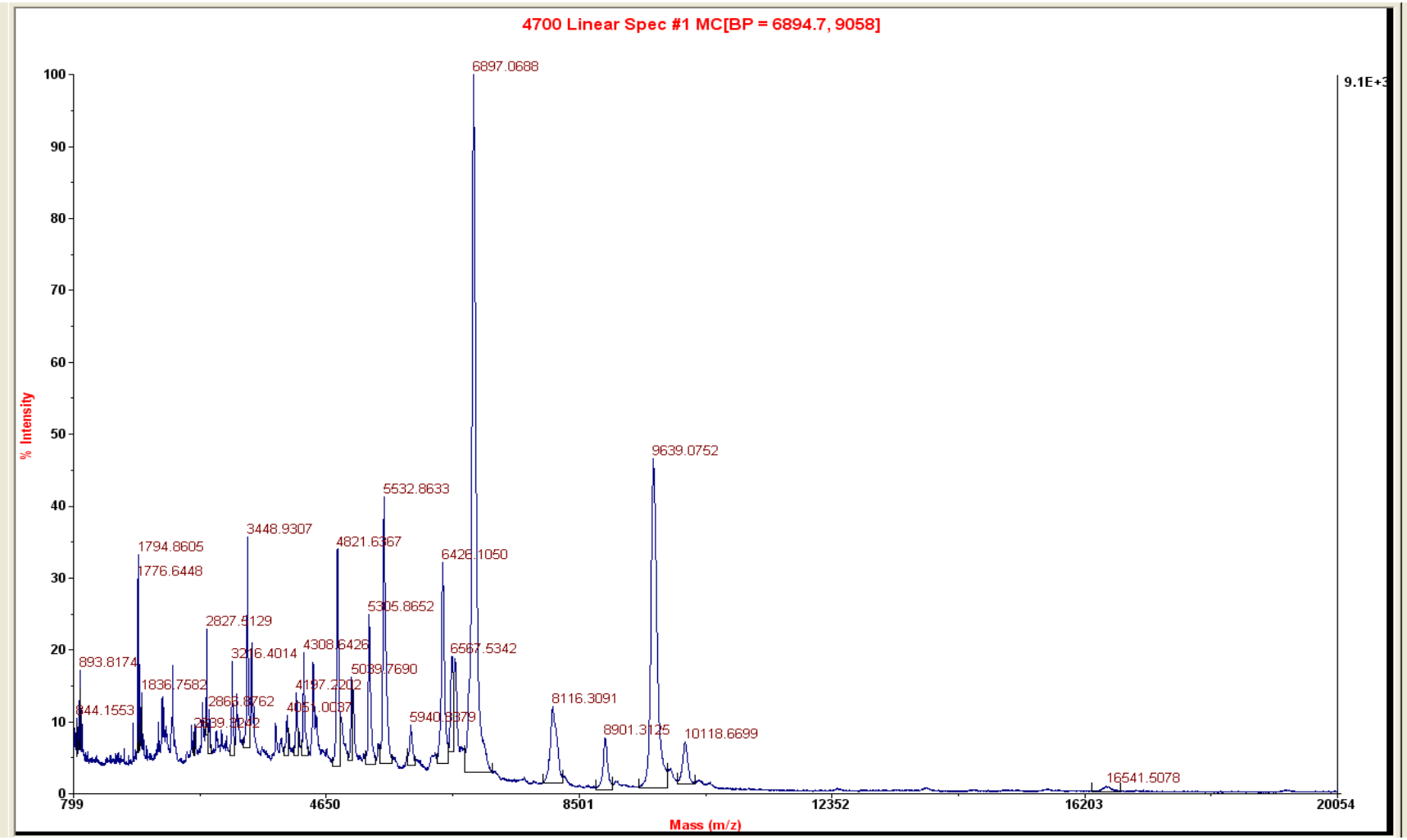

Figure II-h. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 57). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

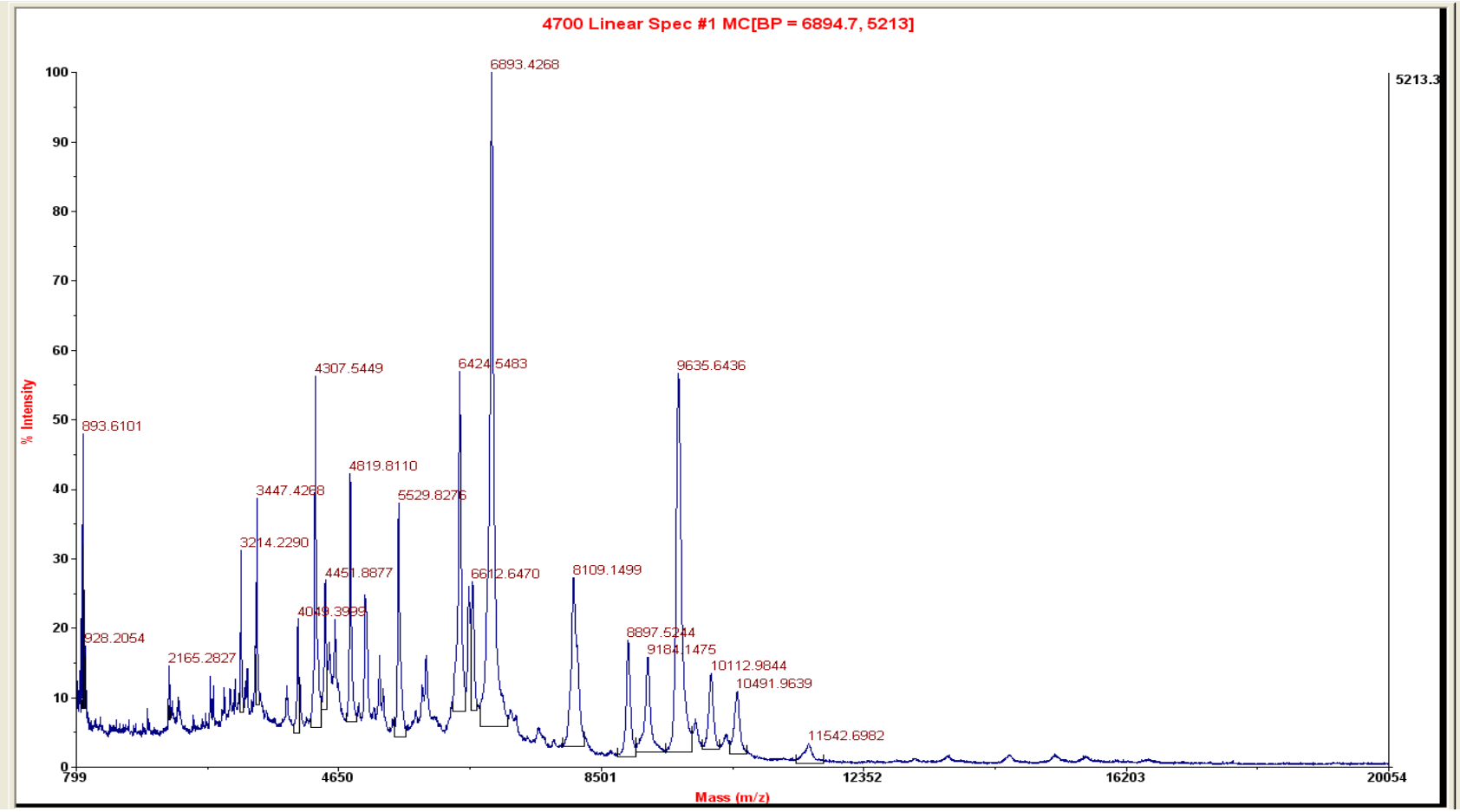

Figure II-i. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 79). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

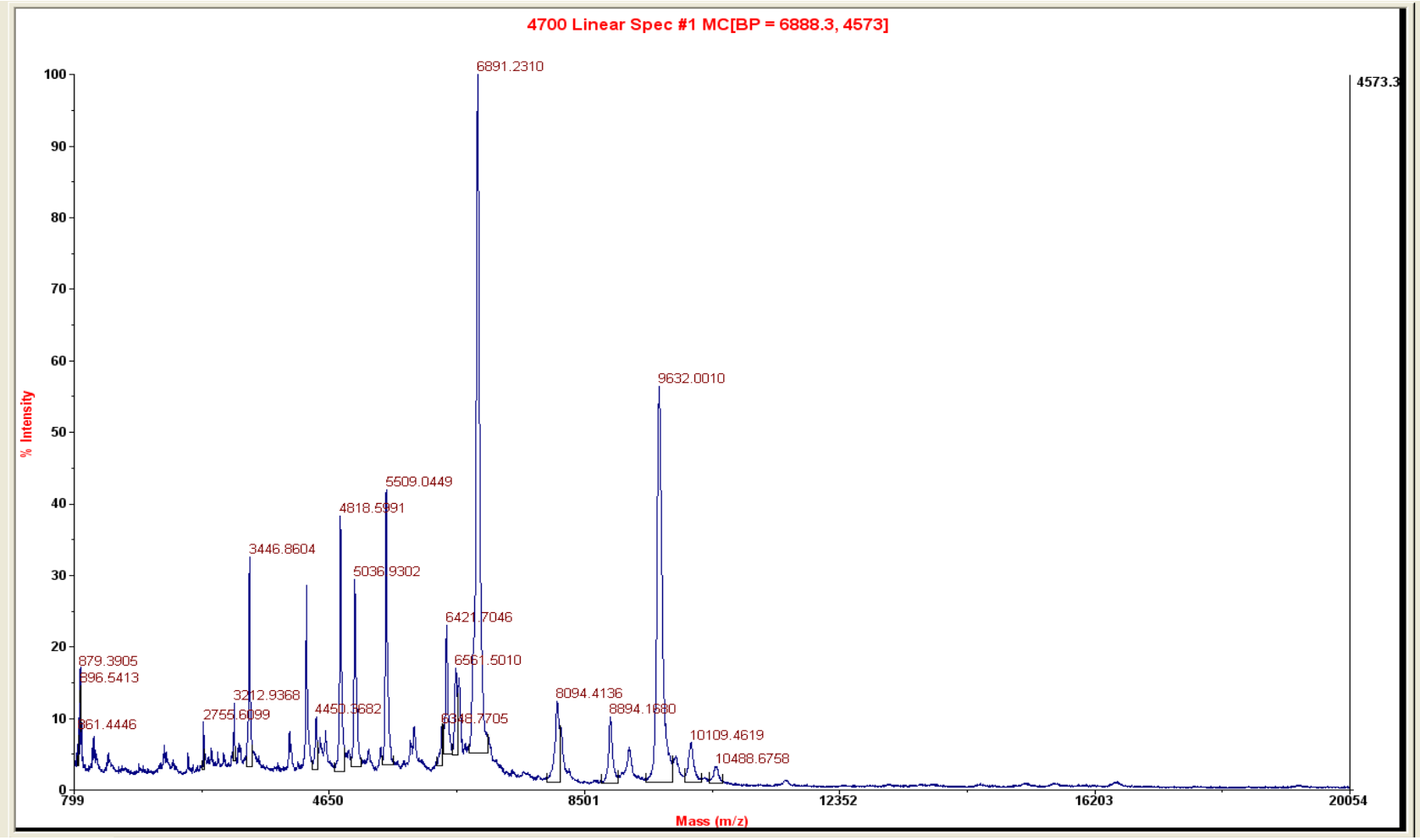

Figure II-j. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 16). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

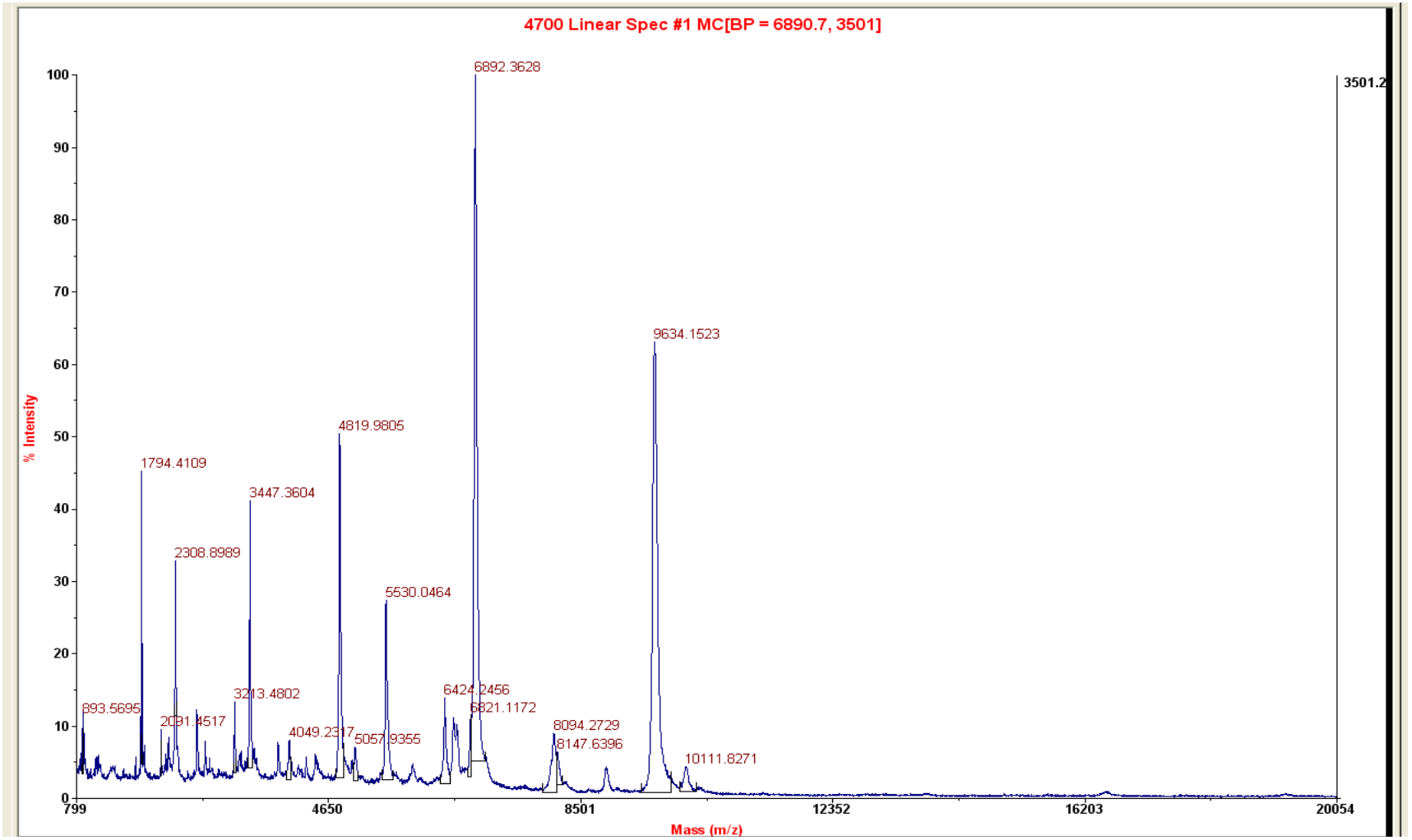

Figure II-k. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 103). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

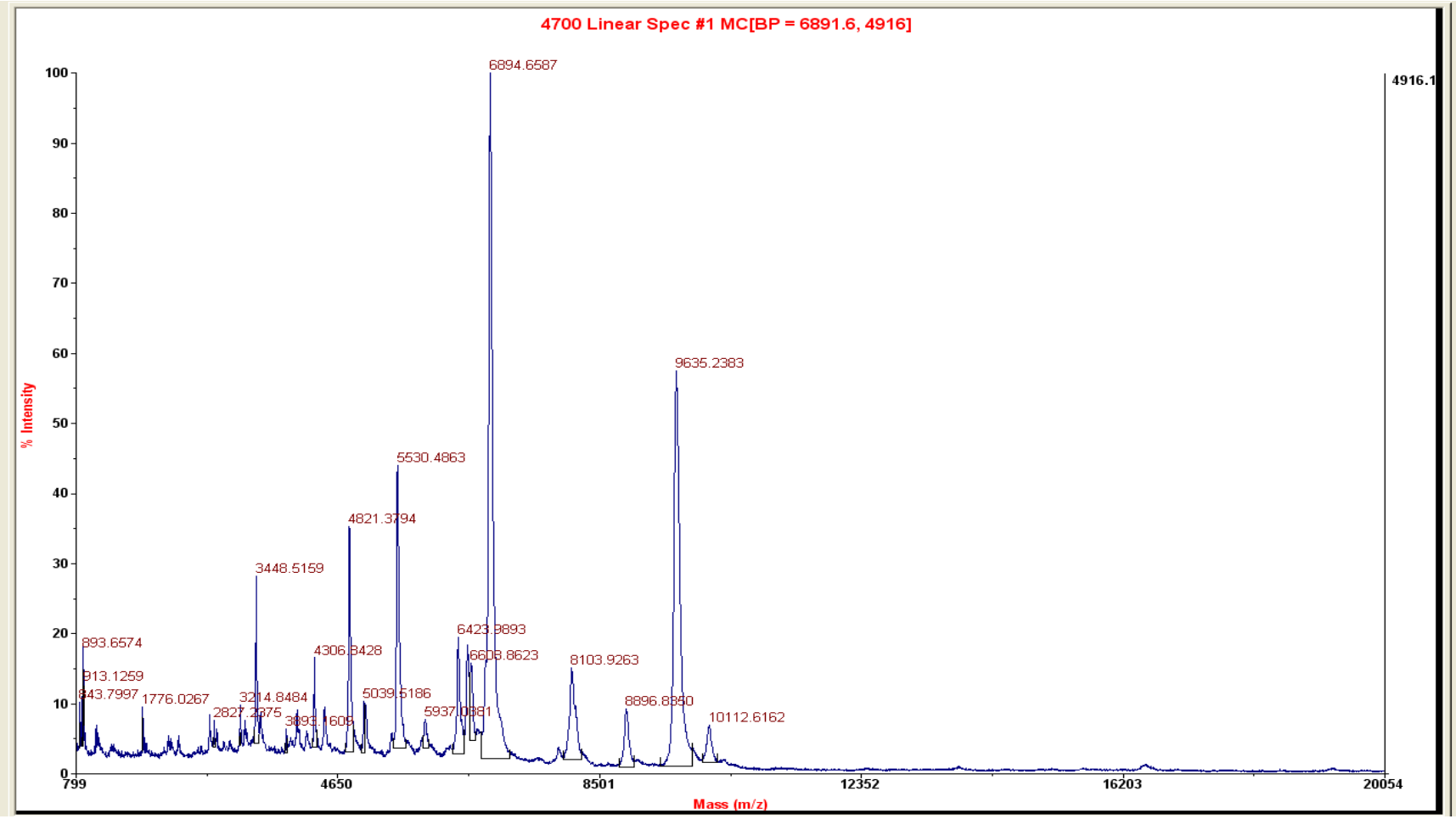

Figure II-l. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 125). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

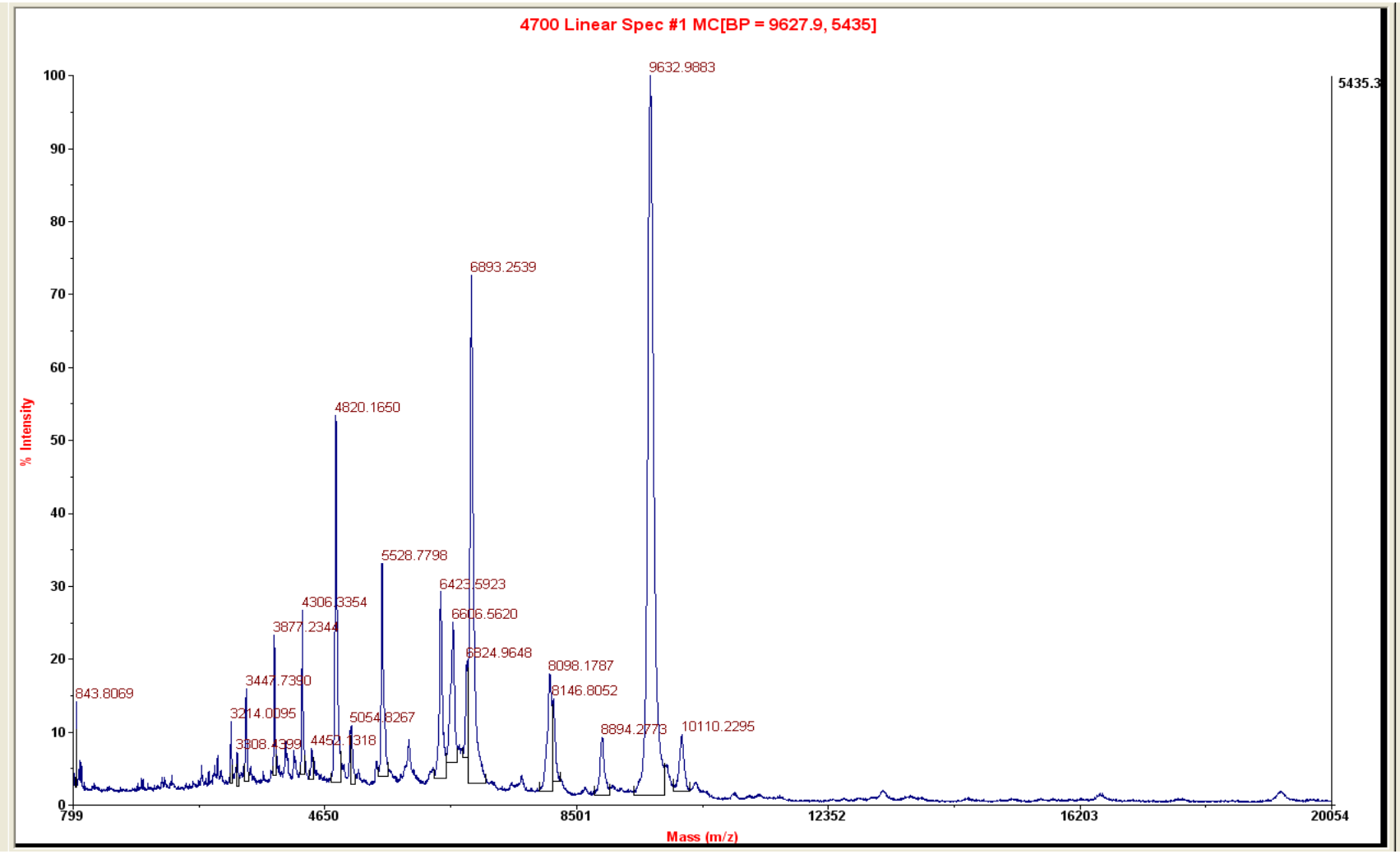

Figure II-m. MALDI-TOF MS spectrum of FAE *S. aureus* isolate (SA 126). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

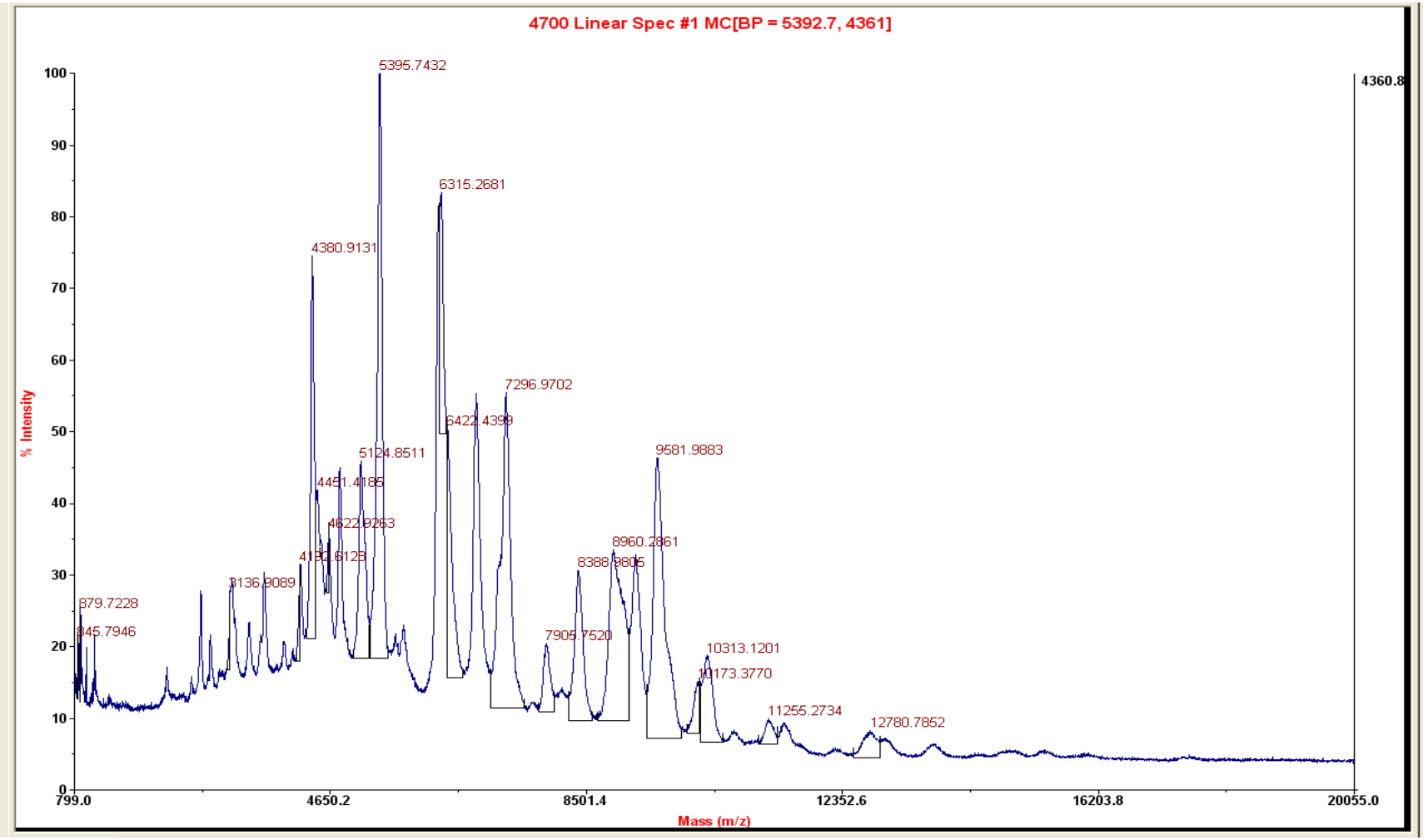

Figure II-n. MALDI-TOF MS spectrum of FAE *Escherichia coli* J96 ATCC 700336). The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.

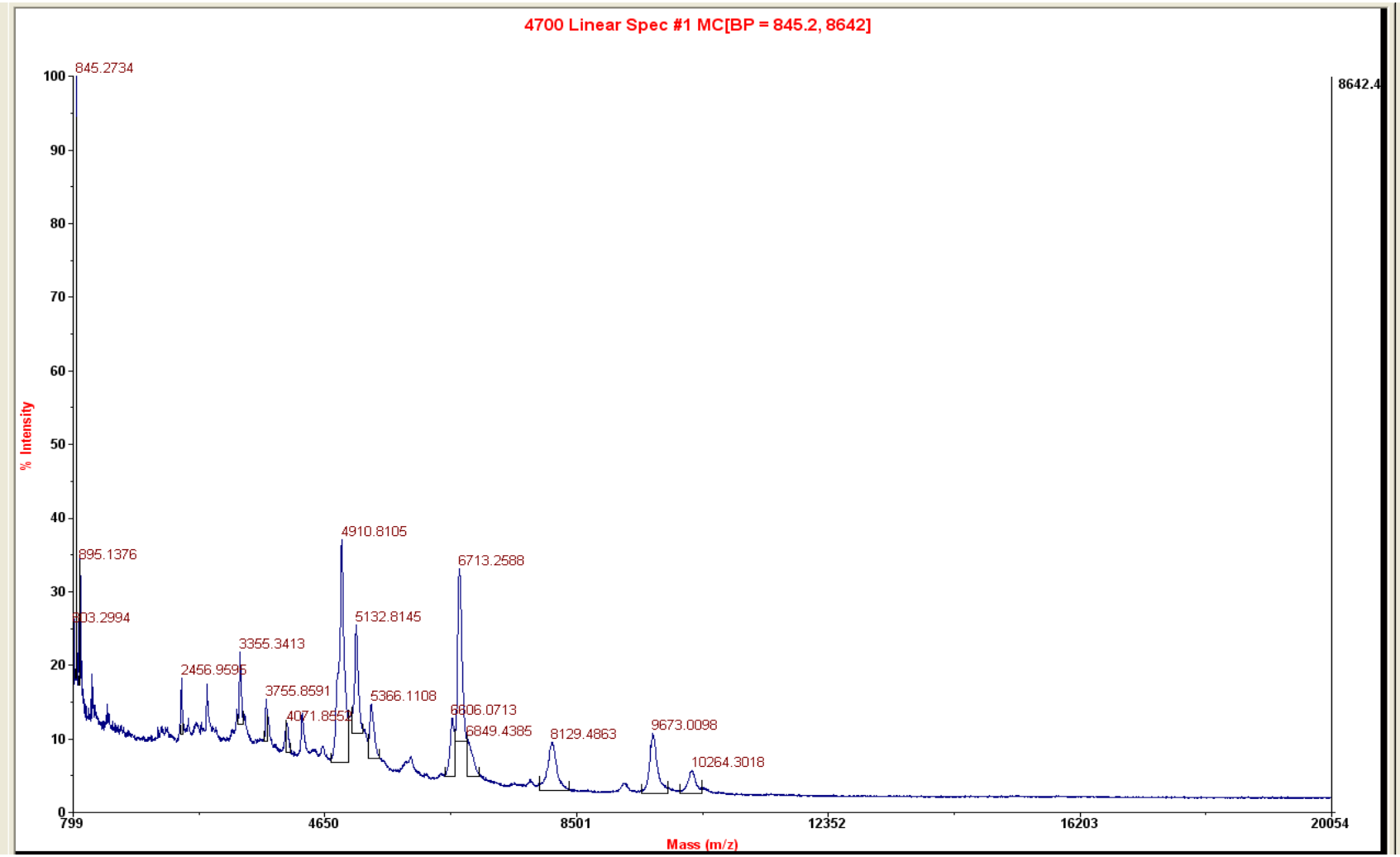

Figure II-o. MALDI-TOF MS spectrum of FAE *Staphylococcus epidermidis* 80. The m/z value shown on the *x*-axis corresponds to the molecular weight of the peak (Da) if it carries a single positive charge. The percentage intensities of the peaks are shown on the *y*-axis.