Bioanalytical Studies of Disease Protein Profiles: MALDI-ToF MS

Prajkta Satish Chivte
prajktachivte4@gmail.com

Follow this and additional works at: https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations

Part of the Analytical Chemistry Commons, and the Biochemistry Commons

Recommended Citation
Chivte, Prajkta Satish, "Bioanalytical Studies of Disease Protein Profiles: MALDI-ToF MS" (2023). Graduate Research Theses & Dissertations. 7307.
https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations/7307

This Dissertation/Thesis is brought to you for free and open access by the Graduate Research & Artistry at Huskie Commons. It has been accepted for inclusion in Graduate Research Theses & Dissertations by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
ABSTRACT

BIOANALYTICAL STUDIES OF DISEASE PROTEIN PROFILES: MALDI-TOF MS

Prajkta Satish Chivte, PhD
Department of Chemistry and Biochemistry
Northern Illinois University, 2023
Elizabeth R. Gaillard, Director

Coronavirus disease-2019 (COVID-19), which is caused by a novel coronavirus named severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has ravaged the world for the past 3 years. Even today, there still exists a need for rapid, accurate, economical and non-invasive diagnostic testing platforms that yield high specificity and sensitivity towards the constantly mutating SARS-CoV-2. Research has consistently indicated saliva to be a more amenable specimen type for early detection of SARS-CoV-2, compared to the oral and nasopharyngeal swabs. Considering the limitations and high demand of the existing COVID-19 testing platforms, this dissertation work studies used MALDI-ToF MS (matrix-assisted laser desorption/ionization-time of flight mass spectrometry) to analyze human gargle samples and compared the resulting spectra against their COVID-19. Several protein standards, including amylase, albumin, serum immunoglobulins, controls such as pre-COVID-19 saliva and heat inactivated viruses like SARS-CoV-2, MERS-CoV, H1N1 and common human coronaviruses, were simultaneously analyzed to provide a relative view of the saliva and viral proteome as they would appear in this workflow. Five potential biomarker peaks, which included host as well as viral proteins were established that
demonstrated high agreement with COVID-19 positive individuals. Overall, the accuracy of this
developed protein profile testing platform was ≥90%.

To further enhance the sensitivity of the assay, amylase, the most abundant protein found
in saliva, was depleted using a rapid and inexpensive amylase depleting device. After depleting
the amylase from the gargle/saliva samples, there was an enhancement in signal intensities of
various peaks as well as the detection of previously unobserved peaks in the MALDI-ToF spectra.
The overall specificity and sensitivity after amylase depletion was 100% and 85.17% respectively
for detecting COVID-19.

MALDI-ToF results were also compared to the RT-qPCR (reverse transcriptase
quantitative polymerase chain reaction) results in a 7-day study which was reported for the first
time with a remarkable correlation between the two techniques. Moreover, attempts were made to
detect not just the SARS-CoV-2 but also the type of variant. As mutations cause mass shifts, these
could be detected using mass spectrometry.

In essence, this research makes a significant contribution to the field of proteomics and
mass spectrometry for the analysis of protein profiles in human saliva to elucidate novel
biomarkers for COVID-19. Protein profiling of body fluids using MALDI-TOF MS could be a
promising PCR-free tool for screening SARS-CoV-2 as well as various other viruses.
ACKNOWLEDGEMENTS

First and foremost, my heartfelt gratitude to my advisor, Dr. Elizabeth R. Gaillard, for her firm commitment, persistence and expertise. Her supervision and mentorship have been the key to my research success. She has helped me broaden my horizons and pushed me to strive for excellence which has made me an independent researcher today. Her insightful feedback, constructive criticism, and invaluable suggestions have truly made my doctoral journey a successful one. I am forever indebted to her for her belief in my potential.

I am grateful to my committee members, Dr. Irina Nesterova, Dr. James Horn and Dr. Shrihari S. Kadkol for their input and support in completing this work. Completion of this research would not have been possible without our collaborators Dr. Pratool Bharti, Dr. Hamed Alhoori, Venkata Devesh Seethi, Dr. Shrihari Kadkol and Joshua Bland. Their enthusiasm, encouragement, and discussions have helped me refine my ideas and methodologies.

I am also immensely grateful to Dr. Michael Vega for mentoring me in various techniques and having countless discussions that have guided me throughout my research journey. I was fortunate to have a lab-mate like Zane LaCasse who is not only an exceptional researcher but also a kind friend. I would also like to thank the administrative staff of the Department of Chemistry and Biochemistry for providing me with the necessary resources,
facilities, and opportunities to pursue my research. A special thank you to the NIU Division of Research and Innovation Partnerships for funding this journey.

Additionally, I extend my heartfelt appreciation to all my friends, Shubham Yadav, Dr. Vinal Pardhi, Vamshi Krishna Dasaraju, Samira Arondekar, Nidhi Shah, Shraddha Suryavanshi, Dr. Sveta Padmanabhan and Aishwarya Jain for their unwavering support and understanding. Their encouragement, motivation, and belief in me have been a source of strength during both the highs and lows of this doctoral journey. Their listening ears, kind words, and occasional distractions have kept me sane and reminded me of the importance of work-life balance.

Furthermore, I would like to express my gratitude to more than 1500+ volunteers who participated in my research study. Without their willingness to donate their saliva samples, this study would not have been possible. Last but not least, I would like to express my profound appreciation to my parents and my sister. Their words of encouragement, faith in my abilities and their sacrifices for my happiness have been the source of my motivation.

In essence, this PhD dissertation has been a ride full of twists and turns with every moment more exciting than the last. I sincerely believe this could not have been accomplished if it weren’t for all the support I received from my well-wishers. Each and every accomplishment in this project has been made possible due to the contributions made by all of you, no matter how small you consider it, it has benefited me greatly.

With sincere appreciation,

Prajkta Chivte
DEDICATION

None of this would have ever been possible if it was not for my Mom and Dad who constantly showered me with love and affection and made me feel confident despite being 8000 miles away.
## TABLE OF CONTENTS

| LIST OF TABLES | iv |
| LIST OF FIGURES | vi |
| LIST OF ABBREVIATIONS | xiv |

### Chapter

1. INTRODUCTION ................................................................. 1

   SARS-CoV-2 ................................................................. 3

   Immune response to SARS-CoV-2 ........................................ 7

   Detection of SARS-CoV-2 .................................................. 9

   Saliva as a diagnostic tool ................................................ 10

   Protein Testing for COVID-19 ............................................. 11

   Mass spectrometry in clinical diagnostics .......................... 14

   MALDI-ToF .................................................................... 17

   MALDI-ToF for COVID-19 ................................................ 29
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC-MS</td>
<td>31</td>
</tr>
<tr>
<td>LC-MS for COVID-19 diagnosis</td>
<td>32</td>
</tr>
<tr>
<td>Machine learning</td>
<td>34</td>
</tr>
<tr>
<td>Dissertation work</td>
<td>35</td>
</tr>
</tbody>
</table>

2. MATERIALS AND METHODS ................................................................. 39

Materials ......................................................................................... 39
Instrumentation ................................................................................ 40
Methods ............................................................................................ 41

Human gargle sample collection and storage .................................. 41
Sample preparation for human gargle samples .............................. 43
Sample preparation for human saliva samples ............................ 44
Sample preparation for human proteins standards/controls .......... 44
MALDI-ToF sample spotting .............................................................. 45
MALDI-ToF instrument parameters ................................................. 45
MALDI-ToF calibration ....................................................................... 46
MALDI-ToF data analysis by Area under the curve (AUC) ............ 46
Receiver Operating Characteristic (ROC) analysis ........................ 47
Sample preparation for gargle samples using an amylase depleting device ................................................................. 48
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample preparation for saliva samples using an amylase depleting device</td>
<td>49</td>
</tr>
<tr>
<td>Sample preparation for human protein standards and controls using an amylase depleting device</td>
<td>50</td>
</tr>
<tr>
<td>SDS-PAGE analysis</td>
<td>50</td>
</tr>
<tr>
<td>Sample preparation for recombinant full-length S protein of SARS-CoV-2 variants</td>
<td>51</td>
</tr>
<tr>
<td>RT-qPCR and Next-generation sequencing</td>
<td>52</td>
</tr>
<tr>
<td>3. MALDI-TOF AS A POTENTIAL DIAGNOSTIC TOOL FOR DETECTION OF COVID-19</td>
<td>53</td>
</tr>
<tr>
<td>Introduction</td>
<td>53</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>56</td>
</tr>
<tr>
<td>4. MALDI-TOF PROTEIN PROFILING OF AMYLASE DEPLETED GARGLE SAMPLES FOR COVID-19 TESTING</td>
<td>73</td>
</tr>
<tr>
<td>Introduction</td>
<td>73</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>76</td>
</tr>
<tr>
<td>5. DISCRIMINATING POWER OF MALDI-TOF TO DETECT SARS-COV-2 AND ITS VARIANTS</td>
<td>96</td>
</tr>
<tr>
<td>Introduction</td>
<td>96</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>99</td>
</tr>
<tr>
<td>6. CONCLUSION AND FUTURE WORK</td>
<td>119</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>123</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>1.1</td>
<td>Summary of MALDI-ToF based assays for detection of COVID-19 using various specimen types</td>
</tr>
<tr>
<td>2.1</td>
<td>Potential protein biomarker ranges identified by MALDI-ToF from gargle samples</td>
</tr>
<tr>
<td>2.2</td>
<td>List of recombinant full-length S proteins of SARS-CoV-2 variants</td>
</tr>
<tr>
<td>3.1</td>
<td>Percent agreement between RT-PCR results and MALDI-ToF results as determined by cut-off AUC values</td>
</tr>
<tr>
<td>3.2</td>
<td>Summary of ROC curve analysis for each potential biomarker as calculated for AUC of peaks within associated m/z ranges</td>
</tr>
<tr>
<td>5.1</td>
<td>Comparison between the theoretical masses in Daltons and the MALDI-ToF m/z of the Wuhan strain of SARS-CoV-2 and four other variants</td>
</tr>
<tr>
<td>5.2</td>
<td>Comparative analysis of S1 fragment peak of SARS-CoV-2 variants analyzed MALDI-ToF MS</td>
</tr>
<tr>
<td>5.3</td>
<td>RT-qPCR results of saliva samples collected from a COVID-19 positive donor for 7 consecutive days</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>8</td>
</tr>
<tr>
<td>1.6</td>
<td>12</td>
</tr>
<tr>
<td>1.7</td>
<td>15</td>
</tr>
<tr>
<td>1.8</td>
<td>16</td>
</tr>
<tr>
<td>1.9</td>
<td>18</td>
</tr>
<tr>
<td>1.10</td>
<td>20</td>
</tr>
</tbody>
</table>
Figure | Page
---|---
1.11 Diagrammatic representation of a TOF mass analyzer where ions are separated on the basis of their m/z .............................................................23
1.12 An illustration of a reflectron ToF mass analyzer ........................................27
1.13 A diagrammatic representation of delayed pulse extraction using the principle of time-lag focusing .................................................................28
1.14 General workflow in the profiling approach using MALDI-TOF mass spectrometry ....28
1.15 Schematics of Liquid Chromatography-Mass Spectrometry ...............................33
1.16 An overview of the biomarker discovery process using the machine learning approach ...35
2.1 An illustration of Shimazu AXIMA Performance hardware ..................................42
2.2 An illustration of salivary amylase depleting device loaded with a gargle sample ........49
3.1 MALDI-ToF mass spectra of a gargle sample from an example donor who tested COVID-19 negative along with the negative control ........................................57
3.2 MALDI-ToF mass spectra of a gargle sample from a COVID-19 negative donor and an overlay of two COVID-19 positive donors ........................................59
3.3 MALDI-ToF mass spectra of human serum derived IgA, IgG and IgM isolated from human serum reduced with 1 M DTT for 10 min ...........................................61
3.4 MALDI-ToF mass spectrum of human salivary α-amylase .....................................62
3.5 MALDI-ToF spectrum of heat inactivated cell lysate and supernatant of Vero E6 cells infected with SARS-CoV-2, which was utilized as the positive control ..................................64

3.6 MALDI-ToF mass spectrum of a COVID-19 positive saliva sample beyond the limit of quantitation for RT-qPCR ........................................................................................................65

3.7 AUC values for the 60 gargle samples for the five potential biomarker peak range ........66

3.8 ROC analysis results on the potential protein biomarker ranges identified by MALDI-ToF from gargle samples ........................................................................................................................................................................69

4.1 Iodine test to determine the presence of starch after multiple water washes ...............78

4.2 MALDI-ToF mass spectra of a COVID-19 positive gargle treated with LBSD-X buffer and an overlay of the same sample treated with the new optimized buffer ..................................79

4.3 MALDI-ToF mass spectra of human salivary α-amylase and an overlay of the same sample after passing through the amylase depleting device .................................................................81

4.4 MALDI-ToF mass spectra of standard HSA and an overlay of the same sample after passing through the amylase depleting device ..............................................................................................82

4.5 MALDI-ToF mass spectra of a standard human serum IgA and an overlay of the same sample after passing through the amylase depleting device .................................................................83

4.6 MALDI-ToF mass spectra of pre-COVID-19 saliva and an overlay of the same sample after passing through the amylase depleting device ..............................................................................................84
4.7 SDS-PAGE analysis of high-abundance proteins in human saliva and an example gargle sample stained with Coomassie-Brilliant-Blue under reducing conditions .........................85

4.8 MALDI-ToF mass spectra of a gargle sample from a COVID-19 positive donor and an overlay of the same sample passed through the amylase depleting device .........................87

4.9 MALDI-ToF mass spectra of gargle samples from a COVID-19 positive donor eluent and COVID-19 negative donor eluent .................................................................88

4.10 SDS-PAGE analysis of COVID-19 positive and COVID-19 negative samples and the respective eluents ...................................................................................................91

4.11 ROC analysis of the biomarker ranges 11,140-11,160 m/z and 78,600-80,500 m/z after amylase depletion ........................................................................................................92

4.12 ROC analysis of the biomarker ranges 11,140-11,160 m/z and 78,600-80,500 m/z before amylase depletion ....................................................................................................93

5.1 An overlay of MALDI-ToF mass spectra of cell lysate and supernatant of cells infected with H1N1 and heat-inactivated cell lysate and supernatant of cells infected with SARS-CoV-2 ........................................................................................................100

5.2 An overlay of MALDI-ToF mass spectra of cell lysate and supernatant of cells infected with HCoV-NL63, HCoV-OC43, HCoV-229E and SARS-CoV-2........................................103

5.3 An overlay of MALDI-ToF mass spectra of cell lysate and supernatant of cells infected with MERS-CoV and SARS-CoV-2 ........................................................................104
### Figure

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4</td>
<td>MALDI-ToF spectrum of fetal bovine serum .................................................................105</td>
</tr>
<tr>
<td>5.5</td>
<td>An overlay of MALDI-ToF mass spectra of the recombinant full-length S protein of Wuhan, Spike D614G Variant, Delta Variant and Omicron variants B.1.1.529 lineage and BA.2 lineage ..........................................................................................................................107</td>
</tr>
<tr>
<td>5.6</td>
<td>An overlay of MALDI-ToF mass spectra of the three COVID-19 positive saliva donors infected with different variants: Delta, Omicron BA.1 and Omicron BA.2 ................................109</td>
</tr>
<tr>
<td>5.7</td>
<td>An overlay of MALDI-ToF mass spectra of the four COVID-19 positive saliva donors infected with different variants: Omicron BA.1 and Omicron BA.2 ........................................110</td>
</tr>
<tr>
<td>5.8</td>
<td>MALDI-ToF mass spectra overlay of gargle samples eluents collected from a COVID-19 positive donor for seven consecutive days ..............................................................113</td>
</tr>
<tr>
<td>5.9</td>
<td>Comparison between MALDI-ToF and RT-qPCR analysis .....................................................115</td>
</tr>
<tr>
<td>5.10</td>
<td>Comparison of DNA sequencing analysis between Day 1 and Day 7 saliva samples showing the differences between the two Omicron variants BA.2.12.1 and BA.4 .............................117</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

ACE2 - Angiotensin-converting enzyme 2
AUC - Area under the curve
COVID-19 - Coronavirus disease 2019
Ct - Cycle threshold
DTT - Dithiothreitol
E Protein - Envelope protein
EUA - Emergency use authorization
FDA - Food and drug administration
HSA - Human serum albumin
IgA - Immunoglobulin A
IgG - Immunoglobulin G
IgM - Immunoglobulin M
IDPH - Illinois Department of Public Health
LoD - Limit of detection
LC-MS - Liquid chromatography mass spectrometry
M Protein - Membrane protein
MALDI-ToF MS - Matrix-assisted laser desorption/ionization-time of flight mass spectrometry
MERS-CoV - Middle eastern respiratory syndrome coronavirus
ML - Machine learning
N Protein - Nucleocapsid protein
NP - Nasopharyngeal
RBD - Receptor binding domain
RNA - Ribonucleic acid
ROC - Receiver operating characteristic
RT-qPCR - Reverse transcriptase quantitative polymerase chain reaction
S Protein - Spike protein
SARS-CoV - Severe acute respiratory syndrome coronavirus
SARS-CoV-2 - Severe acute respiratory syndrome coronavirus type 2
SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis
VEP - Viral envelope protein
WHO - World health organization
CHAPTER 1

INTRODUCTION

A novel form of coronavirus named SARS-CoV-2 (severe acute respiratory syndrome coronavirus type 2), classified under the $\beta$-coronavirus genus of the Coronaviridae family, was first detected in a seafood market of Wuhan, China in December 2019. Within three months of its emergence, the outbreak of this virus was declared a pandemic by the World Health Organization (WHO) and the disease was named coronavirus disease 2019 (COVID-19).\textsuperscript{1} The SARS-CoV-2 virus is primarily transmitted through respiratory droplets of people infected with it. Typical symptoms of COVID-19 include fever, dry cough, fatigue, and anosmia while individuals with co-morbidities are more likely to develop a severe respiratory disease that requires hospitalization.\textsuperscript{2} The individuals infected with SARS-CoV-2 can be broadly classified as asymptomatic, presymptomatic and symptomatic depending on the signs of sickness. While the extent of transmission by asymptomatic individuals remains unclear, percentages of these individuals are reported to range from 20-50%.\textsuperscript{3} For symptomatic individuals, the disease can be described in four stages (Figure 1.1). Stage 1 is a mild phase with fever, malaise and dry cough. Stage 2 is usually characterized by pneumonia and conditions may continue to deteriorate leading to acute respiratory distress syndrome and multiorgan failure (Stage 3). Stage 4 includes the group of people who recover from Stage 3.\textsuperscript{4}
Evidence suggests that the infectivity, as well as the transmissibility of the virus, is increasing as the virus continues to undergo genetic mutations. The continual evolution of SARS-CoV-2 has led to the emergence of several variants of the virus, amongst which Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2) and Omicron (B.1.1.529) are of concern around the globe. As of May 2023, globally more than 765 million people have been infected while more than 6.9 million people have succumbed to COVID-19.
SARS-CoV-2

SARS-CoV-2 is a spherical, enveloped virus with a single positively stranded RNA genome. It is 80-120nm in diameter with multiple outward spike-like projections (Figure 1.2). The other two viruses from the same family, namely SARS-CoV (severe acute respiratory syndrome coronavirus) and MERS-CoV (middle east respiratory syndrome coronavirus) that have caused respiratory outbreaks in the past, share 79% and 50% genome sequence identity with SARS-CoV-2, respectively. The genome of SARS-CoV-2 is ~30,000 nucleotides in length and encodes at least 31 proteins that are classified as structural, non-structural and accessory proteins (Figure 1.3). Together, these proteins are responsible for the structural and functional characteristics of the virus. The four structural proteins are named as: spike (S), membrane (M), envelope (E) and nucleocapsid (N) protein.

The S protein (150-200kDa, UniProtKB P0DTC2) is one of the largest known class I fusion proteins consisting of two subunits, S1 and S2. It is heavily glycosylated and is responsible for initiating the first step of host invasion. The S protein binds to the cellular surface receptor angiotensin-converting enzyme 2 (ACE-2) through the receptor-binding domain (RBD), an essential step for membrane fusion and cellular invasion. Activation of the S protein that exists in a trimeric pre-fusion form requires cleavage between the S1 and S2 subunits by host serine protease TMPRSS2 (transmembrane serine protease 2) and a subsequent conformational change. Figure 1.4 displays the presence of multiple conformational states of the SARS-CoV-2 S protein. The S protein plays a key role in determining the host immune response and thus any mutations in its sequence can result in altered pathogenesis and hence the S protein has become the main target for the design of drugs and vaccines.
Figure 1.2: Basic structure of SARS-CoV-2 virus. Image from Fischetti et al.\textsuperscript{10}

Figure 1.3: Genomic organization of SARS-CoV-2. Image from Redondo et al.\textsuperscript{8}
One such remarkable example of altered pathogenesis is the Omicron sublineages that harbor multiple mutations, especially in the RBD, compared to the RBDs of the SARS-CoV-2 Wuhan-Hu-1 variant. These mutations have resulted in ~2-6-fold higher affinity for ACE-2 receptors as reported by Bowen et al. by performing biolayer interferometry and surface plasmon resonance analyses. The Omicron variants had a 13-fold increase in transmissibility and were 2.8 times more infectious than the Delta variant. These variants were responsible for causing the highest spike in the total number of COVID-19 cases as of May 2023 even though a majority of the world population was fully vaccinated at the time of its emergence. Studies have demonstrated that the Omicron variants showed a significant reduction in the efficacy of major approved COVID-19 vaccines namely mRNA-1273 (Moderna), BNT162b2 (Pfizer-BioNTech), Ad26.COV2.S (Johnson-Johnson) and ChAdOx1 nCoV-19 (Astra Zeneca), which were designed to target the spike protein ACE2 interaction considering the original Wuhan strain.\textsuperscript{11,12}

The most abundant amongst the other structural proteins, the M protein (25-30 kDa, UniProtKB - P0DTC5), is embedded in the viral membrane. It is involved in the assembly of new virions within host cells. The E protein (8-12 kDa, UniProtKB - P0DTC4) is the smallest of all the structural proteins and some of its major functions are pathogenesis, viral assembly and viral release. Both M and E proteins are structurally preserved across various genera of β-coronaviruses. The N protein (45-48 kDa, UniProtKB - P0DTC9) is recruited at the replication-transcription complex and is responsible for genome packaging by guiding the viral genome to the newly assembled replication complex. Several groups have reported high levels of IgG antibody response against the N protein in COVID-19 patients. Moreover, the N protein is a representative antigen for the T-cell response in a vaccine setting, which induces a SARS-CoV-2 specific T-cell
Figure: 1.4: Cryo-electron microscopic structure of S protein of SARS-CoV-2. Panels A, B and C show closed conformations of the S protein trimer and Panels D, E and F show partially open conformations of the S protein trimer. Image from Walls et al.\textsuperscript{13}
proliferation and cytotoxic activity. Hence, considering its high immunogenic nature, the N protein could also be a useful potential target for vaccine development.\textsuperscript{14,15}

The non-structural proteins (NSP) are encoded by the genes located at the 5’ end of the viral genome (NSP1 to NSP16). Each of these proteins performs a specific role during viral replication. There are also nine accessory proteins: ORF3a, 3d, 6, 7a, 7b, 8, 9b, 14, and 10. These play a crucial role in virulence and host interaction. They exhibit the highest degree of variability, even within very closely related coronaviruses.\textsuperscript{16,17}

Immune response to SARS-CoV-2

Once infected with SARS-CoV-2, a highly organized cellular and molecular cascade is initiated to recruit the immune system. The immune system is classified into two categories: innate and adaptive. Within a few hours of infection, the innate immune system triggers type I/III interferon expression, whereas adaptive immunity takes ~6-10 days for activation of B cells and T cells generated specifically against the invading antigen. Virus-specific immunoglobulin IgM and IgA are secreted within 5-7 days while IgG is secreted within 7-10 days after the onset of symptoms (Figure 1.5). These neutralizing antibodies, primarily against the S and N proteins of the virus, are capable of preventing binding with the ACE-2 receptor and thus are the basis for the development of vaccines against SARS-CoV-2.\textsuperscript{18}

The patients that enter Stage 3 (Fig. 1.1) of the disease usually experience a phenomenon called a cytokine storm. A cytokine storm is characterized by elevated levels of interleukin (IL)-1β, IL-6, TNF (tumor necrosis factor), interferon-γ, macrophage inflammatory protein, VEGF
(vascular endothelial growth factor), etc. and is suggested to be the reason for mortality in severe COVID-19 patients. Interestingly, IL-6, a pro-inflammatory cytokine, is one of the key mediators of the acute inflammatory response and, therefore, IL-6 inhibitors were one of the first class of drugs to be considered for COVID-19 therapy. Other clinical abnormalities like elevated levels of C-reactive protein, hypoalbuminemia and renal dysfunction are also observed in severe cases. Numerous pieces of evidence suggest that the disease severity and outcomes are more dependent on the host’s immune responses and pre-existing conditions rather than on the genetic variations of the virus.
Detection of SARS-CoV-2

While the virus is rapidly circulating as well as mutating, enormous efforts have been made to develop speedy but sensitive diagnostic tests for COVID-19 in order to curb the disease. Broadly, there are two major categories of diagnostic tests to detect SARS-CoV-2: i) detection of the viral RNA genome and ii) detection of viral proteins and antibody response against the virus. The current gold standard test for COVID-19 belongs to the first type of diagnostic test which detects the presence of SARS-CoV-2 RNA using reverse transcription quantitative polymerase chain reaction (RT-qPCR), targeting one or more viral genes. While the analytical sensitivity of RT-qPCR can be quite good (with higher limits of detection for viral RNA), surprisingly in the clinical settings, the sensitivity of RT-qPCR testing could be in the range of 50-70%. Nevertheless, factors such as viral load at a given anatomical location, type of specimen and sample collection during the time course of infection play a crucial role in determining the success of the RT-qPCR technique. Also, the diagnostic efficacy between different RT-qPCR test kits may vary significantly depending on the probes and primers used. Moreover, false-negative results can occur due to a wide range of clinical laboratory errors, RNA degradation and infections with variants that contain mutations in the primer/probe binding sites.22 The most common types of specimens for RT-qPCR are nasopharyngeal swabs, oropharyngeal swabs and saliva.

On the other hand, various antigen/antibody detection tests are also being simultaneously developed for COVID-19, some of which provide results on the spot. Point-of-care lateral flow devices and microfluidic devices are some examples. The main advantages of using such methods are miniaturization, high-speed detection, small sample volume, and portability. For instance, the Abbott Diagnostic ID NOW COVID-19 Test delivers results in 13 minutes from a throat or nasal
swab and provides an analytical sensitivity of 125 copies/mL. Another route, serological testing involves measuring the host immune response against the invading pathogens by detecting antigen-specific antibodies. However, as the immune system requires time to build up, this strategy has a major limitation of use in the early phase of infection. Generally, antibody testing has been implemented primarily for healthcare providers in order to evaluate that population’s exposure to SARS-CoV-2.

Although numerous platforms are available for COVID-19 testing, several aspects of the assays make the comparison of test results difficult, including agreeable limits of sensitivity and specificity, appropriate use of standards and controls, and sample handling. This furthermore increases the challenges of making informed healthcare decisions and ultimately controlling the transmission of SARS-CoV-2.

Saliva as a diagnostic tool

Saliva is a complex exocrine secretion produced by salivary glands. Though its main functions are to aid in digestion and protection of the oral cavity, saliva shows immense potential as a diagnostic tool. Gradually, saliva testing is gaining popularity due to rapid and non-invasive sample collection, cost-effectiveness and patient acceptance. Saliva sampling is suitable for all ages and can be performed by non-healthcare professionals and by self-sampling. This reduces the risk of cross-infections and provides a possibility of repeatable collection. Studies have reported the use of saliva for the diagnosis of oral cancer, autoimmune disorders, and viral diseases from Zika, Ebola and SARS-CoV. Pertinent to COVID-19, the ACE-2 receptor is abundantly
expressed in oral epithelial cells and the salivary glands are not only an infection target but also a reservoir for SARS-CoV-2 (Figure 1.6). Additionally, salivary viral loads are reported to be highest at the beginning of the course of infection, thus offering scope for early diagnosis of COVID-19. Saliva also contains growth factors, cytokines and immunoglobulins such as IgA, IgG and IgM, providing multiple biomarkers that can aid in understanding the clinical evolution of the disease.27

The study reported by Azzi et al. provides evidence of saliva being a promising tool in COVID-19 diagnosis by comparing saliva samples with nasopharyngeal swabs of people severely infected with COVID-19.26 Aita et al. reported the use of saliva for monitoring local adaptive immunity by measuring IgA anti-spike SARS-CoV-2 protein.28 Lately, the Food and Drug Administration (FDA) also issued approval for saliva-based testing under emergency use authorization (EUA) to Rutgers’ RUCDR Infinite Biologics and collaborators,29 the Yale School of Public Health30 and the University of Illinois Urbana Champaign.31

Protein Testing for COVID-19

While it is true that molecular diagnosis of COVID-19 primarily depends on the detection of viral RNA, detection of viral proteins and/or detection of host immune proteins in response to the infection can also be employed for the same purposes. A biomolecule such as RNA is highly prone to degradation by ribonucleases which can occur due to improper handling, storage and treatment, and thus demands trained personnel for its analysis. RNA extraction is overall expensive, plus inadequate purification of RNA and inefficient removal of sample
Figure 1.6: Transmission of the SARS-CoV-2 virus and potential diagnostic value of saliva. Inset: a mechanism for entry, replication and release of SARS-CoV-2 from salivary glands. Image from Sapkota et al. 32
matrix/impurities may result in false-negative results. On the other hand, proteins are more stable than RNA and in fact present in higher amounts in viruses. Also, the chances of producing false-positive results in the case of proteins are substantially reduced as they can’t undergo amplification. This improves the overall specificity of a test probing for protein as compared to one testing for viral RNA.  

In specimen types like serum or saliva, proteins such as enzymes, cytokines, immunoglobulins and pathogen proteins can be collectively characterized and quantitated using mass spectrometry and other proteomic techniques. In this way, several proteins can be detected in a single analysis which could lead to establishing multiple biomarkers for a given condition. The field of proteomics was also explored in previous epidemics and pandemics such as swine flu, dengue, chikungunya, zika, Ebola, SARS-CoV and MERS-CoV. A number of studies have reported cytokine storms observed in COVID-19 patients which results in elevated levels of IL-6, IL-2, IL-7, granulocyte-colony stimulating factor, interferon-inducible protein 10, monocyte chemoattractant protein 1, macrophage inflammatory protein 1-α, and tumor necrosis factor-α. Monitoring the levels of interleukins could enhance our comprehension of the pathogenesis of diseases and could eventually help in developing novel therapeutics. Besides, antibodies generated in response to infection may provide a larger window of time for its detection as compared to nucleic acids. Various technologies such as ELISA (enzyme-linked immunosorbent assay), lateral flow devices, magnetic biosensors and separators are also emerging for the detection of viral proteins. One drawback of these techniques, however, is that the signal from highly abundant host proteins can mask the interpretable signals from the viral proteins. Thus, an extremely sensitive technique such as mass spectrometry (MS) coupled with informatics can
enable the prediction of protein patterns from large data sets and ultimately help in biomarker discovery.

Mass spectrometry in clinical diagnostics

Mass spectrometry is an analytical technique that measures the mass-to-charge ratio (m/z) of one or more analytes in a given sample. Here, m is molecular mass in Daltons while z is the integer amount of charge present on the ion. There are three primary components of any mass spectrometer: an ionization source, a mass analyzer and a detector. The raw data obtained from mass spectrometry could provide molecular weight, the abundance of the target analyte and structural information. Mass spectrometry is utilized in an extensive range of fields such as proteomics, metabolomics, forensics, environmental sciences and clinical diagnostics. Particularly after the development of ionization sources such as electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI) and matrix-assisted laser desorption/ionization (MALDI), analysis of large biomolecules like nucleic acids, peptides and proteins became possible. Especially in the last two decades, mass spectrometry-based assays have been implemented in clinical diagnostic settings to achieve multiplexing and high-throughput capacities coupled with greater accuracies compared to traditional methods (Figure 1.7).

MS has the ability to provide abundant pathological information in a single analysis, thereby minimizing technical and supply load on public health resources. While the initial investment for MS equipment is high, the overall assay development, validation and instrument maintenance has proven to be cost-effective. This certainly explains the rising demand for the adoption of MS to save time and labor expense. The two most commonly used MS techniques in
Figure 1.7: Approximate timeline of some important milestones in the evolution of clinical MS. Image from Jannetto et al.40

Clinical laboratories are MALDI-ToF (matrix-assisted laser desorption/ionization-time of flight) and LC-MS (liquid chromatography-mass spectrometry). Figure 1.8 displays the basic workflow followed in clinical labs for the detection of SARS-CoV-2 using MALDI-MS or LC-MS.

The FDA has approved numerous MS-based diagnostic assays in the past ten years for the identification of microbes, newborn screening, quantitative analysis of drugs in serum and vitamin D assays. For instance, Bruker’s MALDI Biotyper CA, a benchtop MALDI-ToF instrument, revolutionized the identification of pathogenic microbes with its ability to accurately characterize more than 400 species of bacteria and yeasts.39 Here, the highly abundant ribosomal proteins present in the microbes along with some species-specific proteins are the contributors to the peaks generated for a given test isolate. The obtained mass spectrum is compared to the reference
Figure 1.8: Schematic overview of the workflow for MS-based proteomics for clinical diagnosis of COVID-19. Image from Mahmud et al.\textsuperscript{35}

spectrum or deconvoluted spectrum to assign a score. Depending on the scores obtained, the organisms can be identified at the family, genus or species level using various algorithms.\textsuperscript{41}

Likewise, certain LC-MS-based assays can successfully screen for more than 50 different metabolic disorders in a single test even before symptoms arise in patients.\textsuperscript{39}

Focusing more on proteins, identifying proteome patterns associated with disease development has shown potential in untargeted biomarker studies. Proteomic analysis in MS can be performed in two ways: top-down proteomics and bottom-up proteomics (the latter often referred to as shotgun proteomics). Top-down proteomics involves the analysis of intact proteins from complex biological matrices. In contrast, bottom-up proteomics involves enzymatic cleavage
of whole protein with a protease (commonly trypsin) to generate a mixture of peptides. The top-down method gives a global overview and the details of the individual proteoforms can be elucidated through scrutiny of the multiple layers of acquired information. Conversely, the bottom-up method allows distinguishing the proteoforms by rebuilding from pieces of information. However, in this method, protein structure characteristics like post-translational modifications are sometimes lost.\textsuperscript{38}

**MALDI-ToF**

Matrix-assisted laser desorption/ionization-time of flight (MALDI-ToF), a soft ionization technique, is commonly used in systems biology and clinical studies to search for new potential biomarkers associated with a disease. The Nobel Prize in Chemistry in 2002 was awarded in part to Koichi Tanaka who developed the soft ionization technique. This invention has provided a unique prospect for the analysis of proteins and other polymers exceeding 200 kDa with improved sensitivity of several orders of magnitude. MALDI analysis is initiated by mixing a sample solution with a matrix material and depositing the mixture on a specially designed MALDI sample target. This spot is irradiated with a laser beam of high irradiance power to simultaneously desorb and ionize the sample and matrix molecules into the gas phase (Figure 1.9). MALDI is optimally combined with a ToF (time of flight) mass analyzer due to its theoretical ability to analyze unlimited mass range and to acquire an entire spectrum from a single laser pulse event. ToF, currently in high demand, is one of the simplest mass-analyzing devices which converts a time spectrum into a mass spectrum.\textsuperscript{42}
An ionization technique such as MALDI has significantly revolutionized the world of large biomolecules and biopolymers analysis. Compared to other ionization techniques, this technique shows high tolerance towards various organic and inorganic contaminants such as complex biological matrices, salts, detergents, buffers, etc. Hence, MALDI is also known to require minimal sample processing and thereby making it a cost and time-efficient technique. MALDI is initiated by mixing the sample solution and the matrix material and depositing the mixture on a target plate. The matrix is used in large molar excess as compared to the sample. After complete evaporation
of the solvent, the dried crystallized mixture is irradiated with a laser beam of high irradiance power and short pulse widths. This results in simultaneous desorption and ionization of the sample+matrix molecules into a gas phase.\textsuperscript{42,44}

The matrix serves as a key ingredient to the success of MALDI. The essential characteristics of a MALDI matrix are: (1) strong absorption of radiation at the laser wavelength; (2) good mixing and solvent compatibilities with the target analyte to form well-defined microcrystals; (3) a low sublimation temperature that allows the formation of an instantaneous high-pressure plume of the matrix+sample clusters during the laser pulse duration; (4) participation in the mechanistic reactions which allows the target molecules to protonate or deprotonate with high efficiency. There are four main groups of matrices that are used in MALDI analysis depending on the nature of the sample analyte. The first group of matrix compounds known as solid organic matrices include the most commonly used matrices that have a light-absorbing aromatic ring(s). A few examples include compounds such as sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid), CHCA (α-Cyano-4-hydroxycinnamic acid), DHB (2,5-Dihydroxybenzoic acid), cinnamic acid, picolinic acid. Figure 1.10 shows the structures of common solid organic matrices used in MALDI applications.\textsuperscript{42}

The second group of matrices is known as liquid organic matrices such as 2-nitrophenyloctyl ether and 3-nitrobenzyl alcohol that are used to overcome the inhomogeneous sample preparation problem. The third group of matrices is the ionic liquids that are formed by an equimolar mixture of a traditional solid organic matrix with an organic base (e.g., butylamine). These cation–anion pairs offer homogeneous sample preparation resulting in improved reproducibility and increased signal intensities. The last group of matrices include inorganic
Figure 1.10: Structures of common organic MALDI matrices. Image from Dass et al.⁴²
materials such as metals (Cu, Co, Al, Mn, W, etc.), metal oxides, and graphite carbon nanotubes, fullerenes, textured silica, etc. These materials are usually dispersed in a suitable nonvolatile liquid such as glycerol or liquid paraffin and they behave as a suitable trap for analyte molecules and a reservoir of laser energy.\textsuperscript{42}

While there have been numerous applications of MALDI across various fields of science, the essential ionization step has not been clearly understood. Multiple diverse views have been expressed on this topic in the mass spectrometry community. However, there are two predominantly discussed theories/models of MALDI ionization mechanisms. The older model known as Coupled Physical and Chemical Dynamics model assumes that there are neutral analyte molecules present in the expanding plume, irrespective of whether the analytes were mixed with the matrix crystals as neutral species or were quantitatively neutralized by their counterions upon cluster dissociation in the case of pre-charged incorporated analyte molecules. Due to photoionization (laser), the matrix undergoes intermolecular matrix reactions resulting in protonated as well as deprotonated matrix ions. Next, there occurs a charge transfer to the neutral analyte molecules in the plume leading to the generation of positive or negative analyte ions.\textsuperscript{45,46}

The more recent model is called the “lucky survivor” model which can be understood in two parts: (1) The sample analytes have a charge in the solution itself which is retained in the mixture with the matrix. (2) These charged sample ions happen to survive neutralization by retaining their charge. The model also states that the ablated clusters of the matrix+analyte could have pre-charged counterions as well, resulting in a total net zero charge and hence not detected by the detector. This assumption is based on the observation that pH-indicator molecules retain their color and charge state upon crystal incorporation for acidic, neutral, or basic matrices. After
desorption, the ablated plume expands where the matrix+analyte cluster loses the neutral matrix and solvent molecules, and more importantly, the counterions undergo proton-transfer neutralization via interaction with analyte sites. This results in most of the charges being neutralized except for a remaining excess matrix ion charge. It is assumed that some of these analyte-containing clusters statistically carry one or more excess protonated or deprotonated matrix ions and subsequently these excess charges are thought to prevent quantitative analyte ion neutralization, leading to the generation of protonated and deprotonated analyte ions. These single-charged analyte ions are hence called “lucky survivors” of the neutralization process. This model is successful in explaining the observations of mostly single-charged ions and for UV as well as IR-MALDI. Moreover, it can be applied to the ionization process of positive and negative ions.45,46

As the irradiating laser beam is pulsed, MALDI is optimally combined with a ToF mass analyzer and hence MALDI-ToF has become a recognized acronym. ToF is the simplest mass analyzer that offers unlimited mass range and acquisition of the entire mass spectrum from a single laser pulse event. This mass analyzer consists of a long field-free tube called the flight tube in which ions are separated, after their initial acceleration by an electric field, on the basis of their mass differences. The ions are accelerated towards the flight tube by a difference of potential applied between an electrode and the extraction grid. In a linear ToF, all the ions (which have the same charge) gain equal kinetic energy and are accelerated toward the mass analyzer. Figure 1.11 gives a general schematic of a linear ToF mass analyzer.42

After detection, the m/z of an ion can be determined by deriving an equation.44 An ion can be considered as having a mass of \( m \) and a total charge of \( q \) which is the number of charges \( z \), multiplied by the charge of an electron \( e \). The ion was accelerated by a potential \( V_s \) and hence its
kinetic energy, $E_k$, is given by:

$$E_k = \frac{mv^2}{2} = qV_s = zeV_s \quad \text{(i)}$$

As all the ions are accelerated with the same kinetic energy, their path in a straight line in the flight tube is determined by their velocity, $v$. Equation (i) can be rearranged as equation (ii) to isolate the ion’s velocity as a function of its mass and charge.

$$v = \left(\frac{2zeV_s}{m}\right)^{1/2} \quad \text{(ii)}$$

The distance of the flight tube is considered as $L$ where it takes time $t$ for an ion to traverse the length of the flight tube after initial acceleration and reach the detector at the velocity of $v$.

$$t = \frac{L}{u} \quad \text{(iii)}$$
By rearranging equations (ii) and (iii), we can derive:

\[ t^2 = \frac{m}{z} \left( \frac{L^2}{2eV_s} \right) \quad \text{(iv)} \]

Equation (iv) can be rearranged to isolate the m/z in equation (v).

\[ \frac{m}{z} = \frac{t^2 eV_s}{L^2} \quad \text{(v)} \]

Therefore, the m/z of an ion is proportional to \( t^2 \), the time it takes for an ion to travel through the drift tube to the detector, squared. In other words, the lower the mass of an ion, the faster it will reach the detector.\(^{44}\)

ToF mass analyzers generally have a high ion transmission efficiency, which offers high sensitivity. Various proteins such as cytochrome C, ribonuclease A, lysozyme and myoglobin have been detected in 100 attomole amounts. Moreover, the speed of analysis of these analyzers is super-fast, providing a broad mass range in microseconds. However, there could be some weak ions in each individual spectrum which makes it difficult to provide the precision of mass and its abundance. Hence, a recorded spectrum is generally an addition of a number of individual spectra.\(^{44}\)

Another interesting characteristic of the ToF analyzers is its easy mass calibration with only two reference points. Generally, in a mass analyzer, the calibration equation relates and converts a physical property that is measured to a mass value. In the case of ToF, the physical property that is measured during analysis is the flight time of the ions. Using the equations mentioned above, we have established that the flight time of an ion is related to its mass and can be also given by:

\[ \left( \frac{m}{z} \right)^{1/2} = \left( \frac{\sqrt{2eV_s}}{L} \right) t \quad \text{(vi)} \]
For a given mass spectrometer, the distance (L) and accelerating potential (Vs) are constant, and hence the parentheses on the right of equation (vi) can be replaced with a constant \( A \). As the relationship between \( m^{1/2} \) and \( t \) is linear, a constant \( B \) is added to produce a simple equation for a straight line. This constant \( B \) allows correction of the measured time zero that may not correspond exactly with the true time zero.\(^{44}\)

\[
\left( \frac{m}{z} \right)^{1/2} = At + B \quad \text{...(vii)}
\]

Hence, the conversion of flight times to mass (i.e., the time spectrum to the mass spectrum) supposes a preliminary calibration with two known standard molecules. Using the standards’ known m/z ratios and their measured flight times, this equation is solved for the two calibration constants \( A \) and \( B \). As long as the points are not too close together, a simple two-point calibration is usually accurate to proceed with analysis. Also, two types of calibrations can be performed here: external and internal calibration. External calibration involves determining the calibration constants, \( A \) and \( B \), from two standards that do not include target analyte molecules in an experiment. Whereas internal calibration is a method in which the flight times of the two standards and unknown/analyte ions are measured in the same spectrum, providing the best possible match of experimental conditions for the three species involved. The highest level of mass accuracy is usually attained through internal calibration.\(^{44}\)

One major drawback of the linear ToF is low mass resolution compared to other mass analyzers. The resolution issue can be overcome by employing a different kind of detector called the reflectron ToF. A general layout of a reflectron ToF is illustrated in Figure 1.12. The ions with the same m/z may enter the ToF with a slight kinetic energy dispersion. The purpose of the
reflectron is to correct these energy differences. This is achieved by electrostatic mirrors/lenses that consist of grids and a series of ring electrodes, each with a progressively increasing repelling potential. Ions of the same m/z with higher kinetic energy will penetrate further into the reflectron compared to those with lower kinetic energy. Therefore, ions with higher kinetic energy will travel to the reflectron faster, spending more time in the reflectron before being deflected back into the flight tube. Consequently, they will reach the detector at the same time as the ions with a lower kinetic energy that traveled slower as they did not penetrate as far into the reflectron. Overall, this energy-correcting device improves the mass resolution of the ToF but imposes a limitation on the mass range.\(^{42,44}\)

Another way of improving the resolution of linear ToF mass analyzers is by incorporating delayed pulse extraction in the form of time-lag focusing. This technique uses dual-stage ion-extraction optics with two distinct sections: extraction and acceleration regions (Figure 1.13). At the time of the ionization pulse, no potential is applied to the extraction region and hence the ions drift in this region in a field-free environment with their initial velocities. After a short delay of a few hundred nanoseconds, the ions are extracted from the source by applying acceleration potential. In this delay period, the slow-moving ions lag behind the fast-moving ions. The ions farther from the extraction grid are subjected to a higher electrical potential as compared to the ions closer to the extraction grid. Due to this difference in the accelerating field, the ions that were lagged are accelerated to a higher velocity. By setting a proper delay time and amplitude of the extraction pulse, all ions of a particular m/z but of varying initial kinetic energy are made to reach the detector at the same time. Such a setup is also useful to enhance knowledge of structure-specific fragment ions. However, in a conventional linear ToF, this feature is absent resulting in a rapid
Figure 1.12: An illustration of a reflectron ToF mass analyzer. The spheres represent two ions with the same m/z, where the brown sphere has a slightly higher kinetic energy than the green sphere. Irrespective of their energy differences, they reach the reflectron detector at the same time. Image created with BioRender.com.

extraction of molecular ions from the ion source before they have a chance to fragment.\textsuperscript{42,45}

MALDI-ToF profile analysis is relatively straightforward as ions are mostly singly charged [M+H]\textsuperscript{+} with a lack of fragmentation. It also offers benefits like low consumption of the sample with high throughput analysis. A basic workflow for MALDI-based profiling includes sample preparation, mass spectrometry acquisition, data preprocessing, statistical analysis and finally identification of peaks using sophisticated data mining techniques (Figure 1.14). MALDI-ToF data preprocessing itself can include multiple steps like data import, quality control, spectrum normalization, peak smoothing, baseline correction, model generation and validation.\textsuperscript{47} Hence, nowadays most MALDI-ToF studies utilize the capabilities of machine learning (ML) for processing and analyzing their data to achieve unbiased analysis.
Figure 1.13: A diagrammatic representation of delayed pulse extraction using the principle of time-lag focusing. Image from Dass et al. 42

Figure 1.14: General workflow in the profiling approach using MALDI-TOF mass spectrometry. Image from Hajduk et al. 47
Immunoglobulins, hemoglobin, insulin-like growth factor-1, C-reactive protein, serum amyloid A and fecal calprotectin cystatin C are only a few examples of the proteins that have been studied qualitatively or quantitatively using MALDI-ToF. However, one of the major limitations of MALDI-ToF is that it is confined to the detection of the most highly abundant protein(s) in a given biological sample; there are significant levels of ion suppression where an ion can suppress the peak signal of other ions in the sample.

MALDI-ToF for COVID-19

Researchers started exploiting the potential of the MALDI-ToF technique for COVID-19 research almost immediately after the pandemic was declared. Although there is currently only one assay by Ethos Laboratories in Newport, KY which received approval from the FDA under EUA (emergency use authorization), various groups around the world are making efforts to adopt MALDI-based testing during the pandemic. Several remarkable methodologies have been published recently which show significantly high sensitivities and specificities equivalent to that of RT-qPCR testing. Table 1.1 lists some important studies utilizing MALDI-MS along with the m/z ranges screened and specimen types. In 2017, Xiu et al. reported the use of MALDI-ToF MS for the screening of human coronaviruses (HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU1), SARS-CoV and MERS-CoV. By employing a similar protocol, Rybicka et al. reported the detection of SARS-CoV-2 using oral and nasal samples. In another study, Zhao et al. demonstrated the application of MALDI-ToF for the detection of nine mutation types with 100% specificity. However, all three above mentioned studies are essentially RT-qPCR assays that use MALDI-ToF for sequencing the end products obtained after gene amplification.
Table 1.1: Summary of MALDI-ToF based assays for detection of COVID-19 using various specimen types.

<table>
<thead>
<tr>
<th>Author</th>
<th>m/z range</th>
<th>Specimen</th>
<th>Biomolecule</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xiu et al.</td>
<td>Up to 12,000</td>
<td>Nasal and throat swabs</td>
<td>RNA</td>
<td>Accuracy= 100%</td>
</tr>
<tr>
<td>Gomila et al.</td>
<td>2,000-25,400</td>
<td>Serum</td>
<td>Peptides</td>
<td>Accuracy= 90%</td>
</tr>
<tr>
<td>Iles et al.</td>
<td>2,000-200,000</td>
<td>Saliva (Gargle)</td>
<td>Proteins</td>
<td>Reported high mass range for the first time</td>
</tr>
<tr>
<td>Dollman et al.</td>
<td>500-3,000</td>
<td>Nasal swabs</td>
<td>Peptides</td>
<td>Peptides ions for S, N and M proteins detected</td>
</tr>
<tr>
<td>Rocca et al.</td>
<td>1,960-20,000</td>
<td>Nasal swabs</td>
<td>Proteins</td>
<td>Accuracy= 68%, Sensitivity= 62%, Specificity= 72%</td>
</tr>
<tr>
<td>Nachtigall et al.</td>
<td>3,000-15,500</td>
<td>Nasal swabs</td>
<td>Not specified</td>
<td>Accuracy=93.9%</td>
</tr>
<tr>
<td>Rybicka et al.</td>
<td>Up to 12,000</td>
<td>Oral and nasal swabs</td>
<td>RNA</td>
<td>Accuracy= 90%</td>
</tr>
<tr>
<td>Lazari et al.</td>
<td>2,000–20,000</td>
<td>Plasma</td>
<td>Proteins</td>
<td>Accuracy=92%, Sensitivity= 93%, Specificity=92%</td>
</tr>
<tr>
<td>Zhao et al.</td>
<td>4,000- 9,000</td>
<td>Pharyngeal swab</td>
<td>RNA</td>
<td>Accuracy= 100%, Specificity=99.9%, 9 mutation types could be detected</td>
</tr>
<tr>
<td>Yan et al.</td>
<td>5,000-30,000</td>
<td>Serum</td>
<td>Peptides</td>
<td>Accuracy= 99%, Sensitivity= 98%, Specificity= 100%</td>
</tr>
<tr>
<td>Tran et al.</td>
<td>2,000-20,000</td>
<td>Nasal swabs</td>
<td>Not specified</td>
<td>Accuracy= 97%</td>
</tr>
<tr>
<td>Deulofeu et al.</td>
<td>5,000-20,000</td>
<td>Nasal swabs</td>
<td>Not specified</td>
<td>Accuracy, Sensitivity and Specificity&gt;90%.</td>
</tr>
<tr>
<td>Wan et al.</td>
<td>Up to 1000</td>
<td>Serum</td>
<td>Serum Metabolites</td>
<td>Accuracy= 93.4%</td>
</tr>
<tr>
<td>Costa et al.</td>
<td>2,000-20,000</td>
<td>Saliva</td>
<td>Proteins</td>
<td>Accuracy= 85.2%, Sensitivity= 85.1%, Specificity=85.3%</td>
</tr>
</tbody>
</table>
Other groups such as Nachtigall et al.\textsuperscript{57}, Tran et al.\textsuperscript{60} and Deulofeu et al.\textsuperscript{61} performed stand-alone MALDI-based assays achieving above 90\% accuracies. For these studies, portions of samples that were collected for RT-qPCR analysis were directly spotted on the MALDI target plate without any sample processing. Hence, the peaks detected in such a spectrum wouldn’t necessarily represent viral RNA or viral proteins. The identification of the species associated with observed peaks (or biomarker peaks) was not included. Hence, these studies lack any interpretation for possible reasons for differences between COVID-19 positive and COVID-19 negative spectra.

Gomila et al.\textsuperscript{53} and Yan et al.\textsuperscript{59} tested the utility of MALDI-ToF for visualizing the serum peptidome in the range of 2,000-25,400 and 5,000-30,000 m/z respectively. Gomila et al. were also able to distinguish between mild and severe COVID-19 patients based on their peptidome signature.\textsuperscript{53} Another interesting study was reported by Iles et al. that demonstrated the use of a wide m/z range (i.e., 2,000-200,000) for the first time. A water gargle (saliva) sample was used as their specimen type. Low-cost and minimal sample processing (acetone precipitation) were optimized for concentrating the proteins. The authors reported detection of host immune proteins (such as IgA and immunoglobulin light chains) along with viral proteins (S1, S2 and viral envelope proteins) in a single spectrum. Despite such a promising method, no direct comparison of MALDI-ToF MS testing with RT-qPCR detection of COVID-19 was performed.\textsuperscript{54}

**LC-MS**

Liquid chromatography-mass spectrometry is the most widely used analytical technique. High-performance liquid chromatography (HPLC) helps in the separation of analytes from a complex mixture based on differences in their affinity toward the stationary phase and mobile phase. It is
also possible to quantitate the target analyte using peak intensity and peak area. Chromatographic separation can offer good resolution but can become challenging if multiple components elute at the same time. However, mass spectrometry can offer high sensitivity to detect component ions based on their m/z values. Thus, the coupling of LC to MS (Figure 1.15) has proven to be an extraordinarily successful pair in the world of analytical chemistry. Moreover, tandem mass spectrometry (MS/MS) can induce fragmentation and provide structural information for the ions being analyzed. For LC-MS, commonly used ion sources in clinical laboratories are electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI) and atmospheric pressure photo-ionization (APPI) in combination with mass analyzers such as quadrupole, time-of-flight (ToF) and ion traps. Some of the well-known applications of LC-MS include biochemical screening for genetic disorders, therapeutic drug monitoring and toxicology, assays for vitamins and related metabolites, and detection of steroids and hormones.64

LC-MS for COVID-19 diagnosis

Similar to MALDI-ToF, LC-MS was also utilized for the detection of SARS-CoV-2 infection. However, LC-MS-based assays focus only on the host and/or viral proteomics for diagnostic purposes rather than viral RNA. One of the most notable studies was reported by Messner et al. where 27 potential biomarkers that were differentially expressed depending on the WHO severity grade of COVID-19 patients were identified. Serum samples of COVID-19 positive patients were analyzed on an ultra-high-pressure high-flow LC system coupled to quadrupole-ToF mass spectrometer. The detected biomarkers included proteins such as IL-6, inflammation modulators, pro-inflammatory factors, coagulation factors and complement factors.65
Ihling et al. reported the use of gargoyle samples for the detection of unique peptides originating from SARS-CoV-2 nucleoprotein using a nano-HPLC system coupled with an orbitrap mass analyzer. While the accuracy of this assay was reported to be 100%, the sample size was only three, whereas Cardozo et al. reported a sensitivity of 84% and specificity of 97% after analyzing 985 respiratory tract samples. Using LC-MS for the analysis of urine samples, Li et al. could differentiate mild and severe COVID-19 patients by studying urine proteome. Apart from the detection of COVID-19, LC-MS has also been employed in understanding protein-protein interactions between humans and SARS-CoV-2 to suggest druggable targets for treating COVID-19.
Machine learning

Machine learning (ML), a branch of artificial intelligence, involves computer programming to improve its own performance by learning through experience. ML algorithms have now become common for the analysis of MS spectra due to their inherent nonlinear data representation and power to process large heterogeneous datasets rapidly, leading to the possible discovery of new biological insights. Several studies have implemented artificial intelligence to recognize complex patterns that classify disease and healthy states. In general, ML consists of two components: training and testing. The first component includes supervised learning by mining complex MS data from samples of known disease status to generate statistical classifiers. The latter component cross-validates these classifiers with the training (unseen) data set. The pattern recognition approach does an exceptional job of identifying multiple biomarkers at once which could have been easily overlooked by manual interpretation. Moreover, the ability to capture slight but significant variations makes it possible to detect diseases at early stages. Recently, Xie et al. published ML methods to detect five metabolic biomarkers using LC-MS/MS for the diagnosis of early lung cancer. Previously mentioned groups such as Deulofeu et al., Tran et al., Lazari et al., Nachtigall et al. and Yan et al. successfully employed ML for evaluating sensitivities and specificities for their MALDI-ToF assays. Commonly used ML classifiers in the field of biology are random forests, genetic algorithms, support vector machines and artificial neural networks. Figure 1.16 gives an overview of ML approach in clinical settings for the discovery of biomarkers.
The wide spread of SARS-CoV-2 caused a global pandemic in December 2019. COVID-19 has led to a dramatic loss of human life and is continuing to affect individuals of all age groups. While there isn’t any direct treatment for COVID-19 as of January 2023, rapid and early diagnosis, isolation if infected, and vaccination against SARS-CoV-2 are the key practices currently used to curtail the transmission of the virus. The sensitivity and specificity of diagnostic assays as well as
the efficacies of the approved vaccines are still under investigation as several variants of the SARS-CoV-2 virus have now emerged.

In response to the ongoing pandemic, researchers are attempting to develop better approaches for accurate testing. The current gold-standard test for COVID-19 diagnosis is RT-qPCR. However, this technique has several limitations including but not limited to the time and cost factors. Testing of entire populations has also led to a worldwide shortage of RT-qPCR supplies which severely affects the diagnosis and research of other diseases as well. In the current study, we aim to establish a new platform for the diagnosis of COVID-19 by performing protein profiling of gargle/saliva samples on MALDI-ToF MS. Saliva sampling is simple and offers several benefits as a diagnostic tool. Moreover, saliva contains antibodies, cytokines and enzymes that can be detected and quantified. Thus, for respiratory diseases such as COVID-19, analysis of saliva proteome serves a dual purpose of exploring the host as well as the viral proteome.

MALDI-ToF protein profiling offers advantages such as simplicity in instrument handling and low sample consumption. The data generated on MALDI-ToF can provide abundant information in a single analysis as compared to RT-qPCR results which are binary in nature. Additionally, the sample preparation for MALDI-ToF is simple and economical, without compromising the speed of analysis.

Our primary objective was to perform protein profiling in a wide mass-to-charge range using MALDI-ToF MS in order to detect and identify key COVID-19 biomarker proteins present in the saliva. Another objective is to minimize sample preparation to achieve rapid and cost-effective testing while maintaining the quality of our analysis. Rather than a mere qualitative analysis to differentiate COVID-19 positives from COVID-19 negatives, our study attempts to
establish new biomarkers for the disease. A promising area for the initial study could be the identification and characterization of the protein peaks detected in the saliva sample using techniques such as 2-D gel electrophoresis, affinity chromatography and bottom-up proteomics on MS. An attempt was made not only to establish certain biomarkers for COVID-19 but also assign putative identities based on previous literature and various protein standards and controls analyzed. More research is required in order to confirm the identity of these biomarkers by protein sequencing.

The third objective was to deplete salivary amylase, the most abundant protein in the saliva as it can obscure the detection of low-abundance proteins by MALDI-TOF and diminish the diagnostic utility of this specimen type. We used a potato starch-based device to deplete salivary amylase from gargle samples through affinity adsorption. Here, amylase depletion led to the enhancement of signal intensities of various peaks as well as the detection of previously unobserved peaks in the MALDI-ToF spectra and more importantly, still detect COVID-19. As biomarkers are usually present in low abundance, they could plausibly be unmasked and detected by utilizing such a technique.

SARS-CoV-2 is constantly evolving, and the majority of mutations have been reported in the S1 fragment of SARS-CoV-2 S protein. The current study hypothesizes that there could be visible mass shifts in this range of 78,000-82,000 m/z on MALDI-ToF when comparing the saliva/gargle samples of individuals infected with different SARS-CoV-2 variants. Such a method will be able to detect the disease as well as the viral subtypes/mutations in a single spectrum. In due course, we intend to achieve a rapid diagnosis of COVID-19 with high accuracy, sensitivity and specificity. The findings of this work could also be extended for the detection of other viruses
that are currently present or that could emerge in the near future. An analytical platform like MALDI-ToF provides time and cost-effective methods with highly accurate results that could easily identify viruses thereby probably preventing a future pandemic by early and rapid detection.
CHAPTER 2

MATERIALS AND METHODS

Materials

All the chemicals used in this dissertation work were of the highest grade available or as indicated. Plastic syringes (10 mL), 8% Bis-Tris Plus acrylamide gels (Invitrogen), NuPAGE MES SDS running buffer (Novex) and SeeBlue Plus2 prestained molecular weight protein ladder (Invitrogen) were purchased from Thermo Scientific (Rockwood). Methanol (ACS grade) and glacial acetic acid (ACS grade) and 30-mL disposable polypropylene beakers were purchased from Fischer Scientific (Waltham, MA). Polyethersulfone membrane filters (0.45 μm) and 50-mL centrifuge tubes were purchased from Celltreat scientific products (Pepperell, MA). Acetone, trifluoroacetic acid, dithiothreitol (DTT), potato starch, iodine, potassium iodide, ammonium bicarbonate, tris(2-carboxyethyl)phosphine (TCEP), octyl ß-D-glucopyranoside, sinapinic acid, H₂O (LC-MS grade; OmniSolv), acetonitrile (LC-MS grade; OmniSolv), fetal bovine serum, human serum antibodies IgA (I4036), IgG (I4506) and IgM (I8260), human salivary α-amylase (A1031), human serum albumin (SRP6182) and ProteoMass apomyoglobin MALDI-MS Standard (A8971) were purchased from Sigma-Aldrich (St. Louis, MO). SDS-PAGE gels were stained using
PhastGel Blue R (Amersham Biosciences, Uppsala Sweden). Milli-Q water used was generated using a Millipore Milli-Q Plus Purepak 2 water purification system (Bellerica, MA, USA) with a resistivity of 18.2 MΩ cm at 25°C. Bottled spring water was purchased from a local grocery store and the brand did not affect the study. The negative control, consisting of pooled human saliva (pre-COVID-19) collected before November 2019 was purchased from Lee Biosolutions (Maryland Heights, MO). The positive control (heat-inactivated SARS-CoV-2) and SARS-CoV-2 full-length S protein of the Wuhan variant (NR-53589), Spike D614G Variant (NR-55343), Delta Variant (NR-55614), B.1.1.529 lineage (NR-56447) and BA.2 lineage (NR-56517) of Omicron variants were obtained from BEI Resources (Manassas, VA). Influenza A H1N1 (N19810), MERS-CoV (NR-50549) and common human coronaviruses such as 229E (NR-52726), OC43 (NR-52725) and NL63 (NR-470) were also obtained from BEI Resources (Manassas, VA). The viral envelope membrane dissolution and protein solubilization LBSD-X buffer was gifted by MAPSciences (Bedford, UK).

**Instrumentation**

In this dissertation work, samples were spotted on a stainless steel MALDI-MS sample plate (Shimadzu, Kyoto, Japan) and the MALDI-ToF analyses were performed using an AXIMA Performance MALDI-ToF MS (Shimadzu Kratos Analytical, Manchester, UK). The AXIMA performance is equipped with three mass detectors (linear, reflectron and CovalX) and capable of detecting a mass range from 1-2,000,000 Da for analysis of various types of biomolecules. In the current study, linear detection was predominately used, which offered a mass range of 1-500,000 Da with a sensitivity of 250 fmol for BSA and a resolution of >5,000 for ACTH (2,456 Da). It is
equipped with a 50-Hz nitrogen (UV) laser and a high energy collision cell for performing tandem MS. Figure 2.1 shows a schematic representation of an AXIMA Performance MALDI-ToF.

Methods

**Human gargle sample collection and storage**

This study was approved by the Institutional Review Board and Institutional Biosafety Committee of Northern Illinois University (NIU) (August 12, 2020, and revised on July 12, 2021, and August 11, 2022) as well as the Institutional Review Board of the University of Illinois at Chicago (February 11, 2021). Informed consent was obtained with the signature of the volunteers. Personal identification was not associated with any sample and collected information was limited to demographics, symptoms, and RT-qPCR results. Sample handling and processing were conducted under a Class II biosafety cabinet. All methods were performed in accordance with the relevant guidelines and regulations.

Samples were collected at a drive-thru gateway testing program for student athletes organized by NIU Athletics at the Yordon Center in DeKalb, Illinois (August and September 2020) and various drive-thru testing sites conducted by the Illinois Department of Public Health (IDPH) at DeKalb, Aurora and Rockford (November 2020 – July 2022). The subjects initially received an NP swab sample collection that was subjected to RT-qPCR testing for detecting SARS-CoV-2. The RT-qPCR results were obtained from the University of Illinois Hospital laboratory in Chicago and IDPH testing laboratories depending on the testing site. The RT-qPCR results were reported as as detected, not detected, or inconclusive. At the time of NP swab collection, consenting
Figure 2.1: An illustration of Shimazu AXIMA Performance hardware. Image from Shimadzu Scientific Instruments.
volunteers were asked to gargle 10 mL of bottled spring water for 30 s and then deposit the gargle into a 50-mL conical centrifuge tube. Saliva samples were also obtained from the Department of Pathology University of Illinois, Chicago. Saliva was collected by having the subject drool about 2 ml of saliva into 50-mL polypropylene tubes. Samples were stored at -20°C until processing and analysis.

Sample preparation for human gargle samples

The gargle samples were prepared and analyzed following the method reported by Iles et al. with minor modifications. Gargle samples were thawed and transferred to a 30-mL disposable polypropylene beaker. Approximately 5 mL of each sample were filtered through a 0.45-µm polyethersulfone membrane filter with the filtrate being collected in the original 50-mL tube. Next, acetone precipitation was conducted on the filtrate by adding 5 mL of chilled acetone to each tube. Samples were centrifuged in a Beckman Coulter Avanti J-E series centrifuge with a JA-20 rotor at 16,000 × g for 30 min at 4°C. The supernatant was discarded, the rim of the tube patted dry, and the pellet resuspended using 100 µL of 1 M DTT in LBSD-X buffer (proprietary) or viral reconstitution buffer reported by Dollman et al. (with some modifications) consisting of 50 mM ammonium bicarbonate, 10% acetonitrile, 50 mM TCEP, and 5 mM octyl β-D-glucopyranoside at pH 7.5. The LBSD-X buffer was prepared fresh daily by adding appropriate amounts of DTT solution before each analysis and the viral reconstitution buffer was made every 30 days. Finally, to recover as much pelleted material as possible, the 100 µL reconstitution buffer was washed down the sides of the tube multiple times thoroughly. Upon thorough reconstitution, the samples were gently vortexed and incubated at room temperature for 15 min.
Sample preparation for human saliva samples

Five hundred μL of the saliva sample was diluted with 2.5 mL of LC-MS grade H₂O and was filtered through a 0.45-μm polyethersulfone membrane filter. Next, 6 mL of chilled acetone was added to the sample followed by overnight incubation at 4°C. The samples were centrifuged in a Beckman Coulter Avanti J-E series centrifuge with a JA-20 rotor at 16,000 × g for 30 min at 4°C. The supernatant was discarded, the rim of the tube patted dry, and the pellet resuspended in 100 μL of viral disruption buffer.

Sample preparation for human proteins standards/controls

Human serum antibodies IgA, IgG and IgM were reduced with 1 M DTT for 10 min to a final concentration of 3 pmol/μL of each antibody on the plate. Human salivary α-amylase was prepared at 100 pmol/μL and human serum albumin was prepared at 200 pmol/μL in LC-MS grade H₂O. Preparation of positive and negative controls for MALDI-ToF analysis followed closely to that of gargle samples. The negative control was prepared by spiking 500 μL of the thawed stock into 10 mL of water in a 50 mL tube. From here, the control was filtered and processed following the gargle sample procedure. The positive control was a heat-inactivated, clarified, and diluted cell lysate and supernatant from Vero E6 cells infected with SARS-CoV-2. This sample (225 μL) was treated with 4X v/v of chilled acetone (900 μL) and incubated overnight at -20°C followed by centrifugation at 10,000 × g for 15 min at 4°C. The pellet was completely dissolved in 25 μL of the reconstitution buffer. The MERS-CoV, H1N1 and the common corona viruses (229E, OC43 and NL63) were also treated similar to the positive control.
**MALDI-ToF sample spotting**

For the assay, a sandwich method of matrix-sample-matrix spotting was employed. Sinapinic acid was used as the matrix and consisted of 20 mg/mL in a 50:50 LC-MS grade H₂O to acetonitrile solution containing 0.1% trifluoroacetic acid. First, 1 μL of the sinapinic acid matrix was spotted in three wells of a 384 well stainless steel MALDI-MS sample plate and allowed to air dry. Then, 1 μL the of sample, control, or standard was spotted in each well, immediately followed by 1 μL of sinapinic acid matrix. The matrix was prepared fresh every 7 days and stored at 4ºC between analyses.

**MALDI-ToF instrument parameters**

The spectral acquisition was performed using a Shimadzu AXIMA Performance MALDI-ToF MS equipped with a nitrogen laser set at 337.1 nm with a pulse width of 3 ns and a maximum repetition rate of 50 Hz. The AXIMA Performance mass spectrometer was operated with the Shimadzu Biotech Launchpad Software (version 2.9.4) and was run in positive-ion linear detection mode. The laser power and repetition rate were set at 100 μJ/pulse and 50 Hz, respectively. Spectra were acquired by summing 5,000 spectra (250 profiles by 20 shots) in a range of 2,000–200,000 m/z per sample by shooting in a raster pattern over the target well. The ion gate was set to blank values below 1,500 m/z. Pulsed extraction was set to 50,000 m/z.
MALDI-ToF calibration

The instrument was calibrated daily using the \([\text{M+H}]^+\) and \([\text{M+2H}]^{2+}\) peaks of ProteoMass Apomyoglobin MALDI-MS Standard which was prepared at 100 pmol/μL in LC-MS grade H₂O and spotted as described above. Signal intensities of the calibrant were recorded throughout the entire analysis to track inter-day instrument performance. Calibration was accepted if the mass deviation was less than 500 mDa.

MALDI-ToF data analysis by Area under the curve (AUC)

The Shimadzu Biotech Launchpad was used to export a text file for each gargle sample that included the spectrum of mass-to-charge values ranging from 2,000 m/z to 200,000 m/z alongside respective ion count intensities. In preprocessing, seven subranges (shown in Table 2.1) were identified where m/z values were presumably indicative of host immune proteins or viral proteins. The values in each of the seven subranges indicated the presence of similar protein masses, hence ion counts in each subrange were coalesced together through the integration of points to calculate the AUC which produced seven features for each data sample. AUC values were computed by leveraging the composite Simpson’s rule for each spectral range. These calculations were carried out by our collaborators Venkata Devesh Reddy Seethi, Dr. Hamed Alhoori and Dr. Pratool Bharti at the Department of Computer Science, Northern Illinois University, Dekalb, US.
Table 2.1: Potential protein biomarker ranges identified by MALDI-ToF from gargoyle samples.

<table>
<thead>
<tr>
<th>Index</th>
<th>Range (m/z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11,140–11,160</td>
</tr>
<tr>
<td>2</td>
<td>23,550–23,800</td>
</tr>
<tr>
<td>3</td>
<td>27,900–29,400</td>
</tr>
<tr>
<td>4</td>
<td>55,500–59,000</td>
</tr>
<tr>
<td>5</td>
<td>66,400–68,100</td>
</tr>
<tr>
<td>6</td>
<td>78,600–80,500</td>
</tr>
<tr>
<td>7</td>
<td>111,500–115,500</td>
</tr>
</tbody>
</table>

**Receiver Operating Characteristic (ROC) analysis**

ROC curve analysis is a commonly used technique in clinical studies to determine sensitivities and specificities at different threshold values of an analyte. In a ROC curve, the true positive rate (sensitivity) on the Y-axis is plotted against the false positive rate (100-specificity) on the X-axis. The calculated AUC values for each biomarker range were used for ROC curve analysis to determine the sensitivities and specificities. This was computed by our collaborators Joshua Bland and Dr. Shrihari S. Kadkol at the University of Illinois, Chicago, US using the Medcalc software (Belgium).
Sample preparation for gargle samples using an amylase removing device

The collected gargle samples were thawed and transferred to a 30-mL disposable polypropylene beaker. Samples were filtered through a 0.45-µm polyethersulfone membrane filter and transferred into two 50-mL tubes (approximately 5 mL of gargle sample in each tube). The filtered gargle sample from one of the tubes was hand-pressed through an amylase depleting device, with modifications to a previously described apparatus. The device comprised 2 g of potato starch in a 10-mL plastic syringe with a 0.45-µm syringe filter fixed at the tip (Figure 2.2). Twenty mL of LC-MS grade H₂O was hand-pressed through the setup to moisten the substrate and remove any water-soluble residues. To optimize the amount of water required to wash off all water-soluble starch residues, we passed 5 mL of LC-MS H₂O (30 mL total) through the amylase removing device and collected the respective filtrates. An iodine reagent (0.15 M I₂ + 0.3 M KI; 10 µL) was added to 1 mL of each filtrate and the absorbance was measured using a UV-Vis spectrometer (UV-2600, Shimadzu). The presence of starch can be confirmed by the iodine reagent, which turns the sample deep blue in the presence of starch.

Thereafter, 5 mL of the filtered gargle sample was hand-pressed through the moistened column. The resultant 5 mL of eluent was collected in a new 50-mL tube (Figure 2.2). Next, acetone precipitation was conducted on both the samples (that were and were not passed through the amylase depleting device) by adding 5 mL of chilled acetone. The samples were centrifuged at 16,000 × g for 30 min at 4°C. The supernatant was discarded, the rim of the tube patted dry, and the pellet resuspended in 50 µL of a viral disruption buffer reported by Dollman et al. The reconstituted pellets were vortexed for 30 s and incubated at room temperature for 15 min.
Figure 2.2: An illustration of salivary amylase depleting device loaded with a gargle sample. Image Created with BioRender.com.

Sample preparation for saliva samples using an amylase depleting device

Five hundred μL of the saliva sample was diluted with 2.5 mL of LC-MS grade H₂O and was filtered through a 0.45-μm polyethersulfone membrane filter. This 3 mL sample was passed through the amylase depleting device similar to the method for gargle samples previously mentioned. Next, 6 mL of chilled acetone was added to the sample followed by overnight incubation at 4°C. The samples were centrifuged at 16,000 × g for 30 min at 4°C. The supernatant
was discarded, the rim of the tube patted dry, and the pellet resuspended in 50 µL viral disruption buffer.

Sample preparation for human protein standards and controls using an amylase depleting device

The negative control, consisting of pooled human saliva (pre-COVID-19) collected before November 2019, was prepared by spiking 500 µL thawed stock into 10 mL of water. Then, the negative control was filtered and processed as a gargle sample and was subjected to the amylase depleting device identically to the gargle samples. Two hundred pmol of human serum IgA, 200 pmol of human serum albumin and 2,000 pmol of human amylase in LC-MS grade H₂O were also passed through the amylase depleting device. The protein standards and their respective eluents were then lyophilized overnight and reconstituted in 10 µL of viral disruption buffer. A final concentration of 20 pmol/µL, 20 pmol/µL and 200 pmol/µL of IgA, HSA and amylase, respectively, were spotted using the sandwich method for MALDI-ToF analysis.

SDS-PAGE analysis

SDS-PAGE analysis was performed on the remaining reconstituted pellets following the standard reducing procedure provided by Invitrogen with 8% Bis-Tris Plus acrylamide gels for 35 min at 150 V with MES SDS running buffer. Gargle samples (20 µL) before and after passing through the amylase depleting device and protein standards of amylase (14 µg), HSA (2.5 µg) and IgA (3.5 µg) were analyzed on the gel. Ten µL pre-stained protein ladder was loaded along with the samples. The gels were stained using a 0.1% w/v PhastGel Blue R solution for 20 min, followed
by destaining with 3:1:7 methanol/glacial acetic acid/Milli-Q water by volume for 45 min. SDS-PAGE gels were stained and destained by gentle agitation on a belly dancer (Stovall Life Science, Greensboro, NC).

**Sample preparation for recombinant full-length S protein of SARS-CoV-2 variants**

We analyzed the purified recombinant forms of the full-length S glycoprotein that were generated by transient transfection of a particular variant into human embryonic kidney HEK293 cells and purified by immobilized metal affinity and gel filtration chromatography. The theoretical mass and concentration for each of these proteins is shown in Table 2.2. Five different recombinant full-length S protein of SARS-CoV-2 variants namely the Wuhan variant, Spike D614G Variant, Delta Variant, B.1.1.529 lineage and BA.2 lineage (Omicron variants) were directly spotted (1 µL each) on the MALDI-MS target plate for analysis.

<table>
<thead>
<tr>
<th>SARS-CoV-2 variant</th>
<th>Theoretical mass (Daltons)</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wuhan</td>
<td>140,699</td>
<td>0.25 µg/µL</td>
</tr>
<tr>
<td>Spike D614G</td>
<td>137,031</td>
<td>1.2 µg/µL</td>
</tr>
<tr>
<td>Delta</td>
<td>139,536</td>
<td>0.25 µg/µL</td>
</tr>
<tr>
<td>B.1.1.529 Lineage (Omicron)</td>
<td>139,878</td>
<td>0.25 µg/µL</td>
</tr>
<tr>
<td>BA.2 Lineage (Omicron)</td>
<td>139,735</td>
<td>0.25 µg/µL</td>
</tr>
</tbody>
</table>
RT-qPCR and Next-generation sequencing

RT-qPCR and Next-generation sequencing experiments for the saliva samples were performed by our collaborators Joshua Bland and Dr. Shrihari S. Kadkol at the Department of Pathology, University of Illinois at Chicago, US.
CHAPTER 3

MALDI-TOF AS A POTENTIAL DIAGNOSTIC TOOL FOR DETECTION OF COVID-19

Introduction

In December 2019, a novel coronavirus named SARS-CoV-2 emerged in Wuhan, China and has rapidly spread around the world since then. As of May 2023, the United States alone has witnessed more than 103 million cases and more than 1.1 million deaths due to COVID-19. While the list of clinical manifestations of COVID-19 predominantly includes fever, dry cough, anosmia and dyspnea, around 40-50% of the population are asymptomatic. At the beginning of the pandemic, only the individuals that exhibited any symptoms were tested for COVID-19 and were asked to quarantine. This resulted in the rapid spread of the virus as the asymptomatic individuals were still capable of transmitting the virus. A report by Johansson et al. concluded that at least 50% of new infections originate from contact with asymptomatic carriers. These reports are difficult to interpret, however, because not all persons who are asymptomatic continue to be asymptomatic throughout the course of infection; percentages of persons who remain asymptomatic are reported to range from 20-50%. 

Facing the ongoing pandemic caused by SARS-CoV-2, early diagnosis of COVID-19 is still of great importance to control community outbreaks. Thus, it is crucial to design rapid,
sensitive, and accurate diagnostic tests to address this public health crisis. The current gold standard test for COVID-19 is molecular testing for the presence of SARS-CoV-2 RNA, which is accomplished by RT-qPCR. As discussed in Chapter 1, the sensitivity and specificity of RT-qPCR can be high, depending on the primer/probe sets used. False positive results, although uncommon, can occur because of contamination due to the template amplification nature of RT-qPCR, high viral loads in samples, and widespread prevalence of infection. False negative results are generally related to sampling issues, nucleic acid degradation, testing in the very early phase soon after exposure or late convalescent phase of the infection, or infections with variants that contain mutations in the primer/probe binding sites. RT-qPCR tests are usually carried out using specimens collected via a nasopharyngeal (NP) or oropharyngeal swab or saliva.\textsuperscript{22,3383}

SARS-CoV-2 and the biomarkers of COVID-19 can also be found in multiple other specimen types. These include tracheal aspirate, sputum, whole blood, plasma, and serum.\textsuperscript{24} Of particular interest in finding alternatives to NP swabs, it has been reported that standardized saliva collection can be adopted to detect SARS-CoV-2 infection, as saliva droplets are a main vehicle of viral transmission and the salivary glands are reported to be a target of infection as well as a reservoir for the virus.\textsuperscript{28,84} Saliva as a specimen for diagnosis provides benefits like rapid, non-invasive collection, cost-effectiveness and patient acceptance. Perhaps of highest diagnostic importance, however, the SARS-CoV-2 load in the saliva is reported to be maximum within the first week of symptom onset, providing an avenue for early diagnosis of COVID-19.\textsuperscript{32}

In addition to the type of specimen collected for a diagnostic test, it is also crucial to decide upon the type of biomolecule to be detected. In this dissertation work, proteins were considered as target analytes owing to factors such as the sensitive nature of RNA demanding trained personnel,
constant mutations in the RNA genome of the SARS-CoV-2 virus and the higher stability profile of proteins compared to RNA.\textsuperscript{5,33} Moreover, proteins are present in higher amounts in the virus. Although proteins cannot be directly amplified like nucleic acids, this considerably reduces the chances of producing false positive results due to contamination. Furthermore, several different proteins can be detected in the same analysis, which increases the number of diagnostic markers and improves the conclusions drawn from a test. For instance, various classes of proteins, like antibodies, cytokines and viral proteins can be collectively analyzed using mass spectrometry and other proteomic techniques for their characterization and quantitation.\textsuperscript{33,34} Beyond merely detecting a single target of interest, such as a specific SARS-CoV-2 protein, proteomics combined with informatics can derive actionable information from large data sets and, thus, help in establishing biomarkers for this disease.\textsuperscript{36} Such holistic information is necessary to understand the pathogenesis of SARS-CoV-2 infection and may lead to novel therapies not only against the virus, but also to help modify the host response to the virus.

Developing a diagnostic test that overcomes the limitations of nucleic acid testing would assist in fulfilling the worldwide demand for reliable, rapid, accurate and cost-effective testing. Recently, several studies utilizing a powerful analytical tool like MALDI-ToF MS as a potential method for the detection of SARS-CoV-2 are reported in Table 1.1. However, these studies were limited to the mass-to-charge range in collected mass spectra (less than 20,000 m/z) and often lacked association of the peaks to any molecular identity. In another study, published recently by Iles et al. MALDI-MS was employed to detect both viral and human proteins from samples of gargled water.\textsuperscript{54} Though this study utilized non-invasive sample collection and investigated a wider m/z range, no direct comparison was made with RT-PCR results in clinical samples.
Here, we employ a similar technique to that reported by Iles et al. to analyze gargle samples to determine SARS-CoV-2 infection. However, this study extends from Iles et al. in that we compare the mass spectral analysis to the RT-qPCR status of the individuals from NP swabs.\(^\text{54}\) Comparing the RT-qPCR status is crucial as it is considered the gold standard for COVID-19 diagnosis. If high accuracy is achieved by comparing the MALDI-ToF results with the RT-qPCR results, we can establish our MS method as a reliable method. The method described in this chapter potentially detects both viral proteins and the antibodies produced against them. For the 60 saliva samples that were analyzed, the area under the curve (AUC) of potential viral protein and host protein peaks was used to detect the presence of SARS-CoV-2 and the host immune response against the virus.

**Results and Discussion**

In this chapter, we report the analysis of 60 water gargle samples using a MALDI-ToF methodology compared to the RT-qPCR status of individuals done on NP swabs. We, however, did not directly analyze the gargle samples by RT-qPCR, and instead relied on the results from NP swabs to classify an individual as positive or negative for SARS-CoV-2. A baseline spectrum of peaks was determined in saliva collected prior to the emergence of COVID-19. This spectrum served as a negative control against which changes in COVID-19 positive individuals could be compared. Figure 3.1 presents the profile for a representative COVID-19 negative sample that is nearly identical in terms of m/z peaks to the spectrum for the negative control. Of particular note are the peaks consistent between the two profiles, for salivary proteins these would be present in both diseased and healthy individuals.\(^\text{85}\)
Figure 3.1: MALDI-ToF mass spectra of a gargle sample from an example donor who tested COVID-19 negative (—red) along with the negative control (—blue). Both profiles closely overlapped each other, and no other signal was detected after 65,000 m/z.\textsuperscript{85}

Distinct differences in MALDI-ToF spectra were observed in individuals who were COVID-19 negative or positive in NP swabs (Figure 3.2). Overall, gargle samples from COVID-19 positive individuals showed higher intensities along with additional peaks in the spectrum. Peaks around 23,000 m/z, 28,000 m/z and 56,000 m/z were present in both positive and negative individuals. Interestingly, these peaks showed higher signal intensities in the COVID-19 positive cases (Figure 3.2-A). Additional peaks that were unique to COVID-19 positive individuals were
present between the ranges of 33,000-51,000 m/z and 65,000-120,000 m/z (Figure 3.2-B and 3.2-C). It is important to note that although the peaks found within these ranges could be used to reasonably separate COVID-19 positive from negative individuals, variation in peak intensities was apparent between COVID-19 positive spectra. We suspect that this may be due to sampling at different time points in the course of the infection and varying host immune responses. Included among the viral and host proteins discussed is a peak located at 11,150 m/z used as a quality control feature that is most likely cystatin A, a resident protein of saliva.\textsuperscript{34} The presence of this peak was used as an internal control to deem a sample as successfully gargled and appropriate to be included for analysis.\textsuperscript{35}

The mass spectra of gargle specimens contain signals from host salivary proteins and viral proteins in COVID-19 positive individuals. Hence, it is crucial to identify peaks that confirm the presence of SARS-CoV-2 in COVID-19 positive individuals along with markers of an immune response against the virus. To begin, we analyzed the mass spectra of pre-COVID-19 saliva and a gargle sample from an individual who tested negative in an NP swab by RT-qPCR (Figure 3.1) to establish a baseline spectrum against which spectra from COVID-19 positive individuals could be compared. The most prominent peak in these spectra was observed near 56,000 m/z. Two smaller peaks are also present near 23,300 m/z and 28,000 m/z. Based on a previous study, we suspected that the peaks near 56,000 m/z and 23,000 m/z most likely represent IgA heavy chain and Ig light chains, respectively.\textsuperscript{54} To support this notion, we analyzed three human serum immunoglobulins (IgA, IgG and IgM) under reducing conditions using the same protocol as in gargle samples (Figure 3.3). The heavy chains for IgA, IgG and IgM were detected at 57,372 m/z, 51,142 m/z and 71,678 m/z, respectively, while the light chains for these antibodies were found between 23,000 m/z and
Figure 3.2: MALDI-ToF mass spectra of a gargoyle sample from a COVID-19 negative donor (—blue) and an overlay of two COVID-19 positive donors (—red, —green). Panel A is the full-range mass spectra of the samples. Panels B and C are specific ranges where differences in mass spectra are prominently observable between COVID-19 positive and COVID-19 negative samples.
24,000 m/z. These results suggest that the peaks around 56,000 m/z and 23,000 m/z in pre-COVID-19 saliva and the gargle samples from a COVID-19 negative individual are indeed IgA heavy chain and Ig light chains. This is not surprising, given basal levels of secretory IgA and light chains are nearly always detected in saliva. Various other peaks were observed throughout the spectra of the immunoglobulin standards, which may represent combinations of heavy and light chains, dimers of heavy chains, and multiply charged ions.

Interestingly, the IgA heavy chain and Ig light chain peaks were almost twice as intense in COVID-19 positive individuals when compared to COVID-19 negative individuals. We suspect that this may be due to a robust immune response against SARS-CoV-2 upon exposure to the virus. Along the same lines, another feature to note is the existence of a peak at 70,603 m/z in COVID-19 positive individuals that could correspond to the heavy chain peak of IgM at 71,678 m/z. This suggests an early response to infection in COVID-19 positive individuals, as it has been reported that viral-specific IgM antibodies are produced first, followed by IgA and IgG.

Additionally, human α-amylase is found in two forms in saliva with molecular weights of around 56 kDa (unglycosylated) and 62 kDa (glycosylated). MALDI-MS data have also been reported for human salivary α-amylase and a Y151M mutant, both having a parent ion peak near 56 kDa. Because these masses fall within the region discussed above regarding heavy chains of IgA, samples of α-amylase were also analyzed by MALDI-MS. The observed spectra exhibit a peak at 56,600 m/z with a shoulder at 58,000 m/z (Figure 3.4), therefore, α-amylase overlaps with the IgA heavy chain peak in the gargle sample spectra.

Apart from the IgM heavy chain peak at 71,000 m/z, additional peaks were observed between 65,000 m/z and 120,000 m/z in gargle samples from COVID-19 positive individuals.
Figure 3.3: MALDI-ToF mass spectra of human serum derived IgA (—blue), IgG (—green) and IgM (—red) isolated from human serum reduced with 1 M DTT for 10 min. Panel A shows the entire mass range collected and Panels B and C show zoomed-in ranges.
Peaks were prominent in the ranges of 66,400–68,100 m/z and 78,600–80,500 m/z (Figure 3.2-C). Comparing the UniProt database (UniProtKB P0DTC2) and the work of Iles et al.\textsuperscript{54}, the peak at 79,837 m/z most likely represents a signal for the S1 fragment of the SARS-CoV-2 spike protein; a peak with a similar m/z was also observed in the positive control spectrum from SARS-CoV-2 (Figure 3.5). Furthermore, the peaks described by Iles et al.\textsuperscript{54} as viral envelope proteins (VEPs) were also visible in COVID-19 positive profiles (Figure 3.2-B). Compared to COVID-19 negative individuals, a signal was consistently observed in the range of 66,000 m/z to 68,000 m/z in gargle samples from COVID-19 positive individuals. By comparing numerous positive spectra, it appears that there are at least two distinct species with closely overlapping m/z envelopes. This peak could represent the S2 fragment of the SARS-CoV-2 S protein, which is predicted by UniProt.
(UniProtKB P0DTC2) to have a mass of 64.5 kDa (unglycosylated) with additional mass arising from extensive glycosylation.\textsuperscript{91} Alternatively, this peak may arise from a fragment of IgA. As such, this signal cannot be unequivocally identified at this time. It should be noted that, in the mass spectrum of viral isolates from cell culture (positive control), an intense peak from bovine serum albumin occurs at 66,600 m/z; this unfortunately would suppress the signal from S2 (if present) into the baseline for the control samples of SARS-CoV-2 (Figure 3.5).\textsuperscript{85}

To estimate the sensitivity of the MALDI-ToF protocol, we tested a saliva sample that contained a very low SARS-CoV-2 viral load by quantitative RT-qPCR, as described previously.\textsuperscript{92} The Ct value of this sample was 36.09 and, as such, the viral load was less than the quantifiable limit of the assay (600 copies/ml). The observed signal in the mass spectrum for the potential biomarker peak found between 78,600 and 80,500 m/z was approximately 3 times the baseline noise level (S/N = 3), a commonly accepted value for finding the limit of detection in MALDI-ToF methods (Figure 3.6). Based on the S/N ratio, we interpreted this sample as being positive for SARS-CoV-2. This result suggests that the MALDI-ToF protocol may indeed be as sensitive as RT-qPCR to detect SARS-CoV-2 in specimens that contain very low viral loads. More studies are necessary to determine the exact limit of detection of the MALDI-ToF protocol.\textsuperscript{85}

In order to compare the results of the NP RT-qPCR samples and the protein profiles collected with gargle samples via MALDI-ToF, the AUC was calculated under the seven peak ranges of interest listed in Table 2.1 for each sample. As stated above, the peak located at 11,150 m/z is most likely to be associated with cystatin A, a resident protein of saliva.\textsuperscript{34} The presence of
Figure 3.5: MALDI-ToF spectrum of heat inactivated cell lysate and supernatant of Vero E6 cells infected with SARS-CoV-2, which was utilized as the positive control. Panel A shows the entire mass range collected and Panel B shows the major peak signals beyond 70,000 m/z. 85
Figure 3.6: MALDI-ToF mass spectrum of a COVID-19 positive saliva sample beyond the limit of quantitation for RT-qPCR. The inset shows the potential biomarker peak in the range of 78,600 – 80,500 m/z having a signal approximately 3 times higher than the baseline (S/N = 3), which is commonly accepted for finding the limit of detection in MALDI-ToF methods.\(^8^5\)

This peak was used as an internal control. AUC values for a given feature were sorted from largest to smallest. Sorting in this way showed a reasonable separation between COVID-19 positive and COVID-19 negative specimens for five out of the seven features, including the peaks of S1, S2/immune protein, immunoglobulin heavy/amylase, immunoglobulin heavy doubly charged, and the biomarker near 112,000 m/z as seen in Figure 3.7.\(^8^5\)
Figure 3.7: AUC values for the 60 gargoyle samples for the five potential biomarker peak range. The sample data files were labeled 1–60 and the AUC for the biomarker range for each file is depicted. The COVID-19 positive files (red) were clustered and labelled as 1–30 while the COVID-19 negative files (blue) were clustered and labelled as 31–60. A marked difference for the AUC in the biomarker ranges of COVID-19 positive versus COVID-19 negative samples can be observed.⁸-five
To establish a cut-off threshold for each potential biomarker as it compares to the COVID-19 status, the AUC values for each range were sorted from high to low, as described above, and compared to the disease status. Cut-offs were made to yield the best separation of negative/positive samples. The samples with AUC values above the threshold were assigned positive and values below were assigned negative. The cut-off values for a given feature chosen under this criterion are reported in Table 3.1, along with the percent agreement between RT-qPCR results and our analysis.85

For all five potential biomarkers mentioned in Table 3.1, we achieved 90% and higher agreement with the RT-qPCR results. Although the identities of the proteins are yet to be confirmed, these m/z ranges certainly exhibit a relationship with the COVID-19 status.85 We refined these observations further with ROC curve analysis. The true positive rate (sensitivity) on the Y-axis was plotted against the false positive rate (100-specificity) on the X-axis at each of the observed AUCs for every peak, shown in Figure 3.8. The analysis clearly demonstrates the ability of the MALDI-ToF assay to discriminate between positive/negative COVID-19 status while providing a quantitative optimization of sensitivity/specificity for each potential biomarker, as summarized in Table 3.2.85

Figure 3.8 represents the ROC plots of true versus false positive rates for each of the seven m/z ranges listed in Table 2.1. Figure 3.8-A shows data for the internal quality control biomarker tentatively assigned as cystatin A. The data points lie closely along the diagonal line with the AUC indicating no discrimination between COVID-19 negative and positive samples, as predicted.
Table 3.1: Percent agreement between RT-PCR results and MALDI-ToF results as determined by cut-off AUC values. Cut-off values for each biomarker peak were based on sorting all specimen AUC values for each feature (m/z peak ranges) and setting a threshold AUC value that separated COVID-19 positive from COVID-19 negative results.\textsuperscript{85}

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Cut-off AUC:</th>
<th>PCR-Positive</th>
<th>PCR-Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ig heavy chain/amylase\textsuperscript{+2} (27,900-29,400 m/z)</td>
<td>5.2E+06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Positive</td>
<td>96.67%</td>
<td>10.00%</td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Negative</td>
<td>3.33%</td>
<td>90.00%</td>
<td></td>
</tr>
<tr>
<td>Ig heavy chain/amylase (55,500-59,000 m/z)</td>
<td>4.0E+07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Positive</td>
<td>96.67%</td>
<td>10.00%</td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Negative</td>
<td>3.33%</td>
<td>90.00%</td>
<td></td>
</tr>
<tr>
<td>Spike protein S2/immune protein (66,400-68,100 m/z)</td>
<td>1.5E+06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Positive</td>
<td>100.00%</td>
<td>6.67%</td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Negative</td>
<td>0.00%</td>
<td>93.33%</td>
<td></td>
</tr>
<tr>
<td>Spike protein S1 (78,600-80,500 m/z)</td>
<td>1.0E+06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Positive</td>
<td>93.33%</td>
<td>10.00%</td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Negative</td>
<td>6.67%</td>
<td>90.00%</td>
<td></td>
</tr>
<tr>
<td>Asymptomatic Biomarker (111,500-115,500 m/z)</td>
<td>1.2E+06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Positive</td>
<td>93.33%</td>
<td>10.00%</td>
<td></td>
</tr>
<tr>
<td>MALDI-ToF Negative</td>
<td>6.67%</td>
<td>90.00%</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.8: ROC analysis results on the potential protein biomarker ranges identified by MALDI-ToF from gargo samples. Panel 6A-G presents the graphs of true versus false positive rate for each of the seven labeled m/z ranges. In any given panel, each point represents a MALDI-ToF specificity/sensitivity value at a cut-off AUC for each potential biomarker range, with the highest specificity/sensitivity and discriminating power having values near the top left of each plot.\textsuperscript{85}
Table 3.2: Summary of ROC curve analysis for each potential biomarker as calculated for AUC of peaks within associated m/z ranges. A peak area threshold is listed for each potential biomarker which optimizes sensitivity and specificity for the MALDI-ToF protocol when compared to COVID-19 status.

<table>
<thead>
<tr>
<th>Potential biomarker m/z range</th>
<th>Peak area threshold</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gargle marker 11,140 – 11,160</td>
<td>&gt;57,89,021.17</td>
<td>0.548</td>
<td>40</td>
<td>76</td>
</tr>
<tr>
<td>Ig light chains 23,500 – 23,800</td>
<td>&gt;46,29,948.31</td>
<td>0.756</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>Ig Heavy chain/Amylase(^+2) 27,900 – 29,400</td>
<td>&gt;43,68,750.14</td>
<td>0.941</td>
<td>96.67</td>
<td>90</td>
</tr>
<tr>
<td>Ig Heavy chain/Amylase 55,500 – 59,000</td>
<td>&gt;3,86,84,701.83</td>
<td>0.961</td>
<td>96.67</td>
<td>90</td>
</tr>
<tr>
<td>Spike protein S2 66,400 – 68,100</td>
<td>&gt;14,29,133.22</td>
<td>0.976</td>
<td>100</td>
<td>93.33</td>
</tr>
<tr>
<td>Spike protein S1 78,600-80,500</td>
<td>&gt;9,37,903.91</td>
<td>0.954</td>
<td>93.33</td>
<td>90</td>
</tr>
<tr>
<td>Asymptomatic biomarker 111,500-115,500</td>
<td>&gt;11,04,743.08</td>
<td>0.933</td>
<td>93.33</td>
<td>90</td>
</tr>
</tbody>
</table>

Figures 3.8C-G plots the data for the potential biomarkers S1, S2/immune protein, immunoglobulin heavy/amylase, immunoglobulin heavy doubly-charged, and the potential biomarker near 112,000 m/z. These five biomarkers show highly sensitive and specific discrimination between the COVID-19 positive and negative individuals with AUCs $\geq 0.933$ and
sensitivities and specificities of 93.33–100.00% and 90.00–93.33%, respectively (Table 3.3). The closer the AUC is to 1 for ROC curve analysis, the better the model for predicting the disease state. Figure 3.8-B displays the graph for the immunoglobulin light chains m/z region that shows a much weaker discrimination ability; this phenomenon was also observed anecdotally from visual examination of the mass spectra. In summary, the ROC analysis indicates that five peaks in the MALDI-ToF spectrum are capable of highly sensitive and specific discrimination of COVID-19 status in individuals.\textsuperscript{85}

It is interesting to note that 89% of the positive samples were from donors who were asymptomatic. This is not surprising given that the student athletes in this study are generally young, healthy individuals. This cohort has been an elusive group to track, particularly early in the pandemic when only symptomatic persons were eligible for RT-qPCR testing. The COVID-19 positive, asymptomatic group was reported to be as contagious as symptomatic persons\textsuperscript{93} towards the beginning of the pandemic and had been suggested to be disproportionately responsible for spreading the virus due to a lack of symptoms. This could be of concern in a setting such as a university where communal housing, dining and recreational facilities are predominant. Recent reports, however, cast doubt on a linear relationship between viral loads, symptoms and transmissibility (i.e., how much of an RT-qPCR detected viral load is replicating virus that can be transmitted to other persons is unclear). Transmissibility will need to be confirmed by viral cultures.\textsuperscript{85}

As this research has demonstrated, a highly sensitive and specific saliva/gargle test was developed for diagnosing COVID-19 infection using MALDI-ToF MS. The method described in this study is relatively rapid and inexpensive and is sensitive enough to detect SARS-CoV-2
infection in samples with very low viral loads compared to RT-qPCR. It has advantages over other tests such as non-invasive sampling and the ability to observe both viral proteins and host response. By comparing the mass spectra in known COVID-19 positive and COVID-19 negative individuals, clear distinctions were observed in the range of 20,000–200,000 m/z, a range not analyzed in previous COVID-19 MS studies. Furthermore, the AUC of putative host and viral protein peaks was used to correlate the MALDI-ToF profiles with the RT-qPCR results from NP swabs sampled in parallel.\(^8^5\) This method was further extended to a larger sample size study, where the data analysis was performed in collaboration with Department of Computer Science, Northern Illinois University, US. ML algorithms were utilized to achieve higher accuracies for the established MALDI-TOF testing platform.\(^9^4\) Chapter 1 referenced various research studies that employed ML algorithms for MALDI-ToF data analysis to detect COVID-19. In line with these studies, our own research aimed to apply similar approaches. We conducted an analysis using 152 human gargle samples, which resulted in an accuracy of 94.21% and more importantly, gave a biological foundation to the decision-making process of the algorithms. Such techniques would strengthen the relationship between AI and clinical diagnostics by providing biomedical researchers and healthcare workers with trustworthy and, most importantly, explainable test results.\(^9^4\)
CHAPTER 4

MALDI-TOF PROTEIN PROFILING OF AMYLASE DEPLETED GARGLE
SAMPLES FOR COVID-19 TESTING

Introduction

Saliva is a complex biological fluid consisting of a mixture of major and minor salivary gland secretions, nasal and bronchial secretions, plasma filtrates, host cells and bacteria. The fluid is slightly acidic and contains various organic (proteins, hormones, nucleic acids, fatty acids) and inorganic (Na+, K+, Cl−, Ca2+, HCO3−, H2PO4−, I−, Mg2+ and NH4+) components.\textsuperscript{95,96,97,98}

Described as a “mirror of the body,” saliva reflects the physiological and pathological states of the body. Saliva has been used for screening\textsuperscript{99,100}, diagnosis\textsuperscript{101,102}, prognosis and monitoring\textsuperscript{103,104} of numerous human diseases. Collecting saliva is fast, non-invasive and can be done by the subjects themselves. Saliva does not need any transport medium and requires minimal processing before testing compared to other specimen types.\textsuperscript{25,105} Major advancements in the field of salivary diagnostics have included RNA-sequencing, point-of-care technologies, liquid biopsies and protein profiling.\textsuperscript{106}
Proteomic analysis of saliva has led to the discovery of ~3000 proteins, including mucins, proline-rich proteins, histatins, statherins, cystatins, interleukins, amylase, albumin and immunoglobulins (among others). A number of these proteins are currently being pursued as biomarkers for diagnosis of oral diseases, various cancers, Srojen’s syndrome, and even autism. Saliva has been used to detect viruses such as hepatitis, human immunodeficiency virus, dengue and, more recently, severe acute respiratory syndrome coronavirus type-2 (SARS-CoV-2). Several studies have demonstrated higher efficiencies of SARS-CoV-2 detection in saliva as compared to nasopharyngeal swabs, owing to the fact that the receptor used for SARS-CoV-2 viral entry, angiotensin-converting enzyme-2 (ACE-2), is abundantly expressed in the oral epithelial cells. In fact, it has been shown that coronavirus disease 2019 (COVID-19) can be detected 1-5 days earlier in saliva samples as compared to nasopharyngeal swabs.

A major hurdle in proteomic analyses of saliva, however, is the presence of high-abundance proteins that can mask the presence of target proteins expressed at a lower level. Salivary α-amylases, albumin and immunoglobulins alone make up to 75% of the total saliva proteome. Consequently, it is crucial to deplete these high-abundance proteins (or conversely, to enrich low-abundance proteins) to improve detection sensitivity. For instance, non-analyte, high-abundance proteins can cause significant ion suppression for low-abundant biomarkers in mass spectrometric techniques. Particularly, in a complex biological matrix, the ionization efficiencies are highly influenced by the abundance, molecular weight and chemistry of the proteins.

Salivary α-amylase is the most abundant of these proteins and is an enzyme that catalyzes the hydrolysis of 1,4-glucosidic linkages in starch and other polysaccharides. It exists in two proteoforms: non-glycosylated (~56 kDa) and glycosylated (~59 kDa). Several techniques
are currently utilized for α-amylase depletion including syringe-based potato starch device, gel filtration, affinity chromatography by lectin ConA, co-precipitation and paper-based chips. Deutsch et al. designed and patented an amylase depleting device to separate salivary α-amylase from the whole saliva by affinity adsorption to potato starch. The authors demonstrated a six-fold amylase reduction as compared to the total amylase in the saliva, and 97% reduced amylase activity after using this device. Xiao et al. and Crosara et al. validated the functionality and practicality of this device via SDS-PAGE and western blotting. In these studies, several identified proteins, which could be potential biomarkers, were protected by amylase removal including desmoplakin, short palate lung and nasal epithelium carcinoma-associated protein 2, mucin-7 and several immunoglobulins’ isoforms.

In chapter 3, we demonstrated the utility of matrix-assisted laser desorption-ionization time of flight mass spectrometry (MALDI-ToF MS) to distinguish COVID-19 (coronavirus disease 2019) positive and negative individuals. One hurdle in the analysis of these protein profiles is the dominant signal by salivary amylase resulting in ion suppression of other proteins. In this chapter, we hypothesize that depletion of salivary α-amylase from gargle samples before MALDI-ToF analysis would unmask and improve the detection of low-abundance proteins that are otherwise difficult to visualize. The simple removal of amylase coupled with the sensitivity of MALDI-ToF could open avenues for the discovery of new protein biomarkers in saliva-based diagnostics. Taking this into account, we combined methods to remove amylase by affinity adsorption with potato starch and salivary protein profiling using MALDI-ToF MS. We also optimized sample preparation involving the viral disruption buffer. In our previous study, we used LBSD-X buffer by MAPSciences (Bedford, UK) as a viral disruption buffer for gargle samples to
test for COVID-19. However, LBSD-X buffer is not readily available, and the exact composition of this buffer remains unknown. Hence, the viral disruption buffer reported by Dollman et al. was used in this study. The buffer has already been used in methods for the detection of SARS-CoV-2 and influenza viruses such as H1N1 and H3N2. Overall, we aimed to develop a simple, rapid and inexpensive technique to deplete salivary amylase and to detect COVID-19 by performing MALDI-ToF protein profiling.

**Results and Discussion**

A key issue in proteomic analysis of any specimen by MALDI-ToF MS is the presence of high-abundance proteins which, in terms of assay development, can significantly mask the detection of other low-abundance proteins. This poses a hurdle to biomarker discovery as these low-abundance proteins might be potential candidates for use in the diagnosis of diseases. While there are various studies to deplete high-abundance proteins from blood, there exists limited data concerning saliva samples. One of the high-abundance proteins in saliva is amylase. Herein, we adopted a simple, cost-effective yet efficient method to deplete salivary amylase in COVID-19 gargle samples before MALDI-ToF MS analysis.

Deutsch et al. designed an amylase depleting device that included a 0.45 μm paper filter at the tip of the syringe. In our initial efforts in preparing and utilizing this apparatus, we experienced difficulty while removing the plunger between starch activation and sample loading. Therefore, we replaced the 0.45 μm paper filter with a 0.45 μm syringe filter (Figure 2.2), minimizing disturbances to the potato starch bed while releasing the plunger. We also passed
increased amounts of water through the device before loading samples in order to remove water-soluble residues present within the potato starch that might contribute to the ion suppression of proteins or interfere with amylase adsorption. The presence of starch can be confirmed by the iodine reagent, which turns the sample deep blue in the presence of starch. However, it can be distinctly observed in Panel A of Figure 4.1 that starch was no longer present after passing 20 mL of water. This was also verified by the UV-Vis spectrometer shown in Figure 4.1-B. Hence, 20 mL of water was passed through the amylase depleting device before passing any gargle sample or standards/controls as it resulted in a significant decrease in water-soluble residues present in starch. These additional washing steps were crucial to prevent any ion suppression of proteins in the MALDI-ToF analysis due to the presence of water-soluble starch residues in the samples. Research concerning the types of starch used and the influence of its structure on the interaction with amylase is poorly understood. Furthermore, it is stated that the Michaelis-Menten model cannot be applied to in-vitro studies of amylase action because of uncertainties in the structure of its substrate (starch), the possible presence of inhibitors and high enzymatic activity.\textsuperscript{131,132}

The LBSD-X buffer by MAPSciences (Bedford, UK) used in our previous study isn’t available in the market, and thus reproducibility of our method is compromised. This motivated us to prepare an in-house buffer that was reported as an efficient viral reconstitution buffer. As mentioned earlier, we made modifications to the viral reconstitution buffer by Dollman et al. We optimized the buffer composition by substituting 2 mM dithiothreitol with 50 mM TCEP (each reducing agent) and by substituting sonication followed by 2-hour incubation with a simple vortex
Figure 4.1: Iodine test to determine the presence of starch after multiple water washes. The gradual color change in Panel A and the overall decrease in the absorbance in the UV-Vis spectra in Panel B confirmed the reduction in water-soluble starch residues present in starch.

and 15-minute incubation at room temperature. The buffer showed promising results as we observed peaks within the ranges established previously (11,140–11,160 m/z; 23,550–23,800 m/z; 27,900–29,400 m/z; 55,500–59,000 m/z; 66,400–68,100 m/z; 78,600–80,500 m/z; and 111,500–115,500 m/z; Figure 4.2). Furthermore, and perhaps more importantly, we could still observe differences between COVID-19 positive and negative gargle samples using the in-house buffer.
Figure 4.2: MALDI-ToF mass spectra of a COVID-19 positive gargle treated with LBSD-X buffer (—red) and an overlay of the same sample treated with the new optimized buffer (—blue). Panel A is the full-range mass spectra of the same positive sample treated with both buffers. The optimized buffer showed peaks in the ranges that were established in our previous study. Panels B and C are specific ranges in mass spectra where the peaks obtained by both buffers closely overlap with each other.
Before testing the amylase depleting device with gargle samples from subjects, we determined whether the procedure affected the detection of other proteins found in saliva. Individual protein standards that exist in high-abundance in saliva, including salivary amylase, HSA and IgA were tested. When standard amylase was passed through the device, the collected eluent showed a negligible signal at the theoretical m/z value of the parent peak confirming efficient removal of amylase (Figure 4.3). Moreover, as anticipated, when HSA and IgA were passed through this device, minimal to no binding/interaction was observed between the starch bed and these two proteins (Figures 4.4 and 4.5). The spectra from these experiments confirmed that the activated starch bed selectively binds amylase and results in negligible sample loss of other proteins.

Figure 4.3: MALDI-ToF mass spectra of human salivary α-amylase (—blue) and an overlay of the same sample after passing through the amylase depleting device (—red). No peak was observed at 56,000 m/z after passing through the device, confirming the binding between the starch bed and salivary amylase.
Figure 4.4: MALDI-ToF mass spectra of standard HSA (—blue) and an overlay of the same sample after passing through the amylase depleting device (—red). The peaks at 66,995 m/z and 33,592 m/z which correspond to the single- and double-charged peaks of HSA, respectively, were observed with similar signal intensities before and after using the amylase depleting device.

Next, we passed pre-COVID-19 saliva, the negative control, through this device. The complete protein profile of the negative control before and after passing through the amylase depleting device has observable differences (Figure 4.6). For instance, the intensity of the peak at 56,216 m/z (approximately the theoretical molecular weight of salivary amylase) was considerably reduced after the sample was passed through the device whereas the peak around 66,000 m/z was enhanced. A low intensity peak between 56,000-58,000 m/z was maintained in the eluent.
Comparing Figures 4.3 and 4.5, this signal coincides with the heavy chain of salivary IgA, which would not be retained on the starch bed.

Figure 4.5: MALDI-ToF mass spectra of a standard human serum IgA (—blue) and an overlay of the same sample after passing through the amylase depleting device (—red). The peaks at 23,397 and 57,197 m/z which correspond to the light chain and heavy chain peaks of IgA respectively, can be observed with similar signal intensities before and after using the amylase depleting device.
In order to further visualize the effects of passing gargle samples through the amylase depleting device, we performed an SDS-PAGE analysis. Figure 4.7 shows the positions of the bands for protein standards HSA, amylase and IgA. For IgA, two bands at \(~55\) kDa and \(~25\) kDa represent heavy and light chains respectively (Lane 4). Comparing the positions of these standard

![Image](image.png)

Figure 4.6: MALDI-ToF mass spectra of pre-COVID-19 saliva (—blue) and an overlay of the same sample after passing through the amylase depleting device (—red). The peak at 56,216 m/z which corresponds to the presence of amylase, was significantly reduced in the sample’s eluent. Interestingly, the peak at 66,867 m/z was significantly enhanced in the eluent.
Figure 4.7: SDS-PAGE analysis of high-abundance proteins in human saliva and an example gargoyle sample stained with Coomassie-Brilliant-Blue under reducing conditions. Lane 1: Standard protein ladder. Lane 2: HSA (2.5 μg), Lane 3: salivary Amylase (14 μg), Lane 4: IgA (3.5 μg), Lane 5: gargle sample before passing through amylase depleting device, Lane 6: same gargle sample after passing through the device (eluent).
proteins, it can be distinctly observed that the band between 50-60 kDa in the gargle sample (Lane 5) decreased in intensity after passing through the amylase depleting device as seen in Lane 6.

Furthermore, almost all the other bands were still observed in Lane 6. This coincides well with the corresponding MALDI-ToF spectra of the negative control, wherein the peak between 55,500–59,000 m/z significantly decreased in intensity upon passing the gargle samples through the amylase depleting device (Figure 4.6). As previously established, we observed that there is selective binding on the starch bed only w.r.t amylase, whereas all the other proteins showed a negligible loss in signal intensity on MALDI-ToF. Hence, we can conclude that the band getting depleted between 50-60 kDa in the SDS PAGE is corresponding to amylase. In Lane 6, we could still observe the light chains of immunoglobulins along with various other bands indicating that the rest of the proteome is conserved. Certainly, there is some protein loss occurring overall in the gargle samples while using the amylase depleting device as we are including an additional step in our analysis. However, when analyzing the eluents on a sensitive instrument such as MALDI-ToF, we can still observe all the other proteins even after depleting amylase.

In gargles from subjects, we observed a substantial signal drop between 55,500–59,000 m/z in the MALDI-ToF profiles of COVID-19 positive samples after passage through the set-up (Figure 4.8). This suggests that the ions contributing to this peak are of amylase and are reduced in number upon employing the amylase depleting device. Moreover, the peak around 66,000 m/z (along with several other peaks) was remarkably enhanced in these eluents, supporting the hypothesis that amylase depletion unmasked signals from multiple proteins in saliva.

Next, we compared the eluents of COVID-19 positive and COVID-19 negative gargle samples (Figure 4.9). As predicted, COVID-19 positive profiles had numerous and higher intensity
Figure 4.8: MALDI-ToF mass spectra of a gargle sample from a COVID-19 positive donor (—red) and an overlay of the same sample passed through the amylase depleting device (—green).

peaks compared to COVID-19 negative profiles, even after depleting amylase. Specifically, the presumed viral protein peaks at the previously established ranges of 66,400–68,100 and 78,600–80,500 m/z were still significantly higher in COVID-19 positive compared to COVID-19 negative gargles suggesting that amylase depletion did not affect the ability to detect viral proteins in the spectra. SDS-PAGE analysis was also performed to compare the clinically confirmed COVID-19
Figure 4.9: MALDI-ToF mass spectra of gangle samples from a COVID-19 positive donor eluent (—red) and COVID-19 negative donor eluent (—green). Panel A is the full-range mass spectra of the samples. Panels B and C show specific ranges where differences in mass spectra are observed between COVID-19 positive eluent and COVID-19 negative eluent samples after amylase depletion.
positive and negative samples (Figure 4.10). The eluents of both of these samples (Lane 3 and Lane 5) showed a reduction in band intensity at around 60kDa which is close to the theoretical weight of salivary amylase. Moreover, the rest of the proteome was protected and could be visible by CBB staining method.

The sample files were normalized by dividing the signal intensities (from 2,000-200,000 m/z) by the peak intensity of the (M+H)+ peak of apomyoglobin which was used to calibrate the instrument daily. Next, the areas of the peaks in the 11,140–11,160 m/z (quality control peak present in saliva irrespective of COVID-19 status) and 78,600-80,500 m/z (S1 spike protein of SARS-CoV-2) ranges were calculated by leveraging the composite Simpson’s rule for each spectral range as described in Chapter 2. Simpson’s rule is a numerical method for approximating the definite integral of a function which is based on approximating the AUC using quadratic polynomials. The formula for Simpson’s rule involves dividing the interval of integration into an even number of subintervals and evaluating the function at the endpoints and midpoints of these subintervals. The approximate value of the integral is then given by a weighted sum of these function values. ROC curve analysis was performed to determine the appropriate cut-off to call a sample COVID-19 positive vs negative. The disease prevalence for generating the ROC plot was assumed to be 10% based on the CDC’s COVID-19 data tracker. As expected, peaks in the 11,140–11,160 m/z range did not discriminate between COVID-19 negative and positive samples (AUC 0.77, threshold of 4.44).
Figure 4.10: SDS-PAGE analysis of COVID-19 positive and COVID-19 negative samples and the respective eluents. Lane 1: Standard protein ladder, Lane 2: COVID-19 positive sample before passing through the amylase depleting device while Lane 3 represents its eluent. Lane 4 represents the COVID-19 negative sample whereas Lane 5 represents its eluent.

Peaks in the range of 78,600-80,500 m/z that represent the S1 spike protein of SARS-CoV-2 showed good discrimination between COVID-19 positive and negative specimens (AUC 0.893, threshold of 10.14) after amylase depletion. Using a threshold area of 10.14, the sensitivity and
specificity of COVID-19 detection was 85.71% and 100% respectively (Figure 4.11-B). The sensitivity of SARS-CoV-2 detection after amylase depletion was somewhat lower than expected when compared to RT-qPCR. It is quite possible that some of the positive samples we selected had low viral loads. IDPH reports positive cases as “detected” but does not report a Ct value to assess the level of the viral load. Further studies with a larger number of samples and comparison with Ct values of positive samples are necessary to estimate the realistic sensitivity of MALDI-ToF method to detect SARS-COV-2.

Figure 4.11: ROC analysis of the biomarker ranges 11,140-11,160 m/z (Panel A) and 78,600-80,500 m/z (Panel B) after amylase depletion. In the plot, each point represents a MALDI-ToF specificity/sensitivity value at a cut-off AUC for each potential biomarker range, with the highest specificity/sensitivity and discriminating power having values near the top left of each plot.
To assess how the amylase depletion process affected the sensitivity and specificity of our method, we analyzed the AUC values of the two peaks (11,140-11,160 m/z and 78,600-80,500 m/z) for the same set of samples before amylase depletion (Figure 4.12). There certainly was an improvement in the detection of COVID-19 after depleting amylase, as the specificity before amylase depletion was 57.12% whereas the sensitivity was still the same (85.71%). This again directs us towards the hypothesis that the signal intensity for the biomarkers of the disease (in this case 78,600-80,500 m/z) were being suppressed due to the abundance of amylase in the gargle samples.

![Figure 4.12](image)

Figure 4.12: ROC analysis of the biomarker ranges 11,140-11,160 m/z (Panel A) and 78,600-80,500 m/z (Panel B) before amylase depletion. In the plot, each point represents a MALDI-ToF specificity/sensitivity value at a cut-off AUC for each potential biomarker range, with the highest specificity/sensitivity and discriminating power having values near the top left of each plot.
Even though five biomarkers were established in our previous study, we only chose to analyze samples based on the 78,600–80,500 m/z range. The biomarkers in the range of 55,500–59,000 m/z and 27,900–29,400 m/z are the singly and double charged peaks of IgA Heavy chain+Amylase. These peaks seem to be depleted after using the amylase depleting device as the starch bed shows affinity toward the amylase. Moreover, as shown in Figures 4.3 and 4.5, when the amylase and standard IgA were passed individually on the amylase depleting device, we could see the selective binding only towards amylase and minimal difference in the intensities for IgA. Hence, when considering a gargle sample, it is important to note the presence of two proteins in the ranges of 55,500–59,000 m/z and 27,900–29,400 m/z amongst which only one protein (amylase) is getting depleted. Overall, there certainly is some protein loss while using this device. But irrespective of this loss, we could see the effects of amylase depletion on the intensities of other proteins. The S2 fragment of the Spike protein of SARS-CoV-2 (66,400–68,100 m/z) was not considered in this study because it overlaps with the albumin (~66,348 kDa) peak present in the host saliva. Nevertheless, some viral proteins are still retained after amylase depletion and have the potential to be used as biomarkers of disease processes in saliva.

Several eluents of gargle samples from amylase-depleted samples showed the presence of additional peaks apart from our established biomarker ranges that were suppressed in non-depleted samples due to the ions from amylase. Further studies are needed to characterize these peaks and determine their utility in diseased and normal conditions.

Finally, previous potato starch column studies were successful in depleting salivary amylase but were not applied towards clinical applications. For the first time, we report a potential application of this method for the diagnosis of COVID-19. Because amylase depletion unmasks
peaks that were otherwise not visible, the quick and cost-efficient method can be employed to discover novel biomarkers for diseases by MALDI-ToF analysis of saliva. Our proof-of-concept study shows that other high abundance proteins can be potentially depleted in a similar manner to improve the diagnostic utility of saliva. Overall, this type of sample processing can result in improved sensitivities of protein analytes present in low abundance. This work was possible due to collaboration with Zane LaCasse, Venkata Devesh R. Seethi, Dr. Hamed Alhoori, Joshua Bland and Dr. Shrihari S. Kadkol.
CHAPTER 5

DISCRIMINATING POWER OF MALDI-TOF TO DETECT SARS-COV-2 AND ITS VARIANTS

Introduction

Detection and identification of viruses primarily depends on cell culture-based methods, electron microscopy, nucleic acid detection and antigen detection. These techniques have a high accuracy rate but are expensive and time-consuming. MALDI-ToF MS is a rapid, sensitive and high throughput clinical tool with leading applications in the area of pathogen identification, disease diagnosis, nucleic acids analysis and metabolite analysis. MALDI MS platforms are now being introduced for the detection of various viruses such as polioviruses, human herpesviruses, human papillomaviruses, hepatitis B and C viruses and more recently SARS-CoV-2.134,135,136,137

Viruses are constantly evolving resulting in diverse pathogenicity that can be observed in their host. Some viral mutations are also known to circumvent the host’s immune system, and thus the antibodies generated after vaccination or from previous infection may not always protect the host.136 The standard method to detect these mutated variants is by performing whole-genome sequencing. However, this technique is expensive, time-consuming and requires expertise and skill to handle its processing. Hence, in addition to genome sequencing, scientists are now attempting
to develop rapid and inexpensive methods for detection of viruses as well as their variants. The benefit of MALDI-TOF technology is that it exhibits characteristics that would allow the identification of individual variants of viruses and the ability to simultaneously detect other viruses in a single analysis. While the initial cost of purchasing a MALDI-ToF instrument is expensive ($87,000-700,000), this is only a one-time cost as MALDI-ToF methods require common lab chemicals and the cost per test is much lower compared to other routinely used molecular diagnostic tools.\textsuperscript{123}

In chapter 3, we established five biomarker ranges by performing protein profiling of human gargle samples to diagnose COVID-19 with $\geq 90\%$ sensitivity and specificity. We also assigned putative identities to the biomarker ranges amongst which the range of 78,600–80,500 m/z was assigned for the S1 fragment of the SARS-CoV-2 S protein.\textsuperscript{85} As most of the mutations have been reported in the S1 fragment of the SARS-CoV-2 virus,\textsuperscript{138} the current study hypothesizes that there could be visible mass shifts in this range of 78,000–82,000 m/z on MALDI-ToF when comparing the saliva/gargle samples of individuals infected with different SARS-CoV-2 variants. Such a method will be able to detect the disease as well as the viral subtypes/mutations in a single spectrum. Another important factor to consider is that the identification based on viral proteins/peptides could offer promising results due to the lack of a need for PCR amplification and better chemical stability of proteins compared to RNA.\textsuperscript{123}

Previously, Zhao et al. employed the MALDI-ToF-based multiplex PCR technique to identify SARS-CoV-2 variants such as the wild-type, B.1.1.7 (Alpha), B.1.351 (Beta), B.1.429 (Epsilon), B.1.526 (Iota), P.1 (Gamma) and B.1.617 (Delta) from oropharyngeal swabs. Here, the mutations for each of the variants were detected with 100\% accuracy in the receptor binding
domain of the spike protein of SARS-CoV-2. Another study reported by Hernandez et al. used a similar platform to detect the variants at a diagnostic sensitivity of ≥93.67%. However, both of these techniques used a combination of two techniques which increases the cost of analysis and were focusing on the detection of nucleic acids. Interestingly, Mann et al. published the results of mass mapping performed on the MALDI-FT-ICR platform of tryptic-digested recombinant S protein of five major variants (Alpha, Beta, Delta, Gamma) of SARS-CoV-2. The peptide signatures generated from each of these variants were used to identify the presence of mutations.

In the current chapter, we are interested in exploring the discriminating power of MALDI-ToF to detect specifically SARS-CoV-2 and its variants from other viruses by performing protein profiling within a wide m/z range. The abundance of viral proteins/peptides is low as compared to the host proteins and the extraction of these viral proteins is known to be complex as compared to bacterial proteins from a given biological matrix. The other reason for the lower utilization of MALDI-TOF MS for virus detection is that most of the established protocols are for the mass range from 2 to 20 kDa, whereas the viral proteins can go beyond 100 kDa, making these protocols incapable of detecting intact viral proteins. Moreover, viral proteins also undergo numerous post-translational modifications in the host’s system which further adds to the mass of the protein. Hence, we aim to utilize our optimized protein extraction method in a wide range of mass detection (20,000-200,000 m/z) by protein profiling method to establish MALDI-ToF as a clinical tool for the SARS-CoV-2 as well as its variants detection.

Another interesting study reported in this chapter is a 7-day study of a COVID-19 positive donor, whose saliva as well as gargle samples were simultaneously collected for seven consecutive days. The gargle samples were processed with the amylase depleting device and analyzed on the
MALDI-ToF whereas the saliva samples were processed for RT-qPCR analysis. Such a comparison was reported for the first time, and it further strengthens our hypothesis that COVID-19 can be detected even after depleting amylase from the gargle samples.

**Results and Discussion**

Before examining the SARS-CoV-2 variants, we were interested in understanding the protein profiles of other viruses. The most common types of viruses that cause seasonal epidemics every winter in the US are the influenza viruses which are further divided into four types: A, B, C and D. Influenza A viruses are the only influenza viruses known to cause flu pandemics and the current subtypes of influenza A viruses that routinely circulate in people include A(H1N1) and A(H3N2). As the flu symptoms can overlap with the COVID-19 symptoms, we were interested in understanding the protein profile of an H1N1 virus. Protein concentration and precipitation were carried out on the cell lysate and the supernatant of Madin-Darby Canine Kidney cells that were infected with the influenza A H1N1 virus. The H1N1 influenza viruses are RNA viruses of 80 to 120 nm in diameter and have a genome size of approximately 13.5 kb. This is less than half of the size of the SARS-CoV-2 virus (29.9 kb). Figure 5.1 shows an overlay of the H1N1 protein profile and the SARS-CoV-2 (positive control). No overlapping peak proteins were detected between these two viruses. This confirmed that we could detect SARS-CoV-2 from the H1N1 infections using our established method.
Figure 5.1: An overlay of MALDI-ToF mass spectra of cell lysate and supernatant of cells infected with H1N1 (—blue) and heat-inactivated cell lysate and supernatant of cells infected with SARS-CoV-2 (—red). Panel A shows a wide range mass spectra whereas Panel B and C show specific ranges where differences in mass spectra are prominently observable.
A total of seven human coronaviruses have been identified as of May 2023. The SARS-CoV, MERS-CoV and SARS-CoV-2 are known to have high transmissibility and have caused serious illness in the infected individuals. But the other four coronaviruses: HCoV-OC43, HCoV-HKU1, HCoV-229E, and HCoV-NL63 are known to cause a mild range of symptoms and are self-limiting to upper respiratory tract infections.\textsuperscript{143} We studied the protein profiles of the cell lysates and supernatants cell lines that were infected with the common human coronaviruses. As mentioned in Chapter 2, these virus infected samples were obtained from BEI Resources, Virginia, US. The \textit{Macaca mulatta} kidney epithelial cells infected with HCoV-NL63, the human ileocecal colorectal adenocarcinoma cells infected with HCoV-OC43, and the human lung fibroblast cells infected with HCoV-229E were processed and analyzed. Figure 5.2 shows an overlay of the three common coronaviruses on MALDI-ToF along with SARS-CoV-2. Each common coronavirus spectrum shows a unique protein profile and hence would make it easy to distinguish from SARS-CoV-2.

Attempts were also made to study the MERS-CoV protein profile. Irradiated cell lysate and supernatant from Vero E6 cells infected with MERS-CoV was analyzed on MALDI-ToF. MERS-CoV has a genome of 30.1 kb that encodes 10 proteins whereas SARS-CoV-2 encodes 29 proteins.\textsuperscript{144} Figure 5.3 shows the overlay of MERS-CoV and SARS-CoV-2 cell lysate and supernatant protein profiles.
Figure 5.2: An overlay of MALDI-ToF mass spectra of cell lysate and supernatant of cells infected with HCoV-NL63 (—green), HCoV-OC43 (—orange), HCoV-229E (—blue) and SARS-CoV-2 (—red). Panel A shows a wide range mass spectra whereas Panel B shows specific ranges where differences in mass spectra are prominently observable.
An important factor to consider in the analysis of proteins from all the cell lysates and supernatants is that the cell growth medium contains fetal bovine serum. The fetal bovine serum is made up of various proteins such as albumin, globulins, bilirubin, hemoglobin, etc. which will be reflected in the spectrum along with the viral proteins. The MALDI-ToF spectrum of fetal bovine serum is presented in Figure 5.4. However, even though fetal bovine serum proteins are present in all the virus-infected samples, we could still observe distinct differences between the
Figure 5.4: MALDI-ToF spectrum of fetal bovine serum.

spectra. This confirms that viral proteins are contributing to the protein profile. Further studies are required in order to establish specific spectral protein profiles for each of these viruses.

Another crucial aspect was to identify the SARS-CoV-2 variants from the collected gargle/saliva samples. As mentioned earlier, we were interested in the S1 fragment of the Spike protein to observe the mass shift and provide any correlation with the type of variant. To begin with, we analyzed the recombinant full-length S proteins of the original Wuhan variant, Spike D614G Variant, Delta Variant, B.1.1.529 lineage and BA.2 lineage (Omicron variants). As
mentioned in Chapter 2, these recombinant S proteins were obtained from BEI Resources, Virginia, US. These five recombinant proteins were produced in human embryonic kidney (HEK293) cells and purified by affinity chromatography. Figure 5.5 shows an overlay of all five recombinant proteins from the SARS-CoV-2 variants. The mass shift between each of the variants is distinctly observable in the spectrum. Moreover, the m/z observed for the variant’s S proteins correlated with the trend of their theoretical masses as seen in Table 5.1 (Protein masses in kDa: Wuhan>Omicron BA.1> Omicron BA.2> Delta>D614G). The theoretical weights for these were calculated from the sequence given by BEI using Expasy Protein mass calculator. The differences in the theoretical weight and the MALDI-TOF m/z values are due to the post-translation modification and particularly glycosylation, as the S protein is highly glycosylated where the modified glycans shield about 40% of the protein surface of the S trimer. Nevertheless, the degree of glycosylation exhibited notable similarity among the five recombinant proteins, primarily owing to their production within a consistent in vitro system (HEK293).

Next, preliminary studies of COVID-19 positive human saliva samples consisting of three different variants were carried out using the amylase depleting device method. These saliva samples were procured from our collaborators at the Department of Pathology, University of Illinois, Chicago, US. Saliva samples were confirmed COVID-19 positive by RT-qPCR and further analyzed by Next-Generation sequencing to identify the variants (Delta, Omicron BA.1 and Omicron BA.2) by our collaborators at the Department of Pathology, University of Illinois Chicago, US.
Figure 5.5: An overlay of MALDI-ToF mass spectra of the recombinant full-length S protein of Wuhan (—red), Spike D614G Variant (—blue), Delta Variant (—pink) and Omicron variants B.1.1.529 lineage (—green) and BA.2 lineage (—orange).

Table 5.1: Comparison between the theoretical masses in Daltons and the MALDI-ToF m/z of the Wuhan strain of SARS-CoV-2 and four other variants.

<table>
<thead>
<tr>
<th>Full-length Spike protein</th>
<th>MALDI-ToF m/z</th>
<th>Theoretical mass in Daltons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wuhan</td>
<td>184,801</td>
<td>140,699</td>
</tr>
<tr>
<td>Spike D614G Variant</td>
<td>177,457</td>
<td>137,031</td>
</tr>
<tr>
<td>Delta Variant</td>
<td>179,784</td>
<td>139,536</td>
</tr>
<tr>
<td>B.1.1.529 Lineage (Omicron Variant)</td>
<td>182,833</td>
<td>139,878</td>
</tr>
<tr>
<td>BA.2 Lineage (Omicron Variant)</td>
<td>180,747</td>
<td>139,735</td>
</tr>
</tbody>
</table>
As predicted, mass shifts were evident in the range of 78,000-82,000 m/z for the COVID-19 positive saliva samples from three different variants when compared with each other (Figures 5.6 and 5.7). The S1 fragment for the Delta variant was observed at 77,365 m/z whereas the Omicron variants BA.1 and BA.2 were observed close to each between the ranges 79,617-81152 m/z and 80,974-80,588 m/z respectively. The m/z ratio of the Delta variant is lower than that of Omicron variants similar to what was seen with the recombinant S protein. Based on the preliminary analysis, visible mass shifts could be observed on MALDI-ToF for detecting COVID-19 variants.

To extend the variant analysis study, COVID-19 positive saliva samples consisting of four different SARS-CoV-2 variants were analyzed. Once again, the saliva samples in this study were obtained from collaborators at the Department of Pathology, University of Illinois, Chicago, US. This included nine Delta variant samples, five Omicron BA.1 variant samples, seven Omicron BA.2 variant samples and seven XBB/BQ variant samples. Once again, the S1 fragment of the S protein in the range of 78,000-82,000 m/z was noted from the MALDI-ToF spectrum for each sample. Table 5.2 displays the summary of the study that includes theoretical weight as well as the observed S1 fragment peak for each saliva sample.

In chapter 3, we demonstrated the use of amylase depleting device on gargle samples (5 mL) to detect the S1 peak and ultimately determine its COVID-19 status. Here, even with the use of saliva samples (0.5 mL), we were able to observe a similar protein profile trend and were successful in detecting the S1 peak. While comparing the variants, at this point we are unable to establish exact ranges for each variant (Table 5.2). The key reason for the failure to assign a range for a given variant could be the post-translational modifications of the S protein. It is to be noted
that the S protein of SARS-CoV-2 is highly glycosylated. Each protomer in the trimeric spike has 22 glycosylation sites and this extensive glycosylation is responsible for shielding the virus from...

Figure 5.6: An overlay of MALDI-ToF mass spectra of the three COVID-19 positive saliva donors infected with different variants: Delta (—green), Omicron BA.1 (—blue) and Omicron BA.2 (—red).
Figure 5.7: An overlay of MALDI-ToF mass spectra of the four COVID-19 positive saliva donors infected with different variants: Omicron BA.1 (blue) and Omicron BA.2 (red).

the human defense system. Hence, the addition of these glycan moieties will certainly add to the overall mass of the protein. Moreover, the extent or the degree of glycosylation could vary based on multiple factors. Some of the studied factors include the 3-D structure of the protein, the enzyme repository of the host cell, the type of tissue or cell, the transit time in the Golgi and the availability of the intracellular sugar-nucleotide donors. Moreover, the physiological state of the host and the
Table 5.2: Comparative analysis of S1 fragment peak of SARS-CoV-2 variants analyzed MALDI-ToF MS.

<table>
<thead>
<tr>
<th>SARS-CoV-2 Variant</th>
<th>Theoretical mass in Daltons</th>
<th>MALDI-ToF m/z</th>
<th>Average</th>
<th>Standard Deviation</th>
<th>Margin of Error at Confidence level of 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>75239</td>
<td>77377</td>
<td>81228</td>
<td>80856</td>
<td>79895</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80595</td>
<td>77629</td>
<td>80463</td>
<td>80305</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80762</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omicron BA.1</td>
<td>75517</td>
<td>79686</td>
<td>81152</td>
<td>79722</td>
<td>79888</td>
</tr>
<tr>
<td></td>
<td></td>
<td>77939</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omicron BA.2</td>
<td>75423</td>
<td>80804</td>
<td>80588</td>
<td>79112</td>
<td>80458</td>
</tr>
<tr>
<td></td>
<td></td>
<td>79688</td>
<td>79162</td>
<td>79559</td>
<td></td>
</tr>
<tr>
<td>Omicron XBB/BQ</td>
<td>Not sequenced</td>
<td>80725</td>
<td>80023</td>
<td>80879</td>
<td>80490</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80051</td>
<td>77672</td>
<td>79537</td>
<td></td>
</tr>
</tbody>
</table>
immune pressure could also lead to altered glycosylation patterns as a means to evade immune recognition\textsuperscript{148,149,150}.

The effect of mutations in the S protein on glycosylation has been the subject of research and several researchers attempted to use advanced mass spectrometry to compare the glycosylation among the variants. For instance, Wang D., et al. reported that D614G variant had a 45% lower amount of complex-type glycans and a 33% higher amount of oligomannose glycans as compared to the wild type SARS-CoV-2.\textsuperscript{151} Similarly, the alpha variant, B.1.1.7, is reported to have 10 variations in the S protein compared to the original Wuhan strain.\textsuperscript{152} Aloor et al. stated that there has been no evidence of mutations at the reported or predicted glycosylation sites on the spike proteins. However, the conformational changes due to the amino acid switch may alter the glycosylation enzyme accessibility and hence result in modification of the glycosylation pattern.\textsuperscript{153}

Next, we extended our research to compare the MALDI-ToF analysis results with RT-qPCR results in a 7-day study. A symptomatic COVID-19 positive donor was asked to donate a gargle and a saliva (drool) sample every morning for seven consecutive days. The saliva samples were sent to the Department of Pathology at University of Illinois, Chicago, US for RT-qPCR analysis whereas the gargle samples were processed using the amylase depleting device. Figure 5.8 shows an overlay of MALDI-ToF analysis of the seven gargle samples from a COVID-19 positive donor after passing through the amylase depleting device. After careful scrutiny of the MALDI-ToF spectra overlay, it was observed that there was an overall gradual decrease in the intensity of the peak signals from Day 1 to Day 6. However, Day 7 showed a sudden upsurge in the signal intensities especially around the S1 biomarker peak (81,280 m/z) which can be correlated to the viral protein concentration. In fact, Day 7 had the highest signal intensity for S1
viral protein peak. This observation was an unusual phenomenon as the viral load should eventually decrease once the host immune system is triggered. We also observed a heightened peak signal for 56,300 m/z on Day 4 which plausibly corresponds to the heavy chain of IgA.

Figure 5.8: MALDI-ToF mass spectra overlay of gargle samples eluents collected from a COVID-19 positive donor for seven consecutive days. Day 1 in presented in (—blue), Day 2 in (—sky blue), Day 3 in (—lilac), Day 4 in (—green), Day 5 in (—orange), Day 6 in (—yellow) and Day 7 in (—red). The gargle samples were processed with the amylase depleting device and their corresponding eluents are shown in the figure.
Next, we evaluated the RT-qPCR results for all seven saliva samples. Remarkably, the RT-qPCR results followed a similar trend to the MALDI-ToF analysis. There was a gradual decrease in the viral copies/mL from Day 1 to Day 6, followed by a sudden surge on Day 7 which was reported to be higher than on Day 1 (Table 5.3).

Table 5.3: RT-qPCR results of saliva samples collected from a COVID-19 positive donor for 7 consecutive days.

<table>
<thead>
<tr>
<th>Day</th>
<th>Average Cycle threshold (Ct)</th>
<th>Viral RNA Copies/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26.71</td>
<td>72,308</td>
</tr>
<tr>
<td>2</td>
<td>30.26</td>
<td>6,612</td>
</tr>
<tr>
<td>3</td>
<td>31.29</td>
<td>3,258</td>
</tr>
<tr>
<td>4</td>
<td>33.51</td>
<td>719</td>
</tr>
<tr>
<td>5</td>
<td>35.14</td>
<td>302</td>
</tr>
<tr>
<td>6</td>
<td>35.11</td>
<td>241</td>
</tr>
<tr>
<td>7</td>
<td>26.28</td>
<td>97,338</td>
</tr>
</tbody>
</table>

While it is difficult to directly compare the results between MALDI-ToF and RT-qPCR analysis owing to the nature of their respective results, we attempted to understand a trend of the
Figure 5.9: Comparison between MALDI-ToF and RT-qPCR analysis. Panel A shows the AUCs of the viral proteins S1 (78,600–80,500) and S2 (66,400–68,100) fragments whereas panel B shows viral loads in copies/mL versus days for a COVID-19 positive donor.

Viral load between these two testing platforms. Panel A of Figure 5.9 shows the AUCs of the biomarker ranges for the S1 and the S2 fragment of spike protein plotted against days. Panel B of Figure 5.9 shows the viral load plotted against the days. When comparing panels A with B of Figure 5.9, we can witness that Day 1 and Day 7 have higher intensities/viral load whereas other days (Day 2-Day 6) showed lower intensities/viral load.

To further explore the strange trend of a sudden upsurge in viral load on Day 7, we analyzed Next-generation sequencing data of Day 1 and Day 7 saliva samples (Figure 5.10). The saliva samples were sent to the Department of Pathology at University of Illinois, Chicago, US. The S1
fragment of the S gene of the SARS-CoV-2 genome was sequenced, and three single nucleotide variations were reported between the two variants. After evaluating the results, it was noticed that the SARS-CoV-2 variant for Day 1 was Omicron BA.2.12.1 whereas for Day 7 was Omicron BA.4. This indicated that the donor was infected with the second variant while he/she was still recovering from the first infection cycle. This could also suggest the unpreparedness of the immune system to tackle the second variant resulting in an increase in the viral load and a decrease in the Ct value on Day 7. Overall, there are seven mutations reported between the S protein sequence of Omicron variants BA.2.12.1 and BA. 4 out of which six alone are in the S1 fragment of the protein. As stated in chapter 1, the S protein and more specifically S1 fragment includes the receptor binding domain crucial for viral entry, and hence mutations in this region could result in altered pathogenesis. In the case of this donor, such a small variation in the sequence was still capable of infecting the host before the previously produced antibodies could fight the second variant. This has so far been the first case to report the presence of two variants in such a short period of time. Interestingly, the donor was fully vaccinated with a booster for COVID-19 six months before testing COVID-19 positive.

To summarize, due to the small sample size, it is difficult to establish the exact ranges for each of the variants using the protein profiling method on MALDI-ToF. However, if successful, MALDI-ToF will prove to be the most efficient platform to detect the disease and the variant in a single spectrum. RT-PCR is used for the detection of SARS-CoV-2 whereas whole-genome
Figure 5.10: Comparison of DNA sequencing analysis between Day 1 and Day 7 saliva samples showing the differences between the two Omicron variants BA.2.12.1 and BA.4.

sequencing is used for variant detection which overall increases the total cost of testing. On the basis of the preliminary results, this method can be further expanded to a larger set of clinical samples. Among all the coronaviruses, SARS-CoV-2 is the only one which showed asymptomatic infection. This further increases the urgency of having rapid, cost-effective and accurate
detection methods for COVID-19 to curb the transmission. MALDI-ToF also offers the possibility of automation and data analysis using ML algorithms for achieving higher accuracies. However, one major limitation of MALDI-ToF in the clinical world is that pathogen identification is limited to the existing database. Currently, we can only compare the analyzed data with the existing database of well-characterized organisms. Efforts are needed to extend the database libraries to detect the maximum number of viruses and their variants.137
CHAPTER 6

CONCLUSION AND FUTURE WORK

While a majority of the global population is now estimated to be immune to SARS-CoV-2 through vaccination or natural infection, the virus continues to mutate and has caused more than 766 million infections worldwide.⁷ These mutations have affected the rate of transmission as well the pathogenicity of this virus. Studies suggest that the emerging variants also influence the efficacies of the therapeutic drugs and vaccines. Moreover, it is estimated that approximately 10% of the infected people have ‘long COVID’ which is referred to as post-acute sequelae of COVID-19 resulting in multisystemic conditions comprising of severe symptoms that follow a SARS-CoV-2 infection. The wealth of studies that have explored the different types of platforms for detecting SARS-CoV-2 is impressive. The presence of SARS-CoV-2 can be determined by several techniques including but not limited to RT-PCR, amplicon-based metagenomic sequencing, and nucleic acid hybridization to detect the viral RNA and serological techniques such as lateral flow test and ELISA to detect the antibodies produced in response to the virus.¹⁵⁶,¹⁵⁷ More recently, mass spectrometry, particularly MALDI-ToF, has been employed by numerous scientists for the diagnosis of COVID-19.
Protein expression anomalies resulted due to disruption, misfolding or mutations are associated with disease onset and progression. These altered protein amounts can serve as biomarkers of the disease not only for early diagnosis but also for drug discovery. In similar lines, MALDI-ToF MS is now being used to study the protein profiles of various body fluids such as serum, urine and saliva to detect diseases. More importantly, analysis of proteins profiles would include host proteins as well as infectious agent’s proteins providing a dual advantage for detecting biomarkers.123,136

This dissertation’s study established a highly sensitive and specific saliva/gargle test for diagnosing COVID-19 infection using MALDI-ToF MS. The method described here is relatively rapid and inexpensive and is sensitive enough to detect SARS-CoV-2 infection in samples with very low viral loads. It has advantages over other tests such as non-invasive sampling and the ability to observe both viral proteins and host response in a single spectrum. Comparison of COVID-19 status (assessed by RT-qPCR results) with MALDI-ToF analysis over a wide range of 2000–200,000 m/z identified five potential biomarkers of COVID-19 infection. Overall, the agreement of these results with RT-qPCR testing on nasopharyngeal swabs was ≥90% for the studied cohort, which consisted of young and largely asymptomatic student athletes. From a clinical standpoint, the results from this pilot study suggest that MALDI-ToF could be used to develop a relatively rapid and inexpensive COVID-19 assay. It is also crucial to further identify the potential biomarker peaks to understand the relationship of the proteins with the course of the disease. Preliminary studies to identify the peaks were performed by examining human immunoglobulins and heat inactivated SARS-CoV-2 virus by this reported assay and comparing peak masses of these controls to those observed for gargle samples. Further confirmatory analysis
such as sequencing of proteins that occur in a healthy and SARS-CoV-2-infected saliva proteome is required by using advanced mass spectrometric techniques such as LC-MS/MS. The verification of the identity of the human immune biomarkers may serve as a useful tool for monitoring the immune response under various conditions and stressors.\textsuperscript{85}

One of the limitations while performing high-throughput MALDI-ToF protein profiling is that it is presently confined to the detection of highly abundant proteins. We recognized that amylase, the most abundant protein found in saliva, can obscure the detection of low-abundance proteins by MALDI-ToF MS. To resolve this issue, an inexpensive device was employed to deplete salivary amylase from human gargle samples and then to further compare the protein profiles of COVID-19 positive and negative samples. This device consisted of a starch bed that selectively captured salivary amylase whereas the rest of the proteome remained unaffected when passed through the amylase removing device. Furthermore, this method resulted in enhanced MALDI-ToF signals of various other peaks which were suppressed by the presence of amylase. Overall, a sensitivity of 85\% and specificity of 100\% were achieved by ROC curve analysis. Biomarkers that are usually present in low abundance could plausibly be detected by utilizing such a technique. Future studies with increased sample sizes are required for the validation of this method.

This work also investigated the discriminating power of MALDI-ToF MS to identify a influenza virus, common coronaviruses and the variants of SARS-CoV-2. The protein profiles of H1N1, MERS-CoV, HCoV-OC43, HCoV-229E, and HCoV-NL63 viruses were distinctly different from the SARS-CoV-2 profile. The variant analysis study couldn’t establish specific ranges of the S1 fragment peak among the different viruses at this point. The major reason for this
could be heavy glycosylation of the S protein as well as the mutations between the variants that affect the glycosylation patterns. The results of this work were also the first of a kind to compare MALDI-ToF results with RT-qPCR results simultaneously in a 7-day study. Comparing the viral loads with the AUCs of the MALDI-ToF biomarker peaks once again confirmed the sensitivity and accuracy of our established method.

While the world is seeking rapid and inexpensive testing platforms for early and accurate detection of the SARS-CoV-2 virus, MALDI-ToF MS could soon become the new gold standard for diagnosis of COVID-19. MALDI-ToF MS demonstrates potential to diagnose various other viruses as well and become an alternative technology in combating the current and future pandemics. However, MALDI-ToF methods undoubtedly do require further optimization and development to be established in routine clinical testing.
REFERENCES


44. de Hoffmann, E. & Stroobant, V. *Mass Spectrometry: Principles and Applications*.


